Reading Notes: Wasserstein Distributionally Robust Shortest Path Problem

CO2MAKER

October 15, 2021

1 Background

Table 1: Mathematical Notations

$\mathcal{G}=(\mathcal{V},\mathcal{A})$	directed and connected network
$\mathcal{V}$	vertex set, $\vert\mathcal{V}\vert=m$
$\mathcal{A}$	arcs set, $\vert\mathcal{A}\vert=n$
$\boldsymbol{\xi}$	the travel time over all arcs
$\boldsymbol{p}$	binary decision variables, $\boldsymbol{p}=\{p_{ij}:p_{ij}\in\{0,1\},(i,j)\in\mathcal{A}\}$

The standard version of SSP is:

$$\min \quad \sum_{(i,j)\in\mathcal{A}} \xi_{ij}p_{ij} \tag{1}$$

$$s.t. \quad \begin{cases} \sum\limits_{j:(i,j)\in\mathcal{A}} p_{ij} - \sum\limits_{j:(i,j)\in\mathcal{A}} p_{ji} = b_i, & \forall i\in\mathcal{V} \\ p_{i,j}\in\{0,1\}, & \forall i,j\in\mathcal{A} \end{cases} \tag{2}$$

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

$$\mathcal{P}=\{\boldsymbol{p}\mid \boldsymbol{p}\text{ satisfies (2)}\} \tag{3}$$

In practice, the travel time $\boldsymbol{\xi}$ variability is unavoidable. Obviously, the travel time $\boldsymbol{\xi}$ has a significant impact on finding an optimal path for travelers. To quantify the reliability of a path, some criteria have been proposed, mean-excess travel time (METT) is one of such criteria.

Definition 1. *The α-reliable METT of path $\boldsymbol{p}$ is defined as:*

$$METT_\alpha(\boldsymbol{p})=\min_{t\in\mathbb{R}}\left\{t+\frac{1}{\alpha}\mathbb{E}_F\left\{h(\boldsymbol{p},t,\boldsymbol{\xi})\right\}\right\} \tag{4}$$

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

$$\min_{\boldsymbol{p}\in\mathcal{P}} \ METT_\alpha(\boldsymbol{p}) \tag{5}$$

However, solving the SSP model in (5) requires the exact distribution function F.

2 Motivation

Usually, the true distribution F in (5) 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 $[N]=\{1,...,N\}$. A natural idea is to adopt the Sample Average Approximation (SAA), F is approximated by an empirical distribution F_N over the sample dataset, i.e.

$$F_N(\boldsymbol{\xi}) = \frac{1}{N}\sum_{i=1}^{N}\mathbb{I}_{\{\hat{\boldsymbol{\xi}}^i\le\boldsymbol{\xi}\}} \tag{6}$$

where $\mathbb{I}_A$ is the indicator of event A. Then the SSP model in (5) is approximately solved by

$$\min_{t\in\mathbb{R},\boldsymbol{p}\in\mathcal{P}}\left\{t+\frac{1}{N}\sum_{i=1}^{N}h(\boldsymbol{p},t,\hat{\boldsymbol{\xi}}^i)\right\} \tag{7}$$

The empirical distribution F_N converges weakly to the true distribution F as N tends to infinity. 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
- the sample $\hat{\xi}^i$ if of low quality
- the distribution F may not be constant and is *time-varying*

SAA model in (7) may exhibit poor out-of-sample performance and is not always reliable. An alternative approach is data-driven robust optimization. The key idea is that the true distribution F is expected to *close* to the empirical distribution F_N with a high probability.

3 Formulation

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

$$d_W(F_1,F_2) = \inf\left\{\int_{\Xi\times\Xi} d(\boldsymbol{\xi}_1,\boldsymbol{\xi}_2)K(d\boldsymbol{\xi}_1,d\boldsymbol{\xi}_2):\right. \tag{8}$$

$$\left.\int_{\Xi}K(\boldsymbol{\xi}_1,d\boldsymbol{\xi}_2)=F_1(\boldsymbol{\xi}_1),\int_{\Xi}K(d\boldsymbol{\xi}_1,\boldsymbol{\xi}_2)=F_2(\boldsymbol{\xi}_2)\right\} \tag{9}$$

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

Then, the Wasserstein ball $\mathcal{F}_N$ is constructed as:

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

where $\epsilon_N\ge 0$ reflects the confidence in the empirical distribution F_N. And the more interested objective is the worst-case METT (w-METT) over the Wasserstein ball $\mathcal{F}_N$, i.e.

$$\text{w-METT} = \sup_{F\in\mathcal{F}_N}\ \text{METT}_\alpha(\boldsymbol{p}) \tag{11}$$

$$= \min_{t\in\mathbb{R}}\left\{t+\frac{1}{\alpha}\sup_{F\in\mathcal{F}_N}\mathbb{E}_F\left\{h(\boldsymbol{p},t,\boldsymbol{\xi})\right\}\right\} \tag{12}$$

Therefore, the corresponding DRSP model is:

$$\min_{\boldsymbol{p}\in\mathcal{P}}\quad \text{w-METT}_\alpha(\boldsymbol{p}) \tag{13}$$

4 Solution

In this paper, they provide two versions of model, with support set and without support set. They transform the DRSP model to a finite mixed 0-1 convex problem. The following shows a lemma which is cited from Zhang et al. (2017),

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

$$\sup_{\boldsymbol{x}\in\mathbb{R}^n} \left\{\boldsymbol{w}^\top \boldsymbol{x} - \lambda\|\boldsymbol{x}\|_p\right\} = \sup_{\boldsymbol{x}\in\mathbb{R}^n} \{(\|\boldsymbol{w}\|_q - \lambda)\|\boldsymbol{x}\|_p\} \tag{14}$$

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

4.1 Without Support Set

Theorem 1. *The w-METT over the Wasserstein ball $\mathcal{F}_N$ can be computed by a finite linear programming problem:*

$$\min_{t,\boldsymbol{s},\lambda} \quad t + \frac{1}{\alpha}\left\{\frac{1}{N}\sum_{i=1}^{N} s_i + \lambda\epsilon_N\right\} \tag{15}$$

$$s.t. \quad \begin{cases} \boldsymbol{p}^\top \hat{\boldsymbol{\xi}}^i - t \le s_i, s_i \ge 0, \forall i \in [N] \\ \|\boldsymbol{p}\|_q \le \lambda \end{cases} \tag{16}$$

Then, the DRSP model is equivalently reformulated as the following mixed 0-1 convex problem:

$$\min_{\boldsymbol{p},t,\boldsymbol{s},\lambda} \quad t + \frac{1}{\alpha}\left\{\frac{1}{N}\sum_{i=1}^{N} s_i + \lambda\epsilon_N\right\} \tag{17}$$

$$s.t. \quad \begin{cases} \boldsymbol{p}^\top \hat{\boldsymbol{\xi}}^i - t \le s_i, s_i \ge 0, \forall i \in [N] \\ \|\boldsymbol{p}\|_q \le \lambda \\ \boldsymbol{p} \in \mathcal{P} \end{cases} \tag{18}$$

Proof. w-METT is equivalent to the following problem:

$$\max_{K(\boldsymbol{\xi},\hat{\boldsymbol{\xi}}^i)\ge 0} \quad \int_\Xi \sum_{i=1}^{N} h(\boldsymbol{p}, t, \boldsymbol{\xi})K(d\boldsymbol{\xi}, \hat{\boldsymbol{\xi}}^i) \tag{19}$$

$$s.t. \quad \begin{cases} \int_\Xi K(d\boldsymbol{\xi}, \hat{\boldsymbol{\xi}}^i) = \frac{1}{N}, \forall i \in [N] \\ \int_\Xi \sum_{i=1}^{N} d(\boldsymbol{\xi}, \hat{\boldsymbol{\xi}}^i)K(d\boldsymbol{\xi}, \hat{\boldsymbol{xi}}^i) \le \epsilon_N \end{cases} \tag{20}$$

Then the Lagrange function is:

$$\mathcal{L}(\boldsymbol{\xi}, \lambda, \boldsymbol{s}) = \int_\Xi \sum_{i=1}^{N} h(\boldsymbol{p}, t, \boldsymbol{\xi})K(d\boldsymbol{\xi}, \hat{\boldsymbol{\xi}}^i) - \int_\Xi \sum_{i=1}^{N} s_i K(d\boldsymbol{\xi}, \hat{\boldsymbol{\xi}}^i) \tag{21}$$

$$- \int_\Xi \sum_{i=1}^{N} \lambda d(\boldsymbol{\xi}, \hat{\boldsymbol{\xi}}^i)K(d\boldsymbol{\xi}, \hat{\boldsymbol{\xi}}^i) + \frac{1}{N}\sum_{i=1}^{N} s_i + \lambda\epsilon_N \tag{22}$$

And then the Lagrange dual function is:

$$g(\lambda, \boldsymbol{s}) = \sup_{\boldsymbol{\xi}\in\Xi} \mathcal{L}(\boldsymbol{\xi}, \lambda, \boldsymbol{s}) \tag{23}$$

$$= \int_{\Xi} \sum_{i=1}^{N} \Big(h(\boldsymbol{p}, t, \boldsymbol{\xi}) - s_i - \lambda d(\boldsymbol{\xi}, \hat{\boldsymbol{\xi}}^i) \Big) K(d\boldsymbol{\xi}, \hat{\boldsymbol{\xi}}^i) + \frac{1}{N} \sum_{i=1}^{N} s_i + \lambda \epsilon_N \tag{24}$$

Consequently, the dual problem is:

$$\min_{\lambda, \boldsymbol{s}} \quad \frac{1}{N} \sum_{i=1}^{N} s_i + \lambda \epsilon_N \tag{25}$$

$$s.t. \quad \begin{cases} h(\boldsymbol{p}, t, \boldsymbol{\xi}) - \lambda d(\boldsymbol{\xi}, \hat{\boldsymbol{\xi}}^i) \leq s_i, \forall \boldsymbol{\xi} \in \Xi i \in [N] \\ \lambda \geq 0 \end{cases} \tag{26}$$

Since $h(\boldsymbol{p}, t, \boldsymbol{\xi}) = [\boldsymbol{\xi}^\top \boldsymbol{p} - t]^+$, then,

$$\sup_{\boldsymbol{\xi} \in \Xi} \Big\{ \boldsymbol{\xi}^\top \boldsymbol{p} - t - \lambda \| \boldsymbol{\xi} - \hat{\boldsymbol{\xi}}^i \|_p \Big\} \leq s_i \tag{27}$$

$$\sup_{\boldsymbol{\xi} \in \Xi} \Big\{ -\lambda \| \boldsymbol{\xi} - \hat{\boldsymbol{\xi}}^i \|_p \Big\} \leq s_i \tag{28}$$

And because $\lambda \geq 0, \boldsymbol{\xi} \in \Xi$, inequality (28) implies $s_i \geq 0$. Denote $\boldsymbol{\Delta u}_i = \boldsymbol{\xi} - \hat{\boldsymbol{\xi}}^i$, then the left hand side in (27) can be re-expressed as:

$$\begin{aligned} & \sup_{\Delta \boldsymbol{u}_i \in \Xi} \quad \Big\{ (\Delta \boldsymbol{u}_i + \hat{\boldsymbol{\xi}}^i)^\top \boldsymbol{p} - t - \lambda \| \Delta \boldsymbol{u}_i \|_p \Big\} && (29) \\ = & \sup_{\Delta \boldsymbol{u}_i \in \Xi} \quad \{ \boldsymbol{p}^\top \Delta \boldsymbol{u}_i - \lambda \| \Delta \boldsymbol{u}_i \|_p \} + \boldsymbol{p}^\top \hat{\boldsymbol{\xi}}^i - t && (30) \\ = & \sup_{\Delta \boldsymbol{u}_i \in \Xi} \quad \{ (\| \boldsymbol{p} \|_q - \lambda) \| \Delta \boldsymbol{u}_i \|_p \} + \boldsymbol{p}^\top \hat{\boldsymbol{\xi}}^i - t && (31) \\ = & \begin{cases} \boldsymbol{p}^\top \hat{\boldsymbol{\xi}}^i - t, & \| \boldsymbol{p} \|_q - \lambda \leq 0 \\ +\infty, & \| \boldsymbol{p} \|_q - \lambda > 0 \end{cases} && (32) \end{aligned}$$

Thus, problem (19) - (20) can be reformulated as:

$$\min_{\lambda, \boldsymbol{s}} \quad \frac{1}{N} \sum_{i=1}^{N} s_i + \lambda \epsilon_N \tag{33}$$

$$s.t. \quad \begin{cases} \boldsymbol{p}^\top \hat{\boldsymbol{\xi}}^i - t \leq s_i, \forall i \in N \\ \| \boldsymbol{p} \|_q - \lambda \leq 0 \\ \lambda \geq 0, s_i \geq 0, \forall i \in [N] \end{cases} \tag{34}$$

And the transformed DRSP model is:

$$\min_{\lambda, \boldsymbol{s}, \boldsymbol{p}} \quad \frac{1}{N} \sum_{i=1}^{N} s_i + \lambda \epsilon_N \tag{35}$$

$$s.t. \quad \begin{cases} \boldsymbol{p}^\top \hat{\boldsymbol{\xi}}^i - t \leq s_i, \forall i \in N \\ \| \boldsymbol{p} \|_q - \lambda \leq 0 \\ \lambda \geq 0, s_i \geq 0, \forall i \in [N] \\ \boldsymbol{p} \in \mathcal{P} \end{cases} \tag{36}$$

The theorem is proved. □

4.2 With Support Set

The traveling time is finite in practice, and thus its support set should not be neglected.

Theorem 2. *Let $\Xi = [\boldsymbol{a}, \boldsymbol{b}]$, then the w-METT over the Wasserstein ball can be computed by a finite convex problem:*

$$\min_{t,\boldsymbol{s},\lambda,\boldsymbol{\gamma}_i,\boldsymbol{\eta}_i} \quad t + \frac{1}{\alpha}\left\{\frac{1}{N}\sum_{i=1}^{N} s_i + \lambda\epsilon_N\right\} \tag{37}$$

$$s.t. \quad \begin{cases} (\boldsymbol{p}+\boldsymbol{\gamma}_i-\boldsymbol{\eta}_i)^\top\hat{\boldsymbol{\xi}}^i - \boldsymbol{\gamma}_i^\top\boldsymbol{a} + \boldsymbol{\eta}_i^\top\boldsymbol{b} - t \le s_i, \forall i \in [N] \\ \|\boldsymbol{\gamma}_i + \boldsymbol{p} - \boldsymbol{\eta}_i\|_q \le \lambda \\ \boldsymbol{\eta}_i \ge 0, \boldsymbol{\gamma}_i \ge 0, s_i \ge 0, \forall i \in [N] \end{cases} \tag{38}$$

Moreover, the DRSP problem is re-expressed as:

$$\min_{\boldsymbol{p},t,\boldsymbol{s},\lambda,\boldsymbol{\gamma}_i,\boldsymbol{\eta}_i} \quad t + \frac{1}{\alpha}\left\{\frac{1}{N}\sum_{i=1}^{N} s_i + \lambda\epsilon_N\right\} \tag{39}$$

$$s.t. \quad \begin{cases} (\boldsymbol{p}+\boldsymbol{\gamma}_i-\boldsymbol{\eta}_i)^\top\hat{\boldsymbol{\xi}}^i - \boldsymbol{\gamma}_i^\top\boldsymbol{a} + \boldsymbol{\eta}_i^\top\boldsymbol{b} - t \le s_i, \forall i \in [N] \\ \|\boldsymbol{\gamma}_i + \boldsymbol{p} - \boldsymbol{\eta}_i\|_q \le \lambda \\ \boldsymbol{\eta}_i \ge 0, \boldsymbol{\gamma}_i \ge 0, s_i \ge 0, \forall i \in [N] \\ \boldsymbol{p} \in \mathcal{P} \end{cases} \tag{40}$$

Proof. w-METT can be reformulated as:

$$\min_{t,\boldsymbol{s},\lambda\ge 0} \quad t + \frac{1}{\alpha}\left\{\frac{1}{N}\sum_{i=1}^{N} s_i + \lambda\epsilon_N\right\} \tag{41}$$

$$s.t. \quad h(\boldsymbol{p}, t, \boldsymbol{\xi}) - \lambda d(\boldsymbol{\xi}, \hat{\boldsymbol{\xi}}^i) \le s_i, \forall \boldsymbol{\xi} \in \Xi, i \in [N] \tag{42}$$

Constraints (42) can be represented as:

$$\sup_{\boldsymbol{\xi}\in\Xi}\left\{\boldsymbol{\xi}^\top\boldsymbol{p} - t - \lambda\|\boldsymbol{\xi} - \hat{\boldsymbol{\xi}}^i\|_p\right\} \le s_i \tag{43}$$

$$\sup_{\boldsymbol{\xi}\in\Xi}\left\{-\lambda\|\boldsymbol{\xi} - \hat{\boldsymbol{\xi}}^i\|_p\right\} \le s_i \tag{44}$$

And denote $\Delta\boldsymbol{u}_i = \boldsymbol{\xi} - \hat{\boldsymbol{\xi}}^i$, the Lagrange function of the lhs of inequality (43):

$$\mathcal{L}(\Delta\boldsymbol{u}_i, \boldsymbol{\gamma}_i, \boldsymbol{\eta}_i) = (\boldsymbol{p}+\boldsymbol{\gamma}_i-\boldsymbol{\eta}_i)^\top(\Delta\boldsymbol{u}_i + \hat{\boldsymbol{\xi}}^i) - \lambda\|\Delta\boldsymbol{u}_i\|_p - \boldsymbol{\gamma}_i^\top\boldsymbol{a} + \boldsymbol{\eta}^\top\boldsymbol{b} - t \tag{45}$$

Then, the Lagrangian dual function is:

$$\begin{aligned} g(\boldsymbol{\gamma}_i, \boldsymbol{\eta}_i) &= \sup_{\Delta\boldsymbol{u}_i} \mathcal{L}(\Delta\boldsymbol{u}_i, \boldsymbol{\gamma}_i, \boldsymbol{\eta}_i) && (46)\\ &= \sup_{\Delta\boldsymbol{u}_i}\left\{(\boldsymbol{p}+\boldsymbol{\gamma}_i-\boldsymbol{\eta}_i)^\top\Delta\boldsymbol{u}_i - \lambda\|\Delta\boldsymbol{u}_i\|_p\right\} + (\boldsymbol{p}+\boldsymbol{\gamma}_i-\boldsymbol{\eta}_i)^\top\hat{\boldsymbol{\xi}}^i - \boldsymbol{\gamma}_i\boldsymbol{a} + \boldsymbol{\eta}_i^\top\boldsymbol{b} - t && (47)\\ &= \sup_{\Delta\boldsymbol{u}_i}\left\{(\|\boldsymbol{p}+\boldsymbol{\gamma}_i-\boldsymbol{\eta}_i\|_q - \lambda)\|\Delta\boldsymbol{u}_i\|_p\right\} + (\boldsymbol{p}+\boldsymbol{\gamma}_i-\boldsymbol{\eta}_i)^\top\hat{\boldsymbol{\xi}}^i - \boldsymbol{\gamma}_i\boldsymbol{a} + \boldsymbol{\eta}_i^\top\boldsymbol{b} - t && (48)\\ &= \begin{cases} (\boldsymbol{p}+\boldsymbol{\gamma}_i-\boldsymbol{\eta}_i)^\top\hat{\boldsymbol{\xi}}^i - \boldsymbol{\gamma}_i\boldsymbol{a} + \boldsymbol{\eta}_i^\top\boldsymbol{b} - t, & \|\boldsymbol{p}+\boldsymbol{\gamma}_i-\boldsymbol{\eta}_i\|_q - \lambda \le 0 \\ +\infty, & \|\boldsymbol{p}+\boldsymbol{\gamma}_i-\boldsymbol{\eta}_i\|_q - \lambda > 0 \end{cases} && (49) \end{aligned}$$

Consequently, the lhs of inequality (43) admits an equivalent problem:

$$\min_{\boldsymbol{\gamma}_i,\boldsymbol{\eta}_i} \quad (\boldsymbol{p}+\boldsymbol{\gamma}_i-\boldsymbol{\eta}_i)^\top \hat{\boldsymbol{\xi}}^i - \boldsymbol{\gamma}_i \boldsymbol{a} + \boldsymbol{\eta}_i^\top \boldsymbol{b} - t \tag{50}$$

$$s.t. \quad \begin{cases} \|\boldsymbol{p}+\boldsymbol{\gamma}_i-\boldsymbol{\eta}_i\|_q - \lambda \le 0 \\ \boldsymbol{\gamma}_i \ge 0, \boldsymbol{\eta}_i \ge 0 \end{cases} \tag{51}$$

Substitute the above problem into constraints (42). And the DRSP model eventually given by:

$$\min_{\boldsymbol{p},t,\boldsymbol{s},\lambda,\boldsymbol{\gamma}_i,\boldsymbol{\eta}_i} \quad t + \frac{1}{\alpha}\left\{\frac{1}{N}\sum_{i=1}^{N} s_i + \lambda\epsilon_N\right\} \tag{52}$$

$$s.t. \quad \begin{cases} (\boldsymbol{p}+\boldsymbol{\gamma}_i-\boldsymbol{\eta}_i)^\top \hat{\boldsymbol{\xi}}^i - \boldsymbol{\gamma}_i \boldsymbol{a} + \boldsymbol{\eta}_i^\top \boldsymbol{b} - t \le s_i, \forall i \in [N] \\ \|\boldsymbol{p}+\boldsymbol{\gamma}_i-\boldsymbol{\eta}_i\|_q - \lambda \le 0 \\ \boldsymbol{p} \in \mathcal{P}, t \ge 0, \lambda \ge 0, s_i \ge 0, \boldsymbol{\gamma}_i \ge 0, \boldsymbol{\eta}_i \ge 0, \forall i \in [N] \end{cases} \tag{53}$$

The theorem is proved. □

5 Conclusion

In this paper, they study a distributionally robust version of classical shortest path problem (DRSP). The ambiguity set is constructed as a Wasserstein ball. And through Lagrangian duality, they transform the problem into tractable convex optimization problem. Experiments demonstrate the advantages of the presented model in terms of the out-of-sample performance and computational complexity.

References

Zhang, Y., Z.-J. Max Shen, and S. Song (2017). Lagrangian relaxation for the reliable shortest path problem with correlated link travel times. *Transportation Research Part B: Methodological 104*, 501–521.