

Wasserstein Distributionally Robust Shortest Path Problem

Zhuolin Wang^a, Keyou You^{a,*}, Shiji Song^a, Yuli Zhang^b

^aDepartment of Automation and BNRist, Tsinghua University, Beijing 100084, P.R. China

^bSchool of Management and Economics, Beijing Institute of Technology, Beijing 100081, P.R. China

Abstract

This paper proposes a data-driven distributionally robust shortest path (DRSP) model where the distribution of the travel time in the transportation network can only be partially observed through a *finite* number of samples. Specifically, we aim to find an optimal path to minimize the worst-case α -reliable mean-excess travel time (METT) over a Wasserstein ball, which is centered at the empirical distribution of the sample dataset and the ball radius quantifies the level of its confidence. In sharp contrast to the existing DRSP models, our model is equivalently reformulated as a tractable mixed 0-1 convex problem, e.g., 0-1 linear program or 0-1 second-order cone program. Moreover, we also explicitly derive the distribution achieving the worst-case METT by simply perturbing each sample. Experiments demonstrate the advantages of our DRSP model in terms of the out-of-sample performance and computational complexity. Finally, our DRSP model is easily extended to solve the DR bi-criteria shortest path problem and the minimum cost flow problem.

Keywords: Distributionally robust shortest path, ambiguity set, Wasserstein metric, METT, mixed 0-1 convex program

1. Introduction

The shortest path problem is one of the most fundamental problems in the transportation network and has broad applications, see e.g. Baxter et al. (2014); Cao et al. (2016); Tilk et al. (2017). To obtain an optimal path, the travel time in each arc of the network is essential. Due to different weather conditions, path capacity, traffic control and etc, travel time is usually subject to large variabilities, which may greatly affect the selection of an optimal path. In fact, travelers are not only concerned with the “nominal” travel time of each path but also its reliability (Bertsekas & Tsitsiklis, 1991; Fosgerau & Engelson, 2011; Nikolova & Stier-Moses, 2014).

In the literature, stochastic shortest path (SSP) models have been proposed to handle the random uncertainty in travel time under different criteria such as effective travel time (Lo & Tung, 2003), percentile travel time (Frank, 1969) and mean-excess travel time (Chen & Zhou, 2010). These models require an exact distribution of travel time

*Corresponding author.

Email addresses: wangz117@mails.tsinghua.edu.cn (Zhuolin Wang), youky@tsinghua.edu.cn (Keyou You), shijis@tsinghua.edu.cn (Shiji Song), zhangyuli@bit.edu.cn (Yuli Zhang)

for finding an optimal path. In practice, it is difficult to obtain the exact distribution since it may be time-varying and can only be estimated through a finite sample dataset (Masson & Denceux, 2006). Thus, a natural method is to approximate the SSP model via the sample-average approximation (SAA) method (Shapiro & Homem-de Mello, 1998; Verweij et al., 2003), where the true distribution is approximated by the discrete empirical distribution over the sample dataset. This method is only applicable to situations where the distribution is time-invariant and a large number of good samples can be generated cheaply. When the sample dataset is of low quality, the empirical distribution may significantly deviate from the true distribution and the SAA method tends to exhibit poor out-of-sample performance. From this perspective, the SAA method is not always reliable.

An alternative approach is to apply the distributionally robust (DR) optimization technique to the shortest path problem (Cheng et al., 2013; Shahabi et al., 2015; Yang & Zhou, 2017; Zhang et al., 2018), leading to a distributionally robust shortest path (DRSP) model. The DRSP model assumes that the true distribution belongs to an ambiguity set of distributions, over which an optimal path is to be found in some worst-case sense, e.g., the one minimizes the worst-case α -reliable mean-excess travel time (METT) (Zhang et al., 2018). For instance, the true distribution in Chassein et al. (2019) is parameterized with a vector that is assumed to lie in some set constructed by exploiting samples. However, it is NP-hard to solve most of the DRSP models (Yu & Yang, 1998) and only a few DRSP models are tractable by some well-defined ambiguity set (Esfahani & Kuhn, 2018).

In general, the ambiguity set should be large enough to include the true distribution with a high probability but can not be too “large” to avoid too conservative decisions. Moreover, the associated DRSP model needs to be as tractable as possible. The moment-based ambiguity set which consists of distributions with specified moment constraints is adopted for the DR optimization problem (Delage & Ye, 2010). Specifically, the DRSP model in Cheng et al. (2013); Zhang et al. (2018) assumes that the ambiguity set contains distributions with the *exactly* known first and second moments. Observe that this may lead to poor decisions if mismatch moments are used for the ambiguity set. In fact, it is often the case that we cannot obtain exact moment information with a finite sample dataset. Thus, one inevitably needs to further introduce uncertainty in the moment, which easily renders their DRSP models intractable. To resolve it, a metric-based ambiguity set has been developed in Erdođan & Iyengar (2006); Hu & Hong (2012); Jiang & Guan (2016); Pflug & Wozabal (2007); Wozabal (2012).

In this paper, we propose a novel data-driven DRSP model with a metric-based ambiguity set for the unknown distribution of travel time, which is defined as a ball centered at the empirical distribution over the finite sample dataset and the ball radius reflects our confidence in the empirical distribution. That is, the higher the quality of the empirical distribution, the smaller the ball radius. This facilitates us to utilize the sample dataset in a flexible way to hedge uncertainty. Then, we further incorporate the support set information into our DRSP model where the travel time is restricted to an interval constructed from the sample dataset. Noting that the empirical distribution is discrete and the true distribution of travel time is usually continuous, we adopt the Wasserstein metric (Kantorovich & Rubinshtein, 1958) to measure the distance between *any* two distributions, which is different from the Kullback-Leibler divergence (Hu & Hong, 2012; Jiang & Guan, 2016) and the Prokhorov metric (Erdođan & Iyengar, 2006).

Then, our DRSP model aims to minimize the worst-case METT over the aforementioned Wasserstein ball. While the moment-based DRSP model in Zhang et al. (2018) is NP-hard, our model can be equivalently reformulated as a solvable and finite mixed 0-1 convex program, e.g., 0-1 linear program (LP) or 0-1 second-order cone program (SOCP). Different from Zhang et al. (2018), the distribution that achieves the worst-case METT can also be explicitly derived by simply perturbing each sample, which is very helpful to assess the quality of the optimal path. Since the Wasserstein ball includes the true distribution with a high probability, the optimal path of our model offers good out-of-sample performance. Moreover, it asymptotically converges to the solution of the SSP model under the true distribution if the size of the sample dataset tends to infinity. Finally, experiment results show that our model achieves a better out-of-sample performance than that of the moment-based DRSP model in Zhang et al. (2018) and the SAA method. Our method is easily extended to solve the robust bi-criteria shortest path problem and the robust minimum cost flow problem over the Wasserstein ball. Both problems are reduced to finite convex programs which can be solved efficiently by existing algorithms. A preliminary version of this work has been presented in Wang et al. (2019), which only introduces the DRSP model over the Wasserstein set. Overall, the main contributions of this paper are summarized below.

- (1) We propose a novel data-driven DRSP model which aims to find an optimal path to minimize the worst-case METT within a Wasserstein ball.
- (2) In contrast to the existing DRSP models, our model is equivalently reformulated as a solvable mixed 0-1 convex problem and the worst-case distribution attaining the worst-case METT can be explicitly derived.
- (3) Besides its good out-of-sample performance and low computational complexity, our model can be easily extended to solve the robust bi-criteria shortest path problem as well as minimum flow cost problems.

The rest of the paper is organized as follows. Section 2 proposes the DRSP model to minimize the worst-case METT over a Wasserstein ball. In Section 3, we reformulate our DRSP model as a mixed 0-1 convex problem. The distribution achieving the worst-case METT is explicitly derived in Section 4. In Section 5, our model is extended to solve other DR problems. In Section 6, we perform numerical experiments to illustrate the performance of our model. Some conclusion remarks are drawn in Section 7.

2. Problem Formulation

2.1. The Shortest Path Problem and Reliability Criteria

Consider a directed and connected network $\mathcal{G} = (\mathcal{V}, \mathcal{A})$ with the vertex set \mathcal{V} and the arcs set \mathcal{A} , where $|\mathcal{V}| = m$ and $|\mathcal{A}| = n$. Let ξ_{ij} denote the travel time on an arc (i, j) in \mathcal{A} and $\boldsymbol{\xi} = \{\xi_{ij} : (i, j) \in \mathcal{A}\}$ be the stacked vector of the travel time over all arcs. A directed path is a sequence of arcs which connect a sequence of vertices in the same direction. Let $\boldsymbol{p} = \{p_{ij} : (i, j) \in \mathcal{A}\}$ represent a directed path from the origin vertex o to the destination vertex d , where p_{ij} is a binary decision variable and is one if arc (i, j) is on the path from vertex o to vertex d .

The standard shortest path problem aims to find an optimal path such that the sum of the travel time along the path is minimized (Ahuja et al., 1995), i.e.,

$$\underset{\mathbf{p}}{\text{minimize}} \quad \sum_{(i,j) \in \mathcal{A}} \xi_{ij} p_{ij} \quad (1)$$

$$\text{subject to} \quad \sum_{j:(i,j) \in \mathcal{A}} p_{ij} - \sum_{j:(j,i) \in \mathcal{A}} p_{ji} = b_i, \forall i \in \mathcal{V}, \quad (2)$$

$$p_{ij} \in \{0, 1\}, \forall (i, j) \in \mathcal{A}, \quad (3)$$

where $b_o = 1, b_d = -1$ and $b_i = 0$ for $i \in \mathcal{V}/\{o, d\}$. The constraint in (2) ensures the flow balance for the origin-destination pair (o, d) . In the sequel, let \mathcal{P} be the set of feasible paths from the original vertex o to the destination vertex d , i.e.,

$$\mathcal{P} = \{\mathbf{p} \mid \mathbf{p} \text{ satisfies (2) and (3)}\}. \quad (4)$$

The shortest path problem considers the network with a known vector of travel time, i.e., the constant vector ξ is assumed to be exactly given. In practice, travel time variability is unavoidable due to uncertain factors, e.g. traffic jams and weather conditions. Obviously, this vector has a significant impact on finding an optimal path for travelers. For example, consider a simple network with only three vertices and three arcs in Figure 1. The traveler aims to find an optimal path \mathbf{p}^* from 1 to 3 with the random travel time ξ_{ij} . Suppose that a traveler observes $\xi_{13} = 3.5, \xi_{12} = 1.5,$

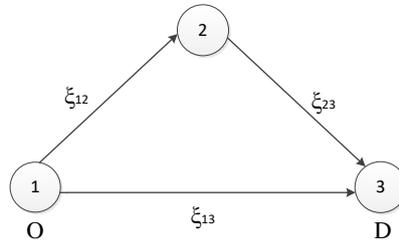


Figure 1: Topology of a simple network where O represents the origin vertex and D represents the destination vertex.

$\xi_{23} = 1.5$. Then an optimal path is obtained via solving (1), i.e., $1 \rightarrow 2 \rightarrow 3$ where $a \rightarrow b$ represents the directed arc from vertex a to vertex b . However, the travel time vector may change to $\xi_{13} = 2.5, \xi_{12} = 2, \xi_{23} = 1.2$ and then the optimal choice is reset as $1 \rightarrow 3$. Thus, the optimal path in the above is not always reliable if the uncertainty in travel time is neglected.

To quantify the reliability of a path, some criteria have been proposed, such as effective travel time (ETT) (Lo & Tung, 2003), percentile travel time (PTT) (Lo & Tung, 2003) and mean-excess travel time (METT) (Chen & Zhou, 2010). Under the assumption that ξ is a random vector with a distribution function F , the α -reliable METT of a path and the corresponding SSP model are defined below.

Definition 1. The α -reliable METT of path \mathbf{p} is defined as

$$\text{METT}_\alpha(\mathbf{p}) = \min_{t \in \mathbb{R}} \left\{ t + \frac{1}{\alpha} \mathbb{E}_F \{h(\mathbf{p}, t, \boldsymbol{\xi})\} \right\}, \quad (5)$$

where $h(\mathbf{p}, t, \boldsymbol{\xi}) = [\boldsymbol{\xi}^T \mathbf{p} - t]^+$ and $[x]^+ = \max\{x, 0\}$. The associated SSP model is given by

$$\underset{\mathbf{p} \in \mathcal{P}}{\text{minimize}} \quad \text{METT}_\alpha(\mathbf{p}). \quad (6)$$

The METT which coincides with the conditional Value-at-Risk (Rockafellar et al., 2000) is one of the most important criteria to evaluate the path reliability. It is able to simultaneously address questions “how much time do I need to allow” and “how bad should I expect from the worst cases?” (Chen & Zhou, 2010). However, solving the SSP model in (6) requires the exact distribution function F .

2.2. Data-driven Robust Shortest Path Problem

Usually, the true distribution F in (6) is unavailable and can only be partially observed through a finite sample dataset $\{\hat{\boldsymbol{\xi}}^i\}_{i \in [N]}$ where $\hat{\boldsymbol{\xi}}^i$ is an independent sample of the random vector of travel time and $[N] = \{1, \dots, N\}$. In this case, a natural idea is to adopt the SAA method. Specifically, F is approximated by an empirical distribution F_N over the sample dataset, i.e.,

$$F_N(\boldsymbol{\xi}) = \frac{1}{N} \sum_{i=1}^N \mathbf{1}_{\{\hat{\boldsymbol{\xi}}^i \leq \boldsymbol{\xi}\}}$$

where $\mathbf{1}_A$ is the indicator of event A . Then the SSP model in (6) is approximately solved by

$$\underset{t \in \mathbb{R}, \mathbf{p} \in \mathcal{P}}{\text{minimize}} \left\{ t + \frac{1}{\alpha} \frac{1}{N} \sum_{i=1}^N h(\mathbf{p}, t, \hat{\boldsymbol{\xi}}^i) \right\}. \quad (7)$$

By Glivenko-Cantelli theorem (Cantelli, 1933; Glivenko, 1933), the empirical distribution F_N converges weakly to the true distribution F as N tends to infinity. This implies that the solution to the SAA model in (7) converges to that of the SSP model in (6). That is, the SAA method is sensible only for the case where the true distribution F can be well approximated by the empirical distribution.

When the size of the sample dataset is small and/or the sample $\hat{\boldsymbol{\xi}}^k$ is of low quality, the empirical distribution F_N may deviate far from the true distribution F . More importantly, the distribution F may not be constant and is *time-varying*. Then, an optimal path of the SAA model in (7) may exhibit poor out-of-sample performance and is not always reliable.

As in Esfahani & Kuhn (2018) for the continuous optimization problem, we adopt a data-driven robust approach to hedge against the path unreliability either from the uncertainty of travel time or its distribution. The key idea is that the true distribution F is expected to “close” to the empirical distribution F_N with a high probability. In particular, we believe that F may belong to an ambiguity set \mathcal{F}_N that is centered at the empirical distribution F_N and its size reflects our confidence in F_N . The higher the confidence of F_N , the smaller the ambiguity set \mathcal{F}_N .

Since the true distribution F is usually continuous and the empirical distribution F_N is discrete, we adopt the Wasserstein metric (Kantorovich & Rubinshtein, 1958) to evaluate their distance, leading to a Wasserstein ball \mathcal{F}_N . Then, we are interested in the worst-case METT (w-METT) over the Wasserstein ball \mathcal{F}_N , i.e.,

$$\begin{aligned} \text{w-METT}_\alpha(\mathbf{p}) &= \sup_{F \in \mathcal{F}_N} \text{METT}_\alpha(\mathbf{p}) \\ &= \min_{t \in \mathbb{R}} \left\{ t + \frac{1}{\alpha} \sup_{F \in \mathcal{F}_N} \mathbb{E}_F \{h(\mathbf{p}, t, \boldsymbol{\xi})\} \right\}, \end{aligned} \quad (8)$$

where the second equality follows from Xie (2018). Then our DRSP model is obtained as

$$\underset{\mathbf{p} \in \mathcal{P}}{\text{minimize}} \quad \text{w-METT}_\alpha(\mathbf{p}). \quad (9)$$

To evaluate our optimal path, we aim to find the worst-case distribution F^* that achieves the w-METT, i.e.,

$$\sup_{F \in \mathcal{F}_N} \mathbb{E}_F \{h(\mathbf{p}, t, \boldsymbol{\xi})\} = \mathbb{E}_{F^*} \{h(\mathbf{p}, t, \boldsymbol{\xi})\}.$$

2.3. Ambiguity Set via Wasserstein Metric

The key of the DRSP model (9) is how to construct the Wasserstein ball

$$\mathcal{F}_N = \{F \in \mathcal{M}(\Xi) : d_W(F_N, F) \leq \epsilon_N\}, \quad (10)$$

where $\epsilon_N \geq 0$ reflects our confidence in the empirical distribution F_N and $\mathcal{M}(\Xi)$ is the set of some probability distributions supported on Ξ , i.e., \mathcal{F}_N contains all distributions within the ϵ_N -Wasserstein distance from F_N , in which the metric d_W is defined as follows.

Definition 2. The Wasserstein metric $d_W: \mathcal{M}(\Xi) \times \mathcal{M}(\Xi) \rightarrow \mathbb{R}_+$ is defined as

$$\begin{aligned} d_W(F_1, F_2) &= \inf \left\{ \int_{\Xi \times \Xi} d(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2) K(d\boldsymbol{\xi}_1, d\boldsymbol{\xi}_2) : \right. \\ &\quad \left. \int_{\Xi} K(\boldsymbol{\xi}_1, d\boldsymbol{\xi}_2) = F_1(\boldsymbol{\xi}_1), \int_{\Xi} K(d\boldsymbol{\xi}_1, \boldsymbol{\xi}_2) = F_2(\boldsymbol{\xi}_2) \right\}, \end{aligned}$$

where (Ξ, d) is a Polish metric space, $K: \Xi \times \Xi \rightarrow \mathbb{R}_+$ is the joint distribution of $F_1 \in \mathcal{M}(\Xi)$ and $F_2 \in \mathcal{M}(\Xi)$. Moreover, $d(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2) = \|\boldsymbol{\xi}_1 - \boldsymbol{\xi}_2\|_p$ where $\|\cdot\|$ represents l_p -norm on \mathbb{R}^n .

Though d_W satisfies the axioms of a metric, it may take the infinity value and thus is not a real distance. We need the following assumption on the set \mathcal{M} (Ambrosio & Gigli, 2013), under which d_W is actually a distance metric.

Assumption 1. For any distribution $F \in \mathcal{M}(\Xi)$, it holds that

$$\int_{\Xi} \|\boldsymbol{\xi}\|_p F(d\boldsymbol{\xi}) < \infty.$$

Assumption 1 requires the first moment of the distribution F to be finite, and only sacrifices little modeling power.

To the best of our knowledge, we are the first to adopt the Wasserstein ball in the DRSP model. We note that a similar line of this research for a traveling salesman problem is studied in Carlsson et al. (2018). Alternative metric-based ambiguity sets have also been adopted in DR optimizations, e.g., Kullback-Leibler set in Hu & Hong (2012); Jiang & Guan (2016), and Prokhorov set in Erdoğan & Iyengar (2006). However, the Kullback-Leibler metric is unable to effectively evaluate the distance between a continuous distribution and a discrete one. In particular, it enforces the associated \mathcal{F}_N to be a set of discrete distributions. The Prokhorov metric easily leads to an intractable shortest path model.

When the sample dataset is reasonably large, the Wasserstein ball \mathcal{F}_N includes the true distribution with a high probability (Esfahani & Kuhn, 2018) and thus the DRSP model is expected to exhibit good out-of-sample performance. Importantly, the proposed DRSP model can be equivalently reformulated as a mixed 0-1 program which is solvable via existing optimization techniques, e.g. the outer approximation decomposition algorithm (Duran & Grossmann, 1986), the branch-and-bound method (Gupta & Ravindran, 1985) and the extended cutting plane method (Westerlund & Pettersson, 1995).

3. Reformulation of the DRSP Model

In this section, we transform the DRSP model (9) over the Wasserstein ball \mathcal{F}_N in (10) with or without the support set to a finite mixed 0-1 convex problem respectively.

Norm	Without Support Set		With Support Set	
	w-METT	DRSP Model	w-METT	DRSP Model
$p = 1$	LP	Mixed 0-1 LP	LP	Mixed 0-1 LP
$p = 2$	LP	Mixed 0-1 SOCP	SOCP	Mixed 0-1 SOCP
$p = \infty$	LP	Mixed 0-1 LP	LP	Mixed 0-1 LP
Otherwise	LP	Mixed 0-1 Convex Program	Mixed 0-1 Convex Program	Mixed 0-1 Convex Program

Table 1: Equivalent problems of the our DRSP model, where p represents the l_p -norm on \mathbb{R}^n .

3.1. The DRSP Model Without Support Set

We first derive equivalent formulations for the w-METT (8) and the proposed DRSP model (9) over the Wasserstein ball \mathcal{F}_N without a support set.

Theorem 1. *Under Assumption 1, the w-METT in (8) over the Wasserstein ball \mathcal{F}_N can be computed by a finite linear programming (LP) problem*

$$\underset{t,s,\lambda}{\text{minimize}} \quad t + \frac{1}{\alpha} \left\{ \frac{1}{N} \sum_{i=1}^N s_i + \lambda \epsilon_N \right\} \quad (11a)$$

$$\text{subject to} \quad \mathbf{p}^T \hat{\boldsymbol{\xi}}^i - t \leq s_i, s_i \geq 0, \quad \forall i \in [N] \quad (11b)$$

$$\|\mathbf{p}\|_q \leq \lambda, \quad (11c)$$

where $\|\cdot\|_q$ is the dual of l_p -norm, i.e., $1/p + 1/q = 1$.

Moreover, the DRSP model (9) is equivalently reformulated as the following mixed 0-1 convex problem

$$\underset{\mathbf{p},t,s,\lambda}{\text{minimize}} \quad t + \frac{1}{\alpha} \left\{ \frac{1}{N} \sum_{i=1}^N s_i + \lambda \epsilon_N \right\} \quad (12)$$

$$\text{subject to} \quad \mathbf{p} \in \mathcal{P}, (11b) \text{ and } (11c).$$

To prove Theorem 1, a lemma is introduced below.

Lemma 1. For any $\mathbf{w} \in \mathbb{R}^n$, it holds that

$$\sup_{\mathbf{x} \in \mathbb{R}^n} \{\mathbf{w}^T \mathbf{x} - \lambda \|\mathbf{x}\|_p\} = \sup_{\mathbf{x} \in \mathbb{R}^n} \{(\|\mathbf{w}\|_q - \lambda) \|\mathbf{x}\|_p\}. \quad (13)$$

Proof. By Lemma 1 in Zhang et al. (2017), the maximum of $\sup_{\|\mathbf{x}\|_p=t} \{\mathbf{w}^T \mathbf{x}\}$ is explicitly given by $t\|\mathbf{w}\|_q$. Then we obtain

$$\begin{aligned} \sup_{\mathbf{x} \in \mathbb{R}^n} \{\mathbf{w}^T \mathbf{x} - \lambda \|\mathbf{x}\|_p\} &= \sup_{t \geq 0} \sup_{\|\mathbf{x}\|_p=t} \{\mathbf{w}^T \mathbf{x} - \lambda \|\mathbf{x}\|_p\} \\ &= \sup_{t \geq 0} \sup_{\|\mathbf{x}\|_p=t} \{\mathbf{w}^T \mathbf{x} - \lambda t\} = \sup_{t \geq 0} \{t\|\mathbf{w}\|_q - \lambda t\} \\ &= \sup_{\|\mathbf{x}\|_p \geq 0} \{(\|\mathbf{w}\|_q - \lambda) \|\mathbf{x}\|_p\} \\ &= \sup_{\mathbf{x} \in \mathbb{R}^n} \{(\|\mathbf{w}\|_q - \lambda) \|\mathbf{x}\|_p\}, \end{aligned}$$

which implies (13). □

Now, we are ready to prove Theorem 1.

Proof of Theorem 1. For any feasible path $\mathbf{p} \in \mathcal{P}$, it is easily seen from (10) that $\sup_{F \in \mathcal{F}_N} \mathbb{E}_F \{h(\mathbf{p}, t, \boldsymbol{\xi})\}$ can be obtained by solving the following conic linear program

$$\begin{aligned} \underset{K(\boldsymbol{\xi}, \hat{\boldsymbol{\xi}}^i) \geq 0}{\text{maximize}} \quad & \int_{\Xi} \sum_{i=1}^N h(\mathbf{p}, t, \boldsymbol{\xi}) K(d\boldsymbol{\xi}, \hat{\boldsymbol{\xi}}^i) \\ \text{subject to} \quad & \int_{\Xi} K(d\boldsymbol{\xi}, \hat{\boldsymbol{\xi}}^i) = \frac{1}{N}, \quad \forall i \in [N] \\ & \int_{\Xi} \sum_{i=1}^N d(\boldsymbol{\xi}, \hat{\boldsymbol{\xi}}^i) K(d\boldsymbol{\xi}, \hat{\boldsymbol{\xi}}^i) \leq \epsilon_N. \end{aligned} \quad (14)$$

We introduce a Lagrangian function for (14), i.e.,

$$L(\boldsymbol{\xi}, \lambda, \mathbf{s}) = \int_{\Xi} \sum_{i=1}^N h(\mathbf{p}, t, \boldsymbol{\xi}) K(d\boldsymbol{\xi}, \hat{\boldsymbol{\xi}}^i) - \int_{\Xi} \sum_{i=1}^N s_i K(d\boldsymbol{\xi}, \hat{\boldsymbol{\xi}}^i) - \int_{\Xi} \sum_{i=1}^N \lambda d(\boldsymbol{\xi}, \hat{\boldsymbol{\xi}}^i) K(d\boldsymbol{\xi}, \hat{\boldsymbol{\xi}}^i) + \frac{1}{N} \sum_{i=1}^N s_i + \lambda \epsilon_N. \quad (15)$$

It follows that the Lagrange dual function can be represented as

$$g(\lambda, \mathbf{s}) = \sup_{\boldsymbol{\xi} \in \Xi} L(\boldsymbol{\xi}, \lambda, \mathbf{s}) = \sup_{\boldsymbol{\xi} \in \Xi} \int_{\Xi} \sum_{i=1}^N (h(\mathbf{p}, t, \boldsymbol{\xi}) - s_i - \lambda d(\boldsymbol{\xi}, \hat{\boldsymbol{\xi}}^i)) K(d\boldsymbol{\xi}, \hat{\boldsymbol{\xi}}^i) + \frac{1}{N} \sum_{i=1}^N s_i + \lambda \epsilon_N. \quad (16)$$

Consequently, the dual problem of (14) is given as

$$\underset{s, \lambda}{\text{minimize}} \quad \frac{1}{N} \sum_{i=1}^N s_i + \lambda \epsilon_N \quad (17a)$$

$$\text{subject to} \quad h(\mathbf{p}, t, \boldsymbol{\xi}) - \lambda d(\boldsymbol{\xi}, \hat{\boldsymbol{\xi}}^i) \leq s_i, \forall \boldsymbol{\xi} \in \Xi, i \in [N] \quad (17b)$$

$$\lambda \geq 0. \quad (17c)$$

Consider the primal problem (14) and its dual problem (17). If $\epsilon_N > 0$, there exists a strictly feasible solution $K = F_N \times F_N$ to (14). Thus, the Slater condition for their strong duality holds (Shapiro, 2001). If $\epsilon_N = 0$, the Wasserstein ball \mathcal{F}_N reduces to a singleton $\{F_N\}$ and (14) changes to a sample average problem $\frac{1}{N} \sum_{i=1}^N h(\mathbf{p}, t, \hat{\boldsymbol{\xi}}^i)$. Indeed, any feasible solution to the dual problem (17) satisfies that $s_i \geq h(\mathbf{p}, t, \hat{\boldsymbol{\xi}}^i)$ with $\boldsymbol{\xi} = \hat{\boldsymbol{\xi}}^i$ and $s_i \geq 0$ with $\boldsymbol{\xi} \neq \hat{\boldsymbol{\xi}}^i$ when λ tends to infinity. Accordingly, the optimal value of problem (17) reduces to the sample average problem $\frac{1}{N} \sum_{i=1}^N h(\mathbf{p}, t, \hat{\boldsymbol{\xi}}^i)$ as well. Overall, there is no duality gap between (14) and (17) under any case. Thus, it is sufficient to solve the dual problem (17).

Since $h(\mathbf{p}, t, \boldsymbol{\xi}) = [\boldsymbol{\xi}^T \mathbf{p} - t]^+$, the constraint in (17b) amounts to

$$\sup_{\boldsymbol{\xi} \in \Xi} \{ \boldsymbol{\xi}^T \mathbf{p} - t - \lambda \|\boldsymbol{\xi} - \hat{\boldsymbol{\xi}}^i\|_p \} \leq s_i, \quad (18)$$

$$\sup_{\boldsymbol{\xi} \in \Xi} \{ -\lambda \|\boldsymbol{\xi} - \hat{\boldsymbol{\xi}}^i\|_p \} \leq s_i. \quad (19)$$

The inequality (19) implies that $s_i \geq 0$ since $\lambda \geq 0$ and $\hat{\boldsymbol{\xi}}^i \in \Xi$. For simplicity, we denote $\Delta \mathbf{u}_i = \boldsymbol{\xi} - \hat{\boldsymbol{\xi}}^i$ and re-express the left-hand side of (18) as

$$\begin{aligned} & \sup_{\Delta \mathbf{u}_i} \{ \mathbf{p}^T (\hat{\boldsymbol{\xi}}^i + \Delta \mathbf{u}_i) - t - \lambda \|\Delta \mathbf{u}_i\|_p \} \\ &= \sup_{\Delta \mathbf{u}_i} \{ \mathbf{p}^T \Delta \mathbf{u}_i - \lambda \|\Delta \mathbf{u}_i\|_p \} + \mathbf{p}^T \hat{\boldsymbol{\xi}}^i - t \\ &= \sup_{\Delta \mathbf{u}_i} \{ (\|\mathbf{p}\|_q - \lambda) \|\Delta \mathbf{u}_i\|_p \} + \mathbf{p}^T \hat{\boldsymbol{\xi}}^i - t \\ &= \begin{cases} \mathbf{p}^T \hat{\boldsymbol{\xi}}^i - t, & \text{if } \|\mathbf{p}\|_q \leq \lambda \\ +\infty, & \text{if } \|\mathbf{p}\|_q > \lambda \end{cases} \end{aligned}$$

where the second equality follows from Lemma 1. Then (14) can be reformulated as

$$\begin{aligned}
& \underset{s, \lambda}{\text{minimize}} && \frac{1}{N} \sum_{i=1}^N s_i + \lambda \epsilon_N \\
& \text{subject to} && \mathbf{p}^T \hat{\boldsymbol{\xi}}^i - t \leq s_i, \quad s_i \geq 0, \quad \forall i \in [N] \\
& && \|\mathbf{p}\|_q \leq \lambda.
\end{aligned} \tag{20}$$

Combining (20) with (8) leads to the equivalence of the w-METT and (11).

Consequently, the DRSP model in (9) can be reformulated as a mixed 0-1 convex problem (12), which completes the proof. \square

Theorem 1 implies that the DRSP model (9) is equivalent to a finite mixed 0-1 convex program. Note that the program in this paper is a discrete optimization problem which is more difficult to solve than the continuous optimization program in Esfahani & Kuhn (2018). Furthermore, different l_p -norms lead to different equivalent problems of the proposed DRSP model. For example, when we adopt l_1 -norm or l_2 -norm in the Wasserstein distance, the DRSP model is a 0-1 LP or 0-1 SOCP problem, respectively. See Table 1 for details.

3.2. The DRSP Model with the Support Set

The travel time is finite in practice and thus its support set should not be neglected. In this subsection, we incorporate this information by assuming that the travel time ξ_{ij} belongs to an empirical interval $[a_{ij}, b_{ij}]$ where $a_{ij} = \min_{k \in [N]} \xi_{ij}^k$ and $b_{ij} = \max_{k \in [N]} \xi_{ij}^k$. Then we derive an equivalent formulation for our DRSP model (9).

Let $\Xi = [\mathbf{a}, \mathbf{b}]$ in this subsection denote the support set where $\mathbf{a} = \{a_{ij} : (i, j) \in \mathcal{A}\}$ and $\mathbf{b} = \{b_{ij} : (i, j) \in \mathcal{A}\}$. Since $\Xi = [\mathbf{a}, \mathbf{b}]$ is compact, any distribution F in \mathcal{F}_N given by (10) in this subsection automatically satisfies Assumption 1. Following this, we derive equivalent problems for w-METT (8) and the proposed DRSP model (9).

Theorem 2. *Let $\Xi = [\mathbf{a}, \mathbf{b}]$, then the w-METT in (8) over the Wasserstein ball (10) can be computed by a finite convex problem*

$$\underset{t, s, \lambda, \boldsymbol{\gamma}_i, \boldsymbol{\eta}_i}{\text{minimize}} \quad t + \frac{1}{\alpha} \left\{ \frac{1}{N} \sum_{i=1}^N s_i + \lambda \epsilon_N \right\} \tag{21a}$$

$$\text{subject to} \quad (\mathbf{p} + \boldsymbol{\gamma}_i - \boldsymbol{\eta}_i)^T \hat{\boldsymbol{\xi}}^i - \boldsymbol{\gamma}_i^T \mathbf{a} + \boldsymbol{\eta}_i^T \mathbf{b} - t \leq s_i \tag{21b}$$

$$\|\boldsymbol{\gamma}_i + \mathbf{p} - \boldsymbol{\eta}_i\|_q \leq \lambda \tag{21c}$$

$$\boldsymbol{\eta}_i \geq 0, \boldsymbol{\gamma}_i \geq 0, s_i \geq 0, \forall i \in [N]. \tag{21d}$$

Moreover, the DRSP problem (9) is re-expressed as

$$\underset{p, t, s, \lambda, \boldsymbol{\gamma}_i, \boldsymbol{\eta}_i}{\text{minimize}} \quad t + \frac{1}{\alpha} \left\{ \frac{1}{N} \sum_{i=1}^N s_i + \lambda \epsilon_N \right\} \tag{22}$$

$$\text{subject to} \quad \mathbf{p} \in \mathcal{P}, \text{ (21b), (21c) and (21d)}.$$

Proof. The strong duality still holds for the inner largest expectation in (8), which allows us to reformulate problem (8) as

$$\underset{t,s,\lambda}{\text{minimize}} \quad t + \frac{1}{\alpha} \left\{ \frac{1}{N} \sum_{i=1}^N s_i + \lambda \epsilon_N \right\} \quad (23a)$$

$$\text{subject to} \quad h(\mathbf{p}, t, \boldsymbol{\xi}) - \lambda d(\boldsymbol{\xi}, \hat{\boldsymbol{\xi}}^i) \leq s_i, \forall \boldsymbol{\xi} \in \Xi, i \in [N] \quad (23b)$$

$$\lambda \geq 0. \quad (23c)$$

The constraint in (23b) can be represented as

$$\sup_{\boldsymbol{\xi} \in \Xi} \left\{ \boldsymbol{\xi}^T \mathbf{p} - t - \lambda \|\boldsymbol{\xi} - \hat{\boldsymbol{\xi}}^i\|_p \right\} \leq s_i \quad (24)$$

$$\sup_{\boldsymbol{\xi} \in \Xi} \left\{ -\lambda \|\boldsymbol{\xi} - \hat{\boldsymbol{\xi}}^i\|_p \right\} \leq s_i, \quad (25)$$

where the inequality in (25) implies that $s_i \geq 0$.

Note that Lemma 1 in Section 3.1 cannot be applied directly to (24). We utilize the Lagrange dual function to solve this problem. For brevity, we denote $\Delta \mathbf{u}_i = \boldsymbol{\xi} - \hat{\boldsymbol{\xi}}^i$ and express the Lagrangian function of $\sup_{\boldsymbol{\xi} \in \Xi} \left\{ \boldsymbol{\xi}^T \mathbf{p} - t - \lambda \|\boldsymbol{\xi} - \hat{\boldsymbol{\xi}}^i\|_p \right\}$ as

$$L(\Delta \mathbf{u}_i, \boldsymbol{\gamma}_i, \boldsymbol{\eta}_i) = (\mathbf{p} + \boldsymbol{\gamma}_i - \boldsymbol{\eta}_i)^T (\Delta \mathbf{u}_i + \hat{\boldsymbol{\xi}}^i) - \lambda \|\Delta \mathbf{u}_i\|_p - \boldsymbol{\gamma}_i^T \mathbf{a} + \boldsymbol{\eta}_i^T \mathbf{b} - t. \quad (26)$$

Then, the Lagrange dual function of (26) is given by

$$\begin{aligned} g(\boldsymbol{\gamma}_i) &= \sup_{\Delta \mathbf{u}_i} L(\Delta \mathbf{u}_i, \boldsymbol{\gamma}_i, \boldsymbol{\eta}_i) \\ &= \sup_{\Delta \mathbf{u}_i} \left\{ (\mathbf{p} + \boldsymbol{\gamma}_i - \boldsymbol{\eta}_i)^T \Delta \mathbf{u}_i - \lambda \|\Delta \mathbf{u}_i\|_p \right\} + (\mathbf{p} + \boldsymbol{\gamma}_i - \boldsymbol{\eta}_i)^T \hat{\boldsymbol{\xi}}^i - \boldsymbol{\gamma}_i^T \mathbf{a} + \boldsymbol{\eta}_i^T \mathbf{b} - t \\ &= \sup_{\Delta \mathbf{u}_i} \left\{ (\|\mathbf{p} + \boldsymbol{\gamma}_i - \boldsymbol{\eta}_i\|_q - \lambda) \|\Delta \mathbf{u}_i\|_p \right\} + (\mathbf{p} + \boldsymbol{\gamma}_i - \boldsymbol{\eta}_i)^T \hat{\boldsymbol{\xi}}^i - \boldsymbol{\gamma}_i^T \mathbf{a} + \boldsymbol{\eta}_i^T \mathbf{b} - t \\ &= \begin{cases} +\infty, & \text{if } \|\mathbf{p} + \boldsymbol{\gamma}_i - \boldsymbol{\eta}_i\|_q > \lambda \\ (\mathbf{p} + \boldsymbol{\gamma}_i - \boldsymbol{\eta}_i)^T \hat{\boldsymbol{\xi}}^i - \boldsymbol{\gamma}_i^T \mathbf{a} + \boldsymbol{\eta}_i^T \mathbf{b} - t, & \text{if } \|\mathbf{p} + \boldsymbol{\gamma}_i - \boldsymbol{\eta}_i\|_q \leq \lambda \end{cases} \end{aligned}$$

where the third equality follows from Lemma 1.

Consequently, $\sup_{\boldsymbol{\xi} \in \Xi} \left\{ \boldsymbol{\xi}^T \mathbf{p} - t - \lambda \|\boldsymbol{\xi} - \hat{\boldsymbol{\xi}}^i\|_p \right\}$ admits an equivalent problem

$$\begin{aligned} \underset{\boldsymbol{\gamma}_i, \boldsymbol{\eta}_i}{\text{minimize}} \quad & (\mathbf{p} + \boldsymbol{\gamma}_i - \boldsymbol{\eta}_i)^T \hat{\boldsymbol{\xi}}^i - \boldsymbol{\gamma}_i^T \mathbf{a} + \boldsymbol{\eta}_i^T \mathbf{b} - t \\ \text{subject to} \quad & \|\mathbf{p} + \boldsymbol{\gamma}_i - \boldsymbol{\eta}_i\|_q \leq \lambda \\ & \boldsymbol{\gamma}_i \geq 0, \quad \boldsymbol{\eta}_i \geq 0, \end{aligned} \quad (27)$$

where the strong duality holds as the uncertainty set is non-empty.

Substituting (27) into constraints in (23b), we obtain that the w-METT (8) can be computed by the program (21) and the DRSP model eventually is given by (22). \square

We also summarize the results in Theorem 2 in Table 1. Different from the NP-hard moment-based DRSP model with support set in Zhang et al. (2018), Table 1 shows that the Wasserstein distance with l_1 -norm, l_2 -norm and l_∞ -norm leads to a tractable problem for our DRSP model, e.g., the mixed 0-1 LP or SOCP problem.

3.3. Asymptotic consistency

We finally discuss the asymptotic consistency of our DRSP model under the following mild condition.

Assumption 2. For the true distribution F , there exists a constant $c > 1$ such that

$$\int_{\Xi} \exp(\|\xi\|_p^c) F(d\xi) \leq \infty$$

Under Assumption 2, the asymptotic consistency of our model can be formalized below.

Theorem 3. Under Assumption 2 and let $\beta_N \in (0, 1)$ with $\sum_{N=1}^{\infty} \beta_N \leq \infty$. Define the radius ϵ_N of the Wasserstein ball as

$$\epsilon_N(\beta_N) = \begin{cases} \left(\frac{\log(c_1 \beta_N^{-1})}{c_2 N} \right)^{1/\max\{n, 2\}} & \text{if } N \geq \frac{\log(c_1 \beta_N^{-1})}{c_2} \\ \left(\frac{\log(c_1 \beta_N^{-1})}{c_2 N} \right)^{1/c} & \text{if } N < \frac{\log(c_1 \beta_N^{-1})}{c_2}, \end{cases}$$

where c_1 and c_2 are two positive constants that depend on the constant c in Assumption 2.

Then the optimal value and the optimal solution of the DRSP model (9) converge to those of the SSP model (6) with probability one as N tends to infinity.

Proof. Since $h(\mathbf{p}, t, \xi)$ is continuous in ξ and there exists $L \geq 0$ with $|h(\mathbf{p}, t, \xi)| \leq L(1 + \|\xi\|)$ for all $\mathbf{p} \in \mathcal{P}$, $t \in \mathbb{R}$ and $\xi \in \Xi$, then the asymptotic consistency of our DRSP model follows from Theorem 3.6 in Esfahani & Kuhn (2018). \square

4. The Worst-case Distribution Achieving the w-METT

In this section we derive the worst-case distribution that attains the supremum of the w-METT in (8) for any feasible pair of $\{\mathbf{p}, t\}$, i.e., $\mathbf{p} \in \mathcal{P}$ and $t \in \mathbb{R}$.

Lemma 2. Given any feasible pair of $\{\mathbf{p}, t\}$, then $\sup_{F \in \mathcal{F}_N} \mathbb{E}_F\{h(\mathbf{p}, t, \xi)\}$ is equivalent to

$$\sup_{\tilde{\xi} \in \mathcal{B}} \left\{ \frac{1}{N} \sum_{i=1}^N h(\mathbf{p}, t, \xi^{(i)}) \right\} \quad (28)$$

where

$$\mathcal{B} = \left\{ (\xi^{(1)}, \dots, \xi^{(N)}) \mid \frac{1}{N} \sum_{i=1}^N d(\xi^{(i)}, \hat{\xi}^i) \leq \epsilon_N, \xi^{(i)} \in \Xi \right\}.$$

Proof. Fix any solution $\{\mathbf{p}, t\}$, it follows from the weak duality that

$$\begin{aligned} \sup_{\tilde{\xi} \in \mathcal{B}} \left\{ \frac{1}{N} \sum_{i=1}^N h(\mathbf{p}, t, \xi^{(i)}) \right\} &\leq \inf_{\lambda \geq 0} \sup_{\xi^{(i)} \in \Xi} \left\{ \frac{1}{N} \sum_{i=1}^N h(\mathbf{p}, t, \xi^{(i)}) - \lambda \left(\frac{1}{N} \sum_{i=1}^N d(\xi^{(i)}, \hat{\xi}^i) - \epsilon_N \right) \right\} \\ &= \inf_{\lambda \geq 0} \left\{ \lambda \epsilon_N + \frac{1}{N} \sum_{i=1}^N \sup_{\xi^{(i)} \in \Xi} \left\{ h(\mathbf{p}, t, \xi^{(i)}) - \lambda d(\xi^{(i)}, \hat{\xi}^i) \right\} \right\} \\ &= \sup_{F \in \mathcal{F}_N} \mathbb{E}_F \{h(\mathbf{p}, t, \xi)\} \end{aligned}$$

where the last equality follows from the program (17) in the proof of Theorem 1. This implies that $\sup_{F \in \mathcal{F}_N} \mathbb{E}_F \{h(\mathbf{p}, t, \xi)\}$ is greater than (28).

Next we show that $\sup_{F \in \mathcal{F}_N} \mathbb{E}_F \{h(\mathbf{p}, t, \xi)\}$ is also less than (28). For any $\varepsilon \geq 0$, it follows from the equivalent program (17) that there exists $\{\tilde{\xi}^{(i)}\}_{i \in [N]} \subseteq \Xi$ such that

$$\sup_{F \in \mathcal{F}_N} \mathbb{E}_F \{h(\mathbf{p}, t, \xi)\} - \varepsilon < \inf_{\lambda \geq 0} \left\{ \lambda \epsilon_N + \frac{1}{N} \sum_{i=1}^N \left\{ h(\mathbf{p}, t, \tilde{\xi}^{(i)}) - \lambda d(\tilde{\xi}^{(i)}, \hat{\xi}^i) \right\} \right\}. \quad (29)$$

If $(\tilde{\xi}^{(1)}, \dots, \tilde{\xi}^{(N)}) \notin \mathcal{B}$ and let $\lambda > 0$, it holds that

$$\lambda \left\{ \epsilon_N - \frac{1}{N} \sum_{i=1}^N d(\tilde{\xi}^{(i)}, \hat{\xi}^i) \right\} < 0.$$

If λ tends to $+\infty$ in (29), it leads to $\sup_{F \in \mathcal{F}_N} \mathbb{E}_F \{h(\mathbf{p}, t, \xi)\} = -\infty$. This contradicts the fact that $\sup_{F \in \mathcal{F}_N} \mathbb{E}_F \{h(\mathbf{p}, t, \xi)\} \geq \mathbb{E}_{F_N} \{h(\mathbf{p}, t, \xi)\} > -\infty$. Thus, $(\tilde{\xi}^{(1)}, \dots, \tilde{\xi}^{(N)}) \in \mathcal{B}$ and

$$\sup_{F \in \mathcal{F}_N} \mathbb{E}_F \{h(\mathbf{p}, t, \xi)\} - \varepsilon < \inf_{\lambda \geq 0} \sup_{\tilde{\xi} \in \mathcal{B}} \left\{ \lambda \epsilon_N + \frac{1}{N} \sum_{i=1}^N \left\{ h(\mathbf{p}, t, \tilde{\xi}^{(i)}) - \lambda d(\tilde{\xi}^{(i)}, \hat{\xi}^i) \right\} \right\}$$

Since for any $(\xi^{(1)}, \dots, \xi^{(N)}) \in \mathcal{B}$ it holds that

$$\lambda \left\{ \epsilon_N - \frac{1}{N} \sum_{i=1}^N d(\xi^{(i)}, \hat{\xi}^i) \right\} \geq 0,$$

it implies that

$$\sup_{F \in \mathcal{F}_N} \mathbb{E}_F \{h(\mathbf{p}, t, \xi)\} - \varepsilon < \sup_{\tilde{\xi} \in \mathcal{B}} \left\{ \frac{1}{N} \sum_{i=1}^N \left\{ h(\mathbf{p}, t, \xi^{(i)}) \right\} \right\}$$

Letting ε tend to zero leads to the desired inequality and hence we obtain that

$$\sup_{F \in \mathcal{F}_N} \mathbb{E}_F \{h(\mathbf{p}, t, \xi)\} = \sup_{\tilde{\xi} \in \mathcal{B}} \left\{ \frac{1}{N} \sum_{i=1}^N h(\mathbf{p}, t, \xi^{(i)}) \right\}.$$

□

Note that the objective function of (28) is continuous and its feasible set \mathcal{B} is compact, the supremum problem admits an optimal solution. That is, “sup” can be replaced by “max”. Since Lemma 2 implies the equivalence of the optimal value of (28) and $\sup_{F \in \mathcal{F}_N} \mathbb{E}_F \{h(\mathbf{p}, t, \xi)\}$ for any feasible pair of $\{\mathbf{p}, t\}$, we can construct a worst-case distribution based on the optimal solution of (28).

Theorem 4. Given any feasible pair of $\{\mathbf{p}, t\}$, let $\xi_{\{\mathbf{p}, t\}} = (\xi_{\{\mathbf{p}, t\}}^{(1)}, \dots, \xi_{\{\mathbf{p}, t\}}^{(N)})$ be an optimal solution of the optimization problem in Lemma 2. Then the following distribution

$$F_{\{\mathbf{p}, t\}}^* = \frac{1}{N} \sum_{i=1}^N \delta_{\xi_{\{\mathbf{p}, t\}}^{(i)}}$$

is the worst-case distribution, i.e.,

$$\sup_{F \in \mathcal{F}_N} \mathbb{E}_F \{h(\mathbf{p}, t, \xi)\} = \mathbb{E}_{F_{\{\mathbf{p}, t\}}^*} \{h(\mathbf{p}, t, \xi)\}.$$

Proof. We first show $F_{\{\mathbf{p}, t\}}^*$ belongs to the Wasserstein ball \mathcal{F}_N . Denote a joint distribution of F_N and $F_{\{\mathbf{p}, t\}}^*$ as

$$\Pi_{\{\mathbf{p}, t\}} = \frac{1}{N} \sum_{i=1}^N \delta_{(\xi_{\{\mathbf{p}, t\}}^{(i)}, \hat{\xi}^i)}.$$

Then by the definition of the Wasserstein metric we obtain that

$$d_W(F_N, F_{\{\mathbf{p}, t\}}^*) \leq \int \|\xi - \xi'\|_p \Pi_{\{\mathbf{p}, t\}}(d\xi, d\xi') = \frac{1}{N} \sum_{i=1}^N \|\xi_{\{\mathbf{p}, t\}}^{(i)} - \hat{\xi}^i\|_p \leq \epsilon_N,$$

where the last inequality holds since $(\xi_{\{\mathbf{p}, t\}}^{(1)}, \dots, \xi_{\{\mathbf{p}, t\}}^{(N)}) \in \mathcal{B}$. Thus, $F_{\{\mathbf{p}, t\}}^*$ is contained in the Wasserstein ball \mathcal{F}_N . Consequently, we obtain that

$$\sup_{F \in \mathcal{F}_N} \mathbb{E}_F \{h(\mathbf{p}, t, \xi)\} \geq \mathbb{E}_{F_{\{\mathbf{p}, t\}}^*} \{h(\mathbf{p}, t, \xi)\} = \frac{1}{N} \sum_{i=1}^N (h(\mathbf{p}, t, \xi_{\{\mathbf{p}, t\}}^{(i)})) = \sup_{F \in \mathcal{F}_N} \mathbb{E}_F \{h(\mathbf{p}, t, \xi)\}$$

where the last equality holds due to Lemma 2. This implies that $F_{\{\mathbf{p}, t\}}^*$ is one of the worst-case distribution in the Wasserstein ball \mathcal{F}_N . Here we complete this proof. \square

Theorem 4 implies that the worst-case distribution has a finite support and the number of support elements is the same as the number of the data samples. As the moment-based DRSP model is only solved by an approximation method, they are unable to provide any result on the worst-case distribution.

5. Extensions

5.1. DR Bi-criteria Shortest Path Problem

Now we extend our DRSP model to the bi-criteria DR shortest path problem, i.e.,

$$\underset{t \in \mathbb{R}, \mathbf{p} \in \mathcal{P}}{\text{minimize}} \left[\begin{array}{c} \sum_{(i,j) \in \mathcal{A}} c_{ij} p_{ij} \\ t + \frac{1}{\alpha} \sup_{F \in \mathcal{F}_N} \mathbb{E}_F \{h(\xi, t, \mathbf{p})\} \end{array} \right], \quad (30)$$

where $\mathbf{c} = \{c_{ij} : (i, j) \in \mathcal{A}\}$ is the deterministic vector of the travel cost over all arcs and c_{ij} is the cost on each arc $(i, j) \in \mathcal{A}$. Moreover, we set $\Xi = \mathbb{R}^n$ in this subsection.

Since the existence of a path simultaneously minimizing both objectives in (30) cannot be guaranteed, we follow the idea of Mangasarian (1994) to seek weakly robust efficient paths with a bearable trade off between two objectives.

Definition 3. Let $f_1(\mathbf{p}) = \mathbf{c}^T \mathbf{p}$, $f_2(\mathbf{p}, t) = t + \frac{1}{\alpha} \mathbb{E}_F \{h(\boldsymbol{\xi}, t, \mathbf{p})\}$, then $\{\bar{\mathbf{p}}, \bar{t}\}$ is a weakly robust efficient solution of (30) if and only if there exists $\bar{\lambda}_1 \geq 0, \bar{\lambda}_2 \geq 0$ and $\bar{\lambda}_1 + \bar{\lambda}_2 \neq 0$ such that

$$\bar{\lambda}_1 f_1(\mathbf{p}) + \bar{\lambda}_2 \sup_{F \in \mathcal{F}_N} f_2(\mathbf{p}, t) \geq \bar{\lambda}_1 f_1(\bar{\mathbf{p}}) + \bar{\lambda}_2 \sup_{F \in \mathcal{F}_N} f_2(\bar{\mathbf{p}}, \bar{t})$$

holds for any feasible pair of $\{\mathbf{p}, t\}$.

Corollary 1. The weakly robust efficient solution to (30) over the Wasserstein ball \mathcal{F}_N can be obtained by solving the following problem

$$\begin{aligned} & \underset{\mathbf{p}, s, \lambda}{\text{minimize}} && \bar{\lambda}_1 f_1(\mathbf{p}) + \bar{\lambda}_2 \left(t + \frac{1}{\alpha} \frac{1}{N} \sum_{i=1}^N s_i + \lambda \epsilon_N \right) \\ & \text{subject to} && \mathbf{p}^T \hat{\boldsymbol{\xi}}^i - t \leq s_i, s_i \geq 0, \forall i \in [N], \\ & && \|\mathbf{p}\|_q \leq \lambda, \mathbf{p} \in \mathcal{P}. \end{aligned} \tag{31}$$

Proof. Since the uncertainty only arises in the second objective function $f_2(\mathbf{p})$, which equals the objective function in our DRSP model, the result follows from Theorem 1. \square

Corollary 1 implies that our DRSP model can be extended to the DR bi-criteria shortest path problem and its weakly robust efficient solutions can be obtained simply by solving a deterministic convex problem.

5.2. DR Minimum Cost Flow Problem

This subsection considers the DR minimum cost flow problem based on the Wasserstein ball, i.e.,

$$\begin{aligned} & \underset{\mathbf{x}}{\text{minimize}} && \sup_{F \in \mathcal{F}_N} \mathbb{E}_F \left\{ \sum_{(i,j) \in \mathcal{A}} c_{ij} x_{ij} \right\} \\ & \text{subject to} && \sum_{j:(i,j) \in \mathcal{A}} x_{ij} - \sum_{j:(j,i) \in \mathcal{A}} x_{ji} = b_i, \forall i \in \mathcal{V}, \\ & && 0 \leq x_{ij} \leq u_{ij}, \forall (i, j) \in \mathcal{A}, \end{aligned} \tag{32}$$

where c_{ij} and $u_{ij} \geq 0$ are the random cost and the capacity of arc (i, j) respectively. $b_i > 0$ is the supply of each vertex i . Let $\mathbf{x} = \{x_{ij} : (i, j) \in \mathcal{A}\}$ represent a flow from the source vertex to the sink vertex, where x_{ij} is the flow on arc (i, j) . Let \mathcal{X} be the set of feasible flows of problem (32).

We are interested in the problem whose distribution of the random cost vector $\mathbf{c} = \{c_{ij} : (i, j) \in \mathcal{A}\} \in \mathbb{R}^n$ belongs to a Wasserstein ball \mathcal{F}_N . We prove that (32) can be reformulated as a modified network flow problem.

Corollary 2. Let $d(\cdot) = \|\cdot\|_\infty$ and $\Xi = \mathbb{R}^n$ in the Wasserstein ball \mathcal{F}_N , the optimal DR flow to (32) can be obtained by solving

$$\begin{aligned} & \underset{\mathbf{x}}{\text{minimize}} && \sum_{(i,j) \in \mathcal{A}} \left(\frac{1}{N} \sum_{k=1}^N \hat{c}_{ij}^k + \epsilon_N \right) x_{ij} \\ & \text{subject to} && \mathbf{x} \in \mathcal{X}, \end{aligned} \tag{33}$$

where $\hat{\mathbf{c}}^k$ is the k th sample of the random cost vector and \mathcal{X} is the set of all feasible flows.

Proof. By Theorem 1, the problem (32) admits an equivalent program

$$\begin{aligned}
& \underset{\mathbf{x}}{\text{minimize}} && \frac{1}{N} \sum_{k=1}^N s_k + \lambda \epsilon_N \\
& \text{subject to} && \sum_{(i,j) \in \mathcal{A}} \hat{c}_{ij}^k x_{ij} \leq s_k, \forall k \in [N], \\
& && \lambda \geq \|\mathbf{x}\|_1, \mathbf{x} \in \mathcal{X}.
\end{aligned} \tag{34}$$

Eliminating the variables s_k and λ from (34) leads to

$$\begin{aligned}
& \underset{\mathbf{x}}{\text{minimize}} && \frac{1}{N} \sum_{k=1}^N \sum_{(i,j) \in \mathcal{A}} \hat{c}_{ij}^k x_{ij} + \epsilon_N \|\mathbf{x}\|_1 \\
& \text{subject to} && \mathbf{x} \in \mathcal{X}.
\end{aligned} \tag{35}$$

Since $x_{ij} \geq 0$, we can rewrite $\|\mathbf{x}\|_1$ as $\sum_{(i,j) \in \mathcal{A}} x_{ij}$ and reformulate (35) as (33), which is a nominal minimum cost flow problem with edge cost $c_{ij} = \frac{1}{N} \sum_{k=1}^N \hat{c}_{ij}^k + \epsilon_N$. \square

Corollary 2 implies that the DR minimum cost flow problem over the Wasserstein ball can be converted to a deterministic network flow problem with arc cost $c_{ij} = \frac{1}{N} \sum_{k=1}^N \hat{c}_{ij}^k + \epsilon_N$.

6. Experiments

Numerical experiments are conducted to validate the performance of our DRSP model in this section. We take the commonly-used l_1 -norm to compute the difference of travel time of two vectors. In view of Table 1, our DRSP model is reduced to a mixed 0-1 LP problem, which can be solved via existing optimization techniques. All experiments are implemented on a 64 bit PC with an Intel Core i5-7500 CPU at 3.4GHz and 8 GB RAM. Cplex 12.6 is used to solve the mixed 0-1 LP problem.

6.1. DR Shortest Path Problems

Experiments on the Eastern Massachusetts (EMA) network (Stabler et al., 2019) with 74 vertices and 258 links are firstly performed to reveal the impact of different parameters on our DRSP model. Travel time of each arc is captured by a random vector $\xi = \{\xi_{ij} : (i, j) \in \mathcal{A}\}$. We find paths from the origin vertex 1 to the destination vertex 74 by solving our DRSP model.

We assume that ξ_{ij} of different arcs is independent across the network and follows a mixture of Gaussian distribution $\mathcal{N}(\mu_{ij}, \mu_{ij} \times 10)$ and uniform distribution $\mathcal{U}(0, \mu_{ij})$ where μ_{ij} is obtained from Stabler et al. (2019). Clearly, both distributions satisfy Assumption 1 with finite first moments. In the experiment, half of the dataset are generated from $\mathcal{N}(\mu_{ij}, \mu_{ij} \times 10)$ and the rest are generated from $\mathcal{U}(0, \mu_{ij})$. Impacts of the Wasserstein radius ϵ_N and the dataset size N on the out-of-sample performance of our DRSP model are tested respectively. We compare our model with the moment-based DRSP model in Zhang et al. (2018) and the SAA method in terms of the out-of-sample performance and computation complexity. Table 2 shows the use of parameters in different experiments.

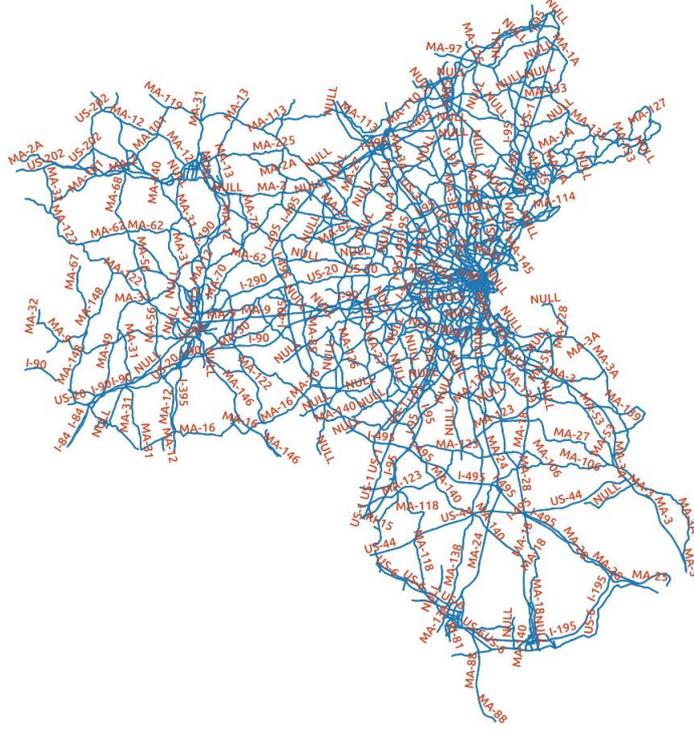


Figure 2: Eastern Massachusetts (EMA) network with 74 vertices and 258 links.

We evaluate the out-of-sample performance by examining the cost of the model under *new* samples, i.e.,

$$\underset{t \in \mathbb{R}}{\text{minimize}} \left\{ t + \frac{1}{\alpha} \mathbb{E}_F \{ h(\mathbf{p}, t, \boldsymbol{\xi}) \} \right\} \quad (36)$$

Noting that the true distribution F is a mixture of the Gaussian distribution and the uniform distribution, and importantly unknown. We are unable to exactly compute (36). Instead, we randomly generate 500 samples from the Gaussian distribution and 500 samples from the uniform distribution as test samples to approximate it, i.e.,

$$\underset{t \in \mathbb{R}}{\text{minimize}} \left\{ t + \frac{1}{\alpha N_T} \sum_{i=1}^{N_T} \{ h(\mathbf{p}, t, \hat{\boldsymbol{\xi}}^i) \} \right\}$$

where $\hat{\boldsymbol{\xi}}^i$ is the i th test sample and N_T is the size of test dataset.

	α	ϵ_N	N
Impact of ϵ_N	0.1	{0, 0.001, 0.005, 0.01, 0.05, 0.1, 0.2, ..., 1}	{30, 100, 300}
Impact of N	0.1	0.1	{30, 50, ..., 290}

Table 2: Parameters in different experiments.

6.2. Performance of the DRSP Model Without Support Set

In this subsection we evaluate the performance of our DRSP model in (9).

6.2.1. Impact of the Wasserstein Radius

We first conduct experiments to test the impact of the Wasserstein radius ϵ_N on the out-of-sample performance of our DRSP model. Parameters α , ϵ_N and N are selected as the first row of Table 2.

Since $\|\mathbf{p}\|_\infty = 1$ for any feasible path \mathbf{p} , one can argue from (11c) that l_1 -norm in the Wasserstein distance is not sensible to our DRSP model without the support set. In this subsection we take l_2 -norm in the Wasserstein distance.

We perform 200 independent experiments and the averaged out-of-sample performance is shown in Figure 3. It reveals that the performance improves until the Wasserstein radius exceeds a certain value, and then deteriorates as the radius increases. As a result, the Wasserstein radius of our DRSP model should be selected based on the size of the sample dataset for good out-of-sample performance.

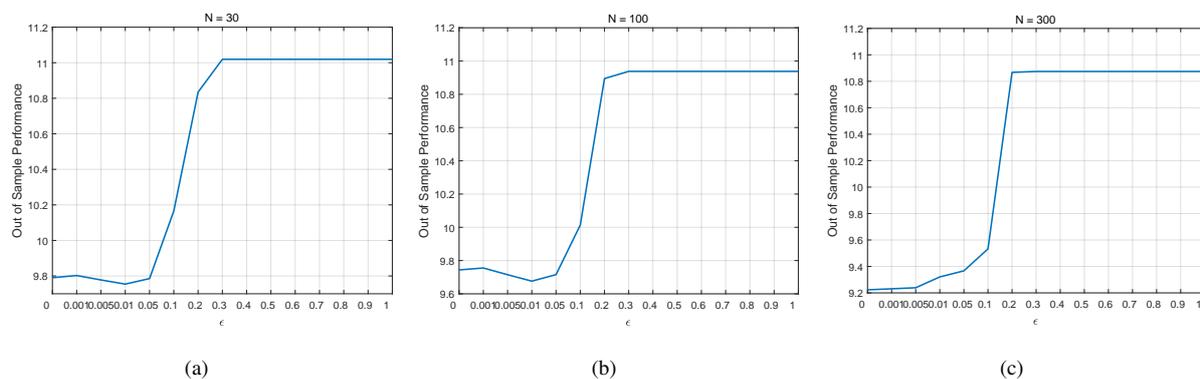


Figure 3: Averaged out-of-sample performance as a function of Wasserstein radius under sample datasets with different sizes. (a) $N = 30$, (b) $N = 100$, (c) $N = 300$.

6.2.2. Impact of the Sample Size

In this subsection we perform experiments on the sample dataset of different sizes to examine the impact of the sample sizes. Parameters in these experiments are set as the second row in Table 2. The averaged out-of-sample performance over 200 independent experiments is presented in Figure 4.

The improvement of the out-of-performance with increasing dataset sizes in Figure 4 validates the asymptotic consistency of our model as shown in Theorem 3.

6.3. Performance of the DRSP Model With Support Set

This subsection validates our DRSP model with the support set $\Xi = [\mathbf{a}, \mathbf{b}]$ in Section 3.2. We test the impact of radius ϵ_N and the sample size N on our DRSP model. All parameters are set as Table 2.

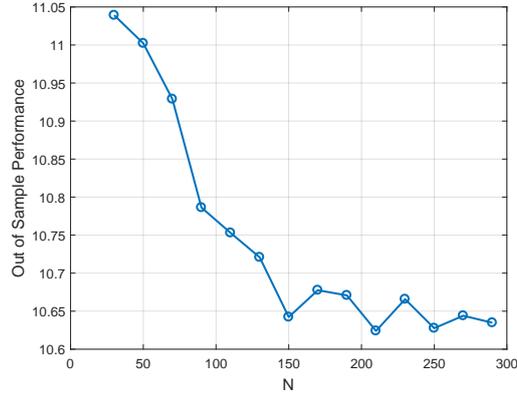


Figure 4: Averaged out-of-sample performance as a function of sample size N for 200 independent experiments, where the support set of travel time is \mathbb{R}^d .

6.3.1. Impact of the Wasserstein Radius

We first test the impact of radius ϵ_N . Figure 5 describes the averaged out-of-sample performance over 200 independent experiments under different Wasserstein radii ϵ_N . Similar to results in Section 6.2.1, the out-of-sample performance obtains its optimal value at a certain point and then deteriorates as the radius increases.

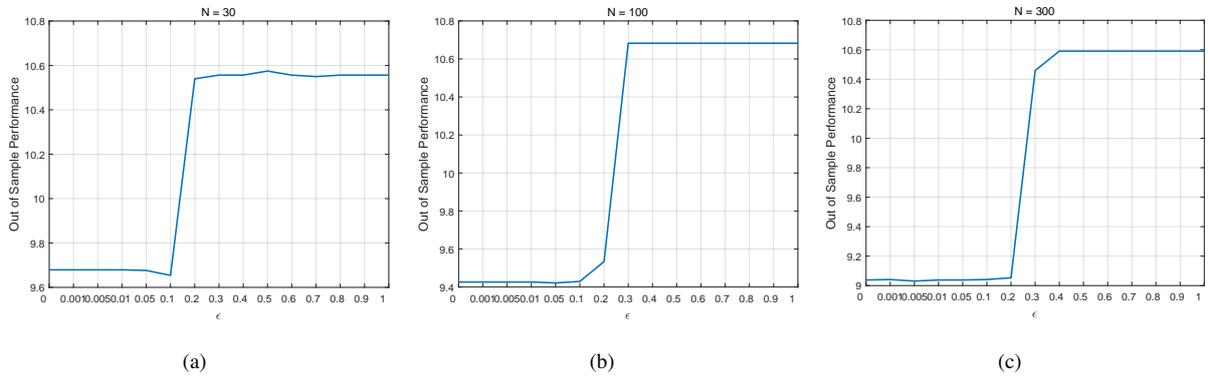


Figure 5: Averaged out-of-sample performance as a function of Wasserstein radius under the sample dataset with different sizes where the support set of travel time is $[a, b]$. (a) $N = 30$, (b) $N = 100$, (c) $N = 300$.

6.3.2. Impact of the Sample Size

Now, we test the impact of the sample size N . Figure 6 shows the out-of-sample performance averaged over 200 independent simulation runs as a function of N , in which the performance improves as N increases.

6.4. Comparisons with the state-of-the-art methods

In this subsection we compare our model with the SAA method and the DRSP model in Zhang et al. (2018) where the ambiguity set is based on moment constraints. We set $\alpha = 0.1$ and $N = \{20, 30, 50, 100, 200, 300, 500\}$ in

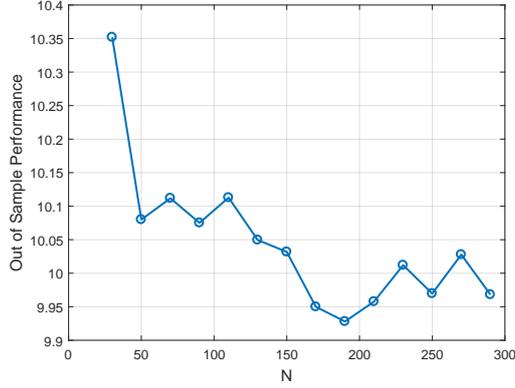


Figure 6: Averaged out-of-sample performance as a function of sample size N averaged over 200 independent simulations, where the support set of travel time is $[a, b]$.

this subsection. Noting that the Wasserstein radius ϵ_N depends on the training data. We tune it for different sample datasets to provide a powerful out-of-sample performance guarantee. Since the moment-based model is an intractable co-positive program problem, Zhang et al. (2018) derives dual approximation methods to provide a lower bound and an upper bound for the DR shortest path problem. They are denoted as M-LB and M-UB for short respectively.

We solve these models and evaluate their out-of-sample performance. For the purpose of comparison, we use the following percentage difference

$$\left(\frac{\text{DR}}{\text{SAA}} - 1 \right) \times 100\%$$

where DR denotes the out-of-sample performance of paths obtained from the DR shortest path model and SAA denotes that of the SAA method.

Comparisons in terms of the out-of-sample performance and the computation time are given in Table 3 and Table 4, respectively. A negative value in Table 3 indicates that the DR performs better than the SAA. One can observe that our DRSP model with the support set (denoted as DRSP-S) exhibits the best out-of-sample performance if the dataset is small, which however needs a longer computational time. As the dataset increases, our DRSP model without the support set also performs better than the SAA method and the moment-based DRSP model. More importantly, it can be solved in an appropriate time even when the sample set is large. Kindly note that both the exact mean and the variance are needed and essential to the moment-based DRSP model. From this perspective, they need more exact model information.

6.5. Real Road Network Experiments

Now we evaluate our DRSP model on a road network with a real dataset. We compare our model with the SAA and the moment-based model. In this experiment, both α and N are the same as these in Section 6.4 and the Wasserstein radius ϵ_N is tuned according to the sample dataset.

Method	Number of Samples						
	20	30	50	100	200	300	500
DRSP	1.6	-0.3	-1.8	-2.5	-3.2	-5.2	-5.4
DRSP-S	-7.1	-6.4	-6.4	-6.1	-6.6	-5.0	-3.3
M-LB	-0.1	-0.2	-0.2	-0.5	-0.6	-0.3	-0.8
M-UB	0.6	-0.5	-1.6	-2.4	-2.0	-0.2	-0.6

Table 3: Percentage differences (in %) between the DR models and SAA in terms of out-of-sample performance.

Method	Number of Samples						
	20	30	50	100	200	300	500
DRSP	0.13	0.16	0.16	0.23	1.93	3.16	3.85
DRSP-S	0.66	1.05	2.02	5.93	18.26	37.48	100.72
M-LB	68.16	113.13	180.66	413.46	844.54	1158.12	2090.82
M-UB	3.40	3.73	3.71	6.44	11.66	23.20	29.77
SAA	0.73	0.83	1.89	2.43	3.98	7.58	11.50

Table 4: Averaged computation time (second) of different methods.

We construct a dataset of travel time from Tsinghua University (THU) to Beijing Capital International Airport (BCIA) captured from the AMAP which provides a live traffic data interface¹. Firstly we select twenty-one paths which individuals usually take from THU to BCIA as illustrated in Figure 7. We set THU as the origin vertex and BCIA as the destination vertex. Moreover, we select several way-points on each path as vertices in the network and then provide an illustrative network of the map in Figure 8.



Figure 7: Map with 21 paths from THU to BCIA.

¹<https://lbs.amap.com/api/webservice/guide/api/direction#driving>

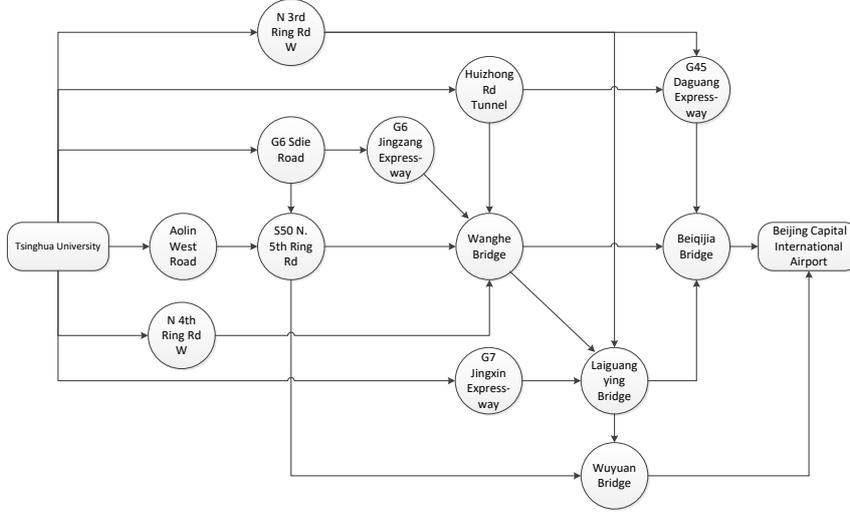


Figure 8: Simplified network between THU and BCIA.

We collect travel time of each arc along paths as shown in Figure 8 and update it every 10 minutes over a week spanning from Sunday morning, January 06, 2019 to Sunday morning, January 13, 2019. Thus, we have collected $24 \cdot 6 \cdot 7 = 1008$ different observations where each data point contains the travel time of all arcs.

Since the true distribution F of the travel time is unknown, it is unable to exactly compute the out-of-sample performance. Similar to Section 6.1, we utilize 300 randomly chosen test samples from the dataset to approximate the out-of-sample performance.

It should be noticed that we do not consider transportation expenses such as tolls for high-speed and bridge maintenance of roads. Thus, the optimal path may charge more than others. Moreover, the time spent on waiting for the traffic light is ignored as well.

Method	Number of Samples						
	20	30	50	100	200	300	500
DRSP	-0.05	-0.5	-2.3	-2.4	-3.5	-3.4	-3.2
DRSP-S	-1.3	-2.1	-4.3	-3.8	-3.3	-4.3	-3.7
M-LB	0.9	0.6	-1.5	-2.1	-2.3	-1.9	-2.0
M-UB	0.8	0.5	-1.9	-1.7	-1.9	-2.1	-1.8

Table 5: Percentage differences (in %) between the DR solutions and SAA solutions for out-of-sample performance.

Results averaged on 200 independent simulations are given in Table 5 and Table 6 which confirm the advantages of our DRSP models as expected.

Method	Number of Samples						
	20	30	50	100	200	300	500
DRSP	0.08	0.10	0.11	0.13	0.21	0.33	0.52
DRSP-S	0.11	0.13	0.17	0.27	0.58	1.01	1.82
M-LB	0.14	0.31	0.30	0.96	2.98	6.93	18.23
M-UB	0.40	0.33	0.37	0.56	0.84	1.27	1.96
SAA	0.21	0.21	0.25	0.47	0.80	1.20	2.15

Table 6: Averaged computation time (second) of different methods in different experiments.

7. Conclusion

We have proposed a data-driven DRSP model for finding optimal paths to minimize the w-METT over a Wasserstein ball. Our DRSP model can be reformulated as a solvable finite convex problem while the DRSP model over the moment-based ambiguity set is an intractable co-positive program. We derived an explicit form of the worst-case distribution in w-METT. Experimental results validate that the proposed DRSP model provides good out-of-sample performance. Moreover, our model can be extended to the DR bi-criteria shortest path problem and the minimum cost flow problem easily.

Acknowledgements

The authors would like to thank the Associate Editor and anonymous reviewers for their very constructive comments, which greatly improved the quality of this work. This work is supported by the National Natural Science Foundation of China under grant number 61722308, U1660202, and 71871023.

References

References

- Ahuja, R. K., Magnanti, T. L., Orlin, J. B., & Weihe, K. (1995). Network flows: theory, algorithms and applications. *ZOR-Methods and Models of Operations Research*, 41, 252–254.
- Ambrosio, L., & Gigli, N. (2013). A users guide to optimal transport. In *Modelling and optimisation of flows on networks* (pp. 1–155). Springer.
- Baxter, M., Elgindy, T., Ernst, A. T., Kalinowski, T., & Savelsbergh, M. W. (2014). Incremental network design with shortest paths. *European Journal of Operational Research*, 238, 675–684.
- Bertsekas, D. P., & Tsitsiklis, J. N. (1991). An analysis of stochastic shortest path problems. *Mathematics of Operations Research*, 16, 580–595.
- Cantelli, F. P. (1933). Sulla determinazione empirica delle leggi di probabilita (On the empirical determination of the laws of probability). *Giorn. Ist. Ital. Attuari*, 4, 421–424.
- Cao, Z., Guo, H., Zhang, J., Niyato, D., & Fastenrath, U. (2016). Finding the shortest path in stochastic vehicle routing: A cardinality minimization approach. *IEEE Transactions on Intelligent Transportation Systems*, 17, 1688–1702.
- Carlsson, J. G., Behroozi, M., & Mihic, K. (2018). Wasserstein distance and the distributionally robust tsp. *Operations Research*, 66, 1603–1624.

- Chassein, A., Dokka, T., & Goerigk, M. (2019). Algorithms and uncertainty sets for data-driven robust shortest path problems. *European Journal of Operational Research*, 274, 671–686.
- Chen, A., & Zhou, Z. (2010). The α -reliable mean-excess traffic equilibrium model with stochastic travel times. *Transportation Research Part B: Methodological*, 44, 493–513.
- Cheng, J., Lisser, A., & Letournel, M. (2013). Distributionally robust stochastic shortest path problem. *Electronic Notes in Discrete Mathematics*, 41, 511–518.
- Delage, E., & Ye, Y. (2010). Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research*, 58, 595–612.
- Duran, M. A., & Grossmann, I. E. (1986). An outer-approximation algorithm for a class of mixed-integer nonlinear programs. *Mathematical Programming*, 36, 307–339.
- Erdođan, E., & Iyengar, G. (2006). Ambiguous chance constrained problems and robust optimization. *Mathematical Programming*, 107, 37–61.
- Esfahani, P. M., & Kuhn, D. (2018). Data-driven distributionally robust optimization using the Wasserstein metric: performance guarantees and tractable reformulations. *Mathematical Programming*, 171, 115–166.
- Fosgerau, M., & Engelson, L. (2011). The value of travel time variance. *Transportation Research Part B: Methodological*, 45, 1–8.
- Frank, H. (1969). Shortest paths in probabilistic graphs. *Operations Research*, 17, 583–599.
- Glivenko, V. (1933). Sulla determinazione empirica delle leggi di probabilita (On the empirical determination of the laws of probability). *Gion. Ist. Ital. Attauri.*, 4, 92–99.
- Gupta, O. K., & Ravindran, A. (1985). Branch and bound experiments in convex nonlinear integer programming. *Management Science*, 31, 1533–1546.
- Hu, Z., & Hong, J. (2012). Kullback-leibler divergence constrained distributionally robust optimization. *Available at Optimization Online*, http://www.optimization-online.org/DB_FILE/2012/11/3677.pdf.
- Jiang, R., & Guan, Y. (2016). Data-driven chance constrained stochastic program. *Mathematical Programming*, 158, 291–327.
- Kantorovich, L., & Rubinshtein, G. (1958). On a space of totally additive functions, *vestn. Vestn Lening. Univ*, 13, 52–59.
- Lo, H. K., & Tung, Y.-K. (2003). Network with degradable links: capacity analysis and design. *Transportation Research Part B: Methodological*, 37, 345–363.
- Mangasarian, O. L. (1994). *Nonlinear programming*. SIAM.
- Masson, M.-H., & Dencoux, T. (2006). Inferring a possibility distribution from empirical data. *Fuzzy Sets and Systems*, 157, 319–340.
- Nikolova, E., & Stier-Moses, N. E. (2014). A mean-risk model for the traffic assignment problem with stochastic travel times. *Operations Research*, 62, 366–382.
- Pflug, G., & Wozabal, D. (2007). Ambiguity in portfolio selection. *Quantitative Finance*, 7, 435–442.
- Rockafellar, R. T., Uryasev, S. et al. (2000). Optimization of conditional value-at-risk. *Journal of Risk*, 2, 21–42.
- Shahabi, M., Unnikrishnan*, A., & Boyles, S. D. (2015). Robust optimization strategy for the shortest path problem under uncertain link travel cost distribution. *Computer-Aided Civil and Infrastructure Engineering*, 30, 433–448.
- Shapiro, A. (2001). On duality theory of conic linear problems. *Nonconvex Optimization and Its Applications*, 57, 135–165.
- Shapiro, A., & Homem-de Mello, T. (1998). A simulation-based approach to two-stage stochastic programming with recourse. *Mathematical Programming*, 81, 301–325.
- Stabler, B., Bar-Gera, H., & Sall, E. (2019). Transportation networks for research. *Available at Optimization Online*, <https://github.com/bstabler/TransportationNetworks.09,24,2019>.
- Tilk, C., Rothenbächer, A.-K., Gschwind, T., & Irnich, S. (2017). Asymmetry matters: Dynamic half-way points in bidirectional labeling for solving shortest path problems with resource constraints faster. *European Journal of Operational Research*, 261, 530–539.
- Verweij, B., Ahmed, S., Kleywegt, A. J., Nemhauser, G., & Shapiro, A. (2003). The sample average approximation method applied to stochastic routing problems: a computational study. *Computational Optimization and Applications*, 24, 289–333.
- Wang, Z., You, K., Song, S., & Shang, C. (2019). Data-driven distributionally robust shortest path problem using the Wasserstein ambiguity set.

- In *2019 IEEE 15th International Conference on Control and Automation (ICCA)* (pp. 1391–1396). IEEE.
- Westerlund, T., & Pettersson, F. (1995). An extended cutting plane method for solving convex minlp problems. *Computers and Chemical Engineering, 19*, 131–136.
- Wozabal, D. (2012). A framework for optimization under ambiguity. *Annals of Operations Research, 193*, 21–47.
- Xie, W. (2018). On distributionally robust chance constrained programs with Wasserstein distance. [arXiv:arXiv:1806.07418](https://arxiv.org/abs/1806.07418).
- Yang, L., & Zhou, X. (2017). Optimizing on-time arrival probability and percentile travel time for elementary path finding in time-dependent transportation networks: Linear mixed integer programming reformulations. *Transportation Research Part B: Methodological, 96*, 68–91.
- Yu, G., & Yang, J. (1998). On the robust shortest path problem. *Computers & Operations Research, 25*, 457–468.
- Zhang, Y., Shen, Z.-J. M., & Song, S. (2017). Lagrangian relaxation for the reliable shortest path problem with correlated link travel times. *Transportation Research Part B: Methodological, 104*, 501–521.
- Zhang, Y., Song, S., Shen, Z. J. M., & Wu, C. (2018). Robust shortest path problem with distributional uncertainty. *IEEE Transactions on Intelligent Transportation Systems, 19*, 1080–1090.