Distributionally Robust Origin Destination Demand Estimation

Jingxing Wang\textsuperscript{a}, Jun Song\textsuperscript{b}, Chaoyue Zhao\textsuperscript{b,}\textsuperscript{*}, and Xuegang (Jeff) Ban\textsuperscript{a}

\textsuperscript{a}Department of Civil and Environmental Engineering, University of Washington, Seattle Washington, United States, 98195

\textsuperscript{b}Department of Industrial and Systems Engineering, University of Washington, Seattle Washington, United States, 98195

\textbf{Keywords:} OD demand estimation, distributionally robust optimization, quasi-sparsity

Gaining a good understanding of the travel demands of a city or region is extremely important for many transportation applications. For stochastic origin-destination (OD) estimation problems, an accurate distribution assumption or observation of OD estimates or data is usually desired but not always available. In this paper, we establish a novel two-stage OD estimation framework based on distributionally robust optimization (DRO) and quasi-sparsity property of large-scale OD demand matrices. The proposed two-stage Distributionally Robust Quasi-Sparsity OD estimation (DR-QSOD) model does not require an accurate or complete distribution assumption of estimates/data. Numerical results demonstrate that DR-QSOD model outperforms stochastic QSOD model in estimating OD demands when the distribution assumption of data is biased. This paper also discusses two different approaches to solve the DR-QSOD model as well as compares their computational efficiency. In addition, DR-QSOD model is shown to keep relatively high quasi-sparsity consistency, which also brings lots of meaningful practical insights.

\section{Introduction}

A good understanding of the travel demands of a city or region is extremely vital for many transportation applications. In order to obtain high-quality origin-destination (OD) demands, many OD estimation methods, i.e. incorporating newly-acquired field data (e.g. traffic link flows) to improve prior OD demands (usually from local planning agencies), have been proposed over the last several decades; see (Bera and Rao, 2011) for a more recent review of OD estimation methods. For a large-scale urban road network, recent studies focus on the sparsity property (Sanandaji and Varaiya, 2014; Menon et al., 2015; Wen et al.,

\textsuperscript{*} Corresponding author. E-mail address: cyzhao@uw.edu.
2018) and quasi-sparsity property (Wang et al., 2022) of its OD demand matrix. The former (i.e., the sparsity property) indicates that the demand of most OD pairs of the network is strictly zero, while the latter (i.e., the quasi-sparsity property) refers to the fact that most OD pairs own very small (i.e., insignificant) travel demands and those demands only contribute to a small portion of total demands in the network. Since this quasi-sparsity property only requires that most OD pairs have very small demands (rather than exactly zero), it is less restrictive than the sparsity property of OD matrices. Wang et al. (2022) formally introduced this quasi-sparsity property and further proposed quasi-sparsity OD (QSOD) estimation models, with extensive discussions about the mathematical properties and numerical results of the QSOD models (Wang et al., 2022). More importantly, as shown in (Wang et al., 2022), the OD demands of city-scale or region-scale real-world transportation networks likely possess the quasi-sparsity property, necessitating further explorations/extensions of quasi-sparsity when developing OD estimation models.

The QSOD models proposed in (Wang et al., 2022) are purely deterministic, ignoring the stochastic features of OD estimation, due to, e.g., the possible (day to day) variability of demands, observations of traffic, or even traffic assignment rules. On the other hand, stochastic modeling of 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 (Ma and Qian, 2018; Shao et al., 2014). A more detailed review of existing stochastic OD estimation models is presented in the next section. 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/estimate in practice. Furthermore, if such assumptions are not satisfied, the resulting OD estimates can be degraded and may not be properly used in real applications. Noteworthy is that a few factors can account for the stochasticity of OD estimation problems, e.g. travel demand variations, uncertainty of the observed data, and/or traffic assignment rules, to name a few. For the purpose of simplicity, this paper only considers the stochastic property of OD demands and observed traffic data. We leave the investigation of the proposed estimation problem with stochastic traffic assignment for future study.

Inspired by the stochastic characteristics of OD estimation problems as well as the potential limited data availability, in this study we propose to 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 \( x \) that minimizes the expected cost under the most adversarial probability distribution, i.e., \( \min_{x \in \mathcal{X}} \max_{P \in \mathcal{P}} \mathbb{E}_{P}[f(x, \xi)] \), where the distribution \( P \) of the random parameter \( \xi \) is not precisely known but assumed to belong to an ambiguity set \( \mathcal{P} \). 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 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. It is noted that the proposed DR-QSOD model is not designed for generating the distribution of OD demand estimates: by considering the data distributional uncertainty, the DR-QSOD model gives the overall estimated OD demands (the ”nominal” demand) and realized day-to-day OD demands (the daily demand) based on daily observed traffic data. Stochastic OD models, on the other hand, usually gives the mean (and variance) of the distribution 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.
Table 1: Comparison between DR-QSOD and stochastic OD models

<table>
<thead>
<tr>
<th>Attributes</th>
<th>DR-QSOD</th>
<th>Stochastic OD models</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assumption</td>
<td>OD demand matrix is quasi-sparse, and distribution parameters follow an ambiguity set</td>
<td>Either input data or output OD estimate, or both follow certain probability distributions (Abareshi, 2020; Hazelton, 2000, 2001, 2003; Li, 2005; Parry and Hazelton, 2012)</td>
</tr>
<tr>
<td>Input</td>
<td>Prior OD demands, day-to-day link flows</td>
<td>Prior OD demands, day-to-day link flows (Hazelton, 2000, 2001, 2003; Li, 2005), routing data (Parry and Hazelton, 2012)</td>
</tr>
<tr>
<td>Output</td>
<td>Overall OD demands and daily OD demands</td>
<td>Only mean (Hazelton, 2000, 2001, 2003; Li, 2005) or mean and variance of OD demand estimates (Ma and Qian, 2018; Parry and Hazelton, 2012; Shao et al., 2014; Yang et al., 2019)</td>
</tr>
<tr>
<td>Solution property</td>
<td>Quasi-sparsity</td>
<td>Not keeping quasi-sparsity</td>
</tr>
</tbody>
</table>

On the other hand, the proposed DR-QSOD model in this paper is a major extension of the deterministic QSOD model in (Wang et al., 2022). The deterministic QSOD framework defines the quasi-sparsity and explores its solution property and quasi-sparsity consistency for both fixed-mapping and bi-level QSOD models. Such deterministic QSOD framework however does not consider the stochastic property of OD estimates or observed data. By leveraging the key advantage of distributional uncertainty of data, the proposed DR-QSOD here allows us to consider quasi-sparsity and stochasticity at the same time. Given the unique property of QSOD model in real applications (Wang et al., 2022), i.e., one only requires the prior OD demands being representative in terms of quasi-sparsity rather than being close to the true OD demands, DR-QSOD further enriches the QSOD framework, making it more practically useful when applying QSOD models and data is inaccurate or incomplete. The numerical experiments demonstrate that when the data distribution is biased, the DR-QSOD model can outperform its stochastic counterpart in terms of OD estimation accuracy; see Section 5 for details.

In this paper, we formulate the quasi-sparsity OD estimation problem as a two-stage DRO problem, and then further reformulate the two-stage DR-QSOD model into a tractable form and apply the Benders Decomposition (BD) method to solve the reformulated problem. Moreover, we integrate the Column-and-Constraint Generation (CCG) method (Zeng and Zhao, 2013) into the BD framework to accelerate convergence. Numerical results show that, compared with its stochastic counterpart, i.e, stochastic QSOD model, the proposed DR-QSOD model can produce a higher-quality OD estimates with lower estimation errors given the distribution of input data is not accurate. Furthermore, by comparing convergence performance of Benders Decomposition with Column-and-Constriant Generation (BD-CCG) with that of the pure BD method, the study further shows that there exists a trade-off between the running time and convergence speed (in terms of relative error change per iteration). It is worthwhile to note that although OD estimation has been attracting much attention and extended to much complex scenarios, e.g. dynamic OD estimation, this paper only focuses on static DR-QSOD models. Dynamic DR-QSOD models can be explored in the
future based on the theoretical results of the static DR-QSOD model in this paper, which be left for future research.

The key contributions of this paper are summarized as follows.

- **DRO modeling technique is integrated with the QSOD model to consider both data distributional uncertainty and the quasi-sparsity property of OD demands.** The use of DRO relaxes the strict distributional assumptions employed in stochastic models, making it possible to consider both the OD stochasticity and distributional robustness at the same time. To our best knowledge, this is the first study to apply the DRO modeling technique on OD estimation problems, which provides an important alternative to the stochastic optimization methods for OD estimation especially when the distribution of OD demands or data is not available. Integrating DRO with the QSOD model also ensures quasi-sparsity in the resulting OD estimates.

- **The proposed DR-QSOD framework can provide two-stage OD demand estimator with less restrictive distributional assumptions, which is practically useful for real applications.** The two-stage DR-QSOD model can produce a nominal OD demand (first-stage decision) accounting for the distributional uncertainty of observed data, and the daily OD estimates (second-stage decision) corresponding to daily observed data, which provides richer and more useful information for decisions and modeling based on OD demands. In addition, the DR-QSOD framework relaxes the distributional assumptions usually imposed on OD demands or observed data in stochastic OD estimation models, which may be practically important for certain applications.

- **A Benders Decomposition (BD) framework is proposed to solve the two-stage DR-QSOD model.** This paper illustrates the detailed steps to build a tractable form of a two-stage DRO model for OD estimation, and proposes two frameworks BD and BD-CCG to solve the tractable form. Numerical tests are conducted to illustrate how different framework setups may affect the convergence performance. This provides practical implications for real-world applications.

This paper is organized as follows. Section 2 reviews the studies of stochastic OD estimation and gives an introduction to distributionally robust optimization. In Section 3 and Section 4, we formally formulate the two-stage DR-QSOD model and demonstrate its solution algorithm, respectively. Section 5 gives numerical tests based on a transportation network. Section 6 summarizes the findings of this paper and points out the future work.

## 2 Literature Review

OD estimation problems have been widely studied in the past few decades, which can be broadly categorized as static OD estimation and dynamic OD estimation. The former considers the long-term average network traffic state, while the latter considers the dynamic (temporal) features of network traffic flows. This paper focuses on static OD estimation problems, which we simply refer to as the “OD estimation problems”. Most existing (static) OD estimation problems belong to the category of deterministic optimization models, i.e. producing a high-quality single point OD demand estimator given the observed data (e.g. traffic counts) and prior information. For a more complete review of OD estimation and especially deterministic OD estimation, one can refer to (Bera and Rao, 2011; Castillo et al., 2015). In the era of Information Technology, the increasing data availability provides great opportunities to account for OD estimation with uncertainty, either in the estimate itself, or the input data, or both. In this section, we start with reviewing an important
category of studies considering such uncertainty, i.e., stochastic OD estimation. Then we will introduce the DRO technique used in this paper, and give a detailed comparison between stochastic optimization and DRO.

2.1 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, Hazelton (2000) proposed a Maximum Likelihood Estimation (MLE) method (Hazelton, 2000) and a Bayesian inference method (Hazelton, 2001) for OD estimation. Following the same assumption, Hazelton (2003) continued to employ a Poisson model to conduct OD estimation (Hazelton, 2003), where the first- and second-order statistical property of link flows were used. Li (2005) 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 (Li, 2005). Parry and Hazelton (2012) proposed an MLE method to estimate OD demands using link counts and sporadic routing data (Parry and Hazelton, 2012). 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. Abareshi (2020) 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 (Abareshi, 2020). Yang et al. (2018) proposed a model to estimate the mean value of OD demands as well as improve network identifiability using multi-day observation sets (Yang et al., 2018).

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 (Clark and Watling, 2005). The daily fluctuations of OD demand also lead to significant impacts to traffic system performance, urban traffic management, route choice of travelers, and etc (Ma and Qian, 2018; Shao et al., 2014). By considering the variations of OD demands, many recent studies started to consider transportation network reliability and uncertainty issues (Shao et al., 2014). For instance, by assuming OD demands follow mutually independent normal distributions, Chen et al. (2002) evaluated travelers’ risk-taking behaviors under travel demand uncertainties (Chen et al., 2002). Clark and Watling (2005) developed a stochastic modeling framework to estimate network travel time reliability with mutually independent Poisson distributions of OD demands (Clark and Watling, 2005). Shao et al. (2006) proposed a reliability-based user equilibrium model under the assumptions that OD demands and path flows are mutually independently normal distributions (Shao et al., 2006). Nakayama and Watling (2014) 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 (Nakayama and Watling, 2014). 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, Shao et al. (2014) used a weighted least squared method to estimate the means and covariance of travel demands (Shao et al., 2014).
Ma and Qian (2018) 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 (Ma and Qian, 2018). 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. Yang et al. (2019) proposed a Generalized Method of Moment (GMM)-based framework to infer the probability density function (pdf) of OD demand variables (Yang et al., 2019). 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, one can refer to recent studies (Shao et al., 2014; Ma and Qian, 2018; Yang et al., 2019).

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.

### 2.2 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 (Scarf, 1958) 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 (Delage and Ye, 2010a; Goh and Sim, 2010a; Wiesemann et al., 2014). 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 (1) presents the formulation for SO, where $x$ represents the decision variable, $\mathcal{X}$ denotes the feasible region for the decision variable, and $\xi$ is a random variable defined on a probability space $(\Omega, \sigma(\Omega), \mathbb{P})$. Here, $\Omega$ is the sample space of $\xi$, $\sigma(\Omega)$ denotes the $\sigma$-algebra of the sample space, and $\mathbb{P}$ represents the probability distribution of $\xi$. The objective of the SO problem is to minimize the expected value of the function $f(x, \xi)$ 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), exhibits a certain level of similarity with SO. It aims to optimize the worst-case function $\sup_{\mathbb{P} \in \mathcal{G}} \mathbb{E}_{\mathbb{P}} f(x, \xi)$ over $x \in \mathcal{X}$. In this context, the term “worst” indicates the supremum function, representing the most unfavorable outcome among all possible probability distributions in the ambiguity set $\mathcal{G}$.
The key difference between SO and DRO lies in how much probability information of variable $\xi$ we have and the way we treat its uncertainties. Different from SO that predefines the probability distribution $P$, DRO (2) assumes the probability distribution $P$ falls into the set $\mathcal{D}$, where $\mathcal{D}$ is a predefined ambiguity set that is assumed to include the true probability distribution of $\xi$. Even both models need to take the expectation over $\xi$ in the objective, DRO relaxes the assumption from a single fixed probability distribution $P$ to a family of probability distributions $P \in \mathcal{D}$, which has very important implication in practice. This is because the exact distribution $P$ in SO is often hard to obtain or estimate with empirical data (Shang and You, 2018), 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 (Delage and Ye, 2010b; Zymler et al., 2011; Goh and Sim, 2010b). These sets can be easily constructed using first or second moment information of the data. Another class of ambiguity sets is based on discrepancy, such as $\phi$ divergences (Hu and Hong, 2012; Ben-Tal et al., 2013) and the Wasserstein metric (Esfahani and Kuhn, 2018; Zhao and Guan, 2018), which centers around a finite-support nominal distribution constructed from data samples.

Though the discrepancy-based ambiguity sets can provide more accurate representations of the distance between the nominal and true distributions, like capturing differences in the shape, spread, and tail behavior of distributions, the computation of the discrepancy often involves solving an optimization problem (e.g., the optimal transport problem) that can be challenging, especially in high-dimensional spaces. In addition, determining an appropriate discrepancy metric and threshold for defining closeness pose another dimension of challenges. The selection of an unsuitable metric or threshold may potentially result in an ambiguity set that is either overly restrictive or too permissive. This, in turn, may lead to suboptimal or excessively conservative solutions. Conversely, moment-based ambiguity sets typically involve a smaller set of parameters (moments) compared to discrepancy-based ambiguity sets that consider the entire distribution. This reduction in dimensionality can simplify the optimization process and make it more feasible for large-scale problems. Therefore, in the context of this paper, we leverage the first moment-based confidence set to construct the ambiguity set based on observed linked flows.

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 our paper, 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 (Babaei et al., 2020; Zhao and Jiang, 2017), supply chain optimization (Liu et al., 2019; Xin and Goldberg, 2015), and chemical engineering (Shang and You, 2018), among others. For a more recent survey on this topic, readers may refer to (Rahimian and Mehrotra, 2019). 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 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 the following sections, 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.

![Diagram of stochastic optimization vs. distributionally robust optimization](image)

**Figure 1**: In SO, the distribution of uncertain parameters in OD estimation problem is predetermined; In DRO, the distribution of uncertain parameters in OD estimation problem is allowed to vary ambiguously within a predefined set.

### 3 Model Formulation

#### 3.1 Preliminaries

This study focuses on 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 stochasticity of the physical systems and the available data (e.g., the stochastic optimization model in the literature or the DRO model employed in this paper). 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.

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 paper 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) 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 paper in Table 2.

Table 2: Table of Notation

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>( d )</td>
<td>nominal demand estimate, “here-and-now” decision for the OD demands</td>
</tr>
<tr>
<td>( t )</td>
<td>daily demand estimate, “wait-and-see” decision for the OD demands</td>
</tr>
<tr>
<td>( \lambda )</td>
<td>weighting parameter in DR-QSOD</td>
</tr>
<tr>
<td>( z_1, z_2, z_3 )</td>
<td>auxiliary variables</td>
</tr>
<tr>
<td>( \xi )</td>
<td>daily network-level traffic link flow</td>
</tr>
<tr>
<td>( A )</td>
<td>dimension indices set for daily network-level traffic link flow</td>
</tr>
<tr>
<td>( W )</td>
<td>dimension indices set for OD demands</td>
</tr>
<tr>
<td>( d^0 )</td>
<td>prior OD demand</td>
</tr>
<tr>
<td>( \mathbb{P} )</td>
<td>probability distribution of ( \xi )</td>
</tr>
<tr>
<td>( \mathbb{D} )</td>
<td>ambiguity set for ( \mathbb{P} )</td>
</tr>
<tr>
<td>( \hat{Q} )</td>
<td>fixed-mapping matrix between OD demands and link flows</td>
</tr>
<tr>
<td>( \bar{\mu} )</td>
<td>mean value of observed link flow</td>
</tr>
<tr>
<td>( \hat{S}, A^*, B, C, E, F, G )</td>
<td>pre-specified coefficient matrices</td>
</tr>
<tr>
<td>( \bar{a}, b, c, e, f, g )</td>
<td>pre-specified coefficient vectors</td>
</tr>
<tr>
<td>( \Omega )</td>
<td>support of ( \xi )</td>
</tr>
<tr>
<td>( M^+ (\Omega) )</td>
<td>set of all probability distributions on a ( \sigma )-algebra of ( \Omega )</td>
</tr>
<tr>
<td>( \xi_{\text{min}} )</td>
<td>predefined lower bound on ( \xi )</td>
</tr>
<tr>
<td>( \xi_{\text{max}} )</td>
<td>predefined upper bound on ( \xi )</td>
</tr>
<tr>
<td>( y )</td>
<td>first-stage decision variable of two-stage DRO</td>
</tr>
<tr>
<td>( x )</td>
<td>second-stage decision variable of two-stage DRO</td>
</tr>
<tr>
<td>( Q(y, \xi) )</td>
<td>second-stage objective</td>
</tr>
<tr>
<td>( \gamma, \beta )</td>
<td>dual variables in dual formulation of two-stage DRO</td>
</tr>
<tr>
<td>( \zeta, \eta, \delta )</td>
<td>dual variables in dual formulation of ( Q(y, \xi) )</td>
</tr>
<tr>
<td>( \bar{v} )</td>
<td>true link flows</td>
</tr>
<tr>
<td>( \bar{d} )</td>
<td>“true” OD demands</td>
</tr>
<tr>
<td>( \varepsilon_0 )</td>
<td>threshold for insignificant OD demand</td>
</tr>
<tr>
<td>( \varepsilon_1 )</td>
<td>relative error in prior OD demands</td>
</tr>
<tr>
<td>( \varepsilon_2 )</td>
<td>relative error in observed link flows</td>
</tr>
</tbody>
</table>
3.2 Model setup

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

\[
\begin{align*}
\min_d & \quad \frac{1}{\lambda} \|d - d^0\|_1 + \sup_{P \in \mathcal{P}} E_P[Q(d, \xi)] \\
\text{s.t.} & \quad d \geq 0,
\end{align*}
\]

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:

\[
\begin{align*}
\min_t & \quad \|t - d\|_1 + \|\hat{Q} \cdot t - \xi\|_1 \\
\text{s.t.} & \quad t \geq 0,
\end{align*}
\]

where \(\hat{Q} \in \mathbb{R}^{|A| \times |W|}\) represents the fixed-mapping matrix between OD demands and link flows. Here \(d\) and \(t\) are the first-stage and second-stage decision variables, which represent, respectively, the nominal demand estimate and the daily demand estimate corresponding to each daily traffic data input. \(Q(d, \xi)\) 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 \(\xi\). Further, if \(d = d^0\), then \(Q(d, \xi)\) will be the same as the deterministic fixed-mapping QSOD model in (Wang et al., 2022). \(\lambda\) in (3) is a parameter that needs to be tuned, which can be interpreted as how reliable the prior OD demand \(d^0\) is. If \(\lambda\) 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 \(d^0\). Note that for Problem (3) and (4), there are no additional constraints except the non-negativity conditions, as the traffic assignment relationship has been written into the objective function.

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 \(\xi\). Therefore, in this paper, we treat \(\xi\) as a random parameter with distributional uncertainty. We let \(\mathbb{P}\) be the probability distribution of the link flow \(\xi\), and \(\mathcal{D}\) be the ambiguity set of \(\mathbb{P}\) defined in (5).

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

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 in this paper is defined based on the first moment of the distribution, and the expectation of \(\bar{S}\xi\) over \(\mathbb{P}\) is upper-bounded by \(\bar{\mu}\). \(M^+(\Omega)\) here stands for the set of all probability distributions on a \(\sigma\)-algebra of \(\Omega\), and \(\Omega\) is the support of \(\xi\) defined below:

\[
\Omega := \{\xi \in \mathbb{R}^{|A|} : \xi_{\text{min}}^\alpha \leq \xi^\alpha \leq \xi_{\text{max}}^\alpha, \forall \alpha \in A\},
\]

where \(\xi_{\text{min}}\) and \(\xi_{\text{max}}\) are predefined lower and upper bounds of \(\xi\).

Notice that the ambiguity set \(\mathcal{D}\) basically defines a family of probability \(\mathbb{P}\), and all possible probability distributions of \(\xi\) 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 $\xi$, therefore the
modeling process of DRO will be less restrict than the typical stochastic optimization.

Similar to the deterministic QSOD model in (Wang et al., 2022), we use the $L_1$-norm in the objectives
of (3) and (4). 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 $\alpha$ matrices, but also make the model easier to solve. Denote $\alpha$ 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 (3) and (4), and transform the two-stage DR-QSOD model into the following form:

$$\min_{d,z_1} \frac{1}{\lambda} \sum_{w \in W} z_w^w + \sup_{P \in \mathcal{G}} E_P[Q(d, z_1, \xi)]$$

s.t. $d_w^w \geq 0, \quad \forall w \in W,$

$$z_1^w \geq d_w^w - (d_0)^w, \quad \forall w \in W,$$

$$z_1^w \geq -d_w^w + (d_0)^w, \quad \forall w \in W,$$

where

$$Q(d, z_1, \xi) = \min_{t, z_2, z_3} 0 \cdot \sum_{w \in W} t_w^w + 1 \cdot \sum_{w \in W} z_2^w + 1 \cdot \sum_{\alpha \in A} z_3^\alpha$$

s.t. $t_w^w \geq 0, \quad \forall w \in W,$

$$z_2^w \geq t_w^w - d_w^w, \quad \forall w \in W,$$

$$z_2^w \geq -t_w^w + d_w^w, \quad \forall w \in W,$$

$$z_3^\alpha \geq [\hat{Q}_t]^\alpha - \xi^\alpha, \quad \forall \alpha \in A,$$

$$z_3^\alpha \geq -[\hat{Q}_t]^\alpha + \xi^\alpha, \quad \forall \alpha \in A.$$
Note that problem (9) and (10) 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 $\xi$, i.e. $\xi_i$, the link flow for day $i$. The key difference between the two-stage model and single-stage DRO model (2) is that the second-stage problem can yield to an estimate corresponding to each realized input data $\xi_i$, which makes it possible to give a daily OD demand as the change of daily link flows. In terms of constraints of the model, the constraint $A^* y \leq b$ in Problem(9) indicates the non-negativity of $d$ and equation $z = \|d - d^0\|_1$. For the constraints of Problem (8), $Bx + Cy \geq e$ indicates the inequality between second-stage decision variable $x$ and first-stage decision variable $y$; $Ex + F \xi \geq f$ indicates the inequality between second-stage decision variable $x$ and uncertain observed data $\xi$; $Gx \geq g$ implies the non-negativity of $t$.

Besides, it is valuable to briefly discuss the connections between the two-stage DR-QSOD model with the deterministic QSOD model and stochastic models (to save space and keep the continuation of the discussion, we present the formulation of the deterministic QSOD model and stochastic QSOD model in Appendix A). First, the two-stage DR-QSOD model is more general than the fixed-mapping QSOD model, a simplified deterministic QSOD model in (Wang et al., 2022); see Eq (22). If we set a relatively small value for $\lambda$ in (3), 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; see Eq (23) for its formulation. 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.

4 Solution Algorithm

4.1 Tractable reformulation

Equation (9)-(10) 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 (9) 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 1.

**Theorem 1.** The two-stage DRO model (9)-(10) is equivalent to the following (tractable) min-max problem (11).

$$\min_{y, \tau \geq 0} \quad \bar{a}^T y + \bar{\mu}^T \tau + \max_{\xi \in \Omega} (Q(y, \xi) - \tau^T S \xi)$$

$$s.t. \quad A^* y \leq b.$$  

**Proof.** Note that the second part of the objective of (9) requires to solve the worst-case expectation $\sup_{\mathbb{P} \in \mathbb{D}} E_{\mathbb{P}}[Q(y, \xi)]$ under the ambiguity set $\mathbb{D}$, and $Q(y, \xi)$ itself is an optimization problem. First we
rewrite \( \sup_{P \in \mathcal{D}} E_P[Q(y, \xi)] \) as the following optimization problem, according to the property and definition of probability distribution and expectation.

\[
\sup_{P \in \mathcal{D}} E_P[Q(y, \xi)] = \max_{\mathcal{P}} \int_{\Omega} Q(y, \xi) \mathbb{P}(d\xi)
\]

\[
s.t. \quad \int_{\Omega} \hat{S}_{\xi} \mathbb{P}(d\xi) \leq \bar{\mu}
\]

\[
\int_{\Omega} \mathbb{P}(d\xi) = 1
\]

Then by using the standard duality theory, we can write the dual problem of (12) as follow:

\[
\min_{\tau \geq 0, \beta} \bar{\mu}^T \tau + \beta
\]

\[
s.t. \quad \tau^T \hat{S}_{\xi} + \beta \geq Q(y, \xi), \forall \xi \in \Omega
\]

where \( \tau \) and \( \beta \) are dual variables associated with the two constraints in (12) respectively. Substituting \( \sup_{P \in \mathcal{D}} E_P[Q(y, \xi)] \) with (13), we can rewrite (9) as follows:

\[
\min_{y} \bar{a}^T y + \min_{\tau \geq 0, \beta} (\bar{\mu}^T \tau + \beta)
\]

\[
s.t. \quad A^* y \leq b
\]

\[
\tau^T \hat{S}_{\xi} + \beta \geq Q(y, \xi), \forall \xi \in \Omega
\]

Now the objective of (14) is a min-min problem, so we write it as the following optimization problem,

\[
\min_{y, \tau \geq 0, \beta} (\bar{a}^T y + \bar{\mu}^T \tau + \beta)
\]

\[
s.t. \quad A^* y \leq b
\]

\[
\tau^T \hat{S}_{\xi} + \beta \geq Q(y, \xi), \forall \xi \in \Omega
\]

As the constraint \( \beta \geq Q(y, \xi) - \tau^T \hat{S}_{\xi} \) must be satisfied for \( \forall \xi \in \Omega \), we further write the above problem (15) as (11). Therefore, the two-stage model is equivalent to the min-max problem.

\[\square\]

### 4.2 Benders decomposition

In this section, we will apply Benders Decomposition (BD) (Zhao and Jiang, 2017) to solve the problem in (11). Firstly, we rewrite (11) as follow:

\[
\min_{y, \tau \geq 0, \beta} (\bar{a}^T y + \bar{\mu}^T \tau + \beta)
\]

\[
s.t. \quad A^* y \leq b
\]

\[
\beta \geq Q(y, \xi) - \tau^T \hat{S}_{\xi}, \forall \xi \in \Omega
\]

Notice that as \( \xi \) in the second constraint of (16) is a continuous parameter and has to fall into the support of \( \Omega \), it is unwise or impossible to enumerate all values of \( \xi \) to construct the realized constraints. However, we can employ BD method to solve the problem. Given a realization of \( \xi \), say \( \xi' \), we call the constraint \( \beta \geq Q(y, \xi') - \tau^T \hat{S}_{\xi'} \) as a Benders cut pertaining to \( \xi' \). 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 $\epsilon$, iteration limit $L$, set of Benders’ cuts $CUT := \emptyset$.

2. For $l = 1, \cdots, L$, repeat the steps:
   (a) Solve the master problem
   \[
   \min_{y, \tau \geq 0, \beta} \quad a^T y + \bar{\mu}^T \tau + \beta \\
   \text{s.t.} \quad A^* y \leq b
   \]
   with the current set of Benders’ cuts in $CUT$ as additional constraints. Record optimal solutions $(y^l, \tau^l, \beta^l)$, and set $LB$ equal to the optimal value.
   (b) Solve the separation problem
   \[
   \max_{\xi \in \Omega} \{ Q(y^l, \xi) - (\tau^l)^T \tilde{S} \xi \}
   \]
   Record optimal solution $\xi^l$ 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 $\beta \geq Q(y^l, \xi^l) - (\tau^l)^T \tilde{S} \xi^l$ 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 (18) we consider the dual problem of $Q(y, \xi)$ as follow:
\[
Q(y, \xi) = \max_{\xi, \eta, \delta} \quad (e - Cy)^T \xi + (f - F \xi)^T \eta + g^T \delta \\
\text{s.t.} \quad B^T \xi + E^T \eta + G^T \delta \leq c, \\
\xi, \eta, \delta \geq 0
\]
where the dual variables $\xi$, $\eta$, and $\delta$ correspond to the three constraints in (10), respectively. As $Q(y, \xi)$ has been converted into its dual form in terms of a maximization problem, we can plug (19) into (18), and recast the separation problem as:
\[
\max_{\xi, \eta, \delta} (e - Cy)^T \xi + (f - F \xi)^T \eta + g^T \delta - (\tau^l)^T \tilde{S} \xi \\
\text{s.t.} \quad B^T \xi + E^T \eta + G^T \delta \leq c, \\
\xi, \eta, \delta \geq 0, \\
\bar{\xi}_\alpha \leq \xi \leq \bar{\xi}_\alpha, \forall \alpha \in A
\]

Solving the problem (20), we then record optimal solution $(\tilde{\xi}^l, \tilde{\xi}^l, \eta^l, \delta^l)$ 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 $\beta \geq (-C^T \xi^l)^T y + e^T \xi^l + (f - F \xi^l)^T \eta^l + g^T \delta^l - (\tilde{\xi}^l)^T \tau$ 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 (17), which is a classic linear program. This problem can be solved by any LP solvers, e.g. simplex, dual-simplex, and interior-point method (Potra and Wright, 2000).
Another is the separation problem, which can be converted into the form of (20). Containing the bi-linear term \((F_\xi)^T \eta\), problem (20) 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 (Birge and Louveaux, 2011).

4.3 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 (17), and finally reaches the optimal solution under the distributional uncertainty of input data \(\xi\). 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 (Zeng and Zhao, 2013), 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. The CCG algorithm exhibits a faster convergence and requires less computational time, attributed to the following factors.

Firstly, the CCG algorithm efficiently identifies significant scenarios through the solution of its sub-problems, leading to a substantial increase in convergence rates. In contrast, the Benders-dual method necessitates numerous iterations to obtain the value function for a specific first-stage decision within the same scenario. Secondly, the CCG algorithm formulates a (large-scale) mixed-integer program as its master problem, preserving the network structure of the nominal model. This structural preservation enables the solver to fully leverage it during computations. Conversely, the Benders-dual method, with its generated cutting planes, hampers the identification and utilization of the inherent structure, potentially impeding computational efficiency.

For further discussions on the Benders-dual method with primal and dual cuts, as well as the CCG method, we refer readers to (Zeng and Zhao, 2013; Zhao and Jiang, 2017).

5 Numerical Tests

In this section, we test the proposed DR-QSOD model using the well-known Sioux Falls network (Morlok, 1973), and compare the performance of DR-QSOD with its deterministic counterpart, fixed-mapping QSOD model in (Wang et al., 2022), and stochastic QSOD model.

The Sioux Falls network (as shown in Figure 2) 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 (Stabler, 2016). 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 \(\varepsilon_0 = 5\) 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 (Wang et al., 2022).

Denote $\bar{d}$ as the “true” OD demands and $d^0 = (1 + \Delta_1) \times \bar{d}$ as the prior OD demands, where $\Delta_1 \sim \text{uniform}(-\epsilon_1, \epsilon_1)$. Given $\bar{d}$ and link attributes of the network, we solve the user equilibrium (UE) problem and derive the true link flows, denoted as $\bar{v}$. Then the observed link flow $\xi$ is generated as $\xi = (1 + \Delta_2) \times \bar{v}$, where $\Delta_2 \sim \text{uniform}(-\epsilon_2, \epsilon_2)$. The two parameters $\epsilon_1, \epsilon_2$ 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 2: Sioux Falls network](image)

### 5.1 Parameter tuning

The first step to conduct the numerical test is to tune a proper value for the weighting parameter $\lambda$ in (3). As previously stated, $\lambda$ 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 $\lambda$ 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 $\lambda$ is. Given this situation, we test various values of $\lambda$ in the DR-QSOD model and compare their OD estimation performance to determine the best choice. Table 3 demonstrates the results of DR-QSOD model performance under various values of $\lambda$ with the setting of $\epsilon_1 = 0.25$ and $\epsilon_2 = 0.1$. As $\lambda$ increases, the Root Mean Squared Errors (RMSE) of OD estimates, a widely-used metric to measure OD estimation errors, shows a trend of decreasing.

When $\lambda \leq 1$, the RMSE does not change across different values of $\lambda$. When $\lambda > 1$, the RMSE of OD estimates decreases while the time required by the BD method to converge grows. For example, when
Table 3: DR-QSOD model performance, running time vs. weighting parameter $\lambda$

<table>
<thead>
<tr>
<th>$\lambda$</th>
<th>RMSE</th>
<th>Running time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.25</td>
<td>142.42</td>
<td>281.21</td>
</tr>
<tr>
<td>0.5</td>
<td>142.42</td>
<td>300.83</td>
</tr>
<tr>
<td>1</td>
<td>142.42</td>
<td>224.94</td>
</tr>
<tr>
<td>1.5</td>
<td>140.37</td>
<td>991.35</td>
</tr>
<tr>
<td>2</td>
<td>133.92</td>
<td>1575.77</td>
</tr>
<tr>
<td>4</td>
<td>129.47</td>
<td>2540.97</td>
</tr>
<tr>
<td>6</td>
<td>131.29</td>
<td>5636.05</td>
</tr>
</tbody>
</table>

$\lambda = 4$, 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 $\lambda$ 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.

5.2 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 3 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 BD 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 4, 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.

5.3 Model performance

Having determined the weighting parameter $\lambda$ 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 5 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 6 shows the capability of keeping OD quasi-sparsity for both DR-QSOD model and deterministic QSOD model. It can be 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 (Wang et al., 2022), i.e. most of estimates will be either equal to its prior value or zero. Table 4 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 (Kleywegt et al., 2002) to recast
the stochastic QSOD model as (21), 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, (21) 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 demands estimated from link flows (\( \xi_i \)) of each individual day.

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

(21)

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. \( \Delta_2' \sim \text{uniform}(-\varepsilon_{l_2}', \varepsilon_{l_2}') \), \( \forall l \in L \). The deviation of the observed link flows \( \xi_i \) we incorporate into the two models were generated based on independent normal distribution, i.e. \( \Delta_2 \sim N(0, (\varepsilon_{l_2})^2) \), \( \forall l \in L \). In this way, it actually simulates a case where observed link flows are biased or its assumed distribution is not accurate enough. By setting \( \varepsilon_2 = 0.1 \), we first generate a link flow set \( \xi = \{\xi_1, \xi_2, \ldots, \xi_5\} \) consisting of 5-day link flows based on the normally distributed deviations. Then we use such \( \xi \) as the input, and solve both DR-QSOD and stochastic-QSOD model. As (21) 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 \( \xi' = \{\xi_1', \xi_2', \ldots, \xi_N'\} \) (generated from independent uniform distributions), and solve the second stage with the fixed \( d \) and \( \xi' \). Table 5 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
Figure 5: RMSE of estimates of DR-IOD model and deterministic QSOD model

Figure 6: Accuracy of OD quasi-sparsity of estimates for DR-QSOD model and deterministic QSOD model
Table 4: Illustration of sparsity property of DR-QSOD model

| $d$ | $d^0$   | $d$   | $|d - d^0|$ |
|-----|---------|-------|-----------|
| 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         |

accurate, DR-QSOD can produce more accurate results than the stochastic QSOD model since DRO models are more robust in dealing with data uncertainty. On the other hand, if the distribution bias does not exist, i.e. the assumed distribution of link flows 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 5. The conclusion is validated by 10 more similar tests as well. When it comes to the computational time, stochastic QSOD model generally has a much better performance. For the case in Table 5, the computational time for stochastic QSOD model is 3.15 seconds while the DR-QSOD model requires 1623.77 seconds. Such big difference is intuitive. According to Eq (21), the reformulated stochastic QSOD model can be converted to a linear programming problem, which is generally easier to solve. However, when the number of days $N$ increases, the dimension of this problem would linearly increase accordingly, which may cause potential memory issue. It is a very valuable topic to explore by our future experiments and study.

6 Conclusions

In this study, we propose a two-stage DR-QSOD model for the OD estimation problem on transportation networks. Our model is robust to the setting when the true distribution of input data or estimates is unknown. We first formulate a two-stage DRO model and reformulate the model into a tractable min-max form. We then develop solution algorithms based on the Benders Decomposition and the Column-and-Constraint Generation. Numerical experiments are conducted to compare the performance of different models and different solution algorithms. Numerical results show that DR-QSOD model produces a more accurate OD estimate in terms of estimation errors (RMSE) when compared with deterministic QSOD model. It
Table 5: RMSE of daily estimated OD demands $t$ for DR-QSOD and stochastic QSOD model

<table>
<thead>
<tr>
<th>Distribution bias</th>
<th>Index of day ($i$)</th>
<th>DR-QSOD</th>
<th>Stochastic-QSOD</th>
<th>DR-QSOD</th>
<th>Stochastic-QSOD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>160.873</td>
<td>167.426</td>
<td>168.865</td>
<td>162.642</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>169.496</td>
<td>182.107</td>
<td>172.326</td>
<td>168.226</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>167.886</td>
<td>182.572</td>
<td>170.111</td>
<td>170.002</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>189.413</td>
<td>194.093</td>
<td>173.072</td>
<td>148.41</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>158.896</td>
<td>174.649</td>
<td>176.702</td>
<td>173.058</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>182.661</td>
<td>182.154</td>
<td>173.105</td>
<td>175.444</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>171.717</td>
<td>164.508</td>
<td>184.829</td>
<td>163.79</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>179.773</td>
<td>190.527</td>
<td>184.846</td>
<td>184.437</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>181.319</td>
<td>185.095</td>
<td>162.125</td>
<td>162.193</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>173.053</td>
<td>169.696</td>
<td>181.217</td>
<td>178.443</td>
</tr>
<tr>
<td>Average</td>
<td></td>
<td>173.509</td>
<td>179.283</td>
<td>174.720</td>
<td>168.665</td>
</tr>
</tbody>
</table>

can also keep relatively high OD quasi-sparsity consistency, i.e. demands between most OD pairs will keep significant/insignificant as they do in the prior OD demands. When comparing with its stochastic counterpart, DR-QSOD model demonstrates its advantage in dealing with data uncertainty: DR-QSOD outperforms stochastic-QSOD model when the distribution of data is inaccurate or hard to obtain.

There are a few directions that are worthwhile to explore based on the DR-QSOD framework. First 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 model in (Wang et al., 2022). It is also interesting to explore other approaches to build the ambiguity set of observed data so that the model can leverage more meaningful statistical information to give more accurate results. This may also benefit from fuzzy methods when constructing fuzzy sets and memberships, which should also be explored. What’s more, as more sources of transportation big data become available to us, it is even more worthwhile to consider multi-source data integration for OD estimation problem as well as apply DR-QSOD to dynamic OD estimation scenarios. The DR-QSOD framework sits in an ideal position to connect emerging data sources with traditional OD estimation theories. The authors may investigate those topics, and the results may be reported in the subsequent research papers.

Acknowledgement

The authors thank the two anonymous reviewers for their constructive comments that helped significantly improve the previous versions of the paper. The work of the first and forth authors is partially supported by the National Science Foundation under grant CMMI-1825053. Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the National Science Foundation.
A Appendix

Quasi-sparsity is a property of large-scale OD demand matrices proposed and defined in (Wang et al., 2022). It states that for demands between OD pairs of large-scale transportation networks, most of them are small yet nonzero demands. An OD demand matrix can be called quasi-sparse if the following two conditions are both satisfied: 1) dominance in terms of the total number of insignificant OD pairs (OD pairs with small OD demands), i.e., the insignificant OD pairs in the OD matrix take an overwhelming portion (in terms of the number) over all the OD pairs, and 2) insignificance in terms of the total travel demands of insignificant OD pairs, i.e., the total OD demands of insignificant OD pairs account for a relatively small portion of the total OD demands in the network. As the quasi-sparsity does not require most small demands be strictly zero, its definition is more general than the sparsity definition. For a more complete discussion of quasi-sparsity, including its definition, derived property and applications, we refer readers to (Wang et al., 2022).

The fixed-mapping deterministic QSOD model (Wang et al., 2022) can be written as follow:

$$\min \|t - d^0\|_1 + \|\hat{Q} \cdot t - \xi\|_1 \text{ s.t. } t \geq 0,$$

(22)

where $t$, $d^0$, and $\xi$ indicate the estimated OD demand, prior OD demands, and observed link flow, respectively; $\hat{Q}$ stands for the fixed-mapping between estimated OD demand $t$ and observed link flow $\xi$.

The stochastic QSOD model can be written as follow:

$$\min_{d,t} \frac{1}{\lambda} \|d - d^0\|_1 + E_P(\|t - d\|_1 + \|\hat{Q} \cdot t - \xi\|_1) \text{ s.t. } d \geq 0,$$

$$t \geq 0,$$

(23)

where the probability of observed data $\xi$, denoted as $P$, is known or predefined. The definition of $d$ and $t$ are the same as the DR-QSOD model discussed above.

References


