

Computationally Efficient Approximations for Distributionally Robust Optimization under Moment and Wasserstein Ambiguity

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 significantly 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 under a piece-wise linear objective function and 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. The approximations also help construct an interval that is tight for most cases and includes the (unknown) optimal value for a large-scale DRO problem, which usually cannot be solved to optimality (or even feasibility in most cases).

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: (i) rich enough to contain the true distribution with high confidence; (ii) small enough to exclude pathological distributions that make DRO solutions overly conservative; (iii) calibrated easily from historical data; and (iv) 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 (Erdoğ an 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., χ^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). Meanwhile, decision-makers usually face the difficulty of exactly estimating an ambiguity set because high-dimensional and correlated uncertainties are involved in the real-world problems, where different decision-makers may also have different understanding and estimates of the ambiguity. Thus, to cope with such situation and further give the decision-makers

more flexibility and freedom in selecting an appropriate ambiguity set from available alternatives, we may better consider different types of ambiguity sets. For instance, we can combine two different types of ambiguity sets to construct a better one that enjoys the advantages of both. In particular, 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 approximation 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 under a piece-wise linear objective function and 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 inner and outer 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. We also investigate the corresponding results by additionally including the first-order moment information in the combined ambiguity set.
4. Our proposed inner and outer approximations together, while quickly finding a feasible solution with small optimality gap, enable us to construct a tight interval that includes the (unknown) optimal value of DRO formulations. Such an interval is very helpful for decision making in many real-world applications with large-scale instances and high-dimensional uncertainties (e.g., energy and transportation), where the corresponding DRO model cannot be solved to optimality (or even feasibility in most cases) by existing methods in reasonable time.
5. We perform extensive computational experiments to demonstrate the effectiveness of our approximations in 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., $\mathbf{x} = (x_1, \dots, x_m)^\top$ and \mathbf{q} . Similarly, matrices

are represented by bold capital symbols, e.g., \mathbf{A} and $\mathbf{\Sigma}$, 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., \mathbf{Q}_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 $\mathbf{A} \bullet \mathbf{B} = \sum_{i,j} A_{ij} B_{ij}$, where \mathbf{A} and \mathbf{B} are two conformal matrices. If a matrix \mathbf{M} is positive semi-definite (PSD), it is indicated by $\mathbf{M} \succeq 0$. 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} and \mathcal{S} for ambiguity set and support, respectively. For any strictly positive integer number n , we use $[n]$ to represent the set $\{1, 2, \dots, n\}$. The identity matrix of size m is denoted by \mathbf{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 $\mathbf{x} \in \mathcal{X}$, which is a convex set in \mathbb{R}^n , to minimize the expectation of a convex function $f(\mathbf{x}, \boldsymbol{\xi})$ with respect to \mathbb{P} . We present this problem as the following stochastic program:

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

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}_{M1} :

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

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}_{M1} (see [Delage and Ye \(2010\)](#)), i.e.,

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

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 $\gamma_1 \geq 0$ and $\gamma_2 \geq 1$ 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}_{M1} describe that (i) the support of $\boldsymbol{\xi}$ is a subset of \mathcal{S} ; (ii) the mean of $\boldsymbol{\xi}$ lies in an ellipsoid of size γ_1 centered at $\boldsymbol{\mu}$; and (iii) 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 good inner and outer approximations of (DRO-M) much more efficiently and obtain high-quality solutions, thereby complementing the existing studies such as Delage and Ye (2010) and Cheng et al. (2018). More importantly, the inner and outer approximations together can help characterize an interval that includes the unknown optimal value for a large-scale instance that may not be solved to feasibility. 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 $f(\mathbf{x}, \boldsymbol{\xi})$ is piecewise linear convex in $\boldsymbol{\xi}$, i.e., $f(\mathbf{x}, \boldsymbol{\xi}) = \max_{k=1}^K \{y_k^0(\mathbf{x}) + y_k(\mathbf{x})^\top \boldsymbol{\xi}\}$ with both $y_k(\mathbf{x}) = (y_k^1(\mathbf{x}), \dots, y_k^m(\mathbf{x}))^\top$ and $y_k^0(\mathbf{x})$ affine in \mathbf{x} for any $k \in [K]$, and \mathcal{S} is polyhedral, i.e., $\mathcal{S} = \{\boldsymbol{\xi} | \mathbf{A}\boldsymbol{\xi} \leq \mathbf{b}\}$ with $\mathbf{A} \in \mathbb{R}^{n \times m}$ and $\mathbf{b} \in \mathbb{R}^n$, with at least one interior point.*

2.1. Lower Bound

The uncertainty characterization in the ambiguity set affects the computational performance of solving the corresponding DRO problems because a large uncertainty space leads to a large solution search space, which further asks for more computational time. We accordingly investigate the moment-based ambiguity set in this section, and realize that the components with the lowest variance play the smallest role in defining the uncertainty and hence are the best candidates for relaxation. Therefore, we use the PCA approach to project high-dimensional and correlated uncertainty onto a lower dimensional space by preserving the components with the highest variance and relaxing the rest of ones.

First, we perform an eigenvalue decomposition on matrix $\boldsymbol{\Sigma}$, i.e., $\boldsymbol{\Sigma} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^\top = \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}})^\top$, where $\mathbf{U} \in \mathbb{R}^{m \times m}$ is an orthogonal transformation matrix and $\boldsymbol{\Lambda} \in \mathbb{R}^{m \times m}$ is a diagonal matrix whose diagonal elements are in non-increasing order. By letting $\boldsymbol{\xi}_1 = (\mathbf{U}\boldsymbol{\Lambda}^{-1/2})^\top(\boldsymbol{\xi} - \boldsymbol{\mu})$, we reformulate (DRO-M) as

$$\min_{\mathbf{x} \in \mathcal{X}} \max_{\mathbb{P}_1 \in \mathcal{D}_{M2}} \mathbb{E}_{\mathbb{P}_1} \left[f \left(\mathbf{x}, \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_1 + \boldsymbol{\mu} \right) \right], \quad (2)$$

where

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

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

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

$$\min_{\mathbf{x}, s, \mathbf{q}, \mathbf{Q}} s + \gamma_2 \mathbf{I}_m \bullet \mathbf{Q} + \sqrt{\gamma_1} \|\mathbf{q}\|_2 \quad (3a)$$

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

$$\mathbf{Q} \succeq 0, \quad \mathbf{x} \in \mathcal{X},$$

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

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

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. Our approximation techniques may be applied to DRO problems with more general objective functions.

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

$$Z_M^*(m) := \min_{\mathbf{x}, s, \boldsymbol{\lambda}, \mathbf{q}, \mathbf{Q}} s + \gamma_2 \mathbf{I}_m \bullet \mathbf{Q} + \sqrt{\gamma_1} \|\mathbf{q}\|_2 \quad (4a)$$

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

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

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

where $\hat{\boldsymbol{\lambda}} = \{\boldsymbol{\lambda}_1, \dots, \boldsymbol{\lambda}_K\}$.

Proof. See Online Supplement A.1 for the detailed proof. \square

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

$$\boldsymbol{\xi} \approx \mathbf{U} \boldsymbol{\Lambda}^{\frac{1}{2}} [\boldsymbol{\xi}_r; \mathbf{0}_{m-m_1}] + \boldsymbol{\mu} = \mathbf{U}_{m \times m_1} \boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}} \boldsymbol{\xi}_r + \boldsymbol{\mu}, \quad (5)$$

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

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

where

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

with

$$\mathcal{S}_r := \left\{ \boldsymbol{\xi}_r \in \mathbb{R}^{m_1} : \mathbf{U}_{m \times m_1} \mathbf{\Lambda}_{m_1}^{\frac{1}{2}} \boldsymbol{\xi}_r + \boldsymbol{\mu} \in \mathcal{S} \right\}. \quad (6c)$$

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

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

where $\mathbf{q}_r \in \mathbb{R}^{m_1}$ and $\mathbf{Q}_r \in \mathbb{R}^{m_1 \times m_1}$. 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 Online Supplement A.2 for the detailed proof. \square

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

$$Z_M^*(m_1) := \min_{\substack{\mathbf{x}, s, \hat{\boldsymbol{\lambda}}, \\ \mathbf{q}_r, \mathbf{Q}_r}} s + \gamma_2 \mathbf{I}_{m_1} \bullet \mathbf{Q}_r + \sqrt{\gamma_1} \|\mathbf{q}_r\|_2 \quad (8a)$$

$$\text{s.t.} \quad \left[\begin{array}{cc} s - y_k^0(\mathbf{x}) - \boldsymbol{\lambda}_k^\top \mathbf{b} - y_k(\mathbf{x})^\top \boldsymbol{\mu} + \boldsymbol{\lambda}_k^\top \mathbf{A} \boldsymbol{\mu} & \frac{1}{2} \left(\mathbf{q}_r + \left(\mathbf{U}_{m \times m_1} \mathbf{\Lambda}_{m_1}^{\frac{1}{2}} \right)^\top (\mathbf{A}^\top \boldsymbol{\lambda}_k - y_k(\mathbf{x})) \right)^\top \\ \frac{1}{2} \left(\mathbf{q}_r + \left(\mathbf{U}_{m \times m_1} \mathbf{\Lambda}_{m_1}^{\frac{1}{2}} \right)^\top (\mathbf{A}^\top \boldsymbol{\lambda}_k - y_k(\mathbf{x})) \right) & \mathbf{Q}_r \end{array} \right] \succeq 0, \quad (8b)$$

$$\mathbf{x} \in \mathcal{X}, \quad \boldsymbol{\lambda}_k \in \mathbb{R}_+^n, \quad \forall k \in [K], \quad (8c)$$

where $\hat{\boldsymbol{\lambda}} = \{\boldsymbol{\lambda}_1, \dots, \boldsymbol{\lambda}_K\}$.

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

Comparing Problems (4) and (8) in terms of size, one can observe that Problem (8) is significantly easier to solve than Problem (4) because: (i) Problem (8) includes fewer decision variables than Problem (4), i.e., $(m_1^2 + m_1 + 2n + 1)$ vs. $(m^2 + m + 2n + 1)$; and (ii) 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., $Z_M^*(m_1)$ 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: (i) it provides a rough approximation for the optimal value of Problem (4), which may not be solved efficiently in practice; and (ii) 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*

$$0 \leq Z_M^*(m) - Z_M^*(m_1) \leq \sqrt{\gamma_2} \sum_{k=1}^K \sqrt{\sum_{i=m_1+1}^m \Lambda_{i,i} \left((\mathbf{A}^\top \boldsymbol{\lambda}_k^* - y_k(\mathbf{x}^*))^\top \mathbf{U}_i \right)^2}, \quad (9)$$

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

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

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

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

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

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

where $s_1^k > 0$ and $s_2^k > 0$ for any $k \in [K]$. As $\bar{\mathbf{x}} = \mathbf{x}^* \in \mathcal{X}$ and $\bar{\boldsymbol{\lambda}}_k = \boldsymbol{\lambda}_k^* \in \mathbb{R}_+^n$, for any $k \in [K]$, due to constraint (8c), we only require $(\bar{\mathbf{x}}, \bar{s}, \bar{\boldsymbol{\lambda}}_k \forall k \in [K], \bar{\mathbf{q}}, \bar{\mathbf{Q}})$ 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 $(\bar{\mathbf{x}}, \bar{s}, \bar{\boldsymbol{\lambda}}_k \forall k \in [K], \bar{\mathbf{q}}, \bar{\mathbf{Q}})$ to (4b) and use $\bar{\mathbf{Y}}^k$ for any $k \in [K]$ to denote the corresponding matrix in (4b). For any given $k \in [K]$, we perform the following decomposition:

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

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

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

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 $\begin{bmatrix} \bar{\mathbf{A}} & \bar{\mathbf{B}} \\ \bar{\mathbf{B}}^\top & \bar{\mathbf{C}} \end{bmatrix} \succeq 0$ because $\bar{\mathbf{A}}$ 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 $(\bar{\mathbf{x}}, \bar{s}, \bar{\boldsymbol{\lambda}}_k \forall k \in [K], \bar{\mathbf{q}}, \bar{\mathbf{Q}})$. That is,

$$\begin{aligned} Z_{\text{M}}^*(m) &\leq \bar{s} + \gamma_2 \mathbf{I}_m \bullet \bar{\mathbf{Q}} + \sqrt{\gamma_1} \|\bar{\mathbf{q}}\| \\ &= Z_{\text{M}}^*(m_1) + \sum_{k=1}^K s_1^k + \gamma_2 \sum_{k=1}^K \frac{s_2^k}{4} \text{trace}(\mathbf{q}_{m-m_1}^k (\mathbf{q}_{m-m_1}^k)^\top). \quad (12) \end{aligned}$$

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

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

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

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

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 .

Before closing this subsection, we can observe that theoretical optimality gap (9) does not explicitly depend on parameter γ_1 . It is because we develop this gap based on a specific feasible solution of Problem (4), i.e., $(\bar{\mathbf{x}}, \bar{s}, \bar{\boldsymbol{\lambda}}_k \forall k \in [K], \bar{\mathbf{q}}, \bar{\mathbf{Q}})$ where $\bar{\mathbf{q}} = (\mathbf{q}_r^{*\top}, \mathbf{0}_{m-m_1}^\top)^\top$. By setting the last $m - m_1$ elements of $\bar{\mathbf{q}}$ to be 0, we eventually obtain the optimality gap that does not explicitly depend on parameter γ_1 . Technically we can construct a different feasible solution and correspondingly develop a possibly better optimality gap, which will explicitly depend on γ_1 though. For instance, we can let $\bar{\mathbf{x}} = \mathbf{x}^*$, $\bar{\boldsymbol{\lambda}}_k = \boldsymbol{\lambda}_k^*$ for any $k \in [K]$, $\bar{\mathbf{q}} = (\mathbf{q}_r^{*\top}, \hat{\mathbf{q}}^\top)^\top$, $\bar{s} = s^* + \sum_{k=1}^K s^k$, and

$$\bar{\mathbf{Q}} = \begin{bmatrix} \mathbf{Q}_r^* & \mathbf{0}_{m_1 \times (m-m_1)} \\ \mathbf{0}_{(m-m_1) \times m_1} & \hat{\mathbf{Q}} \end{bmatrix},$$

where $s^k > 0$ for any $k \in [K]$. Instead of fixing the values of $\hat{\mathbf{q}}$ and $\hat{\mathbf{Q}}$ as what do in the proof of Proposition 3, we can optimize the values of $s^k, \forall k \in [K]$, $\hat{\mathbf{q}}$, and $\hat{\mathbf{Q}}$ together towards minimizing the difference between the objective value corresponding to this feasible solution and the original optimal value $Z^*_M(m)$, leading to a better gap denoted by $Z^1_{\text{gap}}(m_1)$. That is, we solve the following optimization problem

$$\begin{aligned} Z^1_{\text{gap}}(m_1) := & \min_{\substack{s^k, \forall k \in [K], \\ \hat{\mathbf{q}}, \hat{\mathbf{Q}}}} \sum_{k=1}^K s^k + \gamma_2 \mathbf{I}_{m-m_1} \bullet \hat{\mathbf{Q}} + \sqrt{\gamma_1} \|\hat{\mathbf{q}}\|_2 & (13) \\ \text{s.t.} & \begin{bmatrix} s^k & \frac{1}{2} (\mathbf{q}_{m-m_1}^k - \hat{\mathbf{q}})^\top \\ \frac{1}{2} (\mathbf{q}_{m-m_1}^k - \hat{\mathbf{q}}) & \hat{\mathbf{Q}} \end{bmatrix} \succeq 0, \forall k \in [K], \\ & s^k \in \mathbb{R}_+, \forall k \in [K], \hat{\mathbf{q}} \in \mathbb{R}^{m-m_1}, \hat{\mathbf{Q}} \in \mathbb{R}^{(m-m_1) \times (m-m_1)}, \end{aligned}$$

where the SDP constraints are enforced to ensure the constructed solution to be feasible, similar to what we do in the proof of Proposition 3.

Solving Problem (13) clearly can give us a better theoretical error bound than (9). We also notice that (13) can be computationally intensive, particularly when m_1 is large. Thus, we can further develop a more conservative yet computationally tractable error bound, based on Problem (13), by fixing $\hat{\mathbf{Q}}$ at a feasible value. Specifically, the SDP constraints in (13) implies that $\hat{\mathbf{Q}} \succeq \frac{(\mathbf{q}_{m-m_1}^k - \hat{\mathbf{q}})(\mathbf{q}_{m-m_1}^k - \hat{\mathbf{q}})^\top}{4s^k}$ for any $k \in [K]$ by Schur's complement. Thus, we fix $\hat{\mathbf{Q}}$ at the following feasible value by letting

$$\hat{\mathbf{Q}} = \sum_{k=1}^K \frac{(\mathbf{q}_{m-m_1}^k - \hat{\mathbf{q}})(\mathbf{q}_{m-m_1}^k - \hat{\mathbf{q}})^\top}{4s^k}.$$

It follows that we can reduce Problem (13) to the following problem where we only optimize the values of $s^k, \forall k \in [K]$, and $\hat{\mathbf{q}}$:

$$\begin{aligned} Z_{\text{gap}}^2(m_1) := & \min_{s^k, \forall k \in [K], \hat{\mathbf{q}}} \sum_{k=1}^K s^k + \gamma_2 \sum_{k=1}^K \frac{(\mathbf{q}_{m-m_1}^k - \hat{\mathbf{q}})^\top (\mathbf{q}_{m-m_1}^k - \hat{\mathbf{q}})}{4s^k} + \sqrt{\gamma_1} \|\hat{\mathbf{q}}\|_2 \\ \text{s.t.} \quad & s^k \in \mathbb{R}_+, \forall k \in [K], \hat{\mathbf{q}} \in \mathbb{R}^{m-m_1}. \end{aligned} \quad (14)$$

Problem (14) can be reformulated as a second-order conic program, which is significantly more tractable than (13). Meanwhile, based on the above descriptions on how to construct the required feasible solution for deriving the corresponding error bound, we have

$$0 \leq Z_{\text{M}}^*(m) - Z_{\text{M}}^*(m_1) \leq Z_{\text{gap}}^1(m_1) \leq Z_{\text{gap}}^2(m_1) \leq \sqrt{\gamma_2} \sum_{k=1}^K \sqrt{\sum_{i=m_1+1}^m \Lambda_{i,i} \left((\mathbf{A}^\top \boldsymbol{\lambda}_k^* - y_k(\mathbf{x}^*))^\top \mathbf{U}_i \right)^2}.$$

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 $\boldsymbol{\xi}_1$ in the second-moment constraint in $\mathcal{D}_{\text{M}2}$. This is a relaxation of the second-moment constraint, leading to a larger ambiguity set and so an inner approximation of Problem (2):

$$\min_{\mathbf{x} \in \mathcal{X}} \max_{\mathbb{P}_1 \in \mathcal{D}_{\text{M}4}} \mathbb{E}_{\mathbb{P}_1} \left[f \left(\mathbf{x}, \mathbf{U} \boldsymbol{\Lambda}^{\frac{1}{2}} \boldsymbol{\xi}_1 + \boldsymbol{\mu} \right) \right], \quad (15)$$

where

$$\mathcal{D}_{\text{M}4}(\mathcal{S}_1, \gamma_1, \gamma_2) = \left\{ \mathbb{P}_1 \left| \begin{array}{l} \mathbb{P}_1(\boldsymbol{\xi}_1 \in \mathcal{S}_1) = 1 \\ \mathbb{E}_{\mathbb{P}_1}[\boldsymbol{\xi}_1^\top] \mathbb{E}_{\mathbb{P}_1}[\boldsymbol{\xi}_1] \leq \gamma_1 \\ \mathbb{E}_{\mathbb{P}_1}[\boldsymbol{\xi}_r \boldsymbol{\xi}_r^\top] \preceq \gamma_2 \mathbf{I}_{m_1} \end{array} \right. \right\}.$$

THEOREM 3. *If $f(\mathbf{x}, \mathbf{U} \boldsymbol{\Lambda}^{\frac{1}{2}} \boldsymbol{\xi}_1 + \boldsymbol{\mu})$ is \mathbb{P}_1 -integrable for any $\mathbb{P}_1 \in \mathcal{D}_{\text{M}4}$, then Problem (15) has the same optimal value as the following problem:*

$$\min_{\mathbf{x}, \mathbf{s}, \mathbf{q}, \mathbf{Q}_r} s + \gamma_2 \mathbf{I}_{m_1} \bullet \mathbf{Q}_r + \sqrt{\gamma_1} \|\mathbf{q}\|_2 \quad (16a)$$

$$\text{s.t.} \quad s \geq f \left(\mathbf{x}, \mathbf{U} \boldsymbol{\Lambda}^{\frac{1}{2}} \boldsymbol{\xi}_1 + \boldsymbol{\mu} \right) - \boldsymbol{\xi}_r^\top \mathbf{Q}_r \boldsymbol{\xi}_r - \mathbf{q}^\top \boldsymbol{\xi}_1, \forall \boldsymbol{\xi}_1 \in \mathcal{S}_1, \quad (16b)$$

$$\mathbf{Q}_r \succeq 0, \mathbf{x} \in \mathcal{X},$$

where $\mathbf{q} \in \mathbb{R}^m$ and $\mathbf{Q}_r \in \mathbb{R}^{m_1 \times m_1}$. We also have the following: (i) Problem (16) provides an upper bound for the optimal value of Problem (3); (ii) the optimal value of Problem (16) is non-increasing in m_1 ; and (iii) if $m_1 = m$, then Problems (3) and (16) have the same optimal value.

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

PROPOSITION 4. Under Assumption 1, Problem (16) has the same optimal value as the following SDP formulation:

$$\bar{Z}_M^*(m_1) := \min_{\mathbf{x}, s, \hat{\boldsymbol{\lambda}}, \mathbf{q}, \mathbf{Q}_r} s + \gamma_2 \mathbf{I}_{m_1} \bullet \mathbf{Q}_r + \sqrt{\gamma_1} \|\mathbf{q}\|_2 \quad (17a)$$

$$\text{s.t.} \left[\begin{array}{c} s - y_k^0(\mathbf{x}) - \boldsymbol{\lambda}_k^\top \mathbf{b} - y_k(\mathbf{x})^\top \boldsymbol{\mu} + \boldsymbol{\lambda}_k^\top \mathbf{A} \boldsymbol{\mu} \quad \frac{1}{2} \left(\mathbf{q}_1 + \left(\mathbf{U}_{m \times m_1} \boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}} \right)^\top (\mathbf{A}^\top \boldsymbol{\lambda}_k - y_k(\mathbf{x})) \right)^\top \\ \frac{1}{2} \left(\mathbf{q}_1 + \left(\mathbf{U}_{m \times m_1} \boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}} \right)^\top (\mathbf{A}^\top \boldsymbol{\lambda}_k - y_k(\mathbf{x})) \right) \quad \mathbf{Q}_r \end{array} \right] \succeq 0, \quad \forall k \in [K], \quad (17b)$$

$$\begin{aligned} \mathbf{q}_2 + \left(\mathbf{U}_{m \times (m-m_1)} \boldsymbol{\Lambda}_{m-m_1}^{\frac{1}{2}} \right)^\top (\mathbf{A}^\top \boldsymbol{\lambda}_k - y_k(\mathbf{x})) &= \mathbf{0}, \quad \forall k \in [K], \\ \mathbf{x} \in \mathcal{X}, \quad \boldsymbol{\lambda}_k \in \mathbb{R}_+^n, \quad \forall k \in [K], \end{aligned} \quad (17c)$$

where $\hat{\boldsymbol{\lambda}} = \{\boldsymbol{\lambda}_1, \dots, \boldsymbol{\lambda}_K\}$ and $\mathbf{q} = (\mathbf{q}_1^\top \in \mathbb{R}^{m_1}, \mathbf{q}_2^\top \in \mathbb{R}^{m-m_1})^\top$.

Proof. See Online Supplement A.3 for the detailed proof. \square

One can observe that Problem (17) is significantly easier to solve than Problem (4) due to fewer decision variables and lower-dimensional PSD matrices in Problem (17).

2.3.2. Vector Splitting based Upper Bound We derive the second inner approximation by splitting the random vector $\boldsymbol{\xi}_1$ into P pieces, i.e., $\boldsymbol{\xi}_1 = (\boldsymbol{\xi}_{1_1}^\top, \boldsymbol{\xi}_{1_2}^\top, \dots, \boldsymbol{\xi}_{1_P}^\top)^\top$, where $\boldsymbol{\xi}_{1_i} \in \mathbb{R}^{m_i}$, $\forall i \in [P]$, and $\sum_{i=1}^P m_i = m$. Accordingly, we revise the second-moment constraint in \mathcal{D}_{M_2} with respect to these smaller pieces, leading to the following ambiguity set:

$$\mathcal{D}_{M_5}(\mathcal{S}_1, \gamma_1, \gamma_2) = \left\{ \mathbb{P}_1 \left| \begin{array}{l} \mathbb{P}_1(\boldsymbol{\xi}_1 \in \mathcal{S}_1) = 1 \\ \mathbb{E}_{\mathbb{P}_1}[\boldsymbol{\xi}_1^\top] \mathbb{E}_{\mathbb{P}_1}[\boldsymbol{\xi}_1] \leq \gamma_1 \\ \mathbb{E}_{\mathbb{P}_1}[\boldsymbol{\xi}_{1_i} \boldsymbol{\xi}_{1_i}^\top] \preceq \gamma_2 \mathbf{I}_{m_i}, \quad \forall i \in [P] \end{array} \right. \right\}.$$

Set \mathcal{D}_{M_5} is a superset of \mathcal{D}_{M_2} because we ignore the correlations among $\boldsymbol{\xi}_{1_p}$ and $\boldsymbol{\xi}_{1_q}$ for any $p, q \in [P]$ with $p \neq q$. This leads to the following inner approximation of Problem (2):

$$\min_{\mathbf{x} \in \mathcal{X}} \max_{\mathbb{P}_1 \in \mathcal{D}_{M_5}} \mathbb{E}_{\mathbb{P}_1} \left[f \left(\mathbf{x}, \mathbf{U} \boldsymbol{\Lambda}^{\frac{1}{2}} \boldsymbol{\xi}_1 + \boldsymbol{\mu} \right) \right]. \quad (18)$$

THEOREM 4. If $f(\mathbf{x}, \mathbf{U} \boldsymbol{\Lambda}^{\frac{1}{2}} \boldsymbol{\xi}_1 + \boldsymbol{\mu})$ is \mathbb{P}_1 -integrable for any $\mathbb{P}_1 \in \mathcal{D}_{M_5}$, then Problem (18) has the same optimal value as the following problem:

$$\min_{\mathbf{x}, s, \mathbf{q}, \mathbf{Q}} s + \gamma_2 \sum_{i=1}^P \mathbf{I}_{m_i} \bullet \mathbf{Q}_i + \sqrt{\gamma_1} \|\mathbf{q}\|_2 \quad (19a)$$

$$\begin{aligned} \text{s.t.} \quad s &\geq f \left(\mathbf{x}, \mathbf{U} \boldsymbol{\Lambda}^{\frac{1}{2}} \boldsymbol{\xi}_1 + \boldsymbol{\mu} \right) - \sum_{i=1}^P \boldsymbol{\xi}_{1_i}^\top \mathbf{Q}_i \boldsymbol{\xi}_{1_i} - \mathbf{q}^\top \boldsymbol{\xi}_1, \quad \forall \boldsymbol{\xi}_1 \in \mathcal{S}_1, \\ \mathbf{x} &\in \mathcal{X}, \quad \mathbf{Q}_i \succeq 0, \quad \forall i \in [P], \end{aligned} \quad (19b)$$

where $\mathbf{q} \in \mathbb{R}^m$, $\mathbf{Q}_i \in \mathbb{R}^{m_i \times m_i}$ for any $i \in [P]$, and $\hat{\mathbf{Q}} = \{\mathbf{Q}_1, \dots, \mathbf{Q}_P\}$. Furthermore, Problem (19) provides an upper bound for the optimal value of Problem (3).

Proof. See Online Supplement A.4 for the detailed proof. \square

PROPOSITION 5. Under Assumption 1, Problem (19) has the same optimal value as the following SDP formulation:

$$\begin{aligned}
 UB_M^* := \min_{\substack{\mathbf{x}, s, \mathbf{q}, \\ \hat{\mathbf{Q}}, \hat{\boldsymbol{\lambda}}, \hat{s}}} s + \gamma_2 \sum_{i=1}^P \mathbf{I}_{m_i} \bullet \mathbf{Q}_i + \sqrt{\gamma_1} \|\mathbf{q}\|_2 \quad (20) \\
 \text{s.t.} \quad & \begin{bmatrix} s_{ik} & & \frac{1}{2} \left(\mathbf{q}_i + \left(\mathbf{U}_{m \times m_i} \boldsymbol{\Lambda}_{\frac{1}{2} m_i} \right)^\top (\mathbf{A}^\top \boldsymbol{\lambda}_k - y_k(\mathbf{x})) \right)^\top \\ \frac{1}{2} \left(\mathbf{q}_i + \left(\mathbf{U}_{m \times m_i} \boldsymbol{\Lambda}_{\frac{1}{2} m_i} \right)^\top (\mathbf{A}^\top \boldsymbol{\lambda}_k - y_k(\mathbf{x})) \right) & & \mathbf{Q}_i \end{bmatrix} \succeq 0, \\
 & \forall i \in [P], \forall k \in [K], \\
 & \sum_{i=1}^P s_{ik} = s - y_k^0(\mathbf{x}) - \boldsymbol{\lambda}_k^\top \mathbf{b} - y_k(\mathbf{x})^\top \boldsymbol{\mu} + \boldsymbol{\lambda}_k^\top \mathbf{A} \boldsymbol{\mu}, \quad \forall k \in [K], \\
 & \boldsymbol{\lambda}_k \in \mathbb{R}_+^n, \quad \forall k \in [K], \quad \mathbf{x} \in \mathcal{X},
 \end{aligned}$$

where $\mathbf{q}_i \in \mathbb{R}^{m_i}$, $\mathbf{Q}_i \in \mathbb{R}^{m_i \times m_i}$ for any $i \in [P]$, $\hat{\mathbf{Q}} = \{\mathbf{Q}_1, \dots, \mathbf{Q}_P\}$, $\hat{\boldsymbol{\lambda}} = \{\boldsymbol{\lambda}_1, \dots, \boldsymbol{\lambda}_K\}$, and $\hat{s} = \{s_{ik}, \forall i \in [P], \forall k \in [K]\}$.

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

One can observe that Problem (20) is significantly easier to solve than Problem (4) because it has smaller-sized PSD matrices and matrix variables compared to Problem (4). We also note that the size of each piece (i.e., m_i with $i \in [P]$) does not necessarily equal to each other. Thus, our theoretical results, including theoretical reformulations in this section and optimal gap bounds in the following Section 2.4, are general with respect to the size of each piece. Nevertheless, different ways of assigning a random vector to pieces lead to different computational performance. We will explain how we split the random vector in our numerical experiments in Section 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 (17) (resp. Problem (20)). Before that, we present the following lemma that will facilitate the proofs in this section.

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

$$\mathbf{Z} = \begin{bmatrix} s & \mathbf{q}_1^\top & \dots & \mathbf{q}_K^\top \\ \mathbf{q}_1 & \mathbf{Q}_1 & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{q}_K & \mathbf{0} & \dots & \mathbf{Q}_K \end{bmatrix} \succeq 0, \quad (21)$$

where $s \in \mathbb{R}$, $\mathbf{q}_k \in \mathbb{R}^{m_k}$ for any $k \in [K]$ with $\sum_{k=1}^K m_k = m$, $\mathbf{Q}_k \in \mathbb{R}^{m_k \times m_k}$ for any $k \in [K]$, and other components are zero. Inequality (21) holds if and only if there exist $\{s_k\}_{k=1}^K$ with $\sum_{k=1}^K s_k = s$ such that

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

Proof. We prove the following equivalence:

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

Matrix \mathbf{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 $\boldsymbol{\eta} = (\eta_0, \boldsymbol{\eta}_1^\top, \boldsymbol{\eta}_2^\top, \dots, \boldsymbol{\eta}_K^\top)^\top$ 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 Online Supplements A.1 and A.3, respectively, we assume that $\eta_0 = 1$ without loss of generality. Thus, we have

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

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

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

In summary, we have

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

PROPOSITION 6. Suppose that \mathbf{x}^* is an optimal solution of Problem (17) and $\mathbf{Y}_r^{k*} = \begin{bmatrix} \mathbf{Y}_{11}^{k*} & \mathbf{Y}_{12r}^{k*} \\ \mathbf{Y}_{12r}^{k*} & \mathbf{Y}_{22r}^{k*} \end{bmatrix}$ and $\boldsymbol{\theta}_k^*$ are the corresponding dual optimal solutions associated with constraints (17b) and (17c), respectively. Then, it holds that

$$0 \leq \bar{Z}_M^*(m_1) - Z_M^*(m) \leq \sum_{k=1}^K \left((\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}})^\top \mathbf{y}_k(\mathbf{x}^*) \right)^\top \left(\mathbf{Y}_{12r}^{k*}{}^\top, \boldsymbol{\theta}_k^{*\top} \right)^\top \left(\frac{c^k - 1}{c^k} \right), \quad (22)$$

where

$$c^k = \sqrt{1 + \frac{1}{Y_{11}^{k*}} \sum_{i=1}^{m-m_1} \frac{K}{\gamma_2} \theta_{ki}^{*2}}, \quad \forall k \in [K]. \quad (23)$$

Proof. We prove the results by investigating the duals of Problems (4) and (17) where \mathbf{x} is fixed at \mathbf{x}^* . Given $\mathbf{x} = \mathbf{x}^*$, we use \mathbf{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:

$$Z_M^*(m) = \max_{\mathbf{Y}} \sum_{k=1}^K (y_k^0(\mathbf{x}) + y_k(\mathbf{x})^\top \boldsymbol{\mu}) Y_{11}^k + \sum_{k=1}^K \left(\left(\mathbf{U} \boldsymbol{\Lambda}^{\frac{1}{2}} \right)^\top y_k(\mathbf{x}) \right)^\top \mathbf{Y}_{12}^k \quad (24a)$$

$$\text{s.t. } 1 - \sum_{k=1}^K Y_{11}^k = 0, \quad (24b)$$

$$\sqrt{\gamma_1} - \left\| \sum_{k=1}^K \mathbf{Y}_{12}^k \right\|_2 \geq 0, \quad (24c)$$

$$\gamma_2 \mathbf{I}_m - \sum_{k=1}^K \mathbf{Y}_{22}^k \succeq 0, \quad (24d)$$

$$\mathbf{b} Y_{11}^k - \mathbf{A} \boldsymbol{\mu} Y_{11}^k - \mathbf{A} \mathbf{U} \boldsymbol{\Lambda}^{\frac{1}{2}} \mathbf{Y}_{12}^k \geq 0, \quad \forall k \in [K], \quad (24e)$$

$$\mathbf{Y}^k : \begin{bmatrix} Y_{11}^k & \mathbf{Y}_{12}^k{}^\top \\ \mathbf{Y}_{12}^k & \mathbf{Y}_{22}^k \end{bmatrix} \succeq 0, \quad \forall k \in [K]. \quad (24f)$$

Similarly, given $\mathbf{x} = \mathbf{x}^*$, we use \mathbf{Y}_r^k and $\boldsymbol{\theta}_k$ to denote the dual variables of constraints (17b) and (17c), respectively, for any $k \in [K]$, where the dual of Problem (17) can be described as follows:

$$\begin{aligned} \bar{Z}_M^*(m_1) = \max_{\mathbf{Y}, \boldsymbol{\theta}} & \sum_{k=1}^K (y_k^0(\mathbf{x}) + y_k(\mathbf{x})^\top \boldsymbol{\mu}) Y_{11}^k + \sum_{k=1}^K \left(\left(\mathbf{U}_{m \times m_1} \boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}} \right)^\top y_k(\mathbf{x}) \right)^\top \mathbf{Y}_{12r}^k \\ & + \sum_{k=1}^K \left(\left(\mathbf{U}_{m \times (m-m_1)} \boldsymbol{\Lambda}_{m-m_1}^{\frac{1}{2}} \right)^\top y_k(\mathbf{x}) \right)^\top \boldsymbol{\theta}_k \end{aligned} \quad (25a)$$

$$\text{s.t. } 1 - \sum_{k=1}^K Y_{11}^k = 0, \quad (25b)$$

$$\sqrt{\gamma_1} - \left\| \sum_{k=1}^K \left(\mathbf{Y}_{12r}^k{}^\top, \boldsymbol{\theta}_k{}^\top \right)^\top \right\|_2 \geq 0, \quad (25c)$$

$$\gamma_2 \mathbf{I}_{m_1} - \sum_{k=1}^K \mathbf{Y}_{22r}^k \succeq 0, \quad (25d)$$

$$\mathbf{b} Y_{11}^k - \mathbf{A} \boldsymbol{\mu} Y_{11}^k - \mathbf{A} \mathbf{U}_{m \times m_1} \boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}} \mathbf{Y}_{12r}^k - \mathbf{A} \mathbf{U}_{m \times (m-m_1)} \boldsymbol{\Lambda}_{m-m_1}^{\frac{1}{2}} \boldsymbol{\theta}_k \geq 0, \quad \forall k \in [K], \quad (25e)$$

$$\mathbf{Y}_r^k : \begin{bmatrix} Y_{11}^k & \mathbf{Y}_{12r}^k{}^\top \\ \mathbf{Y}_{12r}^k & \mathbf{Y}_{22r}^k \end{bmatrix} \succeq 0, \quad \boldsymbol{\theta}_k \text{ free}, \quad \forall k \in [K]. \quad (25f)$$

Given an optimal solution of Problem (25), i.e., \mathbf{Y}_r^{k*} and $\boldsymbol{\theta}_k^*$ for any $k \in [K]$, we construct a feasible solution of Problem (24), represented by $\bar{\mathbf{Y}}^k = \begin{bmatrix} \bar{Y}_{11}^k & \bar{\mathbf{Y}}_{12}^k{}^\top \\ \bar{\mathbf{Y}}_{12}^k & \bar{\mathbf{Y}}_{22}^k \end{bmatrix}$ for any $k \in [K]$. For any given

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

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

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 $\bar{\mathbf{Y}}^k$ satisfies all the constraints in Problem (24).

First, for the solution $(\bar{\mathbf{Y}}^k \forall k \in [K])$, it satisfies constraint (24b) because $1 - \sum_{k=1}^K Y_{11}^{k*} = 0$ due to constraint (25b). This solution satisfies constraint (24c) because

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

where the first equality is due to the definition of $\bar{\mathbf{Y}}_{12}^k$, the first inequality is because $c^k \geq 1$, and the second inequality is because of constraint (25c). This solution also satisfies constraints (24e) because for any $k \in [K]$, we have $\mathbf{b} \bar{Y}_{11}^k - \mathbf{A} \boldsymbol{\mu} \bar{Y}_{11}^k - \mathbf{A} \mathbf{U} \boldsymbol{\Lambda}^{\frac{1}{2}} \bar{\mathbf{Y}}_{12}^k = \mathbf{b} Y_{11}^{k*} - \mathbf{A} \boldsymbol{\mu} Y_{11}^{k*} - \mathbf{A} \mathbf{U}_{m \times m_1} \boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}} \mathbf{Y}_{12r}^{k*} - \mathbf{A} \mathbf{U}_{m \times (m-m_1)} \boldsymbol{\Lambda}_{m-m_1}^{\frac{1}{2}} \boldsymbol{\theta}_k^* \geq 0$, where the first equality is due to the definition of $\bar{\mathbf{Y}}_{12}^k$ and the first inequality is because of constraints (25e).

Next, in order for $(\bar{\mathbf{Y}}^k \forall k \in [K])$ to satisfy constraint (24d), we require $\gamma_2 \mathbf{I}_m - \sum_{k=1}^K \bar{Y}_{22}^k \succeq 0$, which is equivalent to

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

Since $\sum_{k=1}^K \mathbf{Y}_{22r}^{k*} \preceq \gamma_2 \mathbf{I}_{m_1}$ due to constraint (25d), we require

$$\sum_{k=1}^K w_i^k \leq \gamma_2, \quad \forall i \in [m - m_1]. \quad (26)$$

In addition, in order for $(\bar{\mathbf{Y}}^k \forall k \in [K])$ to satisfy constraints (24f), matrix $\bar{\mathbf{Y}}^k$ must be PSD for any $k \in [K]$. Note that if $c^k \bar{\mathbf{Y}}^k$ is PSD, then $\bar{\mathbf{Y}}^k$ is PSD because $c^k \geq 1$. Therefore, we consider the following decomposition on $c^k \bar{\mathbf{Y}}^k$:

$$c^k \bar{\mathbf{Y}}^k = \begin{bmatrix} \frac{1}{c^k} \mathbf{Y}_{11}^{k*} & \mathbf{Y}_{12r}^{k* \top} & \mathbf{0}_{1 \times (m-m_1)} \\ \mathbf{Y}_{12r}^{k*} & c^k \mathbf{Y}_{22r}^{k*} & \\ \mathbf{0}_{(m-m_1) \times 1} & \mathbf{0}_{(m-m_1) \times (m-m_1)} & \end{bmatrix} + \begin{bmatrix} \left(c^k - \frac{1}{c^k} \right) \mathbf{Y}_{11}^{k*} & \mathbf{0}_{1 \times m_1} & \boldsymbol{\theta}_k^{* \top} \\ \mathbf{0}_{m_1 \times 1} & \mathbf{0}_{m_1 \times m_1} & \mathbf{0} \\ \boldsymbol{\theta}_k^* & \mathbf{0} & c^k w_1^k \\ & & \ddots \\ & & c^k w_{m-m_1}^k \end{bmatrix}, \quad \forall k \in [K], \quad (27)$$

where the first matrix on the RHS of (27) is clearly PSD because \mathbf{Y}_r^{k*} is PSD due to constraints (25f) and we require the second one to be PSD as well so that $\bar{\mathbf{Y}}^k$ can be PSD. By Lemma 1, we equivalently require

$$\begin{bmatrix} y_i^k & \theta_{ki}^* \\ \theta_{ki}^* & c^k w_i^k \end{bmatrix} \succeq 0, \quad \forall i \in [m - m_1], \quad \forall k \in [K], \quad \text{with} \quad \sum_{i=1}^{m-m_1} y_i^k = \left(c^k - \frac{1}{c^k}\right) Y_{11}^{k*}, \quad \forall k \in [K]. \quad (28)$$

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 (26). It follows that, from (28), 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

$$c^k \geq \frac{(\theta_{ki}^*)^2}{w_i y_i^k} \geq \frac{K}{\gamma_2} \frac{(\theta_{ki}^*)^2}{y_i^k}, \quad (29)$$

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 (29) is equivalent to $y_i^k \geq \frac{K}{\gamma_2} \frac{(\theta_{ki}^*)^2}{c^k}$ and we have $\sum_{i=1}^{m-m_1} y_i^k = (c^k - \frac{1}{c^k}) Y_{11}^{k*}$ from (28), we can conclude that

$$\begin{aligned} \sum_{i=1}^{m-m_1} y_i^k &= \left(c^k - \frac{1}{c^k}\right) Y_{11}^{k*} \geq \sum_{i=1}^{m-m_1} \frac{K}{\gamma_2} \frac{(\theta_{ki}^*)^2}{c^k}, \quad \forall k \in [K] \\ \Rightarrow c^k &\geq \sqrt{1 + \frac{1}{Y_{11}^{k*}} \sum_{i=1}^{m-m_1} \frac{K}{\gamma_2} (\theta_{ki}^*)^2}, \quad \forall k \in [K]. \end{aligned} \quad (30)$$

Therefore, we choose the value of c^k such that (30) is satisfied at equality, leading to (23), while the constructed solution ($\bar{\mathbf{Y}}^k \quad \forall k \in [K]$) satisfies all the constraints in Problem (24).

Finally, by Theorem 3, we have $\bar{Z}_M^*(m_1) - Z_M^*(m) \geq 0$. Meanwhile, we have

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

where the inequality holds because the constructed solution ($\bar{\mathbf{Y}}^k \quad \forall k \in [K]$) is feasible for Problem (24), which is a maximization problem. Therefore, it follows that

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

REMARK 2. Optimality gap (22) does not explicitly depend on γ_1 while γ_2 is explicitly there because upper bound (17) captures the complete first-order moment information while dropping some information about the second-order moment.

PROPOSITION 7. Suppose that Assumption 1 holds, $\boldsymbol{\mu} \in \mathcal{S}$, and $\min_{k=1}^K \{y_k^0(\mathbf{x}^*) + y_k(\mathbf{x}^*)^\top \boldsymbol{\mu}\} \geq 0$, where \mathbf{x}^* is an optimal solution of Problem (4), then the relative gap between the optimal values of Problems (20) and (4) is bounded from above by $\sqrt{P}-1$, i.e., $0 \leq UB_M^* - Z_M^*(m) \leq (\sqrt{P}-1)Z_M^*(m)$.

Proof. We reformulate Problem (20) as the following problem:

$$UB_M^* = \min_{\substack{\mathbf{x}, s, \mathbf{q}, \\ \hat{\mathbf{Q}}, \hat{\boldsymbol{\lambda}}} } s + \gamma_2 \sum_{i=1}^P \mathbf{I}_{m_i} \bullet \mathbf{Q}_i + \sqrt{\gamma_1} \|\mathbf{q}\|_2 \quad (31a)$$

$$\text{s.t.} \quad \begin{bmatrix} s - y_k^0(\mathbf{x}) - \boldsymbol{\lambda}_k^\top \mathbf{b} - y_k(\mathbf{x})^\top \boldsymbol{\mu} + \boldsymbol{\lambda}_k^\top \mathbf{A} \boldsymbol{\mu} & \frac{1}{2} \left(\mathbf{q} + \left(\mathbf{U} \boldsymbol{\Lambda}^{\frac{1}{2}} \right)^\top \left(\mathbf{A}^\top \boldsymbol{\lambda}_k - y_k(\mathbf{x}) \right) \right)^\top \\ \frac{1}{2} \left(\mathbf{q} + \left(\mathbf{U} \boldsymbol{\Lambda}^{\frac{1}{2}} \right)^\top \left(\mathbf{A}^\top \boldsymbol{\lambda}_k - y_k(\mathbf{x}) \right) \right) & \mathbf{Q}' \end{bmatrix} \succeq 0, \quad \forall k \in [K], \quad (31b)$$

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

where

$$\mathbf{Q}' = \begin{bmatrix} \mathbf{Q}_1 & \mathbf{0}_{m_1 \times m_2} & \cdots & \mathbf{0}_{m_1 \times m_{P-1}} & \mathbf{0}_{m_1 \times m_P} \\ \mathbf{0}_{m_2 \times m_1} & \mathbf{Q}_2 & \cdots & \mathbf{0}_{m_2 \times m_{P-1}} & \mathbf{0}_{m_2 \times m_P} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0}_{m_{P-1} \times m_1} & \mathbf{0}_{m_{P-1} \times m_2} & \cdots & \mathbf{Q}_{P-1} & \mathbf{0}_{m_{P-1} \times m_P} \\ \mathbf{0}_{m_P \times m_1} & \mathbf{0}_{m_P \times m_2} & \cdots & \mathbf{0}_{m_P \times m_{P-1}} & \mathbf{Q}_P \end{bmatrix}, \quad (32)$$

$\mathbf{Q}_i \in \mathbb{R}^{m_i \times m_i}$ for any $i \in [P]$ with $\sum_{i=1}^P m_i = m$, $\hat{\mathbf{Q}} = \{\mathbf{Q}_1, \dots, \mathbf{Q}_P\}$, and $\hat{\boldsymbol{\lambda}} = \{\boldsymbol{\lambda}_1, \dots, \boldsymbol{\lambda}_K\}$. Let $(\mathbf{x}^*, s^*, \boldsymbol{\lambda}_k^* \forall k \in [K], \mathbf{q}^*, \mathbf{Q}^*)$ denote an optimal solution of Problem (4) with

$$\mathbf{Q}^* = \begin{bmatrix} \mathbf{Q}_1^* & \mathbf{Q}_{m_1 \times m_2}^* & \cdots & \mathbf{Q}_{m_1 \times m_{P-1}}^* & \mathbf{Q}_{m_1 \times m_P}^* \\ \mathbf{Q}_{m_2 \times m_1}^* & \mathbf{Q}_2^* & \cdots & \mathbf{Q}_{m_2 \times m_{P-1}}^* & \mathbf{Q}_{m_2 \times m_P}^* \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{Q}_{m_{P-1} \times m_1}^* & \mathbf{Q}_{m_{P-1} \times m_2}^* & \cdots & \mathbf{Q}_{P-1}^* & \mathbf{Q}_{m_{P-1} \times m_P}^* \\ \mathbf{Q}_{m_P \times m_1}^* & \mathbf{Q}_{m_P \times m_2}^* & \cdots & \mathbf{Q}_{m_P \times m_{P-1}}^* & \mathbf{Q}_P^* \end{bmatrix}. \quad (33)$$

Based on this optimal solution, in the following, we construct a feasible solution of Problem (31), denoted by $(\bar{\mathbf{x}}, \bar{s}, \bar{\mathbf{q}}, \bar{\mathbf{Q}}', \bar{\boldsymbol{\lambda}}_k \forall k \in [K])$.

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

$$\bar{\mathbf{Q}}' = \begin{bmatrix} k_1 \mathbf{Q}_1^* & \mathbf{0}_{m_1 \times m_2} & \cdots & \mathbf{0}_{m_1 \times m_{P-1}} & \mathbf{0}_{m_1 \times m_P} \\ \mathbf{0}_{m_2 \times m_1} & k_2 \mathbf{Q}_2^* & \cdots & \mathbf{0}_{m_2 \times m_{P-1}} & \mathbf{0}_{m_2 \times m_P} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0}_{m_{P-1} \times m_1} & \mathbf{0}_{m_{P-1} \times m_2} & \cdots & k_{P-1} \mathbf{Q}_{P-1}^* & \mathbf{0}_{m_{P-1} \times m_P} \\ \mathbf{0}_{m_P \times m_1} & \mathbf{0}_{m_P \times m_2} & \cdots & \mathbf{0}_{m_P \times m_{P-1}} & k_P \mathbf{Q}_P^* \end{bmatrix}, \quad (34)$$

with $k_i \geq 1$ for any $i \in \{0, 1, 2, \dots, P\}$. In order for this solution to satisfy (31b), we require

$$\begin{bmatrix} k_0 s^* - y_k^0(\mathbf{x}^*) - \boldsymbol{\lambda}_k^{*\top} \mathbf{b} - y_k(\mathbf{x}^*)^\top \boldsymbol{\mu} + \boldsymbol{\lambda}_k^{*\top} \mathbf{A} \boldsymbol{\mu} & \frac{1}{2} \left(\mathbf{q}^* + \left(\mathbf{U} \boldsymbol{\Lambda}^{\frac{1}{2}} \right)^\top \left(\mathbf{A}^\top \boldsymbol{\lambda}_k^* - y_k(\mathbf{x}^*) \right) \right)^\top \\ \frac{1}{2} \left(\mathbf{q}^* + \left(\mathbf{U} \boldsymbol{\Lambda}^{\frac{1}{2}} \right)^\top \left(\mathbf{A}^\top \boldsymbol{\lambda}_k^* - y_k(\mathbf{x}^*) \right) \right) & \bar{\mathbf{Q}}' \end{bmatrix} \succeq 0, \quad \forall k \in [K]. \quad (35)$$

In the following, we find the values of k_i for any $i \in \{0, 1, 2, \dots, P\}$ so that (35) holds. To that end, we construct the following matrix

$$\begin{bmatrix} k_0(s^* - y_k^0(\mathbf{x}^*) - \boldsymbol{\lambda}_k^{*\top} \mathbf{b} - y_k(\mathbf{x}^*)^\top \boldsymbol{\mu} + \boldsymbol{\lambda}_k^{*\top} \mathbf{A} \boldsymbol{\mu}) & \frac{1}{2} \left(\mathbf{q}^* + \left(\mathbf{U} \boldsymbol{\Lambda}^{\frac{1}{2}} \right)^\top (\mathbf{A}^\top \boldsymbol{\lambda}_k^* - y_k(\mathbf{x}^*)) \right)^\top \\ \frac{1}{2} \left(\mathbf{q}^* + \left(\mathbf{U} \boldsymbol{\Lambda}^{\frac{1}{2}} \right)^\top (\mathbf{A}^\top \boldsymbol{\lambda}_k^* - y_k(\mathbf{x}^*)) \right) & \bar{\mathbf{Q}}' \end{bmatrix}, \quad \forall k \in [K]. \quad (36)$$

Note that subtracting (36) from (35) leads to the following matrix:

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

which is PSD because its eigenvalues are non-negative. In fact, $(k_0 - 1)(y_k^0(\mathbf{x}^*) + y_k(\mathbf{x}^*)^\top \boldsymbol{\mu} - \boldsymbol{\lambda}_k^{*\top} (\mathbf{A} \boldsymbol{\mu} - \mathbf{b}))$ is the only non-zero eigenvalue of this matrix that is non-negative because $k_0 \geq 1$, $-\boldsymbol{\lambda}_k^{*\top} (\mathbf{A} \boldsymbol{\mu} - \mathbf{b}) \geq 0$ due to $\mathbf{A} \boldsymbol{\mu} \leq \mathbf{b}$ and $\boldsymbol{\lambda}_k^* \geq 0$, and we have $y_k^0(\mathbf{x}^*) + y_k(\mathbf{x}^*)^\top \boldsymbol{\mu} \geq 0$ according to the assumption $\min_{k=1}^K \{y_k^0(\mathbf{x}^*) + y_k(\mathbf{x}^*)^\top \boldsymbol{\mu}\} \geq 0$. Thus, we choose good values of k_i for any $i \in \{0, 1, 2, \dots, P\}$ to ensure (36) to be a PSD matrix and accordingly will make (35) hold.

Next, by Lemma 1, in order for (36) to be a PSD, we equivalently require

$$\begin{bmatrix} s_i(s^* - y_k^0(\mathbf{x}^*) - \boldsymbol{\lambda}_k^{*\top} \mathbf{b} - y_k(\mathbf{x}^*)^\top \boldsymbol{\mu} + \boldsymbol{\lambda}_k^{*\top} \mathbf{A} \boldsymbol{\mu}) & \frac{1}{2} \left(\mathbf{q}_i^* + \left(\mathbf{U}_{m \times m_i} \boldsymbol{\Lambda}_{m_i}^{\frac{1}{2}} \right)^\top (\mathbf{A}^\top \boldsymbol{\lambda}_k^* - y_k(\mathbf{x}^*)) \right)^\top \\ \frac{1}{2} \left(\mathbf{q}_i^* + \left(\mathbf{U}_{m \times m_i} \boldsymbol{\Lambda}_{m_i}^{\frac{1}{2}} \right)^\top (\mathbf{A}^\top \boldsymbol{\lambda}_k^* - y_k(\mathbf{x}^*)) \right) & k_i \mathbf{Q}_i^* \end{bmatrix} \succeq 0, \quad \forall k \in [K], i \in [P], \quad (37)$$

with $\sum_{i=1}^P s_i = k_0$. Constraints (37) can be satisfied by allowing $s_i \times k_i \geq 1$ for any $i \in [P]$ due to (4b). Then, we let $k_0 = k_1 = \dots = k_P$ and $s_i \times k_i = 1$ for any $i \in [P]$, leading to $k_0 = k_1 = \dots = k_P = \sqrt{P}$.

Finally, we have $UB_M^* \geq Z_M^*(m)$ by Theorem 4. Meanwhile, as Problem (20) is a minimization problem, UB_M^* is no larger than the objective value corresponding to our constructed feasible solution $(\bar{\mathbf{x}}, \bar{s}, \bar{\mathbf{q}}, \bar{\mathbf{Q}}', \bar{\boldsymbol{\lambda}}_k \forall k \in [K])$. That is, we have

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

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

$$0 \leq UB_M^* - Z_M^*(m) \leq (\sqrt{P} - 1) Z_M^*(m). \quad \square$$

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

EXAMPLE 1. Suppose that $\boldsymbol{\mu} = \mathbf{0}$, $\mathcal{S} = \mathbb{R}^m$, $\gamma_1 = +\infty$, and $\gamma_2 = 1$. With $f(\mathbf{x}, \mathbf{U} \boldsymbol{\Lambda}^{\frac{1}{2}} \boldsymbol{\xi}_1 + \boldsymbol{\mu}) = |\mathbf{x}^\top \boldsymbol{\xi}_1|$, Problem (4) can be recast as the following SDP formulation:

$$\begin{aligned} Z_M^*(m) &= \min_{\mathbf{x}, s, \mathbf{Q}} s + \mathbf{I}_m \bullet \mathbf{Q} \\ \text{s.t.} \quad & \begin{bmatrix} s & \frac{\mathbf{x}^\top}{2} \\ \frac{\mathbf{x}}{2} & \mathbf{Q} \end{bmatrix} \succeq 0, \quad \begin{bmatrix} s & \frac{-\mathbf{x}^\top}{2} \\ \frac{-\mathbf{x}}{2} & \mathbf{Q} \end{bmatrix} \succeq 0, \quad \mathbf{x} \in \mathcal{X}. \end{aligned} \quad (38)$$

For fixed $\mathbf{x} \in \mathcal{X}$, optimizing over the remaining decision variables in (38) yields $\mathbf{Q}^* = \frac{\mathbf{x}\mathbf{x}^\top}{4s}$, and $s^* = \frac{\sqrt{\mathbf{x}^\top \mathbf{x}}}{2}$ with objective value $\sqrt{\mathbf{x}^\top \mathbf{x}}$. By Proposition 5, we obtain an upper bound by considering

$$UB_M^* = \min_{\mathbf{x}, s, \hat{\mathbf{Q}}} s + \sum_{i=1}^P \mathbf{I}_i \bullet \mathbf{Q}_i \quad (39)$$

$$\text{s.t. } \begin{bmatrix} s & \frac{\mathbf{x}^\top}{2} \\ \frac{\mathbf{x}}{2} & \text{diag}(\mathbf{Q}_i) \end{bmatrix} \succeq 0, \begin{bmatrix} s & \frac{-\mathbf{x}^\top}{2} \\ \frac{-\mathbf{x}}{2} & \text{diag}(\mathbf{Q}_i) \end{bmatrix} \succeq 0, \mathbf{x} \in \mathcal{X},$$

where $\hat{\mathbf{Q}} = \{\mathbf{Q}_1, \dots, \mathbf{Q}_P\}$ and $\text{diag}(\mathbf{Q}_i)$ is a block diagonal matrix consisting of $\mathbf{Q}_1, \dots, \mathbf{Q}_P$. For fixed $\mathbf{x} \in \mathcal{X}$, where $\mathbf{x} = (\mathbf{x}_1^\top, \dots, \mathbf{x}_P^\top)^\top$ and $\mathbf{x}_i \in \mathbb{R}^{m_i}$ for all $i \in [P]$, optimizing over the remaining decision variables in Problem (39) yields $\mathbf{Q}_i^* = \frac{\mathbf{x}_i \mathbf{x}_i^\top}{4s_i}$ for all $i \in [P]$, and $s^* = \sum_{i=1}^P \frac{\sqrt{\mathbf{x}_i^\top \mathbf{x}_i}}{2}$ with objective value $\sum_{i=1}^P \sqrt{\mathbf{x}_i^\top \mathbf{x}_i}$. Now we let $m_i = \frac{m}{P}$ for any $i \in [P]$ and $\mathcal{X} = \{\mathbf{x} \in \mathbb{R}^m | x_i \geq 1, \forall i \in [m]\}$. It follows that $Z_M^*(m) = \sqrt{m}$ and $UB_M^* = P\sqrt{\frac{m}{P}}$, and so the relative gap between $Z_M^*(m)$ 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.

REMARK 3. The theoretical error bound $\sqrt{P} - 1$ in Proposition 7, albeit achievable in the above example, is usually conservative because it needs to hold valid for arbitrary (including many pathetic and worst-case) DRO problem instances. For instance, if $P > 1$, $\gamma_1 > 0$, and $\|q^*\|_2 > 0$, then the theoretical bound cannot be attained. The above example is comparable to the worst-case distribution induced by solving a DRO problem. Nevertheless, the inner approximation (20) leads to much better computational optimality gaps under various instances, which will be shown in our numerical experiments in Section 4.

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

$$\min_{\mathbf{x} \in \mathcal{X}} \max_{\mathbb{P} \in \mathcal{D}_{C1}} \mathbb{E}_{\mathbb{P}} [f(\mathbf{x}, \boldsymbol{\xi})], \quad (\text{DRO-C})$$

where

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

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

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

where π denotes a joint distribution of $\boldsymbol{\xi}$ and $\hat{\boldsymbol{\xi}}$ with marginals \mathbb{P} and \mathbb{P}_0 , respectively. Intuitively, $W(\mathbb{P}, \mathbb{P}_0)$ represents the minimum expected distance between $\boldsymbol{\xi}$ and $\hat{\boldsymbol{\xi}}$ over all possible joint distributions π . It has been shown that, as $N \rightarrow \infty$, \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}_{C_1} 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}_{C_1} . The larger radius R_0 is, \mathcal{D}_{C_1} 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 $f(\mathbf{x}, \boldsymbol{\xi})$ is piecewise linear convex in $\boldsymbol{\xi}$, i.e., $f(\mathbf{x}, \boldsymbol{\xi}) = \max_{k=1}^K \{y_k^0(\mathbf{x}) + y_k(\mathbf{x})^\top \boldsymbol{\xi}\}$ with both $y_k(\mathbf{x}) = (y_k^1(\mathbf{x}), \dots, y_k^m(\mathbf{x}))^\top$ and $y_k^0(\mathbf{x})$ affine in \mathbf{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:*

$$\begin{aligned} Z_C^*(m) := & \min_{\mathbf{x}, \lambda, \mathbf{Q}, \boldsymbol{\zeta}, \hat{\boldsymbol{\gamma}}} \lambda R_0 + \gamma_2 \boldsymbol{\Sigma} \bullet \mathbf{Q} + \frac{1}{N} \sum_{i=1}^N y_i & (40a) \\ \text{s.t.} & \begin{bmatrix} \mathbf{Q} & \frac{1}{2}(-y_k(\mathbf{x}) + \boldsymbol{\zeta}^i - 2\mathbf{Q}\boldsymbol{\mu}) \\ \frac{1}{2}(-y_k(\mathbf{x}) + \boldsymbol{\zeta}^i - 2\mathbf{Q}\boldsymbol{\mu})^\top & y_i - y_k^0(\mathbf{x}) - \boldsymbol{\zeta}^{i\top} \hat{\boldsymbol{\xi}}^i + \boldsymbol{\mu}^\top \mathbf{Q}\boldsymbol{\mu} \end{bmatrix} \succeq 0, \forall i \in [N], \forall k \in [K], & (40b) \\ & \lambda \in \mathbb{R}_+, \mathbf{x} \in \mathcal{X}, \|\boldsymbol{\zeta}^i\|_* \leq \lambda, \forall i \in [N], \end{aligned}$$

where $\mathbf{Q} \in \mathbb{R}^{m \times m}$, $\boldsymbol{\zeta}^i \in \mathbb{R}^m$ for any $i \in [N]$, $\hat{\boldsymbol{\zeta}} = \{\boldsymbol{\zeta}^1, \dots, \boldsymbol{\zeta}^N\}$, and $\hat{\boldsymbol{\gamma}} = \{y_1, \dots, y_N\}$.

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

As discussed in Section 2, Problem (40) 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 (40) (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 Σ , we first reformulate (DRO-C) as

$$\min_{\mathbf{x} \in \mathcal{X}} \max_{\mathbb{P}_I \in \mathcal{D}_{C2}} \mathbb{E}_{\mathbb{P}_I} \left[f \left(\mathbf{x}, \mathbf{U} \Lambda^{\frac{1}{2}} \boldsymbol{\xi}_I + \boldsymbol{\mu} \right) \right], \quad (41)$$

where

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

under the condition that $f(\mathbf{x}, \mathbf{U} \Lambda^{\frac{1}{2}} \boldsymbol{\xi}_I + \boldsymbol{\mu})$ is \mathbb{P}_I -integrable for any $\mathbb{P}_I \in \mathcal{D}_{C2}$ and $\mathcal{S}_I := \{\boldsymbol{\xi}_I \in \mathbb{R}^m : \mathbf{U} \Lambda^{\frac{1}{2}} \boldsymbol{\xi}_I + \boldsymbol{\mu} \in \mathcal{S}\}$. Next, by the approximation of $\boldsymbol{\xi}$ in (5) due to PCA, we outer approximate (41) as the following problem:

$$\min_{\mathbf{x} \in \mathcal{X}} \max_{\mathbb{P}_r \in \mathcal{D}_{C3}} \mathbb{E}_{\mathbb{P}_r} \left[f \left(\mathbf{x}, \mathbf{U}_{m \times m_1} \Lambda_{m_1}^{\frac{1}{2}} \boldsymbol{\xi}_r + \boldsymbol{\mu} \right) \right], \quad (42a)$$

where

$$\mathcal{D}_{C3}(\mathcal{S}_r, \boldsymbol{\mu}, \gamma_2, \mathbb{P}_0, R_0) = \left\{ \mathbb{P}_r \left| \begin{array}{l} \mathbb{E}_{\mathbb{P}_r} [\boldsymbol{\xi}_r \boldsymbol{\xi}_r^\top] \preceq \gamma_2 \mathbf{I}_{m_1} \\ \exists \pi : \int_{\mathcal{S}_r^2} \left\| \mathbf{U}_{m \times m_1} \Lambda_{m_1}^{\frac{1}{2}} \boldsymbol{\xi}_r + \boldsymbol{\mu} - \hat{\boldsymbol{\xi}} \right\|_1 \pi \left(\mathbf{U}_{m \times m_1} \Lambda_{m_1}^{\frac{1}{2}} \boldsymbol{\xi}_r + \boldsymbol{\mu}, \hat{\boldsymbol{\xi}} \right) \leq R_0 \end{array} \right. \right\} \quad (42b)$$

with

$$\mathcal{S}_r := \left\{ \boldsymbol{\xi}_r \in \mathbb{R}^{m_1} : \mathbf{U}_{m \times m_1} \Lambda_{m_1}^{\frac{1}{2}} \boldsymbol{\xi}_r + \boldsymbol{\mu} \in \mathcal{S} \right\}. \quad (42c)$$

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 (42) has the same optimal value as the following SDP formulation:*

$$Z_C^*(m_1) := \min_{\mathbf{x}, \lambda, \mathbf{Q}_r, \boldsymbol{\zeta}, \hat{\mathbf{y}}} \lambda R_0 + \gamma_2 \mathbf{I}_{m_1} \bullet \mathbf{Q}_r + \frac{1}{N} \sum_{i=1}^N y_i \quad (43a)$$

$$\text{s.t.} \quad \begin{bmatrix} \mathbf{Q}_r & \frac{1}{2} \left((-y_k(\mathbf{x}) + \boldsymbol{\zeta}^i)^\top \mathbf{U}_{m \times m_1} \Lambda_{m_1}^{\frac{1}{2}} \right)^\top \\ \frac{1}{2} (-y_k(\mathbf{x}) + \boldsymbol{\zeta}^i)^\top \mathbf{U}_{m \times m_1} \Lambda_{m_1}^{\frac{1}{2}} & y_i - y_k(\mathbf{x})^\top \boldsymbol{\mu} - y_k^0(\mathbf{x}) + (\boldsymbol{\mu} - \hat{\boldsymbol{\xi}}^i)^\top \boldsymbol{\zeta}^i \end{bmatrix} \succeq 0, \quad \forall i \in [N], \forall k \in [K], \quad (43b)$$

$$\lambda \in \mathbb{R}_+, \mathbf{x} \in \mathcal{X}, \|\boldsymbol{\zeta}^i\|_* \leq \lambda, \forall i \in [N], \quad (43c)$$

where $\mathbf{Q}_r \in \mathbb{R}^{m_1 \times m_1}$, $\boldsymbol{\zeta}^i \in \mathbb{R}^m$ for any $i \in [N]$, $\hat{\boldsymbol{\zeta}} = \{\boldsymbol{\zeta}^1, \dots, \boldsymbol{\zeta}^N\}$, and $\hat{\mathbf{y}} = \{y_1, \dots, y_N\}$. Furthermore, we have the following: (i) Problem (43) provides a lower bound for the optimal value of (DRO-C); (ii) the optimal value of Problem (43) is nondecreasing in m_1 ; and (iii) if $m_1 = m$, then (DRO-C) and (43) have the same optimal value.

Proof. See Online Supplement B.1 for the detailed proof. \square

We show the quality of the outer approximation (43) in the following proposition.

PROPOSITION 9. *It holds that*

$$0 \leq Z^*_C(m) - Z^*_C(m_1) \leq \sqrt{\frac{\gamma_2}{N}} \sum_{i=1}^N \sum_{k=1}^K \sqrt{\mathbf{L}_{m-m_1}^{ik} (\mathbf{L}_{m-m_1}^{ik})^\top}, \quad (44)$$

where $\mathbf{L}_{m-m_1}^{ik} = (-y_k(\mathbf{x}^*) + \boldsymbol{\zeta}^{i*})^\top \mathbf{U}_{m \times (m-m_1)} (\boldsymbol{\Lambda}^{m-m_1})^{\frac{1}{2}}$, \mathbf{x}^* and $\boldsymbol{\zeta}^{i*}$ ($\forall i \in [N]$) denote an optimal solution of Problem (43), 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 Online Supplement B.2 for the detailed proof. \square

3.2. Upper Bound

We further inner approximate Problem (40), leading to an upper bound, by splitting random vector $\boldsymbol{\xi}_I$ into P pieces in the second-moment constraint in \mathcal{D}_{C2} so that $\boldsymbol{\xi}_I = (\boldsymbol{\xi}_{I_1}^\top, \boldsymbol{\xi}_{I_2}^\top, \dots, \boldsymbol{\xi}_{I_P}^\top)^\top$, where $\boldsymbol{\xi}_{I_j} \in \mathbb{R}^{m_j}$, $\forall j \in [P]$, and $\sum_{j=1}^P m_j = m$. This gives rise to an inner approximation

$$\min_{\mathbf{x} \in \mathcal{X}} \max_{\mathbb{P}_I \in \mathcal{D}_{C4}} \mathbb{E}_{\mathbb{P}_I} \left[f \left(\mathbf{x}, \mathbf{U} \boldsymbol{\Lambda}^{\frac{1}{2}} \boldsymbol{\xi}_I + \boldsymbol{\mu} \right) \right], \quad (45)$$

where

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

THEOREM 6. *Under Assumption 2, Problem (45) has the same optimal value as the following SDP formulation:*

$$\begin{aligned} UB_C^* := & \min_{\substack{\mathbf{x}, \lambda, \hat{\mathbf{Q}}, \\ \hat{\boldsymbol{\zeta}}, \hat{\mathbf{y}}, \hat{\mathbf{s}}}} \lambda R_0 + \gamma_2 \sum_{j=1}^P \mathbf{I}_{m_j} \bullet \mathbf{Q}_j + \frac{1}{N} \sum_{i=1}^N y_i & (46) \\ \text{s.t.} & \left[\begin{array}{cc} \mathbf{Q}_j & \frac{1}{2} \left((-y_k(\mathbf{x}) + \boldsymbol{\zeta}^i)^\top \mathbf{U}_{m \times m_j} \boldsymbol{\Lambda}_{m_j}^{\frac{1}{2}} \right)^\top \\ \frac{1}{2} (-y_k(\mathbf{x}) + \boldsymbol{\zeta}^i)^\top \mathbf{U}_{m \times m_j} \boldsymbol{\Lambda}_{m_j}^{\frac{1}{2}} & s_{jik} \end{array} \right] \succeq 0, \forall j \in [P], \forall i \in [N], \\ & \forall k \in [K], \\ & \sum_{j=1}^P s_{jik} = y_i - y_k(\mathbf{x})^\top \boldsymbol{\mu} - y_k^0(\mathbf{x}) + \left(\boldsymbol{\mu} - \hat{\boldsymbol{\xi}}^i \right)^\top \boldsymbol{\zeta}^i, \forall i \in [N], \forall k \in [K], \\ & \lambda \in \mathbb{R}_+, \mathbf{x} \in \mathcal{X}, \|\boldsymbol{\zeta}^i\|_* \leq \lambda, \forall i \in [N], \end{aligned}$$

where $\mathbf{Q}_j \in \mathbb{R}^{m_j \times m_j}$, $\boldsymbol{\zeta}^i \in \mathbb{R}^m$ for any $i \in [N]$, $\hat{\mathbf{Q}} = \{\mathbf{Q}_1, \dots, \mathbf{Q}_P\}$, $\hat{\boldsymbol{\zeta}} = \{\boldsymbol{\zeta}^1, \dots, \boldsymbol{\zeta}^N\}$, $\hat{\mathbf{y}} = \{y_1, \dots, y_N\}$, and $\hat{\mathbf{s}} = \{s_{ijk}, \forall i \in [N], \forall j \in [P], \forall k \in [K]\}$. Furthermore, Problem (46) provides an upper bound for the optimal value of (DRO-C).

Proof. See Online Supplement B.3 for the detailed proof. \square

We show the quality of our derived inner approximation (46) in the following proposition.

PROPOSITION 10. *If $\min_{k=1}^K \{y_k^0(\mathbf{x}^*) + y_k(\mathbf{x}^*)^\top \boldsymbol{\mu}\} \geq 0$ and $\max_{i=1}^N \{(\boldsymbol{\mu} - \hat{\boldsymbol{\xi}}^i)^\top \boldsymbol{\zeta}^{i*}\} \leq 0$, where \mathbf{x}^* and $\boldsymbol{\zeta}^{i*}$ are optimal solutions of Problem (40), then the relative gap between the optimal values of Problems (46) and (40) is bounded from above by $\sqrt{P}-1$, i.e., $0 \leq UB_C^* - Z_C^*(m) \leq (\sqrt{P}-1)Z_C^*(m)$.*

Proof. See Online Supplement B.4 for the detailed proof. \square

REMARK 4. The theoretical upper bound in Proposition 10 is achievable. Indeed, when we enlarge R_0 enough such that \mathcal{D}_{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. Meanwhile, similar to our notes in Remark 3, the theoretical bound in Proposition 10 is also conservative and only achievable in pathetic problem instances. The quality of the inner approximation (46) is usually much more optimistic due to the corresponding numerical results in Section 4.

Before closing this section, we note that the first-order moment constraint, albeit important to reduce solution conservatism, is not included in the combined ambiguity set of (DRO-C). We consider additionally including the first-order moment constraint, leading to a DRO problem with a combined ambiguity set consisting of Wasserstein distance and first- and second-order moment information; see Problem (DRO-C2) in Online Supplement B.5. We correspondingly derive the inner and outer approximations of (DRO-C2), and conduct computational experiments to compare their performance with the approximations of (DRO-C). Based on the results, we observe that adding the first-order moment information does further reduce the conservatism of optimal solutions for both (DRO-C) and its approximations, though such reduction is not significant. This is because the second-order moment information together with the Wasserstein information has already (partially) implied the first-order moment information. Such insight is also explained in Gao and Kleywegt (2017). Due to this, we only keep (DRO-C) in the main body.

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 $\sum_{j \in [n]} d_j \leq m$. 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:

$$\min_{\mathbf{x}, \mathbf{z}} \sum_{i=1}^m c_i x_i + \sum_{i=1}^m \sum_{j=1}^n \xi_{ij} z_{ij} \quad (47a)$$

$$\text{s.t.} \quad \sum_{i=1}^m z_{ij} = d_j, \quad \forall j \in [n], \quad (47b)$$

$$\sum_{j=1}^n z_{ij} = x_i, \quad \forall i \in [m], \quad (47c)$$

$$0 \leq x_i \leq 1, \quad \forall i \in [m], \quad (47d)$$

$$z_{ij} \geq 0, \quad \forall i \in [m], \quad \forall j \in [n]. \quad (47e)$$

Now we derive the DRO counterpart of Problem (47). 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} . The decision \mathbf{x} is decided before the realization of randomness and \mathbf{z} is made as a recourse to specific realizations (Bertsimas et al. 2010). This leads to a two-stage DRO counterpart

$$\min_{\mathbf{x}} \left\{ \mathbf{c}^\top \mathbf{x} + \max_{\mathbb{P} \in \mathcal{D}} \mathbb{E}_{\mathbb{P}} [\mathcal{U}(\mathcal{Q}(\mathbf{z}, \boldsymbol{\xi}))] : (47d) \right\}, \quad (48)$$

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

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

where $\mathcal{Q}(\mathbf{z}, \boldsymbol{\xi}) = \min_{\mathbf{x}} \{\mathbf{z}^\top \boldsymbol{\xi} : (47b), (47c), (47e)\}$. We can apply the proposed inner and outer approximations (i.e., Problems (8), (43), (17), (20), and (46)) to approximate Problem (48) in the context of production-transportation problem, with the details provided in Online Supplement 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 $[0.5\frac{m}{n}, \frac{m}{n}]$, respectively. We consider disutility function $\mathcal{U}(x) = 0.25(e^{2x} - 1)$ 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 $\mathbf{x} = (x_i, i \in [n])^\top$ to minimize the total loss described as follows:

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

where \mathbf{c} represents the wholesale price, \mathbf{v} represents the retail price, \mathbf{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:

$$\min_{\mathbf{x} \geq 0} \max_{\mathbb{P} \in \mathcal{D}} \mathbb{E}_{\mathbb{P}} \left[\max \left\{ (\mathbf{c} - \mathbf{v})^\top \mathbf{x}, (\mathbf{c} - \mathbf{g})^\top \mathbf{x} + (\mathbf{g} - \mathbf{v})^\top \boldsymbol{\xi} \right\} \right]. \quad (49)$$

Note that the procedure of applying the proposed inner and outer approximations and the theoretical bounds to Problem (49) is similar to that for Problem (48) 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 salvage 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 (or even feasibility in most cases) by existing methods in reasonable time. We further investigate the benefits of choosing the components with the largest variability (i.e., leading components) in the PCA approach (as compared to randomly choosing components) and also perform sensitivity analyses with respect to several given parameters.

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 ξ is equally split with $m_1 = m_2 = \dots = m_P$; that is, here we use a homogeneous division of the random vector. This is because such homogeneous division eventually leads to a set of SDP constraints, with all of them having the smallest possible matrices simultaneously, in our proposed formulations. It helps improve the computational performance of the corresponding inner approximation. In addition, when assigning all the components of ξ to pieces, we consider the non-increasing order of the components based on their variance. Specifically, the first piece includes the first m_1 components with the largest variance, the next m_2 components are assigned to the second pieces, and so on. The reason is that in our approximation, the correlations among different pieces are dropped, but we would like to capture the correlations among those important components within each piece. Third, we consider different supports, i.e., $\mathcal{S} \in \{[-2\sigma, 2\sigma], [-3\sigma, 3\sigma], [-4\sigma, 4\sigma]\}$, where σ represents the sample standard deviation of random vector ξ . 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 Wasserstein radius R_0 is set as 30 for the production-transportation problem and as 700 for the newsvendor problem. We will perform sensitivity analyses with respect to other different values of γ_1 , γ_2 , and R_0 . For each combination of the above three variants, we randomly generate five instances¹ and report the average results over them.

In the following Sections 4.2.2 - 4.2.6, 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 (22), we use column “Gap2” in percentage to represent the value of theoretical bound. Whenever needed, we use “LB” and “UB” to denote the lower and upper bounds, respectively. Note that the percentage of ξ '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 ξ , 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.

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 (%)												
(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 (%)												
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

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 reduce the computational time significantly and maintain very high solution quality. For instance, even with a 90% reduction in the dimension of uncertainty space, our approximation solution can give an objective value at around 2% of the original optimal value, as shown in the column “Gap”. 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

Table 3 Lower bound (43) on the production-transportation problem

$\frac{m_1}{m}$ (%)		100%			75%			50%			25%			10%		
Size (m, n)	Orig. (secs)	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 (43) on the newsvendor problem

$\frac{m_1}{m}$ (%)		100%			75%			50%			25%			10%		
Size (n)	Orig. (secs)	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

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.

From Tables 3 and 4, where the column ‘‘Gap2’’ represents the relative gap induced by the theoretical upper bound in (44), we have the similar observations as from Tables 1 and 2. In addition, a comparison among Tables 1 - 4 shows that (i) 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 (ii) theoretical bound (44) 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 (17) in Tables 5 - 6 and report that of (20) and (46) in Tables 7 - 10.

Table 5 Upper bound (17) on the production-transportation problem

$\frac{m_1}{m}$ (%)		100%			75%			50%			25%			10%			
Size (m, n)	Support	Orig. (secs)	Time (secs)	Gap (%)	Gap2 (%)												
(5, 20)	$[-2\sigma, 2\sigma]$	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
	$[-3\sigma, 3\sigma]$	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
	$[-4\sigma, 4\sigma]$	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)	$[-2\sigma, 2\sigma]$	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
	$[-3\sigma, 3\sigma]$	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
	$[-4\sigma, 4\sigma]$	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)	$[-2\sigma, 2\sigma]$	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
	$[-3\sigma, 3\sigma]$	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
	$[-4\sigma, 4\sigma]$	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

From Tables 5 - 6, where the column ‘‘Gap2’’ represents the relative gap induced by the theoretical upper bound in (22), we have similar observations as from Tables 1 and 2. In addition, our approximation (17) performs better when solving the production-transportation problem as compared to solving the newsvendor problem. Meanwhile, the values of Gap2 are always larger than those of Gap, implying that theoretical bound (22) is valid.

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

Table 6 Upper bound (17) 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	$[-2\sigma, 2\sigma]$	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
	$[-3\sigma, 3\sigma]$	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
	$[-4\sigma, 4\sigma]$	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	$[-2\sigma, 2\sigma]$	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
	$[-3\sigma, 3\sigma]$	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
	$[-4\sigma, 4\sigma]$	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	$[-2\sigma, 2\sigma]$	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
	$[-3\sigma, 3\sigma]$	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
	$[-4\sigma, 4\sigma]$	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

Table 7 Upper bound (20) on the production-transportation problem

P		2			4			5			
Size (m, n)	Support	Orig. (secs)	Time (secs)	Gap (%)	Gap2 (%)	Time (secs)	Gap (%)	Gap2 (%)	Time (secs)	Gap (%)	Gap2 (%)
(5, 20)	$[-2\sigma, 2\sigma]$	65.9	11.3	0.18	41.42	5.4	0.52	100	4.4	0.71	123.61
	$[-3\sigma, 3\sigma]$	69.0	10.7	0.07	41.42	5.7	0.35	100	4.5	0.54	123.61
	$[-4\sigma, 4\sigma]$	67.1	11.4	0.10	41.42	5.4	0.39	100	4.3	0.51	123.61
(4, 40)	$[-2\sigma, 2\sigma]$	521.9	79.4	0.15	41.42	32.8	0.40	100	22.9	0.53	123.61
	$[-3\sigma, 3\sigma]$	566.2	86.8	0.13	41.42	35.5	0.43	100	24.7	0.54	123.61
	$[-4\sigma, 4\sigma]$	542.9	83.0	0.17	41.42	33.9	0.47	100	25.1	0.57	123.61
(8, 25)	$[-2\sigma, 2\sigma]$	1531.3	205.2	0.02	41.42	55.8	0.07	100	61.5	0.11	123.61
	$[-3\sigma, 3\sigma]$	1594.0	216.8	0.04	41.42	58.4	0.13	100	67.8	0.18	123.61
	$[-4\sigma, 4\sigma]$	1539.1	207.0	0.04	41.42	56.8	0.16	100	67.5	0.24	123.61

Table 8 Upper bound (20) on the newsvendor problem

P		2			4			5			
Size (n)	Support	Orig. (secs)	Time (secs)	Gap (%)	Gap2 (%)	Time (secs)	Gap (%)	Gap2 (%)	Time (secs)	Gap (%)	Gap2 (%)
100	$[-2\sigma, 2\sigma]$	20.1	1.3	1.11	41.42	0.5	3.05	100	0.5	3.86	123.61
	$[-3\sigma, 3\sigma]$	17.7	1.3	1.26	41.42	0.5	3.20	100	0.5	4.07	123.61
	$[-4\sigma, 4\sigma]$	18.1	1.3	1.11	41.42	0.5	2.99	100	0.5	3.73	123.61
160	$[-2\sigma, 2\sigma]$	162.3	7.3	0.91	41.42	1.3	2.52	100	0.9	3.25	123.61
	$[-3\sigma, 3\sigma]$	169.2	7.4	1.10	41.42	1.2	2.88	100	0.9	3.57	123.61
	$[-4\sigma, 4\sigma]$	169.6	7.9	0.95	41.42	1.2	2.62	100	0.9	3.34	123.61
200	$[-2\sigma, 2\sigma]$	521.6	25.7	0.84	41.42	2.6	2.13	100	1.7	2.62	123.61
	$[-3\sigma, 3\sigma]$	493.4	21.7	0.84	41.42	2.9	2.28	100	1.7	2.86	123.61
	$[-4\sigma, 4\sigma]$	518.1	23.9	0.84	41.42	2.8	2.18	100	1.7	2.81	123.61

pieces as a tool to balance between solution quality and computational time. Second, the quality of upper bound (20) 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. Third, as mentioned in Remarks 3 and 4, although the theoretical error bound $\sqrt{P} - 1$ is very conservative, as represented in the column “Gap2”, the computational optimality gap perform very well, as represented in the column “Gap”. 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.

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 (or even feasibility in most cases) by existing methods in reasonable time. In this case, it can be very useful to quickly find a feasible solution with small optimality

Table 9 Upper bound (46) on the production-transportation problem

P		2			4			5		
Size (m, n)	Orig. (secs)	Time (secs)	Gap (%)	Gap2 (%)	Time (secs)	Gap (%)	Gap2 (%)	Time (secs)	Gap (%)	Gap2 (%)
(5, 20)	716.4	179.3	1.78	41.42	121.4	4.90	100	114.2	6.26	123.61
(4, 40)	5050.3	1028.3	2.06	41.42	521.8	4.56	100	488.4	5.50	123.61
(8, 25)	11383.6	2281.3	0.93	41.42	1017.6	3.25	100	1010.7	3.96	123.61

Table 10 Upper bound (46) on the newsvendor problem

P		2			4			5		
Size (m, n)	Orig. (secs)	Time (secs)	Gap (%)	Gap2 (%)	Time (secs)	Gap (%)	Gap2 (%)	Time (secs)	Gap (%)	Gap2 (%)
100	129.5	55.8	1.66	41.42	37.1	4.35	100	34.1	5.46	123.61
160	798.6	275.7	1.13	41.42	132.5	3.18	100	120.5	4.17	123.61
200	2038.4	615.7	0.93	41.42	251.4	2.70	100	225.3	3.49	123.61

gap, evaluated through an interval that includes the (unknown) optimal value of this model. In this section, we construct such intervals with the help of our proposed inner and outer approximations.

Table 11 Intervals on the production-transportation problem

Size (m, n)	Support	[LB, UB]		[(8),(17)]		[(8),(20)]		
		$(\frac{m_1}{m}\%, P)$	Orig. (secs)	Itv-Time (secs)	Itv-Gap (%)	Itv-Time (secs)	Itv-Gap (%)	
(6, 50)	[-2 σ , 2 σ]	(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	
		(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	
	[-3 σ , 3 σ]	(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	
		[-4 σ , 4 σ]	(25%, 5)	-	265.7	4.94	604.7	2.78
			(50%, 4)	-	3375.2	4.53	2181.1	2.47
			(75%, 2)	-	-	-	-	-
(25%, 5)	-		302.6	10.28	629.1	2.80		
(50%, 4)	-		3697.8	8.21	2667.7	2.21		
(75%, 2)	-		-	-	-	-		
(8, 50)	[-3 σ , 3 σ]	(25%, 5)	-	299.4	15.86	650	2.86	
		(50%, 4)	-	4129.7	14.08	2596.4	2.41	
		(75%, 2)	-	-	-	-	-	
	[-4 σ , 4 σ]	(25%, 5)	-	-	-	-	-	
		(50%, 4)	-	-	-	-	-	
		(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 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 $\frac{UB-LB}{UB} \times 100\%$, 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 ‘‘ $(\frac{m_1}{m}\%, P)$ ’’, $\frac{m_1}{m}\%$ is used to define the lower bound problem and the PCA based upper bound problems, i.e., Problems

Table 12 Intervals on the newsvendor problem

Size (n)	[LB, UB]		Orig. (secs)	[(8), (17)]		[(8), (20)]		
	Support	$(\frac{m_1}{m}\%, P)$		Itv-Time (secs)	Itv-Gap (%)	Itv-Time (secs)	Itv-Gap (%)	
300	[-2 σ , 2 σ]	(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 σ , 3 σ]	(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 σ , 4 σ]	(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 σ , 2 σ]	(25%, 5)	-	36	28.82	35.7	2.38
			(50%, 4)	-	958.2	15.86	423.5	1.82
			(75%, 2)	-	-	-	-	-
[-3 σ , 3 σ]		(25%, 5)	-	34.7	58.55	35.4	2.48	
		(50%, 4)	-	1093.7	20.06	525.8	1.74	
		(75%, 2)	-	-	-	-	-	
[-4 σ , 4 σ]		(25%, 5)	-	38.5	80.60	39.1	2.38	
		(50%, 4)	-	990.7	23.66	430.4	1.55	
		(75%, 2)	-	-	-	-	-	

Table 13 Intervals on the production-transportation problem

Size (m, n)	[LB, UB]		[(43), (46)]	
	$(\frac{m_1}{m}\%, P)$	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

Size (n)	[LB, UB]		[(43), (46)]	
	$(\frac{m_1}{m}\%, P)$	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)	-	-	-

(8), (17), and (43), while P is used to define the vector splitting based upper bound problems, i.e., Problems (20) and (46).

From Tables 11 - 14, we have the following observations. First, when the optimal solution cannot be found for most original cases, an interval that includes the unknown optimal value can be constructed fast by using our proposed inner and outer approximations. Such interval is relatively tight for most cases, as demonstrated by the interval [(8),(20)] in Tables 11 - 12 and the interval [(43), (46)] in Tables 13 - 14. Second, when m_1 increases and P decreases, such interval can be tighter 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 due to (20) are much tighter than the PCA based upper bounds due to (17) and take less computational time.

4.2.5. Benefits of Choosing Leading Components When using the PCA approach to derive our inner and outer approximations, we choose the leading components and relax the rest

of ones. To show the benefits of choosing the leading components in guaranteeing solution quality, we take our derived lower bound (8) together with its corresponding theoretical error bound (9) as an example and compare its performance of solving the newsvendor problem under two cases: (a) choose m_1 arbitrary components; (b) choose m_1 leading components.

If we choose m_1 arbitrary components, then we can follow the same process in Section 2.1 to derive the corresponding outer approximation (denoted by (8') here for short) and theoretical error bound (denoted by (9') here for short). Following the table header of Tables 1 and 2, which show the performance of lower bound (8), we introduce subscripts a and b to use “Gap_a” to denote the relative gap in percentage between the lower bound (8') and the original optimal value, “Gap2_a” to denote the value of theoretical error bound (9') in percentage, use “Gap_b” to denote the relative gap in percentage between the lower bound (8) and the original optimal value, and “Gap2_b” to denote the value of theoretical error bound (9) in percentage. We further use “Det_c” to denote the relative gap in percentage between “Gap_a” and “Gap_b”, i.e., $(\text{Gap}_a - \text{Gap}_b) / \text{Gap}_a \times 100\%$, and use “Det_t” to denote the relative gap in percentage between “Gap2_a” and “Gap2_b”, i.e., $(\text{Gap2}_a - \text{Gap2}_b) / \text{Gap2}_a \times 100\%$. We show the corresponding results in Table 15.

Table 15 Arbitrary vs. leading components on the newsvendor problem

Size (n)	$\frac{m_1}{m}$ (%) Support	100%		75%		50%		25%		10%	
		Det _c (%)	Det _t (%)								
100	$[-2\sigma, 2\sigma]$	0.00	0.00	98.80	88.04	93.84	69.68	78.47	40.56	53.72	16.42
	$[-3\sigma, 3\sigma]$	0.00	0.00	99.08	89.17	94.97	72.70	79.27	42.36	55.57	18.59
	$[-4\sigma, 4\sigma]$	0.00	0.00	99.05	89.25	93.65	69.51	74.69	36.59	59.68	20.72
160	$[-2\sigma, 2\sigma]$	0.00	0.00	99.09	89.56	94.00	69.93	75.59	37.00	52.88	16.41
	$[-3\sigma, 3\sigma]$	0.00	0.00	98.98	88.82	94.47	71.10	79.10	41.23	60.45	20.96
	$[-4\sigma, 4\sigma]$	0.00	0.00	98.79	87.66	92.97	67.85	77.78	40.25	58.43	19.59
200	$[-2\sigma, 2\sigma]$	0.00	0.00	98.85	88.02	93.40	68.60	80.09	42.79	57.00	18.68
	$[-3\sigma, 3\sigma]$	0.00	0.00	98.92	88.50	94.77	71.69	78.49	40.57	57.36	19.18
	$[-4\sigma, 4\sigma]$	0.00	0.00	99.03	88.92	94.88	72.10	81.28	44.06	57.67	19.04

From the table, we can observe that $\text{Gap}_a > \text{Gap}_b$ and $\text{Gap2}_a > \text{Gap2}_b$ for all the instances when $m_1 < m$. It means that choosing m_1 arbitrary components leads to a worse lower bound. For instance, for a newsvendor problem with size 100 and support $[-2\sigma, 2\sigma]$, if we arbitrarily choose 75% of the components, then the values of Gap_a and Gap2_a are 98.80% and 88.04% larger (i.e., worse) than those of Gap_b and Gap2_b , respectively, when we choose 75% of the components with the largest variability (leading components).

4.2.6. Sensitivity Analyses To better analyze the performance of our derived approximations more thoroughly, we conduct sensitivity analyses with respect to parameters γ_1 , γ_2 , and R_0 . We take lower bound (8) with support $[-3\sigma, 3\sigma]$ on the newsvendor problem as an example to conduct the sensitivity analyses with respect to γ_1 and γ_2 and summarize the results in Tables

Table 16 Sensitivity analysis for lower bound (8) on the newsvendor problem with respect to γ_1

$\frac{m_1}{m}$ (%)		100%			75%			50%			25%			10%			
Size (n)	γ_1	Orig. (secs)	Time (secs)	Gap (%)	Gap2 (%)	Time (secs)	Gap (%)	Gap2 (%)	Time (secs)	Gap (%)	Gap2 (%)	Time (secs)	Gap (%)	Gap2 (%)	Time (secs)	Gap (%)	Gap2 (%)
100	0.4	13.2	13.6	0.00	0.00	4.1	0.03	0.78	1.18	0.24	2.32	0.38	1.07	4.68	0.3	2.16	6.24
	0.6	13.7	13.5	0.00	0.00	4.0	0.03	0.78	1.1	0.25	2.33	0.4	1.10	4.69	0.3	2.23	6.25
	0.8	12.8	13.0	0.00	0.00	4.3	0.03	0.78	1.1	0.25	2.33	0.4	1.12	4.69	0.3	2.27	6.26
	1	18.8	19.5	0.00	0.00	5.7	0.03	0.78	1.4	0.25	2.33	0.4	1.13	4.70	0.3	2.29	6.26
	1.2	13.9	14.1	0.00	0.00	4.5	0.03	0.78	1.2	0.25	2.33	0.4	1.13	4.70	0.3	2.29	6.26
160	0.4	129.2	130.0	0.00	0.00	34.4	0.02	0.55	6.2	0.24	1.97	0.9	0.94	3.83	0.5	2.00	5.17
	0.6	136.7	138.5	0.00	0.00	34.8	0.02	0.55	6.4	0.25	1.97	0.9	0.97	3.83	0.4	2.07	5.18
	0.8	145.0	143.4	0.00	0.00	33.7	0.02	0.55	6.2	0.25	1.97	0.9	0.99	3.83	0.5	2.11	5.18
	1	163.6	164.5	0.00	0.00	46.4	0.02	0.55	9.1	0.25	1.98	1.0	0.99	3.83	0.4	2.12	5.18
	1.2	132.1	129.0	0.00	0.00	31.7	0.02	0.55	6.1	0.25	1.98	0.8	0.99	3.83	0.4	2.12	5.18
200	0.4	402.2	402.8	0.00	0.00	88.4	0.02	0.53	16.1	0.18	1.67	1.4	0.77	3.30	0.4	1.57	4.40
	0.6	435.1	434.9	0.00	0.00	90.6	0.02	0.53	15.2	0.19	1.67	1.3	0.80	3.30	0.4	1.62	4.40
	0.8	425.3	425.1	0.00	0.00	103.0	0.02	0.53	14.4	0.19	1.67	1.3	0.81	3.31	0.4	1.65	4.41
	1	483.0	483.5	0.00	0.00	113.9	0.02	0.53	21.5	0.19	1.67	1.8	0.81	3.31	0.4	1.65	4.41
	1.2	371.3	371.6	0.00	0.00	100.6	0.02	0.53	15.2	0.19	1.67	1.3	0.81	3.31	0.4	1.65	4.41

Table 17 Sensitivity analysis for lower bound (8) on the newsvendor problem with respect to γ_2

$\frac{m_1}{m}$ (%)		100%			75%			50%			25%			10%			
Size (n)	γ_2	Orig. (secs)	Time (secs)	Gap (%)	Gap2 (%)	Time (secs)	Gap (%)	Gap2 (%)	Time (secs)	Gap (%)	Gap2 (%)	Time (secs)	Gap (%)	Gap2 (%)	Time (secs)	Gap (%)	Gap2 (%)
100	1	12.2	12.0	0.00	0.00	3.9	0.01	0.42	1.1	0.17	1.54	0.4	0.83	3.25	0.3	1.72	4.36
	2	17.8	17.5	0.00	0.00	5.7	0.02	0.60	1.4	0.24	2.21	0.4	1.19	4.66	0.3	2.46	6.26
	3	13.3	13.4	0.00	0.00	4.04	0.02	0.74	1.1	0.29	2.73	0.4	1.45	5.77	0.3	3.01	7.76
	4	14.1	14.1	0.00	0.00	4.44	0.02	0.86	1.2	0.33	3.18	0.4	1.65	6.72	0.3	3.43	9.03
	5	13.8	13.7	0.00	0.00	4.4	0.03	0.97	1.2	0.37	3.58	0.4	1.83	7.57	0.3	3.80	10.17
160	1	114.0	113.3	0.00	0.00	28.8	0.01	0.38	5.4	0.14	1.29	0.7	0.64	2.66	0.3	1.34	3.61
	2	158.0	159.8	0.00	0.00	45.2	0.02	0.54	8.3	0.20	1.86	0.9	0.91	3.82	0.4	1.92	5.18
	3	140.3	140.5	0.00	0.00	32.7	0.02	0.67	6.0	0.24	2.29	0.8	1.11	4.72	0.4	2.34	6.40
	4	152.7	152.8	0.00	0.00	33.5	0.02	0.78	6.6	0.27	2.67	0.7	1.27	5.49	0.3	2.67	7.45
	5	142.2	142.8	0.00	0.00	38.8	0.03	0.88	7.4	0.30	3.00	0.7	1.40	6.18	0.4	2.95	8.38
200	1	396.4	398.2	0.00	0.00	103.2	0.01	0.36	15.5	0.11	1.06	1.4	0.52	2.26	0.4	1.12	3.09
	2	473.8	473.3	0.00	0.00	125.4	0.02	0.52	22.1	0.15	1.52	1.9	0.74	3.23	0.5	1.60	4.42
	3	452.03	451.8	0.00	0.00	86.7	0.02	0.64	19.1	0.19	1.87	1.3	0.91	3.99	0.4	1.95	5.45
	4	429.3	430.0	0.00	0.00	106.6	0.02	0.74	15.9	0.21	2.18	1.4	1.03	4.63	0.4	2.22	6.34
	5	466.9	465.9	0.00	0.00	98.6	0.03	0.84	14.8	0.23	2.45	1.4	1.14	5.20	0.4	2.45	7.12

Table 18 Sensitivity analysis for lower bound (43) on the newsvendor problem with respect to R_0

$\frac{m_1}{m}$ (%)		100%			75%			50%			25%			10%			
Size (n)	R_0	Orig. (secs)	Time (secs)	Gap (%)	Gap2 (%)	Time (secs)	Gap (%)	Gap2 (%)	Time (secs)	Gap (%)	Gap2 (%)	Time (secs)	Gap (%)	Gap2 (%)	Time (secs)	Gap (%)	Gap2 (%)
100	400	141.2	139.7	0.00	0.00	60.9	0.02	2.12	31.4	0.20	6.38	14.7	1.17	14.88	12.9	1.86	18.09
	500	118.7	118.7	0.00	0.00	69.0	0.03	2.13	30.9	0.21	6.39	15.7	1.18	14.90	14.0	1.87	18.10
	600	128.4	128.9	0.00	0.00	60.4	0.03	2.14	31.0	0.22	6.39	15.7	1.19	14.90	14.0	1.88	18.11
	700	138.5	139.0	0.00	0.00	56.8	0.04	2.14	30.9	0.23	6.40	15.9	1.19	14.91	14.1	1.88	18.12
	800	119.2	118.8	0.00	0.00	63.0	0.04	2.14	32.2	0.23	6.40	15.8	1.20	14.92	14.0	1.88	18.12
160	400	832.6	834.8	0.00	0.00	294.1	0.02	1.74	97.5	0.14	5.35	38.3	0.65	11.31	22.9	1.78	16.13
	500	941.2	943.7	0.00	0.00	332.4	0.02	1.75	110.2	0.14	5.35	43.3	0.66	11.31	25.9	1.79	16.14
	600	911.6	914.4	0.00	0.00	322.1	0.03	1.76	106.8	0.15	5.36	42.6	0.66	11.31	25.1	1.79	16.15
	700	897.2	899.6	0.00	0.00	316.9	0.03	1.76	105.1	0.16	5.37	41.3	0.67	11.33	24.7	1.80	16.16
	800	958.2	960.8	0.00	0.00	338.4	0.04	1.77	112.2	0.16	5.37	44.1	0.68	11.33	26.4	1.81	16.16
200	400	2575.2	2556.7	0.00	0.00	820.3	0.01	1.40	233.5	0.13	4.16	51.1	0.59	10.42	26.9	1.56	14.73
	500	2421.2	2438.7	0.00	0.00	771.5	0.01	1.41	224.8	0.13	4.17	48.2	0.60	10.43	30.1	1.57	14.74
	600	2367.1	2331.6	0.00	0.00	725.7	0.02	1.42	218.5	0.13	4.17	46.1	0.61	10.43	27.0	1.57	14.74
	700	2255.9	2250.3	0.00	0.00	729.0	0.02	1.43	213.7	0.14	4.18	45.5	0.61	10.44	27.7	1.58	14.76
	800	2316.7	2318.1	0.00	0.00	735.1	0.02	1.43	226.8	0.14	4.18	47.2	0.62	10.45	28.1	1.58	14.76

16 and 17, respectively. Similarly, we conduct the sensitivity analysis with respect to R_0 for lower bound (43) on the newsvendor problem and summarize the results in Table 18.

When parameter γ_1 increases, the values of Gap and Gap2 increase very slightly, as shown in Table 16. In contrast, when parameter γ_2 increases, the values of Gap and Gap2 also increase, though at higher rates. It implies that the value of γ_2 and accordingly the second-order moment constraint are more important than the value of γ_1 and the first-order moment constraint, respectively, in terms of affecting the computational performance of solving the corresponding DRO problems. It also further supports our focus on approximating the covariance matrix in an ambiguity set through either the PCA approach or the vector splitting approach in this paper. In addition, we can observe that an increase in R_0 will also lead to slightly increasing Gap and Gap2, as shown in Table 18. Meanwhile, the values of computational time are close for the changes of values in γ_1 , γ_2 , and R_0 , respectively.

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: (i) use PCA to shrink the dimensionality of the uncertainty space; and (ii) 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, the inner and outer approximations together enable us to construct an interval that contains the (unknown) optimal solution for a large-scale DRO problem, which cannot be solved to optimality (or even feasibility in most cases) by existing methods in a reasonable time. Such interval is tight for most cases, as demonstrated by our numerical experiments. 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.

In the future research, we can extend our work from multiple perspectives by addressing the limitations of this paper. First, it would be nice to consider a more general objective function and further develop inner and outer approximations for DRO problems. Currently in this paper, we only consider piece-wise linear objective functions. Second, it would be appealing to derive tighter and thus less conservative theoretical error bounds for our derived approximations, given

that error bounds in Propositions 7 and 10 are not tight in general. Third, by observing our derived approximations perform differently in the context of different applications, it would be very interesting to investigate how the problem structure and the technical insights of deriving the approximations can affect the computational performance of our approximations in solving various application problems.

Acknowledgments

Jianqiang Cheng and Kai Pan are co-corresponding authors. The authors thank the area editor, associate editor, and two referees for their sincere suggestions, which significantly improve the quality of this paper. This work was supported in part by the Office of Naval Research [Grant N00014-20-1-2154]. Ruiwei Jiang was supported in part by the National Science Foundation (ECCS-1845980) and Kai Pan was supported in part by The Research Grants Council of Hong Kong (Project No. 15501319).

Endnotes

1. All the instance data are publicly available at <https://github.com/meysamcheramin1370/Computationally-Efficient-Approximations-for-Distributionally-Robust-Optimization>.

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 Lissner, A. (2014). Distributionally robust stochastic knapsack problem. *SIAM J. Optim.*, 24(3):1485–1506.
- Cheng, J., Leung, J., and Lissner, 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.

Online Supplement for “Computationally Efficient Approximations for Distributionally Robust Optimization under Moment and Wasserstein Ambiguity”

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

Appendix A: Supplement to Section 2

A.1. Proof of Proposition 1

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

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

which are equivalent to $\min_{\mathbf{A}\boldsymbol{\xi} \leq \mathbf{b}, \boldsymbol{\xi} \in \mathbb{R}^m} g_k(\boldsymbol{\xi}) \geq 0$, where $g_k(\boldsymbol{\xi}) = s + \boldsymbol{\xi}^\top \mathbf{q} + \boldsymbol{\xi}^\top \mathbf{Q} \boldsymbol{\xi} - y_k^0(\mathbf{x}) - y_k(\mathbf{x})^\top \boldsymbol{\xi}$, for any $k \in [K]$. Moreover, we consider the Lagrangian dual problem of $\min_{\mathbf{A}\boldsymbol{\xi} \leq \mathbf{b}, \boldsymbol{\xi} \in \mathbb{R}^m} g_k(\boldsymbol{\xi})$, i.e., $\max_{\boldsymbol{\lambda}_k \geq 0} \inf_{\boldsymbol{\xi}} g_k(\boldsymbol{\xi}) + \boldsymbol{\lambda}_k^\top (\mathbf{A}\boldsymbol{\xi} - \mathbf{b})$, where $\boldsymbol{\lambda}_k \in \mathbb{R}^n$. Note that function $g_k(\boldsymbol{\xi})$ is convex in $\boldsymbol{\xi}$ because $\mathbf{Q} \succeq 0$. 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:

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

which are further equivalent to the following constraints:

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

As $\boldsymbol{\xi} = \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_1 + \boldsymbol{\mu}$, we replace $\boldsymbol{\xi}$ with $\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_1 + \boldsymbol{\mu}$ in (EC.2). Thus, we have

$$(\text{EC.2}) \Leftrightarrow \exists \boldsymbol{\lambda}_k \geq 0, (1, \boldsymbol{\xi}_1^\top) \mathbf{Z}_k (1, \boldsymbol{\xi}_1^\top)^\top \geq 0, \quad \forall \boldsymbol{\xi}_1 \in \mathbb{R}^m, \forall k \in [K], \quad (\text{EC.3})$$

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

where

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

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 $(\eta_0, \boldsymbol{\eta}^\top) \mathbf{Z}_k (\eta_0, \boldsymbol{\eta}^\top)^\top = \boldsymbol{\eta}^\top \mathbf{Q} \boldsymbol{\eta} \geq 0$ because \mathbf{Q} is PSD; (2) if $\eta_0 \neq 0$, then we have $(\eta_0, \boldsymbol{\eta}^\top) \mathbf{Z}_k (\eta_0, \boldsymbol{\eta}^\top)^\top = \eta_0^2 (1, \frac{\boldsymbol{\eta}^\top}{\eta_0}) \mathbf{Z}_k (1, \frac{\boldsymbol{\eta}^\top}{\eta_0})^\top \geq 0$ according to (EC.3). Therefore, \Rightarrow holds and we obtain Problem (4) by replacing constraints (3b) with (EC.4). \square

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 $\boldsymbol{\zeta} = \mathbf{U}_{m \times m_1} \boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}} \boldsymbol{\xi}_r + \boldsymbol{\mu}$ and use \mathcal{S}_ζ and \mathcal{D}_ζ to denote its support and ambiguity set, respectively. As $\mathcal{S}_r = \{\boldsymbol{\xi}_r \in \mathbb{R}^{m_1} : \mathbf{U}_{m \times m_1} \boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}} \boldsymbol{\xi}_r + \boldsymbol{\mu} \in \mathcal{S}\}$ and $\mathcal{S}_\zeta = \{\boldsymbol{\zeta} \in \mathbb{R}^m : \boldsymbol{\zeta} = \mathbf{U}_{m \times m_1} \boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}} \boldsymbol{\xi}_r + \boldsymbol{\mu}, \boldsymbol{\xi}_r \in \mathcal{S}_r\}$, we can deduce $\mathcal{S}_\zeta \subset \mathcal{S}$. We also have

$$\begin{aligned} & (\mathbb{E}_{\mathbb{P}_\zeta} [\boldsymbol{\zeta}] - \boldsymbol{\mu})^\top (\mathbf{U}_{m \times m_1} \boldsymbol{\Lambda}_{m_1} \mathbf{U}_{m \times m_1}^\top)^{-1} (\mathbb{E}_{\mathbb{P}_\zeta} [\boldsymbol{\zeta}] - \boldsymbol{\mu}) \\ &= \left(\mathbb{E}_{\mathbb{P}_r} [\mathbf{U}_{m \times m_1} \boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}} \boldsymbol{\xi}_r] \right)^\top (\mathbf{U}_{m \times m_1} \boldsymbol{\Lambda}_{m_1} \mathbf{U}_{m \times m_1}^\top)^{-1} \mathbb{E}_{\mathbb{P}_r} [\mathbf{U}_{m \times m_1} \boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}} \boldsymbol{\xi}_r] \\ &= \mathbb{E}_{\mathbb{P}_r} [\boldsymbol{\xi}_r^\top] \left(\mathbf{U}_{m \times m_1} \boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}} \right)^\top \left(\mathbf{U}_{m \times m_1} \boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}} \left(\mathbf{U}_{m \times m_1} \boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}} \right)^\top \right)^{-1} \left(\mathbf{U}_{m \times m_1} \boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}} \right)^\top \mathbb{E}_{\mathbb{P}_r} [\boldsymbol{\xi}_r] \\ &= \mathbb{E}_{\mathbb{P}_r} [\boldsymbol{\xi}_r^\top] \mathbb{E}_{\mathbb{P}_r} [\boldsymbol{\xi}_r] \\ &\leq \gamma_1, \end{aligned}$$

where the last inequality is due to (6b). Note that

$$\mathbf{U}_{m \times m_1} \boldsymbol{\Lambda}_{m_1} \mathbf{U}_{m \times m_1}^\top = \mathbf{U} \begin{bmatrix} \boldsymbol{\Lambda}_{m_1} & \mathbf{0}_{m_1 \times (m-m_1)} \\ \mathbf{0}_{(m-m_1) \times m_1} & \mathbf{0}_{(m-m_1) \times (m-m_1)} \end{bmatrix} \mathbf{U}^\top \preceq \mathbf{U} \boldsymbol{\Lambda} \mathbf{U}^\top = \boldsymbol{\Sigma}.$$

It follows that

$$(\mathbb{E}_{\mathbb{P}_\zeta} [\boldsymbol{\zeta}] - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbb{E}_{\mathbb{P}_\zeta} [\boldsymbol{\zeta}] - \boldsymbol{\mu}) \leq \gamma_1. \quad (\text{EC.5})$$

Meanwhile, we have

$$\begin{aligned} \mathbb{E}_{\mathbb{P}_\zeta} [(\boldsymbol{\zeta} - \boldsymbol{\mu})(\boldsymbol{\zeta} - \boldsymbol{\mu})^\top] &\preceq \gamma_2 \mathbf{U}_{m \times m_1} \boldsymbol{\Lambda}_{m_1} \mathbf{U}_{m \times m_1}^\top \\ &= \gamma_2 \mathbf{U} \begin{bmatrix} \boldsymbol{\Lambda}_{m_1} & \mathbf{0}_{m_1 \times (m-m_1)} \\ \mathbf{0}_{(m-m_1) \times m_1} & \mathbf{0}_{(m-m_1) \times (m-m_1)} \end{bmatrix} \mathbf{U}^\top \preceq \gamma_2 \mathbf{U} \boldsymbol{\Lambda} \mathbf{U}^\top = \gamma_2 \boldsymbol{\Sigma}. \end{aligned} \quad (\text{EC.6})$$

By $\mathcal{S}_\zeta \subset \mathcal{S}$, (EC.5), and (EC.6), it follows that \mathcal{D}_ζ lies in \mathcal{D}_{M1} , i.e., $\mathcal{D}_\zeta \subset \mathcal{D}_{M1}$, and accordingly

$$\max_{\mathbb{P}_\zeta \in \mathcal{D}_\zeta} \mathbb{E}_{\mathbb{P}_\zeta} [f(\mathbf{x}, \boldsymbol{\zeta})] \leq \max_{\mathbb{P} \in \mathcal{D}_{M1}} \mathbb{E}_{\mathbb{P}} [f(\mathbf{x}, \boldsymbol{\xi})]. \quad (\text{EC.7})$$

By the definition of $\boldsymbol{\zeta}$, we have $\max_{\mathbb{P}_\zeta \in \mathcal{D}_\zeta} \mathbb{E}_{\mathbb{P}_\zeta} [f(\mathbf{x}, \boldsymbol{\zeta})] = \max_{\mathbb{P}_r \in \mathcal{D}_{M3}} \mathbb{E}_{\mathbb{P}_r} [f(\mathbf{x}, \mathbf{U}_{m \times m_1} \boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}} \boldsymbol{\xi}_r + \boldsymbol{\mu})]$ due to change of variable. Then (EC.7) implies the following inequality

$$\max_{\mathbb{P}_r \in \mathcal{D}_{M3}} \mathbb{E}_{\mathbb{P}_r} [f(\mathbf{x}, \mathbf{U}_{m \times m_1} \boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}} \boldsymbol{\xi}_r + \boldsymbol{\mu})] \leq \max_{\mathbb{P} \in \mathcal{D}_{M1}} \mathbb{E}_{\mathbb{P}} [f(\mathbf{x}, \boldsymbol{\xi})],$$

which demonstrates that 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 $\zeta_i = \mathbf{U}_{m \times m_i} \Lambda_{m_i}^{\frac{1}{2}} \xi_{r_i} + \boldsymbol{\mu}$ for any $i \in [2]$, where $\xi_{r_i} \in \mathbb{R}^{m_i}$ for $m_2 > m_1$. The ambiguity set of ζ_i is denoted by \mathcal{D}_{ζ_i} , i.e.,

$$\mathcal{D}_{\zeta_i} = \left\{ \mathbb{P}_{\zeta_i} \mid \zeta_i \sim \mathbb{P}_{\zeta_i}, \zeta_i = \mathbf{U}_{m \times m_i} \Lambda_{m_i}^{\frac{1}{2}} \xi_{r_i} + \boldsymbol{\mu}, \xi_{r_i} \sim \mathbb{P}_{r_i} \in \mathcal{D}_{r_i} \right\}, \forall i \in [2],$$

where \mathcal{D}_{r_i} (defined as (6b)) represents the ambiguity set of ξ_{r_i} for any $i \in [2]$. For any $\zeta_1 \sim \mathbb{P}_{\zeta_1} \in \mathcal{D}_{\zeta_1}$, there exists a $\xi_{r_1} \sim \mathbb{P}_{r_1} \in \mathcal{D}_{r_1}$ such that $\zeta_1 = \mathbf{U}_{m \times m_1} \Lambda_{m_1}^{\frac{1}{2}} \xi_{r_1} + \boldsymbol{\mu} = \mathbf{U}_{m \times m_2} \Lambda_{m_2}^{\frac{1}{2}} \bar{\xi}_{r_2} + \boldsymbol{\mu}$, where $\bar{\xi}_{r_2} = (\xi_{r_1}^\top, \mathbf{0}_{m_2 - m_1}^\top)^\top \in \mathbb{R}^{m_2}$. By using \mathcal{S}_{r_i} (defined as (6c)) to denote the support of ξ_{r_i} for any $i \in [2]$, we have

$$\mathbb{P} \{ \xi_{r_1} \in \mathcal{S}_{r_1} \} = \mathbb{P} \left\{ \mathbf{U}_{m \times m_1} \Lambda_{m_1}^{\frac{1}{2}} \xi_{r_1} + \boldsymbol{\mu} \in \mathcal{S} \right\} = 1,$$

which is equivalent to $\mathbb{P} \{ \mathbf{U}_{m \times m_2} \Lambda_{m_2}^{\frac{1}{2}} \bar{\xi}_{r_2} + \boldsymbol{\mu} \in \mathcal{S} \} = 1$ and implies that $\mathbb{P} \{ \bar{\xi}_{r_2} \in \mathcal{S}_{r_2} \} = 1$ because $\mathbf{U}_{m \times m_1} \Lambda_{m_1}^{\frac{1}{2}} \xi_{r_1} = \mathbf{U}_{m \times m_2} \Lambda_{m_2}^{\frac{1}{2}} \bar{\xi}_{r_2}$. In addition, we have $\mathbb{E}[\bar{\xi}_{r_2}] = \mathbf{0}_{m_2}$ and

$$\mathbb{E} \left[\bar{\xi}_{r_2} \bar{\xi}_{r_2}^\top \right] = \begin{bmatrix} \mathbb{E} [\xi_{r_1} \xi_{r_1}^\top] & \mathbf{0}_{m_1 \times (m_2 - m_1)} \\ \mathbf{0}_{(m_2 - m_1) \times m_1} & \mathbf{0}_{(m_2 - m_1) \times (m_2 - m_1)} \end{bmatrix} \preceq \gamma_2 \mathbf{I}_{m_2}.$$

It follows that the distribution of $\bar{\xi}_{r_2}$ belongs to \mathcal{D}_{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

$$\max_{\mathbb{P}_{\zeta_1} \in \mathcal{D}_{\zeta_1}} \mathbb{E}_{\mathbb{P}_{\zeta_1}} [f(\mathbf{x}, \zeta_1)] \leq \max_{\mathbb{P}_{\zeta_2} \in \mathcal{D}_{\zeta_2}} \mathbb{E}_{\mathbb{P}_{\zeta_2}} f[(\mathbf{x}, \zeta_2)].$$

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. \square

A.3. Proof of Proposition 4

We apply the strong duality theorem to constraints (16b). As function $f(\mathbf{x}, \boldsymbol{\xi})$ is piecewise linear convex, $\boldsymbol{\xi} = \mathbf{U} \Lambda^{\frac{1}{2}} \boldsymbol{\xi}_1 + \boldsymbol{\mu}$, and $\boldsymbol{\xi}_1 = (\boldsymbol{\xi}_r^\top \in \mathbb{R}^{m_1}, \boldsymbol{\xi}_{r_2}^\top \in \mathbb{R}^{m - m_1})^\top$, we reformulate (16b) as

$$s \geq y_k^0(\mathbf{x}) + y_k(\mathbf{x})^\top \left(\mathbf{U} \Lambda^{\frac{1}{2}} (\boldsymbol{\xi}_r^\top, \boldsymbol{\xi}_{r_2}^\top)^\top + \boldsymbol{\mu} \right) - \boldsymbol{\xi}_r^\top \mathbf{Q}_r \boldsymbol{\xi}_r - \mathbf{q}^\top (\boldsymbol{\xi}_r^\top, \boldsymbol{\xi}_{r_2}^\top)^\top, \quad \forall \boldsymbol{\xi}_1 \in \mathcal{S}_1, \forall k \in [K], \quad (\text{EC.8})$$

which are equivalent to $\min_{\mathbf{A}(\mathbf{U} \Lambda^{\frac{1}{2}} \boldsymbol{\xi}_1 + \boldsymbol{\mu}) \leq \mathbf{b}, \boldsymbol{\xi}_1 \in \mathbb{R}^m} g_k(\boldsymbol{\xi}_1) \geq 0$, where $g_k(\boldsymbol{\xi}_1) = s + \mathbf{q}^\top (\boldsymbol{\xi}_r^\top, \boldsymbol{\xi}_{r_2}^\top)^\top + \boldsymbol{\xi}_r^\top \mathbf{Q}_r \boldsymbol{\xi}_r - y_k^0(\mathbf{x}) - y_k(\mathbf{x})^\top (\mathbf{U} \Lambda^{\frac{1}{2}} (\boldsymbol{\xi}_r^\top, \boldsymbol{\xi}_{r_2}^\top)^\top + \boldsymbol{\mu})$, for any $k \in [K]$. Moreover, we consider the Lagrangian dual problem of $\min_{\mathbf{A}(\mathbf{U} \Lambda^{\frac{1}{2}} \boldsymbol{\xi}_1 + \boldsymbol{\mu}) \leq \mathbf{b}, \boldsymbol{\xi}_1 \in \mathbb{R}^m} g_k(\boldsymbol{\xi}_1)$, i.e., $\max_{\boldsymbol{\lambda}_k \geq 0} \inf_{\boldsymbol{\xi}_1} g_k(\boldsymbol{\xi}_1) + \boldsymbol{\lambda}_k^\top (\mathbf{A}(\mathbf{U} \Lambda^{\frac{1}{2}} \boldsymbol{\xi}_1 + \boldsymbol{\mu}) - \mathbf{b})$, where $\boldsymbol{\lambda}_k \in \mathbb{R}^n$. Note that function $g_k(\boldsymbol{\xi}_1)$ is convex in $\boldsymbol{\xi}_1$ because it is a quadratic function that can be written as the general form $f(\mathbf{x}) = \mathbf{x}^\top \mathbf{M} \mathbf{x} + \mathbf{x}^\top \mathbf{b} + c$ where \mathbf{M} is PSD, i.e., $\mathbf{Q} \succeq 0$. Due to Assumption 1, there exists an interior point for the primal problem. It follows that constraints (EC.8) are equivalent to the following ones:

$$\max_{\boldsymbol{\lambda}_k \geq 0} \inf_{\boldsymbol{\xi}_1} g_k(\boldsymbol{\xi}_1) + \boldsymbol{\lambda}_k^\top \left(\mathbf{A} \left(\mathbf{U} \Lambda^{\frac{1}{2}} \boldsymbol{\xi}_1 + \boldsymbol{\mu} \right) - \mathbf{b} \right) \geq 0, \quad \forall k \in [K],$$

which are further equivalent to the following constraints:

$$\begin{aligned} \exists \boldsymbol{\lambda}_k \geq 0, s + \mathbf{q}^\top (\boldsymbol{\xi}_r^\top, \boldsymbol{\xi}_{r_2}^\top)^\top + \boldsymbol{\xi}_r^\top \mathbf{Q}_r \boldsymbol{\xi}_r - y_k^0(\mathbf{x}) - y_k(\mathbf{x})^\top \left(\mathbf{U} \Lambda^{\frac{1}{2}} (\boldsymbol{\xi}_r^\top, \boldsymbol{\xi}_{r_2}^\top)^\top + \boldsymbol{\mu} \right) \\ + \boldsymbol{\lambda}_k^\top \left(\mathbf{A} \left(\mathbf{U} \Lambda^{\frac{1}{2}} (\boldsymbol{\xi}_r^\top, \boldsymbol{\xi}_{r_2}^\top)^\top + \boldsymbol{\mu} \right) - \mathbf{b} \right) \geq 0, \quad \forall \boldsymbol{\xi}_1 \in \mathbb{R}^m, \forall k \in [K]. \end{aligned} \quad (\text{EC.9})$$

Then, we perform the following decomposition:

$$\mathbf{U}\Lambda^{\frac{1}{2}}(\boldsymbol{\xi}_r^\top, \boldsymbol{\xi}_{r2}^\top)^\top + \boldsymbol{\mu} = \mathbf{U}_{m \times m_1} \Lambda_{m_1}^{\frac{1}{2}} \boldsymbol{\xi}_r + \mathbf{U}_{m \times (m-m_1)} \Lambda_{m-m_1}^{\frac{1}{2}} \boldsymbol{\xi}_{r2} + \boldsymbol{\mu}, \quad (\text{EC.10})$$

where $\mathbf{U}_{m \times m_1} \in \mathbb{R}^{m \times m_1}$ and $\Lambda_{m_1}^{\frac{1}{2}} \in \mathbb{R}^{m_1 \times m_1}$ are upper-left submatrices of \mathbf{U} and $\Lambda^{\frac{1}{2}}$, respectively, and $\Lambda_{m-m_1}^{\frac{1}{2}} \in \mathbb{R}^{(m-m_1) \times (m-m_1)}$ and $\mathbf{U}_{m \times (m-m_1)} \in \mathbb{R}^{m \times (m-m_1)}$ are their lower-right submatrices, respectively. By plugging (EC.10) to (EC.9) and defining $\mathbf{q} = (\mathbf{q}_1^\top \in \mathbb{R}^{m_1}, \mathbf{q}_2^\top \in \mathbb{R}^{m-m_1})^\top$, we have

$$\begin{aligned} (\text{EC.9}) \Leftrightarrow \exists \lambda_k \geq 0, s - y_k^0(\mathbf{x}) - \boldsymbol{\lambda}_k^\top \mathbf{b} - y_k(\mathbf{x})^\top \boldsymbol{\mu} + \boldsymbol{\lambda}_k^\top \mathbf{A} \boldsymbol{\mu} + \left(\mathbf{q}_1 + \left(\mathbf{U}_{m \times m_1} \Lambda_{m_1}^{\frac{1}{2}} \right)^\top (\mathbf{A}^\top \boldsymbol{\lambda}_k - y_k(\mathbf{x})) \right)^\top \boldsymbol{\xi}_r \\ + \left(\mathbf{q}_2 + \left(\mathbf{U}_{m \times (m-m_1)} \Lambda_{m-m_1}^{\frac{1}{2}} \right)^\top (\mathbf{A}^\top \boldsymbol{\lambda}_k - y_k(\mathbf{x})) \right)^\top \boldsymbol{\xi}_{r2} + \boldsymbol{\xi}_r^\top \mathbf{Q}_r \boldsymbol{\xi}_r \geq 0, \forall \boldsymbol{\xi}_I \in \mathbb{R}^m, \forall k \in [K], \\ \Leftrightarrow (1, \boldsymbol{\xi}_r^\top) \mathbf{Z}_k (1, \boldsymbol{\xi}_r^\top)^\top + \mathbf{W}_k^\top \boldsymbol{\xi}_{r2} \geq 0, \forall \boldsymbol{\xi}_I \in \mathbb{R}^m, \forall k \in [K], \end{aligned} \quad (\text{EC.11})$$

where

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

and

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

Since $\mathbf{W}_k^\top \boldsymbol{\xi}_{r2}$ in (EC.11) is affine and $\boldsymbol{\xi}_{r2} \in \mathbb{R}^{m-m_1}$, we set $\mathbf{W}_k = \mathbf{0}$ for any $k \in [K]$, which prevents the objective value of the Lagrangian dual problem from going to infinity and accordingly leads to constraints (17c). Thus, we have

$$(\text{EC.9}) \Leftrightarrow \exists \lambda_k \geq 0, (1, \boldsymbol{\xi}_r^\top) \mathbf{Z}_k (1, \boldsymbol{\xi}_r^\top)^\top \geq 0, \forall \boldsymbol{\xi}_r \in \mathbb{R}^{m_1}, \forall k \in [K]; \quad (17c), \quad (\text{EC.12})$$

$$\Leftrightarrow \exists \lambda_k \geq 0, \mathbf{Z}_k \succeq 0, \forall k \in [K]; \quad (17c). \quad (\text{EC.13})$$

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 $(\eta_0 \in \mathbb{R}, \boldsymbol{\eta}^\top \in \mathbb{R}^{m_1})^\top \in \mathbb{R}^{m_1+1}$: (1) if $\eta_0 = 0$, then $(\eta_0, \boldsymbol{\eta}^\top) \mathbf{Z}_k (\eta_0, \boldsymbol{\eta}^\top)^\top = \boldsymbol{\eta}^\top \mathbf{Q}_r \boldsymbol{\eta} \geq 0$ because \mathbf{Q}_r is PSD; (2) if $\eta_0 \neq 0$, then we have $(\eta_0, \boldsymbol{\eta}^\top) \mathbf{Z}_k (\eta_0, \boldsymbol{\eta}^\top)^\top = \eta_0^2 (1, \frac{\boldsymbol{\eta}^\top}{\eta_0}) \mathbf{Z}_k (1, \frac{\boldsymbol{\eta}^\top}{\eta_0})^\top \geq 0$ according to (EC.12). Therefore, \Rightarrow holds and we obtain Problem (17) by replacing constraints (16b) with (EC.13). \square

A.4. Proof of Theorem 4

As \mathbb{P}_I is a probability measure on $(\mathbb{R}^m, \mathbb{B})$, where \mathbb{B} denotes the Borel σ -algebra on \mathbb{R}^m , Problem (18) can be described as the following problem:

$$\min_{\mathbf{x} \in \mathcal{X}} \max_{\mathbb{P}_I \in \mathcal{D}_{M5}} \int_{\mathcal{S}_I} f(\mathbf{x}, \mathbf{U} \Lambda^{\frac{1}{2}} \boldsymbol{\xi}_I + \boldsymbol{\mu}) d\mathbb{P}_I(\boldsymbol{\xi}_I) \quad (\text{EC.14a})$$

$$\text{s.t.} \quad \int_{\mathcal{S}_I} d\mathbb{P}_I(\boldsymbol{\xi}_I) = 1, \quad (\text{EC.14b})$$

$$\int_{\mathcal{S}_I} \begin{bmatrix} \mathbf{I}_m & \boldsymbol{\xi}_I \\ \boldsymbol{\xi}_I^\top & \gamma_i \end{bmatrix} d\mathbb{P}_I(\boldsymbol{\xi}_I) \succeq 0, \quad (\text{EC.14c})$$

$$\int_{\mathcal{S}_I} \boldsymbol{\xi}_{I_i} \boldsymbol{\xi}_{I_i}^\top d\mathbb{P}_I(\boldsymbol{\xi}_I) \leq \gamma_2 \mathbf{I}_{m_i}, \quad \forall i \in [P], \quad (\text{EC.14d})$$

where (EC.14c) is derived due to Schur's complement. In the following, we first formulate the dual of Problem (EC.14) and then we show that strong duality holds.

Considering $s, \begin{bmatrix} \mathbf{W} & \mathbf{w} \\ \mathbf{w}^\top & r \end{bmatrix} \succeq 0$, and $\mathbf{Q}_i \succeq 0$ for any $i \in [P]$ as Lagrangian multipliers of constraints (EC.14b), (EC.14c), and (EC.14d), respectively, we formulate the following problem as the Lagrangian dual problem of (EC.14):

$$\begin{aligned} \min_{\mathbf{x} \in \mathcal{X}} \max_{\mathbb{P}_1 \in \mathcal{D}_{M5}} & s + \mathbf{I}_m \bullet \mathbf{W} + \gamma_1 r + \gamma_2 \sum_{i=1}^P \mathbf{I}_{m_i} \bullet \mathbf{Q}_i \\ & - \int_{\mathcal{S}_1} \left(s - 2\mathbf{w}^\top \boldsymbol{\xi}_1 + \sum_{i=1}^P \boldsymbol{\xi}_{1_i}^\top \mathbf{Q}_i \boldsymbol{\xi}_{1_i} - f(\mathbf{x}, U\Lambda^{\frac{1}{2}}\boldsymbol{\xi}_1 + \boldsymbol{\mu}) \right) d\mathbb{P}_1(\boldsymbol{\xi}_1). \end{aligned}$$

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

$$s - 2\mathbf{w}^\top \boldsymbol{\xi}_1 + \sum_{i=1}^P \boldsymbol{\xi}_{1_i}^\top \mathbf{Q}_i \boldsymbol{\xi}_{1_i} - f(\mathbf{x}, U\Lambda^{\frac{1}{2}}\boldsymbol{\xi}_1 + \boldsymbol{\mu}) \geq 0, \quad \forall \boldsymbol{\xi}_1 \in \mathcal{S}_1.$$

Accordingly, the dual problem of (EC.14) can be described as follows:

$$\min_{\substack{\mathbf{x}, s, \mathbf{W} \\ \mathbf{w}, r, \hat{\mathbf{Q}}}} s + \mathbf{I}_m \bullet \mathbf{W} + \gamma_1 r + \gamma_2 \sum_{i=1}^P \mathbf{I}_{m_i} \bullet \mathbf{Q}_i \quad (\text{EC.15a})$$

$$\text{s.t.} \quad s - 2\mathbf{w}^\top \boldsymbol{\xi}_1 + \sum_{i=1}^P \boldsymbol{\xi}_{1_i}^\top \mathbf{Q}_i \boldsymbol{\xi}_{1_i} - f(\mathbf{x}, U\Lambda^{\frac{1}{2}}\boldsymbol{\xi}_1 + \boldsymbol{\mu}) \geq 0, \quad \forall \boldsymbol{\xi}_1 \in \mathcal{S}_1,$$

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

$$\begin{bmatrix} \mathbf{W} & \mathbf{w} \\ \mathbf{w}^\top & r \end{bmatrix} \succeq 0, \quad (\text{EC.15b})$$

where $\hat{\mathbf{Q}} = \{\mathbf{Q}_1, \dots, \mathbf{Q}_P\}$. We further simplify Problem (EC.15) towards eliminating variables \mathbf{W} and r . To that end, we keep variables \mathbf{Q}_i , for any $i \in [P]$, and s fixed while solving Problem (EC.15) analytically for variables \mathbf{W} , \mathbf{w} , and r . It follows that we solve $\min_{\mathbf{x}, \mathbf{W}, \mathbf{w}, r} \mathbf{I}_m \bullet \mathbf{W} + \gamma_1 r$ analytically for \mathbf{W} , \mathbf{w} , and r . We consider two cases for the optimal solution of r (denoted by r^*) due to constraint (EC.15b), i.e., $r^* > 0$ and $r^* = 0$, as follows.

- If $r^* > 0$, then constraint (EC.15b) can be reformulated as $\mathbf{W} \succeq \frac{\mathbf{w}\mathbf{w}^\top}{r}$ by Schur's complement. As a result, $\mathbf{W}^* = \frac{\mathbf{w}\mathbf{w}^\top}{r}$ is a valid optimal solution because $\min_{\mathbf{x}, \mathbf{W}, \mathbf{w}, r} \mathbf{I}_m \bullet \mathbf{W} + \gamma_1 r$ is a minimization problem. Replacing \mathbf{W}^* by $\frac{\mathbf{w}\mathbf{w}^\top}{r}$ leads to solve a one-dimensional convex optimization problem, i.e., $\min_{r>0} \frac{\mathbf{w}^\top \mathbf{w}}{r} + \gamma_1 r$. 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 $r^* = \frac{\|\mathbf{w}\|_2}{\sqrt{\gamma_1}}$ as the optimal solution of r . If we plug $\mathbf{W}^* = \frac{\mathbf{w}\mathbf{w}^\top}{r}$ and $r^* = \frac{\|\mathbf{w}\|_2}{\sqrt{\gamma_1}}$ in (EC.15a), we obtain the following problem:

$$\min_{\mathbf{x}, s, \mathbf{w}, \hat{\mathbf{Q}}} s + \gamma_2 \sum_{i=1}^P \mathbf{I}_{m_i} \bullet \mathbf{Q}_i + \sqrt{\gamma_1} \|\mathbf{w}\|_2 \quad (\text{EC.16})$$

$$\text{s.t.} \quad s - 2\mathbf{w}^\top \boldsymbol{\xi}_1 + \sum_{i=1}^P \boldsymbol{\xi}_{1_i}^\top \mathbf{Q}_i \boldsymbol{\xi}_{1_i} - f(\mathbf{x}, U\Lambda^{\frac{1}{2}}\boldsymbol{\xi}_1 + \boldsymbol{\mu}) \geq 0, \quad \forall \boldsymbol{\xi}_1 \in \mathcal{S}_1,$$

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

By introducing a new variable $\mathbf{q} = -2\mathbf{w}$, we obtain Problem (19).

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

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

which contradicts constraint (EC.15b). Considering $r^* = 0$ and $\mathbf{w}^* = \mathbf{0}$, $\min_{\mathbf{x}, \mathbf{W}, \mathbf{w}, r} \mathbf{I}_m \bullet \mathbf{W} + \gamma_1 r$ reduces to $\min_{\mathbf{x}, \mathbf{W}} \mathbf{I}_m \bullet \mathbf{W}$ whose optimal solution is clearly $\mathbf{W}^* = \mathbf{0}$ as it is a minimization problem. Here also by replacing $\mathbf{q} = -2\mathbf{w}$, we obtain Problem (19).

Note that our conditions on γ_1 , γ_2 , and \mathbf{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 (18). Therefore, we can conclude that there is no duality gap between Problems (18) and (19) according to the weaker version of Proposition 3.4 in Shapiro (2001).

Finally, to prove Problem (19) provides an upper bound for Problem (3), we can equivalently prove that Problem (18) is an upper bound of Problem (2) since Problems (19) and (3) are equivalent reformulations of Problems (18) and (2), respectively. To that end, we only need to prove $\mathcal{D}_{M2} \subset \mathcal{D}_{M5}$, i.e., any distribution in \mathcal{D}_{M2} also belongs to \mathcal{D}_{M5} . As the first two constraints of \mathcal{D}_{M2} and \mathcal{D}_{M5} are the same, any distribution in \mathcal{D}_{M2} satisfies constraints $\mathbb{P}(\boldsymbol{\xi}_1 \in \mathcal{S}_1) = 1$ and $\mathbb{E}_{\mathbb{P}_1}[\boldsymbol{\xi}_1^\top] \mathbb{E}_{\mathbb{P}_1}[\boldsymbol{\xi}_1] \leq \gamma_1$ in \mathcal{D}_{M5} . Thus, we only require to show any distribution in \mathcal{D}_{M2} satisfies constraint $\mathbb{E}_{\mathbb{P}_1}[\boldsymbol{\xi}_{I_i} \boldsymbol{\xi}_{I_i}^\top] \preceq \gamma_2 \mathbf{I}_{m_i}$, $\forall i \in [P]$, in \mathcal{D}_{M5} . To that end, we let $\boldsymbol{\xi}_1 = (\boldsymbol{\xi}_{I_1}^\top, \boldsymbol{\xi}_{I_2}^\top, \dots, \boldsymbol{\xi}_{I_P}^\top)^\top$, $\boldsymbol{\xi}_{I_i} \in \mathbb{R}^{m_i}$ for any $i \in [P]$, and reformulate the second-order moment constraint of \mathcal{D}_{M2} as the following equivalent constraint:

$$\mathbb{E}_{\mathbb{P}_1} \begin{bmatrix} \boldsymbol{\xi}_{I_1} \boldsymbol{\xi}_{I_1}^\top & \boldsymbol{\xi}_{I_1} \boldsymbol{\xi}_{I_2}^\top & \cdots & \boldsymbol{\xi}_{I_1} \boldsymbol{\xi}_{I_P}^\top \\ \boldsymbol{\xi}_{I_2} \boldsymbol{\xi}_{I_1}^\top & \boldsymbol{\xi}_{I_2} \boldsymbol{\xi}_{I_2}^\top & \cdots & \boldsymbol{\xi}_{I_2} \boldsymbol{\xi}_{I_P}^\top \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\xi}_{I_P} \boldsymbol{\xi}_{I_1}^\top & \boldsymbol{\xi}_{I_P} \boldsymbol{\xi}_{I_2}^\top & \cdots & \boldsymbol{\xi}_{I_P} \boldsymbol{\xi}_{I_P}^\top \end{bmatrix} \succeq \begin{bmatrix} \gamma_2 \mathbf{I}_{m_1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \gamma_2 \mathbf{I}_{m_2} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \gamma_2 \mathbf{I}_{m_P} \end{bmatrix}, \quad (\text{EC.17})$$

which implies $\mathbb{E}_{\mathbb{P}_1}[\boldsymbol{\xi}_{I_i} \boldsymbol{\xi}_{I_i}^\top] \preceq \gamma_2 \mathbf{I}_{m_i}$, $\forall i \in [P]$ by simply considering the diagonal components of the matrices on both sides of (EC.17). That is, any distribution in \mathcal{D}_{M2} , which satisfies $\mathbb{E}_{\mathbb{P}_1}[\boldsymbol{\xi}_1 \boldsymbol{\xi}_1^\top] \preceq \gamma_2 \mathbf{I}_m$, also satisfies $\mathbb{E}_{\mathbb{P}_1}[\boldsymbol{\xi}_{I_i} \boldsymbol{\xi}_{I_i}^\top] \preceq \gamma_2 \mathbf{I}_{m_i}$, $\forall i \in [P]$ in \mathcal{D}_{M5} , i.e., $\mathcal{D}_{M2} \subset \mathcal{D}_{M5}$. \square

Appendix B: Supplement to Section 3

B.1. Proof of Theorem 5

By Theorem 1 in Gao and Kleywegt (2017), Problem (42) has the following strong dual problem:

$$\min_{\mathbf{x} \in \mathcal{X}, \mathbf{Q}_r, \lambda} \left\{ \lambda R_0 + \gamma_2 \mathbf{I}_{m_1} \bullet \mathbf{Q}_r + \int_{\mathbb{R}^{m_1}} \sup_{\boldsymbol{\xi}_r} g(\boldsymbol{\xi}_r, \hat{\boldsymbol{\xi}}) \mathbb{P}_0(d\hat{\boldsymbol{\xi}}) \right\}, \quad (\text{EC.18})$$

where $g(\boldsymbol{\xi}_r, \hat{\boldsymbol{\xi}}) = \max_{k=1}^K \{y_k(\mathbf{x})^\top (\mathbf{U}_{m \times m_1} \boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}} \boldsymbol{\xi}_r + \boldsymbol{\mu}) + y_k^0(\mathbf{x})\} - \boldsymbol{\xi}_r^\top \mathbf{Q}_r \boldsymbol{\xi}_r - \lambda \|\mathbf{U}_{m \times m_1} \boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}} \boldsymbol{\xi}_r + \boldsymbol{\mu} - \hat{\boldsymbol{\xi}}\|_1$, and \mathbf{Q}_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

$$\int_{\mathbb{R}^{m_1}} \sup_{\boldsymbol{\xi}_r} g(\boldsymbol{\xi}_r, \hat{\boldsymbol{\xi}}) \mathbb{P}_0(d\hat{\boldsymbol{\xi}}) = \frac{1}{N} \sum_{i=1}^N \sup_{\boldsymbol{\xi}_r} g(\boldsymbol{\xi}_r, \hat{\boldsymbol{\xi}}^i). \quad (\text{EC.19})$$

Thus, by plugging (EC.19) into (EC.18), (EC.18) can be reformulated as

$$\min_{\mathbf{x} \in \mathcal{X}, \mathbf{Q}_r, \lambda} \left\{ \lambda R_0 + \gamma_2 \mathbf{I}_{m_1} \bullet \mathbf{Q}_r + \frac{1}{N} \sum_{i=1}^N y_i \right\} \quad (\text{EC.20a})$$

$$\text{s.t. } y_i = \sup_{\xi_r} g(\xi_r, \hat{\xi}^i), \quad \forall i \in [N]. \quad (\text{EC.20b})$$

Since Problem (EC.20) is a minimization problem, constraints (EC.20b) can be relaxed to $y_i \geq \sup_{\xi_r} g(\xi_r, \hat{\xi}^i)$ for any $i \in [N]$. Thus, we have

$$\begin{aligned} y_i &\geq \sup_{\xi_r} \left\{ \max_{k=1}^K \left\{ y_k(\mathbf{x})^\top \left(\mathbf{U}_{m \times m_1} \Lambda_{m_1}^{\frac{1}{2}} \xi_r + \boldsymbol{\mu} \right) + y_k^0(\mathbf{x}) \right\} - \xi_r^\top \mathbf{Q}_r \xi_r - \lambda \left\| \mathbf{U}_{m \times m_1} \Lambda_{m_1}^{\frac{1}{2}} \xi_r + \boldsymbol{\mu} - \hat{\xi}^i \right\|_1 \right\}, \quad \forall i \in [N] \\ \Leftrightarrow y_i &\geq \max_{k=1}^K \sup_{\xi_r} \left\{ y_k(\mathbf{x})^\top \left(\mathbf{U}_{m \times m_1} \Lambda_{m_1}^{\frac{1}{2}} \xi_r + \boldsymbol{\mu} \right) + y_k^0(\mathbf{x}) - \xi_r^\top \mathbf{Q}_r \xi_r - \lambda \left\| \mathbf{U}_{m \times m_1} \Lambda_{m_1}^{\frac{1}{2}} \xi_r + \boldsymbol{\mu} - \hat{\xi}^i \right\|_1 \right\}, \quad \forall i \in [N] \\ \Leftrightarrow y_i &\geq \sup_{\xi_r} \left\{ y_k(\mathbf{x})^\top \left(\mathbf{U}_{m \times m_1} \Lambda_{m_1}^{\frac{1}{2}} \xi_r + \boldsymbol{\mu} \right) + y_k^0(\mathbf{x}) - \xi_r^\top \mathbf{Q}_r \xi_r - \lambda \left\| \mathbf{U}_{m \times m_1} \Lambda_{m_1}^{\frac{1}{2}} \xi_r + \boldsymbol{\mu} - \hat{\xi}^i \right\|_1 \right\}, \quad \forall i \in [N], \forall k \in [K]. \end{aligned}$$

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

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

and accordingly we have

$$\begin{aligned} y_i &\geq \sup_{\xi_r} \inf_{\|\hat{\zeta}\|_* \leq 1} \left\{ y_k(\mathbf{x})^\top \left(\mathbf{U}_{m \times m_1} \Lambda_{m_1}^{\frac{1}{2}} \xi_r + \boldsymbol{\mu} \right) + y_k^0(\mathbf{x}) - \xi_r^\top \mathbf{Q}_r \xi_r - \lambda \hat{\zeta}^\top \left(\mathbf{U}_{m \times m_1} \Lambda_{m_1}^{\frac{1}{2}} \xi_r + \boldsymbol{\mu} - \hat{\xi}^i \right) \right\}, \quad \forall k \in [K] \\ \Leftrightarrow y_i &\geq \inf_{\|\hat{\zeta}\|_* \leq 1} \sup_{\xi_r} \left\{ y_k(\mathbf{x})^\top \left(\mathbf{U}_{m \times m_1} \Lambda_{m_1}^{\frac{1}{2}} \xi_r + \boldsymbol{\mu} \right) + y_k^0(\mathbf{x}) - \xi_r^\top \mathbf{Q}_r \xi_r - \lambda \hat{\zeta}^\top \left(\mathbf{U}_{m \times m_1} \Lambda_{m_1}^{\frac{1}{2}} \xi_r + \boldsymbol{\mu} - \hat{\xi}^i \right) \right\}, \quad \forall k \in [K] \\ \Leftrightarrow \exists \hat{\zeta} \text{ s.t. } \|\hat{\zeta}\|_* &\leq 1, y_i \geq \sup_{\xi_r} \left\{ y_k(\mathbf{x})^\top \left(\mathbf{U}_{m \times m_1} \Lambda_{m_1}^{\frac{1}{2}} \xi_r + \boldsymbol{\mu} \right) + y_k^0(\mathbf{x}) - \xi_r^\top \mathbf{Q}_r \xi_r - \lambda \hat{\zeta}^\top \left(\mathbf{U}_{m \times m_1} \Lambda_{m_1}^{\frac{1}{2}} \xi_r + \boldsymbol{\mu} - \hat{\xi}^i \right) \right\}, \\ &\quad \forall k \in [K] \\ \Leftrightarrow \exists \hat{\zeta} \text{ s.t. } \|\hat{\zeta}\|_* &\leq 1, y_i \geq y_k(\mathbf{x})^\top \left(\mathbf{U}_{m \times m_1} \Lambda_{m_1}^{\frac{1}{2}} \xi_r + \boldsymbol{\mu} \right) + y_k^0(\mathbf{x}) - \xi_r^\top \mathbf{Q}_r \xi_r - \lambda \hat{\zeta}^\top \left(\mathbf{U}_{m \times m_1} \Lambda_{m_1}^{\frac{1}{2}} \xi_r + \boldsymbol{\mu} - \hat{\xi}^i \right), \\ &\quad \forall \xi_r \in \mathbb{R}^{m_1}, \forall k \in [K] \\ \Leftrightarrow \exists \hat{\zeta} \text{ s.t. } \|\hat{\zeta}\|_* &\leq 1, \begin{bmatrix} \mathbf{Q}_r & \frac{1}{2} \left((-y_k(\mathbf{x})^\top + \lambda \hat{\zeta}^\top) \mathbf{U}_{m \times m_1} \Lambda_{m_1}^{\frac{1}{2}} \right)^\top \\ \frac{1}{2} \left(-y_k(\mathbf{x})^\top + \lambda \hat{\zeta}^\top \right) \mathbf{U}_{m \times m_1} \Lambda_{m_1}^{\frac{1}{2}} & y_i - y_k(\mathbf{x})^\top \boldsymbol{\mu} - y_k^0(\mathbf{x}) + \lambda \hat{\zeta}^\top \left(\boldsymbol{\mu} - \hat{\xi}^i \right) \end{bmatrix} \succeq 0, \quad \forall k \in [K], \end{aligned} \quad (\text{EC.21})$$

where the first equivalence is due to the convexity of $g(\xi_r, \hat{\xi})$, \mathcal{S}_r , and the feasible region defined by $\|\hat{\zeta}\|_* \leq 1$.

For any given $i \in [N]$, we replace $\lambda \hat{\zeta}$ by ζ^i , and then we can obtain Problem (43) by further replacing (EC.20b) by (EC.21) for any $i \in [N]$.

To prove Problem (43) provides a lower bound for Problem (40), we consider Problem (42) and define $\zeta = \mathbf{U}_{m \times m_1} \Lambda_{m_1}^{\frac{1}{2}} \xi_r + \boldsymbol{\mu}$, denoting its support and ambiguity set by \mathcal{S}_ζ and \mathcal{D}_ζ , respectively. As $\mathcal{S}_r = \{\xi_r \in \mathbb{R}^{m_1} : \mathbf{U}_{m \times m_1} \Lambda_{m_1}^{\frac{1}{2}} \xi_r + \boldsymbol{\mu} \in \mathcal{S}\}$ and $\mathcal{S}_\zeta = \{\zeta \in \mathbb{R}^m : \zeta = \mathbf{U}_{m \times m_1} \Lambda_{m_1}^{\frac{1}{2}} \xi_r + \boldsymbol{\mu}, \xi_r \in \mathcal{S}_r\}$, we can deduce $\mathcal{S}_\zeta \subset \mathcal{S}$. We also have

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

Moreover, we have

$$\min_{\pi} \left\{ \int_{\mathcal{S}^2} \|\zeta - \hat{\xi}\|_1 \pi(\zeta, \hat{\xi}) \right\} = \min_{\pi} \int_{\mathcal{S}^2} \left\| \mathbf{U}_{m \times m_1} \mathbf{\Lambda}_{m_1}^{\frac{1}{2}} \boldsymbol{\xi}_r + \boldsymbol{\mu} - \hat{\xi} \right\|_1 \pi \left(\mathbf{U}_{m \times m_1} \mathbf{\Lambda}_{m_1}^{\frac{1}{2}} \boldsymbol{\xi}_r + \boldsymbol{\mu}, \hat{\xi} \right) \leq R_0.$$

It follows that \mathcal{D}_ζ lies in \mathcal{D}_{C_1} , i.e., $\mathcal{D}_\zeta \subset \mathcal{D}_{C_1}$, and accordingly

$$\max_{\mathbb{P}_\zeta \in \mathcal{D}_\zeta} \mathbb{E}_{\mathbb{P}_\zeta} [f(\mathbf{x}, \zeta)] \leq \max_{\mathbb{P} \in \mathcal{D}_{C_1}} \mathbb{E}_{\mathbb{P}} [f(\mathbf{x}, \boldsymbol{\xi})]. \quad (\text{EC.22})$$

By the definition of ζ , we have $\max_{\mathbb{P}_\zeta \in \mathcal{D}_\zeta} \mathbb{E}_{\mathbb{P}_\zeta} [f(\mathbf{x}, \zeta)] = \max_{\mathbb{P}_r \in \mathcal{D}_{C_3}} \mathbb{E}_{\mathbb{P}_r} [f(\mathbf{x}, \mathbf{U}_{m \times m_1} \mathbf{\Lambda}_{m_1}^{\frac{1}{2}} \boldsymbol{\xi}_r + \boldsymbol{\mu})]$ due to change of variable. Then (EC.22) implies the following inequality

$$\max_{\mathbb{P}_r \in \mathcal{D}_{C_3}} \mathbb{E}_{\mathbb{P}_r} \left[f(\mathbf{x}, \mathbf{U}_{m \times m_1} \mathbf{\Lambda}_{m_1}^{\frac{1}{2}} \boldsymbol{\xi}_r + \boldsymbol{\mu}) \right] \leq \max_{\mathbb{P} \in \mathcal{D}_{C_1}} \mathbb{E}_{\mathbb{P}} [f(\mathbf{x}, \boldsymbol{\xi})],$$

which demonstrates that the optimal value of Problem (42) (i.e., Problem (43)) is a lower bound for that of Problem (DRO-C) (i.e., Problem (40)).

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

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

where \mathcal{D}_{r_i} (defined as (42b)) represents the ambiguity set of $\boldsymbol{\xi}_{r_i}$ for any $i \in [2]$. For any $\zeta_1 \sim \mathbb{P}_{\zeta_1} \in \mathcal{D}_{\zeta_1}$, there exists a $\boldsymbol{\xi}_{r_1} \sim \mathbb{P}_{r_1} \in \mathcal{D}_{r_1}$ such that $\zeta_1 = \mathbf{U}_{m \times m_1} \mathbf{\Lambda}_{m_1}^{\frac{1}{2}} \boldsymbol{\xi}_{r_1} + \boldsymbol{\mu} = \mathbf{U}_{m \times m_2} \mathbf{\Lambda}_{m_2}^{\frac{1}{2}} \bar{\boldsymbol{\xi}}_{r_2} + \boldsymbol{\mu}$, where $\bar{\boldsymbol{\xi}}_{r_2} = (\boldsymbol{\xi}_{r_1}^\top, \mathbf{0}_{m_2 - m_1}^\top)^\top \in \mathbb{R}^{m_2}$. By using \mathcal{S}_{r_i} (defined as (42c)) to denote the support of $\boldsymbol{\xi}_{r_i}$ for any $i \in [2]$, we have

$$\mathbb{P} \{ \boldsymbol{\xi}_{r_1} \in \mathcal{S}_{r_1} \} = \mathbb{P} \left\{ \mathbf{U}_{m \times m_1} \mathbf{\Lambda}_{m_1}^{\frac{1}{2}} \boldsymbol{\xi}_{r_1} + \boldsymbol{\mu} \in \mathcal{S} \right\} = 1,$$

which is equivalent to $\mathbb{P} \{ \mathbf{U}_{m \times m_2} \mathbf{\Lambda}_{m_2}^{\frac{1}{2}} \bar{\boldsymbol{\xi}}_{r_2} + \boldsymbol{\mu} \in \mathcal{S} \} = 1$ and implies that $\mathbb{P} \{ \bar{\boldsymbol{\xi}}_{r_2} \in \mathcal{S}_{r_2} \} = 1$ because $\mathbf{U}_{m \times m_1} \mathbf{\Lambda}_{m_1}^{\frac{1}{2}} \boldsymbol{\xi}_{r_1} = \mathbf{U}_{m \times m_2} \mathbf{\Lambda}_{m_2}^{\frac{1}{2}} \bar{\boldsymbol{\xi}}_{r_2}$. In addition, we have $\mathbb{E}[\bar{\boldsymbol{\xi}}_{r_2}] = \mathbf{0}_{m_2}$ and

$$\mathbb{E} \left[\bar{\boldsymbol{\xi}}_{r_2} \bar{\boldsymbol{\xi}}_{r_2}^\top \right] = \begin{bmatrix} \mathbb{E} [\boldsymbol{\xi}_{r_1} \boldsymbol{\xi}_{r_1}^\top] & \mathbf{0}_{m_1 \times (m_2 - m_1)} \\ \mathbf{0}_{(m_2 - m_1) \times m_1} & \mathbf{0}_{(m_2 - m_1) \times (m_2 - m_1)} \end{bmatrix} \preceq \gamma_2 \mathbf{I}_{m_2}.$$

It follows that the distribution of $\bar{\boldsymbol{\xi}}_{r_2}$ belongs to \mathcal{D}_{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

$$\max_{\mathbb{P}_{\zeta_1} \in \mathcal{D}_{\zeta_1}} \mathbb{E}_{\mathbb{P}_{\zeta_1}} [f(\mathbf{x}, \zeta_1)] \leq \max_{\mathbb{P}_{\zeta_2} \in \mathcal{D}_{\zeta_2}} \mathbb{E}_{\mathbb{P}_{\zeta_2}} [f(\mathbf{x}, \zeta_2)].$$

That is, the optimal value of Problem (42) (i.e., Problem (43)) is nondecreasing in m_1 .

Finally, Problem (42) is equivalent to Problem (41) when $m_1 = m$. Therefore, Problem (43) results in an exact reformulation of Problem (40). \square

B.2. Proof of Proposition 9

By Theorem 5, we have $Z^*_C(m) - Z^*_C(m_1) \geq 0$. Moreover, according to this theorem, Problem (40) and the following problem, i.e., Problem (43) with $m_1 = m$, have the same optimal value.

$$\min_{\mathbf{x}, \lambda, \mathbf{Q}, \zeta, \hat{y}} \lambda R_0 + \gamma_2 \mathbf{I}_m \bullet \mathbf{Q} + \frac{1}{N} \sum_{i=1}^N y_i \quad (\text{EC.23a})$$

$$\text{s.t.} \quad \begin{bmatrix} \mathbf{Q} & \frac{1}{2} \left((-y_k(\mathbf{x}) + \zeta^i)^\top \mathbf{U} \mathbf{\Lambda}^{\frac{1}{2}} \right)^\top \\ \frac{1}{2} (-y_k(\mathbf{x}) + \zeta^i)^\top \mathbf{U} \mathbf{\Lambda}^{\frac{1}{2}} & y_i - y_k(\mathbf{x})^\top \boldsymbol{\mu} - y_k^0(\mathbf{x}) + \zeta^{i\top} (\boldsymbol{\mu} - \hat{\xi}^i) \end{bmatrix} \succeq 0, \quad \forall i \in [N], \forall k \in [K], \quad (\text{EC.23b})$$

$$\lambda \in \mathbb{R}_+, \mathbf{x} \in \mathcal{X}, \|\zeta^i\|_* \leq \lambda, \forall i \in [N], \quad (\text{EC.23c})$$

We use $(\mathbf{x}^*, \lambda^*, \mathbf{Q}_r^*, \boldsymbol{\zeta}^{i*} \forall i \in [N], y_i^* \forall i \in [N])$ to denote an optimal solution of Problem (43). Based on this optimal solution, we construct a feasible solution of Problem (EC.23), represented by $(\bar{\mathbf{x}}, \bar{\lambda}, \bar{\mathbf{Q}}, \bar{\boldsymbol{\zeta}}^i \forall i \in [N], \bar{y}_i \forall i \in [N])$. For clarity, we define

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

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

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

$$\bar{\mathbf{Q}} = \begin{bmatrix} \mathbf{Q}_r^* & \mathbf{0}_{m_1 \times (m-m_1)} \\ \mathbf{0}_{(m-m_1) \times m_1} & \sum_{i=1}^N \sum_{k=1}^K \frac{s_1^{ik}}{4} (\mathbf{L}_{m-m_1}^{ik})^\top \mathbf{L}_{m-m_1}^{ik} \end{bmatrix}, \text{ and} \quad (\text{EC.24})$$

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

where $s_1^{ik} > 0$ and $s_2^{ik} > 0$ for any $i \in [N]$ and $k \in [K]$. As $\bar{\lambda} = \lambda^* \geq 0$, $\bar{\mathbf{x}} = \mathbf{x}^* \in \mathcal{X}$, and $\|\bar{\boldsymbol{\zeta}}^i\|_* = \|\boldsymbol{\zeta}^{i*}\|_* \leq \lambda$, $\forall i \in [N]$ due to constraint (43c), we only require $(\bar{\mathbf{x}}, \bar{\lambda}, \bar{\mathbf{Q}}, \bar{\boldsymbol{\zeta}}^i \forall i \in [N], \bar{y}_i \forall i \in [N])$ to satisfy (EC.23b). 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.23b).

We plug $(\bar{\mathbf{x}}, \bar{\lambda}, \bar{\mathbf{Q}}, \bar{\boldsymbol{\zeta}}^i \forall i \in [N], \bar{y}_i \forall i \in [N])$ to (EC.23b) and use $\bar{\mathbf{Y}}^{ik}$ for any $i \in [N]$ and $k \in [K]$ to denote the corresponding matrix in (EC.23b). For any given $i \in [N]$ and $k \in [K]$, we perform the following decomposition:

$$\begin{aligned} \bar{\mathbf{Y}}^{ik} &= \begin{bmatrix} \mathbf{Q}_r^* & \mathbf{0}_{m_1 \times (m-m_1)} & \frac{1}{2} (\mathbf{L}_{m_1}^{ik})^\top \\ \mathbf{0}_{(m-m_1) \times m_1} & \mathbf{0}_{(m-m_1) \times (m-m_1)} & \mathbf{0}_{(m-m_1) \times 1} \\ \frac{1}{2} \mathbf{L}_{m_1}^{ik} & \mathbf{0}_{1 \times (m-m_1)} & S^{ik} \end{bmatrix} + \begin{bmatrix} \mathbf{0}_{m_1 \times m_1} & \mathbf{0}_{m_1 \times (m-m_1)} & \mathbf{0}_{m_1 \times 1} \\ \mathbf{0}_{(m-m_1) \times m_1} & \sum_{i=1}^N \sum_{k=1}^K \frac{s_1^{ik}}{4} (\mathbf{L}_{m-m_1}^{ik})^\top \mathbf{L}_{m-m_1}^{ik} & \frac{1}{2} (\mathbf{L}_{m-m_1}^{ik})^\top \\ \mathbf{0}_{1 \times m_1} & \frac{1}{2} \mathbf{L}_{m-m_1}^{ik} & \sum_{k=1}^K s_2^{ik} \end{bmatrix} \\ &\succeq \begin{bmatrix} \mathbf{Q}_r^* & \mathbf{0}_{m_1 \times (m-m_1)} & \frac{1}{2} (\mathbf{L}_{m_1}^{ik})^\top \\ \mathbf{0}_{(m-m_1) \times m_1} & \mathbf{0}_{(m-m_1) \times (m-m_1)} & \mathbf{0}_{(m-m_1) \times 1} \\ \frac{1}{2} \mathbf{L}_{m_1}^{ik} & \mathbf{0}_{1 \times (m-m_1)} & S^{ik} \end{bmatrix} + \begin{bmatrix} \mathbf{0}_{m_1 \times m_1} & \mathbf{0}_{m_1 \times (m-m_1)} & \mathbf{0}_{m_1 \times 1} \\ \mathbf{0}_{(m-m_1) \times m_1} & \frac{s_1^{ik}}{4} (\mathbf{L}_{m-m_1}^{ik})^\top \mathbf{L}_{m-m_1}^{ik} & \frac{1}{2} (\mathbf{L}_{m-m_1}^{ik})^\top \\ \mathbf{0}_{1 \times m_1} & \frac{1}{2} \mathbf{L}_{m-m_1}^{ik} & s_2^{ik} \end{bmatrix}. \quad (\text{EC.25}) \end{aligned}$$

The first matrix in (EC.25) is clearly PSD because the elimination of its zero components leads to a PSD matrix due to constraints (43b). 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.23).

Next, we use $\begin{bmatrix} \bar{\mathbf{A}} & \bar{\mathbf{B}} \\ \bar{\mathbf{B}}^\top & \bar{\mathbf{C}} \end{bmatrix}$ to denote the second matrix in (EC.25) by letting $\bar{\mathbf{A}} = \begin{bmatrix} \mathbf{0}_{m_1 \times m_1} & \mathbf{0}_{m_1 \times (m-m_1)} \\ \mathbf{0}_{(m-m_1) \times m_1} & \frac{s_1^{ik}}{4} (\mathbf{L}_{m-m_1}^{ik})^\top \mathbf{L}_{m-m_1}^{ik} \end{bmatrix}$, $\bar{\mathbf{B}}^\top = (\mathbf{0}_{1 \times m_1} \frac{1}{2} \mathbf{L}_{m-m_1}^{ik})$, and $\bar{\mathbf{C}} = s_2^{ik}$. It follows that

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

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 $\begin{bmatrix} \bar{\mathbf{A}} & \bar{\mathbf{B}} \\ \bar{\mathbf{B}}^\top & \bar{\mathbf{C}} \end{bmatrix} \succeq 0$ because $\bar{\mathbf{C}}$ is invertible and positive definite.

In addition, since Problem (EC.23) is a minimization problem, its optimal value is no larger than the objective corresponding to the feasible solution $(\bar{\mathbf{x}}, \bar{\lambda}, \bar{\mathbf{Q}}, \bar{\boldsymbol{\zeta}}^i \forall i \in [N], \bar{y}_i \forall i \in [N])$. That is,

$$\begin{aligned} Z^*_C(m) &\leq \bar{\lambda}R_0 + \gamma_2 \mathbf{I}_m \bullet \bar{\mathbf{Q}} + \frac{1}{N} \sum_{i=1}^N \bar{y}_i \\ &= Z^*_C(m_1) + \gamma_2 \sum_{i=1}^N \sum_{k=1}^K \frac{s_1^{ik}}{4} \text{trace} \left((\mathbf{L}_{m-m_1}^{ik})^\top \mathbf{L}_{m-m_1}^{ik} \right) + \frac{1}{N} \sum_{i=1}^N \sum_{k=1}^K s_2^{ik}. \end{aligned} \quad (\text{EC.26})$$

Due to the condition $s_1^{ik} \times s_2^{ik} \geq 1$, we let $s_1^{ik} = \frac{2}{\sqrt{\gamma_2 N \mathbf{L}_{m-m_1}^{ik} (\mathbf{L}_{m-m_1}^{ik})^\top}}$ and $s_2^{ik} = \frac{\sqrt{\gamma_2 N \mathbf{L}_{m-m_1}^{ik} (\mathbf{L}_{m-m_1}^{ik})^\top}}{2}$ for any $i \in [N]$ and $k \in [K]$, which leads to the smallest possible value of the RHS of (EC.26). Therefore, we have

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

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

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

B.3. Proof of Theorem 6

By Theorem 1 in Gao and Kleywegt (2017), Problem (45) has the following strong dual problem:

$$\min_{\mathbf{x} \in \mathcal{X}, \mathbf{Q}_j \forall j, \lambda} \left\{ \lambda R_0 + \gamma_2 \sum_{j=1}^P \mathbf{I}_{m_j} \bullet \mathbf{Q}_j + \int_{\mathbb{R}^m} \sup_{\boldsymbol{\xi}_I} g(\boldsymbol{\xi}_I, \hat{\boldsymbol{\xi}}) \mathbb{P}_0(d\hat{\boldsymbol{\xi}}) \right\}, \quad (\text{EC.27})$$

where $g(\boldsymbol{\xi}_I, \hat{\boldsymbol{\xi}}) = \max_{k=1}^K \{y_k(\mathbf{x})^\top (\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_I + \boldsymbol{\mu}) + y_k^0(\mathbf{x})\} - \sum_{j=1}^P \boldsymbol{\xi}_{I_j}^\top \mathbf{Q}_j \boldsymbol{\xi}_{I_j} - \lambda \|\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_I + \boldsymbol{\mu} - \hat{\boldsymbol{\xi}}\|_1$, and \mathbf{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

$$\int_{\mathbb{R}^m} \sup_{\boldsymbol{\xi}_I} g(\boldsymbol{\xi}_I, \hat{\boldsymbol{\xi}}) \mathbb{P}_0(d\hat{\boldsymbol{\xi}}) = \frac{1}{N} \sum_{i=1}^N \sup_{\boldsymbol{\xi}_I} g(\boldsymbol{\xi}_I, \hat{\boldsymbol{\xi}}^i). \quad (\text{EC.28})$$

because $g(\boldsymbol{\xi}_I, \hat{\boldsymbol{\xi}})$ is a convex function and $\mathcal{S}_I = \mathbb{R}^m$, which is convex. Thus, by plugging (EC.28) into (EC.27), (EC.27) can be reformulated as

$$\min_{\mathbf{x} \in \mathcal{X}, \mathbf{Q}_j \forall j, \lambda} \left\{ \lambda R_0 + \gamma_2 \sum_{j=1}^P \mathbf{I}_{m_j} \bullet \mathbf{Q}_j + \frac{1}{N} \sum_{i=1}^N y_i \right\} \quad (\text{EC.29a})$$

$$\text{s.t. } y_i = \sup_{\boldsymbol{\xi}_I} g(\boldsymbol{\xi}_I, \hat{\boldsymbol{\xi}}^i), \quad \forall i \in [N]. \quad (\text{EC.29b})$$

Since Problem (EC.29) is a minimization problem, constraints (EC.29b) can be relaxed to $y_i \geq \sup_{\boldsymbol{\xi}_I} g(\boldsymbol{\xi}_I, \hat{\boldsymbol{\xi}}^i)$ for any $i \in [N]$. Thus, we have

$$y_i \geq \sup_{\boldsymbol{\xi}_I} \left\{ \max_{k=1}^K \left\{ y_k(\mathbf{x})^\top (\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_I + \boldsymbol{\mu}) + y_k^0(\mathbf{x}) \right\} - \sum_{j=1}^P \boldsymbol{\xi}_{I_j}^\top \mathbf{Q}_j \boldsymbol{\xi}_{I_j} - \lambda \left\| \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_I + \boldsymbol{\mu} - \hat{\boldsymbol{\xi}}^i \right\|_1 \right\}, \quad \forall i \in [N]$$

$$\begin{aligned}
&\Leftrightarrow y_i \geq \max_{k=1}^K \sup_{\boldsymbol{\xi}_I} \left\{ y_k(\mathbf{x})^\top \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_I + \boldsymbol{\mu} \right) + y_k^0(\mathbf{x}) - \sum_{j=1}^P \boldsymbol{\xi}_{I_j}^\top \mathbf{Q}_j \boldsymbol{\xi}_{I_j} - \lambda \left\| \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_I + \boldsymbol{\mu} - \hat{\boldsymbol{\xi}}^i \right\|_1 \right\}, \forall i \in [N] \\
&\Leftrightarrow y_i \geq \sup_{\boldsymbol{\xi}_I} \left\{ y_k(\mathbf{x})^\top \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_I + \boldsymbol{\mu} \right) + y_k^0(\mathbf{x}) - \sum_{j=1}^P \boldsymbol{\xi}_{I_j}^\top \mathbf{Q}_j \boldsymbol{\xi}_{I_j} - \lambda \left\| \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_I + \boldsymbol{\mu} - \hat{\boldsymbol{\xi}}^i \right\|_1 \right\}, \forall i \in [N], \forall k \in [K].
\end{aligned} \tag{EC.30}$$

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

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

and accordingly we have

$$\begin{aligned}
&y_i \geq \sup_{\boldsymbol{\xi}_I} \inf_{\|\hat{\boldsymbol{\zeta}}\|_* \leq 1} \left\{ y_k(\mathbf{x})^\top \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_I + \boldsymbol{\mu} \right) + y_k^0(\mathbf{x}) - \sum_{j=1}^P \boldsymbol{\xi}_{I_j}^\top \mathbf{Q}_j \boldsymbol{\xi}_{I_j} - \lambda \hat{\boldsymbol{\zeta}}^\top \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_I + \boldsymbol{\mu} - \hat{\boldsymbol{\xi}}^i \right) \right\}, \forall k \in [K] \\
&\Leftrightarrow y_i \geq \inf_{\|\hat{\boldsymbol{\zeta}}\|_* \leq 1} \sup_{\boldsymbol{\xi}_I} \left\{ y_k(\mathbf{x})^\top \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_I + \boldsymbol{\mu} \right) + y_k^0(\mathbf{x}) - \sum_{j=1}^P \boldsymbol{\xi}_{I_j}^\top \mathbf{Q}_j \boldsymbol{\xi}_{I_j} - \lambda \hat{\boldsymbol{\zeta}}^\top \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_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\|_* \leq 1, y_i \geq \sup_{\boldsymbol{\xi}_I} \left\{ y_k(\mathbf{x})^\top \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_I + \boldsymbol{\mu} \right) + y_k^0(\mathbf{x}) - \sum_{j=1}^P \boldsymbol{\xi}_{I_j}^\top \mathbf{Q}_j \boldsymbol{\xi}_{I_j} - \lambda \hat{\boldsymbol{\zeta}}^\top \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_I + \boldsymbol{\mu} - \hat{\boldsymbol{\xi}}^i \right) \right\}, \\
&\hspace{25em} \forall k \in [K] \\
&\Leftrightarrow \exists \hat{\boldsymbol{\zeta}} \text{ s.t. } \left\| \hat{\boldsymbol{\zeta}} \right\|_* \leq 1, y_i \geq y_k(\mathbf{x})^\top \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_I + \boldsymbol{\mu} \right) + y_k^0(\mathbf{x}) - \sum_{j=1}^P \boldsymbol{\xi}_{I_j}^\top \mathbf{Q}_j \boldsymbol{\xi}_{I_j} - \lambda \hat{\boldsymbol{\zeta}}^\top \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_I + \boldsymbol{\mu} - \hat{\boldsymbol{\xi}}^i \right), \\
&\hspace{25em} \forall \boldsymbol{\xi}_I \in \mathbb{R}^m, \forall k \in [K] \\
&\Leftrightarrow \exists \hat{\boldsymbol{\zeta}} \text{ s.t. } \left\| \hat{\boldsymbol{\zeta}} \right\|_* \leq 1, \begin{bmatrix} \mathbf{Q}' & \frac{1}{2} \left(\left(-y_k(\mathbf{x})^\top + \lambda \hat{\boldsymbol{\zeta}}^\top \right) \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}} \right)^\top \\ \frac{1}{2} \left(-y_k(\mathbf{x})^\top + \lambda \hat{\boldsymbol{\zeta}}^\top \right) \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}} & y_i - y_k(\mathbf{x})^\top \boldsymbol{\mu} - y_k^0(\mathbf{x}) + \lambda \hat{\boldsymbol{\zeta}}^\top \left(\boldsymbol{\mu} - \hat{\boldsymbol{\xi}}^i \right) \end{bmatrix} \succeq 0, \forall k \in [K],
\end{aligned} \tag{EC.31}$$

where decision variable \mathbf{Q}' is described as (32) and the first equivalence is due to the convexity of $g(\boldsymbol{\xi}_I, \hat{\boldsymbol{\xi}}^i)$, \mathcal{S}_I , and the feasible region defined by $\|\hat{\boldsymbol{\zeta}}\|_* \leq 1$. For any given $i \in [N]$, we replace $\lambda \hat{\boldsymbol{\zeta}}^\top$ by $\boldsymbol{\zeta}^i$, and then we can reduce Problem (45) to the following problem by further replacing (EC.29b) by (EC.31) for any $i \in [N]$:

$$\min_{\mathbf{x}, \lambda, \mathbf{Q}_j \forall j, \boldsymbol{\zeta}^i \forall i, y_i \forall i} \lambda R_0 + \gamma_2 \sum_{j=1}^P \mathbf{I}_{m_j} \bullet \mathbf{Q}_j + \frac{1}{N} \sum_{i=1}^N y_i \tag{EC.32a}$$

$$\text{s.t. } \begin{bmatrix} \mathbf{Q}' & \frac{1}{2} \left(\left(-y_k(\mathbf{x}) + \boldsymbol{\zeta}^i \right)^\top \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}} \right)^\top \\ \frac{1}{2} \left(-y_k(\mathbf{x}) + \boldsymbol{\zeta}^i \right)^\top \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}} & y_i - y_k(\mathbf{x})^\top \boldsymbol{\mu} - y_k^0(\mathbf{x}) + \boldsymbol{\zeta}^i \top \left(\boldsymbol{\mu} - \hat{\boldsymbol{\xi}}^i \right) \end{bmatrix} \succeq 0, \forall i \in [N], \forall k \in [K], \tag{EC.32b} \\
\lambda \in \mathbb{R}_+, \mathbf{x} \in \mathcal{X}, \left\| \boldsymbol{\zeta}^i \right\|_* \leq \lambda, \forall i \in [N].$$

Finally, by Lemma 1, we reformulate Problem (EC.32) as Problem (46) by decomposing the PSD matrix in (EC.32b) equivalently to K PSD matrices. The proof of the claim that Problem (46) provides an upper bound for Problem (40) is the same as that of Theorem 4 and thus is omitted here. \square

B.4. Proof of Proposition 10

We reformulate Problem (46) as Problem (EC.32). Let $(\mathbf{x}^*, \lambda^*, \mathbf{Q}^*, \boldsymbol{\zeta}^{i^*} \forall i \in [N], y_i^* \forall i \in [N])$ denote an optimal solution of Problem (40) with \mathbf{Q}^* represented by (33). Based on this optimal solution, in the following, we construct a feasible solution of Problem (EC.32), denoted by $(\bar{\mathbf{x}}, \bar{\lambda}, \bar{\mathbf{Q}}_j \forall j \in [P], \bar{\boldsymbol{\zeta}}^i \forall i \in [N], \bar{y}_i \forall i \in [N])$.

First, we let $\bar{\mathbf{x}} = \mathbf{x}^*$, $\bar{\lambda} = \lambda^*$, $\bar{\boldsymbol{\zeta}}^i = \boldsymbol{\zeta}^{i*}$ for any $i \in [N]$, $\bar{y}_i = k_0 y_i^*$ for any $i \in [N]$, and $\bar{\mathbf{Q}}_j = \bar{\mathbf{Q}}'$ (as described in (34)), with $k_j \geq 1$ for any $j \in \{0, 1, 2, \dots, P\}$. In order for this solution to satisfy (EC.32b), we require

$$\begin{bmatrix} \bar{\mathbf{Q}}' & \frac{1}{2} \left((-y_k(\mathbf{x}^*) + \boldsymbol{\zeta}^{i*})^\top \mathbf{U} \boldsymbol{\Lambda}^{\frac{1}{2}} \right)^\top \\ \frac{1}{2} (-y_k(\mathbf{x}^*) + \boldsymbol{\zeta}^{i*})^\top \mathbf{U} \boldsymbol{\Lambda}^{\frac{1}{2}} & k_0 y_i^* - y_k(\mathbf{x}^*)^\top \boldsymbol{\mu} - y_k^0(\mathbf{x}^*) + \boldsymbol{\zeta}^{i* \top} (\boldsymbol{\mu} - \hat{\boldsymbol{\xi}}^i) \end{bmatrix} \succeq 0, \quad \forall i \in [N], \forall k \in [K]. \quad (\text{EC.33})$$

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

$$\begin{bmatrix} \bar{\mathbf{Q}}' & \frac{1}{2} \left((-y_k(\mathbf{x}^*) + \boldsymbol{\zeta}^{i*})^\top \mathbf{U} \boldsymbol{\Lambda}^{\frac{1}{2}} \right)^\top \\ \frac{1}{2} (-y_k(\mathbf{x}^*) + \boldsymbol{\zeta}^{i*})^\top \mathbf{U} \boldsymbol{\Lambda}^{\frac{1}{2}} & k_0 (y_i^* - y_k(\mathbf{x}^*)^\top \boldsymbol{\mu} - y_k^0(\mathbf{x}^*) + \boldsymbol{\zeta}^{i* \top} (\boldsymbol{\mu} - \hat{\boldsymbol{\xi}}^i)) \end{bmatrix}, \quad \forall i \in [N], \forall k \in [K]. \quad (\text{EC.34})$$

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

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

which is PSD because its eigenvalues are non-negative. In fact, $(k_0 - 1) (y_k^0(\mathbf{x}^*) + y_k(\mathbf{x}^*)^\top \boldsymbol{\mu} - \boldsymbol{\zeta}^{i* \top} (\boldsymbol{\mu} - \hat{\boldsymbol{\xi}}^i))$ 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 $y_k^0(\mathbf{x}^*) + y_k(\mathbf{x}^*)^\top \boldsymbol{\mu} \geq 0$ according to the assumption $\min_{k=1}^K \{y_k^0(\mathbf{x}^*) + y_k(\mathbf{x}^*)^\top \boldsymbol{\mu}\} \geq 0$. Thus, we choose good values of k_j for any $j \in \{0, 1, 2, \dots, P\}$ to ensure (EC.34) to be a PSD matrix and accordingly will make (EC.33) hold.

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

$$\begin{bmatrix} k_j \mathbf{Q}_j^* & \frac{1}{2} \left((-y_k(\mathbf{x}^*) + \boldsymbol{\zeta}^{i*})^\top \mathbf{U}_{m \times m_j} \boldsymbol{\Lambda}_{m_j}^{\frac{1}{2}} \right)^\top \\ \frac{1}{2} (-y_k(\mathbf{x}^*) + \boldsymbol{\zeta}^{i*})^\top \mathbf{U}_{m \times m_j} \boldsymbol{\Lambda}_{m_j}^{\frac{1}{2}} & s_j (y_i^* - y_k(\mathbf{x}^*)^\top \boldsymbol{\mu} - y_k^0(\mathbf{x}^*) + \boldsymbol{\zeta}^{i* \top} (\boldsymbol{\mu} - \hat{\boldsymbol{\xi}}^i)) \end{bmatrix} \succeq 0, \quad \forall k \in [K], \forall i \in [N], \forall j \in [P], \quad (\text{EC.35})$$

with $\sum_{j=1}^P s_j = k_0$. Constraints (EC.35) can be satisfied by allowing $s_j \times k_j \geq 1$ for any $j \in [P]$ due to (40b). Then, we let $k_0 = k_1 = \dots = k_P$ and $s_j \times k_j = 1$ for any $j \in [P]$, leading to $k_0 = k_1 = \dots = k_P = \sqrt{P}$.

Finally, we have $UB_C^* \geq Z_C^*(m)$ by Theorem 6. Meanwhile, as Problem (46) is a minimization problem, UB_C^* is no larger than the objective value corresponding to our constructed feasible solution $(\bar{\mathbf{x}}, \bar{\lambda}, \bar{\mathbf{Q}}_j \forall j \in [P], \bar{\boldsymbol{\zeta}}^i \forall i \in [N], \bar{y}_i \forall i \in [N])$. That is, we have

$$UB_C^* \leq \lambda^* R_0 + \gamma_2 \sum_{j=1}^P \mathbf{I}_{m_j} \bullet (\sqrt{P} \mathbf{Q}_j^*) + \frac{1}{N} \sum_{i=1}^N \sqrt{P} y_i^* \leq \sqrt{P} \left(\lambda^* R_0 + \gamma_2 \sum_{j=1}^P \mathbf{I}_{m_j} \bullet \mathbf{Q}_j^* + \sum_{i=1}^N \frac{1}{N} y_i^* \right) = \sqrt{P} Z_C^*(m),$$

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

$$0 \leq UB_C^* - Z_C^*(m) \leq (\sqrt{P} - 1) Z_C^*(m). \quad \square$$

B.5. Further Including the First-Order Moment Information

We develop a DRO problem with the combined ambiguity set that incorporates Wasserstein distance information as well as both the first- and second-order moment information. That is, we consider

$$\min_{\boldsymbol{\alpha} \in \mathcal{X}} \max_{\mathbb{P} \in \mathcal{D}_{C5}} \mathbb{E}_{\mathbb{P}} [f(\boldsymbol{\alpha}, \boldsymbol{\xi})], \quad (\text{DRO-C2})$$

where

$$\mathcal{D}_{C5}(\mathcal{S}, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \gamma_1, \gamma_2, \mathbb{P}_0, R_0) = \left\{ \mathbb{P} \left| \begin{array}{l} (\mathbb{E}_{\mathbb{P}}[\boldsymbol{\xi}] - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbb{E}_{\mathbb{P}}[\boldsymbol{\xi}] - \boldsymbol{\mu}) \leq \gamma_1 \\ \mathbb{E}_{\mathbb{P}} [(\boldsymbol{\xi} - \boldsymbol{\mu})(\boldsymbol{\xi} - \boldsymbol{\mu})^\top] \preceq \gamma_2 \boldsymbol{\Sigma} \\ W(\mathbb{P}, \mathbb{P}_0) \leq R_0 \end{array} \right. \right\}.$$

We derive an SDP reformulation of (DRO-C2), as well as its inner and outer approximations that can be solved more efficiently.

PROPOSITION EC.1. Under Assumption 2, (DRO-C2) can be recast as the following SDP formulation:

$$Z_{\text{C2}}^*(m) := \min_{\substack{\mathbf{x}, \lambda, \mathbf{Q} \\ \hat{\boldsymbol{\zeta}}, \hat{\mathbf{y}}, \mathbf{w}}} \lambda R_0 + \gamma_2 \boldsymbol{\Sigma} \bullet \mathbf{Q} + 2\sqrt{\gamma_1 \mathbf{w}^\top \boldsymbol{\Sigma} \mathbf{w}} + \frac{1}{N} \sum_{i=1}^N y_i \quad (\text{EC.36a})$$

$$\text{s.t.} \begin{bmatrix} \mathbf{Q} & \frac{1}{2}(-y_k(\mathbf{x}) + \boldsymbol{\zeta}^i - 2\mathbf{Q}\boldsymbol{\mu} - 2\mathbf{w}) \\ \frac{1}{2}(-y_k(\mathbf{x}) + \boldsymbol{\zeta}^i - 2\mathbf{Q}\boldsymbol{\mu} - 2\mathbf{w})^\top & y_i - y_k^0(\mathbf{x}) - \boldsymbol{\zeta}^{i\top} \hat{\boldsymbol{\xi}}^i + \boldsymbol{\mu}^\top \mathbf{Q}\boldsymbol{\mu} + 2\mathbf{w}^\top \boldsymbol{\mu} \end{bmatrix} \succeq 0, \forall i \in [N], \forall k \in [K] \quad (\text{EC.36b})$$

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

where $\mathbf{w} \in \mathbb{R}^m$, $\mathbf{Q} \in \mathbb{R}^{m \times m}$, $\boldsymbol{\zeta}^i \in \mathbb{R}^m$ for any $i \in [N]$, $\hat{\boldsymbol{\zeta}} = \{\boldsymbol{\zeta}^1, \dots, \boldsymbol{\zeta}^N\}$, and $\hat{\mathbf{y}} = \{y_1, \dots, y_N\}$.

Proof. Problem (DRO-C2) can be rewritten as

$$\min_{\mathbf{x} \in \mathcal{X}} \max_{\mathbb{P}, \pi} \int_{\mathbb{R}^m} f(\mathbf{x}, \boldsymbol{\xi}) d\mathbb{P}(\boldsymbol{\xi}) \quad (\text{EC.37a})$$

$$\text{s.t.} \int_{\mathbb{R}^m} \begin{bmatrix} \boldsymbol{\Sigma} & (\boldsymbol{\xi} - \boldsymbol{\mu}) \\ (\boldsymbol{\xi} - \boldsymbol{\mu})^\top & \gamma_1 \end{bmatrix} d\mathbb{P}(\boldsymbol{\xi}) \succeq 0, \quad (\text{EC.37b})$$

$$\int_{\mathbb{R}^m} (\boldsymbol{\xi} - \boldsymbol{\mu})(\boldsymbol{\xi} - \boldsymbol{\mu})^\top d\mathbb{P}(\boldsymbol{\xi}) \preceq \gamma_2 \boldsymbol{\Sigma}, \quad (\text{EC.37c})$$

$$\int_{(\mathbb{R}^m)^2} \|\boldsymbol{\xi} - \hat{\boldsymbol{\xi}}\|_1 \pi(\boldsymbol{\xi}, \hat{\boldsymbol{\xi}}) \leq R_0, \quad (\text{EC.37d})$$

where $\mathbb{R}^m = \mathcal{S}$ and (EC.37b) is derived due to Schur's complement.

We let $\begin{bmatrix} \mathbf{W} & \mathbf{w} \\ \mathbf{w}^\top & r \end{bmatrix} \succeq 0$, $\mathbf{Q} \succeq 0$, and $\lambda \in \mathbb{R}_+$ denote the Lagrangian multipliers of constraints (EC.37b), (EC.37c), and (EC.37d), respectively, and thus derive the Lagrangian dual problem of (EC.37) as follows:

$$\min_{\substack{\mathbf{x}, \mathbf{W}, \mathbf{w} \\ r, \mathbf{Q}, \lambda}} \lambda R_0 + \gamma_2 \boldsymbol{\Sigma} \bullet \mathbf{Q} + \boldsymbol{\Sigma} \bullet \mathbf{W} + \gamma_1 r + \int_{\mathbb{R}^m} \sup_{\boldsymbol{\xi}} g(\boldsymbol{\xi}, \hat{\boldsymbol{\xi}}) \mathbb{P}_0(d\hat{\boldsymbol{\xi}}) \quad (\text{EC.38a})$$

$$\text{s.t.} \lambda \in \mathbb{R}_+, \mathbf{x} \in \mathcal{X}, \mathbf{Q} \succeq 0, \quad \begin{bmatrix} \mathbf{W} & \mathbf{w} \\ \mathbf{w}^\top & r \end{bmatrix} \succeq 0, \quad (\text{EC.38b})$$

where $g(\boldsymbol{\xi}, \hat{\boldsymbol{\xi}}) = f(\mathbf{x}, \boldsymbol{\xi}) + 2\mathbf{w}^\top (\boldsymbol{\xi} - \boldsymbol{\mu}) - (\boldsymbol{\xi} - \boldsymbol{\mu})^\top \mathbf{Q} (\boldsymbol{\xi} - \boldsymbol{\mu}) - \lambda \|\boldsymbol{\xi} - \hat{\boldsymbol{\xi}}\|_1$. By following the similar steps in the proof of Lemma 1 in Gao and Kleywegt (2017), we can prove that here the strong duality holds for Problem (EC.38). We then further simplify (EC.38) towards eliminating variables \mathbf{W} and r in the outer minimization problem. To that end, we keep variables \mathbf{x} , \mathbf{Q} , λ , and \mathbf{w} fixed while solving Problem (EC.38) analytically for variables \mathbf{W} and r . It follows that we solve $\min_{\mathbf{W}, r} \boldsymbol{\Sigma} \bullet \mathbf{W} + \gamma_1 r$ analytically for \mathbf{W} and r . We consider two cases for the optimal solution of r (denoted by r^*) due to constraint (EC.38b), i.e., $r^* > 0$ and $r^* = 0$, as follows.

- If $r^* > 0$, then constraint (EC.38b) can be reformulated as $\mathbf{W} \succeq \frac{\mathbf{w}\mathbf{w}^\top}{r}$ by Schur's complement. As a result, $\mathbf{W}^* = \frac{\mathbf{w}\mathbf{w}^\top}{r}$ is a valid optimal solution because $\min_{\mathbf{W}, r} \boldsymbol{\Sigma} \bullet \mathbf{W} + \gamma_1 r$ is a minimization problem. Replacing \mathbf{W}^* by $\frac{\mathbf{w}\mathbf{w}^\top}{r}$ leads to solve a one-dimensional convex optimization problem, i.e., $\min_{r>0} \left(\frac{1}{r}\right) \mathbf{w}^\top \boldsymbol{\Sigma} \mathbf{w} + \gamma_1 r$. 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 $r^* = \sqrt{\left(\frac{1}{\gamma_1}\right) \mathbf{w}^\top \boldsymbol{\Sigma} \mathbf{w}}$ as the optimal solution of r , resulting in $\min_{\mathbf{W}, r} \boldsymbol{\Sigma} \bullet \mathbf{W} + \gamma_1 r = 2\sqrt{\gamma_1 \mathbf{w}^\top \boldsymbol{\Sigma} \mathbf{w}}$.

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

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

which contradicts constraint (EC.38b). Considering $r^* = 0$ and $\mathbf{w}^* = \mathbf{0}$, $\min_{\mathbf{x}, \mathbf{W}, \mathbf{w}, r} \boldsymbol{\Sigma} \bullet \mathbf{W} + \gamma_1 r$ reduces to $\min_{\mathbf{W}} \boldsymbol{\Sigma} \bullet \mathbf{W}$ whose optimal solution is clearly $\mathbf{W}^* = \mathbf{0}$ as it is a minimization problem. As a result, $\min_{\mathbf{W}, r} \boldsymbol{\Sigma} \bullet \mathbf{W} + \gamma_1 r = 2\sqrt{\gamma_1 \mathbf{w}^\top \boldsymbol{\Sigma} \mathbf{w}}$.

Therefore, Problem (EC.38) can be recast as follows:

$$\begin{aligned} \min_{\substack{\mathbf{x}, \mathbf{w} \\ \mathbf{Q}, \lambda}} \quad & \lambda R_0 + \gamma_2 \boldsymbol{\Sigma} \bullet \mathbf{Q} + 2\sqrt{\gamma_1 \mathbf{w}^\top \boldsymbol{\Sigma} \mathbf{w}} + \int_{\mathbb{R}^m} \sup_{\boldsymbol{\xi}} g(\boldsymbol{\xi}, \hat{\boldsymbol{\xi}}) \mathbb{P}_0(\mathbf{d}\hat{\boldsymbol{\xi}}) \\ \text{s.t.} \quad & \lambda \in \mathbb{R}_+, \quad \mathbf{x} \in \mathcal{X}, \quad \mathbf{Q} \succeq 0. \end{aligned} \quad (\text{EC.39})$$

Note that $f(\mathbf{x}, \boldsymbol{\xi}) = \max_{k=1}^K \{y_k^0(\mathbf{x}) + y_k(\mathbf{x})^\top \boldsymbol{\xi}\}$. Thus, we have $g(\boldsymbol{\xi}, \hat{\boldsymbol{\xi}})$ can be rewritten as follows:

$$g(\boldsymbol{\xi}, \hat{\boldsymbol{\xi}}) = \max_{k=1}^K \{y_k^0(\mathbf{x}) + y_k(\mathbf{x})^\top \boldsymbol{\xi}\} + 2\mathbf{w}^\top (\boldsymbol{\xi} - \boldsymbol{\mu}) - (\boldsymbol{\xi} - \boldsymbol{\mu})^\top \mathbf{Q} (\boldsymbol{\xi} - \boldsymbol{\mu}) - \lambda \|\boldsymbol{\xi} - \hat{\boldsymbol{\xi}}\|_1.$$

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., $\mathbb{P}\{\boldsymbol{\xi} = \hat{\boldsymbol{\xi}}^i\} = \frac{1}{N}$ for any $i \in [N]$, we have

$$\int_{\mathbb{R}^m} \sup_{\boldsymbol{\xi}} g(\boldsymbol{\xi}, \hat{\boldsymbol{\xi}}) \mathbb{P}_0(\mathbf{d}\hat{\boldsymbol{\xi}}) = \frac{1}{N} \sum_{i=1}^N \sup_{\boldsymbol{\xi}} g(\boldsymbol{\xi}, \hat{\boldsymbol{\xi}}^i). \quad (\text{EC.40})$$

Thus, by plugging (EC.40) into (EC.39), (EC.39) can be reformulated as

$$\min_{\substack{\mathbf{x}, \mathbf{w} \\ \mathbf{Q}, \lambda}} \left\{ \lambda R_0 + \gamma_2 \boldsymbol{\Sigma} \bullet \mathbf{Q} + 2\sqrt{\gamma_1 \mathbf{w}^\top \boldsymbol{\Sigma} \mathbf{w}} + \frac{1}{N} \sum_{i=1}^N y_i \right\} \quad (\text{EC.41a})$$

$$\begin{aligned} \text{s.t.} \quad & y_i = \sup_{\boldsymbol{\xi}} g(\boldsymbol{\xi}, \hat{\boldsymbol{\xi}}^i), \quad \forall i \in [N], \\ & \lambda \in \mathbb{R}_+, \quad \mathbf{x} \in \mathcal{X}, \quad \mathbf{Q} \succeq 0. \end{aligned} \quad (\text{EC.41b})$$

Since Problem (EC.41) is a minimization problem, constraints (EC.41b) can be relaxed to $y_i \geq \sup_{\boldsymbol{\xi}} g(\boldsymbol{\xi}, \hat{\boldsymbol{\xi}}^i)$ for any $i \in [N]$. Thus, we have

$$\begin{aligned} y_i & \geq \sup_{\boldsymbol{\xi}} \left\{ \max_{k=1}^K \{y_k^0(\mathbf{x}) + y_k(\mathbf{x})^\top \boldsymbol{\xi}\} + 2\mathbf{w}^\top (\boldsymbol{\xi} - \boldsymbol{\mu}) - (\boldsymbol{\xi} - \boldsymbol{\mu})^\top \mathbf{Q} (\boldsymbol{\xi} - \boldsymbol{\mu}) - \lambda \|\boldsymbol{\xi} - \hat{\boldsymbol{\xi}}^i\|_1 \right\}, \quad \forall i \in [N] \\ \Leftrightarrow y_i & \geq \max_{k=1}^K \sup_{\boldsymbol{\xi}} \left\{ y_k^0(\mathbf{x}) + y_k(\mathbf{x})^\top \boldsymbol{\xi} + 2\mathbf{w}^\top (\boldsymbol{\xi} - \boldsymbol{\mu}) - (\boldsymbol{\xi} - \boldsymbol{\mu})^\top \mathbf{Q} (\boldsymbol{\xi} - \boldsymbol{\mu}) - \lambda \|\boldsymbol{\xi} - \hat{\boldsymbol{\xi}}^i\|_1 \right\}, \quad \forall i \in [N] \\ \Leftrightarrow y_i & \geq \sup_{\boldsymbol{\xi}} \left\{ y_k^0(\mathbf{x}) + y_k(\mathbf{x})^\top \boldsymbol{\xi} + 2\mathbf{w}^\top (\boldsymbol{\xi} - \boldsymbol{\mu}) - (\boldsymbol{\xi} - \boldsymbol{\mu})^\top \mathbf{Q} (\boldsymbol{\xi} - \boldsymbol{\mu}) - \lambda \|\boldsymbol{\xi} - \hat{\boldsymbol{\xi}}^i\|_1 \right\}, \quad \forall i \in [N], \quad \forall k \in [K]. \end{aligned}$$

For any given $i \in [N]$, we let $\|\boldsymbol{\xi} - \hat{\boldsymbol{\xi}}^i\|_1 = \sup_{\|\hat{\boldsymbol{\zeta}}\|_* \leq 1} \hat{\boldsymbol{\zeta}}^\top (\boldsymbol{\xi} - \hat{\boldsymbol{\xi}}^i)$, and accordingly we have

$$\begin{aligned} y_i & \geq \sup_{\boldsymbol{\xi}} \inf_{\|\hat{\boldsymbol{\zeta}}\|_* \leq 1} \left\{ y_k^0(\mathbf{x}) + y_k(\mathbf{x})^\top \boldsymbol{\xi} + 2\mathbf{w}^\top (\boldsymbol{\xi} - \boldsymbol{\mu}) - (\boldsymbol{\xi} - \boldsymbol{\mu})^\top \mathbf{Q} (\boldsymbol{\xi} - \boldsymbol{\mu}) - \lambda \hat{\boldsymbol{\zeta}}^\top (\boldsymbol{\xi} - \hat{\boldsymbol{\xi}}^i) \right\}, \quad \forall k \in [K] \\ \Leftrightarrow y_i & \geq \inf_{\|\hat{\boldsymbol{\zeta}}\|_* \leq 1} \sup_{\boldsymbol{\xi}} \left\{ y_k^0(\mathbf{x}) + y_k(\mathbf{x})^\top \boldsymbol{\xi} + 2\mathbf{w}^\top (\boldsymbol{\xi} - \boldsymbol{\mu}) - (\boldsymbol{\xi} - \boldsymbol{\mu})^\top \mathbf{Q} (\boldsymbol{\xi} - \boldsymbol{\mu}) - \lambda \hat{\boldsymbol{\zeta}}^\top (\boldsymbol{\xi} - \hat{\boldsymbol{\xi}}^i) \right\}, \quad \forall k \in [K] \\ \Leftrightarrow \exists \hat{\boldsymbol{\zeta}} \text{ s.t. } & \|\hat{\boldsymbol{\zeta}}\|_* \leq 1, \quad y_i \geq \sup_{\boldsymbol{\xi}} \left\{ y_k^0(\mathbf{x}) + y_k(\mathbf{x})^\top \boldsymbol{\xi} + 2\mathbf{w}^\top (\boldsymbol{\xi} - \boldsymbol{\mu}) - (\boldsymbol{\xi} - \boldsymbol{\mu})^\top \mathbf{Q} (\boldsymbol{\xi} - \boldsymbol{\mu}) - \lambda \hat{\boldsymbol{\zeta}}^\top (\boldsymbol{\xi} - \hat{\boldsymbol{\xi}}^i) \right\}, \end{aligned}$$

where $\mathbf{q}_r \in \mathbb{R}^{m_1}$, $\mathbf{Q}_r \in \mathbb{R}^{m_1 \times m_1}$, $\boldsymbol{\zeta}^i \in \mathbb{R}^m$, $\hat{\boldsymbol{\zeta}} = \{\boldsymbol{\zeta}^1, \dots, \boldsymbol{\zeta}^N\}$, and $\hat{\mathbf{y}} = \{y_1, \dots, y_N\}$. Furthermore, (i) Problem (EC.46) provides a lower bound for the optimal value of (DRO-C2); (ii) the optimal value of Problem (EC.46) is nondecreasing in m_1 ; and (iii) if $m_1 = m$, then (DRO-C2) and (EC.46) have the same optimal value.

Proof. The reformulation proof is similar to that in Proposition EC.1, while the proofs for claims (i), (ii), and (iii) are similar to those in Theorems 2 and 5. Thus we omit them here for brevity. \square

B.5.2. Upper Bound We approximate Problem (EC.36) by splitting $\boldsymbol{\xi}_1$ in \mathcal{D}_{C_6} into P pieces so that $\boldsymbol{\xi}_1 = (\boldsymbol{\xi}_{1_1}^\top, \boldsymbol{\xi}_{1_2}^\top, \dots, \boldsymbol{\xi}_{1_P}^\top)^\top$, where $\boldsymbol{\xi}_{1_j} \in \mathbb{R}^{m_j}$, $\forall j \in [P]$, and $\sum_{j=1}^P m_j = m$. This gives rise to an inner approximation

$$\min_{\mathbf{x} \in \mathcal{X}} \max_{\mathbb{P}_1 \in \mathcal{D}_{C_8}} \mathbb{E}_{\mathbb{P}_1} \left[f \left(\mathbf{x}, \mathbf{U} \boldsymbol{\Lambda}^{\frac{1}{2}} \boldsymbol{\xi}_1 + \boldsymbol{\mu} \right) \right], \quad (\text{EC.47})$$

where

$$\mathcal{D}_{C_8}(\mathcal{S}_1, \boldsymbol{\mu}, \gamma_1, \gamma_2, \mathbb{P}_0, R_0) = \left\{ \mathbb{P}_1 \left| \begin{array}{l} \mathbb{E}_{\mathbb{P}_1} [\boldsymbol{\xi}_1^\top] \mathbb{E}_{\mathbb{P}_1} [\boldsymbol{\xi}_1] \leq \gamma_1 \\ \mathbb{E}_{\mathbb{P}_1} [\boldsymbol{\xi}_{1_j} \boldsymbol{\xi}_{1_j}^\top] \preceq \gamma_2 \mathbf{I}_{m_j}, \forall j \in [P] \\ \exists \pi : \int_{\mathcal{S}^2} \left\| \mathbf{U} \boldsymbol{\Lambda}^{\frac{1}{2}} \boldsymbol{\xi}_1 + \boldsymbol{\mu} - \hat{\boldsymbol{\xi}} \right\|_1 \pi \left(\mathbf{U} \boldsymbol{\Lambda}^{\frac{1}{2}} \boldsymbol{\xi}_1 + \boldsymbol{\mu}, \hat{\boldsymbol{\xi}} \right) \leq R_0 \end{array} \right. \right\}.$$

THEOREM EC.2. Under Assumption 2, Problem (EC.47) has the same optimal value as the following SDP formulation:

$$\begin{aligned} UB_{C_2}^* := & \min_{\substack{\mathbf{x}, \lambda, \mathbf{q} \\ \mathbf{Q}_j, \hat{\boldsymbol{\zeta}}, \hat{\mathbf{y}}, \hat{\mathbf{s}}}} \lambda R_0 + \gamma_2 \sum_{j=1}^P \mathbf{I}_{m_j} \bullet \mathbf{Q}_j + \sqrt{\gamma_1} \|\mathbf{q}\|_2 + \frac{1}{N} \sum_{i=1}^N y_i & (\text{EC.48}) \\ \text{s.t.} & \left[\begin{array}{cc} \mathbf{Q}_j & \frac{1}{2} \left(\mathbf{q}_j + (-y_k(\mathbf{x}) + \boldsymbol{\zeta}^i)^\top \mathbf{U}_{m \times m_j} \boldsymbol{\Lambda}_{m_j}^{\frac{1}{2}} \right)^\top \\ \frac{1}{2} \left(\mathbf{q}_j + (-y_k(\mathbf{x}) + \boldsymbol{\zeta}^i)^\top \mathbf{U}_{m \times m_j} \boldsymbol{\Lambda}_{m_j}^{\frac{1}{2}} \right) & s_{jik} \end{array} \right] \succeq \mathbf{0}, \forall j \in [P], \\ & \forall i \in [N], \forall k \in [K], \\ & \sum_{j=1}^P s_{jik} = y_i - y_k(\mathbf{x})^\top \boldsymbol{\mu} - y_k^0(\mathbf{x}) + \left(\boldsymbol{\mu} - \hat{\boldsymbol{\zeta}}^i \right)^\top \boldsymbol{\zeta}^i, \forall i \in [N], \forall k \in [K], \\ & \lambda \in \mathbb{R}_+, \mathbf{x} \in \mathcal{X}, \|\boldsymbol{\zeta}^i\|_* \leq \lambda, \forall i \in [N], \end{aligned}$$

where $\mathbf{q} = (\mathbf{q}_1^\top \in \mathbb{R}^{m_1}, \dots, \mathbf{q}_P^\top \in \mathbb{R}^{m_P})^\top$, $\mathbf{Q}_j \in \mathbb{R}^{m_j \times m_j}$ for any $j \in [P]$, $\hat{\mathbf{Q}} = \{\mathbf{Q}_1, \dots, \mathbf{Q}_P\}$, $\boldsymbol{\zeta}^i \in \mathbb{R}^m$ for any $i \in [N]$, $\hat{\boldsymbol{\zeta}} = \{\boldsymbol{\zeta}^1, \dots, \boldsymbol{\zeta}^N\}$, $\hat{\mathbf{y}} = \{y_1, \dots, y_N\}$, and $\hat{\mathbf{s}} = \{s_{jik}, \forall i \in [N], \forall j \in [P], \forall k \in [K]\}$. Furthermore, Problem (EC.48) provides an upper bound for the optimal value of (DRO-C2).

Proof. The reformulation proof is similar to that in Proposition EC.1 and Theorem 6, while the proof of the claim that Problem (EC.48) provides an upper bound for Problem (EC.36) is the same as that of Theorem 4. Thus we omit them here for brevity. \square

B.5.3. Computational Experiments To evaluate how the first-order moment information affects computational performance, we solve Problem (DRO-C2) and its approximations in the context of production-transportation problem, and further compare the results with those of solving Problem (DRO-C) and its approximations. We first specify the proposed lower and upper bounds of (DRO-C2) in this context. The outer approximation (EC.46) leads to the following problem:

$$\min_{\substack{\mathbf{x}, \lambda, \mathbf{q}_r \\ \mathbf{Q}_r, \hat{\boldsymbol{\zeta}}, \hat{\mathbf{y}}}} \mathbf{c}^\top \mathbf{x} + \lambda R_0 + \gamma_2 \mathbf{I}_{m_1} \bullet \mathbf{Q}_r + \sqrt{\gamma_1} \|\mathbf{q}_r\|_2 + \frac{1}{N} \sum_{i=1}^N y_i$$

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

$$\lambda \in \mathbb{R}_+, \|\boldsymbol{\zeta}^i\|_* \leq \lambda, \forall i \in [N], \text{ (47d), (EC.49b) - (EC.49d),}$$

where $\mathbf{z}_k \in \mathbb{R}^{mn}$ is a vector whose $((i-1)m+j)$ -th element is z_{ijk} and $\boldsymbol{\zeta}^i \in \mathbb{R}^{mn}$. The inner approximation (EC.48) leads to the following problem:

$$\begin{aligned} \min_{\substack{\mathbf{x}, \mathbf{z}, \lambda, \mathbf{q} \\ \mathbf{Q}, \boldsymbol{\zeta}, \hat{\mathbf{y}}, \hat{\mathbf{s}}}} \quad & \mathbf{c}^\top \mathbf{x} + \lambda R_0 + \gamma_2 \sum_{j=1}^P \mathbf{I}_{m_j} \bullet \mathbf{Q}_j + \sqrt{\gamma_1} \|\mathbf{q}\|_2 + \frac{1}{N} \sum_{i=1}^N y_i \\ \text{s.t.} \quad & \begin{bmatrix} \mathbf{Q}_j & \frac{1}{2} \left(\mathbf{q}_j + \left(-\alpha_k \mathbf{z}_k^\top + \boldsymbol{\zeta}^i \right) \mathbf{U}_{mn \times m_j} \boldsymbol{\Lambda}_{m_j}^{\frac{1}{2}} \right)^\top \\ \frac{1}{2} \left(\mathbf{q}_j + \left(-\alpha_k \mathbf{z}_k^\top + \boldsymbol{\zeta}^i \right) \mathbf{U}_{mn \times m_j} \boldsymbol{\Lambda}_{m_j}^{\frac{1}{2}} \right) & s_{jik} \end{bmatrix} \succeq 0, \forall j \in [P], \\ & \sum_{j=1}^P s_{jik} = y_i - \alpha_k \mathbf{z}_k^\top \boldsymbol{\mu} - \beta_k + \boldsymbol{\zeta}^i \left(\boldsymbol{\mu} - \hat{\boldsymbol{\xi}}^i \right), \forall i \in [N], \forall k \in [K], \\ & \lambda \in \mathbb{R}_+, \|\boldsymbol{\zeta}^i\|_* \leq \lambda, \forall i \in [N], \text{ (47d), (EC.49b) - (EC.49d).} \end{aligned}$$

where $\mathbf{Q}_j \in \mathbb{R}^{m_j \times m_j}$ and $\mathbf{q}_j \in \mathbb{R}^{m_j}$ for any $j \in [P]$ so that $\sum_{j=1}^P m_j = mn$.

First, we report the performance of lower bound (EC.46) and upper bound (EC.48) of Problem (DRO-C2) in Tables EC.1 and EC.2, respectively. By comparing Table EC.1 with Table 3 that reports the lower bound performance of (DRO-C), we can observe that both the computational time and gap are not changed significantly, after additionally including the first-order moment information. By comparing Table EC.2 with Table 9 that reports the upper bound performance of (DRO-C), we can observe that the computational time change from Table 9 to Table EC.2 is not significant, while the computational gap reduction is observable.

Table EC.1 Lower bound (EC.46) on the production-transportation problem

$\frac{m_1}{m}$ (%)	100%		75%		50%		25%		10%	
Size (m, n)	Orig. (secs)	Time Gap (secs) (%)								
(5, 20)	774.6	759.2 0.00	290.3 1.42	99.1 2.67	38.6 4.85	19.5 5.19				
(4, 40)	5260.4	4985.9 0.00	1531.5 1.09	511.1 2.96	140.7 5.35	57.7 5.66				
(8, 25)	12324.4	12464.5 0.00	3998.0 2.04	1205.9 3.82	285.6 4.00	96.2 4.11				

Table EC.2 Upper bound (EC.48) on the production-transportation problem

	P	2		4		5	
Size (m, n)	Orig. (secs)	Time Gap (secs) (%)					
(5, 20)	824.1	202.8 0.25	126.5 0.58	118.1 0.80			
(4, 40)	5454.6	1088.8 0.46	548.3 1.09	519.6 1.33			
(8, 25)	12797.1	2505.0 0.08	1145.4 0.25	1135.5 0.37			

Second, we compare the optimal value difference between the lower (resp. upper) bound of Problem (DRO-C) and the lower (resp. upper) bound of Problem (DRO-C2), as reported in Tables EC.3 and EC.4.

In Tables EC.3 and EC.4, the columns “Orig1” and “Orig2” represent the optimal values of Problems (40) (i.e., original Problem (DRO-C)) and (EC.36) (i.e., original Problem (DRO-C2)), respectively. The column “Obj1” (resp. “Obj2”) represents the objective value of the approximation of Problem (DRO-C) (resp. Problem (DRO-C2)). From Tables EC.3 and EC.4, we can observe that by including the first-order moment information in the combined ambiguity set, the conservatism of the optimal solution can be reduced (leading to a smaller objective value), though very slightly.

Table EC.3 Lower bounds (43) and (EC.46) on the production-transportation problem

$\frac{m_1}{m}$ (%)	100%		75%		50%		25%		10%			
Size (m, n)	Orig1	Orig2	Obj1	Obj2								
(5, 20)	4.18	4.09	4.18	4.09	4.09	4.02	3.97	3.94	3.87	3.87	3.86	3.86
(4, 40)	2.96	2.90	2.96	2.90	2.92	2.88	2.86	2.83	2.78	2.77	2.77	2.77
(8, 25)	7.04	6.93	7.04	6.93	6.94	6.87	6.74	6.71	6.68	6.67	6.66	6.65

Table EC.4 Upper bounds (46) and (EC.48) on the production-transportation problem

	P		2		4		5	
Size (m, n)	Orig1	Orig2	Obj1	Obj2	Obj1	Obj2	Obj1	Obj2
(5, 20)	3.49	3.46	3.51	3.47	3.61	3.49	3.63	3.49
(4, 40)	2.96	2.90	3.03	2.91	3.12	2.92	3.15	2.92
(8, 25)	7.04	6.93	7.12	6.94	7.29	6.95	7.37	6.96

Third, we perform sensitivity analyses with respect to parameters γ_1 and R_0 , where we consider $(m, n) = (5, 20)$. The results are reported in Tables EC.5 and EC.6. When either γ_1 or R_0 increases, both the computational gap induced by lower bound (EC.46) increases slightly, while the computational time change is not significant. From Table EC.6 we can observe that an increase in R_0 slightly increases the computational gap, while an increase in γ_1 slightly decreases it.

Table EC.5 Sensitivity analysis for lower bound (EC.46) with respect to (γ_1, R_0)

$\frac{m_1}{m}$ (%)	100%		75%		50%		25%		10%	
(γ_1, R_0)	Orig. (secs)	Time Gap (secs) (%)	Time (secs)	Gap (%)	Time (secs)	Gap (%)	Time (secs)	Gap (%)	Time (secs)	Gap (%)
(0.4, 30)	802.7	759.1 0.00	287.7	0.87	100.7	3.07	39.7	3.87	19.1	3.96
(0.6, 30)	776.8	760.8 0.00	295.6	1.02	113.1	3.62	40.3	4.57	18.3	4.68
(0.8, 30)	861.0	845.5 0.00	282.8	1.15	109.9	4.07	39.5	5.15	17.0	5.27
(0.4, 40)	842.7	728.9 0.00	273.7	0.97	101.8	3.64	38.3	4.47	19.53	4.72
(0.6, 40)	769.2	728.9 0.00	281.4	1.25	94.3	4.03	37.5	5.05	17.0	5.26
(0.8, 40)	799.9	731.9 0.00	282.8	1.40	94.5	4.53	37.1	5.78	17.2	6.26
(0.4, 50)	760.0	703.6 0.00	276.6	1.18	98.6	4.05	40.08	4.93	17.9	5.11
(0.6, 50)	725.3	731.0 0.00	277.2	1.39	101.2	4.52	38.7	5.47	17.7	5.94
(0.8, 50)	747.7	703.4 0.00	286.4	1.57	105.1	5.12	38.2	6.43	17.1	7.63

Table EC.6 Sensitivity analysis for upper bound (EC.46) with respect to (γ_1, R_0)

P		2		4		5	
(γ_1, R_0)	Orig. (secs)	Time (secs)	Gap (%)	Time (secs)	Gap (%)	Time (secs)	Gap (%)
(0.4,30)	823.3	202.8	0.25	127.2	0.74	119.4	0.77
(0.6,30)	852.6	205.4	0.24	126.2	0.70	117.4	0.74
(0.8,30)	712.3	189.1	0.23	128.1	0.68	120.7	0.72
(0.4,40)	795.8	199.4	0.26	133.1	0.75	121.9	0.77
(0.6,40)	827.6	202.6	0.25	135.3	0.71	126.4	0.75
(0.8,40)	775.5	200.7	0.23	134.4	0.69	122.6	0.73
(0.4,50)	747.0	198.0	0.26	130.4	0.76	124.7	0.78
(0.6,50)	749.2	206.3	0.26	130.3	0.71	125.0	0.75
(0.8,50)	773.5	205.4	0.24	131.7	0.70	124.9	0.74

Appendix C: Supplement to Section 4

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

$$\min_{\substack{\mathbf{x}, \mathbf{z}, s, \\ \boldsymbol{\lambda}, \mathbf{Q}_r, \mathbf{Q}_r}} \mathbf{c}^\top \mathbf{x} + s + \gamma_2 \mathbf{I}_{m_1} \bullet \mathbf{Q}_r + \sqrt{\gamma_1} \|\mathbf{q}_r\|_2 \quad (\text{EC.49a})$$

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

$$\boldsymbol{\lambda}_k \in \mathbb{R}_+^n, \quad \forall k \in [K], \quad (47d),$$

$$\sum_{i=1}^m z_{ijk} = d_j, \quad \forall j \in [n], \quad \forall k \in [K], \quad (\text{EC.49b})$$

$$\sum_{j=1}^n z_{ijk} = x_i, \quad \forall i \in [m], \quad \forall k \in [K], \quad (\text{EC.49c})$$

$$z_{ijk} \geq 0, \quad \forall i \in [m], \quad \forall j \in [n], \quad \forall k \in [K], \quad (\text{EC.49d})$$

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

Second, the outer approximation (43) leads to the following problem:

$$\min_{\mathbf{x}, \mathbf{z}, \lambda, \mathbf{Q}_r, \boldsymbol{\zeta}, \hat{\mathbf{y}}} \mathbf{c}^\top \mathbf{x} + \lambda R_0 + \gamma_2 \mathbf{I}_{m_1} \bullet \mathbf{Q}_r + \frac{1}{N} \sum_{i=1}^N y_i \quad (\text{EC.50})$$

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

$$\lambda \in \mathbb{R}_+, \quad \|\boldsymbol{\zeta}^i\|_* \leq \lambda, \quad \forall i \in [N], \quad (47d), \quad (\text{EC.49b}) - (\text{EC.49d}),$$

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

Third, the inner approximation (17) leads to the following problem:

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

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

$$\begin{aligned} \mathbf{q}_2 + \left(\mathbf{U}_{mn \times (mn-m_1)} \boldsymbol{\Lambda}_{mn-m_1}^{\frac{1}{2}} \right)^\top (\mathbf{A}^\top \boldsymbol{\lambda}_k - \alpha_k \mathbf{z}_k) &= 0, \quad \forall k \in [K], \\ \boldsymbol{\lambda}_k \in \mathbb{R}_+^n, \quad \forall k \in [K], \end{aligned} \quad (47d), \quad (\text{EC.49b}) - (\text{EC.49d}),$$

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

Fourth, the inner approximation (20) leads to the following problem:

$$\begin{aligned} \min_{\mathbf{x}, \mathbf{z}, s, \mathbf{q}, \boldsymbol{\lambda}} \quad & \mathbf{c}^\top \mathbf{x} + s + \gamma_2 \sum_{i=1}^P \mathbf{I}_{m_i} \bullet \mathbf{Q}_i + \sqrt{\gamma_1} \|\mathbf{q}\|_2 \\ \text{s.t.} \quad & \left[\begin{array}{cc} s_{ik} & \frac{1}{2} \left(\mathbf{q}_i + \left(\mathbf{U}_{mn \times m_i} \boldsymbol{\Lambda}_{m_i}^{\frac{1}{2}} \right)^\top (\mathbf{A}^\top \boldsymbol{\lambda}_k - \alpha_k \mathbf{z}_k) \right)^\top \\ \frac{1}{2} \left(\mathbf{q}_i + \left(\mathbf{U}_{mn \times m_i} \boldsymbol{\Lambda}_{m_i}^{\frac{1}{2}} \right)^\top (\mathbf{A}^\top \boldsymbol{\lambda}_k - \alpha_k \mathbf{z}_k) \right) & \mathbf{Q}_i \end{array} \right] \succeq 0, \\ & \forall i \in [P], \quad \forall k \in [K], \\ & \sum_{i=1}^P s_{ik} = s - \beta_k - \boldsymbol{\lambda}_k^\top \mathbf{b} - \alpha_k \mathbf{z}_k^\top \boldsymbol{\mu} + \boldsymbol{\lambda}_k^\top \mathbf{A} \boldsymbol{\mu}, \quad \forall k \in [K], \\ & \boldsymbol{\lambda}_k \in \mathbb{R}_+^n, \quad \forall k \in [K], \end{aligned} \quad (47d), \quad (\text{EC.49b}) - (\text{EC.49d}),$$

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

Fifth, the inner approximation (46) leads to the following problem:

$$\begin{aligned} \min_{\mathbf{x}, \mathbf{z}, \boldsymbol{\lambda}, \mathbf{Q}, \boldsymbol{\zeta}, \hat{\mathbf{y}}} \quad & \mathbf{c}^\top \mathbf{x} + \lambda R_0 + \gamma_2 \sum_{j=1}^P \mathbf{I}_{m_j} \bullet \mathbf{Q}_j + \frac{1}{N} \sum_{i=1}^N y_i \\ \text{s.t.} \quad & \left[\begin{array}{cc} \mathbf{Q}_j & \frac{1}{2} \left((-\alpha_k \mathbf{z}_k^\top + \boldsymbol{\zeta}^{i\top}) \mathbf{U}_{mn \times m_j} \boldsymbol{\Lambda}_{m_j}^{\frac{1}{2}} \right)^\top \\ \frac{1}{2} \left((-\alpha_k \mathbf{z}_k^\top + \boldsymbol{\zeta}^{i\top}) \mathbf{U}_{mn \times m_j} \boldsymbol{\Lambda}_{m_j}^{\frac{1}{2}} \right) & s_{jik} \end{array} \right] \succeq 0, \quad \forall j \in [P], \\ & \forall i \in [N], \quad \forall k \in [K], \\ & \sum_{j=1}^P s_{jik} = y_i - \alpha_k \mathbf{z}_k^\top \boldsymbol{\mu} - \beta_k + \boldsymbol{\zeta}^{i\top} (\boldsymbol{\mu} - \hat{\boldsymbol{\xi}}^i), \quad \forall i \in [N], \quad \forall k \in [K], \\ & \lambda \in \mathbb{R}_+, \quad \|\boldsymbol{\zeta}^i\|_* \leq \lambda, \quad \forall i \in [N], \end{aligned} \quad (47d), \quad (\text{EC.49b}) - (\text{EC.49d}).$$

By Proposition 3, the optimal value gap between Problem (48) with the moment-based ambiguity set and Problem (EC.49) can be described as follows:

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

where \mathbf{z}_k^* and $\boldsymbol{\lambda}_k^*$, $k \in [K]$, are optimal solutions of Problem (EC.49), and $Z_{\text{M}}^*(mn)$ and $Z_{\text{M}}^*(m_1)$ are the optimal values of Problems (48) and (EC.49), respectively. Similarly, by Proposition 9, the optimal value gap between Problem (48) with the combined ambiguity set and Problem (EC.50) can be described as follows:

$$0 \leq Z_{\text{C}}^*(mn) - Z_{\text{C}}^*(m_1) \leq \sqrt{\frac{\gamma_2}{N}} \sum_{i=1}^N \sum_{k=1}^K \sqrt{\mathbf{L}_{mn-m_1}^{ik} (\mathbf{L}_{mn-m_1}^{ik})^\top},$$

where $Z_{\text{C}}^*(m_1)$ is the optimal value of Problem (EC.50) and $\mathbf{L}_{mn-m_1}^{ik} = (-\alpha_k \mathbf{z}_k^{*\top} + \boldsymbol{\zeta}^{i* \top}) \mathbf{U}_{mn \times (mn-m_1)} \boldsymbol{\Lambda}_{mn-m_1}^{\frac{1}{2}}$ with \mathbf{z}_k^* and $\boldsymbol{\zeta}^{i*}$, $i \in [N]$, representing optimal solutions of Problem (EC.50).