THE HONG KONG
POLYTECHNIC UNIVERSITY
香港理工大學
Pao Yue-kong Library
包玉剛圖書館

Copyright Undertaking

This thesis is protected by copyright, with all rights reserved.

By reading and using the thesis, the reader understands and agrees to the following terms:

1. The reader will abide by the rules and legal ordinances governing copyright regarding the use of the thesis.
2. The reader will use the thesis for the purpose of research or private study only and not for distribution or further reproduction or any other purpose.
3. The reader agrees to indemnify and hold the University harmless from and against any loss, damage, cost, liability or expenses arising from copyright infringement or unauthorized usage.

IMPORTANT

If you have reasons to believe that any materials in this thesis are deemed not suitable to be distributed in this form, or a copyright owner having difficulty with the material being included in our database, please contact lbsys@polyu.edu.hk providing details. The Library will look into your claim and consider taking remedial action upon receipt of the written requests.

Pao Yue-kong Library, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong

http://www.lib.polyu.edu.hk

LARGE-SCALE MULTICLASS TRAFFIC ASSIGNMENT: LEVERAGING THE PARETO FRONTIER FOR DISCRETE AND CONTINUOUS PROBLEMS

ZHENGYANG LI

PhD

The Hong Kong Polytechnic University

2025

The Hong Kong Polytechnic University
Department of Civil and Environmental Engineering

Large-Scale Multiclass Traffic Assignment: Leveraging the Pareto Frontier for Discrete and Continuous Problems

Zhengyang Li

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

June, 2025

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 acknowledgment has been made in the text.

(Signed)

Zhengyang Li (Name of student)

Abstract

Modern urban transportation networks offer travelers unprecedentedly diversified travel options and personalized route guidance. For instance, apps like Google Maps and Waze can provide distance, toll charges, speed limits, and real-time traffic congestion information on different routes to travelers. Therefore, in real life, individual travelers may consider different criteria of travel options and make travel decisions in a way that is more customized to their own preferences. Meanwhile, travelers exhibit significant heterogeneity in preferences, such as different values of time (VOT) and risk preferences, influenced by their socio-economic status, trip purposes, or simply personal characteristics. Empirical studies revealed the limitations of homogeneous transport models and underscore the importance of accounting for travelers' heterogeneous preferences in transport models to facilitate a more disaggregate representation of individual choices.

Traffic assignment (TA) is a fundamental tool for planning and managing urban and regional transportation networks. It aims to predict the spatial distribution of traffic flows on the network by analyzing and aggregating the route choices of numerous travelers. Integrating traveler preference heterogeneity into TA can relax the simplistic route choice assumption that all travelers select the shortest route, leading to the development of multiclass traffic assignment (MCTA) models. Compared to the standard TA models, the MCTA models can not only improve the prediction accuracy of network flows but also enable analyzing different travelers' responses to potential traffic planning projects and policies, supporting real-world applications from traditional ones like toll design to emerging challenges like evaluating the impacts of connected and autonomous vehicles. Nevertheless, incorporating preference heterogeneity into TA models also introduces many model and solution challenges, such as increased model complexity and solution inefficiency.

This thesis focuses on the long-standing topic of transportation network equilibrium with heterogeneous users. The main contribution lies in the development of advanced discrete and

continuous MCTA models and algorithms that skillfully leverage both supply-side (Pareto frontier of network paths) and demand-side (preference distribution of users) information, making the developed models and algorithms applicable to large-scale transportation networks. The thesis consists of four key studies:

1. *Modeling user heterogeneity in congested transportation networks: A review of discrete and continuous approaches.* MCTA models can be broadly classified into discrete and continuous, depending on whether heterogeneity is represented via finite classes or continuous distributions. This study reviews 108 studies (1979–2025) on the model developments, solution algorithms, and applications of discrete and continuous MCTA models. This review not only categorizes the literature along discrete and continuous paradigms but also identifies critical research gaps and emerging trends, particularly highlighting recent methodological breakthroughs that achieve single-class computational complexity for certain MCTA problems.

2. *Discrete and continuous multiclass bi-criteria traffic assignment problems: A unified gradient projection algorithm*. This study develops a unified path-based gradient projection algorithm that simultaneously addresses both discrete and continuous multiclass bi-criteria traffic assignment problems. The algorithm incorporates three specialized modules for column generation, equilibration, and convergence testing, each customized for the distinctive characteristics of discrete and continuous formulations. Numerical experiments demonstrate the algorithm's computational efficiency while showing the convergence relationship between discrete and continuous solutions as the number of user classes increases.

3. *Discrete multiclass bi-criteria traffic assignment without class-specific variables: An alternative formulation and a subgradient projection algorithm*. This study introduces an alternative formulation for the discrete MCTA problem that eliminates the need for class-specific variables through innovative exploitation of the order relationship between path preferences and traveler characteristics. This approach recognizes that high-VOT travelers prefer faster, more expensive routes while low-VOT travelers favor slower, cheaper alternatives. The resulting model maintains mathematical equivalence with conventional formulations while achieving dramatic reductions in variable counts, accompanied by a specialized subgradient projection algorithm that efficiently handles the non-differentiable convex objective function.

4. *Continuous multiclass mean-risk traffic assignment problem: both criteria are flow-dependent*. This study extends the continuous MCTA framework to address the challenging

mean-risk traffic assignment problem with flow-dependent criteria. By introducing the novel concept of cumulative path costs and modeling risk-aversion through continuous distributions, we develop finite-dimensional complementarity conditions that avoid the curse of dimensionality. The accompanying computational framework combines path-based column generation with gradient projection methods, demonstrating superior performance in capturing nuanced risk-averse behaviors while maintaining computational tractability for large-scale networks.

Collectively, this thesis contributes to the modeling and computation of transportation network equilibrium with heterogeneous users. It develops advanced mathematical models and solution algorithms for both the discrete and continuous MCTA problems, providing researchers and practitioners with enhanced analytical tools for capturing user heterogeneity in congested transportation systems. Ultimately, these tools can be applied to a wide range of applications to support the planning of innovative urban mobility solutions.

List of Publications

Transportation network modeling

1. **Zhengyang Li**, Zhandong Xu, Anthony Chen*. 2025. Modeling user heterogeneity in congested transportation networks: A review of discrete and continuous approaches and future directions. Submitted. (The content in Chapter 2 is mainly from this paper.)

2. **Zhengyang Li**, Zhandong Xu, Guoyuan Li, Anthony Chen*. 2025. A unified gradient projection algorithm for solving both discrete and continuous multiclass bi-criteria traffic assignment problems. *Computers and Operations Research*, 186, 107320. (The content in Chapter 3 is mainly from this paper.)

3. **Zhengyang Li**, Guoyuan Li, Zhandong Xu, Anthony Chen*. 2023. Multiclass bi-criteria traffic assignment without class-specific variables: An alternative formulation and a sub-gradient projection algorithm. *Transportation Research Part E: Logistics and Transportation Review, 176, 103210.* (The content in Chapter 4 is mainly from this paper.)

4. Zhandong Xu, **Zhengyang Li**, Jun Xie, Anthony Chen*, Xiaobo Liu. 2025. Mean-risk traffic assignment under the continuously distributed risk-aversion factor. Minor revision at *Operations Research*. (The content in Chapter 5 is mainly from this paper.)

5. **Zhengyang Li**, Zhong Zhou, Anthony Chen*. 2025. Path finding under uncertainty: Considering both reliability and unreliability of travel time distributions. Under review at *Transportation Research Part C: Emerging Technologies*.

6. Umer Mansoor, **Zhengyang Li**, Anthony Chen*. 2025. Incorporating safety reliability and unreliability into network equilibrium model: Heterogeneous users with safety-

conscious route choice sets. Under revision at *Transportation Research Part C: Emerging Technologies*.

7. Umer Mansoor, **Zhengyang Li**, Anthony Chen*. 2025. Safety vs. efficiency: Modeling heterogeneous preferences in traffic assignment. Under revision at *Transportation Research Part E: Logistics and Transportation Review*.

8. Seungkyu Ryu, Anthony Chen*, **Zhengyang Li**, Heqing Tan, Xiangdong Xu. 2025. Modeling multi-class traffic assignment problems considering capacity constraints. Under review at *Computers and Operations Research*.

9. Zhandong Xu, Anthony Chen*, Guoyuan Li, **Zhengyang Li**, Xiaobo Liu. 2024. Elastic-demand bi-criteria traffic assignment under the continuously distributed value of time: A two-stage gradient projection algorithm with graphical interpretations. *Transportation Research Part E: Logistics and Transportation Review, 183, 103425.*

10. Guangming Xu, Yangyang Song, Yini Xiao, **Zhengyang Li**, Anthony Chen*. 2023. Capacity-constrained mean-excess equilibrium assignment method for railway networks. *Transportation Research Part C: Emerging Technologies, 156, 104350.*

11. **Zhengyang Li**, Jun Zhao*, Qiyuan Peng. 2021. Train service design in an urban rail transit line incorporating multiple service routes and multiple train compositions. *Transportation Research Part C: Emerging Technologies, 123, 102959*.

Community resilience under disruptions

1. **Zhengyang Li**, Katherine Idziorek, Anthony Chen, Cynthia Chen*. 2025. Untapped capacity of place-based peer-to-peer (P2P) resource sharing for community resilience. *Nature Cities.* 2(1), 47-57.

2. Olivia (Jie) Wang, **Zhengyang Li**, Cynthia Chen*. 2025. The efficacy of decentralized disaster relief resource allocation within communities: The role of community-based sharing captains. Under review at *Transportation Research Part E: Logistics and Transportation Review.*

Acknowledgements

I am deeply grateful to my Ph.D. supervisor, Prof. Anthony Chen, for his guidance over the past five years. Prof. Chen always wants his students to become professional and excellent researchers and consistently puts effort into nurturing us, which is evidenced by the countless meetings we had during my doctoral journey. From him, I learned classical theories and techniques, including network modeling and discrete choice, but more profoundly, his approach to critical thinking and professional scholarship. These lessons will undoubtedly guide my career and personal growth.

My sincere thanks go to Prof. Cynthia Chen at the University of Washington for hosting my visiting study in Seattle. Her support and innovative ideas inspired me to explore a topic involving social networks and peer-to-peer resource sharing, a meaningful departure from my Ph.D. research on transportation network modeling. This exploration not only led to a publication on Nature Cities, but also prompted me to reflect more deeply on the broader impact of research.

I would like to extend special thanks to Dr. Zhandong Xu from Southwest Jiaotong University for his invaluable advice during his postdoctoral period at PolyU. With his solid expertise in traffic assignment algorithms, Dr. Xu patiently addressed many nuanced and technical questions for me. Our discussions and collaborations on related papers significantly enriched the depth and scope of my doctoral research.

My heartfelt thanks go to my friends and colleagues at PolyU and UW: Mingxi Li, Zijian Hu, Shuo Hao, Yiyang Peng, Daniyar Kurmankhojayev, Umer Mansoor, and many others whom I cannot list fully here. Their companionship, positive energy, and the memories we created—both in and out of the academic setting—have been a constant source of inspiration. I feel privileged to have known each of them and am confident that our friendships will last for years to come.

Finally, I wish to express my deepest gratitude to my parents for their unwavering support. Despite not having the opportunity to pursue higher education themselves, they understood its profound value and invested wholeheartedly in my brother's and my education. Whenever pos-

sible, they ensured I had access to better schooling—from a rural school in a village with a mud-field stadium to a well-equipped private school in Chengdu—ultimately paving the way for me to become a first-generation college student. I also thank my girlfriend, Jie Wang, for her patience and encouragement, which sustained me through the challenges of my doctoral research. Without their unconditional support, this journey would not have been possible.

Contents

List of Publications **7**

Acknowledgement **9**

Table of Contents **16**

List of Figures **19**

List of Tables **21**

1 Introduction **22**

1.1 Background 22

1.2 Challenges 24

1.3 Contribution and thesis outline 25

2 Modeling user heterogeneity in congested transportation networks: A review of discrete and continuous approaches **28**

2.1 Introduction 28

2.1.1 Why this review? 30

2.1.2 Scope and chapter organization 31

2.2 Literature identification and selection 31

2.2.1 Literature search 32

2.2.2 Screening and supplementation 33

2.2.3 Annual publications 33

2.3 An example of discrete and continuous multiclass approaches 33

2.3.1 The bi-criteria traffic assignment problem 34

2.3.2 Single-class approach . . . 36
2.3.3 Discrete multiclass approach . . . 37
2.3.4 Continuous multiclass approach . . . 38
2.3.5 A comparison between different approaches . . . 39
2.3.5.1 Model features . . . 39
2.3.5.2 Numerical illustration . . . 39
2.3.5.3 Why prefer the continuous multiclass model? . . . 42
2.4 Continuous multiclass approach . . . 43
2.4.1 Model developments . . . 43
2.4.1.1 Infinite-dimensional formulations . . . 43
2.4.1.2 Finite-dimensional formulations . . . 44
2.4.2 Solution algorithms . . . 47
2.4.3 Applications on different types of user heterogeneity . . . 48
2.4.3.1 Preference heterogeneity . . . 48
2.4.3.2 Choice set heterogeneity . . . 49
2.4.4 A summary table . . . 50
2.5 Discrete multiclass approach . . . 53
2.5.1 Model developments . . . 53
2.5.2 Solution algorithms . . . 54
2.5.3 Applications on different types of user heterogeneity . . . 55
2.5.3.1 Preference heterogeneity . . . 55
2.5.3.2 Information heterogeneity . . . 56
2.5.3.3 Choice set heterogeneity . . . 57
2.5.3.4 Routing principle heterogeneity . . . 57
2.5.3.5 Mixed heterogeneity . . . 58
2.5.4 A summary table . . . 58
2.6 Summary . . . 64

3 Discrete and continuous multiclass bi-criteria traffic assignment problems: A unified gradient projection algorithm 65

3.1 Introduction . . . 66

3.2 Bi-criteria traffic assignment problem with multi-user classes 68
3.2.1 Problem setting 68
3.2.2 Route choices of heterogeneous users 69
3.2.3 Continuous multiclass bi-criteria traffic assignment problem 70
3.2.3.1 Continuous multiclass 70
3.2.3.2 Equilibrium 71
3.2.3.3 Formulation 73
3.2.4 Discrete multiclass bi-criteria traffic assignment problem 76
3.2.4.1 Equilibrium 76
3.2.4.2 Formulation 77
3.3 A unified path-based gradient projection algorithm 78
3.3.1 Solving bi-criteria TAP with continuous multi-user classes 80
3.3.1.1 Column generation 80
3.3.1.2 Decomposition and equilibration: boundary update 81
3.3.1.3 Convergence criterion 83
3.3.2 Solving bi-criteria TAP with discrete multi-user classes 84
3.3.2.1 Column generation 84
3.3.2.2 Decomposition and equilibration: class-specific path flow update 85
3.3.2.3 Convergence criterion 86
3.3.3 Summary 87
3.4 Numerical results 87
3.4.1 Small network 88
3.4.1.1 Solution of the CMBTA model 89
3.4.1.2 Solution of the DMBTA model 89
3.4.1.3 CMBTA versus DMBTA: path flow pattern 91
3.4.2 Networks of different scales 92
3.4.2.1 Results 93
3.4.2.2 Impact of inner loops 94
3.4.2.3 Comparison with existing methods 96
3.5 Conclusions 97

4 Discrete multiclass bi-criteria traffic assignment without class-specific variables: An alternative formulation and a subgradient projection algorithm 99

4.1 Introduction . . . 100

4.2 Discrete multiclass bi-criteria traffic assignment problem . . . 103

4.2.1 Notation list . . . 103

4.2.2 Conventional discrete MBTA model . . . 103

4.2.3 An alternative formulation for the discrete MBTA problem without class-specific variables . . . 107

4.2.3.1 Model formulation . . . 107

4.2.3.2 Solution properties . . . 111

4.2.4 Comparisons between the conventional model and the alternative formulation . . . 115

4.2.4.1 Memory requirement . . . 115

4.2.4.2 Computation time . . . 116

4.3 The path-based subgradient projection algorithm . . . 117

4.3.1 The overall framework . . . 117

4.3.2 Equilibration . . . 119

4.3.2.1 One-dimensional nondifferentiable convex subproblem . . . 119

4.3.2.2 Subdifferential and available second-order information . . . 120

4.3.2.3 Boundary-updating equation . . . 121

4.3.2.4 Implementation . . . 123

4.4 Numerical experiments . . . 123

4.4.1 A small network . . . 124

4.4.1.1 Algorithmic details . . . 125

4.4.1.2 Solution results . . . 127

4.4.2 Large networks . . . 129

4.4.2.1 Number of classes . . . 129

4.4.2.2 Network scale . . . 131

4.4.2.3 Sensitivity analyses . . . 131

4.4.2.4 Implementation discussion . . . 133

4.5 Summary . . . 134

5 Continuous multiclass mean-risk traffic assignment problem: both criteria are flow-dependent 136

5.1 Introduction . . . 137

5.2 Preliminaries . . . 140

5.2.1 Equilibrium characterization for DMRTA . . . 142

5.2.2 Discretization issues . . . 143

5.3 Equilibrium characterization for CMRTA . . . 145

5.3.1 Infinite-dimensional characterization . . . 145

5.3.2 Finite-dimensional characterization . . . 147

5.4 Algorithm . . . 152

5.4.1 Procedure of column generation problem . . . 153

5.4.2 Procedure of restricted master problem . . . 154

5.4.3 Termination criterion . . . 157

5.4.4 Convergence proof . . . 157

5.5 Results . . . 160

5.5.1 Braess network . . . 160

5.5.1.1 Evolution of the Pareto frontier . . . 161

5.5.1.2 Comparison of equilibrium solutions . . . 162

5.5.1.3 Discretization errors on operational decisions . . . 163

5.5.2 Real-world networks . . . 164

5.5.2.1 Exogenous versus endogenous CMRTA. . . . 165

5.5.2.2 CMRTA versus DMRTA. . . . 167

5.6 Summary . . . 170

6 Summary and future research 172

6.1 Summary . . . 172

6.2 Future research . . . 174

Appendix A Discrete and continuous multiclass bi-criteria traffic assignment problems: A unified gradient projection algorithm 176

A.1 MSA and FW algorithms for solving the CMBTA problem . . . 176

A.2 MSA and FW algorithms for solving the DMBTA problem . . . 177

Appendix B Discrete multiclass bi-criteria traffic assignment without class-specific variables: An alternative formulation and a subgradient projection algorithm 179

B.1 Dijkstra-like bi-criteria shortest path algorithm 179

B.2 Gradient projection algorithm for conventional discrete MBTA formulation . . 181

Appendix C Continuous multiclass mean-risk traffic assignment problem: both criteria are flow-dependent 183

C.1 Proof of Proposition 5.3 183

C.2 Formulation equivalence 187

C.2.1 Proof of Proposition 5.4 187

C.2.2 Proof of Proposition 5.5 187

C.3 Column generation algorithm 188

C.3.1 Labeling mean-variance bi-objective algorithm 188

C.3.2 Graphical illustration for identifying supported efficient paths 189

C.4 Path-based algorithm for DMRTA 190

Bibliography 193

List of Figures

1.1 Urban transportation network equilibrium analysis. 23
1.2 Thesis outline 26

2.1 The procedure of literature identification and selection. a. Literature search. b. Screening and supplementation. 32
2.2 Annual publications on modeling user heterogeneity in TA. 34
2.3 The modeling approaches of user heterogeneity in TA. 35
2.4 An example of the different models on a two-link network. 41
2.5 The impact of discretization error on decision-making. 42
2.6 Illustration of the core idea for the finite-dimensional formulations. 45
2.7 Illustration of choice set heterogeneity. (b) shows the choice sets of travelers with different driving ranges. 50
2.8 Illustration of information heterogeneity 56
2.9 Illustration of routing principle heterogeneity 57

3.1 An illustrative network and efficient paths. In (b), the dashed lines passing through a path with travel time c_p^{rs} and toll m_p^{rs} represent the contour of generalized travel cost function $d_p^{rs}(\alpha) = c_p^{rs} + \frac{1}{\alpha} m_p^{rs}$. The slope of the dashed line is determined by α, and its x-intercept corresponds to the generalized cost $d_p^{rs}(\alpha)$ given α. 70
3.2 Characterization of randomly distributed VOT among the population. 71
3.3 Illustration of efficient paths and VOT profile function 72
3.4 Path choices of discrete user classes. 77
3.5 A general framework of the path-based gradient projection algorithm. 79
3.6 An illustration of the boundary updating equation. 83

3.7 A small network reproduced from Slavin et al. (2014). 88
3.8 Solution of the CMBTA model for the small network. 90
3.9 Path flow pattern and path-to-class ratio with the increase of the number of classes. 92
3.10 The convergence curves of the path-based GP algorithms for solving the CMBTA model and DMBTA model. 94
3.11 CPU times of different steps in the path-based GP algorithm. 96
3.12 Convergence curves of the different methods. 97

4.1 An empirical distribution of the value of time (VOT) 101
4.2 Illustration of class-specific variables in the conventional MBTA model. 106
4.3 Illustrations of the TEM profile function g_{rs} and flow-weighted TEM G_{rs}. . . . 108
4.4 Illustrations of efficient paths, path choices, path flows, and boundaries. 109
4.5 Illustration of subdifferential and the interpretation of KKT condition. 112
4.6 The framework of the path-based SGP algorithm. 118
4.7 Illustration of boundary updating equation. 122
4.8 A small network (reproduced from Slavin et al. (2014)) 125
4.9 Iterative detail of the subgradient projection algorithm. 127
4.10 Computation results of the two models for the small network. 128
4.11 The convergence curves of the two models under different numbers of user classes on the Winnipeg network. 130
4.12 The convergence curves of the two models under different networks. 132
4.13 Sensitivity analyses of key inputs/parameters on the Winnipeg network. 133

5.1 Path flow errors by discretization. 144
5.2 Path flow errors under different continuous distributions of RAFs. 144
5.3 Computational burden by discretization. 145
5.4 Route choice based on the Pareto frontier. 148
5.5 Geometric interpretation of cumulative path costs. Note that $\hat{\gamma}_1^w = \hat{\gamma}_2^w$ because $f_2^w = 0$. 150
5.6 Counterexample to monotonicity and pseudo-monotonicity. 152
5.7 The Braess network. Label: {a,b}, where a and b are t_a^0 and c_a, respectively. . . 161
5.8 Pareto frontier evolution for the Braess network. 162

5.9 Equilibrium path flow patterns for different traffic assignment problems. 163
5.10 Effect of discretization on operational planning decisions. 164
5.11 Comparison of convergence performance between exogenous and endogenous CMRTA problems. 166
5.12 Comparison of convergence performance between CMRTA and DMRTA problems. 168
5.13 Link flow difference with varying RAF intervals on Sioux Falls. 170
5.14 The quality-efficiency comparison between DMRTA and CMRTA on Sioux Falls. 170

B.1 The gradient projection algorithm for the conventional MBTA problem. 182

C.1 Justifications of Pareto frontier construction. 190

List of Tables

2.1 A comparison of modeling approaches in BTA 40

2.2 A summary of 32 continuous MCTA studies. 51

2.3 A summary of 76 discrete MCTA studies. 59

3.1 Key features of the path-based GP algorithm for solving the CMBTA and DMBTA problems 87

3.2 Algorithmic parameters. 88

3.3 Iteration details of the path-based GP algorithm for solving the DMBTA model. 91

3.4 Path flow pattern of the DMBTA model. There are in total $3 \times 5 = 15$ class-specific path flow variables. Among them, 6 are positive. 91

3.5 Networks of different scales. The detailed network data is available at https://github.com/bstabler/TransportationNetworks. 92

3.6 The computational results in the networks of different scales. 95

4.1 Notation list 104

4.2 Differences in the model formulation 115

4.3 The required memory for storing the paths using the conventional MBTA model on the Chicago Regional network 116

4.4 The comparison of operation complexity 117

4.5 Algorithmic parameters 124

4.6 Variable comparison of the two models 128

4.7 Large networks. 129

4.8 Comparison of the working path set size and memory consumption of the two models on the Winnipeg network. 130

5.1 Tested real-world networks. 165
5.2 Iterative performance metrics on the Chicago Sketch network. Note: EP, efficient path; AP, added path; UEP, used efficient path; OC, order change. 167
5.3 Path number statistics across four networks at UE. 167
5.4 Comparison of activating variables and required memory on Chicago Sketch. . 169

Chapter 1

Introduction

1.1 Background

Urban transportation networks serve millions of travelers every day and greatly shape how people organize their lives, including residential and work locations, daily activities, vehicle ownership, and route choices. Expanding a transportation network (e.g., building new roads) or managing existing transport infrastructure (e.g., setting road tolls) may affect many people and require substantial investment. Therefore, effective planning and management of transportation networks necessitate transport models that can realistically predict user responses and traffic flow distributions.

Traffic assignment (TA) aims to predict the spatial distribution of traffic flows on urban transportation networks. It is an essential tool for transport planning and management, with predicted flows used to compare different scenarios and conduct cost-benefit analyses. Compared to other networks, such as the electricity network and the internet, transportation networks exhibit two distinctive characteristics: travelers are active decision-makers who optimize their routes according to their own interests, and the network experiences congestion effects that influence travel conditions. As illustrated in Fig. 1.1, TA must therefore explicitly address two interconnected complexities: (1) modeling the route choice behaviors of numerous individual travelers whose collective decisions aggregate into network flows; and (2) accounting for traffic congestion, which emerges on heavily loaded roads and in turn affects travelers' route choices. This creates a complex interplay between numerous travelers and the performance of congested networks.

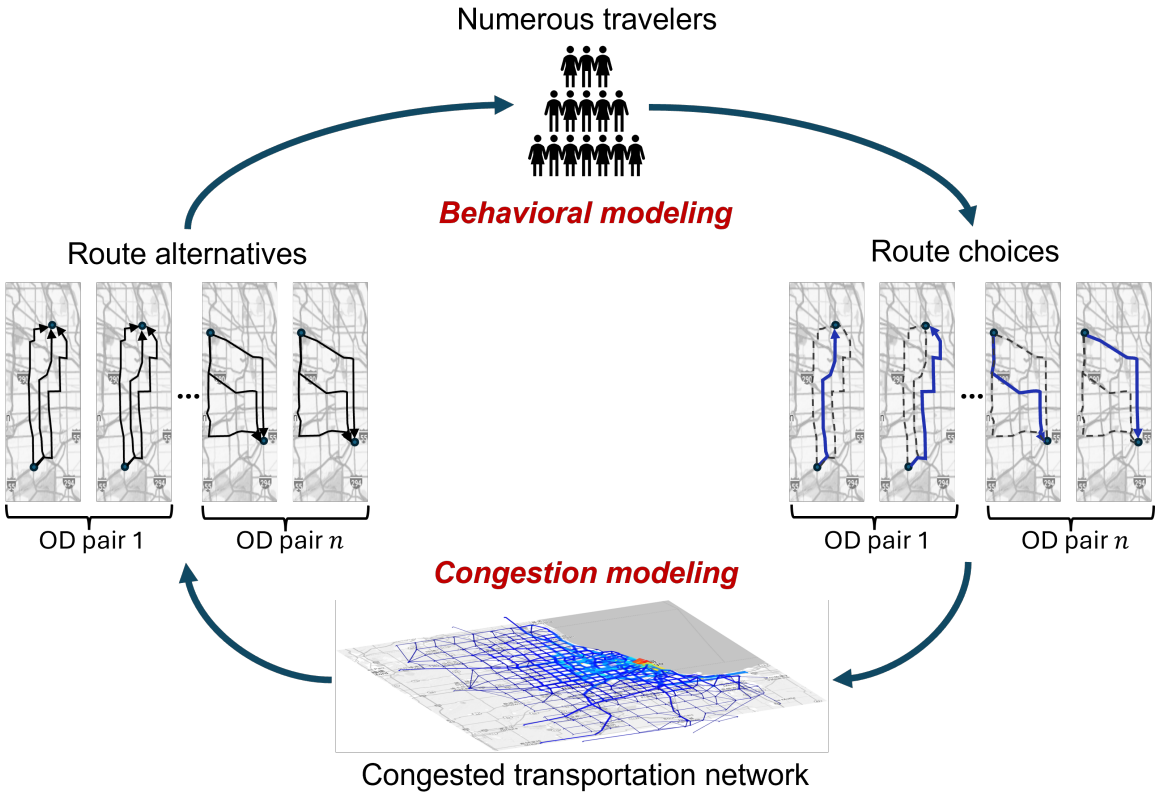


Figure 1.1: Urban transportation network equilibrium analysis.

Network equilibrium models capture this complex interplay by analyzing the stable state where no travelers have incentives to alter their route choices in a congested network. The seminal work by Wardrop (1952) defines:

> *"The journey times on all the routes actually used are equal, and less than those which would be experienced by a single vehicle on any unused route".*

This equilibrium model describes how travelers choose routes in congested urban networks and is known as "user equilibrium" (UE). It lays the foundation for the development of tractable mathematical models (Wardrop, 1952; Beckmann et al., 1956) and effective solution algorithms (e.g., LeBlanc et al., 1975; Jayakrishnan et al., 1994; Patriksson, 2015) and finally becomes a widely applied TA model. Notably, the realism and effectiveness of network equilibrium models largely depend on the behavioral assumptions underlying route choices. It is widely recognized by researchers and practitioners that the UE model has many oversimplified assumptions, including assuming that all travelers consider only travel time and behave homogeneously by choosing the shortest path.

In real life, when making route choices, travelers may not only consider travel time, but also other attributes such as toll, reliability, congestion (e.g., Abdel-Aty et al., 1995; Small et al., 2005; Hossan et al., 2016). Especially in modern urban transportation networks, travelers are offered unprecedentedly diversified travel options and personalized route guidance. For instance, apps like Google Maps and Waze can provide distance, toll charges, speed limits, and real-time traffic information on different routes to travelers. Therefore, individual travelers have a better understanding of the transportation networks and are making travel decisions in a way that is more customized to their preferences.

Moreover, travelers exhibit significant heterogeneity in preferences, such as different values of time (VOT) and risk preferences (Small et al., 2005; Brownstone and Small, 2005; Carrion and Levinson, 2012). This heterogeneity stems from various factors such as travelers' socioeconomic status, trip purpose, or personal characteristics, finally leading to distinct route choice behaviors. Empirical studies reveal the limitations of homogeneous transport models and demonstrate the benefits of accounting for travelers' heterogeneous preferences in transport models to facilitate a more disaggregate representation of individual choices. For example, modeling travelers' VOT as a random variable instead of a constant value can significantly improve the prediction of travel choices (Ben-Akiva et al., 1993).

1.2 Challenges

Compared to the standard UE model, considering travelers' heterogeneous route preferences in network equilibrium models can not only (i) enhance traffic flow prediction accuracy but also (ii) enable analysis of the differential impacts of transport planning and management strategies on diverse user groups. On the other hand, it also mainly brings two challenges, including:

- *Equilibrium modeling.* Modeling travelers' heterogeneous preferences in network equilibrium models is indeed complicated, as it not only involves route choices of heterogeneous users but also the congestion effect on shared network links. For example, modeling a continuously distributed VOT can lead to infinite-dimensional mathematical programs, which are inherently more complex than their finite-dimensional counterparts.

- *Equilibrium computation.* Accounting for travelers' heterogeneous preferences signifi-

cantly complicates equilibrium computation. Solving the resulting network equilibrium problems requires not only leveraging the network structure but also effectively managing user heterogeneity information, making it challenging to develop efficient solution algorithms.

1.3 Contribution and thesis outline

The thesis contributes to network equilibrium modeling and computation by modeling traveler preferences and their heterogeneity in urban transportation networks. It develops advanced discrete and continuous MCTA models and algorithms that skillfully leverage both supply-side (Pareto front of network paths) and demand-side (preference distribution of users) information, making the developed models and algorithms applicable to large-scale transportation networks. Ultimately, these resulting analytical tools can be applied to a wide range of applications to support the planning of innovative urban mobility solutions.

The outline of this thesis is depicted in Fig. 1.2. The research begins with a systematic review, followed by detailed investigations into specific problems, each with distinct motivations and contributions.

Chapter 2 addresses a critical gap in the literature by presenting a systematic review of user heterogeneity modeling in congested transportation networks, with a specialized focus on discrete and continuous multiclass approaches. While this topic has been studied since 1979, research has remained fragmented because user heterogeneity is complex and often intertwined with other issues like congestion pricing and network uncertainty. To synthesize this diverse body of work, this chapter examines 108 studies, analyzing the model developments, solution algorithms, and applications of both multiclass paradigms. It also highlights recent methodological advancements that reduce certain MCTA problems to the complexity of single-class problems.

Chapter 3 compares discrete and continuous bi-criteria (time and toll) MCTA models. Based on whether the VOT is modeled as a discrete or continuous variable, the MCTA models can be classified as discrete or continuous. While both models have been suggested in the literature, their interconnections and differences remain underexplored. This chapter provides a unified path-based gradient projection algorithm for solving both the discrete and continuous MCTA

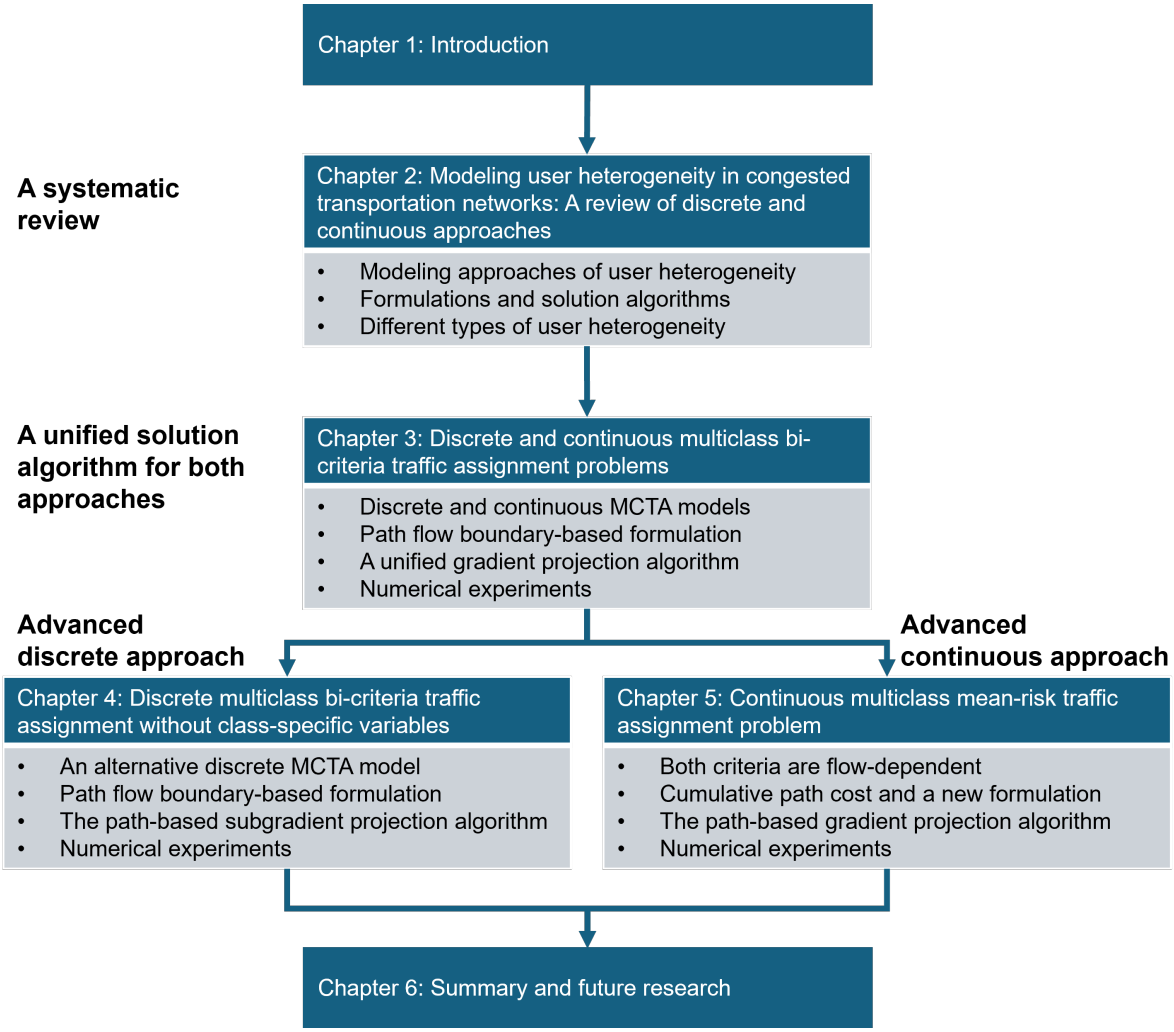


Figure 1.2: Thesis outline

models. In the algorithm, three modules, including column generation, equilibration, and convergence criteria, are customized for discrete and continuous models, respectively. These numerical computations demonstrate that discrete MCTA equilibrium flows converge to continuous MCTA flows as the number of classes increases, and provide insights into the performance and scalability of these two algorithmic approaches.

Chapter 4 introduces an alternative formulation for MCTA without class-specific variables. The conventional discrete MCTA model keeps multiple copies of class-specific variables to model user heterogeneity, which puts a great burden on memory storage and computational speed when dealing with large-scale transportation networks with a large number of user classes. This study provides an alternative formulation for the discrete MCTA problem without class-specific variables by exploiting the order information of paths and travelers, i.e., high-VOT travelers

will prefer fast but expensive paths, while low-VOT travelers will prefer slow but cheap paths. A path-based subgradient projection algorithm is developed to solve the model with a nondifferentiable convex objective function. The numerical results demonstrate that this alternative formulation maintains the same link flow patterns as the conventional model while significantly reducing the number of variables, which can greatly relieve the burden on computer memory and computational speed in solving real transportation networks.

Chapter 5 considers the mean-risk traffic assignment problem with continuously distributed risk-aversion factors. Empirical studies indicate that travelers consider not only travel time but also travel time variability. Further, travelers' preferences towards travel time and variability can be modeled by the risk-aversion factor, which is a random variable that follows a continuous distribution. Under such a case, how to appropriately define the cost of a path for different travelers is the key to network equilibrium modeling and computation. This chapter defines virtual path cost, called cumulative path cost, which provides geometrical interpretations on the Pareto frontier and facilitates characterizing Wardrop's equilibrium condition. General formulations are proposed to model both exogenous and endogenous sources of variability, with the latter dependent on traffic flow. A tailored path-based column generation with the gradient projection algorithm is further developed to solve large-scale problems.

Chapter 6 summarizes this thesis and outlines promising directions for future research in both methodological advancements and emerging applications.

Chapter 2

Modeling user heterogeneity in congested transportation networks: A review of discrete and continuous approaches

Humans are heterogeneous, and so are users of transportation networks. Accurately modeling such user heterogeneity in congested transportation networks gives rise to multiclass traffic assignment (MCTA) models, which improve demand forecasting and enable differentiated impact analysis of policies across user groups. MCTA models can be broadly classified into discrete and continuous, depending on whether heterogeneity is represented via finite classes or continuous distributions. Each paradigm offers unique strengths, applicability, and limitations. Despite substantial progress in both streams, a comprehensive synthesis remains lacking. This chapter reviews 108 studies (1979–2025) on the model developments, solution algorithms, and applications of discrete and continuous MCTA models. We highlight recent methodological advancements that reduce certain MCTA problems to the complexity of single-class traffic assignment.

2.1 Introduction

Traffic assignment (TA) models are fundamental tools in strategic transport planning, enabling authorities to forecast traffic flows and travel times in congested networks to support project appraisal and policy evaluation. The core mechanism of TA is to distribute the predetermined origin-destination (OD) travel demand across a congested network (supply) under assumed route

choice behavior, which can be described as:

$$\texttt{TA}(\underbrace{G, \mathbf{t}(\cdot)}_{\text{Supply}}, \underbrace{\mathbf{q}, \psi}_{\text{Demand}}) = \underbrace{(\mathbf{x}, \mathbf{f})}_{\text{Link/path flow}} , \tag{2.1}$$

where G is the network, $\mathbf{t}(\cdot)$ denotes link performance functions, $\mathbf{q}$ is the OD demand vector, ψ represents route choice behavior, and $\mathbf{x}$ and $\mathbf{f}$ are resulting link and path flow vectors, respectively.

The seminal user equilibrium (UE) principle (Wardrop, 1952) states:

> *"The journey times on all the routes actually used are equal, and less than those which would be experienced by a single vehicle on any unused route."*

This implicitly assumes *homogeneous travelers* with identical preferences (consider travel time only), information/perception (no perception error), and routing principles (minimize travel time). However, the assumption of homogeneous travelers presents shortcomings in both behavioral realism and practical applications:

- *Behavioral realism*. Many empirical studies demonstrated that travelers in the real world are not homogeneous when making route choices (e.g., Lam and Small, 2001; Small et al., 2005; Hossan et al., 2016). Instead, travelers exhibit heterogeneity across multiple dimensions, such as trip purpose, socioeconomic status, and personal experience, that shape their route choices. Therefore, homogeneous models fail to capture these differences, leading to biased flow predictions.

- *Practical applications*. Modern transport policies require disaggregate insights. For example, in build-operate-transfer (BOT) schemes for tolled highways, private operators must estimate the share of users with different values of time (VOT) to assess investment viability, capacity, and toll levels, while governments rely on VOT distributions to evaluate project benefits, costs, and equity impacts on affected populations (Yang and Meng, 2000; Chen and Subprasom, 2007; Chen et al., 2021). Homogeneous TA models cannot provide these essential insights, leading to suboptimal infrastructure decisions.

To address these gaps, multiclass traffic assignment (MCTA) models explicitly represent user heterogeneity on the demand side by introducing different user classes with distinct route choice

behaviors:

$$(\mathbf{q}, \psi) \rightarrow (\mathbf{q}_k, \psi_k)_{k \in K}\,, \tag{2.2}$$

where K is the set of different user classes. From the methodological point of view, two methodological paradigms have emerged:

- *Discrete approach*: divides travelers into a finite number of classes with class-specific demands and route choice behaviors, i.e., K is a discrete set.

- *Continuous approach*: treats user heterogeneity as a continuous random variable across the population, i.e., K is a continuous set.

Both approaches have been developed substantially, and plenty of different MCTA models have been proposed in the literature (e.g., Dafermos, 1981; Leurent, 1993; Dial, 1996; Nagurney and Dong, 2002a; Yang and Huang, 2004; Xie et al., 2021; Xu et al., 2023b; Li et al., 2023b, 2025).

2.1.1 Why this review?

The motivations for this review are twofold.

First, since Dial (1979) firstly modeled users with continuously distributed VOT in a bi-criteria mode-route choice setting, both discrete MCTA models and continuous MCTA models have been explored in the literature. However, despite nearly five decades of development, these two lines of research are seemingly parallel and lack a concerted effort to summarize and analyze the model developments and solution algorithms. Therefore, a synthesis can not only consolidate existing knowledge from MCTA studies but also catalyze methodological innovations.

Second, both the discrete and continuous MCTA approaches have been widely used in different applications, from traditional ones such as toll design (e.g., Dial, 1999a,b) to emerging ones such as modeling the impacts of range anxiety (e.g., Xie et al., 2019; Xu et al., 2024b) and autonomous vehicle (AV) (e.g., Zhang and Nie, 2018). Especially, emerging mobility technologies, such as AVs, will amplify user heterogeneity in VOT, risk perception, safety valuation, and mode acceptance. Hence, a systematic review enables researchers and practitioners to select appropriate modeling approaches to address these evolving dynamics.

In this chapter, we present a systematic review of 108 studies (1979–2025) on MCTA with heterogeneous users, with a focus on discrete and continuous MCTA approaches. The contributions of this work are derived from: (1) a comprehensive archive and synthesis of up-to-date literature on MCTA, (2) a detailed example illustrating the model and numerical similarities and differences between discrete and continuous approaches, (3) an in-depth review of model formulations and solution algorithms for both paradigms, (4) an analysis of different applications of both approaches, and (5) identification of both potential methodological advancements and emerging applications for future research.

2.1.2 Scope and chapter organization

Our scope is limited to user heterogeneity in static TA, excluding dynamic TA models (e.g., Lu et al., 2008; Zhang et al., 2013). We also distinguish this work from MCTA studies focusing on physical heterogeneity (e.g., cars and trucks), which was pioneered by Dafermos (1972) and mainly addresses asymmetric vehicle interactions, as different vehicles (e.g., cars and trucks) contribute unequally to road congestion (i.e., $\mathbf{t}(\cdot)$ in Eq. (2.1)). Interested readers can refer to Chapter 7 in Sheffi (1985) for details and to Boyce et al. (2004) for a review. These studies are omitted in this chapter as they focus on more realistic congestion effects, not on relaxing homogeneity assumptions related to route choice behavior. Furthermore, our analysis is restricted to papers that investigate user heterogeneity as a primary or secondary modeling aspect alongside other relevant modeling issues, focusing specifically on user heterogeneity in route choices.

The remainder of this chapter is organized as follows. Section 2.2 outlines our process for retrieving relevant MCTA studies. Section 2.3 provides a concrete example to elucidate the similarities and distinctions between discrete and continuous multiclass approaches. A detailed review of studies using the continuous and discrete approaches is presented in Sections 2.4 and 2.5, respectively. Section 2.6 offers concluding remarks.

2.2 Literature identification and selection

This section presents our framework for Literature identification and selection, which contains two major steps: (a) literature search and (b) screening and supplementation, as shown in Fig. 2.1.

2.2.1 Literature search

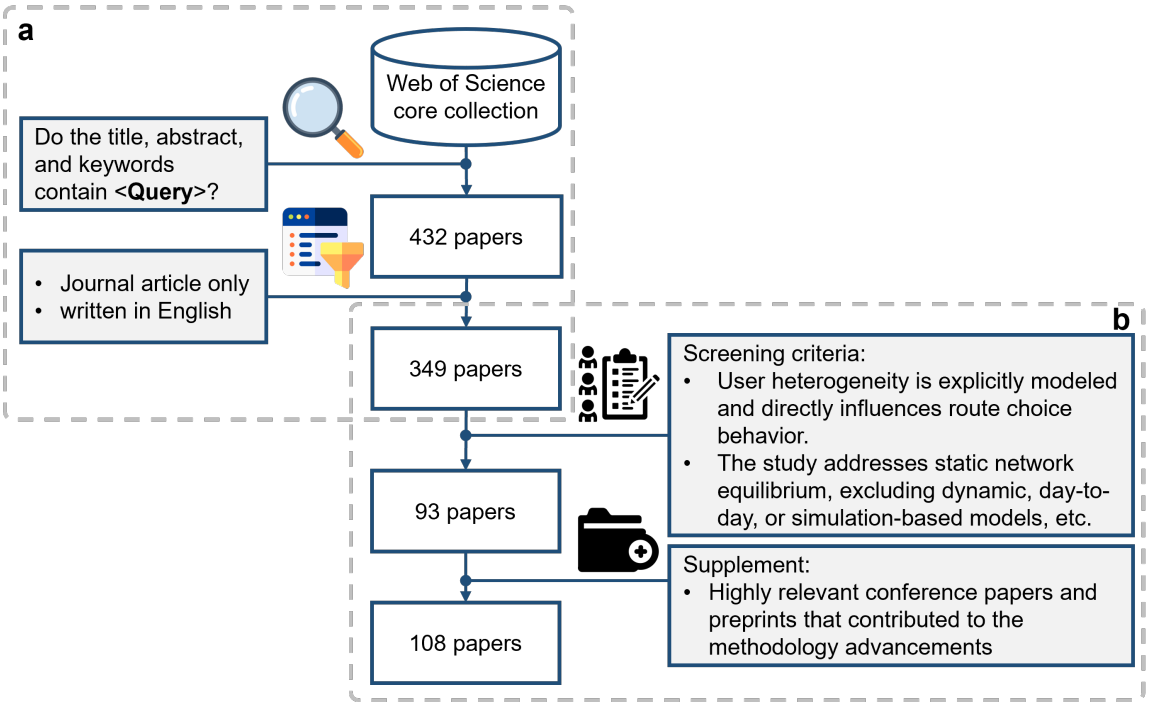


Figure 2.1: The procedure of literature identification and selection. a. Literature search. b. Screening and supplementation.

The literature search process is illustrated in Fig. 2.1a. We conducted our search using the Web of Science (WoS) Core Collection[1]. Due to the diverse terminologies used in studies on modeling user heterogeneity in TA, we iteratively refined and pilot-tested our search query to balance specificity and comprehensiveness, resulting in the following Boolean query:

Query: ("Multiclass" OR "User class" OR "User heterogeneity" OR "Heterogeneous" OR "Value of time") AND ("Traffic assignment" OR "Network equilibrium" OR "Traffic equilibrium" OR "User equilibrium" OR "Route choice" OR "System optimum" OR "System optimal") NOT ("Bottleneck" OR "Transit")

This query was executed on January 17, 2025, targeting the title, abstract, and keyword fields, resulting in 432 papers. After excluding books and conference papers and retaining only journal articles written in English, the search yielded 349 papers for further analysis.

[1] https://www.webofscience.com/wos/woscc/basic-search

2.2.2 Screening and supplementation

The initial search yielded a broad set of publications, many of which were tangential to our review objective. These included empirical studies, discrete choice models, dynamic traffic assignment (Han et al., 2011), day-to-day evolution models (Wei et al., 2016), bottleneck models (Wang et al., 2022a), and simulation-based approaches.

To ensure focus and relevance, we applied two strict inclusion criteria: (1) User heterogeneity is explicitly modeled and directly influences route choice behavior, and (2) the study addresses the static TA problem, excluding dynamic, day-to-day, or simulation-based models, etc. This screening reduced the pool to 93 papers.

To capture seminal methodological contributions not indexed in WoS, we supplemented the corpus with highly relevant conference papers and preprints identified through author expertise and citation tracing. This added 15 studies, resulting in a final set of 108 publications (1979–2025) for in-depth review.

2.2.3 Annual publications

Figure 2.2 shows the annual publication distribution of the collected papers. Following Dial (1979)'s introduction of a bicriteria route choice problem with heterogeneous VOT and Dafermos (1981)'s formalization of the MCTA model, the field received limited attention until the 1990s (1979–1999: emerging stage). During this period, research was dominated by continuous approaches. From 2000 to 2008, publication volume grew steadily, with discrete methods becoming predominant. Since 2009, the field has entered a rapid growth stage, characterized by a surge in applied studies and active exploration of both discrete and continuous approaches. In addition, the majority of the MCTA papers adopted the discrete approach, while a few adopted the continuous approach.

2.3 An example of discrete and continuous multiclass approaches

Prior to the literature review, this section utilizes the widely studied bi-criteria TA (BTA) problem to illustrate the fundamental distinctions between approaches to modeling user heterogeneity, which can be categorized as follows (see Fig. 2.3):

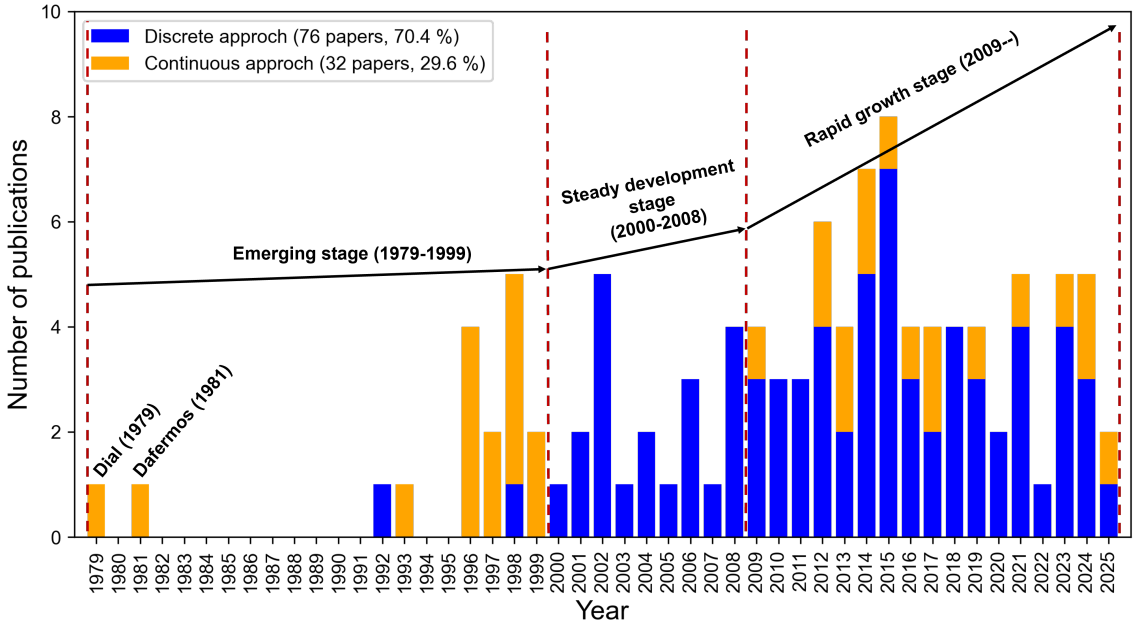


Figure 2.2: Annual publications on modeling user heterogeneity in TA.

1. *Single-class approach*: All travelers are assumed to be homogeneous, possessing identical characteristics (e.g., VOT) and adhering to the same route choice behavior.

2. *Discrete multiclass approach*: Travelers are segmented into a finite number of distinct classes, each defined by specific characteristics (e.g., high vs low VOT). Within each class, travelers exhibit identical route choice behavior.

3. *Continuous multiclass approach*: User heterogeneity is captured via a continuous distribution (e.g., a spectrum of VOT), accounting for nuanced variations across the population. Consequently, each traveler, with their unique characteristics, has their own route choice behavior.

2.3.1 The bi-criteria traffic assignment problem

Consider a transportation network as a directed graph $G = (N, A)$, where N is the set of nodes and A is the set of directed links. Let q^{rs} denote the travel demand for OD pair rs, $rs \in RS \subseteq N \times N$. P_{rs} denote the set of simple paths connecting OD pair rs. The set of all paths is $P = \cup_{rs \in RS} P_{rs}$. Let f_p^{rs} denote the flow on path p connecting OD pair rs. The link flow x_a is the summation of path flows traversing this link, i.e., $x_a = \sum_{rs \in RS} \sum_{p \in P_{rs}} f_p^{rs} \delta_{ap}^{rs}, \forall a \in A$,

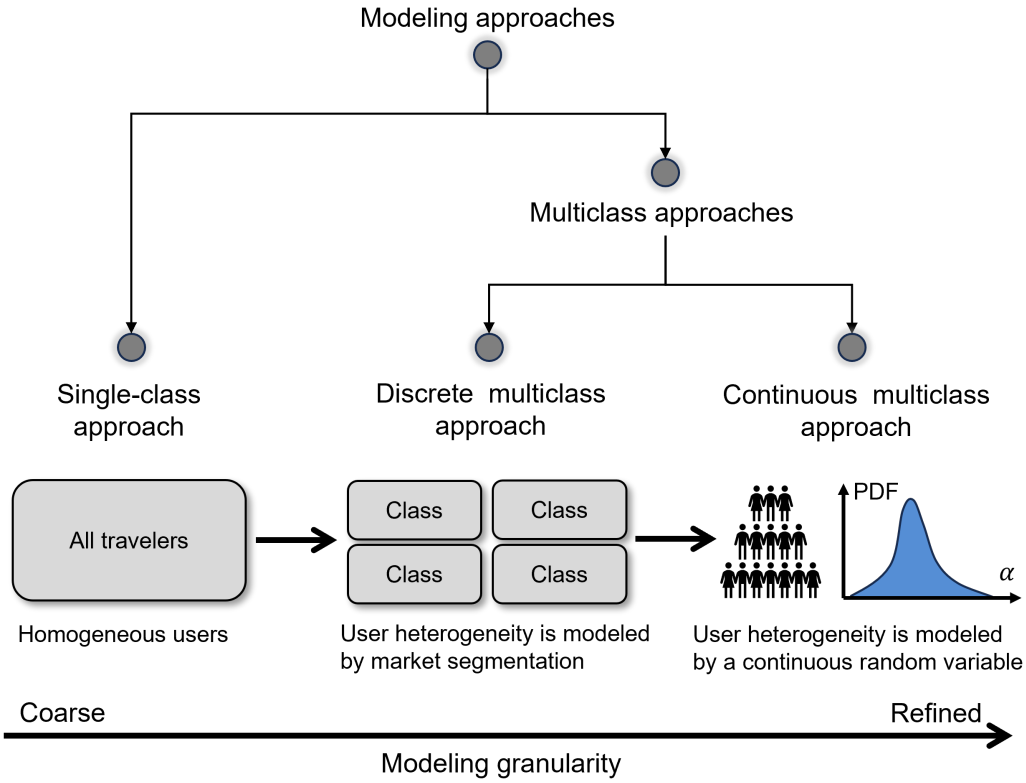


Figure 2.3: The modeling approaches of user heterogeneity in TA.

where $\delta_{ap}^{rs} = 1$ if link a is on path p connecting OD pair rs, 0 otherwise.

For the BTA problem, each link has two attributes: flow-dependent travel time $t_a(\cdot)$, which is continuous and strictly increasing, and exogenously determined toll $\tau_a \geq 0$. The travel time on each link is assumed to depend only on the flow on this link. The path travel time $c_p^{rs}(\mathbf{f})$ and path toll m_p^{rs} are assumed to be additive.

$$c_p^{rs}(\mathbf{f}) = \sum_{a \in A} t_a(\mathbf{f}) \delta_{ap}^{rs}, \quad \forall rs \in RS, \forall p \in P_{rs} \tag{2.3}$$

$$m_p^{rs} = \sum_{a \in A} \tau_a \delta_{ap}^{rs}, \quad \forall rs \in RS, \forall p \in P_{rs} \tag{2.4}$$

Travelers are assumed to choose the path(s) that can minimize their *generalized cost*, defined as a VOT-weighted combination of travel time and toll:

$$d_p^{rs}(\mathbf{f}, \alpha) = c_p^{rs}(\mathbf{f}) + \frac{1}{\alpha} m_p^{rs}, \quad \forall rs \in RS, p \in P_{rs} \tag{2.5}$$

In Eq. (2.5), VOT α denotes the travelers' preferences regarding travel time and monetary

cost. A higher α implies that travelers prioritize travel time savings over monetary cost (i.e., they are less sensitive to tolls).

2.3.2 Single-class approach

The single-class approach assumes that all the travelers have the same VOT α, and the single-class BTA-UE (SCBTA-UE) condition can be defined as follows.

Definition 2.1 (SCBTA-UE). *A feasible path flow vector* $\mathbf{f}^* = (f_p^{rs*})^\top_{\forall rs \in RS, p \in P_{rs}}$ *corresponds to a SCBTA-UE flow if for all OD pairs* $rs \in RS$*, if* $f_p^{rs*} > 0$*, then*

$$c_p^{rs}(\mathbf{f}^*) + \frac{1}{\alpha} m_p^{rs} \leq c_{p'}^{rs}(\mathbf{f}^*) + \frac{1}{\alpha} m_{p'}^{rs}, \quad \forall p' \in P_{rs} \tag{2.6}$$

This equilibrium flow corresponds to an optimal solution to the following convex program:

$$\min Z_1(\mathbf{f}) = \sum_{a \in A} \int_0^{x_a} t_a(v) dv + \sum_{a \in A} \tau_a \frac{1}{\alpha} x_a \tag{2.7a}$$

$$\sum_{p \in P_{rs}} f_p^{rs} = q^{rs}, \quad \forall rs \in RS \tag{2.7b}$$

$$x_a = \sum_{rs \in RS} \sum_{p \in P_{rs}} f_p^{rs} \delta_{ap}^{rs}, \quad \forall a \in A \tag{2.7c}$$

$$f_p^{rs} \geq 0, \quad \forall rs \in RS, p \in P_{rs} \tag{2.7d}$$

In the model, the objective function (2.7a) contains two terms. The first term is the Beckmann transformation (Beckmann et al., 1956) modeling the congestion effect, while the second term models the toll effect. Eq. (2.7b) enforces demand conservation, Eq. (2.7c) defines link flows, and Eq. (2.7d) ensures non-negativity.

This formulation inherits the properties of the standard UE model (Sheffi, 1985). Therefore, the equivalence between the optimal solution of Model (2.7) and Definition 2.1 can be easily proved by checking the Karush–Kuhn–Tucker (KKT) condition. Like the standard UE model, the objective function is strictly convex in $\mathbf{x}$ and the feasible set is convex and compact. Thus, the equilibrium link flow $\mathbf{x}^*$ is unique, while the path flow $\mathbf{f}^*$ is generally not unique in networks with multiple paths per OD pair.

2.3.3 Discrete multiclass approach

The discrete multiclass approach segments travelers into a finite set of classes K, where individuals within the same class $k \in K$ share the same VOT α_k. Let q_k^{rs} denote the demand for class k between OD pair $rs \in RS$, with total demand satisfying $q^{rs} = \sum_{k \in K} q_k^{rs}$. Travelers in class k choose paths to minimize their class-specific generalized cost.

By defining *class-specific path flow* f_p^{rsk} and *class-specific link flow* x_a^k, the discrete multiclass BTA-UE (DMBTA-UE) condition can be established as follows.

Definition 2.2 (DMBTA-UE). *A feasible class-specific path flow vector* $\tilde{\mathbf{f}}^* = (f_p^{rsk*})^\top_{\forall rs \in RS, p \in P_{rs}, k \in K}$ *corresponds to a DMBTA-UE flow if for all OD pairs* $rs \in RS$ *and class* $k \in K$*, if* $f_p^{rsk*} > 0$*, then*

$$c_p^{rs}(\tilde{\mathbf{f}}^*) + \frac{1}{\alpha_k} m_p^{rs} \leq c_{p'}^{rs}(\tilde{\mathbf{f}}^*) + \frac{1}{\alpha_k} m_{p'}^{rs}, \quad \forall p' \in P_{rs} \tag{2.8}$$

The DMBTA-UE flow corresponds to an optimal solution to the following convex program (Yang and Huang, 2004):

$$\min Z_2(\tilde{\mathbf{f}}) = \sum_{a \in A} \int_0^{x_a} t(v) dv + \sum_{a \in A} \tau_a \sum_{k \in K} \frac{1}{\alpha_k} x_a^k \tag{2.9a}$$

$$\sum_{p \in P_{rs}} f_p^{rsk} = q_k^{rs}, \quad \forall rs \in RS, k \in K \tag{2.9b}$$

$$x_a^k = \sum_{rs \in RS} \sum_{p \in P_{rs}} f_p^{rsk} \delta_{ap}^{rs}, \quad \forall a \in A, k \in K \tag{2.9c}$$

$$x_a = \sum_{k \in K} x_a^k, \quad \forall a \in A \tag{2.9d}$$

$$f_p^{rsk} \geq 0, \quad \forall rs \in RS, p \in P_{rs}, k \in K \tag{2.9e}$$

The objective function in Eq. (2.9a) consists of two terms: the Beckmann transformation and the monetary cost of discrete classes. Eq. (2.9b) is the demand conservation constraint for each class. Eq. (2.9c) converts class-specific path flow into class-specific link flows. Eq. (2.9d) merges class-specific link flow into total link flows. Eq. (2.9e) is the nonnegativity constraint of class-specific path flows.

The equivalence between the optimal solution of Model (2.9) and Definition 2.2 can be proved by checking the KKT condition (Yang and Huang, 2004). For the solution property of

the DMBTA model, it is known that only the total link flow $\mathbf{x}^*$ is unique, while the class-specific link flow, path flow, and class-specific path flow are not in general networks (Nagurney, 2000; Nagurney and Dong, 2002a).

2.3.4 Continuous multiclass approach

The continuous multiclass approach treats user heterogeneity as a continuous random variable. Let VOT be a continuous random variable on support $[\underline{\alpha}, \overline{\alpha}]$ with a probability density function (PDF) $h(\cdot)$ and cumulative density function (CDF) $H(\cdot)$. Then, the *demand density* for travelers with VOT α is $q^{rs}h(\alpha)$.

If we follow a similar logic to the discrete multiclass approach, we can introduce *path flows density* $f_p^{rs}(\alpha)$ and *link flows density* $x_a(\alpha)$ with respect to VOT $\alpha \in [\underline{\alpha}, \overline{\alpha}]$. Then the continuous multiclass BTA-UE (CMBTA-UE) condition can be established as follows.

Definition 2.3 (CMBTA-UE). *A feasible path flow density vector* $\tilde{\mathbf{f}}^* = \left(f_p^{rs*}(\alpha)\right)^\top_{\forall rs \in RS, p \in P_{rs}, \alpha \in [\underline{\alpha}, \overline{\alpha}]}$ *corresponds to a CMBTA-UE flow if for all OD pairs* $rs \in RS$ *and VOT* $\alpha \in [\underline{\alpha}, \overline{\alpha}]$*, if* $f_p^{rs*}(\alpha) > 0$*, then*

$$c_p^{rs}(\tilde{\mathbf{f}}^*) + \frac{1}{\alpha} m_p^{rs} \leq c_{p'}^{rs}(\tilde{\mathbf{f}}^*) + \frac{1}{\alpha} m_{p'}^{rs}, \quad \forall p' \in P_{rs} \tag{2.10}$$

The CMBTA-UE flow corresponds to an optimal solution to the following *infinite-dimensional* convex program (Marcotte and Zhu, 1996):

$$\min Z_3(\tilde{\mathbf{f}}) = \sum_{a \in A} \int_0^{x_a} t(v) dv + \sum_{a \in A} \tau_a \int_{\underline{\alpha}}^{\overline{\alpha}} \frac{1}{\alpha} x_a(\alpha) d\alpha \tag{2.11a}$$

$$\sum_{p \in P_{rs}} f_p^{rs}(\alpha) = q^{rs} h(\alpha), \quad \forall rs \in RS, \alpha \in [\underline{\alpha}, \overline{\alpha}] \tag{2.11b}$$

$$x_a(\alpha) = \sum_{rs \in RS} \sum_{p \in P_{rs}} f_p^{rs}(\alpha) \delta_{ap}^{rs}, \quad \forall a \in A, \alpha \in [\underline{\alpha}, \overline{\alpha}] \tag{2.11c}$$

$$x_a = \int_{\underline{\alpha}}^{\overline{\alpha}} x_a(\alpha) d\alpha, \quad \forall a \in A \tag{2.11d}$$

$$f_p^{rs}(\alpha) \geq 0, \quad \forall rs \in RS, p \in P_{rs}, \alpha \in [\underline{\alpha}, \overline{\alpha}] \tag{2.11e}$$

The objective function in Eq. (2.11a) consists of two terms: the Beckmann transformation and the monetary cost of continuous classes. Eq. (2.11b) is the demand conservation constraint

for continuous classes. Eq. (2.11c) converts path flow density into link flow density. Eq. (2.11d) merges link flow density into total link flows. Eq. (2.11e) is the nonnegativity constraint of path flow density. This model has an infinite number of link/path flow density variables and constraints.

The equivalence between the optimal solution of Model (2.11) and Definition 2.3 was proved in Marcotte and Zhu (1996). The solution properties of the CMBTA are similar to those of the DMBTA: only the equilibrium link flow $\mathbf{x}^*$ is unique, while the link flow density, path flow, and path flow density are not in general networks. Different from the discrete model, the equilibrium path flow $\mathbf{f}^*$ can be unique if no two paths associated with a given OD pair have the same monetary cost (Leurent, 1993; Marcotte and Zhu, 1996).

2.3.5 A comparison between different approaches

2.3.5.1 Model features

Table 2.1 compares the single-class, discrete multiclass, and continuous multiclass BTA models across three dimensions: (1) conceptual foundations and granularity, (2) mathematical structure and variables, and (3) equilibrium and uniqueness properties.

The single-class model assumes homogeneous travelers with a fixed VOT α. The discrete multiclass model captures heterogeneity via a finite set of $|K|$ classes, each with distinct demand q_k^{rs} and VOT α_k. The continuous multiclass model achieves the highest refinement by treating VOT as continuously distributed over an interval $[\underline{\alpha}, \overline{\alpha}]$ with density $h(\alpha)$, resulting in an infinite-dimensional model. While all models yield unique equilibrium link flows $\mathbf{x}^*$, path flow uniqueness holds only conditionally in the continuous case.

2.3.5.2 Numerical illustration

We employ a simple two-link network (Fig. 2.4(a)) to demonstrate the behavior of the single-class, discrete multiclass, and continuous multiclass BTA models and to reveal the limitations of the first two approaches. The network consists of two parallel links connecting a single OD pair with total demand $q^{rs} = 1$. Link 1 has a constant travel time of 1, while link 2 has a flow-dependent travel time equal to its flow x_2. A toll of 2 is imposed on link 2. Travelers' VOTs are assumed to follow a uniform distribution over $[\underline{\alpha}, \overline{\alpha}] = [10, 30]$.

Table 2.1: A comparison of modeling approaches in BTA

Approach	Single class	Discrete multiclass	Continuous multiclass
Conceptual foundations and granularity			
Core idea	Homogeneous users	Heterogeneous users with finite classes	Heterogeneous users with continuously distributed VOT
Route cost	Generalized cost	Generalized cost	Generalized cost
Heterogeneity	None	Class-based	Traveler-based
# of user classes	1	$\|K\|$	∞
Granularity	Coarse	Medium	Refined
Mathematical structure and variables			
Path/link variables	f_p^{rs}, x_a	f_p^{rsk}, x_a^k	$f_p^{rs}(\alpha), x_a(\alpha)$
Demand (rs)	q^{rs}	$(q_1^{rs}, q_2^{rs}, \ldots, q_{\|K\|}^{rs})$	$(q^{rs}h(\alpha))_{\alpha\in[\underline{\alpha},\overline{\alpha}]}$
VOT	α	$(\alpha_1, \alpha_2, \ldots, \alpha_{\|K\|})$	$\alpha \in [\underline{\alpha}, \overline{\alpha}]$
Problem dimension	$O(\|P\|)$	$O(\|P\| \cdot \|K\|)$	$O(\|P\| \cdot \infty)$
Objective	$\sum_a \int_0^{x_a} t_a(v)\,dv + \frac{1}{\alpha}\sum_a \tau_a x_a$	$\sum_a \int_0^{x_a} t_a(v)\,dv + \sum_a \tau_a \sum_k \frac{x_a^k}{\alpha_k}$	$\sum_a \int_0^{x_a} t(v)dv + \sum_a \tau_a \int_{\underline{\alpha}}^{\overline{\alpha}} \frac{1}{\alpha} x_a(\alpha)d\alpha$
Equilibrium and uniqueness properties			
Equilibrium	UE	Class-specific UE	Density-specific UE
Class-specific link flow $\tilde{\mathbf{x}}^*$	Not unique	Not unique	Not unique
Link flow $\mathbf{x}^*$	Unique	Unique	Unique
Class-specific path flow $\tilde{\mathbf{f}}^*$	Not unique	Not unique	Not unique
Path flow $\mathbf{f}^*$	Not unique in general	Not unique in general	Not unique in general. Unique if no two paths associated with a given OD pair have the same monetary cost.

For the single-class model, we adopt the expected VOT, $\alpha = 20$ (Fig. 2.4(b)). The discrete multiclass model partitions the VOT support into $|K|$ equal intervals, assigning to each class k the interval mean α_k and demand proportion $q_k^{rs} = 1/|K|$ (see Fig. 2.4(c) for an illustration of two classes). The continuous multiclass model directly uses the uniform VOT density $h(\alpha) = 1/20$ for $\alpha \in [10, 30]$ (Fig. 2.4(d)). All models can be readily solved; see details in the following sections.

For reference, the continuous model (Fig. 2.4(f)) yields equilibrium flows of 0.153 and 0.847 on Links 1 and 2, respectively, with a critical VOT of 13.06 marking the route choice switch.

Flow comparison. Using the continuous multiclass solution as the benchmark, Fig. 2.4(e) reports the equilibrium flow on the tolled link 2 (x_2^*) for varying numbers of classes The single-class model overestimates x_2^* due to its inability to capture VOT heterogeneity. The discrete

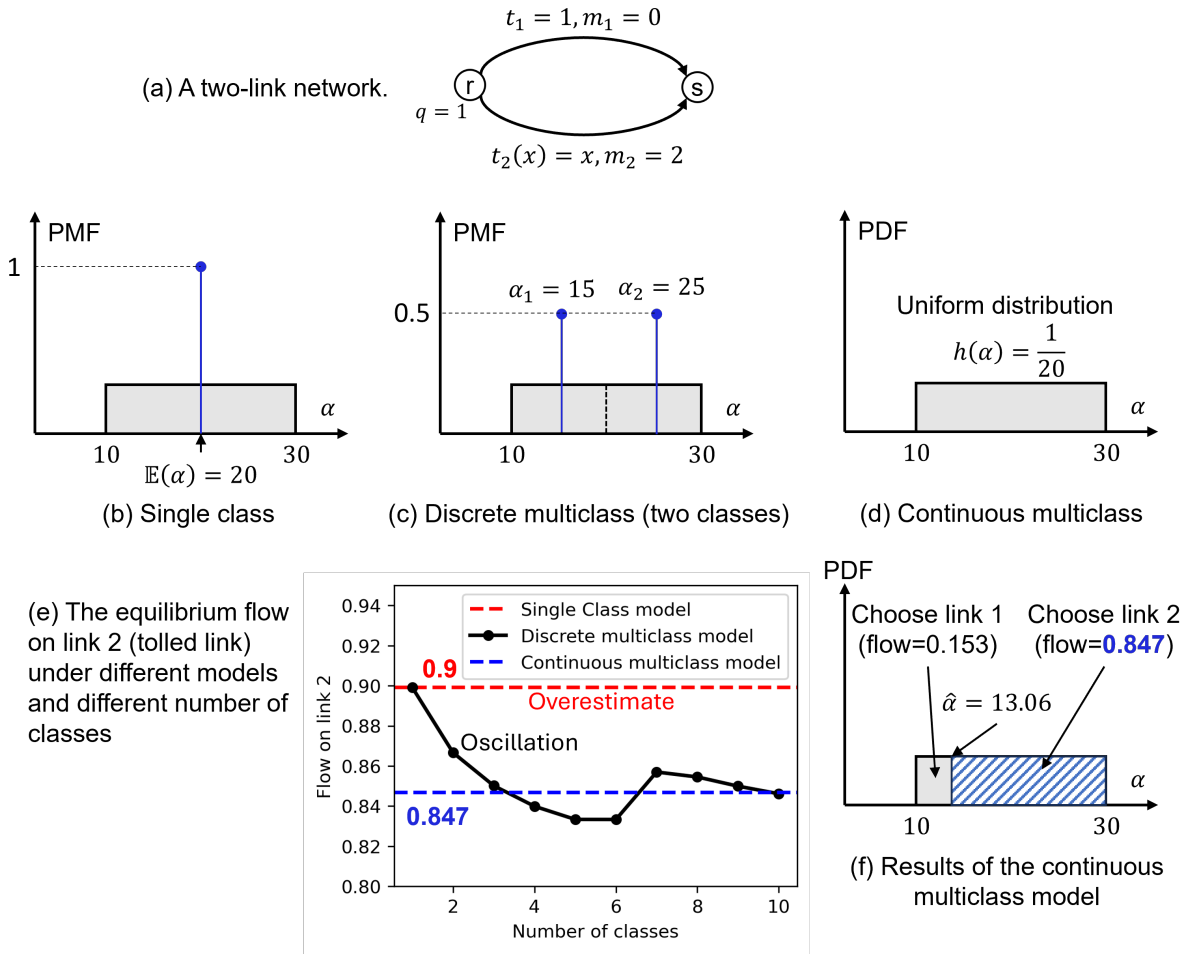


Figure 2.4: An example of the different models on a two-link network.

multiclass results oscillate around the continuous benchmark as $|K|$ increases, reflecting discretization error. This oscillatory behavior implies that the predicted flow—and thus system performance—is sensitive to the arbitrary choice of $|K|$, making it challenging to select an appropriate class structure in practice. However, as $|K| \to \infty$, the link flows of the discrete model converge to that of the continuous model (Li et al., 2025).

Decision-making. Furthermore, DMBTA's discretization error has critical implications on decision-making. Consider a toll increase on link 2 from 2 to 2.2 (see Figs. 2.5(a) and (b)). Both the single-class and continuous multiclass models correctly predict a reduction in total system travel time (TSTT) due to improved congestion relief. However, when $|K| = 6$, the discrete model predicts no change in TSTT, falsely suggesting that the toll adjustment is ineffective for reducing TSTT. This erroneous prediction could lead planners to abandon a beneficial policy, underscoring the risk of relying on coarsely discretized models in decision-making. One may

argue that specifying more user classes will relieve this problem. Then, a subsequent question is how many classes are needed for different MCTA problems? For example, while Rezaee and Hickman (2015) found that 9 classes stabilized link flows in a small tolled network, Slavin et al. (2014) showed that 20 classes are still not satisfactory in toll revenue prediction. Therefore, analysts have to do sensitivity analyses for an appropriate number of classes before decision-making.

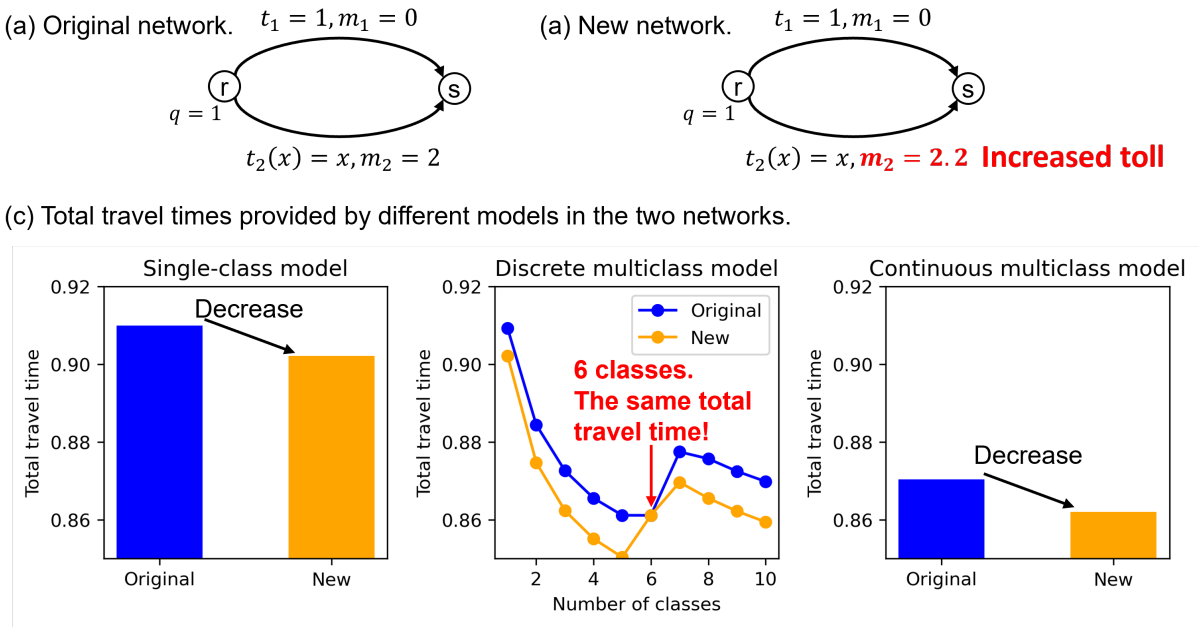


Figure 2.5: The impact of discretization error on decision-making.

In summary, the single-class model fails to account for user heterogeneity, while the discrete multiclass model suffers from discretization errors and is sensitive to the number of classes. In contrast, the continuous model offers a more accurate and reliable solution.

2.3.5.3 Why prefer the continuous multiclass model?

Although the continuous multiclass model (2.11) may appear complex at first glance, several reasons make the continuous multiclass approach preferable to the discrete multiclass approach.

First, many heterogeneous user attributes such as VOT are inherently numerical and continuous, making a continuous representation more natural and behaviorally accurate than arbitrary class segmentation (e.g., "low-VOT" vs. "high-VOT").

Second, discrete models impose artificial boundaries that distort preferences. For instance, with VOT uniformly distributed over $[10, 30]$, dividing travelers into "low-VOT" ($[10, 20]$) and

"high-VOT" ($[20, 30]$) classes groups individuals with VOTs of 10 and 19 together while separating 19 from 21, despite the latter pair being closer in preference. This misrepresentation affects route choice and equilibrium predictions.

Third, discretization introduces errors that are highly sensitive to the number of classes, leading to oscillatory and unreliable results that may mislead planning decisions, as we have demonstrated in Fig. 2.5.

Finally, as we will discuss in Section 2.4, for problems like BTA, the continuous model can be reformulated into a finite-dimensional, computationally efficient program that can even outperforms discrete multiclass approaches when the number of classes is large.

2.4 Continuous multiclass approach

This section reviews the development of models, solution algorithms, and applications of the continuous multiclass approach, concluding with a summary table.

2.4.1 Model developments

The continuous approach was first explored by Dial (1979) for the bicriteria route-mode choice problem in an uncongested network, treating travelers' VOT as a continuous random variable. Since then, both infinite-dimensional formulations and finite-dimensional formulations for MCTA problems have been developed in the literature.

2.4.1.1 Infinite-dimensional formulations

A common feature of infinite-dimensional formulations is the introduction of link and path flow density variables, as illustrated in Section 2.3.4. The primary distinctions between these formulations arise from their specific problem settings, such as whether only travel time is flow-dependent, leading to different mathematical frameworks, such as mathematical programs (MP) or variational inequalities (VI).

Along this line, Dafermos (1981) considered both link attributes to be flow-dependent and provided the first infinite-dimensional VI formulation for the continuous MCTA problem using

link flow density variables:

$$\sum_{a\in A}\int_{\underline{\alpha}}^{\overline{\alpha}}\left(t_a(x_a^*)+\frac{1}{\alpha}m_a(x_a^*)\right)\left[x_a(\alpha)-x_a^*(\alpha)\right]d\alpha\geq 0,\quad \forall\tilde{\mathbf{x}}\in\tilde{X} \tag{2.12}$$

where

$$\tilde{X}=\left\{\tilde{\mathbf{x}}\in\mathbb{R}^{|A|\times[\underline{\alpha},\overline{\alpha}]}\;\middle|\;\begin{array}{c}\sum_{p\in P_{rs}}f_p^{rs}(\alpha)=q^{rs}h(\alpha),\quad \forall rs\in RS,\alpha\in[\underline{\alpha},\overline{\alpha}]\\ f_p^{rs}(\alpha)\geq 0,\quad \forall rs\in RS,p\in P_{rs},\alpha\in[\underline{\alpha},\overline{\alpha}]\\ x_a(\alpha)=\sum_{rs\in RS}\sum_{p\in P_{rs}}f_p^{rs}(\alpha)\delta_{ap},\quad \forall a\in A,\alpha\in[\underline{\alpha},\overline{\alpha}]\end{array}\right\} \tag{2.13}$$

Note that Dafermos (1981) proved the uniqueness of total link flows if the link cost functions are monotone. Subsequent work expanded on this foundation. Under the assumption of exogenous link costs, Marcotte and Zhu (1996) presented an infinite-dimensional MP formulation (2.11), while Marcotte (1999) outlined several other formulations, including infinite-dimensional VI and MP. Dial (1996) and Dial (1997) also adopted link-based VI formulation for algorithm development, which was later applied to congestion pricing problems (Dial, 1999a,b). Meanwhile, Cantarella and Binetti (1998) developed an infinite-dimensional fixed-point (FP) formulation for the bi-criteria SUE problem with continuously distributed VOT, noting that the resulting route choice model is essentially a mixed-logit model. Cantarella and Binetti (1998)'s formulation was also adopted in Meng et al. (2012) and Liu et al. (2013) for toll design. In later years, the application of infinite-dimensional formulations shifted towards policy analysis. They were used to evaluate the existence of optimal toll patterns (Marcotte and Zhu, 2009; Zhu et al., 2012) and were later extended to tradable credit schemes (Zhu et al., 2015). Since then, however, few studies have further explored this specific methodological avenue.

2.4.1.2 Finite-dimensional formulations

For continuous MCTA problem, a few studies focused on developing finite-dimensional formulations, under the key assumption that only link travel time is flow-dependent while tolls are exogenous. These formulations achieve computational tractability by reducing the problem complexity to that of a single-class model!

The core idea is to order the paths in advance and then leverage the geometric information

of the Pareto frontier in the time-cost space to partition the continuum of users into distinct segments. The boundaries between these segments—determined by the slopes between adjacent efficient paths—represent the critical VOT thresholds where optimal path choices change. This concept was pioneered by Dial (1979) for bi-criteria route choice. The following example illustrates this idea.

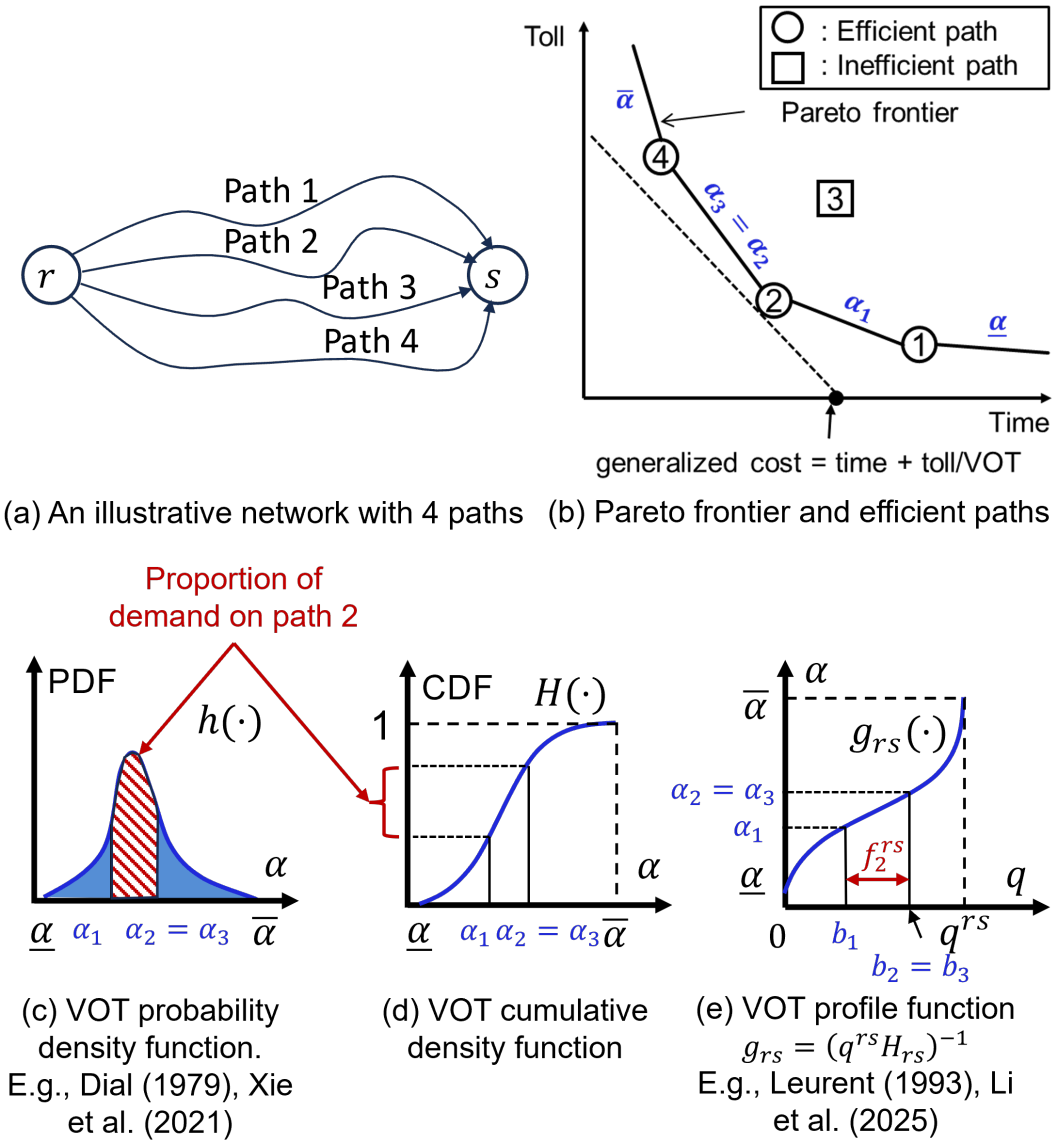


Figure 2.6: Illustration of the core idea for the finite-dimensional formulations.

Example. Consider a network with four paths, three of which are efficient, as shown in Fig. 2.6(a). Their travel time and toll attributes are plotted in Fig. 2.6(b), with paths ordered by increasing toll. Travelers' VOT follows a continuous distribution with PDF $h(\cdot)$ (Fig. 2.6(c)) and CDF $H(\cdot)$ (Fig. 2.6(d)). Fig. 2.6(e) shows the inverse of the demand-weighted CDF function

$g_{rs}(\cdot)$, which is called the VOT profile function (Li et al., 2025). The VOT profile function specifies the VOT for each traveler in an increasing order.

Travelers aim to minimize their generalized costs. Therefore, the optimal path choices may be different for travelers with different VOTs. On the Pareto frontier, the slope between two adjacent efficient paths marks VOT thresholds or boundaries where path choices change. As shown in Fig. 2.6(b), for VOT in $[\underline{\alpha}, \alpha_1)$, path 1 is optimal; for VOT in $[\alpha_1, \alpha_2)$, path 2 is optimal; and for VOT in $[\alpha_3, \overline{\alpha}]$, path 4 is optimal; no one will choose the inefficient path 3 as it never minimizes the generalized cost. Therefore, we can simply set $\alpha_3 = \alpha_2$ to ensure $[\alpha_2, \alpha_3)$ be an empty set.

Consequently, the flow on each path can be calculated in three equivalent ways:

- Using the PDF: $f_p^{rs} = q^{rs} \int_{\alpha_{p-1}}^{\alpha_p} h(v)dv$, see Fig. 2.6(c).
- Using the CDF: $f_p^{rs} = q^{rs} \left(H(\alpha_p) - H(\alpha_{p-1})\right)$, see Fig. 2.6(d).
- Using the VOT profile function: $f_p^{rs} = g_{rs}^{-1}(\alpha_p) - g_{rs}^{-1}(\alpha_{p-1}) = b_p - b_{p-1}$, see Fig. 2.6(e).

The geometric information of the Pareto frontier transforms a complex continuum of travelers into a manageable set of discrete groups, forming the foundation for finite-dimensional formulations in continuous MCTA research. All such formulations adopt one of the aforementioned path flow representations.

Building on this core idea, Leurent (1993) proposed the first finite-dimensional MP for the MCTA problem, conceptually aligned with the VOT profile function (Fig. 2.6)(e)). This work was later extended by Leurent (1998a) for sensitivity analysis using a VI approach. Marcotte (1999) outlined various continuous MCTA formulations, including finite-dimensional MP and FP. Wu and Huang (2014a) and Wu and Huang (2014b) utilized the finite-dimensional MP formulation to investigate the system optimum (SO) and anonymous tolls. Amirgholy and Gonzales (2017) modeled continuously distributed mean-variance preferences and presented a nonlinear complementarity problem (NCP).

More recently, Xie et al. (2021) presented a finite-dimensional MP formulation using PDF and boundary VOTs (Fig. 2.6(c)), which has been extended to cases with nonadditive path costs (Xu et al., 2023b) and elastic demand (Xu et al., 2024a). Li et al. (2025) utilized the flow boundaries (i.e.,b in Fig. 2.6(e)) to build the formulation and compared the properties of discrete and continuous multiclass models.

For illustrative purposes, we provide the finite-dimensional convex program in Li et al. (2025):

$$\min Z_4(\mathbf{b}) = \sum_{a\in A}\int_0^{x_a} t_a(v)dv + \sum_{rs\in RS}\sum_{p\in P_{rs}} m_p^{rs}\int_{b_{p-1}^{rs}}^{b_p^{rs}} \frac{1}{g_{rs}(v)}dv \tag{2.14a}$$

$$0 = b_0^{rs} \le b_1^{rs} \le b_2^{rs} \le \ldots \le b_{|P_{rs}|}^{rs} = q^{rs}, \quad \forall rs \in RS \tag{2.14b}$$

$$f_p^{rs} = b_p^{rs} - b_{p-1}^{rs}, \quad \forall rs \in RS, p \in P_{rs} \tag{2.14c}$$

$$x_a = \sum_{rs\in RS}\sum_{p\in P_{rs}} f_p^{rs}\delta_{ap}^{rs}, \quad \forall a \in A \tag{2.14d}$$

The objective function in Eq. (2.14a) is convex with respect to $\mathbf{b}$ and it consists of two terms. The first term is the Beckmann transformation (Beckmann et al., 1956) to model the congestion effect; the second term is the continuous monetary cost to model the tolling effect on travelers with a continuous distribution of VOT. Eq. (2.14b) ensures the correct ordering of boundaries, which is essential for maintaining flow conservation and the non-negativity of path flows. Eq. (2.14c) transfers boundaries to path flows. Eq. (2.14d) converts path flows into link flows. For equivalence conditions and solution properties, we refer readers to Li et al. (2025).

Compared to the infinite-dimensional formulation (2.11), the finite-dimensional formulation (2.14) offers several significant advantages: (1) **Tractability**: The formulation is finite-dimensional and can be directly solved by commercial optimization solvers; (2) **Model complexity**: The problem complexity (in terms of the number of variables and constraints) is comparable to that of a single-class problem; (3) **Algorithmic flexibility**: The boudnary variables contian rich information, which facilitates the development of efficient path-based solution algorithms.

2.4.2 Solution algorithms

For both infinite-dimensional and finite-dimensional formulations, the solution algorithms are based on a finite-dimensional reduction of the problem. The algorithms can be mainly divided into two categories depending on whehter the paths will be stored or not.

Link-based algorithms typically employ parametric shortest path techniques to construct the Pareto frontier (Fig. 2.6(b)), and iteratively load demand onto efficient paths to compute total link flows and attributes. The foundational work by Dial (1979) introduced a procedure to identify efficient paths and compute flow proportions in an uncongested network. Marcotte and Zhu

(1996) and Marcotte et al. (1996) modified the Frank-Wolfe (FW) algorithm to solve the continuous MCTA model (2.11), where the auxiliary link flow is obtained by a bicriteria loading process. Based on the restricted simplicial decomposition (RSD) method (Lawphongpanich and Hearn, 1984), Dial (1996) and Dial (1997) developed efficient link-based solution algorithms for the continuous MCTA problem, which he called T2 and was regarded as the most efficient method for the continuous MCTA problem in the 1990s. For the SUE problem with a continuously distributed VOT, Cantarella and Binetti (1998) used a method of successive averages (MSA)-type algorithm called cost averaging, which was adopted by Meng et al. (2012) and Liu et al. (2013) to solve the equilibrium model in their toll design problems.

Path-based algorithms explicitly manage path flows and the VOT boundaries or flow boundaries between them (see Figs. 2.6(c)-(e)). Leurent (1993) firstly utilized the MSA to solve the continuous MCTA model. Later, Leurent (1996) proposed a procedure of equalization by transfer (PET) to equilibrate path costs. Wu and Huang (2014b) adopted a path-based gradient projection (GP) algorithm (Jayakrishnan et al., 1994) for the SO-TA problem. A considerable advancement was recently made by Xie et al. (2021) with the single boundary adjustment (SBA) method, which employs a GP framework with Gauss-Seidel decomposition to directly and sequentially optimize VOT boundaries (Fig. 2.6(c)). The SBA method has proven versatile, being adapted for nonlinear time-toll trade-offs (Xu et al., 2023b), elastic demand (Xu et al., 2024a), and electric vehicle range anxiety (Xu et al., 2024b). An alternative boundary-type method is suggested by Li et al. (2025), which optimizes the path flow boundaries in model (2.14) (Fig. 2.6(e)).

2.4.3 Applications on different types of user heterogeneity

In the literature, the continuous multiclass approach has been applied to model preference heterogeneity and choice set heterogeneity, mainly concentrating on cases where user heterogeneity can be represented numerically (e.g., VOT).

2.4.3.1 Preference heterogeneity

Preference heterogeneity captures the variation in how travelers evaluate path costs, such as differing VOT or risk aversion. This is modeled by defining a generalized cost, such as Eq. (2.5),

that is a function of a traveler-specific parameter α. This cost can incorporate various deterministic attributes (e.g., time, toll, reliability) in either additive (Leurent, 1993) or nonadditive (Xu et al., 2023b) forms. The core modeling insight is that a continuous distribution of α across the population leads to varied path cost evaluations and path choices.

As summarized in Table 2.2, modeling preference heterogeneity—particularly via a continuously distributed VOT—constitutes the dominant application of the continuous multiclass approach. This includes models with linear time-cost trade-offs (Leurent, 1993; Dial, 1996, 1997; Xie et al., 2021; Li et al., 2025), as well as extensions to nonlinear trade-offs (Xu et al., 2023b) and heterogeneous risk preferences in mean-risk models (Amirgholy and Gonzales, 2017).

2.4.3.2 Choice set heterogeneity

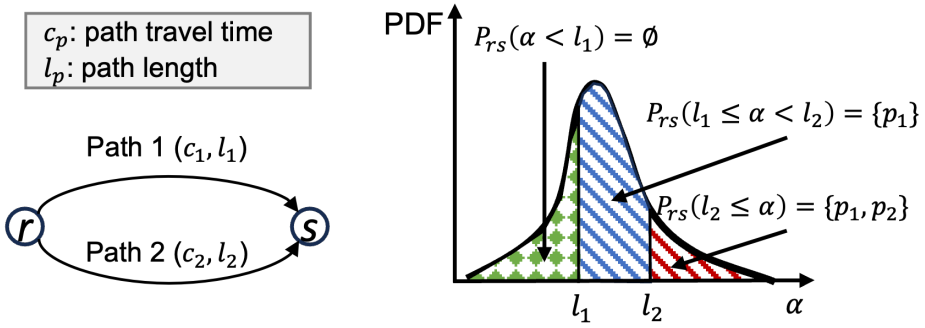
Choice set heterogeneity arises when travelers have access to different feasible sets of paths due to physical (e.g., ability to walk or cycle), informational (e.g., recommended routes by different apps), or behavioral constraints (e.g., range anxiety), etc. In other words, rather than a fixed path set P_{rs} for all travelers, the path set for heterogeneous travelers, denoted as $P_{rs}(\alpha)$, depends on a user-specific parameter α.

Few continuous MCTA studies addressed choice set heterogeneity. A prominent example is the modeling of range anxiety among electric vehicle (EV) drivers, where a driver's maximum driving range, α, defines their feasible path set:

$$P_{rs}(\alpha) = \{p \in P_{rs} | l_p \leq \alpha\} . \tag{2.15}$$

where l_p is the length of path p.

Xie et al. (2017) considered the case that maximum driving range α follows a continuous distribution. By ranking paths according to their length, it is also possible to find the range boundaries to segment the continuum of travelers (Fig. 2.7) to build finite-dimensional formulations (Xu et al., 2024b). This framework was later extended by Xie et al. (2019) to incorporate both range anxiety and mental accounting (e.g., budget constraints).



(a) An illustrative network (b) Driving range distribution

Figure 2.7: Illustration of choice set heterogeneity. (b) shows the choice sets of travelers with different driving ranges.

2.4.4 A summary table

Table 2.2 summarizes all the collected continuous MCTA studies based on the discussed aspects. For the ease of reading, we provide the abbreviations of terminologies:

- **Heterogeneity factor**: Value of time (VOT), risk preference (RP).
- **Equilibrium type**: All-or-nothing (AoN), user equilibrium (UE), stochastic user equilibrium (SUE).
- **Formulation**: Infinite-dimensional (Inf.-dim.), mathematical programming (MP), variational inequality (VI), nonlinear complementarity problems (NCP), fixed-point (FP), not mentioned or not applicable (NA).
- **Solution algorithm**: Link-based (L-), path-based (P-), method of successive averages (MSA), parallel tangent (PARTAN), Frank-Wolfe (FW), procedure of equalization by transfer (PET), restricted simplicial decomposition (RSD), gradient projection (GP), projected gradient (PG), single boundary adjustment (SBA).
- **Remarks**: Additional features or applications (e.g., elastic demand, congestion pricing).

Table 2.2: A summary of 32 continuous MCTA studies.

Study	Heter. factor	Equil. type	Formulation	Solution algorithm	Remarks
Preference heterogeneity (29 papers)					
Dial (1979)	VOT	AoN	NA	Parametric path finding	
Dafermos (1981)	VOT	UE	Inf.-dim. VI	NA	First formulation for continuous MCTA
Leurent (1993)	VOT	UE	MP	P-MSA	First MP formulation for continuous MCTA, Elastic demand
Marcotte et al. (1996)	VOT	UE	Inf.-dim.z VI	L-FW	
Leurent (1996)	VOT	UE	Inf.-dim. VI	P-PET	Elastic demand, side constraint
Marcotte and Zhu (1996)	VOT	UE	Inf.-dim. VI, Inf.-dim. MP	L-FW	
Dial (1996), Dial (1997)	VOT	UE	Inf.-dim. VI	L-RSD	Ingenious algorithm design.
Marcotte and Zhu (1997)	VOT	UE	Inf.-dim. VI	L-FW	
Marcotte (1999)	VOT	UE	Inf.-dim. VI, Inf.-dim. MP, FP, MP	NA	Four differente formulations
Cantarella and Binetti (1998)	VOT	SUE	Inf.-dim. FP	L-Cost averaging	Used simulation for discrete choice.
Leurent (1998b)	VOT	UE	Inf.-dim. VI, MP	NA	Discussion of multicriteria (Inf.-dim. VI) and bi-criteria (MP)
Leurent (1998a)	VOT	UE	VI	P-PET	Elastic demand, side constraint, sensitivity analysis
Dial (1999a), Dial (1999b)	VOT	UE	Inf.-dim. VI	L-RSD	Pricing/toll design
Marcotte and Zhu (2009)	VOT	UE	Inf.-dim. VI	NA	Toll design
Zhu et al. (2012)	VOT	UE	Inf.-dim. VI	NA	Toll design
Meng et al. (2012)	VOT	SUE	Inf.-dim. FP	L-Cost averaging	Elastic demand, asymmetric link travel time functions
Wang and Huang (2013)	VOT	UE	MP	NA	Elastic demand, SO
Liu et al. (2013)	VOT	SUE	Inf.-dim. FP	L-Cost averaging	Elastic demand, asymmetric link travel time functions

Wu and Huang (2014a)	VOT	UE	MP	NA	Pricing/toll design, SO
Wu and Huang (2014b)	VOT	UE	MP	P-GP	Pricing/toll design, SO
Zhu et al. (2015)	VOT	UE	Inf.-dim. VI	NA	Tradable credit scheme design
Zhao et al. (2016)	VOT	UE	MP	P-AoN	Greenhouse gas emission charge design
Amirgholy and Gonzales (2017)	RP	UE	NCP	P-Bi-stage method	Mean-standard travel time
Xie et al. (2021)	VOT	UE	MP	P-SBA-GP	An efficient algorithm to update the VOT boundaires.
Xu et al. (2023b)	Indifference curve	UE	MP	P-SBA-GP	Nonlinear trade-off between time and toll.
Xu et al. (2024a)	VOT	UE	MP	P-Two-stage SBA-GP	Elastic demand
Li et al. (2025)	VOT	UE	MP	P-SBA-GP	Comparison of different approaches
Choice set heterogeneity (3 papers)					
Xie et al. (2017)	Driving range	UE	MP, VI	P-PG	Trip chain
Xie et al. (2019)	Driving range, mental account	UE	MP	L-PARTAN FW	
Xu et al. (2024b)	Driving range	UE	VI	P-GP	Path-based solution algorithm.

2.5 Discrete multiclass approach

This section reviews the development of models, solution algorithms, and applications of the discrete multiclass approach, concluding with a summary table.

2.5.1 Model developments

The discrete multiclass approach originates from the modeling of different vehicle classes, where Dafermos (1972) introduced two sets of class-specific variables for trucks and passenger cars to model the resulting network equilibrium. In the pool of studies we collected, the earliest one focusing on user heterogeneity is Lam and Huang (1992), which incorporated different VOTs within a combined trip distribution and traffic assignment model.

The core of the discrete multiclass approach lies in the introduction of multiple copies of class-specific link and path flow variables, which can capture the route choices and flows of each user class, making it a direct extension of the single-class TA framework (Sheffi, 1985). Consequently, discrete multiclass models retain key mathematical properties of their single-class counterparts, such as convex objective functions (for MP) and unique equilibrium link flow patterns. For a detailed analysis of these properties, we refer interested readers to (Nagurney, 2000; Nagurney and Dong, 2002a).

This mathematical tractability has facilitated the wide application of discrete multiclass models to various problems, including modeling different VOTs (e.g., Nagurney and Dong, 2002a; Yang and Huang, 2004; Li et al., 2023b), different risk preferences (e.g., Yin and Ieda, 2001; Lo et al., 2006; Shao et al., 2008; Wu and Nie, 2013; Xu et al., 2014), different information levels (e.g., Yang and Meng, 2001; Miwa et al., 2010), different choice sets (e.g., Acemoglu et al., 2018), etc. Furthermore, all typical formulations that are applicable to single-class models can be adopted for discrete MCTA models, such as MP (Yang and Huang, 2004), VI (Nagurney, 2000; Nagurney and Dong, 2002a), NCP (Papadopoulos et al., 2021), FP (Ying, 2015), and gap function (Lo and Chen, 2000; Kashmiri and Lo, 2024).

Noteworthy, the introduction of class-specific variables increases the number of variables and constraints, as illustrated in Section 2.3.3, thereby increasing model complexity. Consequently, the solution time typically scales directly with the number of user classes (Ye, 2010; Li et al., 2025). To address this limitation, Li et al. (2023b) recently proposed an alternative formulation

for the multiclass BTA problem that uses boundary variables instead of class-specific link/path variables. This leads to a formulation similar to model (2.14) but with a nondifferentiable convex objective function. This formulation is proved to be equivalent to the conventional model (2.9) but offers significant advantages: (1) **Model complexity**: The problem complexity (in terms of the number of variables and constraints) is comparable to that of a single-class problem; (2) **Algorithmic flexibility**: The boudnary variables contian rich information, which facilitates the development of efficient path-based solution algorithms.

2.5.2 Solution algorithms

The mathematical properties of discrete MCTA models closely mirror those of single-class models, allowing for the straightforward adaptation of established traffic assignment algorithms. Standard methods such as the link-based Frank-Wolfe algorithm (LeBlanc et al., 1975) and path-based gradient projection (Jayakrishnan et al., 1994) can be applied with minimal modification. The primary adjustment involves implementing a class-specific loading process, which causes computational effort to scale linearly with the number of user classes (Li et al., 2025). While Table 2.3 comprehensively lists over 70 relevant studies, the following discussion highlights some representative algorithms.

Link-based algorithms have been widely applied. For instance, Lam and Huang (1992) employed the FW algorithm alongside Evans' algorithm for the combined trip distribution and TA problem. The modified projection method (MPM) has been frequently used for VI formulations of MCTA problems (Nagurney, 2000; Nagurney and Dong, 2002a,b). Similarly, Yang and Huang (2004) applied the FW algorithm to solve MCTA with heterogeneous VOTs, while Di et al. (2008) utilized a link-based MSA for a multiclass reliability-based traffic assignment problem. These studies represent the primary link-based approaches found in the literature.

Path-based algorithms are typically preferred for problems with complex route cost structures. Yin and Ieda (2001) employed a path-based gradient descent method to solve the multiclass reliability-based TA problem. Chen et al. (2011c) and Chen et al. (2012c) applied the path equilibration (PE) algorithm (Dafermos and Sparrow, 1969) to shift flow from the costliest path to the cheapest in the multiclass mean-risk TA problem. Other path-based methods include GP for problems with heterogeneous risk attitudes (Nie, 2011), partial linearization (PL) for multiclass SUE (Chen et al., 2012a), and disaggregate simplicial decomposition (DSD) for SUE with

heterogeneous knowledge levels (Xie and Liu, 2014). A recent advancement by Li et al. (2023b) introduced an alternative formulation with a subgradient projection (SGP) algorithm, inspired by the boundary adjustment idea for the multiclass BTA problem.

2.5.3 Applications on different types of user heterogeneity

Compared to the continuous multiclass approach, the discrete method accommodates a wider spectrum of user heterogeneity, including both numerical attributes like VOT and categorical distinctions such as CAV versus HDV users.

2.5.3.1 Preference heterogeneity

As summarized in Table 2.3, preference heterogeneity is the most extensively studied dimension of user variation in MCTA literature. The dominant focus is heterogeneous VOT, which has been applied across diverse contexts, including combined trip distribution and TA (Lam and Huang, 1992), elastic demand (Nagurney and Dong, 2002a), EV charging behavior (Mi et al., 2023), and ridesharing flow modeling (Li et al., 2023a). Notably, a considerable portion of studies specifically investigates toll design problems to steer equilibrium flows toward system optimum in the presence of VOT heterogeneity (Yang and Huang, 2004; Clark et al., 2009; Guo and Yang, 2009, 2010; Lindberg and Engelson, 2015; Fakhrmoosavi et al., 2021).

A second major stream addresses heterogeneous risk preferences in stochastic networks. Studies in this vein model how different risk attitudes—represented by metrics such as travel time budget (Lo et al., 2006; Shao et al., 2006a) or percentile travel times (Nie, 2011; Wu and Nie, 2013)—influence route choice. These reliability-based MCTA models have been effectively applied to network assessment (Yin and Ieda, 2001) and design (Szeto and and, 2016).

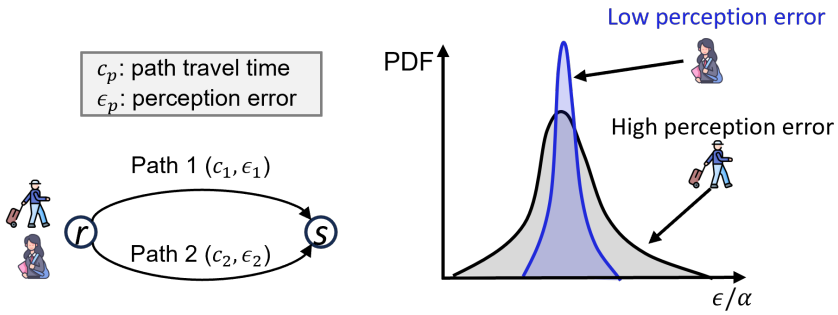
Other, less prevalent aspects of preference heterogeneity include safety concerns (Yang et al., 2015; Huang et al., 2020; Han et al., 2023), travel budget constraints (Bao et al., 2015), bicycling preferences for slope and environment (Ryu et al., 2021), and variations in traveler indifference curves (Wang and Ehrgott, 2013).

2.5.3.2 Information heterogeneity

Information heterogeneity arises from variations in how travelers perceive travel costs, influenced by differences in information quality, experience, or system access. This is modeled by introducing a perception error ϵ_p to the deterministic path cost c_p, resulting in a perceived path cost:

$$T_p(\alpha) = c_p + \frac{1}{\alpha}\epsilon_p, \quad p \in P_{rs} \tag{2.16}$$

where α is the scaling parameter denoting travelers' information level. The error term may follow different distributions, such as Normal (Daganzo and Sheffi, 1977), Gumbel (Fisk, 1980), or Weibull (Kitthamkesorn and Chen, 2013). In addition, the perceived path cost can be additive (e.g., Daganzo and Sheffi, 1977; Fisk, 1980) or multiplicative (Kitthamkesorn and Chen, 2013). In short, this formulation captures the inherent randomness in perceived costs due to varying information levels, as illustrated in Fig. 2.8.



(a) An illustrative network (b) Distribution of perception error

Figure 2.8: Illustration of information heterogeneity

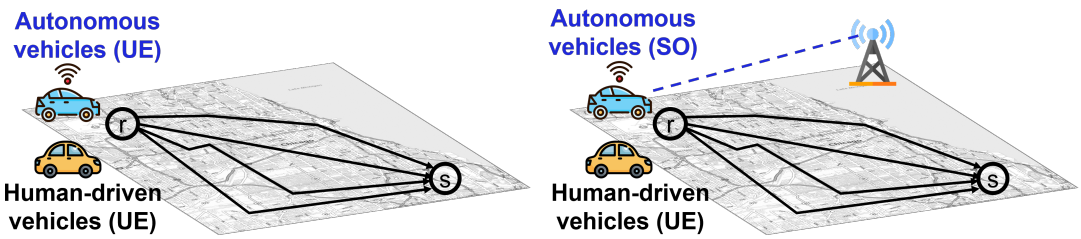
Early SUE models assumed a homogeneous information level across all travelers (Daganzo and Sheffi, 1977; Fisk, 1980). Subsequent research has relaxed this assumption. Yang (1998) introduced two distinct scaling parameters to differentiate between users with and without advanced traveler information systems (ATIS), a framework later extended to analyze ATIS adoption rates (Yang and Meng, 2001). Further developments include trip-length-based scaling parameters (Miwa et al., 2010), incorporation of road overlaps (Chen et al., 2012a), and models with heterogeneous reference points (Site and Filippi, 2011). More recent studies have explored the impact of real-time information sources (Zeng et al., 2017), heterogeneous knowledge levels (Xie and Liu, 2014), and varying unit information costs (Jiang et al., 2020).

2.5.3.3 Choice set heterogeneity

In discrete MCTA, Zhang and Yang (2015) modeled travelers' heterogeneous choice inertia by specifying different route sets for various travelers. Acemoglu et al. (2018) examined the scenario in which different navigation applications provide users with distinct link sets, demonstrating that offering more information could potentially make all users worse off.

2.5.3.4 Routing principle heterogeneity

Routing principle heterogeneity arises when travelers employ different decision-making rules, thereby relaxing the homogeneous behavioral assumption of Wardrop's user equilibrium principle. This framework is essential for capturing diverse behaviors, such as the contrast between rational utility maximizers and travelers using simple heuristics (e.g., choosing the familiar route). It is particularly relevant for modeling mixed traffic flows involving autonomous vehicles (AVs) and human-driven vehicles (HDVs), as illustrated in Fig. 2.9.



(a) Homogeneous routing principle (b) Heterogeneous routing principles

Figure 2.9: Illustration of routing principle heterogeneity

Recent studies, often motivated by AV technologies, have modeled this heterogeneity through distinct routing strategies. A common approach contrasts UE with SO routing. For instance, Zhang and Nie (2018) modeled a mix of SO-controlled and UE-controlled AVs, while Kashmiri and Lo (2022) distinguished between subscriber SO travelers and non-subscriber UE travelers. Other studies incorporate more nuanced behavioral differences: Wang et al. (2019) differentiated HDVs following a cross-nested logit rule from CAVs following UE, and Antúnez and Levin (2024) introduced eco-routing travelers who minimize fuel consumption or emissions. Espinosa Mireles de Villafranca and Roncoli (2024) further classified vehicles as altruistic (SO) or selfish (UE). Although these studies often involve different vehicle types, their primary focus remains on capturing fundamental differences in route choice principles.

2.5.3.5 Mixed heterogeneity

Finally, several discrete MCTA studies integrate multiple types of user heterogeneity, reflecting the complex interplay of factors in real-world travel behavior. For instance, some combine preference heterogeneity with other dimensions: Huang and Li (2007) and Jaber and O'mahony (2009) merged VOT classes with information heterogeneity, while Li et al. (2017) incorporated heterogeneous reference points, VOT, and perception errors. Other studies blend routing principles with preference heterogeneity, such as Zhang et al. (2008) and Kashmiri and Lo (2024), who modeled mixed equilibrium with heterogeneous VOT. Ye (2010) combined preference and choice set heterogeneity by considering different VOTs and link access for various vehicle types.

2.5.4 A summary table

Table 2.3 summarizes the reviewed MCTA studies. The following abbreviations are newly introduced compared to Table 2.2:

- **Heterogeneity factor**: Scaling parameter (SP).
- **Equilibrium type**: Mixed equilibrium (ME).
- **Solution algorithm**: Modified projection method (MPM), partial linearization (PL), disaggregate simplicial decomposition (DSD), path equilibration (PE).

Table 2.3: A summary of 76 discrete MCTA studies.

Study	Heter. factor	Equil. type	Formulation	Solution algorithm	Remarks
Preference heterogeneity (55 papers)					
Lam and Huang (1992)	VOT	UE	MP	L-FW, Evans	Combined trip distribution and traffic assignment.
Nagurney (2000)	Weights	UE	VI	L-MPM	
Yin and Ieda (2001)	RP	UE	NCP, Gap func.	P-Gradient descent	Network assessment.
Yang and Zhang (2002)	VOT	UE	MP	L-FW	Elastic demand, Pricing/toll design, bi-level
Nagurney and Dong (2002b)	Weights	UE	VI	L-MPM	Residential/employment location decision
Nagurney and Dong (2002a)	Weights	UE	VI	L-MPM	Elastic demand, flow-dependent weights
Yang et al. (2002)	VOT	UE	MP	L-FW	Elastic demand
Nagurney et al. (2002)	Weights	UE	VI	L-MPM	Enviromental considerations
Nagurney et al. (2003)	Weights	UE	VI	L-MPM	Telecommuting versus commuting
Yin and Yang (2004)	VOT	UE	VI	NA	Pricing/toll design
Yang and Huang (2004)	VOT	UE	MP	L-FW	System optimum, Pricing/toll design
De Cea et al. (2005)	VOT	UE	VI	L-Evans algorithm	A general implementation framework
Shao et al. (2006b)	RP	SUE	VI	P-Heuristic with projection	Travel time budget
Lo et al. (2006)	RP	UE	MP	NA	Travel time budget
Florian (2006)	VOT	UE	MP	NA	Explicit choice of tolled facilities
Shao et al. (2008)	RP	SUE	VI	P-Heuristic with projection	Elastic demand
Di et al. (2008)	RP	UE	NA	L-MSA	Travel time reliability
Han and Yang (2008)	VOT	UE	VI	NA	Price of anarchy

Clark et al. (2009)	VOT	UE	MP	NA	Elastic demand, Pricing/toll design
Guo and Yang (2009)	VOT	UE	MP	NA	System optimum, price of anarchy
Guo and Yang (2010)	VOT	UE	MP	NA	Congestion pricing and refunding
Chen et al. (2011c)	RP	UE	Gap func.	P-PE	Shortest path acceleration
Nie (2011)	RP	UE	VI	P-GP	Percentile travel time
Wang et al. (2012)	VOT	UE	VI	NA	Tradable credit scheme design
Chen et al. (2012c)	RP	UE	VI	P-PE	Vulnerability analysis
Jia et al. (2012)	VOT, RP	UE	Gap func.	Software: DTALite	
Wu and Nie (2013)	RP	UE	VI	P-GP	Percentile travel time
Wang and Ehrgott (2013)	Indifference curve	UE	MP	NA	Bi-objective, time surplus maximization
Fu et al. (2014)	RP	UE	Gap func.	L-MSA	Private car and bus
Nikolova and Stier-Moses (2014)	RP	UE	MP, VI	NA	Exogenous and endogenous standard deviations of travel time.
Xu et al. (2014)	RP	UE	VI	P-Modified alternating direction method	Mean-excess travel time.
Ge et al. (2014)	VOT	SUE	VI	NA	Elastic demand, Signal timing design
Ying (2015)	VOT	SUE	FP	NA	Residential choice
Yang et al. (2015)	Safety concerns	UE	VI	NA	Speed selection, speed limit, route choice
Sun et al. (2015)	RP, Uncertainty budgets	SUE	NA	P-MSA	Elastic demand
Rezaee and Hickman (2015)	VOT	UE	MP, VI	Software:EMME/4	
Bao et al. (2015)	VOT, Travel budget	UE	VI	L-MSA	
Lindberg and Engelson (2015)	VOT	UE	MP	NA	Sensitivity analysis
Szeto and and (2016)	RP	UE	Equilibrium constraints	NA	Reliable network design

Sun et al. (2016)	VOT	UE	MP	L-MSA	Parking choice model
Liu and Huang (2016)	RP	UE	VI	L-MSA	Pricing/toll design
Xu et al. (2018)	RP	UE	Gap func.	P-MSA	Passenger assignment in a railway network
Noruzoliaee et al. (2018)	VOT	UE	NCP	NA	AV and HV price design
Sun et al. (2019)	RP	UE	Entropy maximization	P-PL	Tri-criteria route choice
Zill et al. (2019)	VOT	SUE	MP	P-Bi-conjugate FW	Route choice set generation, path-size logit
Huang et al. (2020)	Safety concerns	UE	Gap func.	P-MSA	Consider crash risk
Zhang et al. (2021)	VOT	UE	VI	NA	Elastic demand
Papadopoulos et al. (2021)	VOT	UE	NCP	NA	Pricing-and-routing schemes for freight
Fakhrmoosavi et al. (2021)	VOT, RP	UE	Equilibrium constraints	L-MSA	Pricing/toll design
Ryu et al. (2021)	Biking experience	SUE	NA	P-Two-stage method	Bicycle traffic assignment
Xu et al. (2023a)	RP	UE	Gap func.	P-MSA	Passenger assignment in a railway network
Mi et al. (2023)	VOT	UE	MP	NA	Fast chargers are more expensive than ordinary chargers.
Li et al. (2023b)	VOT	UE	MP	P-SBA-SGP	An alternative formulation for MCTA
Li et al. (2023a)	VOT	SUE	VI	NA	Ridesharing with compensation optimization, side constraints
Wang et al. (2025)	VOT	UE	MP	P-Partial dualization algorithm	Entropy maximization
Information heterogeneity (8 papers)					
Yang (1998)	SP	SUE	MP	L-FW	High information quality (with ATIS) versus low information quality (without ATIS).
Yang and Meng (2001)	SP	SUE	MP	L-MSA	Analyzing the adoption rate of ATIS.

Miwa et al. (2010)	SP	SUE	MP	P-PL	OD-based scaling parameter
Site and Filippi (2011)	SP	SUE	FP	NA	Different reference points and scaling parameters
Chen et al. (2012a)	SP	SUE	MP	P-PL	Scaling effect and overlapping problem
Xie and Liu (2014)	Knowledge level	SUE	VI	P-DSD	Inertial degrees and knowledge structures.
Zeng et al. (2017)	SP	SUE	MP	P-MSA	ATIS and VMS
Jiang et al. (2020)	Information cost	SUE	Nonlinear equations	P-MSA	
Choice set heterogeneity (2 papers)					
Zhang and Yang (2015)	Route set	UE	VI	NA	Inertia UE
Acemoglu et al. (2018)	Link set	UE	MP	NA	Informational paradox and price of anarchy
Routing principle heterogeneity (5 papers)					
Zhang and Nie (2018)	UE and SO	ME	VI	L-FW	Control ratio determination
Wang et al. (2019)	CNL (HDV user) and UE (CAV user)	ME	VI	P-Route-swapping	Cross-nested logit and UE
Kashmiri and Lo (2022)	UE (non-subscriber) and SO (subscriber)	ME	NCP, Gap func.	NA	Routing of AVs
Espinosa Mireles de Villafranca and Roncoli (2024)	UE and SO	ME	MP	NA	
Antúnez and Levin (2024)	Eco-UE and UE	ME	VI	L-MSA	Emisson paradox
Mixed heterogeneity (6 papers)					
Huang and Li (2007)	VOT, SP	SUE	FP	L-MSA	Multiclass, multicriteira Logit SUE.
Zhang et al. (2008)	VOT, UE and Cournot-Nash equilibrium	ME	VI	NA	SO and toll design
Jaber and O'mahony (2009)	VOT, SP	ME	NCP	NA	Congestion pricing

Ye (2010)	VOT, Link set (vehicle types)	UE	NA	L-FW	Shortest path acceleration
Li et al. (2017)	VOT, SP	SUE	VI	L-MSA	Toll design
Kashmiri and Lo (2024)	VOT, UE (non-subscriber) and SO (subscriber)	ME	NCP, Gap func.	L-FW	Vehicle routing

2.6 Summary

This chapter has presented a systematic review of MCTA models, a critical field of study for enhancing the behavioral realism and policy relevance of TA by explicitly modeling user heterogeneity. Spanning nearly five decades of research (1979–2025), our analysis of 108 studies synthesizes the developments across two primary methodological paradigms: the discrete approach, which segments users into a finite number of classes, and the continuous approach, which represents heterogeneity via continuous distributions.

The discrete approach has proven to be a robust and versatile framework, directly extending single-class TA models through class-specific flow variables. Its mathematical tractability and straightforward implementation have made it the dominant choice for modeling various heterogeneity types, particularly, including preference, information, choice set, and routing principle, and mixed heterogeneity. In contrast, the continuous approach offers greater behavioral fidelity by capturing the full spectrum of user characteristics, though it demands more specialized solution algorithms and formulations. Our review highlights how recent methodological advances, particularly finite-dimensional formulations for continuous MCTA, have significantly bridged the computational gap between these approaches, enabling some continuous problems to be solved with single-class-level complexity.

Chapter 3

Discrete and continuous multiclass bi-criteria traffic assignment problems: A unified gradient projection algorithm

The multiclass bi-criteria traffic assignment (MBTA) problem considers travelers' bi-criteria (time and toll) route choice behaviors and user heterogeneity. The MBTA models can be classified as discrete or continuous based on whether the value of time (VOT) is modeled as a discrete or continuous variable. While both models have been suggested in the literature, their connections and differences remain underexplored. This study compares the discrete MBTA (DMBTA) and continuous MBTA (CMBTA) models and proposes a unified path-based gradient projection (GP) algorithm framework to solve both models. In the unified framework, three modules, including column generation, decomposition and equilibration, and convergence criteria, are customized for discrete and continuous models, respectively. With appropriate algorithmic designs, both problems can be solved effectively by the path-based GP algorithm. Extensive numerical experiments show that the equilibrium flow of the DMBTA model will fluctuate when the number of classes is small, and it will converge to the equilibrium flow of the CMBTA model with the increase in the number of classes. Additionally, in all test networks, the CMBTA model requires CPU time comparable to the DMBTA model with five classes and maintains a working path set smaller than that of the DMBTA model with three classes, demonstrating that the CMBTA model can achieve a refined solution with modest computational time and memory requirements.

3.1 Introduction

In urban transportation networks, travel time and toll cost are critical attributes influencing travelers' route choices between origin-destination (OD) pairs. The value of time (VOT), which linearly combines these two attributes into generalized costs, plays a central role in modeling how individuals trade off time and cost in response to toll charges. Empirical studies have shown that VOT varies significantly across individuals based on socioeconomic characteristics, trip purposes, and personal traits (Small and Yan, 2001; Small et al., 2005; Brownstone and Small, 2005; Cirillo and Axhausen, 2006). Predicting the spatial distribution of traffic flows for heterogeneous users with diverse VOTs in congested networks gives rise to the *multiclass bi-criteria traffic assignment* (MBTA) problem addressed in this study.

There are two different approaches for modeling user heterogeneity in the MBTA problem: discrete multiclass and continuous multiclass. The discrete multiclass approach divides travelers into a finite number of classes with class-specific demands and VOTs (e.g., Nagurney, 2000; Nagurney and Dong, 2002a; Yang and Huang, 2004; Huang and Li, 2007; Chen et al., 2010; Li et al., 2023b). This approach straightforwardly extends standard traffic assignment models (Beckmann et al., 1956; Sheffi, 1985; Patriksson, 2015) into discrete MBTA (DMBTA) models, allowing existing solution methods (including link-based, path-based, or bush-based methods, see Perederieieva et al. (2015) for a comprehensive review) to be readily adapted. Consequently, it is widely adopted in the literature. For instance, Yang and Huang (2004) investigated the multiclass bi-criteria system optimum problem with discrete VOTs and suggested using the link-based Frank-Wolfe (FW) algorithm to solve it, while Nagurney (2000) and Nagurney and Dong (2002a) considered flow-dependent criteria and weights for multiclass, multi-criteria networks. The discrete approach has also been applied to other traffic assignment problems, such as reliability-based multiclass traffic assignment models (Nie, 2011; Chen et al., 2011c; Xu et al., 2014) and bicycle traffic assignment (Ryu et al., 2021). Recently, Li et al. (2023b) focused on the DMBTA problem and proposed an alternative formulation with a nondifferentiable objective function, which led to the development of a subgradient projection algorithm. It is worth noting, however, that while the DMBTA model is a straightforward extension of the standard traffic assignment model, it introduces discretization errors compared to the continuous multiclass, and its computational cost increases significantly with the number of classes (Ye, 2010; Li et al.,

2023b).

The continuous multiclass approach treats VOT as a continuous random variable across the population (e.g., Leurent, 1993; Dial, 1996, 1997; Marcotte, 1999; Wu and Huang, 2014a; Calderone et al., 2017; Xie et al., 2021). This method is considered more general and refined, as it accounts for an infinite number of user classes, each with a unique VOT (Marcotte, 1999). In a non-congested network, Dial (1979) proposed a parametric shortest path algorithm and a loading procedure to assign travelers with continuously distributed VOT onto the network. Leurent (1993) proposed perhaps the first path-based mathematical programming (MP) model for the continuous MBTA (CMBTA) problem, which is solved by the method of successive averages (MSA). Dial (1996) formulated the CMBTA problem as a variational inequality problem and proposed a link-based algorithm called T2 (Dial, 1997). Marcotte and Zhu (1996) modified the link-based FW algorithm to solve the CMBTA model. Marcotte (1999) discussed different formulations for the CMBTA problem, including both finite and infinite-dimensional formulations. Wu and Huang (2014a) investigated the system optimum problem where the travelers' VOTs are continuously distributed. Calderone et al. (2017) extended the work of Leurent (1993) to further consider the possible negative VOT values. More recently, Xie et al. (2021) developed a path-based algorithm to solve the CMBTA problem, which was later extended to consider nonadditive path cost (Xu et al., 2023b) and elastic demand (Xu et al., 2024a).

Although both the DMBTA model and CMBTA model have been explored in the literature, there is no research to systematically compare these two models theoretically and numerically. Many questions are still unclear: How big is the gap between the DMBTA model and the CMBTA model? How many classes are required for the discrete method to approximate the continuous method? Which method has more appealing computational performance in terms of CPU time and memory consumption?

This research aims to answer these questions by analyzing the connections and differences between the CMBTA and DMBTA models and developing a unified path-based gradient projection (GP) algorithm framework for solving both models. Specifically, in the unified path-based GP algorithm, the column generation process, decomposition scheme, equilibration process, and detailed implementation issues are customized and compared for both problems. With appropriate algorithmic designs, both problems can be solved effectively by the path-based GP algorithm.

Numerical results reveal that the equilibrium flow of the DMBTA model fluctuates when the number of classes is small but converges to the equilibrium flow of the CMBTA model as the number of classes increases. Specifically, across various networks, the DMBTA model with 50 classes achieves a mean squared error (MSE) of link flows below 30, indicating a close approximation to the CMBTA model's equilibrium flow. Additionally, the CMBTA model requires CPU time comparable to the DMBTA model with five classes and maintains a working path set smaller than that of the DMBTA model with three classes. These findings demonstrate that the CMBTA model achieves a refined solution with modest computational time and compact memory requirements.

The rest of this chapter is structured as follows. Section 3.2 introduced the discrete and continuous approaches for the MBTA problem. Section 3.3 presented the unified path-based GP framework with corresponding customizations for the CMBTA problem and the DMBTA problem. The numerical experiments are conducted in Section 3.4. Section 3.5 concludes the chapter and identifies some future directions.

3.2 Bi-criteria traffic assignment problem with multi-user classes

3.2.1 Problem setting

A transportation network can be modeled as a directed graph $G = (N, A)$, where N is the set of nodes and A is the set of directed links. Let q^{rs} denote the travel demand for OD pair rs, $rs \in RS \subseteq N \times N$. P_{rs} denote the set of simple paths connecting OD pair rs. The set of all paths is $P = \cup_{rs \in RS} P_{rs}$. Let f_p^{rs} denote the flow on path p connecting OD pair rs. The link flow x_a is the summation of path flows traversing this link:

$$x_a = \sum_{rs \in RS} \sum_{p \in P_{rs}} f_p^{rs} \delta_{ap}^{rs}, \quad \forall a \in A \tag{3.1}$$

where $\delta_{ap}^{rs} = 1$ if link a is on path p connecting OD pair rs, 0 otherwise.

Each link has two attributes: flow-dependent travel time $t_a(\cdot)$ and exogenously determined toll τ_a. The path travel time c_p^{rs} and path toll m_p^{rs} are the sums of link travel times and link tolls,

respectively.

$$c_p^{rs} = \sum_{a \in A} t_a(x_a)\delta_{ap}^{rs}, \quad \forall rs \in RS, \forall p \in P_{rs} \tag{3.2}$$

$$m_p^{rs} = \sum_{a \in A} \tau_a \delta_{ap}^{rs}, \quad \forall rs \in RS, \forall p \in P_{rs} \tag{3.3}$$

In this thesis, we assume the link travel time functions $t_a(\cdot)$ are positive and strictly increasing functions, i.e., $t_a'(\cdot) > 0$.

3.2.2 Route choices of heterogeneous users

Travelers are making route choice considering two criteria, i.e., time and toll. To simplify the problem, transport modelers (e.g., Dial, 1979; Nagurney, 2000) define the general cost of a path as the linear combination of time and toll:

$$d_p^{rs}(\alpha) = c_p^{rs} + \frac{1}{\alpha} m_p^{rs}, \quad \forall rs \in RS, p \in P_{rs} \tag{3.4}$$

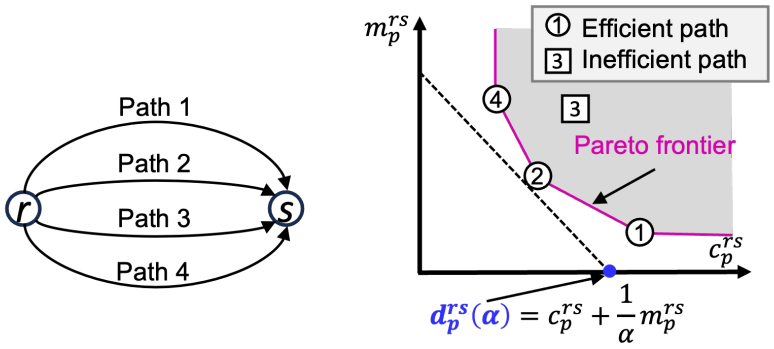
In Eq. (3.4), VOT α denotes the travelers' preferences regarding travel time and monetary cost. A large value of α indicates that travelers assign a greater weight to a unit of travel time compared to a unit of monetary cost.

The MBTA problem assumes that travelers are heterogeneous in terms of VOT α. Based on whether α is modeled as a discrete random variable or a continuous random variable, the resulting MBTA problem can be classified as discrete or continuous, respectively. Noteworthy, no matter how α is modeled, travelers with different VOTs are assumed to choose the path that minimizes their generalized cost, which is one of the (supported) *efficient paths* in the space of time and toll.

Definition 3.1 (Efficient paths)**.** *Given an OD pair rs and path set P_{rs}, the paths $p \in P_{rs}$ that can minimize $d_p^{rs}(\alpha)$ in Eq. (3.4) given some $\alpha > 0$ are called the efficient paths. The set of efficient paths is:*

$$\left\{ p \in P_{rs} \middle| \exists \alpha > 0, d_p^{rs}(\alpha) \leq d_{p'}^{rs}(\alpha), \forall p' \in P_{rs} \right\} \tag{3.5}$$

Fig. 3.1(a) shows an illustrative network with 4 paths, and Fig. 3.1(b) shows these 4 paths in the space of time (x-axis) and toll (y-axis). In Fig. 3.1(b), the slope of the dashed line passing a path is determined by α, and the intersection of the dashed line with the x-axis is the generalized



(a) An illustrative network (b) Generalized cost and efficient paths

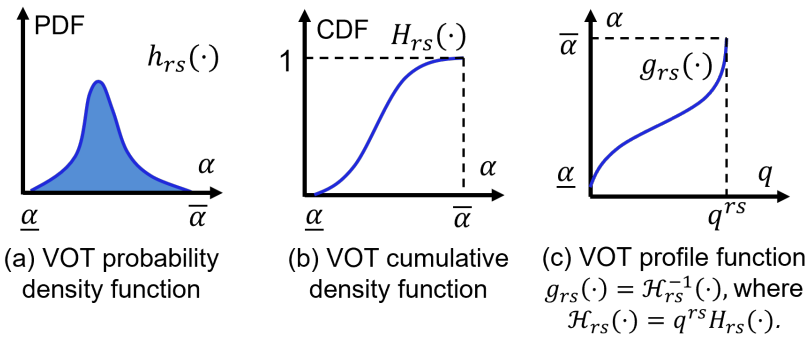
Figure 3.1: An illustrative network and efficient paths. In (b), the dashed lines passing through a path with travel time c_p^{rs} and toll m_p^{rs} represent the contour of generalized travel cost function $d_p^{rs}(\alpha) = c_p^{rs} + \frac{1}{\alpha} m_p^{rs}$. The slope of the dashed line is determined by α, and its x-intercept corresponds to the generalized cost $d_p^{rs}(\alpha)$ given α.

cost $d_p^{rs}(\alpha)$ of this path given α. Among these paths, paths 1, 2, and 4 can minimize $d_p^{rs}(\alpha)$ in Eq. (3.4) given some $\alpha > 0$, hence efficient; path 3 is not. The efficient paths form a Pareto frontier. The complexity of the MBTA problem lies in that (i) travelers with different VOTs may choose different paths on the Pareto frontier, determining the flow distribution within the network; (ii) the path travel time is flow-dependent, which in turn alters the shape of the Pareto frontier and subsequently affects travelers' choices.

3.2.3 Continuous multiclass bi-criteria traffic assignment problem

3.2.3.1 Continuous multiclass

The CMBTA problem assumes that, for each OD pair rs, α is a continuous random variable on support $\mathcal{A} = [\underline{\alpha}, \overline{\alpha}]$ with a probability density function (PDF) $h_{rs}(\cdot)$ (see Fig. 3.2(a)) and cumulative density function (CDF) $H_{rs}(\cdot)$ (see Fig. 3.2(b)). $H_{rs}(\cdot)$ is assumed to be invertible, which forbids a single traveler from having multiple VOTs. In this study, we further scale $H_{rs}(\cdot)$ by the travel demand, i.e., $\mathcal{H}_{rs}(\cdot) = q_{rs} H_{rs}(\cdot)$, and use the inverse of the scaled function $g_{rs}(\cdot) = \mathcal{H}_{rs}^{-1}(\cdot)$ to characterize the VOTs among the travel demand. We call $g_{rs}(\cdot)$ as *VOT profile function*, which is a positive and increasing function with respect to flow, i.e., $g'_{rs}(\cdot) > 0$, as shown in Fig. 3.2(c). The VOT profile function specifies the VOT for each traveler in an increasing order, implying an infinite number of user classes if we differentiate travelers by VOT.



(a) VOT probability density function

(b) VOT cumulative density function

(c) VOT profile function $g_{rs}(\cdot) = \mathcal{H}_{rs}^{-1}(\cdot)$, where $\mathcal{H}_{rs}(\cdot) = q^{rs} H_{rs}(\cdot)$.

Figure 3.2: Characterization of randomly distributed VOT among the population.

3.2.3.2 Equilibrium

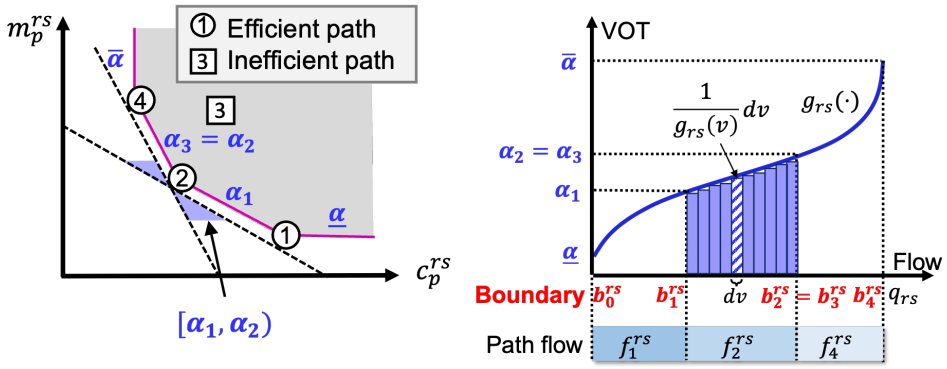
Without loss of generality, for any origin-destination (OD) pair rs, we assume that the paths are (weakly) increasingly ordered by toll.

$$m_1^{rs} \leq m_2^{rs} \leq ... \leq m_{|P_{rs}|}^{rs}, \quad \forall rs \in RS \tag{3.6}$$

The geometry of the Pareto frontier provides valuable information for determining the path choices of travelers with different VOTs. On the Pareto frontier, the slope of the line connecting two adjacent efficient paths is the threshold where the path choice will change. As shown in Fig. 3.3(a), when $\alpha < \alpha_1$, path 1 minimizes the generalized cost; when $\alpha = \alpha_1$, path 1 and path 2 are equally optimal; when α exceeds α_1 but smaller than α_2, path 2 gives the minimum generalized cost. In other words, each efficient path has its *dominant VOT set*. The dominant VOT sets of paths 1, 2, and 4 are $[\underline{\alpha}, \alpha_1)$, $[\alpha_1, \alpha_2)$, and $[\alpha_3, \overline{\alpha}]$, respectively. Notably, for the inefficient paths, the dominant VOT set is empty. In this example, we simply set $\alpha_3 = \alpha_2$ to let path 3's dominant VOT set be $[\alpha_2, \alpha_3) = \emptyset$.

With the dominant VOT sets established, we can determine the path flows using the VOT profile function. As shown in Fig. 3.3(b), travelers within intervals $[b_0^{rs}, b_1^{rs})$, $[b_1^{rs}, b_2^{rs})$, $[b_2^{rs}, b_3^{rs})$, and $[b_3^{rs}, b_4^{rs}]$ will choose path 1, path 2, path 3, and path 4, respectively. We refer to these dividing points b_p^{rs} as *flow boundaries*. For a given OD pair rs with $|P_{rs}|$ paths, there will be $|P_{rs}| + 1$ boundaries while the two ends are fixed, i.e., $b_0^{rs} = 0$ and $b_{|P_{rs}|}^{rs} = q^{rs}$. These boundaries and path flows have a unique mapping.

$$f_p^{rs} = g_{rs}^{-1}(\alpha_p) - g_{rs}^{-1}(\alpha_{p-1}) = b_p^{rs} - b_{p-1}^{rs}, \quad \forall rs \in RS, p \in P_{rs} \tag{3.7}$$



(a) Dominant VOT sets of paths (b) Path flow assignment

Figure 3.3: Illustration of efficient paths and VOT profile function

Notably, because the paths are ordered increasingly by tolls as per Eq. (3.6), the VOTs on the Pareto frontier are also arranged in increasing order, i.e., $\underline{\alpha} \leq \alpha_1 \leq \alpha_2 \leq ... \leq \overline{\alpha}$. Moreover, since travelers are similarly arranged by their VOTs in an increasing order ($g'_{rs}(\cdot) > 0$ as illustrated in Fig. 3.3(b)), the corresponding flow boundaries also follow an increasing order:

$$0 = b_0^{rs} \leq b_1^{rs} \leq b_2^{rs} \leq ... \leq b_{|P_{rs}|}^{rs} = q^{rs}, \quad \forall rs \in RS \tag{3.8}$$

The physical interpretation of this ordering is that travelers with different VOTs are pre-matched to their respective path choices with different tolls.

Instead of taking path flows as decision variables, the boundaries can be utilized as decision variables to define the CMBTA user equilibrium (UE) condition.

Definition 3.2 (CMBTA-UE). *A feasible path flow boundary vector* $\mathbf{b}^* = (b_p^{rs*})^\top_{\forall rs \in RS, p \in P_{rs}}$ *corresponds to a CMBTA-UE flow if for any OD pair* rs*, if* $b_p^{rs*} - b_{p-1}^{rs*} > 0$ *(i.e.,* $f_p^{rs*} > 0$*), then*

$$\underbrace{c_p^{rs}(\mathbf{b}^*) + \frac{1}{g_{rs}(b)} m_p^{rs} \leq c_{p'}^{rs}(\mathbf{b}^*) + \frac{1}{g_{rs}(b)} m_{p'}^{rs}, \quad \forall p' \in P_{rs}}_{(i)}, \underbrace{b \in \left[b_{p-1}^{rs*}, b_p^{rs*}\right]}_{(ii)} \tag{3.9}$$

In words, the CMBTA-UE condition states that: *(i) used paths must be efficient; (ii) efficient paths receive fair flows.* Or alternatively: *no traveler can reduce his/her generalized cost by unilaterally switching to another path.*

3.2.3.3 Formulation

The CMBTA-UE problem can be formulated as an equivalent convex programming problem:

$$\text{CMBTA} \quad \min Z_1(\mathbf{b}) = \underbrace{\sum_{a\in A}\int_0^{x_a} t_a(v)dv}_{\text{Beckmann transformation}} + \underbrace{\sum_{rs\in RS}\sum_{p\in P_{rs}} m_p^{rs}\int_{b_{p-1}^{rs}}^{b_p^{rs}} \frac{1}{g_{rs}(v)}dv}_{\text{Continuous monetary cost}} \tag{3.10a}$$

$$0 = b_0^{rs} \le b_1^{rs} \le b_2^{rs} \le \ldots \le b_{|P_{rs}|}^{rs} = q^{rs}, \quad \forall rs \in RS \tag{3.10b}$$

$$f_p^{rs} = b_p^{rs} - b_{p-1}^{rs}, \quad \forall rs \in RS, p \in P_{rs} \tag{3.10c}$$

$$x_a = \sum_{rs\in RS}\sum_{p\in P_{rs}} f_p^{rs}\delta_{ap}^{rs}, \quad \forall a \in A \tag{3.10d}$$

The objective function in Eq. (3.10a) is convex with respect to $\mathbf{b}$ and it consists of two terms. The first term is the Beckmann transformation (Beckmann et al., 1956) to model the congestion effect; the second term is the continuous monetary cost to model the tolling effect on travelers with a continuous distribution of VOT. See Fig. 3.3(b) for a graphical illustration. Eq. (3.10b) ensures the correct ordering of boundaries, which is essential for maintaining flow conservation and the non-negativity of path flows. Eq. (3.10c) transfers boundaries to path flows. Eq. (3.10d) converts path flows into link flows.

Noteworthy, the objective function in Eq. (3.10a) can be expressed as:

$$\begin{aligned}
\min Z_1(\mathbf{b}) &= \sum_{a\in A}\int_0^{x_a} t_a(v)dv + \sum_{rs\in RS}\sum_{p\in P_{rs}} m_p^{rs}\int_{b_{p-1}^{rs}}^{b_p^{rs}} \frac{1}{g_{rs}(v)}dv \\
&= \sum_{a\in A}\int_0^{x_a} t_a(v)dv + \sum_{rs\in RS}\sum_{p\in P_{rs}} \left(m_p^{rs}\int_0^{b_p^{rs}} \frac{1}{g_{rs}(v)}dv - m_p^{rs}\int_0^{b_{p-1}^{rs}} \frac{1}{g_{rs}(v)}dv \right) \\
&= \underbrace{\sum_{a\in A}\int_0^{x_a} t_a(v)dv}_{\text{Convex}} + \sum_{rs\in RS}\sum_{p=1,\ldots,|P_{rs}|-1} \underbrace{\underbrace{(m_p^{rs} - m_{p+1}^{rs})}_{\le 0}\underbrace{\int_0^{b_p^{rs}} \frac{1}{g_{rs}(v)}dv}_{\text{Concave}}}_{\text{Convex}} + \underbrace{m_{|P_{rs}|}^{rs}\int_0^{q^{rs}} \frac{1}{g_{rs}(v)}dv}_{\text{Constant}}
\end{aligned}$$

The first term, the Beckmann transformation, is known to be convex with respect to link flow x_a because its second derivative $t_a'(x_a) > 0$. The link flow $x_a = \sum_{rs\in RS}\sum_{p\in P_{rs}}(b_p^{rs} - b_{p-1}^{rs})\delta_{ap}^{rs}$ is a linear function of $\mathbf{b}$. Therefore, the first term is convex with respect to $\mathbf{b}$. For the second term, the VOT profile function $g_{rs}(\cdot)$ is positive and strictly increasing, implying that $\frac{1}{g_{rs}(\cdot)}$ is positive and strictly decreasing. Thus, the integral $\int_0^{b_p^{rs}} \frac{1}{g_{rs}(v)}dv$ is concave with respect to b_p^{rs},

as its second derivative $-\frac{g'_{rs}(b_p^{rs})}{(g_{rs}(b_p^{rs}))^2} \leq 0$. However, $m_p^{rs} - m_{p+1}^{rs} \leq 0$ as per Eq. (3.6), making the product $(m_p^{rs} - m_{p+1}^{rs}) \int_0^{b_p^{rs}} \frac{1}{g_{rs}(v)} dv$ convex. The third term, $\sum_{rs \in RS} m_{|P_{rs}|}^{rs} \int_0^{q^{rs}} \frac{1}{g_{rs}(v)} dv$, is constant because q^{rs} is the fixed demand for each O-D pair rs, and thus convex. Consequently, $Z_1(\mathbf{b})$ is convex with respect to $\mathbf{b}$.

Theorem 3.1 (Equivalence)**.** *A solution* $\mathbf{b}^*$ *of the CMBTA model (3.10) corresponds to a CMBTA-UE flow in Definition 3.2.*

Proof. By attaching Lagrange multipliers λ_p^{rs} to the boundary order constraint (3.10b), the Lagrangian of the CMBTA model (3.10) is given by:

$$L(\mathbf{b}, \lambda) = \sum_{a \in A} \int_0^{x_a} t_a(v) dv + \sum_{rs \in RS} \sum_{p \in P_{rs}} \int_{b_{p-1}^{rs}}^{b_p^{rs}} \frac{m_p^{rs}}{g_{rs}(v)} dv + \sum_{rs \in RS} \sum_{p \in P_{rs}} \lambda_p^{rs} \left(b_{p-1}^{rs} - b_p^{rs} \right) \quad (3.11)$$

The KKT conditions read:

$$\text{Stationarity} \qquad \frac{\partial L(\mathbf{b}, \lambda)}{\partial b_p^{rs}} = 0, \quad \forall rs \in RS, \forall p \in P_{rs} \qquad (3.12)$$

$$\text{Primal feasibility} \qquad b_p^{rs} - b_{p-1}^{rs} \geq 0, \quad \forall rs \in RS, \forall p \in P_{rs} \qquad (3.13)$$

$$\text{Dual feasibility} \qquad \lambda_p^{rs} \geq 0, \quad \forall rs \in RS, \forall p \in P_{rs} \qquad (3.14)$$

$$\text{Complementary slackness} \qquad \lambda_p^{rs} \left(b_p^{rs} - b_{p-1}^{rs} \right) = 0, \quad \forall rs \in RS, \forall p \in P_{rs} \qquad (3.15)$$

We check the stationarity condition (3.12):

$$\begin{aligned} \frac{\partial L(\mathbf{b}, \lambda)}{\partial b_p^{rs}} &= \sum_{a \in A} \frac{\partial \sum_{a \in A} \int_0^{x_a} t_a(v) dv}{\partial x_a} \times \frac{\partial x_a}{\partial b_p^{rs}} + \frac{\partial \left(\int_{b_{p-1}^{rs}}^{b_p^{rs}} \frac{m_p^{rs}}{g_{rs}(v)} dv + \int_{b_p^{rs}}^{b_{p+1}^{rs}} \frac{m_{p+1}^{rs}}{g_{rs}(v)} dv \right)}{\partial b_p^{rs}} - \lambda_p^{rs} + \lambda_{p+1}^{rs} \\ &= \sum_{a \in A} t_a \left(x_a \right) \left(\delta_{ap}^{rs} - \delta_{ap+1}^{rs} \right) + \frac{1}{g_{rs} \left(b_p^{rs} \right)} \left(m_p^{rs} - m_{p+1}^{rs} \right) - \lambda_p^{rs} + \lambda_{p+1}^{rs} \\ &= c_p^{rs} - c_{p+1}^{rs} + \frac{1}{g_{rs} \left(b_p^{rs} \right)} \left(m_p^{rs} - m_{p+1}^{rs} \right) - \lambda_p^{rs} + \lambda_{p+1}^{rs} = 0, \quad \forall rs \in RS, p = 1, ..., |P_{rs}| - 1 \end{aligned} \quad (3.16)$$

which implies that:

$$c_p^{rs} + \frac{m_p^{rs}}{g_{rs} \left(b_p^{rs} \right)} - \lambda_p^{rs} = c_{p+1}^{rs} + \frac{m_{p+1}^{rs}}{g_{rs} \left(b_p^{rs} \right)} - \lambda_{p+1}^{rs}, \quad \forall rs \in RS, p = 1, ..., |P_{rs}| - 1 \quad (3.17)$$

Recall that for any OD pair rs, the paths are increasingly ordered by tolls, and g_{rs} is an increasing function. We have:

$$\begin{cases} m_p^{rs} \leq m_{p+1}^{rs} \\ g_{rs}(b_p^{rs}) \leq g_{rs}(b_{p+1}^{rs}) \end{cases} \qquad \forall rs \in RS, p = 1, ..., |P_{rs}| - 1 \tag{3.18}$$

since $b_p^{rs} \leq b_{p+1}^{rs}$ according to the primal feasibility condition (3.13). The order information Eq. (3.18) helps us to further derive the following inequalities from Eq. (3.17):

$$\begin{cases} ... \geq c_{p-1}^{rs} + \frac{m_{p-1}^{rs}}{g_{rs}(b)} - \lambda_{p-1}^{rs} \geq c_p^{rs} + \frac{m_p^{rs}}{g_{rs}(b)} - \lambda_p^{rs} \\ c_p^{rs} + \frac{m_p^{rs}}{g_{rs}(b)} - \lambda_p^{rs} \leq c_{p+1}^{rs} + \frac{m_{p+1}^{rs}}{g_{rs}(b)} - \lambda_{p+1}^{rs} \leq ... \end{cases} \qquad \forall rs \in RS, p = 2, ..., |P_{rs}|-1, b \in \left[b_{p-1}^{rs}, b_p^{rs}\right] \tag{3.19}$$

Further, for any OD pair rs, if $b_p^{rs} - b_{p-1}^{rs} > 0$ (path p carries positive flows), then $\lambda_p^{rs} = 0$ according to the complementary slackness condition (3.15). Eq. (3.19) becomes:

$$c_p^{rs} + \frac{m_p^{rs}}{g_{rs}\left(b\right)} \leq c_{p'}^{rs} + \frac{m_{p'}^{rs}}{g_{rs}\left(b\right)} - \lambda_{p'}^{rs}, \quad \forall p' \in P_{rs}, b \in \left[b_{p-1}^{rs}, b_p^{rs}\right] \tag{3.20}$$

As the dual variables are nonnegative as per Eq. (3.14), we can safely drop them to obtain:

$$c_p^{rs} + \frac{m_p^{rs}}{g_{rs}\left(b\right)} \leq c_{p'}^{rs} + \frac{m_{p'}^{rs}}{g_{rs}\left(b\right)}, \quad \forall p' \in P_{rs}, b \in \left[b_{p-1}^{rs}, b_p^{rs}\right] \tag{3.21}$$

This completes the proof. □

Theorem 3.2 (Uniqueness)**.** *(i) If the travel time functions $t_a(\cdot)$ are strictly increasing, then the equilibrium link flow $\mathbf{x}^*$ is unique. (ii) Further, for any OD pair rs, if every path $p \in P_{rs}$ has a unique toll m_p^{rs}, then the solution $\mathbf{b}^*$ of the CMBTA model (3.10) is unique, and so is the path flow $\mathbf{f}^*$.*

Proof. See Theorem 3 in Leurent (1993). □

The CMBTA model (3.10) is related to Leurent (1993) and Calderone et al. (2017) but with three improvements. First, while both previous studies utilize path flows as decision variables, leading to complex formulations that complicate equilibrium analysis and algorithm development, our approach offers a cleaner version. By parameterizing the original path flow variables

as boundary variables, we unify the flow conservation constraints and non-negativity of path flows into a series of simple box constraints (3.10b), thereby simplifying both model formulation and algorithm development. Second, we define the VOT profile function to streamline the model and enhance interpretability. This function enables the CMBTA model (3.10) to have an objective function similar to that of the DMBTA model (3.23), differing only in the integration versus summation of monetary costs. Finally, we present an alternative, concise proof of equivalence.

Marcotte (1999) and Xie et al. (2021) also explored the CMBTA problem using boundary variables. Their approaches utilized VOT boundaries (α_p in Fig. 3.3(a)) and PDF functions (see Fig. 3.2(a)), resulting in more complex model formulations. In contrast, this study leverages the path flow boundaries and VOT profile function to characterize the problem, ultimately leading to a simpler formulation.

3.2.4 Discrete multiclass bi-criteria traffic assignment problem

3.2.4.1 Equilibrium

Given an origin-destination (OD) pair rs, the DMBTA model assumes VOT α is molded as a discrete random variable. Equivalently, the DMBTA model discretizes the travel demand q^{rs} into $|K|$ class-specific demands $\{q_k^{rs} \mid k \in K\}$, where $q^{rs} = \sum_{k \in K} q_k^{rs}$. Travelers with similar VOTs are grouped into the same class, with the class-specific VOT α_k^{rs} defined as the mean VOT of travelers within that class. Each class is assumed to select the path that minimizes its generalized cost.

Fig. 3.4 provides an example with 3 classes of travelers. According to their VOTs, classes 1 and 2 opt for path 1, while travelers in class 3 choose path 4. Notably, although path 2 is also efficient, no traveler will choose it due to the class VOT specifications in this scenario. This contrasts with Fig. 3.3(a), where all efficient paths are used. This distinction also highlights a difference between the CMBTA and DMBTA. In CMBTA, all the efficient paths that can minimize $d_p^{rs}(\alpha)$ for some $\alpha \in [\underline{\alpha}, \overline{\alpha}]$ will be used. In contrast, some of the paths in the middle of the Pareto frontier may not be used in DMBTA, depending on the specification of $\{\alpha_k^{rs}, k \in K\}$.

By defining class-specific path flow variable f_p^{rsk}, which denotes the flow for the OD pair rs

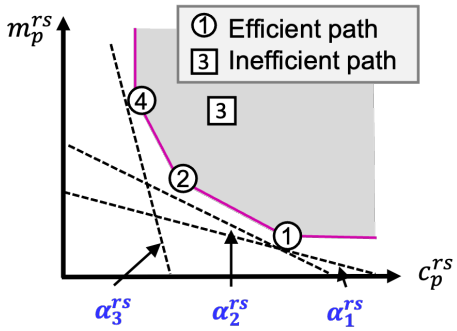


Figure 3.4: Path choices of discrete user classes.

and class k on path p, the DMBTA-UE condition can be established as follows.

Definition 3.3 (DMBTA-UE). *A feasible class-specific path flow vector* $\tilde{\mathbf{f}}^* = (f_p^{rsk*})^\top_{\forall rs \in RS, k \in K, p \in P_{rsk}}$ *corresponds to a DMBTA-UE flow if for all OD pair* rs *and class* k*, if* $f_p^{rsk*} > 0$*, then*

$$c_p^{rs}(\tilde{\mathbf{f}}^*) + \frac{1}{\alpha_k^{rs}} m_p^{rs} \leq c_{p'}^{rs}(\tilde{\mathbf{f}}^*) + \frac{1}{\alpha_k^{rs}} m_{p'}^{rs}, \quad \forall p' \in P_{rs} \tag{3.22}$$

3.2.4.2 Formulation

The DMBTA-UE problem can be formulated as an equivalent convex programming problem (Yang and Huang, 2004):

$$\text{DMBTA} \quad \min Z_2(\tilde{\mathbf{f}}) = \underbrace{\sum_{a \in A} \int_0^{x_a} t_a(v) dv}_{\text{Beckmann transformation}} + \underbrace{\sum_{rs \in RS} \sum_{p \in P_{rsk}} m_p^{rs} \sum_{k \in K} \frac{1}{\alpha_k^{rs}} f_p^{rsk}}_{\text{Discrete monetary cost}} \tag{3.23a}$$

$$\sum_{p \in P_{rsk}} f_p^{rsk} = q_k^{rs}, \quad \forall rs \in RS, k \in K \tag{3.23b}$$

$$x_a^k = \sum_{rs \in RS} \sum_{p \in P_{rsk}} f_p^{rsk} \delta_{ap}^{rs}, \quad \forall a \in A, k \in K \tag{3.23c}$$

$$x_a = \sum_{k \in K} x_a^k, \quad \forall a \in A \tag{3.23d}$$

$$f_p^{rsk} \geq 0, \quad \forall rs \in RS, k \in K, p \in P_{rsk} \tag{3.23e}$$

The objective function in Eq. (3.23a) consists of two terms: the Beckmann transformation and the discrete monetary cost. The latter models the tolling effect on travelers with a discrete set of VOTs. In the objective function, the first term is known to be convex with respect to link flow

x_a and the link flow x_a is a linear function of $\tilde{\mathbf{f}}$. Therefore, it is convex with respect to $\tilde{\mathbf{f}}$. The second term is a linear function of $\tilde{\mathbf{f}}$. Thus, the objective function, as the sum of a convex term and a linear term, is convex with respect to $\tilde{\mathbf{f}}$. Eq. (3.23b) is the flow conservation constraint for each class. Eq. (3.23c) converts class path flow into class link flows. Eq. (3.23d) merges class link flow into total link flows. Eq. (3.23e) is the nonnegativity constraint of class path flows. For the solution property of the DMBTA model, it is known that only the total link flow $\mathbf{x}^*$ is unique, while all other variables are not in general networks (Nagurney, 2000; Nagurney and Dong, 2002a).

Although the DMBTA model looks simpler than the CMBTA model, it has two shortcomings: (1) discretization error and (2) massive variables. For the discretization error, there are no universal guidelines to specify how many classes are needed to keep the discretization error in an acceptable range. In this case, increasing the number of classes is the only way to obtain more refined results. At the same time, compared to single-class traffic assignment models, the DMBTA model needs to keep $|K|$ copies of variables. Especially, there will be $|P| \times |K|$ class path flow variables, which will consume a huge amount of computer memory as well as longer CPU time in practical transportation networks.

3.3 A unified path-based gradient projection algorithm

In this section, we employ the framework of the path-based gradient projection (GP) algorithm (Bertsekas, 1976; Jayakrishnan et al., 1994; Chen et al., 2012b, 2014) to develop customized algorithms for solving the CMBTA model (3.10) and the DMBTA model (3.23), as shown in Fig. 3.5. For a general convex programming problem defined as $\min_{\mathbf{x}\in X} Z(\mathbf{x})$, the GP algorithm updates the variables at iteration n by integrating the projection method (Goldstein, 1964; Levitin and Polyak, 1966) with Newton's method:

$$\mathbf{x}^{n+1} = \texttt{Proj}_X \left(\mathbf{x}^n - \beta(n)[\nabla^2 Z(\mathbf{x}^n)]^{-1} \nabla Z(\mathbf{x}^n)\right) \tag{3.24}$$

where $\mathbf{x}^n$ is the variable vector at iteration n; $\texttt{Proj}_X(\cdot)$ denotes the projection operator onto feasible set X; $\nabla Z(\mathbf{x}^n)$ and $\nabla^2 Z(\mathbf{x}^n)$ are the Jacobian and Hessian of $Z(\cdot)$ at $\mathbf{x}^n$; $\beta(n)$ is a step size satisfying Armijo rule, and 1 is a robust choice for the path-based GP algorithm

(Jayakrishnan et al., 1994).

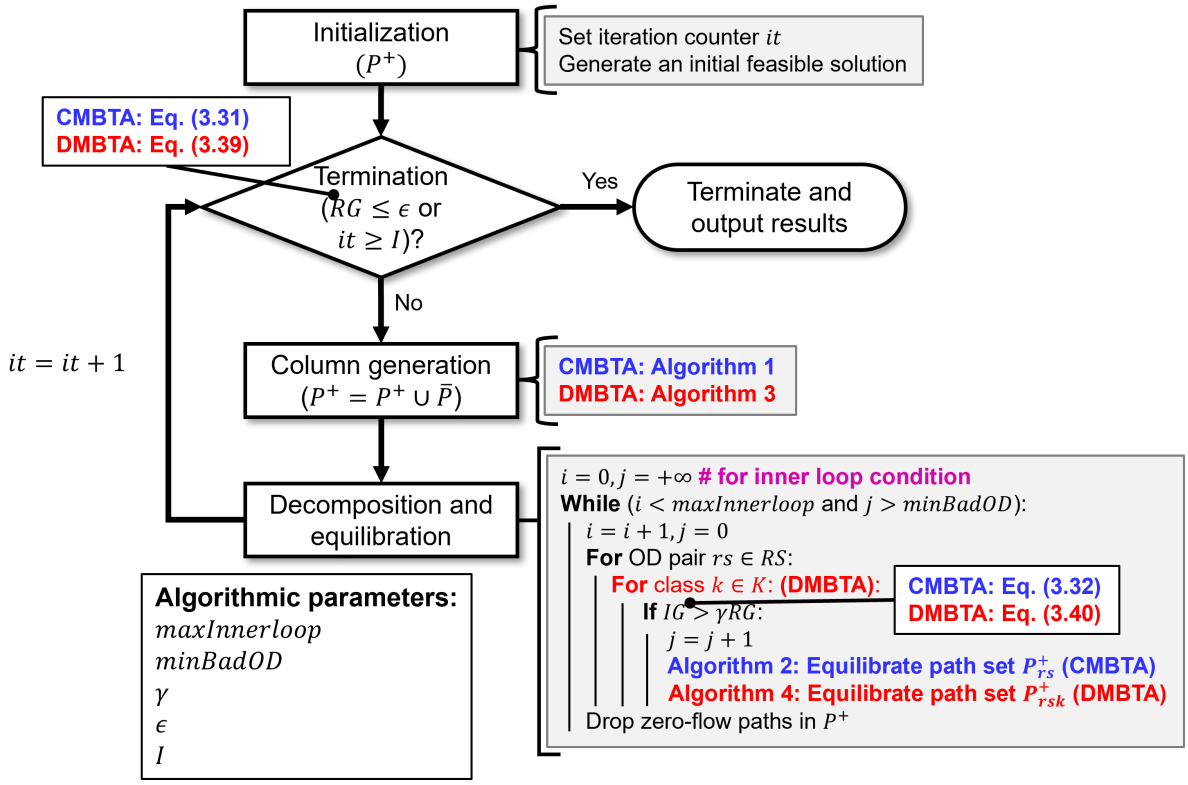


Figure 3.5: A general framework of the path-based gradient projection algorithm.

To efficiently implement the GP algorithm for solving path-based TA formulations, several additional features need to be incorporated (Ryu et al., 2014; Perederieieva et al., 2015; Galligari and Sciandrone, 2019):

1. **Column generation**. In large transportation networks, the number of paths can increase exponentially. Therefore, the path-based algorithms typically maintain a working path set $P^+ \subseteq P$ and utilize the column generation method to generate auxiliary paths $p \in \bar{P}$ to improve the working path set as needed (Perederieieva et al., 2015; Jiang, 2024), i.e., $P^+ = P^+ \cup \bar{P}$.

2. **Decomposition and equilibration**. Due to the separability of the feasible set, the original TA problem is usually decomposed by OD pairs and solved sequentially using a Gauss-Seidel method, which is also called the one-at-a-time equilibration strategy (Chen and Jayakrishnan, 1998).

3. **Inner loop**. In traffic assignment, column generation involves finding the auxiliary paths for all OD pairs at each iteration. This process can be time-consuming and may result in

auxiliary paths that replicate those already present in the working path set. Therefore, it is effective to incorporate a strategy called inner loop to fully equilibrate the working path set before performing the next column generation (Perederieieva et al., 2015; Xie et al., 2018; Xu et al., 2020; Tan et al., 2022; Zhang et al., 2023).

For both CMBTA and DMBTA problems, we can implement an all-or-nothing assignment to generate an initial feasible solution as well as a working path set. Specifically, in the DMBTA case, the class-specific shortest path is generated for each class and each OD pair, with the corresponding class-specific OD demand assigned onto it. For the CMBTA problem, the bi-criteria shortest path algorithm is used to generate the set of efficient paths between each OD pair; the OD demand is then distributed among these paths based on the VOT PDF and the dominant VOT range associated with each path.

The main differences between solving the CMBTA problem (3.10) and the DMBTA problem (3.23) lie in the column generation and the decomposition and equilibration processes. Besides, the convergence criteria of the two models also have distinct formulas. We now provide the details on these differences in the path-based GP framework.

3.3.1 Solving bi-criteria TAP with continuous multi-user classes

3.3.1.1 Column generation

The detailed procedure is shown in Algorithm 1. The CMBTA model (3.10) is built on efficient paths, as shown in Fig. 3.3(a) and Definition 3.2. Therefore, at each iteration, the column generation process will generate auxiliary efficient paths for each OD pair to improve the working path set. For this purpose, this study adopts the *Dijkstra-like bi-criteria shortest path algorithm* developed by Sedeño-Noda and Raith (2015) as a building block to generate one-to-all efficient paths[1]. It is worth mentioning that, when adding new paths, we need to maintain the path toll order as per Eq. (3.6), which is executed by lines 5-8 in Algorithm 1. At the same time, a temporary boundary value is assigned to the newly added path (cf. line 7) for the following equilibration operation.

[1]According to our preliminary experiments, this label-setting method is around 10 times faster than Dial's label-correcting method (Dial, 1979, 1997) in the Chicago Sketch network (https://github.com/bstabler/TransportationNetworks).

Algorithm 1 Column generation for the CMBTA model.

Input: Network $G = (N, A)$, link travel times $\{t_a, \forall a \in A\}$, and link tolls $\{\tau_a, \forall a \in A\}$.
Output: Working path set $P^+ = \cup_{rs \in RS} P_{rs}^+$ and boundaries $\left\{b_p^{rs}, \forall rs \in RS, p \in P_{rs}^+\right\}$.

1: **for** origin $r \in R$ **do**
2: Call the *Dijkstra-like bi-criteria shortest path algorithm* (Sedeño-Noda and Raith, 2015) to generate the one-to-all efficient path tree T_r from r.
3: **for** destination $s \in S$ **do**
4: Get the auxiliary efficient paths set $\bar{P}_{rs} = T_r(s)$.
5: **for** path $p \in \bar{P}_{rs} \setminus P_{rs}^+$ **do**
6: Find the largest path index p' in P_{rs}^+ such that $m_{p'}^{rs} \leq m_p^{rs}$.
7: Add path p into P_{rs}^+ at position p' and set boundary $b_p^{rs} = b_{p'}^{rs}$.
8: **end for**
9: **end for**
10: **end for**

3.3.1.2 Decomposition and equilibration: boundary update

The CMBTA model (3.10) can be decomposed as a series of one-dimensional subproblems with respect to b_p^{rs}:

$$\min \bar{Z}_1(b_p^{rs}) = \sum_{a \in A} \int_0^{x_a} t_a(v)dv + \int_{b_{p-1}^{rs}}^{b_p^{rs}} \frac{1}{g_{rs}(v)} m_p^{rs} dv + \int_{b_p^{rs}}^{b_{p+1}^{rs}} \frac{1}{g_{rs}(v)} m_{p+1}^{rs} dv \tag{3.25a}$$

$$f_p^{rs} = b_p^{rs} - b_{p-1}^{rs}, \quad \forall rs \in RS, p \in P_{rs} \tag{3.25b}$$

$$x_a = \sum_{rs \in RS} \sum_{p \in P_{rs}} f_p^{rs} \delta_{ap}^{rs}, \quad \forall a \in A \tag{3.25c}$$

$$b_{p-1}^{rs} \leq b_p^{rs} \leq b_{p+1}^{rs} \tag{3.25d}$$

for all $rs \in RS$ and $p \in 1, ..., |P_{rs}| - 1$. The only constraint (3.25d) is a simple box constraint.

The first- and second-order derivatives of the objective function (3.25a) are given by:

$$v_p^{rs} = \frac{\partial \bar{Z}_1(b_p^{rs})}{\partial b_p^{rs}} = c_p^{rs} - c_{p+1}^{rs} + \frac{1}{g_{rs}\left(b_p^{rs}\right)} \left(m_p^{rs} - m_{p+1}^{rs}\right), \quad \forall rs \in RS, p \in 1, ..., |P_{rs}| - 1 \tag{3.26}$$

$$s_p^{rs} = \frac{\partial^2 \bar{Z}_1(b_p^{rs})}{\partial \left(b_p^{rs}\right)^2} = \sum_{a \in A} t_a'\left(x_a\right) \left(\delta_{ap}^{rs} - \delta_{ap+1}^{rs}\right)^2 - \frac{g_{rs}'\left(b_p^{rs}\right)}{g_{rs}^2\left(b_p^{rs}\right)} \left(m_p^{rs} - m_{p+1}^{rs}\right), \quad \forall rs \in RS, p \in 1, ..., |P_{rs}| - 1 \tag{3.27}$$

Therefore, at iteration n, the boundary updating equation for subproblems reads:

$$\begin{aligned} b_p^{rs}(n+1) &= \texttt{Proj}_{[b_{p-1}^{rs}, b_{p+1}^{rs}]}\left(b_p^{rs}(n) - \frac{\beta(n)}{s_p^{rs}(n)} v_p^{rs}(n)\right) \\ &= \begin{cases} \min\left\{b_p^{rs}(n) - \frac{\beta(n)}{s_p^{rs}(n)} v_p^{rs}(n), b_{p+1}^{rs}\right\}, & \text{if } v_p^{rs} < 0 \\ \max\left\{b_p^{rs}(n) - \frac{\beta(n)}{s_p^{rs}(n)} v_p^{rs}(n), b_{p-1}^{rs}\right\}, & \text{if } v_p^{rs} > 0 \end{cases} \\ &\forall rs \in RS, p \in 1, ..., |P_{rs}| - 1, \end{aligned} \tag{3.28}$$

Fig. 3.6 provides a graphical illustration for the boundary updating equation. Note that, according to the first-order derivative (3.26), if

$$v_p^{rs} < 0 \Rightarrow g_{rs}\left(b_p^{rs}\right) < \frac{m_p^{rs} - m_{p+1}^{rs}}{c_{p+1}^{rs} - c_p^{rs}}, \tag{3.29}$$

which means the VOT $g_{rs}(b_p^{rs})$ corresponding to boundary b_p^{rs} is smaller than the slope between path p and path $p+1$, as shown in Fig. 3.6(a). Then, we need to increase $g_{rs}(b_p^{rs})$ by shifting boundary b_p^{rs} forward, i.e., shifting flow from path $p+1$ to path p, as shown in Fig. 3.6(b). At the same time, the box constraint (3.25d) imposes an upper bound for b_p^{rs}.

Alternatively, if

$$v_p^{rs} > 0 \Rightarrow g_{rs}\left(b_p^{rs}\right) > \frac{m_p^{rs} - m_{p+1}^{rs}}{c_{p+1}^{rs} - c_p^{rs}}, \tag{3.30}$$

which means the VOT $g_{rs}(b_p^{rs})$ corresponding to boundary b_p^{rs} is larger than the slope between path p and path $p+1$, as shown in Fig. 3.6(c). Then, we need to decrease $g_{rs}(b_p^{rs})$ by shifting boundary b_p^{rs} backward, i.e., shifting flow from path p to path $p+1$, as shown in Fig. 3.6(d). Similarly, the box constraint (3.25d) imposes an lower bound for b_p^{rs}.

In short, the essence of the boundary updating equation is to use the projected Newton's method to shift the boundary between path p and $p+1$ subjected to box constraint (3.25d). Algorithm 2 describes the procedure for equilibrating the working path set P_{rs}^{+} connecting OD pair rs in detail.

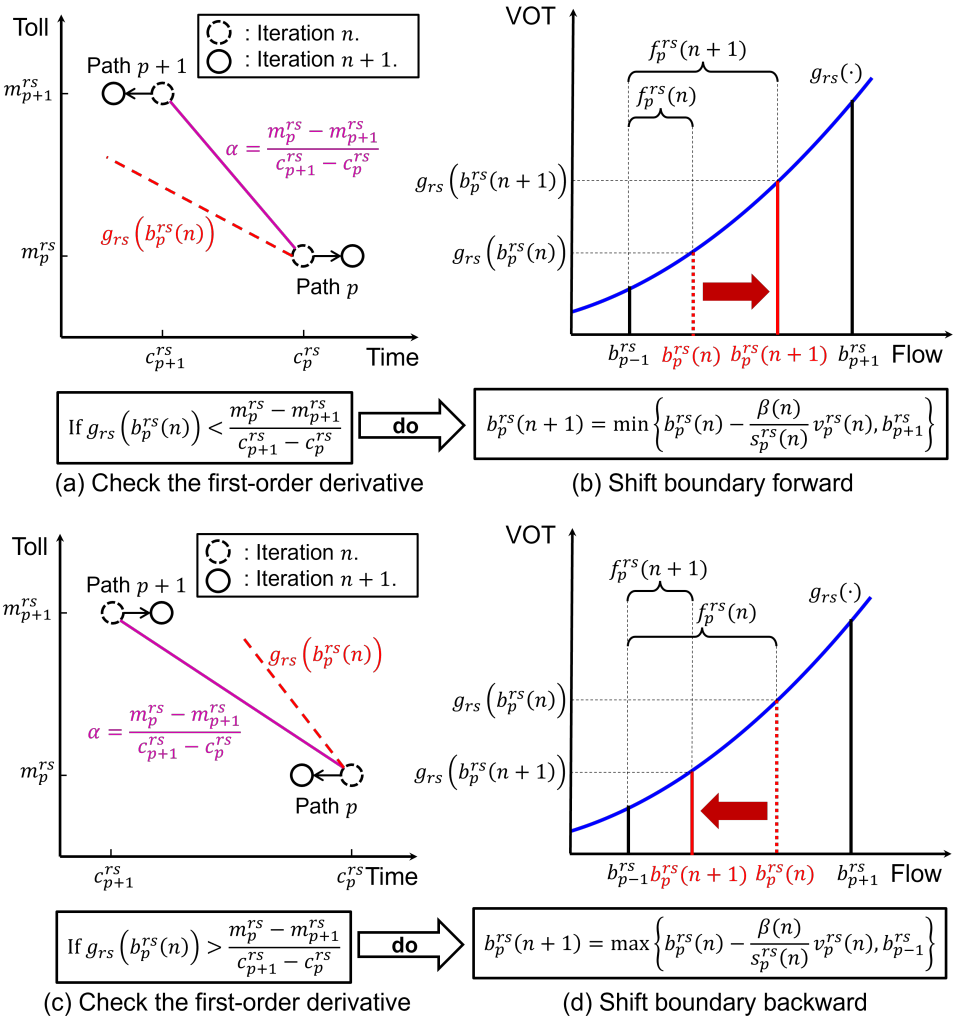


Figure 3.6: An illustration of the boundary updating equation.

3.3.1.3 Convergence criterion

To measure the convergence, we use the relative gap to gauge the "distance" between the current solution and the CMBTA-UE condition, which is defined as:

$$RG = 1 - \frac{\sum_{rs \in RS} \sum_{\bar{p} \in \bar{P}_{rs}} \left(\left(b_{\bar{p}}^{rs} - b_{\bar{p}-1}^{rs}\right) c_{\bar{p}}^{rs} + \int_{b_{\bar{p}-1}^{rs}}^{b_{\bar{p}}^{rs}} \frac{1}{g_{rs}(v)} m_{\bar{p}}^{rs} dv \right)}{\sum_{rs \in RS} \sum_{p \in P_{rs}^{+}} \left(\left(b_p^{rs} - b_{p-1}^{rs}\right) c_p^{rs} + \int_{b_{p-1}^{rs}}^{b_p^{rs}} \frac{1}{g_{rs}(v)} m_p^{rs} dv \right)} \tag{3.31}$$

where $\bar{P}_{rs}$ and P_{rs}^{+} are the auxiliary path set and the working path set for OD pair rs, respectively.

For OD pair rs with working path set P_{rs}^{+}, the inner gap is defined as the maximum absolute

Algorithm 2 Equilibrate path set P_{rs}^{+}.

Input: Working path set P_{rs}^{+} and boundaries $\left\{b_p^{rs}(n), \forall rs \in RS, p \in P_{rs}^{+}\right\}$.
Output: Working path set P_{rs}^{+} and new boundaries $\left\{b_p^{rs}(n+1), \forall rs \in RS, p \in P_{rs}^{+}\right\}$.

1: **for** path $p \in 1, ..., |P_{rs}^{+}| - 1$ **do**
2: Compute path time c_p^{rs} and c_{p+1}^{rs} based on Eq. (3.2).
3: Calculate v_p^{rs} and s_p^{rs} based on Eq. (3.26) and Eq. (3.27), respectively.
4: Calculate boundary $b_p^{rs}(n+1)$ based on Eq. (3.28), and set $\Delta = b_p^{rs}(n+1) - b_p^{rs}(n)$.
5: Update path flows $f_p^{rs} = f_p^{rs} + \Delta$ and $f_{p+1}^{rs} = f_{p+1}^{rs} - \Delta$.
6: Update link flow x_a, link time $t_a(x_a)$, and link time derivative $t_a'(x_a)$ of each link a on path p and path $p+1$.
7: **end for**

value of first-order derivatives:

$$IG_{rs} = \max\left\{|v_p^{rs}|, \forall p = 1, ..., |P_{rs}^{+}| - 1\right\}, \quad \forall rs \in RS \tag{3.32}$$

3.3.2 Solving bi-criteria TAP with discrete multi-user classes

3.3.2.1 Column generation

For the DMBTA problem, the column generation process is similar to that of the CMBTA problem, as shown in Algorithm 3. The difference is that an auxiliary path for each class and each OD pair will be generated in DMBTA (lines 5-8).

Algorithm 3 Column generation for the DMBTA model.

Input: Network $G = (N, A)$, link travel times $\{t_a, \forall a \in A\}$, and link tolls $\{\tau_a, \forall a \in A\}$.
Output: Working path set $P^{+} = \cup_{rs \in RS, k \in K} P_{rsk}^{+}$.

1: **for** each origin $r \in R$ **do**
2: Call the *Dijkstra-like bi-criteria shortest path algorithm* (Sedeño-Noda and Raith, 2015) to generate the one-to-all efficient path tree T_r from r.
3: **for** each destination $s \in S$ **do**
4: Get the auxiliary efficient paths set $\bar{P}_{rs} = T_r(s)$.
5: **for** each class $k \in K$ **do**
6: Get an auxiliary path $\bar{p} = \arg\min_{p \in \bar{P}_{rs}} d_p^{rs}(\alpha_k^{rs})$.
7: $P_{rsk}^{+} = P_{rsk}^{+} \cup \{\bar{p}\}$.
8: **end for**
9: **end for**
10: **end for**

3.3.2.2 Decomposition and equilibration: class-specific path flow update

The DMBTA model (3.23)'s feasible set has an equality constraint (3.23b) and a non-negativity constraint (3.23e). To simplify the projection operation, we can eliminate the class flow conservation constraints (3.23b) by specifying a reference path $\bar{p}_{rsk}$ (the one with minimum generalized cost) for each OD pair and each user class. The flow of this path can be represented as

$$f_{\bar{p}_{rsk}}^{rsk} = q_k^{rs} - \sum_{p \in P_{rsk} \setminus \{\bar{p}_{rsk}\}} f_p^{rsk}, \quad \forall rs \in RS, k \in K \tag{3.33}$$

and this equation then can be moved to the objective function. Now, the reformulated DMBTA problem only has nonnegativity constraints:

$$\min \bar{Z}_2(\tilde{\mathbf{f}}) \tag{3.34a}$$

$$f_p^{rsk} \geq 0, \quad \forall rs \in RS, k \in K, p \in P_{rsk} \setminus \{\bar{p}_{rsk}\} \tag{3.34b}$$

Also, based on the spirit of Gauss-Seidel method, we can decompose the reformulated model (3.34) as a series of one-dimensional subproblems with respect to f_p^{rsk}:

$$\min \bar{Z}_2\left(f_p^{rsk}\right) \tag{3.35a}$$

$$f_p^{rsk} \geq 0 \tag{3.35b}$$

for all $rs \in RS$, $k \in K$, $p \in P_{rsk} \setminus \{\bar{p}_{rsk}\}$.

As the only constraint (3.35b) is a simple nonnegativity constraint, solving subproblem (3.35) is actually shifting flow between path p and reference path $\bar{p}_{rsk}$ connecting OD pair rs of user class k. The path-based GP algorithm utilizes the projected Newton's method to solve subproblem (3.35), where the first- and second-order derivatives of the objective function are given by:

$$v_p^{rsk} = \frac{\partial \bar{Z}_2\left(f_p^{rsk}\right)}{f_p^{rsk}} = c_p^{rs} - c_{\bar{p}_{rsk}}^{rs} + \frac{1}{\alpha_k}\left(m_p^{rs} - m_{\bar{p}_{rsk}}^{rs}\right), \quad \forall rs \in RS, k \in K, p \in P_{rsk} \setminus \{\bar{p}_{rsk}\} \tag{3.36}$$

$$s_p^{rsk} = \frac{\partial^2 \bar{Z}_2\left(f_p^{rsk}\right)}{\partial\left(f_p^{rsk}\right)^2} = \sum_{a \in A} t_a'\left(x_a\right)\left(\delta_{ap}^{rs} - \delta_{a\bar{p}_{rsk}}^{rs}\right)^2, \quad \forall rs \in RS, k \in K, p \in P_{rsk} \setminus \{\bar{p}_{rsk}\} \tag{3.37}$$

At iteration n, the flow shift from path p to reference path $\bar{p}_{rsk}$ is:

$$\Delta = \begin{cases} \max\left\{\frac{\beta(n)}{s_p^{rsk}(n)} v_p^{rsk}(n), -f_{\bar{p}_{rsk}}^{rsk}(n)\right\}, & \text{if } v_p^{rsk} < 0 \\ \min\left\{\frac{\beta(n)}{s_p^{rsk}(n)} v_p^{rsk}(n), f_p^{rsk}(n)\right\}, & \text{if } v_p^{rsk} > 0 \end{cases} \quad \forall rs \in RS, k \in K, p \in P_{rsk} \setminus \{\bar{p}_{rsk}\}, \tag{3.38}$$

Algorithm 4 describes the procedure for equilibrating the working path set P_{rsk}^+ connecting OD pair rs for user class k in detail.

Algorithm 4 Equilibrate path set P_{rsk}^+.

Input: Working path set P_{rsk}^+ and class path flows $\left\{f_p^{rsk}(n), \forall rs \in RS, k \in K, p \in P_{rsk}^+\right\}$.
Output: Working path set P_{rsk}^+ and new class path flows $\left\{f_p^{rsk}(n+1), \forall rs \in RS, k \in K, p \in P_{rsk}^+\right\}$.

1: Find the minimum generalized cost path $\bar{p}$ in P_{rsk}^+.
2: **for** path $p \in P_{rsk}^+ \setminus \{\bar{p}\}$ **do**
3: Compute path time c_p^{rs} and c_{p+1}^{rs} based on Eq. (3.2).
4: Calculate v_p^{rs} and s_p^{rs} based on Eq. (3.36) and Eq. (3.37), respectively.
5: Calculate shift flow Δ based on Eq. (3.38).
6: Update class path flows $f_p^{rsk} = f_p^{rsk} - \Delta$ and $f_{\bar{p}}^{rsk} = f_{\bar{p}}^{rsk} + \Delta$.
7: Update link flow x_a, link time $t_a(x_a)$, and link time derivative $t_a'(x_a)$ of each link a on path p and path $\bar{p}$.
8: **end for**

3.3.2.3 Convergence criterion

For the DMBTA model, the relative gap is defined as:

$$RG = 1 - \sum_{rs \in RS} \sum_{k \in K} \sum_{p \in P_{rsk}^+} \frac{f_p^{rsk} d_{\bar{p}_{rsk}}^{rs}\left(\alpha_k^{rs}\right)}{f_p^{rsk} d_p^{rs}\left(\alpha_k^{rs}\right)} \tag{3.39}$$

where $\bar{p}_{rsk}$ and P_{rsk}^+ are the minimum generalized cost path and the working path set connecting OD pair rs for user class k, respectively.

For working path set P_{rsk}^+, the inner gap is defined as the maximum absolute value of first-order derivatives:

$$IG_{rsk} = \max\left\{|v_p^{rsk}|, \forall p \in P_{rsk}^+ \setminus \{\bar{p}_{rsk}\}\right\}, \quad \forall rs \in RS, k \in K \tag{3.40}$$

3.3.3 Summary

Here we briefly summarize the key features of the path-based GP algorithm for solving the CMBTA and DMBTA problems in Table 3.1.

Table 3.1: Key features of the path-based GP algorithm for solving the CMBTA and DMBTA problems

	CMBTA	**DMBTA**
Feasible set simplification	No	Eliminate class flow conservation constraint (3.23b)
Column generation		
-Path-finding algorithm	Bi-criteria parametric shortest path algorithm	Bi-criteria parametric shortest path algorithm
-Working path set	OD-based: P_{rs}^{+}; Ordered by path tolls	OD- and class-based: P_{rsk}^{+}
Decomposition		
-Decomposition scheme	Gauss-Seidel method; Single-dimensional subproblem (3.25) with respect to b_p^{rs}	Gauss-Seidel method; Single-dimensional subproblem (3.35) with respect to f_p^{rsk}
Equilibration		
-Flow updating method	Projected Newton's method; Boundary updating equation (3.28)	Projected Newton's method; class-specific path flow updating equation (3.38)
Convergence measure		
-Relative gap	Eq. (3.31)	Eq. (3.39)
-Inner gap	Eq. (3.32) for path set P_{rs}^{+}	Eq. (3.40) for path set P_{rsk}^{+}

3.4 Numerical results

In this section, we utilize a small network and several networks of different scales to evaluate the models and solution algorithms. All algorithms are coded in Python 3.11.4, and the implementation shares code wherever possible to ensure a fair comparison of computational performance. The experiments were conducted on a PC with an Intel(R) Core(TM) i7-10700 CPU running at 2.9 GHz and equipped with 16 GB of RAM, operating on Windows 10.

The link performance function adopts the well-known Bureau of Public Roads (BPR) function:

$$t_a(x_a) = FFTT_a \left(1 + 0.15 \left(\frac{x_a}{Cap_a}\right)^4\right) \tag{3.41}$$

where $FFTT_a$ is the free-flow travel time of link a, and Cap_a denotes the capacity of link a. The other algorithmic parameters shown in Fig. 3.5 are listed in Table 3.2.

Table 3.2: Algorithmic parameters.

Parameter	Value	Description
$maxInnerloop$	1 for Sec. 3.4.1; 100 for Sec. 3.4.2	
$minBadOD$	$\|RS\|/3000$	
γ	0.5	Inner gap tolerance
ϵ	10^{-10}	Relative gap tolerance
I	30	Maximum number of iterations

3.4.1 Small network

The small network is shown in Fig. 3.7. In this network, two out of seven links are subject to tolls. As a result, the three available paths consist of a non-toll path, a low-toll path, and a high-toll path. As this network only has one OD pair with a demand of 100, the superscript/subscript rs is omitted for simplicity.

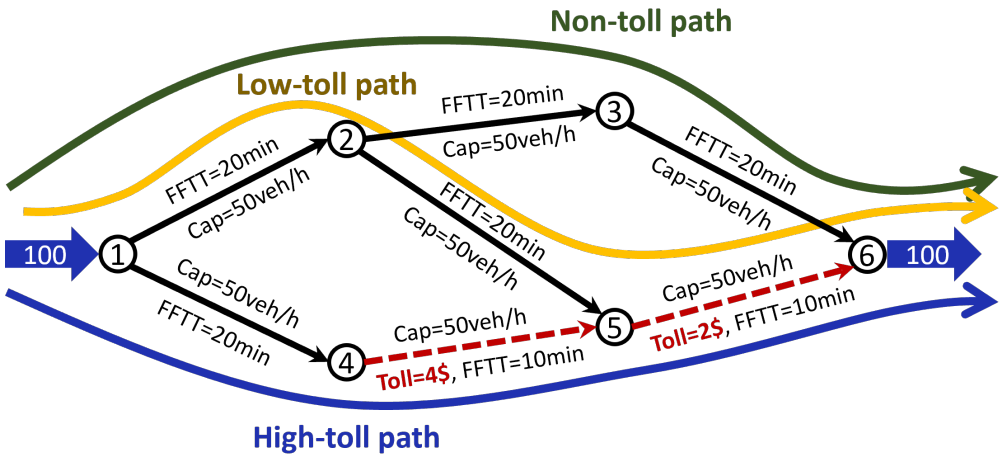


Figure 3.7: A small network reproduced from Slavin et al. (2014).

We consider a uniformly distributed VOT $\alpha \sim U[5, 30]$. Correspondingly, the VOT profile function is:

$$H(u) = \frac{u-5}{30-5} \Rightarrow \mathcal{H}(u) = 100\frac{u-5}{30-5} \Rightarrow g(v) = 0.25v + 5 \tag{3.42}$$

In the following, we first solve the two models to illustrate the details of the path-based GP algorithm. Notably, the DMBTA problem can be considered an approximation of the CMBTA problem. Hence, we also compare the path flow results from both models.

3.4.1.1 Solution of the CMBTA model

Given an initial feasible solution $\mathbf{b} = (10, 20)$, where the corresponding path flow is $\mathbf{f} = (10, 10, 80)$, we solve the CMBTA model. As shown in Fig. 3.8a, the path-based gradient projection (GP) algorithm converges in just 4 iterations, demonstrating its efficiency. The iteration process in the (b_1, b_2) space exhibits a ladder-like pattern. This behavior occurs because the CMBTA model (3.10) is decomposed into a series of one-dimensional subproblems (3.25) based on the Gauss-Seidel method, where each update involves only one variable. Additionally, during the first iteration, updating b_1 using Newton's method yields $\mathbf{b} = (27.55, 20)$, which falls outside the feasible region (i.e., where $b_1 \leq b_2$). To ensure feasibility, the projection operator projects b_1 to the boundary of the feasible set ($b_1 = b_2$).

Fig. 3.8b and Fig. 3.8c illustrate the Pareto frontier, the VOT profile function, and the boundaries and path flows at equilibrium, enabling us to verify the CMBTA-UE condition as defined in Definition 3.2. For instance, in Fig. 3.8b, the slope between the non-toll path and the low-toll path is calculated as

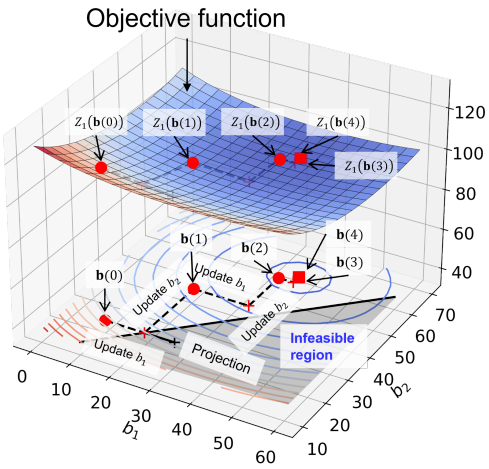
$$\frac{m_1 - m_2}{c_2 - c_1} = \frac{0 - 2}{0.9771 - 1.1155} = 14.45 = 0.25 \times 37.80 + 5 = g(b_1), \tag{3.43}$$

which matches the value of VOT profile function $g(\cdot)$ at boundary b_1 in Fig. 3.8c. This consistency demonstrates the validity of the CMBTA model and confirms that the algorithm successfully achieves the CMBTA-UE condition.

3.4.1.2 Solution of the DMBTA model

We also show the results of the DMBTA model. To align with the continuous case, we divide the support $[5, 30]$ evenly into 5 classes and calculate the class-specific VOTs and demands. The class-specific VOTs are $[7.5, 12.5, 17.5, 22.5, 27.5]$ and the class-specific demand are $[20, 20, 20, 20, 20]$. It's important to note that since the DMBTA model (3.23) involves more variables, visualizing the results can be challenging. Therefore, we will only present the iteration details in Table 3.3 and the final results in Table 3.4.

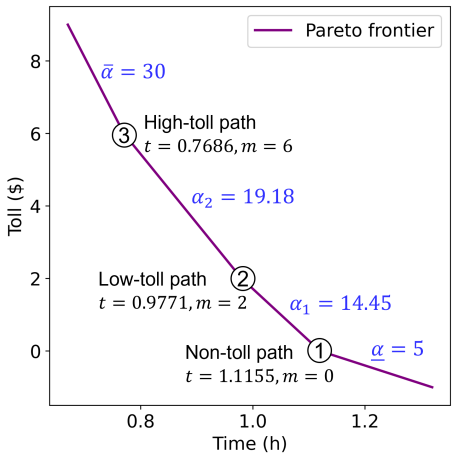
Table 3.3 demonstrates the efficiency of the path-based GP algorithm, which converges in only seven iterations. At iteration 0, the all-or-nothing assignment generates a class-specific path for each class, forming an initial working path set P^+ with five paths. In subsequent iterations,



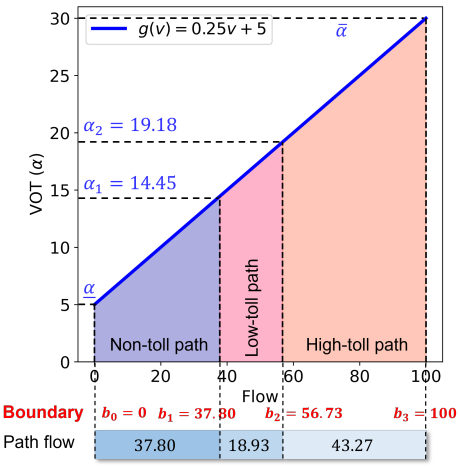
Iteration details

Iteration	$\mathbf{b} = (b_1, b_2)$	Obj. Val.	RG
0	(10.00, 20.00)	112.93	3.87E-01
Update b_1 Projection	(27.55, 20.00) (20.00, 20.00)		
1	(20.00, 40.77)	101.04	1.41E-01
Update b_1	(34.02, 40.77)		
2	(34.02, 54.25)	97.83	4.51E-03
Update b_1	(37.69, 54.25)		
3	(37.69, 56.72)	97.72	2.28E-06
Update b_1	(37.80, 56.72)		
4	(37.80, 56.73)	97.72	1.75E-11

(a) Iteration details of the path-based GP algorithm for solving the CMBTA model given initial solution $(b_1, b_2) = (10, 20)$.



(b) Pareto frontier of paths at the equilibrium.



(c) VOT profile function and equilibrium flow.

Figure 3.8: Solution of the CMBTA model for the small network.

the column generation process introduces additional class-specific paths. After each equilibration process, a dropping operation removes zero-flow paths. As a result, the size of the working path set initially increases but later decreases as redundant paths are eliminated, ultimately stabilizing at six class-specific paths, corresponding to the positive class-specific path flow variables reported in Table 3.4. Notably, the DMBTA model theoretically contains $3 \times 5 = 15$ class-specific path flow variables, while only six are needed to achieve equilibrium.

Table 3.4 reveals a key feature of the DMBTA model: it maintains multiple copies of variables to capture the differentiated choices of travelers from different classes. For instance, class 1 travelers select the non-toll path, while class 4 travelers choose the high-toll path. However,

Table 3.3: Iteration details of the path-based GP algorithm for solving the DMBTA model.

Iteration	$\lvert P^+\rvert$	Obj. Val.	RG
0	5	104.09	1.51E-01
1	8	101.31	8.88E-02
2	7	98.32	7.02E-03
3	7	98.06	4.64E-03
4	6	97.82	1.59E-03
5	6	97.75	6.63E-05
6	6	97.75	1.50E-08
7	6	97.75	7.95E-16

Table 3.4: Path flow pattern of the DMBTA model. There are in total $3 \times 5 = 15$ class-specific path flow variables. Among them, 6 are positive.

Path	Class 1	Class 2	Class 3	Class 4	Class 5	Total flow
Non-toll path	20	20	0	0	0	40
Low-toll path	0	0	18.79	0	0	18.79
High-toll path	0	0	1.21	20	20	41.21

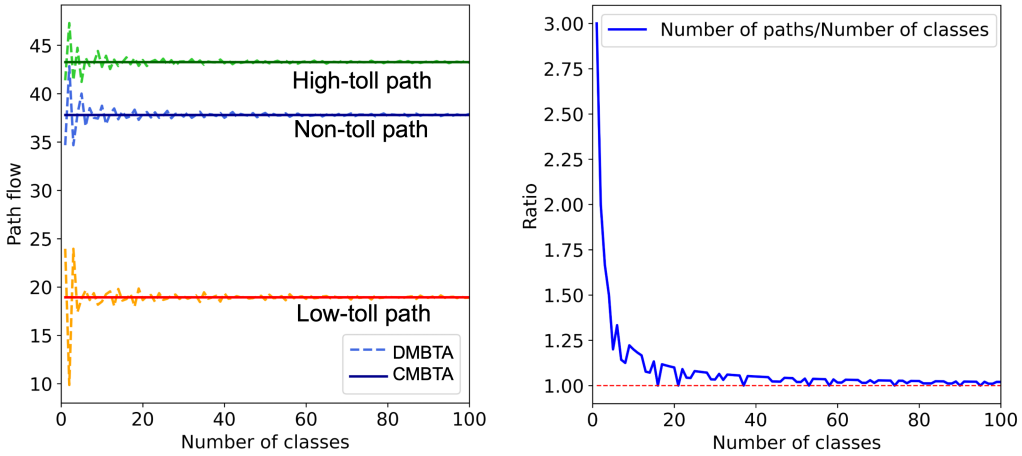
class 3 travelers display varied choices, with some choosing the low-toll path and others selecting the high-toll path. This is somewhat inconsistent with the modeling target of multiclass, where each class is expected to make the same decision. In contrast, the CMBTA model ensures that each infinitesimal class of travelers selects a single path, using only one set of boundary variables to differentiate the choices of different travelers seamlessly.

3.4.1.3 CMBTA versus DMBTA: path flow pattern

We compare the path flow patterns of the CMBTA and DMBTA models. Using the CMBTA solution as a benchmark, we increase the number of classes in the DMBTA model from 1 to 100. The results are presented in Fig. 3.9.

Fig. 3.9(a) reveals two key observations: (i) when the number of classes in the DMBTA model is small (e.g., $|K| \leq 10$), the path flow pattern exhibits significant fluctuations due to discretization errors; (ii) as the number of classes increases, the discretization error diminishes, and the DMBTA model's path flow pattern converges toward that of the CMBTA model.

Fig. 3.9(b) illustrates the ratio of the number of paths $|P^+|$ to the number of classes $|K|$ in the DMBTA model. When the number of classes is small, the path-to-class ratio is larger than 1, indicating that travelers within the same class select multiple paths, e.g., class 3 selects two paths in Table 3.4. However, as the number of classes increases, the path-to-class ratio



(a) Equilibrium path flow of the two models

(b) Path-to-class ratio $\frac{|P^+|}{|K|}$

Figure 3.9: Path flow pattern and path-to-class ratio with the increase of the number of classes.

approaches 1. This trend reflects specifying more classes will lead to more consistent choices within each class. Suppose that we can discretize travelers into infinite classes, each class would select a single path that minimizes its generalized cost, aligning with the equilibrium condition for the CMBTA model as defined in Definition 3.2.

3.4.2 Networks of different scales

In this subsection, we evaluate the computational performance of the path-based GP algorithm for solving the CMBTA and DMBTA models across networks of varying scales, as presented in Table 3.5. Since the original network instances lack tolls, we assign a toll equal to 0.1 times the free-flow travel time on the 10% most congested links under the UE flow pattern, with tolls set to 0 on all other links. Similar to the small network analysis, we assume a uniformly distributed VOT $\alpha \sim U[5, 30]$. For the DMBTA model, the VOT support $[5, 30]$ is evenly divided into $|K|$ classes to determine class-specific VOTs and demands.

Table 3.5: Networks of different scales. The detailed network data is available at https://github.com/bstabler/TransportationNetworks.

Network	# of zones	# of nodes	# of links	# of tolled links	# of OD pairs
Sioux Falls	24	24	76	8	528
Anaheim	38	416	914	91	1,406
Winnipeg	147	1,052	2,836	284	4,344
Chicago Sketch	387	933	2,950	295	93,135

3.4.2.1 Results

Fig. 3.10 illustrates the convergence curves of the path-based GP algorithm for solving the CMBTA and DMBTA models with varying numbers of classes. For the CMBTA model, the path-based GP algorithm consistently achieves high-precision equilibrium solutions (RG < 1e-10) within 30 iterations across all cases. In contrast, for the DMBTA model in the Chicago Sketch network, the path-based GP algorithm fails to reach this precision level (RG < 1e-10) within 30 iterations when $|K| \geq 10$. But still, it achieves a satisfactory precision level of RG < 1e-8.

For the DMBTA model, it is expected that CPU time increases with the number of classes. Notably, across all networks, the CPU time required to solve the CMBTA model is comparable to that of the DMBTA model with 3 to 5 classes. This is to say, spending a relatively short CPU time, the CMBTA model can produce a "perfectly" refined solution that the DMBTA model can hardly achieve.

Additional computational results are presented in Table 3.6. As expected, in the DMBTA model, the number of working paths $|P^+|$ increases with the number of classes. In contrast, the CMBTA model maintains a compact working path set, consistently smaller than that of the DMBTA model with 3 classes across all networks. This compactness explains why the CPU time for the CMBTA model is consistently comparable to that of the DMBTA model with 3–5 classes, as the path-based GP algorithm iterates over the entire working path set to update variables. This feature of the CMBTA model is particularly advantageous for real-world applications involving large-scale transportation networks, where storing numerous paths could significantly strain CPU time and memory resources.

Notably, the DMBTA model may retain duplicate paths across different classes. Therefore, we also examined the number of unique paths, which also exhibits an increasing trend with the number of classes, as shown in Table 3.6. This is expected: the more the number of classes, the more the efficient paths on the Pareto frontiers are likely to be used (see Fig. 3.4). Additionally, the CMBTA model retains a relatively high number of unique paths. For example, in the Chicago sketch network, the number of unique paths in the CMBTA model exceeds that of the DMBTA model with $|K| = 50$. This observation aligns with Fig. 3.3(a), which shows that all efficient paths within the range $[\underline{\alpha}, \overline{\alpha}]$ on the Pareto frontier may be used, whereas some paths remain unused in the DMBTA model, as depicted in Fig. 3.4. In summary, the GP algorithm generates

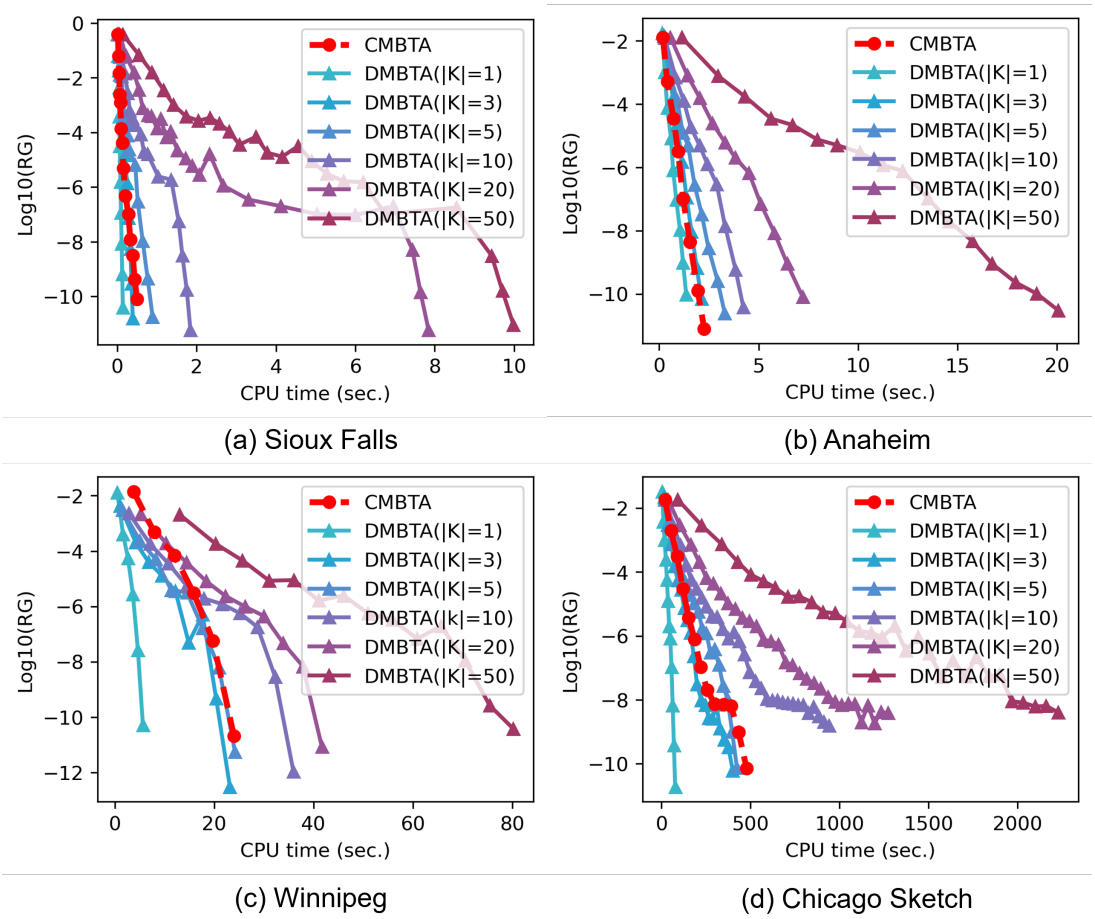


Figure 3.10: The convergence curves of the path-based GP algorithms for solving the CMBTA model and DMBTA model.

numerous "redundant" class-specific paths when solving the DMBTA model, while producing a compact yet diverse working path set for the CMBTA model.

Using the equilibrium link flow of the CMBTA model as the benchmark, we report the mean squared error (MSE) of the DMBTA model's link flows in the last column of Table 3.6. As expected, the MSE decreases as the number of user classes increases. Notably, across all networks, the MSE falls below 30 when $|K| = 50$, indicating that the link flow pattern of the DMBTA model with $|K| = 50$ can serve as a satisfactory approximation of the CMBTA model's link flow pattern.

3.4.2.2 Impact of inner loops

The inner loop is an effective strategy for reducing CPU time (Xie et al., 2018; Galligari and Sciandrone, 2019; Xu et al., 2020; Tan et al., 2022). This strategy is particularly beneficial for

Table 3.6: The computational results in the networks of different scales.

Network	Model	$\|P^+\|$	# of unique paths	CPU time (sec.)	MSE of Link flow
Sioux Falls	CMBTA	722	722	0.50	0.0
	DMBTA (|K|=1)	638	638	0.14	23,634.7
	DMBTA (|K|=3)	1,770	697	0.39	5,610.3
	DMBTA (|K|=5)	2,900	709	0.89	9,592.2
	DMBTA (|K|=10)	5,810	738	1.84	1,324.9
	DMBTA (|K|=20)	11,737	731	7.84	97.5
	DMBTA (|K|=50)	28,609	749	9.98	21.1
Anaheim	CMBTA	2,353	2,353	2.27	0.0
	DMBTA (|K|=1)	1,540	1,540	1.36	38,238.4
	DMBTA (|K|=3)	4,426	1,966	2.13	5,745.2
	DMBTA (|K|=5)	7,424	2,239	3.30	1,276.5
	DMBTA (|K|=10)	14,452	2,344	4.23	646.5
	DMBTA (|K|=20)	28,458	2,375	7.22	158.8
	DMBTA (|K|=50)	72,005	2,529	20.08	26.6
Winnipeg	CMBTA	11,006	11,006	23.98	0.0
	DMBTA (|K|=1)	4,524	4,524	5.67	9,471.0
	DMBTA (|K|=3)	13,308	7,336	23.19	769.8
	DMBTA (|K|=5)	21,914	8,342	24.22	296.6
	DMBTA (|K|=10)	43,710	9,284	35.97	69.3
	DMBTA (|K|=20)	87,642	10,419	41.69	21.9
	DMBTA (|K|=50)	218,232	11,165	80.25	4.3
Chicago Sketch	CMBTA	247,555	247,555	478.61	0.0
	DMBTA (|K|=1)	96,639	96,639	77.41	41,012.6
	DMBTA (|K|=3)	284,125	157,155	400.88	5,913.2
	DMBTA (|K|=5)	469,555	178,297	424.19	1,491.7
	DMBTA (|K|=10)	988,056	213,704	941.48	413.0
	DMBTA (|K|=20)	1,987,777	234,259	1,273.67	92.6
	DMBTA (|K|=50)	4,938,920	245,393	2,229.20	12.6

the MBTA problem, which requires executing a bi-criteria shortest path algorithm for each origin during the column generation process. To assess the impact of the inner loop, we evaluated the performance of the path-based GP algorithm for solving the CMBTA and DMBTA models under different inner loop settings in the Winnipeg network.. The results are presented in Fig. 3.11.

As illustrated in Fig. 3.11, when only one equilibration process is executed per iteration (i.e., the maximum number of inner loops is 1), the column generation process consumes a significant portion of the computational time for both problems. In contrast, when the maximum number of inner loops is set to 100, the path-based GP algorithm allocates more time to the equilibration process in each iteration, significantly accelerating convergence for both models. Notably, the CMBTA model exhibits greater acceleration with the inner loop, converging in only seven iterations compared to the DMBTA model.

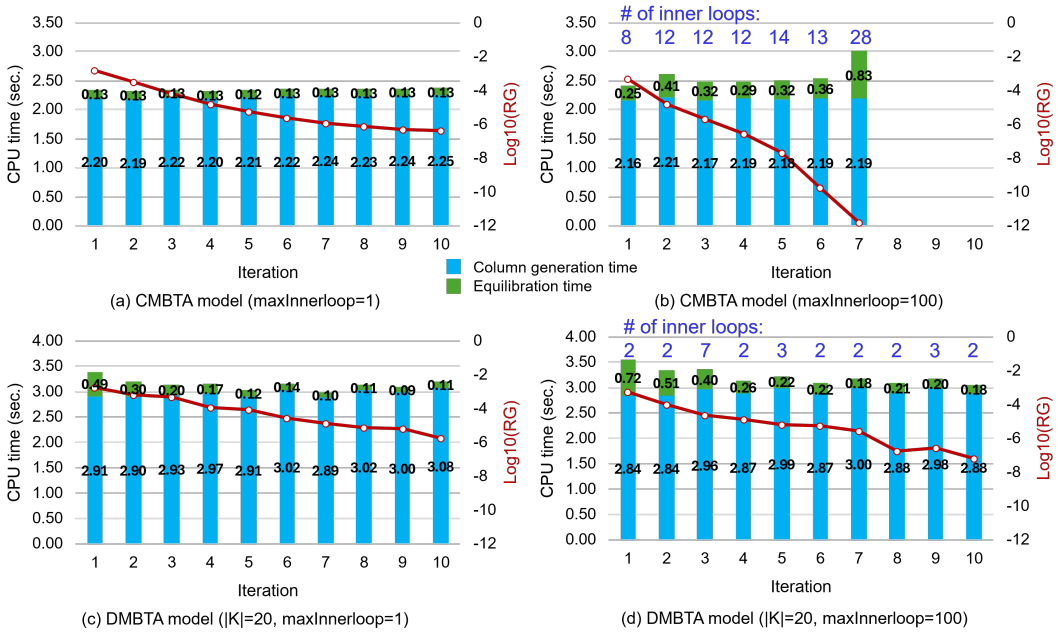


Figure 3.11: CPU times of different steps in the path-based GP algorithm.

3.4.2.3 Comparison with existing methods

Finally, we compare the proposed unified GP algorithm with existing methods, including the MSA and the FW algorithm. It is worth noting that the MSA was first applied to the CMBTA problem by Leurent (1993), and its extension to the discrete version (DMBTA) is straightforward. The FW algorithm has also been adapted to solve both the CMBTA problem (Marcotte and Zhu, 1996) and the DMBTA problem (Yang and Huang, 2004). Detailed algorithmic procedures are provided in A.1 and A.2, respectively. The maximum number of iterations for MSA and FW is set to 300. A comparative summary of the results is presented in Fig. 3.12.

The results demonstrate that across all tested networks and for both models, the GP algorithm achieves the highest performance, followed by the FW algorithm, while the MSA performs the least effectively. This outcome is consistent with expectations, as the GP algorithm is well-established for its rapid convergence and ability to attain high solution accuracy (Xie et al., 2018; Galligari and Sciandrone, 2019; Tan et al., 2022; Xu et al., 2024a). In contrast, the two link-based algorithms, FW and MSA, exhibit limitations in reaching comparable equilibrium precision, with FW converging to an approximate accuracy of 10^{-6} and MSA only reaching about 10^{-4}. The notably strong performance of the GP algorithm further confirms its suitability for addressing real-world instances of both the continuous and discrete MBTA problems.

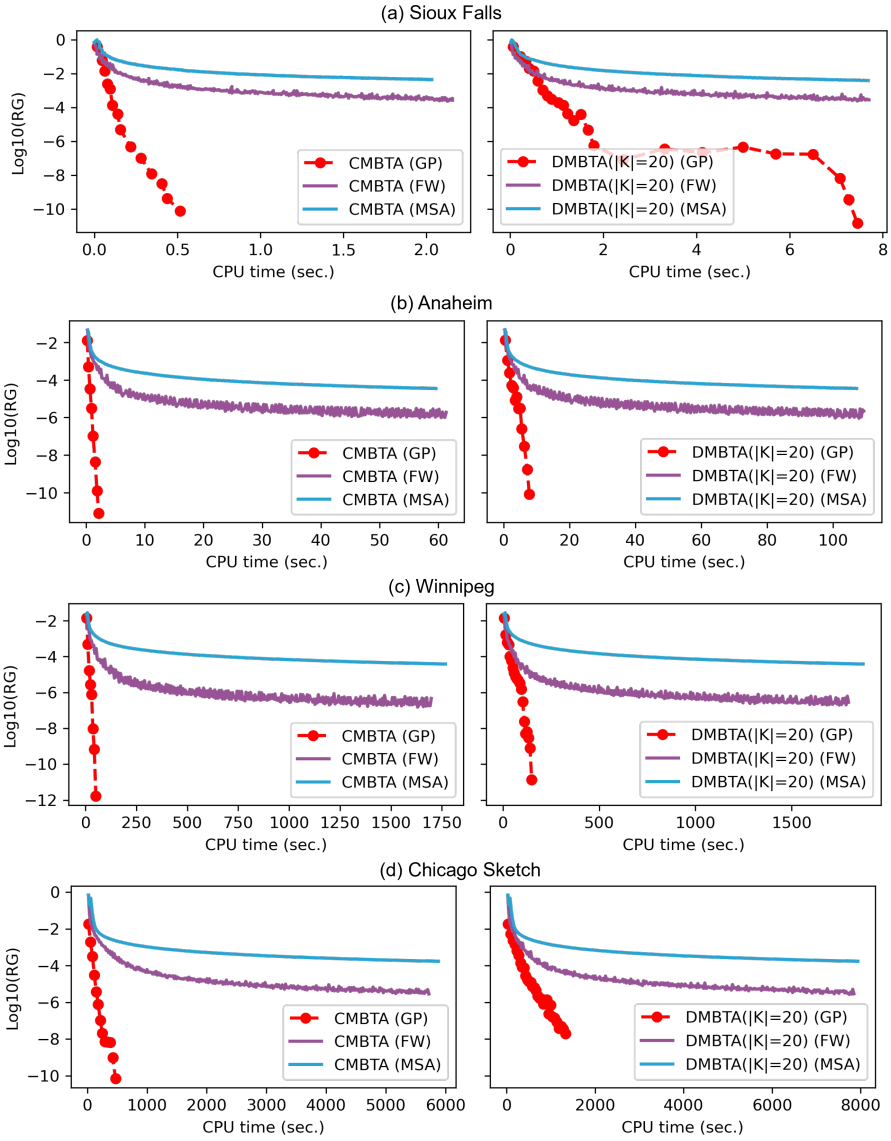


Figure 3.12: Convergence curves of the different methods.

3.5 Conclusions

This study provides a unified path-based GP algorithm framework to address both CMBTA and DMBTA problems. The framework incorporates three tailored modules—column generation, decomposition and equilibration, and convergence criteria—designed for discrete and continuous cases. Numerical results demonstrate that the path-based GP algorithm can successfully solve both problems efficiently. Notably, the equilibrium flow of the DMBTA model exhibits fluctuations when the number of classes is small but converges to the equilibrium flow of the CMBTA model as the number of classes increases. Furthermore, the CMBTA model requires

significantly less computation time and memory resources to achieve a highly precise solution, making it more suitable for real-world applications.

Future research could explore the following directions. First, the provided framework could be extended to multi-criteria traffic assignment by incorporating factors such as travel time reliability, enabling the study of how individuals balance time, toll, and reliability trade-offs. Second, the current framework assumes toll costs are flow-independent. Extending the path-based GP algorithm to address continuous multiclass problems with both travel time and toll costs being flow-dependent is of significant interest. Finally, the CMBTA and DMBTA models could be applied to investigate road congestion pricing, identifying first- or second-best tolls across transportation networks. Evaluating their modeling and computation performance differences in various applications would provide valuable insights for practical implementation.

Chapter 4

Discrete multiclass bi-criteria traffic assignment without class-specific variables: An alternative formulation and a subgradient projection algorithm

In this chapter, we focus on the discrete multiclass bi-criteria (time and toll) traffic assignment (DMBTA) problem. The conventional DMBTA model keeps multiple copies of class-specific variables to model user heterogeneity, which puts a great burden on memory storage and computational speed when dealing with real transportation networks. This chapter proposes an alternative formulation for the DMBTA problem without class-specific variables by exploiting the order information of paths and travelers, i.e., *high value of time (VOT) travelers will prefer fast but expensive paths, while low VOT travelers will prefer slow but cheap paths*. We prove the equivalence of the alternative formulation to the conventional DMBTA model. To solve the alternative formulation with a nondifferentiable convex objective function, a path-based subgradient projection algorithm is developed utilizing the subgradient and available second-order information. We adopt a small network and several large networks to examine the detailed features and the computational performance of the proposed formulation and algorithm, respectively. The results show that the alternative formulation provides the same link flow pattern as that of the conventional DMBTA model but uses much fewer variables, which can greatly relieve the burden on computer memory and computational speed in solving real transportation networks.

4.1 Introduction

As the final step of the 4-step transport planning model, the traffic assignment problem aims to distribute travelers on a transportation network based on predefined path choice assumptions. The resulting traffic assignment models are the fundamental tools for many transport applications, including network capacity assessment (Chen et al., 2002b; Liu et al., 2020; Du et al., 2022), road network design (Yang and H. Bell, 1998; Chen et al., 2011a; Fan et al., 2022; Wang et al., 2022b), multimodal network modeling (Du et al., 2022; Li et al., 2022), emission estimation (Chen et al., 2011b; Xu et al., 2015; Tidswell et al., 2021), etc. While the classic traffic assignment problem assumes that travelers choose the path with minimum travel time (Wardrop, 1952; Beckmann et al., 1956), the time-toll bi-criteria traffic assignment (BTA) problem models a more realistic path choice behavior that travelers will make a trade-off between time and toll. The BTA problem gains its special position in the field of transportation by not only providing a tool to examine the famous marginal cost theory that regulates user-equilibrium (UE) flow to system-optimal (SO) flow but also serving a lot of revenue-related real-world applications such as build-operator-transfer (BOT) project evaluation (Yang and Meng, 2000; Chen and Subprasom, 2007; Chen et al., 2021) and congestion pricing (Yang and Huang, 2005; Han and Yang, 2008; Meng et al., 2012; Ohazulike et al., 2013; Han et al., 2019).

Dial (1979) is one of the first to model the travelers' path choice behavior facing time and toll trade-off, where the travel time and toll of a path are linearly combined with a conversion factor called the value of time (VOT) to form a generalized cost[1]. VOT reflects travelers' trade-off between travel time and toll, which is influenced by travelers' socio-economic status, trip purposes, inherent preference, etc. It is generally believed that VOT is a traveler-specific value that is randomly distributed among the population (Small and Yan, 2001; Small et al., 2005; Brownstone and Small, 2005; Cirillo and Axhausen, 2006). In practice, since VOT is an abstract factor that cannot be observed directly, traffic engineers usually create histograms of travelers' social-economic status, e.g., income, to infer an empirical distribution of VOT, as exemplified in Fig. 4.1.

From the modeling side, two research lines have developed in the literature to account for

[1]It should be mentioned that there are also some studies that consider nonlinear (or nonadditive) trade-off between time and toll. Interested readers may refer to Gabriel and Bernstein (1997), Larsson et al. (2002), Chen et al. (2010), Wang and Ehrgott (2013), and Xu et al. (2023b).

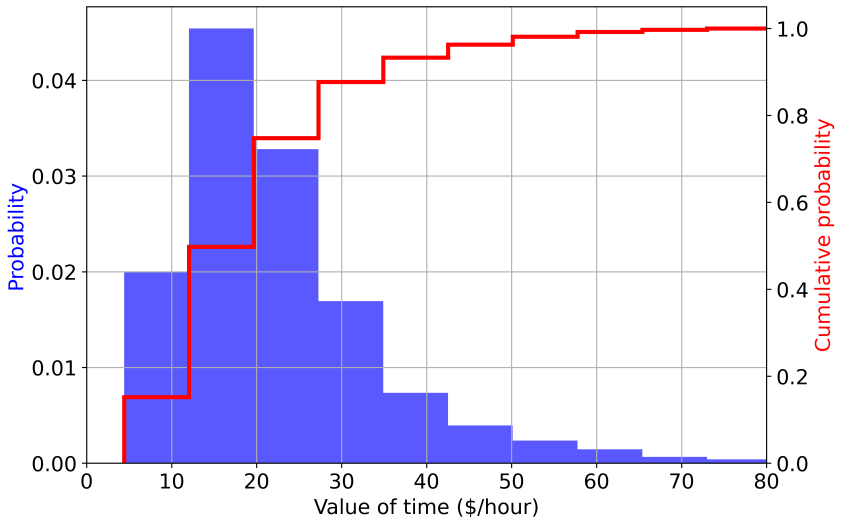


Figure 4.1: An empirical distribution of the value of time (VOT)

the user heterogeneity of VOT in the BTA problem. The first one assumes the VOT follows a continuous distribution (Dial, 1979, 1996, 1997; Leurent, 1993; Marcotte, 1999; Wu and Huang, 2014a; Xie et al., 2021; Xu et al., 2023b), while the second one discretizes travelers into multiple user classes (Dafermos, 1981; Nagurney, 2000; Nagurney and Dong, 2002a; Yang and Huang, 2004; Ryu et al., 2016, 2021). Although assuming a continuous distribution is theoretically more appealing, due to the simplicity of collecting data, formulating models, and developing solution algorithms, the second line seems more applicable in transport planning practice, which has been implemented in most of the mainstream transport planning software (e.g., CUBE[2], EMME[3], TransCAD[4], PTV[5]).

This chapter falls into the scope of the second line, i.e., assuming the VOT follows a discrete distribution. Along this line, Dafermos (1981) first suggested discretizing the whole population into a finite number of user classes with class-specific VOTs. From the formulation's point of view, this method creates multiple copies of class-specific link flows and class-specific path flows to model the path choices of travelers from different user classes. The resulting multiclass BTA (MBTA) model is a direct extension of the single-class traffic assignment models (Sheffi, 1985), which can be readily solved by existing traffic assignment algorithms without too many additional methodological efforts. For example, Yang and Huang (2004) investigated the multi-

[2] https://www.bentley.com/software/cube/
[3] https://www.inrosoftware.com/en/products/emme/
[4] https://www.caliper.com/tcovu.htm
[5] https://company.ptvgroup.com/en/

class bi-criteria SO problem in a tolled network and suggested using the link-based Frank-Wolfe algorithm (Frank and Wolfe, 1956) to solve the problem. Chen et al. (2011c) presented a path-based algorithm to solve the reliability-based MBTA problem. However, class-specific variables are both a blessing and a curse. The conventional MBTA model needs to keep a massive number of class-specific link/path flow variables, which puts a great burden on the computer memory and computational speed when solving large transportation networks. To partially relieve the memory concern, Ryu et al. (2016) suggested creating a universal path set for the path-based algorithms to reduce memory consumption due to multiple user classes. Yet, a huge incidence matrix is still needed to keep the class-path incidence relationship. For the speed problem, the computation time of the conventional MBTA model is approximately proportional to the number of user classes (Ye, 2010) as there are more copies of variables to optimize, but no relief has been proposed until now.

Since the aforementioned computational problems (i.e., memory storage and computational speed) of the conventional MBTA model originate from class-specific link/path flow variables, can we drop the class-specific variables but still model the MBTA problem? This chapter addresses this problem by proposing an alternative formulation for the MBTA problem without class-specific variables. Further, to solve the alternative formulation with a nondifferentiable convex objective function, we develop an efficient path-based subgradient projection (SGP) algorithm by utilizing the subgradient and available second-order information. The main contributions of this chapter are:

- Propose an alternative formulation for the MBTA problem without class-specific variables. The alternative formulation exploits the order information of paths and travelers, i.e., *high VOT travelers will prefer fast but expensive paths, while low VOT travelers will prefer slow but cheap paths*, to pre-align travelers with their preferable path choices, which allows us to use boundary variables (cutting each origin-destination (OD) demand into path flows) instead of class-specific link/path flow variables to model the MBTA problem.

- Develop a path-based subgradient projection (SGP) algorithm to solve the alternative formulation with a nondifferentiable convex objective function. The path-based SGP algorithm utilizes the subgradient and the available second-order information to activate the projected quasi-Newton's method to update the boundaries between adjacent path pairs

sequentially, which shows high efficiency and fast convergence in real networks.

- Test the proposed formulation and algorithm in a small network and several large networks. The results show that the alternative formulation provides the same link flow pattern as that of the conventional MBTA model but uses much fewer variables, which can greatly reduce the memory consumption and computation time of the MBTA problem. Notably, even without class-specific variables, the class-specific link/path information does not vanish. It is also possible to extract class-specific link/path flows from the boundary solution when class-specific information is needed.

The remainder of this chapter is organized as follows. Section 4.2 proposes an alternative formulation for the MBTA problem, analyzes its properties, and compares it with the conventional MBTA model. Section 4.3 presents a path-based SGP algorithm for solving the alternative MBTA formulation. Section 4.4 utilizes a small network and several large networks to demonstrate the features of the proposed alternative MBTA formulation and the path-based SGP algorithm. Finally, the conclusions and future research directions are discussed in Section 4.5.

4.2 Discrete multiclass bi-criteria traffic assignment problem

In this section, we first provide the notation list for ease of reading. Then, the conventional MBTA model and its characteristics are revisited. After that, an alternative formulation for the MBTA problem is provided followed by the solution properties. Finally, the two models are comprehensively compared.

4.2.1 Notation list

The notations used in this chapter are listed in Table 4.1.

4.2.2 Conventional discrete MBTA model

In a transport network with link tolls, travelers between each origin-destination (OD) pair rs may face several path choices with distinct travel times and tolls, where the final path choice is subject to a complex decision mechanism based on travelers' social-economic status, trip

Table 4.1: Notation list

Symbol	Description
Sets	
G	A transportation network, $G = (N, A)$.
N	Set of nodes indexed by n.
A	Set of directed links indexed by a.
R	Set of origins indexed by r, $R \subseteq N$.
S	Set of destinations indexed by s, $S \subseteq N$.
P	Set of paths indexed by p.
P_{rs}	Set of paths connecting origin-destination (OD) pair rs, $P_{rs} \subseteq P$.
K	Set of user classes indexed by k.
Inputs and parameters	
δ_{ap}^{rs}	$\delta_{ap}^{rs} = 1$ if link a is on path p connecting OD pair rs, 0 otherwise.
q_k^{rs}	Travel demand of user class k from r to s.
q^{rs}	Total travel demand from r to s, $q^{rs} = \sum_{k \in K} q_k^{rs}$.
τ_a	The toll on link a, $\tau_a \geq 0$, unit: \$.
m_p^{rs}	The toll on path p connecting OD pair rs, $m_p^{rs} = \sum_{a \in A} \tau_a \delta_{ap}^{rs}$.
α_k^{rs}	The value of time (VOT) of user class k from r to s.
β_k^{rs}	The time equivalence of money (TEM) of user class k from r to s, $\beta_k^{rs} = 1/\alpha_k^{rs}$.
Intermediate variables	
x_a	Traffic flow on link a.
x_a^k	Traffic flow of user class k on link a.
f_p^{rs}	Traffic flow on path p connecting OD pair rs.
$t_a(x_a)$	Travel time function on link a, which is a positive, continuously differentiable, and strictly increasing function with respect to its own flow x_a.
c_p^{rs}	Travel time on path p connecting OD pair rs, $c_p^{rs} = \sum_{a \in A} t_a(x_a) \delta_{ap}^{rs}$.
$d_p^{rs}(\beta)$	Generalized cost on path p connecting OD pair rs given TEM β.
Decision variables	
f_p^{rsk}	Traffic flow of user class k on path p connecting OD pair rs.
b_p^{rs}	Accumulated flow of the first p paths connecting OD pair rs, $b_p^{rs} = \sum_{i=1,\ldots,p} f_i^{rs}$, which is also the flow *boundary* between path p and path $p+1$. See Fig. 4.4(c) for illustration.

purpose, characteristics, etc. To model travelers' path choice behavior under such a condition, Dial (1979, 1996, 1997) simplified the problem by assuming that any rational traveler with value to time (VOT) α will choose the path with the minimum *generalized money cost*, which is defined as a linear combination of time and toll:

$$d_p^{rs}(\alpha)_{money} = \alpha c_p^{rs} + m_p^{rs}, \quad \forall p \in P_{rs} \tag{4.1}$$

where $d_p^{rs}(\alpha)_{money}$ denotes the generalized money cost of path p connecting OD pair rs; c_p^{rs} is the travel time on path p connecting OD pair rs; m_p^{rs} is the toll on path p connecting OD pair rs.

Noteworthy, it is a more common practice in the literature to form an equivalently[6] *generalized time cost* by associating the time equivalence of money (TEM) $\beta = \frac{1}{\alpha}$ to path toll:

$$d_p^{rs}(\beta)_{time} = c_p^{rs} + \beta m_p^{rs}, \quad \forall p \in P_{rs} \tag{4.2}$$

This chapter also complies with this convention. For notation simplicity, we call Eq. (4.2) as generalized cost and omit the subscript $time$ hereafter.

To model the user heterogeneity in the time-toll tradeoff, the most common method in academia (Dafermos, 1981; Nagurney, 2000; Nagurney and Dong, 2002a; Yang and Huang, 2004; Ryu et al., 2016, e.g.,) and practice (De Cea et al., 2005; Ye, 2010; Slavin et al., 2014) is to segment the travel demand between each OD pair rs into a finite number of classes with class-specific demands $\{q_k^{rs}, k \in K\}$ and class-specific TEMs $\{\beta_k^{rs}, k \in K\}$. Each traveler of class $k \in K$ is assumed to choose a path that can minimize generalized cost $d_p^{rs}(\beta_k^{rs})$. The corresponding multiclass bi-criteria traffic equilibrium (MBTE) problem has an equivalently mathematical programming (MP) model (Yang and Huang, 2004):

$$\min Z_1(\mathbf{f}) = \underbrace{\sum_{a \in A} \int_0^{x_a} t_a(v)dv}_{\text{Beckmann transformation}} + \underbrace{\sum_{rs \in RS} \sum_{p \in P_{rs}} \sum_{k \in K} m_p^{rs} \beta_k^{rs} f_p^{rsk}}_{\text{Multiclass external cost}} \tag{4.3a}$$

$$q_k^{rs} = \sum_{p \in P_{rs}} f_p^{rsk}, \quad \forall rs \in RS, k \in K \tag{4.3b}$$

$$x_a^k = \sum_{rs \in RS} \sum_{p \in P_{rs}} f_p^{rsk} \delta_{ap}^{rs}, \quad \forall a \in A, k \in K \tag{4.3c}$$

$$x_a = \sum_{k \in K} x_a^k, \quad \forall a \in A \tag{4.3d}$$

$$f_p^{rsk} \geq 0, \quad \forall rs \in RS, p \in P_{rs}, k \in K \tag{4.3e}$$

The objective function in Eq. (4.3a) consists of two terms. The first term is the well-known Beckmann transformation (Beckmann et al., 1956) to model the congestion nature of the transportation network. The second term is the multiclass external cost to model the tolling effect

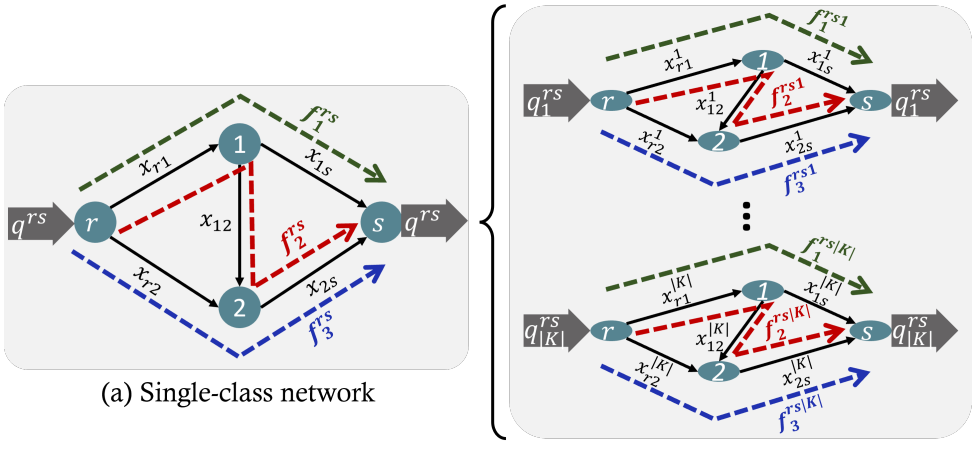
[6]The equivalence here is from the perspective of the network equilibrium solution.

on travelers from different classes. Eq. (4.3b) is the flow conservation constraint for each user class. Eq. (4.3c) converts class-specific path flows into class-specific link flows. Eq. (4.3d) merges class-specific link flows into total link flows. Eq. (4.3e) is the nonnegativity constraint of class-specific path flows. For the solution property of the conventional MBTA model, it is known that only the total link flow $\mathbf{x} = \{x_a^*, a \in A\}$ is unique, while all other variables are not in general networks (Dafermos, 1981; Nagurney, 2000; Nagurney and Dong, 2002a). Furthermore, it is easy to check the Karush–Kuhn–Tucker (KKT) conditions of the above convex programming model, resulting in the following equilibrium condition:

Condition 4.1 (MBTE)**.** *For any OD pair rs and user class k, if path p's class-specific flow is positive, i.e., $f_p^{rsk} > 0$, then*

$$c_p^{rs} + \beta_k^{rs} m_p^{rs} \leq c_{p'}^{rs} + \beta_k^{rs} m_{p'}^{rs}, \quad \forall p' \in P_{rs} \tag{4.4}$$

To handle the user heterogeneity in terms of time-toll tradeoff, the conventional MBTA model (4.3) adopts the strategy of *increasing variable dimension*. To be more specific, the conventional MBTA model has to keep $|K|$ copies of class-specific link/path flow variables to record each user class's link/path flow pattern, as shown in Fig. 4.2. Especially, real transportation networks usually have numerous OD pairs and a formidable number of paths. Thus, each extra user class will require a considerable amount of computer memory and longer computation time as there are much more variables to store and solve.



(a) Single-class network

(b) Multiclass network

Figure 4.2: Illustration of class-specific variables in the conventional MBTA model.

4.2.3 An alternative formulation for the discrete MBTA problem without class-specific variables

4.2.3.1 Model formulation

Alternatively, it is also possible to develop an MBTA model without class-specific variables, which can greatly relieve the MBTA problem's computation burden. The key is to exploit the *order information* of paths and travelers to pre-aline travelers with their preferable path choices. Without loss of generality, for any OD pair rs, we assume that the paths are (weakly) decreasingly ordered by toll.

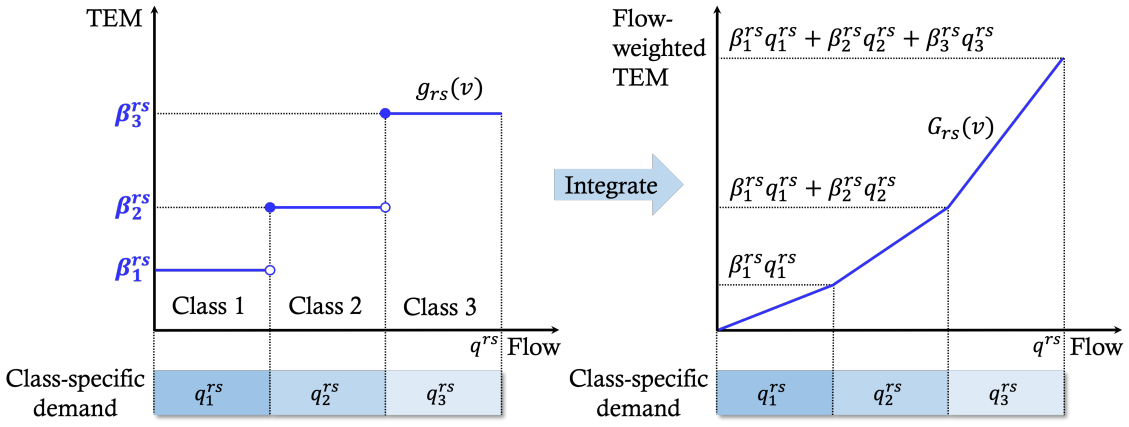
$$\text{(Expensive to cheap)} \quad m_1^{rs} \geq m_2^{rs} \geq ... \geq m_{|P_{rs}|}^{rs}, \quad \forall rs \in RS \tag{4.5}$$

Further, given the class-specific demands $\{q_k^{rs}, k \in K\}$ and class-specific TEMs $\{\beta_k^{rs}, k \in K\}$, we can obtain an (weakly) increasing step function $g_{rs}(v)$, which depicts the *TEM profile* (y-axis) among travelers (x-axis). See Fig. 4.3(a) for a 3-class example.

$$\text{(“Rich” to “poor”)} \quad g_{rs}(v) = \begin{cases} \beta_1^{rs}, & \text{if } v \in [0, q_1^{rs}) \\ \beta_2^{rs}, & \text{if } v \in [q_1^{rs}, q_1^{rs} + q_2^{rs}) \\ ... & \\ \beta_k^{rs}, & \text{if } v \in \left[\sum_{i=1}^{k-1} q_i^{rs}, \sum_{i=1}^{k} q_i^{rs}\right) \\ ... & \\ \beta_{|K|}^{rs}, & \text{if } v \in \left[\sum_{i=1}^{|K|-1} q_i^{rs}, \sum_{i=1}^{|K|} q_i^{rs}\right] \end{cases}, \quad \forall rs \in RS \tag{4.6}$$

where $\beta_1^{rs} \leq \beta_2^{rs} \leq ... \leq \beta_{|K|}^{rs}$ for all OD pairs $rs \in RS$. The essence of Eq. (4.6) is to order each OD pair's class-specific demands from “rich” (low TEM) to “poor” (high TEM). Note that this TEM profile function is a very general representation that can be obtained from any empirical/discrete VOT distribution; it can also be obtained by discretizing the continuous distributions if needed.

Now, we discuss travelers' path choice behaviors. Recall the assumption of the MBTA problem that each traveler considers to choose a path that can minimize his/her generalized cost, i.e., Eq. (4.2). In other words, travelers will only consider the *efficient paths* as defined in Dial



(a) ***Traveler TEM order***: the TEM profile function g_{rs}

(b) Flow-weighted TEM G_{rs}

Figure 4.3: Illustrations of the TEM profile function g_{rs} and flow-weighted TEM G_{rs}.

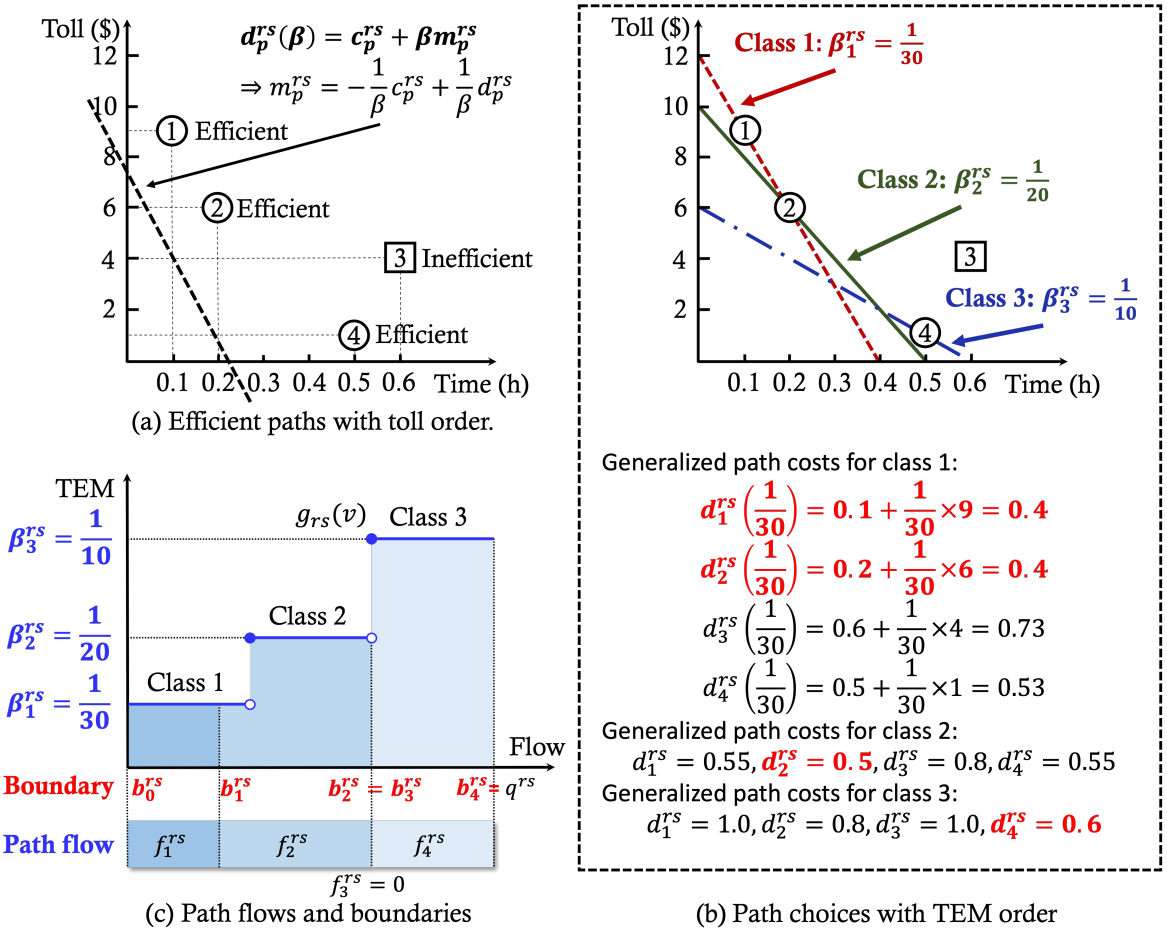
(1979).

Definition 4.1 (Efficient path)**.** *Given a set of paths P_{rs} with different times and tolls between OD pair $rs \in RS$, a path p is said to be efficient when it can minimize the generalized cost (4.2) given some $\beta > 0$, i.e., $p = \arg\min_{p' \in P_{rs}} d_{p'}^{rs}(\beta)$.*

Given a set of efficient paths, it is intuitive to observe that *low-TEM ("rich") travelers will prefer fast but expensive paths, while high-TEM ("poor") travelers will prefer slow but cheap paths*. Let's examine this through an example.

- **Efficient paths with toll order.** Fig. 4.4(a) shows a 4-path example that follows the path toll order requirement Eq. (4.5), i.e., $m_1^{rs} \geq m_2^{rs} \geq m_3^{rs} \geq m_4^{rs}$. Among these paths, only paths 1, 2, and 4 can minimize the generalized cost $d_p^{rs}(\beta)$ given some $\beta > 0$, hence efficient. By contrast, for path 3, there always exists another path with a lower generalized cost given any $\beta > 0$, hence inefficient.

- **Path choices with TEM order.** Consider the simple example depicted in Fig. 4.3(a) with three user classes. We further assume that $\beta_1^{rs} = \frac{1}{30} < \beta_2^{rs} = \frac{1}{20} < \beta_3^{rs} = \frac{1}{10}$, as shown in Fig. 4.4(b). The equations in Fig. 4.4(b) illustrate that for user class 1, both paths 1 and 2 can minimize the generalized cost ($d_1^{rs} = d_2^{rs} = 0.4$); for user class 2, only path 2 can minimize the generalized cost ($d_2^{rs} = 0.5$); for user class 3, only path 4 can minimize the generalized cost ($d_4^{rs} = 0.6$). That is, low-TEM travelers to high-TEM travelers (e.g.,



(a) Efficient paths with toll order.

(c) Path flows and boundaries

(b) Path choices with TEM order

Figure 4.4: Illustrations of efficient paths, path choices, path flows, and boundaries.

user classes 1, 2, and 3) will choose expensive efficient paths to cheap efficient paths (e.g., paths 1, 2, and 4) in sequence. No one chooses the inefficient path.

- **Path flows and boundaries.** Based on the path choice analysis in Fig. 4.4(b), a possible path flow pattern $(f_1^{rs}, f_2^{rs}, f_3^{rs}, f_4^{rs})$ is shown in Fig. 4.4(c). Here "possible" means we have not considered the congestion effect yet. Hence, it is sufficient to show that the essence of the multiclass bi-criteria path flow assignment is to determine the *boundaries* (or break points) that separate the TEM-ordered class-specific demands into path flows with toll order.

$$f_p^{rs} = b_p^{rs} - b_{p-1}^{rs}, \quad \forall rs \in RS, p \in P_{rs} \tag{4.7}$$

Instead of utilizing the path flow as decision variables, the alternative formulation takes the boundaries as decision variables. For an OD pair rs with $|P_{rs}|$ paths, there will be $|P_{rs}| - 1$ independent boundary variables because the two outmost are fixed. Further, given a feasible

boundary pattern, the multiclass external cost in Eq. (4.3a) can be rewritten as:

$$\sum_{rs\in RS}\sum_{p\in P_{rs}} \underbrace{m_p^{rs}}_{\text{Toll}} \int_{b_{p-1}^{rs}}^{b_p^{rs}} \underbrace{g_{rs}(v)}_{\text{TEM}} \underbrace{dv}_{\text{Flow}} = \sum_{rs\in RS}\sum_{p\in P_{rs}} m_p^{rs}\left(G_{rs}\left(b_p^{rs}\right) - G_{rs}\left(b_{p-1}^{rs}\right)\right) \tag{4.8}$$

where $G_{rs}(v)$ is the integral of $g_{rs}(v)$, and we call it *flow-weighted TEM*, which is a convexly piecewise linear function as illustrated in Fig. 4.3(b).

$$G_{rs}(v) = \begin{cases} \beta_1^{rs} v, & \text{if } v \in [0, q_1^{rs}) \\ \beta_2^{rs}\left(v - q_1^{rs}\right) + \beta_1^{rs} q_1^{rs}, & \text{if } v \in [q_1^{rs}, q_1^{rs} + q_2^{rs}) \\ \ldots & \\ \beta_k^{rs}\left(v - \sum_{i=1}^{k-1} q_i^{rs}\right) + \sum_{i=1}^{k-1} \beta_i^{rs} q_i^{rs}, & \text{if } v \in \left[\sum_{i=1}^{k-1} q_i^{rs}, \sum_{i=1}^{k} q_i^{rs}\right) \\ \ldots & \\ \beta_{|K|}^{rs}\left(v - \sum_{i=1}^{|K|-1} q_i^{rs}\right) + \sum_{i=1}^{|K|-1} \beta_i^{rs} q_i^{rs}, & \text{if } v \in \left[\sum_{i=1}^{|K|-1} q_i^{rs}, \sum_{i=1}^{|K|} q_i^{rs}\right] \end{cases} \tag{4.9}$$

Based on the order information of paths and travelers and the boundary variables, the alternative formulation of the MBTA problem reads:

$$\min Z_2(\mathbf{b}) = \underbrace{\sum_{a\in A}\int_0^{x_a} t_a(v)dv}_{\text{Beckmann transformation}} + \underbrace{\sum_{rs\in RS}\sum_{p\in P_{rs}} m_p^{rs}\left(G_{rs}\left(b_p^{rs}\right) - G_{rs}\left(b_{p-1}^{rs}\right)\right)}_{\text{Multiclass external cost}} \tag{4.10a}$$

$$f_p^{rs} = b_p^{rs} - b_{p-1}^{rs}, \quad \forall rs \in RS, p \in P_{rs} \tag{4.10b}$$

$$x_a = \sum_{rs\in RS}\sum_{p\in P_{rs}} f_p^{rs}\delta_{ap}^{rs}, \quad \forall a \in A \tag{4.10c}$$

$$0 = b_0^{rs} \le b_1^{rs} \le b_2^{rs} \le \ldots \le b_{|P_{rs}|}^{rs} = q^{rs}, \quad \forall rs \in RS \tag{4.10d}$$

The objective function in Eq. (4.10a) is convex (Leurent, 1993) and consists of two terms. The first term is the well-known Beckmann transformation to model the congestion effect; the second term is the multiclass external cost to model the tolling effect on travelers from different user classes. The two terms work collaboratively to allocate travelers onto the transportation network according to their' TEMs. Boundaries cut the OD demand into path flows through Eq. (4.10b). Eq. (4.10c) converts path flows to link flows. Eq. (4.10d) holds the correct order of

boundaries, which ensures flow conservation and the nonnegativity of path flows. To see the convexity of the objective function, we can alternatively write the objective function as:

$$\min Z_2(\mathbf{b}) = \sum_{a\in A}\int_0^{x_a} t_a(v)dv + \sum_{rs\in RS}\sum_{p=1,...,|P_{rs}|-1} \underbrace{\left(m_p^{rs} - m_{p+1}^{rs}\right)}_{\geq 0} G_{rs}\left(b_p^{rs}\right) + \underbrace{m_{|P_{rs}|}^{rs} G_{rs}\left(b_{|P_{rs}|}^{rs}\right)}_{\text{Constant}} \tag{4.11}$$

which is the summation of Beckmann transformation and convexly piecewise linear functions, hence convex but nondifferentiable.

4.2.3.2 Solution properties

Here below, we analyze the solution properties of the alternative formulation (4.10).

Theorem 4.1 (Equivalence)**.** *A solution* $\mathbf{b}^*$ *of the alternative formulation (4.10) corresponds to an MBTE flow in Condition 4.1.*

Proof. By attaching Lagrange multipliers λ_p^{rs} to the boundary order constraint (4.10d), the Lagrangian of model (4.10) is given by:

$$L(\mathbf{b},\lambda) = \sum_{a\in A}\int_0^{x_a} t_a(v)dv + \sum_{rs\in RS}\sum_{p\in P_{rs}} m_p^{rs}\left(G_{rs}\left(b_p^{rs}\right) - G_{rs}\left(b_{p-1}^{rs}\right)\right) + \sum_{rs\in RS}\sum_{p\in P_{rs}} \lambda_p^{rs}\left(b_{p-1}^{rs} - b_p^{rs}\right) \tag{4.12}$$

The KKT conditions for the nondifferentiable convex program (4.10) read:

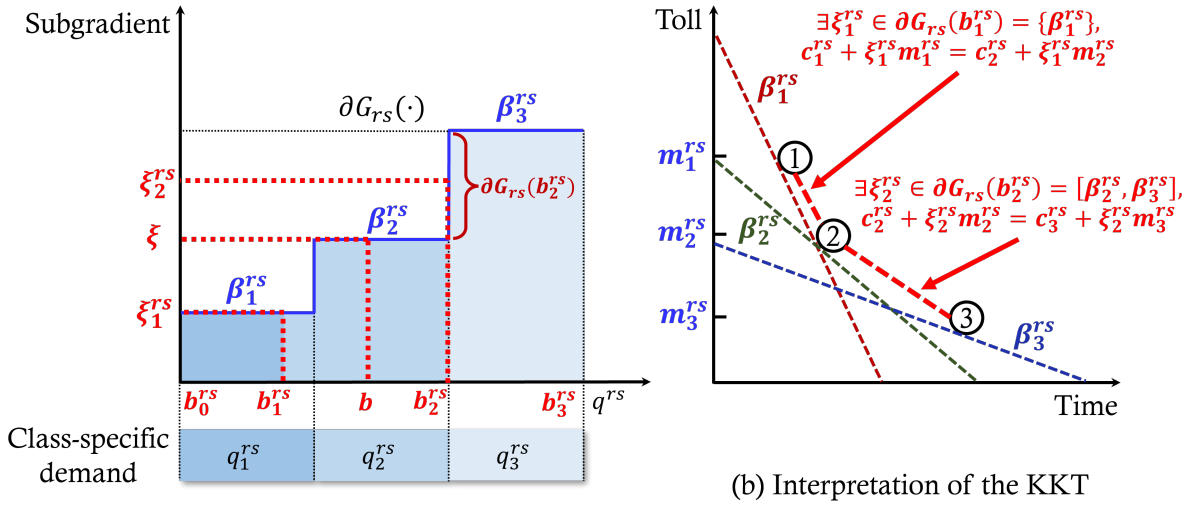
Stationarity	$0 \in \dfrac{\partial L(\mathbf{b},\lambda)}{\partial b_p^{rs}}, \quad \forall rs \in RS, \forall p = 1, ..., \lvert P_{rs}\rvert - 1$	(4.13)
Primal feasibility	$b_p^{rs} - b_{p-1}^{rs} \geq 0, \quad \forall rs \in RS, \forall p \in P_{rs}$	(4.14)
Dual feasibility	$\lambda_p^{rs} \geq 0, \quad \forall rs \in RS, \forall p \in P_{rs}$	(4.15)
Complementary slackness	$\lambda_p^{rs}\left(b_p^{rs} - b_{p-1}^{rs}\right) = 0, \quad \forall rs \in RS, \forall p = P_{rs}$	(4.16)

We check the stationarity condition (4.13):

$$
\begin{aligned}
0 \in \frac{\partial L(\mathbf{b},\lambda)}{\partial b_p^{rs}} &= \sum_{a\in A} \frac{\partial \sum_{a\in A}\int_0^{x_a} t_a(v)dv}{\partial x_a} \times \frac{\partial x_a}{\partial b_p^{rs}} + \frac{\partial \left(m_p^{rs}-m_{p+1}^{rs}\right) G_{rs}\left(b_p^{rs}\right)}{\partial b_p^{rs}} - \lambda_p^{rs} + \lambda_{p+1}^{rs} \\
&= \sum_{a\in A} t_a\left(x_a\right)\left(\delta_{ap}^{rs}-\delta_{ap+1}^{rs}\right) + \frac{\partial G_{rs}\left(b_p^{rs}\right)}{\partial b_p^{rs}}\left(m_p^{rs}-m_{p+1}^{rs}\right) - \lambda_p^{rs} + \lambda_{p+1}^{rs} \\
&= c_p^{rs} - c_{p+1}^{rs} + \partial G_{rs}\left(b_p^{rs}\right)\left(m_p^{rs}-m_{p+1}^{rs}\right) - \lambda_p^{rs} + \lambda_{p+1}^{rs}, \quad \forall rs \in RS, p = 1,...,|P_{rs}|-1
\end{aligned}
\tag{4.17}
$$

where $\partial G_{rs}\left(b_p^{rs}\right)$ is the subdifferential of G_{rs} at b_p^{rs}, as illustrated by the blue lines in Fig. 4.5(a).

$$
\partial G_{rs}\left(b_p^{rs}\right) = \begin{cases} \left[\beta_k^{rs}, \beta_{k+1}^{rs}\right] & \text{if } b_p^{rs} = \sum_{i=0}^{k} q_k^{rs}, \forall k = 1,...,|K|-1 \\ \left\{g_{rs}\left(b_p^{rs}\right)\right\} & \text{otherwise} \end{cases} \tag{4.18}
$$



(a) Illustration of subdifferential $\partial G_{rs}(\cdot)$

(b) Interpretation of the KKT conditions

Figure 4.5: Illustration of subdifferential and the interpretation of KKT condition.

The stationary condition Eq. (4.17) implies that there exists a subgradient $\xi_p^{rs} \in \partial G_{rs}\left(b_p^{rs}\right)$ such that:

$$
c_p^{rs} + \xi_p^{rs} m_p^{rs} - \lambda_p^{rs} = c_{p+1}^{rs} + \xi_p^{rs} m_{p+1}^{rs} - \lambda_{p+1}^{rs}, \quad \forall rs \in RS, p = 1,...,|P_{rs}|-1 \tag{4.19}
$$

Recall that for any OD pair rs, the paths are (weakly) decreasingly ordered by tolls as per Eq. (4.5). In addition, travelers are (weakly) increasingly ordered by TEMs as per Eq. (4.6).

Therefore, if $b_p^{rs} - b_{p-1}^{rs} > 0$ (path p carries positive flows), we have:

$$\xi_{p-1}^{rs} \leq \xi \leq \xi_p^{rs}, \quad \forall rs \in RS, p = 1, ..., |P_{rs}|, b \in \left(b_{p-1}^{rs}, b_p^{rs}\right), \xi \in \partial G_{rs}(b), \tag{4.20}$$

as illustrated by the red dotted lines in Fig. 4.5(a).

The order information Eq. (4.20) helps us to further derive the following inequalities from Eq. (4.19):

$$\begin{aligned} ... \geq c_{p-1}^{rs} + \xi m_{p-1}^{rs} - \lambda_{p-1}^{rs} \geq & c_p^{rs} + \xi m_p^{rs} - \lambda_p^{rs} \leq c_{p+1}^{rs} + \xi m_{p+1}^{rs} - \lambda_{p+1}^{rs} \leq ... \\ & \forall rs \in RS, b \in \left(b_{p-1}^{rs}, b_p^{rs}\right), \xi \in \partial G_{rs}(b) \end{aligned} \tag{4.21}$$

Further, for any OD pair rs, if $b_p^{rs} - b_{p-1}^{rs} > 0$ (path p carries positive flows), then $\lambda_p^{rs} = 0$ according to the complementary slackness condition (4.16). Eq. (4.21) becomes:

$$c_p^{rs} + \xi m_p^{rs} \leq c_{p'}^{rs} + \xi m_{p'}^{rs} - \lambda_{p'}^{rs}, \quad \forall rs \in RS, p' \in P_{rs}, b \in \left(b_{p-1}^{rs}, b_p^{rs}\right), \xi \in \partial G_{rs}(b) \tag{4.22}$$

As the dual variables are nonnegative as per Eq. (4.15), we can safely drop them to obtain:

$$c_p^{rs} + \xi m_p^{rs} \leq c_{p'}^{rs} + \xi m_{p'}^{rs}, \quad \forall rs \in RS, p' \in P_{rs}, \underbrace{b \in \left(b_{p-1}^{rs}, b_p^{rs}\right), \xi \in \partial G_{rs}(b)}_{\text{TEM of travelers choosing path } p} \tag{4.23}$$

Eq. (4.23) indicates that for any OD pair rs, if path p carries positive flows (i.e., $b_p^{rs} - b_{p-1}^{rs} > 0$), then all the travelers choosing path p (i.e., $b \in \left(b_{p-1}^{rs}, b_p^{rs}\right)$) can not find any other path with a smaller generalized cost (Eq. (4.2)). In other words, travelers in each user class choose the path that can minimize their generalized cost, which is equivalent to Condition 4.1. This completes the proof. □

To better illustrate the idea, we use an example in Fig. 4.5 to interpret the equivalence condition. As illustrated by Fig. 4.5(a), since all three paths are used at equilibrium, $\lambda_1^{rs} = \lambda_2^{rs} = \lambda_3^{rs} = 0$ as per the complementary slackness condition. Therefore, Eq. (4.19) reads

$$\begin{cases} \exists \xi_1^{rs} \in \partial G_{rs}\left(b_1^{rs}\right) = \{\beta_1^{rs}\} \Rightarrow c_1^{rs} + \xi_1^{rs} m_1^{rs} = c_2^{rs} + \xi_1^{rs} m_2^{rs} \\ \exists \xi_2^{rs} \in \partial G_{rs}\left(b_2^{rs}\right) = [\beta_2^{rs}, \beta_3^{rs}] \Rightarrow c_2^{rs} + \xi_2^{rs} m_2^{rs} = c_3^{rs} + \xi_2^{rs} m_3^{rs} \end{cases}, \tag{4.24}$$

as shown in Fig. 4.5(b). The physical meanings are that (i) the used paths must provide the minimum generalized cost given some β_k^{rs} for $k \in K$, and (ii) the slope between two adjacent used paths p and $p+1$ should be within the subdifferential of flow-weighted TEM function G_{rs} at the boundary b_p^{rs} between them.

Theorem 4.2 (Solution uniqueness)**.** *If the link travel time functions $t_a(x_a)$ are strictly increasing, then the link flows are unique at equilibrium; however, as in the conventional MBTA model, the path flows of the alternative formulation are not unique in general transportation networks.*

Proof. We discuss the Beckmann transformation and the multiclass external cost term separately. It is well-known that the Beckmann transformation is strictly convex with respect to link flows if the link travel time functions $t_a(x_a)$ are strictly increasing (Sheffi, 1985). Therefore, the link flows are unique. The Hessian matrix of the Beckmann transformation with respect to boundaries $\mathbf{b}$ is:

$$\nabla^2_{\mathbf{b}}\text{Beck}(\mathbf{b}) = \underbrace{(\Delta\Lambda)^T \overbrace{\nabla^2_{\mathbf{x}}\text{Beck}(\mathbf{x})}^{\succ 0} \Delta\Lambda}_{\succeq 0} \tag{4.25}$$

where Δ is the link-path incidence matrix and Λ is the path-boundary incidence matrix; $\mathbf{x} = \Delta\mathbf{f}$ and $\mathbf{f} = \Lambda\mathbf{b}$ correspond to Eq. (4.10c) and (4.10b), respectively. We further have:

$$rank\left(\nabla^2_{\mathbf{b}}\text{Beck}(\mathbf{b})\right) \leq \min\left\{rank\left(\Delta\Lambda\right), rank\left(\nabla^2_{\mathbf{x}}\text{Beck}(\mathbf{x})\right)\right\} \leq |A| \tag{4.26}$$

In general transportation networks, $\nabla^2_{\mathbf{b}}\text{Beck}(\mathbf{b})$ is a positive semidefinite matrix but without full rank, because $\nabla^2_{\mathbf{b}}\text{Beck}(\mathbf{b})$ is a $|P|$ dimensional square matrix and $|P| >> |A|$. That is, the Beckmann term is weakly convex with respect to $\mathbf{b}$ in general transportation networks.

For the piecewise linear convex function G_{rs}, there is no Hessian. But we can easily find two points b_1 and b_2 belong to the same piece such that:

$$G_{rs}(ub_1 + (1-u)b_2) = uG_{rs}(b_1) + (1-u)G_{rs}(b_2), \quad \forall u \in [0,1], \tag{4.27}$$

which means G_{rs} is weakly convex with respect to b. Therefore, the objective function (4.10a) is the summation of weakly convex functions, which is weakly convex with respect to $\mathbf{b}$ in general transportation networks. Hence, boundaries $\mathbf{b}$ are not unique in general transportation networks, and neither are the path flows, as they have a unique conversion as per Eq. (4.10b). □

To conclude this subsection, the alternative formulation (4.10) models the same equilibrium problem as the conventional MBTA model (4.3). Also, like the conventional MBTA model, only the link flow solution of the alternative formulation is unique in general transportation networks.

4.2.4 Comparisons between the conventional model and the alternative formulation

Compared to the conventional MBTA model (4.3), the alternative formulation (4.10) drops the class index k but adds a piecewise linear function in the objective function, resulting in the following differences in the formulation, as shown in Table 4.2.

Table 4.2: Differences in the model formulation

	Conventional model (4.3)	Alternative formulation (4.10)
Decision variables	Class-specific path flow f_p^{rsk}	Boundary b_p^{rs}
Objective function	Convex differentiable	Convex nondifferentiable
Constraints	Flow conservation (4.3b); nonnegativity (4.3e)	Boundary order (4.10d)

Interestingly, despite the formulation differences, the last subsection demonstrates that the two formulations solve the same equilibrium problem and have the same solution properties. In this case, we definitely prefer the more computation-friendly one. Here below, we analyze these two formulations from the computational point of view. As this chapter extends the path-based gradient projection (GP) algorithm (see Sec. 4.3) to solve the alternative formulation (4.10), we discuss these two formulations in this algorithmic setting.

4.2.4.1 Memory requirement

For path-based traffic assignment algorithms, storing the paths is the most memory-consuming operation. Usually, a transportation network $G = (N, A)$ has 10 to 20 percent of nodes that are zones, and the total number of OD pairs is $|R| \cdot |S|$. The expected number of paths between an OD pair, N_p, is typically less than 5 to achieve the UE condition, depending on the chosen path-based solution algorithm (Chen et al., 2002a). Further, each path in a traffic network has roughly $O(|N|^{1/2})$ links (Jayakrishnan et al., 1994). Assuming the paths are stored as linked lists[7], we need to store $O(|R| \cdot |S| \cdot N_p \cdot |N|^{1/2})$ links for single-class traffic assignment problems. For the

[7]Linked lists can facilitate path-level operations such as dropping duplicate paths, though the most memory-saving data structure is tree (Jayakrishnan et al., 1994).

conventional MBTA model (4.3), we need to further store the paths for each user class, yielding in total $O(|R| \cdot |S| \cdot N_p \cdot |N|^{1/2} \cdot |K|)$ links. Taking the well-known Chicago Regional network[8] as an example, the memory requirements under a different number of user classes are shown in Table 4.3. Hence, in practice, transport planners usually keep the number of user classes in a controllable range (usually less than 20), which is the main reason that restricts the fineness of the MBTA results. On the contrary, as the alternative formulation (4.10) does not contain class-specific variables, we only need to store $O(|R| \cdot |S| \cdot N_p \cdot |N|^{1/2})$ links, the same magnitude as the single-class traffic assignment model, which is a very desirable feature for transport planning.

Table 4.3: The required memory for storing the paths using the conventional MBTA model on the Chicago Regional network

$\|K\|$	$\|N\|$	$\|A\|$	$\|R\| \cdot \|S\|$	N_p	# of paths	# of links per path	# of links	RAM (GB)
1	12,982	39,018	2,296,227	2	4,592,454	113.94	5.23E+08	3.90
2	12,982	39,018	2,296,227	2	9,184,908	113.94	1.05E+09	7.80
5	12,982	39,018	2,296,227	2	22,962,270	113.94	2.62E+09	19.49
10	12,982	39,018	2,296,227	2	45,924,540	113.94	5.23E+09	38.99
20	12,982	39,018	2,296,227	2	91,849,080	113.94	1.05E+10	77.97
50	12,982	39,018	2,296,227	2	229,622,700	113.94	2.62E+10	194.93

[a] We make a conservative estimation that each OD pair only has 2 paths, i.e., $N_p = 2$.
[b] Each link is assumed to be stored as a pointer occupying 8 bytes in a 64-bit operating system.

4.2.4.2 Computation time

For traffic assignment problems, it is well-known that generating the shortest paths is computationally expensive during the solution process. This is especially true for the MBTA problem, where bi-criteria shortest path calculation is involved. For the conventional MBTA model (4.3) and the alternative formulation (4.10), the path generation process is the same. Assuming we are using the Dijkstra algorithm with a Fibonacci heap (Fredman and Tarjan, 1987) data structure (the complexity is $O\left(|A| + |N| \log |N|\right)$), at each iteration, we need to generate the shortest path tree for each origin and each user class, which will result in $O(|R| \cdot (|A| + |N| \log |N|) \cdot |K|)$ operations. In addition, the other time-consuming modules of the path-based GP algorithm are (1) calculating the first derivative of each path, (2) calculating the second derivative of each path, and (3) updating the link flows along with path flows (Jayakrishnan et al., 1994). For the conventional MBTA model, each of these modules will take $O(|R| \cdot |S| \cdot N_p \cdot |N|^{1/2} \cdot |K|)$ operations

[8]The Chicago regional network data is available at https://github.com/bstabler/TransportationNetworks

per iteration. For the alternative formulation, the complexity reduces to $O(|R| \cdot |S| \cdot N_p \cdot |N|^{1/2})$, the same level as the single-class traffic assignment problem. We summarize the complexities of the above operations in Table 4.4. Therefore, the alternative formulation (4.10) has advantages in equilibrating the path flows at each iteration.

Table 4.4: The comparison of operation complexity

Operation	Conventional model (4.3)	Alternative formulation (4.10)
Shortest path	$O(\|R\| \cdot (\|A\| + \|N\| \log \|N\|) \cdot \|K\|)$	$O(\|R\| \cdot (\|A\| + \|N\| \log \|N\|) \cdot \|K\|)$
First derivative	$O(\|R\| \cdot \|S\| \cdot N_p \cdot \|N\|^{1/2} \cdot \|K\|)$	$O(\|R\| \cdot \|S\| \cdot N_p \cdot \|N\|^{1/2})$
Second derivative	$O(\|R\| \cdot \|S\| \cdot N_p \cdot \|N\|^{1/2} \cdot \|K\|)$	$O(\|R\| \cdot \|S\| \cdot N_p \cdot \|N\|^{1/2})$
Update link flow	$O(\|R\| \cdot \|S\| \cdot N_p \cdot \|N\|^{1/2} \cdot \|K\|)$	$O(\|R\| \cdot \|S\| \cdot N_p \cdot \|N\|^{1/2})$

4.3 The path-based subgradient projection algorithm

The alternative formulation (4.10) is built on the order information of paths and travelers, which calls for a path-based algorithm. Besides, the objective function contains nondifferentiable convex functions G_{rs}, requiring a nonsmooth optimization method. In view of this, this chapter extends the path-based gradient projection (GP) algorithm (Bertsekas, 1976; Jayakrishnan et al., 1994) by allowing subgradients, resulting in a path-based subgradient (SGP) algorithm.

4.3.1 The overall framework

The framework of the path-based SGP algorithm is shown in Fig. 4.6. To initialize the algorithm, the simplest way is to assign all the class-specific demands to the corresponding minimum generalized cost path between each OD pair. In order to avoid enumerating a massive number of paths in general transportation networks. The SGP algorithm only keeps a *working path set* $\tilde{P} \subseteq P$ containing positive-flow paths. At each iteration, a *column generation* process will generate an auxiliary path set to determine the descent direction. For the MBTA problem, the column generation process tries to find the minimum generalized cost path for each OD pair and each user class. A "brute-force" method is often used to update the generalized link costs and run the traditional one-to-all shortest path algorithm (e.g., Dijkstra algorithm) for each origin and each user class, where the complexity is proportional to the number of user classes $|K|$, as discussed in Sec. 4.2.4.2. Instead, this chapter adopts the *Dijkstra-like bi-criteria shortest path*

algorithm[9] (see B.1 for pseudo-code) developed by Sedeño-Noda and Raith (2015) to generate efficient paths for each origin and all user classes in one pass, which shows better performance compared to the traditional method. It is worth mentioning that, when adding new paths to the working path set, we need to maintain the path toll order as per Eq. (4.5). At the same time, a temporary boundary value is assigned to the newly added path for the following equilibration operation.

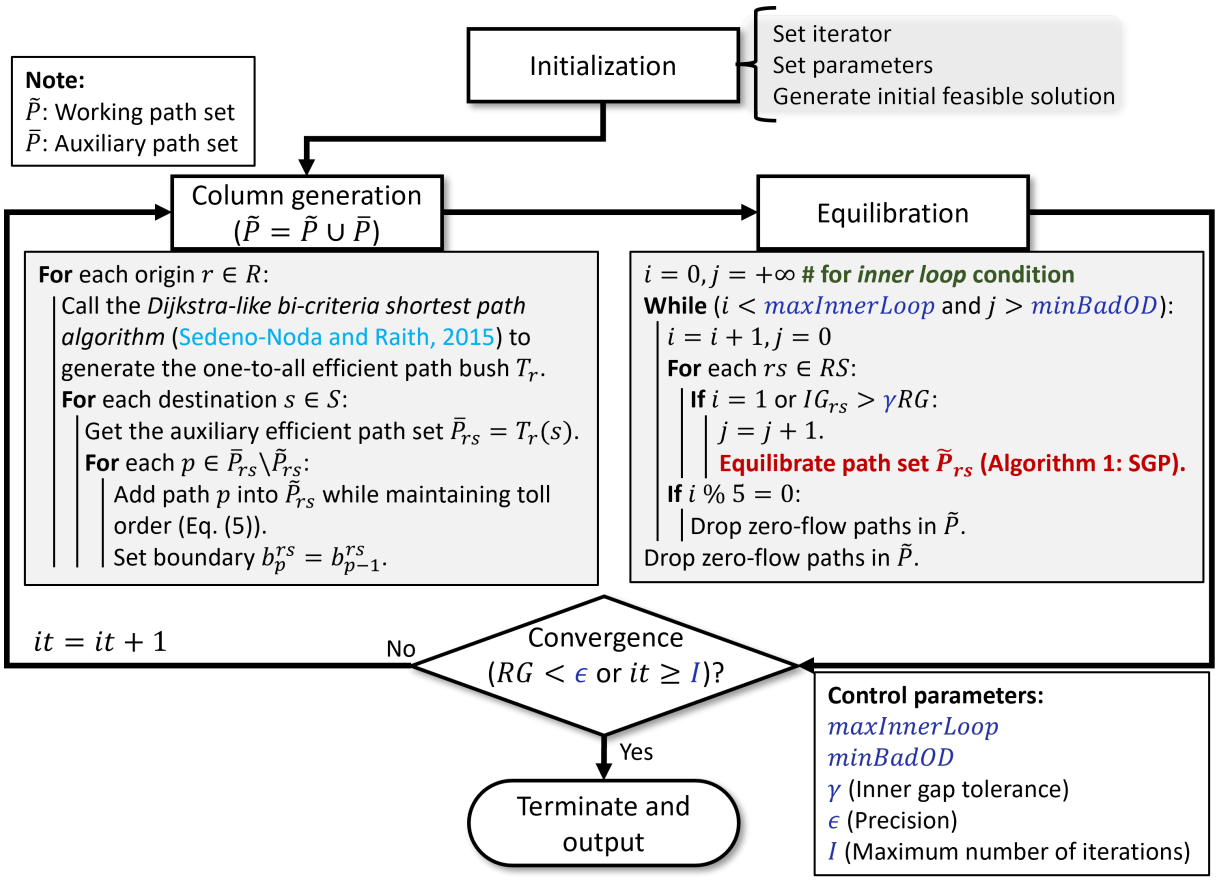


Figure 4.6: The framework of the path-based SGP algorithm.

After that, based on the spirit of the Gauss-Seidel method or the one-at-a-time flow updating strategy (Chen and Jayakrishnan, 1998), the equilibration process will sequentially equilibrate the working path set between each OD pair. In the alternative formulation (4.10), the decision variables are boundaries. Equilibrating the path set between an OD pair rs is essentially updating the boundary between each adjacent path pair in $\tilde{P}_{rs}$, in which the physical meaning is to shift traffic flow between the two adjacent paths, as shown in Fig. 4.4(c). This innovative boundary-

[9]This algorithm utilized a Dijkstra-like label-setting method to generate one-to-all efficient paths and has a complexity of $O\left(C\left(|A| + |N|\log|N|\right)\right)$, where C is a constant denoting the number of efficient paths.

updating idea was recently proposed by Xie et al. (2021) to solve the continuous MBTA problem, which updates the TEM boundaries under a probability density function (PDF) to adjust the corresponding path flows. Although their formulation cannot address the case in this chapter where TEM is discretely distributed, this chapter borrows the boundary-updating concept to develop the path-based SGP algorithm.

In addition, given the fact that the column generation process usually consumes a considerable amount of time, it is effective to incorporate a greedy strategy called *inner loop* in the equilibration process to fully equilibrate the path set before performing the next column generation (Xie et al., 2018; Xu et al., 2020; Tan et al., 2022), where the precision of working path set $\tilde{P}_{rs}$ is evaluated by:

$$IG_{rs} = 1 - \frac{\sum_{k \in K} q_k^{rs} d_{\dot{p}}^{rs}(\beta_k^{rs})}{\sum_{\tilde{p} \in \tilde{P}_{rs}} \left(\left(b_{\tilde{p}}^{rs} - b_{\tilde{p}-1}^{rs} \right) c_{\tilde{p}}^{rs} + m_{\tilde{p}}^{rs} \left(G_{rs} \left(b_{\tilde{p}}^{rs} \right) - G_{rs} \left(b_{\tilde{p}-1}^{rs} \right) \right) \right)} \tag{4.28}$$

where $\dot{p}$ is the minimum generalized cost path in working path set $\tilde{P}_{rs}$ between OD pair rs for user class k.

After equilibration, the relative gap will be used to measure the solution precision, which is defined as:

$$RG = 1 - \frac{\sum_{rs \in RS} \sum_{k \in K} q_k^{rs} d_{\bar{p}}^{rs}(\beta_k^{rs})}{\sum_{rs \in RS} \sum_{\tilde{p} \in \tilde{P}_{rs}} \left(\left(b_{\tilde{p}}^{rs} - b_{\tilde{p}-1}^{rs} \right) c_{\tilde{p}}^{rs} + m_{\tilde{p}}^{rs} \left(G_{rs} \left(b_{\tilde{p}}^{rs} \right) - G_{rs} \left(b_{\tilde{p}-1}^{rs} \right) \right) \right)} \tag{4.29}$$

where $\bar{p}$ is the minimum generalized cost path in auxiliary path set $\bar{P}_{rs}$ between OD pair rs for user class k.

4.3.2 Equilibration

4.3.2.1 One-dimensional nondifferentiable convex subproblem

The efficient implementation of the path-based traffic assignment algorithm usually relies on divide-and-conquer strategies such as the Gauss-Seidel method or the one-at-a-time flow updating strategy (Chen and Jayakrishnan, 1998). Based on this idea, the alternative formulation (4.10) can be decomposed as a series of one-dimensional nondifferentiable convex subproblems

with respect to b_p^{rs}:

$$\min \bar{Z}_2(b_p^{rs}) = \sum_{a\in A}\int_0^{x_a} t_a(v)dv + m_p^{rs}\left(G_{rs}\left(b_p^{rs}\right) - \underbrace{G_{rs}\left(b_{p-1}^{rs}\right)}_{\text{Constant}}\right) + m_{p+1}^{rs}\left(\underbrace{G_{rs}\left(b_{p+1}^{rs}\right)}_{\text{Constant}} - G_{rs}\left(b_p^{rs}\right)\right)$$

$$\Leftrightarrow \sum_{a\in A}\int_0^{x_a} t_a(v)dv + G_{rs}\left(b_p^{rs}\right)\left(m_p^{rs} - m_{p+1}^{rs}\right) \tag{4.30a}$$

$$f_p^{rs} = b_p^{rs} - b_{p-1}^{rs}, \quad \forall rs \in RS, p \in P_{rs} \tag{4.30b}$$

$$x_a = \sum_{rs\in RS}\sum_{p\in P_{rs}} f_p^{rs}\delta_{ap}^{rs}, \quad \forall a \in A \tag{4.30c}$$

$$b_{p-1}^{rs} \le b_p^{rs} \le b_{p+1}^{rs} \tag{4.30d}$$

for all $rs \in RS$ and $p = 1, ..., |P_{rs}| - 1$.

As the only constraint (4.30d) is a simple box constraint, solving subproblem (4.30) is just moving a single boundary b_p^{rs} forward or backward within a box $\left[b_{p-1}^{rs}, b_{p+1}^{rs}\right]$, where the physical meaning is shifting flow between path p and $p+1$ connecting OD pair rs, as shown in Fig. 4.4(c).

4.3.2.2 Subdifferential and available second-order information

For subproblem (4.30), the objective function contains a differentiable convex term, i.e., the Beckmann transformation, and a nondifferentiable convex term, i.e., $G_{rs}\left(b_p^{rs}\right)\left(m_p^{rs} - m_{p+1}^{rs}\right)$. For the Beckmann term, the first-order derivative is:

$$\begin{aligned}\frac{\partial \sum_{a\in A}\int_0^{x_a} t_a(u)du}{\partial b_p^{rs}} &= \sum_{a\in A}\frac{\partial \sum_{a\in A}\int_0^{x_a} t_a(u)du}{\partial x_a}\times\frac{\partial x_a}{\partial b_p^{rs}} = \sum_{a\in A} t_a\left(x_a\right)\left(\delta_{ap}^{rs} - \delta_{ap+1}^{rs}\right)\\ &= c_p^{rs} - c_{p+1}^{rs}, \quad \forall rs \in RS, p = 1, \ldots, |P_{rs}| - 1\end{aligned} \tag{4.31}$$

And the second-order derivative is:

$$\begin{aligned}s_p^{rs} &= \frac{\partial \sum_{a\in A} t_a\left(x_a\right)\left(\delta_{ap}^{rs} - \delta_{ap+1}^{rs}\right)}{\partial b_p^{rs}} = \sum_{a\in A}\frac{d \sum_{a\in A} t_a\left(x_a\right)\left(\delta_{ap}^{rs} - \delta_{ap+1}^{rs}\right)}{\partial x_a}\times\frac{\partial x_a}{\partial b_p^{rs}}\\ &= \sum_{a\in A} t_a'\left(x_a\right)\left(\delta_{ap}^{rs} - \delta_{ap+1}^{rs}\right)^2, \quad \forall rs \in RS, p = 1, \ldots, |P_{rs}| - 1\end{aligned} \tag{4.32}$$

which is the sum of the travel time functions' first derivatives on the non-common links between

path p and $p+1$.

Recall that for the nondifferentiable convex term G_p^{rs}, the subdifferetntial at b_p^{rs} is:

$$\partial G_{rs}\left(b_p^{rs}\right) = \begin{cases} \left[\beta_k^{rs}, \beta_{k+1}^{rs}\right] & \text{if } b_p^{rs} = \sum_{i=0}^{k} q_k^{rs}, \forall k = 1, ..., |K| - 1 \\ \left\{g_{rs}\left(b_p^{rs}\right)\right\} & \text{otherwise} \end{cases} \tag{4.33}$$

Therefore, the subdifferential of the objective function $\bar{Z}_2$ at b_p^{rs} is:

$$\partial \bar{Z}_2\left(b_p^{rs}\right) = \begin{cases} \left[c_p^{rs} - c_{p+1}^{rs} + \beta_k^{rs}\left(m_p^{rs} - m_{p+1}^{rs}\right), c_p^{rs} - c_{p+1}^{rs} + \beta_{k+1}^{rs}\left(m_p^{rs} - m_{p+1}^{rs}\right)\right], \\ \quad \text{if } b_p^{rs} = \sum_{i=0}^{k} q_k^{rs}, \forall k = 1, ..., |K| - 1 \\ \left\{c_p^{rs} - c_{p+1}^{rs} + g_{rs}\left(b_p^{rs}\right)\left(m_p^{rs} - m_{p+1}^{rs}\right)\right\}, \\ \quad \text{otherwise} \end{cases}$$
$$\forall rs \in RS, p = 1, \ldots, |P_{rs}| - 1 \tag{4.34}$$

To determine a descent direction for subproblem (4.30), we can take any of the subderivative v_p^{rs} from $\partial \bar{Z}_2\left(b_p^{rs}\right)$ (Nesterov, 2003). Specifically, for the non-optimal solution, we take the steepest subderivative for simplicity, as shown in Fig. 4.7(b) and Fig. 4.7(c).

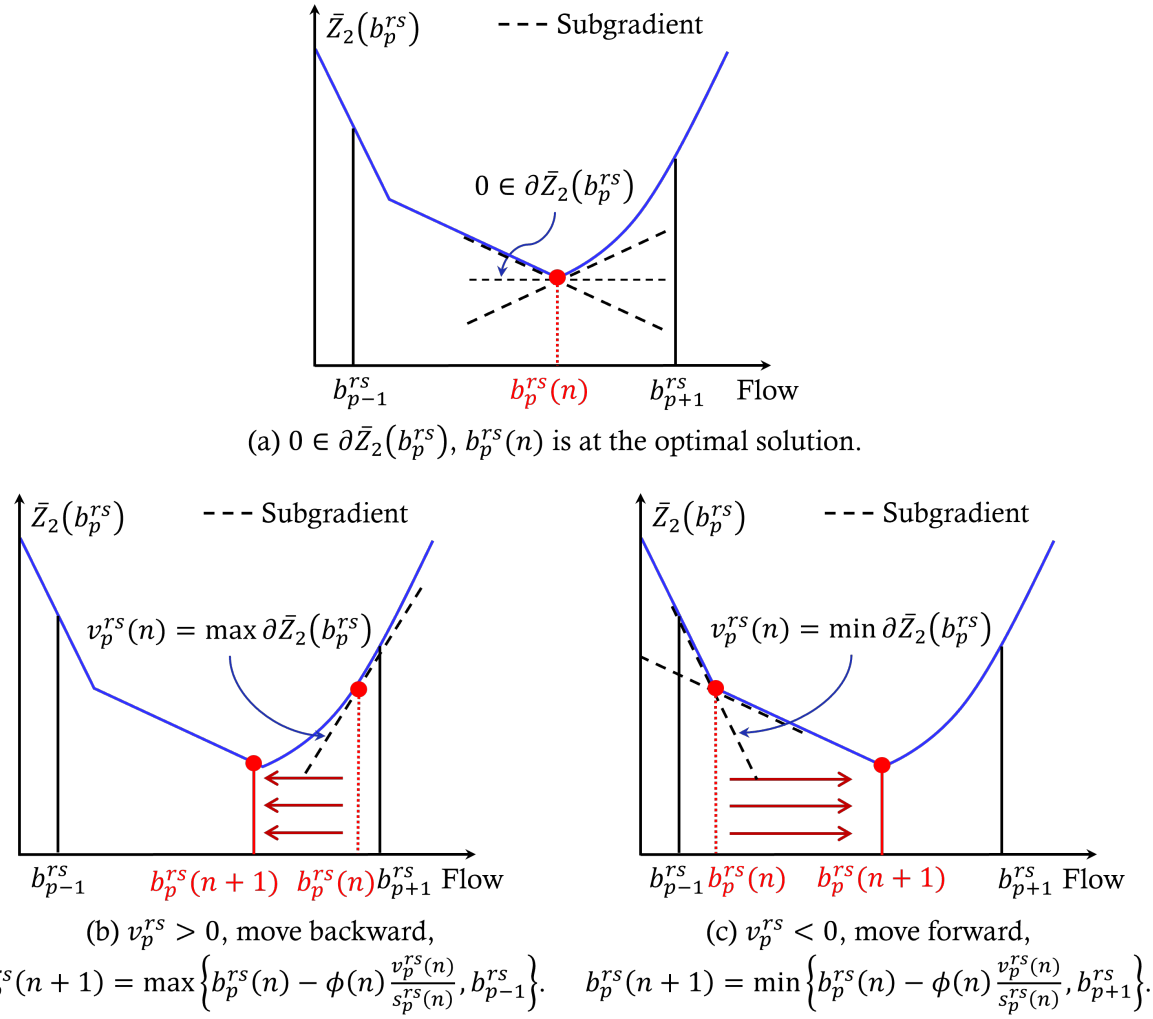
$$v_p^{rs} = \begin{cases} \max \partial \bar{Z}_2\left(b_p^{rs}\right), & \text{if } \min \partial \bar{Z}_2\left(b_p^{rs}\right) > 0 \\ \min \partial \bar{Z}_2\left(b_p^{rs}\right), & \text{if } \max \partial \bar{Z}_2\left(b_p^{rs}\right) < 0 \end{cases}, \quad \forall rs \in RS, p = 1, ..., |P_{rs}| - 1 \tag{4.35}$$

4.3.2.3 Boundary-updating equation

Now, we are ready to derive the iterative boundary-updating equation. Combining the quasi-Newton's method and the projection method, the boundary-updating equation at iteration n is given by:

$$b_p^{rs}(n+1) = \begin{cases} b_p^{rs}(n) & \text{if } 0 \in \partial \bar{Z}_2\left(b_p^{rs}\right) \\ \max\left\{b_p^{rs}(n) - \phi(n)\frac{v_p^{rs}(n)}{s_p^{rs}(n)}, b_{p-1}^{rs}\right\}, & \text{if } v_p^{rs} > 0 \\ \min\left\{b_p^{rs}(n) - \phi(n)\frac{v_p^{rs}(n)}{s_p^{rs}(n)}, b_{p+1}^{rs}\right\}, & \text{if } v_p^{rs} < 0 \end{cases} \quad \forall rs \in RS, p = 1, ..., |P_{rs}| - 1 \tag{4.36}$$

This boundary-updating equation utilizes the subderivative and the available second-order



(a) $0 \in \partial \bar{Z}_2(b_p^{rs})$, $b_p^{rs}(n)$ is at the optimal solution.

(b) $v_p^{rs} > 0$, move backward,
$b_p^{rs}(n+1) = \max\left\{b_p^{rs}(n) - \phi(n)\frac{v_p^{rs}(n)}{s_p^{rs}(n)}, b_{p-1}^{rs}\right\}$.

(c) $v_p^{rs} < 0$, move forward,
$b_p^{rs}(n+1) = \min\left\{b_p^{rs}(n) - \phi(n)\frac{v_p^{rs}(n)}{s_p^{rs}(n)}, b_{p+1}^{rs}\right\}$.

Figure 4.7: Illustration of boundary updating equation.

information to update the boundary using a quasi-Newton method. As shown in Fig. 4.7, if $0 \in \bar{Z}_2\left(b_p^{rs}\right)$, which means $b_p^{rs}(n)$ is already at the optimal solution to subproblem (4.30), there is no need to adjust the boundary; if $v_p^{rs} > 0$, boundary b_p^{rs} will move backward, lower bounded by b_{p-1}^{rs}; if $v_p^{rs} < 0$, boundary b_p^{rs} will move forward, upper bounded by b_{p+1}^{rs}.

In addition, in the boundary updating equation (4.36), we set step size as

$$\phi(n) = \bar{\alpha}\eta^j, \tag{4.37}$$

where j is the smallest nonnegative integer that validates the Armijo rule (Armijo, 1966; Bertsekas, 1976; Jayakrishnan et al., 1994):

$$\text{Armijo rule:} \quad \bar{Z}_2\left(b_p^{rs}(n)\right) - \bar{Z}_2\left(b_p^{rs}(n+1)\right) \geq \sigma v_p^{rs}\left(b_p^{rs}(n) - b_p^{rs}(n+1)\right) \tag{4.38}$$

where $\bar{\alpha}$ is the upper bound of step size, which is set as 1 in this chapter since we are using a "Newton-like" direction; $\eta \in (0,1)$ is the backtracking ratio, which is typically set as $\frac{1}{2}$; $\sigma \in (0,1)$ is acceptable reduction rate and is set as 10^{-4} in this chapter.

4.3.2.4 Implementation

In the end, Algorithm 5 describes the procedure for equilibrating the working path set $\tilde{P}_{rs}$ connecting OD pair rs in detail.

Algorithm 5 (SGP) Equilibrate path set $\tilde{P}_{rs}$.

Input: Working path set $\tilde{P}_{rs}$ and boundaries $\left\{b_p^{rs}(n), \forall rs \in RS, \forall p \in \tilde{P}_{rs}\right\}$.

Output: Working path set $\tilde{P}_{rs}$ and new boundaries $\left\{b_p^{rs}(n+1), \forall rs \in RS, \forall p \in \tilde{P}_{rs}\right\}$.

for each $p = 1, ..., \left|\tilde{P}_{rs}\right| - 1$ **do**
    Update path times c_p^{rs} and c_{p+1}^{rs}.
    Calculate subdifferential $\partial \bar{Z}_2\left(b_p^{rs}\right)$, subderivative v_p^{rs}, and available second-order derivative s_p^{rs} based on Eq. (4.34), Eq. (4.35), and Eq. (4.32), respectively.
    Update boundary b_p^{rs} according to Eq. (4.36) and set $\Delta = b_p^{rs}(n+1) - b_p^{rs}(n)$.
    Update path flows $f_p^{rs} = f_p^{rs} + \Delta$ and $f_{p+1}^{rs} = f_{p+1}^{rs} - \Delta$.
    Update link flow x_a, link time $t_a(x_a)$, and link time derivative $t_a'(x_a)$ of each link a on path p and path $p+1$.
    if $f_{p+1}^{rs} = 0$ and $v_{p+1}^{rs} > 0$ **then**
        Drop path $p+1$.
        $p = p - 1$.
    end if
    end for

4.4 Numerical experiments

In this section, we utilize (1) a small network and (2) several large networks to examine the features of the proposed alternative formulation as well as the path-based SGP algorithm and their computational performance, respectively. All algorithms are coded by Python 3.9.12 with PyPy 3.9[10] as interpreter. The implementations share codes wherever possible to ensure a fair comparison of the computational performance. All the experiments are conducted on a PC with Intel(R) Core(TM) i7-10700 CPU 2.9 GHz and 16 GB RAM using the Windows 10 operating system.

[10] https://www.pypy.org/index.html

In all the networks, the link performance function adopts the ubiquitous Bureau of Public Roads (BPR) function:

$$t_a(x_a) = FFTT_a\left(1 + 0.15\left(\frac{x_a}{Cap_a}\right)^4\right) \tag{4.39}$$

where $FFTT_a$ and Cap_a denote the free-flow travel time and the capacity of link a, respectively.

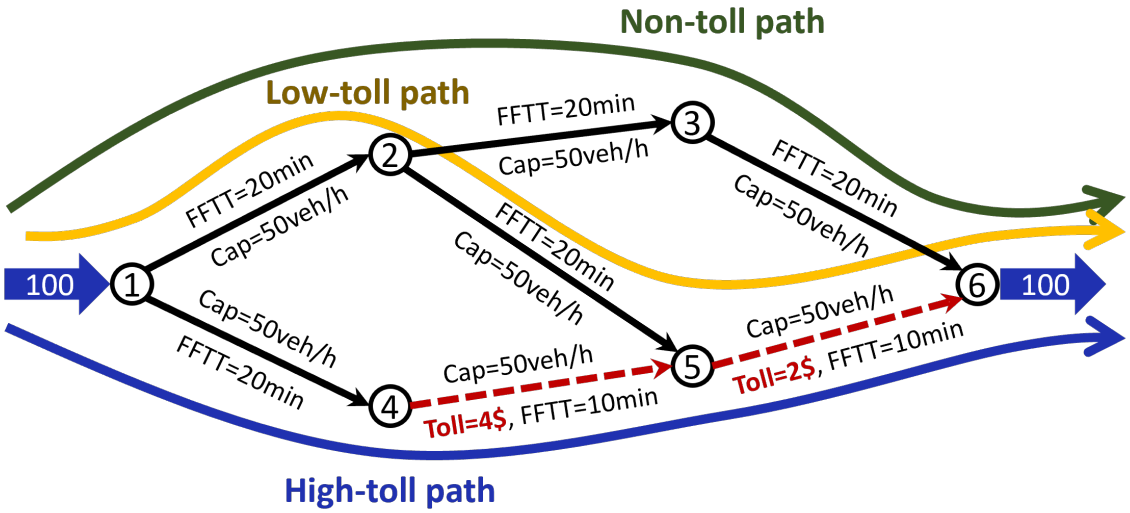
For the path-based SGP algorithm, the parameter settings (see Fig. 4.6) are shown in Table 4.5. To clearly show the algorithmic operations per iteration in the small network, the $maxInnerLoop$ is set as 1 in Sec. 4.4.1. For efficiency consideration, the $maxInnerLoop$ is set as 30 for large networks in Sec. 4.4.2. Except for this, all other parameters are fixed throughout this section. For the conventional MBTA model, we use the path-based GP algorithm (Jayakrishnan et al., 1994) to solve (see B.2), and the parameter settings are the same as the path-based SGP algorithm.

Table 4.5: Algorithmic parameters

Parameter	Value	Description
$maxInnerLoop$	1 in Sec. 4.4.1; 30 in Sec. 4.4.2	
$minBadOD$	$\|RS\|/1000$	
γ	0.1	Inner gap tolerance
ϵ	10^{-10}	Precision
I	30	Maximum number of iterations

4.4.1 A small network

As shown in Fig. 4.8, the small network adapted from Caliper Corporation (Slavin et al., 2014) has 6 nodes, 7 links, and 3 paths. Only two links have tolls; all other links are free. Accordingly, we can divide the 3 paths into a non-toll path, a low-toll path, and a high-toll path. As illustrated by the path indices, we sort these paths in descending order of tolls according to the toll order requirement, i.e., Eq. (4.5). In this network, as only one OD pair (1,7) with demand 100 is considered, the superscript/subscript rs of parameters and variables is omitted for simplicity. Finally, we evenly divide the travelers into 3 classes, i.e., $\mathbf{q} = (33.33, 33.33, 33.33)$, with class-specific TEM as $\beta = (1/30, 1/20, 1/10)$. Correspondingly, the TEM profile function g and the



Path characteristics

Index	Path	Nodes	Free-flow travel time	# of tolled links	Toll ($)
1	High-toll path	1 → 4 → 5 → 6	40 min (0.67 hour)	2	6
2	Low-toll path	1 → 2 → 5 → 6	50 min (0.83 hour)	1	2
3	Non-toll path	1 → 2 → 3 → 6	60 min (1.00 hour)	0	0

Figure 4.8: A small network (reproduced from Slavin et al. (2014))

flow-weighted TEM G in Eqs. (4.6) and (4.9) take the following form:

$$g(v) = \begin{cases} \frac{1}{30}, & \text{if } v \in [0, 33.33) \\ \frac{1}{20}, & \text{if } v \in [33.33, 66.67) \\ \frac{1}{10}, & \text{if } v \in [66.67, 100.00] \end{cases} \Rightarrow G(v) = \begin{cases} \frac{1}{30}v, & \text{if } v \in [0, 33.33) \\ \frac{1}{20}v - 0.56, & \text{if } v \in [33.33, 66.67) \\ \frac{1}{10}v - 3.89, & \text{if } v \in [66.67, 100.00] \end{cases} \tag{4.40}$$

4.4.1.1 Algorithmic details

We first present the details of the path-based SGP algorithm. The path-based SGP algorithm generates an initial solution by assigning the class-specific demand to the corresponding minimum generalized cost path. Therefore, in the small network, the initial solution is $\mathbf{b} = (0.00, 33.33, 66.67, 100)$ and the corresponding flows and travel times of the high-toll path, low-toll path, and non-toll path are $\mathbf{f} = (33.33, 33.33, 33.33)$ and $\mathbf{c} = (0.76, 1.08, 1.18)$, respectively. The objective value and relative gap are 96.79 and 3.61e-2, respectively. At each iteration, the path-based SGP algorithm will update b_1 and b_2 sequentially based on the Gauss-Seidel method:

- *Update* b_1. According to the flow-weighted TEM function $G(v)$ in Eq. (4.40), $b_1 = 33.33$ is a nondifferentiable point. The subdifferential of the objective function (4.10a) with respect to b_1 is $\partial \bar{Z}_2(b_1) = [-0.186, -0.119]$ according to Eq. (4.34). In this case, we take the steepest subgradient $v_1 = -0.186$ as per Eq. (4.35). In addition, the second derivative (4.32) of the Beckmann term is $s_1 = 0.0124$. Therefore, the boundary-updating equation (4.36) will move b_1 forward (see Fig. 4.7(c) for illustration) by:

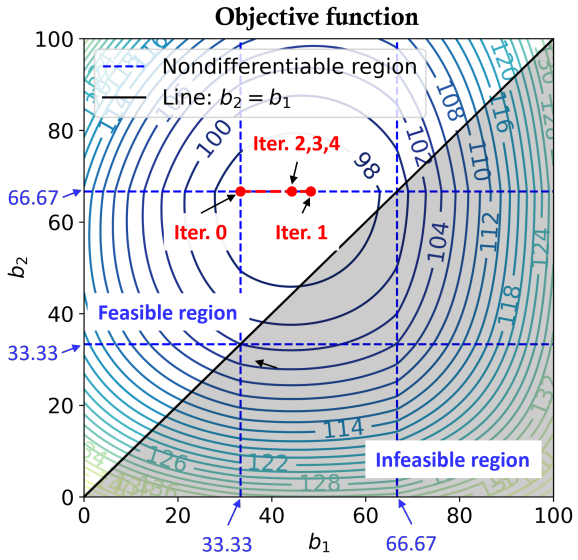
$$b_1 = \min\left\{b_1 - \frac{v_1}{s_1}, b_2\right\} = \min\left\{33.33 - \frac{-0.186}{0.0124}, 66.67\right\} = 48.31 \tag{4.41}$$

  After updating b_1, the boundary, path flow, and path travel time become $\mathbf{b} = (0.00, 48.31, 66.67, 100)$, $\mathbf{f} = (48.31, 18.35, 33.33)$, and $\mathbf{c} = (0.81, 0.97, 1.08)$, respectively.

- *Update* b_2. According to the flow-weighted TEM function $G(v)$ in Eq. (4.40), $b_2 = 66.67$ is also a nondifferentiable point. The subdifferential of the alternative formulation with respect to b_2 is $\partial \bar{Z}_2(b_2) = [-0.0065, 0.0935]$ according to Eq. (4.34). Here, we notice that $0 \in \partial \bar{Z}_2(b_2)$. Therefore, b_2 is already the optimal solution to its subproblem (see Fig. 4.7(a) for illustration). No need to update b_2.

- *Convergence check*. After one iteration, the boundary, path flow, and path travel time become $\mathbf{b} = (0.00, 48.31, 66.67, 100)$, $\mathbf{f} = (48.31, 18.35, 33.33)$, and $\mathbf{c} = (0.81, 0.97, 1.08)$, respectively. The objective value is 96.24, and the relative gap is 5.76e-03, larger than the given precision, i.e., 10^{-10}. The path-based SGP algorithm continues.

Due to the limit of space, we only provide the operation details for the first iteration, for interested readers to replicate our experiments. The remaining details are provided in Fig. 4.9. In this example, the path-based SGP algorithm is very efficient, and it only takes 4 iterations to achieve the predefined precision, i.e., 10^{-10}. Different from the conventional traffic assignment problems, the objective function of the alternative formulation contains some nondifferentiable "lines", i.e., $b_1 = 33.33$, $b_1 = 66.67$, $b_2 = 33.33$, and $b_2 = 66.67$. Therefore, the alternative formulation calls for nondifferentiable optimization methods, e.g., the SGP algorithm proposed in this chapter.



Iteration details

Iteration	Operation	(b_1, b_2)	Obj. Val.	RG
0	NA	(33.33, 66.67)	96.79	3.61e-02
1	Update b_1	(48.31, 66.67)	96.24	5.76e-03
	Update b_2	(48.31, 66.67)		
2	Update b_1	(44.25, 66.67)	96.16	8.63e-06
	Update b_2	(44.25, 66.67)		
3	Update b_1	(44.25, 66.67)	96.16	4.74e-10
	Update b_2	(44.25, 66.67)		
4	Update b_1	(44.25, 66.67)	96.16	0.00
	Update b_2	(44.25, 66.67)		

Note: the RG in iteration 4 is less than machine precision 1e-16.

Figure 4.9: Iterative detail of the subgradient projection algorithm.

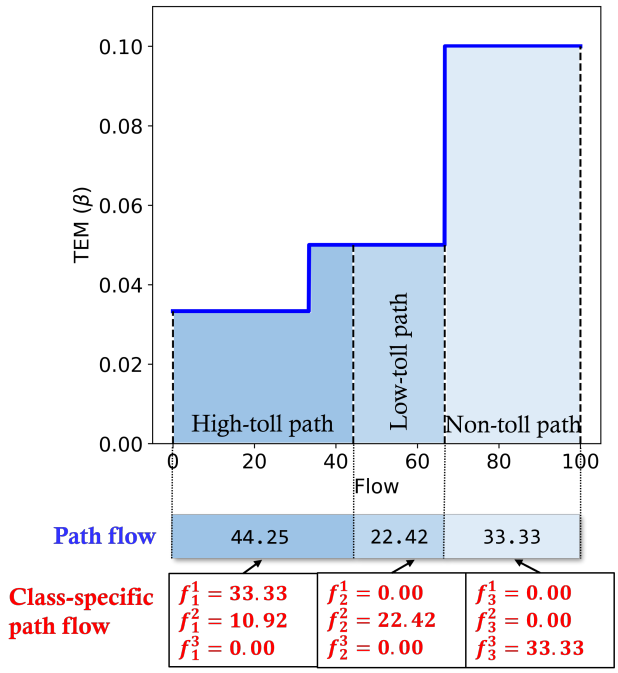
4.4.1.2 Solution results

Fig. 4.10(a) and Fig. 4.10(b) show the solution details of the conventional MBTA model and the alternative formulation for the small network, respectively. Both models provide the same link flows[11] and path flow patterns. Note that, this small network has a special topology in which each path has one or more "*unique*" links, e.g., link (2, 5) only belongs to the low-toll path and $x_{(2,5)} = f_2$. Therefore, these "*unique*" links[12] impose uniqueness on path flows, explaining why the two models provide the same path flow pattern. However, the uniqueness of path flows is not guaranteed in general transportation networks.

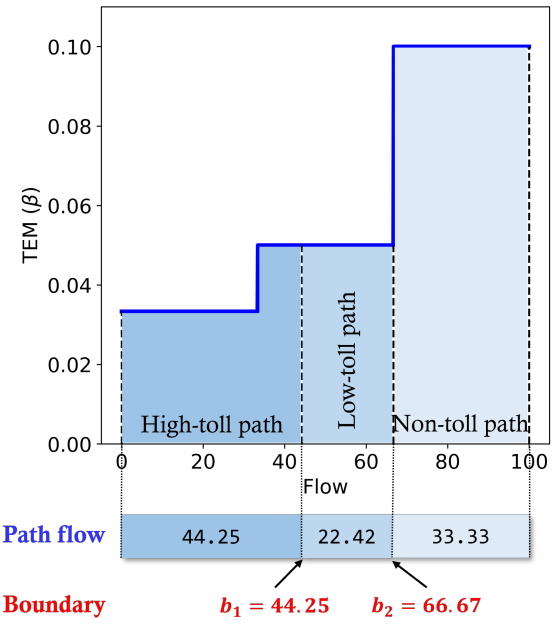
Besides, although the alternative formulation (4.10) only provides the boundary solution or the aggregated path flow, it is straightforward to extract the class-specific path flows as we know the order information of the paths and travelers, as shown in Fig. 4.10(c). Just like initializing a transportation simplex tableau (Dantzig, 1951), we (i) start with the northwestern corner cell, (ii) allocate as many flows as possible, consistent with the available path flow and class-specific demand, and (iii) move one cell to right if there is remaining class-specific demand; otherwise, move one cell down. In a nutshell, the alternative formulation can provide the same level of information, i.e., class-specific path flows and the corresponding class-specific link flows, as

[11] We do not provide detailed values here as the path flows are already the same, and link flows can be easily derived from path flows.

[12] Mathematically, the Beckmann term's Hessian matrix $\nabla^2_{\mathbf{b}}\text{Beck}(\mathbf{b})$ in Eq. (4.25) is positive definite in this small network.



(a) Solution of the conventional MBTA model



(b) Solution of the alternative formulation

Paths ordered by toll

Travelers ordered by TEM

	High-toll path	Low-toll path	Non-toll path	Sum
Class 1 flow	33.33	0.00	0.00	33.33
Class 2 flow	10.92	22.42	0.00	33.33
Class 3 flow	0.00	0.00	33.33	33.33
Sum	44.25	22.42	33.33	100.00

(c) Extract class-specific path flows

Figure 4.10: Computation results of the two models for the small network.

the conventional MBTA model. This is a very useful feature when class-specific information is needed, such as the select-link analysis in transport planning software.

Table 4.6: Variable comparison of the two models

Model	# of decision variables	# of activating decision variables	# of intermediate variables
Conventional MBTA model	9 (f_p^{rsk})	4	21 (x_a^k) + 7 (x_a)
Alternative formulation	2 (b_p^{rs})	2	3 (f_p^{rs}) + 7 (x_a)

As shown in Table. 4.6, the alternative formulation has much fewer decision variables and intermediate variables than the conventional MBTA model. It is worth mentioning that the path-based traffic assignment algorithms usually only keep the activating (class-specific) path flow variables (i.e., positive-flow paths) to reduce the computation burden. For instance, there

are only 4 activating class-specific path flow variables as shown in Fig. 4.10(a). But still, the alternative formulation outperforms the conventional MBTA model by a large margin. Therefore, the alternative formulation can save a great amount of computer memory when dealing with large-scale transportation networks. And it is also faster as there are fewer variables to manipulate and solve.

4.4.2 Large networks

In this section, we focus on analyzing the computational performance of the two formulations and solution algorithms on more realistic transportation networks, as shown in Table 4.7. The detailed network information is available at https://github.com/bstabler/TransportationNetworks. Because the original network instances from the website do not contain tolls, we set toll as 0.1 $\$/km$ on the top 10% congested links under the user equilibrium flow pattern and 0 on other links. Further, to generate user classes, we evenly divide the original demand between each OD pair into $|K|$ classes, and the class-specific VOTs are evenly spaced over [5, 35] to generate the class-specific TEMs.

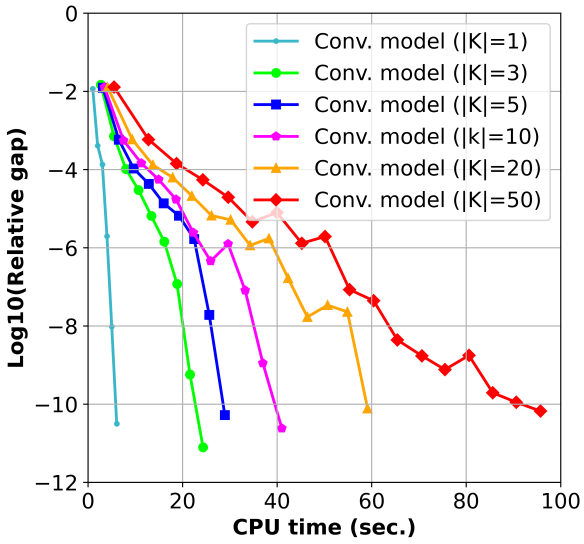
Table 4.7: Large networks.

Network	# of zones	# of nodes	# of links	# of tolled links	# of OD pairs	# of trips
Sioux Falls	24	24	76	8	528	360,600
Anaheim	38	416	914	91	1,406	104,694
Winnipeg	147	1,052	2,836	284	4,344	64,775
Chicago Sketch	387	933	2,950	295	93,135	1,260,907

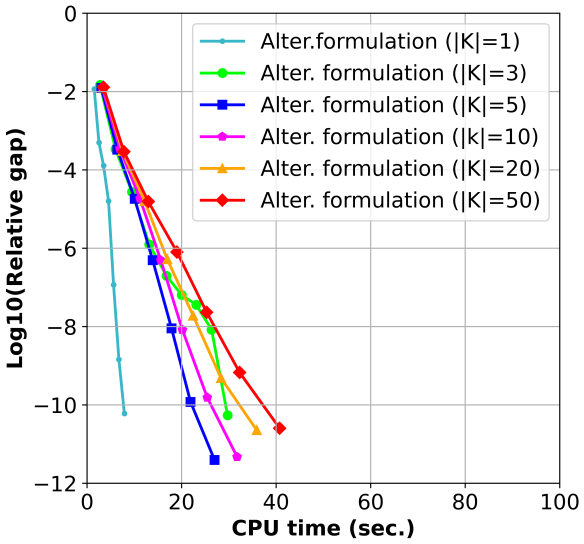
4.4.2.1 Number of classes

Using the Winnipeg network, we first test the performance of the alternative formulation and the path-based SGP algorithm under a different number of user classes. Fig. 4.11 shows that when the number of user classes is small (i.e., $|K| \leq 5$), the two formulations and solution algorithms are matched rivals; when the number of user classes gets larger (i.e., $|K| \geq 5$), the conventional model begins to lose the game. To uncover the reasons behind this, we shall further check the working path set sizes of the two models. As shown in Table 4.8, for the conventional MBTA model, the number of generated paths is proportional to the number of classes. Therefore, each extra user class will substantially increase the number of paths, thereby increasing the operations

at each iteration and slowing down computation time. This phenomenon matches our discussion in Sec. 4.2.4.2. For the alternative formulation, the increase in computation time is mainly due to (i) a mild increase in the path set size and (ii) the flow-weighted TEM function G_{rs} will contain more pieces, which requires more effort in function evaluation. But the computation time increase in the alternative formulation is much more modest than in the conventional model. Therefore the alternative formulation is obviously quicker when the number of classes is large.



(a) Conventional MBTA model's convergence curves.



(b) Alternative formulation's convergence curves.

Figure 4.11: The convergence curves of the two models under different numbers of user classes on the Winnipeg network.

Table 4.8: Comparison of the working path set size and memory consumption of the two models on the Winnipeg network.

	Conventional model (CM)		**Alternative formulation (AF)**		**Comparison**	
# of classes	# of paths	RAM (MB)	# of paths	RAM (MB)	$\frac{\text{\# of paths (CM)}}{\text{\# of paths (AF)}}$	$\frac{\text{RAM (CM)}}{\text{RAM (AF)}}$
1	4516	1.35	4620	1.43	0.99	0.95
3	13319	3.65	7426	2.40	1.79	1.52
5	22026	6.39	8668	2.72	2.55	2.35
10	43710	12.68	9475	2.92	4.62	4.35
20	87280	25.31	10387	3.15	8.40	8.04
50	218183	60.66	11129	3.33	19.56	18.20

As shown in Table 4.8, from the memory consumption point of view, the alternative formulation outperforms the conventional MBTA formulation significantly. Although nowadays computer memory is no longer a big issue, it's still gratifying news for practitioners that there is

a formulation that can reduce the memory requirement of the MBTA problem by one order of magnitude.

4.4.2.2 Network scale

A practical traffic assignment model and the solution algorithm should also demonstrate their scalability under different networks. In this subsection, we utilize the 4 networks shown in Table 4.7 to test the scalability of the proposed methods. In the MBTA literature, given that the number of user classes ranges from several to 439 (De Cea et al., 2005), we adopt a middle ground of 50 user classes here. The results are shown in Fig. 4.12.

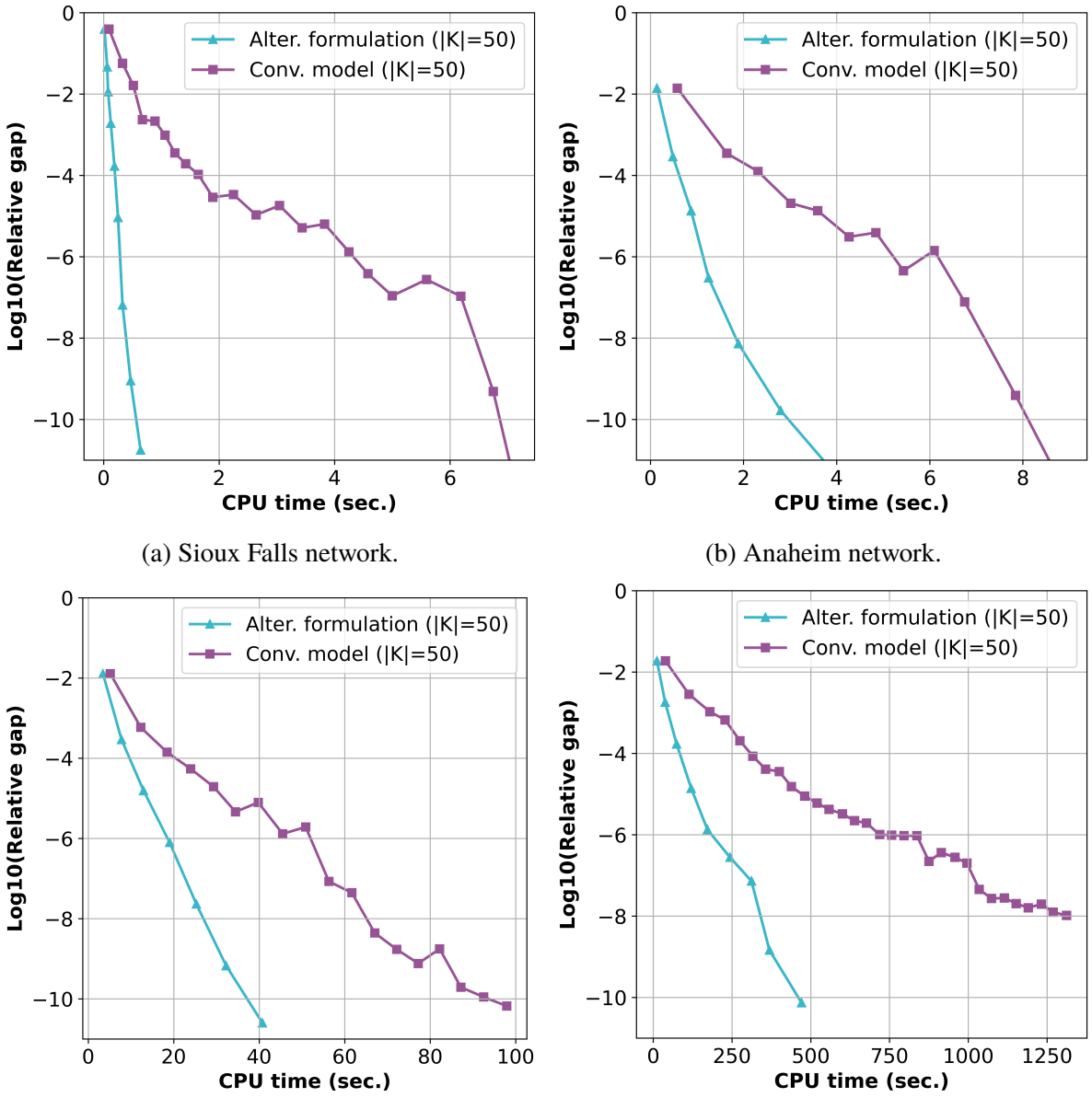
As shown in Fig. 4.12, in general, as the network size increases, the required computation time increases. The alternative formulation and the SGP algorithm outperform the conventional MBTA model in all the networks. It is worth mentioning that a relative gap of 10^{-4} is already enough to get a stable equilibrium flow for practical applications(Boyce et al., 2004), which means the path-based SGP algorithm only needs no more than 5 iterations in all the test networks. Although we do not want to repeatedly show the memory consumption as in Table 4.8, it is fair enough to claim that the alternative formulation and the path-based SGP algorithm advance the computation of the MBTA problem from both speed and memory significantly, especially when the number of user classes is large.

4.4.2.3 Sensitivity analyses

We further utilize the Winnipeg network to conduct sensitivity analysis on key inputs/parameters to examine the performance of the alternative MBTA formulation and the path-based SGP algorithm. Same as the last subsection, we fix the number of user classes as 50 (i.e., $|K| = 50$).

From the demand side, we test the impact of congestion level. We multiply all the OD demands by a coefficient ranging from 0.2, 0.4, 0.6, ..., to 1.8 to set different congestion levels. The results of the proposed methods are shown in Fig. 4.13a. It can be seen that when the congestion level is low (i.e., ≤ 0.4), the SGP algorithm only needs a few iterations to converge. When the congestion level increases, as expected, more iterations are needed to reach the predefined precision, resulting in a longer CPU time. In addition, the CPU time is approximately proportional to the congestion level.

From the network side, we test the impact of toll magnitude. We set the link tolls as 0.0



(a) Sioux Falls network.

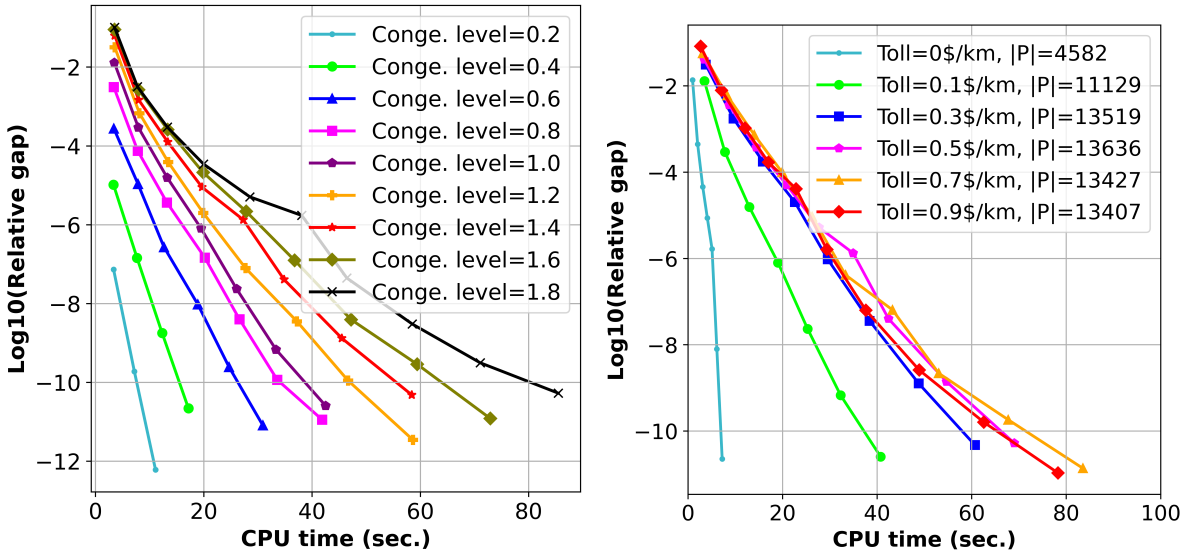
(b) Anaheim network.

(c) Winnipeg network.

(d) Chicago Sketch network.

Figure 4.12: The convergence curves of the two models under different networks.

$\$/km$ (i.e., no toll), 0.1 $\$/km$, 0.3 $\$/km$, 0.5 $\$/km$, 0.7 $\$/km$, and 0.9 $\$/km$ on the top 10% congested links under the user equilibrium flow and 0 on other links. The results are shown in Fig. 4.13b. The results indicate that the toll magnitude obviously impacts the algorithm's convergence. When there is no toll, equivalent to the standard user equilibrium case, the SGP algorithm takes about 7 seconds to converge. When the toll increases to 0.1 $\$/km$, the CPU time increases significantly (about 6 times) because travelers start to face the time-toll trade-off. But after tolls reach 0.3 $\$/km$, the toll increase has a minor effect on the CPU time. To see the



(a) Convergence curves of the SGP algorithm under different congestion levels.

(b) Convergence curves of the SGP algorithm under different toll magnitudes.

Figure 4.13: Sensitivity analyses of key inputs/parameters on the Winnipeg network.

reason behind this, we also include the path set size $|P|$ as part of the legend in Fig. 4.13b. It shows that after tolls reach 0.3 $\$/km$, the path set size tends to be stable despite the toll increase. Therefore, the CPU times needed for equilibrating the path sets are also similar.

4.4.2.4 Implementation discussion

The experiments in this section demonstrate that the proposed alternative formulation and path-based SGP algorithm have obvious speed and memory advantages over the conventional MBTA model and algorithm (see Table 4.8, Fig. 4.11, and Fig. 4.12). Therefore, it can be an alternative option to be included as a new tool in the existing traffic assignment software/packages. Meanwhile, it should also be noted that the MBTA problem is much more computationally intensive compared to the standard single criterion traffic assignment problem. As shown in Figure 4.13b, the Toll=0.1 $/km case requires 7 times more CPU time than the Toll=0 $/km case. Especially in practice, it is usually megacities (i.e., huge traffic networks) suffering from traffic congestion that need to design road tolls to regulate the traffic flows, where the CPU time for solving the MBTA problem is expected to be much longer. To further speed up the proposed methods, practitioners can consider combining the proposed methods with other techniques, such as parallel computing (Chen et al., 2020) and network partition (Jafari et al., 2017), to gain additional savings.

4.5 Summary

In this chapter, we examined the characteristics of the multiclass bi-criteria traffic assignment (MBTA) problem and proposed an alternative formulation. While the conventional MBTA model duplicates the link/path flow variables to track the class-specific information, the alternative formulation does not need to keep class-specific variables by well exploiting the order information of paths and travelers. That is, *low-TEM (high-VOT) travelers will prefer fast but expensive paths, while high-TEM (low-VOT) travelers will prefer slow but cheap paths.* This chapter proved the equivalence between the alternative formulation and the MBTA problem and also found out that the alternative formulation has a similar solution property as the conventional MBTA model.

Further, to deal with the nondifferentiable convex objective function, this chapter developed a path-based subgradient projection (SGP) algorithm to solve the alternative formulation. The path-based SGP algorithm utilized the subgradient and available second-order information to update the boundary variables by quasi-Newton's method, which shows high efficiency in the numerical experiments.

In the numerical experiments, a small network and several large networks were employed to examine the features of the alternative formulation and the path-based SGP algorithm. The results showed that although the alternative formulation does not contain class-specific link/path flow variables, it provides the same link flows and can further extract the same level of class-specific link/path flows as the conventional MBTA model. However, compared to the conventional MBTA model, the alternative formulation keeps much fewer variables, making the path-based SGP algorithm superior in computation speed and dominant in memory consumption when dealing with practical transportation networks and a large number of user classes.

This work also opens many other directions that would be interesting to explore. First, following many previous works (Dial, 1979, 1996, 1997, etc.), this chapter also adopts the assumption that travelers value time and toll linearly. Nevertheless, some researchers, e.g., Wang and Ehrgott (2013), argued that the nonlinear valuation of time and toll could be more realistic. Therefore, future research could consider the nonlinear trade-off between time and toll using the proposed methods. Second, this chapter only considers two criteria, i.e., time and toll, in the path choice behavior. Actually, travelers' path choice behaviors are affected by many criteria, such as travel

time, travel time reliability, and monetary cost (Abdel-Aty et al., 1995). Therefore, it would be meaningful to extend the proposed methodology to consider more criteria, so as to develop more realistic models for traffic assignment. Third, one of the criteria (i.e., toll) in this chapter is flow-independent. It would be challenging but rewarding to relax this assumption, which allows the practitioners to model more complex route choice behaviors of travelers in urban transportation networks. Last but not least, as this chapter utilized order information to improve the computation of the static MBTA problem, future research could consider exploring and utilizing the order information in the bi-criteria dynamic traffic assignment problem (Lu et al., 2008, 2009).

Chapter 5

Continuous multiclass mean-risk traffic assignment problem: both criteria are flow-dependent

This chapter generalizes the single-class problem of Nikolova and Stier-Moses (2014) ["A Mean-Risk Model for the Traffic Assignment Problem with Stochastic Travel Times." Operations Research 62(2): 366-382] by considering the continuously distributed risk-aversion factor, termed the *continuous mean-risk traffic assignment (CMRTA)* problem. In CMRTA, travelers categorized into infinitely many user classes play a congestion routing game based on their risk attitude toward the mean-standard deviation trade-off, where both criteria are flow-dependent. We identify and verify the following finite-dimensional complementarity condition: All used paths must have an equal and minimum "virtual" cost, called the cumulative path cost, which enables the characterization of Wardrop's equilibrium for CMRTA. Based on this result, we present general formulations to model CMRTA problems with exogenous and endogenous variability sources. A tailored path-based column generation with the gradient projection algorithm is then proposed to solve large-scale problems. The results show that CMRTA is more computationally efficient and less memory-intensive than the traditional discrete multiclass method.

5.1 Introduction

The user equilibirum (UE) is a state in which no traveler can decrease their travel cost by unilaterally switching to an alternative route in transportation networks. In this paradigm, travelers are modeled as rational agents who strategically select routes that minimize their travel cost based on traffic conditions. Although conventional UE models use deterministic travel times (Wardrop, 1952), real-world transportation networks are prone to uncertainty. Sources of travel time uncertainty can be classified as *exogenous* (e.g., weather and construction) or *endogenous* to traffic flows (e.g., congestion-induced delays). Both uncertainty types are prevalent in the real world and markedly influence travelers' routing behaviors (Nikolova and Stier-Moses, 2014).

To measure a route's travel time reliability, researchers have widely adopted a method that uses the linear combination of the mean and standard deviation (stdev) of path travel times through a user-specific *risk-aversion factor* (RAF, denoted by γ), i.e., mean + $\gamma\times$ stdev. The stochastic travel time mean and stdev are generally obtained by determining how uncertainties from different sources propagate along the route. The RAF, known as the *reliability ratio* (Jackson and Jucker, 1982; Zang et al., 2022), is the ratio of the value associated with reducing travel time variability (value of reliability) to the value associated with reducing travel time (value of time). Note that the RAF is not fixed but varies by individual. A higher RAF indicates greater risk aversion toward travel time uncertainty.

Empirical studies have shown that the RAF is affected by various factors, such as socioeconomics, trip purposes, and personality traits, and that it exhibits significant heterogeneity across populations (Small et al., 2005; Bhat and Sardesai, 2006). To capture this heterogeneity, many studies have employed discrete choice models to analyze mean-risk trade-offs through stated preference and/or revealed preference data (Carrion and Levinson, 2012). The mixed logit model (McFadden and Train, 2000), which is particularly powerful for capturing taste heterogeneity through random parameters, has been widely applied to estimate the continuous distributions of risk preferences (Liu et al., 2007; Li et al., 2010; Hossan et al., 2016).

Moreover, an alternative discrete distribution can be adopted by grouping users with similar, identifiable characteristics (e.g., male/female (Lam and Small, 2001) and commuting/non-commuting trips (de Dios Ortúzar and Willumsen, 2011). However, for more abstract characteristics such as the RAF, the discretization method leads to arbitrary choices in both group

thresholds and group numbers, which reduces its effectiveness in capturing user heterogeneity (Hossan et al., 2016). For example, a population with RAFs ranging from 0 to 2 can be discretely segmented into low (0-1) and high (1-2) categories. Between groups, it characterizes someone with an RAF of 0.9 identically to someone with an RAF of 0.1 but differently from someone with an RAF of 1.1, even though 0.9 is closer to 1.1. Thus, within every group, taking a representative RAF value (e.g., 0.5 for low) may force users to select the same path, even though their actual RAFs may produce different path choices.

To achieve a more tractable equilibrium analysis, previous studies have focused on a discrete distribution, either using a single-value RAF (Prakash et al., 2018) or a discrete RAF set (Nie, 2011). This discrete method is a simplification of the continuous method, but this also raises the question of how many discrete classes are needed to adequately capture variations across a continuous distribution. Researchers and practitioners have considered this question in terms of the value of time in a tolled network via transportation planning software (EMME and TransCAD). Specifically, Rezaee and Hickman (2015) found that 9 classes were sufficient to generate a stable link flow pattern on a synthetic network with 15 regular links and 2 tolled links. Moreover, Slavin et al. (2014) found that even 20 classes yielded unsatisfactory toll revenues compared with a 100-class instance on a small network with 7 regular links and 2 tolled links. As discussed in Section 5.5, discretization errors can be significant enough to substantially affect optimal operational planning decisions. These issues highlight the potential bias in discrete representations and the inherent difficulty in predetermining a suitable class number.

In sum, using a continuous RAF distribution is more suitable for equilibrium analysis. Although a fine-grained discretization method can represent user heterogeneity, every additional class potentially requires the computation and storage of a separate set of class-specific paths and flows. This issue creates significant computational costs when there are many user classes (see Section 5.2.2). In contrast, computing the equilibrium under continuous distributions avoids the overhead of class discretization while maintaining solution quality for general networks.

This chapter extends the single-class model of Nikolova and Stier-Moses (2014) by considering continuously distributed RAFs, termed the *continuous mean-risk traffic assignment (CM-RTA)* problem. While Nikolova and Stier-Moses (2014) concluded that their results could readily accommodate a discrete set of RAFs, the adoption of a continuous distribution creates significant methodological and computational difficulties because it may involve infinitely many user

classes. Motivated by these observations, we propose general mathematical formulations and a tailored solution algorithm for addressing CMRTA problems. The main contributions of this chapter are as follows.

We identify and verify the following finite-dimensional complementarity condition: All used paths connecting the same origin-destination (OD) pair must have an equal and minimum "virtual" cost—termed the *cumulative path cost*—which equivalently characterizes UE for CMRTA. The cumulative path cost enables a geometric interpretation from the mean-stdev Pareto frontier and enables route choice analysis across infinitely many user classes. The importance of this finding extends beyond CMRTA, offering an effective tool for analyzing bi-objective congestion routing games with continuous preference distributions.

Next, we present general mathematical formulations for both exogenous and endogenous CMRTA problems based on the finite-dimensional complementarity condition. These methods mainly have two significant advantages over existing methods. In one sense, the proposed models operate in the space of total path flows and maintain a structure analogous to that of single-class problems in Nikolova and Stier-Moses (2014) while avoiding the dimensional expansion often required for discrete multiclass analysis. In another sense, both the mean and stdev criteria can be flow-dependent, extending beyond previous studies that have typically employed at least one flow-independent criterion, such as travel toll (Leurent, 1993) or travel distance (Xie et al., 2017). These advances lead to a more generalizable modeling framework with dimensional efficiency and flexible applicability.

The third contribution concerns how to solve CMRTA problems. A core difficulty, as identified by Nikolova and Stier-Moses (2014), is in computing equilibrium under mean-stdev objectives owing to the nonadditive path cost structure, which invalidates traditional Dijkstra-like path identification algorithms (Dijkstra, 1959; Sedeño-Noda and Raith, 2015). This complexity is significantly amplified in CMRTA because path costs are defined for infinitely many classes. In addition, existing algorithms for equilibrating path flows often depend on the fixed path order by one flow-independent criterion (Xie et al., 2021; Xu et al., 2024b). However, the path ranking in the endogenous CMRTA problem inherently depends on flow distributions, which renders existing methods inapplicable.

Therefore, we develop a tailored path-based *column generation with the gradient projection (CG-GP)* algorithm for handling large-scale networks. The algorithm employs an innovative

two-stage column generation procedure to identify all efficient paths with nonadditive costs along the Pareto frontier. To achieve equilibrium, we apply the quasi-Newton method to equalize cumulative path costs between paired paths, where these costs are obtained from the Pareto frontier.

Finally, we perform comprehensive numerical experiments to demonstrate the effect of continuously distributed RAFs on network flows and computational efficiency. Using the Braess network, we visualize the iterative evolution of the Pareto frontier and compare equilibrium solutions between CMRTA models and previous studies. In real-world networks, the CG-GP algorithm achieves high-quality solutions with modest computational times. Our results evidence that CMRTA outperforms conventional discrete multiclass methods in computational efficiency and memory utilization. Moreover, we show that the discretization method may still underrepresent the network's path diversity, even when many user classes are considered.

The rest of the chapter is organized as follows: Section 5.2 introduces the multiclass mean-risk traffic assignment problem and details the potential limitations of employing a discrete distribution of RAFs. Section 5.3 presents the formulations for CMRTA, and the solution algorithm is given in Section 5.4. Numerical results are reported in Section 5.5 and Section 5.6 concludes the chapter.

5.2 Preliminaries

We consider a transportation network represented by a directed graph $G(N, A)$, where N and A are the set of nodes and links, respectively. The network comprises a set of OD pairs, $W \subseteq N \times N$. Every OD pair $w \in W$ is associated with the travel demand q_w and a set of simple paths K_w. Travelers are regarded as players in a congestion game to minimize their experienced path costs, making route choice decisions that result in a path flow vector $\mathbf{f} = (f_k^w)_{k \in K_w, w \in W}$. The feasible set of $\mathbf{f}$ can be written as $\Omega = \{\mathbf{f} : f_k^w \geq 0 \text{ for all } k \in K_w, w \in W \text{ and } \sum_{k \in K_w} f_k^w = q_w \text{ for all } w \in W\}$. The total flow on link a is the summation over all traversed paths, i.e., $x_a(\mathbf{f}) = \sum_{w \in W} \sum_{k \in K_w} \delta_{ak} f_k^w$, where $\delta_{ak} = 1$ if link a is included by path k, and $\delta_{ak} = 0$ otherwise. Let $\mathbf{x} = (x_a)_{a \in A}$ be the vector of link flows.

Under network uncertainty, the stochastic travel time of a link $a \in A$ is measured by $\tilde{t}_a(x_a) = e_a(x_a) + \epsilon_a(x_a)$, where $e_a(\cdot)$ is the mean travel time, and $\epsilon_a(\cdot)$ is the noise travel time. We assume

that $e_a(x_a)$ is a continuous and increasing function of x_a and that $\epsilon_a(x_a)$ has zero expectation. Moreover, the stdev is denoted by $s_a(x_a)$. Two types of congestion games can be defined based on the functional settings of $s_a(x_a)$:

(i) Exogenous variability: $s_a(x_a)$ is flow-independent and given exogenously as a constant (also simply written as s_a in this case). It relates to external factors that are independent of traffic flows, such as the weather, construction, and events (Ordónez and Stier-Moses, 2010).

(ii) Endogenous variability: $s_a(x_a)$ is a continuous and increasing function of x_a, making it endogenous to the congested network. This scenario is more general than the exogenous case for modeling congestion effects (van Lint et al., 2008).

Supposing that link travel times are independent of each other, the mean and stdev of path travel times are expressed as follows:

$$\mu_k^w(\mathbf{f}) = \sum_{a\in A} \delta_{ak} \cdot e_a\left(x_a(\mathbf{f})\right), \tag{5.1}$$

$$\sigma_k^w(\mathbf{f}) = \sqrt{\sum_{a\in A} \delta_{ak} \cdot s_a^2\left(x_a(\mathbf{f})\right)}. \tag{5.2}$$

The path cost is defined as the linear combination of the mean-stdev objectives through the users' RAF (denoted by γ). That is,

$$C_k^w(\mathbf{f};\gamma) = \mu_k^w(\mathbf{f}) + \gamma \cdot \sigma_k^w(\mathbf{f}). \tag{5.3}$$

Equation (5.3) reflects the nonadditive cost structure because the stdev along a path is not separable across traversed links (Gabriel and Bernstein, 1997). Nikolova and Stier-Moses (2014) concluded that accounting for travelers' heterogeneous risk attitudes can be readily achieved by using a discrete distribution of γ. Based on the discretization method, we consider its application to the discrete mean-risk traffic assignment (DMRTA) problem. The following subsections characterize the equilibrium condition of DMRTA and detail potential limitations using an example.

5.2.1 Equilibrium characterization for DMRTA

Next, we introduce more notations to define the DMRTA problem with a finite number of user classes. We denote M_w as the discrete set of user classes for OD pair w. Every class $m \in M_w = \{1, 2, \cdots, |M_w|\}$ is associated with a distinct class-specific RAF γ_m and a divided sub-demand q_{wm}, with the users' cost objective given by $\mu_k^w + \gamma_m \cdot \sigma_k^w$. The demand conservation is ensured by $\sum_{m \in M_w} q_{wm} = q_w$. We then define f_k^{wm} as the class-specific path flow contributed by class m. Therefore, the total flow on path k is $f_k^w = \sum_{m \in M_w} f_k^{wm}$. Let $\check{\mathbf{f}} = (f_k^{wm})_{k \in K_w, m \in M_w, w \in W}$ denote the class-specific path flow vector, and its feasible set can be expressed as $\check{\Omega} = \{\check{\mathbf{f}} : f_k^{wm} \geq 0$ for all $k \in K_w, m \in M_w, w \in W$ and $\sum_{k \in K_w} f_k^{wm} = q_{wm}$ for all $m \in M_w, w \in W\}$. The UE condition for accommodating a discrete number of user classes is as follows:

Definition 5.1. *A class-specific path flow vector $\check{\mathbf{f}} \in \check{\Omega}$ is at UE for DMRTA when $C_k^w(\check{\mathbf{f}}; \gamma_m) \leq C_{k'}^w(\check{\mathbf{f}}; \gamma_m)$ for all $k,\ k' \in K_w$, $m \in M_w$, $w \in W$ with $f_k^{wm} > 0$.*

The complementarity condition is widely known to play a key role in characterizing UE, underlying both its formulations and algorithms. Building upon Definition 5.1, we define the complementarity condition to capture the relationship between the path flow pattern and the users' objective of route choice.

Proposition 5.1. *A class-specific path flow vector $\check{\mathbf{f}} \in \check{\Omega}$ of DMRTA is at UE by Definition 5.1 if and only if it satisfies the following complementarity condition:*

$$C_k^w\left(\check{\mathbf{f}}; \gamma_m\right) - u_{wm} \quad \geq 0, \tag{5.4a}$$

$$f_k^{wm} \cdot \left[C_k^w\left(\check{\mathbf{f}}; \gamma_m\right) - u_{wm}\right] \quad = 0, \tag{5.4b}$$

for all $k \in K_w$, $m \in M_w$ and $w \in W$, where u_{wm} denotes the minimal class-specific OD cost.

Proposition 5.1 establishes that for a given class m of OD pair w, the used paths ($f_k^{wm} > 0$) have the equal cost u_{wm}, while paths with higher costs remain unused, thus aligning with Definition 5.1. Based on the above result, DMRTA can be formulated as a variational inequality (VI)—that is, finding $\check{\mathbf{f}}^* \in \check{\Omega}$ such that

$$\sum_{w \in W} \sum_{m \in M_w} \sum_{k \in K_w} C_k^w(\check{\mathbf{f}}^*; \gamma_m) \cdot (f_{km}^w - f_{km}^{w*}) \geq 0, \quad \forall \check{\mathbf{f}} \in \check{\Omega}. \tag{5.5}$$

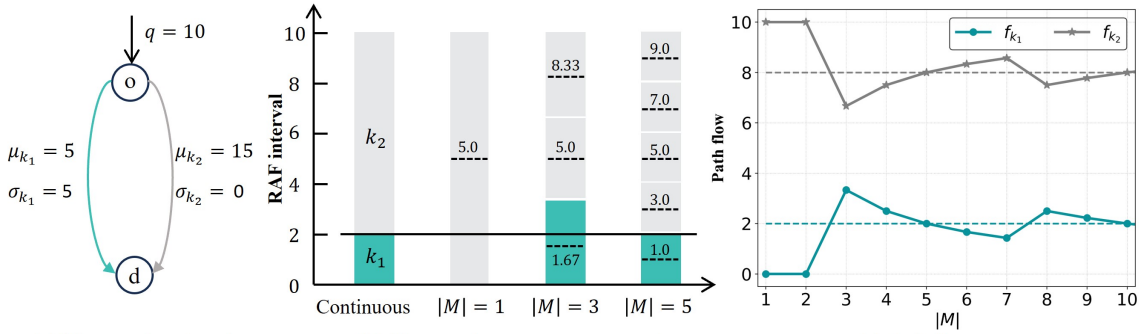
The solution existence for VI (5.5) is guaranteed given that (i) the class-specific path cost is continuous, and (ii) the feasible set $\check{\Omega}$ is compact and convex (Hartman and Stampacchia, 1966). For the exogenous DMRTA problem with constant s_a, we can reformulate VI (5.5) as a convex mathematical program, whose first-order optimality conditions match Condition (5.4). In contrast, the endogenous DMRTA problem with flow-dependent $s_a(x_a)$ may prevent such a transformation because of the asymmetric Jacobian of the path cost vector (see Proposition 5.2 in Nikolova and Stier-Moses, 2014). Regarding uniqueness, $\mathbf{x}^*$ is uniquely determined in the exogenous model, while no such guarantee can be made for the endogenous model.

5.2.2 Discretization issues

This section employs a two-path network (see Figure 5.1(a)) to demonstrate the limitations of DMRTA regarding discretization errors and computational burden. For simplicity, the OD index w is omitted. Suppose that the RAF values range from 0 to 10, where users with RAFs in $[0, 2]$ prefer path k_1, while those in $[2, 10]$ prefer path k_2. Therefore, RAF = 2 is the "*breakpoint*" where optimal route choice changes. To construct DMRTA instances, we discretize the $[0, 10]$ range into equally sized intervals, using every interval's conditional expectation as the class-specific RAF.

Figures 5.1(b) and 5.1(c) illustrate the path flow assignment under a uniform distribution of RAFs. The single-class instance ($|M| = 1$) seriously underrepresents user heterogeneity, thus leaving path k_1 unused. However, the three-class instance ($|M| = 3$) overestimates flows on path k_1. The case of $|M| = 5$ matches the continuous result only because its first interval coincidentally aligns with $[0, 2]$. We can observe that path flow errors oscillate between negative and positive values, potentially skewing network performance evaluations. Although the path flow error is zero at $|M| = 5$, increasing the class number instead introduces new errors. This unpredictability makes it difficult to predetermine an optimal class number in congested networks, where preferred path choice intervals are not known in advance.

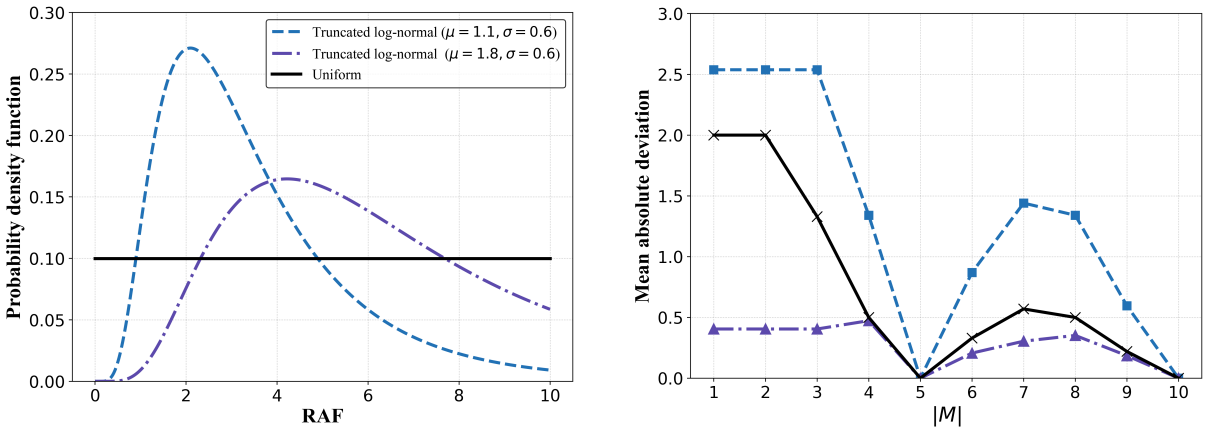
As depicted in Figure 5.2, we further evaluate the potential impacts of discretization under different RAF distributions. The truncated log-normal distribution $Logn(\mu = 1.8, \sigma = 0.6)$, with more users being highly risk-averse, produces the smallest, most stable errors. As most users have high RAF values far from the breakpoint (RAF = 2), their route choices on k_2 remain the same across different discrete groupings. In contrast, the distribution $Logn(\mu = 1.1, \sigma =$



(a) Two-path network (b) Flow assignment on two paths (c) Path flow errors

Figure 5.1: Path flow errors by discretization.

0.6), characterized by a high density of low-RAF users near the breakpoint of 2, yields larger flow errors. When $|M|$ is small, many users near the breakpoint are misclassified, which causes more flow assignment errors. The uniform distribution produces moderate errors owing to the evenly distributed RAF. These results indicate that discretization errors may be driven by the distribution shape and the alignment of its density with the breakpoints.

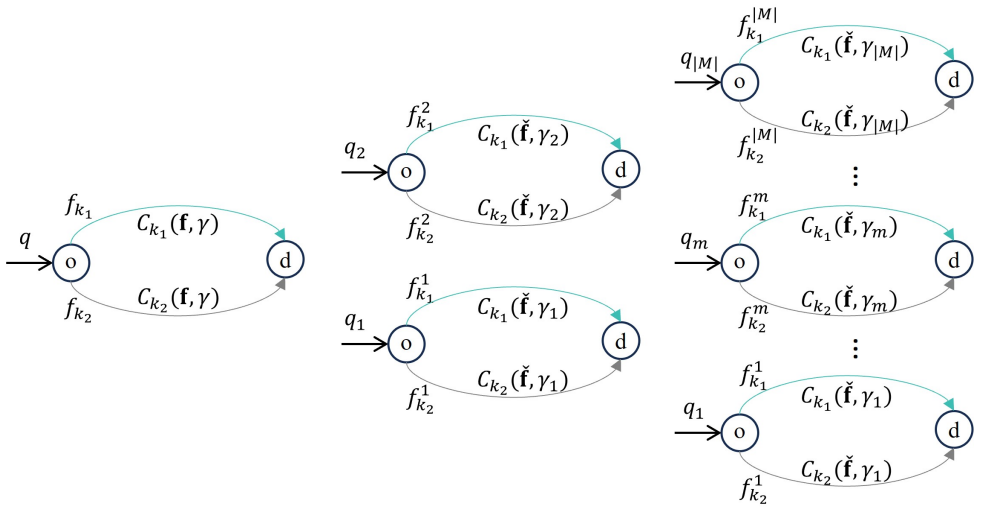


(a) Three continuous distributions (b) Mean absolute deviation of path flow

Figure 5.2: Path flow errors under different continuous distributions of RAFs.

Although the aforementioned results show that finer discretization generally reduces errors, this method also significantly increases computational costs. Figure 5.3 shows how DMRTA's network representation expands across three instances (1, 2, and $|M|$ classes). The single-class case only needs to adjust the total flows on two paths to achieve equilibrium. However, as more user classes are introduced, each requires independent flow equilibration across two paths, given its specific sub-demand and cost objective. This requirement can give rise to a linear increase in the number of OD pairs with every extra class. The computational implications of this expansion

are twofold: (i) It necessitates greater storage resources to manage the growing set of class-specific paths, and (ii) it may increase the runtime for path manipulations. Even with a relatively modest increase in $|M|$, the computational burden can become intensive in general networks, which often contain exponentially many paths and a multitude of OD pairs.



(a) Single class (b) Two classes (c) $|M|$ classes

Figure 5.3: Computational burden by discretization.

5.3 Equilibrium characterization for CMRTA

This section addresses the equilibrium characterization for the CMRTA problem, which inherently accommodates infinitely many differentiated classes. We consider the RAF between an OD pair $w \in W$ to follow a continuous probability distribution on a compact set $\Upsilon_w = [\gamma_{lb}^w, \gamma_{ub}^w]$, where $0 \leq \gamma_{lb}^w < \gamma_{ub}^w$. In what follows, we first present an infinite-dimensional VI based on the direct extension from Proposition 5.1, and then we introduce a new finite-dimensional equilibrium condition, along with mathematical formulations.

5.3.1 Infinite-dimensional characterization

Let $d_w(\gamma)$ and $D_w(\gamma)$ denote the probability density function and cumulative distribution function of γ between OD pair w, respectively. Assume that $d_w(\gamma)$ is positive in its domain. We introduce $\breve{f}_k^w(\gamma)$ as the (class-specific) path flow density with RAF γ using path k between OD pair w. Thus, the total flow on path k is $f_k^w = \int_{\gamma_{lb}^w}^{\gamma_{ub}^w} \breve{f}_k^w(\gamma) d\gamma$. We then denote the infinite-dimensional

path flow density vector $\breve{\mathbf{f}} = (\breve{f}_k^w(\gamma))_{\gamma\in\Upsilon_w,k\in K_w,w\in W}$, which is constrained by the set $\breve{\Omega} = \{\breve{\mathbf{f}} : \breve{f}_k^w(\gamma) \geq 0 \text{ for all } \gamma \in \Upsilon_w, k \in K_w, w \in W \text{ and } \sum_{k\in K_w} \breve{f}_k^w(\gamma) = q_w \cdot d_w(\gamma) \text{ for all } \gamma \in \Upsilon_w, w \in W\}$. The UE and its complementarity condition can be restated below.

Definition 5.2. *A path flow density vector $\breve{\mathbf{f}} \in \breve{\Omega}$ is at UE for CMRTA when $C_k^w(\breve{\mathbf{f}};\gamma) \leq C_{k'}^w(\breve{\mathbf{f}};\gamma)$ for all $k,\ k' \in K_w$, $\gamma \in \Upsilon_w$, $w \in W$ with $\breve{f}_k^w(\gamma) > 0$.*

Proposition 5.2. *A path flow density vector $\breve{\mathbf{f}} \in \breve{\Omega}$ of CMRTA is at UE by Definition 5.2 if and only if it satisfies the following complementarity condition:*

$$C_k^w\left(\breve{\mathbf{f}};\gamma\right) - u_w(\gamma) \quad \geq 0, \tag{5.6a}$$

$$\breve{f}_k^w(\gamma) \cdot \left[C_k^w\left(\breve{\mathbf{f}};\gamma\right) - u_w(\gamma)\right] \quad = 0, \tag{5.6b}$$

for all $k \in K_w$, $\gamma \in \Upsilon_w$ and $w \in W$, where $u_w(\gamma)$ is the minimum cost for users with γ from OD pair w.

That is, equilibrium is attained when for every possible value of γ, the flow density $\breve{f}_k^w(\gamma)$ is assigned to path k with minimum costs connecting OD pair w. Based on Proposition 5.2, we can formulate the CMRTA problem as an infinite-dimensional VI: finding $\breve{\mathbf{f}}^* \in \breve{\Omega}$ such that

$$\sum_{w\in W}\sum_{k\in K_w}\int_{\gamma_{lb}^w}^{\gamma_{ub}^w} C_k^w\left(\breve{\mathbf{f}}^*;\gamma\right)\cdot\left[\breve{f}_k^w(\gamma) - \breve{f}_k^{w*}(\gamma)\right]d\gamma \geq 0, \quad \forall \breve{\mathbf{f}} \in \breve{\Omega}. \tag{5.7}$$

We can establish equivalence between Condition (5.6) and VI (5.7) by applying similar arguments as used in Dial (1996) and Marcotte and Zhu (1996). The key difference in formulation between CMRTA and DMRTA is their handling of γ. For every $w \in W$, VI (5.7) integrates over a continuous interval Υ_w, while VI (5.5) sums over discrete classes in M_w. This indicates that VI (5.5) is simply a discretized approximation of VI (5.7). However, the infinite-dimensional nature poses significant theoretical challenges and computational difficulties in both analysis and implementation. We now attempt to propose a finite-dimensional method that equivalently characterizes the equilibrium of CMRTA.

5.3.2 Finite-dimensional characterization

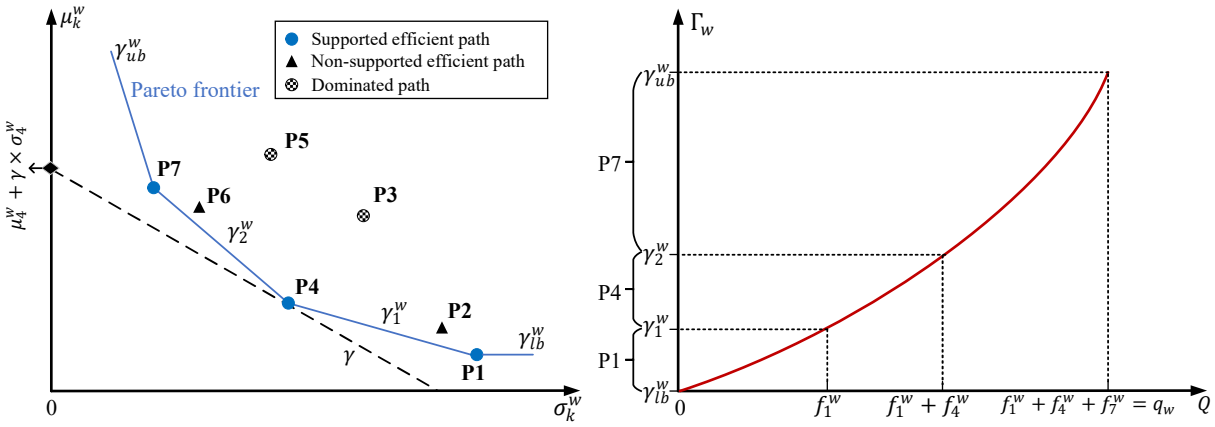
To realize finite-dimensional characterization, we leverage the Pareto frontier defined by mean-stdev efficient paths. We first consider the route choice under fixed μ_k^w and σ_k^w for all paths. We say a path $k \in K_w$ is efficient if there does not exist another path $k' \neq k$ such that (i) $\mu_{k'}^w \leq \mu_k^w$, (ii) $\sigma_{k'}^w \leq \sigma_k^w$, and (iii) at least one inequality in (i) or (ii) strictly holds. As an illustration, consider a network with seven paths between a single OD pair. As depicted in Figure 5.4(a), P3 and P5 are dominated by superior alternatives (e.g., P4 dominates P3 in both objectives). The remaining five efficient paths can be further categorized into two types: *supported* or *non-supported* (Ehrgott, 2005). To obtain the path cost incurred by a γ-user, one can analyze the intercepts on y-axis of the tangent line with a negative slope γ (e.g., dashed line for P4). Three supported efficient paths—namely P1, P4, and P7—form the Pareto frontier (solid blue line). Every path is optimal for a specific RAF interval: Users prefer P1 for $\gamma \in [\gamma_{lb}^w, \gamma_1^w]$, P4 for $\gamma \in [\gamma_1^w, \gamma_2^w]$, and P7 for $\gamma \in [\gamma_2^w, \gamma_{ub}^w]$. Therefore, the values γ_1^w and γ_2^w are breakpoints where users change their optimal route choices.

We now clarify the connection between path flows and potential breakpoints in general networks. Note that the Pareto frontier suggests less risk-averse users prefer paths with lower μ_k^w but higher σ_k^w. The ordered path set $\hat{K}_w$ is arranged in nonincreasing order of stdev such that $\sigma_1^w \geq \cdots \geq \sigma_k^w \geq \cdots \geq \sigma_{|\hat{K}_w|}^w \geq \sigma_{|\hat{K}|+1}^w = 0$, where $\sigma_{|\hat{K}_w|+1}^w$ is introduced for convenience. Hereafter, we use the path set $\hat{K}_w = \{1, \ldots, |\hat{K}_w|\}$, instead of K_w, when order information is needed. Let $U_w = q_w \cdot D_w(\gamma)$ and the inverse of U_w is defined as $\Gamma_w = U_w^{-1} : Q \to \gamma$. We refer to $\Gamma_w(\cdot)$ as the cumulative flow function, which maps cumulative path flows $Q \in [0, q_w]$ to the corresponding RAF breakpoint $\gamma \in [\gamma_{lb}^w, \gamma_{ub}^w]$. In our example (see Figure 5.4(b)), $\gamma_1^w = \Gamma_w(f_1^w)$ and $\gamma_2^w = \Gamma_w(f_1^w + f_4^w)$, which indicates that users in $[\gamma_1^w, \gamma_2^w]$ select P4.

For any $w \in W$, we define the cumulative flow Q_k^w for any path with index k as

$$Q_k^w(\mathbf{f}) = \begin{cases} 0 & \text{if } k = 0, \\ \sum_{i=1}^{k} f_k^w & \text{if } k \in \hat{K}_w, \end{cases} \tag{5.8}$$

where $k = 0$ is included to represent the initial case. For every $k \in \{0\} \cup \hat{K}_w$, we denote $\hat{\gamma}_k^w(\mathbf{f}) = \Gamma_w(Q_k^w(\mathbf{f}))$ as the RAF breakpoint for index k. Hence, $\hat{\gamma}_k^w(0) = \gamma_{lb}^w$ and $\hat{\gamma}_k^w(q_w) = \gamma_{ub}^w$.



(a) Supported and non-supported efficient paths (b) Flow assignment on three supported efficient paths

Figure 5.4: Route choice based on the Pareto frontier.

Given the constant values of μ_k^w and σ_k^w for all paths, the flow distribution among supported efficient paths can be readily identified by the Pareto frontier. However, the problem becomes considerably more complex when the mean and stdev objectives are inherently coupled with users' route choices in a congested network. As such, a supported efficient path may lose its efficiency because of the flow redistribution, and vice versa. The complexity is particularly high in the endogenous case, where the path order may vary with path flow patterns. We now turn to the CMRTA problem, with its UE condition restated below.

Definition 5.3. *A path flow vector* $\mathbf{f}$ *is at UE for CMRTA when* $C_k^w(\mathbf{f};\gamma) \leq C_{k'}^w(\mathbf{f};\gamma)$ *for all* $k, k' \in \hat{K}_w, \gamma \in [\hat{\gamma}_{k-1}^w(\mathbf{f}), \hat{\gamma}_k^w(\mathbf{f})] \neq \emptyset, w \in W$ *with* $f_k^w > 0$.

The aforementioned definition indicates that at equilibrium, a minimum-cost path k with positive flows is optimal for users whose RAFs are in the non-empty interval $[\hat{\gamma}_{k-1}^w(\mathbf{f}), \hat{\gamma}_k^w(\mathbf{f})]$. Definition 5.3 improves on Definitions 5.1 and 5.2 by intrinsically grouping users with identical path choices, thus eliminating the need for class-specific path flows. However, different γ-users within the same group still incur disparate path costs even if they choose the same path. Thus, deriving equivalent complementarity conditions for Definition 5.3 is not as straightforward as for the other two definitions. Importantly, the following result enables the UE condition to be characterized with finite dimensions.

Proposition 5.3. *A path flow vector* $\mathbf{f} \in \Omega$ *of CMRTA is at UE by Definition 5.3 if and only if it*

satisfies the following complementarity condition:

$$H_k^w(\mathbf{f}) - h_w \geq 0, \tag{5.9a}$$

$$f_k^w \cdot [H_k^w(\mathbf{f}) - h_w] = 0, \tag{5.9b}$$

for all $k \in \hat{K}_w, w \in W$*, where* h_w *is the minimal OD cost, and* $H_k^w(\mathbf{f})$ *denotes the cumulative path cost by*

$$H_k^w(\mathbf{f}) = \mu_k^w(\mathbf{f}) + \sum_{i=k}^{|\hat{K}_w|} \hat{\gamma}_i^w(\mathbf{f}) \cdot \left[\sigma_i^w(\mathbf{f}) - \sigma_{i+1}^w(\mathbf{f})\right]. \tag{5.10}$$

Proposition 5.3 indicates that all used paths from the same OD pair w share equal and minimum cumulative costs h_w, while paths with higher cumulative costs remain unused. The proof to establish the equivalence between Definition 5.3 and Condition (5.9) is given in Appendix C.1.

Figure 5.5 depicts the geometric interpretation of the cumulative path cost based on the Pareto frontier. This example consists of two efficient, used paths (P1 and P3) and one dominated, unused path (P2). As per Equation (5.10), the cumulative cost of a path k is the sum of its initial vertical quantity of mean travel time and a sequence of vertical increases by the negative slopes of tangent lines between adjacent paths. We can easily verify that the cumulative costs on P1 and P3 are identical, while P2, which is above the Pareto frontier, incurs a higher cumulative cost by an excessive quantity Δ. Therefore, Proposition 5.3 effectively captures how heterogeneous users' route choice behaviors collectively induce the equilibrium path flow pattern.

With Condition (5.9), we now present equivalent formulations for both exogenous and endogenous CMRTA problems, with detailed proofs given in Appendix C.2. The following result shows that the exogenous problem can be characterized by a mathematical program.

Proposition 5.4 (Exogenous model)**.** *A vector* $\mathbf{f}^* \in \Omega$ *is an equilibrium path flow of the exogenous CMRTA problem with constant stdev if and only if it is the optimal solution to the following program:*

$$\min_{\mathbf{f}\in\Omega} \quad Z = \sum_{a\in A} \int_0^{x_a(\mathbf{f})} e_a(z)dz + \sum_{w\in W} \sum_{k\in\hat{K}_w} \sigma_k^w \int_{Q_{k-1}^w(\mathbf{f})}^{Q_k^w(\mathbf{f})} \Gamma_w(z)dz. \tag{5.11}$$

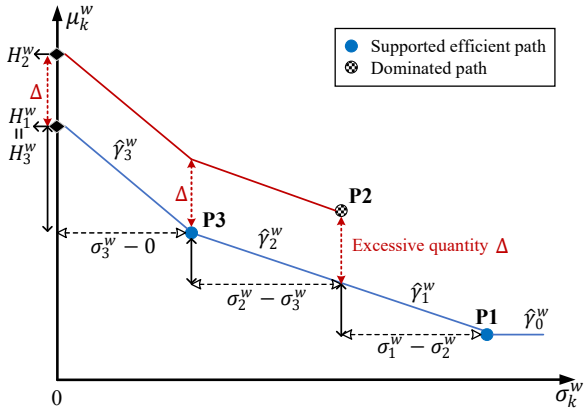


Figure 5.5: Geometric interpretation of cumulative path costs. Note that $\hat{\gamma}_1^w = \hat{\gamma}_2^w$ because $f_2^w = 0$.

The first term on the right-hand side of Equation (5.11) represents the Beckmann term associated with mean travel times. The second term employs integration to represent the total stdev of path travel times. The objective function Z is continuously convex with respect to $\mathbf{f}$ because functions $e_a(\cdot)$ and $\Gamma_w(\cdot)$ are continuous and increasing. Given that the feasible set Ω of path flows is compact and convex, Problem (5.11) is convex, thus ensuring a global minimum.

In terms of solution properties, the polyhedral structure of Ω guarantees the existence of a solution. As Problem (5.11) is strictly convex to link flows, the uniqueness of $\mathbf{x}^*$ still holds. This result further leads to the unique Pareto frontier for every OD pair at UE, as stated below.

Theorem 5.1. *The equilibrium Pareto frontier of every OD pair w is unique for Problem* (5.11).

Proof. Given unique $\mathbf{x}^*$, the mean path travel times for all paths are also unique. Accordingly, the Pareto frontier must be unique because the coordinate (σ_k^w, μ_k^w) of every path $k \in K_w$ is uniquely determined. □

Generally, the path flow solution $\mathbf{f}^*$ is not unique because multiple path flow solutions can yield the same link flow solution $\mathbf{x}^*$. However, the following result demonstrates the uniqueness of $\mathbf{f}^*$ in CMRTA when exogenous σ_k^w values are unique within every OD pair.

Corollary 5.1.1. *If the path order strictly decreases by the stdev of path travel times for any $w \in W$, every path's coordinate (σ_k^w, μ_k^w) of that OD pair remains distinct on the Pareto frontier, and then the path flow solution $\mathbf{f}^*$ is unique, with travel demand being assigned to a unique set of supported efficient paths.*

We next turn to the endogenous CMRTA problem, which is stated as an equivalent VI as follows.

Proposition 5.5 (Endogenous model)**.** *A vector* $\mathbf{f}^* \in \Omega$ *is an equilibrium path flow of the endogenous CMRTA problem with flow-dependent stdev if and only if it satisfies the following VI:*

$$\mathbf{H}(\mathbf{f}^*) \cdot (\mathbf{f} - \mathbf{f}^*) \geq 0, \quad \forall \mathbf{f} \in \Omega, \tag{5.12}$$

where $\mathbf{H}(\mathbf{f}^*) = (H_k^w(\mathbf{f}^*))_{k \in \hat{K}_w, w \in W}$ *denotes the vector of cumulative path costs under* $\mathbf{f}^*$.

Because Ω is compact and convex, and the cumulative cost mapping $\mathbf{H}$ is continuous, VI (5.12) guarantees the existence of a solution (Hartman and Stampacchia, 1966). The uniqueness of a VI arises from a monotone mapping, which generally does not hold in an endogenous CMRTA problem.

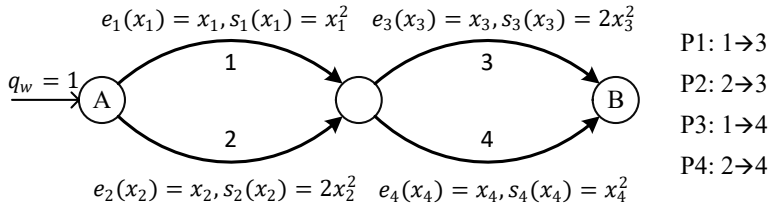
Proposition 5.6. *The mapping* $\mathbf{H}$ *of VI* (5.12) *is not monotone:* $\exists\, \mathbf{f}_1, \mathbf{f}_2 \in \Omega$, $[\mathbf{H}(\mathbf{f}_1) - \mathbf{H}(\mathbf{f}_2)] \cdot (\mathbf{f}_1 - \mathbf{f}_2) < 0$.

Proof. Consider a single OD pair with four links and four paths, as shown in Figure 5.6. Suppose that $q_w = 1$ and $\Gamma_w(Q) = 2Q$. Given that $\mathbf{f}_1 = (0.3, 0.4, 0.3, 0)$ and $\mathbf{f}_2 = (0.6, 0.1, 0.1, 0.2)$, we have $\mathbf{H}(\mathbf{f}_1) = (2.9738, 2.7659, 1.6422, 1.3648)$ and $\mathbf{H}(\mathbf{f}_2) = (3.0942, 2.5751, 1.8776, 1.0025)$ as per Equation (5.10). This violates monotonicity because $[\mathbf{H}(\mathbf{f}_1) - \mathbf{H}(\mathbf{f}_2)] \cdot (\mathbf{f}_1 - \mathbf{f}_2) = -0.0262 < 0$. □

Proposition 5.7. *The mapping* $\mathbf{H}$ *of VI* (5.12) *is not pseudo-monotone:* $\exists\, \mathbf{f}_1, \mathbf{f}_2 \in \Omega$, $\mathbf{H}(\mathbf{f}_2) \cdot (\mathbf{f}_1 - \mathbf{f}_2) \geq 0$ *while* $\mathbf{H}(\mathbf{f}_1) \cdot (\mathbf{f}_1 - \mathbf{f}_2) < 0$.

Proof. Consider the path flow patterns in Figure 5.6 again, we have $\mathbf{H}(\mathbf{f}_2) \cdot (\mathbf{f}_1 - \mathbf{f}_2) = 0.0193 \geq 0$ and $\mathbf{H}(\mathbf{f}_1) \cdot (\mathbf{f}_1 - \mathbf{f}_2) = -0.0069 < 0$. □

The non-monotonicity property arises from the non-convexity of the square-root function, along with the flow intersections on common links by multiple paths. As such, the endogenous CMRTA model may be unable to produce unique link flow solutions, which aligns with the observations for the single-class DMRTA problem in Nikolova and Stier-Moses (2014).



	f_k^w	μ_k^w	σ_k^w	Order
P1	0.3	1.3	1.0440	1
P2	0.4	1.1	1.0309	2
P3	0.3	0.9	0.3711	3
P4	0	0.7	0.3324	4

$\mathbf{f}_1$

	f_k^w	μ_k^w	σ_k^w	Order
P1	0.6	1.4	1.0957	1
P2	0.1	1.0	0.9964	2
P3	0.1	1.0	0.4982	3
P4	0.2	0.6	0.2012	4

$\mathbf{f}_2$

Figure 5.6: Counterexample to monotonicity and pseudo-monotonicity.

5.4 Algorithm

When solving CMRTA problems in large-scale networks, the central difficulties are handling the exponential number of paths and performing path flow adjustments to achieve UE by Condition (5.9). To this end, we outline a path-based CG-GP algorithm, which allows us to split the original problem into a pricing problem and a restricted master problem at every major iteration. The former identifies new potential improving columns to expand the variable space, while the latter optimizes within a subset of the variables. The full CG-GP procedure is detailed below.

Algorithm 6 The CG-GP algorithm for solving CMRTA problems

Input: Network $G(N, A)$, OD demand vector $\mathbf{q} = (q_w)_{w \in W}$, link performance functions $(e_a(\cdot))_{a \in A}$ and $(s_a(\cdot))_{a \in A}$.
Output: Equilibrium path flow $\mathbf{f}^*$.

1: **Initialization.** Initialize counter $n = 0$, assign total demand to path k with minimum mean cost under the zero-flow network and set the working path set $\mathcal{K}_w = \{k\}$ for any $w \in W$.
2: **Solving column generation problem.** Set $n = n + 1$ and add new paths to $\mathcal{K}_w$ for each $w \in W$.
3: **Termination criterion.** Stop the algorithm if the *outer convergence* is achieved and no new column is identified.
4: **Solving restricted master problem.** Initialize the inner counter $l = 0$. (i) Set $l = l + 1$; (ii) perform flow equilibrations among paths contained in the current working set for each OD pair; and (iii) stop it if *inner convergence* is achieved, otherwise return to (i).
5: **Dropping.** Remove unused paths (columns) for each OD pair, and return to Step 2.

The following sections detail the procedures for column generation (Section 5.4.1) and the restricted master problem(Section 5.4.2), followed by termination criteria in Section 5.4.3, and the convergence proof of CG-GP in Section 5.4.4.

5.4.1 Procedure of column generation problem

Column generation is widely adopted in traffic assignment algorithms to avoid path enumeration. However, a notable distinction of CMRTA exists involving heterogeneous risk-averse users who may show preferences for different paths. This behavior contrasts with conventional methods that seek the shortest path for homogeneous users, as the objective here is to identify multiple paths and collectively minimize expected total costs. Specifically, fixing $\mu_k^w(\mathbf{f}^n)$ and $\sigma_k^w(\mathbf{f}^n)$ for all paths induced by $\mathbf{f}^n$ at iteration n, we can express the pricing problem as follows:

$$\min_{\mathbf{f}\in\Omega} \quad \bar{Z} = \sum_{w\in W}\sum_{k\in\hat{K}_w}\left(\mu_k^w(\mathbf{f}^n)\cdot f_k + \int_{Q_{k-1}^w(\mathbf{f})}^{Q_k^w(\mathbf{f})} \sigma_k^w(\mathbf{f}^n)\cdot\Gamma_w(z)dz\right). \tag{5.13}$$

Although Problem (5.13) can be decomposable by the OD pair, solving it remains difficult because of the exponential number of paths. The non-separability of path costs also presents challenges when directly employing existing Dijkstra-like algorithms. The idea of our method involves initially finding a mean-stdev efficient path set $\tilde{K}_w$ under $\mathbf{f}^n$ and then identifying a supported efficient path set $\bar{\mathcal{K}}_w(\subseteq \tilde{K}_w)$. The former task can be completed by running the labeling mean-variance bi-objective algorithm thanks to the additive structure of μ_k^w and $(\sigma_k^w)^2$ (see Appendix C.3.1). As the square-root function is monotonically increasing, it preserves path dominance relationships in terms of μ_k^w and σ_k^w.

The next task is to find a subset of support efficient paths on the Pareto frontier. As shown in Figure 5.4(a), starting from and ending at specific paths, two critical conditions exist: (a) The negative slope of the tangent line between adjacent supported efficient paths must lie within the support interval $(\gamma_{lb}^w, \gamma_{ub}^w)$, and (b) the consecutive negative slopes of the tangent line are strictly increasing. Appendix C.3.2 provides an illustrative example of the iterative process and these two conditions. The sequential procedure to generate new paths for a single OD pair is detailed below.

The time complexity of Algorithm 7 is $O(|\tilde{K}_w|\cdot\log(|\tilde{K}_w|))$ because of the sorting step. Specifically, when the lower-bound user is risk-neutral ($\gamma_{lb}^w = 0$), the starting path is the first path in $\tilde{K}_w$ with the lowest mean but the largest stdev. Conversely, for the upper-bound user being infinitely risk-averse ($\gamma_{ub}^w \to \infty$), the ending path is the last path in $\tilde{K}_w$ with the lowest stdev but the highest mean. As all efficient paths can be found simultaneously by using the one-to-all mean-variance bi-objective algorithm once, new paths are generated sequentially for OD

Algorithm 7 Column generation algorithm for an OD pair.

Input: Efficient path set $\tilde{K}_w$ using the mean-variance bi-objective algorithm.
Output: Expanded working path set $\mathcal{K}_w$.

1: **Initialization.** Initialize path set $\bar{\mathcal{K}}_w = \emptyset$, a RAF slope set $\bar{\Upsilon}_w = \emptyset$.
2: **Sorting.** Arrange $\tilde{K}_w$ in descending order of σ_k^w.
3: **Starting and ending paths setup.** By sequentially scanning each path in $\tilde{K}_w$, identify starting path $\underline{k}$ with shortest costs $\mu_{\underline{k}}^w(\mathbf{f}^n) + \gamma_{lb}^w \cdot \sigma_{\underline{k}}^w(\mathbf{f}^n)$ and the maximum index, ending path $\overline{k}$ with shortest costs $\mu_{\overline{k}}^w(\mathbf{f}^n) + \gamma_{ub}^w \cdot \sigma_{\overline{k}}^w(\mathbf{f}^n)$ and the minimum index. Set $\bar{\mathcal{K}}_w = \{\underline{k}\}$ and $\bar{\Upsilon}_w = \{\gamma_{lb}^w\}$.
4: **Sequential slope verification.** Initialize $k = \underline{k} + 1$. (i) Compute the negative tangent $\gamma_k^w = (\mu_k^w(\mathbf{f}^n) - \mu_{k'}^w(\mathbf{f}^n))/(\sigma_{k'}^w(\mathbf{f}^n) - \sigma_k^w(\mathbf{f}^n))$, where k' is the last element of $\bar{\mathcal{K}}_w$. (ii) If $\gamma_k^w \in (\gamma_{lb}^w, \gamma_{ub}^w)$, then repeat to pop back $\bar{\mathcal{K}}_w$ and $\bar{\Upsilon}_w$ when $\gamma_k^w \leq \gamma_{k'}^w$ (the last element in $\bar{\Upsilon}_w$), update γ_k^w between k and the last element in $\bar{\mathcal{K}}_w$ until $\gamma_k^w > \gamma_{k'}^w$, append k to $\bar{\mathcal{K}}_w$ and γ_k^w to $\bar{\Upsilon}_w$. (iii) Stop if $k = \overline{k}$ and proceed to Step 5; otherwise update $k = k + 1$ and return to (i).
5: **Adding.** Update $\mathcal{K}_w = \mathcal{K}_w \cup \bar{\mathcal{K}}_w$ with $f_k^w = 0$ for any new path $k \in \bar{\mathcal{K}}_w \setminus \mathcal{K}_w$.

pairs sharing the same origin.

5.4.2 Procedure of restricted master problem

We next turn to solving the restricted master problem defined on the working path set. Our basic idea is to decompose it into a sequence of single-variate paired-path problems between every OD pair while holding flows on other paths constant. Given an OD pair w, two major concerns then arise: (i) how to choose paired paths from the working path set $\mathcal{K}_w$, and (ii) how to adjust flows between paired paths? Here, we focus on a given OD pair and omit the subscript w for simplicity.

For Question (i), we select the shortest and longest paths that have the largest gap of cumulative costs, denoted by $\bar{k}$ and k, respectively. Flows are expected to shift from k to $\bar{k}$ to equalize their cumulative cost. This method is inspired by Dafermos and Sparrow (1969) for the standard traffic assignment problem, and it has been recently used in acyclic bush-based algorithms (Dial, 2006; Xu et al., 2022). To ensure convergence, we define an *effective* paired-path problem as follows:

Definition 5.4. *A paired-path problem formed by $\bar{k}$ and k is said to be effective if (i) $f_k > 0$ and (ii) $\tilde{H}_{\bar{k}k} = H_k - H_{\bar{k}} > 0$.*

Conditions (i) and (ii) in Definition 5.4 ensure that both the flow shift and descent are

bounded away from zero, indicating the potential for improving the solution at non-equilibrium. At the inner iteration l of outer iteration n, let $\phi_{\bar{k}k} = f_{\bar{k}}^{(nl)} + f_k^{(nl)}$ denote total flows on paired paths and is treated as constant. Accordingly, the flow on path $\bar{k}$ can be expressed as $f_{\bar{k}} = \phi_{\bar{k}k} - f_k$, and the feasible set for f_k is given by $\Omega_{\bar{k}k} = [0, \phi_{\bar{k}k}]$. With these definitions, the paired-path problems decomposed from the original exogenous and endogenous models are detailed below.

The paired-path problem with exogenous stdev is formulated as follows:

$$\min_{f_k \in \Omega_{\bar{k}k}} \quad Z = \sum_{a \in A_k \cup A_{\bar{k}}} \int_0^{x_a(f_k)} e_a(z)dz + \sum_{i=\min\{k,\bar{k}\}}^{\max\{k,\bar{k}\}-1} \sigma_i \int_{Q_{i-1}(f_k)}^{Q_i(f_k)} \Gamma(z)dz + \tilde{Z}, \quad (5.14)$$

where A_k is the set of links on path k, $x_a(f_k) = \tilde{x}_a + \delta_{ak} f_k + \delta_{a\bar{k}} \cdot (\phi_{\bar{k}k} - f_k)$, $\tilde{x}_a$ is the constant link flows contributed by other paths, and $\tilde{Z}$ denotes the constant part decomposed from Equation (5.11) by unaffected link flows and cumulative path flows.

The paired-path problem with endogenous stdev is expressed as the following VI:

$$\tilde{H}_{\bar{k}k}(f_k^*) \cdot (f_k - f_k^*) \geq 0, \quad \forall f_k \in \Omega_{\bar{k}k}, \quad (5.15)$$

where

$$\tilde{H}_{\bar{k}k}(f_k) = \begin{cases} \mu_k(f_k) - \mu_{\bar{k}}(f_k) - \sum_{i=\bar{k}}^{k-1} \hat{\gamma}_i(f_k) \cdot [\sigma_i(f_k) - \sigma_{i+1}(f_k)] & \text{if } k > \bar{k}, \\ \mu_k(f_k) - \mu_{\bar{k}}(f_k) + \sum_{i=k}^{\bar{k}-1} \hat{\gamma}_i(f_k) \cdot [\sigma_i(f_k) - \sigma_{i+1}(f_k)] & \text{if } \bar{k} > k. \end{cases} \quad (5.16)$$

Note that the first-order derivative of the objective function (5.14) to f_k also yields the cumulative cost difference defined in Equation (5.16). Once the paired-path problem is constructed, we proceed to solve Question (ii) based on the gradient projection algorithm. The algorithm updates f_k by

$$f_k^{(r+1)} = \min\left\{\max\left\{f_k^{(r)} - \alpha^{(r)} \cdot \tilde{H}_{\bar{k}k}(f_k^{(r)}), 0\right\}, \phi_{\bar{k}k}\right\}, \quad (5.17)$$

where r is the iteration counter, and $\alpha^{(r)}$ is a positive stepsize determined by the line search at iteration r.

We use the Barzilai–Borwein stepsize to adaptively approximate the Hessian using two suc-

cessive gradients (Barzilai and Borwein, 1988):

$$\alpha_{\text{BB}}^{(r)} = \left(f_k^{(r)} - f_k^{(r-1)}\right) / \left(\tilde{H}_{\bar{k}k}(f_k^{(r)}) - \tilde{H}_{\bar{k}k}(f_k^{(r-1)})\right). \tag{5.18}$$

This adaptive stepsize scheme is suitable for nonadditive CMRTA problems, especially in the endogenous case, to avoid computing complex second-order derivatives (Du et al., 2021). To guarantee convergence, we also integrate a backtracking technique to adjust the stepsize when necessary. Specifically, let $\alpha^{(r)} = \alpha_{\text{BB}}^{(r)} \cdot \nu^{\eta}$, where ν is the backtracking ratio (set as 1/2), and η denotes the smallest non-negative integer that satisfies a certain condition. For Problem (5.14), the condition can be defined by the well-known Armijo rule to ensure a sufficient decrease in the objective function value (Armijo, 1966). For VI (5.15), we simply impose a strict decreasing condition for the cumulative cost difference. We can now present the procedure for solving an effective paired-path problem.

Algorithm 8 Gradient projection algorithm for an effective paired-path problem.

Input: A working path set $\mathcal{K}_w$, initial path flows $\{f_k^{(0)} : k \in \mathcal{K}_w\}$.
Output: Update path flows f_k and $f_{\bar{k}}$.

1: **Initialization**: Update mean and stdev for each path $k \in \mathcal{K}_w$, select paired-paths $\bar{k}$ and k, set $r = 0$, initialize $\alpha^{(0)} = 0.1$.
2: **Descent direction**: Update the cumulative cost difference $\tilde{H}_{\bar{k}k}(f_k^{(r)})$ based on Eq. (5.16).
3: **Termination check**: If $\tilde{H}_k(f_k^{(r)}) = 0$ or $(\tilde{H}_k(f_k^{(r)}) > 0$ and $f_k^{(r)} = 0)$, stop the algorithm; otherwise move to Step 4.
4: **Line search with backtracking**: Update $\alpha^{(r)}$ based on Eq. (5.18) and backtrack if necessary.
5: **Update**: (i) Update $f_k^{(r)}$ as per Eq. (5.17), $f_{\bar{k}}^{(r)} = \phi_{\bar{k}k} - f_k^{(r)}$; (ii) update $e_a(x_a)$ and $s_a(x_a)$ for each link a on path k and $\bar{k}$; (iii) and set $r = r + 1$. Return to Step 2.

For every inner iteration l, we perform the above procedure sequentially for all OD pairs. As captured in Step 3, the algorithm stops when either the cumulative cost difference reaches zero or the longest path k still incurs higher costs, even after all flows have been shifted to the shortest path $\bar{k}$.

5.4.3 Termination criterion

To evaluate the solution quality, we use the relative gap to assess the proximity of the current solution to UE. For the outer loop given $\mathbf{f}^n$, we define

$$RG_{outer} = 1 - \frac{\sum_{w \in W} q_w \cdot h_w(\mathbf{f}^n)}{\sum_{w \in W} \sum_{k \in \mathcal{K}_w} f_k^w \cdot H_k^w(\mathbf{f}^n)}, \tag{5.19}$$

where $h_w(\mathbf{f}^n)$ is the minimum cumulative cost obtained by column generation at iteration n.

For the inner loop given $\mathbf{f}^l$, the relative gap is rewritten as follows:

$$RG_{inner} = 1 - \frac{\sum_{w \in W} \sum_{k \in \mathcal{K}_w} q_w \cdot H_{min}^w(\mathbf{f}^l)}{\sum_{w \in W} \sum_{k \in \mathcal{K}_w} f_k^w \cdot H_k^w(\mathbf{f}^l)}, \tag{5.20}$$

where $H_{min}^w(\mathbf{f}^l) = \min_{k \in \mathcal{K}_w} H_k^w(\mathbf{f}^l)$ is the minimum cumulative cost at inner iteration l.

Even though both relative gaps share similar functional forms, they show distinct differences yet preserve close relationships. Specifically, RG_{outer} determines convergence across the entire original variable set Ω by column generation, while RG_{inner} determines convergence within the restricted working path set. Hence, the CG-GP algorithm terminates when $RG_{outer} = RG_{inner} \leq \epsilon$ and no new columns are identified to further improve the solution.

5.4.4 Convergence proof

We establish that Algorithm 6 (CG-GP) converges to the optimal solution satisfying UE condition (5.9) concerning the exogenous Problem (5.11). In the endogenous scenario, establishing convergence poses challenges because of the non-monotonic nature. In such instances, CG-GP may be viewed as a heuristic. Recall that Algorithm 6 has two building blocks: Algorithm 7 identifies new variables with the potential to improve the current solution, while Algorithm 8 solves a restricted version of the original problem using a subset of variables. To prove the overall convergence, we first demonstrate that Algorithm 8 converges to the optimal solution of the restricted master problem (5.14). Then, we show that Algorithm 6 can converge to the globally optimal solution of Problem (5.11) with a finite number of iterations by incorporating Algorithm 7.

Theorem 5.2 (Convergence theorem (Zangwill, 1969))**.** *Let the point-to-set map $E : \Phi \to \Phi$ determine an algorithm that, given a point $\mathbf{y}^0 \in \Phi$, generates the sequence $\{\mathbf{y}^j\}_{j=1}^{\infty}$. In addition,*

let a solution set $\Phi^* \subseteq \Phi$ be given. Suppose that: (a) All points $\mathbf{y}^j$ are in a compact set $S \subseteq \Phi$; (b) there is a continuous function $z : \Phi \to \Re$ such that (i) if $\mathbf{y}$ is not a solution, then, for any $\bar{\mathbf{y}} \in E(\mathbf{y})$, $z(\bar{\mathbf{y}}) < z(\mathbf{y})$, and (ii) if $\mathbf{y}$ is a solution, then either the algorithm terminates or, for any $\bar{\mathbf{y}} \in E(\mathbf{y})$, $z(\bar{\mathbf{y}}) \leq z(\mathbf{y})$; and (c) the map E is closed at any point $\mathbf{y} \in \Phi \setminus \Phi^*$. Then, every limit point of $\{\mathbf{y}^j\}$ belongs to the solution set Φ^*.

We begin by formalizing the algorithmic map used to solve Problem (5.14). Let Ω_n denote the feasible set based on all working paths at outer iteration n. The gradient projection algorithm for solving Problem (5.14) can be defined as a point-to-set mapping $E : \Omega_n \to \Omega_n$, which produces a sequence $\{\mathbf{f}^l\}$ according to the composition of the OD-based point-to-set mappings $E = E_1 \circ E_2 \circ \cdots \circ E_{|W|}$. For every map $E_w : \Omega_{\bar{k}k} \to \Omega_{\bar{k}k}$, Algorithm 8 updates path flows with $\left(f_k^{(r+1)}\right) = E_w\left(f_k^{(r)}\right)$ as per Equation (5.17). By verifying conditions (a)–(c) of Theorem 5.2, we can formally state the convergence of Algorithm 8 below.

Theorem 5.3. *For the restricted master problem at outer iteration n, given any sequence $\{\mathbf{f}^l\}$ such that $\mathbf{f}^0 \in \Omega_n$ and $\mathbf{f}^{l+1} \in E(\mathbf{f}^l)$, every limit point of this sequence converges to the equilibrium.*

Proof. We first verify Conditions (a)-(c) in Zangwill's theorem concerning the map E_w for a single OD pair w. Consequently, we omit the index w for notional convenience.

For Condition (a), the feasible domain of f_k is $\Omega_{\bar{k}k} = [0, \phi_{\bar{k}k}]$, which is a compact set. The update rule by Eq. (5.17) ensures that the new point $f_k^{(r+1)}$ is always contained within the compact set $\Omega_{\bar{k}k}$. Hence, Condition (a) is satisfied because all points $f_k^{(r)}$ are in a compact set.

For condition (b), we consider two feasible points $f_{k[1]}$ and $f_{k[2]}$ such that $f_{k[2]} \in E_w(f_{k[1]})$. Suppose that $f_{k[1]}$ is not a solution to the paired-path problem, then the derivative $\tilde{H}_k(f_{k[1]}) \neq 0$ and $f_{k[1]} > 0$. The BB stepsize scheme with the Amoij rule ensures the strict reduction on the objective function value, i.e., $Z(f_{k[2]}) < Z(f_{k[1]})$. If $f_{k[1]}$ is a solution, then Algorithm 8 must induce a zero-flow shift because either the two paths achieve zero cost difference ($\tilde{H}_k(f_{k[1]}) = 0$), or path k is unused with a higher cost ($\tilde{H}_k(f_{k[1]}) > 0$ and $f_{k[1]} = 0$). The algorithm is terminated (as per Step 3 in Algorithm 8) and yields $Z(f_{k[2]}) = Z(f_{k[1]})$. It follows that $Z(f_{k[2]}) \leq Z(f_{k[1]})$, with equality holds only when $f_{k[1]}$ is a solution. This verifies Condition (b) to show the monotonically decreasing of the sequence until the algorithm is terminated.

For Condition (c) to show E_w is closed, we need to prove that at any non-optimal point

$f^*_{k[1]} \in \Omega_{\bar{k}k}$ if for any two sequences $\{f^{(r)}_{k[1]}\}$ and $\{f^{(r)}_{k[2]}\}$ satisfying $\lim_{r\to\infty}\{f^{(r)}_{k[1]}\} = f^*_{k[1]}$, $f^{(r)}_{k[2]} \in E\left(f^{(r)}_{k[1]}\right)$ for any r and $\lim_{r\to\infty}\{f^{(r)}_{k[2]}\} = f^*_{k[2]}$, then it follows that $f^*_{k[2]} \in E\left(f^*_{k[1]}\right)$ (Bazaraa et al., 2013). According to the update rule by Eq. (5.17), E_w is a continuous function of f_k because (i) $\tilde{H}_{\bar{k}k}(f_k)$ is continuous, (ii) the stepsize $\alpha^{(r)}$ is well-defined for each iteration r, and (iii) the min and max operators still preserve continuity regarding a continuous function and the constant (i.e., 0 and $\phi_{k\bar{k}}$). Given that $f^{(r)}_{k[2]} \in E_w(f^{(r)}_{k[1]})$ for all r, and E_w is a point-to-set map, we have $\lim_{r\to\infty}\{f^{(r)}_{k[2]}\} \subseteq \lim_{r\to\infty}\{E_w(f^{(r)}_{k[1]})\} = E_w(f^*_{k[1]})$, where the set limit is taken in the sense of Painlevé-Kuratowski convergence. Since $\lim_{r\to\infty}\{f^{(r)}_{k[2]}\} = f^*_{k[2]}$ is a singleton set by assumption, it leads to $f^*_{k[2]} \in E_w(f^*_{k[1]})$, which satisfies the condition for closeness.

Given that the algorithm map E is the composition of E_w for all $w \in W$, it follows from Zangwill's theorem that any sequence $\{\mathbf{f}^l\}$ generated by E converges to $\mathbf{f}^*_n$, which corresponds to the equilibrium solution for the restricted master problem at outer iteration n. □

We are now ready to present the overall convergence proof of CG-GP for solving Problem (5.11) by incorporating column generation (Algorithm 7).

Theorem 5.4. *The sequence of* $\{\mathbf{f}^*_n\}$ *generated by Algorithm 6 converges to equilibrium for Problem* (5.11) *after a finite number of iterations.*

Proof. Let $\mathbf{f}^*$ denote the optimality that satisfies the UE condition (5.9) regarding Problem (5.11), and $\mathcal{K}^n_w$ the working path set at iteration n. Suppose that $Z(\mathbf{f}^*_n) > Z(\mathbf{f}^*)$ at iteration n, the algorithm moves to $n+1$ to yield $\mathcal{K}^{n+1}_w = \mathcal{K}^n_{w+} \cup \bar{\mathcal{K}}^{n+1}_w$, where $\mathcal{K}^n_{w+} = \{k \mid f^{n*}_k > 0, k \in \mathcal{K}^n_w\}$ is the set of used working paths after accurately solving the restricted master problem, and $\bar{\mathcal{K}}^{n+1}_w$ is the set of generated new paths by Algorithm 7. It's worth emphasizing that $\cup_{w\in W}\bar{\mathcal{K}}^{n+1}_w \neq \emptyset$ since currently there must exists promising new columns outside $\mathcal{K}^n_{w+}$. Given that solving restricted master problem decreases the objective function value by invoking Theorem 5.3, it follows that $Z(\mathbf{f}^*_{n+1} \mid \cup_{w\in W}\mathcal{K}^{n+1}_w) < Z(\mathbf{f}^*_n \mid \cup_{w\in W}\mathcal{K}^n_w) < \cdots < Z(\mathbf{f}^*_1 \mid \cup_{w\in W}\mathcal{K}^1_w)$. Recall that the objective $Z(\mathbf{f})$ is convex to induce a unique minimum for a given working path set, it yields $\cup_{w\in W}\mathcal{K}^{n+1}_w \nsubseteq \cup_{w\in W}\mathcal{K}^j_w$ for any $j = 1, \cdots, n$, ensuring the algorithm explores a distinct set of working paths at each iteration. We further introduce B_w as the set of all possible OD-based column combinations (a subset of K_w), then the set of all possible column combinations across all OD pairs can be defined by $B = \prod_{w\in W} B_w$, where $|B| = \prod_{w\in W}(2^{|K_w|} - 1)$ is finite. Because (i) the objective of Problem (5.11) is bounded below by its optimal value $Z(\mathbf{f}^*)$ and (ii)

the maximum number of iterations is bounded by $|B|$, it leads to $\{Z(\mathbf{f}_n^*)\} \to Z(\mathbf{f}^*)$ after a finite number of iterations. □

5.5 Results

This section presents the numerical analysis by applying the CG-GP algorithm to solve exogenous and endogenous CMRTA problems. We also compare the results for continuously and discretely distributed RAFs. To construct a DMRTA problem, we generate random samples from the continuous RAF distribution and divide them into $|M|$ groups using equally spaced percentile cutoffs. The conditional expectation within every percentile group is set as the class-specific RAF value. As Nikolova and Stier-Moses (2014) did not propose an algorithm for the DMRTA problem, we extend the path-based CG-GP algorithmic framework to solve it. The essential idea is to iteratively update the working path set for every predetermined user class and then perform class-specific path flow adjustments for equilibration. The detailed procedure and an analysis of its differences and complexity from Algorithm 6 are given in Appendix C.4.

All algorithms are implemented in C++, with shared code wherever possible to achieve a fair computational performance comparison. Moreover, the termination criteria for both RG_{inner} and RG_{outer} are set to 1e-10. All results are generated by a workstation with an Intel Core i7-11700 processor running at 3.4 GHz and 32 GB RAM. The next two sections separately analyze the results for small and large networks.

5.5.1 Braess network

Figure 5.7 depicts the Braess network, which contains four nodes and five links. Its only OD pair (A, D) is assumed to have 80 trips and to be connected by three paths (P1, P2, and P3). We follow the approach of Lo et al. (2006) to evaluate the stochastic link capacity degradation due to day-to-day events, such as traffic jams and accidents. The link capacity is assumed to follow a uniform distribution ranging from a value below the nominal capacity (c_a)—multiplied by a degradation factor (η_a)—up to the nominal capacity itself. We employ the widely used BPR function (Bureau of Public Roads, 1964) to model link travel times, with the mean and variance

expressed as follows:

$$
\begin{aligned}
e_a(x_a) &= t_a^0 + \frac{\alpha t_a^0 x_a^\beta (1-\eta^{1-\beta})}{c_a^\beta (1-\eta_a)(1-\beta)}, \\
s_a^2(x_a) &= \alpha^2 (t_a^0)^2 x_a^{2\beta} \cdot \left(\frac{1-\eta_a^{1-2\beta}}{c_a^{2\beta}(1-\eta_a)(1-2\beta)} - \left(\frac{1-\eta_a^{1-\beta}}{c_a^\beta (1-\eta_a)(1-\beta)} \right)^2 \right),
\end{aligned}
$$

where t_a^0 is the free-flow travel time, $\alpha = 0.15$ and $\beta = 4$ are the parameters of the BPR function, and the degradation factor η_a is set as 0.5.

For the exogenous model, the constant stdev of travel times for five links are 8, 0, 4, 6, and 0 (mins). To facilitate comparison, we restrict the flow-dependent stdev to Link 2 and Link 5 in the endogenous model. Therefore, P2 maintains a constant stdev. The RAFs follow a truncated log-normal distribution $Logn(\mu = 1, \sigma = 1)$ such that $\gamma_{lb}^w = 0$ and $\gamma_{ub}^w = 2$.

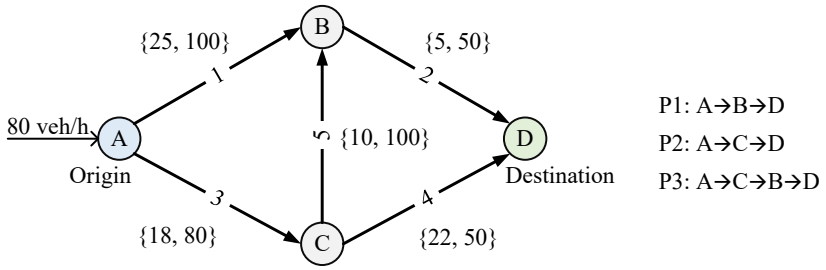


Figure 5.7: The Braess network. Label: {a,b}, where a and b are t_a^0 and c_a, respectively.

5.5.1.1 Evolution of the Pareto frontier

Figure 5.8 illustrates the evolution of the Pareto frontier over three iterations for solving the Braess network. It demonstrates that previously Pareto-efficient (inefficient) paths can become Pareto-inefficient (efficient) owing to the flow redistribution, and vice versa. Initially (Iter 0), all trips are assigned to P1 in both models, which makes P2 the sole path on the Pareto frontier. At Iter 1, P2 is added to the working path set. Equalizing the cumulative costs between P1 and P2 enables both paths to carry positive flows. As such, P1 and P3 are Pareto-efficient, while the previously efficient P2 becomes inefficient as it carries overloaded flows. After including P3 in Iter 2, equilibrium is achieved when all three paths lie on the Pareto frontier with equal cumulative costs.

Furthermore, in the exogenous case with constant stdev, the coordinates of all paths move vertically. In contrast, the endogenous case demonstrates larger spatial location variations, with

the order of paths P2 and P3 changing iteratively. This result suggests that an endogenous model may require increased computational effort to achieve an equilibrium.

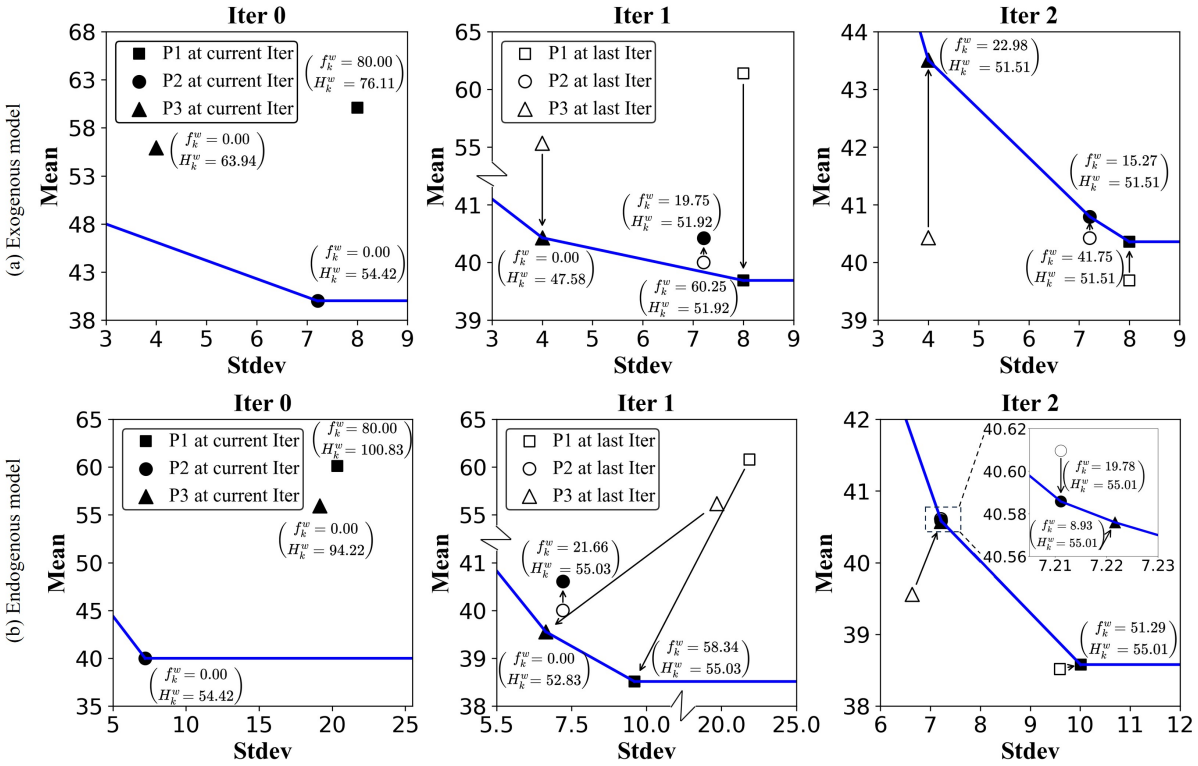


Figure 5.8: Pareto frontier evolution for the Braess network.

5.5.1.2 Comparison of equilibrium solutions

Figure 5.9 compares equilibrium flow patterns across different traffic assignment problems, including standard traffic assignment (Beckmann et al., 1956), mean-risk traffic assignment with risk-neutral or infinitely risk-averse users (Nikolova and Stier-Moses, 2014), and the CMRTA of this paper. We find that network uncertainty and risk attitudes significantly affect route choice patterns. In standard traffic assignment (Figure 5.9(a)), P1 is the only preferred path without considering network uncertainty and risk heterogeneity. For homogeneous risk-neutral users (Figure 5.9(b)), both P1 and P2 become preferable with identical and minimum mean travel times. Moreover, the route choice for homogeneous infinitely risk-averse users differs between exogenous and endogenous models. In the former case (Figure 5.9(c)), all users select P3 because it has the minimum constant stdev. In the latter case (Figure 5.9(d)), P2 receives approximately 75% of the travel demand to attain equal stdev with P3.

For CMRTA models (Figures 5.9(e) and 5.9(f)), all three paths carry flows at equilibrium, indicating that heterogeneous risk-averse travelers may take a broader range of paths. In addition, while path flow distributions are comparable between the exogenous and endogenous models, the order of RAF intervals differs on P2 and P3. For example, the most risk-averse user ($\gamma = 2$) prefers P2 in Figure 5.9(e) but P3 in Figure 5.9(f). This result indicates that endogenous stdev can significantly affect the mapping between users' risk attitudes and their preferred routes.

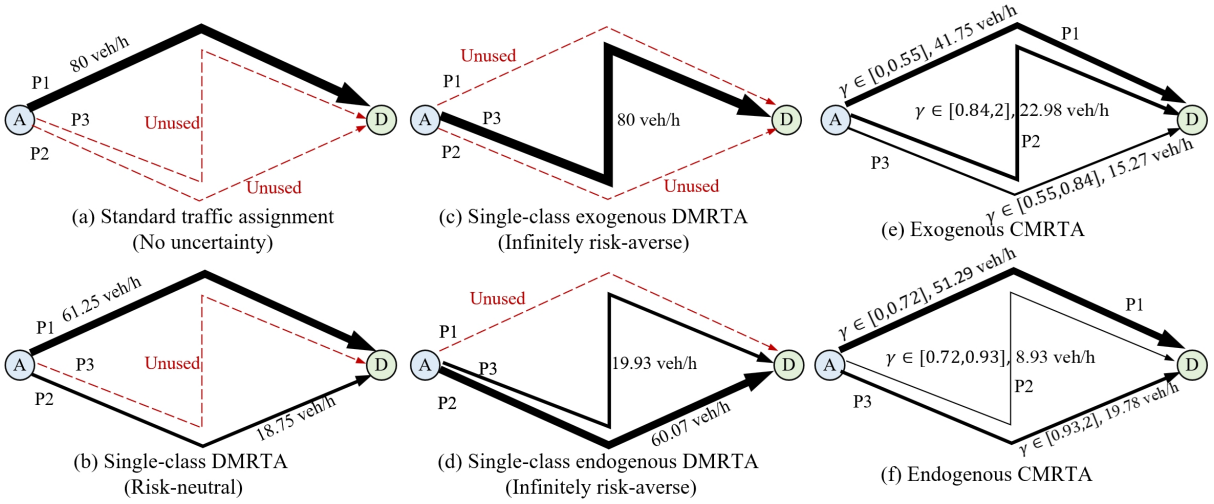
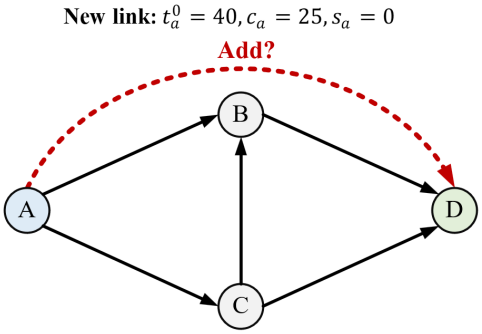


Figure 5.9: Equilibrium path flow patterns for different traffic assignment problems.

5.5.1.3 Discretization errors on operational decisions

Using the Braess network, we also show how discretization errors in DMRTA can lead to poor operational decisions. Consider a planner assessing the addition of a new reliable link P4 with zero stdev that directly connects the OD pair, aiming to reduce total mean travel time (Nikolova and Stier-Moses, 2015). As shown in Figure 5.10(a), the before-and-after analysis under the exogenous CMRTA model shows that adding this link increases the total mean travel times from 3307.65 to 3314.62 minutes, indicating that the project should be rejected.

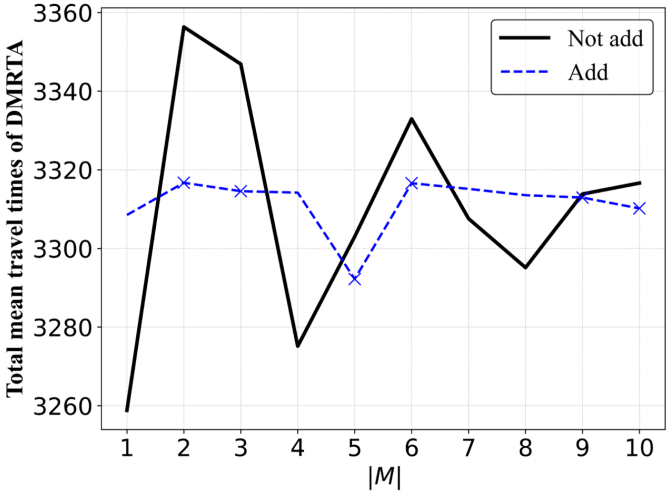
However, Figure 5.10(b) shows that DMRTA incorrectly predicts that the new link will reduce mean travel times for levels $|M| \in \{2, 3, 5, 6, 9, 10\}$. Thus, practitioners using commonly adopted levels (e.g., $|M| = 9$ from Rezaee and Hickman (2015)) face a substantial risk of poor planning decisions. Furthermore, DMRTA produces inconsistent recommendations that heavily depend on the discretization level $|M|$, which underscores the advantage of CMRTA in avoiding inconsistencies due to discretization.



Total mean travel times of CMRTA

Not add	Add
3307.65	3314.62

(a) Before-and-after analysis of CMRTA



(b) Before-and-after analysis of DMRTA

Figure 5.10: Effect of discretization on operational planning decisions.

5.5.2 Real-world networks

This section evaluates the effectiveness and scalability of CG-GP by applying it to four real-world networks of different sizes. Table 5.1 lists detailed topology information for tested networks, which are available at *https://github.com/bstabler/TransportationNetworks*. Following Mirchandani and Soroush (1987), we consider the stochastic link free-flow travel time owing to various nonstandard events, such as weather and traffic delays. Suppose that the link free-flow travel times follow a Gamma distribution, characterized by the shape parameter ω and scale parameter θ. By employing the BRP link performance function, we can express the mean and variance of link travel times as follows:

$$e_a(x_a) = \omega\theta\left(1+\alpha\left(x_a/c_a\right)^\beta\right), \tag{5.21}$$

$$s_a^2(x_a) = \omega\theta^2\left(1+\alpha\left(x_a/c_a\right)^\beta\right)^2, \tag{5.22}$$

where parameters α, β, and c_a are extracted from the source network file.

We set $w = t_a^0/\theta$ to ensure the mean equals the nominal free-flow time (t_a^0), and θ takes a random value from [0,1]. For the exogenous model, the constant variance of link travel times is set as the free-flow time t_a^0 multiplied by a randomly selected factor from the interval [0,1]. By default, RAFs are uniformly distributed between 0 and 2 for all OD pairs.

Table 5.1: Tested real-world networks.

Network	$\|N\|$	$\|A\|$	$\|W\|$	Trips
Sioux Falls	24	76	528	360,600
Anaheim	416	914	1,406	104,694
Barcelona	1,020	2,522	7,865	184,680
Chicago Sketch	933	2,950	93,135	1,260,907

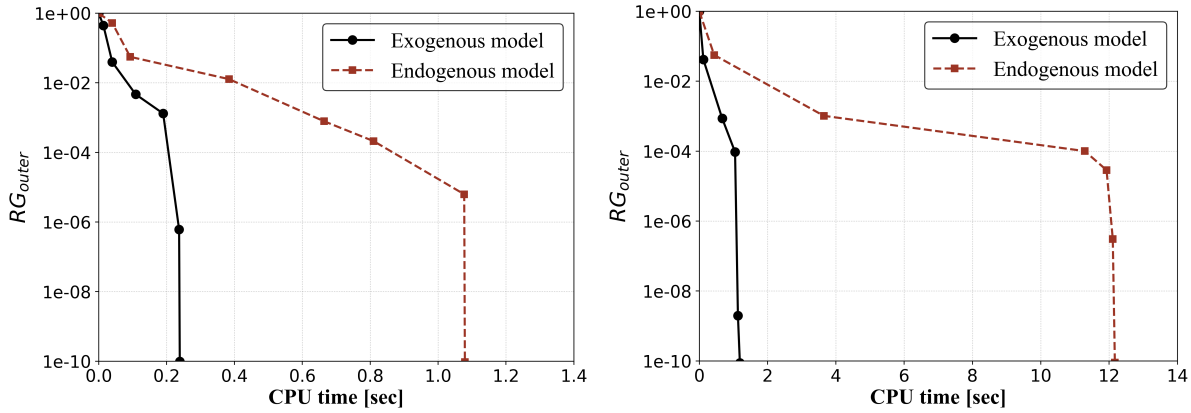
5.5.2.1 Exogenous versus endogenous CMRTA.

We first compare the convergence performance and path number statistics between exogenous and endogenous CMRTA models. Figure 5.11 illustrates the convergence performance of the CG-GP algorithm in CPU time against RG_{outer} for solving four large networks. Every marked point denotes the RG_{outer} value at an outer iteration, demonstrating that the algorithm terminates within 10 iterations across all cases. The results show that CG-GP can obtain high-quality solutions with modest computational times. Both endogenous and exogenous instances exhibit an increase in the number of iterations and computational times as the network size increases.

Our experimental results also show that the endogenous model needs more computational time than the exogenous case, especially when higher solution accuracy is demanded. As shown in Tables 5.2 and 5.3, this performance difference can be attributed to two factors: (i) The endogenous model generates and manages a notably larger number of efficient paths, and (ii) additional computational efforts are required to handle endogenously determined stdev owing to the flow redistribution. Notably, the CG-GP algorithm achieves faster convergence rates at smaller relative gaps than earlier iterations. This behavior aligns with the finding that fewer new paths are added to the working set in later iterations, thus reducing computational times compared with earlier iterations that involve broader solution space exploration.

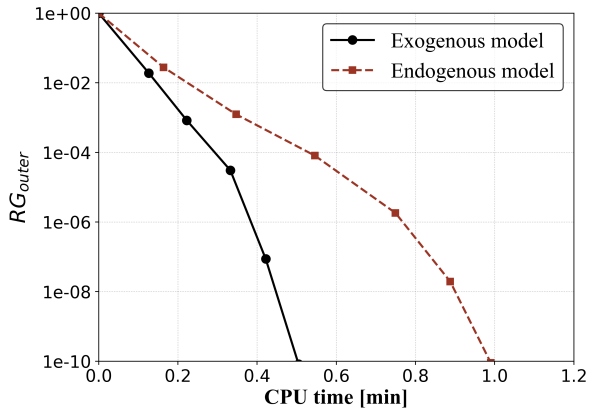
Table 5.2 lists the computational details of every iteration when solving the Chicago Sketch network. The results show commonalities for the pricing problem and distinctions for the restricted master problem. For the column generation phase, the efficient path (EP) count shows marginal variation over iterations for both models, although the endogenous model requires more computational time because of its higher EP count. The algorithm's convergence progress is seen in two indicators: the number of added paths (APs) approaching zero, and RG_{outer} consistently decreasing by approximately one order of magnitude per iteration.

Regarding the restricted master problem, computational efforts differ significantly between

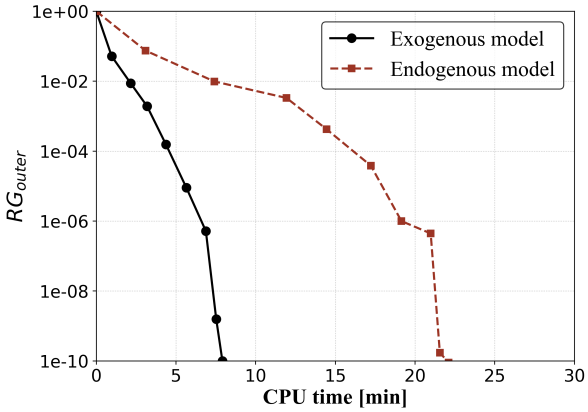


(a) Sioux Falls.

(b) Anaheim



(c) Barcelona



(d) Chicago Sketch

Figure 5.11: Comparison of convergence performance between exogenous and endogenous CM-RTA problems.

models. Specifically, the endogenous model is substantially more demanding, requiring nearly 1e+08 projections compared with the exogenous model. Because fewer paths are generated in every outer iteration, the number of inner iterations (l) decreases, resulting in reduced projection counts and computational time spent on the restricted master problem. Under flow-dependent stdev, the order changes (OCs) fall to zero when RG_{outer} reaches approximately 1e-05. These patterns show that the endogenous model requires significantly greater computational resources in both phases compared with its exogenous counterpart.

As shown in Table 5.3, the ratio between UEPs to EPs decreases as the network size increases, ranging from over 55% for Sioux Falls to approximately 15% for Chicago Sketch. This trend indicates that although larger networks have more potential EPs, the proportion of used paths is lower. Despite this decline, the models maintain comparable path use, as evidenced by a stable UEP/OD ratio of approximately 1.5 across all networks. Moreover, the endogenous model

Table 5.2: Iterative performance metrics on the Chicago Sketch network. Note: EP, efficient path; AP, added path; UEP, used efficient path; OC, order change.

	n	Column generation problem					Restricted master problem				
		#. EP	#. AP	#. UEP	Time (s)	RG_{outer}	l	Projection	#. OC	Time (s)	RG_{inner}
Exogenous	1	895,185	90,611	115,305	24.59	5.16e-02	607	1.88e+07	0	33.35	9.98e-11
	2	869,356	43,355	130,029	21.71	8.73e-03	674	3.64e+07	0	49.68	9.92e-11
	3	900,229	22,095	137,338	23.37	1.92e-03	446	3.13e+07	0	38.38	9.96e-11
	4	887,902	6,550	140,606	22.96	1.55e-04	555	3.84e+07	0	48.08	9.94e-11
	5	898,593	1,320	141,702	24.31	9.03e-06	601	4.19e+07	0	52.40	9.90e-11
	6	898,135	76	141,771	23.17	5.18e-07	597	4.15e+07	0	50.34	9.96e-11
	7	898,134	6	141,777	23.28	1.57e-09	204	1.19e+07	0	15.66	9.99e-11
	8	898,134	0	141,777	23.28	9.99e-11	-	-	-	-	-
Endogenous	1	814,798	94,755	114,846	20.24	7.49e-02	2,311	3.59e+07	62,919	164.82	9.99e-11
	2	979,668	44,046	128,277	28.26	9.82e-03	1,426	7.47e+07	25,819	231.00	9.79e-11
	3	1,017,076	26,449	137,312	30.85	3.31e-03	1,162	8.04e+07	5,842	240.69	9.88e-11
	4	1,044,273	8,757	142,309	32.24	4.23e-04	529	4.02e+07	2,383	118.45	9.98e-11
	5	1,047,776	2,702	144,036	32.46	3.83e-05	600	4.52e+07	2	134.50	9.88e-11
	6	1,049,161	303	144,265	32.52	9.98e-07	376	2.78e+07	0	82.43	9.94e-11
	7	1,049,028	91	144,324	32.41	4.42e-07	341	2.62e+07	0	77.72	9.30e-11
	8	1,049,177	3	144,327	32.59	1.72e-10	9	5.80e+05	0	1.80	9.04e-11
	9	1,049,177	0	144,327	32.62	9.04e-11	-	-	-	-	-

generally exhibits a slightly larger number of EPs than the exogenous model. This difference may be attributed to the wider range of distinguishable stdev resulting from flow-dependent variability.

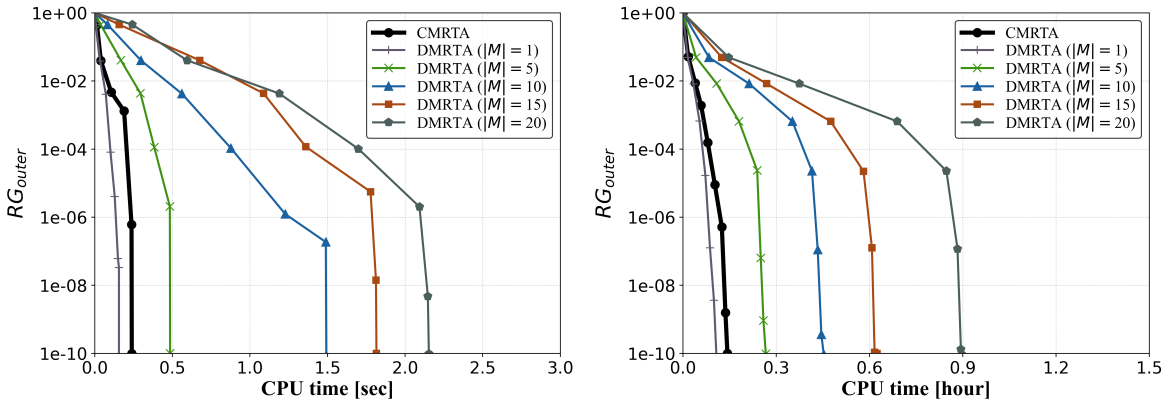
Table 5.3: Path number statistics across four networks at UE.

Networks	Exogenous model				Endogenous model			
	#. EP	#. UEP	#. UEP / #. EP	#. UEP / #. OD	#. EP	#. UEP	#. UEP / #. EP	#. UEP / #. OD
Sioux Falls	1,294	715	55.26%	1.35	1,276	730	57.21%	1.38
Anaheim	5,252	2,252	42.88%	1.60	5,830	2,310	39.62%	1.64
Barcelona	75,960	13,834	18.21%	1.76	84,215	13,839	16.43%	1.76
Chicago Sketch	898,134	141,777	15.79%	1.52	1,049,177	144,327	13.76%	1.55

5.5.2.2 CMRTA versus DMRTA.

This part demonstrates the superior performance of CMRTA over its counterpart (DMRTA) in three aspects: computational efficiency, memory requirements, and link flow solution accuracy. Our analysis considers five discrete class numbers ($|M| \in \{1, 5, 10, 15, 20\}$) using two representative networks: Sioux Falls and Chicago Sketch. Specifically, we only consider the exogenous model because of its ability to generate a unique link flow solution, which offers a meaningful basis for comparative analysis.

As illustrated in Figure 5.12, the convergence comparison shows CMRTA's increased computational efficiency over DMRTA, particularly considering many classes on a large network. Both methods demonstrate comparable convergence trajectories and show faster convergence when approaching the equilibrium. Single-class DMRTA ($|M| = 1$) shows faster performance than CMRTA because CMRTA has to handle more working paths (see the number of decision variables in Table 5.4). However, as the number of discrete classes rises, the computational time for DMRTA rises proportionally due to additional class-specific flow equilibration operations. This performance difference becomes particularly notable in the larger Chicago Sketch network, where CMRTA converges within 10 minutes, while DMRTA, with 20 classes, requires almost 1 hour. Such a significant performance disparity evidences the scalability and superiority of CMRTA, despite its handling of infinitely many classes.



(a) Sioux Falls

(b) Chicago Sketch

Figure 5.12: Comparison of convergence performance between CMRTA and DMRTA problems.

Table 5.4 compares the computational resource requirements and path diversity characteristics between DMRTA and CMRTA on the Chicago Sketch network. The results reveal three critical inefficiencies in the discrete method. First, DMRTA is prone to increased computational redundancy with more user classes. The duplication ratio, calculated as the ratio between the number of non-zero decision variables (i.e., class-based path flows) and the number of intermediate variables (i.e., path flows) minus one, reflects the extent of redundancy in the DMRTA model. This ratio increases dramatically from 0 at $|M| = 1$ (no redundancy) to 12.49 at $|M| = 20$. Second, computational resource consumption increases unsustainably. At $|M| = 20$, DMRTA requires about 13 times more decision variables and 12.86 times more memory than CMRTA owing to variable duplication. Third, DMRTA underrepresents path diversity despite the added

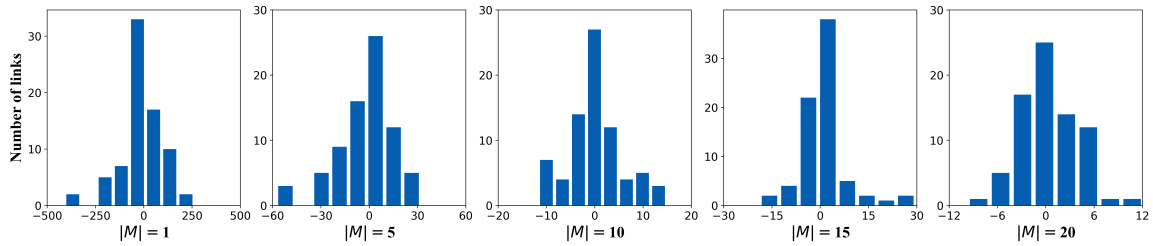
computational cost. Even with 20 classes, DMRTA's used paths remain below CMRTA's level.

Table 5.4: Comparison of activating variables and required memory on Chicago Sketch.

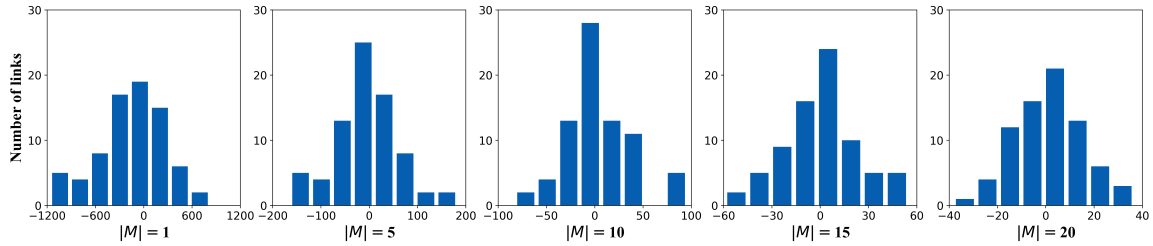
	DMRTA			CMRTA		$\frac{\text{DMRTA}}{\text{CMRTA}}$ Ratio	
$\|M\|$	#. Decision variables (f_k^{wm})	#. Intermediate variables (f_k^w)	RAM (MB)	#. Decision variables (f_k^w)	RAM (MB)	Decision variables	RAM
1	93,791	93,791	19.38	141,771	30.08	0.66	0.64
5	467,851	130,548	96.74			3.30	3.22
10	935,422	135,785	193.45			6.60	6.43
15	1,402,982	137,794	290.14			9.90	9.65
20	1,870,561	138,672	386.85			13.19	12.86

We also analyze the link flow differences in DMRTA relative to CMRTA on the Sioux Falls network using two RAF intervals: $[0, 2]$ and $[0, 20]$. Figure 5.13 shows histograms of equilibrium link flow differences between the two models across network links. First, although increasing $|M|$ generally reduces errors, the discretization method can lead to asymmetrically distributed negative or positive flow errors. This indicates that DMRTA may spatially underestimate or overestimate link flows compared with CMRTA. Second, the accuracy of DMRTA is highly sensitive to the RAF interval width. A broader RAF range (e.g., $[0, 20]$) generates larger errors than a narrower one (e.g., $[0, 2]$) given the same $|M|$. These results suggest that predetermining an appropriate number of discrete classes for accurate flow estimation is difficult and requires comprehensive sensitivity analyses for specific application scenarios.

We then perform quality-efficiency analysis comparing DMRTA and CMRTA. Taking link flows of CMRTA as the benchmark, we use the mean squared error (MSE) to quantify the quality of DMRTA link flows, and we measure computational efficiency as the runtime ratio of DMRTA to CMRTA required to achieve the termination criteria. As shown in Figure 5.14, increasing discretization levels $|M|$ improves the solution quality (MSE decreases from approximately 10,000 to 15) but reduces the solution efficiency (runtime ratios increase from approximately 1 time to nearly 12 times), showing a clear quality-efficiency trade-off. In contrast, CMRTA offers greater computational efficiency without sacrificing solution quality, making it the preferred method for practical implementations.



(a) RAF interval: [0,2]



(b) RAF interval: [0,20]

Figure 5.13: Link flow difference with varying RAF intervals on Sioux Falls.

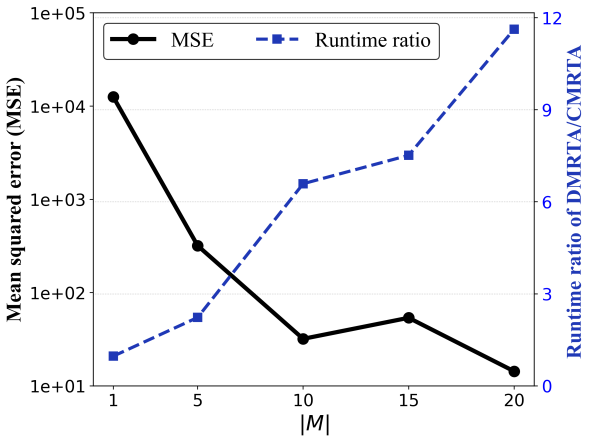


Figure 5.14: The quality-efficiency comparison between DMRTA and CMRTA on Sioux Falls.

5.6 Summary

This chapter examines the CMRTA problem, where travelers make route decisions based on their own RAF toward the mean-stdev trade-off under network uncertainty. The RAF among the travel population follows a continuous probabilistic distribution, which enables the representation of infinitely many potential user classes.

To characterize the equilibrium of CMRTA, we propose a novel finite-dimensional complementarity condition based on the concept of cumulative path costs, thus facilitating both model formulation and algorithm development for nonadditive CMRTA problems. The proposed mathematical formulations (in the space of total path flows) simultaneously circumvent

the dimensional expansion inherent in DMRTA and maintain the flexibility to incorporate dual flow-dependent criteria within a generalized routing game framework. To address large-scale problems, we develop a tailored path-based CG-GP algorithm, comprising two components for solving the pricing problem and restricted master problem at every major iteration. Extensive computational experiments across different networks verify the efficiency of the CG-GP algorithm in obtaining high-quality solutions for CMRTA. Our analysis also shows that DMRTA can underestimate the number of used paths, with discretization errors significantly affecting optimal operational decisions. Furthermore, CMRTA demonstrates superior computational and resource efficiency compared with DMRTA, particularly in large-scale instances with many discrete user classes.

This study also inspires many research directions that merit exploration. First, researchers can customize the developed methodological tools to other problems involving two flow-dependent criteria (e.g., congestion pricing with the flow-dependent toll scheme). Second, a potential remedy can be used to extend our model to incorporate link travel time correlations, where travel times on a link may be spatially correlated with other links. Third, risk-seeking travelers (with negative RAF values) are not considered in this study and thus merit further investigation. Last, future research will explore the empirical estimation of RAF distributions, followed by end-to-end validation to assess the practical advantages of CMRTA over DMRTA in real-world applications.

Chapter 6

Summary and future research

6.1 Summary

This thesis contributes to modeling travelers' route preference and its heterogeneity in transportation networks, with a specialized focus on discrete and continuous multiclass methodologies. The core innovation lies in strategically leveraging the Pareto frontier of network paths (supply-side) in conjunction with distributions of user preferences (demand-side) to create tractable and scalable models. This approach has enabled the development of generalized modeling frameworks, efficient mathematical formulations, and computationally practical algorithms that bridge theoretical sophistication with real-world applicability.

The research journey began with a systematic review (Chapter 2), which synthesized 108 studies spanning 1979-2025 and focused on the distinctions between discrete and continuous MCTA approaches. The review revealed that while discrete approaches offer mathematical tractability and wide applicability across various heterogeneity types (preference, information, choice set, and routing principle), continuous approaches provide superior behavioral fidelity. Crucially, it identified that recent finite-dimensional formulations have begun narrowing the computational gap between these paradigms—a finding that directly motivated the subsequent methodological contributions of this thesis.

A key contribution was the development of a unified path-based gradient projection algorithm (Chapter 3) capable of solving both continuous and discrete multiclass bi-criteria problems within a single computational framework. The algorithm incorporates three tailored modules—column generation, decomposition and equilibration, and convergence criteria—specifically

designed for each paradigm. Numerical experiments demonstrated not only the framework's computational efficiency but also that the discrete model's equilibrium flows converge to their continuous counterpart as the number of user classes increases, providing valuable insights for model selection in practical applications.

Building on these insights, the thesis introduced an alternative formulation for the discrete multiclass problem (Chapter 4) that eliminates class-specific variables by exploiting the fundamental order structure of paths and traveler preferences. The formulation distinguishes between low-TEM (high-VOT) travelers who prefer fast but expensive paths and high-TEM (low-VOT) travelers who favor slow but cheap paths. We proved this formulation's equivalence to conventional models while developing a specialized subgradient projection algorithm to handle its non-differentiable convex objective. Numerical validation across various networks demonstrated that this approach reduces variable counts by up to 90% while maintaining identical link flow patterns, achieving superior computational performance, particularly beneficial for networks with numerous user classes.

Finally, the research was extended to the more complex continuous multiclass mean-risk problem (Chapter 5), where both travel time and reliability are flow-dependent. By introducing the novel concept of cumulative path costs and modeling risk-aversion factors as continuous distributions, we developed finite-dimensional complementarity conditions that avoid dimensional explosion. The tailored column generation with gradient projection (CG-GP) algorithm effectively solves these challenging problems, with computational experiments revealing that discrete approximations may underestimate utilized paths when overlooking fine-grained user heterogeneity, while the continuous approach shows superior computational efficiency in large-scale scenarios.

Collectively, these contributions provide a cohesive and powerful toolkit. They demonstrate that the strategic integration of supply and demand information enables the development of multiclass models that are not only behaviorally richer but also computationally more efficient than traditional approaches, thereby facilitating more accurate and equitable planning for future urban mobility systems.

6.2 Future research

This thesis focuses on transportation network equilibrium models and algorithms with heterogeneous users. This issue is a common problem in many transportation problems. However, the modeling methodology still needs further advancement. Numerous questions, problems, and proposals remain to be investigated. Future research may be pursued through the following directions, which embody a blend of theoretical and practical efforts.

- *Advanced continuous MCTA models and algorithms*. The development of advanced continuous MCTA models and algorithms represents a promising research frontier that remains relatively underexplored compared to discrete approaches. Especially, current continuous MCTA models mainly considered VOT distribution (e.g., Leurent, 1993; Dial, 1996, 1997; Xie et al., 2021; Xu et al., 2023b), with a few considering the continuously distributed range anxiety (Xie et al., 2017, 2019; Xu et al., 2024b). Significant potential exists to model other types of heterogeneity. For example, while Lo et al. (2006) presented a TTB-based discrete MCTA model to capture risk preferences, subsequent research could adopt a continuous distribution to model the risk preferences of travelers, which would be more challenging but better reflect real-world population characteristics.

- *Hybrid user heterogeneity: discrete–continuous methods*. Combining discrete and continuous heterogeneity representations presents a significant opportunity. Current MCTA models typically adopt one approach, yet real-world behavior often requires hybrid methods. For instance, Lam and Small (2001) found distinct VOR preferences by gender, suggesting models that segment groups discretely (e.g., by gender) while modeling attributes continuously (e.g., VOR) within groups. Such hybrid models would balance the simplicity of discrete methods with the precision of continuous approaches. These are particularly relevant for emerging technologies, such as mixed fleets of connected autonomous vehicles (CAVs) and human-driven vehicles (HDVs), where CAV users may have distinct VOT distributions due to non-driving activities (e.g., work or leisure). Hybrid models could better capture these differences, enhancing predictions of traffic patterns and policy impacts.

- *Incorporating multidimensional travel choices*. While discrete MCTA models have suc-

cessfully incorporated multidimensional travel choices, such as combined trip distribution and TA by Lam and Huang (1992), and elastic demand TA by Nagurney and Dong (2002a), continuous MCTA models remain limited in this regard. Only Leurent (1993) and Xu et al. (2024a) have tackled elastic demand TA with continuously distributed VOT, but these studies defined aggregate demand functions based on average travel costs, missing the individual-level elasticity. Furthermore, other traveler dimensions, such as travel choice, destination choice, and mode choice, remain unexplored in continuous MCTA methods. This gap highlights the need to enhance continuous MCTA models to integrate multidimensional choices to improve traffic flow predictions.

- *Modeling user heterogeneity in the era of emerging transport technologies*. The rise of emerging transport technologies, such as EVs, CAVs, low-altitude economy technologies (e.g., flying cars, drones), and other emerging innovations introduces new dimensions of user heterogeneity in transportation networks. For instance, CAV users may exhibit fundamentally different VOT distributions compared to HDV users; travelers may show different safety concerns and willingness to pay for using flying cars. Additionally, the asymmetric interactions between these new technologies—such as differences in driving behavior and congestion effect—must be explicitly modeled to capture their impact on traffic flows. Future research should explicitly model the user heterogeneity in TA amid emerging transport technologies to better predict traffic flow patterns, optimize network performance in mixed-fleet environments, and support the safe and efficient integration of next-generation mobility solutions.

By addressing these gaps and pursuing these opportunities, MCTA research will keep pace with the times and contribute to new theoretical developments and practical applications for transportation planning and management.

Appendix A

Discrete and continuous multiclass bi-criteria traffic assignment problems: A unified gradient projection algorithm

A.1 MSA and FW algorithms for solving the CMBTA problem

Our implementation of the MSA and FW algorithms for solving the CMBTA problem is shown in Algorithm 9. Interested readers may refer to Leurent (1993) and Marcotte and Zhu (1996) for more details on link-based model formulation and solution methods. The main difference between MSA and FW lies in the step size determination. All other algorithmic routines are the same for MSA and FW.

Algorithm 9 MSA and FW algorithm for solving the CMBTA problem.

Initialization: Call the all-or-nothing algorithm (same as line 4–line 16) to obtain the link flow vector $\mathbf{x}^0$, link moment vector $\mathbf{u}^0$. Set counter $it = 1$.
Update: Update link cost vector $\mathbf{t}^{it}$.
Direction finding:
Set auxiliary link flow vector $\bar{\mathbf{x}}^{it} = \mathbf{0}$ and auxiliary link moment vector $\bar{\mathbf{u}}^{it} = \mathbf{0}$.
for each origin $r \in R$ **do**
    Call the *Dijkstra-like bi-criteria shortest path algorithm* (Sedeño-Noda and Raith, 2015) to generate the one-to-all efficient path tree T_r from r.
    for each destination $s \in S$ **do**
        Get the auxiliary efficient paths set $\bar{P}_{rs} = T_r(s)$.
        for each efficient path $p \in \bar{P}_{rs}$ **do**
            Determine the upper bound VOT α_p and lower bound VOT α_{p-1} on the Pareto frontier.
            for each link $a \in p$ **do**
                Accumulate the auxiliary link flow $\bar{x}_a^n = \bar{x}_a^n + q^{rs}\left(H_{rs}(\alpha_p) - H_{rs}(\alpha_{p-1})\right)$ and auxiliary link moment $\bar{u}_a^{it} = \bar{u}_a^{it} + q^{rs}\int_{\alpha_{p-1}}^{\alpha_p} \frac{1}{\alpha} h_{rs}(\alpha) d\alpha$.
            end for
        end for
    end for
end for
Step size determination (choose MSA or FW): For MSA, step size $\beta_{it} = \frac{1}{it}$. For FW, the step size is obtained by solving a one-dimensional optimization problem:

$$\beta_{it} = \arg\min_{\beta\in[0,1]} \left(\sum_{a\in A} \int_0^{x_a^n + \beta(\bar{x}_a^n - x_a^n)} t_a(v)dv + \tau_a \cdot (u_a^n + \beta(\bar{u}_a^n - u_a^n)) \right).$$

Move: $\mathbf{x}^{it+1} = \mathbf{x}^{it} + \beta_{it}(\bar{\mathbf{x}}^{it} - \mathbf{x}^{it})$, $\mathbf{u}^{it+1} = \mathbf{u}^{it} + \beta_{it}(\bar{\mathbf{u}}^{it} - \mathbf{u}^{it})$.
Termination test: Calculate the relative gap

$$RG = 1 - \frac{\sum_{a\in A}(\bar{x}_a^{it} t_a^{it} + \bar{u}_a^{it}\tau_a)}{\sum_{a\in A}(x_a^{it} t_a^{it} + u_a^{it}\tau_a)}.$$

If $RG < \epsilon$ or $it \geq I$, terminate; otherwise, set $it = it + 1$ and go to **Update**.

A.2 MSA and FW algorithms for solving the DMBTA problem

Our implementation of the MSA and FW algorithms for solving the DMBTA problem is shown in Algorithm 10. Interested readers can refer to Yang and Huang (2004) for the link-based formulation for the DMBTA problem. The main difference between MSA and FW lies in the step

size determination. All other algorithmic routines are the same for MSA and FW.

Algorithm 10 MSA and FW algorithm for solving the DMBTA problem.

1: ***Initialization:*** Call the all-or-nothing algorithm (same as line 4–line 16) to obtain the class-specific link flow vector $\tilde{\mathbf{x}}^0 = (x_a^{k,0})_{\forall a \in A, k \in K}$. Set counter $it = 1$.
2: ***Update:*** Update link cost vector $\mathbf{t}^{it}$.
3: ***Direction finding:***
4: Set auxiliary class-specific link flow vector $\tilde{\mathbf{y}}^{it} = \mathbf{0}$.
5: **for** each origin $r \in R$ **do**
6: Call the *Dijkstra-like bi-criteria shortest path algorithm* (Sedeño-Noda and Raith, 2015) to generate the one-to-all efficient path tree T_r from r.
7: **for** each destination $s \in S$ **do**
8: Get the auxiliary efficient paths set $\bar{P}_{rs} = T_r(s)$.
9: **for** each class $k \in K$ **do**
10: Get an auxiliary path $\bar{p} = \arg\min_{p \in \bar{P}_{rs}} d_p^{rs}(\alpha_k)$
11: **for** each link $a \in \bar{p}$ **do**
12: Accumulate the auxiliary class-specific link flow $y_a^k = y_a^k + q_k^{rs}$.
13: **end for**
14: **end for**
15: **end for**
16: **end for**
17: ***Step size determination (choose MSA or FW):*** For MSA, step size $\beta_{it} = \frac{1}{it}$. For FW, the step size is obtained by solving a one-dimensional optimization problem:

$$\beta_{it} = \arg\min_{\beta \in [0,1]} \sum_{a \in A} \int_0^{x_a^{it} + \beta(y_a^{it} - x_a^{it})} t_a(v)dv + \sum_{a \in A} \sum_{k \in K} \frac{1}{\alpha_k} \tau_a \left(x_a^{k,it} + \beta(y_a^{k,it} - x_a^{k,it}) \right).$$

18: ***Move:*** $\tilde{\mathbf{x}}^{it+1} = \tilde{\mathbf{x}}^{it} + \beta_{it}(\tilde{\mathbf{y}}^{it} - \tilde{\mathbf{x}}^{it})$.
19: ***Termination test:*** Calculate the relative gap

$$RG = 1 - \frac{\sum_{a \in A} y_a^{it} t_a^{it} + \sum_{a \in A} \sum_{k \in K} \frac{1}{\alpha_k} \tau_a y_a^{k,it}}{\sum_{a \in A} x_a^{it} t_a^{it} + \sum_{a \in A} \sum_{k \in K} \frac{1}{\alpha_k} \tau_a x_a^{k,it}}.$$

If $RG < \epsilon$ or $it \geq I$, terminate; otherwise, set $it = it + 1$ and go to **Update**.

Appendix B

Discrete multiclass bi-criteria traffic assignment without class-specific variables: An alternative formulation and a subgradient projection algorithm

B.1 Dijkstra-like bi-criteria shortest path algorithm

Algorithm 11 describes our implementation of the *Dijkstra-like bi-criteria shortest path algorithm*, which was developed by Sedeño-Noda and Raith (2015). This algorithm utilizes a Dijkstra-like label-setting method to identify the efficient path set $\bar{P}_{rs}$ between all OD pairs rs originating from the same origin r. It has a complexity of $O\left(C\left(|A| + |N|\log|N|\right)\right)$, where C is a constant denoting the number of efficient paths. Compared to its original version in Sedeño-Noda and Raith (2015), we further utilize the minimum TEM $\underline{\beta}$ and maximum TEM $\overline{\beta}$ among all user classes as a bound to restrict the searching process, which can terminate the algorithm faster.

Algorithm 11 Dijkstra-like bi-criteria shortest path algorithm.

Input: Network topology $G = (N, A)$; link travel time t_{ij} and link toll τ_{ij}, $\forall (I, j) \in A$; origin node r; Minimum TEM $\underline{\beta}$ and Maximum TEM $\overline{\beta}$.

Output: Efficient path set $\left\{\bar{P}_{rs}, \forall s \in S\right\}$ originating from r for $\beta \in \left[\underline{\beta}, \overline{\beta}\right]$.

Set node time $t_i = \infty$, node toll $m_i = \infty$, predecessor $pred_i = 0$, candidate predecessor $Cpred_i = 0$, reduced toll $RM_i = \infty$, node ratio $\theta_i = \infty$, last ratio $L\theta_i = 0, \forall i \in N$.
Set link reduced time $RT_{ij} = \infty$, link reduced toll $RM_{ij} = \infty, \forall (i, j) \in A$.
Set efficient path set $\bar{P}_{rs} = \emptyset, \forall s \in S$.
Set the generalized cost on each link as $t_{ij} + \underline{\beta}\tau_{ij}, \forall (i, j) \in A$.
Call the Dijkstra algorithm to compute the shortest path tree, node time label, and node toll label $\{(pred_i, t_i, m_i), \forall i \in N\}$.
for $s \in N$ **do** ▷ Record the initial shortest paths.
    Retrive shortest path p from r to s using tree $\{pred_i, \forall i \in N\}$, and add p to $\bar{P}_{rs}$.
end for
for $j \in N - \{r\}$ **do** ▷ Scan predecessors.
    Compute $(\theta_j, RM_j, Cpred_j) = \text{lex min}_{i \in j^-} \left\{\left(-\frac{t_i + t_{ij} - t_j}{m_i + \tau_{ij} - m_j}, m_i + \tau_{ij} - m_j, i\right) : m_i + \tau_{ij} - m_j < 0\right\}$.
end for
Create a heap SE.
for $i \in N - \{r\}$ **do**
    if $Cpred_i \neq 0$ **then** Push $(\theta_i, RM_i, Cpred_i)$ into SE.
    end if
end for
while $SE \neq \emptyset$ **do** ▷ Node index: i (predecessor)–>j (this node)–>k (successor).
    $(\theta_j, RM_j, Cpred_j) =$ pop-min(SE). ▷ Remove the lexicographically minimum label from SE.
    if $\theta_j \geq \overline{\beta}$ **then** Continue.
    end if
    Set $i = Cpred_j, t_j = t_i + t_{ij}, m_j = m_i + \tau_{ij}$.
    if $\theta_j > L\theta_j$ **then**
        Set $pred_j = i$.
        Retrieve path p from s to j using tree $\{pred_i, \forall i \in N\}$, and add p to $\bar{P}_{rj}$.
    end if
    Set $L\theta_j = \theta_j, \theta_j = \infty, Cpred_j = 0$.
    Compute $(\theta_j, RM_j, Cpred_j) = \text{lex min}_{i \in j^-} \left\{\left(-\frac{t_i + t_{ij} - t_j}{m_i + \tau_{ij} - m_j}, m_i + \tau_{ij} - m_j, i\right) : m_i + \tau_{ij} - m_j < 0\right\}$.
    if $Cpred_j \neq 0$ **then** push $(\theta_j, RM_j, Cpred_j)$ into SE.
    end if
    for $k \in j^+$ **do** ▷ Scan successors.
        $RT_{jk} = t_j + t_{jk} - t_k, RM_{jk} = m_j + \tau_{jk} - m_k$.
        if $RM_{jk} < 0$ **then**
            if $\left(-\frac{RT_{jk}}{RM_{jk}}, RM_{jk}\right) < (\theta_k, RM_k)$ **then** ▷ Lexicographically smaller.
                if $Cpred_k = 0$ **then**
                    Push $\left(-\frac{RT_{jk}}{RM_{jk}}, RM_{jk}, k\right)$ into SE.
                else
                    Decrease-key $\left(-\frac{RT_{jk}}{RM_{jk}}, RM_{jk}, k\right)$ in SE.
                end if
                $\theta_k = -\frac{RT_{jk}}{RM_{jk}}, RM_k = RM_{jk}, Cpred_k = j$.
            end if
        end if
    end for
end while

B.2 Gradient projection algorithm for conventional discrete MBTA formulation

The detailed gradient projection algorithm for solving the conventional MBTA formulation (4.3) is presented in Fig. B.1. Same as the subgradient projection algorithm in Fig. 4.6, we initialize the algorithm by assigning all the class-specific demands to the corresponding minimum generalized cost path for each OD pair and each user class. Differently, in the column generation process, there is no need to consider the path toll order anymore. Instead, we directly add the minimum generalized cost path to the working path set for each OD pair and each user class. In the equilibration process, the GP algorithm equilibrates the working path set of each OD pair and each user class sequentially, using the one-at-a-time flow updating strategy. The convergence test also uses the relative gap to measure the precision of the solution, though the functional form is slightly different from Eq. (4.29).

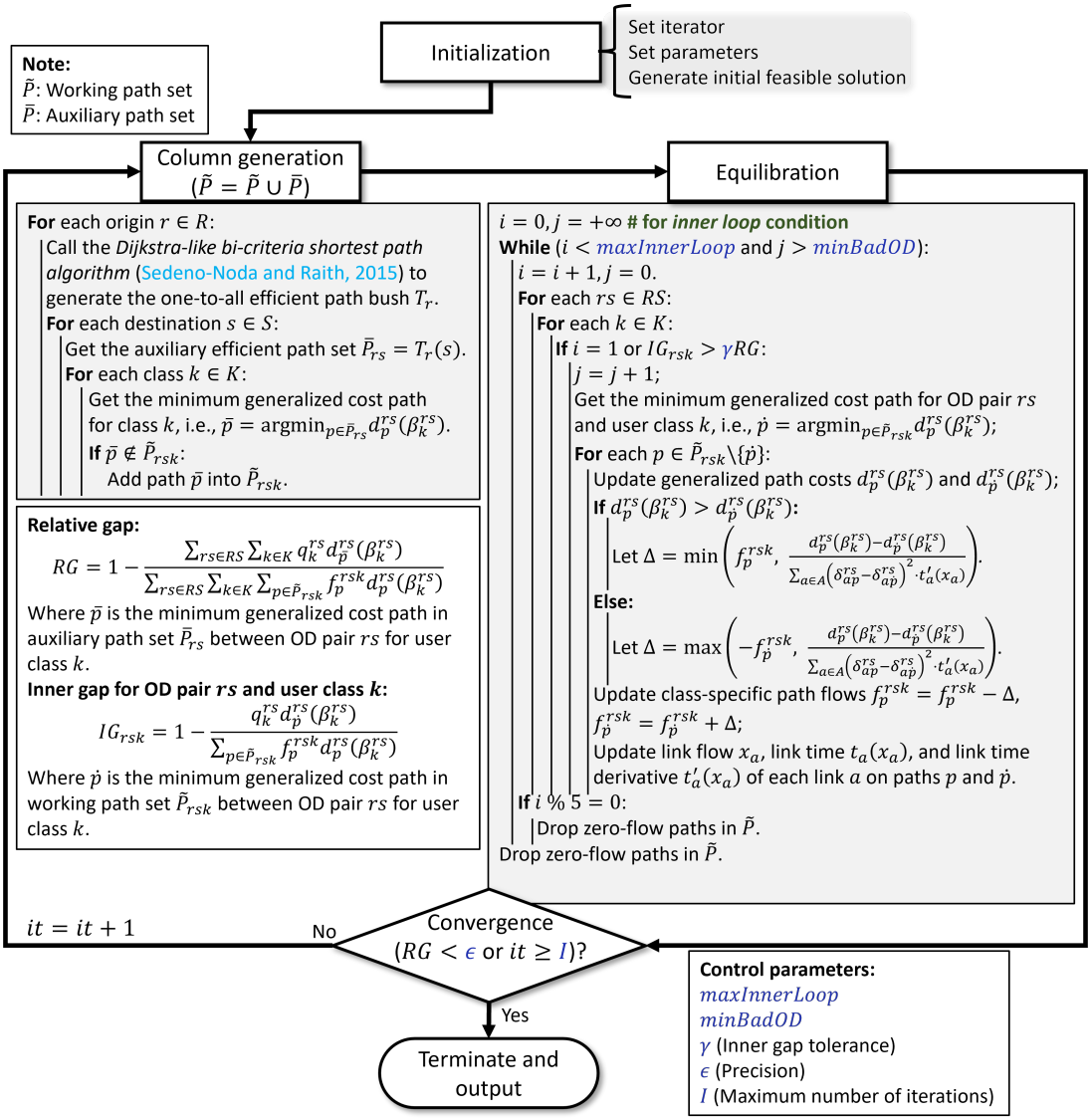


Figure B.1: The gradient projection algorithm for the conventional MBTA problem.

Appendix C

Continuous multiclass mean-risk traffic assignment problem: both criteria are flow-dependent

C.1 Proof of Proposition 5.3

The equivalence between Definition 5.3 and Condition (5.9) is established in both directions.

For necessity, we need to show that Condition (5.9) implies Definition 5.3. Without loss of generality, we consider two paths $k, k' \in \hat{K}_w$ such that $f_k^w > 0$. It follows that $[\hat{\gamma}_{k-1}^w(\mathbf{f}), \hat{\gamma}_k^w(\mathbf{f})] \neq \emptyset$. Then, we need to prove that $C_k^w(\mathbf{f};\gamma) \leq C_{k'}^w(\mathbf{f};\gamma)$ for any $\gamma \in [\hat{\gamma}_{k-1}^w(\mathbf{f}), \hat{\gamma}_k^w(\mathbf{f})]$ and $k' \neq k$. According to Condition (5.9), it leads to

$$H_k^w(\mathbf{f}) = \mu_k^w(\mathbf{f}) + \textstyle\sum_{i=k}^{|\hat{K}_w|} \hat{\gamma}_i^w(\mathbf{f}) \cdot \left[\sigma_i^w(\mathbf{f}) - \sigma_{i+1}^w(\mathbf{f})\right] - h_w = 0, \tag{C.1}$$

$$H_{k'}^w(\mathbf{f}) = \mu_{k'}^w(\mathbf{f}) + \textstyle\sum_{i=k'}^{|\hat{K}_w|} \hat{\gamma}_i^w(\mathbf{f}) \cdot \left[\sigma_i^w(\mathbf{f}) - \sigma_{i+1}^w(\mathbf{f})\right] - h_w \geq 0. \tag{C.2}$$

Combing Eqs. (C.1) and (C.2) together yields

$$\mu_k^w(\mathbf{f}) - \mu_{k'}^w(\mathbf{f}) - \sum_{i=k'}^{|\hat{K}_w|-1} \hat{\gamma}_i^w(\mathbf{f}) \cdot \left[\sigma_i^w(\mathbf{f}) - \sigma_{i+1}^w(\mathbf{f})\right] + \sum_{i=k}^{|\hat{K}_w|-1} \hat{\gamma}_i^w(\mathbf{f}) \cdot \left[\sigma_i^w(\mathbf{f}) - \sigma_{i+1}^w(\mathbf{f})\right] \leq 0. \tag{C.3}$$

There are two cases in terms of the order between two paths: (i) $k' > k$ and (ii) $k' < k$. For

case (i) such that $\sigma_k^w(\mathbf{f}) \geq \sigma_{k'}^w(\mathbf{f})$, Eq (C.3) can be simplified as

$$\mu_k^w(\mathbf{f}) - \mu_{k'}^w(\mathbf{f}) + \sum_{k \leq i < k'} \hat{\gamma}_i^w(\mathbf{f}) \cdot \left[\sigma_i^w(\mathbf{f}) - \sigma_{i+1}^w(\mathbf{f})\right] \leq 0. \tag{C.4}$$

For any $\gamma \in [\hat{\gamma}_{k-1}^w(\mathbf{f}), \hat{\gamma}_k^w(\mathbf{f})]$, we have $\gamma \leq \hat{\gamma}_i^w(\mathbf{f})$ for each $i \geq k$. Given that $\sigma_i^w(\mathbf{f}) \geq \sigma_{i+1}^w(\mathbf{f})$ by path order, we have the following relationship as per Eq. (C.4):

$$\mu_k^w(\mathbf{f}) - \mu_{k'}^w(\mathbf{f}) + \sum_{k \leq i < k'} \gamma \cdot \left[\sigma_i^w(\mathbf{f}) - \sigma_{i+1}^w(\mathbf{f})\right] \leq 0. \tag{C.5}$$

Simplifying Eq. (C.5) yields

$$\mu_k^w(\mathbf{f}) - \mu_{k'}^w(\mathbf{f}) + \gamma \cdot \sigma_k^w(\mathbf{f}) - \gamma \cdot \sigma_{k'}^w(\mathbf{f}) \leq 0 \Rightarrow C_k^w(\mathbf{f}; \gamma) \leq C_{k'}^w(\mathbf{f}; \gamma). \tag{C.6}$$

For case (ii) such that $\sigma_k^w(\mathbf{f}) \leq \sigma_{k'}^w(\mathbf{f})$, Eq (C.3) can be simplified as

$$\mu_k^w(\mathbf{f}) - \mu_{k'}^w(\mathbf{f}) - \sum_{k' \leq i < k} \hat{\gamma}_i^w(\mathbf{f}) \cdot \left[\sigma_i^w(\mathbf{f}) - \sigma_{i+1}^w(\mathbf{f})\right] \leq 0. \tag{C.7}$$

For any $\gamma \in [\hat{\gamma}_{k-1}^w(\mathbf{f}), \hat{\gamma}_k^w(\mathbf{f})]$, we have $\gamma \geq \hat{\gamma}_i^w(\mathbf{f})$ for each $i \leq k$. Accordingly, Eq. (C.7) yields the following relationship:

$$\mu_k^w(\mathbf{f}) - \mu_{k'}^w(\mathbf{f}) - \sum_{k' \leq i < k} \gamma \cdot \left[\sigma_i^w(\mathbf{f}) - \sigma_{i+1}^w(\mathbf{f})\right] \leq 0. \tag{C.8}$$

Simplifying Eq. (C.8) also yields

$$\mu_k^w(\mathbf{f}) - \mu_{k'}^w(\mathbf{f}) + \gamma \cdot \sigma_k^w(\mathbf{f}) - \gamma \cdot \sigma_{k'}^w(\mathbf{f}) \leq 0 \Rightarrow C_k^w(\mathbf{f}; \gamma) \leq C_{k'}^w(\mathbf{f}; \gamma). \tag{C.9}$$

For sufficiency, we need to show that Definition 5.3 implies Condition (5.9). Given an equilibrium path flow $\mathbf{f}$, we need to prove that all used paths have an equal cumulative cost, while paths with higher cumulative costs remain unused. Given an OD pair w, we split the set $\hat{K}_w$ into two sets: a used path set $\hat{K}_w^1 \subseteq \hat{K}_w$ and an unused path set $\hat{K}_w^2 \subseteq \hat{K}_w$. Therefore, we have $f_k^w > 0$ for all $k \in \hat{K}_w^1$, $f_k^w = 0$ for all $k \in \hat{K}_w^2$, and $\hat{K}_w = \hat{K}_w^1 \cup \hat{K}_w^2$. Note that all paths in $\hat{K}_w^1$ must lie on the Pareto frontier. Let $\Delta_k^w \geq 0$ denote the vertical quantity relative to the Pareto

frontier for all paths in $\hat{K}_w^2$. There are two cases regarding the number of used paths, i.e., Case (i) for $|\hat{K}_w^1| = 1$ and Case (ii) for $|\hat{K}_w^1| > 1$.

For the case of $|\hat{K}_w^1| = 1$, i.e., only one path is used. Let i denote the index of that path. Therefore, $\hat{\gamma}_k^w = \gamma_{lb}^w$ for $k < i$, and $\hat{\gamma}_k^w = \gamma_{ub}^w$ for $k > i$. It follows that

$$H_i^w(\mathbf{f}) = \mu_i^w(\mathbf{f}) + \sum_{k=i}^{|\hat{K}_w|} \gamma_{ub}^w \cdot \left[\sigma_k^w(\mathbf{f}) - \sigma_{k+1}^w(\mathbf{f})\right] = \mu_i^w(\mathbf{f}) + \gamma_{ub}^w \cdot \sigma_i^w(\mathbf{f}). \tag{C.10}$$

Consider an unused path $k \in \hat{K}_w^2$ such that $k > i$, we have

$$H_k^w(\mathbf{f}) = \mu_k^w(\mathbf{f}) + \gamma_{ub}^w \cdot \sigma_k^w(\mathbf{f}) = \mu_i^w(\mathbf{f}) + \gamma_{ub}^w \cdot [\sigma_i^w(\mathbf{f}) - \sigma_k^w(\mathbf{f})] + \gamma_{ub}^w \cdot \sigma_k^w(\mathbf{f}) + \Delta_k^w \tag{C.11a}$$

$$= \mu_i^w(\mathbf{f}) + \gamma_{ub}^w \cdot \sigma_i^w(\mathbf{f}) + \Delta_k^w = H_i^w(\mathbf{f}) + \Delta_k^w \geq H_i^w(\mathbf{f}). \tag{C.11b}$$

Consider an unused path $k \in \hat{K}_w^2$ such that $k < i$, we have

$$H_k^w(\mathbf{f}) = \mu_k^w(\mathbf{f}) + \sum_{j=k}^{i-1} \gamma_{lb}^w \cdot \left[\sigma_j^w(\mathbf{f}) - \sigma_{j+1}^w(\mathbf{f})\right] + \sum_{j=i}^{|\hat{K}_w|} \gamma_{ub}^w \cdot \left[\sigma_j^w(\mathbf{f}) - \sigma_{j+1}^w(\mathbf{f})\right] \tag{C.12a}$$

$$= \mu_k^w(\mathbf{f}) + \gamma_{lb}^w \cdot [\sigma_k^w(\mathbf{f}) - \sigma_i^w(\mathbf{f})] + \gamma_{ub}^w \cdot \sigma_i^w(\mathbf{f}) \tag{C.12b}$$

$$= \mu_i^w(\mathbf{f}) + \gamma_{ub}^w \cdot \sigma_i^w(\mathbf{f}) + \Delta_k^w = H_i^w(\mathbf{f}) + \Delta_k^w \geq H_i^w(\mathbf{f}). \tag{C.12c}$$

Consequently, the above results show that $H_k^w(\mathbf{f}) \geq H_i^w(\mathbf{f})$ for all $k \in \hat{K}_w^2$. Paths in $\hat{K}_w^2$ with higher cumulative costs, i.e., $\Delta_k^w > 0$, strictly lie above the Pareto frontier and are not chosen by travelers.

For the case of $|\hat{K}_w^1| > 1$, without loss of generality, let i and j are two adjacent used paths in $\hat{K}_w^1$. Note that $\hat{\gamma}_k^w = \hat{\gamma}_i^w$ for all paths $k \in \{i+1, \cdots, j-1\} \subseteq \hat{K}_w^2$. We have the following

relationship:

$$H_j^w(\mathbf{f}) = \mu_j^w(\mathbf{f}) + \sum_{l=j}^{|\hat{K}_w|} \hat{\gamma}_l^w(\mathbf{f}) \cdot \left[\sigma_l^w(\mathbf{f}) - \sigma_{l+1}^w(\mathbf{f})\right] \tag{C.13a}$$

$$= \mu_i^w(\mathbf{f}) + \hat{\gamma}_i^w \cdot \left[\sigma_i^w(\mathbf{f}) - \sigma_j^w(\mathbf{f})\right] + \sum_{l=j}^{|\hat{K}_w|} \hat{\gamma}_l^w(\mathbf{f}) \cdot \left[\sigma_l^w(\mathbf{f}) - \sigma_{l+1}^w(\mathbf{f})\right] \tag{C.13b}$$

$$= \mu_i^w(\mathbf{f}) + \sum_{k=i}^{j-1} \hat{\gamma}_i^w \cdot \left[\sigma_k^w(\mathbf{f}) - \sigma_{k+1}^w(\mathbf{f})\right] + \sum_{l=j}^{|\hat{K}_w|} \hat{\gamma}_l^w(\mathbf{f}) \cdot \left[\sigma_l^w(\mathbf{f}) - \sigma_{l+1}^w(\mathbf{f})\right] \tag{C.13c}$$

$$= \mu_i^w(\mathbf{f}) + \sum_{l=i}^{|\hat{K}_w|} \hat{\gamma}_l^w(\mathbf{f}) \cdot \left[\sigma_l^w(\mathbf{f}) - \sigma_{l+1}^w(\mathbf{f})\right] = H_i^w(\mathbf{f}). \tag{C.13d}$$

The above derivations can be applied to any paired adjacent paths in $\hat{K}_w^1$, indicating that all used paths have identical cumulative costs. We now proceed to consider an unused path $k \in \{i+1, \cdots, j-1\} \subseteq \hat{K}_w^2$, it leads to the following relationship:

$$H_k^w(\mathbf{f}) = \mu_k^w(\mathbf{f}) + \hat{\gamma}_i^w \cdot \left[\sigma_k^w(\mathbf{f}) - \sigma_j^w(\mathbf{f})\right] + \sum_{l=j}^{|\hat{K}_w|} \hat{\gamma}_l^w(\mathbf{f}) \cdot \left[\sigma_l^w(\mathbf{f}) - \sigma_{l+1}^w(\mathbf{f})\right] \tag{C.14a}$$

$$= \mu_j^w(\mathbf{f}) + \sum_{l=j}^{|\hat{K}_w|} \hat{\gamma}_l^w(\mathbf{f}) \cdot \left[\sigma_l^w(\mathbf{f}) - \sigma_{l+1}^w(\mathbf{f})\right] + \Delta_k^w \tag{C.14b}$$

$$= H_j^w(\mathbf{f}) + \Delta_k^w \Rightarrow H_k^w(\mathbf{f}) \geq H_j^w(\mathbf{f}). \tag{C.14c}$$

The above derivation can be extended to every unused path between any paired adjacent used paths. For unused paths ordered before the first or after the last path in $\hat{K}_w^1$, the relationship holds by invoking the derivations in Case (1). Thus, the results indicate that the cumulative cost of any unused path is no less than that of used paths.

In summary, all used efficient paths must have identical cumulative costs, while paths with higher cumulative costs than those of used paths must carry no flows. This completes the proof.

C.2 Formulation equivalence

C.2.1 Proof of Proposition 5.4

We need to show that the optimality of the convex Problem (5.11) satisfies Condition (5.9). Let us construct the Lagrangian function of Problem (5.11) as follows:

$$L(\mathbf{f},\mathbf{h}) = \sum_{a\in A}\int_0^{x_a(\mathbf{f})} e_a(z)dz + \sum_{w\in W}\sum_{k\in\hat{K}_w}\sigma_k^w\int_{Q_{k-1}^w(\mathbf{f})}^{Q_k^w(\mathbf{f})}\Gamma_w(z)dz + \sum_{w\in W}h_w\left(q_w - \sum_{k\in K_w} f_k^w\right) \tag{C.15}$$

where h_w denotes the multiplier associated with the flow conservation constraint.

To minimize L with respect to $\mathbf{f}$, he following KKT conditions have to hold:

$$f_k^w\frac{\partial L(\mathbf{f},\mathbf{u})}{\partial f_k^w} = 0, \quad \forall k\in\hat{K}_w, w\in W, \tag{C.16a}$$

$$\frac{\partial L(\mathbf{f},\mathbf{u})}{\partial f_k^w} \geq 0, \quad \forall k\in\hat{K}_w, w\in W, \tag{C.16b}$$

The first-order derivative of L with respect to f_k^w is derived as

$$\frac{\partial L(\mathbf{f},\mathbf{u})}{\partial f_k^w} = \sum_{a\in A}\frac{\partial\int_0^{x_a}e_a(z)dz}{\partial x_a}\cdot\frac{\partial x_a}{\partial f_k^w} + \sigma_k^w\cdot\Gamma_w\left(Q_k^w\right) \tag{C.17a}$$

$$+\sum_{i=k+1}^{|\hat{K}_w|}\sigma_i^w\cdot\left[\Gamma_w\left(Q_i^w\right)-\Gamma_w\left(Q_{i-1}^w\right)\right] \tag{C.17b}$$

$$= \sum_{a\in A}\delta_{ak}e_a(x_a) + \sum_{i=k}^{|\hat{K}_w|-1}\hat{\gamma}_i^w\cdot\left(\sigma_i^w-\sigma_{i+1}^w\right) + \sigma_{|\hat{K}_w|}^w\cdot\hat{\gamma}_{|\hat{K}_w|}^w \tag{C.17c}$$

$$= \mu_k^w + \sum_{i=k}^{|\hat{K}_w|}\hat{\gamma}_i^w\cdot\left(\sigma_i^w-\sigma_{i+1}^w\right). \tag{C.17d}$$

The KKT condition of Problem (5.11) is equivalent to Condition (5.9), which completes the proof.

C.2.2 Proof of Proposition 5.5

The proof is completed by showing the equivalence between Condition (5.9) and VI (5.12) in both directions.

For necessity, we need to show that Condition (5.9) implies VI (5.12). At equilibrium, all flows are assigned to paths with minimum and equal cumulative costs. Fixing cumulative path cost vector at $\mathbf{H}(\mathbf{f}^*)$, the total cumulative path costs can not be reduced by any changes in path flows. Equivalently, for any $\mathbf{f} \in \Omega$,it yields the following relationship:

$$\mathbf{H}(\mathbf{f}^*) \cdot \mathbf{f} \geq \mathbf{H}(\mathbf{f}^*) \cdot \mathbf{f}^* \Rightarrow \mathbf{H}(\mathbf{f}^*) \cdot (\mathbf{f} - \mathbf{f}^*) \geq 0. \tag{C.18}$$

For sufficiency, we need to show that VI (5.12) implies Condition (5.9). Suppose that $\mathbf{f}^*$ is a solution to VI (5.12) but it is not an equilibrium flow. Then, there exists at least two paths $k, k' \in K_w$ such that $f_k^{w*} > 0$ and $H_k^w(\mathbf{f}^*) > H_{k'}^w(\mathbf{f}^*)$. Consider a flow δ shifted from path k to k', it leads to the reduction on total cumulative costs by $\delta \cdot [H_k^w(\mathbf{f}^*) - H_{k'}^w(\mathbf{f}^*)] > 0$. Let $\mathbf{f}$ denote the resulting path flow vector by flow adjustments, we have

$$\mathbf{H}(\mathbf{f}^*) \cdot \mathbf{f} < \mathbf{H}(\mathbf{f}^*) \cdot \mathbf{f}^* \Rightarrow \mathbf{H}(\mathbf{f}^*) \cdot (\mathbf{f} - \mathbf{f}^*) < 0, \tag{C.19}$$

which is a contradiction to the VI.

This completes the proof.

C.3 Column generation algorithm

C.3.1 Labeling mean-variance bi-objective algorithm

The one-to-all labeling mean-variance bi-objective algorithm is presented below.

The algorithm initializes by assigning a label to the origin node r. Each iteration involves removing the foremost label from the sorted temporary label list SEL, based on mean as the primary criterion and variance as the secondary criterion. Subsequently, a new label $\bar{l}$ is generated for every outgoing link from the source node. If the label is dominated by existing ones, it is discarded; otherwise, it is retained while eliminating all dominated labels. Termination occurs when SEL becomes empty. To get a set of non-dominated paths between OD pair $w = (r, d)$, one simply traces each efficient label $l \in L_d$ backward to its predecessor label until reaching node r.

As per Sedeno-Noda and Colebrook (2019), the algorithm's complexity is roughly $O(R \times$

Algorithm 12 One-to-all labeling mean-variance bi-objective algorithm.

Input: Fixed $\{e_a(\mathbf{f}^n)\}$ and $\{s_a(\mathbf{f}^n)\}$ for all $a \in A$, origin node r.
Output: A set of efficient labels L_i for each node $i \in N$.

1: **Initialization.** Let L_i be the label set for node i, where a label is defined to include four attributes (i.e., source node, mean costs from r, variance costs from r, predecessor label). Initialize a label $l = [r, 0, 0, \text{NULL}]$ and a scan list $SEL = \{l\}$. Set $L_r = \{l\}$ and $L_i = \emptyset, \forall i \in N \setminus \{r\}$. We use $l_{[n]}$ to denote the nth attribute of label l.
2: Select the first label l from SEL in the lexicographical order, and delete it from SEL.
3: **for** any outgoing link $a = (a^-, a^+)$ from the source node of l **do**
4: Initialize a new label $\bar{l} = \{a^+, l_{[2]} + e_a, l_{[3]} + s_a^2, l\}$.
5: If new label $\bar{l}$ is not dominated by any label in L_{a^+}, then update $L_{a^+} = L_{a^+} \cup \{\bar{l}\}$, and $SEL = SEL \cup \{\bar{l}\}$.
6: If there exists label $l' \in L_{a^+}$ is dominated by new label $\bar{l}$, then update $L_{a^+} = L_{a^+} \setminus \{l'\}$, and $SEL = SEL \setminus \{l'\}$.
7: **end for**
8: If $SEL = \emptyset$, stop; otherwise go to Step 2.

$\log |N| + |A| \times R_{\text{max}})$, where $R = \sum_{i=1}^{|N|} R_i$ denotes the total number of non-dominated labels, R_i represents the number of non-dominated labels at node i, and $R_{\text{max}} = \max_{i=1,\ldots,|N|} R_i$.

C.3.2 Graphical illustration for identifying supported efficient paths

The column generation algorithm 7 mainly consists of two sequential phases: (1) Starting and ending path setup, and (2) Sequential slop verification. Figure C.1 provides graphical justifications of the procedure with nine efficient paths. In Figure C.1(a) for Phrase (1), P2 and P3 yield the shortest cost for γ_{lb}^w, while P7 and P8 do so for γ_{ub}^w. Furthermore, P2 and P8 are excluded because no potential flows will be assigned to them. Additionally, since the frontier should be constructed between P3 and P7, P1 and P9 are also excluded. We are now ready to consider Phrase (2) with candidate paths P4, P5, and P6, as depicted in Figure C.1(b). This process begins with P3, it includes P4 and P6 as the sequential slope is increasing, excluding P5 as the slope between P4 and P5 is greater than γ_{ub}^w. For the slope between P6 and P7, the non-increasing slope prompts a comparison with previously included paths (P6 and P4) until Condition (b) is met, leading to the exclusion of P6 and P4. The procedure concludes with the selection of P3 for slopes in the range $[\gamma_{lb}^w, \text{slope}_{37}]$, and P7 for slopes in $[\text{slope}_{37}, \gamma_{ub}^w]$.

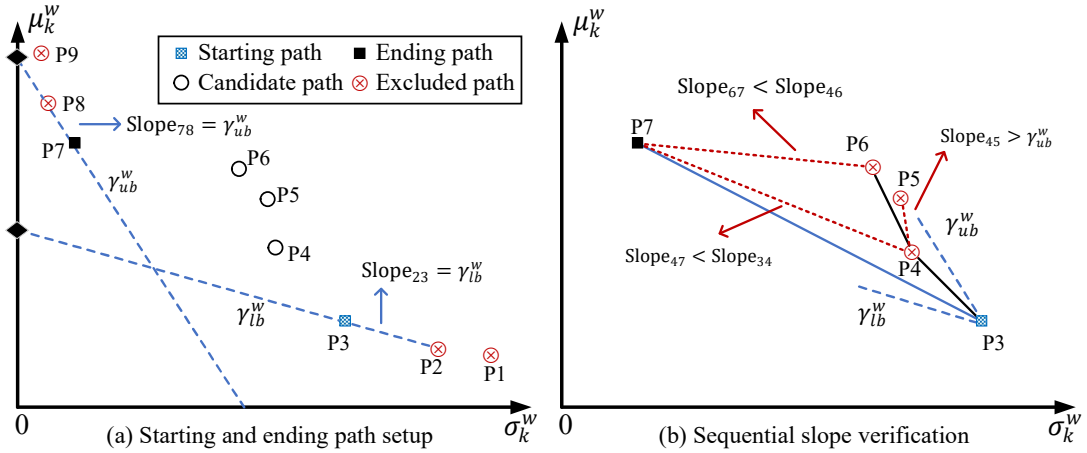


Figure C.1: Justifications of Pareto frontier construction.

C.4 Path-based algorithm for DMRTA

The CG-GP algorithm follows the same fundamental logic when solving either CMRTA or DMRTA problems. However, the primary distinctions between Algorithm 6 (CMRTA) and Algorithm 13 (DMRTA) lie in two key steps: Step 2 (Solving column generation problem) and Step 4 (Solving restricted master problem). The following two remarks provide a detailed analysis of the algorithmic complexity.

Remark (Solving column generation problem). In Step 2, the procedure for solving both DMRTA and CMRTA problems consists of two stages. The first stage, which is common to both problems, runs the one-to-all bi-objective labeling algorithm to find all possible efficient paths for OD pairs with the same origin. The time complexity of this stage is discussed in Appendix C.3.1. The second stage generates promising paths for considered user classes given the set of efficient paths $\tilde{K}_w$ for each OD pair. For CMRTA, the time complexity is $O(|\tilde{K}_w| \cdot \log(|\tilde{K}_w|))$ (cf. Section 5.4.1). For DMRTA, the procedure directly searches for the minimum-cost path for each user class, resulting in the time complexity of $O(|\tilde{K}_w| \cdot |M_w|)$. Notably, generating new paths for the DMRTA problem would incur higher computational time for a sufficiently large number of $|M_w|$.

Remark (Solving restricted master problem). In Step 4, we apply the gradient projection algorithm to sequentially solve the paired path problem defined upon the working path set. The key distinction between CMRTA and DMRTA lies in the selection of paired paths: CMRTA uses

paths with the largest cumulative cost difference, while DMRTA uses those with the largest path cost difference. The one-time projection involves traversing stored links on paired paths four times: twice for computing first-order derivatives (Step 4-2) and twice for updating affected link flows (Step 4-5). Supposing there are two working paths for all OD pairs, with each path containing approximately $\sqrt{|N|}$ links (Jayakrishnan et al. 1994). Under these conditions, the time complexity for a one-time projection across all OD pairs is roughly $O(4|W|\sqrt{|N|})$ for CMRTA and $O(4\sum_w |M_w|\sqrt{|N|})$ for DMRTA. This analysis reveals that DMRTA's time complexity scales linearly with the number of user classes, potentially resulting in higher computational costs as the number of classes increases.

Algorithm 13 The CG-GP algorithm for solving DMRTA problems.

Input: Network $G(N, A)$, a set of user classes M_w for each $w \in W$, class-specific OD demand vector $\check{\mathbf{q}} = (q_{wm})_{w \in W, m \in M_w}$, and link performance functions $(e_a(\cdot))_{a \in A}$ and $(s_a(\cdot))_{a \in A}$.

Output: Equilibrium class-specific path flow $\check{\mathbf{f}}^*$.

1: **Initialization.** Initialize counter $n = 0$, assign class-specific demand to minimum-mean path k for each $m \in M_w, w \in W$, i.e., $f_k^{wm} = q_{wm}$. Set the working path set $\mathcal{K}_{wm} = \{k\}$ for class m of OD pair w.

2: **Solving column generation problem.** (i) Set $n = n + 1$; (ii) Obtain the set of efficient paths $\tilde{K}_w$ for OD pair w by one-to-all labeling bi-objective algorithm in Appendix C.3.1; (iii) For each $m \in M_w, w \in W$, identify path $k = \arg\min_{k'} \{k' | C_{k'}^w(\check{\mathbf{f}}^n; \gamma_m), k' \in \tilde{K}_w\}$, and update $\mathcal{K}_{wm} = \mathcal{K}_{wm} \cup \{k\}$.

3: **Termination criterion.** The algorithm is terminated when achieving a satisfactory outer convergence, which is measured by the following relative gap:

$$RG_{outer} = 1 - \frac{\sum_{w \in W} \sum_{m \in M_w} q_{wm} \cdot u_{wm}(\check{\mathbf{f}}^n)}{\sum_{w \in W} \sum_{m \in M_w} \sum_{k \in \mathcal{K}_{wm}} f_k^{wm} \cdot C_k^w(\check{\mathbf{f}}^n; \gamma_m)}, \tag{C.20}$$

where $u_{wm}(\check{\mathbf{f}}^n)$ is the minimum path cost obtained by column generation at iteration n.

4: **Solving restricted master problem.** Initialize the inner counter $l = 0$. (i) Set $l = l + 1$; (ii) For each $m \in M_w, w \in W$, update f_k^{wm} with gradient projection as follows:

4-1. Initialize k and $\bar{k}$ as the shortest-cost path and longest-cost path respectively, set iteration counter $r = 0$ and stepsize $\alpha^{(0)} = 0.1$.

4-2. Update path cost difference $\tilde{C}_{\bar{k}k}^{wm(r)} = C_{\bar{k}}^w(f_k^{wm(r)}; \gamma_m) - C_k^w(f_k^{wm(r)}; \gamma_m)$.

4-3. If $\tilde{C}_{\bar{k}k}^{wm}(f_k^{wm(r)}) = 0$ or $(\tilde{C}_{\bar{k}k}^{wm(r)} > 0$ and $f_k^{wm(r)} = 0)$, then stop updating f_k^{wm}; otherwise move to Step 4-4.

4-4. Update $f_k^{wm(r+1)} = \min\left\{\max\left\{f_k^{wm(r)} - \alpha^{(r)} \cdot \tilde{C}_{\bar{k}k}^{wm}(f_k^{wm(r)}), 0\right\}, \phi_{k\bar{k}}^{wm}\right\}$, where $\alpha^{(r)}$ is the positive stepsize determined by line search as in Section 5.4.2, and $\phi_{k\bar{k}}^{wm} = f_k^{wm(0)} + f_{\bar{k}}^{wm(0)}$ denotes the total flows on paired paths. Set $f_{\bar{k}}^{wm(r+1)} = \phi_{k\bar{k}}^{wm} - f_k^{wm(r+1)}$.

4-5. Update $e_a(x_a)$ and $s_a(x_a)$ for each link a on path k and $\bar{k}$.

4-6. Set $r = r + 1$, and return to Step 4-2.

(iii) Compute the relative gap for inner convergence by

$$RG_{inner} = 1 - \frac{\sum_{w \in W} \sum_{m \in M_w} \sum_{k \in \mathcal{K}_{wm}} q_{wm} \cdot C_{min}^w(\check{\mathbf{f}}^l; \gamma_m)}{\sum_{w \in W} \sum_{m \in M_w} \sum_{k \in \mathcal{K}_{wm}} f_k^{wm} \cdot C_k^w(\mathbf{f}^l; \gamma_m)}, \tag{C.21}$$

where $C_{min}^w(\check{\mathbf{f}}^l; \gamma_m) = \min_{k \in \mathcal{K}_{wm}} C_k^w(\check{\mathbf{f}}^l; \gamma_m)$ is the minimum cumulative cost at inner iteration l. Stop and go to Step 5 if *inner convergence* is achieved, otherwise return to (i).

5: **Dropping.** Remove unused paths for each OD pair, and return to Step 2.

Bibliography

Abdel-Aty, M., Kitamura, R., Jovanis, P.P., 1995. Investigating effect of travel time variability on route choice using repeated-measurement stated preference data. 1493.

Acemoglu, D., Makhdoumi, A., Malekian, A., Ozdaglar, A., 2018. Informational braess' paradox: The effect of information on traffic congestion. Operations Research 66, 893–917.

Amirgholy, M., Gonzales, E.J., 2017. Efficient frontier of route choice for modeling the equilibrium under travel time variability with heterogeneous traveler preferences. Economics of Transportation 11-12, 1–14.

Antúnez, R.C., Levin, M.W., 2024. How does eco-routing affect total system emissions? City network predictions from user equilibrium models. IEEE Transactions on Intelligent Transportation Systems 25, 1408–1417.

Armijo, L., 1966. Minimization of functions having lipschitz continuous first partial derivatives. Pacific Journal of Mathematics 16, 1–3.

Bao, Y., Gao, Z., Xu, M., Sun, H., Yang, H., 2015. Travel mental budgeting under road toll: An investigation based on user equilibrium. Transportation Research Part A: Policy and Practice 73, 1–17.

Barzilai, J., Borwein, J.M., 1988. Two-point step size gradient methods. IMA journal of numerical analysis 8, 141–148.

Bazaraa, M.S., Sherali, H.D., Shetty, C.M., 2013. Nonlinear programming: theory and algorithms. John wiley & sons.

Beckmann, M., McGuire, C.B., Winsten, C.B., 1956. Studies in the Economics of Transportation. Yale University Press, New Haven.

Ben-Akiva, M., Bolduc, D., Bradley, M., 1993. Estimation of travel choice models with randomly distributed values of time. Transportation Research Record 1413, 88–97.

Bertsekas, D.P., 1976. On the goldstein-levitin-polyak gradient projection method. IEEE Transactions on Automatic Control 21, 174–184.

Bhat, C.R., Sardesai, R., 2006. The impact of stop-making and travel time reliability on commute mode choice. Transportation Research Part B: Methodological 40, 709–730.

Boyce, D., Ralevic-Dekic, B., Bar-Gera, H., 2004. Convergence of traffic assignments: how much is enough? Journal of Transportation Engineering 130, 49–55.

Brownstone, D., Small, K.A., 2005. Valuing time and reliability: assessing the evidence from road pricing demonstrations. Transportation Research Part A: Policy and Practice 39, 279–293.

Bureau of Public Roads, 1964. Traffic assignment manual. US Department of Commerce, Urban Planning Division,Washington, DC. .

Calderone, D., Dong, R., Sastry, S.S., 2017. External-cost continuous-type wardrop equilibria in routing games, in: 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC), IEEE. pp. 1–6.

Cantarella, G.E., Binetti, M., 1998. Stochastic equilibrium traffic assignment with value-of-time distributed among users. International Transactions in Operational Research 5, 541–553.

Carrion, C., Levinson, D., 2012. Value of travel time reliability: A review of current evidence. Transportation research part A: policy and practice 46, 720–741.

Chen, A., Jayakrishnan, R., 1998. A path-based gradient projection algorithm: effects of equilibration with a restricted path set under two flow update policies. In: Proceedings of the Transportation Research Board Annual Meeting .

Chen, A., Lee, D.H., Jayakrishnan, R., 2002a. Computational study of state-of-the-art path-based traffic assignment algorithms. Mathematics and Computers in Simulation 59, 509–518.

Chen, A., Oh, J.S., Park, D., Recker, W., 2010. Solving the bicriteria traffic equilibrium problem with variable demand and nonlinear path costs. Applied Mathematics and Computation 217, 3020–3031.

Chen, A., Pravinvongvuth, S., Xu, X., Ryu, S., Chootinan, P., 2012a. Examining the scaling effect and overlapping problem in logit-based stochastic user equilibrium models. Transportation Research Part A: Policy and Practice 46, 1343–1358.

Chen, A., Ryu, S., Xu, X., Choi, K., 2014. Computation and application of the paired combinatorial logit stochastic user equilibrium problem. Computers & operations research 43, 68–77.

Chen, A., Subprasom, K., 2007. Analysis of regulation and policy of private toll roads in a build-operate-transfer scheme under demand uncertainty. Transportation Research Part A: Policy and Practice 41, 537–558.

Chen, A., Yang, H., Lo, H.K., Tang, W.H., 2002b. Capacity reliability of a road network: an assessment methodology and numerical results. Transportation Research Part B: Methodological 36, 225–252.

Chen, A., Zhou, Z., Chootinan, P., Ryu, S., Yang, C., Wong, S., 2011a. Transport network design problem under uncertainty: a review and new developments. Transport Reviews 31, 743–768.

Chen, A., Zhou, Z., Ryu, S., 2011b. Modeling physical and environmental side constraints in traffic equilibrium problem. International Journal of Sustainable Transportation 5, 172–197.

Chen, A., Zhou, Z., Xu, X., 2012b. A self-adaptive gradient projection algorithm for the nonadditive traffic equilibrium problem. Computers & Operations Research 39, 127–138.

Chen, B.Y., Lam, W.H., Sumalee, A., Li, Q., Li, Z.C., 2012c. Vulnerability analysis for large-scale and congested road networks with demand uncertainty. Transportation Research Part A: Policy and Practice 46, 501–516.

Chen, B.Y., Lam, W.H., Sumalee, A., Shao, H., 2011c. An efficient solution algorithm for solving multi-class reliability-based traffic assignment problem. Mathematical and Computer Modelling 54, 1428–1439.

Chen, X., Liu, Z., Zhang, K., Wang, Z., 2020. A parallel computing approach to solve traffic assignment using path-based gradient projection algorithm. Transportation Research Part C: Emerging Technologies 120, 102809.

Chen, Z., Ye, H., Liu, B., Xue, W., 2021. Analysis of road capacity and franchise price decision delegation in toll road bot project. Transportation Research Part E: Logistics and Transportation Review 146, 102213.

Cirillo, C., Axhausen, K.W., 2006. Evidence on the distribution of values of travel time savings from a six-week diary. Transportation Research Part A: policy and practice 40, 444–457.

Clark, A., Sumalee, A., Shepherd, S., Connors, R., 2009. On the existence and uniqueness of first best tolls in networks with multiple user classes and elastic demand. International Journal of Control 5, 141–157.

Dafermos, S., 1981. A multicriteria route-mode choice traffic equilibrium model. Lefschetz Center for Dynamical Systems, Brown University, Providence, RI .

Dafermos, S.C., 1972. The traffic assignment problem for multiclass-user transportation networks. Transportation science 6, 73–87.

Dafermos, S.C., Sparrow, F.T., 1969. The traffic assignment problem for a general network. Journal of Research of the National Bureau of Standards B 73, 91–118.

Daganzo, C.F., Sheffi, Y., 1977. On stochastic models of traffic assignment. Transportation science 11, 253–274.

Dantzig, G.B., 1951. Application of the simplex method to a transportation problem. Activity Analysis and Production and Allocation .

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, 293–317.

Di, S., Pan, C., Ran, B., 2008. Stochastic multiclass traffic assignment with consideration of risk-taking behaviors. Transportation research record 2085, 111–123.

Dial, R.B., 1979. A model and algorithm for multicriteria route-mode choice. Transportation Research Part B: Methodological 13, 311–316.

Dial, R.B., 1996. Bicriterion traffic assignment: Basic theory and elementary algorithms. Transportation science 30, 93–111.

Dial, R.B., 1997. Bicriterion traffic assignment: Efficient algorithms plus examples. Transportation Research Part B: Methodological 31, 357–379.

Dial, R.B., 1999a. Network-optimized road pricing: Part I: A parable and a model. Operations Research 47, 54–64.

Dial, R.B., 1999b. Network-optimized road pricing: Part II: Algorithms and examples. Operations Research 47, 327–336.

Dial, R.B., 2006. A path-based user-equilibrium traffic assignment algorithm that obviates path storage and enumeration. Transportation Research Part B: Methodological 40, 917–936.

Dijkstra, E.W., 1959. A note on two problems in connexion with graphs, in: Numerische Mathematik. volume 1, p. 269–271.

de Dios Ortúzar, J., Willumsen, L.G., 2011. Modelling transport. John wiley & sons.

Du, M., Tan, H., Chen, A., 2021. A faster path-based algorithm with Barzilai-Borwein step size for solving stochastic traffic equilibrium models. European Journal of Operational Research 290, 982–999.

Du, M., Zhou, J., Chen, A., Tan, H., 2022. Modeling the capacity of multimodal and intermodal urban transportation networks that incorporate emerging travel modes. Transportation Research Part E: Logistics and Transportation Review 168, 102937.

Ehrgott, M., 2005. Multicriteria optimization. volume 491. Springer Science & Business Media.

Fakhrmoosavi, F., Zockaie, A., Abdelghany, K., 2021. Incorporating travel time reliability in equitable congestion pricing schemes for heterogeneous users and bimodal networks. Transportation Research Record 2675, 754–768.

Fan, Y., Ding, J., Liu, H., Wang, Y., Long, J., 2022. Large-scale multimodal transportation network models and algorithms-part I: The combined mode split and traffic assignment problem. Transportation Research Part E: Logistics and Transportation Review 164, 102832.

Fisk, C., 1980. Some developments in equilibrium traffic assignment. Transportation Research Part B: Methodological 14, 243–255.

Florian, M., 2006. Network equilibrium models for analyzing toll highways, in: Mathematical and computational models for congestion charging. Springer, pp. 105–115.

Frank, M., Wolfe, P., 1956. An algorithm for quadratic programming. Naval research logistics quarterly 3, 95–110.

Fredman, M.L., Tarjan, R.E., 1987. Fibonacci heaps and their uses in improved network optimization algorithms. Journal of the ACM (JACM) 34, 596–615.

Fu, X., Lam, W.H., Chen, B.Y., 2014. A reliability-based traffic assignment model for multimodal transport network under demand uncertainty. Journal of Advanced Transportation 48, 66–85.

Gabriel, S.A., Bernstein, D., 1997. The traffic equilibrium problem with nonadditive path costs. Transportation Science 31, 337–348.

Galligari, A., Sciandrone, M., 2019. A computational study of path-based methods for optimal traffic assignment with both inelastic and elastic demand. Computers & Operations Research 103, 158–166.

Ge, X.Y., Li, Z.C., Lam, W.H., Choi, K., 2014. Energy-sustainable traffic signal timings for a congested road network with heterogeneous users. IEEE Transactions on Intelligent Transportation Systems 15, 1016–1025.

Goldstein, A.A., 1964. Convex programming in hilbert space. Bulletin of the American Mathematical Society 70, 709–710.

Guo, X., Yang, H., 2009. User heterogeneity and bi-criteria system optimum. Transportation Research Part B: Methodological 43, 379–390.

Guo, X., Yang, H., 2010. Pareto-improving congestion pricing and revenue refunding with multiple user classes. Transportation Research Part B: Methodological 44, 972–982.

Han, C., Xu, G., Pervez, A., Gao, F., Huang, H., Pei, X., Zhang, Y., 2023. Modeling traveler's speed-route joint choice behavior with heterogeneous safety concern. Analytic Methods in Accident Research 37, 100253.

Han, D., Yang, H., 2008. The multi-class, multi-criterion traffic equilibrium and the efficiency of congestion pricing. Transportation Research Part E: Logistics and Transportation Review 44, 753–773.

Han, L., Ukkusuri, S., Doan, K., 2011. Complementarity formulations for the cell transmission model based dynamic user equilibrium with departure time choice, elastic demand and user heterogeneity. Transportation Research Part B: Methodological 45, 1749–1767.

Han, L., Zhu, C., Wang, D.Z., Sun, H., Tan, Z., Meng, M., 2019. Discrete-time dynamic road congestion pricing under stochastic user optimal principle. Transportation Research Part E: Logistics and Transportation Review 131, 24–36.

Hartman, P., Stampacchia, G., 1966. On some non-linear elliptic differential-functional equations. Acta mathematica 115, 271–310.

Hossan, M.S., Asgari, H., Jin, X., 2016. Investigating preference heterogeneity in value of time (VOT) and value of reliability (VOR) estimation for managed lanes. Transportation Research Part A: Policy and Practice 94, 638–649.

Huang, H., Han, C., Xu, G., Jiang, M., Wong, S., Mazharul Haque, M., 2020. Incorporating safety reliability into route choice model: Heterogeneous crash risk aversions. Analytic Methods in Accident Research 25, 100112.

Huang, H.J., Li, Z.C., 2007. A multiclass, multicriteria logit-based traffic equilibrium assignment model under atis. European Journal of Operational Research 176, 1464–1477.

Jaber, X., O'mahony, M., 2009. Mixed stochastic user equilibrium behavior under traveler information provision services with heterogeneous multiclass, multicriteria decision making. Journal of Intelligent Transportation Systems 13, 188–198.

Jackson, W.B., Jucker, J.V., 1982. An empirical study of travel time variability and travel choice behavior. Transportation Science 16, 460–475.

Jafari, E., Pandey, V., Boyles, S.D., 2017. A decomposition approach to the static traffic assignment problem. Transportation Research Part B: Methodological 105, 270–296.

Jayakrishnan, R., Tsai, W.T., Prashker, J.N., Rajadhyaksha, S., 1994. A faster path-based algorithm for traffic assignment. Transportation Research Board 1443, 75–83.

Jia, A., Zhou, X., Rouphail, N.M., 2012. Traffic mobility impact of mileage-based user fees on traveler route choice behavior and network performance: Planning-level traffic equilibrium-based approach. Transportation research record 2302, 164–173.

Jiang, G., Fosgerau, M., Lo, H.K., 2020. Route choice, travel time variability, and rational inattention. Transportation Research Part B: Methodological 132, 188–207.

Jiang, Y., 2024. A variational inequality formulation for stochastic user equilibrium with a bounded choice set. Computers & Operations Research 167, 106677.

Kashmiri, F.A., Lo, H.K., 2022. Routing of autonomous vehicles for system optimal flows and average travel time equilibrium over time. Transportation Research Part C: Emerging Technologies 143, 103818.

Kashmiri, F.A., Lo, H.K., 2024. Routing of multi-modal autonomous vehicles for system optimal flows and average travel cost equilibrium over time. Transportation research part C: emerging technologies 159, 104483.

Kitthamkesorn, S., Chen, A., 2013. A path-size weibit stochastic user equilibrium model. Transportation Research Part B: Methodological 57, 378–397. URL: https://www.sciencedirect.com/science/article/pii/S0191261513000982, doi:https://doi.org/10.1016/j.trb.2013.06.001.

Lam, T.C., Small, K.A., 2001. The value of time and reliability: measurement from a value pricing experiment. Transportation Research Part E: Logistics and Transportation Review 37, 231–251.

Lam, W.H., Huang, H.J., 1992. A combined trip distribution and assignment model for multiple user classes. Transportation Research Part B: Methodological 26, 275–287.

Larsson, T., Lindberg, P.O., Patriksson, M., Rydergren, C., 2002. On traffic equilibrium models with a nonlinear time/money relation, in: Transportation planning. Springer, pp. 19–31.

Lawphongpanich, S., Hearn, D.W., 1984. Simplical decomposition of the asymmetric traffic assignment problem. Transportation Research Part B: Methodological 18, 123–133.

LeBlanc, L.J., Morlok, E.K., Pierskalla, W.P., 1975. An efficient approach to solving the road network equilibrium traffic assignment problem. Transportation research 9, 309–318.

Leurent, F., 1993. Cost versus time equilibrium over a network. European Journal of Operational Research 71, 205–221.

Leurent, F., 1996. The theory and practice of a dual criteria assignment model with a continuously distributed value-of-time, in: ISTTT, Pergamon. pp. pp–455.

Leurent, F., 1998a. Sensitivity and error analysis of the dual criteria traffic assignment model. Transportation Research Part B: Methodological 32, 189–204.

Leurent, F.M., 1998b. Multicriteria assignment modeling: making explicit the determinants of mode or path choice, in: Equilibrium and Advanced Transportation Modelling. Springer, pp. 153–174.

Levitin, E.S., Polyak, B.T., 1966. Constrained minimization methods. USSR Computational mathematics and mathematical physics 6, 1–50.

Li, T., Sun, H., Wu, J., Ge, Y.e., 2017. Optimal toll of new highway in the equilibrium framework of heterogeneous households' residential location choice. Transportation Research Part A: Policy and Practice 105, 123–137.

Li, T., Xu, M., Sun, H., Xiong, J., Dou, X., 2023a. Stochastic ridesharing equilibrium problem with compensation optimization. Transportation Research Part E: Logistics and Transportation Review 170, 102999.

Li, X., Xie, C., Bao, Z., 2022. A multimodal multicommodity network equilibrium model with service capacity and bottleneck congestion for china-europe containerized freight flows. Transportation Research Part E: Logistics and Transportation Review 164, 102786.

Li, Z., Hensher, D.A., Rose, J.M., 2010. Willingness to pay for travel time reliability in passenger transport: A review and some new empirical evidence. Transportation Research Part E: Logistics and Transportation Review 46, 384–403.

Li, Z., Li, G., Xu, Z., Chen, A., 2023b. Multiclass bi-criteria traffic assignment without class-specific variables: An alternative formulation and a subgradient projection algorithm. Transportation Research Part E: Logistics and Transportation Review 176, 103210.

Li, Z., Xu, Z., Li, G., Chen, A., 2025. A unified gradient projection algorithm for solving both discrete and continuous multiclass bi-criteria traffic assignment problems. Computers & Operations Research , 107320.

Lindberg, P.O., Engelson, L., 2015. Tolled multi-class traffic equilibria and toll sensitivities. EURO Journal on Transportation and Logistics 4, 197–222.

Liu, B.q., Huang, C.c., 2016. Multi-class time reliability-based congestion pricing model based on a degradable transportation network. Applied Mathematical Modelling 40, 3483–3497.

Liu, H.X., He, X., Recker, W., 2007. Estimation of the time-dependency of values of travel time and its reliability from loop detector data. Transportation Research Part B: Methodological 41, 448–461.

Liu, P., Liu, J., Ong, G.P., Tian, Q., 2020. Flow pattern and optimal capacity in a bi-modal traffic corridor with heterogeneous users. Transportation Research Part E: Logistics and Transportation Review 133, 101831.

Liu, Z., Meng, Q., Wang, S., 2013. Speed-based toll design for cordon-based congestion pricing scheme. Transportation Research Part C: Emerging Technologies 31, 83–98.

Lo, H.K., Chen, A., 2000. Traffic equilibrium problem with route-specific costs: formulation and algorithms. Transportation Research Part B: Methodological 34, 493–513.

Lo, H.K., Luo, X., Siu, B.W., 2006. Degradable transport network: travel time budget of travelers with heterogeneous risk aversion. Transportation Research Part B: Methodological 40, 792–806.

Lu, C.C., Mahmassani, H.S., Zhou, X., 2008. A bi-criterion dynamic user equilibrium traffic assignment model and solution algorithm for evaluating dynamic road pricing strategies. Transportation Research Part C: Emerging Technologies 16, 371–389.

Lu, C.C., Mahmassani, H.S., Zhou, X., 2009. Equivalent gap function-based reformulation and solution algorithm for the dynamic user equilibrium problem. Transportation Research Part B: Methodological 43, 345–364.

Marcotte, P., 1999. Reformulations of a bicriterion equilibrium model, in: Reformulation: nonsmooth, piecewise smooth, semismooth and smoothing methods. Springer, pp. 269–291.

Marcotte, P., Nguyen, S., Tanguay, K., 1996. Implementation of an efficient algorithm for the multiclass traffic assignment problem, in: Transportation and Traffic Theory. Proceedings of the 13th International Symposium on Transportation and Traffic Theory, Pergamon, Oxford, Citeseer. pp. 217–236.

Marcotte, P., Zhu, D., 1996. An efficient algorithm for a bicriterion traffic assignment problem, in: Advanced methods in transportation analysis. Springer, pp. 63–73.

Marcotte, P., Zhu, D., 2009. Existence and computation of optimal tolls in multiclass network equilibrium problems. Operations Research Letters 37, 211–214.

Marcotte, P., Zhu, D.L., 1997. Equilibria with infinitely many differentiated classes of customers. Complementarity and variational problems, state of art , 234–258.

McFadden, D., Train, K., 2000. Mixed MNL models for discrete response. Journal of applied Econometrics 15, 447–470.

Meng, Q., Liu, Z., Wang, S., 2012. Optimal distance tolls under congestion pricing and continuously distributed value of time. Transportation Research Part E: Logistics and Transportation Review 48, 937–957.

Mi, H., Chen, S., Ping, J., Yan, Z., 2023. Traffic equilibrium considering heterogeneity across electric vehicles. IEEE Transactions on Intelligent Transportation Systems 24, 14944–14956.

Mirchandani, P., Soroush, H., 1987. Generalized traffic equilibrium with probabilistic travel times and perceptions. Transportation Science 21, 133–152.

Miwa, T., Okada, Y., Morikawa, T., 2010. Applying a structured dispersion parameter to multiclass stochastic user equilibrium assignment model. Transportation research record 2196, 142–149.

Nagurney, A., 2000. A multiclass, multicriteria traffic network equilibrium model. Mathematical and Computer Modelling 32, 393–411.

Nagurney, A., Dong, J., 2002a. A multiclass, multicriteria traffic network equilibrium model with elastic demand. Transportation Research Part B: Methodological 36, 445–469.

Nagurney, A., Dong, J., 2002b. Urban location and transportation in the information age: a multiclass, multicriteria network equilibrium perspective. Environment and Planning B: Planning and Design 29, 53–74.

Nagurney, A., Dong, J., Mokhtarian, P.L., 2002. Traffic network equilibrium and the environment: A multicriteria decision-making perspective, in: Computational methods in decision-making, economics and finance. Springer, pp. 501–523.

Nagurney, A., Dong, J., Mokhtarian, P.L., 2003. A space time network for telecommuting versus commuting decision making. Papers in Regional Science 82, 451–473.

Nesterov, Y., 2003. Introductory lectures on convex optimization: A basic course. volume 87. Springer Science & Business Media.

Nie, Y.M., 2011. Multi-class percentile user equilibrium with flow-dependent stochasticity. Transportation Research Part B: Methodological 45, 1641–1659.

Nikolova, E., Stier-Moses, N.E., 2014. A mean-risk model for the traffic assignment problem with stochastic travel times. Operations Research 62, 366–382.

Nikolova, E., Stier-Moses, N.E., 2015. The burden of risk aversion in mean-risk selfish routing, in: Proceedings of the Sixteenth ACM Conference on Economics and Computation, pp. 489–506.

Noruzoliaee, M., Zou, B., Liu, Y., 2018. Roads in transition: Integrated modeling of a manufacturer-traveler-infrastructure system in a mixed autonomous/human driving environment. Transportation Research Part C: Emerging Technologies 90, 307–333.

Ohazulike, A.E., Still, G., Kern, W., van Berkum, E.C., 2013. An origin-destination based road pricing model for static and multi-period traffic assignment problems. Transportation Research Part E: Logistics and Transportation Review 58, 1–27.

Ordónez, F., Stier-Moses, N.E., 2010. Wardrop equilibria with risk-averse users. Transportation Science 44, 63–86.

Papadopoulos, A.A., Kordonis, I., Dessouky, M.M., Ioannou, P.A., 2021. Personalized pareto-improving pricing-and-routing schemes for near-optimum freight routing: An alternative approach to congestion pricing. Transportation Research Part C: Emerging Technologies 125, 103004.

Patriksson, M., 2015. The traffic assignment problem: models and methods. Courier Dover Publications.

Perederieieva, O., Ehrgott, M., Raith, A., Wang, J.Y., 2015. A framework for and empirical study of algorithms for traffic assignment. Computers & Operations Research 54, 90–107.

Prakash, A.A., Seshadri, R., Srinivasan, K.K., 2018. A consistent reliability-based user-equilibrium problem with risk-averse users and endogenous travel time correlations: Formulation and solution algorithm. Transportation Research Part B: Methodological 114, 171–198.

Rezaee, H., Hickman, M., 2015. Investigating the distribution of the value of time via multiclass traffic assignment. Transportation Research Record 2530, 65–72.

Ryu, S., Chen, A., Choi, K., 2014. A modified gradient projection algorithm for solving the elastic demand traffic assignment problem. Computers & operations research 47, 61–71.

Ryu, S., Chen, A., Choi, K., 2016. Solving the stochastic multi-class traffic assignment problem with asymmetric interactions, route overlapping, and vehicle restrictions. Journal of Advanced Transportation 50, 255–270.

Ryu, S., Chen, A., Su, J., Choi, K., 2021. A multi-class, multi-criteria bicycle traffic assignment model. International Journal of Sustainable Transportation 15, 524–540.

Sedeno-Noda, A., Colebrook, M., 2019. A biobjective dijkstra algorithm. European Journal of Operational Research 276, 106–118.

Sedeño-Noda, A., Raith, A., 2015. A dijkstra-like method computing all extreme supported non-dominated solutions of the biobjective shortest path problem. Computers & Operations Research 57, 83–94.

Shao, H., Lam, W.H., Meng, Q., Tam, M.L., 2006a. Demand-driven traffic assignment problem based on travel time reliability. Transportation Research Record 1985, 220–230.

Shao, H., Lam, W.H., Tam, M.L., 2006b. A reliability-based stochastic traffic assignment model for network with multiple user classes under uncertainty in demand. Networks and Spatial Economics 6, 173–204.

Shao, H., Lam, W.H., Tam, M.L., Yuan, X.M., 2008. Modelling rain effects on risk-taking behaviours of multi-user classes in road networks with uncertainty. Journal of Advanced Transportation 42, 265–290.

Sheffi, Y., 1985. Urban transportation networks: Equilibrium analysis with mathematical programming methods. volume 6. Prentice-Hall, Englewood Cliffs, NJ.

Site, P.D., Filippi, F., 2011. Stochastic user equilibrium and value-of-time analysis with reference-dependent route choice. European Journal of Transport and Infrastructure Research 11.

Slavin, H., Brandon, J., Lam, J., Sundaram, S., 2014. A brief note on toll road assignment methods. Technical Report. Caliper Corporation.

Small, K.A., Winston, C., Yan, J., 2005. Uncovering the distribution of motorists' preferences for travel time and reliability. Econometrica 73, 1367–1382.

Small, K.A., Yan, J., 2001. The value of "value pricing" of roads: Second-best pricing and product differentiation. Journal of Urban Economics 49, 310–336.

Sun, C., Cheng, L., Zhu, S., Chu, Z., 2015. Multiclass stochastic user equilibrium model with elastic demand: Considering systematic and accidental delays. Transportation Research Record 2497, 1–11.

Sun, C., Cheng, L., Zhu, S., Han, F., Chu, Z., 2019. Multi-criteria user equilibrium model considering travel time, travel time reliability and distance. Transportation Research Part D: Transport and Environment 66, 3–12.

Sun, Y., Fan, W., Schonfeld, P., 2016. Static parking choice model with consideration of parking duration. Transportation Research Record 2543, 134–142.

Szeto, W.Y., and, A.B.W., 2016. Reliable network design under supply uncertainty with probabilistic guarantees. Transportmetrica A: Transport Science 12, 504–532.

Tan, H., Du, M., Chen, A., 2022. Accelerating the gradient projection algorithm for solving the non-additive traffic equilibrium problem with the barzilai-borwein step size. Computers & Operations Research , 105723.

Tidswell, J., Downward, A., Thielen, C., Raith, A., 2021. Minimising emissions in traffic assignment with non-monotonic arc costs. Transportation Research Part B: Methodological 153, 70–90.

van Lint, J., van Zuylen, H.J., Tu, H., 2008. Travel time unreliability on freeways: Why measures based on variance tell only half the story. Transportation Research Part A: Policy and Practice 42, 258–277.

Espinosa Mireles de Villafranca, A., Roncoli, C., 2024. Network traffic management via exclusive roads for altruistic vehicles under mixed traffic equilibrium. Transportation Research Interdisciplinary Perspectives 25, 101042.

Wang, J., Peeta, S., He, X., 2019. Multiclass traffic assignment model for mixed traffic flow of human-driven vehicles and connected and autonomous vehicles. Transportation Research Part B: Methodological 126, 139–168.

Wang, J.Y., Ehrgott, M., 2013. Modelling route choice behaviour in a tolled road network with a time surplus maximisation bi-objective user equilibrium model. Transportation Research Part B: Methodological 57, 342–360. doi:https://doi.org/10.1016/j.trb.2013.05.011.

Wang, Q., Feng, L., Li, J., Xie, J., Nie, Y.M., 2025. Entropy maximization in multi-class traffic assignment. Transportation Research Part B: Methodological 192, 103136.

Wang, T., Liao, P., Tang, T.Q., Huang, H.J., 2022a. Deterministic capacity drop and morning commute in traffic corridor with tandem bottlenecks: a new manifestation of capacity expansion paradox. Transportation research part E: logistics and transportation review 168, 102941.

Wang, X., Huang, H.J., 2013. Bi-criteria system optimum traffic assignment in networks with continuous value of time. Promet-Traffic&Transportation 25, 119–125.

Wang, X., Yang, H., Zhu, D., Li, C., 2012. Tradable travel credits for congestion management with heterogeneous users. Transportation Research Part E: Logistics and Transportation Review 48, 426–437.

Wang, Y., Liu, H., Fan, Y., Ding, J., Long, J., 2022b. Large-scale multimodal transportation network models and algorithms-part II: Network capacity and network design problem. Transportation Research Part E: Logistics and Transportation Review 167, 102918.

Wardrop, J.G., 1952. Road paper. some theoretical aspects of road traffic research. Proceedings of the institution of civil engineers 1, 325–362.

Wei, F., Jia, N., Ma, S., 2016. Day-to-day traffic dynamics considering social interaction: from individual route choice behavior to a network flow model. Transportation Research Part B: Methodological 94, 335–354.

Wu, W.X., Huang, H.J., 2014a. Finding anonymous tolls to realize target flow pattern in networks with continuously distributed value of time. Transportation Research Part B: Methodological 65, 31–46.

Wu, W.X., Huang, H.J., 2014b. A path-based gradient projection algorithm for the cost-based system optimum problem in networks with continuously distributed value of time. Journal of Applied Mathematics 2014.

Wu, X., Nie, Y., 2013. Solving the multiclass percentile user equilibrium traffic assignment problem: a computational study. Transportation research record 2334, 75–83.

Xie, C., Liu, Z., 2014. On the stochastic network equilibrium with heterogeneous choice inertia. Transportation Research Part B: Methodological 66, 90–109.

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.

Xie, C., Wu, X., Boyles, S., 2019. Traffic equilibrium with a continuously distributed bound on travel weights: The rise of range anxiety and mental account. Annals of Operations Research 273, 279–310.

Xie, J., Nie, Y., Liu, X., 2018. A greedy path-based algorithm for traffic assignment. Transportation Research Record 2672, 36–44.

Xie, J., Wang, Q., Nie, Y.M., 2021. An efficient algorithm for continuous bi-criteria traffic assignment. Preprint online URL: https://doi.org/10.13140/RG.2.2.24929.43364.

Xu, G., Liu, W., Yang, H., 2018. A reliability-based assignment method for railway networks with heterogeneous passengers. Transportation Research Part C: Emerging Technologies 93, 501–524.

Xu, G., Xiao, Y., Song, Y., Li, Z., Chen, A., 2023a. Capacity-constrained mean-excess equilibrium assignment method for railway networks. Transportation Research Part C: Emerging Technologies 156, 104350.

Xu, X., Chen, A., Cheng, L., 2015. Reformulating environmentally constrained traffic equilibrium via a smooth gap function. International Journal of Sustainable Transportation 9, 419–430.

Xu, X., Chen, A., Zhou, Z., Cheng, L., 2014. A multi-class mean-excess traffic equilibrium model with elastic demand. Journal of Advanced Transportation 48, 203–222.

Xu, Z., Chen, A., Li, G., Li, Z., Liu, X., 2024a. Elastic-demand bi-criteria traffic assignment under the continuously distributed value of time: A two-stage gradient projection algorithm

with graphical interpretations. Transportation Research Part E: Logistics and Transportation Review 183, 103425.

Xu, Z., Chen, A., Liu, X., 2023b. Time and toll trade-off with heterogeneous users: A continuous time surplus maximization bi-objective user equilibrium model. Transportation Research Part B: Methodological 173, 31–58.

Xu, Z., Peng, Y., Li, G., Chen, A., Liu, X., 2024b. Range-constrained traffic assignment for electric vehicles under heterogeneous range anxiety. Transportation Research Part C: Emerging Technologies 158, 104419.

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.

Xu, Z., Xie, J., Liu, X., Nie, Y.M., 2022. Hyperbush algorithm for strategy-based equilibrium traffic assignment problems. Transportation science 56, 877–903.

Yang, H., 1998. Multiple equilibrium behaviors and advanced traveler information systems with endogenous market penetration. Transportation Research Part B: Methodological 32, 205–218.

Yang, H., H. Bell, M.G., 1998. Models and algorithms for road network design: a review and some new developments. Transport Reviews 18, 257–278.

Yang, H., Huang, H.J., 2004. The multi-class, multi-criteria traffic network equilibrium and systems optimum problem. Transportation Research Part B: Methodological 38, 1–15.

Yang, H., Huang, H.J., 2005. Mathematical and economic theory of road pricing.

Yang, H., Meng, Q., 2000. Highway pricing and capacity choice in a road network under a build–operate–transfer scheme. Transportation Research Part A: Policy and Practice 34, 207–222.

Yang, H., Meng, Q., 2001. Modeling user adoption of advanced traveler information systems: dynamic evolution and stationary equilibrium. Transportation Research Part A: Policy and Practice 35, 895–912.

Yang, H., Tang, W.H., Cheung, W.M., Meng, Q., 2002. Profitability and welfare gain of private toll roads in a network with heterogeneous users. Transportation Research Part A: Policy and Practice 36, 537–554.

Yang, H., Ye, H., Li, X., Zhao, B., 2015. Speed limits, speed selection and network equilibrium. Transportation Research Part C: Emerging Technologies 51, 260–273.

Yang, H., Zhang, X., 2002. Multiclass network toll design problem with social and spatial equity constraints. Journal of Transportation Engineering 128, 420–428.

Ye, X., 2010. Accelerated procedure of multiclass highway traffic assignment for maryland statewide transportation model. Transportation Research Record 2176, 84–91.

Yin, Y., Ieda, H., 2001. Assessing performance reliability of road networks under nonrecurrent congestion. Transportation Research Record 1771, 148–155.

Yin, Y., Yang, H., 2004. Optimal tolls with a multiclass, bicriterion traffic network equilibrium. Transportation Research Record 1882, 45–52.

Ying, J.Q., 2015. Optimization for multiclass residential location models with congestible transportation networks. Transportation Science 49, 452–471.

Zang, Z., Xu, X., Qu, K., Chen, R., Chen, A., 2022. Travel time reliability in transportation networks: A review of methodological developments. Transportation Research Part C: Emerging Technologies 143, 103866.

Zangwill, W.I., 1969. Nonlinear programming: a unified approach. volume 52. Prentice-Hall Englewood Cliffs, NJ.

Zeng, M., Long, K., Han, L.D., Yang, X., 2017. Optimization of degradable road network considering VMS information and heterogeneous ATIS users. Journal of Computing in Civil Engineering 31, 04017048.

Zhang, F., Lu, J., Hu, X., 2021. Tradable credit scheme design with transaction cost and equity constraint. Transportation Research Part E: Logistics and Transportation Review 145, 102133.

Zhang, J., Yang, H., 2015. Modeling route choice inertia in network equilibrium with heterogeneous prevailing choice sets. Transportation Research Part C: Emerging Technologies 57, 42–54.

Zhang, K., Mahmassani, H.S., Lu, C.C., 2013. Dynamic pricing, heterogeneous users and perception error: Probit-based bi-criterion dynamic stochastic user equilibrium assignment. Transportation Research Part C: Emerging Technologies 27, 189–204.

Zhang, K., Nie, Y.M., 2018. Mitigating the impact of selfish routing: An optimal-ratio control scheme (orcs) inspired by autonomous driving. Transportation Research Part C: Emerging Technologies 87, 75–90.

Zhang, K., Zhang, H., Cheng, Q., Chen, X., Wang, Z., Liu, Z., 2023. A customized two-stage parallel computing algorithm for solving the combined modal split and traffic assignment problem. Computers & Operations Research , 106193.

Zhang, X., Yang, H., Huang, H.J., 2008. Multiclass multicriteria mixed equilibrium on networks and uniform link tolls for system optimum. European Journal of Operational Research 189, 146–158.

Zhao, L., Xu, X., Gao, H.O., Wang, J., Xie, Y., 2016. A bi-level model for GHG emission charge based on a continuous distribution of travelers' value of time (VOT). Transportation Research Part D: Transport and Environment 47, 371–382.

Zhu, D., Li, C., Chen, G., 2012. Existence of strongly valid tolls for multiclass network equilibrium problems. Acta Mathematica Scientia 32, 1093–1101.

Zhu, D.L., Yang, H., Li, C.M., Wang, X.L., 2015. Properties of the multiclass traffic network equilibria under a tradable credit scheme. Transportation Science 49, 519–534.

Zill, J.C., Camargo, P., Daisy, N.S., Veitch, T., 2019. Toll choice and stochastic user equilibrium: ticking all the boxes. Transportation Research Record 2673, 930–940.