Computationally Efficient Approximations for Distributionally Robust Optimization

Meysam Cheramin, Jianqiang Cheng
Department of Systems and Industrial Engineering, University of Arizona, Tucson, AZ 85721, USA, meysamcheramin@email.arizona.edu, jqcheng@email.arizona.edu

Ruiwei Jiang
Department of Industrial and Operations Engineering, University of Michigan, Ann Arbor, MI 48109, USA, ruiwei@umich.edu

Kai Pan
Department of Logistics and Maritime Studies, Faculty of Business, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong, kai.pan@polyu.edu.hk

Distributionally robust optimization (DRO) is a modeling framework in decision making under uncertainty where the probability distribution of a random parameter is unknown while its partial information (e.g., statistical properties) is available. In this framework, the unknown probability distribution is assumed to lie in an ambiguity set consisting of all distributions that are compatible with the available partial information. Although DRO bridges the gap between stochastic programming and robust optimization, one of its limitations is that its models for large-scale problems can be siginificantly difficult to solve, especially when the uncertainty is of high dimension. In this paper, we propose computationally efficient inner and outer approximations for DRO problems with a moment-based ambiguity set and a combined ambiguity set including Wasserstein distance and moment information. In these approximations, we split a random vector into smaller pieces, leading to smaller matrix constraints. In addition, we use principal component analysis to shrink uncertainty space dimensionality. We quantify the quality of the developed approximations by deriving theoretical bounds on their optimality gap. We display the practical applicability of the proposed approximations in a production-transportation problem and a multi-product newsvendor problem. The results demonstrate that these approximations dramatically reduce the computational time while maintaining high solution quality.

Key words: stochastic programming, distributionally robust optimization, moment information, Wasserstein distance, principal component analysis, semidefinite programming

1. Introduction

Uncertainty poses significant challenge to decision making in many real-world problems. To overcome such challenge, advanced optimization approaches have been developed to model uncertainty from various perspectives. Among them, stochastic programming (SP), robust optimization (RO), and distributionally robust optimization (DRO) prevail nowadays. SP assumes that a decision maker has complete knowledge about the probability distribution of the uncertain parameters, whereas the distribution may not be precisely estimated due to limited data availability (Shapiro et al. 2009). RO assumes the uncertain parameters run in a given set, and it hedges against the worst-case possible scenario within this set, leading to potentially conservative decisions (Ben-Tal and Nemirovski 1998, Bertsimas and Sim 2004). Scarf (1958) introduced the first DRO model by relaxing the complete-knowledge assumption in SP and reducing the conservativeness of RO. DRO models uncertainty through a distributional ambiguity set that specifies available information of the probability distribution of the uncertain parameters. In addition, DRO searches for an optimal solution that concerns the worst-case distribution in the ambiguity set. Thus, the performance of DRO is less conservative than RO; see Rahimian and Mehrotra (2019) for more details.

The performance of DRO highly depends on the ambiguity set. An ideal ambiguity set possesses four properties: (1) rich enough to contain the true distribution with high confidence; (2) small enough to exclude pathological distributions that make DRO solutions overly conservative; (3) calibrated easily from historical data; and (4) leading to a structured DRO model that is computationally tractable (Esfahani and Kuhn 2018). There are several different types of ambiguity sets. Moment-based ambiguity sets contain distributions that share the same moment information (Delage and Ye 2010). Distance-based ambiguity sets contain distributions that are close to a reference distribution with respect to a predetermined probability discrepancy metric. Probability discrepancies that have been extensively studied include Wasserstein distance (Esfahani and Kuhn 2018), phi-divergence (Ben-Tal et al. 2013, Hu and Hong 2013, Gotoh et al. 2018), and Prokhorov metric (Erdoğan and Iyengar 2006). Structural ambiguity sets contain distributions that share the same structural properties such as monotonicity, symmetry, and unimodality (Li et al. 2019). Hypothesis-test-based ambiguity sets contain distributions that pass a hypothesis test (e.g., $\mathcal{X}^2$-test, G-test) based on a given historical dataset and confidence level (Bertsimas et al. 2018a,b). Finally, likelihood-based ambiguity sets contain distributions that achieve a given level of likelihood evaluated under historical data (Wang et al. 2016).

Due to high complexity of the uncertainty involved in real-world problems, none of the individual ambiguity sets can perfectly perform under all circumstances. For example, moment-based ambiguity sets do not guarantee asymptotic consistency, i.e., they do not converge to the true distribution of the uncertain parameters even if the number of historical data points increases to infinity (Chen et al. 2019, Liu et al. 2019). Accordingly, we may combine two different types of ambiguity sets to construct a better one that enjoys the advantages of both. For example, we may consider a combined moment and Wasserstein ambiguity set. This combination can help exclude pathological distributions and result in a less conservative DRO model, which is also asymptotically consistent. Such benefits can be significant when the uncertainty is highly complex (Wang et al. 2018, Gao and Kleywegt 2017).

Many DRO problems can be reformulated or approximated by conic programming problems, including semidefinite programming (SDP), second-order cone programming (SOCP), copositive programming (CP), and completely positive programming (CPP). For example, Delage and Ye (2010) showed that the DRO model with support, mean, and covariance information can be reformulated as an SDP formulation; Natarajan et al. (2010) reformulated a class of robust expected utility models with known mean and covariance matrix as SOCP formulations; Li et al. (2019) reformulated chance constraints under unimodal distributions with known first and second moments as SOCP formulations; El Ghaoui et al. (2003) derived SDP and SOCP formulations for computing robust Value-at-Risk with various ambiguity sets. More SDP reformulations can be found in Cheng et al. (2014, 2016) and Zhang et al. (2018), and more SOCP reformulations were proposed by Li et al. (2018) and Mieth and Dvorkin (2018). Moreover, Hanasusanto and Kuhn (2018) proposed CP and CPP reformulations and approximations of two-stage DRO linear programs over Wasserstein ambiguity sets.

Although SDP formulations are polynomially solvable in theory, many of them require significant computational efforts, especially when the problem is complicated in its nature and the uncertainty is high-dimensional and/or correlated. For instance, solving large-scale SDP problems in practice can be computationally challenging because many high-dimensional matrix constraints may be present (Yang and Wu 2019). To overcome such challenges, several studies have developed approximate solution approaches to trade-off between solution quality and computational burden, including branch-and-bound, cutting-plane, interior point, and delayed constraint generation algorithms (Niu et al. 2019, Vandenberghe and Boyd 1996). In addition, Cheng et al. (2018) used principal component analysis (PCA), which represents the data variability by employing a linear combination of orthogonal eigenmodes (Wold et al. 1987), to consider only the dominant random variables and shrink the dimension of the uncertainty, leading to smaller-size SDP matrix constraints. In this paper, we provide a comprehensive study to derive computationally efficient approaches to solve DRO formulations with various types of ambiguity sets. We summarize our contribution as follows:

1. We derive computationally efficient inner and outer approximations of DRO problems with a moment-based ambiguity set accounting for the support, mean, and covariance of the uncertainty. The inner approximation is based on splitting a random vector into smaller sub-vectors and is parameterized by the number of split pieces. Such approximation appears to be new in the DRO literature. The outer approximation generalizes Cheng et al. (2018).
2. We quantify the quality of our approximations by deriving theoretical bounds on the gap between the optimal value of the DRO problems and those of their approximations. These

theoretical bounds guide us to select specific numbers of split pieces for reaching a predetermined error bound. They also allow us to trade-off between solution quality and computational burden of solving DRO formulations.

3. We extend the inner and outer approximations, as well as their theoretical bounds, to a combined ambiguity set that contain covariance information and the Wasserstein distance between the true distribution and an empirical distribution.
4. We perform extensive computational experiments to demonstrate our approximations on solving DRO formulations. Notably, while commercial solvers were unable to even find a feasible solution to most large-size instances, our inner and outer approximations quickly found solutions with optimality guarantee.

The remainder of this paper is organized as follows. In Section 2 (resp. Section 3), we study DRO with the moment-based ambiguity set (resp. the combined ambiguity set), propose its inner and outer approximations, and derive theoretical bounds of their optimality gaps. In Section 4, we perform extensive computational experiments on distributionally robust multiproduct newsvendor and production-transportation problems to evaluate the theoretical results and demonstrate the strength of the proposed approximations. Finally, Section 5 concludes the paper.

Notation

In this paper, scalar values are denoted by non-bold symbols, e.g., s and γ_1, while vectors are denoted in the column form by bold symbols, e.g., $\boldsymbol{x} = (x_1, \ldots, x_m)^\top$ and $\boldsymbol{q}$. Similarly, matrices are represented by bold capital symbols, e.g., $\boldsymbol{A}$ and $\boldsymbol{\Sigma}$, and the size of a matrix is indicated by $r \times c$, where r and c indicate the numbers of rows and columns, respectively. Italic subscripts indicate indices, e.g., s_k, while non-italic ones represent simplified specifications, e.g., $\boldsymbol{Q}_{\mathrm{r}}$. We use $\mathbb{E}_{\mathbb{P}}[\cdot]$ to represent expectation over distribution $\mathbb{P}$ and use "$\bullet$" to denote the inner product defined by $\boldsymbol{A} \bullet \boldsymbol{B} = \sum_{i,j} A_{ij} B_{ij}$, where $\boldsymbol{A}$ and $\boldsymbol{B}$ are two conformal matrices. If a matrix $\boldsymbol{M}$ is positive semi-definite (PSD), it is indicated by $\boldsymbol{M} \succeq 0$. Symbols $\|\cdot\|_1$ and $\|\cdot\|_2$ denote L1-Norm and L2-Norm, respectively. Symbol $\|\cdot\|_*$ represents the dual norm of $\|\cdot\|_1$. We reserve symbols $\mathcal{D}_{\cdot}$ and $\mathcal{S}_{\cdot}$ for ambiguity set and support, respectively. For any strictly positive integer number n, we use $[n]$ to represent the set $\{1, 2, \ldots, n\}$. The identity matrix of size m is denoted by $\boldsymbol{I}_m$. Symbols $\mathbf{0}_m$ and $\mathbf{0}_{r\times c}$ represent a zero vector of size m and a zero matrix of size $r \times c$, respectively.

2. Moment-based Ambiguity Set

In this section, we introduce a DRO problem with a moment-based ambiguity set. To solve it towards practical uses, we first recast it as an SDP formulation. In view of the computational challenge of solving the SDP problem in practice, we develop its inner and outer approximations that can be solved more efficiently. Moreover, we derive theoretical bounds for their optimality gaps as compared to the original DRO problem, leading to a quality measurement.

Given distribution $\mathbb{P}$ of the random vector $\boldsymbol{\xi} \in \mathbb{R}^m$, we seek an $\boldsymbol{x} \in \mathcal{X}$, which is a convex set in $\mathbb{R}^n$, to minimize the expectation of a convex function $f(\boldsymbol{x}, \boldsymbol{\xi})$ with respect to $\mathbb{P}$. We present this problem as the following stochastic program:

$$\min_{\boldsymbol{x} \in \mathcal{X}} \mathbb{E}_{\mathbb{P}}\left[f(\boldsymbol{x}, \boldsymbol{\xi})\right]. \tag{1}$$

Here $\mathbb{P}$ is assumed to be known exactly, which though in practice may not be precisely estimated due to limited data availability (e.g., missing data, lack of data, and expensive data acquirement). Nevertheless, some partial information of $\boldsymbol{\xi}$ (e.g., mean and covariance) can be easily obtained from historical data. Therefore, instead of solving Problem (1) with a given distribution, we may seek a risk-averse solution that hedges against all the possible distributions that share such available information, leading to the following DRO model with the available information collected in a distributional ambiguity set $\mathcal{D}_{\mathrm{M1}}$:

$$\min_{\boldsymbol{x} \in \mathcal{X}} \max_{\mathbb{P} \in \mathcal{D}_{\mathrm{M1}}} \mathbb{E}_{\mathbb{P}}\left[f(\boldsymbol{x}, \boldsymbol{\xi})\right]. \tag{DRO-M}$$

Depending on different available information, the ambiguity set can be different. In this section, we focus on moment information of $\boldsymbol{\xi}$ in $\mathcal{D}_{\mathrm{M1}}$ (see Delage and Ye (2010)), i.e.,

$$\mathcal{D}_{\mathrm{M1}}(\mathcal{S}, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \gamma_1, \gamma_2) = \left\{ \mathbb{P} \,\middle|\, \begin{array}{l} \mathbb{P}(\boldsymbol{\xi} \in \mathcal{S}) = 1 \\ \left(\mathbb{E}_{\mathbb{P}}[\boldsymbol{\xi}] - \boldsymbol{\mu}\right)^\top \boldsymbol{\Sigma}^{-1} \left(\mathbb{E}_{\mathbb{P}}[\boldsymbol{\xi}] - \boldsymbol{\mu}\right) \le \gamma_1 \\ \mathbb{E}_{\mathbb{P}}\left[(\boldsymbol{\xi} - \boldsymbol{\mu})(\boldsymbol{\xi} - \boldsymbol{\mu})^\top\right] \preceq \gamma_2 \boldsymbol{\Sigma} \end{array} \right\},$$

which specifies the support ($\mathcal{S}$), mean ($\boldsymbol{\mu}$), and covariance of random variable $\boldsymbol{\xi}$ that could be derived using available historical data. We assume that $\mathcal{S}$ is a convex set, $\boldsymbol{\mu}$ lies in the strict interior of $\mathcal{S}$, and $\boldsymbol{\Sigma}$ is a positive definite matrix. Parameters $\gamma_1 \ge 0$ and $\gamma_2 \ge 1$ are derived from historical data to control the size of the ambiguity set and the conservatism of optimal solutions. The three constraints in $\mathcal{D}_{\mathrm{M1}}$ describe that (1) the support of $\boldsymbol{\xi}$ is a subset of $\mathcal{S}$; (2) the mean of $\boldsymbol{\xi}$ lies in an ellipsoid of size γ_1 centered at $\boldsymbol{\mu}$; and (3) the centered second-order moment matrix is bounded by $\gamma_2 \boldsymbol{\Sigma}$ in a PSD sense.

Although (DRO-M) admits a convex reformulation (e.g., SDP reformulation; see Delage and Ye (2010)), as discussed above, solving it in practice can be very challenging. Instead, we can solve a good inner or outer approximation of (DRO-M) much more efficiently and obtain high-quality solutions. In the following, we derive an outer (resp. inner) approximation of (DRO-M), leading to a lower (resp. upper) bound, in Section 2.1 (resp. Section 2.3). We make the following assumption in this section for practical purpose.

Assumption 1. *Function* $f(\boldsymbol{x}, \boldsymbol{\xi})$ *is piecewise linear convex in* $\boldsymbol{\xi}$*, i.e.,* $f(\boldsymbol{x}, \boldsymbol{\xi}) = \max_{k=1}^{K}\left\{y_k^0(\boldsymbol{x}) + y_k(\boldsymbol{x})^\top \boldsymbol{\xi}\right\}$ *with both* $y_k(\boldsymbol{x}) = (y_k^1(\boldsymbol{x}), \ldots, y_k^m(\boldsymbol{x}))^\top$ *and* $y_k^0(\boldsymbol{x})$ *affine in* $\boldsymbol{x}$ *for any* $k \in [K]$*, and* $\mathcal{S}$ *is polyhedral, i.e.,* $\mathcal{S} = \{\boldsymbol{\xi} | \boldsymbol{A}\boldsymbol{\xi} \leq \boldsymbol{b}\}$ *with* $\boldsymbol{A} \in \mathbb{R}^{n \times m}$ *and* $\boldsymbol{b} \in \mathbb{R}^n$*, with at least one interior point.*

2.1. Lower Bound

We use PCA, which approximates a high-dimensional matrix by a lower-dimensional one, to reduce the size of (DRO-M) while maintaining high solution quality. First, we perform an eigenvalue decomposition on matrix $\boldsymbol{\Sigma}$, i.e., $\boldsymbol{\Sigma} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^\top = \boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}}(\boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}})^\top$, where $\boldsymbol{U} \in \mathbb{R}^{m \times m}$ is an orthogonal transformation matrix and $\boldsymbol{\Lambda} \in \mathbb{R}^{m \times m}$ is a diagonal matrix whose diagonal elements are in decreasing order. By letting $\boldsymbol{\xi}_{\mathrm{I}} = (\boldsymbol{U}\boldsymbol{\Lambda}^{-1/2})^\top(\boldsymbol{\xi} - \boldsymbol{\mu})$, we reformulate (DRO-M) as

$$\min_{\boldsymbol{x} \in \mathcal{X}} \max_{\mathbb{P}_{\mathrm{I}} \in \mathcal{D}_{\mathrm{M2}}} \mathbb{E}_{\mathbb{P}_{\mathrm{I}}}\left[f\left(\boldsymbol{x}, \boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}} + \boldsymbol{\mu}\right)\right], \tag{2}$$

where

$$\mathcal{D}_{\mathrm{M2}}(\mathcal{S}_{\mathrm{I}}, \gamma_1, \gamma_2) = \left\{ \mathbb{P}_{\mathrm{I}} \middle| \begin{array}{l} \mathbb{P}_{\mathrm{I}}(\boldsymbol{\xi}_{\mathrm{I}} \in \mathcal{S}_{\mathrm{I}}) = 1 \\ \mathbb{E}_{\mathbb{P}_{\mathrm{I}}}[\boldsymbol{\xi}_{\mathrm{I}}^\top]\mathbb{E}_{\mathbb{P}_{\mathrm{I}}}[\boldsymbol{\xi}_{\mathrm{I}}] \leq \gamma_1 \\ \mathbb{E}_{\mathbb{P}_{\mathrm{I}}}[\boldsymbol{\xi}_{\mathrm{I}}\boldsymbol{\xi}_{\mathrm{I}}^\top] \preceq \gamma_2 \boldsymbol{I}_m \end{array} \right\},$$

with $\mathcal{S}_{\mathrm{I}} := \left\{\boldsymbol{\xi}_{\mathrm{I}} \in \mathbb{R}^m : \boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}} + \boldsymbol{\mu} \in \mathcal{S}\right\}$.

Theorem 1. *If* $f(\boldsymbol{x}, \boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}} + \boldsymbol{\mu})$ *is* $\mathbb{P}_{\mathrm{I}}$*-integrable for any* $\mathbb{P}_{\mathrm{I}} \in \mathcal{D}_{\mathrm{M2}}$*, then* (DRO-M) *has the same optimal value as the following problem:*

$$\min_{\boldsymbol{x}, s, \boldsymbol{q}, \boldsymbol{Q}} \; s + \gamma_2 \boldsymbol{I}_m \bullet \boldsymbol{Q} + \sqrt{\gamma_1}\,\|\boldsymbol{q}\|_2 \tag{3a}$$

$$\text{s.t.} \quad s \geq f\left(\boldsymbol{x}, \boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}} + \boldsymbol{\mu}\right) - \boldsymbol{\xi}_{\mathrm{I}}^\top \boldsymbol{q} - \boldsymbol{\xi}_{\mathrm{I}}^\top \boldsymbol{Q}\boldsymbol{\xi}_{\mathrm{I}}, \; \forall \boldsymbol{\xi}_{\mathrm{I}} \in \mathcal{S}_{\mathrm{I}}, \tag{3b}$$

$$\boldsymbol{Q} \succeq 0, \; \boldsymbol{x} \in \mathcal{X},$$

where $\boldsymbol{q} \in \mathbb{R}^m$ *and* $\boldsymbol{Q} \in \mathbb{R}^{m \times m}$*.*

Proof. The result is deduced from Lemma 1 in Delage and Ye (2010). □

Problem (3) reduces to a SDP formulation with regard to a wide range of objective functions and support of uncertainty, which are specified in Assumption 1.

Proposition 1. *Under Assumption 1,* (DRO-M) *has the same optimal value as the following SDP formulation:*

$$Z_{\mathrm{M}}^*(m) := \min_{\boldsymbol{x}, s, \boldsymbol{\lambda}_k, \boldsymbol{q}, \boldsymbol{Q}} \; s + \gamma_2 \boldsymbol{I}_m \bullet \boldsymbol{Q} + \sqrt{\gamma_1}\,\|\boldsymbol{q}\|_2 \tag{4a}$$

$$\text{s.t.} \begin{bmatrix} s - y_k^0(\boldsymbol{x}) - \boldsymbol{\lambda}_k^\top \boldsymbol{b} - y_k(\boldsymbol{x})^\top \boldsymbol{\mu} + \boldsymbol{\lambda}_k^\top \boldsymbol{A}\boldsymbol{\mu} & \frac{1}{2}\left(\boldsymbol{q} + \left(\boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top \left(\boldsymbol{A}^\top \boldsymbol{\lambda}_k - y_k(\boldsymbol{x})\right)\right)^\top \\ \frac{1}{2}\left(\boldsymbol{q} + \left(\boldsymbol{U}\Lambda^{\frac{1}{2}}\right)^\top \left(\boldsymbol{A}^\top \boldsymbol{\lambda}_k - y_k(\boldsymbol{x})\right)\right) & \boldsymbol{Q} \end{bmatrix} \succeq 0, \quad \forall k \in [K], \quad (4b)$$

$$\boldsymbol{x} \in \mathcal{X}, \ \boldsymbol{\lambda}_k \in \mathbb{R}_+^n, \ \forall k \in [K].$$

Proof. See Appendix A.1 for the detailed proof. □

Next, to derive a lower bound, we approximate $\boldsymbol{\xi}$ by capturing the dominant variability of $\boldsymbol{U}\Lambda^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}}$ through considering only the first m_1 random variables of $\boldsymbol{\xi}_{\mathrm{I}}$, i.e.,

$$\boldsymbol{\xi} \approx \boldsymbol{U}\Lambda^{\frac{1}{2}}[\boldsymbol{\xi}_{\mathrm{r}}; \boldsymbol{0}_{m-m_1}] + \boldsymbol{\mu} = \boldsymbol{U}_{m\times m_1}\boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{r}} + \boldsymbol{\mu}, \tag{5}$$

where $\boldsymbol{U}_{m\times m_1} \in \mathbb{R}^{m\times m_1}$ and $\boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}} \in \mathbb{R}^{m_1\times m_1}$ are upper-left submatrices of $\boldsymbol{U}$ and $\boldsymbol{\Lambda}^{\frac{1}{2}}$, respectively, and $\boldsymbol{\xi}_{\mathrm{r}} \in \mathbb{R}^{m_1}$ consists of the first m_1 entries of $\boldsymbol{\xi}_{\mathrm{I}}$. As the uncertainty of the last $(m - m_1)$ entries of $\boldsymbol{\xi}_{\mathrm{I}}$ vanishes, this yields a relaxation of (DRO-M):

$$\min_{\boldsymbol{x}\in\mathcal{X}} \max_{\mathbb{P}_{\mathrm{r}}\in\mathcal{D}_{\mathrm{M3}}} \mathbb{E}_{\mathbb{P}_{\mathrm{r}}}\left[f\left(\boldsymbol{x}, \boldsymbol{U}_{m\times m_1}\boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{r}} + \boldsymbol{\mu}\right)\right], \tag{6a}$$

where

$$\mathcal{D}_{\mathrm{M3}}(\mathcal{S}_{\mathrm{r}}, \gamma_1, \gamma_2) = \left\{ \mathbb{P}_{\mathrm{r}} \middle| \begin{array}{l} \mathbb{P}_{\mathrm{r}}(\boldsymbol{\xi}_{\mathrm{r}} \in \mathcal{S}_{\mathrm{r}}) = 1 \\ \mathbb{E}_{\mathbb{P}_{\mathrm{r}}}[\boldsymbol{\xi}_{\mathrm{r}}^\top]\mathbb{E}_{\mathbb{P}_{\mathrm{r}}}[\boldsymbol{\xi}_{\mathrm{r}}] \leq \gamma_1 \\ \mathbb{E}_{\mathbb{P}_{\mathrm{r}}}[\boldsymbol{\xi}_{\mathrm{r}}\boldsymbol{\xi}_{\mathrm{r}}^\top] \preceq \gamma_2 \boldsymbol{I}_{m_1} \end{array} \right\} \tag{6b}$$

with

$$\mathcal{S}_{\mathrm{r}} := \left\{\boldsymbol{\xi}_{\mathrm{r}} \in \mathbb{R}^{m_1} : \boldsymbol{U}_{m\times m_1}\boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{r}} + \boldsymbol{\mu} \in \mathcal{S}\right\}. \tag{6c}$$

THEOREM 2. *If $f(\boldsymbol{x}, \boldsymbol{U}_{m\times m_1}\Lambda_{m_1}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{r}} + \boldsymbol{\mu})$ is $\mathbb{P}_{\mathrm{r}}$-integrable for any $\mathbb{P}_{\mathrm{r}} \in \mathcal{D}_{\mathrm{M3}}$, then Problem (6) has the same optimal value as the following problem:*

$$\begin{aligned} \min_{\boldsymbol{x}, s, \boldsymbol{q}_{\mathrm{r}}, \boldsymbol{Q}_{\mathrm{r}}} \quad & s + \gamma_2 \boldsymbol{I}_{m_1} \bullet \boldsymbol{Q}_{\mathrm{r}} + \sqrt{\gamma_1}\,\|\boldsymbol{q}_{\mathrm{r}}\|_2 \\ \text{s.t.} \quad & s \geq f\left(\boldsymbol{x}, \boldsymbol{U}_{m\times m_1}\boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{r}} + \boldsymbol{\mu}\right) - \boldsymbol{\xi}_{\mathrm{r}}^\top \boldsymbol{q}_{\mathrm{r}} - \boldsymbol{\xi}_{\mathrm{r}}^\top \boldsymbol{Q}_{\mathrm{r}}\boldsymbol{\xi}_{\mathrm{r}}, \ \forall \boldsymbol{\xi}_{\mathrm{r}} \in \mathcal{S}_{\mathrm{r}}, \\ & \boldsymbol{Q}_{\mathrm{r}} \succeq 0, \ \boldsymbol{x} \in \mathcal{X}, \end{aligned} \tag{7}$$

where $\boldsymbol{q}_{\mathrm{r}} \in \mathbb{R}^{m_1}$ and $\boldsymbol{Q}_{\mathrm{r}} \in \mathbb{R}^{m_1\times m_1}$. Furthermore, we have the following: (i) Problem (7) provides a lower bound for the optimal value of (DRO-M); (ii) the optimal value of Problem (7) is nondecreasing in m_1; and (iii) if $m_1 = m$, then (DRO-M) and (7) have the same optimal value.

Proof. See Appendix A.2 for the detailed proof. □

Proposition 2. *Under Assumption 1, Problem (7) has the same optimal value as the following SDP formulation*

$$Z^*_{\mathrm{M}}(m_1) := \min_{\substack{\boldsymbol{x},s,\boldsymbol{\lambda}_k,\\ \boldsymbol{q}_{\mathrm{r}},\boldsymbol{Q}_{\mathrm{r}}}} \; s + \gamma_2 \boldsymbol{I}_{m_1} \bullet \boldsymbol{Q}_{\mathrm{r}} + \sqrt{\gamma_1}\,\|\boldsymbol{q}_{\mathrm{r}}\|_2 \tag{8a}$$

$$\text{s.t.} \begin{bmatrix} s - y_k^0(\boldsymbol{x}) - \boldsymbol{\lambda}_k^\top b - y_k(\boldsymbol{x})^\top \boldsymbol{\mu} + \boldsymbol{\lambda}_k^\top \boldsymbol{A}\boldsymbol{\mu} & \frac{1}{2}\left(\boldsymbol{q}_{\mathrm{r}} + \left(\boldsymbol{U}_{m\times m_1}\boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}}\right)^\top (\boldsymbol{A}^\top \boldsymbol{\lambda}_k - y_k(\boldsymbol{x}))\right)^\top \\ \frac{1}{2}\left(\boldsymbol{q}_{\mathrm{r}} + \left(\boldsymbol{U}_{m\times m_1}\boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}}\right)^\top (\boldsymbol{A}^\top \boldsymbol{\lambda}_k - y_k(\boldsymbol{x}))\right) & \boldsymbol{Q}_{\mathrm{r}} \end{bmatrix} \succeq 0, \quad \forall k \in [K], \tag{8b}$$

$$\boldsymbol{x} \in \mathcal{X}, \;\; \boldsymbol{\lambda}_k \in \mathbb{R}^n_+, \;\; \forall k \in [K]. \tag{8c}$$

Proof. The proof is similar with that of Proposition 1 and thus is omitted here. □

Comparing Problems (4) and (8) in terms of size, one can observe that Problem (8) is significantly easier to solve than Problem (4) because (1) Problem (8) includes fewer decision variables that Problem (4) does, i.e., $(m_1^2 + m_1 + 2n + 1)$ vs. $(m^2 + m + 2n + 1)$, and (2) the size of PSD matrices in Problem (8) is smaller than in Problem (4), i.e., $(m_1+1)\times(m_1+1)$ vs. $(m+1)\times(m+1)$.

2.2. Lower Bound Quality

To measure the quality of our derived lower bound, i.e., $Z^*_{\mathrm{M}}(m_1)$ in (8), we develop a theoretical upper bound for the gap between the optimal values of Problems (4) and (8). This upper bound brings two benefits: (1) it provides a rough approximation for the optimal value of Problem (4), which may not be solved efficiently in practice; and (2) it determines how many principal components are required to reach a preferred gap between the original and approximated optimal values, indicating a trade-off between solution quality and computational time.

Proposition 3. *It holds that*

$$0 \le {Z^*}_{\mathrm{M}}(m) - {Z^*}_{\mathrm{M}}(m_1) \le \sqrt{\gamma_2} \sum_{k=1}^{K} \sqrt{\sum_{i=m_1+1}^{m} \boldsymbol{\Lambda}_{i,i}\left(\left(\boldsymbol{A}^\top \boldsymbol{\lambda}_k^* - y_k(\boldsymbol{x}^*)\right)^\top \boldsymbol{U}_i\right)^2}, \tag{9}$$

where $\boldsymbol{U}_i$ represents the i^{th} column of matrix $\boldsymbol{U}$, and $\boldsymbol{x}^*$ and $\boldsymbol{\lambda}_k^*$ ($\forall k \in [K]$) are optimal solutions of Problem (8).

Proof. By Theorem 2, we have ${Z^*}_{\mathrm{M}}(m) - {Z^*}_{\mathrm{M}}(m_1) \ge 0$. Meanwhile, when $m_1 = m$, Problem (8) is equivalent to Problem (4). We use $(\boldsymbol{x}^*, s^*, {\boldsymbol{\lambda}_k}^* \; \forall k \in [K], \boldsymbol{q}_{\mathrm{r}}^*, \boldsymbol{Q}_{\mathrm{r}}^*)$ to denote an optimal solution of Problem (8). Based on this optimal solution, we construct a feasible solution of Problem (4), represented by $\left(\bar{\boldsymbol{x}}, \bar{s}, \bar{\boldsymbol{\lambda}}_k \; \forall k \in [K], \bar{\boldsymbol{q}}, \bar{\boldsymbol{Q}}\right)$. For clarity, we define

$$S^k = s^* - y_k^0(\boldsymbol{x}^*) - {\boldsymbol{\lambda}_k^*}^\top b - y_k(\boldsymbol{x}^*)^\top \boldsymbol{\mu} + {\boldsymbol{\lambda}_k^*}^\top \boldsymbol{A}\boldsymbol{\mu}, \;\; \forall k \in [K], \text{ and}$$

$$\boldsymbol{q}_c^k = \left(\boldsymbol{U}_{m\times c}\boldsymbol{\Lambda}^{c\frac{1}{2}}\right)^\top (\boldsymbol{A}^\top\boldsymbol{\lambda}_k^* - y_k(\boldsymbol{x}^*)),\ \forall k\in[K],\ \forall c\in\{m_1, m-m_1, m\},$$

where $\boldsymbol{\Lambda}^{m_1}\in\mathbb{R}^{m_1\times m_1}$ and $\boldsymbol{\Lambda}^{m-m_1}\in\mathbb{R}^{(m-m_1)\times(m-m_1)}$ represent the upper-left and lower-right submatrices of $\boldsymbol{\Lambda}$, respectively.

First, we let $\bar{\boldsymbol{x}}=\boldsymbol{x}^*$, $\bar{\boldsymbol{\lambda}}_k=\boldsymbol{\lambda}_k^*$ for any $k\in[K]$, $\bar{\boldsymbol{q}}=\left(\boldsymbol{q}_r^{*\top}, \boldsymbol{0}_{m-m_1}^\top\right)^\top$, $\bar{s}=s^*+\sum_{k=1}^K s_1^k$, and

$$\bar{\boldsymbol{Q}} = \begin{bmatrix} \boldsymbol{Q}_{\mathrm{r}}^* & \boldsymbol{0}_{m_1\times(m-m_1)} \\ \boldsymbol{0}_{(m-m_1)\times m_1} & \sum_{k=1}^K \frac{s_2^k}{4}\boldsymbol{q}_{m-m_1}^k\left(\boldsymbol{q}_{m-m_1}^k\right)^\top \end{bmatrix},$$

where $s_1^k>0$ and $s_2^k>0$ for any $k\in[K]$. As $\bar{\boldsymbol{x}}=\boldsymbol{x}^*\in\mathcal{X}$ and $\bar{\boldsymbol{\lambda}}_k=\boldsymbol{\lambda}_k^*\in\mathbb{R}_+^n$, for any $k\in[K]$, due to constraint (8c), we only require $(\bar{\boldsymbol{x}}, \bar{s}, \bar{\boldsymbol{\lambda}}_k\ \forall k\in[K], \bar{\boldsymbol{q}}, \bar{\boldsymbol{Q}})$ to satisfy (4b). Thus, we will find the values of s_1^k and s_2^k for any $k\in[K]$ that enable this solution to satisfy (4b).

We plug $(\bar{\boldsymbol{x}}, \bar{s}, \bar{\boldsymbol{\lambda}}_k\ \forall k\in[K], \bar{\boldsymbol{q}}, \bar{\boldsymbol{Q}})$ to (4b) and use $\bar{\boldsymbol{Y}}^k$ for any $k\in[K]$ to denote the corresponding matrix in (4b). For any given $k\in[K]$, we perform the following decomposition:

$$\begin{aligned}\bar{\boldsymbol{Y}}^k &= \begin{bmatrix} S^k & \frac{1}{2}\left(\boldsymbol{q}_r^*+\boldsymbol{q}_{m_1}^k\right)^\top & \boldsymbol{0}_{1\times(m-m_1)} \\ \frac{1}{2}\left(\boldsymbol{q}_r^*+\boldsymbol{q}_{m_1}^k\right) & \boldsymbol{Q}_{\mathrm{r}}^* & \boldsymbol{0}_{m_1\times(m-m_1)} \\ \boldsymbol{0}_{(m-m_1)\times 1} & \boldsymbol{0}_{(m-m_1)\times m_1} & \boldsymbol{0}_{(m-m_1)\times(m-m_1)} \end{bmatrix} + \begin{bmatrix} \sum_{k=1}^K s_1^k & \boldsymbol{0}_{1\times m_1} & \frac{1}{2}\boldsymbol{q}_{m-m_1}^k{}^\top \\ \boldsymbol{0}_{m_1\times 1} & \boldsymbol{0}_{m_1\times m_1} & \boldsymbol{0}_{m_1\times(m-m_1)} \\ \frac{1}{2}\boldsymbol{q}_{m-m_1}^k & \boldsymbol{0}_{(m-m_1)\times m_1} & \sum_{k=1}^K \frac{s_2^k}{4}\boldsymbol{q}_{m-m_1}^k\left(\boldsymbol{q}_{m-m_1}^k\right)^\top \end{bmatrix} \\ &\succeq \begin{bmatrix} S^k & \frac{1}{2}\left(\boldsymbol{q}_r^*+\boldsymbol{q}_{m_1}^k\right)^\top & \boldsymbol{0}_{1\times(m-m_1)} \\ \frac{1}{2}\left(\boldsymbol{q}_r^*+\boldsymbol{q}_{m_1}^k\right) & \boldsymbol{Q}_{\mathrm{r}}^* & \boldsymbol{0}_{m_1\times(m-m_1)} \\ \boldsymbol{0}_{(m-m_1)\times 1} & \boldsymbol{0}_{(m-m_1)\times m_1} & \boldsymbol{0}_{(m-m_1)\times(m-m_1)} \end{bmatrix} + \begin{bmatrix} s_1^k & \boldsymbol{0}_{1\times m_1} & \frac{1}{2}\boldsymbol{q}_{m-m_1}^k{}^\top \\ \boldsymbol{0}_{m_1\times 1} & \boldsymbol{0}_{m_1\times m_1} & \boldsymbol{0}_{m_1\times(m-m_1)} \\ \frac{1}{2}\boldsymbol{q}_{m-m_1}^k & \boldsymbol{0}_{(m-m_1)\times m_1} & \frac{s_2^k}{4}\boldsymbol{q}_{m-m_1}^k\left(\boldsymbol{q}_{m-m_1}^k\right)^\top \end{bmatrix}. \end{aligned} \tag{10}$$

The first matrix in (10) is clearly PSD because the elimination of its zero components leads to a PSD matrix due to constraint (8b). Now we find the values of s_1^k and s_2^k to make the second matrix PSD as well, and then accordingly the constructed solution is feasible for (4).

Next, we use $\begin{bmatrix} \bar{\boldsymbol{A}} & \bar{\boldsymbol{B}} \\ \bar{\boldsymbol{B}}^\top & \bar{\boldsymbol{C}} \end{bmatrix}$ to denote second matrix in (10) by letting $\bar{\boldsymbol{A}}=s_1^k$, $\bar{\boldsymbol{B}}^\top=\left(\boldsymbol{0}_{1\times m_1}\ \frac{1}{2}\boldsymbol{q}_{m-m_1}^k{}^\top\right)$, and $\bar{\boldsymbol{C}}=\begin{bmatrix} \boldsymbol{0}_{m_1\times m_1} & \boldsymbol{0}_{m_1\times(m-m_1)} \\ \boldsymbol{0}_{(m-m_1)\times m_1} & \frac{s_2^k}{4}\boldsymbol{q}_{m-m_1}^k\left(\boldsymbol{q}_{m-m_1}^k\right)^\top \end{bmatrix}$. It follows that

$$\begin{aligned}\bar{\boldsymbol{C}}-\bar{\boldsymbol{B}}\bar{\boldsymbol{A}}^{-1}\bar{\boldsymbol{B}}^\top &= \begin{bmatrix} \boldsymbol{0}_{m_1\times m_1} & \boldsymbol{0}_{m_1\times(m-m_1)} \\ \boldsymbol{0}_{(m-m_1)\times m_1} & \frac{s_2^k}{4}\boldsymbol{q}_{m-m_1}^k\left(\boldsymbol{q}_{m-m_1}^k\right)^\top \end{bmatrix} - \frac{1}{s_1^k}\left(\boldsymbol{0}_{1\times m_1}\ \frac{1}{2}\boldsymbol{q}_{m-m_1}^k{}^\top\right)^\top\left(\boldsymbol{0}_{1\times m_1}\ \frac{1}{2}\boldsymbol{q}_{m-m_1}^k{}^\top\right) \\ &= \begin{bmatrix} \boldsymbol{0}_{m_1\times m_1} & \boldsymbol{0}_{m_1\times(m-m_1)} \\ \boldsymbol{0}_{(m-m_1)\times m_1} & \left(\frac{s_2^k}{4}-\frac{1}{4s_1^k}\right)\boldsymbol{q}_{m-m_1}^k\left(\boldsymbol{q}_{m-m_1}^k\right)^\top \end{bmatrix}, \end{aligned} \tag{11}$$

which is PSD if $s_1^k\times s_2^k\geq 1$. Thus, we let $s_1^k\times s_2^k\geq 1$ hold for any $k\in[K]$ and by the properties of Schur complement, we have $\begin{bmatrix} \bar{\boldsymbol{A}} & \bar{\boldsymbol{B}} \\ \bar{\boldsymbol{B}}^\top & \bar{\boldsymbol{C}} \end{bmatrix}\succeq 0$ because $\bar{\boldsymbol{A}}$ is invertible and positive definite.

In addition, since Problem (4) is a minimization problem, its optimal value is no larger than the objective corresponding to the feasible solution $\left(\bar{\boldsymbol{x}}, \bar{s}, \bar{\boldsymbol{\lambda}}_k\ \forall k\in[K], \bar{\boldsymbol{q}}, \bar{\boldsymbol{Q}}\right)$. That is,

$$\begin{aligned} Z^*{}_{\mathrm{M}}(m) &\leq \bar{s}+\gamma_2\boldsymbol{I}_m\bullet\bar{\boldsymbol{Q}}+\sqrt{\gamma_1}\,\|\bar{\boldsymbol{q}}\| \\ &= Z^*{}_{\mathrm{M}}(m_1)+\sum_{k=1}^K s_1^k+\gamma_2\sum_{k=1}^K\frac{s_2^k}{4}\mathrm{trace}\left(\boldsymbol{q}_{m-m_1}^k(\boldsymbol{q}_{m-m_1}^k)^\top\right). \end{aligned} \tag{12}$$

Due to the condition $s_1^k \times s_2^k \geq 1$, we let $s_1^k = \frac{\sqrt{\gamma_2 (\boldsymbol{q}_{m-m_1}^k)^\top \boldsymbol{q}_{m-m_1}^k}}{2}$ and $s_2^k = \frac{2}{\sqrt{\gamma_2 (\boldsymbol{q}_{m-m_1}^k)^\top \boldsymbol{q}_{m-m_1}^k}}$, which leads to the smallest possible value of the RHS of (12). Therefore, we have

$$\begin{aligned} Z^*{}_{\mathrm{M}}(m) &\leq Z^*{}_{\mathrm{M}}(m_1) + \sum_{k=1}^K s_1^k + \gamma_2 \sum_{k=1}^K \frac{s_2^k}{4} \operatorname{trace}\left(\boldsymbol{q}_{m-m_1}^k (\boldsymbol{q}_{m-m_1}^k)^\top\right) \\ &= Z^*{}_{\mathrm{M}}(m_1) + \sqrt{\gamma_2} \sum_{k=1}^K \frac{\sqrt{(\boldsymbol{q}_{m-m_1}^k)^\top \boldsymbol{q}_{m-m_1}^k}}{2} + \sqrt{\gamma_2} \operatorname{trace}\left(\sum_{k=1}^K \frac{\boldsymbol{q}_{m-m_1}^k (\boldsymbol{q}_{m-m_1}^k)^\top}{2\sqrt{(\boldsymbol{q}_{m-m_1}^k)^\top \boldsymbol{q}_{m-m_1}^k}} \right). \end{aligned}$$

Finally, since $\sum_{k=1}^K \frac{\sqrt{(\boldsymbol{q}_{m-m_1}^k)^\top \boldsymbol{q}_{m-m_1}^k}}{2}$ is equal to $\operatorname{trace}\left(\sum_{k=1}^K \frac{\boldsymbol{q}_{m-m_1}^k (\boldsymbol{q}_{m-m_1}^k)^\top}{2\sqrt{(\boldsymbol{q}_{m-m_1}^k)^\top \boldsymbol{q}_{m-m_1}^k}} \right)$, we have

$$0 \leq Z^*{}_{\mathrm{M}}(m) - Z^*{}_{\mathrm{M}}(m_1) \leq \sqrt{\gamma_2} \left(\sum_{k=1}^K \sqrt{ \sum_{i=m_1+1}^m \boldsymbol{\Lambda}_{i,i} \left((\boldsymbol{A}^\top \boldsymbol{\lambda}_k^* - y_k(\boldsymbol{x}^*))^\top \boldsymbol{U}_i \right)^2 } \right). \quad \square$$

Remark 1. Note that Cheng et al. (2018) derived a similar upper bound by specifically considering $\gamma_1 = 0$ and $\gamma_2 = 1$, while (9) applies to general values of γ_1 and γ_2.

2.3. Upper Bounds

We further develop computationally efficient inner approximations for Problem (2), leading to upper bounds of its optimal value. Specifically, we derive two inner approximations in Sections 2.3.1 and 2.3.2.

2.3.1. PCA based Upper Bound Similar to Section 2.1, we utilize PCA to consider only the first m_1 entries of $\boldsymbol{\xi}_{\mathrm{I}}$ in the second-moment constraint in $\mathcal{D}_{\mathrm{M2}}$. This is a relaxation of the second-moment constraint, leading to a larger ambiguity set and so an inner approximation of Problem (2):

$$\min_{\boldsymbol{x}\in\mathcal{X}} \max_{\mathbb{P}_{\mathrm{I}}\in\mathcal{D}_{\mathrm{M4}}} \mathbb{E}_{\mathbb{P}_{\mathrm{I}}} \left[f\left(\boldsymbol{x}, \boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}} \boldsymbol{\xi}_{\mathrm{I}} + \boldsymbol{\mu} \right) \right], \tag{13}$$

where

$$\mathcal{D}_{\mathrm{M4}}(\mathcal{S}_{\mathrm{I}}, \gamma_1, \gamma_2) = \left\{ \mathbb{P}_{\mathrm{I}} \middle| \begin{array}{l} \mathbb{P}_{\mathrm{I}}(\boldsymbol{\xi}_{\mathrm{I}} \in \mathcal{S}_{\mathrm{I}}) = 1 \\ \mathbb{E}_{\mathbb{P}_{\mathrm{I}}}\left[\boldsymbol{\xi}_{\mathrm{I}}^\top\right] \mathbb{E}_{\mathbb{P}_{\mathrm{I}}}\left[\boldsymbol{\xi}_{\mathrm{I}}\right] \leq \gamma_1 \\ \mathbb{E}_{\mathbb{P}_{\mathrm{I}}}\left[\boldsymbol{\xi}_{\mathrm{r}} \boldsymbol{\xi}_{\mathrm{r}}^\top\right] \preceq \gamma_2 \boldsymbol{I}_{m_1} \end{array} \right\}.$$

Theorem 3. *If $f(x, \boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}} + \boldsymbol{\mu})$ is $\mathbb{P}_{\mathrm{I}}$-integrable for any $\mathbb{P}_{\mathrm{I}} \in \mathcal{D}_{\mathrm{M4}}$, then Problem (13) has the same optimal value as the following problem:*

$$\begin{aligned} \min_{\boldsymbol{x},s,\boldsymbol{q},\boldsymbol{Q}_{\mathrm{r}}} \quad & s + \gamma_2 \boldsymbol{I}_{m_1} \bullet \boldsymbol{Q}_{\mathrm{r}} + \sqrt{\gamma_1} \|\boldsymbol{q}\|_2 && \text{(14a)} \\ \text{s.t.} \quad & s \geq f\left(\boldsymbol{x}, \boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}} \boldsymbol{\xi}_{\mathrm{I}} + \boldsymbol{\mu} \right) - \boldsymbol{\xi}_{\mathrm{r}}^\top \boldsymbol{Q}_{\mathrm{r}} \boldsymbol{\xi}_{\mathrm{r}} - \boldsymbol{q}^\top \boldsymbol{\xi}_{\mathrm{I}}, \ \forall \boldsymbol{\xi}_{\mathrm{I}} \in \mathcal{S}_{\mathrm{I}}, && \text{(14b)} \\ & \boldsymbol{Q}_{\mathrm{r}} \succeq 0, \ \boldsymbol{x} \in \mathcal{X}, \end{aligned}$$

where $\boldsymbol{q}\in\mathbb{R}^m$ *and* $\boldsymbol{Q}_{\mathrm{r}}\in\mathbb{R}^{m_1\times m_1}$*. We also have the following: (i) Problem* (14) *provides an upper bound for the optimal value of Problem* (3)*; (ii) the optimal value of Problem* (14) *is non-increasing in* m_1*; and (iii) if* $m_1=m$*, then Problems* (3) *and* (14) *have the same optimal value.*

Proof. The proof is similar with that of Theorem 2 and thus is omitted here. □

PROPOSITION 4. *Under Assumption 1, Problem* (14) *has the same optimal value as the following SDP formulation:*

$$\overline{Z}^*_{\mathrm{M}}(m_1):=\min_{\substack{\boldsymbol{x},s,\boldsymbol{\lambda}_k,\\ \boldsymbol{q},\boldsymbol{Q}_{\mathrm{r}}}}\ s+\gamma_2\boldsymbol{I}_{m_1}\bullet\boldsymbol{Q}_{\mathrm{r}}+\sqrt{\gamma_1}\,\|\boldsymbol{q}\|_2 \tag{15a}$$

$$\text{s.t.}\ \begin{bmatrix} s-y_k^0(\boldsymbol{x})-\boldsymbol{\lambda}_k^\top\boldsymbol{b}-y_k(\boldsymbol{x})^\top\boldsymbol{\mu}+\boldsymbol{\lambda}_k^\top\boldsymbol{A}\boldsymbol{\mu} & \frac{1}{2}\left(\boldsymbol{q}_1+\left(\boldsymbol{U}_{m\times m_1}\boldsymbol{\Lambda}^{\frac{1}{2}}_{m_1}\right)^\top\left(\boldsymbol{A}^\top\boldsymbol{\lambda}_k-y_k(\boldsymbol{x})\right)\right)^\top \\ \frac{1}{2}\left(\boldsymbol{q}_1+\left(\boldsymbol{U}_{m\times m_1}\boldsymbol{\Lambda}^{\frac{1}{2}}_{m_1}\right)^\top\left(\boldsymbol{A}^\top\boldsymbol{\lambda}_k-y_k(\boldsymbol{x})\right)\right) & \boldsymbol{Q}_{\mathrm{r}}\end{bmatrix}\succeq 0,\quad \forall k\in[K], \tag{15b}$$

$$\boldsymbol{q}_2+\left(\boldsymbol{U}_{m\times(m-m_1)}\boldsymbol{\Lambda}^{\frac{1}{2}}_{m-m_1}\right)^\top\left(\boldsymbol{A}^\top\boldsymbol{\lambda}_k-y_k(\boldsymbol{x})\right)=\boldsymbol{0},\ \forall k\in[K], \tag{15c}$$

$$\boldsymbol{x}\in\mathcal{X},\ \boldsymbol{\lambda}_k\in\mathbb{R}^n_+,\ \forall k\in[K],$$

where $\boldsymbol{q}=(\boldsymbol{q}_1^\top\in\mathbb{R}^{m_1},\boldsymbol{q}_2^\top\in\mathbb{R}^{m-m_1})^\top$.

Proof. See Appendix A.3 for the detailed proof. □

One can observe that Problem (15) is significantly easier to solve than Problem (4) due to fewer decision variables and lower-dimensional PSD matrices in Problem (15).

2.3.2. Vector Splitting based Upper Bound We derive the second inner approximation by splitting the random vector ξ_{I} into P pieces, i.e., $\boldsymbol{\xi}_{\mathrm{I}}=\left(\boldsymbol{\xi}_{\mathrm{I}_1}^\top,\boldsymbol{\xi}_{\mathrm{I}_2}^\top,\ldots,\boldsymbol{\xi}_{\mathrm{I}_P}^\top\right)^\top$, where $\boldsymbol{\xi}_{\mathrm{I}_i}\in\mathbb{R}^{m_i}$, $\forall i\in[P]$, and $\sum_{i=1}^P m_i=m$. Accordingly, we revise the second-moment constraint in $\mathcal{D}_{\mathrm{M2}}$ with respect to these smaller pieces, leading to the following ambiguity set:

$$\mathcal{D}_{\mathrm{M5}}\left(\mathcal{S}_{\mathrm{I}},\gamma_1,\gamma_2\right)=\left\{\mathbb{P}_{\mathrm{I}}\ \middle|\ \begin{array}{l}\mathbb{P}_{\mathrm{I}}\left(\boldsymbol{\xi}_{\mathrm{I}}\in\mathcal{S}_{\mathrm{I}}\right)=1\\ \mathbb{E}_{\mathbb{P}_{\mathrm{I}}}\left[\boldsymbol{\xi}_{\mathrm{I}}^\top\right]\mathbb{E}_{\mathbb{P}_{\mathrm{I}}}\left[\boldsymbol{\xi}_{\mathrm{I}}\right]\le\gamma_1\\ \mathbb{E}_{\mathbb{P}_{\mathrm{I}}}\left[\boldsymbol{\xi}_{\mathrm{I}_i}\boldsymbol{\xi}_{\mathrm{I}_i}^\top\right]\preceq\gamma_2\boldsymbol{I}_{m_i},\ \forall i\in[P]\end{array}\right\}.$$

$\mathcal{D}_{\mathrm{M5}}$ is a superset of $\mathcal{D}_{\mathrm{M2}}$ because we ignore the correlations among $\boldsymbol{\xi}_{\mathrm{I}_p}$ and $\boldsymbol{\xi}_{\mathrm{I}_q}$ for any $p,q\in[P]$ with $p\neq q$. This leads to the following inner approximation of Problem (2):

$$\min_{\boldsymbol{x}\in\mathcal{X}}\ \max_{\mathbb{P}_{\mathrm{I}}\in\mathcal{D}_{\mathrm{M5}}}\ \mathbb{E}_{\mathbb{P}_{\mathrm{I}}}\left[f\left(\boldsymbol{x},\boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}}+\boldsymbol{\mu}\right)\right]. \tag{16}$$

THEOREM 4. *If $f(\boldsymbol{x}, \boldsymbol{U}\Lambda^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}} + \boldsymbol{\mu})$ is $\mathbb{P}_{\mathrm{I}}$-integrable for any $\mathbb{P}_{\mathrm{I}} \in \mathcal{D}_{\mathrm{M5}}$, then Problem (16) has the same optimal value as the following problem:*

$$\min_{\boldsymbol{x},s,\boldsymbol{q},\boldsymbol{Q}_i} \; s + \gamma_2 \sum_{i=1}^{P} \boldsymbol{I}_{m_i} \bullet \boldsymbol{Q}_i + \sqrt{\gamma_1}\,\|\boldsymbol{q}\|_2 \tag{17a}$$

$$\text{s.t.} \quad s \geq f\left(\boldsymbol{x}, \boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}} + \boldsymbol{\mu}\right) - \sum_{i=1}^{P} \boldsymbol{\xi}_{\mathrm{I}_i}^\top \boldsymbol{Q}_i \boldsymbol{\xi}_{\mathrm{I}_i} - \boldsymbol{q}^\top \boldsymbol{\xi}_{\mathrm{I}}, \; \forall \boldsymbol{\xi}_{\mathrm{I}} \in \mathcal{S}_{\mathrm{I}}, \tag{17b}$$

$$\boldsymbol{x} \in \mathcal{X}, \quad \boldsymbol{Q}_i \succeq 0, \quad \forall i \in [P],$$

where $\boldsymbol{q} \in \mathbb{R}^m$ and $\boldsymbol{Q}_i \in \mathbb{R}^{m_i \times m_i}$ for any $i \in [P]$. Furthermore, Problem (17) provides an upper bound for the optimal value of Problem (3).

Proof. See Appendix A.4 for the detailed proof. □

PROPOSITION 5. *Under Assumption 1, Problem (17) has the same optimal value as the following SDP formulation:*

$$UB_{\mathrm{M}}^* := \min_{\substack{\boldsymbol{x},s,\boldsymbol{q},\\ \boldsymbol{Q}_i,\boldsymbol{\lambda}_k,s_{ik}}} \; s + \gamma_2 \sum_{i=1}^{P} \boldsymbol{I}_{m_i} \bullet \boldsymbol{Q}_i + \sqrt{\gamma_1}\,\|\boldsymbol{q}\|_2 \tag{18}$$

$$\text{s.t.} \begin{bmatrix} s_{ik} & \frac{1}{2}\left(\boldsymbol{q}_i + \left(\boldsymbol{U}_{m\times m_i}\boldsymbol{\Lambda}_{m_i}^{\frac{1}{2}}\right)^\top \left(\boldsymbol{A}^\top\boldsymbol{\lambda}_k - y_k(\boldsymbol{x})\right)\right)^\top \\ \frac{1}{2}\left(\boldsymbol{q}_i + \left(\boldsymbol{U}_{m\times m_i}\boldsymbol{\Lambda}_{m_i}^{\frac{1}{2}}\right)^\top \left(\boldsymbol{A}^\top\boldsymbol{\lambda}_k - y_k(\boldsymbol{x})\right)\right) & \boldsymbol{Q}_i \end{bmatrix} \succeq 0,$$

$$\forall i \in [P], \forall k \in [K],$$

$$\sum_{i=1}^{P} s_{ik} = s - y_k^0(\boldsymbol{x}) - \boldsymbol{\lambda}_k^\top \boldsymbol{b} - y_k(\boldsymbol{x})^\top \boldsymbol{\mu} + \boldsymbol{\lambda}_k^\top \boldsymbol{A}\boldsymbol{\mu}, \; \forall k \in [K],$$

$$\boldsymbol{\lambda}_k \in \mathbb{R}_+^n, \; \forall k \in [K], \; \boldsymbol{x} \in \mathcal{X},$$

where $\boldsymbol{q}_i \in \mathbb{R}^{m_i}$ and $\boldsymbol{Q}_i \in \mathbb{R}^{m_i \times m_i}$ for any $i \in [P]$.

Proof. The proof is similar with that of Proposition 4 and thus is omitted here. □

One can observe that Problem (18) is significantly easier to solve than Problem (4) because it has smaller-sized PSD matrices and matrix variables compared to Problem (4).

2.4. Upper Bound Quality

To measure the quality of our derived upper bounds in Section 2.3, we derive a theoretical bound for the gap between the optimal values of Problem (4) and Problem (15) (resp. Problem (18)). Before that, we present the following lemma that will facilitate the proofs in this section.

LEMMA 1. *Consider the following PSD matrix with dimension $(m+1) \times (m+1)$:*

$$\boldsymbol{Z} = \begin{bmatrix} s & \boldsymbol{q}_1^\top & \cdots & \boldsymbol{q}_K^\top \\ \boldsymbol{q}_1 & \boldsymbol{Q}_1 & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{q}_K & \boldsymbol{0} & \cdots & \boldsymbol{Q}_K \end{bmatrix} \succeq 0, \tag{19}$$

where $s\in\mathbb{R}$, $\boldsymbol{q}_k\in\mathbb{R}^{m_k}$ *for any* $k\in[K]$ *with* $\sum_{k=1}^K m_k=m$, $\boldsymbol{Q}_k\in\mathbb{R}^{m_k\times m_k}$ *for any* $k\in[K]$, *and other components are zero. Inequality* (19) *holds if and only if there exist* $\{s_k\}_{k=1}^K$ *with* $\sum_{k=1}^K s_k=s$ *such that*

$$\begin{bmatrix} s_k & \boldsymbol{q}_k^\top \\ \boldsymbol{q}_k & \boldsymbol{Q}_k\end{bmatrix}\succeq 0,\ \forall k\in[K].$$

Proof. We prove the following equivalence:

$$\boldsymbol{Z}=\begin{bmatrix} s & \boldsymbol{q}_1^\top & \cdots & \boldsymbol{q}_K^\top \\ \boldsymbol{q}_1 & \boldsymbol{Q}_1 & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{q}_K & \boldsymbol{0} & \cdots & \boldsymbol{Q}_K\end{bmatrix}\succeq 0 \quad\Leftrightarrow\quad \begin{bmatrix} s_k & \boldsymbol{q}_k^\top \\ \boldsymbol{q}_k & \boldsymbol{Q}_k\end{bmatrix}\succeq 0,\ \forall k\in[K],\ \textit{with}\ \sum_{k=1}^K s_k=s.$$

Matrix $\boldsymbol{Z}$ is PSD if and only if $\boldsymbol{\eta}^\top\mathbf{Z}\boldsymbol{\eta}\geq 0$ for any $\boldsymbol{\eta}\in\mathbb{R}^m$, where $\boldsymbol{\eta}=(\eta_0,\boldsymbol{\eta}_1^\top,\boldsymbol{\eta}_2^\top,\cdots,\boldsymbol{\eta}_K^\top)^\top$ and $\boldsymbol{\eta}_k\in\mathbb{R}^{m_k}$ for any $k\in[K]$ with $\sum_{k=1}^K m_k=m$. Similar to the proofs of Propositions 1 and 4 in Appendices A.1 and A.3, respectively, we assume that $\eta_0=1$ without loss of generality. Thus, we have

$$\begin{aligned}\boldsymbol{Z}\succeq 0 \quad&\Leftrightarrow\quad \left(1,\boldsymbol{\eta}_1^\top,\boldsymbol{\eta}_2^\top,\cdots,\boldsymbol{\eta}_K^\top\right)\mathbf{Z}\left(1,\boldsymbol{\eta}_1^\top,\boldsymbol{\eta}_2^\top,\cdots,\boldsymbol{\eta}_K^\top\right)^\top\geq 0,\ \forall\boldsymbol{\eta}_k\in\mathbb{R}^{m_k},k\in[K]\\ &\Leftrightarrow\quad s+\sum_{k=1}^K\left(2\boldsymbol{q}_k^\top\boldsymbol{\eta}_k+\boldsymbol{\eta}_k^\top\boldsymbol{Q}_k\boldsymbol{\eta}_k\right)\geq 0,\ \forall\boldsymbol{\eta}_k\in\mathbb{R}^{m_k},k\in[K]\\ &\Leftrightarrow\quad s+\sum_{k=1}^K\inf_{\forall\boldsymbol{\eta}_k\in\mathbb{R}^{m_k},k\in[K]}\left\{2\boldsymbol{q}_k^\top\boldsymbol{\eta}_k+\boldsymbol{\eta}_k^\top\boldsymbol{Q}_k\boldsymbol{\eta}_k\right\}\geq 0.\end{aligned}$$

There exists $s_k\in\mathbb{R}$ for any $k\in[K]$ such that $\sum_{k=1}^K s_k=s$, by which we further have

$$\begin{aligned} s+\sum_{k=1}^K\inf_{\forall\boldsymbol{\eta}_k\in\mathbb{R}^{m_k},k\in[K]}\left\{2\boldsymbol{q}_k^\top\boldsymbol{\eta}_k+\boldsymbol{\eta}_k^\top\boldsymbol{Q}_k\boldsymbol{\eta}_k\right\}\geq 0 \quad&\Leftrightarrow\quad s_k+\inf_{\forall\boldsymbol{\eta}_k\in\mathbb{R}^{m_k}}\left\{2\boldsymbol{q}_k^\top\boldsymbol{\eta}_k+\boldsymbol{\eta}_k^\top\boldsymbol{Q}_k\boldsymbol{\eta}_k\right\}\geq 0,\ \forall k\in[K]\\ &\Leftrightarrow\quad s_k+2\boldsymbol{q}_k^\top\boldsymbol{\eta}_k+\boldsymbol{\eta}_k^\top\boldsymbol{Q}_k\boldsymbol{\eta}_k\geq 0,\ \forall\boldsymbol{\eta}_k\in\mathbb{R}^{m_k},k\in[K]\\ &\Leftrightarrow\quad \left(1,\boldsymbol{\eta}_k^\top\right)\begin{bmatrix} s_k & \boldsymbol{q}_k^\top \\ \boldsymbol{q}_k & \boldsymbol{Q}_k\end{bmatrix}\left(1,\boldsymbol{\eta}_k^\top\right)^\top\geq 0,\ \forall\boldsymbol{\eta}_k\in\mathbb{R}^{m_k},k\in[K]\\ &\Leftrightarrow\quad \begin{bmatrix} s_k & \boldsymbol{q}_k^\top \\ \boldsymbol{q}_k & \boldsymbol{Q}_k\end{bmatrix}\succeq 0,\ \forall k\in[K].\end{aligned}$$

In summary, we have

$$\boldsymbol{Z}\succeq 0 \quad\Leftrightarrow\quad \exists s_k\in\mathbb{R},\forall k\in[K],\ \text{such that}\ \sum_{k=1}^K s_k=s\ \text{and}\ \begin{bmatrix} s_k & \boldsymbol{q}_k^\top \\ \boldsymbol{q}_k & \boldsymbol{Q}_k\end{bmatrix}\succeq 0,\ \forall k\in[K]. \quad\square$$

Proposition 6. *Suppose that* $\boldsymbol{x}^*$ *is an optimal solution of Problem* (15) *and* ${\boldsymbol{Y}_{\mathrm{r}}^k}^*=\begin{bmatrix} {Y_{11}^k}^* & {\boldsymbol{Y}_{12\mathrm{r}}^k}^{*\top} \\ {\boldsymbol{Y}_{12\mathrm{r}}^k}^* & {\boldsymbol{Y}_{22\mathrm{r}}^k}^*\end{bmatrix}$ *and* $\boldsymbol{\theta}_k^*$ *are the corresponding dual optimal solutions associated with constraints* (15b) *and* (15c), *respectively. Then, it holds that*

$$0\leq\overline{Z}_{\mathrm{M}}^*(m_1)-{Z^*}_{\mathrm{M}}(m)\leq\sum_{k=1}^K\left(\left(\boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top\boldsymbol{y}_k(\boldsymbol{x}^*)\right)^\top\left({\boldsymbol{Y}_{12\mathrm{r}}^{k^*}}^\top,\boldsymbol{\theta}_k^{*\top}\right)^\top\left(\frac{c^k-1}{c^k}\right), \tag{20}$$

where

$$c^k=\sqrt{1+\frac{1}{Y_{11}^{k*}}\sum_{i=1}^{m-m_1}\frac{K}{\gamma_2}{\theta_{ki}^*}^2},\ \forall k\in[K]. \tag{21}$$

Proof. We prove the results by investigating the duals of Problems (4) and (15) where $\boldsymbol{x}$ is fixed at $\boldsymbol{x}^*$. Given $\boldsymbol{x}=\boldsymbol{x}^*$, we use $\boldsymbol{Y}^k$ to denote the dual variable of constraints (4b) for any $k\in[K]$, where the dual of Problem (4) can be described as follows:

$$Z_{\mathrm{M}}^*(m)=\max_{\boldsymbol{Y}} \sum_{k=1}^{K}\left(y_k^0(\boldsymbol{x})+y_k(\boldsymbol{x})^\top\boldsymbol{\mu}\right)Y_{11}^k+\sum_{k=1}^{K}\left(\left(\boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top y_k(\boldsymbol{x})\right)^\top \boldsymbol{Y}_{12}^k \tag{22a}$$

$$\text{s.t.}\quad 1-\sum_{k=1}^{K}Y_{11}^k=0, \tag{22b}$$

$$\sqrt{\gamma_1}-\left\|\sum_{k=1}^{K}\boldsymbol{Y}_{12}^k\right\|_2\geq 0, \tag{22c}$$

$$\gamma_2\boldsymbol{I}_m-\sum_{k=1}^{K}\boldsymbol{Y}_{22}^k\succeq 0, \tag{22d}$$

$$\boldsymbol{b}Y_{11}^k-\boldsymbol{A}\boldsymbol{\mu}Y_{11}^k-\boldsymbol{A}\boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{Y}_{12}^k\geq 0,\ \forall k\in[K], \tag{22e}$$

$$\boldsymbol{Y}^k:\begin{bmatrix} Y_{11}^k & \boldsymbol{Y}_{12}^{k\ \top} \\ \boldsymbol{Y}_{12}^k & \boldsymbol{Y}_{22}^k\end{bmatrix}\succeq 0,\ \forall k\in[K]. \tag{22f}$$

Similarly, given $\boldsymbol{x}=\boldsymbol{x}^*$, we use $\boldsymbol{Y}_{\mathrm{r}}^k$ and $\boldsymbol{\theta}_k$ to denote the dual variables of constraints (15b) and (15c), respectively, for any $k\in[K]$, where the dual of Problem (15) can be described as follows:

$$\begin{aligned}\overline{Z}_{\mathrm{M}}^*(m_1)=\max_{\boldsymbol{Y},\boldsymbol{\theta}}\ &\sum_{k=1}^{K}\left(y_k^0(\boldsymbol{x})+y_k(\boldsymbol{x})^\top\boldsymbol{\mu}\right)Y_{11}^k+\sum_{k=1}^{K}\left(\left(\boldsymbol{U}_{m\times m_1}\boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}}\right)^\top y_k(\boldsymbol{x})\right)^\top\boldsymbol{Y}_{12\mathrm{r}}^k\\ &+\sum_{k=1}^{K}\left(\left(\boldsymbol{U}_{m\times(m-m_1)}\boldsymbol{\Lambda}_{m-m_1}^{\frac{1}{2}}\right)^\top y_k(\boldsymbol{x})\right)^\top\boldsymbol{\theta}_k\end{aligned} \tag{23a}$$

$$\text{s.t.}\quad 1-\sum_{k=1}^{K}Y_{11}^k=0, \tag{23b}$$

$$\sqrt{\gamma_1}-\left\|\sum_{k=1}^{K}\left(\boldsymbol{Y}_{12\mathrm{r}}^{k\ \top},\boldsymbol{\theta}_k^\top\right)^\top\right\|_2\geq 0, \tag{23c}$$

$$\gamma_2\boldsymbol{I}_{m_1}-\sum_{k=1}^{K}\boldsymbol{Y}_{22\mathrm{r}}^k\succeq 0, \tag{23d}$$

$$\boldsymbol{b}Y_{11}^k-\boldsymbol{A}\boldsymbol{\mu}Y_{11}^k-\boldsymbol{A}\boldsymbol{U}_{m\times m_1}\boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}}\boldsymbol{Y}_{12\mathrm{r}}^k-\boldsymbol{A}\boldsymbol{U}_{m\times(m-m_1)}\boldsymbol{\Lambda}_{m-m_1}^{\frac{1}{2}}\boldsymbol{\theta}_k\geq 0,\ \forall k\in[K], \tag{23e}$$

$$\boldsymbol{Y}_{\mathrm{r}}^k:\begin{bmatrix} Y_{11}^k & \boldsymbol{Y}_{12\mathrm{r}}^{k\ \top} \\ \boldsymbol{Y}_{12\mathrm{r}}^k & \boldsymbol{Y}_{22\mathrm{r}}^k\end{bmatrix}\succeq 0,\ \boldsymbol{\theta}_k \text{ free},\ \forall k\in[K]. \tag{23f}$$

Given an optimal solution of Problem (23), i.e., $\boldsymbol{Y}_{\mathrm{r}}^{k*}$ and $\boldsymbol{\theta}_k^*$ for any $k\in[K]$, we construct a feasible solution of Problem (22), represented by $\bar{\boldsymbol{Y}}^k=\begin{bmatrix} \bar{Y}_{11}^k & \bar{\boldsymbol{Y}}_{12}^{k\ \top} \\ \bar{\boldsymbol{Y}}_{12}^k & \bar{\boldsymbol{Y}}_{22}^k\end{bmatrix}$ for any $k\in[K]$. For any given

$k \in [K]$, we let $\bar{Y}_{11}^{k} = Y_{11}^{k*}$, $\bar{\boldsymbol{Y}}_{12}^{k} = \frac{1}{c^k}\left({\boldsymbol{Y}_{12\mathrm{r}}^{k*}}^\top, {\boldsymbol{\theta}_k^*}^\top\right)^\top$ with $c^k \geq 1$, and

$$\bar{\boldsymbol{Y}}_{22}^{k} = \begin{bmatrix} \boldsymbol{Y}_{22\mathrm{r}}^{k*} & 0 & \cdots & 0 \\ 0 & w_1^k & \mathbf{0} & \vdots \\ \vdots & \mathbf{0} & \ddots & 0 \\ 0 & \cdots & 0 & w_{m-m_1}^k \end{bmatrix},$$

where the value of c^k as well as the value of w_i^k for any $i \in [m - m_1]$ will be determined later so that $\bar{\boldsymbol{Y}}^k$ satisfies all the constraints in Problem (22).

First, for the solution ($\bar{\boldsymbol{Y}}^k$ $\forall k \in [K]$), it satisfies constraint (22b) because $1 - \sum_{k=1}^{K} Y_{11}^{k*} = 0$ due to constraint (23b). This solution satisfies constraint (22c) because

$$\left\|\sum_{k=1}^{K} \bar{\boldsymbol{Y}}_{12}^{k}\right\|_2 = \left\|\sum_{k=1}^{K} \frac{1}{c^k}\left({\boldsymbol{Y}_{12\mathrm{r}}^{k*}}^\top, {\boldsymbol{\theta}_k^*}^\top\right)^\top\right\|_2 \leq \left\|\sum_{k=1}^{K} \left({\boldsymbol{Y}_{12\mathrm{r}}^{k*}}^\top, {\boldsymbol{\theta}_k^*}^\top\right)^\top\right\|_2 \leq \sqrt{\gamma_1},$$

where the first equality is due to the definition of $\bar{\boldsymbol{Y}}_{12}^{k}$, the first inequality is because $c^k \geq 1$, and the second inequality is because of constraint (23c). This solution also satisfies constraints (22e) because for any $k \in [K]$, we have $\boldsymbol{b}\bar{Y}_{11}^{k} - \boldsymbol{A}\boldsymbol{\mu}\bar{Y}_{11}^{k} - \boldsymbol{A}\boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\bar{\boldsymbol{Y}}_{12}^{k} = \boldsymbol{b}Y_{11}^{k*} - \boldsymbol{A}\boldsymbol{\mu}Y_{11}^{k*} - \boldsymbol{A}\boldsymbol{U}_{m\times m_1}\boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}}\boldsymbol{Y}_{12\mathrm{r}}^{k*} - \boldsymbol{A}\boldsymbol{U}_{m\times(m-m_1)}\boldsymbol{\Lambda}_{m-m_1}^{\frac{1}{2}}\boldsymbol{\theta}_k^* \geq 0$, where the first equality is due to the definition of $\bar{\boldsymbol{Y}}_{12}^{k}$ and the first inequality is because of constraints (23e).

Next, in order for ($\bar{\boldsymbol{Y}}^k$ $\forall k \in [K]$) to satisfy constraint (22d), we require $\gamma_2 \boldsymbol{I}_m - \sum_{k=1}^{K} \bar{Y}_{22}^{k} \succeq 0$, which is equivalent to

$$\gamma_2 \boldsymbol{I}_m - \sum_{k=1}^{K} \begin{bmatrix} \boldsymbol{Y}_{22\mathrm{r}}^{k*} & 0 & \cdots & 0 \\ 0 & w_1^k & \mathbf{0} & \vdots \\ \vdots & \mathbf{0} & \ddots & 0 \\ 0 & \cdots & 0 & w_{m-m_1}^k \end{bmatrix} \succeq 0 \;\Leftrightarrow\; \begin{bmatrix} \sum_{k=1}^{K} \boldsymbol{Y}_{22\mathrm{r}}^{k*} & 0 & \cdots & 0 \\ 0 & \sum_{k=1}^{K} w_1^k & \mathbf{0} & \vdots \\ \vdots & \mathbf{0} & \ddots & 0 \\ 0 & \cdots & 0 & \sum_{k=1}^{K} w_{m-m_1}^k \end{bmatrix} \preceq \begin{bmatrix} \gamma_2 \boldsymbol{I}_{m_1} & & & \\ & \gamma_2 & & \\ & & \vdots & \\ & & & \gamma_2 \end{bmatrix}.$$

Since $\sum_{k=1}^{K} \boldsymbol{Y}_{22\mathrm{r}}^{k*} \preceq \gamma_2 \boldsymbol{I}_{m_1}$ due to constraint (23d), we require

$$\sum_{k=1}^{K} w_i^k \leq \gamma_2, \;\forall i \in [m - m_1]. \tag{24}$$

In addition, in order for ($\bar{\boldsymbol{Y}}^k$ $\forall k \in [K]$) to satisfy constraints (22f), matrix $\bar{\boldsymbol{Y}}^k$ must be PSD for any $k \in [K]$. Note that if $c^k\bar{\boldsymbol{Y}}^k$ is PSD, then $\bar{\boldsymbol{Y}}^k$ is PSD because $c^k \geq 1$. Therefore, we consider the following decomposition on $c^k\bar{\boldsymbol{Y}}^k$:

$$c^k\bar{\boldsymbol{Y}}^k = \begin{bmatrix} \frac{1}{c^k} Y_{11}^{k*} & {\boldsymbol{Y}_{12\mathrm{r}}^{k*}}^\top & \mathbf{0}_{1\times(m-m_1)} \\ \boldsymbol{Y}_{12\mathrm{r}}^{k*} & c^k \boldsymbol{Y}_{22\mathrm{r}}^{k*} & \\ \mathbf{0}_{(m-m_1)\times 1} & & \mathbf{0}_{(m-m_1)\times(m-m_1)} \end{bmatrix} + \begin{bmatrix} \left(c^k - \frac{1}{c^k}\right) Y_{11}^{k*} & \mathbf{0}_{1\times m_1} & {\boldsymbol{\theta}_k^*}^\top \\ \mathbf{0}_{m_1\times 1} & \mathbf{0}_{m_1\times m_1} & \mathbf{0} \\ \boldsymbol{\theta}_k^* & \mathbf{0} & \begin{matrix} c^k w_1^k & & \\ & \ddots & \\ & & c^k w_{m-m_1}^k \end{matrix} \end{bmatrix}, \;\forall k \in [K], \tag{25}$$

where the first matrix on the RHS of (25) is clearly PSD because $\boldsymbol{Y}_{\mathrm{r}}^{k*}$ is PSD due to constraints (23f) and we require the second one to be PSD as well so that $\bar{\boldsymbol{Y}}^k$ can be PSD. By Lemma 1, we equivalently require

$$\begin{bmatrix} y_i^k & \theta_{ki}^* \\ \theta_{ki}^* & c^k w_i^k \end{bmatrix} \succeq 0,\ \forall i \in [m-m_1],\ \forall k \in [K],\ \text{with}\ \sum_{i=1}^{m-m_1} y_i^k = \left(c^k - \frac{1}{c^k}\right) Y_{11}^{k*},\ \forall k \in [K]. \tag{26}$$

Now, for a given $i \in [m-m_1]$, we let $w_i^k = w_i$ for any $k \in [K]$, and then we have $w_i \leq \frac{\gamma_2}{K}$ due to (24). It follows that, from (26), we equivalently require $c^k w_i y_i^k \geq (\theta_{ki}^*)^2$ for any $k \in [K]$ and $i \in [m-m_1]$. Therefore, for any given $k \in [K]$, we have

$$c^k \geq \frac{(\theta_{ki}^*)^2}{w_i y_i^k} \geq \frac{K}{\gamma_2} \frac{(\theta_{ki}^*)^2}{y_i^k}, \tag{27}$$

where the first inequality is due to $c^k w_i y_i^k \geq (\theta_{ki}^*)^2$ and the second inequality is because of $w_i \leq \frac{\gamma_2}{K}$. Since (27) is equivalent to $y_i^k \geq \frac{K}{\gamma_2} \frac{(\theta_{ki}^*)^2}{c^k}$ and we have $\sum_{i=1}^{m-m_1} y_i^k = \left(c^k - \frac{1}{c^k}\right) Y_{11}^{k*}$ from (26), we can conclude that

$$\begin{aligned} \sum_{i=1}^{m-m_1} y_i^k = \left(c^k - \frac{1}{c^k}\right) Y_{11}^{k*} \geq \sum_{i=1}^{m-m_1} \frac{K}{\gamma_2} \frac{(\theta_{ki}^*)^2}{c^k},\ \forall k \in [K] \\ \Rightarrow c^k \geq \sqrt{1 + \frac{1}{Y_{11}^{k*}} \sum_{i=1}^{m-m_1} \frac{K}{\gamma_2} (\theta_{ki}^*)^2},\ \forall k \in [K]. \end{aligned} \tag{28}$$

Therefore, we choose the value of c^k such that (28) is satisfied at equality, leading to (21), while the constructed solution ($\bar{\boldsymbol{Y}}^k\ \forall k \in [K]$) satisfies all the constraints in Problem (22).

Finally, by Theorem 3, we have $\overline{Z}_{\mathrm{M}}^*(m_1) - Z_{\mathrm{M}}^*(m) \geq 0$. Meanwhile, we have

$$\begin{aligned} \overline{Z}_{\mathrm{M}}^*(m_1) = & \sum_{k=1}^K \left(y_k^0(\boldsymbol{x}^*) + y_k(\boldsymbol{x}^*)^\top \boldsymbol{\mu}\right) Y_{11}^{k*} + \sum_{k=1}^K \left(\left(\boldsymbol{U}_{m\times m_1} \boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}}\right)^\top y_k(\boldsymbol{x}^*)\right)^\top \boldsymbol{Y}_{12\mathrm{r}}^{k*} \\ & + \sum_{k=1}^K \left(\left(\boldsymbol{U}_{m\times(m-m_1)} \boldsymbol{\Lambda}_{m-m_1}^{\frac{1}{2}}\right)^\top y_k(\boldsymbol{x}^*)\right)^\top \boldsymbol{\theta}_k^*, \\ Z_{\mathrm{M}}^*(m) \geq & \sum_{k=1}^K \left(y_k^0(\boldsymbol{x}^*) + y_k(\boldsymbol{x}^*)^\top \boldsymbol{\mu}\right) Y_{11}^{k*} + \sum_{k=1}^K \left(\left(\boldsymbol{U}_{m\times m_1} \boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}}\right)^\top y_k(\boldsymbol{x}^*)\right)^\top \frac{\boldsymbol{Y}_{12\mathrm{r}}^{k*}}{c^k} \\ & + \sum_{k=1}^K \left(\left(\boldsymbol{U}_{m\times(m-m_1)} \boldsymbol{\Lambda}_{m-m_1}^{\frac{1}{2}}\right)^\top y_k(\boldsymbol{x}^*)\right)^\top \frac{\boldsymbol{\theta}_k^*}{c^k}, \end{aligned}$$

where the inequality holds because the constructed solution ($\bar{\boldsymbol{Y}}^k\ \forall k \in [K]$) is feasible for Problem (22), which is a maximization problem. Therefore, it follows that

$$0 \leq \overline{Z}_{\mathrm{M}}^*(m_1) - Z_{\mathrm{M}}^*(m) \leq \sum_{k=1}^K \left(\left(\boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top y_k(\boldsymbol{x}^*)\right)^\top \left({\boldsymbol{Y}_{12\mathrm{r}}^{k*}}^\top, {\boldsymbol{\theta}_k^*}^\top\right)^\top \left(\frac{c^k - 1}{c^k}\right). \quad \square$$

Proposition 7. *Suppose that Assumption 1 holds,* $\boldsymbol{\mu} \in S$*, and* $\min_{k=1}^{K}\left\{y_k^0(\boldsymbol{x}^*) + y_k(\boldsymbol{x}^*)^\top \boldsymbol{\mu}\right\} \geq 0$*, where* $\boldsymbol{x}^*$ *is an optimal solution of Problem (4), then the relative gap between the optimal values of Problems (18) and (4) is bounded from above by* $\sqrt{P}-1$*, i.e.,* $0 \leq UB_{\mathrm{M}}^* - Z_{\mathrm{M}}^*(m) \leq (\sqrt{P}-1) Z_{\mathrm{M}}^*(m)$*.*

Proof. We reformulate Problem (18) as the following problem:

$$UB_{\mathrm{M}}^* = \min_{\substack{\boldsymbol{x},s,\boldsymbol{q},\\ \boldsymbol{Q}_i,\boldsymbol{\lambda}_k}} \; s + \gamma_2 \sum_{i=1}^{P} \boldsymbol{I}_{m_i} \bullet \boldsymbol{Q}_i + \sqrt{\gamma_1}\, \|\boldsymbol{q}\|_2 \tag{29a}$$

$$\text{s.t.} \quad \begin{bmatrix} s - y_k^0(\boldsymbol{x}) - \boldsymbol{\lambda}_k^\top b - y_k(\boldsymbol{x})^\top \boldsymbol{\mu} + \boldsymbol{\lambda}_k^\top \boldsymbol{A}\boldsymbol{\mu} & \frac{1}{2}\left(\boldsymbol{q} + \left(\boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top \left(\boldsymbol{A}^\top \boldsymbol{\lambda}_k - y_k(\boldsymbol{x})\right)\right)^\top \\ \frac{1}{2}\left(\boldsymbol{q} + \left(\boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top \left(\boldsymbol{A}^\top \boldsymbol{\lambda}_k - y_k(\boldsymbol{x})\right)\right) & \boldsymbol{Q}^{'} \end{bmatrix} \succeq 0, \; \forall k \in [K], \tag{29b}$$

$$\boldsymbol{x} \in \mathcal{X}, \; \boldsymbol{\lambda}_k \geq 0, \; \forall k \in [K],$$

where

$$\boldsymbol{Q}^{'} = \begin{bmatrix} \boldsymbol{Q}_1 & \boldsymbol{0}_{m_1 \times m_2} & \cdots & \boldsymbol{0}_{m_1 \times m_{P-1}} & \boldsymbol{0}_{m_1 \times m_P} \\ \boldsymbol{0}_{m_2 \times m_1} & \boldsymbol{Q}_2 & \cdots & \boldsymbol{0}_{m_2 \times m_{P-1}} & \boldsymbol{0}_{m_2 \times m_P} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \boldsymbol{0}_{m_{P-1} \times m_1} & \boldsymbol{0}_{m_{P-1} \times m_2} & \cdots & \boldsymbol{Q}_{P-1} & \boldsymbol{0}_{m_{P-1} \times m_P} \\ \boldsymbol{0}_{m_P \times m_1} & \boldsymbol{0}_{m_P \times m_2} & \cdots & \boldsymbol{0}_{m_P * m_{P-1}} & \boldsymbol{Q}_P \end{bmatrix} \tag{30}$$

and $\boldsymbol{Q}_i \in \mathbb{R}^{m_i \times m_i}$ for any $i \in [P]$ with $\sum_{i=1}^{P} m_i = m$. Let $(\boldsymbol{x}^*, s^*, \boldsymbol{\lambda}_k^* \; \forall k \in [K], \boldsymbol{q}^*, \boldsymbol{Q}^*)$ denote an optimal solution of Problem (4) with

$$\boldsymbol{Q}^* = \begin{bmatrix} \boldsymbol{Q}_1^* & \boldsymbol{Q}_{m_1 \times m_2}^* & \cdots & \boldsymbol{Q}_{m_1 \times m_{P-1}}^* & \boldsymbol{Q}_{m_1 \times m_P}^* \\ \boldsymbol{Q}_{m_2 \times m_1}^* & \boldsymbol{Q}_2^* & \cdots & \boldsymbol{Q}_{m_2 \times m_{P-1}}^* & \boldsymbol{Q}_{m_2 \times m_P}^* \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \boldsymbol{Q}_{m_{P-1} \times m_1}^* & \boldsymbol{Q}_{m_{P-1} \times m_2}^* & \cdots & \boldsymbol{Q}_{P-1}^* & \boldsymbol{Q}_{m_{P-1} \times m_P}^* \\ \boldsymbol{Q}_{m_P \times m_1}^* & \boldsymbol{Q}_{m_P \times m_2}^* & \cdots & \boldsymbol{Q}_{m_P \times m_{P-1}}^* & \boldsymbol{Q}_P^* \end{bmatrix}. \tag{31}$$

Based on this optimal solution, in the following, we construct a feasible solution of Problem (29), denoted by $\left(\bar{\boldsymbol{x}}, \bar{s}, \bar{\boldsymbol{q}}, \bar{\boldsymbol{Q}}', \bar{\boldsymbol{\lambda}}_k \; \forall k \in [K]\right)$.

First, we let $\bar{\boldsymbol{x}} = \boldsymbol{x}^*$, $\bar{s} = k_0 s^*$, $\bar{\boldsymbol{q}} = \boldsymbol{q}^*$, $\bar{\boldsymbol{\lambda}}_k = \boldsymbol{\lambda}_k^*$ for any $k \in [K]$, and

$$\bar{\boldsymbol{Q}}' = \begin{bmatrix} k_1 \boldsymbol{Q}_1^* & \boldsymbol{0}_{m_1 \times m_2} & \cdots & \boldsymbol{0}_{m_1 \times m_{P-1}} & \boldsymbol{0}_{m_1 \times m_P} \\ \boldsymbol{0}_{m_2 \times m_1} & k_2 \boldsymbol{Q}_2^* & \cdots & \boldsymbol{0}_{m_2 \times m_{P-1}} & \boldsymbol{0}_{m_2 \times m_P} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \boldsymbol{0}_{m_{P-1} \times m_1} & \boldsymbol{0}_{m_{P-1} \times m_2} & \cdots & k_{P-1} \boldsymbol{Q}_{P-1}^* & \boldsymbol{0}_{m_{P-1} \times m_P} \\ \boldsymbol{0}_{m_P \times m_1} & \boldsymbol{0}_{m_P \times m_2} & \cdots & \boldsymbol{0}_{m_P \times m_{P-1}} & k_P \boldsymbol{Q}_P^* \end{bmatrix}, \tag{32}$$

with $k_i \geq 1$ for any $i \in \{0, 1, 2, \ldots, P\}$. In order for this solution to satisfy (29b), we require

$$\begin{bmatrix} k_0 s^* - y_k^0(\boldsymbol{x}^*) - {\boldsymbol{\lambda}_k^*}^\top \boldsymbol{b} - y_k(\boldsymbol{x}^*)^\top \boldsymbol{\mu} + {\boldsymbol{\lambda}^*}_k^\top \boldsymbol{A}\boldsymbol{\mu} & \frac{1}{2}\left(\boldsymbol{q}^* + \left(\boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top \left(\boldsymbol{A}^\top \boldsymbol{\lambda}_k^* - y_k(\boldsymbol{x}^*)\right)\right)^\top \\ \frac{1}{2}\left(\boldsymbol{q}^* + \left(\boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top \left(\boldsymbol{A}^\top \boldsymbol{\lambda}_k^* - y_k(\boldsymbol{x}^*)\right)\right) & \bar{\boldsymbol{Q}}' \end{bmatrix} \succeq 0, \; \forall k \in [K]. \tag{33}$$

In the following, we find the values of k_i for any $i \in \{0, 1, 2, \ldots, P\}$ so that (33) holds. To that end, we construct the following matrix

$$\begin{bmatrix} k_0\left(s^* - y_k^0(\boldsymbol{x}^*) - {\boldsymbol{\lambda}_k^*}^\top \boldsymbol{b} - y_k(\boldsymbol{x}^*)^\top \boldsymbol{\mu} + {\boldsymbol{\lambda}^*}_k^\top \boldsymbol{A}\boldsymbol{\mu}\right) & \frac{1}{2}\left(\boldsymbol{q}^* + \left(\boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top \left(\boldsymbol{A}^\top \boldsymbol{\lambda}_k^* - y_k(\boldsymbol{x}^*)\right)\right)^\top \\ \frac{1}{2}\left(q^* + \left(\boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top \left(\boldsymbol{A}^\top \boldsymbol{\lambda}_k^* - y_k(\boldsymbol{x}^*)\right)\right) & \bar{\boldsymbol{Q}}' \end{bmatrix}, \; \forall k \in [K]. \tag{34}$$

Note that subtracting (34) from (33) leads to the following matrix:

$$\begin{bmatrix} (k_0-1)\left(y_k^0(\boldsymbol{x}^*)+y_k(\boldsymbol{x}^*)^\top\boldsymbol{\mu}-\boldsymbol{\lambda}_k^{*\top}(\boldsymbol{A}\boldsymbol{\mu}-\boldsymbol{b})\right) & \mathbf{0}_{1\times m} \\ \mathbf{0}_{m\times 1} & \mathbf{0}_{m\times m} \end{bmatrix} \succeq 0,\ \forall k\in[K],$$

which is PSD because its eigenvalues are non-negative. In fact, $(k_0-1)\left(y_k^0(\boldsymbol{x}^*)+y_k(\boldsymbol{x}^*)^\top\boldsymbol{\mu}-\boldsymbol{\lambda}_k^{*\top}(\boldsymbol{A}\boldsymbol{\mu}-\boldsymbol{b})\right)$ is the only non-zero eigenvalue of this matrix that is non-negative because $k_0\geq 1$, $-\boldsymbol{\lambda}_k^{*\top}(\boldsymbol{A}\boldsymbol{\mu}-\boldsymbol{b})\geq 0$ due to $\boldsymbol{A}\boldsymbol{\mu}\leq\boldsymbol{b}$ and $\boldsymbol{\lambda}_k^*\geq 0$, and we have $y_k^0(\boldsymbol{x}^*)+y_k(\boldsymbol{x}^*)^\top\boldsymbol{\mu}\geq 0$ according to the assumption $\min_{k=1}^K\left\{y_k^0(\boldsymbol{x}^*)+y_k(\boldsymbol{x}^*)^\top\boldsymbol{\mu}\right\}\geq 0$. Thus, we choose good values of k_i for any $i\in\{0,1,2,\dots,P\}$ to ensure (34) to be a PSD matrix and accordingly will make (33) hold.

Next, by Lemma 1, in order for (34) to be a PSD, we equivalently require

$$\begin{bmatrix} s_i(s^*-y_k^0(\boldsymbol{x}^*)-\boldsymbol{\lambda}_k^{*\top}\boldsymbol{b}-y_k(\boldsymbol{x}^*)^\top\boldsymbol{\mu}+\boldsymbol{\lambda}^{*\top}_k\boldsymbol{A}\boldsymbol{\mu}) & \frac{1}{2}\left(\boldsymbol{q}_i^*+\left(\boldsymbol{U}_{m\times m_i}\boldsymbol{\Lambda}_{m_i}^{\frac{1}{2}}\right)^\top(\boldsymbol{A}^\top\boldsymbol{\lambda}_k^*-y_k(\boldsymbol{x}^*))\right)^\top \\ \frac{1}{2}\left(\boldsymbol{q}_i^*+\left(\boldsymbol{U}_{m\times m_i}\boldsymbol{\Lambda}_{m_i}^{\frac{1}{2}}\right)^\top(\boldsymbol{A}^\top\boldsymbol{\lambda}_k^*-y_k(\boldsymbol{x}^*))\right) & k_i\boldsymbol{Q}_i^* \end{bmatrix}\succeq 0,\ \forall k\in[K], i\in[P], \quad (35)$$

with $\sum_{i=1}^P s_i=k_0$. Constraints (35) can be satisfied by allowing $s_i\times k_i\geq 1$ for any $i\in[P]$ due to (4b). Then, we let $k_0=k_1=\cdots=k_P$ and $s_i\times k_i=1$ for any $i\in[P]$, leading to $k_0=k_1=\cdots=k_P=\sqrt{P}$.

Finally, we have $UB_{\mathrm{M}}^*\geq Z_{\mathrm{M}}^*(m)$ by Theorem 4. Meanwhile, as Problem (18) is a minimization problem, UB_{M}^* is no larger than the objective value corresponding to our constructed feasible solution $\left(\bar{\boldsymbol{x}},\bar{s},\bar{\boldsymbol{q}},\bar{\boldsymbol{Q}}',\bar{\boldsymbol{\lambda}}_k\ \forall k\in[K]\right)$. That is, we have

$$UB_{\mathrm{M}}^*\leq\sqrt{P}s^*+\gamma_2\sum_{i=1}^P\boldsymbol{I}_{m_i}\bullet\left(\sqrt{P}\boldsymbol{Q}_i^*\right)+\sqrt{\gamma_1}\,\|\boldsymbol{q}^*\|_2\leq\sqrt{P}\left(s^*+\gamma_2\sum_{i=1}^P\boldsymbol{I}_{m_i}\bullet\boldsymbol{Q}_i^*+\sqrt{\gamma_1}\,\|\boldsymbol{q}^*\|_2\right)=\sqrt{P}Z_{\mathrm{M}}^*(m),$$

where the second inequality holds because $P\geq 1$. Therefore, we have

$$0\leq UB_{\mathrm{M}}^*-Z_{\mathrm{M}}^*(m)\leq(\sqrt{P}-1)Z_{\mathrm{M}}^*(m).\quad\square$$

We observe that the theoretical upper bound in Proposition 7 is achievable through the following example.

Example 1. For simplicity, suppose that $\boldsymbol{\mu}=\mathbf{0}$, $\mathcal{S}=\mathbb{R}^m$, $\gamma_1=+\infty$, and $\gamma_2=1$. With $f(\boldsymbol{x},\boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}}+\boldsymbol{\mu})=|\boldsymbol{x}^\top\boldsymbol{\xi}_{\mathrm{I}}|$, Problem (4) can be recast as the following SDP formulation:

$$\begin{aligned} Z_{\mathrm{M}}^*(m)=\min_{\boldsymbol{x},s,\boldsymbol{Q}}\quad & s+\boldsymbol{I}_m\bullet\boldsymbol{Q} \\ \text{s.t.}\quad & \begin{bmatrix} s & \frac{\boldsymbol{x}^\top}{2} \\ \frac{\boldsymbol{x}}{2} & \boldsymbol{Q}\end{bmatrix}\succeq 0,\ \begin{bmatrix} s & \frac{-\boldsymbol{x}^\top}{2} \\ \frac{-\boldsymbol{x}}{2} & \boldsymbol{Q}\end{bmatrix}\succeq 0,\ \boldsymbol{x}\in\mathcal{X}. \end{aligned}\quad (36)$$

For fixed $\boldsymbol{x}\in\mathcal{X}$, optimizing over the remaining decision variables in (36) yields $\boldsymbol{Q}^*=\frac{\boldsymbol{x}\boldsymbol{x}^\top}{4s}$, and $s^*=\frac{\sqrt{\boldsymbol{x}^\top\boldsymbol{x}}}{2}$ with objective value $\sqrt{\boldsymbol{x}^\top\boldsymbol{x}}$. By Proposition 5, we obtain an upper bound by considering

$$\begin{aligned} UB_{\mathrm{M}}^*=\min_{\boldsymbol{x},s,\boldsymbol{Q}_i}\quad & s+\sum_{i=1}^P\boldsymbol{I}_i\bullet\boldsymbol{Q}_i \\ \text{s.t.}\quad & \begin{bmatrix} s & \frac{\boldsymbol{x}^\top}{2} \\ \frac{\boldsymbol{x}}{2} & \mathrm{diag}(\boldsymbol{Q}_i)\end{bmatrix}\succeq 0,\ \begin{bmatrix} s & \frac{-\boldsymbol{x}^\top}{2} \\ \frac{-\boldsymbol{x}}{2} & \mathrm{diag}(\boldsymbol{Q}_i)\end{bmatrix}\succeq 0,\ \boldsymbol{x}\in\mathcal{X}, \end{aligned}\quad (37)$$

where $\text{diag}(\boldsymbol{Q}_i)$ is a block diagonal matrix consisting of $\boldsymbol{Q}_1,\dots,\boldsymbol{Q}_P$. For fixed $\boldsymbol{x}\in\mathcal{X}$, where $\boldsymbol{x}=(\boldsymbol{x}_1^\top,\dots,\boldsymbol{x}_P^\top)^\top$ and $\boldsymbol{x}_i\in\mathbb{R}^{m_i}$ for all $i\in[P]$, optimizing over the remaining decision variables in Problem (37) yields $\boldsymbol{Q}_i^*=\frac{\boldsymbol{x}_i\boldsymbol{x}_i^\top}{4s_i}$ for all $i\in[P]$, and $s^*=\sum_{i=1}^P\frac{\sqrt{\boldsymbol{x}_i^\top\boldsymbol{x}_i}}{2}$ with objective value $\sum_{i=1}^P\sqrt{\boldsymbol{x}_i^\top\boldsymbol{x}_i}$. Now we let $m_i=\frac{m}{P}$ for any $i\in[P]$ and $\mathcal{X}=\{\boldsymbol{x}\in\mathbb{R}^m|x_i\geq 1,\forall i\in[m]\}$. It follows that $Z_{\mathrm{M}}^*(m)=\sqrt{m}$ and $UB_{\mathrm{M}}^*=P\sqrt{\frac{m}{P}}$, and so the relative gap between $Z_{\mathrm{M}}^*(m)$ and UB_{M}^* is

$$\frac{P\sqrt{\frac{m}{P}}-\sqrt{m}}{\sqrt{m}}=\sqrt{P}-1,$$

attaining the theoretical upper bound in Proposition 7.

3. Combined Ambiguity Set

We consider the combined ambiguity set that incorporates both Wasserstein distance and moment information. Like in the last section, we derive an SDP reformulation of the corresponding DRO problem, as well as its inner and outer approximations that can be solved more efficiently. Furthermore, we bound the gaps between the optimal values of the DRO problem and its approximations.

Formally, we consider DRO problem

$$\min_{\boldsymbol{x}\in\mathcal{X}}\ \min_{\mathbb{P}\in\mathcal{D}_{\mathrm{C1}}}\ \mathbb{E}_{\mathbb{P}}\left[f\left(\boldsymbol{x},\boldsymbol{\xi}\right)\right], \tag{DRO-C}$$

where

$$\mathcal{D}_{\mathrm{C1}}\left(\mathcal{S},\boldsymbol{\mu},\boldsymbol{\Sigma},\gamma_2,\mathbb{P}_0,R_0\right)=\left\{\mathbb{P}\left|\begin{array}{l}\mathbb{E}_{\mathbb{P}}\left[\left(\boldsymbol{\xi}-\boldsymbol{\mu}\right)\left(\boldsymbol{\xi}-\boldsymbol{\mu}\right)^\top\right]\preceq\gamma_2\boldsymbol{\Sigma}\\ W\left(\mathbb{P},\mathbb{P}_0\right)\leq R_0\end{array}\right.\right\}.$$

In this combined ambiguity set, $\mathbb{P}_0$ denotes a reference distribution. For example, $\mathbb{P}_0$ is an empirical distribution of $\boldsymbol{\xi}$ generated by N i.i.d. samples $\{\hat{\boldsymbol{\xi}}^i:i\in[N]\}$ of $\boldsymbol{\xi}$, i.e., $\mathbb{P}_0\{\boldsymbol{\xi}=\hat{\boldsymbol{\xi}}^i\}=\frac{1}{N}$ for all $i\in[N]$. In addition, $W(\mathbb{P},\mathbb{P}_0)$ denotes the type-1 Wasserstein distance between $\mathbb{P}$ and $\mathbb{P}_0$ defined through

$$W\left(\mathbb{P},\mathbb{P}_0\right):=\min_{\pi}\left\{\int_{\mathcal{S}^2}\|\boldsymbol{\xi}-\hat{\boldsymbol{\xi}}\|_1\pi\left(\boldsymbol{\xi},\hat{\boldsymbol{\xi}}\right)\right\},$$

where π denotes a joint distribution of $\boldsymbol{\xi}$ and $\hat{\boldsymbol{\xi}}$ with marginals $\mathbb{P}$ and $\mathbb{P}_0$, respectively. Intuitively, $W\left(\mathbb{P},\mathbb{P}_0\right)$ represents the minimum expected distance between $\boldsymbol{\xi}$ and $\hat{\boldsymbol{\xi}}$ over all possible joint distributions π. It has been shown that, as $N\to\infty$, $\mathbb{P}_0$ converges to the true distribution of $\boldsymbol{\xi}$ almost surely (Van der Vaart 2000). As a result, if we select the value of $R_0>0$ appropriately, then the Wasserstein ball centered at $\mathbb{P}_0$ with radius R_0 will include such true distribution with high confidence. Besides the Wasserstein ball, $\mathcal{D}_{\mathrm{C1}}$ designates that the centered second-order moment matrix of $\boldsymbol{\xi}$ is bounded by $\gamma_2\boldsymbol{\Sigma}$. We notice that R_0 controls the conservatism degree of $\mathcal{D}_{\mathrm{C1}}$. The larger radius R_0 is, $\mathcal{D}_{\mathrm{C1}}$ has higher confidence to contain the true distribution of $\boldsymbol{\xi}$, while it leads to a more conservative optimal solution to (DRO-C). In contrast, when R_0 decreases to zero, (DRO-C) reduces to an ambiguity-free stochastic program with regard to $\mathbb{P}_0$. For (DRO-C), we consider a setting slightly stronger than that in Assumption 1.

Assumption 2. *Function* $f(\boldsymbol{x},\boldsymbol{\xi})$ *is piecewise linear convex in* $\boldsymbol{\xi}$*, i.e.,* $f(\boldsymbol{x},\boldsymbol{\xi}) = \max_{k=1}^{K}\left\{y_k^0(\boldsymbol{x}) + y_k(\boldsymbol{x})^\top \boldsymbol{\xi}\right\}$ *with both* $y_k(\boldsymbol{x}) = (y_k^1(\boldsymbol{x}),\dots,y_k^m(\boldsymbol{x}))^\top$ *and* $y_k^0(\boldsymbol{x})$ *affine in* $\boldsymbol{x}$ *for any* $k\in[K]$*. Additionally,* $\mathcal{S}=\mathbb{R}^m$*.*

Proposition 8. *Under Assumption 2, (DRO-C) can be recast as the following SDP formulation:*

$$Z_{\mathrm{C}}^*(m) := \min_{\boldsymbol{x},\lambda,\boldsymbol{Q},\boldsymbol{\zeta}^i,y_i} \quad \lambda R_0 + \gamma_2 \boldsymbol{\Sigma}\bullet\boldsymbol{Q} + \frac{1}{N}\sum_{i=1}^{N} y_i \tag{38a}$$

$$\text{s.t.} \begin{bmatrix} \boldsymbol{Q} & \frac{1}{2}\left(-y_k(\boldsymbol{x})+\boldsymbol{\zeta}^i - 2\boldsymbol{Q}\boldsymbol{\mu}\right) \\ \frac{1}{2}\left(-y_k(\boldsymbol{x})+\boldsymbol{\zeta}^i - 2\boldsymbol{Q}\boldsymbol{\mu}\right)^\top & y_i - y_k^0(\boldsymbol{x}) - {\boldsymbol{\zeta}^i}^\top\hat{\boldsymbol{\xi}}^i + \boldsymbol{\mu}^\top\boldsymbol{Q}\boldsymbol{\mu}\end{bmatrix} \succeq 0,\ \forall i\in[N],\ \forall k\in[K], \tag{38b}$$

$$\lambda\in\mathbb{R}_+,\ \boldsymbol{x}\in\mathcal{X},\ \left\|\boldsymbol{\zeta}^i\right\|_* \le \lambda,\ \forall i\in[N],$$

where $\boldsymbol{Q}\in\mathbb{R}^{m\times m}$ *and* $\boldsymbol{\zeta}^i\in\mathbb{R}^m$ *for any* $i\in[N]$*.*

Proof. The result is deduced from Corollary 1 in Gao and Kleywegt (2017). □

As discussed in Section 2, Problem (38) can be computationally difficult when $\boldsymbol{\xi}$ is high-dimensional and/or correlated, leading to many large-scale PSD constraints. We derive more efficiently solvable outer and inner approximations of Problem (38) (i.e., (DRO-C)) in Sections 3.1 and 3.2, leading to lower and upper bounds, respectively, while theoretically showing their quality.

3.1. Lower Bound

By performing the eigenvalue decomposition on matrix $\boldsymbol{\Sigma}$, we first reformulate (DRO-C) as

$$\min_{\boldsymbol{x}\in\mathcal{X}} \max_{\mathbb{P}_{\mathrm{I}}\in\mathcal{D}_{\mathrm{C2}}} \mathbb{E}_{\mathbb{P}_{\mathrm{I}}}\left[f\left(\boldsymbol{x},\boldsymbol{U}\Lambda^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}}+\boldsymbol{\mu}\right)\right], \tag{39}$$

where

$$\mathcal{D}_{\mathrm{C2}}\left(\mathcal{S}_{\mathrm{I}},\boldsymbol{\mu},\gamma_2,\mathbb{P}_0,R_0\right) = \left\{\mathbb{P}_{\mathrm{I}}\left|\begin{array}{l}\mathbb{E}_{\mathbb{P}_{\mathrm{I}}}\left[\boldsymbol{\xi}_{\mathrm{I}}\boldsymbol{\xi}_{\mathrm{I}}^\top\right]\preceq\gamma_2\boldsymbol{I}_m \\ \int_{\mathcal{S}^2}\left\|\boldsymbol{U}\Lambda^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}}+\boldsymbol{\mu}-\hat{\boldsymbol{\xi}}\right\|_1\pi\left(\boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}}+\boldsymbol{\mu},\hat{\boldsymbol{\xi}}\right)\le R_0\end{array}\right.\right\}$$

under the condition that $f(\boldsymbol{x},\boldsymbol{U}\Lambda^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}}+\boldsymbol{\mu})$ is $\mathbb{P}_{\mathrm{I}}$-integrable for any $\mathbb{P}_{\mathrm{I}}\in\mathcal{D}_{\mathrm{C2}}$. Next, by the approximation of $\boldsymbol{\xi}$ in (5) due to PCA, we outer approximate (39) as the following problem:

$$\min_{\boldsymbol{x}\in\mathcal{X}} \max_{\mathbb{P}_{\mathrm{r}}\in\mathcal{D}_{\mathrm{C3}}} \mathbb{E}_{\mathbb{P}_{\mathrm{r}}}\left[f\left(\boldsymbol{x},\boldsymbol{U}_{m\times m_1}\boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{r}}+\boldsymbol{\mu}\right)\right], \tag{40a}$$

where

$$\mathcal{D}_{\mathrm{C3}}\left(\mathcal{S}_{\mathrm{r}},\boldsymbol{\mu},\gamma_2,\mathbb{P}_0,R_0\right) = \left\{\mathbb{P}_{\mathrm{r}}\left|\begin{array}{l}\mathbb{E}_{\mathbb{P}_{\mathrm{r}}}\left[\boldsymbol{\xi}_{\mathrm{r}}\boldsymbol{\xi}_{\mathrm{r}}^\top\right]\preceq\gamma_2\boldsymbol{I}_{m_1} \\ \int_{\mathcal{S}^2}\left\|\boldsymbol{U}_{m\times m_1}\boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{r}}+\boldsymbol{\mu}-\hat{\boldsymbol{\xi}}\right\|_1\pi\left(\boldsymbol{U}_{m\times m_1}\boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{r}}+\boldsymbol{\mu},\hat{\boldsymbol{\xi}}\right)\le R_0\end{array}\right.\right\} \tag{40b}$$

with

$$\mathcal{S}_{\mathrm{r}} := \left\{\boldsymbol{\xi}_{\mathrm{r}}\in\mathbb{R}^{m_1} : \boldsymbol{U}_{m\times m_1}\boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{r}}+\boldsymbol{\mu}\in\mathcal{S}\right\} = \mathbb{R}^{m_1}. \tag{40c}$$

Note that $\hat{\boldsymbol{\xi}}$ is a given data point and thus it is not approximated following what we do for $\boldsymbol{\xi}$.

THEOREM 5. *Under Assumption 2, Problem (40) has the same optimal value as the following SDP formulation:*

$$Z_{\mathrm{C}}^{*}(m_1) := \min_{\boldsymbol{x},\lambda,\boldsymbol{Q}_{\mathrm{r}},\boldsymbol{\zeta}^i,y_i} \quad \lambda R_0 + \gamma_2 \boldsymbol{I}_{m_1} \bullet \boldsymbol{Q}_{\mathrm{r}} + \frac{1}{N}\sum_{i=1}^{N} y_i \tag{41a}$$

$$\text{s.t.} \quad \begin{bmatrix} \boldsymbol{Q}_{\mathrm{r}} & \frac{1}{2}\left(\left(-y_k(\boldsymbol{x})+\boldsymbol{\zeta}^i\right)^\top \boldsymbol{U}_{m\times m_1}\boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}}\right)^\top \\ \frac{1}{2}\left(-y_k(\boldsymbol{x})+\boldsymbol{\zeta}^i\right)^\top \boldsymbol{U}_{m\times m_1}\boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}} & y_i - y_k(\boldsymbol{x})^\top \boldsymbol{\mu} - y_k^0(\boldsymbol{x}) + \left(\boldsymbol{\mu}-\hat{\boldsymbol{\xi}}^i\right)^\top \boldsymbol{\zeta}^i \end{bmatrix} \succeq 0, \quad \forall i \in [N],\ \forall k \in [K], \tag{41b}$$

$$\lambda \in \mathbb{R}_+,\ \boldsymbol{x} \in \mathcal{X},\ \left\|\boldsymbol{\zeta}^i\right\|_* \le \lambda,\ \forall i \in [N], \tag{41c}$$

where $\boldsymbol{Q}_{\mathrm{r}} \in \mathbb{R}^{m_1\times m_1}$ and $\boldsymbol{\zeta}^i \in \mathbb{R}^m$. Furthermore, we have the following: (i) Problem (41) provides a lower bound for the optimal value of (DRO-C); (ii) the optimal value of Problem (41) is nondecreasing in m_1; and (iii) if $m_1 = m$, then (DRO-C) and (41) have the same optimal value.

Proof. See Appendix B.1 for the detailed proof. □

We show the quality of the outer approximation (41) in the following proposition.

PROPOSITION 9. *It holds that*

$$0 \le Z^*{}_{\mathrm{C}}(m) - Z^*{}_{\mathrm{C}}(m_1) \le \sqrt{\frac{\gamma_2}{N}} \sum_{i=1}^{N}\sum_{k=1}^{K} \sqrt{\boldsymbol{L}_{m-m_1}^{ik}\left(\boldsymbol{L}_{m-m_1}^{ik}\right)^\top}, \tag{42}$$

where $\boldsymbol{L}_{m-m_1}^{ik} = \left(-y_k(\boldsymbol{x}^*)+\boldsymbol{\zeta}^{i^*}\right)^\top \boldsymbol{U}_{m\times(m-m_1)}(\boldsymbol{\Lambda}^{m-m_1})^{\frac{1}{2}}$, $\boldsymbol{x}^*$ and $\boldsymbol{\zeta}^{i^*}$ ($\forall i \in [N]$) denote an optimal solution of Problem (41), and $\boldsymbol{\Lambda}^{m-m_1} \in \mathbb{R}^{(m-m_1)\times(m-m_1)}$ denotes the lower-right submatrix of $\boldsymbol{\Lambda}$.

Proof. See Appendix B.2 for the detailed proof. □

3.2. Upper Bound

We further inner approximate Problem (38), leading to an upper bound, by splitting random vector $\boldsymbol{\xi}_{\mathrm{I}}$ into P pieces in the second-moment constraint in $\mathcal{D}_{\mathrm{C2}}$ so that $\boldsymbol{\xi}_{\mathrm{I}} = \left(\boldsymbol{\xi}_{\mathrm{I}_1}^\top, \boldsymbol{\xi}_{\mathrm{I}_2}^\top, \ldots, \boldsymbol{\xi}_{\mathrm{I}_P}^\top\right)^\top$, where $\boldsymbol{\xi}_{\mathrm{I}_j} \in \mathbb{R}^{m_j}$, $\forall j \in [P]$, and $\sum_{j=1}^{P} m_j = m$. This gives rise to an inner approximation

$$\min_{\boldsymbol{x}\in\mathcal{X}} \ \max_{\mathbb{P}_{\mathrm{I}}\in\mathcal{D}_{\mathrm{C4}}} \ \mathbb{E}_{\mathbb{P}_{\mathrm{I}}}\left[f\left(\boldsymbol{x}, \boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}} + \boldsymbol{\mu}\right)\right], \tag{43}$$

where

$$\mathcal{D}_{\mathrm{C4}}\left(\mathcal{S}_{\mathrm{I}}, \boldsymbol{\mu}, \gamma_2, \mathbb{P}_0, R_0\right) = \left\{ \mathbb{P}_{\mathrm{I}} \left| \begin{array}{l} \mathbb{E}_{\mathbb{P}_{\mathrm{I}}}\left[\boldsymbol{\xi}_{\mathrm{I}_j}\boldsymbol{\xi}_{\mathrm{I}_j}^\top\right] \preceq \gamma_2 \boldsymbol{I}_{m_j},\ \forall j \in [P] \\ \int_{\mathcal{S}^2} \left\|\boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}} + \boldsymbol{\mu} - \hat{\boldsymbol{\xi}}\right\|_1 \pi\left(\boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}} + \boldsymbol{\mu}, \hat{\boldsymbol{\xi}}\right) \le R_0 \end{array} \right. \right\}.$$

Theorem 6. *Under Assumption 2, Problem (43) has the same optimal value as the following SDP formulation:*

$$
\begin{aligned}
UB_{\mathrm{C}}^{*} := \min_{\substack{\boldsymbol{x},\lambda,\boldsymbol{Q}_j,\\ \boldsymbol{\zeta}^i, y_i}} \ & \lambda R_0 + \gamma_2 \sum_{j=1}^{P} \boldsymbol{I}_{m_j} \bullet \boldsymbol{Q}_j + \frac{1}{N}\sum_{i=1}^{N} y_i \qquad (44)\\
\text{s.t.} \ & \begin{bmatrix} \boldsymbol{Q}_j & \frac{1}{2}\left(\left(-y_k(\boldsymbol{x})+\boldsymbol{\zeta}^i\right)^\top \boldsymbol{U}_{m\times m_j}\boldsymbol{\Lambda}_{m_j}^{\frac{1}{2}}\right)^\top \\ \frac{1}{2}\left(-y_k(\boldsymbol{x})+\boldsymbol{\zeta}^i\right)^\top \boldsymbol{U}_{m\times m_j}\boldsymbol{\Lambda}_{m_j}^{\frac{1}{2}} & s_{jik} \end{bmatrix} \succeq 0, \ \forall j\in[P], \ \forall i\in[N], \\
& \qquad \forall k \in [K],\\
& \sum_{j=1}^{P} s_{jik} = y_i - y_k(\boldsymbol{x})^\top \boldsymbol{\mu} - y_k^0(\boldsymbol{x}) + \left(\boldsymbol{\mu} - \hat{\boldsymbol{\xi}}^i\right)^\top \boldsymbol{\zeta}^i, \ \forall i\in[N], \ \forall k\in[K],\\
& \lambda \in \mathbb{R}_+, \ \boldsymbol{x}\in\mathcal{X}, \ \left\|\boldsymbol{\zeta}^i\right\|_* \le \lambda, \ \forall i \in [N],
\end{aligned}
$$

where $\boldsymbol{Q}_j \in \mathbb{R}^{m_j\times m_j}$ and $\boldsymbol{\zeta}^i \in \mathbb{R}^m$ for any $i\in[N]$. Furthermore, Problem (44) provides an upper bound for the optimal value of (DRO-C).

Proof. See Appendix B.3 for the detailed proof. □

We show the quality of our derived inner approximation (44) in the following proposition.

Proposition 10. *If $\min_{k=1}^{K}\left\{y_k^0(\boldsymbol{x}^*) + y_k(\boldsymbol{x}^*)^\top \boldsymbol{\mu}\right\} \ge 0$ and $\max_{i=1}^{N}\{(\boldsymbol{\mu}-\hat{\boldsymbol{\xi}}^i)^\top \boldsymbol{\zeta}^{i^*}\} \le 0$, where $\boldsymbol{x}^*$ and $\boldsymbol{\zeta}^{i^*}$ are optimal solutions of Problem (38), then the relative gap between the optimal values of Problems (44) and (38) is bounded from above by $\sqrt{P}-1$, i.e., $0 \le UB_{\mathrm{C}}^* - Z_{\mathrm{C}}^*(m) \le (\sqrt{P}-1)Z_{\mathrm{C}}^*(m)$.*

Proof. See Appendix B.4 for the detailed proof. □

Remark 2. The theoretical upper bound in Proposition 10 is achievable. Indeed, when we enlarge R_0 enough such that $\mathcal{D}_{\mathrm{C4}}$ degenerates to a moment-based ambiguity set, we can follow the same setting of Example 1 to achieve the theoretical upper bound in Proposition 10.

4. Computational Experiments

We perform extensive computational experiments to demonstrate the effectiveness of our proposed inner and outer approximations in two applications: production-transportation and multi-product newsvendor problems. The mathematical models are implemented in MATLAB R2017a (ver. 9.2) by the modeling language CVX (ver. 2.1) (Grant and Boyd 2008, 2014) with the Mosek solver (8.0.0.60) on a PC with 64-bit Windows Operating System, an Intel(R) Core(TM) i7-7700 CPU @ 3.60 GHz processor, and a 16 GB RAM. The time limit for each run is set at 36 hours. In Section 4.1, we specify the proposed lower and upper bounds in the context of the two aforementioned applications. In Section 4.2, we explain how to randomly generate test instances and report the numerical results together with analyses.

4.1. Computational Setup

In this section, we specify the proposed lower and upper bounds, as well as the theoretical upper bounds for their gaps with the original DRO model, in the context of production-transportation and multi-product newsvendor problems.

4.1.1. Production-Transportation Problem A deterministic production-transportation problem aims to minimize the total production and transportation cost by making production and transportation decisions while satisfying all customer demands. Suppose there are n customers with demand d_j ($\forall j \in [n]$) and m suppliers, each with normalized capacity 1, and $\sum_j d_j \leq m$. We use x_i and z_{ij} to respectively denote the amount of goods produced by supplier i and the amount of goods shipped from supplier i to customer j. Moreover, we use c_i and ξ_{ij} to denote the unit production cost by supplier i and the unit transportation cost to customer j from this supplier, respectively. Thus, this problem can be formulated as follows:

$$\min_{\boldsymbol{x},\boldsymbol{z}} \quad \sum_{i=1}^{m} c_i x_i + \sum_{i=1}^{m}\sum_{j=1}^{n} \xi_{ij} z_{ij} \tag{45a}$$

$$\text{s.t.} \quad \sum_{i=1}^{m} z_{ij} = d_j, \ \forall j \in [n], \tag{45b}$$

$$\sum_{j=1}^{n} z_{ij} = x_i, \ \forall i \in [m], \tag{45c}$$

$$0 \leq x_i \leq 1, \ \forall i \in [m], \tag{45d}$$

$$z_{ij} \geq 0, \ \forall i \in [m], \ \forall j \in [n]. \tag{45e}$$

Now we derive the DRO counterpart of Problem (45). Specifically, we assume that $\boldsymbol{\xi}$ is random and its probability distribution $\mathbb{P}$ is unknown but it belongs to a predefined distributional ambiguity set $\mathcal{D}$. In addition, the decision $\boldsymbol{x}$ is decided before the realization of randomness and $\boldsymbol{z}$ is made as a recourse to specific realizations (Bertsimas et al. 2010). This leads to a two-stage DRO counterpart

$$\min_{\boldsymbol{x}} \left\{ \boldsymbol{c}^\top \boldsymbol{x} + \max_{\mathbb{P}\in\mathcal{D}} \ \mathbb{E}_{\mathbb{P}}\left[\mathcal{U}\left(\mathcal{Q}\left(\boldsymbol{z},\boldsymbol{\xi}\right)\right)\right] : (45\text{d}) \right\}, \tag{46}$$

where $\mathcal{U}(\cdot)$ is a convex nondecreasing disutility function used to incorporate risk considerations into the second-stage cost. In particular, we define

$$\mathcal{U}\left(\mathcal{Q}\left(\boldsymbol{z},\boldsymbol{\xi}\right)\right) = \max_{k\in[K]} \left\{\alpha_k \mathcal{Q}\left(\boldsymbol{z},\boldsymbol{\xi}\right) + \beta_k\right\},$$

where $\mathcal{Q}\left(\boldsymbol{z},\boldsymbol{\xi}\right) = \min_{\boldsymbol{z}} \left\{\boldsymbol{z}^\top \boldsymbol{\xi} : (45\text{b}), (45\text{c}), (45\text{e})\right\}$. We can apply the proposed inner and outer approximations (i.e., Problems (8), (41), (15), (18), and (44)) to approximate Problem (46) in the context of production-transportation problem, with the details provided in Appendix C.

We follow Bertsimas et al. (2010) to randomly generate the locations of m suppliers and n customers from a unit square considering ξ_{ij} as the distance between supplier i and customer j. We estimate the mean, standard deviation, and covariance matrix of $\boldsymbol{\xi}$ by using 10,000 independent samples, generated from independent uniform distributions on intervals $[0.5\xi_{ij}, 1.5\xi_{ij}]$, $\forall i \in [m]$ and $j \in [n]$. We let $\bar{c}$ denote the average transportation cost and generate production cost c_i and demand d_j uniformly on the intervals $[0.5\bar{c}, 1.5\bar{c}]$ and $\left[0.5\frac{m}{n}, \frac{m}{n}\right]$, respectively. We consider disutility function $\mathcal{U}(x) = 0.25\left(e^{2x} - 1\right)$ while approximating it by an equidistant linear approximation with five segments on the interval $[0, 1]$.

4.1.2. Multi-Product Newsvendor Problem Given n products and the demand ξ_i for each $i \in [n]$, a deterministic multi-product newsvendor problem determines a nonnegative ordering amount $\boldsymbol{x} = (x_i, i \in [n])^\top$ to minimize the total loss described as follows:

$$\begin{aligned} f(\boldsymbol{x}, \boldsymbol{\xi}) &= \boldsymbol{c}^\top \boldsymbol{x} - \boldsymbol{v}^\top \min(\boldsymbol{x}, \boldsymbol{\xi}) - \boldsymbol{g}^\top (\boldsymbol{x} - \boldsymbol{\xi})_+ \\ &= (\boldsymbol{c} - \boldsymbol{v})^\top \boldsymbol{x} + (\boldsymbol{v} - \boldsymbol{g})^\top (\boldsymbol{x} - \boldsymbol{\xi})_+ \\ &= \max\left\{ (\boldsymbol{c} - \boldsymbol{v})^\top \boldsymbol{x}, (\boldsymbol{c} - \boldsymbol{g})^\top \boldsymbol{x} + (\boldsymbol{g} - \boldsymbol{v})^\top \boldsymbol{\xi} \right\}, \end{aligned}$$

where $\boldsymbol{c}$ represents the wholesale price, $\boldsymbol{v}$ represents the retail price, $\boldsymbol{g}$ represents the salvage price, and the minimum and nonnegativity operator are applied componentwise. Now we consider that demand $\boldsymbol{\xi}$ is uncertain and its probability distribution belongs to a distributional ambiguity set $\mathcal{D}$. The DRO counterpart of the multi-product newsvendor problem to minimize the expected total loss against the worst-case distribution in $\mathcal{D}$ can be described as follows:

$$\min_{\boldsymbol{x} \geq 0} \max_{\mathbb{P} \in \mathcal{D}} \ \mathbb{E}_{\mathbb{P}} \left[\max\left\{ (\boldsymbol{c} - \boldsymbol{v})^\top \boldsymbol{x}, (\boldsymbol{c} - \boldsymbol{g})^\top \boldsymbol{x} + (\boldsymbol{g} - \boldsymbol{v})^\top \boldsymbol{\xi} \right\} \right]. \tag{47}$$

Note that the procedure of applying the proposed inner and outer approximations and the theoretical bounds to Problem (47) is similar to that for Problem (46) and thus is omitted here.

The mean and standard deviation of $\boldsymbol{\xi}$ are randomly picked from the intervals $[0, 10]$ and $[0, 2]$, respectively. To generate the covariance matrix, first we randomly generate a correlation matrix by the MATLAB function "gallery('randcorr',n)" and then convert it to a covariance matrix. We follow Xu et al. (2018) to set the wholesale, retail, and savage prices as $c_i = 0.1(5 + i - 1)$, $v_i = 0.15(5 + i - 1)$, and $g_i = 0.05(5 + i - 1)$ for any $i \in [n]$, respectively.

4.2. Computational Results

We first evaluate the performance of our proposed lower and upper bounds and then show how they can help construct a tight interval, which includes unknown optimal solutions of large-sized DRO problems that cannot be solved to optimality by existing methods in reasonable time.

4.2.1. Instance Generation and Table Header Description We perform our experiments to solve various instances. First, we consider different levels of problem size, namely small, medium and large, by varying m and n in the production-transportation problem and varying n in the newsvendor problem. Second, we consider different levels of approximation. In particular, for PCA-based lower and upper bounds, we consider different values of $\frac{m_1}{m}$ in $\{100\%, 75\%, 50\%, 25\%, 10\%\}$. For vector splitting based upper bounds, we consider $P \in \{2, 4, 5\}$, by which the random vector $\boldsymbol{\xi}$ is equally split. Third, we consider different supports, i.e., $\mathcal{S} \in \{[-2\boldsymbol{\sigma}, 2\boldsymbol{\sigma}], [-3\boldsymbol{\sigma}, 3\boldsymbol{\sigma}], [-4\boldsymbol{\sigma}, 4\boldsymbol{\sigma}]\}$, where $\boldsymbol{\sigma}$ represents the sample standard deviation of random vector $\boldsymbol{\xi}$. In addition, γ_1 and γ_2 are set as 1 and 2, respectively, for the moment-based ambiguity set, the number of i.i.d data samples $N = 10$ for the combined ambiguity set, the Wasserstain radius R_0 is set as 30 for the production-transportation problem and as 700 for the newsvendor problem. For each combination of the above three variants, we randomly generate five instances and report the average results over them.

In the following Sections 4.2.2 - 4.2.4, we will use tables to report our results and here we describe several table headers that are shared by most of the tables. Column "Size" reports the values of m and n in the production-transportation problem and n in the newsvendor problem, indicating different levels of the problem size. Column "Orig." represents the computational time in seconds required to solve the original DRO problem and column "Time" represents the computational time in seconds required to solve the corresponding inner or outer approximations. Column "Gap" represents the relative gap in percentage between a lower or upper bound and the original optimal value. Here, the relative gap between two values is defined as their difference divided by the maximum. As such relative gaps are theoretically bounded from above, e.g., (9) and (20), we use column "Gap2" in percentage to represent the value of theoretical bound. Whenver needed, we use "LB" and "UB" to denote the lower and upper bounds, respectively. Note that the percentage of $\boldsymbol{\xi}$'s components utilized to construct lower and upper bounds, i.e., $\frac{m_1}{m} \times 100\%$, is represented by $\frac{m_1}{m}(\%)$ and P represents the number of split pieces of $\boldsymbol{\xi}$, with each piece having the same size.

4.2.2. Lower Bound Performance We summarize the lower bounds of the DRO problem with the moment-based ambiguity set on both applications in Tables 1 and 2, while Tables 3 and 4 report the results for the DRO problem with the combined ambiguity set.

From Tables 1 and 2, where the column "Gap2" represents the relative gap induced by the theoretical upper bound in (9), we have the following observations. First, when the number of principal components m_1 increases, both Gap and Gap2 decrease and the computational time increases. When m_1 increases to m, we obtain the lower bound equivalent to the original optimal value but in a large computational time. In practice, thus we can leverage the number of principal components as a tool to trade-off between solution quality and computational time. Second, when the problem size increases, the original problem becomes more difficult to solve, while our approximations

Table 1 Lower bound (8) on the production-transportation problem

$\frac{m_1}{m}$(%)			100%			75%			50%			25%			10%		
Size (m,n)	Support	Orig. (secs)	Time (secs)	Gap (%)	Gap2 (%)	Time (secs)	Gap (%)	Gap2 (%)	Time (secs)	Gap (%)	Gap2 (%)	Time (secs)	Gap (%)	Gap2 (%)	Time (secs)	Gap (%)	Gap2 (%)
(5,20)	$[-2\sigma,2\sigma]$	66.3	66.2	0.00	0.00	21.3	0.95	11.26	3.9	3.42	18.00	1.1	4.73	19.72	0.5	5.07	19.49
	$[-3\sigma,3\sigma]$	73.6	73.8	0.00	0.00	21.8	1.11	10.83	4.7	2.92	16.90	1.2	4.96	19.74	0.5	5.50	20.06
	$[-4\sigma,4\sigma]$	67.3	67.4	0.00	0.00	22.5	1.10	11.95	4.4	3.48	18.89	1.2	5.66	21.26	0.5	5.93	21.21
(4,40)	$[-2\sigma,2\sigma]$	560.9	559.6	0.00	0.00	180.8	0.53	7.48	40.3	1.73	12.52	3.8	3.47	15.05	0.8	3.94	15.38
	$[-3\sigma,3\sigma]$	551.5	549.3	0.00	0.00	181.3	0.58	8.25	39.5	2.28	14.23	4.4	3.74	17.29	0.87	4.63	18.01
	$[-4\sigma,4\sigma]$	542.9	543.2	0.00	0.00	173.9	0.60	7.67	40.2	1.93	12.76	4.4	3.53	15.04	0.9	3.86	15.04
(8,25)	$[-2\sigma,2\sigma]$	1553.1	1553.5	0.00	0.00	392.0	2.55	13.33	68.9	3.50	13.81	7.2	3.71	13.63	1.0	3.83	13.57
	$[-3\sigma,3\sigma]$	1558.3	1553.1	0.00	0.00	440.1	1.71	11.00	76.6	3.09	13.22	7.3	3.60	13.28	0.9	3.80	13.17
	$[-4\sigma,4\sigma]$	1612.2	1610.8	0.00	0.00	465.2	1.40	11.08	92.4	3.07	15.70	7.4	3.96	14.94	0.9	4.16	14.80

Table 2 Lower bound (8) on the newsvendor problem

$\frac{m_1}{m}$(%)			100%			75%			50%			25%			10%		
Size (n)	Support	Orig. (secs)	Time (secs)	Gap (%)	Gap2 (%)	Time (secs)	Gap (%)	Gap2 (%)	Time (secs)	Gap (%)	Gap2 (%)	Time (secs)	Gap (%)	Gap2 (%)	Time (secs)	Gap (%)	Gap2 (%)
100	$[-2\sigma,2\sigma]$	19.0	18.9	0.00	0.00	5.1	0.02	0.65	1.0	0.26	2.37	0.3	1.18	4.82	0.2	2.23	6.29
	$[-3\sigma,3\sigma]$	17.2	17.4	0.00	0.00	4.6	0.03	0.77	1.0	0.24	2.27	0.3	1.05	4.63	0.2	1.98	6.14
	$[-4\sigma,4\sigma]$	17.1	17.2	0.00	0.00	4.4	0.02	0.57	0.9	0.22	2.17	0.3	0.95	4.40	0.2	2.22	6.32
160	$[-2\sigma,2\sigma]$	175.0	175.7	0.00	0.00	42.1	0.02	0.58	6.0	0.21	1.83	0.5	0.90	3.67	0.3	2.00	5.04
	$[-3\sigma,3\sigma]$	171.2	171.4	0.00	0.00	42.2	0.02	0.60	6.1	0.19	1.82	0.5	0.94	3.85	0.2	1.96	5.19
	$[-4\sigma,4\sigma]$	154.1	154.0	0.00	0.00	43.3	0.02	0.57	6.2	0.21	1.82	0.5	0.83	3.53	0.2	1.75	4.81
200	$[-2\sigma,2\sigma]$	518.0	519.0	0.00	0.00	118.5	0.02	0.52	18.3	0.19	1.66	1.0	0.83	3.34	0.3	1.71	4.48
	$[-3\sigma,3\sigma]$	494.8	494.1	0.00	0.00	124.3	0.02	0.51	18.1	0.15	1.52	0.9	0.79	3.37	0.3	1.50	4.40
	$[-4\sigma,4\sigma]$	521.9	522.1	0.00	0.00	120.3	0.02	0.54	19.4	0.15	1.51	0.9	0.80	3.35	0.2	1.82	4.70

reduce the computational time significantly and maintain very high solution quality. Third, when comparing the values of Gap and Gap2, we can observe that the latter is always larger than the former, demonstrating that theoretical bound (9) is valid. Meanwhile, the quality of theoretical bound (9) is sensitive to different problems and datasets, as Gap2 is closer to Gap in the newsvendor problem as compared to in the production-transportation problem. Similarly, the lower bound (8) performs better when applied to the newsvendor application, as the relative gap, i.e., Gap, is generally smaller than that in the production-transportation problem.

Table 3 Lower bound (41) on the production-transportation problem

$\frac{m_1}{m}$ (%)		100%			75%			50%			25%			10%		
Size (m,n)	Orig. (secs)	Time (secs)	Gap (%)	Gap2 (%)	Time (secs)	Gap (%)	Gap2 (%)	Time (secs)	Gap (%)	Gap2 (%)	Time (secs)	Gap (%)	Gap2 (%)	Time (secs)	Gap (%)	Gap2 (%)
(5,20)	704.5	704.4	0.00	0.00	265.3	1.22	32.93	94.1	4.11	57.25	37.2	6.88	63.55	16.4	7.22	62.93
(4,40)	4787.7	4792.9	0.00	0.00	1700.9	1.24	34.02	497.5	2.93	48.95	134.6	5.13	56.38	55.1	5.68	55.96
(8,25)	13503.8	13401.5	0.00	0.00	4132.6	2.21	36.65	1196.8	3.94	42.46	233.2	4.78	44.68	79.0	4.98	42.43

Table 4 Lower bound (41) on the newsvendor problem

$\frac{m_1}{m}$ (%)		100%			75%			50%			25%			10%		
Size (n)	Orig. (secs)	Time (secs)	Gap (%)	Gap2 (%)	Time (secs)	Gap (%)	Gap2 (%)	Time (secs)	Gap (%)	Gap2 (%)	Time (secs)	Gap (%)	Gap2 (%)	Time (secs)	Gap (%)	Gap2 (%)
100	128.9	128.8	0.00	0.00	58.2	0.02	1.93	28.7	0.21	6.69	14.6	1.04	14.52	12.9	2.57	20.96
160	859.3	862.5	0.00	0.00	333.5	0.02	1.82	107.4	0.18	5.64	42.1	0.71	10.87	23.0	1.94	16.52
200	2234.7	2227.4	0.00	0.00	811.8	0.01	1.42	216.8	0.14	4.58	63.0	0.72	10.01	30.4	1.68	14.56

From Tables 3 and 4, where the column "Gap2" represents the relative gap induced by the theoretical upper bound in (42), we have the similar observations as from Tables 1 and 2. In addition, a comparison among Tables 1 - 4 shows that (1) solving DRO problems with the combined ambiguity set and their outer approximations takes more computational time than solving those with the moment-based ambiguity set, and (2) theoretical bound (42) is more conservative than theoretical bound (9) and can be improved in our future studies.

4.2.3. Upper Bound Performance We report performance of the upper bound (15) in Tables 5 - 6 and report that of (18) and (44) in Tables 7 - 10.

Table 5 Upper bound (15) on the production-transportation problem

$\frac{m_1}{m}$(%)			100%			75%			50%			25%			10%		
Size (m,n)	Support	Orig. (secs)	Time (secs)	Gap (%)	Gap2 (%)	Time (secs)	Gap (%)	Gap2 (%)	Time (secs)	Gap (%)	Gap2 (%)	Time (secs)	Gap (%)	Gap2 (%)	Time (secs)	Gap (%)	Gap2 (%)
(5, 20)	$[-2\boldsymbol{\sigma}, 2\boldsymbol{\sigma}]$	60.4	60.7	0.00	0.00	24.4	3.28	8.50	5.7	6.68	14.80	1.8	7.60	16.29	1.1	7.60	16.48
	$[-3\boldsymbol{\sigma}, 3\boldsymbol{\sigma}]$	70.9	71.1	0.00	0.00	24.5	1.25	6.19	5.9	4.98	12.68	1.8	8.91	20.14	1.1	10.60	23.80
	$[-4\boldsymbol{\sigma}, 4\boldsymbol{\sigma}]$	68.9	68.7	0.00	0.00	23.8	3.66	9.35	5.9	10.83	16.60	1.8	16.08	25.03	1.1	17.41	29.25
(4, 40)	$[-2\boldsymbol{\sigma}, 2\boldsymbol{\sigma}]$	524.8	525.3	0.00	0.00	175.2	2.37	6.20	43.4	5.40	11.21	7.6	7.08	14.82	3.0	7.46	13.53
	$[-3\boldsymbol{\sigma}, 3\boldsymbol{\sigma}]$	477.7	477.7	0.00	0.00	189.7	4.37	9.56	42.6	9.17	16.27	8.1	13.34	23.27	3.1	15.79	27.59
	$[-4\boldsymbol{\sigma}, 4\boldsymbol{\sigma}]$	565.0	566.0	0.00	0.00	175.7	7.33	11.74	38.6	14.03	19.69	8.0	18.88	26.56	3.1	20.52	29.76
(8, 25)	$[-2\boldsymbol{\sigma}, 2\boldsymbol{\sigma}]$	1470.2	1468.7	0.00	0.00	479.6	0.98	5.40	100.1	1.77	9.34	15.3	2.06	9.86	4.1	2.11	10.78
	$[-3\boldsymbol{\sigma}, 3\boldsymbol{\sigma}]$	1634.5	1632.6	0.00	0.00	491.6	2.06	7.09	104.8	5.57	16.16	15.7	6.73	19.33	4.2	7.14	20.66
	$[-4\boldsymbol{\sigma}, 4\boldsymbol{\sigma}]$	1562.0	1563.3	0.00	0.00	483.5	6.03	11.82	104.8	10.93	21.38	16.4	12.61	23.84	4.1	13.53	26.24

Table 6 Upper bound (15) on the newsvendor problem

$\frac{m_1}{m}$(%)			100%			75%			50%			25%			10%		
Size (n)	Support	Orig. (secs)	Time (secs)	Gap (%)	Gap2 (%)	Time (secs)	Gap (%)	Gap2 (%)	Time (secs)	Gap (%)	Gap2 (%)	Time (secs)	Gap (%)	Gap2 (%)	Time (secs)	Gap (%)	Gap2 (%)
100	$[-2\boldsymbol{\sigma}, 2\boldsymbol{\sigma}]$	18.8	19.0	0.00	0.00	4.7	5.04	12.12	1.0	19.35	25.88	0.3	40.63	46.47	0.3	53.46	58.93
	$[-3\boldsymbol{\sigma}, 3\boldsymbol{\sigma}]$	17.0	17.0	0.00	0.00	4.7	10.48	20.37	1.0	39.22	48.44	0.3	77.8	86.09	0.3	97.94	105.47
	$[-4\boldsymbol{\sigma}, 4\boldsymbol{\sigma}]$	18.6	18.7	0.00	0.00	4.9	19.29	35.54	1.0	70.79	86.16	0.3	158.74	172.17	0.3	213.95	226.36
160	$[-2\boldsymbol{\sigma}, 2\boldsymbol{\sigma}]$	181.8	181.5	0.00	0.00	44.2	4.45	10.43	6.3	17.53	23.16	0.7	35.25	40.34	0.3	48.71	53.31
	$[-3\boldsymbol{\sigma}, 3\boldsymbol{\sigma}]$	174.3	175.0	0.00	0.00	41.6	9.73	19.07	6.0	37.45	46.33	0.6	81.59	89.55	0.3	110.04	117.66
	$[-4\boldsymbol{\sigma}, 4\boldsymbol{\sigma}]$	149.2	149.0	0.00	0.00	43.8	17.85	31.48	5.7	73.73	86.37	0.7	156.36	167.72	0.3	212.74	222.98
200	$[-2\boldsymbol{\sigma}, 2\boldsymbol{\sigma}]$	501.3	501.2	0.00	0.00	138.4	4.65	10.21	20.5	18.95	24.17	1.4	38.26	42.89	0.5	51.69	56.0
	$[-3\boldsymbol{\sigma}, 3\boldsymbol{\sigma}]$	495.3	495.4	0.00	0.00	126.7	9.04	16.53	19.3	35.06	42.09	1.4	73.70	79.95	0.5	102.89	108.66
	$[-4\boldsymbol{\sigma}, 4\boldsymbol{\sigma}]$	537.5	536.7	0.00	0.00	145.7	18.39	30.61	17.0	78.17	89.52	1.3	165.07	175.67	0.5	222.91	232.47

From Tables 5 - 6, where the column "Gap2" represents the relative gap induced by the theoretical upper bound in (20), we have similar observations as from Tables 1 and 2. In addition, our approximation (15) performs better when solving the production-transportation problem as compared to solving the newsvendor problem. More importantly, theoretical bound (20) is not conservative at all because Gap2 is very close to Gap, which is the relative gap obtained numerically.

From Tables 7 - 10, we have the following observations. First, when P increases, the computational time decreases and the Gap increases. In practice, thus we can leverage the number of split pieces as a tool to balance between solution quality and computational time. Second, the quality of upper bound (18) is sensitive to different problems and datasets because it performs better (i.e., with smaller Gap) on the production-transportation problem than on the newsvendor problem. In addition, by comparing Tables 5 - 6 and Tables 7 - 10, we can observe that the vector splitting based upper bounds are much tighter than the PCA based upper bounds.

Table 7 Upper bound (18) on the production-transportation problem

P			2		4		5	
Size (m,n)	Support	Orig. (secs)	Time (secs)	Gap (%)	Time (secs)	Gap (%)	Time (secs)	Gap (%)
(5, 20)	$[-2\boldsymbol{\sigma}, 2\boldsymbol{\sigma}]$	65.9	11.3	0.18	5.4	0.52	4.4	0.71
	$[-3\boldsymbol{\sigma}, 3\boldsymbol{\sigma}]$	69.0	10.7	0.07	5.7	0.35	4.5	0.54
	$[-4\boldsymbol{\sigma}, 4\boldsymbol{\sigma}]$	67.1	11.4	0.10	5.4	0.39	4.3	0.51
(4, 40)	$[-2\boldsymbol{\sigma}, 2\boldsymbol{\sigma}]$	521.9	79.4	0.15	32.8	0.40	22.9	0.53
	$[-3\boldsymbol{\sigma}, 3\boldsymbol{\sigma}]$	566.2	86.8	0.13	35.5	0.43	24.7	0.54
	$[-4\boldsymbol{\sigma}, 4\boldsymbol{\sigma}]$	542.9	83.0	0.17	33.9	0.47	25.1	0.57
(8, 25)	$[-2\boldsymbol{\sigma}, 2\boldsymbol{\sigma}]$	1531.3	205.2	0.02	55.8	0.07	61.5	0.11
	$[-3\boldsymbol{\sigma}, 3\boldsymbol{\sigma}]$	1594.0	216.8	0.04	58.4	0.13	67.8	0.18
	$[-4\boldsymbol{\sigma}, 4\boldsymbol{\sigma}]$	1539.1	207.0	0.04	56.8	0.16	67.5	0.24

Table 8 Upper bound (18) on the newsvendor problem

P			2		4		5	
Size (n)	Support	Orig. (secs)	Time (secs)	Gap (%)	Time (secs)	Gap (%)	Time (secs)	Gap (%)
100	$[-2\boldsymbol{\sigma}, 2\boldsymbol{\sigma}]$	20.1	1.3	1.11	0.5	3.05	0.5	3.86
	$[-3\boldsymbol{\sigma}, 3\boldsymbol{\sigma}]$	17.7	1.3	1.26	0.5	3.20	0.5	4.07
	$[-4\boldsymbol{\sigma}, 4\boldsymbol{\sigma}]$	18.1	1.3	1.11	0.5	2.99	0.5	3.73
160	$[-2\boldsymbol{\sigma}, 2\boldsymbol{\sigma}]$	162.3	7.3	0.91	1.3	2.52	0.9	3.25
	$[-3\boldsymbol{\sigma}, 3\boldsymbol{\sigma}]$	169.2	7.4	1.10	1.2	2.88	0.9	3.57
	$[-4\boldsymbol{\sigma}, 4\boldsymbol{\sigma}]$	169.6	7.9	0.95	1.2	2.62	0.9	3.34
200	$[-2\boldsymbol{\sigma}, 2\boldsymbol{\sigma}]$	521.6	25.7	0.84	2.6	2.13	1.7	2.62
	$[-3\boldsymbol{\sigma}, 3\boldsymbol{\sigma}]$	493.4	21.7	0.84	2.9	2.28	1.7	2.86
	$[-4\boldsymbol{\sigma}, 4\boldsymbol{\sigma}]$	518.1	23.9	0.84	2.8	2.18	1.7	2.81

Table 9 Upper bound (44) on the production-transportation problem

P		2		4		5	
Size (m,n)	Orig. (secs)	Time (secs)	Gap (%)	Time (secs)	Gap (%)	Time (secs)	Gap (%)
(5, 20)	716.4	179.3	1.78	121.4	4.90	114.2	6.26
(4, 40)	5050.3	1028.3	2.06	521.8	4.56	488.4	5.50
(8, 25)	11383.6	2281.3	0.93	1017.6	3.25	1010.7	3.96

Table 10 Upper bound (44) on the newsvendor problem

P		2		4		5	
Size (m,n)	Orig. (secs)	Time (secs)	Gap (%)	Time (secs)	Gap (%)	Time (secs)	Gap (%)
(5, 20)	129.5	55.8	1.66	37.1	4.35	34.1	5.46
(4, 40)	798.6	275.7	1.13	132.5	3.18	120.5	4.17
(8, 25)	2038.4	615.7	0.93	251.4	2.70	225.3	3.49

4.2.4. Interval Performance In many real-world applications with large-scale models and high-dimensional uncertainties (e.g., energy and transportation), we may not be able to solve a DRO model to optimality by existing methods in reasonable time. In this case, it can be very useful to quickly find a feasible solution with small optimality gap, evaluated through a tight interval that includes the (unknown) optimal value of this model. In this section, we construct such tight intervals with the help of our proposed inner and outer approximations.

Table 11 Intervals on the production-transportation problem

[LB, UB]				[(8),(15)]		[(8),(18)]	
Size (m,n)	Support	$(\frac{m_1}{m}\%, P)$	Orig. (secs)	Itv-Time (secs)	Itv-Gap (%)	Itv-Time (secs)	Itv-Gap (%)
(6, 50)	$[-2\boldsymbol{\sigma}, 2\boldsymbol{\sigma}]$	(25%, 5)	-	100.0	7.98	203.5	3.53
		(50%, 4)	-	1017.9	5.74	730.5	2.32
		(75%, 2)	-	5626.6	2.56	3706.4	0.93
	$[-3\boldsymbol{\sigma}, 3\boldsymbol{\sigma}]$	(25%, 5)	-	106.2	13.60	219.6	3.31
		(50%, 4)	-	1071.3	10.75	813.3	2.42
		(75%, 2)	-	5955.1	6.39	3913.1	1.23
	$[-4\boldsymbol{\sigma}, 4\boldsymbol{\sigma}]$	(25%, 5)	-	102.7	20.37	233.2	3.54
		(50%, 4)	-	1093.9	16.40	780.7	2.87
		(75%, 2)	-	5564.6	6.82	4124.3	1.09
(8, 50)	$[-2\boldsymbol{\sigma}, 2\boldsymbol{\sigma}]$	(25%, 5)	-	265.7	4.94	604.7	2.78
		(50%, 4)	-	3375.2	4.53	2181.1	2.47
		(75%, 2)	-	-	-	-	-
	$[-3\boldsymbol{\sigma}, 3\boldsymbol{\sigma}]$	(25%, 5)	-	302.6	10.28	629.1	2.80
		(50%, 4)	-	3697.8	8.21	2667.7	2.21
		(75%, 2)	-	-	-	-	-
	$[-4\boldsymbol{\sigma}, 4\boldsymbol{\sigma}]$	(25%, 5)	-	299.4	15.86	650	2.86
		(50%, 4)	-	4129.7	14.08	2596.4	2.41
		(75%, 2)	-	-	-	-	-

The interval results of the DRO problem with the moment-based ambiguity set are summarized in Tables 11 and 12, while Tables 13 and 14 report the results for the DRO problem with the

Table 12 Intervals on the newsvendor problem

[LB, UB]				[(8), (15)]		[(8), (18)]	
Size (n)	Support	$(\frac{m_1}{m}\%, P)$	Orig. (secs)	Itv-Time (secs)	Itv-Gap (%)	Itv-Time (secs)	Itv-Gap (%)
300	$[-2\boldsymbol{\sigma}, 2\boldsymbol{\sigma}]$	$(25\%, 5)$	-	10.3	33.06	11.1	2.99
		$(50\%, 4)$	-	227.0	14.31	124.5	1.98
		$(75\%, 2)$	-	1703.2	3.24	982.4	0.71
	$[-3\boldsymbol{\sigma}, 3\boldsymbol{\sigma}]$	$(25\%, 5)$	-	11.1	51.19	13.3	2.62
		$(50\%, 4)$	-	244.7	17.90	144.2	1.76
		$(75\%, 2)$	-	1693.3	3.79	782.1	0.60
	$[-4\boldsymbol{\sigma}, 4\boldsymbol{\sigma}]$	$(25\%, 5)$	-	11.1	84.47	11.9	2.72
		$(50\%, 4)$	-	251.9	26.62	126.4	1.85
		$(75\%, 2)$	-	1814.3	7.00	950.8	0.63
400	$[-2\boldsymbol{\sigma}, 2\boldsymbol{\sigma}]$	$(25\%, 5)$	-	36	28.82	35.7	2.38
		$(50\%, 4)$	-	958.2	15.86	423.5	1.82
		$(75\%, 2)$	-	-	-	-	-
	$[-3\boldsymbol{\sigma}, 3\boldsymbol{\sigma}]$	$(25\%, 5)$	-	34.7	58.55	35.4	2.48
		$(50\%, 4)$	-	1093.7	20.06	525.8	1.74
		$(75\%, 2)$	-	-	-	-	-
	$[-4\boldsymbol{\sigma}, 4\boldsymbol{\sigma}]$	$(25\%, 5)$	-	38.5	80.60	39.1	2.38
		$(50\%, 4)$	-	990.7	23.66	430.4	1.55
		$(75\%, 2)$	-	-	-	-	-

combined ambiguity set. The first row of each table, indicated by [LB,UB], shows that each interval is constructed by which lower and upper bounds. Column "Itv-Time" reports the time spent to construct each interval, which equals to the summation of the computational times needed to find the lower and upper bounds. Column "Itv-Gap", calculated by $\frac{\text{UB}-\text{LB}}{\text{UB}} \times 100\%$, demonstrates how tight the interval [LB, UB] is. Symbol "-" indicates that no optimal solution of the original DRO problem or its approximations can be found within the time limit. In Column "$\big(\frac{m_1}{m}\%, P\big)$", $\frac{m_1}{m}\%$ is used to define the lower bound problem and the PCA based upper bound problems, i.e., Problems (8), (15), and (41), while P is used to define the vector splitting based upper bound problems, i.e., Problems (18) and (44).

Table 13 Intervals on the production-transportation problem

[LB,UB]			[(41), (44)]	
Size (m, n)	$(\frac{m_1}{m}\%, P)$	Orig. (secs)	Itv-Time (secs)	Itv-Gap (%)
$(8, 30)$	$(25\%, 5)$	-	2144.6	7.08
	$(50\%, 4)$	-	4204.3	6.32
	$(75\%, 2)$	-	14043.1	3.14
$(8, 40)$	$(25\%, 5)$	-	5153.1	7.38
	$(50\%, 4)$	-	10290.3	6.98
	$(75\%, 2)$	-	40784.1	2.73

Table 14 Intervals on the newsvendor problem

[LB, UB]			[(41), (44)]	
Size (n)	$(\frac{m_1}{m}\%, P)$	Orig. (secs)	Itv-Time (secs)	Itv-Gap (%)
240	$(25\%, 5)$	-	503.7	3.34
	$(50\%, 4)$	-	891.1	2.44
	$(75\%, 2)$	-	2853.4	0.88
320	$(25\%, 5)$	-	1018.8	3.36
	$(50\%, 4)$	-	2275.7	2.28
	$(75\%, 2)$	-	7946.2	0.77
400	$(25\%, 5)$	-	3078.8	3.24
	$(50\%, 4)$	-	7780.3	2.19
	$(75\%, 2)$	-	-	-

From Tables 11 and 14, we have the following observations. First, when the optimal solution cannot be found, a tight interval that includes the unknown optimal value can be constructed fast by using our proposed inner and outer approximations. Second, when m_1 increases and P decreases, a tighter interval can be constructed but it costs more computational time. In practice, thus we can leverage the number of principal components and split pieces as a tool to balance between the interval tightness and computational time. Third, from Tables 11 and 12, we can observe that the vector splitting based upper bounds are much tighter than the PCA based upper bounds and take less computational time.

5. Conclusions

In this paper, we proposed computationally efficient inner and outer approximations for DRO problems with two types of ambiguity sets: the moment-based ambiguity set and combined ambiguity set. We approximated the original DRO problems mainly through two approaches: (1) use PCA to shrink the dimensionality of the uncertainty space, and (2) split the random parameter vector into smaller pieces, both of which lead to smaller PSD matrix constraints. Furthermore, we derived theoretical bounds on the gap between the optimal values of DRO problems and their approximations. Such bounds help determine the required numbers of split pieces and principal components to reach a predetermined error bound. Our proposed approximations enable decision-makers to better balance the trade-off between solution quality and computational time by leveraging the appropriate numbers of split pieces and principal components. Meanwhile, they help construct very tight intervals, which contain unknown optimal solutions of the DRO problems that cannot be solved to optimality by existing methods. Finally, we demonstrated the significant effectiveness of the proposed approximations in solving the distributionally robust production-transportation and multi-product newsvendor problems. The results showed that our approximations significantly reduce the computational time while maintaining high solution quality, with the strengths of our derived theoretical bounds well justified.

References

Ben-Tal, A., Den Hertog, D., De Waegenaere, A., Melenberg, B., and Rennen, G. (2013). Robust solutions of optimization problems affected by uncertain probabilities. *Management Sci.*, 59(2):341–357.

Ben-Tal, A. and Nemirovski, A. (1998). Robust convex optimization. *Math. Oper. Res.*, 23(4):769–805.

Bertsimas, D., Doan, X. V., Natarajan, K., and Teo, C.-P. (2010). Models for minimax stochastic linear optimization problems with risk aversion. *Math. Oper. Res.*, 35(3):580–602.

Bertsimas, D., Gupta, V., and Kallus, N. (2018a). Data-driven robust optimization. *Math. Programming*, 167(2):235–292.

Bertsimas, D., Gupta, V., and Kallus, N. (2018b). Robust sample average approximation. *Math. Programming*, 171(1-2):217–282.

Bertsimas, D. and Sim, M. (2004). The price of robustness. *Oper. Res.*, 52(1):35–53.

Chen, X., Lin, Q., and Xu, G. (2019). Distributionally robust optimization with confidence bands for probability density functions. *arXiv preprint arXiv:1901.02169*.

Cheng, J., Chen, R. L.-Y., Najm, H. N., Pinar, A., Safta, C., and Watson, J.-P. (2018). Distributionally robust optimization with principal component analysis. *SIAM J. Optim.*, 28(2):1817–1841.

Cheng, J., Delage, E., and Lisser, A. (2014). Distributionally robust stochastic knapsack problem. *SIAM J. Optim.*, 24(3):1485–1506.

Cheng, J., Leung, J., and Lisser, A. (2016). New reformulations of distributionally robust shortest path problem. *Comput. & Oper. Res.*, 74:196–204.

Delage, E. and Ye, Y. (2010). Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Oper. Res.*, 58(3):595–612.

El Ghaoui, L. E., Oks, M., and Oustry, F. (2003). Worst-case value-at-risk and robust portfolio optimization: A conic programming approach. *Oper. Res.*, 51(4):543–556.

Erdoğan, E. and Iyengar, G. (2006). Ambiguous chance constrained problems and robust optimization. *Math. Programming*, 107(1-2):37–61.

Esfahani, P. M. and Kuhn, D. (2018). Data-driven distributionally robust optimization using the wasserstein metric: Performance guarantees and tractable reformulations. *Math. Programming*, 171(1-2):115–166.

Gao, R. and Kleywegt, A. J. (2017). Distributionally robust stochastic optimization with dependence structure. *arXiv preprint arXiv:1701.04200*.

Gotoh, J.-y., Kim, M. J., and Lim, A. E. (2018). Robust empirical optimization is almost the same as mean–variance optimization. *Oper. Res. Lett.*, 46(4):448–452.

Grant, M. and Boyd, S. (2014). CVX: Matlab software for disciplined convex programming, version 2.1.

Grant, M. C. and Boyd, S. P. (2008). Graph implementations for nonsmooth convex programs. In *Recent Advances in Learning and Control*, pages 95–110. Springer.

Hanasusanto, G. A. and Kuhn, D. (2018). Conic programming reformulations of two-stage distributionally robust linear programs over wasserstein balls. *Oper. Res.*, 66(3):849–869.

Hu, Z. and Hong, L. J. (2013). Kullback-Leibler divergence constrained distributionally robust optimization. *Available at Optimization Online*.

Li, B., Jiang, R., and Mathieu, J. L. (2019). Ambiguous risk constraints with moment and unimodality information. *Math. Programming*, 173(1-2):151–192.

Li, Z., Wu, W., Zhang, B., and Tai, X. (2018). Kullback–Leibler divergence-based distributionally robust optimisation model for heat pump day-ahead operational schedule to improve PV integration. *IET Gener., Trans. & Distrib.*, 12(13):3136–3144.

Liu, Y., Yuan, X., and Zhang, J. (2019). Discrete approximation scheme in distributionally robust optimization. http://www.optimization-online.org/DB_FILE/2019/01/7040.pdf.

Mieth, R. and Dvorkin, Y. (2018). Data-driven distributionally robust optimal power flow for distribution systems. *IEEE Control Syst. Lett.*, 2(3):363–368.

Natarajan, K., Sim, M., and Uichanco, J. (2010). Tractable robust expected utility and risk models for portfolio optimization. *Math. Finance*, 20(4):695–731.

Niu, S., Song, S., Ding, J.-Y., Zhang, Y., and Chiong, R. (2019). Distributionally robust single machine scheduling with the total tardiness criterion. *Comput. & Oper. Res.*, 101:13–28.

Rahimian, H. and Mehrotra, S. (2019). Distributionally robust optimization: A review. *arXiv preprint arXiv:1908.05659*.

Scarf, H. (1958). A min-max solution of an inventory problem. *Studies in the Math. Theory of Inventory and Production*.

Shapiro, A. (2001). On duality theory of conic linear problems. In *Semi-infinite programming*, pages 135–165. Springer.

Shapiro, A., Dentcheva, D., and Ruszczyński, A. (2009). *Lectures on Stochastic Programming: Modeling and Theory*. SIAM.

Van der Vaart, A. W. (2000). *Asymptotic Statistics*, volume 3. Cambridge university press.

Vandenberghe, L. and Boyd, S. (1996). Semidefinite programming. *SIAM Rev.*, 38(1):49–95.

Wang, C., Gao, R., Qiu, F., Wang, J., and Xin, L. (2018). Risk-based distributionally robust optimal power flow with dynamic line rating. *IEEE Trans. on Power Systems*, 33(6):6074–6086.

Wang, Z., Glynn, P. W., and Ye, Y. (2016). Likelihood robust optimization for data-driven problems. *Comput. Management Sci.*, 13(2):241–261.

Wold, S., Esbensen, K., and Geladi, P. (1987). Principal Component Analysis. *Chemom. Intell. Lab. Syst.*, 2(1-3):37–52.

Xu, H., Liu, Y., and Sun, H. (2018). Distributionally robust optimization with matrix moment constraints: Lagrange duality and cutting plane methods. *Math. Programming*, 169(2):489–529.

Yang, Y. and Wu, W. (2019). A distributionally robust optimization model for real-time power dispatch in distribution networks. *IEEE Trans. on Smart Grid*, 10(4):3743–3752.

Zhang, Y., Shen, S., and Erdogan, S. A. (2018). Solving 0–1 semidefinite programs for distributionally robust allocation of surgery blocks. *Optim. Lett.*, 12(7):1503–1521.

Appendix A: Supplement to Section 2

A.1. Proof of Proposition 1

We apply the strong duality theorem to constraints (3b). As function $f(\boldsymbol{x},\boldsymbol{\xi})$ is piecewise linear convex, we reformulate constraints (3b) as:

$$s \geq y_k^0(\boldsymbol{x}) + y_k(\boldsymbol{x})^\top \boldsymbol{\xi} - \boldsymbol{\xi}^\top \boldsymbol{q} - \boldsymbol{\xi}^\top \boldsymbol{Q}\boldsymbol{\xi}, \quad \forall \boldsymbol{\xi} \in \mathcal{S}, \forall k \in [K], \tag{EC.1}$$

which are equivalent to $\min_{\boldsymbol{A\xi}\leq \boldsymbol{b},\boldsymbol{\xi}\in\mathbb{R}^m} g_k(\boldsymbol{\xi}) \geq 0$, where $g_k(\boldsymbol{\xi}) = s + \boldsymbol{\xi}^\top \boldsymbol{q} + \boldsymbol{\xi}^\top \boldsymbol{Q}\boldsymbol{\xi} - y_k^0(\boldsymbol{x}) - y_k(\boldsymbol{x})^\top \boldsymbol{\xi}$, for any $k \in [K]$. Moreover, we consider the Lagrange dual problem of $\min_{\boldsymbol{A\xi}\leq \boldsymbol{b},\boldsymbol{\xi}\in\mathbb{R}^m} g_k(\boldsymbol{\xi})$, i.e., $\max_{\boldsymbol{\lambda}_k\geq 0}\inf_{\boldsymbol{\xi}} g_k(\boldsymbol{\xi}) + \boldsymbol{\lambda}_k^\top(\boldsymbol{A\xi}-\boldsymbol{b})$, where $\boldsymbol{\lambda}_k \in \mathbb{R}^n$. Note that function $g_k(\boldsymbol{\xi})$ is convex in $\boldsymbol{\xi}$ because $\boldsymbol{Q} \succeq 0$. Due to Assumption 1, there exists an interior point for the primal problem. It follows that constraints (EC.1) are equivalent to the following ones:

$$\max_{\boldsymbol{\lambda}_k\geq 0}\inf_{\boldsymbol{\xi}} \; g_k(\boldsymbol{\xi}) + \boldsymbol{\lambda}_k^\top(\boldsymbol{A\xi}-\boldsymbol{b}) \geq 0, \quad \forall k \in [K],$$

which are further equivalent to the following constraints:

$$\exists \boldsymbol{\lambda}_k \geq 0,\; s + \boldsymbol{\xi}^\top \boldsymbol{q} + \boldsymbol{\xi}^\top \boldsymbol{Q}\boldsymbol{\xi} - y_k^0(\boldsymbol{x}) - y_k(\boldsymbol{x})^\top \boldsymbol{\xi} + \boldsymbol{\lambda}_k^\top(\boldsymbol{A\xi}-\boldsymbol{b}) \geq 0, \quad \forall \boldsymbol{\xi} \in \mathbb{R}^m, \forall k \in [K]. \tag{EC.2}$$

As $\boldsymbol{\xi} = \boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}} + \boldsymbol{\mu}$, we replace $\boldsymbol{\xi}$ with $\boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}} + \boldsymbol{\mu}$ in (EC.2). Thus, we have

$$\text{(EC.2)} \Leftrightarrow \exists \boldsymbol{\lambda}_k \geq 0,\; \left(1, \boldsymbol{\xi}_{\mathrm{I}}^\top\right) \mathbf{Z}_k \left(1, \boldsymbol{\xi}_{\mathrm{I}}^\top\right)^\top \geq 0,\; \forall \boldsymbol{\xi}_{\mathrm{I}} \in \mathbb{R}^m,\; \forall k \in [K], \tag{EC.3}$$

$$\Leftrightarrow \exists \boldsymbol{\lambda}_k \geq 0,\; \mathbf{Z}_k \succeq 0,\; \forall k \in [K], \tag{EC.4}$$

where

$$\mathbf{Z}_k = \begin{bmatrix} s - y_k^0(\boldsymbol{x}) - \boldsymbol{\lambda}_k^\top \boldsymbol{b} - y_k(\boldsymbol{x})^\top \boldsymbol{\mu} + \boldsymbol{\lambda}_k^\top \boldsymbol{A}\boldsymbol{\mu} & \frac{1}{2}\left(\boldsymbol{q} + \left(\boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top \left(\boldsymbol{A}^\top \boldsymbol{\lambda}_k - y_k(\boldsymbol{x})\right)\right)^\top \\ \frac{1}{2}\left(\boldsymbol{q} + \left(\boldsymbol{U}\Lambda^{\frac{1}{2}}\right)^\top \left(\boldsymbol{A}^\top \boldsymbol{\lambda}_k - y_k(\boldsymbol{x})\right)\right) & \boldsymbol{Q} \end{bmatrix},$$

and the first equivalence holds due to the definition of $\mathbf{Z}_k$. For the second equivalence, clearly $\Leftarrow$ follows from the definition of a PSD matrix. To prove $\Rightarrow$, we consider two possible cases for any $(\eta_0 \in \mathbb{R}, \boldsymbol{\eta}^\top \in \mathbb{R}^m)^\top \in \mathbb{R}^{m+1}$: (1) if $\eta_0 = 0$, then $(\eta_0, \boldsymbol{\eta}^\top)\mathbf{Z}_k(\eta_0, \boldsymbol{\eta}^\top)^\top = \boldsymbol{\eta}^\top \boldsymbol{Q}\boldsymbol{\eta} \geq 0$ because $\boldsymbol{Q}$ is PSD; (2) if $\eta_0 \neq 0$, then we have $(\eta_0, \boldsymbol{\eta}^\top)\mathbf{Z}_k(\eta_0, \boldsymbol{\eta}^\top)^\top = \eta_0^2 \left(1, \frac{\boldsymbol{\eta}^\top}{\eta_0}\right) \mathbf{Z}_k \left(1, \frac{\boldsymbol{\eta}^\top}{\eta_0}\right)^\top \geq 0$ according to (EC.3). Therefore, $\Rightarrow$ holds and we obtain Problem (4) by replacing constraints (3b) with (EC.4). □

A.2. Proof of Theorem 2

The proof of the deterministic reformulation (7) is the same as that of Theorem 1 and thus is omitted here. With (7), we define $\boldsymbol{\zeta} = \boldsymbol{U}_{m\times m_1}\boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{r}} + \boldsymbol{\mu}$ and use $\mathcal{S}_\zeta$ and $\mathcal{D}_\zeta$ to denote its support and ambiguity set, respectively. As $\mathcal{S}_{\mathrm{r}} = \{\boldsymbol{\xi}_{\mathrm{r}} \in \mathbb{R}^{m_1} : \boldsymbol{U}_{m\times m_1}\Lambda_{m_1}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{r}} + \boldsymbol{\mu} \in \mathcal{S}\}$ and $\mathcal{S}_\zeta = \{\boldsymbol{\zeta} \in \mathbb{R}^m : \boldsymbol{\zeta} = \boldsymbol{U}_{m\times m_1}\boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{r}} + \boldsymbol{\mu}, \boldsymbol{\xi}_{\mathrm{r}} \in \mathcal{S}_{\mathrm{r}}\}$, we can deduce $\mathcal{S}_\zeta \subset \mathcal{S}$. We also have $\mathbb{E}_{\mathbb{P}_\zeta}[\boldsymbol{\zeta}] = \boldsymbol{\mu}$ and

$$\begin{aligned} \mathbb{E}_{\mathbb{P}_\zeta}\left[(\boldsymbol{\zeta}-\boldsymbol{\mu})(\boldsymbol{\zeta}-\boldsymbol{\mu})^\top\right] &\preceq \gamma_2 \boldsymbol{U}_{m\times m_1}\boldsymbol{\Lambda}_{m_1}\boldsymbol{U}_{m\times m_1}^\top \\ &= \gamma_2 \boldsymbol{U}\begin{bmatrix} \boldsymbol{\Lambda}_{m_1} & \mathbf{0}_{m_1\times(m-m_1)} \\ \mathbf{0}_{(m-m_1)\times m_1} & \mathbf{0}_{(m-m_1)\times(m-m_1)} \end{bmatrix}\boldsymbol{U}^\top \preceq \gamma_2 \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^\top = \gamma_2 \boldsymbol{\Sigma}. \end{aligned}$$

It follows that $\mathcal{D}_\zeta$ lies in $\mathcal{D}_{\mathrm{M1}}$, i.e., $\mathcal{D}_\zeta \subset \mathcal{D}_{\mathrm{M1}}$, and accordingly

$$\max_{\mathbb{P}_\zeta \in \mathcal{D}_\zeta} \mathbb{E}_{\mathbb{P}_\zeta}\left[f(\boldsymbol{x},\boldsymbol{\zeta})\right] \leq \max_{\mathbb{P}\in\mathcal{D}_{\mathrm{M1}}} \mathbb{E}_{\mathbb{P}}\left[f(\boldsymbol{x},\boldsymbol{\xi})\right].$$

Therefore, the optimal value of Problem (7) (i.e., Problem (6)) is a lower bound for that of Problem (3) (i.e., Problem (2)).

To show the monotonicity result, we define $\boldsymbol{\zeta}_i = \boldsymbol{U}_{m\times m_i}\boldsymbol{\Lambda}_{m_i}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{r}_i} + \boldsymbol{\mu}$ for any $i\in[2]$, where $\boldsymbol{\xi}_{\mathrm{r}_i}\in\mathbb{R}^{m_i}$ for $m_2 > m_1$. The ambiguity set of $\boldsymbol{\zeta}_i$ is denoted by $\mathcal{D}_{\zeta_i}$, i.e.,

$$\mathcal{D}_{\zeta_i} = \left\{\mathbb{P}_{\zeta_i} | \boldsymbol{\zeta}_i \sim \mathbb{P}_{\zeta_i}, \boldsymbol{\zeta}_i = \boldsymbol{U}_{m\times m_i}\boldsymbol{\Lambda}_{m_i}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{r}_i} + \boldsymbol{\mu}, \boldsymbol{\xi}_{\mathrm{r}_i} \sim \mathbb{P}_{\mathrm{r}_i} \in \mathcal{D}_{\mathrm{r}_i}\right\}, \forall i \in [2],$$

where $\mathcal{D}_{\mathrm{r}_i}$ (defined as (6b)) represents the ambiguity set of $\boldsymbol{\xi}_{\mathrm{r}_i}$ for any $i\in[2]$. For any $\boldsymbol{\zeta}_1 \sim \mathbb{P}_{\zeta_1} \in \mathcal{D}_{\zeta_1}$, there exists a $\boldsymbol{\xi}_{\mathrm{r}_1} \sim \mathbb{P}_{\mathrm{r}_1} \in \mathcal{D}_{\mathrm{r}_1}$ such that $\boldsymbol{\zeta}_1 = \boldsymbol{U}_{m\times m_1}\boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{r}_1} + \boldsymbol{\mu} = \boldsymbol{U}_{m\times m_2}\boldsymbol{\Lambda}_{m_2}^{\frac{1}{2}}\bar{\boldsymbol{\xi}}_{\mathrm{r}_2} + \boldsymbol{\mu}$, where $\bar{\boldsymbol{\xi}}_{\mathrm{r}_2} = \left(\boldsymbol{\xi}_{\mathrm{r}_1}^\top, \mathbf{0}_{m_2-m_1}^\top\right)^\top \in \mathbb{R}^{m_2}$. By using $\mathcal{S}_{\mathrm{r}_i}$ (defined as (6c)) to denote the support of $\boldsymbol{\xi}_{\mathrm{r}_i}$ for any $i\in[2]$, we have

$$\mathbb{P}\left\{\boldsymbol{\xi}_{\mathrm{r}_1} \in \mathcal{S}_{\mathrm{r}_1}\right\} = \mathbb{P}\left\{\boldsymbol{U}_{m\times m_1}\boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{r}_1} + \boldsymbol{\mu} \in \mathcal{S}\right\} = 1,$$

which is equivalent to $\mathbb{P}\{\boldsymbol{U}_{m\times m_2}\boldsymbol{\Lambda}_{m_2}^{\frac{1}{2}}\bar{\boldsymbol{\xi}}_{\mathrm{r}_2} + \boldsymbol{\mu} \in \mathcal{S}\} = 1$ and implies that $\mathbb{P}\{\bar{\boldsymbol{\xi}}_{\mathrm{r}_2} \in \mathcal{S}_{\mathrm{r}_2}\} = 1$ because $\boldsymbol{U}_{m\times m_1}\boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{r}_1} = \boldsymbol{U}_{m\times m_2}\boldsymbol{\Lambda}_{m_2}^{\frac{1}{2}}\bar{\boldsymbol{\xi}}_{\mathrm{r}_2}$. In addition, we have $\mathbb{E}[\bar{\boldsymbol{\xi}}_{\mathrm{r}_2}] = \mathbf{0}_{m_2}$ and

$$\mathbb{E}\left[\bar{\boldsymbol{\xi}}_{\mathrm{r}_2}\bar{\boldsymbol{\xi}}_{\mathrm{r}_2}^\top\right] = \begin{bmatrix} \mathbb{E}\left[\boldsymbol{\xi}_{\mathrm{r}_1}\boldsymbol{\xi}_{\mathrm{r}_1}^\top\right] & \mathbf{0}_{m_1\times(m_2-m_1)} \\ \mathbf{0}_{(m_2-m_1)\times m_1} & \mathbf{0}_{(m_2-m_1)\times(m_2-m_1)} \end{bmatrix} \preceq \gamma_2 \boldsymbol{I}_{m_2}.$$

It follows that the distribution of $\bar{\boldsymbol{\xi}}_{\mathrm{r}_2}$ belongs to $\mathcal{D}_{\mathrm{r}_2}$ and thus $\mathbb{P}_{\zeta_1} \in \mathcal{D}_{\zeta_2}$. Therefore, we have $\mathcal{D}_{\zeta_1} \subset \mathcal{D}_{\zeta_2}$ and

$$\max_{\mathbb{P}_{\zeta_1}\in\mathcal{D}_{\zeta_1}} \mathbb{E}_{\mathbb{P}_{\zeta_1}}\left[f\left(\boldsymbol{x},\boldsymbol{\zeta}_1\right)\right] \leq \max_{\mathbb{P}_{\zeta_2}\in\mathcal{D}_{\zeta_2}} \mathbb{E}_{\mathbb{P}_{\zeta_2}} f\left[(\boldsymbol{x},\boldsymbol{\zeta}_2)\right].$$

That is, the optimal value of Problem (7) is nondecreasing in m_1.

Finally, Problem (6) is equivalent to Problem (2) when $m_1 = m$. Then, Problem (7) results in an exact reformulation of Problem (2) by Theorem 1. □

A.3. Proof of Proposition 4

We apply the strong duality theorem to constraints (14b). As function $f(\boldsymbol{x},\boldsymbol{\xi})$ is piecewise linear convex, $\boldsymbol{\xi} = \boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}} + \boldsymbol{\mu}$, and $\boldsymbol{\xi}_{\mathrm{I}} = \left(\boldsymbol{\xi}_{\mathrm{r}}^\top \in \mathbb{R}^{m_1}, \boldsymbol{\xi}_{\mathrm{r}2}^\top \in \mathbb{R}^{m-m_1}\right)^\top$, we reformulate (14b) as

$$s \geq y_k^0(\boldsymbol{x}) + y_k(\boldsymbol{x})^\top \left(\boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\left(\boldsymbol{\xi}_{\mathrm{r}}^\top, \boldsymbol{\xi}_{\mathrm{r}2}^\top\right)^\top + \boldsymbol{\mu}\right) - \boldsymbol{\xi}_{\mathrm{r}}^\top\boldsymbol{Q}_{\mathrm{r}}\boldsymbol{\xi}_{\mathrm{r}} - \boldsymbol{q}^\top\left(\boldsymbol{\xi}_{\mathrm{r}}^\top, \boldsymbol{\xi}_{\mathrm{r}2}^\top\right)^\top, \quad \forall \boldsymbol{\xi}_{\mathrm{I}} \in \mathcal{S}_{\mathrm{I}},\ \forall k \in [K], \tag{EC.5}$$

which are equivalent to $\min_{\boldsymbol{A}(\boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}}+\boldsymbol{\mu})\leq\boldsymbol{b},\boldsymbol{\xi}_{\mathrm{I}}\in\mathbb{R}^m} g_k(\boldsymbol{\xi}_{\mathrm{I}}) \geq 0$, where $g_k(\boldsymbol{\xi}_{\mathrm{I}}) = s + \boldsymbol{q}^\top(\boldsymbol{\xi}_{\mathrm{r}}^\top, \boldsymbol{\xi}_{\mathrm{r}2}^\top)^\top + \boldsymbol{\xi}_{\mathrm{r}}^\top\boldsymbol{Q}_{\mathrm{r}}\boldsymbol{\xi}_{\mathrm{r}} - y_k^0(\boldsymbol{x}) - y_k(\boldsymbol{x})^\top(\boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}}(\boldsymbol{\xi}_{\mathrm{r}}^\top, \boldsymbol{\xi}_{\mathrm{r}2}^\top)^\top + \boldsymbol{\mu})$, for any $k\in[K]$. Moreover, we consider the Lagrange dual problem of $\min_{\boldsymbol{A}(\boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}}+\boldsymbol{\mu})\leq\boldsymbol{b},\boldsymbol{\xi}_{\mathrm{I}}\in\mathbb{R}^m} g_k(\boldsymbol{\xi}_{\mathrm{I}})$, i.e., $\max_{\boldsymbol{\lambda}_k\geq 0}\inf_{\boldsymbol{\xi}_{\mathrm{I}}} g_k(\boldsymbol{\xi}_{\mathrm{I}}) + \boldsymbol{\lambda}_k^\top(\boldsymbol{A}(\boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}} + \boldsymbol{\mu}) - \boldsymbol{b})$, where $\boldsymbol{\lambda}_k \in \mathbb{R}^n$. Note that function $g_k(\boldsymbol{\xi}_{\mathrm{I}})$ is convex in $\boldsymbol{\xi}_{\mathrm{I}}$ because it is a quadratic function that can be written as the general form $f(\boldsymbol{x}) = \boldsymbol{x}^\top\boldsymbol{M}\boldsymbol{x} + \boldsymbol{x}^\top\boldsymbol{b} + c$ where $\boldsymbol{M}$ is PSD, i.e., $\boldsymbol{Q}\succeq 0$. Due to Assumption 1, there exists an interior point for the primal problem. It follows that constraints (EC.5) are equivalent to the following ones:

$$\max_{\boldsymbol{\lambda}_k\geq 0}\ \inf_{\boldsymbol{\xi}_{\mathrm{I}}}\ g_k(\boldsymbol{\xi}_{\mathrm{I}}) + \boldsymbol{\lambda}_k^\top\left(\boldsymbol{A}\left(\boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}} + \boldsymbol{\mu}\right) - \boldsymbol{b}\right) \geq 0, \quad \forall k \in [K],$$

which are further equivalent to the following constraints:

$$\begin{aligned}&\exists \boldsymbol{\lambda}_k \geq 0, s+\boldsymbol{q}^\top\left(\boldsymbol{\xi}_{\mathrm{r}}^\top,\boldsymbol{\xi}_{\mathrm{r}2}^\top\right)^\top+\boldsymbol{\xi}_{\mathrm{r}}^\top\boldsymbol{Q}_{\mathrm{r}}\boldsymbol{\xi}_{\mathrm{r}}-y_k^0(\boldsymbol{x})-y_k(\boldsymbol{x})^\top\left(\boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\left(\boldsymbol{\xi}_{\mathrm{r}}^\top,\boldsymbol{\xi}_{\mathrm{r}2}^\top\right)^\top+\boldsymbol{\mu}\right)\\&\quad+\boldsymbol{\lambda}_k^\top\left(\boldsymbol{A}\left(\boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\left(\boldsymbol{\xi}_{\mathrm{r}}^\top,\boldsymbol{\xi}_{\mathrm{r}2}^\top\right)^\top+\boldsymbol{\mu}\right)-\boldsymbol{b}\right)\geq 0,\quad \forall \boldsymbol{\xi}_{\mathrm{I}}\in\mathbb{R}^m,\forall k\in[K].\end{aligned}\tag{EC.6}$$

Then, we perform the following decomposition:

$$\boldsymbol{U}\Lambda^{\frac{1}{2}}\left(\boldsymbol{\xi}_{\mathrm{r}}^\top,\boldsymbol{\xi}_{\mathrm{r}2}^\top\right)^\top+\boldsymbol{\mu}=\boldsymbol{U}_{m\times m_1}\boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{r}}+\boldsymbol{U}_{m\times(m-m_1)}\boldsymbol{\Lambda}_{m-m_1}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{r}2}+\boldsymbol{\mu},\tag{EC.7}$$

where $\boldsymbol{U}_{m\times m_1}\in\mathbb{R}^{m\times m_1}$ and $\boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}}\in\mathbb{R}^{m_1\times m_1}$ are upper-left submatrices of $\boldsymbol{U}$ and $\boldsymbol{\Lambda}^{\frac{1}{2}}$, respectively, and $\boldsymbol{\Lambda}_{m-m_1}^{\frac{1}{2}}\in\mathbb{R}^{(m-m_1)\times(m-m_1)}$ and $\boldsymbol{U}_{m\times(m-m_1)}\in\mathbb{R}^{m\times(m-m_1)}$ are their lower-right submatrices, respectively. By plugging (EC.7) to (EC.6) and defining $\boldsymbol{q}=(\boldsymbol{q}_1^\top\in\mathbb{R}^{m_1},\boldsymbol{q}_2^\top\in\mathbb{R}^{m-m_1})^\top$, we have

$$\begin{aligned}\text{(EC.6)}\Leftrightarrow&\exists\boldsymbol{\lambda}_k\geq 0,\ s-y_k^0(\boldsymbol{x})-\boldsymbol{\lambda}_k^\top\boldsymbol{b}-y_k(\boldsymbol{x})^\top\boldsymbol{\mu}+\boldsymbol{\lambda}_k^\top\boldsymbol{A}\boldsymbol{\mu}+\left(\boldsymbol{q}_1+\left(\boldsymbol{U}_{m\times m_1}\boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}}\right)^\top\left(\boldsymbol{A}^\top\boldsymbol{\lambda}_k-y_k(\boldsymbol{x})\right)\right)^\top\boldsymbol{\xi}_{\mathrm{r}}\\&+\left(\boldsymbol{q}_2+\left(\boldsymbol{U}_{m\times(m-m_1)}\boldsymbol{\Lambda}_{m-m_1}^{\frac{1}{2}}\right)^\top\left(\boldsymbol{A}^\top\boldsymbol{\lambda}_k-y_k(\boldsymbol{x})\right)\right)^\top\boldsymbol{\xi}_{\mathrm{r}2}+\boldsymbol{\xi}_{\mathrm{r}}^\top\boldsymbol{Q}_{\mathrm{r}}\boldsymbol{\xi}_{\mathrm{r}}\geq 0,\ \forall\boldsymbol{\xi}_{\mathrm{I}}\in\mathbb{R}^m,\ \forall k\in[K],\end{aligned}$$

$$\Leftrightarrow\left(1,\boldsymbol{\xi}_{\mathrm{r}}^\top\right)\mathbf{Z}_k\left(1,\boldsymbol{\xi}_{\mathrm{r}}^\top\right)^\top+\boldsymbol{W}_k^\top\boldsymbol{\xi}_{\mathrm{r}2}\geq 0,\ \forall\boldsymbol{\xi}_{\mathrm{I}}\in\mathbb{R}^m,\ \forall k\in[K],\tag{EC.8}$$

where

$$\mathbf{Z}_k=\begin{bmatrix}s-y_k^0(\boldsymbol{x})-\boldsymbol{\lambda}_k^\top\boldsymbol{b}-y_k(\boldsymbol{x})^\top\boldsymbol{\mu}+\boldsymbol{\lambda}_k^\top\boldsymbol{A}\boldsymbol{\mu} & \frac{1}{2}\left(\boldsymbol{q}_1+\left(\boldsymbol{U}_{m\times m_1}\boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}}\right)^\top\left(\boldsymbol{A}^\top\boldsymbol{\lambda}_k-y_k(\boldsymbol{x})\right)\right)^\top\\ \frac{1}{2}\left(\boldsymbol{q}_1+\left(\boldsymbol{U}_{m\times m_1}\boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}}\right)^\top\left(\boldsymbol{A}^\top\boldsymbol{\lambda}_k-y_k(\boldsymbol{x})\right)\right) & \boldsymbol{Q}_{\mathrm{r}}\end{bmatrix}$$

and

$$\boldsymbol{W}_k=\left(\boldsymbol{q}_2+\left(\boldsymbol{U}_{m\times(m-m_1)}\boldsymbol{\Lambda}_{m-m_1}^{\frac{1}{2}}\right)^\top\left(\boldsymbol{A}^\top\boldsymbol{\lambda}_k-y_k(\boldsymbol{x})\right)\right).$$

Since $\boldsymbol{W}_k^\top\boldsymbol{\xi}_{\mathrm{r}2}$ in (EC.8) is affine and $\boldsymbol{\xi}_{\mathrm{r}2}\in\mathbb{R}^{m-m_1}$, we set $\boldsymbol{W}_k=\mathbf{0}$ for any $k\in[K]$, which prevents the objective value of the Lagrange dual problem from going to infinity and accordingly leads to constraints (15c). Thus, we have

$$\text{(EC.6)}\Leftrightarrow\exists\boldsymbol{\lambda}_k\geq 0,\ \left(1,\boldsymbol{\xi}_{\mathrm{r}}^\top\right)\mathbf{Z}_k\left(1,\boldsymbol{\xi}_{\mathrm{r}}^\top\right)^\top\geq 0,\ \forall\boldsymbol{\xi}_{\mathrm{r}}\in\mathbb{R}^{m_1},\ \forall k\in[K];\ \text{(15c)},\tag{EC.9}$$

$$\Leftrightarrow\exists\boldsymbol{\lambda}_k\geq 0,\ \mathbf{Z}_k\succeq 0,\ \forall k\in[K];\ \text{(15c)}.\tag{EC.10}$$

The first equivalence holds due to the definition of $\boldsymbol{Z}_k$. For the second equivalence, clearly $\Leftarrow$ follows from the definition of a PSD matrix. To prove $\Rightarrow$, we consider two possible cases for any $(\eta_0\in\mathbb{R},\boldsymbol{\eta}^\top\in\mathbb{R}^{m_1})^\top\in\mathbb{R}^{m_1+1}$: (1) if $\eta_0=0$, then $(\eta_0,\boldsymbol{\eta}^\top)\mathbf{Z}_k(\eta_0,\boldsymbol{\eta}^\top)^\top=\boldsymbol{\eta}^\top\boldsymbol{Q}_{\mathrm{r}}\boldsymbol{\eta}\geq 0$ because $\boldsymbol{Q}_{\mathrm{r}}$ is PSD; (2) if $\eta_0\neq 0$, then we have $(\eta_0,\boldsymbol{\eta}^\top)\mathbf{Z}_k(\eta_0,\boldsymbol{\eta}^\top)^\top=\eta_0^2\left(1,\frac{\boldsymbol{\eta}^\top}{\eta_0}\right)\mathbf{Z}_k\left(1,\frac{\boldsymbol{\eta}^\top}{\eta_0}\right)^\top\geq 0$ according to (EC.9). Therefore, $\Rightarrow$ holds and we obtain Problem (15) by replacing constraints (14b) with (EC.10). □

A.4. Proof of Theorem 4

As $\mathbb{P}_{\mathrm{I}}$ is a probability measure on $(\mathbb{R}^m, \mathbb{B})$, where $\mathbb{B}$ denotes the Borel σ-algebra on $\mathbb{R}^m$, Problem (16) can be described as the following problem:

$$\min_{\boldsymbol{x}\in\mathcal{X}} \max_{\mathbb{P}_{\mathrm{I}}\in\mathcal{D}_{\mathrm{M5}}} \int_{\mathcal{S}_{\mathrm{I}}} f\left(\boldsymbol{x}, \boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}}+\boldsymbol{\mu}\right) \boldsymbol{d}\,\mathbb{P}_{\mathrm{I}}(\boldsymbol{\xi}_{\mathrm{I}}) \tag{EC.11a}$$

$$\text{s.t.} \quad \int_{\mathcal{S}_{\mathrm{I}}} \boldsymbol{d}\,\mathbb{P}_{\mathrm{I}}(\boldsymbol{\xi}_{\mathrm{I}}) = 1, \tag{EC.11b}$$

$$\int_{\mathcal{S}_{\mathrm{I}}} \begin{bmatrix} \boldsymbol{I}_m & \boldsymbol{\xi}_{\mathrm{I}} \\ \boldsymbol{\xi}_{\mathrm{I}}^\top & \gamma_1 \end{bmatrix} \boldsymbol{d}\,\mathbb{P}_{\mathrm{I}}(\boldsymbol{\xi}_{\mathrm{I}}) \succeq 0, \tag{EC.11c}$$

$$\int_{\mathcal{S}_{\mathrm{I}}} \boldsymbol{\xi}_{\mathrm{I}_i}\boldsymbol{\xi}_{\mathrm{I}_i}^\top \boldsymbol{d}\,\mathbb{P}_{\mathrm{I}}(\boldsymbol{\xi}_{\mathrm{I}}) \preceq \gamma_2 \boldsymbol{I}_{m_i}, \ \forall i\in[P], \tag{EC.11d}$$

where (EC.11c) is derived due to Schur's complement. In the following, we first formulate the dual of Problem (EC.11) and then we show that strong duality holds.

Considering s, $\begin{bmatrix} \boldsymbol{W} & \boldsymbol{w} \\ \boldsymbol{w}^\top & r \end{bmatrix} \succeq 0$, and $\boldsymbol{Q}_i \succeq 0$ for any $i\in[P]$ as Lagrangian multipliers of constraints (EC.11b), (EC.11c), and (EC.11d), respectively, we formulate the following problem as the Lagrange dual problem of (EC.11):

$$\begin{aligned} \min_{\boldsymbol{x}\in\mathcal{X}} \max_{\mathbb{P}_{\mathrm{I}}\in\mathcal{D}_{\mathrm{M5}}} \ & s + \boldsymbol{I}_m \bullet \boldsymbol{W} + \gamma_1 r + \gamma_2 \sum_{i=1}^{P} \boldsymbol{I}_{m_i} \bullet \boldsymbol{Q}_i \\ & - \int_{\mathcal{S}_{\mathrm{I}}} \left(s - 2\boldsymbol{w}^\top \boldsymbol{\xi}_{\mathrm{I}} + \sum_{i=1}^{P} \boldsymbol{\xi}_{\mathrm{I}_i}^\top \boldsymbol{Q}_i \boldsymbol{\xi}_{\mathrm{I}_i} - f\left(\boldsymbol{x}, \boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}}+\boldsymbol{\mu}\right)\right) \boldsymbol{d}\,\mathbb{P}_{\mathrm{I}}(\boldsymbol{\xi}_{\mathrm{I}}). \end{aligned}$$

To prevent the objective value of the Lagrange dual problem from going to infinity, we require

$$s - 2\boldsymbol{w}^\top \boldsymbol{\xi}_{\mathrm{I}} + \sum_{i=1}^{P} \boldsymbol{\xi}_{\mathrm{I}_i}^\top \boldsymbol{Q}_i \boldsymbol{\xi}_{\mathrm{I}_i} - f\left(\boldsymbol{x}, \boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}}+\boldsymbol{\mu}\right) \geq 0, \ \forall \boldsymbol{\xi}_{\mathrm{I}} \in \mathcal{S}_{\mathrm{I}}.$$

Accordingly, the dual problem of (EC.11) can be described as follows:

$$\min_{\substack{\boldsymbol{x},s,\boldsymbol{W} \\ \boldsymbol{w},r,\boldsymbol{Q}_i}} \quad s + \boldsymbol{I}_m \bullet \boldsymbol{W} + \gamma_1 r + \gamma_2 \sum_{i=1}^{P} \boldsymbol{I}_{m_i} \bullet \boldsymbol{Q}_i \tag{EC.12a}$$

$$\begin{aligned} \text{s.t.} \quad & s - 2\boldsymbol{w}^\top \boldsymbol{\xi}_{\mathrm{I}} + \sum_{i=1}^{P} \boldsymbol{\xi}_{\mathrm{I}_i}^\top \boldsymbol{Q}_i \boldsymbol{\xi}_{\mathrm{I}_i} - f\left(\boldsymbol{x}, \boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}}+\boldsymbol{\mu}\right) \geq 0, \ \forall \boldsymbol{\xi}_{\mathrm{I}} \in \mathcal{S}_{\mathrm{I}}, \\ & \boldsymbol{x}\in\mathcal{X}, \quad \boldsymbol{Q}_i \succeq 0, \quad \forall i\in[P], \\ & \begin{bmatrix} \boldsymbol{W} & \boldsymbol{w} \\ \boldsymbol{w}^\top & r \end{bmatrix} \succeq 0. \end{aligned} \tag{EC.12b}$$

We further simplify Problem (EC.12) towards eliminating variables $\boldsymbol{W}$ and r. To that end, we keep variables $\boldsymbol{Q}_i$, for any $i\in[P]$, and s fixed while solving Problem (EC.12) analytically for variables $\boldsymbol{W}$, $\boldsymbol{w}$, and r. It follows that we solve $\min_{\boldsymbol{x},\boldsymbol{W},\boldsymbol{w},r} \boldsymbol{I}_m \bullet \boldsymbol{W} + \gamma_1 r$ analytically for $\boldsymbol{W}$, $\boldsymbol{w}$, and r. We consider two cases for the optimal solution of r (denoted by r^*) due to constraint (EC.12b), i.e., $r^* > 0$ and $r^* = 0$, as follows.

- If $r^* > 0$, then constraint (EC.12b) can be reformulated as $\boldsymbol{W} \succeq \frac{\boldsymbol{w}\boldsymbol{w}^\top}{r}$ by Schur's complement. As a result, $\boldsymbol{W}^* = \frac{\boldsymbol{w}\boldsymbol{w}^\top}{r}$ is a valid optimal solution because $\min_{\boldsymbol{x},\boldsymbol{W},\boldsymbol{w},r} \boldsymbol{I}_m \bullet \boldsymbol{W} + \gamma_1 r$ is a minimization problem. Replacing $\boldsymbol{W}^*$ by $\frac{\boldsymbol{w}\boldsymbol{w}^\top}{r}$ leads to solve a one-dimensional convex optimization problem, i.e., $\min_{r>0} \frac{\boldsymbol{w}^\top\boldsymbol{w}}{r} + \gamma_1 r$. By applying the necessary first-order optimality condition to this problem, i.e., setting the derivative of the objective function over r to zero, we have $r^* = \frac{\|\boldsymbol{w}\|_2}{\sqrt{\gamma_1}}$ as the optimal solution of r. If we plug $\boldsymbol{W}^* = \frac{\boldsymbol{w}\boldsymbol{w}^\top}{r}$ and $r^* = \frac{\|\boldsymbol{w}\|_2}{\sqrt{\gamma_1}}$ in (EC.12a), we obtain the following problem:

$$\begin{aligned}
\min_{\boldsymbol{x}, s, \boldsymbol{w}, \boldsymbol{Q}_i} \quad & s + \gamma_2 \sum_{i=1}^{P} \boldsymbol{I}_{m_i} \bullet \boldsymbol{Q}_i + \sqrt{\gamma_1}\,\|2\boldsymbol{w}\|_2 && \text{(EC.13)}\\
\text{s.t.} \quad & s - 2\boldsymbol{w}^\top \boldsymbol{\xi}_{\mathrm{I}} + \sum_{i=1}^{P} \boldsymbol{\xi}_{\mathrm{I}_i}^\top \boldsymbol{Q}_i \boldsymbol{\xi}_{\mathrm{I}_i} - f\left(\boldsymbol{x}, \boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}} + \boldsymbol{\mu}\right) \geq 0, \ \forall \boldsymbol{\xi}_{\mathrm{I}} \in \mathcal{S}_{\mathrm{I}}, \\
& \boldsymbol{x} \in \mathcal{X}, \quad \boldsymbol{Q}_i \succeq 0, \quad \forall i \in [P].
\end{aligned}$$

By introducing a new variable $\boldsymbol{q} = -2\boldsymbol{w}$, we obtain Problem (17).

- If $r^* = 0$, then we let $\boldsymbol{w}^*$ denote the optimal solution of $\boldsymbol{w}$ and we must have $\boldsymbol{w}^* = \boldsymbol{0}$. Otherwise, we have $\boldsymbol{w}^{*\top}\boldsymbol{w}^* > 0$, and by defining $\boldsymbol{Z} = \left(\boldsymbol{w}^{*\top}, \eta\right)^\top$ with $\eta < \dfrac{-\boldsymbol{w}^{*\top}\boldsymbol{W}^*\boldsymbol{w}^*}{2\boldsymbol{w}^{*\top}\boldsymbol{w}^*}$, we further have

$$\boldsymbol{Z}^\top \begin{bmatrix} \boldsymbol{W}^* & \boldsymbol{w}^* \\ \boldsymbol{w}^{*\top} & 0 \end{bmatrix} \boldsymbol{Z} = \boldsymbol{w}^{*\top}\boldsymbol{W}^*\boldsymbol{w}^* + 2\eta \boldsymbol{w}^{*\top}\boldsymbol{w}^* < 0,$$

  which contradicts constraint (EC.12b). Considering $r^* = 0$ and $\boldsymbol{w}^* = \boldsymbol{0}$, $\min_{\boldsymbol{x}, \boldsymbol{W}, \boldsymbol{w}, r} \boldsymbol{I}_m \bullet \boldsymbol{W} + \gamma_1 r$ reduces to $\min_{\boldsymbol{x}, \boldsymbol{W}} \boldsymbol{I}_m \bullet \boldsymbol{W}$ whose optimal solution is clearly $\boldsymbol{W}^* = \boldsymbol{0}$ as it is a minimization problem. Here also by replacing $\boldsymbol{q} = -2\boldsymbol{w}$, we obtain Problem (17).

Note that our conditions on γ_1, γ_2, and $\boldsymbol{I}_{m_i}$ for any $i \in [P]$ are sufficient to ensure that the Dirac measure lies in the relative interior of the feasible set of Problem (16). Therefore, we can conclude that there is no duality gap between Problems (16) and (17) according to the weaker version of Proposition 3.4 in Shapiro (2001).

Finally, to prove Problem (17) provides an upper bound for Problem (3), we can equivalently prove that Problem (16) is an upper bound of Problem (2) since Problems (17) and (3) are equivalent reformulations of Problems (16) and (2), respectively. To that end, we only need to prove $\mathcal{D}_{\mathrm{M2}} \subset \mathcal{D}_{\mathrm{M5}}$, i.e., any distribution in $\mathcal{D}_{\mathrm{M2}}$ also belongs to $\mathcal{D}_{\mathrm{M5}}$. As the first two constraints of $\mathcal{D}_{\mathrm{M2}}$ and $\mathcal{D}_{\mathrm{M5}}$ are the same, any distribution in $\mathcal{D}_{\mathrm{M2}}$ satisfies constraints $\mathbb{P}(\boldsymbol{\xi}_{\mathrm{I}} \in \mathcal{S}_{\mathrm{I}}) = 1$ and $\mathbb{E}_{\mathbb{P}_{\mathrm{I}}}\left[\boldsymbol{\xi}_{\mathrm{I}}^\top\right]\mathbb{E}_{\mathbb{P}_{\mathrm{I}}}\left[\boldsymbol{\xi}_{\mathrm{I}}\right] \leq \gamma_1$ in $\mathcal{D}_{\mathrm{M5}}$. Thus, we only require to show any distribution in $\mathcal{D}_{\mathrm{M2}}$ satisfies constraint $\mathbb{E}_{\mathbb{P}_{\mathrm{I}}}\left[\boldsymbol{\xi}_{\mathrm{I}_i}\boldsymbol{\xi}_{\mathrm{I}_i}^\top\right] \preceq \gamma_2 \boldsymbol{I}_{m_i}, \ \forall i \in [P]$, in $\mathcal{D}_{\mathrm{M5}}$. To that end, we let $\boldsymbol{\xi}_{\mathrm{I}} = \left(\boldsymbol{\xi}_{\mathrm{I}_1}^\top, \boldsymbol{\xi}_{\mathrm{I}_2}^\top, \ldots, \boldsymbol{\xi}_{\mathrm{I}_P}^\top\right)^\top$, $\boldsymbol{\xi}_{\mathrm{I}_i} \in \mathbb{R}^{m_i}$ for any $i \in [P]$, and reformulate the second-order moment constraint of $\mathcal{D}_{\mathrm{M2}}$ as the following equivalent constraint:

$$\mathbb{E}_{\mathbb{P}_{\mathrm{I}}}\begin{bmatrix} \boldsymbol{\xi}_{\mathrm{I}_1}\boldsymbol{\xi}_{\mathrm{I}_1}^\top & \boldsymbol{\xi}_{\mathrm{I}_1}\boldsymbol{\xi}_{\mathrm{I}_2}^\top & \cdots & \boldsymbol{\xi}_{\mathrm{I}_1}\boldsymbol{\xi}_{\mathrm{I}_P}^\top \\ \boldsymbol{\xi}_{\mathrm{I}_2}\boldsymbol{\xi}_{\mathrm{I}_1}^\top & \boldsymbol{\xi}_{\mathrm{I}_2}\boldsymbol{\xi}_{\mathrm{I}_2}^\top & \cdots & \boldsymbol{\xi}_{\mathrm{I}_2}\boldsymbol{\xi}_{\mathrm{I}_P}^\top \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\xi}_{\mathrm{I}_P}\boldsymbol{\xi}_{\mathrm{I}_1}^\top & \boldsymbol{\xi}_{\mathrm{I}_P}\boldsymbol{\xi}_{\mathrm{I}_2}^\top & \cdots & \boldsymbol{\xi}_{\mathrm{I}_P}\boldsymbol{\xi}_{\mathrm{I}_P}^\top \end{bmatrix} \preceq \begin{bmatrix} \gamma_2 \boldsymbol{I}_{\mathrm{m}_1} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \gamma_2 \boldsymbol{I}_{\mathrm{m}_2} & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \gamma_2 \boldsymbol{I}_{\mathrm{m}_P} \end{bmatrix}, \qquad \text{(EC.14)}$$

which implies $\mathbb{E}_{\mathbb{P}_{\mathrm{I}}}\left[\boldsymbol{\xi}_{\mathrm{I}_i}\boldsymbol{\xi}_{\mathrm{I}_i}^\top\right] \preceq \gamma_2 \boldsymbol{I}_{m_i}, \ \forall i \in [P]$ by simply considering the diagonal components of the matrices on both sides of (EC.14). That is, any distribution in $\mathcal{D}_{\mathrm{M2}}$, which satisfies $\mathbb{E}_{\mathbb{P}_{\mathrm{I}}}[\boldsymbol{\xi}_{\mathrm{I}}\boldsymbol{\xi}_{\mathrm{I}}^\top] \preceq \gamma_2 \boldsymbol{I}_m$, also satisfies $\mathbb{E}_{\mathbb{P}_{\mathrm{I}}}\left[\boldsymbol{\xi}_{\mathrm{I}_i}\boldsymbol{\xi}_{\mathrm{I}_i}^\top\right] \preceq \gamma_2 \boldsymbol{I}_{m_i}, \ \forall i \in [P]$ in $\mathcal{D}_{\mathrm{M5}}$ i.e., $\mathcal{D}_{\mathrm{M2}} \subset \mathcal{D}_{\mathrm{M5}}$. □

Appendix B: Supplement to Section 3

B.1. Proof of Theorem 5

By Theorem 1 in Gao and Kleywegt (2017), Problem (40) has the following strong dual problem:

$$\min_{\boldsymbol{x}\in\mathcal{X},\boldsymbol{Q}_{\rm r},\lambda}\left\{\lambda R_0+\gamma_2\boldsymbol{I}_{m_1}\bullet\boldsymbol{Q}_{\rm r}+\int_{\mathbb{R}^{m_1}}\sup_{\boldsymbol{\xi}_{\rm r}} g(\boldsymbol{\xi}_{\rm r},\hat{\boldsymbol{\xi}})\,\mathbb{P}_0(\mathbf{d}\hat{\boldsymbol{\xi}})\right\},\tag{EC.15}$$

where $g(\boldsymbol{\xi}_{\rm r},\hat{\boldsymbol{\xi}})=\max_{k=1}^{K}\{y_k(\boldsymbol{x})^\top(\boldsymbol{U}_{m\times m_1}\boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}}\boldsymbol{\xi}_{\rm r}+\boldsymbol{\mu})+y_k^0(\boldsymbol{x})\}-\boldsymbol{\xi}_{\rm r}^\top\boldsymbol{Q}_{\rm r}\boldsymbol{\xi}_{\rm r}-\lambda\|\boldsymbol{U}_{m\times m_1}\boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}}\boldsymbol{\xi}_{\rm r}+\boldsymbol{\mu}-\hat{\boldsymbol{\xi}}\|_1$, and $\boldsymbol{Q}_{\rm r}$ and λ are the Lagrangian multipliers of the primal second-order moment and Wasserstein constraints, respectively. As $\mathbb{P}_0$ denotes an empirical distribution of $\boldsymbol{\xi}$ generated by i.i.d. samples $\{\hat{\boldsymbol{\xi}}^i:i\in N\}\subseteq\mathcal{S}$ from the $\mathbb{P}$, i.e., $\mathbb{P}\{\boldsymbol{\xi}=\hat{\boldsymbol{\xi}}^i\}=\frac{1}{N}$, we have

$$\int_{\mathbb{R}^{m_1}}\sup_{\boldsymbol{\xi}_{\rm r}}\ g(\boldsymbol{\xi}_{\rm r},\hat{\boldsymbol{\xi}})\,\mathbb{P}_0(\mathbf{d}\hat{\boldsymbol{\xi}})=\frac{1}{N}\sum_{i=1}^{N}\sup_{\boldsymbol{\xi}_{\rm r}}g(\boldsymbol{\xi}_{\rm r},\hat{\boldsymbol{\xi}}^i),\tag{EC.16}$$

because $g(\boldsymbol{\xi}_{\rm r},\hat{\boldsymbol{\xi}})$ is a convex function and $\mathcal{S}_{\rm r}=\mathbb{R}^{m_1}$, which is convex. Thus, by plugging (EC.16) into (EC.15), (EC.15) can be reformulated as

$$\min_{\boldsymbol{x}\in\mathcal{X},\boldsymbol{Q}_{\rm r},\lambda}\left\{\lambda R_0+\gamma_2\boldsymbol{I}_{m_1}\bullet\boldsymbol{Q}_{\rm r}+\frac{1}{N}\sum_{i=1}^{N}y_i\right\}\tag{EC.17a}$$

$$\text{s.t.}\quad y_i=\sup_{\boldsymbol{\xi}_{\rm r}}g(\boldsymbol{\xi}_{\rm r},\hat{\boldsymbol{\xi}}^i),\quad\forall i\in[N].\tag{EC.17b}$$

Since Problem (EC.17) is a minimization problem, constraints (EC.17b) can be relaxed to $y_i\geq\sup_{\boldsymbol{\xi}_{\rm r}}g(\boldsymbol{\xi}_{\rm r},\hat{\boldsymbol{\xi}}^i)$ for any $i\in[N]$. Thus, we have

$$y_i\geq\sup_{\boldsymbol{\xi}_{\rm r}}\left\{\max_{k=1}^{K}\left\{y_k(\boldsymbol{x})^\top\left(\boldsymbol{U}_{m\times m_1}\boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}}\boldsymbol{\xi}_{\rm r}+\boldsymbol{\mu}\right)+y_k^0(\boldsymbol{x})\right\}-\boldsymbol{\xi}_{\rm r}^\top\boldsymbol{Q}_{\rm r}\boldsymbol{\xi}_{\rm r}-\lambda\left\|\boldsymbol{U}_{m\times m_1}\boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}}\boldsymbol{\xi}_{\rm r}+\boldsymbol{\mu}-\hat{\boldsymbol{\xi}}^i\right\|_1\right\},\ \forall i\in[N]$$

$$\Leftrightarrow y_i\geq\max_{k=1}^{K}\ \sup_{\boldsymbol{\xi}_{\rm r}}\left\{y_k(\boldsymbol{x})^\top\left(\boldsymbol{U}_{m\times m_1}\boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}}\boldsymbol{\xi}_{\rm r}+\boldsymbol{\mu}\right)+y_k^0(\boldsymbol{x})-\boldsymbol{\xi}_{\rm r}^\top\boldsymbol{Q}_{\rm r}\boldsymbol{\xi}_{\rm r}-\lambda\left\|\boldsymbol{U}_{m\times m_1}\boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}}\boldsymbol{\xi}_{\rm r}+\boldsymbol{\mu}-\hat{\boldsymbol{\xi}}^i\right\|_1\right\},\ \forall i\in[N]$$

$$\Leftrightarrow y_i\geq\sup_{\boldsymbol{\xi}_{\rm r}}\left\{y_k(\boldsymbol{x})^\top\left(\boldsymbol{U}_{m\times m_1}\boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}}\boldsymbol{\xi}_{\rm r}+\boldsymbol{\mu}\right)+y_k^0(\boldsymbol{x})-\boldsymbol{\xi}_{\rm r}^\top\boldsymbol{Q}_{\rm r}\boldsymbol{\xi}_{\rm r}-\lambda\left\|\boldsymbol{U}_{m\times m_1}\boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}}\boldsymbol{\xi}_{\rm r}+\boldsymbol{\mu}-\hat{\boldsymbol{\xi}}^i\right\|_1\right\},\forall i\in[N],\ \forall k\in[K].$$

For any given $i\in[N]$, we let

$$\left\|\boldsymbol{U}_{m\times m_1}\boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}}\boldsymbol{\xi}_{\rm r}+\boldsymbol{\mu}-\hat{\boldsymbol{\xi}}^i\right\|_1=\sup_{\|\hat{\zeta}\|_*\leq1}\hat{\boldsymbol{\zeta}}^\top\left(\boldsymbol{U}_{m\times m_1}\boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}}\boldsymbol{\xi}_{\rm r}+\boldsymbol{\mu}-\hat{\boldsymbol{\xi}}^i\right),$$

and accordingly we have

$$y_i\geq\sup_{\boldsymbol{\xi}_{\rm r}}\inf_{\|\hat{\zeta}\|_*\leq1}\left\{y_k(\boldsymbol{x})^\top\left(\boldsymbol{U}_{m\times m_1}\boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}}\boldsymbol{\xi}_{\rm r}+\boldsymbol{\mu}\right)+y_k^0(\boldsymbol{x})-\boldsymbol{\xi}_{\rm r}^\top\boldsymbol{Q}_{\rm r}\boldsymbol{\xi}_{\rm r}-\lambda\hat{\boldsymbol{\zeta}}^\top\left(\boldsymbol{U}_{m\times m_1}\boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}}\boldsymbol{\xi}_{\rm r}+\boldsymbol{\mu}-\hat{\boldsymbol{\xi}}^i\right)\right\},\ \forall k\in[K]$$

$$\Leftrightarrow y_i\geq\inf_{\|\hat{\zeta}\|_*\leq1}\sup_{\boldsymbol{\xi}_{\rm r}}\left\{y_k(\boldsymbol{x})^\top\left(\boldsymbol{U}_{m\times m_1}\boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}}\boldsymbol{\xi}_{\rm r}+\boldsymbol{\mu}\right)+y_k^0(\boldsymbol{x})-\boldsymbol{\xi}_{\rm r}^\top\boldsymbol{Q}_{\rm r}\boldsymbol{\xi}_{\rm r}-\lambda\hat{\boldsymbol{\zeta}}^\top\left(\boldsymbol{U}_{m\times m_1}\boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}}\boldsymbol{\xi}_{\rm r}+\boldsymbol{\mu}-\hat{\boldsymbol{\xi}}^i\right)\right\},\ \forall k\in[K]$$

$$\Leftrightarrow\exists\hat{\boldsymbol{\zeta}}\ \text{s.t.}\ \left\|\hat{\boldsymbol{\zeta}}\right\|_*\leq1,\ y_i\geq\sup_{\boldsymbol{\xi}_{\rm r}}\left\{y_k(\boldsymbol{x})^\top\left(\boldsymbol{U}_{m\times m_1}\boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}}\boldsymbol{\xi}_{\rm r}+\boldsymbol{\mu}\right)+y_k^0(\boldsymbol{x})-\boldsymbol{\xi}_{\rm r}^\top\boldsymbol{Q}_{\rm r}\boldsymbol{\xi}_{\rm r}-\lambda\hat{\boldsymbol{\zeta}}^\top\left(\boldsymbol{U}_{m\times m_1}\boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}}\boldsymbol{\xi}_{\rm r}+\boldsymbol{\mu}-\hat{\boldsymbol{\xi}}^i\right)\right\},\ \forall k\in[K]$$

$$\Leftrightarrow\exists\hat{\boldsymbol{\zeta}}\ \text{s.t.}\ \left\|\hat{\boldsymbol{\zeta}}\right\|_*\leq1,\ y_i\geq y_k(\boldsymbol{x})^\top\left(\boldsymbol{U}_{m\times m_1}\boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}}\boldsymbol{\xi}_{\rm r}+\boldsymbol{\mu}\right)+y_k^0(\boldsymbol{x})-\boldsymbol{\xi}_{\rm r}^\top\boldsymbol{Q}_{\rm r}\boldsymbol{\xi}_{\rm r}-\lambda\hat{\boldsymbol{\zeta}}^\top\left(\boldsymbol{U}_{m\times m_1}\boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}}\boldsymbol{\xi}_{\rm r}+\boldsymbol{\mu}-\hat{\boldsymbol{\xi}}^i\right),\ \forall\boldsymbol{\xi}_{\rm r}\in\mathbb{R}^{m_1},\ \forall k\in[K]$$

$$\Leftrightarrow\exists\hat{\boldsymbol{\zeta}}\ \text{s.t.}\ \left\|\hat{\boldsymbol{\zeta}}\right\|_*\leq1,\ \begin{bmatrix}\boldsymbol{Q}_{\rm r} & \frac{1}{2}\left(\left(-y_k(\boldsymbol{x})^\top+\lambda\hat{\boldsymbol{\zeta}}^\top\right)\boldsymbol{U}_{m\times m_1}\boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}}\right)^\top\\ \frac{1}{2}\left(-y_k(\boldsymbol{x})^\top+\lambda\hat{\boldsymbol{\zeta}}^\top\right)\boldsymbol{U}_{m\times m_1}\boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}} & y_i-y_k(\boldsymbol{x})^\top\boldsymbol{\mu}-y_k^0(\boldsymbol{x})+\lambda\hat{\boldsymbol{\zeta}}^\top\left(\boldsymbol{\mu}-\hat{\boldsymbol{\xi}}^i\right)\end{bmatrix}\succeq0,\ \forall k\in[K],\tag{EC.18}$$

where the first equivalence is due to the convexity of $g(\boldsymbol{\xi}_{\mathrm{r}}, \hat{\boldsymbol{\xi}})$, $\mathcal{S}_{\mathrm{r}}$, and the feasible region defined by $\|\hat{\boldsymbol{\zeta}}\|_* \leq 1$. For any given $i \in [N]$, we replace $\lambda\hat{\boldsymbol{\zeta}}$ by $\boldsymbol{\zeta}^i$, and then we can obtain Problem (41) by further replacing (EC.17b) by (EC.18) for any $i \in [N]$.

To prove Problem (41) provides a lower bound for Problem (38), we consider Problem (40) and define $\boldsymbol{\zeta} = \boldsymbol{U}_{m\times m_1}\boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{r}} + \boldsymbol{\mu}$, denoting its support and ambiguity set by $\mathcal{S}_\zeta$ and $\mathcal{D}_\zeta$, respectively. As $\mathcal{S}_{\mathrm{r}} = \{\boldsymbol{\xi}_{\mathrm{r}} \in \mathbb{R}^{m_1} : \boldsymbol{U}_{m\times m_1}\Lambda_{m_1}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{r}} + \boldsymbol{\mu} \in \mathcal{S}\} = \mathbb{R}^{m_1}$ and $\mathcal{S}_\zeta = \{\boldsymbol{\zeta} \in \mathbb{R}^m : \boldsymbol{\zeta} = \boldsymbol{U}_{m\times m_1}\boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{r}} + \boldsymbol{\mu}, \boldsymbol{\xi}_{\mathrm{r}} \in \mathcal{S}_{\mathrm{r}}\} = \mathbb{R}^{m_1}$, we can deduce $\mathcal{S}_\zeta \subset \mathcal{S}$. We also have $\mathbb{E}_{\mathbb{P}_\zeta}[\boldsymbol{\zeta}] = \boldsymbol{\mu}$ and

$$\begin{aligned}\mathbb{E}_{\mathbb{P}_\zeta}\left[(\boldsymbol{\zeta}-\boldsymbol{\mu})(\boldsymbol{\zeta}-\boldsymbol{\mu})^\top\right] &\preceq \gamma_2 \boldsymbol{U}_{m\times m_1}\boldsymbol{\Lambda}_{m_1}\boldsymbol{U}_{m\times m_1}^\top \\ &= \gamma_2 \boldsymbol{U}\begin{bmatrix}\boldsymbol{\Lambda}_{m_1} & \mathbf{0}_{m_1\times(m-m_1)} \\ \mathbf{0}_{(m-m_1)\times m_1} & \mathbf{0}_{(m-m_1)\times(m-m_1)}\end{bmatrix}\boldsymbol{U}^\top \preceq \gamma_2\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^\top = \gamma_2\boldsymbol{\Sigma}.\end{aligned}$$

It follows that $\mathcal{D}_\zeta$ lies in $\mathcal{D}_{\mathrm{C1}}$, i.e., $\mathcal{D}_\zeta \subset \mathcal{D}_{\mathrm{C1}}$, and accordingly

$$\max_{\mathbb{P}_\zeta \in \mathcal{D}_\zeta} \mathbb{E}_{\mathbb{P}_\zeta}\left[f(\boldsymbol{x}, \boldsymbol{\zeta})\right] \leq \max_{\mathbb{P}\in\mathcal{D}_{\mathrm{C1}}} \mathbb{E}_{\mathbb{P}}\left[f(\boldsymbol{x}, \boldsymbol{\xi})\right].$$

Therefore, the optimal value of Problem (40) (i.e., Problem (41)) is a lower bound for that of Problem (DRO-C) with $p = 1$ (i.e., Problem (38)).

To show the monotonicity result, we define $\boldsymbol{\zeta}_i = \boldsymbol{U}_{m\times m_i}\boldsymbol{\Lambda}_{m_i}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{r}_i} + \boldsymbol{\mu}$ for any $i \in [2]$, where $\boldsymbol{\xi}_{\mathrm{r}_i} \in \mathbb{R}^{m_i}$ for $m_2 > m_1$. The ambiguity set of $\boldsymbol{\zeta}_i$ is denoted by $\mathcal{D}_{\zeta_i}$, i.e.,

$$\mathcal{D}_{\zeta_i} = \left\{\mathbb{P}_{\zeta_i} | \boldsymbol{\zeta}_i \sim \mathbb{P}_{\zeta_i}, \boldsymbol{\zeta}_i = \boldsymbol{U}_{m\times m_i}\boldsymbol{\Lambda}_{m_i}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{r}_i} + \boldsymbol{\mu}, \boldsymbol{\xi}_{\mathrm{r}_i} \sim \mathbb{P}_{\mathrm{r}_i} \in \mathcal{D}_{\mathrm{r}_i}\right\}, \forall i \in [2],$$

where $\mathcal{D}_{\mathrm{r}_i}$ (defined as (40b)) represents the ambiguity set of $\boldsymbol{\xi}_{\mathrm{r}_i}$ for any $i \in [2]$. For any $\boldsymbol{\zeta}_1 \sim \mathbb{P}_{\zeta_1} \in \mathcal{D}_{\zeta_1}$, there exists a $\boldsymbol{\xi}_{\mathrm{r}_1} \sim \mathbb{P}_{\mathrm{r}_1} \in \mathcal{D}_{\mathrm{r}_1}$ such that $\boldsymbol{\zeta}_1 = \boldsymbol{U}_{m\times m_1}\boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{r}_1} + \boldsymbol{\mu} = \boldsymbol{U}_{m\times m_2}\boldsymbol{\Lambda}_{m_2}^{\frac{1}{2}}\bar{\boldsymbol{\xi}}_{\mathrm{r}_2} + \boldsymbol{\mu}$, where $\bar{\boldsymbol{\xi}}_{\mathrm{r}_2} = \left(\boldsymbol{\xi}_{\mathrm{r}_1}^\top, \mathbf{0}_{m_2-m_1}^\top\right)^\top \in \mathbb{R}^{m_2}$. By using $\mathcal{S}_{\mathrm{r}_i}$ (defined as (40c)) to denote the support of $\boldsymbol{\xi}_{\mathrm{r}_i}$ for any $i \in [2]$, we have

$$\mathbb{P}\left\{\boldsymbol{\xi}_{\mathrm{r}_1} \in \mathcal{S}_{\mathrm{r}_1}\right\} = \mathbb{P}\left\{\boldsymbol{U}_{m\times m_1}\boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{r}_1} + \boldsymbol{\mu} \in \mathcal{S}\right\} = 1,$$

which is equivalent to $\mathbb{P}\{\boldsymbol{U}_{m\times m_2}\boldsymbol{\Lambda}_{m_2}^{\frac{1}{2}}\bar{\boldsymbol{\xi}}_{\mathrm{r}_2} + \boldsymbol{\mu} \in \mathcal{S}\} = 1$ and implies that $\mathbb{P}\{\bar{\boldsymbol{\xi}}_{\mathrm{r}_2} \in \mathcal{S}_{\mathrm{r}_2}\} = 1$ because $\boldsymbol{U}_{m\times m_1}\boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{r}_1} = \boldsymbol{U}_{m\times m_2}\boldsymbol{\Lambda}_{m_2}^{\frac{1}{2}}\bar{\boldsymbol{\xi}}_{\mathrm{r}_2}$. In addition, we have $\mathbb{E}[\bar{\boldsymbol{\xi}}_{\mathrm{r}_2}] = \mathbf{0}_{m_2}$ and

$$\mathbb{E}\left[\bar{\boldsymbol{\xi}}_{\mathrm{r}_2}\bar{\boldsymbol{\xi}}_{\mathrm{r}_2}^\top\right] = \begin{bmatrix}\mathbb{E}\left[\boldsymbol{\xi}_{\mathrm{r}_1}\boldsymbol{\xi}_{\mathrm{r}_1}^\top\right] & \mathbf{0}_{m_1\times(m_2-m_1)} \\ \mathbf{0}_{(m_2-m_1)\times m_1} & \mathbf{0}_{(m_2-m_1)\times(m_2-m_1)}\end{bmatrix} \preceq \gamma_2\boldsymbol{I}_{m_2}.$$

It follows that the distribution of $\bar{\boldsymbol{\xi}}_{\mathrm{r}_2}$ belongs to $\mathcal{D}_{\mathrm{r}_2}$ and thus $\mathbb{P}_{\zeta_1} \in \mathcal{D}_{\zeta_2}$. Therefore, we have $\mathcal{D}_{\zeta_1} \subset \mathcal{D}_{\zeta_2}$ and

$$\max_{\mathbb{P}_{\zeta_1}\in\mathcal{D}_{\zeta_1}} \mathbb{E}_{\mathbb{P}_{\zeta_1}}\left[f\left(\boldsymbol{x}, \boldsymbol{\zeta}_1\right)\right] \leq \max_{\mathbb{P}_{\zeta_2}\in\mathcal{D}_{\zeta_2}} \mathbb{E}_{\mathbb{P}_{\zeta_2}} f\left[(\boldsymbol{x}, \boldsymbol{\zeta}_2)\right].$$

That is, the optimal value of Problem (40) (i.e., Problem (41)) is nondecreasing in m_1.

Finally, Problem (40) is equivalent to Problem (39) when $m_1 = m$. Therefore, Problem (41) results in an exact reformulation of Problem (38). $\square$

B.2. Proof of Proposition 9

By Theorem 5, we have $Z^*{}_{\mathrm{C}}(m) - Z^*{}_{\mathrm{C}}(m_1) \geq 0$. Moreover, according to this theorem, Problem (38) and the following problem, i.e., Problem (41) with $m_1 = m$, have the same optimal value.

$$\min_{\boldsymbol{x},\lambda,\boldsymbol{Q},\boldsymbol{\zeta}^i,y_i} \ \lambda R_0 + \gamma_2 \boldsymbol{I}_m \bullet \boldsymbol{Q} + \frac{1}{N}\sum_{i=1}^{N} y_i \tag{EC.19a}$$

$$\text{s.t.} \quad \begin{bmatrix} \boldsymbol{Q} & \frac{1}{2}\left(\left(-y_k(\boldsymbol{x}) + \boldsymbol{\zeta}^i\right)^\top \boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top \\ \frac{1}{2}\left(-y_k(\boldsymbol{x}) + \boldsymbol{\zeta}^i\right)^\top \boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}} & y_i - y_k(\boldsymbol{x})^\top \boldsymbol{\mu} - y_k^0(\boldsymbol{x}) + {\boldsymbol{\zeta}^i}^\top\left(\boldsymbol{\mu} - \hat{\boldsymbol{\xi}}^i\right) \end{bmatrix} \succeq 0, \quad \forall i \in [N], \ \forall k \in [K], \tag{EC.19b}$$

$$\lambda \in \mathbb{R}_+, \ \boldsymbol{x} \in \mathcal{X}, \ \left\|\boldsymbol{\zeta}^i\right\|_* \leq \lambda, \ \forall i \in [N], \tag{EC.19c}$$

We use $(\boldsymbol{x}^*, \lambda^*, \boldsymbol{Q}_{\mathrm{r}}^*, {\boldsymbol{\zeta}^i}^* \ \forall i \in [N], y_i^* \ \forall i \in [N])$ to denote an optimal solution of Problem (41). Based on this optimal solution, we construct a feasible solution of Problem (EC.19), represented by $(\bar{\boldsymbol{x}}, \bar{\lambda}, \bar{\boldsymbol{Q}}, \bar{\boldsymbol{\zeta}}^i \ \forall i \in [N], \bar{y}_i \ \forall i \in [N])$. For clarity, we define

$$S^{ik} = y_i^* - y_k(\boldsymbol{x}^*)^\top \boldsymbol{\mu} - y_k^0(\boldsymbol{x}^*) + {\boldsymbol{\zeta}^i}^{*\top}\left(\boldsymbol{\mu} - \hat{\boldsymbol{\xi}}^i\right), \ \forall i \in [N], \ \forall k \in [K], \text{ and}$$

$$\boldsymbol{L}_c^{ik} = \left(-y_k(\boldsymbol{x}^*) + {\boldsymbol{\zeta}^i}^*\right)^\top \boldsymbol{U}_{m\times c}\left(\boldsymbol{\Lambda}^c\right)^{\frac{1}{2}}, \ \forall i \in [N], \ \forall k \in [K], \ \forall c \in \{m_1, m - m_1, m\},$$

where $\boldsymbol{\Lambda}^{m_1} \in \mathbb{R}^{m_1 \times m_1}$ and $\boldsymbol{\Lambda}^{m-m_1} \in \mathbb{R}^{(m-m_1)\times(m-m_1)}$ represent the upper-left and lower-right submatrices of $\boldsymbol{\Lambda}$, respectively.

First, we let $\bar{\boldsymbol{x}} = \boldsymbol{x}^*$, $\bar{\lambda} = \lambda^*$, $\bar{\boldsymbol{\zeta}}^i = {\boldsymbol{\zeta}^i}^*$ for any $i \in [N]$,

$$\bar{\boldsymbol{Q}} = \begin{bmatrix} \boldsymbol{Q}_{\mathrm{r}}^* & \mathbf{0}_{m_1\times(m-m_1)} \\ \mathbf{0}_{(m-m_1)\times m_1} & \sum_{i=1}^{N}\sum_{k=1}^{K}\frac{s_1^{ik}}{4}\left(\boldsymbol{L}_{m-m_1}^{ik}\right)^\top \boldsymbol{L}_{m-m_1}^{ik} \end{bmatrix}, \text{ and} \tag{EC.20}$$

$$\bar{y}_i = y_i^* + \sum_{k=1}^{K} s_2^{ik}, \ \forall i \in [N],$$

where $s_1^{ik} > 0$ and $s_2^{ik} > 0$ for any $i \in [N]$ and $k \in [K]$. As $\bar{\lambda} = \lambda^* \geq 0$, $\bar{\boldsymbol{x}} = \boldsymbol{x}^* \in \mathcal{X}$, and $\|\bar{\boldsymbol{\zeta}}^i\|_* = \|{\boldsymbol{\zeta}^i}^*\|_* \leq \lambda$, $\forall i \in [N]$ due to constraint (41c), we only require $(\bar{\boldsymbol{x}}, \bar{\lambda}, \bar{\boldsymbol{Q}}, \bar{\boldsymbol{\zeta}}^i \ \forall i \in [N], \bar{y}_i \ \forall i \in [N])$ to satisfy (EC.19b). Thus, we will find the values of s_1^{ik} and s_2^{ik} for any $i \in [N]$ and $k \in [K]$ that enable this solution to satisfy (EC.19b).

We plug $(\bar{\boldsymbol{x}}, \bar{\lambda}, \bar{\boldsymbol{Q}}, \bar{\boldsymbol{\zeta}}^i \ \forall i \in [N], \bar{y}_i \ \forall i \in [N])$ to (EC.19b) and use $\bar{\boldsymbol{Y}}^{ik}$ for any $i \in [N]$ and $k \in [K]$ to denote the corresponding matrix in (EC.19b). For any given $i \in [N]$ and $k \in [K]$, we perform the following decomposition:

$$\begin{aligned} \bar{\boldsymbol{Y}}^{ik} &= \begin{bmatrix} \boldsymbol{Q}_{\mathrm{r}}^* & \mathbf{0}_{m_1\times(m-m_1)} & \frac{1}{2}\left(\boldsymbol{L}_{m_1}^{ik}\right)^\top \\ \mathbf{0}_{(m-m_1)\times m_1} & \mathbf{0}_{(m-m_1)\times(m-m_1)} & \mathbf{0}_{(m-m_1)\times 1} \\ \frac{1}{2}\boldsymbol{L}_{m_1}^{ik} & \mathbf{0}_{1\times(m-m_1)} & S^{ik} \end{bmatrix} + \begin{bmatrix} \mathbf{0}_{m_1\times m_1} & \mathbf{0}_{m_1\times(m-m_1)} & \mathbf{0}_{m_1\times 1} \\ \mathbf{0}_{(m-m_1)\times m_1} & \sum_{i=1}^{N}\sum_{k=1}^{K}\frac{s_1^{ik}}{4}\left(\boldsymbol{L}_{m-m_1}^{ik}\right)^\top \boldsymbol{L}_{m-m_1}^{ik} & \frac{1}{2}\left(\boldsymbol{L}_{m-m_1}^{ik}\right)^\top \\ \mathbf{0}_{1\times m_1} & \frac{1}{2}\boldsymbol{L}_{m-m_1}^{ik} & \sum_{k=1}^{K} s_2^{ik} \end{bmatrix} \\ &\succeq \begin{bmatrix} \boldsymbol{Q}_{\mathrm{r}}^* & \mathbf{0}_{m_1\times(m-m_1)} & \frac{1}{2}\left(\boldsymbol{L}_{m_1}^{ik}\right)^\top \\ \mathbf{0}_{(m-m_1)\times m_1} & \mathbf{0}_{(m-m_1)\times(m-m_1)} & \mathbf{0}_{(m-m_1)\times 1} \\ \frac{1}{2}\boldsymbol{L}_{m_1}^{ik} & \mathbf{0}_{1\times(m-m_1)} & S^{ik} \end{bmatrix} + \begin{bmatrix} \mathbf{0}_{m_1\times m_1} & \mathbf{0}_{m_1\times(m-m_1)} & \mathbf{0}_{m_1\times 1} \\ \mathbf{0}_{(m-m_1)\times m_1} & \frac{s_1^{ik}}{4}\left(\boldsymbol{L}_{m-m_1}^{ik}\right)^\top \boldsymbol{L}_{m-m_1}^{ik} & \frac{1}{2}\left(\boldsymbol{L}_{m-m_1}^{ik}\right)^\top \\ \mathbf{0}_{1\times m_1} & \frac{1}{2}\boldsymbol{L}_{m-m_1}^{ik} & s_2^{ik} \end{bmatrix}. \end{aligned} \tag{EC.21}$$

The first matrix in (EC.21) is clearly PSD because the elimination of its zero components leads to a PSD matrix due to constraints (41b). Now we find the values of s_1^{ik} and s_2^{ik} to make the second matrix PSD as well, and then accordingly the constructed solution is feasible for (EC.19).

Next, we use $\begin{bmatrix} \bar{\boldsymbol{A}} & \bar{\boldsymbol{B}} \\ \bar{\boldsymbol{B}}^\top & \bar{\boldsymbol{C}} \end{bmatrix}$ to denote the second matrix in (EC.21) by letting $\bar{\boldsymbol{A}} = \begin{bmatrix} \mathbf{0}_{m_1\times m_1} & \mathbf{0}_{m_1\times(m-m_1)} \\ \mathbf{0}_{(m-m_1)\times m_1} & \frac{s_1^{ik}}{4}\left(\boldsymbol{L}_{m-m_1}^{ik}\right)^\top \boldsymbol{L}_{m-m_1}^{ik} \end{bmatrix}$, $\bar{\boldsymbol{B}}^\top = \begin{pmatrix} \mathbf{0}_{1\times m_1} & \frac{1}{2}\boldsymbol{L}_{m-m_1}^{ik} \end{pmatrix}$, and $\bar{\boldsymbol{C}} = s_2^{ik}$. It follows that

$$\begin{aligned}\bar{\boldsymbol{A}} - \bar{\boldsymbol{B}}\bar{\boldsymbol{C}}^{-1}\bar{\boldsymbol{B}}^\top &= \begin{bmatrix} \mathbf{0}_{m_1\times m_1} & \mathbf{0}_{m_1\times(m-m_1)} \\ \mathbf{0}_{(m-m_1)\times m_1} & \frac{s_1^{ik}}{4}\left(\boldsymbol{L}_{m-m_1}^{ik}\right)^\top \boldsymbol{L}_{m-m_1}^{ik} \end{bmatrix} - \frac{1}{s_2^{ik}}\begin{pmatrix} \mathbf{0}_{1\times m_1} & \frac{1}{2}\boldsymbol{L}_{m-m_1}^{ik} \end{pmatrix}^\top \begin{pmatrix} \mathbf{0}_{1\times m_1} & \frac{1}{2}\boldsymbol{L}_{m-m_1}^{ik} \end{pmatrix} \\ &= \begin{bmatrix} \mathbf{0}_{m_1\times m_1} & \mathbf{0}_{m_1\times(m-m_1)} \\ \mathbf{0}_{(m-m_1)\times m_1} & \left(\frac{s_1^{ik}}{4} - \frac{1}{4s_2^{ik}}\right)\left(\boldsymbol{L}_{m-m_1}^{ik}\right)^\top \boldsymbol{L}_{m-m_1}^{ik} \end{bmatrix},\end{aligned}$$

which is PSD if $s_1^{ik}\times s_2^{ik}\geq 1$. Thus, we let $s_1^{ik}\times s_2^{ik}\geq 1$ hold for any $i\in[N]$ and $k\in[K]$ and by the properties of Schur complement, we have $\begin{bmatrix} \bar{\boldsymbol{A}} & \bar{\boldsymbol{B}} \\ \bar{\boldsymbol{B}}^\top & \bar{\boldsymbol{C}} \end{bmatrix}\succeq 0$ because $\bar{\boldsymbol{C}}$ is invertible and positive definite.

In addition, since Problem (EC.19) is a minimization problem, its optimal value is no larger than the objective corresponding to the feasible solution $(\bar{\boldsymbol{x}}, \bar{\lambda}, \bar{\boldsymbol{Q}}, \bar{\boldsymbol{\zeta}}^i\ \forall i\in[N], \bar{y}_i\ \forall i\in[N])$. That is,

$$\begin{aligned} Z^*{}_{\mathrm{C}}(m) &\leq \bar{\boldsymbol{\lambda}}R_0 + \gamma_2 \boldsymbol{I}_m\bullet\bar{\boldsymbol{Q}} + \frac{1}{N}\sum_{i=1}^N \bar{y}_i \\ &= Z^*{}_{\mathrm{C}}(m_1) + \gamma_2\sum_{i=1}^N\sum_{k=1}^K \frac{s_1^{ik}}{4}\mathrm{trace}\left(\left(\boldsymbol{L}_{m-m_1}^{ik}\right)^\top \boldsymbol{L}_{m-m_1}^{ik}\right) + \frac{1}{N}\sum_{i=1}^N\sum_{k=1}^K s_2^{ik}. \end{aligned} \tag{EC.22}$$

Due to the condition $s_1^{ik}\times s_2^{ik}\geq 1$, we let $s_1^{ik} = \frac{2}{\sqrt{\gamma_2 N \boldsymbol{L}_{m-m_1}^{ik}\left(\boldsymbol{L}_{m-m_1}^{ik}\right)^\top}}$ and $s_2^{ik} = \frac{\sqrt{\gamma_2 N \boldsymbol{L}_{m-m_1}^{ik}\left(\boldsymbol{L}_{m-m_1}^{ik}\right)^\top}}{2}$ for any $i\in[N]$ and $k\in[K]$, which leads to the smallest possible value of the RHS of (EC.22). Therefore, we have

$$\begin{aligned} Z^*{}_{\mathrm{C}}(m) &\leq Z^*{}_{\mathrm{C}}(m_1) + \gamma_2\sum_{i=1}^N\sum_{k=1}^K \frac{s_1^{ik}}{4}\mathrm{trace}\left(\left(\boldsymbol{L}_{m-m_1}^{ik}\right)^\top \boldsymbol{L}_{m-m_1}^{ik}\right) + \frac{1}{N}\sum_{i=1}^N\sum_{k=1}^K s_2^{ik} \\ &= Z^*{}_{\mathrm{C}}(m_1) + \sqrt{\frac{\gamma_2}{N}}\mathrm{trace}\left(\sum_{i=1}^N\sum_{k=1}^K \frac{\left(\boldsymbol{L}_{m-m_1}^{ik}\right)^\top \boldsymbol{L}_{m-m_1}^{ik}}{2\sqrt{\boldsymbol{L}_{m-m_1}^{ik}\left(\boldsymbol{L}_{m-m_1}^{ik}\right)^\top}}\right) + \sqrt{\frac{\gamma_2}{N}}\sum_{i=1}^N\sum_{k=1}^K \frac{\sqrt{\boldsymbol{L}_{m-m_1}^{ik}\left(\boldsymbol{L}_{m-m_1}^{ik}\right)^\top}}{2}. \end{aligned}$$

Finally, since $\mathrm{trace}\left(\sum_{i=1}^N\sum_{k=1}^K \frac{\left(\boldsymbol{L}_{m-m_1}^{ik}\right)^\top \boldsymbol{L}_{m-m_1}^{ik}}{2\sqrt{\boldsymbol{L}_{m-m_1}^{ik}\left(\boldsymbol{L}_{m-m_1}^{ik}\right)^\top}}\right)$ is equal to $\sum_{i=1}^N\sum_{k=1}^K \frac{\sqrt{\boldsymbol{L}_{m-m_1}^{ik}\left(\boldsymbol{L}_{m-m_1}^{ik}\right)^\top}}{2}$, we have

$$0\leq Z^*{}_{\mathrm{C}}(m) - Z^*{}_{\mathrm{C}}(m_1)\leq \sqrt{\frac{\gamma_2}{N}}\sum_{i=1}^N\sum_{k=1}^K\sqrt{\boldsymbol{L}_{m-m_1}^{ik}\left(\boldsymbol{L}_{m-m_1}^{ik}\right)^\top}. \quad \square$$

B.3. Proof of Theorem 6

By Theorem 1 in Gao and Kleywegt (2017), Problem (43) has the following strong dual problem:

$$\min_{\boldsymbol{x}\in\mathcal{X},\boldsymbol{Q}_j\forall j,\lambda}\left\{\lambda R_0 + \gamma_2\sum_{j=1}^P \boldsymbol{I}_{m_j}\bullet\boldsymbol{Q}_j + \int_{\mathbb{R}^m}\sup_{\boldsymbol{\xi}_{\mathrm{I}}} g(\boldsymbol{\xi}_{\mathrm{I}},\hat{\boldsymbol{\xi}})\,\mathbb{P}_0(\mathbf{d}\hat{\boldsymbol{\xi}})\right\}, \tag{EC.23}$$

where $g(\boldsymbol{\xi}_{\mathrm{I}},\hat{\boldsymbol{\xi}}) = \max_{k=1}^K\{y_k(\boldsymbol{x})^\top(\boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}}+\boldsymbol{\mu}) + y_k^0(\boldsymbol{x})\} - \sum_{j=1}^P \boldsymbol{\xi}_{\mathrm{I}_j}^\top\boldsymbol{Q}_j\boldsymbol{\xi}_{\mathrm{I}_j} - \lambda\|\boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}}+\boldsymbol{\mu}-\hat{\boldsymbol{\xi}}\|_1$, and $\boldsymbol{Q}_j$ for any $j\in[P]$ and λ are the Lagrangian multipliers of the primal second-order moment and Wasserstein constraints, respectively. As $\mathbb{P}_0$ denotes an empirical distribution of $\boldsymbol{\xi}$ generated by i.i.d. samples $\{\hat{\boldsymbol{\xi}}^i : i\in N\}\subseteq\mathcal{S}$ from the $\mathbb{P}$, i.e., $\mathbb{P}\{\boldsymbol{\xi} = \hat{\boldsymbol{\xi}}^i\} = \frac{1}{N}$, we have

$$\int_{\mathbb{R}^m}\sup_{\boldsymbol{\xi}_{\mathrm{I}}} g(\boldsymbol{\xi}_{\mathrm{I}},\hat{\boldsymbol{\xi}})\,\mathbb{P}_0(\mathbf{d}\hat{\boldsymbol{\xi}}) = \frac{1}{N}\sum_{i=1}^N\sup_{\boldsymbol{\xi}_{\mathrm{I}}} g(\boldsymbol{\xi}_{\mathrm{I}},\hat{\boldsymbol{\xi}}^i). \tag{EC.24}$$

because $g(\boldsymbol{\xi}_{\mathrm{I}},\hat{\boldsymbol{\xi}})$ is a convex function and $\mathcal{S}_{\mathrm{I}}=\mathbb{R}^m$, which is convex. Thus, by plugging (EC.24) into (EC.23), (EC.23) can be reformulated as

$$\min_{\boldsymbol{x}\in\mathcal{X},\boldsymbol{Q}_j\forall j,\lambda} \left\{\lambda R_0+\gamma_2\sum_{j=1}^{P}\boldsymbol{I}_{m_j}\bullet\boldsymbol{Q}_j+\frac{1}{N}\sum_{i=1}^{N}y_i\right\} \tag{EC.25a}$$

$$\text{s.t.}\quad y_i=\sup_{\boldsymbol{\xi}_{\mathrm{I}}} g(\boldsymbol{\xi}_{\mathrm{I}},\hat{\boldsymbol{\xi}}^i),\quad \forall i\in[N]. \tag{EC.25b}$$

Since Problem (EC.25) is a minimization problem, constraints (EC.25b) can be relaxed to $y_i\geq\sup_{\boldsymbol{\xi}_{\mathrm{I}}} g(\boldsymbol{\xi}_{\mathrm{I}},\hat{\boldsymbol{\xi}}^i)$ for any $i\in[N]$. Thus, we have

$$\begin{aligned}
&y_i\geq\sup_{\boldsymbol{\xi}_{\mathrm{I}}}\left\{\max_{k=1}^{K}\left\{y_k(\boldsymbol{x})^\top\left(\boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}}+\boldsymbol{\mu}\right)+y_k^0(\boldsymbol{x})\right\}-\sum_{j=1}^{P}\boldsymbol{\xi}_{\mathrm{I}_j}^\top\boldsymbol{Q}_j\boldsymbol{\xi}_{\mathrm{I}_j}-\lambda\left\|\boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}}+\boldsymbol{\mu}-\hat{\boldsymbol{\xi}}^i\right\|_1\right\},\ \forall i\in[N]\\
\Leftrightarrow &y_i\geq\max_{k=1}^{K}\sup_{\boldsymbol{\xi}_{\mathrm{I}}}\left\{y_k(\boldsymbol{x})^\top\left(\boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}}+\boldsymbol{\mu}\right)+y_k^0(\boldsymbol{x})-\sum_{j=1}^{P}\boldsymbol{\xi}_{\mathrm{I}_j}^\top\boldsymbol{Q}_j\boldsymbol{\xi}_{\mathrm{I}_j}-\lambda\left\|\boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}}+\boldsymbol{\mu}-\hat{\boldsymbol{\xi}}^i\right\|_1\right\},\ \forall i\in[N]\\
\Leftrightarrow &y_i\geq\sup_{\boldsymbol{\xi}_{\mathrm{I}}}\left\{y_k(\boldsymbol{x})^\top\left(\boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}}+\boldsymbol{\mu}\right)+y_k^0(\boldsymbol{x})-\sum_{j=1}^{P}\boldsymbol{\xi}_{\mathrm{I}_j}^\top\boldsymbol{Q}_j\boldsymbol{\xi}_{\mathrm{I}_j}-\lambda\left\|\boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}}+\boldsymbol{\mu}-\hat{\boldsymbol{\xi}}^i\right\|_1\right\},\ \forall i\in[N],\ \forall k\in[K].
\end{aligned} \tag{EC.26}$$

For any given $i\in[N]$, we let

$$\left\|\boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}}+\boldsymbol{\mu}-\hat{\boldsymbol{\xi}}^i\right\|=\sup_{\|\hat{\zeta}\|_*\leq1}\hat{\boldsymbol{\zeta}}^\top\left(\boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}}+\boldsymbol{\mu}-\hat{\boldsymbol{\xi}}^i\right),$$

and accordingly we have

$$\begin{aligned}
&y_i\geq\sup_{\boldsymbol{\xi}_{\mathrm{I}}}\inf_{\|\hat{\zeta}\|_*\leq1}\left\{y_k(\boldsymbol{x})^\top\left(\boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}}+\boldsymbol{\mu}\right)+y_k^0(\boldsymbol{x})-\sum_{j=1}^{P}\boldsymbol{\xi}_{\mathrm{I}_j}^\top\boldsymbol{Q}_j\boldsymbol{\xi}_{\mathrm{I}_j}-\lambda\hat{\boldsymbol{\zeta}}^\top\left(\boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}}+\boldsymbol{\mu}-\hat{\boldsymbol{\xi}}^i\right)\right\},\ \forall k\in[K]\\
\Leftrightarrow &y_i\geq\inf_{\|\hat{\zeta}\|_*\leq1}\sup_{\boldsymbol{\xi}_{\mathrm{I}}}\left\{y_k(\boldsymbol{x})^\top\left(\boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}}+\boldsymbol{\mu}\right)+y_k^0(\boldsymbol{x})-\sum_{j=1}^{P}\boldsymbol{\xi}_{\mathrm{I}_j}^\top\boldsymbol{Q}_j\boldsymbol{\xi}_{\mathrm{I}_j}-\lambda\hat{\boldsymbol{\zeta}}^\top\left(\boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}}+\boldsymbol{\mu}-\hat{\boldsymbol{\xi}}^i\right)\right\},\ \forall k\in[K]\\
\Leftrightarrow &\exists\hat{\boldsymbol{\zeta}}\text{ s.t. }\left\|\hat{\boldsymbol{\zeta}}\right\|_*\leq1,\ y_i\geq\sup_{\boldsymbol{\xi}_{\mathrm{I}}}\left\{y_k(\boldsymbol{x})^\top\left(\boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}}+\boldsymbol{\mu}\right)+y_k^0(\boldsymbol{x})-\sum_{j=1}^{P}\boldsymbol{\xi}_{\mathrm{I}_j}^\top\boldsymbol{Q}_j\boldsymbol{\xi}_{\mathrm{I}_j}-\lambda\hat{\boldsymbol{\zeta}}^\top\left(\boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}}+\boldsymbol{\mu}-\hat{\boldsymbol{\xi}}^i\right)\right\},\\
&\qquad\forall k\in[K]\\
\Leftrightarrow &\exists\hat{\boldsymbol{\zeta}}\text{ s.t. }\left\|\hat{\boldsymbol{\zeta}}\right\|_*\leq1,\ y_i\geq y_k(\boldsymbol{x})^\top\left(\boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}}+\boldsymbol{\mu}\right)+y_k^0(\boldsymbol{x})-\sum_{j=1}^{P}\boldsymbol{\xi}_{\mathrm{I}_j}^\top\boldsymbol{Q}_j\boldsymbol{\xi}_{\mathrm{I}_j}-\lambda\hat{\boldsymbol{\zeta}}^\top\left(\boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}}+\boldsymbol{\mu}-\hat{\boldsymbol{\xi}}^i\right),\\
&\qquad\forall\boldsymbol{\xi}_{\mathrm{I}}\in\mathbb{R}^m,\forall k\in[K]\\
\Leftrightarrow &\exists\hat{\boldsymbol{\zeta}}\text{ s.t. }\left\|\hat{\boldsymbol{\zeta}}\right\|_*\leq1,\ \begin{bmatrix} Q' & \frac{1}{2}\left(\left(-y_k(\boldsymbol{x})^\top+\lambda\hat{\boldsymbol{\zeta}}^\top\right)\boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top\\ \frac{1}{2}\left(-y_k(\boldsymbol{x})^\top+\lambda\hat{\boldsymbol{\zeta}}^\top\right)\boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}} & y_i-y_k(\boldsymbol{x})^\top\boldsymbol{\mu}-y_k^0(\boldsymbol{x})+\lambda\hat{\boldsymbol{\zeta}}^\top\left(\boldsymbol{\mu}-\hat{\boldsymbol{\xi}}^i\right)\end{bmatrix}\succeq0,\ \forall k\in[K],
\end{aligned} \tag{EC.27}$$

where decision variable $\boldsymbol{Q}'$ is described as (30) and the first equivalence is due to the convexity of $g(\boldsymbol{\xi}_{\mathrm{I}},\hat{\boldsymbol{\xi}})$, $\mathcal{S}_{\mathrm{I}}$, and the feasible region defined by $\|\hat{\zeta}\|_*\leq1$. For any given $i\in[N]$, we replace $\lambda\hat{\boldsymbol{\zeta}}$ by $\boldsymbol{\zeta}^i$, and then we can reduce Problem (43) to the following problem by further replacing (EC.25b) by (EC.27) for any $i\in[N]$:

$$\min_{\substack{\boldsymbol{x},\lambda,\boldsymbol{Q}_j\forall j,\\ \boldsymbol{\zeta}^i\forall i,y_i\forall i}} \lambda R_0+\gamma_2\sum_{j=1}^{P}\boldsymbol{I}_{m_j}\bullet\boldsymbol{Q}_j+\frac{1}{N}\sum_{i=1}^{N}y_i \tag{EC.28a}$$

$$\text{s.t.}\quad \begin{bmatrix} Q' & \frac{1}{2}\left(\left(-y_k(\boldsymbol{x})+\boldsymbol{\zeta}^i\right)^\top\boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top\\ \frac{1}{2}\left(-y_k(\boldsymbol{x})+\boldsymbol{\zeta}^i\right)^\top\boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}} & y_i-y_k(\boldsymbol{x})^\top\boldsymbol{\mu}-y_k^0(\boldsymbol{x})+\boldsymbol{\zeta}^{i\top}\left(\boldsymbol{\mu}-\hat{\boldsymbol{\xi}}^i\right)\end{bmatrix}\succeq0,\ \forall i\in[N],\ \forall k\in[K], \tag{EC.28b}$$

$$\lambda\in\mathbb{R}_+,\ \boldsymbol{x}\in\mathcal{X},\ \left\|\boldsymbol{\zeta}^i\right\|_*\leq\lambda,\ \forall i\in[N].$$

Finally, by Lemma 1, we reformulate Problem (EC.28) as Problem (44) by decomposing the PSD matrix in (EC.28b) equivalently to K PSD matrices. The proof of the claim that Problem (44) provides an upper bound for Problem (38) is the same as that of Theorem 4 and thus is omitted here. $\square$

B.4. Proof of Proposition 10

We reformulate Problem (44) as Problem (EC.28). Let $(\boldsymbol{x}^*, \lambda^*, \boldsymbol{Q}^*, \boldsymbol{\zeta}^{i^*}\ \forall i \in [N], y_i^*\ \forall i \in [N])$ denote an optimal solution of Problem (38) with $\boldsymbol{Q}^*$ represented by (31). Based on this optimal solution, in the following, we construct a feasible solution of Problem (EC.28), denoted by $(\bar{\boldsymbol{x}}, \bar{\lambda}, \bar{\boldsymbol{Q}}_j\ \forall j \in [P], \bar{\boldsymbol{\zeta}}^i\ \forall i \in [N], \bar{y}_i\ \forall i \in [N])$.

First, we let $\bar{\boldsymbol{x}} = \boldsymbol{x}^*$, $\bar{\lambda} = \lambda^*$, $\bar{\boldsymbol{\zeta}}^i = \boldsymbol{\zeta}^{i^*}$ for any $i \in [N]$, $\bar{y}_i = k_0 y_i^*$ for any $i \in [N]$, and $\bar{\boldsymbol{Q}}_j = \bar{\boldsymbol{Q}}'$ (as described in (32)), with $k_j \geq 1$ for any $j \in \{0, 1, 2, \ldots, P\}$. In order for this solution to satisfy (EC.28b), we require

$$\begin{bmatrix} \bar{\boldsymbol{Q}}' & \frac{1}{2}\left(\left(-y_k(\boldsymbol{x}^*)+\boldsymbol{\zeta}^{i^*}\right)^\top \boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top \\ \frac{1}{2}\left(-y_k(\boldsymbol{x}^*)+\boldsymbol{\zeta}^{i^*}\right)^\top \boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}} & k_0 y_i^* - y_k(\boldsymbol{x}^*)^\top \boldsymbol{\mu} - y_k^0(\boldsymbol{x}^*) + \boldsymbol{\zeta}^{i^*\top}\left(\boldsymbol{\mu} - \hat{\boldsymbol{\xi}}^i\right) \end{bmatrix} \succeq 0,\ \forall i \in [N],\ \forall k \in [K]. \tag{EC.29}$$

In the following, we find the values of k_j for any $j \in \{0, 1, 2, \ldots, P\}$ so that (EC.29) holds. To that end, we construct the following matrix

$$\begin{bmatrix} \bar{\boldsymbol{Q}}' & \frac{1}{2}\left(\left(-y_k(\boldsymbol{x}^*)+\boldsymbol{\zeta}^{i^*}\right)^\top \boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top \\ \frac{1}{2}\left(-y_k(\boldsymbol{x}^*)+\boldsymbol{\zeta}^{i^*}\right)^\top \boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}} & k_0\left(y_i^* - y_k(\boldsymbol{x}^*)^\top \boldsymbol{\mu} - y_k^0(\boldsymbol{x}^*) + \boldsymbol{\zeta}^{i^*\top}\left(\boldsymbol{\mu} - \hat{\boldsymbol{\xi}}^i\right)\right) \end{bmatrix},\ \forall i \in [N],\ \forall k \in [K]. \tag{EC.30}$$

Note that subtracting (EC.30) from (EC.29) leads to the following matrix:

$$\begin{bmatrix} \mathbf{0}_{m\times m} & \mathbf{0}_{m\times 1} \\ \mathbf{0}_{1\times m} & (k_0-1)\left(y_k^0(\boldsymbol{x}^*)+y_k(\boldsymbol{x}^*)^\top \boldsymbol{\mu} - \boldsymbol{\zeta}^{i^*\top}\left(\boldsymbol{\mu}-\hat{\boldsymbol{\xi}}^i\right)\right) \end{bmatrix} \succeq 0,\ \forall i \in [N],\ \forall k \in [K],$$

which is PSD because its eigenvalues are non-negative. In fact, $(k_0 - 1)(y_k^0(\boldsymbol{x}^*) + y_k(\boldsymbol{x}^*)^\top \boldsymbol{\mu} - \boldsymbol{\zeta}^{i^*\top}(\boldsymbol{\mu} - \hat{\boldsymbol{\xi}}^i))$ is the only non-zero eigenvalue of this matrix that is non-negative because $k_0 \geq 1$, $-\boldsymbol{\zeta}^{i^*\top}(\boldsymbol{\mu} - \hat{\boldsymbol{\xi}}^i) \geq 0$ due to the assumption $\max_{i=1}^N \{\boldsymbol{\zeta}^{i^*\top}(\boldsymbol{\mu} - \hat{\boldsymbol{\xi}}^i)\} \leq 0$, and we have $y_k^0(\boldsymbol{x}^*) + y_k(\boldsymbol{x}^*)^\top \boldsymbol{\mu} \geq 0$ according to the assumption $\min_{k=1}^K \{y_k^0(\boldsymbol{x}^*) + y_k(\boldsymbol{x}^*)^\top \boldsymbol{\mu}\} \geq 0$. Thus, we choose good values of k_j for any $j \in \{0, 1, 2, \ldots, P\}$ to ensure (EC.30) to be a PSD matrix and accordingly will make (EC.29) hold.

Next, by Lemma 1, in order for (EC.30) to be a PSD, we equivalently require

$$\begin{bmatrix} k_j \boldsymbol{Q}_j^* & \frac{1}{2}\left(\left(-y_k(\boldsymbol{x}^*)+\boldsymbol{\zeta}^{i^*}\right)^\top \boldsymbol{U}_{m\times m_j}\boldsymbol{\Lambda}_{m_j}^{\frac{1}{2}}\right)^\top \\ \frac{1}{2}\left(-y_k(\boldsymbol{x}^*)+\boldsymbol{\zeta}^{i^*}\right)^\top \boldsymbol{U}_{m\times m_j}\boldsymbol{\Lambda}_{m_j}^{\frac{1}{2}} & s_j\left(y_i^* - y_k(\boldsymbol{x}^*)^\top \boldsymbol{\mu} - y_k^0(\boldsymbol{x}^*) + \boldsymbol{\zeta}^{i^*\top}\left(\boldsymbol{\mu} - \hat{\boldsymbol{\xi}}^i\right)\right) \end{bmatrix} \succeq 0,\ \forall k \in [K],\ \forall i \in [N],\ \forall j \in [P], \tag{EC.31}$$

with $\sum_{j=1}^P s_j = k_0$. Constraints (EC.31) can be satisfied by allowing $s_j \times k_j \geq 1$ for any $j \in [P]$ due to (38b). Then, we let $k_0 = k_1 = \cdots = k_P$ and $s_j \times k_j = 1$ for any $j \in [P]$, leading to $k_0 = k_1 = \cdots = k_P = \sqrt{P}$.

Finally, we have $UB_{\text{C}}^* \geq Z_{\text{C}}^*(m)$ by Theorem 6. Meanwhile, as Problem (44) is a minimization problem, UB_{C}^* is no larger than the objective value corresponding to our constructed feasible solution $(\bar{\boldsymbol{x}}, \bar{\lambda}, \bar{\boldsymbol{Q}}_j\ \forall j \in [P], \bar{\boldsymbol{\zeta}}^i\ \forall i \in [N], \bar{y}_i\ \forall i \in [N])$. That is, we have

$$UB_{\text{C}}^* \leq \lambda^* R_0 + \gamma_2 \sum_{j=1}^P \boldsymbol{I}_{m_j} \bullet \left(\sqrt{P}\boldsymbol{Q}_j^*\right) + \frac{1}{N}\sum_{i=1}^N \sqrt{P} y_i^* \leq \sqrt{P}\left(\lambda^* R_0 + \gamma_2 \sum_{j=1}^p \boldsymbol{I}_{m_j} \bullet \boldsymbol{Q}_j^* + \sum_{i=1}^N \frac{1}{N} y_i^*\right) = \sqrt{P} Z_{\text{C}}^*(m),$$

where the second inequality holds because $P \geq 1$. Therefore, we have

$$0 \leq UB_{\text{C}}^* - Z_{\text{C}}^*(m) \leq (\sqrt{P} - 1) Z_{\text{C}}^*(m). \quad \square$$

Appendix C: Supplement to Section 4

First, the outer approximation (8) leads to the following problem:

$$\min_{\substack{\boldsymbol{x},\boldsymbol{z},s,\\ \boldsymbol{\lambda}_k,\boldsymbol{q}_{\mathrm{r}},\boldsymbol{Q}_{\mathrm{r}}}} \quad \boldsymbol{c}^\top \boldsymbol{x} + s + \gamma_2 \boldsymbol{I}_{m_1} \bullet \boldsymbol{Q}_{\mathrm{r}} + \sqrt{\gamma_1}\,\|\boldsymbol{q}_{\mathrm{r}}\|_2 \tag{EC.32a}$$

$$\text{s.t.} \quad \begin{bmatrix} s - \beta_k - \boldsymbol{\lambda}_k^\top \boldsymbol{b} - \alpha_k \boldsymbol{z}_k^\top \boldsymbol{\mu} + \boldsymbol{\lambda}_k^\top \boldsymbol{A}\boldsymbol{\mu} & \frac{1}{2}\left(\boldsymbol{q}_{\mathrm{r}} + \left(\boldsymbol{U}_{mn\times m_1}\boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}}\right)^\top \left(\boldsymbol{A}^\top \boldsymbol{\lambda}_k - \alpha_k z_k\right)\right)^\top \\ \frac{1}{2}\left(\boldsymbol{q}_{\mathrm{r}} + \left(\boldsymbol{U}_{mn\times m_1}\boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}}\right)^\top \left(\boldsymbol{A}^\top \boldsymbol{\lambda}_k - \alpha_k \boldsymbol{z}_k\right)\right) & \boldsymbol{Q}_{\mathrm{r}} \end{bmatrix} \succeq 0, \quad \forall k \in [K],$$

$$\boldsymbol{\lambda}_k \in \mathbb{R}^n_+, \ \forall k \in [K], \ (45\text{d}),$$

$$\sum_{i=1}^m z_{ijk} = d_j, \ \forall j \in [n], \ \forall k \in [K], \tag{EC.32b}$$

$$\sum_{j=1}^n z_{ijk} = x_i, \ \forall i \in [m], \ \forall k \in [K], \tag{EC.32c}$$

$$z_{ijk} \geq 0, \ \forall i \in [m], \ \forall j \in [n], \ \forall k \in [K], \tag{EC.32d}$$

where $\boldsymbol{z}_k \in \mathbb{R}^{mn}$ is a vector whose $((i-1)m+j)$-th element is z_{ijk}.

Second, the outer approximation (41) leads to the following problem:

$$\min_{\boldsymbol{x},\boldsymbol{z},\lambda,\boldsymbol{Q}_{\mathrm{r}},\boldsymbol{\zeta}^i,y_i} \quad \boldsymbol{c}^\top \boldsymbol{x} + \lambda R_0 + \gamma_2 \boldsymbol{I}_{m_1} \bullet \boldsymbol{Q}_{\mathrm{r}} + \frac{1}{N}\sum_{i=1}^N y_i \tag{EC.33}$$

$$\text{s.t.} \quad \begin{bmatrix} \boldsymbol{Q}_{\mathrm{r}} & \frac{1}{2}\left(\left(-\alpha_k \boldsymbol{z}_k^\top + \boldsymbol{\zeta}^{i\top}\right)\boldsymbol{U}_{mn\times m_1}\boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}}\right)^\top \\ \frac{1}{2}\left(-\alpha_k \boldsymbol{z}_k^\top + \boldsymbol{\zeta}^{i\top}\right)\boldsymbol{U}_{mn\times m_1}\boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}} & y_i - \alpha_k \boldsymbol{z}_k^\top \boldsymbol{\mu} - \beta_k + \boldsymbol{\zeta}^{i\top}\left(\boldsymbol{\mu} - \hat{\boldsymbol{\xi}}^i\right) \end{bmatrix} \succeq 0, \ \forall i \in [N], \ \forall k \in [K],$$

$$\lambda \in \mathbb{R}_+, \ \left\|\boldsymbol{\zeta}^i\right\|_* \leq \lambda, \ \forall i \in [N], \ (45\text{d}), \ (\text{EC.32b}) - (\text{EC.32d}),$$

where $\boldsymbol{\zeta}^i \in \mathbb{R}^{mn}$.

Third, the inner approximation (15) leads to the following problem:

$$\min_{\boldsymbol{x},\boldsymbol{z},s,\boldsymbol{\lambda}_k,\boldsymbol{q},\boldsymbol{Q}_{\mathrm{r}}} \quad \boldsymbol{c}^\top \boldsymbol{x} + s + \gamma_2 \boldsymbol{I}_{m_1} \bullet \boldsymbol{Q}_{\mathrm{r}} + \sqrt{\gamma_1}\,\|\boldsymbol{q}\|_2$$

$$\text{s.t.} \quad \begin{bmatrix} s - \beta_k - \boldsymbol{\lambda}_k^\top \boldsymbol{b} - \alpha_k \boldsymbol{z}_k^\top \mu + \boldsymbol{\lambda}_k^\top \boldsymbol{A}\boldsymbol{\mu} & \frac{1}{2}\left(\boldsymbol{q}_1 + \left(\boldsymbol{U}_{mn\times m_1}\boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}}\right)^\top \left(\boldsymbol{A}^\top \boldsymbol{\lambda}_k - \alpha_k \boldsymbol{z}_k\right)\right)^\top \\ \frac{1}{2}\left(\boldsymbol{q}_1 + \left(\boldsymbol{U}_{mn\times m_1}\boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}}\right)^\top \left(\boldsymbol{A}^\top \boldsymbol{\lambda}_k - \alpha_k \boldsymbol{z}_k\right)\right) & \boldsymbol{Q}_{\mathrm{r}} \end{bmatrix} \succeq 0, \ \forall k \in [K],$$

$$\boldsymbol{q}_2 + \left(\boldsymbol{U}_{mn\times(mn-m_1)}\boldsymbol{\Lambda}_{mn-m_1}^{\frac{1}{2}}\right)^\top \left(\boldsymbol{A}^\top \boldsymbol{\lambda}_k - \alpha_k \boldsymbol{z}_k\right) = 0, \ \forall k \in [K],$$

$$\boldsymbol{\lambda}_k \in \mathbb{R}^n_+, \ \forall k \in [K], \ (45\text{d}), \ (\text{EC.32b}) - (\text{EC.32d}),$$

where $\boldsymbol{q} = (\boldsymbol{q}_1^\top \in \mathbb{R}^{m_1}, \boldsymbol{q}_2^\top \in \mathbb{R}^{mn-m_1})^\top$.

Fourth, the inner approximation (18) leads to the following problem:

$$\min_{\boldsymbol{x},\boldsymbol{z},s,\boldsymbol{q},\boldsymbol{Q}_i,\boldsymbol{\lambda}_k} \quad \boldsymbol{c}^\top \boldsymbol{x} + s + \gamma_2 \sum_{i=1}^P \boldsymbol{I}_{m_i} \bullet \boldsymbol{Q}_i + \sqrt{\gamma_1}\,\|\boldsymbol{q}\|_2$$

$$\text{s.t.} \quad \begin{bmatrix} s_{ik} & \frac{1}{2}\left(\boldsymbol{q}_i + \left(\boldsymbol{U}_{mn\times m_i}\boldsymbol{\Lambda}_{m_i}^{\frac{1}{2}}\right)^\top \left(\boldsymbol{A}^\top \boldsymbol{\lambda}_k - \alpha_k \boldsymbol{z}_k\right)\right)^\top \\ \frac{1}{2}\left(\boldsymbol{q}_i + \left(\boldsymbol{U}_{mn\times m_i}\boldsymbol{\Lambda}_{m_i}^{\frac{1}{2}}\right)^\top \left(\boldsymbol{A}^\top \boldsymbol{\lambda}_k - \alpha_k \boldsymbol{z}_k\right)\right) & \boldsymbol{Q}_i \end{bmatrix} \succeq 0,$$

$$\forall i \in [P],\ \forall k \in [K],$$

$$\sum_{i=1}^{P} s_{ik} = s - \beta_k - \boldsymbol{\lambda}_k^\top \boldsymbol{b} - \alpha_k \boldsymbol{z}_k^\top \boldsymbol{\mu} + \boldsymbol{\lambda}_k^\top \boldsymbol{A}\boldsymbol{\mu},\ \forall k \in [K],$$
$$\boldsymbol{\lambda}_k \in \mathbb{R}_+^n,\ \forall k \in [K],\ (45d),\ (EC.32b) - (EC.32d),$$

where $\boldsymbol{Q}_i \in \mathbb{R}^{m_i \times m_i}$ and $\boldsymbol{q}_i \in \mathbb{R}^{m_i}$ for any $i \in [P]$ so that $\sum_{i=1}^{P} m_i = mn$.

Fifth, the inner approximation (44) leads to the following problem:

$$\min_{\substack{\boldsymbol{x},\boldsymbol{z},\lambda,\boldsymbol{Q}_j,\\ \boldsymbol{\zeta}^i, y_i, s_{ijk}}} \ \boldsymbol{c}^\top \boldsymbol{x} + \lambda R_0 + \gamma_2 \sum_{j=1}^{P} \boldsymbol{I}_{m_j} \bullet \boldsymbol{Q}_j + \frac{1}{N}\sum_{i=1}^{N} y_i$$

$$\text{s.t.} \quad \begin{bmatrix} \boldsymbol{Q}_j & \frac{1}{2}\left(\left(-\alpha_k \boldsymbol{z}_k^\top + \boldsymbol{\zeta}^{i\top}\right)\boldsymbol{U}_{mn\times m_j}\boldsymbol{\Lambda}_{m_j}^{\frac{1}{2}}\right)^\top \\ \frac{1}{2}\left(-\alpha_k \boldsymbol{z}_k^\top + \boldsymbol{\zeta}^{i\top}\right)\boldsymbol{U}_{mn\times m_j}\boldsymbol{\Lambda}_{m_j}^{\frac{1}{2}} & s_{jik} \end{bmatrix} \succeq 0,\ \forall j \in [P],$$
$$\forall i \in [N],\ \forall k \in [K],$$

$$\sum_{j=1}^{P} s_{jik} = y_i - \alpha_k \boldsymbol{z}_k^\top \boldsymbol{\mu} - \beta_k + \boldsymbol{\zeta}^{i\top}\left(\boldsymbol{\mu} - \hat{\boldsymbol{\xi}}^i\right),\ \forall i \in [N],\ \forall k \in [K],$$
$$\lambda \in \mathbb{R}_+,\ \left\|\boldsymbol{\zeta}^i\right\|_* \le \lambda,\ \forall i \in [N],\ (45d),\ (EC.32b) - (EC.32d).$$

By Proposition 3, the optimal value gap between Problem (46) with the moment-based ambiguity set and Problem (EC.32) can be described as follows:

$$0 \le Z^*{}_{\mathrm{M}}(mn) - Z^*{}_{\mathrm{M}}(m_1) \le \sqrt{\gamma_2}\left(\sum_{k=1}^{K}\sqrt{\sum_{i=m_1+1}^{mn} \boldsymbol{\Lambda}_{i,i}\left[\left(\boldsymbol{A}^\top \boldsymbol{\lambda}_k^* - \alpha_k \boldsymbol{z}_k^*\right)^\top \boldsymbol{U}_i\right]^2}\right),$$

where $\boldsymbol{z}_k^*$ and $\boldsymbol{\lambda}_k^*$, $k \in [K]$, are optimal solutions of Problem (EC.32), and $Z^*{}_{\mathrm{M}}(mn)$ and $Z^*{}_{\mathrm{M}}(m_1)$ are the optimal values of Problems (46) and (EC.32), respectively. Similarly, by Proposition 9, the optimal value gap between Problem (46) with the combined ambiguity set and Problem (EC.33) can be described as follows:

$$0 \le Z^*{}_{\mathrm{C}}(mn) - Z^*{}_{\mathrm{C}}(m_1) \le \sqrt{\frac{\gamma_2}{N}}\sum_{i=1}^{N}\sum_{k=1}^{K}\sqrt{\boldsymbol{L}_{mn-m_1}^{ik}(\boldsymbol{L}_{mn-m_1}^{ik})^\top},$$

where $Z^*{}_{\mathrm{C}}(m_1)$ is the optimal value of Problem (EC.33) and $\boldsymbol{L}_{mn-m_1}^{ik} = (-\alpha_k \boldsymbol{z}_k^{*\top} + \boldsymbol{\zeta}^{i*\top})\boldsymbol{U}_{mn\times(mn-m_1)}\boldsymbol{\Lambda}_{mn-m_1}^{\frac{1}{2}}$ with $\boldsymbol{z}_k^*$ and $\boldsymbol{\zeta}^{i*}$, $i \in [N]$, representing optimal solutions of Problem (EC.33).