Online Supplement of the Paper "Multi-reference Distributionally Robust Resource Allocation with Trust-aided Parametric Data Fusion"

Yanru Guo, Bo Zhou, Ruiwei Jiang, Siqian Shen, and Xi (Jessie) Yang

I. Proof for Theorem I

A. Problem Description and Notation

The notation we use in this appendix is consistent with the paper. Let K denote the number of regions in the resource allocation problem. Let $\boldsymbol{c}^u = [c_1^u, ..., c_K^u]^{\mathrm{T}}$ and $\boldsymbol{c}^o = [c_1^o, ..., c_K^o]^{\mathrm{T}}$, $\boldsymbol{c}^u$ and $\boldsymbol{c}^o \in \mathbb{R}^K$ be the vectors of the unit penalty costs of unmet demand and overmet demand in each subregion, respectively. The allocation plan should be decided under a resource budget $B > 0$. The demands in all subregions for wildfire suppression resources are uncertain and are captured by the random vector $\boldsymbol{\xi} = [\xi_1, ..., \xi_K]^{\mathrm{T}}$, where $\boldsymbol{\xi} \in \mathbb{R}_+^K$ obeys a probability distribution $\mathbb{P}$. We define decision variable $x_k \geq 0$ for all $k \in [K]$ as the amount of resource assigned to region k.

For simplicity, we let $\ell(\boldsymbol{x}, \boldsymbol{\xi}) = (\boldsymbol{c}^u)^{\mathrm{T}}(\boldsymbol{\xi} - \boldsymbol{x})^+ + (\boldsymbol{c}^o)^{\mathrm{T}}(\boldsymbol{x} - \boldsymbol{\xi})^+$ and then we have the DRO model:

$$\inf_{\boldsymbol{x} \in \mathbb{X}} \left\{ \sup_{\mathbb{P} \in \mathcal{P}} \mathop{\mathbb{E}}_{\boldsymbol{\xi} \sim \mathbb{P}} [\ell(\boldsymbol{x}, \boldsymbol{\xi})] \right\}, \tag{1}$$

where the definition of ambiguity set $\mathcal{P}$ is given by:

$$\mathcal{P} := \left\{ \mathbb{P} \in \mathcal{M}(\Xi) : d_W(\mathbb{P}, \hat{\mathbb{P}}_N) \leq \varepsilon \right\}. \tag{2}$$

Here, the ambiguity set $\mathcal{P}$ is a Wassertein ambiguity set centered at the empirical distribution $\hat{\mathbb{P}}_N$, and is constructed based on the Wasserstein metric using L_1-norm. We denote Ξ as the Cartesian product of closed convex sets Ξ_k, where $\xi_k \in \Xi_k$ for all $k \in [K]$. We define the Wasserstein metric on the space $\mathcal{M}(\Xi)$ of all probability distributions $\mathbb{P}$ supported on Ξ with $\mathbb{E}^{\mathbb{P}}[\|\boldsymbol{\xi}\|] = \int_\Xi \|\boldsymbol{\xi}\| \mathbb{P}(d\boldsymbol{\xi}) < \infty$ and $d_W : \mathcal{M}(\Xi) \times \mathcal{M}(\Xi) \to \mathbb{R}_+$ via:

$$d_W(\mathbb{P}, \hat{\mathbb{P}}_N) = \inf \left\{ \int_{\Xi^2} \|\boldsymbol{\xi} - \hat{\boldsymbol{\xi}}\| \Pi(d\boldsymbol{\xi}, d\hat{\boldsymbol{\xi}}) \right\}, \tag{3}$$

where $\|\cdot\|$ represents L_1-norm, and Π is the joint distribution of $\boldsymbol{\xi}$ and $\hat{\boldsymbol{\xi}}$ with marginal distributions $\mathbb{P}$ and $\hat{\mathbb{P}}_N$ respectively and $\mathbb{P}, \hat{\mathbb{P}}_N \in \mathcal{M}(\Xi)$.

The optimization objective of (1) is to minimize the worst-case expected losses, where we try to minimize the maximum loss taken over the ambiguity set $\mathcal{P}$. The ambiguity set can be viewed as a Wasserstein ball with a radius ε centered at the empirical distribution $\hat{\mathbb{P}}_N$. In our problem, the empirical distribution is sampled from a normal distribution we construct based on information provided by different sources. For details, please refer to Section II C in the original paper.

B. DRO Reformulation

We show how to reformulate the MR-DRO based on the way of constructing the empirical distribution used in the ambiguity set in this section.

Proof: By using definition (3), we can re-express the worst-case expectation we want to minimizi in the DRO model (1) as

$$\sup_{\mathbb{P} \in \mathcal{P}} \mathop{\mathbb{E}}_{\boldsymbol{\xi} \sim \mathbb{P}} [\ell(\boldsymbol{\xi})] = \begin{cases} \sup_{\Pi, \mathbb{P}} \int_\Xi \ell(\boldsymbol{\xi}) \mathbb{P}(d\boldsymbol{\xi}) \\ \text{s.t.} \quad \int_{\Xi^2} \|\boldsymbol{\xi} - \hat{\boldsymbol{\xi}}\| \Pi(d\boldsymbol{\xi}, d\hat{\boldsymbol{\xi}}) \leq \varepsilon \end{cases} \tag{4}$$

$$= \begin{cases} \sup_{\mathbb{P}_i \in \mathcal{M}(\Xi)} \frac{1}{N} \sum_{i=1}^N \int_\Xi \ell(\boldsymbol{\xi}) \mathbb{P}_i(d\boldsymbol{\xi}) \\ \text{s.t.} \quad \frac{1}{N} \sum_{i=1}^N \int_\Xi \|\boldsymbol{\xi} - \hat{\boldsymbol{\xi}}_i\| \mathbb{P}_i(d\boldsymbol{\xi}) \leq \varepsilon, \end{cases} \tag{5}$$

where Π is the joint distribution of $\boldsymbol{\xi}$ and $\hat{\boldsymbol{\xi}}$ with marginals $\mathbb{P}$ and $\hat{\mathbb{P}}_N$. We drop the minimization problem in the constraint of (4) since the minimization of the Wasserstein metric $d_W(\mathbb{P}, \hat{\mathbb{P}}_N)$ is less than equal to radius ε is equivalent as (4) has feasible solution. The second equality (5) means that any probability distribution Π of $\boldsymbol{\xi}$ and $\hat{\boldsymbol{\xi}}$ can be constructed from the marginal distribution $\hat{\mathbb{P}}_N$ of $\hat{\boldsymbol{\xi}}$ and the conditional distribution $\mathbb{P}_i$ of $\boldsymbol{\xi}$ given $\hat{\boldsymbol{\xi}} = \hat{\boldsymbol{\xi}}_i$, for all $i \in [N]$. Following the standard duality argument [1], we obtain

$$\sup_{\mathbb{P} \in \mathcal{P}} \mathop{\mathbb{E}}_{\xi \sim \mathbb{P}} [\ell(\boldsymbol{\xi})] = \sup_{\mathbb{P}_i \in \mathcal{M}(\Xi)} \Big[\inf_{\lambda \geq 0} \frac{1}{N} \int_\Xi \ell(\boldsymbol{\xi}) \mathbb{P}_i(d\boldsymbol{\xi}) \tag{6}$$

$$+ \lambda (\varepsilon - \frac{1}{N} \sum_{i=1}^N \int_\Xi \|\boldsymbol{\xi} - \hat{\boldsymbol{\xi}}_i\| \mathbb{P}_i(d\boldsymbol{\xi})) \Big]$$

$$\leq \inf_{\lambda \geq 0} \Big[\sup_{\mathbb{P}_i \in \mathcal{M}(\Xi)} \lambda \varepsilon + \frac{1}{N} \sum_{i=1}^N \int_\Xi (\ell(\boldsymbol{\xi}) - \lambda \|\boldsymbol{\xi} - \hat{\boldsymbol{\xi}}_i\| \mathbb{P}_i(d\boldsymbol{\xi})) \Big] \tag{7}$$

$$= \inf_{\lambda \geq 0} \lambda \varepsilon + \frac{1}{N} \sum_{i=1}^N \sup_{\boldsymbol{\xi} \in \Xi} (\ell(\boldsymbol{\xi}) - \lambda \|\boldsymbol{\xi} - \hat{\boldsymbol{\xi}}_i\|), \tag{8}$$

where (7) holds because of the max-min inequality, and (8) follows from the fact that $\mathcal{M}(\Xi)$ contains all the Dirac distributions supported on Ξ. Meanwhile, the loss function in our problem is additively separable with respect to the temporal structure of $\boldsymbol{\xi}$, that is,

$$\ell(\boldsymbol{\xi}) := \sum_{k=1}^K \max_{j \in [J]} \ell_{jk}(\xi_k),$$

where $\ell_{jk} : \mathbb{R} \to \bar{\mathbb{R}}$ is a measurable function for any $j \in [J]$ and $k \in [K]$. Since we use L_1-norm to define the Wasserstein metric, $\|\cdot\|_{\mathrm{K}}$ reduces to L_1-norm on $\mathbb{R}^K$. Now, (8) can be written with the interchange of the summation and the maximization as

$$
\begin{aligned}
\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\boldsymbol{\xi}\sim\mathbb{P}}[\ell(\boldsymbol{\xi})] &= \inf_{\lambda\geq 0} \quad \lambda\varepsilon + \frac{1}{N}\sum_{i=1}^{N}\sup_{\boldsymbol{\xi}\in\Xi}(\ell(\boldsymbol{\xi}) - \lambda\|\boldsymbol{\xi} - \hat{\boldsymbol{\xi}}_i\|) \\
&= \inf_{\lambda\geq 0} \quad \lambda\varepsilon + \frac{1}{N}\sum_{i-1}^{N}\sum_{k=1}^{K}\sup_{\xi_k\in\Xi_k}\big(\max_{j=1,\dots,J}\ell_{jk}(\xi_k) \\
&\qquad - \lambda\|\xi_k - \hat{\xi}_{ik}\|\big).
\end{aligned} \tag{9}
$$

After introducing auxiliary variables in (10), we have

$$
\begin{cases}
\inf_{\lambda,s_{ik}} \quad \lambda\varepsilon + \frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K} s_{ik} \\
\text{s.t.} \quad \sup_{\xi_k\in\Xi_k}(\ell_{jk}(\xi_k) - \lambda\|\xi_k - \hat{\xi}_{ik}\|) \leq s_{ik} \\
\qquad \forall i\in[N], j\in[J], k\in[K] \\
\quad \lambda \geq 0
\end{cases} \tag{10}
$$

$$
\leq \begin{cases}
\inf_{\lambda,s_{ik},z_{ijk}} \quad \lambda\varepsilon + \frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K} s_{ik} \\
\text{s.t.} \quad \sup_{\xi_k\in\Xi_k}(\ell_{jk}(\xi_k) - \langle z_{ijk}, \xi_k\rangle) + \langle z_{ijk}, \hat{\xi}_{ik}\rangle \leq s_{ik} \\
\qquad \forall i\in[N], j\in[J], k\in[K] \\
\quad \|z_{ijk}\|_* \leq \lambda \quad \forall i\in[N], j\in[J], k\in[K]
\end{cases} \tag{11}
$$

$$
= \begin{cases}
\inf_{\lambda,s_{ik},z_{ijk}} \quad \lambda\varepsilon + \frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K} s_{ik} \\
\text{s.t.} \quad [-\ell_{jk} + \chi_{\Xi_k}]^*(-z_{ijk}) + \langle z_{ijk}, \hat{\xi}_{ik}\rangle \leq s_{ik} \\
\qquad \forall i\in[N], j\in[J], k\in[K] \\
\quad \|z_{ijk}\|_* \leq \lambda \quad \forall i\in[N], j\in[J], k\in[K],
\end{cases} \tag{12}
$$

where the inequality holds as an equality provided that Ξ_k and $\{\ell_{jk}\}_{j\in[J]}$ satisfy the convexity assumption for all $k\in[K]$. Finally, by [2], the conjugate of $-\ell_{jk} + \chi_{\Xi_k}$ can be replaced by the inf-convolution of the conjugates of $-\ell_{jk}$ and χ_{Ξ_k}. By definition of the conjugacy operator, we have

$$
\begin{aligned}
[-\ell_{jk}]^*(z) = [-a_{jk}]^*(z) &= \sup_{\xi}\langle z, \xi_k\rangle + \langle a_{jk}, \xi_k\rangle + b_{jk} \\
&= \begin{cases} b_{jk} & \text{if} \quad z = -a_{jk}, \\ \infty & \text{else,} \end{cases}
\end{aligned}
$$

and

$$
\sigma_{\Xi_k(\nu)} = \begin{cases} \sup_{\xi_k} & \langle \nu, \xi_k\rangle \\ \text{s.t.} & C_k\xi_k \leq d_k \end{cases} = \begin{cases} \inf_{\gamma\geq 0} & \langle \gamma, d\rangle \\ \text{s.t.} & C_k^T\gamma = \nu, \end{cases}
$$

where the last equality follows from strong duality, which holds as the uncertainty set is non-empty. After bringing this form to the original Model (1), we obtain the following equivalent LP (13), which is (7) in the original paper.

$$
\inf_{\boldsymbol{x},\lambda,s_{ik},\gamma_{ijk}} \lambda\varepsilon + \frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K} s_{ik} \tag{13a}
$$

$$
\text{s.t.} \quad x \in \mathbb{X}, \tag{13b}
$$

$$
\begin{aligned}
b_{jk} + \langle a_{jk}, \hat{\xi}_{ik}\rangle + \langle \gamma_{ijk}, d_k - C_k\hat{\xi}_{ik}\rangle \leq s_{ik}, \\
i\in[N], j\in[J], k\in[K],
\end{aligned} \tag{13c}
$$

$$
\|C_k^T\gamma_{ijk} - a_{jk}\|_* \leq \lambda, \forall i\in[N], j\in[J], k\in[K], \tag{13d}
$$

$$
\gamma_{ijk} \geq 0, \quad \forall i\in[N], j\in[J], k\in[K]. \tag{13e}
$$

■

References

[1] D. Bertsimas and J. N. Tsitsiklis, *Introduction to linear optimization*. Athena scientific Belmont, MA, 1997, vol. 6.

[2] R. T. Rockafellar and R. J.-B. Wets, *Variational analysis*. Springer Science & Business Media, 2009, vol. 317.