Quasi-sparsity Based Origin-Destination Demand Estimation

Jingxing Wang

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2023

Reading Committee:

Xuegang (Jeff) Ban, Chair

Chaoyue Zhao

Brice Nichols

Program Authorized to Offer Degree:

Civil and Environmental Engineering

University of Washington

Abstract

Quasi-sparsity Based Origin-Destination Demand Estimation

Jingxing Wang

Chair of the Supervisory Committee:
Xuegang (Jeff) Ban
Department of Civil and Environmental Engineering

A good knowledge of the Origin-Destination (OD) demand matrix has been always important in various transportation applications, including simulation studies, transportation planning, traffic operations and control, and etc. For a large real network, the OD demand matrix may have certain quasi-sparsity property, i.e., the majority of the OD pairs have small demands while only a small portion of OD pairs have large demands. Inspired by Compressed Sensing technique, this dissertation proposes a Quasi-Sparsity Origin-Destination (QSOD) framework to explore such quasi-sparsity property of large-scale OD demand matrices. Three QSOD models (the fixed-mapping QSOD model, the bi-level QSOD model, and the distributionally robust QSOD model) are established under such QSOD framework. The results theoretically and numerically demonstrate that under certain conditions the estimated OD demands will share the same quasi-sparsity with the prior OD demands, and the estimated demands of most OD pairs (of a large-size network) will be equal to their prior values or zeros (or a very small value). Such findings provide important practical insights for OD estimation: one may only require the prior OD demands can capture the relative magnitude of the true OD demands of the network, which makes it much easier to prepare prior OD matrix in practice. The comparison between QSOD models and other existing OS estimation studies, and the integration of multi-sourced data for OD estimation under the QSOD framework are also discussed in this study.

TABLE OF CONTENTS

Page

List of Figures iii

List of Tables iv

Glossary v

Chapter 1: Introduction 3
1.1 Motivation 3
1.2 Objectives and Contributions 7
1.3 Dissertation Outline 10

Chapter 2: Literature Review 12
2.1 OD Estimation Methods 12
2.2 Compressed Sensing and Sparsity-based OD Estimation 14
2.3 OD Estimation with Uncertainty 17

Chapter 3: Methodological Framework 23
3.1 Quasi-sparse OD Demand Matrix 23
3.2 Illustration of the Quasi-sparsity Property of OD Demands 25
3.3 Overview of QSOD Modeling Framework 28

Chapter 4: Fixed-mapping Quasi-sparsity OD Estimation Model 31
4.1 Model Formulation 31
4.2 OD Quasi-sparsity Consistency of the Fixed-mapping QSOD Model 35
4.3 Solution Sparsity of the Fixed-mapping QSOD model 38

Chapter 5: Bi-level Quasi-Sparsity OD Estimation Model 42
5.1 Reformulation of the Bi-level QSOD Model 43

5.2 OD Quasi-sparsity Consistency of the Bi-level QSOD Model 45
5.3 Solution Sparsity of the Bi-level QSOD Model 46
5.4 Algorithm for the Bi-level QSOD Model 48

Chapter 6: Numerical Experiments of QSOD Models 51
6.1 A Small Network 51
6.2 Downtown Seattle Network 57
6.3 Comparisons with Existing OD Estimation Methods 61

Chapter 7: Distributionally Robust QSOD Estimation Model 72
7.1 Model Formulation 72
7.2 Solution Algorithm 78

Chapter 8: Numerical Experiments of DR-QSOD Model 84
8.1 Parameter Tuning 85
8.2 Algorithm Comparison 86
8.3 Model Performance 88

Chapter 9: Discussions and Practical Insights 94
9.1 OD Representativeness 94
9.2 Data Availability and Stochasticity 95
9.3 Multi-sourced Data Integration under QSOD Framework 95

Chapter 10: Summary and Future Work 99

Bibliography 102

Appendix A: The upper bound of r_2 113

Appendix B: Differentiability of $\Psi(d)$ 114

Appendix C: Definition of F1-score, accuracy, and RMSE 118

Appendix D: Discussion about the solution closeness of QSOD and GLS 120

LIST OF FIGURES

Figure Number Page

1.1 Demonstration of TAZ, travel demand between zones, and OD demand matrix 4

2.1 Graphic illustration of stochastic optimization and distributionally robust optimization. 22

3.1 OD quasi-sparsity of two data sources (App-based data vs. PSRC OD demands) [5] 27

3.2 QSOD modeling framework 29

4.1 Graphic illustration for OD quasi-sparsity consistency 37

6.1 A five node network 52

6.2 Downtown Seattle network 58

6.3 Convergence pattern for solving bi-level QSOD model (large network) 59

6.4 RMSE and F1-score of the GLS model vs. errors used in the weighting matrices 66

8.1 Sioux Falls network 85

8.2 Algorithm convergence comparison (relative errors) 87

8.3 Algorithm convergence comparison (running time) 88

8.4 RMSE of estimates of DR-QSOD model and deterministic QSOD model . . 90

8.5 Accuracy of OD quasi-sparsity of estimates for DR-QSOD model and deterministic QSOD model 90

D.1 Comparison across QSOD, OLS, and GLS objective function (one-dimension, $d^0 = 50$) 121

LIST OF TABLES

Table Number Page

1.1 Comparison between DR-QSOD and stochastic OD models 8

3.1 Examples of OD demands with quasi-sparsity property 26

4.1 List of notation (deterministic QSOD models) 33

6.1 Link specification 52
6.2 True OD demands 53
6.3 Comparison between OD true values, priors, and estimates (small network) . 55
6.4 Results for quasi-sparsity consistency (small network) 56
6.5 Results for QSOD solution sparsity (small network) 57
6.6 Results for OD quasi-sparsity consistency (large network example) 60
6.7 Results for QSOD solution sparsity (large network example) 61
6.8 Comparison between OD true values, priors, and estimates for three models . 69
6.9 Comparison between OD true values, priors, and estimates for QSOD and GLS (with outliers) 70
6.10 Comparison between QSOD and GLS when prior ODs are not representative (Fixed-mapping) 71

7.1 List of notation (DR-QSOD model) 73

8.1 DR-QSOD model performance, running time vs. weighting parameter λ . . . 86
8.2 Illustration of sparsity property of DR-QSOD model 91
8.3 RMSE of daily estimated OD demands t for DR-QSOD and stochastic QSOD model 93

C.1 Confusion matrix of OD quasi-sparsity consistency 118

GLOSSARY

OD: Origin-Destination

PSRC: Puget Sound Regional Council

QSOD: Quasi-sparsity based OD estimation

DRO: Distributionally Robust Optimization

DR-QSOD: Distributionally Robust Quasi-sparsity based OD estimation

GLS: Generalized Least Squared

SOD: Sparsity-based OD estimation

LP: Linear Program

UE: User Equilibrium

SUE: Stochastic User Equilibrium

MLE: Maximum Likelihood Estimation

CODE: Compressed Origin-Destination Estimation

SO: Stochastic Optimization

MPO: Metropolitan Planning Organization

TAZ: Traffic Analysis Zone

ADMM: Alternating Direction Method of Multipliers

BPR: Bureau of Public Roads

FFT: Free Flow Time

RMSE: Root Mean Squared Error

OLS: Ordinary Least Squared

BD: Benders Decomposition

CCG: Column-and-Constraint Generation

DOT: Department of Transportation

ACKNOWLEDGMENTS

First of all, I want to express my deepest gratitude to my Ph.D advisor Dr. Jeff Ban. Jeff has been always considerate, caring, and supportive during my time at University of Washington. Every time I was lost in research explorations or career development, Jeff will be one who provides the most generous guidance and suggestions to me. His passions, scrupulousness, and modesty in working will benefit and inspire me a lot in my future career journey.

I would like to thank all the supports and help from peers in intelligent Urban Transportation Systems (iUTS) lab, including Wan Li, Nazib Siddique, Rong Fan, Qiangqiang Guo, Feilong Wang, Yiran Zhang, and many others. Being in a friendly and supportive research lab for the past several years, I gained tremendous supports and priceless friendship from those lovely colleagues. Many thanks should also go to those research collaborators from other departments and other universities, including C2SMART center at New York University, University of North Carolina at Chapel Hill, Department of Industrial and Systems Engineering at UW.

Additionally, this endeavor would not have been possible without the generous support from the PhD committee, including Dr. Maryam Fazel, who provides insightful suggestions for refining deterministic QSOD models in the dissertation, Dr. Chaoyue Zhao, who collaborates with and guides me in exploring distirbutionally robust QSOD model, and Mr. Brice Nichols, who provides practical insights and suggestions for my Ph.D research. Special thanks also should be sent to Dr. Shu Lu, who put great efforts in supervising me developing and deep-diving deterministic QSOD models.

I could not have undertaken this journey without the tremendous supports and love from my family, especially my parents. Consistent understanding and supports from my parents motivate me to bravely conquer any challenges I have ever met. With their care, I always feel beloved even I am thousands of miles away from them over the years.

Finally, words cannot express my gratitude to my fiancée, Ms. Linzi Xu. I deeply thank the constant support and love from her over the past five years, for early mornings and late nights, for the ups and downs in the life. I cannot get here without her being on my side. But most of all, thank her for being my best friend.

Chapter 1

INTRODUCTION

1.1 Motivation

Origin-Destination (OD) demands of a city or a region are essential inputs to many transportation related applications, including transportation planning, traffic operations and control, traffic simulation studies, new mobility services, and so on. For a real-world transportation network, it is extremely difficult, if not impossible, to directly observe or collect the "true" OD demands, while certain datasets (e.g., travel surveys, traffic volumes, or emerging transportation "big" data) can provide a partial picture (i.e., "sample") of the OD demands of the network. Traditionally, OD demand matrices (a matrix with its element in the i-th row and the j-th column representing the travel demand from zone i to zone j, see Figure 1.1) have been directly estimated using travel survey data [76, 78, 104, 114], or estimated indirectly from traffic surveillance data, such as traffic volumes [109, 18, 120, 37, 20, 127]. The former has been widely used by transportation planning agencies to produce estimated OD matrices for a region [31, 29, 30]. The latter, i.e., OD demand estimation models, often use the estimated OD by planning agencies (as the prior OD) to further improve it by using additional traffic data. Both types of models have been extensively investigated over the past several decades, for which (indirect) OD estimation still remains an active research area in transportation and is also the focus of this thesis.

Two features of the OD demands on a real-world transportation network are crucial for OD demand estimation models. The first is that for a real urban or regional network, the number of OD pairs is usually much larger than the number of links (or nodes). For example, the King County network (the area that includes the City of Seattle in the State of Washington, USA) has about 15,000 nodes, 40,000 links, and nearly 500,000 OD pairs. When only

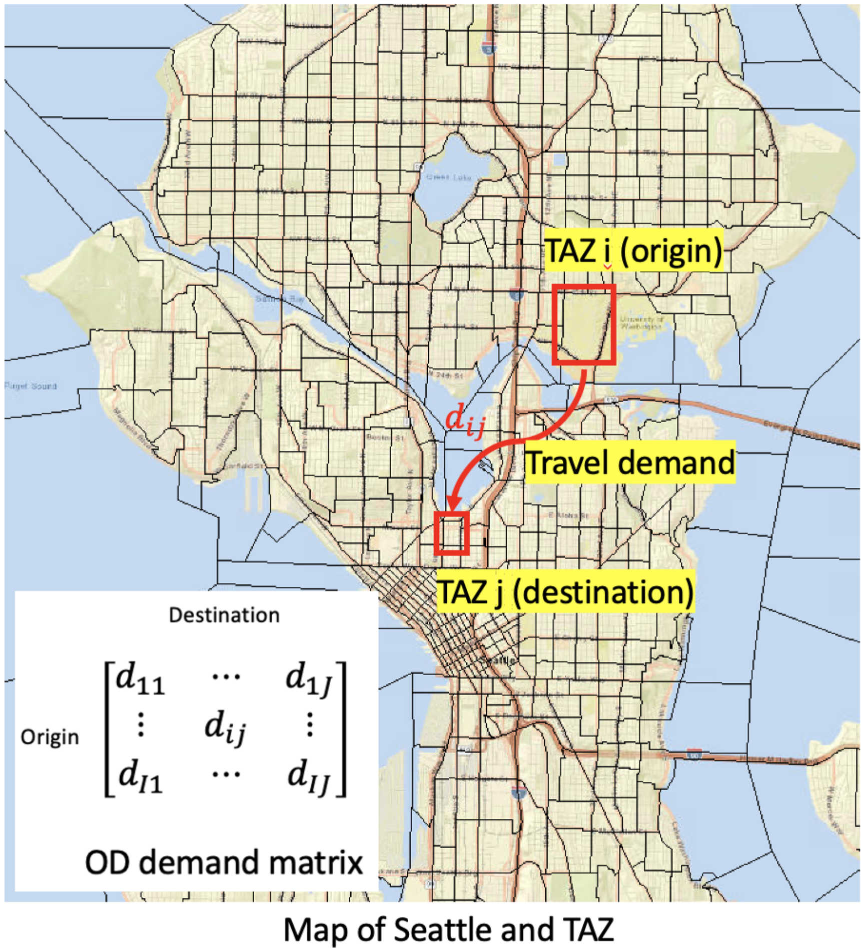


Figure 1.1: Demonstration of TAZ, travel demand between zones, and OD demand matrix

link-based information (e.g., link volume data) is used for OD demand reconstruction, the number of unknowns (the same as the number of OD pairs) will be much greater than the number of inputs (related to the available data points, i.e., the number of links), making the estimation problem "under-determined" [32, 82]. Second, given such large number of OD pairs in the network, most of them have very small demands in reality. This is intuitively understandable: the actual travels between many OD pairs in real world, especially those that are far away from or not very attractive to each other, can be quite small. This can be partially verified using estimated real-world OD demand matrices and observed transportation big data. For example, the King County OD matrix produced by the Puget Sound Regional Council (PSRC) has nearly 85% of the OD pairs with fewer than 2 trips per day while those trips contribute to about 15% of the total demands. These small yet nonzero OD demands are called insignificant OD demands in this thesis and a demand matrix with most OD pairs having insignificant demands is called a **quasi-sparse** demand matrix. Formal definition of quasi-sparsity is given in Section 3.1. Note that such insignificant travel demands, albeit small, cannot be ignored (i.e., treated as zero) since they represent real travel needs of a "minority" of the population and should be respected and considered in transportation applications.

The "under-determined" (also called "ill-posedness") issue of the OD estimation problem has long been recognized in the OD estimation literature. One way to address this issue is to incorporate a *prior* OD demand matrix to make the OD estimation problem well-posed, including those using prior OD demands of a full set of OD pairs in the network [18, 19, 120], and those only using part of the prior OD demands, e.g., studies exploring the OD observability issue [21, 22, 121]. Indeed, as such prior OD demands are readily available for many cities or regions (i.e., the estimated OD demands from planning agencies), incorporating them into OD estimation to further improve them makes a lot of sense. An implicit assumption of those OD estimation models is that the prior OD is reasonably "close" (accurate) to the true OD demand, which however was never defined rigorously and may not be easily verified in practice. Another way to address the under-determined issues is, under the assumption that the true OD demand is sparse, to explore the sparsity of an OD demand matrix (actually modeled as a vector) [97]. In fact, the sparsity-based OD estimation models have been recently investigated, which usually apply the compressed sensing technique [38] to recover the sparse OD; see detailed reviews in Chapter 2. Noteworthy is that some sparsity-based OD demand models [77] also incorporated the prior OD demands to improve estimation results.

The quasi-sparsity feature of OD demands however has not been widely recognized or explored in the OD estimation literature. This study concerns about the OD estimation problem when the OD demands present the quasi-sparsity property rather than the sparsity property, i.e., most OD pairs have small demands which are not exactly zero. In this case, sparsity-based OD estimation models cannot directly apply and techniques to explore sparsity, such as compressed sensing, cannot be used either. This study proposes to use the prior OD demand as the bridge to connect quasi-sparsity and sparsity. That is, to explore the quasi-sparsity property for OD estimation, we can study the sparsity of the deviation between the estimated OD demand and the prior OD demand. To illustrate, let d denote the estimated OD demand of a network, d^0 the prior OD demand (e.g., estimated by transportation planning agencies), and $d = d^0 + \eta$, with η the deviation between the two. As mentioned above, under the quasi-sparsity property of OD demands, it is generally not true that either d or d^0 is sparse. We can instead study the sparsity of η to help explore the quasi-sparsity of OD demands, leading to the quasi-sparsity based OD (QSOD) estimation models. Such conversion also enables us to take advantage of both the prior OD information and the powerful compressed sensing technique, leading to superb performance of QSOD as shown later in this thesis.

Another important aspect of OD estimation studies lies in its stochaticity, due to, e.g, the possible (day to day) variability of demands and/or observations of traffic. Stochastic modeling in OD estimation has been extensively studied, which has attracted more attention recently due to the increasing data availability. For example, given multi-day network-level observed link flow data, one can estimate the mean and covariance of OD demands given certain probabilistic assumptions [71, 102]. A more detailed review of existing stochastic OD estimation models is presented in the next chapter. We can see that most existing stochastic models impose certain assumptions on the distributions of observed data or OD estimates. Those assumptions sometimes can be too restrictive or difficult to measure or estimate in practice. Furthermore, if such assumptions are not satisfied, the resulting OD estimates can be degraded significantly and may not be properly used in real applications.

Inspired by the stochastic characteristics of OD estimation problems as well as changing data availability, in this thesis, we also explore QSOD models under the Distributionally

Robust Optimization (DRO) framework, denoted as distributionally robust QSOD (DR-QSOD) models. DRO is a modeling framework, where the objective is to find a decision $\mathbf{x}$ that minimizes the expected cost under the most adversarial probability distribution, i.e., $\min_{\mathbf{x}\in\mathcal{X}} \max_{\mathbb{P}\in\mathcal{D}} \mathbb{E}_{\mathbb{P}}[f(\mathbf{x},\xi)]$, where the distribution $\mathbb{P}$ of the random parameter ξ is not precisely known but assumed to belong to an ambiguity set $\mathcal{D}$. Unlike stochastic models, DRO models do not assume a given distribution for the input data or estimates, but take data variability into consideration for a set of distributions such that the estimate will be robust to distributional uncertainty of data. Table 1.1 presents the comparison between the proposed DR-QSOD model and existing stochastic OD estimation models. The key difference lies in whether we need to put strong assumptions on the data (e.g. link flows) and OD estimates. Besides, the outputs of the two categories of models may vary, as the DR-QSOD model gives the overall estimated OD demands and realized day-to-day OD demands based on daily variations, while stochastic OD models usually gives the mean (and variance) of the distrbution of OD estimates. Therefore, DR-QSOD and stochastic OD models can be regarded as two alternative modeling methods for OD estimation under stochasticity. Which modeling method to use in practice depends on the data availability and the purpose of the application, e.g., what kinds of results one may expect.

1.2 Objectives and Contributions

In order to fill the above identified gaps for large-scale OD demand estimation, this dissertation research aims to achieve the following objectives:

- formally define the quasi-sparsity property of large-scale OD demand, develop the QSOD framework, and based the previous two to explore the OD quasi-sparsity consistency as well as QSOD solution property for QSOD models.

- aware of the increasing availability and variations of observed data, leverage the distributionally robust optimization framework to consider the data stochasticity of QSOD models.

Table 1.1: Comparison between DR-QSOD and stochastic OD models

Attributes	DR-QSOD	Stochastic OD models
Assumption	OD demand matrix is quasi-sparse, and distribution parameters follow an ambiguity set	Either input data or output OD estimate, or both follow certain probability distributions [1, 51, 52, 53, 62, 85]
Input	Prior OD demands, day-to-day link flows	Prior OD demands, day-to-day link flows [51, 52, 53, 62], routing data [85]
Output	Overall OD demands and daily OD demands	Only mean [51, 52, 53, 62] or mean and variance of OD demand estimates [71, 85, 102, 122]
Solution property	Quasi-sparsity	Not keeping quasi-sparsity

- compare the QSOD models with other existing OD estimation models, summarize the findings in exploring OD quasi-sparsity property, and provide useful recommendations and practical insights for real OD estimation applications.

This first objective establishes the basics for the entire QSOD framework. The insignificant OD pair and OD quasi-sparsity are first defined, followed by the development of three QSOD models: 1) fixed-mapping QSOD model (Chapter 4), 2) bi-level QSOD model (Chapter 5), and 3) DR-QSOD model (Chapter 7). The OD quasi-sparsity consistency (between estimated OD demands and prior OD demands), and the solution property of fixed-mapping and bi-level QSOD models will be discussed in details. These discussions provide practical insights for real OD estimation: one just needs to prepare a prior OD demand matrix that shares the same quasi-sparsity with (rather than very close to) true OD demand matrix (see

Remark 1).

The second objective aims to equip the QSOD framework with the capability to consider stochastic features of data and integrate multiple data sources. The DR-QSOD model is mainly used to take into account the stochastic features and robustness given multiple-day observed data. Using multi-day link flow data as an example, this study demonstrates how to leverage distributionally robust optimization context to integrate real data in QSOD framework, considering the data variability and uncertainty at the same time.

The third objective will first compare the QSOD models with existing deterministic OD estimation models (e.g. OD estimation based on generalized least squared [120], and strict sparsity assumption [77]), in terms of OD quasi-sparsity consistency, OD estimation errors, and computational efficiency. Regarding the stochastic scenario, this study will further compare DR-QSOD model with its counterparts, deterministic model and stochastic QSOD models. More important is to summarize the main findings during the explorations for QSOD models, and offer practical insights to OD estimation stakeholders for their future applications

The main contributions of this dissertation are summarized as four folds:

- It defines the insignificant OD pairs, quasi-sparsity concept of an OD demand matrix and proposes the QSOD estimation framework, which also theoretically and numerically demonstrates that the QSOD framework can maintain the OD quasi-sparsity consistency between the prior OD matrix and the estimated OD matrix for both fixed-mapping case and the bi-level case under certain conditions. Besides, the discussions for the sparsity property of the proposed QSOD models (e.g., the solution sparsity and connection with the prior OD demands) provides deeper understanding of the QSOD models as well as the use of the L_1-norm (i.e., compressed sensing).

- By comparing with existing OD estimation models, in particular Generalized Least Squared model (GLS) and Sparsity-based OD (SOD) model, it reveals the advantages of the QSOD model in solution quality and computational efficiency. The QSOD model outperforms other models in maintaining OD quasi-sparsity consistency as well as keeping relatively low estimation errors. In addition, solving the QSOD model is much less demanding than OD estimation models using the squared-deviations, e.g., GLS and SOD. With the L_1-norm used in the objective function, the QSOD model is essentially a linear program (LP) (the fixed-mapping case) or a linear bi-level program with linear equilibrium constraints (the bi-level case), either of which is much easier to solve computationally compared with existing OD estimation models (e.g. GLS).

- In the stochastic scenario, the proposed DR-QSOD model cannot only include the stochastic features of observed data but also keep robust to the variations in data incorporated in OD estimation problem. By properly defining the uncertainty set of observed data and reformulating the DR-QSOD model, this study converts the original DR-QSOD model into a tractable form and proposes certain algorithm to solve it. As one of the few earliest studies that employed distributionally robust optimization in OD estimation problem, this study demonstrates the possibility to consider both stochasticity and robustness at the same time in estimating OD demands. The sample case (i.e. considering stochastic features in link flows) shown in the current study can be further extended to the uncertainties of other data sources (e.g. travel time data) or conditions (e.g. traffic assignment rules).

- Under the concept of OD quasi-sparsity, the QSOD model only requires that the prior OD demands are representative (i.e., capturing the magnitude of the true OD demands), not necessarily close to, the true OD demands. This helps prepare the prior OD demands when applying the compressed sensing technique to OD estimation problems in practice.

1.3 Dissertation Outline

We organize the remaining of this dissertation as follow. In Chapter 2, we conduct a comprehensive literature review regarding OD estimation and its current progress, including traditional OD estimation methods (Section 2.1), OD estimation concerning sparsity (Section 2.2), stochastic and distributionally robust optimization in OD estimation (Section 2.3). Then in Chapter 3, we give a formal definition for significant OD pairs, quasi-sparsity property of OD demand matrices, and also present the general formulation for the QSOD modeling framework. We then discuss the fixed-mapping QSOD models in detail, including OD quasi-sparsity consistency and solution sparsity property in Chapter 4. Similarly, in Chapter 5, we deep dive the bi-level QSOD model and its property, based on the knowledge of the fixed-mapping QSOD model. Extensive numerical examples are demonstrated in Chapter 6 to show the performance of fixed-mapping and bi-level QSOD model and their comparison with traditional OD estimation methods, including GLS methods and Sparsity-based OD estimation method. In Chapter 7, the formulation of DR-QSOD model is discussed based on the concept of distributionally robust optimization and quasi-sparsity. In addition, the reformulation and solution algorithm for the complex DR-QSOD model is illustrated as well. In Chapter 8, we present the numerical tests for the DR-QSOD model, including its comparison with its stochastic and deterministic counter part. In Chapter 9, we discuss the practical insights and considerations for real-world implementation. Finally, Chapter 10 concludes the dissertation and point out the future work.

Chapter 2
LITERATURE REVIEW

OD estimation problem has been extensively studied for past decades due to its importance in academic research and engineering applications in real-world. In this chapter, we first present a review for traditional OD estimation studies in Section 2.1. Then based on the observation of OD quasi-sparsity, we look into an important technique to address under-determined problem, compressed sensing, and its applications in sparsity-based OD estimation studies in Section 2.2. In Section 2.3, we revisit stochastic OD estimation studies, introduce the concept of distributionally robust optimization and discuss various ways to deal with data/variable uncertainty in OD estimation literature.

2.1 OD Estimation Methods

OD estimation, i.e. using newly acquired field data (e.g. link traffic flows) to update (or adjust) an old/sampled OD demand matrix, has attracted much attention over the past decades. Traditionally, OD matrices were directly obtained from large-scale sampled surveys (e.g. roadside surveys, household travel surveys) conducted every ten to twenty years. Such approach is often costly in time and budgets. In addition, its output, i.e the OD demand matrix for a region or city, usually becomes outdated by the time when the data was collected and processed [12], hence may not properly reflect the most recent patterns of travel demands. In the era of Information Technologies, with more surveillance devices deployed into the network, using the newest acquired field data, e.g. link traffic counts, for OD demand estimation becomes possible. The existing OD estimation models can be categorized into two classes: static and dynamic OD estimation, based on whether they are time independent or not. The static models assume the traffic flows are time-independent and an average OD demand is determined for long-term planning purposes. The dynamic models, on the other hand, assume flows may vary with time, whose results are primarily used for short-term strategies, including route guidance, traffic control, and etc [12]. Although the dynamic OD estimation problems are attracting more attention recently, the static OD estimation problem still merit further exploration due to its importance in theoretical and practical applications. In this section, the majority of past studies reviewed here will be static OD estimation methods, which is also aligned with proposed QSOD models.

As one of the earliest studies in using link counts for OD estimation, Van Zuylen and Willumsen [109] assumed the averaged link counts follow a Poisson distribution, and then used the Entropy Maximization (EM) principle to find a most likely OD matrix that has the greatest number of micro-states $W\{d_w\}$ as well as satisfies the link flow constraints. The number of micro-states is defined in (2.1), where TN denotes the total demands of the OD matrix, d_w denoted the demand of OD pair w where w belongs to OD pair set W. However, such EM-based models have the disadvantage of not considering the uncertainties in traffic counts and prior matrix which can be erroneous when traffic counts are not accurate [12].

$$W\{d_w\} = \frac{TN!}{\prod_w d_w!} \tag{2.1}$$

Later on, a group of quadratic optimization models for OD estimation were proposed, which share the similar formulation as (2.2)

$$\begin{aligned} \min_{d\in\mathbb{R}^{|W|}, v\in\mathbb{R}^{|A|}} \quad & f(d,v) = (d-d^0)^T U (d-d^0) + (v-v^0)^T V (v-v^0) \\ \text{s.t. } & (d,v) \in F = \left\{(d,v) \;\middle|\; \Psi(d) = v,\ d \geq 0, v \geq 0\right\} \end{aligned} \tag{2.2}$$

where U and V can be interpreted as the weights for OD demand errors and link flow errors, respectively; d and v (d^0 and v^0) are the estimates (prior values) for OD demands and link flows, respectively; $\Psi(\cdot)$ represents the traffic assignment rule that assigns OD demands d to link flows v. These studies include but do not limit to, MVN Bayesian inference by [72][51], Aitken generalized least squared (GLS) model by [18] GLS model under user equilibrium (UE) [120][119], GLS model under stochastic user equilibrium (SUE), etc.

Also by assuming the elements of observed sampled OD matrix follows independent Poisson distribution with unknown means, [105] proposed a Maximum Likelihood method to estimate these means, yielding an estimation of the "true" origin-destination matrix which is consistent with the observed link volumes. The maximum likelihood (ML) function can be further re-written as (2.3), where d_w and d_w^0 denote the estimated and sampled (or prior) demand for OD pair w; ρ_w is the sampling factor for OD pair w. In this study, a special case where the total demands generated from (and attracted by) each OD pair are assumed known, was also investigated. This further extends the usage of such ML-based models.

$$\min f(d, d^0) = \min_{d} \sum_{w \in W} (\rho_w d_w - d_w^0 \log d_w) \tag{2.3}$$

Recently, [23] proposed a hierarchical optimization framework to estimate travel demands using Gamma random variables in a Bayesian context. Three sub-problems are considered at the same time: 1) a Wardrop minimum variance (WMV) assignment model, which is used to derive the route choice probabilities, (2) a least squares problem, used to obtain the OD sample data, and (3) a maximum likelihood problem to estimate the posterior modes. This study did not only use Bayesian methods in estimating OD demand but also include the UE constraints under the congested conditions. With a new Wardrop-minimum variance optimization problem embedded in the entire framework, this method allows to obtain unique link flows disaggregated by OD pairs when conducting OD estimation. Besides, its use of Gamma distribution and its corresponding conjugate priors also helps the establishment of Bayesian models.

2.2 Compressed Sensing and Sparsity-based OD Estimation

There are a handful of recent studies that explored OD sparsity for OD estimation [97, 77, 115], which we refer to as sparsity-based OD estimation models in this dissertation. All of these models applied compressed sensing that is designed to find sparse solutions of an under-determined linear system where unknowns are much more than the inputs [38, 16]. Compressed sensing often comes with solving the under-determined system $A \cdot x = b$, where

$A \in \mathbb{R}^{m\times n}$, $x \in \mathbb{R}^n$, $b \in \mathbb{R}^m$, and $m \leq n$. If x is the coefficient of the signal or image in a known basis which happens to sparsely represent the signal or image, it can be recovered by solving the L_0 optimization problem $\min_x \|x\|_0$, s.t. $A \cdot x = b$ [39]. The solution of the problem can be well approximated by (and sometimes equivalent to) the solution of the L_1 optimization problem $\min_x \|x\|_1$, s.t. $A \cdot x = b$, e.g., by the method of Basic Pursuit [27]. The use of L_1-minimization in solving such under-determined equation guarantees that the solution x^* is sparse enough, which also leads to various applications in many other fields. However, its applications in transportation area started to attract attention only recently.

The early studies employing compressed sensing focused on inferring the OD flow anomalies [24] or reconstructing the anomalous components in OD demands [75]. Later on, compressed sensing was applied into the path flow or OD flow estimation [97][77][115]. [97] proposed the Compressed Origin-Destination Estimation (CODE) model to estimate OD demands and their path allocations, based on link flow data. Their work relied on the fact that for a given OD pair in a urban network, there are many alternative paths while only a small fraction of paths may carry traffic flow due to the preferences of travelers on travel times, trip lengths, number of turns, etc. CODE assumed the path set is sparse and recovered the path flow by applying compressed sensing and using the L_1-norm minimization, under three scenarios of variants (noise-less, noisy, and weighted). As one of the first few studies directly employing compressed sensing idea in OD estimation problem, CODE showed great capability in dealing with the under-determined issue and reconstructing sparse path flows. However, limitations do exist for this model. First, the OD recovery process that only relied on the link flows did not make full use of prior information (e.g., prior OD demands) such that the corresponding estimation results might not be accurate if the observed link flows are biased or with large noise. Second, CODE only showed the sparsity property under the fixed-mapping traffic assignment, which may not directly apply in practice. The fixed-mapping can be replaced by the UE traffic assignment [113], to better capture the relation between OD demands and link flows, which however will lead to a more complex bi-level formulation; see Chapter 5 for more detailed discussions.

Following CODE, [77] explored fine-grained OD estimation with automated zoning and sparsity regularisation. They first proposed an algorithm to automatically design traffic analysis zones, based on which to establish the OD estimation problem. In order to address the "ill-posedness" issue of large scale OD estimation problems, they assumed the sparsity of the OD demand matrix then added an L_1-regularizer (with hyper-parameter λ) into the GLS model objective [19, 120]. The resulting estimator hence owned different levels of sparsity per changing of hyper-parameter λ. However, the imposed assumption on the sparsity of the OD demand matrix can be too restrictive in practice: many OD pairs may have insignificant (i.e. small) but nonzero OD demands. Recently, also inspired by Compressed Sensing, [115] proposed a sparse OD matrix estimation method to address the under-determined issue of OD estimation by accounting for the demand volatility and link flow correlations. An objective function similar to [77] was used in the study. They considered the link flow correlations by incorporating non-diagonal weighting matrix for link flow deviation in the objective. A novel strategic UE model (I-STRUE) was also used to address the traffic demand assignment problem. The L_1-regularizer of the estimated OD demand was introduced to maintain the OD matrix sparsity while how to determine the weight for L_1-regularizer is still an issue, similar to [77]. In addition to the applications in OD estimation on road networks, compressed sensing was also used to estimate the sparse OD demands on transit routes recently. [60] estimated the transit OD demands using compressed sensing by incorporating automatic passenger count (APC) data on a single transit trip. Similarly, an L_1-regularizer of transit OD demands was considered in the model with subject to the APC constraints, which forced the solution to be sparse. However, there is still a weighting term for L_1-regularizer to be further determined.

To summarize, existing studies that directly used compressed sensing assumed demands from most OD pairs or path flows are strictly zero, which however may not be the case in practice (except the one for estimating transit OD demands in [60]). In addition, although all of these works applied the compressed sensing method in guaranteeing a sparse solution, few of them ever investigated theoretically how sparse the solution was and how the characteristics of prior information (e.g. prior OD demands and observed link flows) impacted the solution property under the OD estimation scenario. This dissertation aims to fill these gaps by exploring the quasi-sparsity feature of OD demands for OD estimation.

2.3 OD Estimation with Uncertainty

In the era of information technology, the increasing data availability provides great opportunities to account for the OD estimation with uncertainty, either in the estimate itself, or in the input data, or both. In this section, we start with reviewing an important category of studies considering the uncertainty - stochastic OD estimation. Then we introduce the literature on distributionally robust optimization (DRO) approach that we used in this study, and give a detailed comparison between stochastic optimization (SO) and DRO.

2.3.1 Stochastic OD estimation

The early studies of stochastic OD estimation generally assumed OD demands follow mutually independent Poisson distribution and used day-to-day link flow data as the input to estimate OD demands. For example, [51] proposed a Maximum Likelihood Estimation (MLE) method and a Bayesian inference method [52] for OD estimation. Following the same assumption, [53] continued to employ a Poisson model to conduct OD estimation, where the first- and second-order statistical property of link flows were used. [62] also applied Bayesian inference for a transportation network to estimate the population means of traffic flows, reconstruct traffic flows, and predict future traffic flows, through the Expectation-Maximization (EM) algorithm. [85] proposed an MLE method to estimate OD demands using link counts and sporadic routing data. In addition to assuming independent Poisson distributions for OD demands, the study further assumed a normal distribution for the link counts and a binomial probability for the monitored vehicle routing counts. [1] proposed a maximum probability model to estimate the OD demand matrix in the network, where the observed traffic counts of links and the target OD demands are independent discrete random variables with known probabilities. [123] proposed a model to estimate the mean value of

OD demands as well as improve network identifiability using multi-day observation sets.

It is well recognized that the OD demands within the same period of a day fluctuate from day to day, due to daily variations in activity patterns [28]. The daily fluctuations of OD demand also lead to significant impacts to traffic system performance, urban traffic management, route choice of travelers, and etc [71, 102]. By considering the variations of OD demands, many recent studies started to consider transportation network reliability and uncertainty issues [102]. For instance, by assuming OD demands follow mutually independent normal distributions, [25] evaluated travelers' risk-taking behaviors under travel demand uncertainties. [28] developed a stochastic modeling framework to estimate network travel time reliability with mutually independent Poisson distributions of OD demands. [101] proposed a reliability-based user equilibrium model under the assumptions that OD demands and path flows are mutually independently normal distributions. [81] explored an internally-consistent network equilibrium model, which considered two potential sources of flow variability: daily variation in route choice and daily variation in OD demands. In the study, OD demands were assumed to follow multiple distributions, including Binomial, Poisson, Beta-binomial, and negative binomial distribution. To summarize, the above-mentioned studies demonstrate the need to capture the variability of OD demands rather than a single point estimator, which lets the stochastic OD estimation draw an increasing amount of attention.

It is noted that the majority of stochastic OD estimation studies discussed above only concerned about the means of OD demands. Because of the increasing data quality and quantity as well as the needs in other applications, e.g. stochastic network equilibrium, how to better capture the probability distribution, i.e. estimating the means and covariance of OD demands, start to attract much more attention recently. For instance, by assuming multivariate normal (MVN) distributions for OD demands, [102] used a weighted least squared method to estimate the means and covariance of travel demands. [71] proposed a theoretical framework for estimating the means and variance-covariance matrix of OD demand by considering the day-to-day variation induced by travelers' independent route choices. The probability distributions of link/path flow and their travel cost were also estimated where the variance comes from three sources: O-D demand, route choice, and unknown errors. [122] proposed a Generalized Method of Moment (GMM)-based framework to infer the probability density function (pdf) of OD demand variables. Instead of just providing point estimates, their study revealed large sample statistical properties of the proposed estimator, which served as a theoretical foundation for assessing estimation quality, constructing confidence region, and testing model adequacy. The studies reviewed above give us a glimpse of stochastic OD estimation methods, which however are far from exclusive or comprehensive. For a more comprehensive review of such studies, one can refer to other recent studies [102, 71, 122].

To summarize, stochastic OD estimation methods consider stochasticity in data and/or modeling components of the OD estimation problem. In order to mathematically formulate and solve those models, it is common to impose certain statistical assumptions about the distributions on either the estimates, input data, or both. However, there might be issues when implementing those methods in practice. For example, the distributions of estimates or data may be hard to derive or estimate. Data are often used to estimate the distributions. But if the data is incomplete or limited, it is hard or impossible to rely on those data to construct such distributions. If the assumptions on distributions are not satisfied in real applications, the resulting OD estimates from such models may not be accurate or reliable. What's more, most stochastic OD estimation models can only produce an estimate representing the overall demands. Estimating the daily OD demands corresponding to the observed daily data seems difficult for many of those models.

2.3.2 Distributionally robust optimization

DRO is a framework that addresses scenarios involving parameter uncertainty, where the probability distribution for the parameter is unknown. The concept of DRO was first introduced in the 1950s [98] when it was employed to determine the order quantity in a newsvendor problem, with the goal of maximizing the worst-case expected profit. While DRO gained some initial attention during that time, it truly gained prominence in the 2010s when researchers developed various methods for constructing ambiguity sets for uncertain data and derived tractable formulations for DRO [34, 48, 116]. Since then, DRO has seen a surge in research and applications, with growing interest and exploration of its potential across multiple fields.

Before delving further into the unique properties of DRO, it is beneficial to revisit the general formulation of stochastic optimization (SO). Problem (2.4) presents the formulation for SO, where $\mathbf{x}$ represents the decision variable, $\mathcal{X}$ denotes the feasible region for the decision variable, and ξ is a random variable defined on a probability space $(\Omega, \sigma(\Omega), \mathbb{P})$. Here, Ω is the sample space of ξ, $\sigma(\Omega)$ denotes the σ-algebra of the sample space, and $\mathbb{P}$ represents the probability distribution of ξ. The objective of the SO problem is to minimize the expected value of the function $f(\mathbf{x}, \xi)$ over the feasible set $\mathcal{X}$. A common prerequisite for SO is that the probability distribution $\mathbb{P}$ is known or easily estimable from available data. The formulation of single-stage DRO, as depicted in (2.5), exhibits a certain level of similarity with SO. It aims to optimize the worst-case function $sup_{\mathbb{P}\in\mathcal{D}} E_{\mathbb{P}} f(\mathbf{x}, \xi)$ over $\mathbf{x} \in \mathcal{X}$. In this context, the term "worst" indicates the supremum function, representing the most unfavorable outcome among all possible probability distributions in the ambiguity set $\mathcal{D}$.

$$\begin{aligned} SO: &\inf_{\mathbf{x}} E_{\mathbb{P}} f(\mathbf{x}, \xi) \\ \text{s.t. } &\mathbf{x} \in \mathcal{X} \end{aligned} \tag{2.4}$$

$$\begin{aligned} DRO: &\inf_{\mathbf{x}} \sup_{\mathbb{P}\in\mathcal{D}} E_{\mathbb{P}} f(\mathbf{x}, \xi) \\ \text{s.t. } &\mathbf{x} \in \mathcal{X} \end{aligned} \tag{2.5}$$

The key difference between SO and DRO lies in how much probability information of variable ξ we have and the way we treat its uncertainties. Different from SO that predefines the probability distribution $\mathbb{P}$, DRO (2.5) assumes the probability distribution $\mathbb{P}$ falls into the set $\mathcal{D}$, where $\mathcal{D}$ is a predefined ambiguity set that is assumed to include the true probability distribution of ξ. Even both models need to take the expectation over ξ in the objective, DRO relaxes the assumption from a single fixed probability distribution $\mathbb{P}$ to a family of probability distributions $\mathbb{P} \in \mathcal{D}$, which has very important implication in practice. This is because the exact distribution $\mathbb{P}$ in SO is often hard to obtain or estimate with empirical data [100], especially when data is not complete or sufficient. In these cases, however, the ambiguity set $\mathcal{D}$ may be estimated more easily. Various types of ambiguity sets $\mathcal{D}$ have been extensively studied in previous DRO research. One commonly used type of ambiguity set is the moment-based confidence set [35, 128, 49]. These sets can be easily constructed using first or second moment information of the data. Another class of ambiguity sets is based on distance measures, such as ϕ divergences [55, 11] and the Wasserstein metric [43, 125], which centers around a finite-support nominal distribution constructed from data samples. In this dissertation, we use the moment-based confidence set to build the ambiguity set.

In essence, SO and DRO are two alternative modeling approaches for addressing stochasticity or uncertainty inherent in data or models. Each approach has its strengths and limitations, making it more suitable for certain situations while less effective for others. In this dissertation, we further provide a comprehensive comparison of these two approaches in the numerical section.

The DRO framework has found numerous applications in various areas such as power system reliability [3, 126], supply chain optimization [65, 117], and chemical engineering [100], among others. For a more recent survey on this topic, readers may refer to [89]. However, as far as we know, it has not yet been employed in OD estimation problems, while the flexibility of the DRO approach in handling uncertainty makes it a promising alternative to SO-based methods for OD estimation. In practice, obtaining exact and accurate distributions of uncertain parameters, such as traffic link flows in OD estimation, can be extremely difficult or even impossible. Instead, we may only have prior or partial information, which allows us to establish an appropriate ambiguity set relatively easily. The distinction between how DRO and SO handle uncertainty in OD estimation is illustrated in Figure 2.1. As depicted in the figure, SO requires a predetermined distribution of uncertain parameters in OD estimation, while DRO provides more flexibility by allowing the distribution of uncertain parameters to vary ambiguously within a predefined set. In Chapter 7, we will demonstrate how to construct a two-stage DR-QSOD model that considers the uncertainty in link flow data and explain how to efficiently solve it.

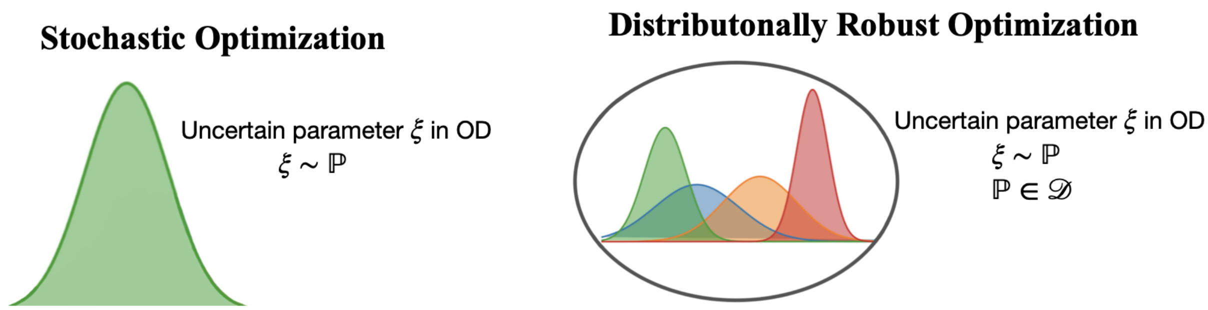


Figure 2.1: Graphic illustration of stochastic optimization and distributionally robust optimization.

Chapter 3

METHODOLOGICAL FRAMEWORK

A transportation network consists of a set of nodes N, a set of directed links A, and a set of OD pairs $W \subset N \times N$. $|A|$ and $|W|$ denote the cardinalities of set A and W, respectively. We use $\mathbb{R}^n$, $\mathbb{R}^n_+$, $\mathbb{R}^n_{++}$to denote the real n-dimensional space, its nonnegative orthant, and its strictly positive orthant, respectively.

In the following, we first present the mathematical definitions of the quasi-sparsity property, and then provide some observations/data to illustrate this property of OD demands on large-size transportation networks. We conclude this chapter by presenting the overall QSOD modeling framework of the dissertation.

3.1 Quasi-sparse OD Demand Matrix

Denote ϵ_0, α, and β as three non-negative, user-specified thresholds. Two quasi-sparsity related definitions about an OD demand matrix, denoted as d, are given as follows. Note that although we follow the convention to call d an OD matrix, it is actually modeled as a vector hereafter in this study.

Definition 1 (An insignificant OD pair) An insignificant OD pair is an OD pair with small (i.e., insignificant) travel demand. Denote w_1 an insignificant OD pair if $0 \leq d_{w_1} \leq \epsilon_0$ (e.g., $\epsilon_0 = 5$), where d_{w_1} is the demand of OD pair w_1.

Definition 2 (A quasi-sparse OD demand matrix) Denote W_1 the set of all insignificant OD pairs of W, i.e., $W_1 = \{w \in W | 0 \leq d_w \leq \epsilon_0\}$. Then an OD demand matrix is quasi-sparse if both of the following conditions hold:

- *Condition 1 (Dominance in terms of the total number of insignificant OD pairs)*: $|W_1| \geq \alpha |W|$, where $\alpha \in (0.5, 1)$ is a user-specified large ratio, e.g. 85%, representing the *quasi-sparsity level* of the network. It indicates that the insignificant OD pairs in the OD matrix take an overwhelming portion (in terms of the number) over all the OD pairs.

- *Condition 2 (Insignificance in terms of the total travel demands of insignificant OD pairs)*: The total OD demands of insignificant OD pairs account for a relatively small portion of the total OD demands in the network. This can be expressed as: $\sum_{w_1 \in W_1} d_{w_1} \leq \beta \sum_{w \in W} d_w$, where β is a user-specified small threshold (e.g., $\beta = 0 \sim 15\%$).

The quasi-sparsity property defined here is similar to sparsity (where most OD pairs are assumed to have exactly zero demands). Concept-wise, quasi-sparsity is more general than sparsity. This can be easily seen: a sparse matrix is a special case of a quasi-sparse matrix with $\epsilon_0 = 0, \beta = 0$, and the quasi-sparsity level in Condition 1 reduces to the sparsity level of the (now sparse) matrix. Further, the QSOD models presented later in this dissertation apply the same compressed sensing technique developed specifically for exploring sparsity, by using the prior OD as the bridge. Therefore, the quasi-sparsity property here can be considered as an alternative way to define and model "sparsity" when OD demand estimation is concerned.

For a given transportation network, this research concerns about three OD demand matrices: the ground-truth OD matrix ($\bar{d}$), the prior OD matrix (d^0), and the estimated OD matrix ($\hat{d}$ or d^*) by the proposed models. The above definitions can be applied to any of the three types of OD matrices, which also provide measures to quantify an OD matrix in terms of their quasi-sparsity. The QSOD models presented in Chapter 4 and Chapter 5 can help answer the *quasi-sparsity consistency* question, i.e., under what conditions the same quasi-sparsity property will hold for the prior OD matrix and the estimated OD matrix under the same thresholds $(\epsilon_0, \alpha, \beta)$ and same W_1. To help draw insights regarding the true OD matrix, we make the following definition to connect the true and prior OD demand matrices.

Definition 3 (Representative prior OD demand matrix) Assume the quasi-sparsity property holds for the true OD demand matrix, defined by threshold parameters $(\epsilon_0, \alpha, \beta)$ and the set of insignificant OD pairs W_1. The prior OD demand matrix is *representative* of the true OD matrix (in terms capturing the magnitude of OD demands) if the above quasi-sparsity property also holds for the prior OD demand matrix on the same thresholds $(\epsilon_0, \alpha, \beta)$ and the same insignificant OD pair set W_1.

Note that the "ground truth" OD demand is normally not available, especially for real-world large scale networks. The above definition provides qualifications of the prior OD when used in the QSOD models, which also helps draw connections between the estimated OD and the true OD. We thus assume the prior OD is representative in this dissertation unless noted otherwise. However, even the prior OD is not representative, we show that the fixed-mapping and bi-level QSOD models can still produce similar or better results compared with existing OD estimation methods; see Table 6.10 in Section 6.3.2 and the discussion therein. Also note that similar assumptions, typically in the form of "the prior OD is reasonably close to the ground-truth OD", are often imposed explicitly or implicitly in the literature when prior OD is used for OD estimation. However, definitions or qualifications like Definition 3 above for the prior OD have rarely been proposed in the past. In this sense, not only does Definition 3 actually relax previous assumptions imposed on the prior OD, it also provides more clear (also more practical) guidance regarding how to prepare the prior OD. See more discussions on this in Section 6.3.2.

3.2 Illustration of the Quasi-sparsity Property of OD Demands

Intuitively, the OD demand matrix of a large-size transportation network is generally quasi-sparse: the demands between most OD pairs (those that are far away or not attractive to each other) are small. Nevertheless, the true OD demand matrices of real-world large-size networks are never directly available, making it extremely difficult, if not possible, to directly verify the above statement. However, we may obtain some reasonable (and partial) validation of this via two ways: the estimated OD demand matrices of real-world networks by planning agencies and some observations from transportation big data that reflect the OD demands of a large network. Table 3.1 shows some quasi-sparsity related statistics of several estimated

OD demand matrices by local Metropolitan Planning Organizations (MPOs) [106, 31] and generated using app-based transportation big data [5]. Using $\epsilon_0 = 2.0$ as the threshold ($\epsilon_0 = 1.0$ for the Chicago regional network as it has relatively small demands), the four matrices present quasi-sparsity (Definition 1 and Definition 2): the majority of OD pairs are insignificant OD pairs (e.g., $\alpha = 0.9$) and they only contribute a small portion of total OD demands (e.g., $\beta = 0.15$).

Table 3.1: Examples of OD demands with quasi-sparsity property

Network	No. of zones	% of insig-OD pairs	% of insig-OD demands
Philadelphia	1525	96.2%	11.4%
Chicago-regional	1790	93.3% ($\epsilon_0 = 1$)	15.27%
Puget Sound Region (PSRC)	3700	92.64%	14.29%
Puget Sound Region (App-based data)	3700	95.21%	7.29%

The last two rows in Table 3.1 are the estimated demand matrices for the Puget Sound region. We can further check whether the insignificant (or significant) OD pairs of the two demand matrices are the same. For this, we can compute the *F1-score* (see its definition in Appendix C) in terms of the insignificant OD pairs (i.e., W_1) between the two demand matrices. Results show that the $F1$-score is 0.966, indicating that not only do the demand matrices from the two data sources present quasi-sparsity, this property is defined almost on an identical set of insignificant OD pairs. This very high $F1$-score is particularly telling: the two demand matrices were produced by different methods on completely different data sources (survey data vs. app-based data); the absolute values and even scales (e.g., maximum values of the demands) of the two matrices are drastically different (their correlation coefficient is only about 0.4). However, both present the quasi-sparsity property on almost the same set of parameters and W_1. This finding thus leads us to reasonably suspect that the quasi-sparsity property likely exists in real-world large-scale transportation networks,

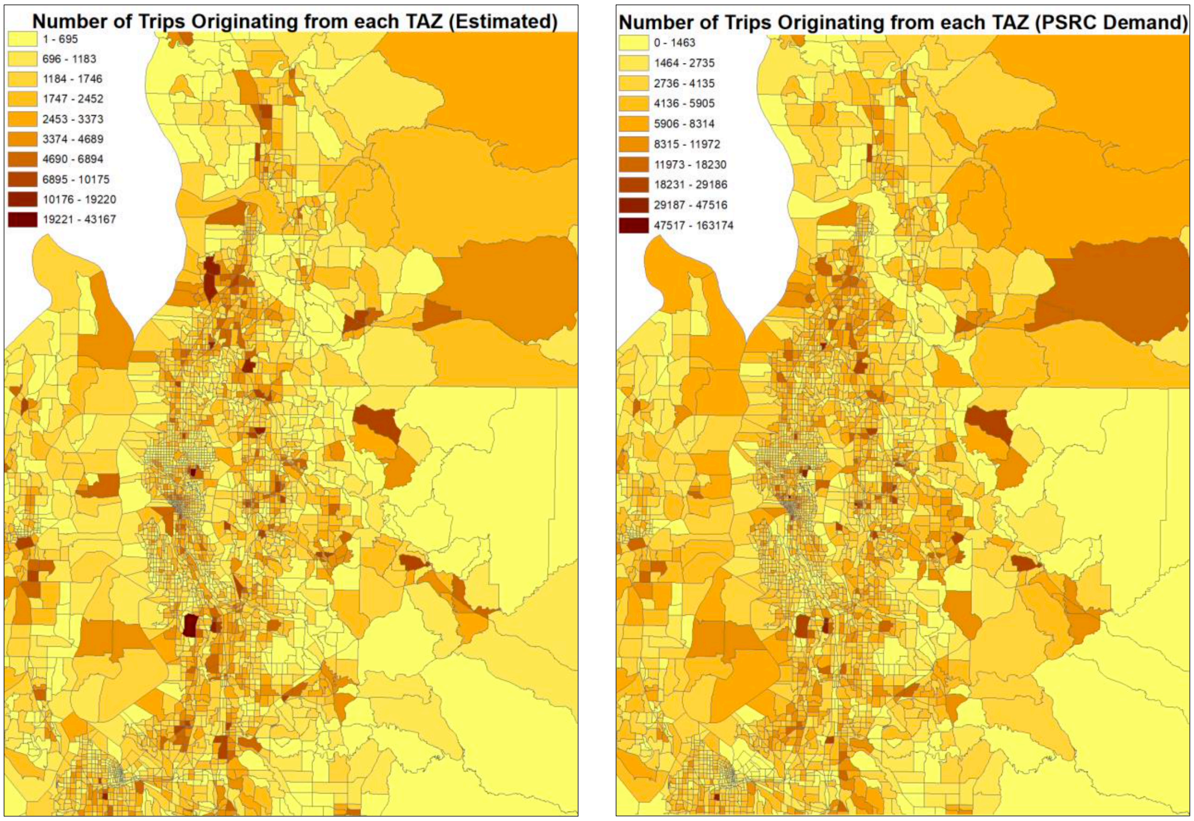


Figure 3.1: OD quasi-sparsity of two data sources (App-based data vs. PSRC OD demands) [5]

although we could only show this for the Puget Sound region and future research should certainly investigate the finding for more regions. Therefore, it is both scientifically interesting and practically important to formally define and study the concept of quasi-sparsity and explore OD estimation problems with this property.

In addition, it is worthwhile to point out investigating "OD quasi-sparsity" rather than "OD sparsity" is more practically important and reasonable in OD estimation studies. First, according to the four examples in Table 3.1, we observe the real large-scale OD demand matrices do show quasi-sparsity rather than pure sparsity, i.e. the majority of those insignificant OD pairs own small but non-zero demands. This further indicates that OD quasi-sparsity widely exists in real world scenarios. Second, considering those small but non-zero demands might be generated by the minority groups of population (in terms of annual income, race, education level, home locations), it is even more valuable to retain and pay attention to those insignificant demands in pursuit of transportation equity. What's more, not only do the overall OD demands of all transportation modes demonstrate quasi-sparsity property, it is noticed that travel demands in public transportation mode, e.g. subway, public transit, also show such quasi-sparsity property.

We conclude this section by visualizing the quasi-sparsity of the two demand matrices in Figure 3.1. It can be seen that, although the absolute values of the two demands are very different, the demands produced using app-based data (left sub-figure) shares a similar pattern with the PSRC-estimated OD demands (right sub-figure): the majority of traffic analysis zones (TAZ) generates very small number of trips per day while only a small portion of TAZs have significant OD demands.

3.3 Overview of QSOD Modeling Framework

Having formally defined the quasi-sparsity property of OD demand matrices and illustrated the quasi-sparsity observation from the real-world data, we now discuss the QSOD modeling framework in this section. Figure 3.2 below presents a diagram of the QSOD framework. The QSOD modeling framework is inspired by wildly-existing quasi-sparsity property of large-scale OD matrices, and the increasing availability of transportation data. This framework consists of three quasi-sparsity OD models: the fixed-mapping QSOD model, the bi-level QSOD model, and the distirbutionally robust QSOD model, which correspond to the discussions in Chapter 4, Chapter 5, and Chapter 7, respectively. The three QSOD models share strong connections because of assumptions and compressed sensing technique, but also inherent differences due to underlying structures and application scenarios.

First, based on the assumptions of prior OD representativeness and fixed-mapping traffic assignment, the fixed-mapping QSOD model lays out the foundation for the QSOD modeling framework. Many extended explorations on the other two models are relying on the

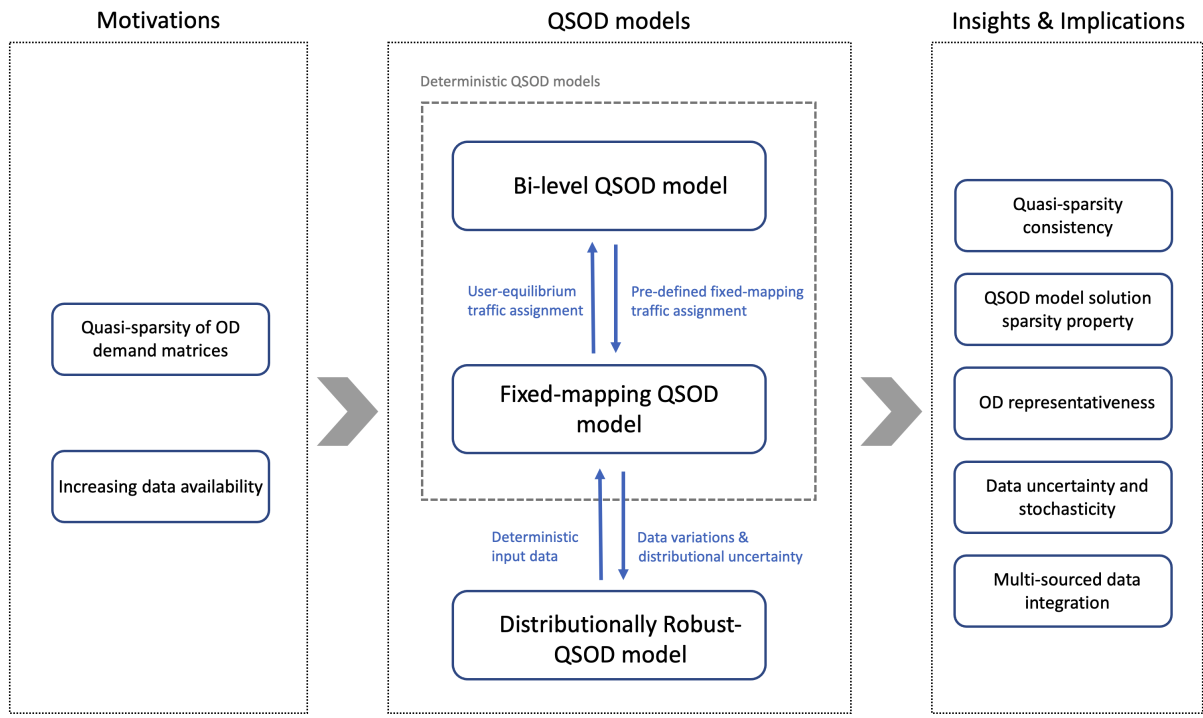


Figure 3.2: QSOD modeling framework

discussions of the fixed-mapping QSOD model. Second, if one replace the pre-defined fixed-mapping traffic assignment with the widely-adopted user-equilibrium traffic assignment, the model becomes the bi-level QSOD model. With a more complex traffic assignment rule, the bi-level QSOD model can mimic a more real OD estimation modeling scenario, therefore could be more suitable for real use case. Both the fixed-mapping QSOD model and bi-level QSOD model belong to the deterministic model category. Third, if one needs to consider data variations and its distributional uncertainty, the QSOD model framework yields to the DR-QSOD model, where the uncertain distributions of input data are dealt with distritbuionally robust optimization. With DRO incorporated, the QSOD framework gets enriched in considering data uncertainty and stochasticity at the same time. Again, the three models are closely connected with each other. The fixed-mapping can be regarded as a special case of the bi-level QSOD model where the traffic assignment is fixed; see Chapter 4 for more information. The fixed-mapping QSOD model can be derived from the DR-QSOD model if one sets a proper weighting parameter λ for the DR-QSOD model; see Chapter 7 for more discussions on this. Based on this framework, this dissertation will investigate some properties of the QSOD models, and explore the practical insights behind these properties, including quasi-sparsity consistency, QSOD model solution sparsity property, OD representativeness, data uncertainty and stochasticity, multi-sourced data integration, and so on.

Chapter 4

FIXED-MAPPING QUASI-SPARSITY OD ESTIMATION MODEL

4.1 Model Formulation

Many existing OD estimation models share the similar generic form as follows [72, 18, 120, 51, 121]:

$$\begin{aligned} \min_{d\in\mathbb{R}^{|W|},v\in\mathbb{R}^{|A|}} \quad & f(d,v) = g_1(d,d^0) + g_2(v,v^0) \\ \text{s.t.} \quad & (d,v) \in F = \big\{(d,v) \;\big|\; \Psi(d) = v,\ d \geq 0, v \geq 0\big\}, \end{aligned} \tag{4.1}$$

where $g_1\colon \mathbb{R}^{|W|} \to \mathbb{R}$ and $g_2\colon \mathbb{R}^{|A|} \to \mathbb{R}$ are two general functions with proper weights, $d \in \mathbb{R}^{|W|}$ and $v \in \mathbb{R}^{|A|}$ are the estimates of OD demands and link flows; $d^0 \in \mathbb{R}^{|W|}$ and $v^0 \in \mathbb{R}^{|A|}$ are the prior OD demands and observed link flows; and $\Psi(\cdot)$ denotes the link flow/OD demand mapping, which may be fixed [18, 19, 121] or follows the user equilibrium (UE) condition [120, 119].

Inspired by (4.1), our proposed deterministic QSOD model has the form as shown below in (4.2), in which the general functions are defined as the L_1-norm (denoted as $\|\cdot\|_1$) of the the differences between estimates and prior (or observed) information. Here the weighing parameters among various OD demand/link flow deviations are set to one for the QSOD models, which are different from existing OD estimation models. More detailed discussions on this will be given in Chapter 6.

$$\begin{aligned} \min_{d\in\mathbb{R}^{|W|},v\in\mathbb{R}^{|A|}} \quad & f(d,v) = \|d-d^0\|_1 + \|v-v^0\|_1 \\ \text{s.t.} \quad & (d,v) \in F = \big\{(d,v) \;\big|\; \Psi(d) = v,\ d \geq 0, v \geq 0\big\}. \end{aligned} \tag{4.2}$$

Model (4.2) itself is a bi-level structure if Ψ is the UE mapping. We call this QSOD model the "bi-level QSOD" model. Assuming the true OD demands are known, one can derive the mapping between the OD demands and the link flows (by simply solving the UE problem under the true OD demands), denoted as $\hat{Q} \in \mathbb{R}^{|A| \times |W|}$. For analysis purpose, one may replace Ψ with $\hat{Q}$, leading to linear constraints in (4.2) (i.e. $\hat{Q} \cdot d = v,\ d \geq 0, v \geq 0$). We call this QSOD model the "fixed-mapping QSOD" model. In practice, $\hat{Q}$ cannot be obtained in advance (because the true OD demands are not available) and thus the fixed-mapping QSOD model may not be applied directly. However, it is often helpful to conduct mathematical analysis of the fixed-mapping QSOD model to shed useful insights on the bi-level QSOD model [121]. In this dissertation, for the deterministic case, we will first explore the fixed-mapping QSOD model in Chapter 4, and then discuss the bi-level QSOD model with a slight modification, in Chapter 5.

Note that the equality constraints in (4.2) reflects the exact UE condition (if Ψ is the UE mapping) or its fixed-mapping version (if Ψ is pre-calculated using the true OD), which are also the mostly widely used by the existing (optimization-based) OD estimation models. There is an alternative way to model this constraint as $\|\Psi(d) - v\|_1 \leq \delta$, with δ denoting certain tolerance. [97] proposed and applied this inequality constraint in the CODE model. The inequality constraint can actually help the quasi-sparsity consistency conditions in Theorem 4.1 in Section 4.2 as intuitively it "enlarges" the constraint set F (S in Theorem 4.1) and thus reduces the value of r_1, making the two conditions in Theorem 4.1 easier to hold. In the QSOD models in this dissertation, we adopt the exact UE or fixed-mapping formulation as shown in (4.2) to (i) simplify the model and discussion as selecting a proper value of δ adds another layer of complexity to the QSOD models; (ii) be consistent with the current OD estimation literature; (iii) conduct theoretical investigations of the resulting QSOD models especially the bi-level QSOD model (the inequality constraints likely make the underlying UE more of the bounded rational UE (BRUE) type [73, 36], which is much harder to study mathematically and solve numerically). We omit more detailed discussions and numerical results of QSOD models constructed using the inequality constraints.

We first assume a fixed mapping between d and v, expressed as $\hat{Q} \cdot d = v, d \geq 0, v \geq 0$, where $\hat{Q}$ is the fixed-mapping matrix reflecting the relation between OD demands and link flows. Note that $\hat{Q} \geq 0$ always holds based on its physical meaning, so the constraint $v \geq 0$ is in fact redundant; we keep it in the formulation to be consistent with the general models in which the constraint $v \geq 0$ may not be automatically satisfied. Then the QSOD framework (4.2) can be specified as the fixed-mapping QSOD model (4.3), where S denotes the feasible set

$$\begin{aligned} \min_{d \in \mathbb{R}^{|W|}, v \in \mathbb{R}^{|A|}} \quad & f(d,v) = \|d - d^0\|_1 + \|v - v^0\|_1 \\ \text{s.t.} \quad & (d,v) \in S = \{(d,v) | \hat{Q} \cdot d = v, d \geq 0, v \geq 0\}. \end{aligned} \tag{4.3}$$

Due to (4.3), $\hat{Q} \cdot d = v$ holds between the estimated OD demand d and the estimated link flow v. Since $\hat{Q}$ is generated by the true demand, this equation also holds between the true OD demand and the true link flow. However, the equation does not apply to the initial OD demand d^0 and the observed link flow v^0, i.e., $\hat{Q} \cdot d^0 = v^0$ does not hold in general. We next discuss the OD quasi-sparsity consistency and the solution sparsity of the fixed-mapping QSOD model. In addition, we list the notation used in Chapter 4-6 in Table 4.1.

Table 4.1: List of notation (deterministic QSOD models)

Notation	**Meaning**
W	set of all OD pairs
A	set of all links
$\mathcal{R}$	set of all routes
W_1	set of insignificant OD pairs
d, d^0	estimated OD demands, prior OD demands, respectively
v, v^0	estimated link flows, observed link flows, respectively

Continued on next page

Table 4.1: List of notation (deterministic QSOD models) (Continued)

Notation	**Meaning**
$(\hat{d}, \hat{v})$	optimal solution for the fixed-mapping QSOD model
(d^*, v^*)	local solution for the bi-level QSOD model
$(\epsilon_0, \alpha, \beta)$	three user-specified thresholds
Ψ	general function describing the link flow/OD demand mapping
$\hat{Q}$	fixed mapping matrix between d and v
K	set of OD demand matrices that satisfy the OD quasi-sparsity property
F	general feasible set of QSOD models
S	feasible set of the fixed-mapping QSOD model
w_1	index for insignificant OD pairs
w, a	index for any OD pair, link, respectively
r_1	the L_1 distance from d^0 to the boundary of K
r_2	the L_1 distance from (d^0, v^0) to S (or S^*)
$\hat{r}_2$	the L_1 distance from d^0 to $\hat{d}$ (or d^*)
$B_r(d^0)$	the L_1 ball centered around d^0 with radius r
e_i	unit vector
u_d, u_v	auxiliary variables
W_0^d	set of OD pairs for which the prior OD demand is zero
$c(\cdot)$	link cost function
$G(d)$	the set of feasible link flows under UE conditions given d
q	vector of route flows
Λ	OD-route incidence matrix

Continued on next page

Table 4.1: List of notation (deterministic QSOD models) (Continued)

Notation	**Meaning**
Δ	link-route incidence matrix
Q^*	Jacobian matrix of UE mapping function Ψ at d^*
$N_F(d^*, v^*)$	normal cone to the set F at (d^*, v^*)
$T_F(d^*, v^*)$	tangent cone to the set F at (d^*, v^*)
S^{**}	the approximated feasible set for the bi-level QSOD model
t	parameter for prox-linear method
ϵ_1, ϵ_2	maximum relative error for prior OD demands and observed link flows respectively
ϵ	lower bound for OD demands in the bi-level QSOD model

4.2 OD Quasi-sparsity Consistency of the Fixed-mapping QSOD Model

The use of the L_1-norm, i.e., the compressed sensing technique, in the objective of the fixed-mapping QSOD model (4.3), will lead to the sparsity of the deviations between the estimated OD demands and the prior OD demands (and the sparsity of the deviations between the estimated link flows and observed link flows). Then one may wonder, whether and under what conditions, the estimated OD will "inherit" the same quasi-sparsity as the prior OD. An affirmative answer to the question implies that if the prior OD demand matrix is representative of the true OD demand matrix, the estimated OD matrix will correctly capture the quasi-sparsity of the true OD demand matrix. The OD quasi-sparsity consistency studied in this section aims to answer this question.

First we express the OD quasi-sparsity consistency in a formal way. Recall that W_1 is the insignificant OD pair set of the prior OD demand matrix, based on given thresholds $(\epsilon_0, \alpha, \beta)$, i.e., $|W_1| \geq \alpha|W|$. Then let K be the set of OD demand matrices that satisfy the quasi-sparsity property for the given W_1, namely,

$$K = \{d \mid d_w \leq \epsilon_0,\ w \in W_1;\ \sum_{w \in W_1} d_w \leq \beta \sum_{w \in W} d_w\}. \tag{4.4}$$

We do not include the constraint $d \geq 0$ in the definition of K, because all OD demand matrices considered in this dissertation are required to be nonnegative, and the quasi-sparsity property is characterized by the constraints in (4.4).

Let $(\hat{d}, \hat{v})$ be an optimal solution to (4.3). Mathematically, the *OD quasi-sparsity consistency* question is: if the prior OD matrix d^0 belong to K, will the estimated OD matrix $\hat{d}$ also belong to K? To better understand the question visually, we graphically express the question in Figure 4.1, and give some definitions as follows.

Suppose $d^0 \in K$; let r_1 be the L_1 distance from d^0 to the boundary of K, that is,

$$r_1 = \max\{r | B_r(d^0) \subset K\}, \tag{4.5}$$

where $B_r(d^0) = \{p | \|p - d^0\|_1 \leq r\}$ is the L_1 ball centered around d^0 with radius r. Let r_2 be the L_1 distance from (d^0, v^0) to the set S in (4.3), that is,

$$r_2 = \min_{(d,v) \in S} \|d - d^0\|_1 + \|v - v^0\|_1 = \|\hat{d} - d^0\|_1 + \|\hat{v} - v^0\|_1, \tag{4.6}$$

where the second equality holds because $(\hat{d}, \hat{v})$ is an optimal solution of (4.3). Then, let $\hat{r}_2$ be the L_1 distance from d^0 to $\hat{d}$:

$$\hat{r}_2 = \|\hat{d} - d^0\|_1. \tag{4.7}$$

See Figure 4.1 for an illustration of the definitions of $r_1, r_2, \hat{r}_2$. Based on these definitions, we have the following OD quasi-sparsity result.

Theorem 4.1. *Let* $(\hat{d}, \hat{v})$ *be an optimal solution to* (4.3). *Suppose prior OD demands* d^0 *belong to* K*, the OD pair set with quasi-sparsity. If* $\hat{r}_2 \leq r_1$*, then the estimated OD demands* $\hat{d}$ *also belong to* K*.*

Proof. The definition of r_1 in (4.5) implies $B_{r_1}(d^0) \subset K$. The condition $\hat{r}_2 \leq r_1$ then implies $\hat{d} \in B_{r_1}(d^0) \subset K$. □

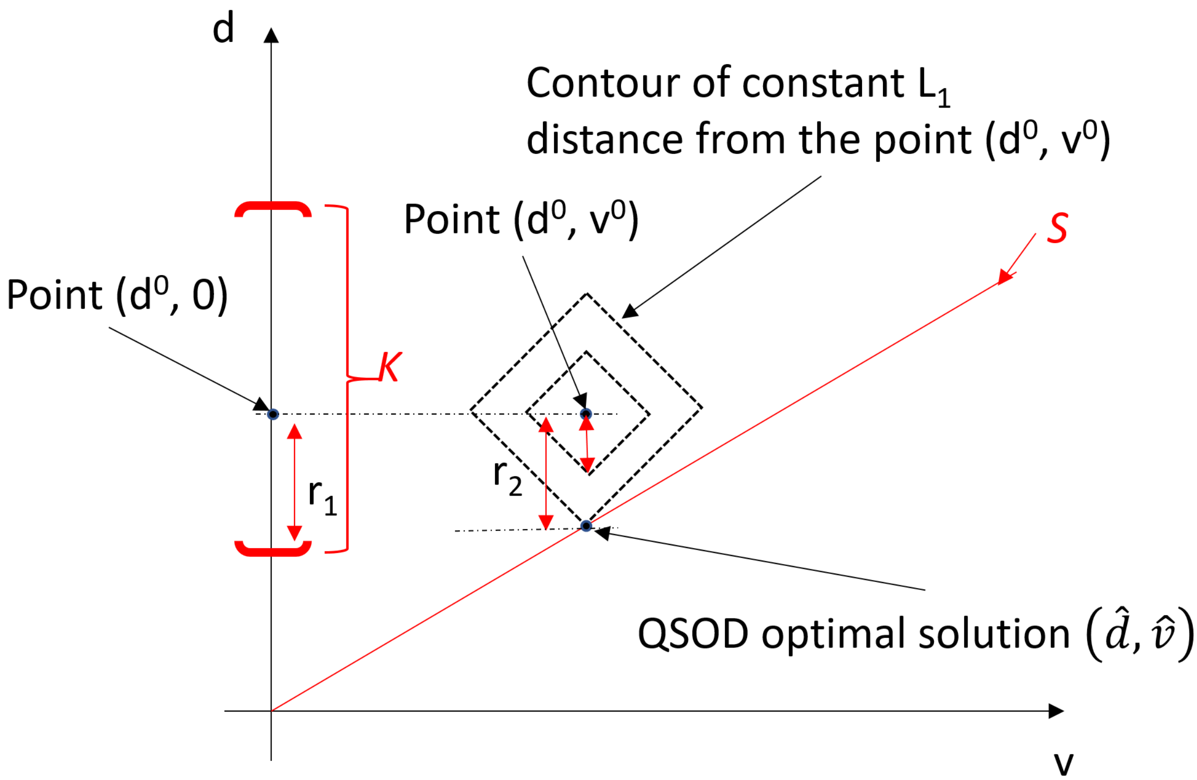


Figure 4.1: Graphic illustration for OD quasi-sparsity consistency

The values r_2 and $\hat{r}_2$ can be readily obtained after (4.3) is solved; clearly, $r_2 \geq \hat{r}_2$, so $\hat{r}_2 \leq r_1$ holds whenever $r_2 \leq r_1$ holds. To apply Theorem 4.1 without solving (4.3), we can calculate an upper bound of r_2, see Appendix A.

The value r_1 can be computed using a formula (4.8) from [107, Corollary 2.1]. Write the set K in (4.4) in the form $K = \{x | A^* x \leq b\}$, with $A^* \in \mathbb{R}^{m \times |W|}$ and $b \in \mathbb{R}^m$ given by

$$A^* = \begin{bmatrix} a_{11} & \cdots & a_{1|W|} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{m|W|} \end{bmatrix} \quad \text{and} \quad b = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}.$$

We assume without loss of generality that each row of A contains some nonzero entries. For the given point $d^0 \in K$, the L_1 distance from d^0 to the boundary of K is

$$r_1 = \min_{1\leq i\leq |W|} \left(\min_{1\leq k\leq m,\ a_{ki}\neq 0} \frac{b_k - (A^* d^0)_k}{|a_{ki}|} \right) = \min_{1\leq k\leq m} \left(\frac{[b_k - (A^* d^0)_k]}{\max_{1\leq i\leq |W|} |a_{ki}|} \right). \tag{4.8}$$

Remark 4.1. Theorem 4.1 has important implications to OD estimation and merits further discussions. First, it states that as long as the prior OD demand satisfies the OD quasi-sparsity property, under the stated conditions, the estimated OD demand will satisfy the same quasi-sparsity property, defined on the same set of insignificant OD pairs and by the same threshold parameters. As a result, if the prior OD is representative, the estimated OD matrix will have the same quasi-sparsity property as the true (and prior) OD matrix under Theorem 4.1. Since the quasi-sparsity property here only concerns about the magnitude of the OD demands and not the actual values, as long as the prior OD matrix can correctly capture the magnitude of the true OD matrix, the estimated OD matrix will also correctly capture the magnitude. This implies that the QSOD model here prefers that the prior OD matrix is "representative" of, but not necessarily "close" to, the true OD demands. Furthermore, numerical experiments in Chapter 6 show that when the estimated OD matrix satisfies the same quasi-sparsity property as the prior (and true) OD matrix, the estimation error is also reasonably small. Therefore when preparing for the prior OD demand matrix in practice, one should focus more on the representativeness of the prior OD demands, which is expected to be less restrictive than requiring the prior OD is close to the true OD .

4.3 Solution Sparsity of the Fixed-mapping QSOD model

In this section we consider the solution sparsity of (4.3). We are interested in how the use of L_1-norm leads to a sparse deviation between the solution of (4.3) and the prior OD demand d^0 and the observed link flow v^0, and what type of sparsity is presented in the solution to (4.3). Theorem 4.2 below provides insights related to these questions.

The proof of Theorem 4.2 uses Lemma 4.1 below, which gives a solution sparsity result for an extreme point of the set of optimal solutions to the problem that minimizes the L_1

distance from a given point to a nonempty polyhedral convex set. By definition, an extreme point of a polyhedral convex set is a point in this set that is not a convex combination of two other points in this set different from the point itself; see [13] and [94]. The proof of this lemma is in [67].

Lemma 4.1. *Let $M \in \mathbb{R}^{m\times n}$, $N \in \mathbb{R}^{k\times n}$, $x^0 \in \mathbb{R}^n$, $b \in \mathbb{R}^m$ and $h \in \mathbb{R}^k$ be fixed. Suppose the following problem has at least one feasible solution:*

$$\min \quad \|x - x^0\|_1 \quad s.t. \quad Mx = b, \quad Nx \geq h. \tag{4.9}$$

Then, the set of optimal solutions to (4.9) *is a nonempty, bounded polyhedral convex set that has at least one extreme point. Let x^* be an extreme point of the set of optimal solutions to* (4.9)*, and define the following index sets,*

$$E = \{i = 1, \cdots, k \mid (Nx)_i = h_i\}, \quad J_P = \{j = 1, \cdots, n \mid x^*_j > x^0_j\},$$
$$J_Z = \{j = 1, \cdots, n \mid x^*_j = x^0_j\}, \qquad J_N = \{j = 1, \cdots, n \mid x^*_j < x^0_j\};$$

we have

$$\text{Rank} \begin{bmatrix} M_{:,J_P} & M_{:,J_N} \\ N_{E,J_P} & N_{E,J_N} \end{bmatrix} + |J_Z| = n, \tag{4.10}$$

where $M_{:,J_P}$ and $M_{:,J_N}$ are submatrices of M containing columns with indices in J_P and J_N respectively, and N_{E,J_P} and N_{E,J_N} are submatrices of N containing rows with indices in E and columns with indices in J_P and J_N respectively.

Theorem 4.2 states that the set of optimal solutions of (4.3) has at least one extreme point, and that any extreme point $(\hat{d}, \hat{v})$ satisfying a solution sparsity results in (4.12). The set W_0 there contains indices $w \in W$ for which $\hat{d}_w = 0$. The sets W_P, W_Z and W_N are defined based on the relation between $\hat{d}$ and d^0, and they are mutually exclusive and form a partition of W. Because $d^0 \geq 0$, we have $W_0 \cap W_P = \emptyset$. The subsets of the link set A are defined in a similar way based on $\hat{v}$. The proof of Theorem 4.2 is given in [111].

Theorem 4.2. *The set of optimal solutions of* (4.3) *is a nonempty, bounded polyhedral convex set that has at least one extreme point. Let $(\hat{d}, \hat{v})$ be an extreme point of the set of* *optimal solutions of* (4.3)*, and define the following index sets*

$$\begin{aligned} & W_0 = \{w \in W: \ \hat{d}_w = 0\}, \\ & W_P = \{w \in W: \ \hat{d}_w - d_w^0 > 0\}, \quad A_P = \{a \in A: \ \hat{v}_a - v_a^0 > 0\}, \\ & W_Z = \{w \in W: \ \hat{d}_w - d_w^0 = 0\}, \quad A_Z = \{a \in A: \ \hat{v}_a - v_a^0 = 0\}, \\ & W_N = \{w \in W: \ \hat{d}_w - d_w^0 < 0\}, \quad A_N = \{a \in A: \ \hat{v}_a - v_a^0 < 0\}. \end{aligned} \tag{4.11}$$

Then,

$$|W_0 \cup W_Z| + |A_Z| \geq |W|. \tag{4.12}$$

The above result is particularly interesting for large-size networks where $|W| >> |A|$ normally holds. In those situations, $|A_Z|$ is much smaller than $|W_0 \cup W_Z|$, and (4.12) implies $|W_0 \cup W_Z| \approx |W|$, that is, $|W_0 \cup W_Z|$ is "approximately equal to" $|W|$. This indicates that the solution of the fixed-mapping QSOD model (4.3) has the following property when $|W| >> |A|$: for most OD pairs, the estimated OD demands are either zero or the same as the prior OD demands. The QSOD solution thus reflects both the sparsity feature (in terms of zero demand) and the use of the prior OD (in terms of OD entries identical to the prior OD), which adds to the current OD estimation literature when compared with SOD models. This theorem also provides the theoretical explanation why the fixed-mapping QSOD model helps maintain the OD quasi-sparsity consistency; see Remark 5.2 in Chapter 5. Note that Theorem 4.2 does not imply that the prior OD and the estimated OD are almost identical. This is because, similar to the results of SOD models, the OD entries that are not zero or equal to the prior OD (analogous to the non-zero entries in the demand matrix produced by SOD models), although very small in terms of their numbers (roughly the size of the number of links in the network) as shown in (4.12), represent a significant portion of the total demands of the network. Therefore we expect that the estimated OD is different from (and improves) the prior OD, as illustrated, e.g., in Table 6.3 in Chapter 6. Finally, define

$$W_0^d = \{w \in W: \ d_w^0 = 0\}, \tag{4.13}$$

as the set of OD pairs for which the prior OD demand is zero. The following corollary of Theorem 4.2 gives a relation between the cardinalities of W_0 and W_0^d, that is, between the number of zero elements in the estimated demand matrix and the number of zero elements in the prior OD demand matrix.

Corollary 4.1. *Let* $(\hat{d}, \hat{v})$ *be an extreme point to the set of optimal solutions to* (4.3)*, and define index sets as in* (4.11) *and* W_0^d *as in* (4.13)*. Then,*

$$|W_0| \geq |W_0^d| - |A_Z|. \tag{4.14}$$

Proof. Because $W_0 \cup W_Z$ is the union of the two disjoint sets W_0 and $W_Z \setminus W_0$, and W_0^d is disjoint from $W_Z \setminus W_0$, (4.14) follows from (4.12). □

Again, for OD estimation on a large network in which $|W| >> |A|$, we can ignore the cardinalities of A_Z in (4.14) to obtain $|W_0| \gtrsim |W_0^d|$, where $\gtrsim$ stands for "approximately greater than". This implies that the estimated OD flow $\hat{d}$ have the similar sparsity level as the prior OD demand d^0.

Chapter 5

BI-LEVEL QUASI-SPARSITY OD ESTIMATION MODEL

We analyze the bi-level QSOD model in this chapter, where the upper-level problem tries to minimize the sum of deviations (in terms of L_1-norm) between the estimates and their prior values (or observed values), while the lower-level problem presents a user equilibrium assignment between the OD demands and corresponding link flows. Derived from the QSOD framework (4.2), the bi-level QSOD model has the following formulation:

$$\begin{aligned} \min_{d\in\mathbb{R}^{|W|}, v\in\mathbb{R}^{|A|}} \quad & f(d,v) = \|d-d^0\|_1 + \|v-v^0\|_1 \\ \text{s.t. } \ & (d,v) \in F = \big\{(d,v) \;\big|\; \Psi(d) = v,\ d \geq \epsilon, v \geq 0\big\}, \end{aligned} \tag{5.1}$$

where $\Psi(d)$ is the user-equilibrium link flow under the OD demand d, and ϵ is a (very) small positive value representing the lower bound of d, which will be discussed next. More specifically, $\Psi(d)$ is the solution to the following traffic user equilibrium problem:

$$v \in G(d) \quad \text{and} \quad \langle c(v), v'-v\rangle \geq 0 \quad \text{for each} \quad v' \in G(d). \tag{5.2}$$

Here $c(\cdot) : \mathbb{R}_+^{|A|} \to \mathbb{R}^{|A|}$ is the link cost function, and $G(d)$ is the set of feasible link flows under d:

$$G(d) = \left\{ v \in \mathbb{R}^{|A|} \mid \text{ for some } q \in \mathbb{R}_+^{|\mathcal{R}|},\ \Lambda q = d, \Delta q = v \right\}, \tag{5.3}$$

where Λ and Δ are the OD-route incidence matrix and the link-route incidence matrix respectively, and q denotes the route flow vector (recall that $\mathcal{R}$ is the set of all routes). We make the following assumption to ensure that Ψ is a well-defined single valued function.

Assumption 5.1. The link cost function c is continuous on $\mathbb{R}_+^{|A|}$, and is strictly monotone on $\mathbb{R}_+^{|A|}$ in the sense that $\langle c(v) - c(v'), v - v'\rangle > 0$ for any $v,\ v' \in \mathbb{R}_+^{|A|}$ with $v \neq v'$.

As shown in [103], Assumption 5.1 guarantees that the problem (5.2) has a unique solution in $\mathbb{R}_+^{|A|}$ for each $d \in \mathbb{R}_+^{|W|}$, and therefore guarantees that $\Psi(d)$ is well defined for each $d \in \mathbb{R}_+^{|W|}$. The user equilibrium condition can be expressed as a few formulations, such as the non-linear optimization problem [10], the nonlinear complementarity problem [4, 6], and the variational inequality (VI) formulation [33, 46]. In this dissertation, we employ its variational inequality form. Note that the definition of $\Psi(d)$ here guarantees that $v = \Psi(d) \geq 0$ for any $d \geq 0$, so the constraint $v \geq 0$ in (5.1) is in fact redundant. We keep this constraint in the formulation to make it convenient for comparison with its first-order approximation.

Note that in (5.1) we require d to be lower-bounded by a small value ϵ (e.g., $\epsilon = 10^{-5}$ or $\epsilon = 10^{-6}$). In sensitivity analysis for traffic user equilibrium problems it is common to assume $d > 0$ to compute the Jacobian matrix of Ψ at a given point d because Ψ is not defined for d with negative entries and is therefore not differentiable at any d with zero entries. The influence of a small value ϵ to the solution of (5.1) is very small in real applications, especially for large networks. For example, the downtown Seattle network discussed in Chapter 6 has over 10,000 OD pairs and more than 580,000 total trips and we use $\epsilon = 10^{-5}$.

In the remainder of the chapter, we start with the reformulation of the bi-level QSOD model. Then we leverage the findings for the fixed-mapping QSOD model (Chapter 4) to investigate the OD quasi-sparsity consistency and solution sparsity of the bi-level QSOD model.

5.1 Reformulation of the Bi-level QSOD Model

Bi-level problems or mathematical programs with equilibrium constraints (MPEC) have been studied extensively in the fields of transportation [118] and mathematical optimization [69]. In this section we give the relation between a local solution to the bi-level QSOD model (5.1) and an optimal solution to a first-order approximation of (5.1) at a point (d^*, v^*) satisfying the following assumption.

Assumption 5.2. Let $(d^*, v^*) \in \mathbb{R}^{|W|+|A|}$ belong to F in (5.1), and define index sets

$$W_\epsilon^* = \{w \in W : d_w^* = \epsilon\} \quad \text{and} \quad A_0^* = \{a \in A : v_a^* = 0\}. \tag{5.4}$$

Suppose that there exists an open neighborhood $\mathcal{O}$ of d^* in $\mathbb{R}_{++}^{|W|}$ such that the function Ψ is continuously differentiable on $\mathcal{O}$, with the Jacobian matrix of Ψ at d^* given by

$$Q^* = \nabla\Psi(d^*). \tag{5.5}$$

The following problem is a first-order approximation of (5.1) at (d^*, v^*):

$$\begin{aligned} \min_{d,v} \quad & f(d,v) = \|d - d^0\|_1 + \|v - v^0\|_1 \\ \text{s.t.} \quad & (d,v) \in S^{**} = \left\{(d,v) \;\middle|\; Q^*(d-d^*) + v^* = v;\ d \geq \epsilon,\ v \geq 0\right\}. \end{aligned} \tag{5.6}$$

The matrix Q^* from (5.5) has a special property that the rows in Q^* corresponding to links $a \in A_0^*$ are entirely zero. This is a consequence of the facts that $(\Psi(d))_a \geq 0$ for any nonnegative $d \in \mathbb{R}^{|W|}$, $(\Psi(d^*))_a = 0$, and Ψ is differentiable at d^*. More discussions about Assumption 5.2 are presented in Appendix B, where we give a condition under which it holds and a formula from [66] to compute the Jacobian matrix Q^* in (5.5). Note that, different from the traffic assignment matrix $\hat{Q}$ defined in (4.3), the matrix Q^* may contain some negative entries, see, e.g., [45] for a related discussion. The matrix Q^* computed for the second example in [66] also contains negative entries.

Part (1) of Theorem 5.1 gives a necessary condition for a point (d^*, v^*) satisfying Assumption 5.2 to be a local solution to (5.1), and part (2) of it gives a sufficient condition for it to be a strict local solution to (5.1). Both conditions are based on the first-order approximation problem (5.6). The point (d^*, v^*) is said to be a local solution to (5.1) if $f(d^*, v^*) \leq f(d, v)$ for any (d, v) in a neighborhood of (d^*, v^*) in F, and it is said to be a strict local solution to (5.1) if $f(d^*, v^*) < f(d, v)$ for any $(d, v) \neq (d^*, v^*)$ in a neighborhood of (d^*, v^*) in F, see [83, Chapter 12]. In the proof we use the fact that f is a piecewise affine function, meaning that it is continuous and takes values from finitely many affine functions of (d, v) on $\mathbb{R}^{|W|+|A|}$.

As a result, f is B-differentiable at the given point (d^*, v^*), which means that there exists a positively homogenuous function $Df(d^*, v^*)$ from $\mathbb{R}^{|W|+|A|}$ to $\mathbb{R}$, such that

$$f(d^* + d, v^* + v) - f(d^*, v^*) - Df(d^*, v^*)(d, v) = o(\|(d, v)\|).$$

The requirement that $Df(d^*, v^*)$ is positively homogenuous means that $Df(d^*, v^*)(\tau d, \tau v) = \tau Df(d^*, v^*)(d, v)$ for any $\tau \geq 0$ and $(d, v) \in \mathbb{R}^{|W|+|A|}$. B-differentiabilty is a generalization of differentiabiliy and is introduced in [91]; the quantity $Df(d^*, v^*)(d, v))$ is called the B-derivative of f at (d^*, v^*) for the direction (d, v), which is also the directional derivative of f at (d^*, v^*) for (d, v). See [92, 44, 99] for detailed discussions about piecewise affine functions and B-differentiable functions. Readers can refer to [111] for the proof of Theorem 5.1.

Theorem 5.1. *For any point $(d^*, v^*) \in \mathbb{R}^{|W|+|A|}$ satisfying Assumption 5.2, the following statements hold.*

(1) If (d^*, v^*) is a local solution to (5.1)*, then (d^*, v^*) is an optimal solution to* (5.6)*.*

(2) If (d^*, v^*) is the unique optimal solution to (5.6)*, then (d^*, v^*) is a strict local solution to* (5.1)*.*

Theorem 5.1 enables us to analyze properties of local solutions to the bi-level QSOD model (5.1) using similar methods for the fixed-mapping QSOD model in Chapter 4, by considering the corresponding formulation (5.6) based on these local solutions.

5.2 OD Quasi-sparsity Consistency of the Bi-level QSOD Model

In this section we give a condition for a local solution to (5.1) to belong to the set K defined in (4.4) given that d^0 belongs to K, see Theorem 5.2 below.

Theorem 5.2. *Let (d^*, v^*) be a local solution of* (5.1) *that satisfies Assumption 5.2. Suppose prior OD demands d^0 belong to K, the OD pair set with quasi-sparsity. Let r_1 be as defined in* (4.5)*, r_2 be the optimal value of* (5.6)*, and $\hat{r}_2 = \|d^* - d^0\|_1$. Then $\hat{r}_2 \leq r_2$. Moreover, if $\hat{r}_2 \leq r_1$, then the estimated OD demands d^* also belong to K.*

Proof. By Theorem 5.1, (d^*, v^*) is an optimal solution to (5.6). Thus, $r_2 = \|d^* - d^0\|_1 + \|v^* - v^0\|_1$, which implies $\hat{r}_2 \le r_2$. Moreover, if $\hat{r}_2 \le r_1$, then d^* belongs to the closed ball around d^0 with radius r_1, and therefore belongs to K by the definition of r_1. □

The reason why we define r_2 as the optimal value of (5.6) instead of (5.1) is that we do not know whether (d^*, v^*) is a global solution to (5.1). If (d^*, v^*) is a global solution to (5.1), then the optimal value of (5.1) is also r_2; otherwise, the optimal value of (5.1) may be smaller than r_2.

5.3 Solution Sparsity of the Bi-level QSOD Model

In this section, we apply Lemma 4.1 and Theorem 5.1 to obtain a solution sparsity result for (5.1). We refer the readers to [111] for the proof of the theorem. In the proof we use the fact that rows in the matrix Q^* defined in Assumption 5.2 corresponding to links $a \in A_0^*$ are zero.

Theorem 5.3. *Suppose $(d^*, v^*) \in \mathbb{R}^{|W|+|A|}$ satisfies Assumption 5.2 and is the unique optimal solution to (5.6). Define the index sets W_ϵ^* and A_0^* as in (5.4), and define*

$$\begin{aligned} W_P^* &= \{w \in W : \; d_w^* - d_w^0 > 0\}, \quad A_P^* = \{a \in A : \; v_a^* - v_a^0 > 0\}, \\ W_Z^* &= \{w \in W : \; d_w^* - d_w^0 = 0\}, \quad A_Z^* = \{a \in A : \; v_a^* - v_a^0 = 0\}, \\ W_N^* &= \{w \in W : \; d_w^* - d_w^0 < 0\}, \quad A_N^* = \{a \in A : \; v_a^* - v_a^0 < 0\}. \end{aligned} \tag{5.7}$$

Then, (d^*, v^*) is a strict local solution to (5.1), and satisfies

$$|W_\epsilon^* \cup W_Z^*| + |A_Z^* \setminus A_0^*| \ge |W|. \tag{5.8}$$

Again, based on the fact that $|W| >> |A|$ in a typical large-size traffic network, we can conclude from the above theorem that $|W_\epsilon^* \cup W_Z^*| \approx |W|$. This means that the estimated OD demand from the bi-level problem (5.1) for a majority of OD pairs is either equal to the small value ϵ or equal to the prior OD demand of that OD pair, when the number of OD pairs is much greater than the number of links. We also have the following two remarks.

Remark 5.1. The above results show that, under the stated assumptions and conditions, the solution of the bi-level QSOD model has very similar properties as those of the fixed-mapping QSOD model, when compared with the prior OD demands. In particular, when applying the bi-level QSOD model to OD estimation that assumes the prior demand matrix is quasi-sparse, the estimated OD demand matrix (i.e., the solution of the bi-level QSOD model) is also quasi-sparse.

Remark 5.2. Theorem 4.2 and Theorem 5.3 have important practical implications. They indicate that the QSOD models (either the fixed-mapping case or the bi-level case) can produce an estimated OD matrix with demands for some OD pairs (most OD pairs for large-size networks, i.e., $|W| >> |A|$) either the same as the prior OD demands or zero (or very close to zero for the bi-level QSOD model). Since we can reasonably assume that most OD pairs with zero (or very close to zero) estimated demand are also insignificant OD pairs in the prior OD matrix, this implies that those OD pairs in the estimated OD matrix will continue to be insignificant as in the prior OD matrix. As a result, the two theorems attest that most OD pairs in the estimated OD matrix will remain significant or insignificant as in the prior OD matrix. This is true even the conditions in Theorem 4.1 or Theorem 5.2 do not hold (thus quasi-sparsity consistency does not hold strictly). It provides the theoretical explanation why the QSOD models have superb performance of maintaining quasi-sparsity consistency (i.e., higher accuracy and F1-score values as shown in Chapter 6 below) compared with other OD estimation models.

5.4 Algorithm for the Bi-level QSOD Model

Solving the QSOD model (4.3) for the fixed-mapping case is straightforward, as one only needs to solve the equivalent linear program below,

$$\begin{aligned} \min_{d,u_d,u_v} \quad & \sum_{w\in W}(u_d)_w + \sum_{a\in A}(u_v)_a \\ \text{s.t.} \quad & d \geq 0, \\ & u_d \geq d - d^0,\ u_d \geq d^0 - d, \\ & u_v \geq \hat{Q}d - v^0,\ u_v \geq v^0 - \hat{Q}d, \end{aligned} \tag{5.9}$$

where $u_d \in \mathbb{R}^{|W|}$ and $u_v \in \mathbb{R}^{|A|}$ are auxiliary variables. In any optimal solution to (5.9), the relations $(u_d)_w = |d_w - d_w^0|$ and $(u_v)_a = |(\hat{Q}d)_a - v_a^0|$ hold element-wise.

The bi-level QSOD model, however, is much more challenging to solve as the lower-level problem in (5.1) contains complementarity conditions hence the mapping function $\Psi(d)$ is usually hard to differentiate. In this section, we discuss the algorithm to solve the bi-level (UE) QSOD model for relatively large-size problems, which usually include a large number of variables.

We first reformulate (5.1) as follows:

$$\min_{d\geq\epsilon} \quad \|d - d^0\|_1 + \|\Psi(d) - v^0\|_1. \tag{5.10}$$

The non-linearity and non-convexity of $\Psi(d)$ bring the difficulty for the optimization of (5.10). Here we apply the prox-linear method (Algorithm 1), which was investigated recently for composite optimization problems [17, 61, 40, 41].

In each iteration, the algorithm approximates $\Psi(d)$ at d_k by its linearization $\Psi(d_k) + \nabla\Psi(d_k)(d-d_k)$, where $\Psi(d_k)$ is the user-equilibrium link flows given OD demands d_k, $\nabla\Psi(d_k)$ represents the Jacobian matrix of $\Psi(\cdot)$ at d_k. For the calculation of $\nabla\Psi(d_k)$, we refer to the work by [66]. Detailed discussions can be found in Appendix B.

As each iteration in Algorithm 1 requires to solve a new large-scale optimization problem,

Algorithm 1 Prox-linear method

Initialize: A starting point $d_0 \in \mathbb{R}_{++}^{|W|}$ and a real $t > 0$.

Step k: $(k \geq 0)$ Compute

$$d_{k+1} = \operatorname*{argmin}_{d \geq \epsilon} \|d - d^0\|_1 + \|\Psi(d_k) + \nabla\Psi(d_k)(d - d_k) - v^0\|_1 + \frac{1}{2t}\|d - d_k\|_2^2 \tag{5.11}$$

Stop: when $\|d_{k+1} - d_k\|_2 \leq \epsilon^* \cdot \|d_k\|_2$ or $k \geq$ maximum allowed iterations

we further make some reformulations to help efficiently solve (5.11). Let

$$A_k = \nabla\Psi(d_k), \tag{5.12}$$

$$b_k = \nabla\Psi(d_k)d_k + v^0 - \Psi(d_k). \tag{5.13}$$

Problem (5.11) is then re-written as:

$$\min_{d \geq \epsilon} \|A_k d - b_k\|_1 + \|d - d^0\|_1 + \frac{1}{2t}\|d - d_k\|_2^2. \tag{5.14}$$

For the selection of t, we need to make sure $1/t \geq ||A_k||_1$. Let $s_1 = A_k d - b_k$ and $s_2 = d - d^0$. We split s_1 and s_2 into their nonnegative and nonpositive parts. That is, $s_1 = s_1^+ - s_1^-$ and $s_2 = s_2^+ - s_2^-$, where $s_1^+ = \max\{s_1, 0\}$, $s_1^- = \max\{-s_1, 0\}$, $s_2^+ = \max\{s_2, 0\}$, $s_2^- = \max\{-s_2, 0\}$. We can reformulate (5.14) into the following quadratic programming problem:

$$\begin{aligned} \min_{d, s_1^+, s_1^-, s_2^+, s_2^-} \quad & \mathbf{1}^T s_1^+ + \mathbf{1}^T s_1^- + \mathbf{1}^T s_2^+ + \mathbf{1}^T s_2^- + \frac{1}{2t}\|d - d_k\|_2^2 \\ \text{s.t.} \quad & s_1^+ - s_1^- = A_k d - b_k \\ & s_2^+ - s_2^- = d - d^0 \\ & d \geq \epsilon, s_1^+, s_1^-, s_2^+, s_2^- \geq 0. \end{aligned} \tag{5.15}$$

Problem (5.15) can be written in the following standard form

$$\begin{aligned} \min_x \quad & \frac{1}{2}x^T Q x + c^T x \\ \text{s.t.} \quad & Ax = b, \quad x \geq 0. \end{aligned} \tag{5.16}$$

with $x = (d, s_1^+, s_1^-, s_2^+, s_2^-) \in \mathbb{R}^{3|W|+2|A|}$, where $d \geq \epsilon$,

$$Q = \begin{bmatrix} \frac{1}{t}\mathbf{I}_{|W|} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \quad c = \begin{bmatrix} -\frac{1}{t}d_k & \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{1} \end{bmatrix} \tag{5.17}$$

and

$$A = \begin{bmatrix} A_k & -\mathbf{I}_{|A|} & \mathbf{I}_{|A|} & \mathbf{0} & \mathbf{0} \\ \mathbf{I}_{|W|} & \mathbf{0} & \mathbf{0} & -\mathbf{I}_{|W|} & \mathbf{I}_{|W|} \end{bmatrix}, \quad b = \begin{bmatrix} b_k \\ d^0 \end{bmatrix}. \tag{5.18}$$

So far, each iteration of Algorithm 1 requires to solve no more than a quadratic programming problem. We can apply standard interior point method to solve this problem to a certain accuracy. However for large-scale problems, applying interior point method may face the issue of slow convergence. Here we use the Alternating Direction Method of Multipliers (ADMM) [15] to solve (5.15) in each iteration of the prox-linear algorithm for the downtown Seattle example. Because of the complexity of the bi-level model, we could not prove the convergence of the prox-linear algorithm; instead some convergence performance plots are shown numerically in Section 6.2.

Chapter 6

NUMERICAL EXPERIMENTS OF QSOD MODELS

To test the proposed fixed-mapping and bi-level QSOD models and validate the theoretical results, two testing networks are used in this chapter. The first one is a small network (5 nodes and 16 links) and the second one is a real (and larger) network in downtown Seattle. We show how the OD quasi-sparsity consistency is kept by the QSOD models, and the properties of the QSOD solutions that are consistent with the theoretical findings in the last two chapters. Detailed comparisons between QSOD and existing OD estimation methods (e.g., GLS in [120] and the SOD estimation model in [77]) are presented in Section 6.3.

6.1 A Small Network

The small network (depicted in Figure 6.1) contains 5 nodes and 16 directed links, where each node can generate and receive travel demands, resulting in 20 OD pairs (we only consider demands between different nodes). Table 6.1 gives the link specification of the network. To test the QSOD models, we assume the true OD matrix is known (shown in Table 6.2), which simulates the quasi-sparsity property of the OD matrix. The set of the significant OD pairs is $W \setminus W_1 = \{(2,4),(2,5),(3,5)\}$, where W_1 is the insignificant OD pair set (assuming $\epsilon_0 = 5, \alpha = 85\%, \beta = 15\%$ in the definitions in Section 3.1).

Given the true OD demands and link specification, the travel demands can be assigned to links based on the Wardrop's UE principle [113]. The link cost function follows the Bureau of Public Roads (BPR) type formula [84]:

$$c_a(v_a) = FFT_a \times \{1 + (\frac{v_a}{C_a^0})^4\}. \tag{6.1}$$

where $c_a(v_a)$ denotes the travel time (cost) for link a given link flow v_a, FFT_a is the free flow time (cost) for link a, and C_a^0 is the capacity of link a, as shown in Table 6.1. To obtain the

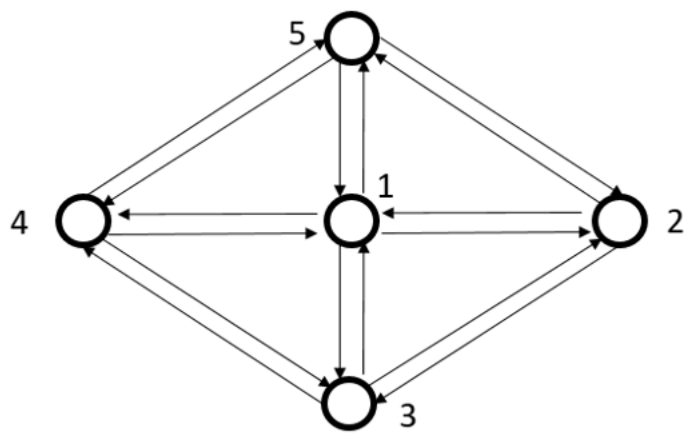


Figure 6.1: A five node network

Table 6.1: Link specification

Link	(1,2) (2,1)	(1,3) (3,1)	(1,4) (4,1)	(1,5) (5,1)	(2,3) (3,2)	(3,4) (4,3)	(4,5) (5,4)	(2,5) (5,2)
FFT_a	8	5	8	5	12	14	12	14
C_a^0	1200	1000	1000	1200	600	800	600	800

prior OD matrix and observed link flows, we add random errors to the true OD demand ($\bar{d}$) and link flow ($\bar{v}$) as follows:

$$d^0 = \bar{d} \times (1 + \Delta_1) \tag{6.2}$$

$$v^0 = \bar{v} \times (1 + \Delta_2) \tag{6.3}$$

Table 6.2: True OD demands

OD nodes	1 (D)	2	3	4	5
1 (O)		1	4	2	2
2	1		4	1500	1600
3	3	4		1	1600
4	2	1	1		5
5	4	3	2	2	

where $\Delta_1 \sim uniform(-\epsilon_1, \epsilon_1)$ and $\Delta_2 \sim uniform(-\epsilon_2, \epsilon_2)$. Element-wise, the error (as measured to be proportional to the true OD) of the prior OD demand matrix has an upper bound of ϵ_1, compared with the true OD demand matrix. Similarly, ϵ_2 is the upper bound of the error term for observed link flows. In most of the numerical experiments in this section, we set ϵ_1 and ϵ_2 in the range of $(0, 1)$, except some in Section 6.3.

6.1.1 OD quasi-sparsity consistency

To demonstrate how the OD quasi-sparsity consistency is maintained by the QSOD models, we first show the results of solving the fixed-mapping and bi-level QSOD models, and then discuss the OD quasi-sparsity consistency theorems for the models.

In order to solve the fixed-mapping QSOD model (4.3), given the true OD matrix and link specification, the first step is to establish the mapping matrix $\hat{Q}$ between link flows and OD demands under the UE condition. As pointed out in the literature [8, 9, 68], for traffic assignment problems with BPR-type link cost functions, link flows are uniquely determined while route flows may not be unique. Here we apply the Origin-Based Algorithm [9] to solve the possible route flows under the UE condition and further build the link/OD demand proportional matrix $\hat{Q}$. Then the fixed-mapping QSOD model can be solved by any LP solver, through proper reformulation such as (5.9). For the bi-level QSOD model, as the dimension of the problem is quite small for this network (20 OD pairs and 16 links), it does not require the use of the prox-linear method in Section 5.4. We can formulate the UE constraints into a link-node nonlinear complimentary problem [7, 6], which can be solved using the MPEC solver in GAMS [6]. We also use $\epsilon = 10^{-5}$ for the bi-level QSOD model.

Assuming $\epsilon_1 = 0.2, \epsilon_2 = 0.02$ (at most 20% and 2% errors are introduced to obtain prior OD demands and observed link flows, respectively), Table 6.3 shows the results of solving the fixed-mapping and bi-level QSOD models. It is observed that the estimated OD demands have a much lower Root Mean Squared Error (RMSE, see its definition in Appendix C) than the prior OD demands (the RMSE of OD estimates from the fixed-mapping and bi-level QSOD models reduce by 51.12% and 36.20% respectively, compared with prior OD demands), showing QSOD's capability to improve OD demand quality. Furthermore, the estimated OD demands for both fixed-mapping and bi-level QSOD models satisfy the quasi-sparsity property based on the same W_1 (when $\epsilon_0 = 5$): all OD pairs in the estimated demand matrix remain insignificant/significant as they do in the true/prior OD demand matrices, implying the quasi-sparsity property of the OD demand is maintained by the two QSOD models.

We then check the OD quasi-sparsity consistency conditions for the fixed-mapping case and bi-level case. Table 6.4 presents six random trials conducted based on the small network. The first three trials are for the fixed-mapping case, showing respectively that 1) conditions in Theorem 4.1 do not satisfy, quasi-sparsity consistency still holds, 2) conditions in Theorem 4.1 satisfy, quasi-sparsity consistency holds, and 3) conditions in Theorem 4.1 do not satisfy, quasi-sparsity consistency does not hold either. Similar to the fixed-mapping case, Trial 4 - Trial 6 are to validate the quasi-sparsity consistency theorem (Theorem 5.2) for the bi-level QSOD model. Various levels of errors were introduced into the prior OD demands and links flows for the six trials. They show that Theorem 4.1 and Theorem 5.2 are the sufficient conditions for OD quasi-sparsity consistency, which means once the required conditions are satisfied, the derived estimated OD demands will keep the same quasi-sparsity feature as the prior OD demands. On the other hand, the quasi-sparsity consistency may still hold even

Table 6.3: Comparison between OD true values, priors, and estimates (small network)

O-D	True demands $\bar{d}$	Prior demands d^0	Estimated demands (fixed-mapping)	Estimated demands (bi-level)
1-2	1	1.260	0.768	1.260
1-3	4	5.000	5.000	5.000
1-4	2	1.872	1.872	1.872
1-5	2	1.525	1.525	1.525
2-1	1	1.231	1.231	1.231
2-3	4	3.432	3.432	3.432
2-4	1500	1194.807	1464.500	1300.802
2-5	1600	1831.703	1831.703	1831.703
3-1	3	3.600	0	3.600
3-2	4	3.995	3.995	3.995
3-4	1	1.199	1.199	1.199
3-5	1600	1890.221	1615.623	1574.862
4-1	2	1.735	1.735	0.00001
4-2	1	0.686	0.686	0.00001
4-3	1	0.557	0.557	0.00001
4-5	5	5.000	5.000	0.00001
5-1	4	4.175	4.175	4.175
5-2	3	3.577	3.577	0.00001
5-3	2	2.043	1.336	2.043
5-4	2	2.105	2.029	2.105
RMSE		107.48	52.54	68.57

those conditions do not hold. We conduct 20 more testing cases by randomly generating ϵ_1 and ϵ_2, which further confirms the above findings. In particular, the results show that if $\epsilon_1 \leq 0.5, \epsilon_2 \leq 0.5$, the quasi-sparsity consistency holds for most cases. Details are omitted here to save space.

Table 6.4: Results for quasi-sparsity consistency (small network)

Trial	Model	r_1	r_2	$\hat{r}_2$	$r_2 \leq r_1$ (or $\hat{r}_2 \leq r_1$)	$d^0 \in K$	$\hat{d} \in K$ (or $d^* \in K$)
1	Fixed-mapping	0.954	147.391	4.332	No	Yes	Yes
2	Fixed-mapping	0.996	0.930	0.882	Yes	Yes	Yes
3	Fixed-mapping	0.025	116.232	110.271	No	Yes	No
4	Bi-level	0.984	12.457	7.292	No	Yes	Yes
5	Bi-level	0.700	0.723	0.314	Yes	Yes	Yes
6	Bi-level	0.775	176.572	28.894	No	Yes	No

6.1.2 QSOD solution sparsity

The QSOD solution sparsity in this dissertation focuses on how the prior information (e.g., prior ODs and observed link flows) will affect the estimated OD demands. Specifically, according to Theorem 4.2, Theorem 5.3 and their corollaries, the estimated OD demands from QSOD models will be mostly equal to either the prior OD demands or zero (or the small value ϵ for the bi-level QSOD model).

Revisiting Table 6.3, for the fixed-mapping case, among the 20 OD pairs of the network, 14 OD pairs have the estimated demands exactly equal to the prior OD demands (including 13 insignificant OD pairs and 1 significant OD pair). For the bi-level case, 13 OD pairs, including 12 insignificant OD pairs and one significant OD pair, have the estimated demands equal to the prior OD demands. Such results demonstrate that the proposed QSOD framework can capture the sparse deviation between the estimated demand and prior demand, which can ascribe to the use of the L_1-norm.

In addition, the number of OD pairs with estimated demands equal to prior OD demands or zeros (or the small value ϵ for the bi-level QSOD model), and the number of links with estimated flows equal to observed link flows or zeros are counted for the example in Table 6.3. The result is shown in Table 6.5. It shows that $|W_0 \cup W_Z| + |A_Z| \geq |W|$ holds for the fixed-mapping case and $|W_\epsilon^* \cup W_Z^*| + |A_Z^* \setminus A_0^*| \geq |W|$ holds for the bi-level case, which validate Theorem 4.2 and Theorem 5.3. We also tested 20 more cases for the small network by assigning different random ϵ_1, ϵ_2 values. The results further confirm these two theorems about QSOD solution sparsity. Detailed results are omitted here for brevity.

Table 6.5: Results for QSOD solution sparsity (small network)

QSOD scenario	$\lvert W_0 \cup W_Z\rvert$ or $\lvert W_\epsilon^* \cup W_Z^*\rvert$	$\lvert A_Z\rvert$ or $\lvert A_Z^* \setminus A_0^*\rvert$	$\lvert W\rvert$
Fixed-mapping case	13	7	20
Bi-level case	18	11	20

6.2 Downtown Seattle Network

The larger size road network studied here is the downtown Seattle network (depicted in Figure 6.2). This network consists of 927 directed links, 413 nodes, and 10,948 OD pairs, covering the downtown area of Seattle in the State of Washington of the U.S. The link cost function follows the same formula as in (6.1), and the free flow time of a link is calculated by the link length and speed limit. It may not be considered as a "large" network in practice, which is however used here to test the proposed QSOD models on a more general and larger network (compared with the five-node network in the previous subsection). For this, the OD demand matrix provided by PSRC is modified so that it can mimic the OD quasi-sparsity property of the entire King County region. The modified OD demand matrix satisfies the quasi-sparsity feature defined in Section 2 ($\epsilon_0 = 2, \alpha = 85\%, \beta = 15\%$): a large portion

(87.9%) of OD pairs own small demands ($\epsilon_0 = 2$) and they contribute to a relatively small portion (10.6%) of the total OD demands of the network. This modified OD matrix is treated as the “true” OD demand matrix for numerical experiments in this section.

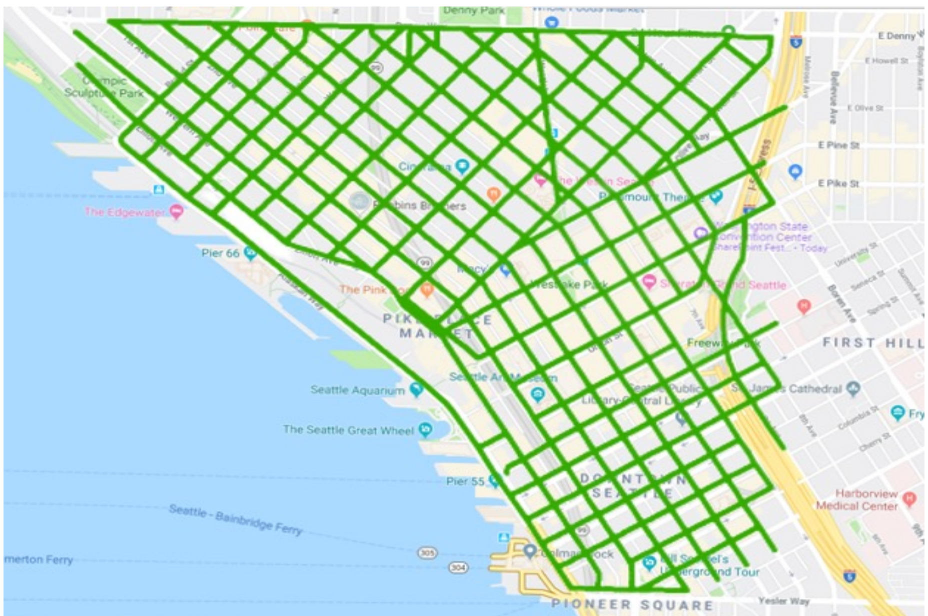


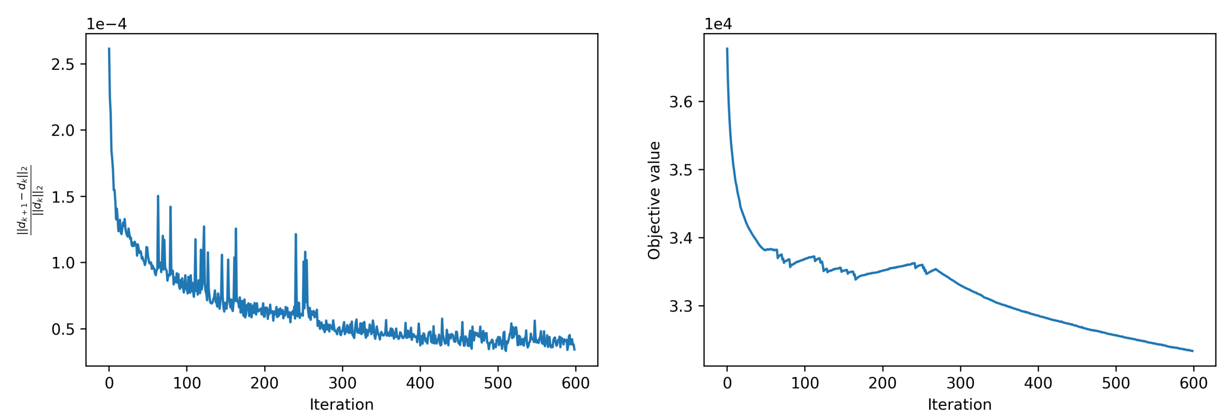
Figure 6.2: Downtown Seattle network

6.2.1 OD quasi-sparsity consistency

As this network example contains more than 10,000 OD pairs, it is not practical to show the estimated demand of each OD pair like the small network example in Table 6.3. Instead two more metrics (in addition to RMSE) are employed to evaluate the OD quasi-sparsity consistency for this network. The first is *F1-score*, see [88], which evaluates the performance of OD estimation models in maintaining the quasi-sparsity property of the true OD demand matrix. The second is the *accuracy* that describes the ratio of OD pairs in the estimated OD matrix that can keep the same significant/insignificant property as in the true OD matrix.

Detailed definitions of the two metrics are provided in Appendix C. For either of them, the performance of the model becomes better as the metric is getting closer to 1.

Similar to the small network example, the fixed-mapping QSOD model for this larger network can still be solved as an LP, after the mapping matrix $\hat{Q}$ is constructed by solving the UE problem. On the other hand, solving the bi-level QSOD model is quite challenging because of the large dimension of the problem. The downtown Seattle network actually presents a large-scale problem for bi-level models, as each iteration of the Prox-linear algorithm (i.e., Algorithm 1) involves almost 35,000 variables and 12,000 constraints. Algorithm 1 can only solve the bi-level QSOD model to certain level of accuracy. Figure 6.3 illustrates the convergence performance of the algorithm. Figure 6.3a shows the relative difference (in terms of L_2-norm) between the estimated OD demands for two consecutive iterations. Figure 6.3b demonstrates the trend of QSOD model objective value vs. iterations. Both figures indicate that this algorithm starts to converge after 600 iterations. After running 600 iterations of the prox-linear algorithm, we obtain an approximate solution of the bi-level QSOD model on the downtown Seattle network. The solution keeps 10,754 OD pairs insignificant/significant as they do in the prior ODs, which yields an accuracy of 98.23%.



(a) Relative L_2-norm of $(d_{k+1} - d_k)$ vs. iterations

(b) Objective value vs. iterations

Figure 6.3: Convergence pattern for solving bi-level QSOD model (large network)

We next analyze how the QSOD model retains the OD quasi-sparsity property of the estimated OD demand matrix (compared with the true/prior OD demand matrix) with various levels of errors introduced to construct the prior OD matrix (the errors in observed link flows ϵ_2 are fixed as 0.05). Table 6.6 indicates that as more errors introduced, the estimated OD matrix will have more OD pairs losing their insignificance/significance property (i.e. becoming significant from insignificant, or vice versa) as indicated by the decreasing accuracy and F1-score values.

When $\epsilon_1 = 0.3$, among the 10,948 total OD pairs, 10,635 OD pairs from fixed-mapping QSOD model and 10,269 OD pairs from bi-level QSOD model keep significant/insignificant as they do in the true OD matrix. This implies the potential of QSOD to keep the quasi-sparsity property of OD demand estimation. Besides, conditions are also checked to validate Theorem 4.1 and Theorem 5.2. Results show that neither $r_2 \leq r_1$ nor $\hat{r}_2 \leq r_1$ holds. Admittedly, the OD quasi-sparsity consistency conditions do not hold exactly (i.e., F-1 score is less than 1) for this larger network. The high values of F1-score and accuracy however do imply that the quasi-sparsity feature of the OD demand matrix is still largely maintained. This confirms the theoretical findings in Theorems 4.2 and 5.3, and the discussions therein.

Table 6.6: Results for OD quasi-sparsity consistency (large network example)

		Fixed-mapping			Bi-level		
Trial	Error (ϵ_1)	RMSE	Accuracy	F1-score	RMSE	Accuracy	F1-score
1	0.1	3.9438	0.9783	0.9875	1.1793	0.9436	0.9669
2	0.2	6.0051	0.9737	0.9849	1.2520	0.9413	0.9656
3	0.3	7.8925	0.9714	0.9836	1.3045	0.9380	0.9636

6.2.2 QSOD solution sparsity

The results of QSOD solution sparsity for all the trials in Table 6.6 are investigated. To illustrate, Table 6.7 shows the results for Trial 1. For the fixed-mapping case, it can be seen that Theorem 4.2 has been validated, as $|W_0 \cup W_Z| + |A_Z| \geq |W|$. In addition, for this larger size example, we notice that $|W_0 \cup W_Z| \approx |W|$ since $|W| >> |A|$, which is also consistent with the discussion after Theorem 4.2. For the bi-level case, as noted earlier, the obtained solution is just an approximate solution, which still shows the trend in obtaining a number of $|W_\epsilon^* \cup W_Z^*|$ fairly close (about 70%) to $|W|$.

Table 6.7: Results for QSOD solution sparsity (large network example)

QSOD scenario	$\lvert W_0 \cup W_Z\rvert$ or $\lvert W_\epsilon^* \cup W_Z^*\rvert$	$\lvert A_Z\rvert$ or $\lvert A_Z^* \setminus A_0^*\rvert$	$\lvert W\rvert$
Fixed-mapping case	10550	443	10948
Bi-level case	7596	291	10948

While considering Corollary 4.1, the inequality (4.14) ($|W_0| \geq |W_0^d| - |A_Z|$) can be easily validated, as in the prior OD demands $|W_0^d| = 0$, so the left hand side is always not less than the right hand side. We also manually let some insignificant OD demands be zero, or even all insignificant OD demands as zero (i.e., the OD demand matrix becomes sparse). Results validate Corollary 4.1 as well. For example, when 3000 insignificant OD demands in prior ODs are set as zero, the number of OD pairs with zero demand in the estimated OD demand matrix is 3064. We then have $3064 \geq 3000 - 919$. This indicates the number of zeros in the estimated OD demand matrix is always lower-bounded by the number of zeros in the prior OD demand matrix negating the number of links with flows equal to prior values.

6.3 Comparisons with Existing OD Estimation Methods

To further test the performance of the proposed QSOD models, we compare the model results with those of two existing OD estimation methods: the GLS model in [120] and the SOD estimation model in [77].

Both GLS and SOD involve some weighing matrices in the objective functions. For example, the GLS model in [120] has the form as:

$$\begin{aligned} \min_{d\in\mathbb{R}^{|W|}, v\in\mathbb{R}^{|A|}} \quad & f(d,v) = (d-d^0)^T U^{-1}(d-d^0) + (v-v^0)^T V^{-1}(v-v^0) \\ \text{s.t.} \quad & (d,v) \in F = \left\{(d,v) \;\middle|\; \Psi(d) = v,\ d \geq 0, v \geq 0\right\} \end{aligned} \tag{6.4}$$

where U and V stand for the variance-covariance matrix of OD demand errors and link flow errors, respectively. Usually when assuming no correlations between the errors of OD demands of any two OD pairs or link flows, the two matrices become diagonal matrices. Notice that in practice, neither U nor V could be obtained or estimated easily. Here to make it comparable with the QSOD model, we first assume them to be identity matrices with proper dimension, which reduces to an Ordinary Least Squared (OLS) model, then demonstrate the results when proper weighing matrices are implemented (GLS). In addition, as the results for SOD (with proper hyper-parameter λ for the L_1-regularizer) are very close to those for GLS (the differences are within 0.01% for all metrics for the testing cases we have conducted), we omit the results of SOD for brevity. The small difference between the two can be intuitively understood since GLS is a special case of SOD when the coefficient λ for the L_1-regularizer of OD demands in the objective function is zero [77].

6.3.1 Result comparison between QSOD and OLS/GLS.

Table 6.8 compares the estimation results on the small network for three models: QSOD, OLS (unweighted GLS) and GLS. Using $\epsilon_0 = 5$, the "ground truth" OD demands ($\bar{d}$) are added 50% errors to generate the prior OD demands (d^0). In order to make sure prior ODs share the same quasi-sparsity property with true ODs, the demands of those insignificant OD pairs in d^0 are capped at 5.

For the OLS model, the weighing matrix U and V in (6.4) are assumed as diagonal matrices with all ones on their diagonals. Although such treatment imposes strong assumptions on the models where errors of different OD demands/link flows are uncorrelated and equal, we test it to help understand how L_1-norm (used in QSOD) and L_2-norm (used in OLS and GLS) work differently in estimating OD demands. For the GLS model, we follow a similar weighing method in [120], but use the prior OD demands (and observed link flows) rather than "ground truth" OD demands (and true link flows) to calculate the diagonal of matrix of U (and V).

In the fixed-mapping case, all three models maintain the OD quasi-sparsity, and QSOD and GLS result in much better RMSE than OLS. In particular, since the prior demands of significant OD pairs (e.g., 2-4, 2-5, and 3-5) deviate a lot from their true values, the OLS model sacrifices the demands of insignificant pairs to match the prior demands of significant ones, leading to a larger RMSE.

In the bi-level case, the RMSE for the three models are comparable but OLS and GLS cannot maintain the OD quasi-sparsity consistency as QSOD. For example, in the GLS model, many insignificant OD demands are zeros and three of them (i.e. the demands for OD pairs 3-1, 3-4, and 4-5) become significant while QSOD can keep the insignificance/significance for all OD pairs. This means QSOD models guarantee the estimated OD demand has at least the same magnitude as its true/prior value while keeping the estimation error (RMSE) relatively small. The L_2-norm based models (e.g. GLS), although with slightly lower RMSE, may greatly change the magnitude of the demands of some OD pairs (e.g. the demand for OD pair 4-5 changes from 5 to 207.99 in the bi-level GLS model), which may not be desirable. Also noteworthy is that the computational time of the bi-level QSOD model (on a server with 8-core CPU and 32 GB memory) is much smaller than that of the bi-level GLS/OLS model, implying that solving the bi-level QSOD model is computationally less demanding. We tested 20 more cases and similar results can be found. Details are omitted here.

The use of weighing matrices in GLS is intuitive. According to [120], in the GLS model, the weights for OD demand deviation and link flow deviation should be $\frac{1}{(d^0 * \epsilon_1)^2}$ and $\frac{1}{(v^0 * \epsilon_2)^2}$ respectively. That is, once certain OD demand (or link flow) has relatively larger values or errors, the weight ($\frac{1}{(d^0 * \epsilon_1)^2}$) makes sure the corresponding deviation has a smaller weight in the objective function. When minimizing the objective, such OD demand will be "forced" less to be close to the prior data, while those with smaller values or errors (i.e., with larger weights) will be "forced" more to be close to the prior data. This is a standard way to improve the robustness of the estimation model (i.e., less sensitive to outliers in prior data). We have observed that GLS, by adding proper weights, can help produce better estimation results than OLS. For the QSOD models, they can produce good OD estimation results (comparable RSME and much better F1-score) even without any weighing terms. Also noteworthy is that with proper weights, the fixed-mapping GLS has similar performance with fixed-mapping QSOD in terms of RMSE and F1-score. One may explain this similarity via the closeness of the solutions of the two models. More discussions can be found in Appendix D.

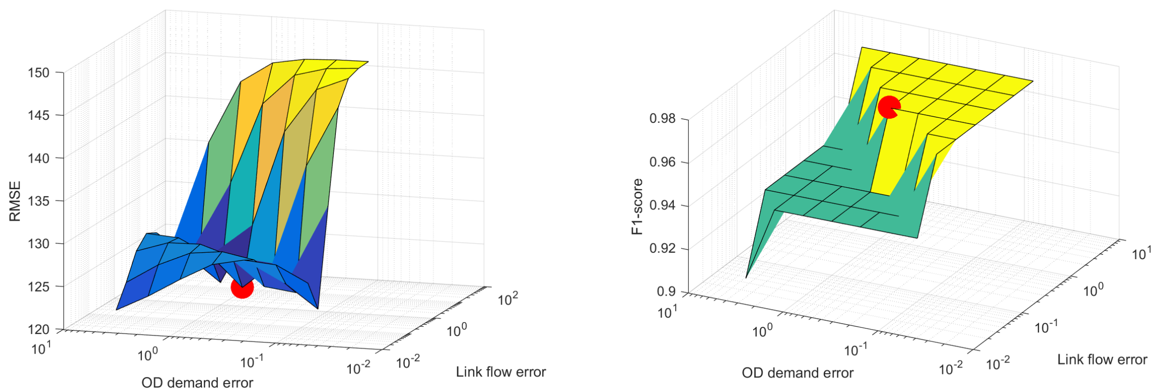
We have also conducted numerical tests using the downtown Seattle network, which are only done for the fixed-mapping case, as currently the bi-level GLS model cannot be solved for this network due to its large dimension. This further illustrates that computationally the bi-level QSOD model is much easier to solve than the bi-level GLS model. The results for the downtown Seattle network are quite similar to those for the small network. For example, when 30% errors are introduced into the prior OD demands, the results produced by QSOD, OLS, and GLS have the RMSE of 5.594, 8.328, and 2.140, respectively. The F1-scores of QSOD and GLS are very close to each other (QSOD: 0.9861, GLS: 0.998), and the F1-score of OLS (0.8525) is much lower than the other two. The dramatic difference of the F1-scores between GLS and OLS indicates that setting "proper weights" for GLS is critical. However this may not be done easily in practice; see discussions in the next subsection.

6.3.2 Discussions

Based on the results shown above, the proposed QSOD models can produce comparatively small RMSEs as GLS or SOD, but have the similar (fixed-mapping QSOD) or much better (bi-level QSOD) ability in keeping OD quasi-sparsity consistency. Besides, the QSOD models also share other advantages over GLS (or SOD), as discussed below.

First, the QSOD models do not require proper weights in the objective function. Setting up the proper weighing matrices U and V in GLS not only requires extra work, which also may not be readily obtained or estimated in practice. On the other hand, if the weighing matrices are not set properly, the estimation result of GLS may deteriorate substantially, as shown in the last subsection when compared with the results of OLS. Figure 6.4a and Figure 6.4b further demonstrate how the error terms in weighing matrices affect the OD estimation results of the GLS model. In this example, at most 50% and 40% errors were added to generate prior OD demands and observed link flows respectively, i.e., $\epsilon_1 = 0.5, \epsilon_2 = 0.4$. Then in the GLS model, the weights for OD demand deviation and link flow deviation should be $\frac{1}{(d^0 * 0.5)^2}$ and $\frac{1}{(v^0 * 0.4)^2}$ respectively. However, in practice, the error terms (i.e., 0.5 and 0.4) cannot be known in advance. Here we test when the errors are not properly set (i.e. ϵ_1 in $\frac{1}{(d^0 * \epsilon_1)^2}$ deviates from 0.5, ϵ_2 in $\frac{1}{(v^0 * \epsilon_2)^2}$ deviates from 0.4), how the model performance changes. Figure 6.4a (RMSE surface plot) indicates that the weighing scheme based on the true error terms (marked as the red point symbol) can lead to OD estimates with almost the smallest OD demand or link flow errors. However, deviations in the error terms can result in much worse results (i.e., much larger RMSE). Figure 6.4b (F1-score surface plot) shows that improper weights can easily degrade the OD quasi-sparsity consistency as well (F1-score decreases from 0.97 to 0.90). As the error terms (ϵ_1, ϵ_2) cannot be easily obtained in advance, how to properly set the weighing matrices in GLS remains a challenge. The QSOD model, however, does not require the use of such weighing matrices.

Second, the QSOD model is more robust to outliers in prior data and works well even the prior OD d^0 is not representative. As pointed out in the literature [80, 42, 108, 96], the L_2-norm based regression is much more sensitive to outliers than the L_1-norm regression. Such weakness of the GLS model also shows in the OD estimation problem. Table 6.9 gives an example where outliers exist in prior OD demand. In this example, at most 50% errors were added to generate prior OD demands, except OD pair 2-4, 2-5, and 3-5 where the priors are only 10% of their true values (i.e., they are the "outliers"). This treatment tries to simulate the case where outliers exist in prior information. We can see that QSOD outperforms GLS in this case. In the bi-level case, in particular, GLS fails to keep the OD quasi-sparsity consistency (with the F_1-score only about 0.7), and its estimated OD matrix



(a) RMSE

(b) F1-score

Figure 6.4: RMSE and F1-score of the GLS model vs. errors used in the weighting matrices

is very likely to deviate from the true OD matrix due to the outliers in prior OD demands. On the other hand, QSOD shows its robustness to the outliers in prior data. We also test when the prior OD is not representative, how the estimated OD demands from QSOD and GLS perform. We randomly set eight (out of 17) insignificant OD pairs as significant for prior ODs by amplifying their demands (thus the prior OD is not representative of the true OD anymore), and repeat this test for 20 times. Table 6.10 indicates that when prior ODs are not representative, the QSOD models can mostly show similar or better performance in reducing RMSE and keeping OD quasi-sparsity consistency than GLS. In particular, the QSOD model outperforms GLS in 16 (out of 20) trials in terms of F1 score, demonstrating its capability in keeping quasi-sparsity consistency, even when the prior OD demand is not representative.

Third, solving QSOD models is computationally less demanding than solving the GLS and SOD models, especially for the bi-level case. The fixed-mapping QSOD (4.3) can be converted into an LP and readily solved by standard LP solution algorithms such as the Simplex method. However, the fixed-mapping GLS (or SOD) model is a quadratic program, which in theory is harder to solve than an LP. More noticeable is the advantage of the bi-level QSOD model in computation efficiency. Such a model can be converted into an LP with linear equilibrium constraints, which is much less complex than the bi-level GLS model. The computational time shown in Table 6.8 confirms that the bi-level QSOD model requires much less time (0.673 vs. 1.522 seconds) to solve compared with the bi-level GLS model for the small network. For the larger downtown Seattle network, the bi-level GLS model cannot run properly due to memory and other computational issues, while the bi-level QSOD model can produce (approximate) optimal solutions.

Fourth, compared with the SOD model, QSOD models are hyperparameter-free, which means they do not require a hyperparameter tuning step to obtain a better solution. When conducting the numerical experiments, implementing the hyper-parameter tuning for the SOD model may require more efforts compared with implementing the QSOD model. Such a tuning process is not cost-effective either in real applications, since each time a new network (or any new data input to an existing network) is used, a new round of hyperparameter tuning will be needed.

Lastly, we briefly comment on the challenges of applying the proposed QSOD models to practical OD estimation applications. These challenges may also apply to other types of optimization-based OD estimation models, such as GLS models and SOD models (prior OD may not be used). For the fixed-mapping QSOD model, $\hat{Q}$ needs to be computed using true OD demand that is unavailable for practical applications. For the bi-level QSOD model, the bi-level model is generally hard to solve on large scale problems since the UE mapping itself changes during the optimization process. Both issues may be resolved by using a reasonably estimated prior OD d^0: an estimate to $\hat{Q}$ or the UE mapping can be produced(denoted as Q) using d^0, solve the fixed-mapping or bi-level QSOD model using Q to obtain a solution d, use d to generate an updated estimate of Q, and repeat the process of updating d and Q. This is practically doable as such prior OD is often available for practical applications (i.e., by planning agencies). However, it often difficult to prove that such a process can lead to convergent results, leaving it largely a heuristic method. The practicability of using the prior OD d^0 also merits further discussions. It may be considered as both a limitation and an advantage: a limitation because we need d^0 to be reasonably estimated (ideally representative for QSOD models or close to the true OD for GLS models) and an advantage because there are indeed practically estimated d^0 by planning agencies which should be leveraged. In fact, most optimization-based OD estimation models (the only exception we are aware of is the CODE model in [97]) apply d^0 as the input, aiming to improve it using additional data sources. Therefore, the use of d^0 in QSOD models (and other models too) is not impractical, as long as it is (ideally) representative of the true OD demand. In case a reasonably estimated d^0 is not available (ensuring representativeness of d^0 is not trivial as discussed in the next section), one may exercise the similar heuristic idea shown above for $\hat{Q}$, i.e., starting with some best initial d^0 and iteratively solving QSOD to obtain improved estimates.

Table 6.8: Comparison between OD true values, priors, and estimates for three models

			Fixed-mapping			Bi-level		
O-D	$\bar{d}$	d^0	QSOD	OLS	GLS	QSOD	OLS	GLS
1-2	1	1.5	1.5	1.13	1.36	1.5	0.861	0
1-3	4	5	5	4.623	3.67	4.777	2.27	0
1-4	2	3	3	0	3.10	3	0	0
1-5	2	3	1.86	0	2.96	3	0	0
2-1	1	1.5	0	0	1.49	1.5	0	0
2-3	4	5	5	0	4.44	0.00001	0	0
2-4	1500	2250	1503.80	1747.43	1501.82	2250.00	1856.72	1713.69
2-5	1600	2400	1596.81	2004.91	1596.33	1810.80	2049.55	2086.07
3-1	3	4.5	4.5	0	4.58	4.5	0	16.02
3-2	4	5	5	4.503	4.01	5	4.503	5
3-4	1	1.5	1.903	0	1.37	1.5	0	7.93
3-5	1600	2400	1603.75	1836.94	1603.26	1151.28	2069.57	1001.68
4-1	2	3	3	2.69	1.85	3	4.01	3.33
4-2	1	1.5	0.50	0.82	1.07	1.5	7.42	2.74
4-3	1	1.5	1.5	0.81	1.09	1.5	3.99	1.5
4-5	5	5	4.98	4.99	4.98	5	12.71	207.99
5-1	4	5	5	4.32	4.20	0.285	0	0
5-2	3	4.5	3.02	3.45	2.59	0.00001	0	0
5-3	2	3	1	1.94	1.94	0.00001	0	0
5-4	2	3	3	2.51	2.51	2	0	0
RMSE			1.605	118.606	1.287	200.926	165.835	184.582
F1-score			1	1	1	1	0.889	0.757
Computational time (seconds)			0.1926	0.1777	0.1757	0.689	4.803	1.522

Table 6.9: Comparison between OD true values, priors, and estimates for QSOD and GLS (with outliers)

			Fixed-mapping		Bi-level	
O-D	$\bar{d}$	d^0	QSOD	GLS	QSOD	GLS
1-2	1	0.59	0.59	0.57	0.59	2.00
1-3	4	2.51	4.70	4.46	4	6.00
1-4	2	1.17	1.17	1.22	2	0
1-5	2	1.74	1.74	1.81	2	0
2-1	1	1.40	1.40	1.43	1	3.27
2-3	4	3.81	3.81	4.52	3.44	0
2-4	1500	150	1506.68	1499.424	1878.32	1783.23
2-5	1600	160	1596.81	1584.94	1600	2071.29
3-1	3	4.03	4.03	4.558	3	18.77
3-2	4	2.12	4.01	4.01	2.16	4.01
3-4	1	1.05	1.05	1.19	1	4.39
3-5	1600	160	1603.93	1595.99	1600	1002.40
4-1	2	2.72	1.93	2.39	2.60	0
4-2	1	1.08	1.08	0.95	1.08	3.01
4-3	1	0.58	0.58	0.68	0.58	0
4-5	5	2.70	4.98	4.98	5	118.73
5-1	4	3.38	4.37	3.65	2.58	9.02
5-2	3	4.39	3.35	3.51	3	0
5-3	2	1.30	1.30	1.87	1.30	0
5-4	2	1.13	2.01	2.01	1.13	2.01
RMSE			1.921	3.3521	84.596	183.403
F1-score			1	1	1	0.703

Table 6.10: Comparison between QSOD and GLS when prior ODs are not representative (Fixed-mapping)

Model	QSOD		GLS	
Trial	RMSE	F1-score	RMSE	F1-score
1	19.728	0.828	34.537	0.786
2	12.309	0.867	40.282	0.828
3	11.946	0.867	16.825	0.867
4	16.988	0.867	38.419	0.938
5	15.697	0.867	14.431	0.903
6	19.064	0.903	16.857	0.828
7	15.681	0.903	14.486	0.828
8	12.297	0.903	40.282	0.867
9	16.473	0.903	39.666	0.828
10	12.324	0.828	40.283	0.786
11	14.841	0.903	18.513	0.903
12	16.504	0.903	16.453	0.903
13	14.802	0.903	30.253	0.903
14	19.123	0.903	19.005	0.867
15	17.121	0.867	15.295	0.903
16	17.025	0.867	38.822	0.828
17	16.217	0.903	15.549	0.903
18	16.345	0.903	13.002	0.903
19	15.898	0.903	32.273	0.828
20	14.093	0.903	17.543	0.938
Mean	15.724	0.885	25.639	0.867

Chapter 7

DISTRIBUTIONALLY ROBUST QSOD ESTIMATION MODEL

Having explored the deterministic QSOD models in Chapter 4-6, in this chapter, we investigate the QSOD framework from a stochastic perspective. The focus of this study is the (static) OD demand estimation problem, i.e. given prior OD demands and observed data within a period, e.g. link traffic flows, how to estimate a more accurate static representation of OD demands, including an "averaged" demand and the daily demand that corresponds to the daily observed data. For OD estimation, stochasticity may come from three sources. First is the stochasticity of the physical system, i.e., the transportation network, which is naturally stochastic due to various randomness related to physical systems (such as incidents) and human behaviors (such as choices of routes, modes, etc.). Second, because of the stochastic nature of the physical system, the collected and available data (e.g., day by day link flows) are also random to reflect the stochasticity of the physical system. The third is the stochastic structure of the modeling methods that can capture the stocahsticity of the physical systems and the available data (e.g., the stochastic optimization model in the literature or the DRO model employed in this chapter). In this sense, SO and DRO are two alternative modeling methods to model and estimate stochastic OD demands. They both apply to similar but distinct scenarios, depending on how accurate knowledge we have about the distribution information of the input data or OD demands. Next, we start with some DRO basics to build the DR-QSOD model.

7.1 Model Formulation

Differing from SO models where certain distributions are assumed for OD estimates or/and observed data, DRO framework allows the distributions of OD estimates or/and observed data to run within an ambiguity set. Under the DRO framework, we introduce two sets of decision variables d and t in this thesis to stand for the OD demands in different modeling stages. Variable d ("here-and-now" decision) represents the "nominal" (or "averaged") OD demands, which is a "representative" OD demand taking into account the stochasticity and robustness of observed data at the same time. It reflects the general patterns of the OD demands of the entire network. On the other hand, variable t ("wait-and-see" decision), which will be solved after obtaining the value of d, stands for the OD demand corresponding to the observed link flow data on each day. The use of a two-stage model rather than a single-stage model (2.5) in this case is natural, as we not only expect an overall OD demand to account for the prior OD demands and the data distributional uncertainty, but also give an OD estimate corresponding to the traffic link flows on each individual day. In addition, since the DRO model does not require a probability assumption on the observed data, it is generally easier to use in practice. We give the notation of the DR-QSOD model (mainly for Chapter 7 and Chapter 8) in Table 7.1.

Table 7.1: List of notation (DR-QSOD model)

Notation	**Meaning**
d	"here-and-now" decision for the OD demands
t	"wait-and-see" decision for the OD demands
λ	weighting parameter in DR-QSOD
z_1, z_2, z_3	auxiliary variables
ξ	daily network-level traffic link flow
A	dimension indices set for daily network-level traffic link flow
W	dimension indices set for OD demands
d^0	prior OD demand

Continued on next page

Table 7.1: List of notation (DR-QSOD model) (Continued)

Notation	**Meaning**
$\mathbb{P}$	probability distribution of ξ
$\mathcal{D}$	ambiguity set for $\mathbb{P}$
$\hat{Q}$	fixed-mapping matrix between OD demands and link flows
$\bar{\mu}$	mean value of observed link flow
$\bar{S}, A^*, B, C, E, F, G$	pre-specified coefficient matrices
$\bar{a}, b, c, e, f, g$	pre-specified coefficient vectors
Ω	support of ξ
$M^+(\Omega)$	set of all probability distributions on a σ-algebra of Ω
ξ_{min}	predefined lower bound on ξ
ξ_{max}	predefined upper bound on ξ
y	first-stage decision variable of two-stage DRO
x	second-stage decision variable of two-stage DRO
$Q(y, \xi)$	second-stage objective
γ, β	dual variables in dual formulation of two-stage DRO
ζ, η, δ	dual variables in dual formulation of $Q(y, \xi)$
$\bar{v}$	true link flows
$\bar{d}$	"true" OD demands
ϵ_0	threshold for insignificant OD demand
ϵ_1	relative error in prior OD demands
ϵ_2	relative error in observed link flows

Now we formally establish the two-stage distributionally robust quasi-sparsity OD estimation model as follow:

$$\begin{aligned} \min_{d} \quad & \frac{1}{\lambda}\|d-d^0\|_1 + \sup_{\mathbb{P}\in\mathcal{D}} E_{\mathbb{P}}[Q(d,\xi)] \\ s.t. \quad & d \geq 0, \end{aligned} \tag{7.1}$$

where $d^0 \in \mathbb{R}^{|W|}$ represents the prior OD demands, which is usually provided by local planning agencies; $\xi \in \mathbb{R}^{|A|}$ stands for the daily network-level traffic link flow, which can be collected by widely-deployed detectors; and $Q(d, \xi)$ equals to the optimal value of the following problem:

$$\begin{aligned} \min_{t} \quad & \|t-d\|_1 + \|\hat{Q}\cdot t - \xi\|_1 \\ s.t. \quad & t \geq 0, \end{aligned} \tag{7.2}$$

where $\hat{Q} \in \mathbb{R}^{|A|\times|W|}$ represents the fixed-mapping matrix between OD demands and link flows. Here $Q(d, \xi)$ can be regarded as the optimal value of sum of the L_1-deviations between estimates and their prior or observed values, given d and ξ. Further, if $d = d^0$, then $Q(d, \xi)$ will be the same as the deterministic fixed-mapping QSOD model in Chapter 4. λ in (7.1) is a parameter that needs to be tuned, which can be interpreted as how reliable the prior OD demand d^0 is. If λ is small, the term $\|d - d^0\|_1$ will have a larger weight, which means the model will penalize more on this deviation so that the first-stage decision variable d will be forced more to be close to d^0.

Since there often exist daily variations within traffic flow, it is generally hard to obtain an accurate distribution of the daily network-level traffic link flow ξ. Therefore, in this dissertation, we treat ξ as a random parameter with distributional uncertainty. We let $\mathbb{P}$ be the probability distribution of the link flow ξ, and $\mathcal{D}$ be the ambiguity set of $\mathbb{P}$ defined in (7.3).

$$\mathcal{D} := \{\mathbb{P} \in M^+(\Omega) : E_{\mathbb{P}}[\bar{S}\xi] \leq \bar{\mu}\}, \tag{7.3}$$

where $\bar{\mu}$ stands for the mean value of observed link flows, $\bar{S}$ represents the pre-specified coefficient matrix ($\bar{S}$ is set as an identity matrix in this study). That is, the ambiguity set is defined based on the first moment of the distribution. $M^+(\Omega)$ here stands for the set of all probability distributions on a σ-algebra of Ω, and Ω is the support of ξ defined below:

$$\Omega := \{\xi \in \mathbb{R}^{|A|} : \xi^{\alpha}_{min} \leq \xi^{\alpha} \leq \xi^{a}_{max}, \forall \alpha \in A\}, \tag{7.4}$$

where ξ_{min} and ξ_{max} are predefined lower and upper bounds of ξ.

Notice that the ambiguity set $\mathcal{D}$ basically defines a family of probability $\mathbb{P}$, and all possible probability distributions of ξ are restricted by the first-moment constraint $E_{\mathbb{P}}[\bar{S}\xi] \leq \bar{\mu}$. Certainly there might exist other ways to define the ambiguity, which can be explored in future research. The term "ambiguity" further indicates that we only know part, rather than all, of probability information regarding ξ, therefore the modeling process of DRO will be less restrict than the typical stochastic optimization.

Similar to the deterministic QSOD model in Chapter 4, we use the L_1-norm in the objectives of (7.1) and (7.2). The use of L_1-norm does not only help keep the quasi-sparsity property of OD demand matrices, but also make the model easier to solve. Denote α and w as any link and OD pair in the set of A and W, respectively. To make the model structure more clear, we introduce auxiliary variables $z_1 \in \mathbb{R}^{|W|}$, $z_2 \in \mathbb{R}^{|W|}$, and $z_3 \in \mathbb{R}^{|A|}$ to replace the L_1-norm in (7.1) and (7.2), and transform the two-stage DR-QSOD model into the following form:

$$\begin{aligned} \min_{d,z_1} \quad & \frac{1}{\lambda} \sum_{w \in W} z_1^w + \sup_{\mathbb{P} \in \mathcal{D}} E_{\mathbb{P}}[Q(d, z_1, \xi)] \\ s.t. \quad & d^w \geq 0, \quad \forall w \in W, \\ & z_1^w \geq d^w - (d^0)^w, \quad \forall w \in W, \\ & z_1^w \geq -d^w + (d^0)^w, \quad \forall w \in W, \end{aligned} \tag{7.5}$$

where

$$Q(d, z_1, \xi) = \min_{t,z_2,z_3} \quad 0 \cdot \sum_{w \in W} t^w + 1 \cdot \sum_{w \in W} z_2^w + 1 \cdot \sum_{\alpha \in A} z_3^\alpha$$
$$\begin{aligned} s.t. \quad & t^w \geq 0 \quad \forall w \in W, \\ & z_2^w \geq t^w - d^w, \quad \forall w \in W, \\ & z_2^w \geq -t^w + d^w, \quad \forall w \in W, \\ & z_3^\alpha \geq [\hat{Q}t]^\alpha - \xi^\alpha, \quad \forall \alpha \in A, \\ & z_3^\alpha \geq -[\hat{Q}t]^\alpha + \xi^\alpha, \quad \forall \alpha \in A. \end{aligned} \tag{7.6}$$

The two problems above can be further expressed in a compact form:

$$\begin{aligned} \min_y \quad & \bar{a}^T y + \sup_{\mathbb{P} \in \mathcal{D}} E_{\mathbb{P}}[Q(y, \xi)] \\ s.t. \quad & A^* y \leq b, \end{aligned} \tag{7.7}$$

where

$$\begin{aligned} Q(y, \xi) = & \min_x c^T x \\ s.t. \quad & Bx + Cy \geq e, \\ & Ex + F\xi \geq f, \\ & Gx \geq g, \end{aligned} \tag{7.8}$$

where $y = [d, z_1]^T$ represents the first-stage decision variables, $x = [t, z_2, z_3]^T$ represents the second-stage decision variables, $A^* = \begin{bmatrix} -\mathbf{I} & \mathbf{0} \\ \mathbf{I} & -\mathbf{I} \\ -\mathbf{I} & -\mathbf{I} \end{bmatrix}$, $B = \begin{bmatrix} -\mathbf{I} & \mathbf{I} & \mathbf{0} \\ \mathbf{I} & \mathbf{I} & \mathbf{0} \end{bmatrix}$, $C = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{I} & \mathbf{0} \end{bmatrix}$, $E = \begin{bmatrix} -\hat{Q} & \mathbf{0} & \mathbf{I} \\ \hat{Q} & \mathbf{0} & \mathbf{I} \end{bmatrix}$, $F = \begin{bmatrix} \mathbf{I} \\ -\mathbf{I} \end{bmatrix}$, $G = \begin{bmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} \end{bmatrix}$, $\bar{a} = [\mathbf{0} \quad \frac{1}{\lambda}\mathbf{1}]^T$, $b = [\mathbf{0} \quad d^0 \quad -d^0]^T$, $c = [\mathbf{0} \quad \mathbf{1} \quad \mathbf{1}]^T$, $e = \mathbf{0}$, $f = \mathbf{0}$, $g = \mathbf{0}$, where $\mathbf{I}$, $\mathbf{0}$, and $\mathbf{1}$ stand for the identity matrices, all-zero matrices/vectors, and all-one vectors with proper dimensions, respectively.

Note that problem (7.7) and (7.8) are in the classical form of a two-stage DRO model: the first-stage problem tries to minimize the sum of $\bar{a}^T y$ and the worst-case expectation of $Q(y, \xi)$

over all possible probability distributions $\mathbb{P}$; the second-stage problem aims to minimize $c^T x$ given y and the realization of ξ, i.e. ξ_i, the link flow for day i. The key difference between the two-stage model and single-stage DRO model (2.5) is that the second-stage problem can yield to an estimate corresponding to each realized input data ξ_i, which makes it possible to give a daily OD demand as the change of daily link flows.

Besides, there exist certain connections between the two-stage DR-QSOD model with the deterministic QSOD model and stochastic models. First, the two-stage DR-QSOD model is more general than the fixed-mapping QSOD model in Chapter 4. If we set a relatively small value for λ in (7.1), the first decision variable d will equal to its prior value d^0 because of the large weight for $\|d - d^0\|_1$. In this case, the second-stage problem is exactly the same as the fixed-mapping QSOD model: given a single point link flow data, the second-stage model output an OD demand estimate. Second, DR-QSOD model is statistically more flexible than the stochastic OD estimation models. Unlike stochastic models generally require strong assumptions on estimates or data, DR-QSOD model works well for the case where the stochastical information of data is not complete. For example, if there is only very limited of data or lots of required data is unobserved, DR-QSOD model will work better comparing with stochastic models, as an exact distribution information regarding the data is not required in the DR-QSOD framework.

7.2 Solution Algorithm

7.2.1 Tractable reformulation

Equation (7.7)-(7.8) discussed above present a clear two-stage DRO framework, which however is challenging to solve in practice. First of all, the probability $\mathbb{P}$ is not known or pre-defined, which makes the first-stage problem (7.7) not tractable. Second, the two-stage framework itself is generally more difficult to solve compared with the single-stage framework. In this section, we discuss the method to solve the above two-stage DR-QSOD model. First of all, we try to convert the two-stage DRO problem into a tractable min-max problem, as discussed in Theorem 7.1.

Theorem 7.1. *The two-stage DRO model* (7.7)-(7.8) *is equivalent to the following (tractable) min-max problem* (7.9)*.*

$$\begin{aligned} \min_{y,\tau\geq 0} \quad & \bar{a}^T y + \bar{\mu}^T \tau + \max_{\xi\in\Omega}(Q(y,\xi) - \tau^T \bar{S}\xi) \\ s.t. \quad & A^* y \leq b. \end{aligned} \tag{7.9}$$

Proof. Note that the second part of the objective of (7.7) requires to solve the worst-case expectation $\sup_{\mathbb{P}\in\mathcal{D}} E_{\mathbb{P}}[Q(y,\xi)]$ under the ambiguity set $\mathcal{D}$, and $Q(y,\xi)$ itself is an optimization problem. First we rewrite $\sup_{\mathbb{P}\in\mathcal{D}} E_{\mathbb{P}}[Q(y,\xi)]$ as the following optimization problem, according to the property and definition of probability distribution and expectation.

$$\begin{aligned} \sup_{\mathbb{P}\in\mathcal{D}} E_{\mathbb{P}}[Q(y,\xi)] = \max_{\mathbb{P}} & \int_{\Omega} Q(y,\xi)\mathbb{P}(d\xi) \\ s.t. \quad & \int_{\Omega} \bar{S}\xi\mathbb{P}(d\xi) \leq \bar{\mu} \\ & \int_{\Omega} \mathbb{P}(d\xi) = 1 \end{aligned} \tag{7.10}$$

Then by using the standard duality theory, we can write the dual problem of (7.10) as follow:

$$\begin{aligned} \min_{\tau\geq 0,\beta} \quad & \bar{\mu}^T \tau + \beta \\ s.t. \quad & \tau^T \bar{S}\xi + \beta \geq Q(y,\xi), \forall \xi \in \Omega \end{aligned} \tag{7.11}$$

where τ and β are dual variables associated with the two constraints in (7.10) respectively. Substituting $\sup_{\mathbb{P}\in\mathcal{D}} E_{\mathbb{P}}[Q(y,\xi)]$ with (7.11), we can rewrite (7.7) as follows:

$$\begin{aligned} \min_{y} \quad & \bar{a}^T y + \min_{\tau\geq 0,\beta}(\bar{\mu}^T \tau + \beta) \\ s.t. \quad & A^* y \leq b \\ & \tau^T \bar{S}\xi + \beta \geq Q(y,\xi), \forall \xi \in \Omega \end{aligned} \tag{7.12}$$

Now the objective of (7.12) is a min-min problem, so we write it as the following optimization problem,

$$\begin{aligned} \min_{y,\tau\geq 0,\beta} \quad & (\bar{a}^T y + \bar{\mu}^T \tau + \beta) \\ s.t. \quad & A^* y \leq b \\ & \tau^T \bar{S}\xi + \beta \geq Q(y,\xi), \forall \xi \in \Omega. \end{aligned} \tag{7.13}$$

As the constraint $\beta \geq Q(y,\xi) - \tau^T \bar{S}\xi$ must be satisfied for $\forall \xi \in \Omega$, we further write the above problem (7.13) as (7.9). Therefore, the two-stage model is equivalent to the min-max problem. □

7.2.2 Benders decomposition

In this section, we will apply Benders Decomposition (BD) [126] to solve the problem in (7.9). Firstly, we rewrite (7.9) as follow:

$$\begin{aligned} \min_{y,\tau\geq 0,\beta} \quad & (\bar{a}^T y + \bar{\mu}^T \tau + \beta) \\ s.t. \quad & A^* y \leq b \\ & \beta \geq Q(y,\xi) - \tau^T \bar{S}\xi, \forall \xi \in \Omega. \end{aligned} \tag{7.14}$$

Notice that as ξ in the second constraint of (7.14) is a continuous parameter and has to fall into the support of Ω, it is unwise or impossible to enumerate all values of ξ to construct the realized constraints. However we can employ BD method to solve the problem. Given a realization of ξ, say ξ', we call the constraint $\beta \geq Q(y,\xi') - \tau^T \bar{S}\xi'$ as a Benders cut pertaining to ξ'. The BD method will start to optimize the problem with some relaxed Benders cuts, and then iteratively add more Benders cuts till the stopping criterion is satisfied. We demonstrate the BD framework as follow:

1. Initialization: lower bound $LB := -\infty$, upper bound $UB := \infty$, tolerance ϵ, iteration limit L, set of Benders' cuts $CUT := \varnothing$.

2. For $l = 1, \cdots, L$, repeat the steps:

(a) Solve the master problem

$$\begin{aligned} \min_{y,\tau \geq 0, \beta} \quad & \bar{a}^T y + \bar{\mu}^T \tau + \beta \\ s.t. \quad & A^* y \leq b \end{aligned} \tag{7.15}$$

with the current set of Benders' cuts in CUT as additional constraints. Record optimal solutions (y^l, τ^l, β^l), and set LB equal to the optimal value.

(b) Solve the separation problem

$$\max_{\xi \in \Omega} \{Q(y^l, \xi) - (\tau^l)^T \bar{S} \xi\} \tag{7.16}$$

Record optimal solution ξ^l and the optimal value V^l, ans set UB equal to $LB - \beta^l + V^l$.

(c) If $|UB - LB|/LB < \epsilon$ or $\beta^l \geq V^l$, then RETURN and output y^l as an optimal solution; otherwise, go the the next step.

(d) Add a Benders' cut $\beta \geq Q(y, \xi^l) - \tau^T \bar{S} \xi^l$ into set CUT.

In Step 2(b), $Q(y^l, \xi)$ itself is a minimization linear program. In order to solve the max-min problem (7.16) we consider the dual problem of $Q(y, \xi)$ as follow:

$$\begin{aligned} Q(y, \xi) = \max_{\zeta, \eta, \delta} \quad & (e - Cy)^T \zeta + (f - F\xi)^T \eta + g^T \delta \\ s.t. \quad & B^T \zeta + E^T \eta + G^T \delta \leq c, \\ & \zeta, \eta, \delta \geq 0 \end{aligned} \tag{7.17}$$

where the dual variables ζ, η, and δ correspond to the three constraints in (7.8), respectively. As $Q(y, \xi)$ has been converted into its dual form in terms of a maximization problem, we can plug (7.17) into (7.16), and recast the separation problem as:

$$\begin{aligned} \max_{\xi, \zeta, \eta, \delta} \quad & (e - Cy^l)^T \zeta + (f - F\xi)^T \eta + g^T \delta - (\tau^l)^T \bar{S} \xi \\ s.t. \quad & B^T \zeta + E^T \eta + G^T \delta \leq c, \\ & \zeta, \eta, \delta \geq 0, \\ & \xi^\alpha_{min} \leq \xi^\alpha \leq \xi^\alpha_{max}, \forall \alpha \in A \end{aligned} \tag{7.18}$$

Solving the problem (7.18), we then record optimal solution $(\xi^l, \zeta^l, \eta^l, \delta^l)$ and the optimal value V^l, ans set UB equal to $LB - \beta^l + V^l$ (Step 2(b)). Then we add a Benders' cut $\beta \geq (-C^T\zeta^l)^T y + e^T\zeta^l + (f - F\xi^l)^T\eta^l + g^T\delta^l - (\bar{S}\xi^l)^T\tau$ into set CUT (Step 2(d)).

In summary, using the BD framework to solve the DR-QSOD model essentially requires to solve two problems iteratively. One is the master problem (7.15), which is a classic linear program. This problem can be solved by any LP solvers, e.g. simplex, dual-simplex, and interior-point method [87]. Another is the separation problem, which can be converted into the form of (7.18). Containing the bi-linear term $(F\xi)^T\eta$, problem (7.18) is a nonlinear program (NLP) with nonlinear objective and linear constraints. This problem can be usually solved by NLP solvers. In some cases, if the objective can be formulated into a quadratic function, it can be efficiently dealt with by quadratic program solvers. Regarding the convergence property, the BD method belongs to a classic L-shaped algorithm, and it has been proved that such BD algorithm finitely converges to an optimal solution when it exists or proves the infeasibility of problem. For a more detailed discussion of the BD algorithm convergence, one can refer to the study [14].

7.2.3 Incorporating Column-and-Constraint Generation

The existence of distributionally uncertain input data with limited probabilistic information brings great difficulties in solving the DR-QSOD model. In order to address this issue, the BD method discussed above iteratively adds dual cuts into the master problem (7.15), and finally reaches the optimal solution under the distributional uncertainty of input data ξ. However, the BD method with pure dual cuts has notorious weakness of slow convergence, as each iteration only adds one constraint into the master problem. Given this situation, we consider introducing the Column-and-Constraint Generation (CCG) method [124], and try to leverage the advantage of two methods (BD with dual cuts and CCG) to solve the model. As a cutting plane procedure, the CCG method defines the generated cutting planes by a set of created recourse decision variables in the forms of constraints of the recourse problem, which may help speed up the convergence process, compared with the BD method with the pure dual cuts. More often, both dual cuts and the primal cuts (from CCG process) are employed for solving DRO models. We refer reader to [124, 126] for more discussions of BD method with primal & dual cuts and the CCG method.

Chapter 8

NUMERICAL EXPERIMENTS OF DR-QSOD MODEL

In this section, we test the proposed DR-QSOD model using the well-known Sioux Falls network [79], and compare the performance of DR-QSOD with its deterministic counterpart, fixed-mapping QSOD model 4.3, and stochastic QSOD model.

The Sioux Falls network (as shown in Figure 8.1) contains 76 links and 24 nodes, contributing to 552 OD pairs ($|W| = 552$, $|L| = 76$). The network attributes and the "reference" OD demands are given in [106]. In order to mimic the quasi-sparsity property of large-scale OD demand matrices, we slightly modified the reference OD demands and constructed the so-called "true" OD demands for the network. Using $\epsilon_0 = 5$ as the threshold to define the insignificant OD demand, there are 330 insignificant OD pairs in the "true" OD demand matrix and those insignificant demands only contribute to 15% of total demands. Those properties satisfy the definition of quasi-sparse OD matrices in Section 3.1.

Denote $\bar{d}$ as the "true" OD demands and $d^0 = (1 + \Delta_1) \times \bar{d}$ as the prior OD demands, where $\Delta_1 \sim uniform(-\epsilon_1, \epsilon_1)$. Given $\bar{d}$ and link attributes of the network, we solve the user equilibrium problem and derive the true link flows, denoted as $\bar{v}$. Then the observed link flow ξ is generated as $\xi = (1 + \Delta_2) \times \bar{v}$, where $\Delta_2 \sim uniform(-\epsilon_2, \epsilon_2)$. The two parameters ϵ_1, ϵ_2 reflect how much relative errors are in the prior OD demands and observed link flows, respectively. In the following experiments, we assume $\epsilon_1 > \epsilon_2$, because in reality the errors for the link flows are relatively small as link flows are usually counted by traffic detectors, which are more reliable.

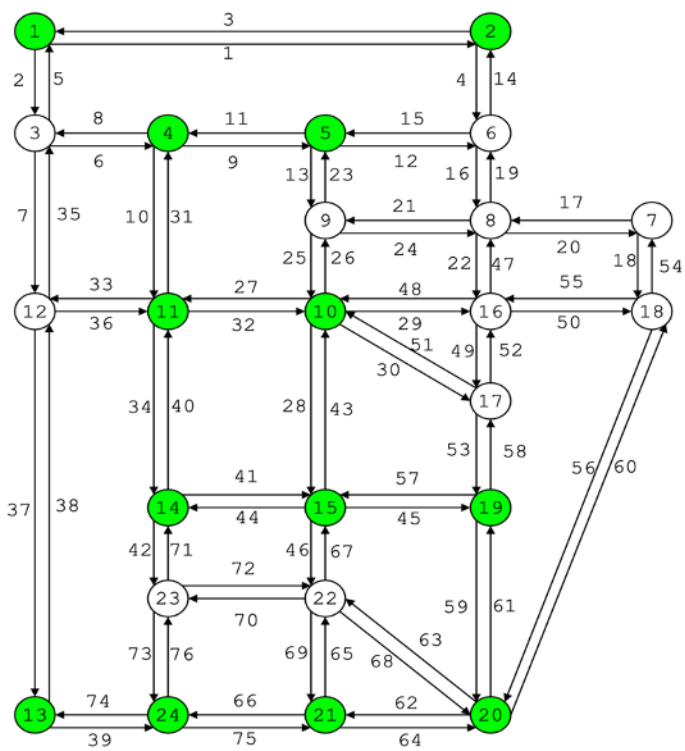


Figure 8.1: Sioux Falls network

8.1 Parameter Tuning

The first step to conduct the numerical test is to tune a proper value for the weighting parameter λ in (7.1). As previously stated, λ is a critical parameter in the two-stage DR-QSOD model, as it can be interpreted as the reliability of the prior OD demands d^0. If λ is small, DR-QSOD model will rely more on the prior OD demands therefore produce an estimate closer to d^0. However in real applications, it is hard or impossible to determine in advance what the best value of λ is. Given this situation, we test various values of λ in the DR-QSOD model and compare their OD estimation performance to determine the best choice. Table 8.1 demonstrates the results of DR-QSOD model performance under various values of λ with the setting of $\epsilon_1 = 0.25$ and $\epsilon_2 = 0.1$. As λ increases, the Root Mean Squared Errors (RMSE) of OD estimates shows a trend of decreasing.

Table 8.1: DR-QSOD model performance, running time vs. weighting parameter λ

λ	RMSE	Running time (s)
0.25	142.42	281.21
0.5	142.42	300.83
1	142.42	224.94
1.5	140.37	991.35
2	133.92	1575.77
4	129.47	2540.97
6	131.29	5636.05

When $\lambda \leq 1$, the RMSE does not change across different values of λ. When $\lambda > 1$, the RMSE of OD estimates decreases while the time required by the BD method to converge grows. For example, when $\lambda = 4$, the RMSE decreases by 9.1% but the converge time increases by 10 times, compared to the case of $\lambda = 1$. Based on the two factors (RMSE and running time), we select $\lambda = 2$ for the remaining numerical tests in order to obtain a good OD estimate within an acceptable time frame. It is worthwhile to note that the best value of λ may vary for different network examples. In real applications, one may need to tune this parameter properly to obtain a better first-stage OD estimate.

8.2 Algorithm Comparison

In this section, we compare the convergence performance of the BD method with pure dual cuts, and the BD-CCG combination method. Figure 8.2 show the relation between relative error $|(UB - LB)/LB|$ and the number of iterations. It can be seen that in the early stage, especially when iteration number is less than 20, the BD-CCG combination methods converge faster than the DB method with dual cuts, and the more CCG rounds we include the quicker the converge will be. From the perspective of running time required to converge, see Figure 8.3, the BD method with pure dual cuts shows fairly good performance compared with the BD-CCG combination methods. In the early stage, the BD dual cuts method requires much less running time per iteration, which is intuitive as each iteration only adds one additional constraint into the master problem. The BD-CCG combination method, on the other hand, needs great mount of time even in the early stage. The more rounds of CCG procedures included, the longer time it requires to converge. This is because in each iteration, the CCG procedure not only introduces a set of new constraints into the master problem (versus only one constraint like BD dual cuts method), but also creates a number of recourse variables in the master problem thus dramatically increases the dimension of the problem, even it is a linear program. In conclusion, there exists a trade-off between the running time and convergence speed (in terms of relative error change per iteration). The experiment demonstrates that BD with 1-round CCG results in the best convergence performance, indicating that it is desired to include a few rounds of CCG procedures into traditional BD dual cuts method in order to speed up convergence. In practice, the number of CCG rounds needs to be further tested.

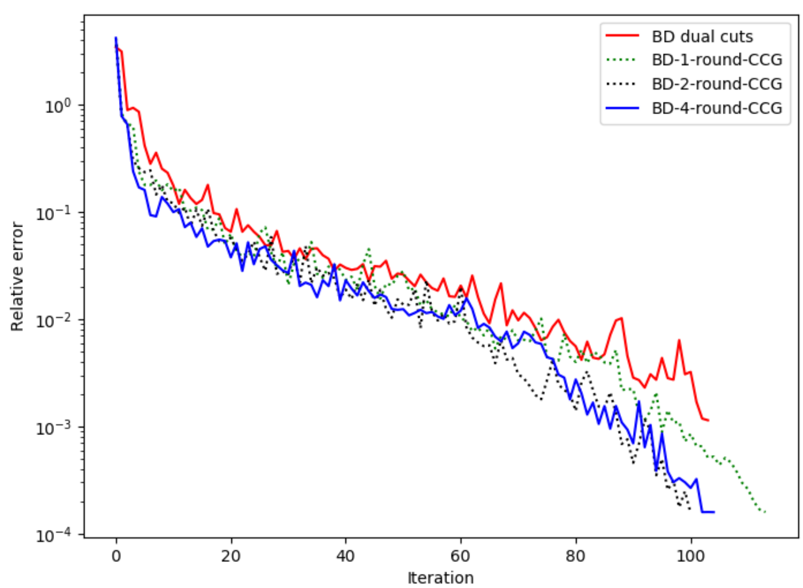


Figure 8.2: Algorithm convergence comparison (relative errors)

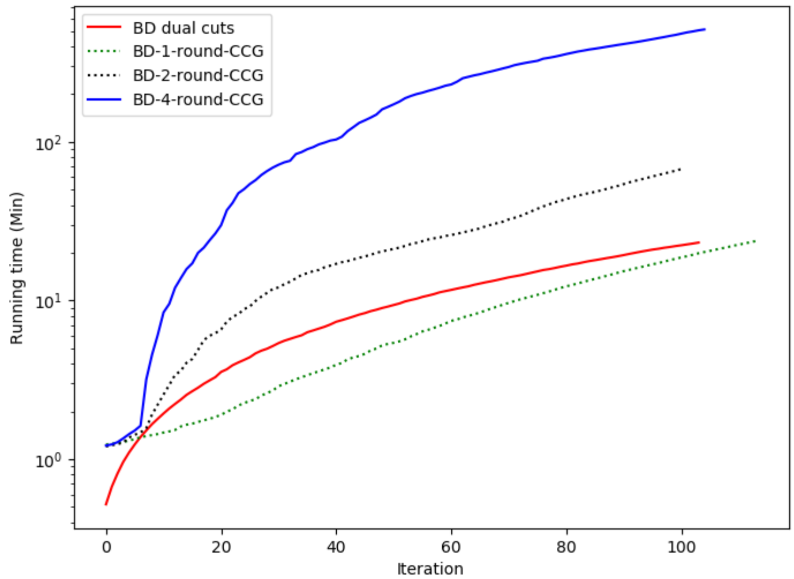


Figure 8.3: Algorithm convergence comparison (running time)

8.3 *Model Performance*

Having determined the weighting parameter λ and proper solution algorithm for the two-stage DR-QSOD model, now we start to compare the estimation performance of DR-QSOD model with that of deterministic QSOD model. Figure 8.4 compares the RMSE of OD estimates between DR-QSOD model and deterministic QSOD model. First, it can be seen that the two-stage DR-QSOD model produces an OD estimate d (first-stage decision variable) with the lowest RMSE (133.17 in blue solid line), compared to either the deterministic QSOD with input of any single day link flow (squares with red dash line) or an averaged link flow (140.58 in red solid line). Second, the RMSE of daily estimated OD demands t from the DR-QSOD model (stars with blue dash line) are mostly (25 out of 30) smaller than that of the deterministic QSOD model. In addition, the averaged RMSE of 30-day daily estimated OD demands for DR-QSOD model (172.15) is lower than that for deterministic QSOD model (178.79). This implies that by considering the stochastic feature of input data (i.e. link flow), DR-QSOD model can help generate the OD estimate with better accuracy.

Similar to the deterministic QSOD model, DR-QSOD model shows its great potentials in keeping OD quasi-sparsity consistency, i.e. demands of most OD pairs will keep significant/insignificant as they do in the prior OD demands. Figure 8.5 shows the capability of keeping OD quasi-sparsity for both DR-QSOD model and deterministic QSOD model. It can been seen that the two-stage DR-QSOD model can output first- and second-stage OD estimates with high insignificance consistency with an accuracy around 0.95, although slightly lower than that of deterministic QSOD model. Besides, the DR-QSOD model shares the similar sparsity solution property as the deterministic QSOD as well: over 552 OD pairs in this network, the majority (544) of OD pairs have the first-stage OD estimate equal to its prior value or zero. This indicate that DR-QSOD model share the similar solution property as the deterministic fixed-mapping QSOD, see (4.3), i.e. most of estimates will be either equal to its prior value or zero. Table 8.2 further illustrate such solution sparsity property: with 20 random sample OD pairs, 19 OD pairs have the estimate either equal to zero or its prior value. Both numerical tests indicate DR-QSOD model inherits the quasi-sparsity consistency and solution sparsity property from deterministic QSOD model.

We next compare the performance of DR-QSOD to that of an stochastic QSOD model. As discussed above, the two-stage DR-QSOD model allows certain levels of uncertainty of data, making it possible to estimate the daily OD demand given the realization of observed link flow. The stochastic QSOD model can also estimate daily OD flows, but with strict distribution assumptions for observed data. Here we compare the estimation performance of the DR-QSOD model and the stochastic QSOD model when data distributions are unknown or inaccurate.

First, we utilize the sample average approximation (SAA) approach [59] to recast the stochastic QSOD model as (8.1), where N denotes number of samples (or, number of days/ time slots) considered in this model; t_i corresponds to the OD demand of day i. Similar to the two-stage DR-QSOD model, (8.1) also yields to two sets of OD demands: one is d representing the overall OD demands within this period; another is t_i denoting the OD

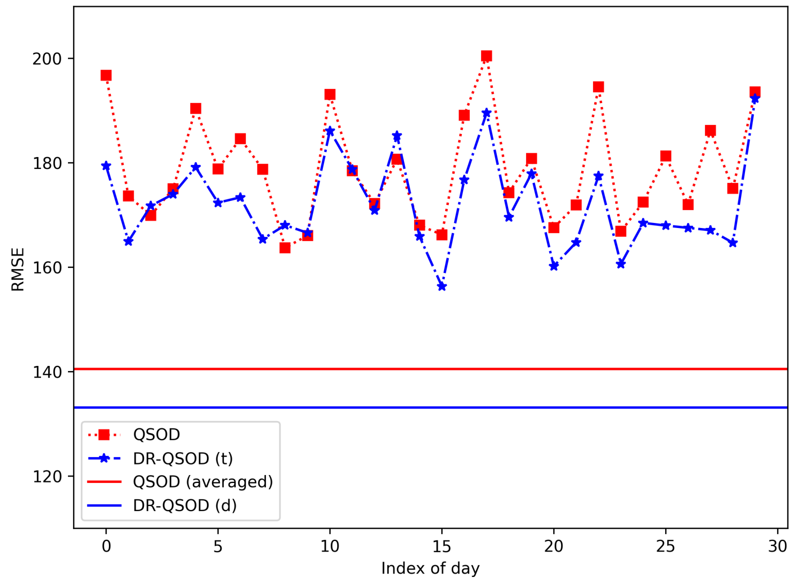


Figure 8.4: RMSE of estimates of DR-QSOD model and deterministic QSOD model

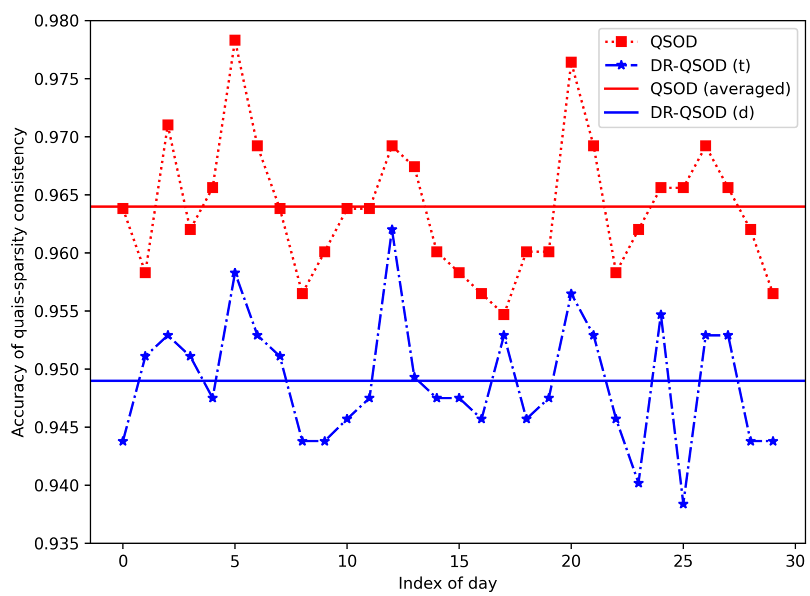


Figure 8.5: Accuracy of OD quasi-sparsity of estimates for DR-QSOD model and deterministic QSOD model

Table 8.2: Illustration of sparsity property of DR-QSOD model

$\bar{d}$	d^0	d	$\|d - d^0\|$
1	1.036	0	1.036
1	0.964	0.964	0
5	5.000	5.000	0
2	1.706	1.706	0
3	3.470	3.470	0
5	5.000	5.000	0
800	861.389	802.837	58.552
5	4.151	4.151	0
1300	1313.440	1313.440	0
5	4.569	4.569	0
2	1.750	1.750	0
5	5.000	5.000	0
3	3.745	3.745	0
5	3.861	3.861	0
5	5.000	5.000	0
4	4.206	4.206	0
1	0.941	0	0.941
3	2.675	0	2.675
3	3.262	3.262	0
1	0.978	0.978	0

demands estimated from link flows (ξ_i) of each individual day.

$$\begin{aligned} \min_{d,t_i} \frac{1}{\lambda}\|d-d^0\|_1 + & \frac{1}{N}\sum_{i=1}^{N}(\|t_i - d\|_1 + \|\hat{Q}t_i - \xi_i\|_1) \\ s.t. \quad & d \geq 0 \\ & t_i \geq 0, \forall i = 1, ..., N \end{aligned} \tag{8.1}$$

In order to test how the inaccuracy of distribution assumptions impacts both DR-QSOD model and stochastic QSOD model, we assume the link flow deviation (from the "true" link flow) for each link follows independent uniform distributions across different days, i.e. $\Delta_2' \sim uniform(-\epsilon_2^l, \epsilon_2^l), \forall l \in L$. The deviation of the observed link flows ξ_i we incorporate into the two models were generated based on independent normal distribution, i.e. $\Delta_2 \sim N(0, (\epsilon_2^l)^2), \forall l \in L$. In this way, it actually simulates a case where observed link flows are biased or its assumed distribution is not accurate enough. By setting $\epsilon_2 = 0.1$, we first generate a link flow set $\xi = \{\xi_1, \xi_2, ..., \xi_5\}$ consisting of 5-day link flows based on the normally distributed deviations. Then we use such ξ as the input, and solve both DR-QSOD and stochastic-QSOD model. As (8.1) has been converted into a linear program, we can readily solve it using any LP solver. Once we obtain the overall OD demands for both models (denote d' as the optimal overall OD demands for DR-QSOD, and d'' as the optimal overall OD demands for stochastic QSOD), we test how they perform respectively in the real-time stage (i.e., the second stage). That is, we fix the first stage d for both models, simulate a new set of samples $\xi' = \{\xi_1', \xi_2', ..., \xi_N'\}$ (generated from independent uniform distributions), and solve the second stage with the fixed d and ξ'. Table 8.3 demonstrates the RMSE of estimated daily OD demands from both models. It can be seen that when distribution bias exists, i.e. the assumed link flow distribution does not exact matches that of real link flows, the averaged RMSE of OD estimates from 10 sample days is lower than that of the stochastic QSOD model. We conducted 10 more similar tests as above, and obtained the same conclusion. This indicates that when the distribution of data is unknown or the assumed distribution for the data is not accurate, DR-QSOD can produce more accurate results than the stochastic QSOD model since DRO models are more robust in dealing with

Table 8.3: RMSE of daily estimated OD demands t for DR-QSOD and stochastic QSOD model

Distribution bias	Yes		No	
Index of day (i)	DR-QSOD	Stochastic-QSOD	DR-QSOD	Stochastic-QSOD
1	160.873	167.426	168.865	162.642
2	169.496	182.107	172.326	168.226
3	167.886	182.572	170.111	170.002
4	189.413	194.093	173.072	148.41
5	158.896	174.649	176.702	173.058
6	182.661	182.154	173.105	175.444
7	171.717	164.508	184.829	163.79
8	179.773	190.527	184.846	184.437
9	181.319	185.095	162.125	162.193
10	173.053	169.696	181.217	178.443
Average	173.509	179.283	174.720	168.665

data uncertainty. On the other hand, if the distribution bias does not exist, i.e. the assumed distribution of link flowss is the same as that of “true” link flows, the stochastic QSOD model outperforms DR-QSOD model as it produces a lower averaged RMSE, as shown in Table 8.3. The conclusion is validated by 10 more similar tests as well.

Chapter 9
DISCUSSIONS AND PRACTICAL INSIGHTS

9.1 OD Representativeness

Data representativeness has been attracting increasing amount of attention in both transportation research fields and practical applications. The assumption of the representativeness of prior OD demands is the foundation of the QSOD framework in this dissertation. Unlike existing OD estimation studies usually require prior OD demands be as close as possible to the true OD demands, QSOD models proposed in this study only require the prior travel demand of each OD pair share the same magnitude as its true value. It is well-known that the exact and accurate OD demands of real transportation networks, especially for large-scale metropolitan areas are usually difficult, if not impossible, to obtain. The prior OD representativeness assumption actually relaxes the requirements in OD estimation, which is practically useful in real applications. Under the QSOD framework, planning agencies do not need to pay much attention to the closeness between prior OD demands (which is the input for OD estimation) and true OD demands. Numerical experiments in Chapter 6 further demonstrate that even the prior OD demands are not representative, QSOD models can outperform the existing OD estimation models by producing more accurate OD demands

Under the representativeness assumption, this dissertation also demonstrates that the estimated OD demands can share the same OD quasi-sparsity as the prior OD demands for certain QSOD models. It is such representativeness assumption that plays the role of a bridge connecting true OD demands, prior OD demands, and estimated OD demands. With such OD quasi-sparsity consistency, QSOD models guarantee that the output OD demands will strictly follow the same quasi-sparsity as the true OD demands, which is also important for real applications.

9.2 Data Availability and Stochasticity

As discussed earlier, an increasing data availability also enables the proposed QSOD framework to consider stochasticity in OD estimation process. Such benefits can be understood from two folds. First, the special structure of the two-stage DR-QSOD model can output an overall OD estimate demand matrix which correpsonds to all observed data, as well as the realized OD demand matrix for each individual day. Compared with the single-stage model, like [120, 119], this model structure in fact provides more flexibility for OD estimation practitioners: one can estimate both overall OD estimates and daily OD estimate at the same time. The DR-QSOD model offers the equivalent capability to, if not better than, the stochastic OD estimation models [71] in describing OD stochasticity. Numerical results shows that DR-QSOD model can beat the corresponding stochastic-QSOD model when certain conditions exist, e.g. the distribution assumption is wrong or biased.

Another fold is about the DR-QSOD model itself. We conducted additional numerical results to understand how various levels of data valaibility would impact the model performance. Numerical results show that as more number of days of observed link flow data is available to be used, DR-QSOD model will produce better OD estimation results. The reasoning of such pattern is straightforward. As more data becomes available, theoretically it is easier to use sample data to describe the distribution of the population of the input data. Therefore, the estimated OD demands will be likely more accurate once more data is available.

9.3 Multi-sourced Data Integration under QSOD Framework

It is well known that link flow data of partial (or full) link set of transportation networks is a common and widely-used data source for most OD estimation studies, due to widely deployed traffic count detectors. The QSOD framework proposed in this dissertation also highly relies on the availability and accuracy of the network-level traffic link flow information. As more and more data sources, e.g. travel time, traffic turning proportions at intersection, OD demand products from app-based data, become available to transportation applications, it is much worthwhile to investigate how such new data sources can be incorporated into OD estimation studies and applications. The QSOD framework in this dissertation is equipped with such capability to incorporate multi-sourced data. In this section, we demonstrate two examples (OD demands from app-based data and travel time data) to leverage the framework and newly-acquired data to conduct OD estimation.

9.3.1 Integration of travel demands from app-based data

App-based data, as a passively-generated transportation data, has been attracting more and more attention in recent years [26, 110]. It is widely admitted that the travel demands provided by app-based data is only a portion of the travel demands generated by population. However, given partial travel demands provided by app-based data, we propose a variant of the fixed-mapping QSOD model as follow. For given OD pair w, denote u_w the travel demands captured by app-based data, d_w the target travel demand for population, λ_w the scaling factor between the two, where $\lambda_w = \frac{d_w}{u_w} \geq 0$. Following the QSOD framework, we have

$$\begin{aligned} &\min_{\lambda_w, v_a} \sum_{w \in W} |\lambda_w u_w - d_w^0| + \sum_{a \in A} |v_a - v_a^0| \\ &\text{s.t. } [v_a] = \hat{Q} \cdot [\lambda_w u_w] \end{aligned} \tag{9.1}$$

where $[\cdot]$ stands for a vector.

Note that parameter u_w represents the travel demands inferred from app-based data, which can be zero due to data availability. Based on this, one can impose a mathematical treatment: $u_w = \delta$, e.g. $\delta = 10e-4.$, if its original observed value $u_w^0 = 0$.

Unlike the original fixed-mapping QSOD model (4.3) directly providing an OD demand estimate given prior OD demands and observed link flows, the QSOD model variant (9.1) discussed here outputs a set of scaling factors $[\lambda_w]$ corresponding to each individual observed OD demands u_w from app-based data. Although the value of λ_w does not directly indicate the value of estimated OD demand itself, it does demonstrate an example for incorporating additional data sources, in addition to observed link flows and prior OD demands. What's more, such treatment has some physical meanings and implications. First, λ_w itself is the scaling factor between d_w and u_w, and its reciprocal stands for the portion of total demands which was captured by app-based data for given OD pair w. From the perspective of data representativeness, λ_w also indicates how reliable app-based data is in terms of monitoring the travel demands. Second, the estimates of the scaling factor λ_w make it possible to fully use the OD information from transportation big data. Once the estimates of scaling factors are obtained from (9.1) based on historical data (e.g. OD demands generated from app-based data), one can utilize such parameters to estimate the population-level OD demands given newly-observed transportation big data.

9.3.2 Integration of travel time

In addition to travel demands inferred from app-based data, travel time can be incorporated into the QSOD framework to further improve OD estimate quality. Travel time (usually modeled as travel cost) is a very important attribute in traffic assignment related problem. As the QSOD framework also involves the traffic assignment problem in various forms (e.g. fixed-mapping, and user-equilibrium), it provides us a great opportunity to make use of travel time for OD estimation. According to Waldrop's first principle regarding the network equilibrium [113], for given origin and destination, all used routes have the equal and least travel time under UE conditions. If we can obtain reliable and relatively accurate travel time data for certain key OD pairs, we can incorporate such information into the QSOD model.

Next, we present a QSOD model variant by using travel time. To simplify the discussion, we consider the fixed-mapping case as follow:

$$\begin{aligned} \min_{d} & \|d - d^0\|_1 + \|\hat{Q}d - v^0\|_1 + \sum_{p \in P} |\tau_p(\hat{Q}d) - \tau_p^0| \\ s.t. \quad & d \geq 0 \end{aligned} \tag{9.2}$$

where $\tau_p(\hat{Q}d) = \sum_{a \in p} c_a(\hat{Q}d)$.

Note that τ_p^0 is the travel time of path p, which is obtained from other data sources; P denotes the set of paths we consider, and it mainly consists of key paths within a city or area; $\tau_p(\hat{Q}d)$ represents the travel time of path p, which is the sum of link travel time of the path p; $c_a(\cdot)$ stands for the travel time function of link a, which is a function of link flow vector $\hat{Q}d$ (one can leverage the BPR function to characterize $c_a(\cdot)$ here).

By incorporating the travel time deviation into the objective of (9.2), the framework will penalize the objective if the resulting travel time is far from the observed travel time, thus further helps produce better estimated OD demands that matching the travel time. Similarly, one can apply the travel time deviation item for user-equilibrium case. The difference only lies in the traffic assignment stage. It is worthwhile to mention that travel time data, especially for those key OD pairs in an area, is much easier to obtain that before. To name a few, with the help of the licence plate recognition system, traffic management/operations departments, e.g. Department of Transportation (DOTs) are able to monitor the travel times of certain OD pairs. More recently, with the increasing availability of app-based data, it is more convenient for researchers and transportation practitioners to infer the network travel time of individuals, with a larger geographic and longitudinal coverage. Based on this, we expect more applications and studies by combining the proposed QSOD framework and those emerging transportation data sources.

Chapter 10
SUMMARY AND FUTURE WORK

In this dissertation, we proposed a quasi-sparsity based OD (QSOD) estimation framework that includes the fixed-mapping QSOD model, the bi-level QSOD model, and the DR-QSOD model. We defined insignificant OD demands and OD pairs, and based on which the quasi-sparsity property of an OD demand matrix. By applying the method of compressed sensing, the QSOD models explored the sparsity of the deviation between the prior OD matrix and the estimated OD matrix. For deterministic QSOD models, i.e. the fixed-mapping QSOD mdoel and bi-level QSOD model, we studied the OD quasi-sparsity consistency, i.e., under what conditions, the estimated OD demand matrix will hold the same quasi-sparsity property as the prior demand matrix. We also investigated the solution sparsity property of the QSOD models and found that, for large-size networks, the estimated demands of most OD pairs will be either equal to its prior values or zero (or the small value ϵ for the bi-level QSOD case). For the stochastic case, we studies the DR-QSOD model, especially formulated it into a two-stage formulation and evaluated how the distributional uncertainty would impact the OD estimate result, compared with its stochastic counterpart.

The proposed framework was tested using three example networks: a 5-node network and the downtown Seattle network for both the fixed-mapping QSOD model and the bi-level QSOD model, and Sioux Falls network for the DR-QSOD model. Theoretical and numerical results showed that for the deterministic QSOD models, the estimated OD demand matrix will share the same quasi-sparsity properties as the prior OD demand matrix once certain conditions are satisfied. Even those conditions are not satisfied, the proposed QSOD models, by focusing on the quasi-sparsity of OD demand matrix, can help maintain relatively high OD quasi-sparsity consistency with comparable estimation errors when compared with existing models, which are also less computationally demanding. For the DR-QSOD model, results demonstrated its great capability in dealing with the distributional uncertainty for the input data, compared with stochastic QSOD model. Algorithms for each of these QSOD models were discussed in detail and validated with numerical experiments.

In addition to three QSOD models mentioned above, this study also discussed the increasing availability of transportation data, and particularly explored some insightful directions based on data and QSOD framework, e.g. data representativeness, multi-sourced data integration for QSOD model, data distributional uncertainty and stochasticity, and etc. The quasi-sparsity of OD demands, the proposed QSOD models, and the theoretical and numerical results in this dissertation are expected to provide some insights on OD estimation for real-world transportation networks. For example, the prior OD demand matrix is often assumed to be as close as possible to the true OD demand matrix. This of course is extremely difficult to achieve or even to check in practice. Our results show that we may only need to prepare the prior OD matrix such that it is representative (Definition 3 in Section 3.1). This way, the estimated OD demand matrix will share the same quasi-sparsity property with the prior OD (and the true OD as well) exactly (Theorem 4.1 and Theorem 5.2) or approximately (Theorem 4.2 and Theorem 5.3), with reasonable estimation errors.

Future research may first refine Theorem 4.1 and Theorem 5.2 with sharper conditions (e.g., sufficient and necessary conditions), and generalize the conditions in the theorems for cases where the insignificant OD pair set W_1 may not be fixed. Second, we added a small positive lower bound (ϵ) to d and d^0 in the bi-level QSOD model to make the analysis easier. Mathematically, it is interesting (and challenging) to investigate whether we can relax this (i.e., setting $\epsilon = 0$) and if so whether the quasi-sparsity consistency and solution sparsity of the bi-level QSOD model still hold. Third is to provide theoretical proofs for quasi-sparsity consistency as well as solution sparsity property of DR-QSOD model, based on the discussions of deterministic QSOD models in Chapter 4 and 5. Fourth, it is also interesting to explore other approaches to build the ambiguity set of observed data for DR-QSOD model so that the model can leverage more meaningful statistical information to give more accurate results. Fifth, we only considered prior OD demands and link flows as the input data to the QSOD models, and briefly introduced the idea for incorporating other data sources. Conducting solid numerical results for the multi-sourced integration strategy in Section 9.3 is of great importance both scientifically and practically. Sixth, the OD observability problem [21, 22, 54, 121], i.e. how to better estimate the OD demands when partial of prior ODs are available, is worthwhile to explore. This is also very related to the sensitivity analysis of QSOD models, which will be helpful for the practical applications of related models. Seventh, the current study only covers the investigations for the static QSOD framework. The dynamic QSOD estimation problem, where variables (e.g., OD demands and link flows) change with time and traffic dynamics are explicitly considered, is worthwhile to explore in the future. Eighth, preparing a prior OD demand matrix that is representative of the true OD matrix is not trivial either. This is closely related to recent research efforts on using emerging transportation big data for OD demand estimation and other related human mobility pattern analyses, for which data representativeness or bias was identified as one of the major issues that need to be carefully addressed [5, 26]. This is an emerging area of research and results could help prepare representative prior OD demand matrices. Last but not least, future research should apply the proposed QSOD models to OD estimation applications on large-scale networks. The QSOD framework proposed in this dissertation sits in an ideal position to connect emerging trasnportation data sources with traditional OD estimation theories. The authors may investigate those topics, and the results may be reported in the subsequent research papers and conferences.

BIBLIOGRAPHY

[1] Maryam Abareshi. Maximum probability od matrix estimation in large-sized networks. *Iranian Journal of Numerical Analysis and Optimization*, 10(2):105–131, 2020.

[2] Gautam Appa. On the uniqueness of solutions to linear programs. *Journal of the Operational Research Society*, 53(10):1127–1132, 2002.

[3] Sadra Babaei, Ruiwei Jiang, and Chaoyue Zhao. Distributionally robust distribution network configuration under random contingency. *IEEE Transactions on Power Systems*, 35(5):3332–3341, 2020.

[4] Jeff X Ban, Henry X Liu, Michael C Ferris, and Bin Ran. A general mpcc model and its solution algorithm for continuous network design problem. *Mathematical and Computer Modelling*, 43(5-6):493–505, 2006.

[5] X Ban, C Chen, F Wang, J Wang, and Y Zhang. Promises of data from emerging technologies for transportation applications: Puget sound region case study (no. fhwahep-19-026). federal highway administration report fhwa-or-15-01, 2018.

[6] Xuegang Ban, Henry X Liu, Jiangang Lu, and Michael C Ferris. Decomposition scheme for continuous network design problem with asymmetric user equilibria. *Transportation research record*, 1964(1):185–192, 2006.

[7] Xuegang Jeff Ban. *Quasi-variational inequality formulations and solution approaches for dynamic user equilibria*. The University of Wisconsin-Madison, 2005.

[8] H Bar-Gera and D Boyce. Route flow entropy maximization in origin-based traffic assignment. In *14th International Symposium on Transportation and Traffic Theory-Transportation Research Institute*, 1999.

[9] Hillel Bar-Gera. Origin-based algorithm for the traffic assignment problem. *Transportation Science*, 36(4):398–417, 2002.

[10] Martin Beckmann, Charles B McGuire, and Christopher B Winsten. Studies in the economics of transportation. Technical report, 1956.

[11] Aharon Ben-Tal, Dick den Hertog, Anja De Waegenaere, Bertrand Melenberg, and Gijs Rennen. Robust solutions of optimization problems affected by uncertain probabilities. *Management Science*, 59(2):341–357, 2013.

[12] Sharminda Bera and KV Rao. Estimation of origin-destination matrix from traffic counts: the state of the art. 2011.

[13] Dimitris Bertsimas and John N Tsitsiklis. *Introduction to linear optimization*, volume 6. Athena Scientific Belmont, MA, 1997.

[14] John R Birge and Francois Louveaux. *Introduction to Stochastic Programming*. Springer Science & Business Media, 2011.

[15] Stephen Boyd, Neal Parikh, and Eric Chu. *Distributed optimization and statistical learning via the alternating direction method of multipliers*. Now Publishers Inc, 2011.

[16] Emmanuel J Candès et al. Compressive sampling. In *Proceedings of the international congress of mathematicians*, volume 3, pages 1433–1452. Madrid, Spain, 2006.

[17] Coralia Cartis, Nicholas IM Gould, and Philippe L Toint. On the evaluation complexity of composite function minimization with applications to nonconvex nonlinear programming. *SIAM Journal on Optimization*, 21(4):1721–1739, 2011.

[18] Ennio Cascetta. Estimation of trip matrices from traffic counts and survey data: a generalized least squares estimator. *Transportation Research Part B: Methodological*, 18(4-5):289–299, 1984.

[19] Ennio Cascetta and Sang Nguyen. A unified framework for estimating or updating origin/destination matrices from traffic counts. *Transportation Research Part B: Methodological*, 22(6):437–455, 1988.

[20] Enrique Castillo, Aida Calviño, Hong K Lo, José María Menéndez, and Zacarías Grande. Non-planar hole-generated networks and link flow observability based on link counters. *Transportation Research Part B: Methodological*, 68:239–261, 2014.

[21] Enrique Castillo, Antonio J Conejo, José María Menéndez, and Pilar Jiménez. The observability problem in traffic network models. *Computer-Aided Civil and Infrastructure Engineering*, 23(3):208–222, 2008.

[22] Enrique Castillo, Pilar Jimenez, Jose Maria Menendez, and Antonio J Conejo. The observability problem in traffic models: algebraic and topological methods. *IEEE Transactions on Intelligent Transportation Systems*, 9(2):275–287, 2008.

[23] Enrique Castillo, José María Menéndez, Santos Sánchez-Cambronero, Aida Calviño, and José María Sarabia. A hierarchical optimization problem: estimating traffic flow using gamma random variables in a bayesian context. *Computers & Operations Research*, 41:240–251, 2014.

[24] Sanjay Chawla, Yu Zheng, and Jiafeng Hu. Inferring the root cause in road traffic anomalies. In *2012 IEEE 12th International Conference on Data Mining*, pages 141–150. IEEE, 2012.

[25] Anthony Chen, Zhaowang Ji, and Will Recker. Travel time reliability with risk-sensitive travelers. *Transportation Research Record*, 1783(1):27–33, 2002.

[26] Cynthia Chen and Xuegang Jeff Ban. Transportation big data: Promises, issues, and implications. Technical report, 2018.

[27] Scott Shaobing Chen, David L Donoho, and Michael A Saunders. Atomic decomposition by basis pursuit. *SIAM review*, 43(1):129–159, 2001.

[28] Stephen Clark and David Watling. Modelling network travel time reliability under stochastic demand. *Transportation Research Part B: Methodological*, 39(2):119–140, 2005.

[29] Greater Niagara-Buffalo Regional Transportation Council. Moving forward 2050, 2018.

[30] New York Metropolitan Transportation Council. New york best practice model, 2015.

[31] Puget Sound Regional Council. Puget sound 4k model, 2015.

[32] M Cremer and H Keller. A new class of dynamic methods for the identification of origin-destination flows. *Transportation Research Part B: Methodological*, 21(2):117–132, 1987.

[33] Stella Dafermos. Traffic equilibrium and variational inequalities. *Transportation Science*, 14(1):42–54, 1980.

[34] Erick Delage and Yinyu Ye. Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations research*, 58(3):595–612, 2010.

[35] Erick Delage and Yinyu Ye. Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research*, 58(3):595–612, 2010.

[36] Xuan Di, Henry X Liu, Jong-Shi Pang, and Xuegang Jeff Ban. Boundedly rational user equilibria (brue): mathematical formulation and solution sets. *Procedia-Social and Behavioral Sciences*, 80:231–248, 2013.

[37] Michael P Dixon and Laurence R Rilett. Real-time od estimation using automatic vehicle identification and traffic count data. *Computer-Aided Civil and Infrastructure Engineering*, 17(1):7–21, 2002.

[38] David L Donoho. Compressed sensing. *IEEE Transactions on information theory*, 52(4):1289–1306, 2006.

[39] David Leigh Donoho. Method and apparatus for compressed sensing, January 12 2010. US Patent 7,646,924.

[40] Dmitriy Drusvyatskiy and Adrian S Lewis. Error bounds, quadratic growth, and linear convergence of proximal methods. *Mathematics of Operations Research*, 43(3):919–948, 2018.

[41] Dmitriy Drusvyatskiy and Courtney Paquette. Efficiency of minimizing compositions of convex functions and smooth maps. *Mathematical Programming*, 178(1-2):503–558, 2019.

[42] Lars Eldén. Perturbation theory for the least squares problem with linear equality constraints. *SIAM Journal on Numerical Analysis*, 17(3):338–350, 1980.

[43] Peyman Mohajerin Esfahani and Daniel Kuhn. Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations. *Mathematical Programming*, 171(1–2):115–166, 2018.

[44] Francisco Facchinei and Jong-Shi Pang. *Finite-dimensional variational inequalities and complementarity problems*. Springer Science & Business Media, 2007.

[45] Caroline Fisk. More paradoxes in the equilibrium assignment problem. *Transportation Research Part B: Methodological*, 13(4):305–309, 1979.

[46] CS Fisk. On combining maximum entropy trip matrix estimation with user optimal assignment. *Transportation Research Part B: Methodological*, 22(1):69–73, 1988.

[47] Byron J Gajewski, Laurence R Rilett, Michael P Dixon, and Clifford H Spiegelman. Robust estimation of origin-destination matrices. *Journal of Transportation and Statistics*, 5(2/3):37–56, 2002.

[48] Joel Goh and Melvyn Sim. Distributionally robust optimization and its tractable approximations. *Operations research*, 58(4-part-1):902–917, 2010.

[49] Joel Goh and Melvyn Sim. Distributionally robust optimization and its tractable approximations. *Operations Research*, 58(4-part-1):902–917, 2010.

[50] Osman Güler, Alan J Hoffman, and Uriel G Rothblum. Approximations to solutions to systems of linear inequalities. *SIAM Journal on Matrix Analysis and Applications*, 16(2):688–696, 1995.

[51] Martin L Hazelton. Estimation of origin–destination matrices from link flows on uncongested networks. *Transportation Research Part B: Methodological*, 34(7):549–566, 2000.

[52] Martin L Hazelton. Estimation of origin–destination trip rates in leicester. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 50(4):423–433, 2001.

[53] Martin L Hazelton. Some comments on origin–destination matrix estimation. *Transportation Research Part A: Policy and Practice*, 37(10):811–822, 2003.

[54] Shou-Ren Hu, Srinivas Peeta, and Chun-Hsiao Chu. Identification of vehicle sensor locations for link-based network traffic applications. *Transportation Research Part B: Methodological*, 43(8-9):873–894, 2009.

[55] Zhaolin Hu and L. Jeff Hong. Kullback-Leibler divergence constrained distributionally robust optimization. *Optimization Online preprint*, 2012.

[56] Peter J Huber. Robust estimation of a location parameter. In *Breakthroughs in statistics*, pages 492–518. Springer, 1992.

[57] Yasunori Iada and Jun-ichi Takayama. Comparison study of model formulations on od matrix estimation from observed link flows. In *Research for Tomorrow's Transport Requirements. Proceedings of the Fourth World Conference on Transport ResearchUniversity of British Columbia, Vancouver*, 1986.

[58] Diethard Klatte and Gisbert Thiere. Error bounds for solutions of linear equations and inequalities. *Zeitschrift für Operations Research*, 41(2):191–214, 1995.

[59] Anton J Kleywegt, Alexander Shapiro, and Tito Homem-de Mello. The sample average approximation method for stochastic discrete optimization. *SIAM Journal on Optimization*, 12(2):479–502, 2002.

[60] Pramesh Kumar, Alireza Khani, and Gary A Davis. Transit route origin–destination matrix estimation using compressed sensing. *Transportation Research Record*, 2673(10):164–174, 2019.

[61] Adrian S Lewis and Stephen J Wright. A proximal method for composite minimization. *Mathematical Programming*, 158(1-2):501–546, 2016.

[62] Baibing Li. Bayesian inference for origin-destination matrices of transport networks using the em algorithm. *Technometrics*, 47(4):399–408, 2005.

[63] Pei-Wei Lin and Gan-Len Chang. Robust estimation of the dynamic origin-destination matrix for a freeway network. In *IEEE International Conference on Networking, Sensing and Control, 2004*, volume 2, pages 862–867. IEEE, 2004.

[64] Pei-Wei Lin and Gang-Len Chang. Robust model for estimating freeway dynamic origin–destination matrix. *Transportation Research Record*, 1923(1):110–118, 2005.

[65] Ming Liu, Xin Liu, Feng Chu, Feifeng Zheng, and Chengbin Chu. Distributionally robust inventory routing problem to maximize the service level under limited budget. *Transportation Research Part E: Logistics and Transportation Review*, 126:190–211, 2019.

[66] Shu Lu. Sensitivity of static traffic user equilibria with perturbations in arc cost function and travel demand. *Transportation science*, 42(1):105–123, 2008.

[67] Shu Lu. Solution sparsity of minimizing the l_1 norm. Manuscript, 2022.

[68] Shu Lu and Yu Marco Nie. Stability of user-equilibrium route flow solutions for the traffic assignment problem. *Transportation Research Part B: Methodological*, 44(4):609–617, 2010.

[69] Zhi-Quan Luo, Jong-Shi Pang, and Daniel Ralph. *Mathematical programs with equilibrium constraints*. Cambridge University Press, 1996.

[70] Zhi-Quan Luo, Jong-Shi Pang, and Daniel Ralph. *Mathematical programs with equilibrium constraints*. Cambridge University Press, 1996.

[71] Wei Ma and Zhen Sean Qian. Statistical inference of probabilistic origin-destination demand using day-to-day traffic data. *Transportation Research Part C: Emerging Technologies*, 88:227–256, 2018.

[72] Michael J Maher. Inferences on trip matrices from observations on link volumes: a bayesian statistical approach. *Transportation Research Part B: Methodological*, 17(6):435–447, 1983.

[73] Hani S Mahmassani and Gang-Len Chang. On boundedly rational user equilibrium in transportation systems. *Transportation science*, 21(2):89–99, 1987.

[74] Olvi Mangasarian. Uniqueness of solution in linear programming. Technical report, University of Wisconsin-Madison Department of Computer Sciences, 1978.

[75] Morteza Mardani and Georgios B Giannakis. Robust network traffic estimation via sparsity and low rank. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 4529–4533. IEEE, 2013.

[76] William A Martin and Nancy A McGuckin. *Travel estimation techniques for urban planning*, volume 365. National Academy Press Washington, DC, 1998.

[77] Aditya Krishna Menon, Chen Cai, Weihong Wang, Tao Wen, and Fang Chen. Fine-grained od estimation with automated zoning and sparsity regularisation. *Transportation Research Part B: Methodological*, 80:150–172, 2015.

[78] Robert B Mitchell and Chester Rapkin. *Urban traffic: A function of land use*. Columbia University Press, 1954.

[79] E. K. Morlok. Development and application of a highway network design model. *Federal Highway Administration*, 1-2, 1973.

[80] Frederick Mosteller, John Wilder Tukey, et al. *Data analysis and regression: a second course in statistics.* 1977.

[81] Shoichiro Nakayama and David Watling. Consistent formulation of network equilibrium with stochastic flows. *Transportation Research Part B: Methodological*, 66:50–69, 2014.

[82] Nancy L Nihan and Gary A Davis. Application of prediction-error minimization and maximum likelihood to estimate intersection od matrices from traffic counts. *Transportation Science*, 23(2):77–90, 1989.

[83] J. Nocedal and S. J. Wright. *Numerical Optimization.* Springer Series in Operations Research. Springer-Verlag, New York, 2006. Second Edition.

[84] United States. Bureau of Public Roads. *Traffic assignment manual for application with a large, high speed computer*, volume 2. US Department of Commerce, Bureau of Public Roads, Office of Planning, Urban . . . , 1964.

[85] Katharina Parry and Martin L Hazelton. Estimation of origin–destination matrices from link counts and sporadic routing data. *Transportation Research Part B: Methodological*, 46(1):175–188, 2012.

[86] Javier Pena, Juan Vera, and Luis Zuluaga. An algorithm to compute the hoffman constant of a system of linear constraints. *arXiv preprint arXiv:1804.08418*, 2018.

[87] Florian A. Potra and Stephen J. Wright. Interior-point methods. *Journal of Computational and Applied Mathematics*, 124(1):281–302, 2000.

[88] David Martin Powers. Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation. 2011.

[89] Hamed Rahimian and Sanjay Mehrotra. Distributionally robust optimization: A review. *arXiv preprint arXiv:1908.05659*, 2019.

[90] D. Ralph and Wright S.J. Some properties of regularization and penalization schemes for mpecs. *Optimization Methods and Software*, 19:527—556, 2004.

[91] Stephen M. Robinson. Local structure of feasible sets in nonlinear programming, Part III: Stability and sensitivity. *Mathematical Programming Studies*, 30:45–66, 1987.

[92] Stephen M. Robinson. An implicit-function theorem for a class of nonsmooth functions. *Mathematics of Operations Research*, 16:292–309, 1991.

[93] Stephen M Robinson and Shu Lu. Solution continuity in variational conditions. *Journal of Global Optimization*, 40(1-3):405–415, 2008.

[94] R. Tyrrell Rockafellar. *Convex Analysis*. Princeton University Press, Princeton, NJ, 1970.

[95] R Tyrrell Rockafellar and Roger J-B Wets. *Variational analysis*, volume 317. Springer Science & Business Media, 2009.

[96] Peter J Rousseeuw and Annick M Leroy. *Robust regression and outlier detection*, volume 589. John wiley & sons, 2005.

[97] Borhan M Sanandaji and Pravin P Varaiya. Compressive origin-destination matrix estimation. *arXiv preprint arXiv:1404.3263*, 2014.

[98] Herbert Scarf. A min-max solution of an inventory problem. *Studies in the mathematical theory of inventory and production*, 1958.

[99] Stefan Scholtes. *Introduction to Piecewise Differentiable Equations*. SpringerBriefs in Optimization. Springer, 2012.

[100] Chao Shang and Fengqi You. Distributionally robust optimization for planning and scheduling under uncertainty. *Computers & Chemical Engineering*, 110:53–68, 2018.

[101] H Shao, Hing Keung William Lam, Q Meng, and ML Tam. A demand driven travel time reliability-based traffic assignment problem. In *Transportation Research Board. Meeting*, 2006.

[102] Hu Shao, William HK Lam, Agachai Sumalee, Anthony Chen, and Martin L Hazelton. Estimation of mean and covariance of peak hour origin–destination demands from day-to-day traffic counts. *Transportation Research Part B: Methodological*, 68:52–75, 2014.

[103] M. J. Smith. The existence, uniqueness and stability of traffic equilibria. *Transportation Research Part B*, 13(3):295–304, 1979.

[104] Arthur B Sosslau, Amin B Hassam, Maurice M Carter, and George V Wickstrom. Quick-response urban travel estimation techniques and transferable parameters. user's guide. Technical report, 1978.

[105] Heinz Spiess. A maximum likelihood model for estimating origin-destination matrices. *Transportation Research Part B: Methodological*, 21(5):395–412, 1987.

[106] Ben Stabler. Transportation networks. https://github.com/bstabler/TransportationNetworks, 2016.

[107] Hans JH Tuenter. Minimum l1-distance projection onto the boundary of a convex set: Simple characterization. *Journal of optimization theory and applications*, 112(2):441–445, 2002.

[108] Charles F Van Loan and Gene H Golub. *Matrix computations*. Johns Hopkins University Press Baltimore, 1983.

[109] Henk J Van Zuylen and Luis G Willumsen. The most likely trip matrix estimated from traffic counts. *Transportation Research Part B: Methodological*, 14(3):281–293, 1980.

[110] Feilong Wang and Cynthia Chen. On data processing required to derive mobility patterns from passively-generated mobile phone data. *Transportation Research Part C: Emerging Technologies*, 87:58–74, 2018.

[111] Jingxing Wang, Shu Lu, Hongsheng Liu, and Xuegang Ban. Transportation origin-destination demand estimation with quasi-sparsity. *Transportation science*, 57(2):289–312, 2023.

[112] Po-Wei Wang and Chih-Jen Lin. Iteration complexity of feasible descent methods for convex optimization. *The Journal of Machine Learning Research*, 15(1):1523–1548, 2014.

[113] John Glen Wardrop and James Ivor Whitehead. Correspondence. some theoretical aspects of road traffic research. *Proceedings of the Institution of Civil Engineers*, 1(5):767–768, 1952.

[114] Edward Weiner. *Urban transportation planning in the United States: an historical overview*. US Department of Transportation, 1997.

[115] Tao Wen, Chen Cai, Lauren Gardner, Steven Travis Waller, Vinayak Dixit, and Fang Chen. Estimation of sparse o–d matrix accounting for demand volatility. *IET Intelligent Transport Systems*, 12(9):1020–1026, 2018.

[116] Wolfram Wiesemann, Daniel Kuhn, and Melvyn Sim. Distributionally robust convex optimization. *Operations Research*, 62(6):1358–1376, 2014.

[117] Linwei Xin and David A Goldberg. Distributionally robust inventory control when demand is a martingale. *arXiv preprint arXiv:1511.09437*, 2015.

[118] H. Yang and M.G.H Bell. Transport bilevel programming problems: recent methodological advances. *Transportation Research Part B*, 35:1–4, 2001.

[119] Hai Yang, Yasunori Iida, and Tsuna Sasaki. The equilibrium-based origin-destination matrix estimation problem. *Transportation Research Part B: Methodological*, 28(1):23–33, 1994.

[120] Hai Yang, Tsuna Sasaki, Yasunori Iida, and Yasuo Asakura. Estimation of origin-destination matrices from link traffic counts on congested networks. *Transportation Research Part B: Methodological*, 26(6):417–434, 1992.

[121] Yudi Yang and Yueyue Fan. Data dependent input control for origin–destination demand estimation using observability analysis. *Transportation Research Part B: Methodological*, 78:385–403, 2015.

[122] Yudi Yang, Yueyue Fan, and Johannes O Royset. Estimating probability distributions of travel demand on a congested network. *Transportation Research Part B: Methodological*, 122:265–286, 2019.

[123] Yudi Yang, Yueyue Fan, and Roger JB Wets. Stochastic travel demand estimation: Improving network identifiability using multi-day observation sets. *Transportation Research Part B: Methodological*, 107:192–211, 2018.

[124] Bo Zeng and Long Zhao. Solving two-stage robust optimization problems using a column-and-constraint generation method. *Operations Research Letters*, 41(5):457–461, 2013.

[125] Chaoyue Zhao and Yongpei Guan. Data-driven risk-averse stochastic optimization with Wasserstein metric. *Operations Research Letters*, 46(2):262–267, 2018.

[126] Chaoyue Zhao and Ruiwei Jiang. Distributionally robust contingency-constrained unit commitment. *IEEE Transactions on Power Systems*, 33(1):94–102, 2017.

[127] Xuesong Zhou and Hani S Mahmassani. Dynamic origin-destination demand estimation using automatic vehicle identification data. *IEEE Transactions on intelligent transportation systems*, 7(1):105–114, 2006.

[128] Steve Zymler, Daniel Kuhn, and Berç Rustem. Distributionally robust joint chance constraints with second-order moment information. *Mathematical Programming*, 137(1–2):167–198, 2011.

Appendix A

THE UPPER BOUND OF R_2

When testing the conditions for Theorem 4.1, if the calculation of r_2 is hard or not necessary, one can turn to finding the upper bound of r_2 instead. According to the Hoffman' error bound theorem [44], there exists a constant $H(\hat{Q})$ that depends only on the matrix $\hat{Q}$, such that the L_2 distance from (d^0, v^0) to S is bounded above by $H(\hat{Q}) \cdot (\|\hat{Q}d^0 - v^0\| + \|min(0, (d^0, v^0))\|)$, where $H(\hat{Q})$ is called the Hoffman's constant. If we use r_3 to denote such an L_2 distance, it turns out that $r_3 \leq H(\hat{Q}) \cdot (\|\hat{Q}d^0 - v^0\| + \|min(0, (d^0, v^0))\|)$. Since for any vector x, it is always true that $\|x\|_1 \leq \sqrt{n}\|x\|_2, \forall x \in \mathbb{R}^n$. Based on the definition of r_2, we have $r_2 \leq \sqrt{n}r_3 \leq \sqrt{n}H(\hat{Q}) \cdot (\|\hat{Q}d^0 - v^0\| + \|min(0, (d^0, v^0))\|)$. This means $\sqrt{n}H(\hat{Q}) \cdot (\|\hat{Q}d^0 - v^0\| + \|min(0, (d^0, v^0))\|)$ can be considered as an upper bound of r_2. According to Theorem 4.1, if r_1 is greater than this upper bound, it can be stated that $\hat{d} \in K$ and the OD quasi-sparsity consistency holds. Unfortunately, calculating the Hoffman's constant is very challenging [58]. We could however estimate the Hoffman's constant via some algorithmic approaches based on the characterization of matrix $\hat{Q}$ [50, 112, 86]. Details are omitted here to save space.

Appendix B
DIFFERENTIABILITY OF $\Psi(D)$

Recall that, under Assumption 5.1, for each $d \in \mathbb{R}_+^{|W|}$ there is a unique user-equilibrium link flow denoted by $\Psi(d)$, which is the unique solution to the following traffic user equilibrium problem

$$v \in G(d) \quad \text{and} \quad \langle c(v), v' - v\rangle \geq 0 \text{ for each } v' \in G(d). \tag{B.1}$$

Here, $c : \mathbb{R}_+^{|A|} \to \mathbb{R}^{|A|}$ is the link cost function, and $G(d)$ is the set of feasible link flows defined as

$$G(d) = \left\{ v \in \mathbb{R}^{|A|} \mid \; v = \Delta q \; \text{ for some } q \in \mathbb{R}_+^{|\mathcal{R}|} \; \text{ satisfying } \Lambda q = d \right\}, \tag{B.2}$$

where Λ and Δ are the OD-route incidence matrix and the link-route incidence matrix respectively. In this section, we consider a point $d^* \in \mathbb{R}^{|W|}$ with $d^* > 0$, and give a condition that guarantees Ψ to be continuously differentiable on an open neighborhood of d^* in $\mathbb{R}_{++}^{|W|}$. See [66] for more references on sensitivity analysis of the traffic user equilibrium problem.

The traffic user equilibrium problem can also be written in a route-based formulation

$$q \in H(d) \quad \text{and} \quad \langle C(q), q' - q\rangle \geq 0 \text{ for each } q' \in H(d), \tag{B.3}$$

where $C(q) = \Delta^T c(\Delta q)$ is the route cost function and $H(d)$ is the set of feasible route flows under d defined as

$$H(d) = \left\{ q \in \mathbb{R}_+^{|\mathcal{R}|} \mid \; \Lambda q = d \right\}.$$

A solution to (B.3) is called a user-equilibrium route flow under d. The formulation (B.3) is a characterization of the Wardrop traffic equilibrium, in the sense that $q \in H(d)$ solves (B.3) if and only if $q_p = 0$ for any non-user-optimal route p. Finally, (B.1) and (B.3) are related in that v is a solution to (B.1) if and only if there exists a solution q of (B.3) with $v = \Delta q$, and that $q \in H(d)$ solves (B.3) if and only if Δq solves (B.1).

Now, let $v^* = \Psi(d^*)$, and let $q^* \in \mathbb{R}_+^{|\mathcal{R}|}$ be a user-equilibrium route flow under d^*. Since v^* is the only user-equilibrium link flow under Assumption 5.1, we have $v^* = \Delta q^*$. It is possible to have multiple user-equilibrium route flows corresponding to v^*, so the choice of q^* is not unique. Since $d^* > 0$, for each OD pair w there exists at least one route p connecting w with $q_p^* > 0$; we choose one such route for each w and denote it by $p(w)$. We then divide the set $\mathcal{R}$ of all routes into the following four subsets:

$$\begin{aligned}
&\mathcal{R}^{00} \text{ is the set of all user-optimal routes } p \text{ with } q_p^* = 0,\\
&\mathcal{R}^{01} \text{ is the set of all non-user-optimal routes,}\\
&\mathcal{R}^{1} = \{p(w), w \in W\},\\
&\mathcal{R}^{2} \text{ is the set of all routes } p \text{ with } q_p^* > 0 \text{ not in } \mathcal{R}^1.
\end{aligned} \tag{B.4}$$

We denote the numbers of routes in $\mathcal{R}^{00}$, $\mathcal{R}^{01}$, $\mathcal{R}^{1}$ and $\mathcal{R}^{2}$ by $\pi_{00}, \pi_{01}, \pi_1$, and π_2 respectively; note that $\pi_1 = |W|$. We then partition the matrices Λ and Δ by column as

$$\begin{bmatrix} \Lambda \\ \Delta \end{bmatrix} = \begin{bmatrix} \Lambda_{00} & \Lambda_{01} & \Lambda_1 & \Lambda_2 \\ \Delta_{00} & \Delta_{01} & \Delta_1 & \Delta_2 \end{bmatrix} \tag{B.5}$$

corresponding to routes in $\mathcal{R}^{00}, \mathcal{R}^{01}, \mathcal{R}^{1}$, and $\mathcal{R}^{2}$ respectively. While different choices of q^* may result in different $\mathcal{R}^{00}, \mathcal{R}^{1}, \mathcal{R}^{2}$, they all lead to the same $\mathcal{R}^{01}$ and $\mathcal{R}^{00} \cup \mathcal{R}^{1} \cup \mathcal{R}^{2}$ because the route cost $C(q^*)$ is the same under different choices of q^*.

The following theorem provides a condition under which Ψ is continuously differentiable in a neighborhood of d^*, and a formula to compute the Jacobian matrix $\nabla\Psi(d)$ for d in this neighborhood. The theorem assumes the link cost function c to be continuously differentiable on an open neighborhood of v^* in $\mathbb{R}^{|A|}$, while the domain of the link cost function c here is $\mathbb{R}_+^{|A|}$. There are two ways to handle situations in which some entries in v^* are zero. One way is to extend the domain of c to include an open neighborhood of v^* in $\mathbb{R}^{|A|}$, to evaluate the Jacobian matrix $\nabla c(v^*)$. Another way is to first remove the links $a \in A$ with v_a^* from the network to compute $\nabla\Psi(d)$ using (B.7) for the reduced network, and then add zero rows to $\nabla\Psi(d)$ corresponding to those a back; this works because the links with zero flow remain to have zero flow under small perturbation of d under condition (2) in the theorem. As long as c can be extended to a continuously differentiable function on a neighborhood of v^* in $\mathbb{R}^{|A|}$, these two treatments will give the same Jacobian matrix $\nabla\Psi(d)$ for the original network, because the rows in $\tilde{A}$ and Δ_1 corresponding to links a with $v_a^* = 0$ are all zero.

Theorem B.1. *Suppose Assumption 5.1 holds. Let $d^* > 0$, $v^* = \Psi(d^*)$, and q^* be a user-equilibrium route flow under d^*. Suppose that the link cost function c is continuously differentiable on an open neighborhood V of v^* in $\mathbb{R}^{|A|}$, and write $\tilde{L} = \nabla c(v^*) \in \mathbb{R}^{|A|\times|A|}$. Let $\tilde{A}$ be a matrix whose columns form a basis for the column space of the matrix $[\Delta_2 - \Delta_1\Lambda_2, \Delta_{00} - \Delta_1\Lambda_{00}]$. Suppose that the following two conditions hold:*

(1) The matrix $\tilde{A}^T\tilde{L}\tilde{A}$ is nonsingular.

(2) There exist $u_{\pi_{00}} \in \mathbb{R}^{\pi_{00}}$ and $v_{\pi_2} \in \mathbb{R}^{\pi_2}$ such that $u_{\pi_{00}} > 0$ and

$$(\Delta_{00} - \Delta_1\Lambda_{00})u_{\pi_{00}} + (\Delta_2 - \Delta_1\Lambda_2)v_{\pi_2} = 0. \tag{B.6}$$

Then, Ψ is continuously differentiable on an open neighborhood $\mathcal{O}$ of d^* in $\mathbb{R}^{|W|}_{++}$ with

$$\nabla\Psi(d) = \tilde{A}\left(\tilde{A}^T\nabla c(\Psi(d))\tilde{A}\right)^{-1}\tilde{A}^T\left[-\nabla c(\Psi(d))\Delta_1\right] + \Delta_1 \quad \textit{for each } d \in \mathcal{O}. \tag{B.7}$$

Proof. By [66, Section 2.5], condition (2) here implies that the critical cone K defined in that paper (the critical cone K is different from the set K in this dissertation) is the column space of $\tilde{A}$. Condition (1) here implies Condition 1.1 in [66] holds, with the matrix $\tilde{A}$ here in place of the matrix A in that paper; Condition 1.2 in [66] holds because the matrix B there is empty. Thus, by [66, Theorem 2.3], Ψ is differentiable at d^* under assumptions given in this theorem, and the Jacobian matrix $\nabla\Psi(d^*)$ is given by (B.7) with $\tilde{L} = \nabla c(\Psi(d^*))$. It remains to show that Ψ is continuously differentiable on an open neighborhood $\mathcal{O}$ of d^* in $\mathbb{R}^{|W|}_{++}$. By [66, Proposition 2.2], condition (2) here guarantees the existence of a user-equilibrium route flow $q^{**} \in H(d^*)$ with $\Delta q^{**} = v^*$ and $q_p^{**} > 0$ for all user-optimal routes $p \in \mathcal{R}^{00} \cup \mathcal{R}^1 \cup \mathcal{R}^2$. On the other hand, by [66, Theorem 2.3], there exists a neighborhood D' of d^* in $\mathbb{R}^{|W|}_{++}$ such that $\Psi(\cdot)$ is continuous on D' and $\Psi(d) \in V$ for each $d \in D'$. Now apply [44, Corollary 3.2.5(a)] to the set-valued map

$$(d, v) \to \{q \in \mathbb{R}^{|\mathcal{R}|} : \ q \geq 0, \ \Lambda q = d, \ \Lambda q = v\}$$

and the base point (d^*, v^*, q^{**}), to find an open neighborhood D'' of d^* in D', such that for each $d \in D''$ there exists a user-equilibrium route flow $q(d)$ with $q(d^*) = q^{**}$ and $q_p(d) > 0$ for all $p \in \mathcal{R}^{00} \cup \mathcal{R}^1 \cup \mathcal{R}^2$. Moreover, with the continuity of Ψ, we can shrink D'' further to ensure that non-user-optimal routes under d^* continue to be non-user-optimal under any $d \in D''$. It follows that $q_p(d) = 0$ for all $p \in \mathcal{R}^{01}$ and $d \in D''$. Lastly, by shrinking D'' further if necessary, we can ensure that $\tilde{A}^T \nabla c(\Psi(d))\tilde{A}$ is nonsingular for each $d \in D''$.

Now, apply [66, Theorem 2.3] to an arbitrary $d \in D''$ and the corresponding user-equilibrium arc flow $\Psi(d)$, using $q(d)$ to partition the route set as $\mathcal{R} = \bar{\mathcal{R}}^{00} \cup \bar{\mathcal{R}}^{01} \cup \bar{\mathcal{R}}^1 \cup \bar{\mathcal{R}}^2$. Because $q_p(d) > 0$ for $p \in \mathcal{R}^{00} \cup \mathcal{R}^1 \cup \mathcal{R}^2$, we can let $\bar{\mathcal{R}}^1 = \mathcal{R}^1$. We have $\bar{\mathcal{R}}^{00} = \emptyset$ because all user-optimal routes are used under $q(d)$, $\bar{\mathcal{R}}^{01} = \mathcal{R}^{01}$ because non-user-optimal (user-optimal) routes under d^* remain to be non-user-optimal (user-optimal) under d, and $\bar{\mathcal{R}}^2 = \mathcal{R}^2 \cup \mathcal{R}^{00}$ consists of routes p with $q_p(d) > 0$ not in $\mathcal{R}^1$. Let $\bar{\Lambda}_2$ and $\bar{\Delta}_2$ be submatrices of Λ and Δ corresponding to routes in $\bar{\mathcal{R}}^2$; the matrix $[\bar{\Delta}_2 - \Delta_1 \bar{\Lambda}_2]$ contains the same columns as $[\Delta_2 - \Delta_1 \Lambda_2, \Delta_{00} - \Delta_1 \Lambda_{00}]$, so columns of the same matrix $\tilde{A}$ defined in the statement of the present theorem form a basis for the column space of $[\bar{\Delta}_2 - \Delta_1 \bar{\Lambda}_2]$. The first condition for Ψ to be differentiable at d holds because $\tilde{A}^T \nabla c(\Psi(d))\tilde{A}$ is nonsingular for each $d \in D''$ by the choice of D'', and the second condition also holds because $\bar{\mathcal{R}}^{00}$ is empty. Consequently, Ψ is differentiable at d with

$$\nabla \Psi(d) = \tilde{A}\Big(\tilde{A}^T \nabla c(\Psi(d))\tilde{A}\Big)^{-1} \tilde{A}^T \Big[-\nabla c(\Psi(d))\Delta_1 \Big] + \Delta_1,$$

which is continuous with respect to d due to the continuous differentiability of c and the continuity of Ψ. Letting $\mathcal{O} = D''$ completes the proof. □

Appendix C

DEFINITION OF F1-SCORE, ACCURACY, AND RMSE

To define F1-score, it is necessary to introduce the confusion matrix that is widely used in statistical classification. As shown in Table C.1, for a given network, we define the set of insignificant/significant OD pairs based on the true OD demand matrix (denoted as W_1 and $W \setminus W_1$ in this dissertation) as the ground truth, while the set of the insignificant/significant OD pairs under the estimated OD demands as the prediction.

Table C.1: Confusion matrix of OD quasi-sparsity consistency

		Partitions in the true OD matrix (true)	
		Insignificant(+)	Significant(-)
Partitions in the estimated OD matrix (prediction)	Insignificant(+)	True positives (TP)	False positives (FP)
	Significant(-)	False negatives (FN)	True negatives (TN)

The confusion matrix can be used to define the True Positive Rate (TPR, or recall/sensitivity) and precision, which can then be used to define F1-score. We give their formulas as follows.

$$TPR = \frac{TP}{TP + FN} \tag{C.1}$$

$$precision = \frac{TP}{TP + FP} \tag{C.2}$$

$$F_1 = \frac{2 \times precision \times TPR}{precision + TPR} \tag{C.3}$$

TPR indicates the portion of insignificant OD pairs in the true OD demands that still keep insignificant in the estimated OD demands, as calculated in (C.1). Precision means the portion of insignificant OD pairs in the estimated OD demands that are insignificant in the true OD demands, as calculated in (C.2). F1-score is the harmonic mean of TPR and precision, which is widely used as the performance measure for classification algorithms since it is less sensitive to imbalanced dataset by considering both TPR and precision. As the OD demand dataset is highly imbalanced (with a large portion of insignificant OD pairs and a small portion of significant OD pairs), F1-score is an ideal measure to evaluate the quasi-sparsity consistency of the proposed QSOD models. The classification becomes better as the F1-score gets closer to 1, indicating more OD pairs in the estimated OD demand matrix will remain insignificant or significant as the true OD matrix.

Accuracy is defined in (C.4) based on the confusion matrix.

$$Accuracy = \frac{TP + TN}{Population} \tag{C.4}$$

Lastly, Root Mean Square Error (RMSE) is also used to evaluate how close the estimated value is to the "ground truth" value:

$$RMSE(d, \bar{d}) = \sqrt{\frac{\sum_{w \in W} (d_w - \bar{d}_w)^2}{|W|}} \tag{C.5}$$

where d is the estimated OD demands and $\bar{d}$ is the true OD demands.

Appendix D

DISCUSSION ABOUT THE SOLUTION CLOSENESS OF QSOD AND GLS

Here we discuss, for the fixed-mapping models, why the solution of GLS gets much closer to the QSOD solution after adding proper weights, when compared with the OLS solution (with no weight added). Denote (d_1, v_1) and (d_2, v_2) as the unique optimal solution and f_1 and f_2 as the objective function for the fixed-mapping QSOD model and OLS model, respectively. According to [93, Theorem 2], if both f_1 and f_2 are strongly monotone with modulus m and Lipschitzian with modules M, we have the following inequality:

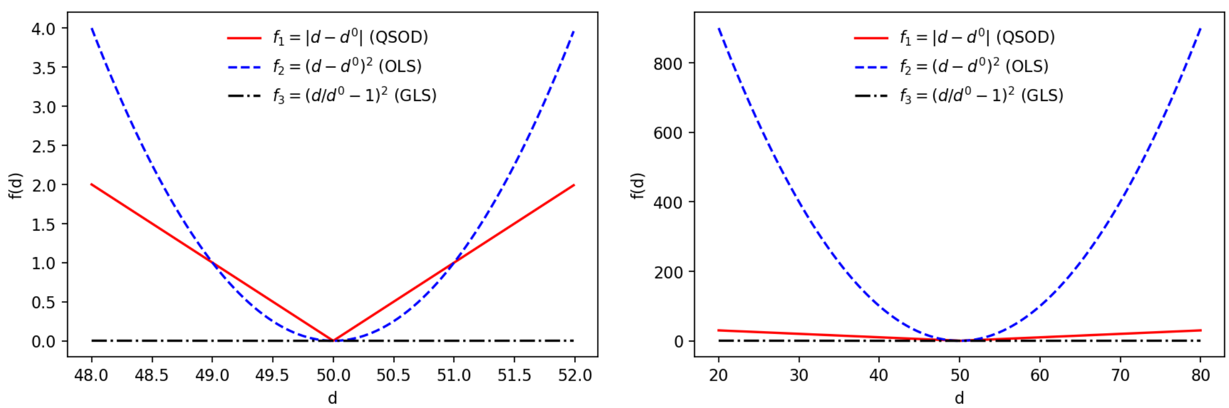
$$||(d_1, v_1) - (d_2, v_2)|| \leq \theta^{-1}\mu||f_1(d_1, v_1) - f_2(d_1, v_1)|| \tag{D.1}$$

where $\mu = M^{-2}m$ and $\theta = (1 - [1 - (m/M)^2]^{1/2})^{-1}$. (D.1) in fact gives an upper bound for the difference between the optimal solutions of the two models (QSOD and OLS), and such bound is only determined by (d_1, v_1), f_1, f_2, θ and μ, where the last two parameters rely on the collection of single-valued functions F which contain f_1 and f_2. Assume such function collection F also includes the GLS objective f_3 and denote the optimal solution for the GLS model as (d_3, v_3). Similarly, we will have:

$$||(d_1, v_1) - (d_3, v_3)|| \leq \theta^{-1}\mu||f_1(d_1, v_1) - f_3(d_1, v_1)|| \tag{D.2}$$

Notice that the RHS of (D.1) and (D.2) only differs in the OLS and GLS objective function (f_2 and f_3 respectively). If we only illustrate the problem in one dimension (as well as ignore the link flow deviations as they are similar to the OD demand deviations), then the QSOD, OLS, and GLS objective will have the following form respectively: $f_1 = |d - d^0|$, $f_2 = (d - d^0)^2$, and $f_3 = (d - d^0)^2/(\epsilon_1 d^0)^2$. Further we can ignore ϵ_1 in f_3 as ϵ_1 is a constant for all OD pairs. Therefore f_3 can be written as $(d/d^0 - 1)^2$. Taking $d^0 = 50$, Figure D.1

shows the plots of the three functions in a range of d that include d^0. It can be seen that in the range that closes to d^0 (e.g. $d \in (48, 52)$), $|f_1 - f_2|$ is less than or equal to $|f_1 - f_3|$. When d deviates a lot from d^0 (as shown in Figure D.1b), $|f_1 - f_2|$ will be much greater than $|f_1 - f_3|$. In the OD estimation problem, once the estimate d (or v) goes far from its prior value d^0 (or v^0), the RHS of (D.2) will be much less than that of (D.1), which indicates that $||(d_1, v_1) - (d_3, v_3)||$ has a small upper bound compared with $||(d_1, v_1) - (d_2, v_2)||$. This can partially explain the solution closeness between QSOD and GLS (compared with QSOD and OLS).



(a) $d \in (48, 52)$ (b) $d \in (20, 80)$

Figure D.1: Comparison across QSOD, OLS, and GLS objective function (one-dimension, $d^0 = 50$)