Optimized Dimensionality Reduction for Moment-based Distributionally Robust Optimization

Shiyi Jiang[a], Jianqiang Cheng[b], Kai Pan[a], Zuo-Jun Max Shen[c,d]
[a]Faculty of Business, The Hong Kong Polytechnic University, Kowloon, Hong Kong
[b]College of Engineering, University of Arizona, Tucson, AZ 85721, USA
[c]College of Engineering, University of California, Berkeley, California 94720, USA
[d]Faculty of Engineering and Faculty of Business and Economics, The University of Hong Kong, Hong Kong
Contact: shiyi.jiang@connect.polyu.hk (SJ), jqcheng@arizona.edu (JC),
kai.pan@polyu.edu.hk (KP), maxshen@berkeley.edu (Z-JMS)

Moment-based distributionally robust optimization (DRO) provides an optimization framework to integrate statistical information with traditional optimization approaches. Under this framework, one assumes that the underlying joint distribution of random parameters runs in a distributional ambiguity set constructed by moment information and makes decisions against the worst-case distribution within the set. Although most moment-based DRO problems can be reformulated as semidefinite programming (SDP) problems that can be solved in polynomial time, solving high-dimensional SDPs is still time-consuming. Unlike existing approximation approaches that first reduce the dimensionality of random parameters and then solve the approximated SDPs, we propose an optimized dimensionality reduction (ODR) approach. We first show that the ranks of the matrices in the SDP reformulations are small, by which we are then motivated to integrate the dimensionality reduction of random parameters with the subsequent optimization problems. Such integration enables two outer and one inner approximations of the original problem, all of which are low-dimensional SDPs that can be solved efficiently, providing two lower bounds and one upper bound correspondingly. More importantly, these approximations can theoretically achieve the optimal value of the original high-dimensional SDPs. As these approximations are nonconvex SDPs, we develop modified Alternating Direction Method of Multipliers (ADMM) algorithms to solve them efficiently. We demonstrate the effectiveness of our proposed ODR approach and algorithm in solving multiproduct newsvendor and conditional value at risk (CVaR) problems. Numerical results show significant advantages of our approach on the computational time and solution quality over the three best possible benchmark approaches. Our approach can obtain an optimal or near-optimal (mostly within 0.1%) solution and reduce the computational time by up to three orders of magnitude.

Key words: distributionally robust optimization, dimensionality reduction, principal component analysis, semidefinite programming, data-driven optimization

---

1. Introduction

Distributionally robust optimization (DRO) is a modeling framework that integrates statistical information with traditional optimization methods (Scarf 1958, Delage and Ye 2010). Under this framework, one assumes that the underlying joint distribution of random parameters runs in a distributional ambiguity set inferred from given data or prior belief and then optimizes their decisions against the worst-case distribution within the set.

To solve different applications, researchers study the DRO under various distributional ambiguity sets. The ambiguity set plays a crucial role in connecting statistical information and optimization modeling, providing a flexible framework for modeling uncertainties and incorporating partial information of random parameters into the model, such as information from historical data and prior belief. Moreover, the performance of DRO depends significantly on the distributional ambiguity set (Mohajerin Esfahani and Kuhn 2018). This paper focuses on moment-based ambiguity sets, which include all distributions satisfying certain moment constraints. Examples of such constraints include restricting all distributions to have the exact mean and covariance matrix (Scarf 1958), bounding the first and second moments (Delage and Ye 2010), or placing the first and second moments in a convex set (Ghaoui et al. 2003). Moment-based DRO has been extensively studied because it may be more tractable than other ambiguity sets and has a wide range of applications in industry, including but not limited to newsvendor problems (Gallego and Moon 1993, Yue et al. 2006, Natarajan et al. 2018), portfolio optimization problems (Ghaoui et al. 2003, Goldfarb and Iyengar 2003, Zymler et al. 2013, Rujeerapaiboon et al. 2016, Li 2018, Lotfi and Zenios 2018), knapsack problems (Cheng et al. 2014), transportation problems (Zhang et al. 2017, Ghosal and Wiesemann 2020), reward-risk ratio optimization (Liu et al. 2017), scheduling problems (Shehadeh et al. 2020), and machine learning (Lanckriet et al. 2002, Farnia and Tse 2016).

Moment-based DRO model can be reformulated as a semi-infinite program (Xu et al. 2018), which is generally intractable. Three approaches are mainly used to solve such a reformulation: (i) the cutting plane/surface method (Gotoh and Konno 2002, Mehrotra and Papp 2014), by which a solution is first obtained by considering a subset of the distributional ambiguity set and cuts are then added iteratively until converging to an optimal solution; (ii) the dual method (Delage and Ye 2010, Bertsimas et al. 2019), by which the inner optimization problem (e.g., a minimization problem) is dualized and integrated with the outer optimization problem (e.g., a maximization problem); (iii) the analytical method (Scarf 1958, Popescu 2007), by which the worst-case distribution is obtained and its properties are analyzed. Among these methods, the dual method is the most popular. Most literature focuses on convex reformulations of different moment-based DRO problems, mainly including second-order cone programming (SOCP) (Ghaoui et al. 2003, Lotfi and Zenios 2018, Goldfarb and Iyengar 2003) and semidefinite programming (SDP) (Ghaoui et al. 2003, Delage and Ye 2010, Cheng et al. 2014).

While SOCPs can be solved efficiently, SDPs still require significant computational time to obtain an optimal solution when they have high dimensions (Burer and Monteiro 2003, Cheng et al. 2018). Thus, it is of great interest to study efficient algorithms for solving SDPs in the context of moment-based DRO. Besides generic methods (e.g., the interior point methods) that solve the SDPs, two types of algorithms can speed up solving SDP reformulations of moment-based

DRO: low-rank SDP algorithms and dimensionality reduction methods. First, as the interior-point method is intolerably time-consuming when solving high-dimensional SDPs, some studies develop efficient low-rank algorithms by exploiting the low-rank properties of SDP constraints (Burer and Monteiro 2003). These algorithms rarely have theoretical guarantees but are practically efficient. Second, dimensionality reduction techniques stem from the field of statistics to represent the data with the important information while omitting the trivial one. In the context of moment-based DRO, such techniques can be extended to reduce the dimension of random parameters and approximate the high-dimensional SDP reformulations using low-dimensional SDPs (Cheng et al. 2018, Cheramin et al. 2022), thereby reducing the computational time significantly.

However, both the general SDP algorithms and existing dimensionality reduction methods may not perform well in the context of moment-based DRO. The general SDP algorithm aims to solve more general SDPs and may fail to consider the specific structure of the moment-based DRO models. The existing dimensionality reduction methods fail to consider the subsequent optimization problems when reducing the dimensionality space. For example, Cheng et al. (2018) and Cheramin et al. (2022) first use the PCA to choose the random parameters corresponding to the largest eigenvalues and then solve the low-dimensional SDP problem with the chosen random parameters. Such a sequential process may not provide an optimal solution of the original problem because the aim of leveraging data is to reduce the dimensionality space by focusing on only the statistical information, rather than optimizing the subsequent SDP problems. Therefore, in this paper, we integrate the dimensionality reduction with subsequent SDP problems, which leads to an *optimized dimensionality reduction (ODR) approach for moment-based DRO*. This idea echoes the recently emerging framework that integrates machine learning with decision-making (Bertsimas and Kallus 2020, Elmachtoub and Grigas 2022). We summarize our contributions as follows:

1. We prove the low-rank property of SDP reformulations of moment-based DRO problems. Specifically, we show that the ranks of matrices in SDP reformulations are less than the number of SDP constraints plus one.
2. Different from the PCA approximation approaches (Cheng et al. 2018, Cheramin et al. 2022) that first reduce the dimensionality and then solve approximation problems, we integrate the dimensionality reduction with the subsequent optimization problems and provide an optimized dimensionality reduction approach.
3. With the ODR approach, we develop two outer and one inner approximations for the original problem, leading to three low-dimensional SDP problems that can be solved efficiently. More importantly, these low-dimensional approximations can achieve the optimal value of the original high-dimensional SDP.

4. The low-dimensional SDP problems are nonconvex with bilinear terms and we develop modified Alternating Direction Method of Multipliers (ADMM) algorithms to solve them efficiently. We apply the ODR approach and ADMM algorithms to solve multiproduct newsvendor and conditional value at risk (CVaR) problems. We compare our ODR approach with three benchmark approaches: the Mosek solver, low-rank algorithm (Burer and Monteiro 2003), and PCA approximations (Cheramin et al. 2022). The results demonstrate that our ODR approach significantly outperforms them in terms of computational time and solution quality. Our approach can obtain an optimal or near-optimal (mostly within 0.1%) solution and reduce the computational time by up to three orders of magnitude. More importantly, our approach is not sensitive to the dimension m of random parameters, while the benchmark approaches perform much worse when m is larger.

The remainder of this paper is organized as follows. In Section 2, we review related literature. In Section 3, we present the SDP reformulation of moment-based DRO problems. In Section 4, we prove the low-rank property of the SDP reformulation and propose the first outer approximation under the ODR approach, leading to the lower bound for the original problem. In Sections 5 and 6, we provide the upper bound and the second lower bound for the original problem, respectively. In Section 8, we perform extensive numerical experiments on multiproduct newsvendor and CVaR problems. Section 9 concludes the paper. All proofs are presented in the Appendix.

Notation

We use non-bold symbols to denote scalar values, e.g., s and γ_1, and bold symbols to denote vectors, e.g., $\mathbf{x} = (x_1, \ldots, x_n)^\top$ 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., $\mathbf{Q}_\mathrm{r}$. We use $\mathbb{E}_\mathbb{P}[\cdot]$ to represent expectation over distribution $\mathbb{P}$ and use "$\bullet$" to denote the inner product defined by $\mathbf{A} \bullet \mathbf{B} = \sum_{i,j} A_{ij} B_{ij}$, where $\mathbf{A}$ and $\mathbf{B}$ are two conformal matrices. For any matrix $\mathbf{M}$, we use $\mathbf{M} \succeq 0$ (resp. $\mathbf{M} \succ 0$) to indicate that it is positive semi-definite (PSD) (resp. positive definite). Symbols $\|\cdot\|_1$ and $\|\cdot\|_2$ denote L1-Norm and L2-Norm, respectively. For any integer number $n \geq 1$, we use $[n]$ to denote the set $\{1, 2, \ldots, n\}$. 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. Symbols $\mathbf{1}_m$ and $\mathbf{1}_{r\times c}$ represent a one vector of size m and a one matrix of size $r \times c$, respectively.

2. Literature Review

We review related literature from four streams: moment-based DRO, dimensionality reduction, low-rank SDP algorithms, and the integration of machine learning with decision-making.

2.1. Moment-based DRO

Extensive studies provide the theories and applications of moment-based DRO (see Rahimian and Mehrotra 2019 and Lin et al. 2022 for detailed review). Scarf (1958) and Gallego and Moon (1993) are among the first to introduce the moment-based DRO framework, under which they consider fixed mean and variance of random parameters and analytically obtain the worst-case distribution in the ambiguity set, thereby obtaining the analytical optimal solution. Ghaoui et al. (2003) study distributionally robust portfolio optimization problem, where the value at risk (VaR) is minimized against the worst-case distribution in the ambiguity set. The problem is reformulated into an SOCP if the mean and covariance matrix are given and into an SDP if they belong to bounded convex ambiguity sets. There are a series of follow-up studies (Delage and Ye 2010, Li 2018, Lotfi and Zenios 2018, Zymler et al. 2013). Among them, Delage and Ye (2010) consider a general moment-based DRO framework with an ambiguity set considering the information of support, mean, and covariance matrix, leading to an SDP reformulation. We consider the same framework in this paper.

2.2. Dimensionality Reduction

To efficiently represent high-dimensional data, dimensionality reduction techniques are proposed in the literature to maintain the important information from the data and omit the trivial information. Principal component analysis (PCA) provides a high-quality representation of the data with as much information as possible by maintaining the random variables with the largest eigenvalues (Abdi and Williams 2010). Several variants of PCA are further studied in the literature: kernel PCA (Schölkopf et al. 1997), robust PCA (Candès et al. 2011), and scaled PCA (Huang et al. 2022). They are largely applied in the field of machine learning: independent component analysis (Comon 1994), latent semantic analysis (Landauer et al. 1998), and locality preserving projections (He and Niyogi 2003). As the PCA may still lose useful information, sufficient dimensionality reduction is proposed to represent the information from the data using a linear combination of the original random variables. A series of specific methods are studied in machine learning, such as the sliced inverse regression (Li 1991), sliced average variance estimation (Li and Zhu 2007), principal Hessian direction (Li 1992), and minimum average variance estimation (Yin and Li 2011).

These dimensionality reduction techniques are rarely used to support decision-making in mathematical optimization. Cheng et al. (2018) and Cheramin et al. (2022) are among the first to reduce the dimensionality space of random variables that are modeled in a moment-based DRO framework. They adopt the PCA to maintain the important random variables in the ambiguity set and reduce the size of the subsequent SDP reformulation. However, they consider only the statistical information and fail to consider the structure information of the subsequent SDP reformulation when performing dimensionality reduction. We resolve this issue in this paper.

2.3. Low-rank SDP Algorithms

As the moment-based DRO model is reformulated as an SDP (Vandenberghe and Boyd 1996) in this paper, SDP algorithms are important to solve it. Commercial optimization solvers (e.g., Mosek and Gurobi) use the interior point method to solve SDPs. Although this method can converge very fast (Helmberg 2002), its computation is very expensive. More importantly, as a general algorithm, it does not exploit useful structural properties of the SDP constraints. To solve this issue, Burer and Monteiro (2003) are among the first to propose low-rank algorithms (Lemon et al. 2016) to solve general SDPs. Specifically, they analyze the low-rank property of the SDP constraints and transform the convex SDP into a nonconvex optimization problem with a smaller size, which is further solved by augmented Lagrangian methods. In addition, Yurtsever et al. (2021) consider trace-constrained SDPs and show that the SDP constraints are weakly constrained, by which a low-rank approximation is proposed and efficiently solved. Our paper solves an integrated optimization problem that incorporates both the dimensionality reduction of random parameters and SDP formulation, under which we also exploit the low-rank property of our formulation. Numerical results show that our proposed ODR approach performs better than the low-rank algorithm in Burer and Monteiro (2003).

2.4. Integration of Machine Learning with Decision-making

Bertsimas and Kallus (2020) summarize that many optimization problems have three primitives: (i) data on uncertain parameters, (ii) auxiliary data on associated covariates, and (iii) a structured optimization concerning decisions, constraints, and objective functions. Traditional approaches first build machine learning models to perform parameter estimation and then solve the optimization problem with the estimated parameters, while a good prediction may not lead to a good decision. Thus, Bertsimas and Kallus (2020) and Elmachtoub and Grigas (2022) integrate the parameter estimation with optimization problems. Similar ideas are reflected in early attempts in Liyanage and Shanthikumar (2005) and See and Sim (2010) that solve inventory management problems. More relevant applications are recently studied. For instance, Ban and Rudin (2019) and Zhang et al. (2023) integrate feature data within the newsvendor problem; Liu et al. (2021) integrate travel-time predictors with order-assignment optimization to provide last-mile delivery services; Kallus and Mao (2023) propose a new random forest algorithm that considers the downstream optimization problem; Zhu et al. (2022) develop a joint estimation and robustness optimization framework; Qi et al. (2023) and Ho-Nguyen and Kılınç-Karzan (2022) provide an end-to-end framework to integrate prediction and optimization.

3. SDP Reformulation

Given the distribution $\mathbb{P}$ of a random vector $\boldsymbol{\xi} \in \mathbb{R}^m$, the following stochastic programming (SP) formulation seeks an $\mathbf{x} \in \mathcal{X} \subseteq \mathbb{R}^n$ to minimize the expectation of an objective function $f(\mathbf{x}, \boldsymbol{\xi})$:

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

As the distribution $\mathbb{P}$ is often unknown, we assume that $\mathbb{P}$ belongs to a distributional ambiguity set $\mathcal{D}_{\mathrm{M0}}$ constructed by statistical information estimated from historical data, and then minimize $f(\mathbf{x}, \boldsymbol{\xi})$ against the worst-case distribution instead. It leads to the following DRO formulation:

$$\min_{\mathbf{x}\in\mathcal{X}} \max_{\mathbb{P}\in\mathcal{D}_{\mathrm{M0}}} \mathbb{E}_{\mathbb{P}}\left[f\left(\mathbf{x}, \boldsymbol{\xi}\right)\right]. \tag{2}$$

We consider moment-based statistical information (Delage and Ye 2010) is included in the set $\mathcal{D}_{\mathrm{M0}}$ as follows:

$$\mathcal{D}_{\mathrm{M0}}\left(\mathcal{S}, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \gamma_1, \gamma_2\right) = \left\{ \mathbb{P} \;\middle|\; \begin{array}{l} \mathbb{P}\left(\boldsymbol{\xi} \in \mathcal{S}\right) = 1 \\ \left(\mathbb{E}_{\mathbb{P}}\left[\boldsymbol{\xi}\right] - \boldsymbol{\mu}\right)^\top \boldsymbol{\Sigma}^{-1}\left(\mathbb{E}_{\mathbb{P}}\left[\boldsymbol{\xi}\right] - \boldsymbol{\mu}\right) \le \gamma_1 \\ \mathbb{E}_{\mathbb{P}}\left[\left(\boldsymbol{\xi} - \boldsymbol{\mu}\right)\left(\boldsymbol{\xi} - \boldsymbol{\mu}\right)^\top\right] \preceq \gamma_2 \boldsymbol{\Sigma} \end{array} \right\},$$

which describes that (i) the support of $\boldsymbol{\xi}$ is $\mathcal{S}$, (ii) the mean of $\boldsymbol{\xi}$ lies in an ellipsoid of size γ_1 centered at $\boldsymbol{\mu}$, and (iii) the covariance of $\boldsymbol{\xi}$ is bounded from above by $\gamma_2\boldsymbol{\Sigma}$, with $\gamma_1 \ge 0$, $\gamma_2 \ge 1$, and $\boldsymbol{\Sigma} \succ 0$. We perform eigenvalue decomposition on the covariance matrix $\boldsymbol{\Sigma}$ as follows:

$$\boldsymbol{\Sigma} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^\top = \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top,$$

where $\mathbf{U} \in \mathbb{R}^{m\times m}$ is an orthogonal matrix and $\boldsymbol{\Lambda} \in \mathbb{R}^{m\times m}$ is a diagonal matrix. Without loss of generality, we assume that the diagonal elements of $\boldsymbol{\Lambda}$ are arranged in a nonincreasing order. By letting $\boldsymbol{\xi} = \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}} + \boldsymbol{\mu}$, we can reformulate Problem (2) as:

$$\Theta_{\mathrm{M}}(m) := \min_{\mathbf{x}\in\mathcal{X}} \max_{\mathbb{P}_{\mathrm{I}}\in\mathcal{D}_{\mathrm{M}}} \mathbb{E}_{\mathbb{P}_{\mathrm{I}}}\left[f\left(\mathbf{x}, \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}} + \boldsymbol{\mu}\right)\right], \tag{3}$$

where

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

with $\mathcal{S}_{\mathrm{I}} := \{\boldsymbol{\xi}_{\mathrm{I}} \in \mathbb{R}^m \mid \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}} + \boldsymbol{\mu} \in \mathcal{S}\}$. Similar to Cheng et al. (2018) and Cheramin et al. (2022), we make the following assumption throughout the paper.

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 $y_k(\mathbf{x}) = (y_k^1(\mathbf{x}), \ldots, y_k^m(\mathbf{x}))^\top$ and $y_k^0(\mathbf{x})$ affine in $\mathbf{x}$ for any $k \in [K]$, and $\mathcal{S}$ is polyhedral, i.e., $\mathcal{S} = \{\boldsymbol{\xi} \mid \mathbf{A}\boldsymbol{\xi} \le \mathbf{b}\}$ with $\mathbf{A} \in \mathbb{R}^{l\times m}$ and $\mathbf{b} \in \mathbb{R}^l$, with at least one interior point.*

PROPOSITION 1 **(Cheramin et al. 2022)**. *Under Assumption 1, Problem (3) has the same optimal value as the following SDP formulation:*

$$\Theta_{\mathrm{M}}(m) = \min_{\mathbf{x},s,\hat{\boldsymbol{\lambda}},\mathbf{q},\mathbf{Q}} \; s + \gamma_2 \mathbf{I}_m \bullet \mathbf{Q} + \sqrt{\gamma_1}\,\|\mathbf{q}\|_2 \tag{4a}$$

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

$$\boldsymbol{\lambda}_k \in \mathbb{R}_+^l, \; \forall k \in [K], \; \mathbf{x} \in \mathcal{X}, \tag{4c}$$

where $\hat{\boldsymbol{\lambda}} = \{\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_K\}$, $\mathbf{q} \in \mathbb{R}^m$, *and* $\mathbf{Q} \in \mathbb{R}^{m \times m}$.

Although Problem (4) is a convex program when $\mathbf{x}$ is given, it can be difficult to solve because a large m leads to high-dimensional SDP constraints at size $m+1$. As such SDP constraints originate from the covariance matrix $\boldsymbol{\Sigma}$, early attempts in Cheng et al. (2018) and Cheramin et al. (2022) exploit the statistical information $\boldsymbol{\Sigma}$ to address the computational challenge while maintaining solution quality. Specifically, they adopt the PCA, a dimensionality reduction method commonly used in statistical learning, to capture the dominant variability of $\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}}$ by maintaining the first $m_1 (\leq m)$ components of $\boldsymbol{\xi}_{\mathrm{I}}$ and fixing its other components at 0; that is,

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

where $\boldsymbol{\xi}_{\mathrm{r}} \in \mathbb{R}^{m_1}$ and $\mathbf{U}_{m\times m_1} \in \mathbb{R}^{m\times m_1}$ and $\boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}} \in \mathbb{R}^{m_1\times m_1}$ are upper-left submatrices of $\mathbf{U}$ and $\boldsymbol{\Lambda}$, respectively. That is, the m_1 components of $\boldsymbol{\xi}_{\mathrm{I}}$ corresponding to the largest eigenvalues are maintained as uncertain and the other components are fixed at their means. With a lower-dimensional random vector $\boldsymbol{\xi}_{\mathrm{r}}$, we can have a relaxation of Problem (3):

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

where

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

with

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

Meanwhile, the corresponding SDP formulation of Problem (6) has SDP constraints with smaller size at $m_1 + 1$ and can be solved more efficiently than Problem (4), leading to an efficient "*PCA approximation*." Specifically, Cheramin et al. (2022) show that the following PCA approximation

$$\Theta_{\mathrm{M}}(m_1) = \min_{\substack{\mathbf{x},s,\hat{\boldsymbol{\lambda}},\\ \mathbf{q}_{\mathrm{r}},\mathbf{Q}_{\mathrm{r}}}} s + \gamma_2 \mathbf{I}_{m_1} \bullet \mathbf{Q}_{\mathrm{r}} + \sqrt{\gamma_1}\,\|\mathbf{q}_{\mathrm{r}}\|_2 \tag{7a}$$

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

$$\forall k \in [K], \tag{7b}$$

$$\boldsymbol{\lambda}_k \in \mathbb{R}_+^l, \ \forall k \in [K], \mathbf{x} \in \mathcal{X}, \tag{7c}$$

where $\hat{\boldsymbol{\lambda}} = \{\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_K\}$, $\mathbf{q}_\mathrm{r} \in \mathbb{R}^{m_1}$, and $\mathbf{Q}_\mathrm{r} \in \mathbb{R}^{m_1 \times m_1}$, provides a *lower bound* for the optimal value of Problem (3) (i.e., Problem (4)). The PCA approximation that leads to an *upper bound* for the optimal value of Problem (3) can be similarly derived. Hereafter, we call the problem whose optimal value is a lower bound of the original Problem (3) as an *outer approximation*. In contrast, the problem generating an upper bound is called an *inner approximation* of Problem (3).

However, relying on only the statistical information (i.e., dominant variability) to choose the components and reducing the high-dimensional uncertainty space may not lead to the best approximation performance. Although Cheramin et al. (2022) provide a performance guarantee to bound the gap between the original and approximated objective values, it is difficult to close the gap when reducing the dimensionality of $\boldsymbol{\xi}_\mathrm{I}$. Such a difficulty is not surprising because maintaining only the largest statistical variability in the PCA approximations does not capture the optimality conditions of the original problems (e.g., Problem (3)). We provide an example as follows to illustrate that choosing the components of $\boldsymbol{\xi}_\mathrm{I}$ corresponding to the largest eigenvalues can be even worse than choosing the components corresponding to the least eigenvalues.

EXAMPLE 1. Given $\mathbf{x} \in \mathcal{X}$, we consider the $\mathrm{CVaR}_{1-\alpha}$ of a cost function $g(\mathbf{x}, \boldsymbol{\xi})$ formulated as the following optimization problem (Rockafellar and Uryasev 2000):

$$\min_{t \in \mathbb{R}} \ t + \frac{1}{\alpha} \mathbb{E}_\mathbb{P}\left[g(\mathbf{x}, \boldsymbol{\xi}) - t\right]^+,$$

where $\alpha \in (0,1)$ is a risk tolerance level and function $[\cdot]^+ := \max\{0, \cdot\}$. For brevity, we let $g(\mathbf{x}, \boldsymbol{\xi}) = \mathbf{x}^\top \boldsymbol{\xi}$, $\mathcal{X} = \{\mathbf{x} \in \mathbb{R}_+^m \mid \sum_{i=1}^m x_i = 1\}$, $\mathcal{D} = \{\mathbb{P} \mid \mathbb{P}(\boldsymbol{\xi} \in \mathcal{S}) = 1, \ \mathbb{E}_\mathbb{P}[\boldsymbol{\xi}] = \boldsymbol{\mu}, \ \mathbb{E}_\mathbb{P}[(\boldsymbol{\xi} - \boldsymbol{\mu})(\boldsymbol{\xi} - \boldsymbol{\mu})^\top] \preceq \boldsymbol{\Sigma}\}$, $\mathcal{S}$ is compact, and $\boldsymbol{\mu}$ is in the interior of $\mathcal{S}$. The distributionally robust counterpart of the above CVaR problem can be formulated as

$$\begin{aligned} &\min_{\mathbf{x} \in \mathcal{X}} \max_{\mathbb{P} \in \mathcal{D}} \min_{t \in \mathbb{R}} \ t + \frac{1}{\alpha} \mathbb{E}_\mathbb{P}\left[g\left(\mathbf{x}, \boldsymbol{\xi}\right) - t\right]^+ \\ = &\min_{\mathbf{x} \in \mathcal{X}, \ t \in \mathbb{R}} \max_{\mathbb{P} \in \mathcal{D}} \ t + \frac{1}{\alpha} \mathbb{E}_\mathbb{P}\left[g\left(\mathbf{x}, \boldsymbol{\xi}\right) - t\right]^+, \end{aligned} \tag{8}$$

where the equality holds by the Sion's minimax theorem (Sion 1958) because $t + (1/\alpha)\mathbb{E}_\mathbb{P}[g(\mathbf{x}, \boldsymbol{\xi}) - t]^+$ is convex in t, concave (specifically, linear) in $\mathbb{P}$, and $\mathcal{D}$ is compact. By Proposition 1, Problem (8) has the same optimal value with the following SDP formulation:

$$\min_{\substack{\mathbf{x}, s, t, \boldsymbol{\lambda}_1, \\ \boldsymbol{\lambda}_2, \mathbf{q}, \mathbf{Q}}} \ s + \mathbf{I}_m \bullet \mathbf{Q} \tag{9a}$$

$$\text{s.t.} \quad \begin{bmatrix} s - t - \boldsymbol{\lambda}_1^\top \mathbf{b} + \boldsymbol{\lambda}_1^\top \mathbf{A}\boldsymbol{\mu} & \frac{1}{2}\left(\mathbf{q} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top \mathbf{A}^\top \boldsymbol{\lambda}_1\right)^\top \\ \frac{1}{2}\left(\mathbf{q} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top \mathbf{A}^\top \boldsymbol{\lambda}_1\right) & \mathbf{Q} \end{bmatrix} \succeq 0, \tag{9b}$$

$$\begin{bmatrix} s - \left(1 - \frac{1}{\alpha}\right)t - \boldsymbol{\lambda}_2^\top \mathbf{b} - \left(\frac{1}{\alpha}\mathbf{x}\right)^\top \boldsymbol{\mu} + \boldsymbol{\lambda}_2^\top \mathbf{A}\boldsymbol{\mu} & \frac{1}{2}\left(\mathbf{q} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top \left(\mathbf{A}^\top \boldsymbol{\lambda}_2 - \frac{1}{\alpha}\mathbf{x}\right)\right)^\top \\ \frac{1}{2}\left(\mathbf{q} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top \left(\mathbf{A}^\top \boldsymbol{\lambda}_2 - \frac{1}{\alpha}\mathbf{x}\right)\right) & \mathbf{Q} \end{bmatrix} \succeq 0, \tag{9c}$$

$$\mathbf{x} \in \mathcal{X},\ t \in \mathbb{R},\ \boldsymbol{\lambda}_1 \in \mathbb{R}^l_+,\ \boldsymbol{\lambda}_2 \in \mathbb{R}^l_+.$$

Let $\alpha = 0.05$, $\mathcal{S} = \{\boldsymbol{\xi} \in \mathbb{R}^3 \mid 0 \le \xi_1 \le 2, 1 \le \xi_2 \le 3, 2 \le \xi_3 \le 4\}$, $\boldsymbol{\mu} = [1, 2, 3]$, $\boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 2 \end{bmatrix}$ with eigenvalues 1, 3, and 2. Solving Problem (9) gives the optimal value 2 with $t = 2$. When we follow Cheng et al. (2018) and Cheramin et al. (2022) to perform PCA approximation over Problem (8) by capturing only one component of the three components in $\boldsymbol{\xi}$, we observe the following:

- Choosing the components corresponding the largest and second largest eigenvalues 3 and 2, respectively, the PCA approximations both give the optimal value at 1 with $t = 1$.
- Choosing the component corresponding the least eigenvalue 1, the PCA approximation gives the optimal value at 2 with $t = 2$.

Example 1 shows that performing dimensionality reduction (i.e., from $\boldsymbol{\xi}$ to $\boldsymbol{\xi}_{\mathrm{r}}$) using the components with the largest variability may not produce a good optimal value from the *subsequent* PCA approximation (i.e., an SDP) and it can be even worse than using the components with the least variability. To solve this issue, we integrate the dimensionality reduction with the subsequent approximation in the following sections, leading to an *optimized dimensionality reduction (ODR) approach*. Correspondingly, we obtain efficient lower and upper bounds in the following Sections 4–6 and more importantly, the bounds can achieve the optimal value of the original Problem (3).

4. Lower Bound

We extend the dimensionality reduction method (i.e., PCA) in (5) by introducing a decision variable $\mathbf{B}$ within a certain set $\mathcal{B}_{m_1} \subseteq \mathbb{R}^{m \times m_1}$ such that

$$\boldsymbol{\xi} = \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}} + \boldsymbol{\mu} \approx \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\boldsymbol{\xi}_{\mathrm{r}} + \boldsymbol{\mu}, \tag{10}$$

where $\mathbf{B}$ will be optimized in the subsequent PCA approximation, i.e., optimized dimensionality reduction. Note that when $\mathbf{B} = \begin{bmatrix} \mathbf{I}_{m_1} \\ \mathbf{0}_{(m-m_1)\times m_1} \end{bmatrix}$, (10) reduces to (5). By adopting (10), when we still reduce the dimensionality space from m to m_1, we allow $\mathbf{B}\boldsymbol{\xi}_{\mathrm{r}}$ to take linear combinations of the original components of $\boldsymbol{\xi}_{\mathrm{I}}$, instead of taking only the components corresponding to the largest eigenvalues. Therefore, we would like to choose a good (even an optimal) $\mathbf{B}$ to obtain a better lower bound for Problem (3) than Problem (7).

Given any $m_1 \in [m]$ and $\mathbf{B} \in \mathcal{B}_{m_1}$, we obtain a relaxation of Problem (3) by extending Problem (6). If the relaxation provides a lower bound for the optimal value of Problem (3), then we may choose the best $\mathbf{B} \in \mathcal{B}_{m_1}$ such that we obtain the largest possible lower bound. Thus, we build the following *integrated dimensionality reduction and optimization* problem:

$$\Theta_{\mathrm{L}}(m_1) = \max_{\mathbf{B}\in\mathcal{B}_{m_1}} \min_{\mathbf{x}\in\mathcal{X}} \max_{\mathbb{P}_{\mathrm{r}}\in\mathcal{D}_{\mathrm{L}}} \mathbb{E}_{\mathbb{P}_{\mathrm{r}}}\left[f\left(\mathbf{x}, \mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{B}\boldsymbol{\xi}_{\mathrm{r}} + \boldsymbol{\mu}\right)\right], \tag{11}$$

where $\mathcal{D}_{\mathrm{L}}$ is defined in (6b) with

$$\mathcal{S}_{\mathrm{r}} := \left\{\boldsymbol{\xi}_{\mathrm{r}} \in \mathbb{R}^{m_1} \,\middle|\, \mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{B}\boldsymbol{\xi}_{\mathrm{r}} + \boldsymbol{\mu} \in \mathcal{S}\right\}. \tag{12}$$

To ensure Problem (11) provides a lower bound for the optimal value of Problem (3), we need to guarantee that $\mathbf{B}\boldsymbol{\xi}_{\mathrm{r}}$ should maintain partial or all information of $\boldsymbol{\xi}_{\mathrm{I}}$ while not introducing additional statistical information. That is, the distributional ambiguity set of $\mathbf{B}\boldsymbol{\xi}_{\mathrm{r}}$ should be contained in that of $\boldsymbol{\xi}_{\mathrm{I}}$, i.e., $\mathcal{D}_{\mathrm{M}}$. It follows that

$$\mathbb{P}(\mathbf{B}\boldsymbol{\xi}_{\mathrm{r}} \in \mathcal{S}_{\mathrm{I}}) = 1, \tag{13a}$$

$$\mathbb{E}_{\mathbb{P}}[(\mathbf{B}\boldsymbol{\xi}_{\mathrm{r}})^\top]\mathbb{E}_{\mathbb{P}}[\mathbf{B}\boldsymbol{\xi}_{\mathrm{r}}] = \mathbb{E}_{\mathbb{P}}[\boldsymbol{\xi}_{\mathrm{r}}^\top](\mathbf{B}^\top\mathbf{B})\mathbb{E}_{\mathbb{P}}[\boldsymbol{\xi}_{\mathrm{r}}] \leq \gamma_1, \tag{13b}$$

$$\mathbb{E}_{\mathbb{P}}[\mathbf{B}\boldsymbol{\xi}_{\mathrm{r}}(\mathbf{B}\boldsymbol{\xi}_{\mathrm{r}})^\top] = \mathbf{B}\mathbb{E}_{\mathbb{P}}[\boldsymbol{\xi}_{\mathrm{r}}\boldsymbol{\xi}_{\mathrm{r}}^\top]\mathbf{B}^\top \preceq \gamma_2\mathbf{I}_m. \tag{13c}$$

Comparing the representation of $\mathcal{D}_{\mathrm{L}}$ with (13), we define $\mathcal{B}_{m_1} := \{\mathbf{B} \in \mathbb{R}^{m\times m_1} \mid \mathbf{B}^\top\mathbf{B} = \mathbf{I}_{m_1}\}$ and will show that Problem (11) provides a lower bound for Problem (3) under this definition (see Theorem 1). Before presenting this theorem, we prepare the following two lemmas.

LEMMA 1. *When $\mathbf{B} \in \mathbb{R}^{m\times m_1}$, the following three constraints are equivalent: (i) $\begin{bmatrix}\mathbf{I}_m & \mathbf{B}\\ \mathbf{B}^\top & \mathbf{I}_{m_1}\end{bmatrix} \succeq 0$, (ii) $\mathbf{B}\mathbf{B}^\top \preceq \mathbf{I}_m$, and (iii) $\mathbf{B}^\top\mathbf{B} \preceq \mathbf{I}_{m_1}$.*

Lemma 1 shows that both $\mathbf{B}\mathbf{B}^\top \preceq \mathbf{I}_m$ and $\mathbf{B}^\top\mathbf{B} \preceq \mathbf{I}_{m_1}$ can be reformulated as an SDP constraint $\begin{bmatrix}\mathbf{I}_m & \mathbf{B}\\ \mathbf{B}^\top & \mathbf{I}_{m_1}\end{bmatrix} \succeq 0$. Although this SDP constraint has a high dimension at $m + m_1$, it is very sparse and usually does not create additional computational challenges.

LEMMA 2. *For any matrix $\mathbf{V} \in \mathbb{R}^{m\times n}$ and symmetric matrices $\mathbf{X} \in \mathbb{R}^{m\times m}$ and $\mathbf{Y} \in \mathbb{R}^{m\times m}$, we have: (i) If $\mathbf{X} \succeq \mathbf{Y}$, then $\mathbf{V}^\top\mathbf{X}\mathbf{V} \succeq \mathbf{V}^\top\mathbf{Y}\mathbf{V}$; (ii) If $n = m$ and $\mathbf{V}$ is invertible, then $\mathbf{X} \succeq \mathbf{Y}$ is equivalent to $\mathbf{V}^\top\mathbf{X}\mathbf{V} \succeq \mathbf{V}^\top\mathbf{Y}\mathbf{V}$.*

Lemma 2 shows that a PSD matrix (e.g., $\mathbf{X} - \mathbf{Y}$) remains PSD if it is pre-multiplied by an arbitrary matrix with appropriate dimensions (e.g., $\mathbf{V}^\top$) and post-multiplied by this arbitrary matrix's transpose (e.g., $\mathbf{V}$). Furthermore, if this arbitrary matrix is invertible, then the original PSD matrix is equivalent to the matrix after the pre-multiplication and post-multiplication. With Lemmas 1 and 2, we are now ready to present the following theorem.

THEOREM 1. *The following three conclusions hold: (i) Problem (11) provides a lower bound for the optimal value of Problem (3), i.e., $\Theta_{\rm L}(m_1) \le \Theta_{\rm M}(m)$ for any $m_1 \le m$; (ii) the optimal value of Problem (11) is nondecreasing in m_1, i.e., $\Theta_{\rm L}(m_1) \le \Theta_{\rm L}(m_2)$ for any $m_1 < m_2 \le m$; and (iii) when $m_1 = m$, Problem (3) and Problem (11) have the same optimal value, i.e., $\Theta_{\rm L}(m) = \Theta_{\rm M}(m)$.*

Theorem 1 shows that we obtain a lower bound for the optimal value of Problem (3) when reducing the dimensionality space of $\boldsymbol{\xi}_{\rm I}$ *while optimizing the choice of* $\mathbf{B} \in \mathcal{B}_{m_1}$ in Problem (11). When the reduced dimensionality (i.e., m_1) is higher, we obtain a better lower bound. We maintain the optimal value of Problem (3) if the dimensionality space is not reduced (i.e., $m_1 = m$). Note that the conclusions in Theorem 1 are similar to Theorem 2 in Cheramin et al. (2022), both demonstrating the validity of dimensionality reduction in solving the moment-based DRO problems. However, here by optimizing the choice of $\mathbf{B} \in \mathcal{B}_{m_1}$, Problem (11) provides a better lower bound than Problem (6) (i.e., the PCA approximation in Cheramin et al. 2022) does because the latter problem is a special case of the former problem. More importantly, we may expect to close the gap between $\Theta_{\rm L}(m_1)$ and $\Theta_{\rm M}(m)$ when we choose a small m_1. To that end, we follow the PCA approximation (7) to reformulate Problem (11) as the following SDP formulation:

$$\Theta_{\rm L}(m_1) = \max_{\mathbf{B} \in \mathcal{B}_{m_1}} \underline{\Theta}(m_1, \mathbf{B}), \tag{14}$$

where

$$\underline{\Theta}(m_1, \mathbf{B}) := \min_{\substack{\mathbf{x}, s, \hat{\boldsymbol{\lambda}},\\ \mathbf{q}_{\rm r}, \mathbf{Q}_{\rm r}}} \; s + \gamma_2 \mathbf{I}_{m_1} \bullet \mathbf{Q}_{\rm r} + \sqrt{\gamma_1}\, \|\mathbf{q}_{\rm r}\|_2 \tag{15a}$$

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

$$\mathbf{x} \in \mathcal{X};\; \hat{\boldsymbol{\lambda}} = \{\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_K\},\; \boldsymbol{\lambda}_k \in \mathbb{R}_+^l,\; \forall k \in [K];\; \mathbf{q}_{\rm r} \in \mathbb{R}^{m_1};\; \mathbf{Q}_{\rm r} \in \mathbb{R}^{m_1 \times m_1}. \tag{15c}$$

Now we would like to find a $m_1 < m$ such that $\Theta_{\rm L}(m_1)$ in Problem (14) is close (even equal) to $\Theta_{\rm M}(m)$ in Problem (4). Note that, if $\Theta_{\rm L}(m_1) = \Theta_{\rm M}(m)$, then comparing the SDP constraints between (4) and (14) shows that the rank of $\mathbf{Q}$ in the optimal solution of Problem (4) can be smaller than m. Specifically, we have the following conclusion holds.

THEOREM 2. *Consider $K < m$ and any optimal solution $(\mathbf{x}^*, s^*, \hat{\boldsymbol{\lambda}}^*, \mathbf{q}^*, \mathbf{Q}^*)$ of Problem (4) with $S_k = s^* - y_k^0(\mathbf{x}^*) - \boldsymbol{\lambda}_k^{*\top}\mathbf{b} - y_k(\mathbf{x}^*)^\top \boldsymbol{\mu} + \boldsymbol{\lambda}_k^{*\top}\mathbf{A}\boldsymbol{\mu}$ for any $k \in [K]$. We can always construct another optimal solution $(\mathbf{x}^*, s^*, \hat{\boldsymbol{\lambda}}^*, \mathbf{q}', \mathbf{Q}')$ of Problem (4) such that $\mathrm{rank}(\mathbf{Q}') \le K$, $\mathbf{q}' = \mathbf{V}\delta$, $\mathbf{Q}' = \mathbf{V}\mathbf{Y}_{11}\mathbf{V}^\top$, and $(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}})^\top(\mathbf{A}^\top \boldsymbol{\lambda}_k^* - y_k(\mathbf{x}^*)) = \mathbf{V}\boldsymbol{\nu}_k$ for any $k \in [K]$, where $\mathbf{Y}_{11} \in \mathbb{R}^{K \times K}$, $\mathbf{Y}_{11} \succeq 0$, $\mathbf{V} = [\mathbf{v}_k, \forall k \in [K]] \in \mathbb{R}^{m \times K}$ with orthonormal vectors $\mathbf{v}_k \in \mathbb{R}^m$, $\boldsymbol{\delta} \in \mathbb{R}^K$, and $\boldsymbol{\nu}_k \in \mathbb{R}^K$ for any $k \in [K]$.*

As K is the number of pieces formulating the piecewise linear function $f(\mathbf{x},\boldsymbol{\xi})$, Theorem 2 shows that the rank of $\mathbf{Q}'$ that optimizes Problem (4) can be small. Note that when $K \geq m$, we have $\mathrm{rank}(\mathbf{Q}') \leq m \leq K$, thereby no need to consider this case in Theorem 2. Therefore, given any $m_1 \in [m]$ and $\mathbf{B} \in \mathcal{B}_{m_1}$, the rank of the optimal $\mathbf{Q}_\mathrm{r}$ in Problem (15) may also be small. With an optimized $\mathbf{B} \in \mathbb{R}^{m\times m_1}$ in Problem (14), we then would like to choose a small m_1 and find a $\mathbf{B} \in \mathcal{B}_{m_1}$ such that $\Theta_\mathrm{L}(m_1)$ can be close to $\Theta_\mathrm{M}(m)$. Specifically, we have the following conclusion holds.

THEOREM 3. *Consider the optimal solution* $(\mathbf{x}^*, s^*, \hat{\boldsymbol{\lambda}}^*, \mathbf{q}', \mathbf{Q}')$ *of Problem* (4), $S_k(\forall k \in [K])$, $\mathbf{V}$, δ, $\boldsymbol{\nu}_k(\forall k \in [K])$, *and* $\mathbf{Y}_{11}$ *that are defined in Theorem* 2. *When* $m_1 \geq K$, *there exists a feasible solution* $\mathbf{B}^\dagger = [\mathbf{V},\mathbf{C}]$ *in Problem* (14) *with* $\mathbf{C} \in \mathbb{R}^{m\times(m_1-K)}$ *and* $[\mathbf{V},\mathbf{C}]^\top[\mathbf{V},\mathbf{C}] = \mathbf{I}_{m_1}$ *and given this* $\mathbf{B}^\dagger$, *there exists a feasible solution* $(\mathbf{x}^\dagger = \mathbf{x}^*, s^\dagger = s^*, \hat{\boldsymbol{\lambda}}^\dagger = \hat{\boldsymbol{\lambda}}^*, \mathbf{q}_\mathrm{r}^\dagger = (\delta^\top, \mathbf{0}_{m_1-K}^\top)^\top, \mathbf{Q}_\mathrm{r}^\dagger = \begin{bmatrix} \mathbf{Y}_{11} & \mathbf{0}_{K\times(m_1-K)} \\ \mathbf{0}_{(m_1-K)\times K} & \mathbf{0}_{(m_1-K)\times(m_1-K)} \end{bmatrix})$ *in Problem* (15) *such that the corresponding objective value equals the optimal value of Problem* (4), $\Theta_\mathrm{M}(m)$.

Theorem 3 shows that we may reduce the dimensionality space of the random parameters from m to K while maintaining high-quality solutions. Meanwhile, when $m_1 \geq K$, we can always find a feasible solution of Problems (14) and (15) such that the corresponding objective value is equal to the optimal value of the original Problem (4). More importantly, the SDP constraints in Problem (14) have smaller sizes (i.e., $m_1 + 1$) than those in Problem (4) (i.e., $m + 1$), potentially reducing computational challenges because K is usually small (e.g., $K = 2$ in the distributionally robust CVaR problem in Example 1). We used to conjecture that this constructed feasible solution is an optimal solution of Problems (14) and (15) such that the optimal value of Problem (14) is equal to the optimal value of Problem (4) when $m_1 \geq K$. Most numerical experiments (see Section 8) show this conjecture may be correct, but we find a counter-example.

Now we provide an example as follows to illustrate that the optimal value of Problem (15) with $\mathbf{B} = \mathbf{V}$ is strictly less than the optimal value of Problem (4), which means that the constructed feasible solution ($\mathbf{B} = \mathbf{V}$) is not optimal.

EXAMPLE 2. We consider an instance of Problem (4), where $m = n = 4$, $\gamma_1 = 1$, $\gamma_2 = 2$, $\mathbf{A} = \mathbf{0}_{l\times m}$, $\mathbf{b} = \mathbf{0}_l$, $\boldsymbol{\mu} = \mathbf{1}_m$, $\boldsymbol{\Sigma} = \mathbf{I}_m$, $K = 3$, $y_k^0(\mathbf{x}) = 0$ $(\forall k \in [K])$, $y_k(\mathbf{x}) = \mathbf{W}_k\mathbf{x}$ $(\forall k \in [K])$ with $\mathbf{W}_1 = \begin{bmatrix} 1&0&0&0\\0&0&0&0\\0&0&0&0\\0&0&0&1 \end{bmatrix}$, $\mathbf{W}_2 = \begin{bmatrix} 0&0&0&0\\0&1&0&0\\0&0&0&0\\0&0&0&2 \end{bmatrix}$, and $\mathbf{W}_3 = \begin{bmatrix} 0&0&0&0\\0&0&0&0\\0&0&1&0\\0&0&0&1 \end{bmatrix}$, and $\mathcal{X} = \{\mathbf{x} \in \mathbb{R}^4 \mid x_1 = x_3 = x_4 = 1, x_2 \in \{-7, 1\}\}$, then Problem (4) becomes

$$\min_{\mathbf{x}\in\mathcal{X},s,\mathbf{q},\mathbf{Q}} \left\{ s + 2\mathbf{I}_m \bullet \mathbf{Q} + \|\mathbf{q}\|_2 \;\middle|\; \begin{bmatrix} s - \mathbf{x}^\top\mathbf{W}_k^\top\mathbf{1}_m & \frac{1}{2}(\mathbf{q} - \mathbf{W}_k\mathbf{x})^\top \\ \frac{1}{2}(\mathbf{q} - \mathbf{W}_k\mathbf{x}) & \mathbf{Q} \end{bmatrix} \succeq 0,\ \forall k \in [K] \right\}. \tag{16}$$

Solving Problem (16) gives the optimal value 5.9882 with $\mathbf{x} = [1,1,1,1]^\top$, $\mathbf{Q} = \begin{bmatrix} 0.0911 & -0.0558 & -0.0354 & -0.0558 \\ -0.0558 & 0.1115 & -0.0558 & 0.1115 \\ -0.0354 & -0.0558 & 0.0911 & -0.0558 \\ -0.0558 & 0.1115 & -0.0558 & 0.1115 \end{bmatrix}$, and $\text{rank}(\mathbf{Q}) = 2$. By Theorem 2, we can correspondingly obtain a feasible $\mathbf{V} = \begin{bmatrix} 0.7071 & -0.5774 & -0.1543 \\ 0 & 0.5774 & -0.3086 \\ 0 & 0 & 0.9258 \\ 0.7071 & 0.5774 & 0.1543 \end{bmatrix}$. Now given $m_1 = K = 3$ and $\mathbf{B} = \mathbf{V}$, Problem (15) becomes

$$\min_{\mathbf{x}\in\mathcal{X},s,\mathbf{q}_{\mathrm{r}},\mathbf{Q}_{\mathrm{r}}} \left\{ s + 2\mathbf{I}_{m_1} \bullet \mathbf{Q}_{\mathrm{r}} + \|\mathbf{q}_{\mathrm{r}}\|_2 \;\middle|\; \begin{bmatrix} s - \mathbf{x}^\top \mathbf{W}_k^\top \mathbf{1}_m & \frac{1}{2}\left(\mathbf{q}_{\mathrm{r}} - \mathbf{V}^\top \mathbf{W}_k \mathbf{x}\right)^\top \\ \frac{1}{2}\left(\mathbf{q}_{\mathrm{r}} - \mathbf{V}^\top \mathbf{W}_k \mathbf{x}\right) & \mathbf{Q}_{\mathrm{r}} \end{bmatrix} \succeq 0,\ \forall k \in [K] \right\}. \tag{17}$$

Solving Problem (17) gives the optimal value 5.1139 with $\mathbf{x} = [1,-7,1,1]^\top$. That is, the optimal value of Problem (15) with $\mathbf{B} = \mathbf{V}$ is strictly less than the optimal value of Problem (4).

Theorem 3 and Example 2 show that while the optimized dimensionality reduction maintains very high-quality solutions (mostly the optimal solutions as shown in our later numerical experiments in Section 8), we may still potentially lose some useful information that achieves the optimal solution of the original problem. To resolve this issue, we will also derive an upper bound and a new lower bound for the optimal value of the original problem in the later sections.

Note that Problem (14) is a nonconvex optimization problem due to the max-min operator. That is, we develop a low-dimensional nonconvex optimization technique to solve the original high-dimensional SDP problem, which can be significantly difficult to solve because of the large sizes of SDP matrices. To further efficiently solve Problem (14), we first reformulate it into a bilinear SDP problem (see Proposition 2) under the following assumption and then propose efficient algorithms (see Section 7) to solve it.

ASSUMPTION 2. *The set $\mathcal{X}$ is convex with at least one interior point. More specifically, we consider the convex set $\mathcal{X}$ in a generic SDP form: $\mathcal{X} = \{\mathbf{x} \in \mathbb{R}^n \mid \sum_{i=1}^n (\mathbf{\Delta}_i x_i) + \mathbf{\Delta}_0 \succeq 0\}$, where $\mathbf{\Delta}_i \in \mathbb{R}^{\tau\times\tau}$ for any $i \in \{0,1,\ldots,n\}$ and some $\tau \geq 1$.*

We use $\mathbf{a}_{ij}\mathbf{x} + a_{ij}^0$ ($\forall i \in [\tau], j \in [\tau]$) to denote the elements of the matrix $\sum_{i=1}^n (\mathbf{\Delta}_i x_i) + \mathbf{\Delta}_0$, where $\mathbf{a}_{ij}^\top \in \mathbb{R}^n$. We let $y_k^0(\mathbf{x}) = \mathbf{w}_k^0\mathbf{x} + d_k^0$ and $y_k(\mathbf{x}) = (\mathbf{w}_k^1\mathbf{x} + d_k^1, \ldots, \mathbf{w}_k^m\mathbf{x} + d_k^m)^\top = \mathbf{W}_k\mathbf{x} + \mathbf{d}_k$ for any $k \in [K]$, where $(\mathbf{w}_k^i)^\top \in \mathbb{R}^n$ for any $i \in \{0,1,\ldots,m\}$ and $k \in [K]$, $\mathbf{W}_k \in \mathbb{R}^{m\times n}$ for any $k \in [K]$, and $\mathbf{d}_k \in \mathbb{R}^m$ for any $k \in [K]$. The following proposition holds.

PROPOSITION 2. *Under Assumption 2, Problem (14) has the same optimal value as the following bilinear SDP formulation:*

$$\Theta_{\mathrm{L}}(m_1) = \max_{\substack{t_k,\mathbf{p}_k,\mathbf{P}_k,\forall k\in[K],\\ \mathbf{Z},\mathbf{B}}} \sum_{k=1}^K \left(t_k d_k^0 + \left(t_k \boldsymbol{\mu}^\top + \mathbf{p}_k^\top \left(\mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{B} \right)^\top \right) \mathbf{d}_k \right) - \sum_{i=1}^\tau \sum_{j=1}^\tau z_{ij} a_{ij}^0 \tag{18a}$$

$$\text{s.t.} \qquad 1 - \sum_{k=1}^K t_k = 0,\ \sqrt{\gamma_1} - \left\| \sum_{k=1}^K \mathbf{p}_k \right\|_2 \geq 0,\ \gamma_2 \mathbf{I}_{m_1} - \sum_{k=1}^K \mathbf{P}_k \succeq 0, \tag{18b}$$

$$t_k\left(\mathbf{A}\boldsymbol{\mu}-\mathbf{b}\right)^\top+\mathbf{p}_k^\top\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\right)^\top\mathbf{A}^\top\leq 0,\ \forall k\in[K], \tag{18c}$$

$$\sum_{k=1}^{K}\left(t_k\mathbf{w}_k^0+\left(t_k\boldsymbol{\mu}^\top+\mathbf{p}_k^\top\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\right)^\top\right)\mathbf{W}_k\right)-\sum_{i=1}^{\tau}\sum_{j=1}^{\tau}z_{ij}\mathbf{a}_{ij}=0, \tag{18d}$$

$$\begin{bmatrix} t_k & \mathbf{p}_k^\top \\ \mathbf{p}_k & \mathbf{P}_k \end{bmatrix}\succeq 0,\ \forall k\in[K],\ \mathbf{B}^\top\mathbf{B}=\mathbf{I}_{m_1},\ \mathbf{Z}\succeq 0, \tag{18e}$$

where $\mathbf{p}_k\in\mathbb{R}^{m_1}$ ($k\in[K]$), $\mathbf{P}_k\in\mathbb{R}^{m_1\times m_1}$ ($k\in[K]$), $\mathbf{Z}\in\mathbb{R}^{\tau\times\tau}$, $\mathbf{B}\in\mathbb{R}^{m\times m_1}$, and z_{ij} is the element of the matrix $\mathbf{Z}$. In addition, $\mathbf{Z}$ is the dual variable of the constraint $\sum_{i=1}^{n}(\boldsymbol{\Delta}_i x_i)+\boldsymbol{\Delta}_0\succeq 0$ in $\mathcal{X}$ and $\begin{bmatrix} t_k & \mathbf{p}_k^\top \\ \mathbf{p}_k & \mathbf{P}_k \end{bmatrix}$ ($\forall k\in[K]$) are the dual variables of constraints (15b).

As solving Problem (18) may not provide the optimal value of Problem (4), we further quantify the gap between the optimal value of Problem (18) (i.e., Problem (14)) and the original Problem (4) (i.e., Problem (3)) and show that this gap is bounded from above by a constant, as shown in the following proposition.

PROPOSITION 3. *Given any $m_1\in[m]$ and $\mathbf{B}'$ such that $(\mathbf{B}')^\top\mathbf{B}'=\mathbf{I}_{m_1}$, we use $(\mathbf{x}^*,s^*,\hat{\boldsymbol{\lambda}}^*,\mathbf{q}_{\mathrm{r}}^*,\mathbf{Q}_{\mathrm{r}}^*)$ to denote an optimal solution of Problems (14) and (15). We let $P=\sum_{k=1}^{K}(\gamma_2/4)\mathbf{I}_m\bullet\mathbf{M}_k$ and $S=\min\{S_k,\ \forall k\in[K]\}$, where*

$$\mathbf{M}_k=\left(\mathbf{B}'\mathbf{q}_{\mathrm{r}}^*+\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top\left(\mathbf{A}^\top\boldsymbol{\lambda}_k^*-y_k(\mathbf{x}^*)\right)\right)\left(\mathbf{B}'\mathbf{q}_{\mathrm{r}}^*+\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top\left(\mathbf{A}^\top\boldsymbol{\lambda}_k^*-y_k(\mathbf{x}^*)\right)\right)^\top,\ \forall k\in[K],$$
$$S_k=s^*-y_k^0(\mathbf{x}^*)-\boldsymbol{\lambda}_k^{*\top}\mathbf{b}-y_k(\mathbf{x}^*)^\top\boldsymbol{\mu}+\boldsymbol{\lambda}_k^{*\top}\mathbf{A}\boldsymbol{\mu},\ \forall k\in[K].$$

Then, it holds that

$$0\leq\Theta_{\mathrm{M}}(m)-\Theta_{\mathrm{L}}(m_1)\leq\frac{P}{S}\mathbb{1}_{\{\sqrt{P}-S<0\}}+(2\sqrt{P}-S)\mathbb{1}_{\{\sqrt{P}-S\geq 0\}}.$$

Proposition 3 shows that, without solving the original Problem (4) that could be difficult to solve, we can obtain a theoretically guaranteed high-quality solution after solving Problem (14), which has a much smaller size than Problem (4). That is, the gap between the original optimal value and the lower bound is bounded by a constant. In the next section, we will further develop an upper bound for the optimal value of Problem (4) while *closing the gap* between them.

5. Upper Bound

We develop an inner approximation for Problem (3) by relaxing the second-moment constraint in $\mathcal{D}_{\mathrm{M}}$ while optimizing the choice of components in $\boldsymbol{\xi}_{\mathrm{I}}$ to be relaxed, leading to the best possible upper bound for the optimal value of Problem (3). Specifically, given $m_1\in[m]$, we build the following optimized inner approximation of Problem (3):

$$\Theta_{\mathrm{U}}(m_1):=\min_{\mathbf{B}\in\mathcal{B}_{m_1}}\ \overline{\Theta}(m_1,\mathbf{B}), \tag{19}$$

where

$$\overline{\Theta}(m_1,\mathbf{B})=\min_{\mathbf{x}\in\mathcal{X}}\ \max_{\mathbb{P}_{\mathrm{I}}\in\mathcal{D}_{\mathrm{U}}}\ \mathbb{E}_{\mathbb{P}_{\mathrm{I}}}\left[f\left(\mathbf{x},\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}}+\boldsymbol{\mu}\right)\right] \tag{20}$$

with

$$\mathcal{D}_{\mathrm{U}}(\mathcal{S}_{\mathrm{I}},\gamma_1,\gamma_2)=\left\{\mathbb{P}_{\mathrm{I}}\;\middle|\;\begin{array}{l}\mathbb{P}_{\mathrm{I}}\left(\boldsymbol{\xi}_{\mathrm{I}}\in\mathcal{S}_{\mathrm{I}}\right)=1\\ \mathbb{E}_{\mathbb{P}_{\mathrm{I}}}\left[\boldsymbol{\xi}_{\mathrm{I}}^{\top}\right]\mathbb{E}_{\mathbb{P}_{\mathrm{I}}}\left[\boldsymbol{\xi}_{\mathrm{I}}\right]\leq\gamma_1\\ \mathbb{E}_{\mathbb{P}_{\mathrm{I}}}\left[\mathbf{B}^{\top}\boldsymbol{\xi}_{\mathrm{I}}\boldsymbol{\xi}_{\mathrm{I}}^{\top}\mathbf{B}\right]\preceq\gamma_2\mathbf{I}_{m_1}\end{array}\right\}. \tag{21}$$

The second-moment constraint in $\mathcal{D}_{\mathrm{U}}$ is relaxed from $\mathbb{E}_{\mathbb{P}_{\mathrm{I}}}[\boldsymbol{\xi}_{\mathrm{I}}\boldsymbol{\xi}_{\mathrm{I}}^{\top}]\preceq\gamma_2\mathbf{I}_m$. Intuitively, the feasible region defined by the ambiguity set $\mathcal{D}_{\mathrm{U}}$ is larger than that by $\mathcal{D}_{\mathrm{M}}$. Therefore, we have several conclusions based on this new ambiguity set $\mathcal{D}_{\mathrm{U}}$.

THEOREM 4. *The following three conclusions hold: (i) Problem (19) provides an upper bound for the optimal value of Problem (3), i.e.,* $\Theta_{\mathrm{U}}(m_1)\geq\Theta_{\mathrm{M}}(m)$ *for any* $m_1\leq m$*; (ii) the optimal value of Problem (19) is nonincreasing in* m_1*, i.e.,* $\Theta_{\mathrm{U}}(m_1)\geq\Theta_{\mathrm{U}}(m_2)$ *for any* $m_1<m_2\leq m$*; and (iii) when* $m_1=m$*, Problem (19) and Problem (3) have the same optimal value, i.e.,* $\Theta_{\mathrm{U}}(m)=\Theta_{\mathrm{M}}(m)$*.*

Theorem 4 shows that Problem (19) provides a valid upper bound for the optimal value of Problem (3), $\Theta_{\mathrm{M}}(m)$, and the upper bound is closer to $\Theta_{\mathrm{M}}(m)$ if less information is relaxed in $\mathcal{D}_{\mathrm{U}}$.

PROPOSITION 4. *Under Assumption 1, Problem (20) has the same optimal value as the following SDP formulation:*

$$\overline{\Theta}(m_1,\mathbf{B})=\min_{\substack{\mathbf{x},s,\hat{\boldsymbol{\lambda}},\\ \mathbf{q},\mathbf{Q}_{\mathrm{r}},\hat{\mathbf{u}}}}\; s+\gamma_2\mathbf{I}_{m_1}\bullet\mathbf{Q}_{\mathrm{r}}+\sqrt{\gamma_1}\,\|\mathbf{q}\|_2 \tag{22a}$$

$$\text{s.t.}\quad \begin{bmatrix} s-y_k^0(\mathbf{x})-\boldsymbol{\lambda}_k^{\top}\mathbf{b}-y_k(\mathbf{x})^{\top}\boldsymbol{\mu}+\boldsymbol{\lambda}_k^{\top}\mathbf{A}\boldsymbol{\mu} & \frac{1}{2}\mathbf{u}_k^{\top}\\ \frac{1}{2}\mathbf{u}_k & \mathbf{Q}_{\mathrm{r}}\end{bmatrix}\succeq 0,\ \forall k\in[K], \tag{22b}$$

$$\mathbf{q}+\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^{\top}\left(\mathbf{A}^{\top}\boldsymbol{\lambda}_k-y_k(\mathbf{x})\right)=\mathbf{B}\mathbf{u}_k,\ \forall k\in[K], \tag{22c}$$

$$\mathbf{x}\in\mathcal{X},\ \mathbf{q}\in\mathbb{R}^m,\ \mathbf{Q}_{\mathrm{r}}\in\mathbb{R}^{m_1\times m_1}, \tag{22d}$$

$$\hat{\boldsymbol{\lambda}}=\{\boldsymbol{\lambda}_1,\ldots,\boldsymbol{\lambda}_K\},\ \boldsymbol{\lambda}_k\in\mathbb{R}_+^l,\ \hat{\mathbf{u}}=\{\mathbf{u}_1,\ldots,\mathbf{u}_K\},\ \mathbf{u}_k\in\mathbb{R}^{m_1},\ \forall k\in[K]. \tag{22e}$$

Proposition 4 shows that Problem (19) can be updated by replacing its inner optimization Problem (20) with Problem (22). With the updated Problem (19), we have the following conclusion.

THEOREM 5. *Consider the optimal solution* $(\mathbf{x}^*,s^*,\hat{\boldsymbol{\lambda}}^*,\mathbf{q}',\mathbf{Q}')$ *of Problem (4),* S_k $(\forall k\in[K])$, $\mathbf{V}$, $\boldsymbol{\delta}$, $\mathbf{Y}_{11}$, *and* $\boldsymbol{\nu}_k$ $(\forall k\in[K])$ *that are defined in Theorem 2. If* $m_1\geq K$*, then* $\Theta_{\mathrm{U}}(m_1)=\Theta_{\mathrm{M}}(m)$*. Specifically, when* $m_1=K$*, there exist optimal* $\mathbf{B}=\mathbf{V}$ *and* $\mathbf{Q}_{\mathrm{r}}=\mathbf{Y}_{11}$ *in Problem (19) such that* $\Theta_{\mathrm{U}}(m_1)=\Theta_2(m_1,\mathbf{V})$*.*

Theorem 5 shows that when $m_1\geq K$, the optimal value of Problem (19) is always equal to the optimal value of the original problem, $\Theta_{\mathrm{M}}(m)$. We may interpret the insights as follows. Comparing the inner-approximation Problem (19) and the original Problem (3), we can notice that they differ only in the second-moment constraints in their ambiguity sets. When $m_1=K$, we relax the original second-moment constraint from $\mathbb{E}_{\mathbb{P}_{\mathrm{I}}}[\boldsymbol{\xi}_{\mathrm{I}}\boldsymbol{\xi}_{\mathrm{I}}^{\top}]\preceq\gamma_2\mathbf{I}_m$ to $\mathbf{B}^{\top}\mathbb{E}_{\mathbb{P}_{\mathrm{I}}}[\boldsymbol{\xi}_{\mathrm{I}}\boldsymbol{\xi}_{\mathrm{I}}^{\top}]\mathbf{B}\preceq\gamma_2\mathbf{I}_K$ with

$\mathbf{B} \in \mathcal{B}_K$. That is, this relaxation eventually does not lead to a different optimal value. Specifically, under a worst-case distribution $\mathbb{P}_{\mathrm{I}}^*$ generated by solving Problem (3), we have $\mathbb{E}_{\mathbb{P}_{\mathrm{I}}^*}[\boldsymbol{\xi}_{\mathrm{I}}\boldsymbol{\xi}_{\mathrm{I}}^\top] \preceq \gamma_2 \mathbf{I}_m$ may be equivalent to $\mathbf{B}^\top \mathbb{E}_{\mathbb{P}_{\mathrm{I}}^*}[\boldsymbol{\xi}_{\mathrm{I}}\boldsymbol{\xi}_{\mathrm{I}}^\top]\mathbf{B} \preceq \gamma_2 \mathbf{I}_K$. Such equivalence largely depends on the property provided by Theorem 2, which states that the rank of an optimal solution of $\mathbf{Q}$ of Problem (4) is not larger than K. Note that the variable $\mathbf{Q}$ in Problem (4) is a dual variable with respect to the second-moment constraint $\mathbb{E}_{\mathbb{P}_{\mathrm{I}}}[\boldsymbol{\xi}_{\mathrm{I}}\boldsymbol{\xi}_{\mathrm{I}}^\top] \preceq \gamma_2 \mathbf{I}_m$, indicating that the rank of $\mathbb{E}_{\mathbb{P}_{\mathrm{I}}^*}[\boldsymbol{\xi}_{\mathrm{I}}\boldsymbol{\xi}_{\mathrm{I}}^\top]$ may not be large. Specifically, we have the following proposition holds.

PROPOSITION 5. *For any PSD matrix $\mathbf{X} \in \mathbb{R}^{m\times m}$ such that $\operatorname{rank}(\mathbf{X}) \le m_1 \le m$, we have the following equivalence holds:*

$$\mathbf{X} \preceq \mathbf{I}_m \iff \mathbf{B}^\top \mathbf{X}\mathbf{B} \preceq \mathbf{I}_{m_1},\ \forall \mathbf{B} \in \mathcal{B}_{m_1}.$$

COROLLARY 1. *For any PSD matrix $\mathbf{X} \in \mathbb{R}^{m\times m}$ and $\operatorname{rank}(\mathbf{X}) \le m_1 \le m$, there exists a $\mathbf{B} \in \mathcal{B}_{m_1}$ such that $\mathbf{X} \preceq \mathbf{I}_m$ is equivalent to $\mathbf{B}^\top \mathbf{X}\mathbf{B} \preceq \mathbf{I}_{m_1}$.*

In the context of solving Problem (3) and its inner-approximation Problem (19), Proposition 5 and Corollary 1 show that there exist a worst-case distribution $\mathbb{P}_{\mathrm{I}}^* \in \mathcal{D}_{\mathrm{M}}$ such that the rank of $\mathbb{E}_{\mathbb{P}_{\mathrm{I}}^*}[\boldsymbol{\xi}_{\mathrm{I}}\boldsymbol{\xi}_{\mathrm{I}}^\top]$ is not larger than K and an optimal solution $\mathbf{B}^* \in \mathcal{B}_{m_1}$ such that $\mathbb{E}_{\mathbb{P}_{\mathrm{I}}^*}[\boldsymbol{\xi}_{\mathrm{I}}\boldsymbol{\xi}_{\mathrm{I}}^\top] \preceq \gamma_2 \mathbf{I}_m$ is equivalent to $(\mathbf{B}^*)^\top \mathbb{E}_{\mathbb{P}_{\mathrm{I}}^*}[\boldsymbol{\xi}_{\mathrm{I}}\boldsymbol{\xi}_{\mathrm{I}}^\top]\mathbf{B}^* \preceq \gamma_2 \mathbf{I}_K$. As such, even when we use a relaxed second-moment constraint, Problem (19) with $m_1 \ge K$ does not lose the optimality.

6. Lower Bound Revisited

Given that Problem (19) with $m_1 = K$ maintains the optimal value of the original Problem (3), we can further perform dimensionality reduction based on Problem (19) as we do in Section 4, thereby obtaining a new lower bound for the optimal value of Problem (3). Specifically, we consider $K \le m$ and recall that $\mathcal{B}_K = \{\mathbf{B} \in \mathbb{R}^{m\times K} \mid \mathbf{B}^\top\mathbf{B} = \mathbf{I}_K\}$. Given any $m_1 \in [K]$, we consider

$$\min_{\mathbf{B}\in\mathcal{B}_K} \min_{\mathbf{x}\in\mathcal{X}} \max_{\mathbb{P}_{\mathrm{I}}\in\mathcal{D}_{\mathrm{L2}}} \mathbb{E}_{\mathbb{P}_{\mathrm{I}}}\left[f\left(\mathbf{x}, \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}} + \boldsymbol{\mu}\right)\right] \tag{23}$$

with

$$\mathcal{D}_{\mathrm{L2}}(\mathcal{S}_{\mathrm{I}}, \gamma_1, \gamma_2) = \left\{ \mathbb{P}_{\mathrm{I}} \;\middle|\; \begin{array}{l} \mathbb{P}_{\mathrm{I}}(\boldsymbol{\xi}_{\mathrm{I}} \in \mathcal{S}_{\mathrm{I}}) = 1 \\ \mathbb{E}_{\mathbb{P}_{\mathrm{I}}}\left[\boldsymbol{\xi}_{\mathrm{I}}^\top\right]\mathbb{E}_{\mathbb{P}_{\mathrm{I}}}\left[\boldsymbol{\xi}_{\mathrm{I}}\right] \le \gamma_1 \\ \mathbb{E}_{\mathbb{P}_{\mathrm{I}}}\left[\mathbf{B}_1^\top\boldsymbol{\xi}_{\mathrm{I}}\boldsymbol{\xi}_{\mathrm{I}}^\top\mathbf{B}_1\right] \preceq \gamma_2 \mathbf{I}_{m_1} \\ \mathbb{E}_{\mathbb{P}_{\mathrm{I}}}\left[\mathbf{B}_2^\top\boldsymbol{\xi}_{\mathrm{I}}\boldsymbol{\xi}_{\mathrm{I}}^\top\mathbf{B}_2\right] \preceq \mathbf{0}_{(K-m_1)\times(K-m_1)} \end{array} \right\},$$

$\mathbf{B} = [\mathbf{B}_1, \mathbf{B}_2]$, $\mathbf{B}_1 \in \mathbb{R}^{m\times m_1}$, and $\mathbf{B}_2 \in \mathbb{R}^{m\times(K-m_1)}$. To obtain the above ambiguity set $\mathcal{D}_{\mathrm{L2}}$, we shrink the ambiguity set $\mathcal{D}_{\mathrm{U}}$ of Problem (19) by replacing the second-moment constraint $\mathbb{E}_{\mathbb{P}_{\mathrm{I}}}[\mathbf{B}^\top\boldsymbol{\xi}_{\mathrm{I}}\boldsymbol{\xi}_{\mathrm{I}}^\top\mathbf{B}] \preceq \gamma_2\mathbf{I}_K$ with $\mathbb{E}_{\mathbb{P}_{\mathrm{I}}}[\mathbf{B}_1^\top\boldsymbol{\xi}_{\mathrm{I}}\boldsymbol{\xi}_{\mathrm{I}}^\top\mathbf{B}_1] \preceq \gamma_2\mathbf{I}_{m_1}$ and $\mathbb{E}_{\mathbb{P}_{\mathrm{I}}}[\mathbf{B}_2^\top\boldsymbol{\xi}_{\mathrm{I}}\boldsymbol{\xi}_{\mathrm{I}}^\top\mathbf{B}_2] \preceq \mathbf{0}_{(K-m_1)\times(K-m_1)}$. The constraint $\mathbb{E}_{\mathbb{P}_{\mathrm{I}}}[\mathbf{B}_2^\top\boldsymbol{\xi}_{\mathrm{I}}\boldsymbol{\xi}_{\mathrm{I}}^\top\mathbf{B}_2] \preceq \mathbf{0}_{(K-m_1)\times(K-m_1)}$ implies that we project the random vector $\boldsymbol{\xi}_{\mathrm{I}}$ to the space spanned by $\mathbf{B}_2$ and the second-moment value of the projected random vector is fixed at 0. By doing so, we may slightly lose some information to characterize the distribution $\mathbb{P}_{\mathrm{I}}$, but we can obtain a formulation with an even smaller size than Problem (19) and maintain high-quality solutions. Specifically, the following theorem holds.

THEOREM 6. *Under Assumption 1, by dualizing the inner maximization problem of Problem (23), we obtain the following SDP formulation:*

$$\Theta_{\mathrm{L2}}(m_1) = \min_{\substack{\mathbf{x},s,\hat{\boldsymbol{\lambda}},\\ \mathbf{q},\mathbf{Q}_{\mathrm{r}}',\hat{\mathbf{u}}',\hat{\mathbf{u}}'',\\ \mathbf{B}_1,\mathbf{B}_2}} \quad s + \gamma_2 \mathbf{I}_{m_1} \bullet \mathbf{Q}_{\mathrm{r}}' + \sqrt{\gamma_1}\,\|\mathbf{q}\|_2 \tag{24a}$$

$$\text{s.t.} \quad \begin{bmatrix} s - y_k^0(\mathbf{x}) - \boldsymbol{\lambda}_k^\top \mathbf{b} - y_k(\mathbf{x})^\top \boldsymbol{\mu} + \boldsymbol{\lambda}_k^\top \mathbf{A}\boldsymbol{\mu} & \frac{1}{2}(\mathbf{u}_k')^\top \\ \frac{1}{2}\mathbf{u}_k' & \mathbf{Q}_{\mathrm{r}}' \end{bmatrix} \succeq 0,\ \forall k \in [K], \tag{24b}$$

$$\mathbf{q} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top \left(\mathbf{A}^\top \boldsymbol{\lambda}_k - y_k(\mathbf{x})\right) = \mathbf{B}_1 \mathbf{u}_k' + \mathbf{B}_2 \mathbf{u}_k'',\ \forall k \in [K], \tag{24c}$$

$$\mathbf{x} \in \mathcal{X},\ \mathbf{q} \in \mathbb{R}^m,\ \mathbf{Q}_{\mathrm{r}}' \in \mathbb{R}^{m_1 \times m_1}, \tag{24d}$$

$$\mathbf{B}_1 \in \mathbb{R}^{m \times m_1}, \mathbf{B}_2 \in \mathbb{R}^{m \times (K-m_1)},\ [\mathbf{B}_1, \mathbf{B}_2]^\top [\mathbf{B}_1, \mathbf{B}_2] = \mathbf{I}_K, \tag{24e}$$

$$\hat{\boldsymbol{\lambda}} = \{\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_K\},\ \boldsymbol{\lambda}_k \in \mathbb{R}_+^l,\ \forall k \in [K], \tag{24f}$$

$$\hat{\mathbf{u}}' = \{\mathbf{u}_1', \ldots, \mathbf{u}_K'\},\ \mathbf{u}_k' \in \mathbb{R}^{m_1},\ \forall k \in [K], \tag{24g}$$

$$\hat{\mathbf{u}}'' = \{\mathbf{u}_1'', \ldots, \mathbf{u}_K''\},\ \mathbf{u}_k'' \in \mathbb{R}^{K-m_1},\ \forall k \in [K]. \tag{24h}$$

In addition, the following three conclusions hold: (i) Problem (24) provides a lower bound for the optimal value of Problem (4), i.e., $\Theta_{\mathrm{L2}}(m_1) \le \Theta_{\mathrm{M}}(m)$ for any $m_1 \le K$; (ii) the optimal value of Problem (24) is nondecreasing in m_1, i.e., $\Theta_{\mathrm{L2}}(m_1) \le \Theta_{\mathrm{L2}}(m_2)$ for any $m_1 < m_2 \le K$; and (iii) when $m_1 = K$, Problem (24) and Problem (4) have the same optimal value, i.e., $\Theta_{\mathrm{L2}}(K) = \Theta_{\mathrm{M}}(m)$.

Recall that the lower bound provided by Problem (14) may not achieve the optimal value of the original Problem (4) when reducing the dimensionality to K. However, the new lower bound provided by Problem (24) achieves the optimal value of the original problem when $m_1 = K$.

7. Efficient Algorithm

In Sections 4–6, we provide two outer approximations (i.e., Problems (18) and (24)) leading to lower bounds for the optimal value of Problem (3) and an inner approximation (i.e., Problem (19)) leading to an upper bound. Although these approximations have matrices with smaller sizes than the original problem (i.e., Problem (4)), they are nonconvex with bilinear terms. We derive the Alternating Direction Method of Multipliers (ADMM) based on Hajinezhad and Shi (2018) to solve them efficiently.

Consider the following nonconvex optimization problem with bilinear terms:

$$\min_{\mathbf{X},\mathbf{Y},\mathbf{Z}} \left\{ g(\mathbf{X},\mathbf{Y},\mathbf{Z}) = f(\mathbf{Z}) + r_1(\mathbf{X}) + r_2(\mathbf{Y}) \;\middle|\; \mathbf{Z} = \mathbf{X}\mathbf{Y} \right\}, \tag{25}$$

where $\mathbf{X} \in \mathbb{R}^{P\times K}$, $\mathbf{Y} \in \mathbb{R}^{K\times Q}$, and $\mathbf{Z} \in \mathbb{R}^{P\times Q}$, and the augmented Lagrangian for Problem (25):

$$\mathcal{L}_0\left(\mathbf{Y};(\mathbf{X},\mathbf{Z});\mathbf{\Lambda}\right) = f(\mathbf{Z}) + r_1(\mathbf{X}) + r_2(\mathbf{Y}) + \mathbf{\Lambda} \bullet (\mathbf{Z} - \mathbf{X}\mathbf{Y}) + \frac{\rho}{2}\|\mathbf{Z} - \mathbf{X}\mathbf{Y}\|_{\mathrm{F}}^2$$

where $\mathbf{\Lambda} \in \mathbb{R}^{P\times Q}$ represents a matrix with the Lagrangian multipliers, $\rho > 0$ is the penalty parameter, and $\|\cdot\|_{\mathrm{F}}^2$ represents the Frobenius norm. When both $r_1(\mathbf{X})$ and $r_2(\mathbf{Y})$ are convex functions, Hajinezhad and Shi (2018) derive the following ADMM Algorithm 1 and show that any limit point of the sequence generated by Algorithm 1 is a stationary solution for Problem (25) under certain assumptions (i.e., Assumption A in Hajinezhad and Shi 2018).

Algorithm 1 Generic ADMM for Problem (25)

Initialize: $\mathbf{Y}^0, \mathbf{\Lambda}^0$
Repeat: update $(\mathbf{X},\mathbf{Z}), \mathbf{Y}$ and $\mathbf{\Lambda}$ alternatingly by

$$(\mathbf{X},\mathbf{Z})^{i+1} = \arg\min_{(\mathbf{X},\mathbf{Z})} \mathcal{L}_0\left((\mathbf{X},\mathbf{Z});\mathbf{Y}^i,\mathbf{\Lambda}^i\right);$$

$$\mathbf{Y}^{i+1} = \arg\min_{\mathbf{Y}} \mathcal{L}_0\left(\mathbf{Y};(\mathbf{X},\mathbf{Z})^{i+1},\mathbf{\Lambda}^i\right);$$

$$\mathbf{\Lambda}^{i+1} = \mathbf{\Lambda}^i + \rho\left(\mathbf{Z}^{i+1} - \mathbf{X}^{i+1}\mathbf{Y}^{i+1}\right)$$

Until Convergence.

Based on Algorithm 1 and its convergence property, we derive detailed ADMM algorithms for Problems (18) and (24) (i.e., lower bounds) and Problem (19) (i.e., upper bound), leading to three ADMM algorithms. Because (i) these algorithms are similar and all converge to stationary solutions and (ii) the formulation for Problem (19) is simpler than those for Problems (18) and (24), we only introduce the algorithmic details for solving Problem (19) thereafter.

7.1. ADMM for Problem (19)

Recall that Problem (19) is formulated as follows:

$$\Theta_{\mathrm{U}}(m_1) = \min_{\substack{\mathbf{B},\mathbf{x},s,\hat{\boldsymbol{\lambda}},\\ \mathbf{q},\mathbf{Q}_{\mathrm{r}},\hat{\mathbf{u}}}} \; s + \gamma_2 \mathbf{I}_{m_1} \bullet \mathbf{Q}_{\mathrm{r}} + \sqrt{\gamma_1}\,\|\mathbf{q}\|_2 \tag{26a}$$

$$\text{s.t.} \quad \begin{bmatrix} s - y_k^0(\mathbf{x}) - \boldsymbol{\lambda}_k^\top \mathbf{b} - y_k(\mathbf{x})^\top \boldsymbol{\mu} + \boldsymbol{\lambda}_k^\top \mathbf{A}\boldsymbol{\mu} & \frac{1}{2}\mathbf{u}_k^\top \\ \frac{1}{2}\mathbf{u}_k & \mathbf{Q}_{\mathrm{r}} \end{bmatrix} \succeq 0,\ \forall k \in [K], \tag{26b}$$

$$\mathbf{q} + \left(\mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}\right)^\top \left(\mathbf{A}^\top \boldsymbol{\lambda}_k - y_k(\mathbf{x})\right) = \mathbf{B}\mathbf{u}_k,\ \forall k \in [K] \tag{26c}$$

$$\mathbf{x} \in \mathcal{X},\ \mathbf{B}^\top\mathbf{B} = \mathbf{I}_{m_1},\ \mathbf{B} \in \mathbb{R}^{m\times m_1},\ \mathbf{q} \in \mathbb{R}^m,\ \mathbf{Q}_{\mathrm{r}} \in \mathbb{R}^{m_1\times m_1}, \tag{26d}$$

$$\hat{\boldsymbol{\lambda}} = \{\boldsymbol{\lambda}_1,\ldots,\boldsymbol{\lambda}_K\},\ \boldsymbol{\lambda}_k \in \mathbb{R}_+^l,\ \hat{\mathbf{u}} = \{\mathbf{u}_1,\ldots,\mathbf{u}_K\},\ \mathbf{u}_k \in \mathbb{R}^{m_1},\ \forall k \in [K]. \tag{26e}$$

Note that $\mathbf{B}^\top\mathbf{B} = \mathbf{I}_{m_1}$ is a nonconvex constraint. When performing the ADMM algorithm, we have to solve a nonconvex optimization problem in each iteration. To solve this issue, we introduce $\tilde{\mathbf{u}}_k = \mathbf{B}\mathbf{u}_k$ for any $k \in [K]$, $\mathbf{C} = \mathbf{B}$, and $\tilde{\mathbf{B}} = \mathbf{C}^\top\mathbf{B}$, and rewrite Problem (26) as follows:

$$\Theta_{\mathrm{U}}(m_1) = \min_{\substack{\mathbf{B},\mathbf{C},\tilde{\mathbf{B}},\\ \mathbf{x},s,\mathbf{q},\mathbf{Q}_{\mathrm{r}},\\ \boldsymbol{\lambda}_k,\mathbf{u}_k,\tilde{\mathbf{u}}_k,\forall k\in[K]}} \; s + \gamma_2 \mathbf{I}_{m_1} \bullet \mathbf{Q}_{\mathrm{r}} + \sqrt{\gamma_1}\,\|\mathbf{q}\|_2 \tag{27a}$$

$$\text{s.t.} \quad \begin{bmatrix} s - y_k^0(\mathbf{x}) - \boldsymbol{\lambda}_k^\top \mathbf{b} - y_k(\mathbf{x})^\top \boldsymbol{\mu} + \boldsymbol{\lambda}_k^\top \mathbf{A}\boldsymbol{\mu} & \frac{1}{2}\mathbf{u}_k^\top \\ \frac{1}{2}\mathbf{u}_k & \mathbf{Q}_\mathrm{r} \end{bmatrix} \succeq 0, \ \forall k \in [K], \tag{27b}$$

$$\mathbf{q} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top \left(\mathbf{A}^\top \boldsymbol{\lambda}_k - y_k(\mathbf{x})\right) = \tilde{\mathbf{u}}_k, \ \forall k \in [K], \tag{27c}$$

$$\mathbf{x} \in \mathcal{X}, \ \tilde{\mathbf{B}} = \mathbf{I}_{m_1}, \ \boldsymbol{\lambda}_k \in \mathbb{R}_+^l, \ \forall k \in [K], \tag{27d}$$

$$\tilde{\mathbf{u}}_k = \mathbf{B}\mathbf{u}_k, \ \forall k \in [K], \ \mathbf{C} = \mathbf{B}, \ \tilde{\mathbf{B}} = \mathbf{C}^\top \mathbf{B}, \tag{27e}$$

where $\mathbf{u}_k \in \mathbb{R}^{m_1}$, $\tilde{\mathbf{u}}_k \in \mathbb{R}^m, \forall k \in [K]$, $\mathbf{q} \in \mathbb{R}^m$, $\mathbf{Q}_\mathrm{r} \in \mathbb{R}^{m_1 \times m_1}$, $\mathbf{B} \in \mathbb{R}^{m \times m_1}$, $\mathbf{C} \in \mathbb{R}^{m \times m_1}$, and $\tilde{\mathbf{B}} \in \mathbb{R}^{m_1 \times m_1}$. We then consider the following augmented Lagrangian function for Problem (27):

$$\begin{aligned} &\mathcal{L}_\mathrm{U}\left(\mathbf{B};\ \mathbf{x}, s, \mathbf{q}, \mathbf{Q}_\mathrm{r}, \boldsymbol{\lambda}_k, \tilde{\mathbf{u}}_k, \mathbf{u}_k, \forall k \in [K], \mathbf{C}, \tilde{\mathbf{B}};\ \boldsymbol{\beta}_k, \forall k \in [K], \boldsymbol{\Lambda}_{\mathrm{U}1}, \boldsymbol{\Lambda}_{\mathrm{U}2}\right) \\ =& s + \gamma_2 \mathbf{I}_{m_1} \bullet \mathbf{Q}_\mathrm{r} + \sqrt{\gamma_1}\|\mathbf{q}\|_2 + \sum_{k=1}^K \left(\boldsymbol{\beta}_k^\top (\tilde{\mathbf{u}}_k - \mathbf{B}\mathbf{u}_k)\right) + \boldsymbol{\Lambda}_{\mathrm{U}1} \bullet \left(\tilde{\mathbf{B}} - \mathbf{C}^\top \mathbf{B}\right) + \boldsymbol{\Lambda}_{\mathrm{U}2} \bullet (\mathbf{C} - \mathbf{B}) \\ &+ \sum_{k=1}^K \frac{\rho}{2} (\tilde{\mathbf{u}}_k - \mathbf{B}\mathbf{u}_k)^\top (\tilde{\mathbf{u}}_k - \mathbf{B}\mathbf{u}_k) + \frac{\rho}{2}\left\|\tilde{\mathbf{B}} - \mathbf{C}^\top \mathbf{B}\right\|_\mathrm{F}^2 + \frac{\rho}{2}\|\mathbf{C} - \mathbf{B}\|_\mathrm{F}^2 + \mathbb{1}((27\mathrm{b}) - (27\mathrm{d})), \end{aligned}$$

where $\boldsymbol{\beta}_k \in \mathbb{R}^m$ ($\forall k \in [K]$), $\boldsymbol{\Lambda}_{\mathrm{U}1} \in \mathbb{R}^{m_1 \times m_1}$, and $\boldsymbol{\Lambda}_{\mathrm{U}2} \in \mathbb{R}^{m \times m_1}$ are Lagrangian multipliers, $\rho > 0$ is the penalty parameter, and $\mathbb{1}((27\mathrm{b}) - (27\mathrm{d}))$ is an indicator function that takes 0 if $(\mathbf{B}, \mathbf{C}, \tilde{\mathbf{B}}, \mathbf{x}, s, \mathbf{q}, \mathbf{Q}_\mathrm{r}, \boldsymbol{\lambda}_k, \mathbf{u}_k, \tilde{\mathbf{u}}_k, \forall k \in [K])$ satisfies constraints (27b)–(27d) and takes $+\infty$ otherwise.

Algorithm 2 ADMM for Problem (27)

Initialize: $\mathbf{B}^0, \boldsymbol{\beta}_k^0, \forall k \in [K], \boldsymbol{\Lambda}_{\mathrm{U}1}^0, \boldsymbol{\Lambda}_{\mathrm{U}2}^0$

Repeat: update $(\mathbf{x}, s, \mathbf{q}, \mathbf{Q}_\mathrm{r}, \boldsymbol{\lambda}_k, \tilde{\mathbf{u}}_k, \mathbf{u}_k, \forall k \in [K], \mathbf{C}, \tilde{\mathbf{B}})$, $\mathbf{B}$ and $(\boldsymbol{\beta}_k, \forall k \in [K], \boldsymbol{\Lambda}_{\mathrm{U}1}, \boldsymbol{\Lambda}_{\mathrm{U}2})$ alternatingly by

$$\begin{aligned} &\left(\mathbf{x}, s, \mathbf{q}, \mathbf{Q}_\mathrm{r}, \boldsymbol{\lambda}_k, \tilde{\mathbf{u}}_k, \mathbf{u}_k, \forall k \in [K], \mathbf{C}, \tilde{\mathbf{B}}\right)^{i+1} \\ =& \mathop{\arg\min}_{(\mathbf{x}, s, \mathbf{q}, \mathbf{Q}_\mathrm{r}, \boldsymbol{\lambda}_k, \tilde{\mathbf{u}}_k, \mathbf{u}_k, \forall k \in [K], \mathbf{C}, \tilde{\mathbf{B}})} \mathcal{L}_\mathrm{U}\left(\mathbf{x}, s, \mathbf{q}, \mathbf{Q}_\mathrm{r}, \boldsymbol{\lambda}_k, \tilde{\mathbf{u}}_k, \mathbf{u}_k, \forall k \in [K], \mathbf{C}, \tilde{\mathbf{B}};\ \mathbf{B}^i,\ (\boldsymbol{\beta}_k, \forall k \in [K], \boldsymbol{\Lambda}_{\mathrm{U}1}, \boldsymbol{\Lambda}_{\mathrm{U}2})^i\right); \end{aligned} \tag{28}$$

$$\mathbf{B}^{i+1} = \mathop{\arg\min}_{\mathbf{B}} \mathcal{L}_\mathrm{U}\left(\mathbf{B};\ (\mathbf{x}, s, \mathbf{q}, \mathbf{Q}_\mathrm{r}, \boldsymbol{\lambda}_k, \tilde{\mathbf{u}}_k, \mathbf{u}_k, \forall k \in [K], \mathbf{C}, \tilde{\mathbf{B}})^{i+1},\ (\boldsymbol{\beta}_k, \forall k \in [K], \boldsymbol{\Lambda}_{\mathrm{U}1}, \boldsymbol{\Lambda}_{\mathrm{U}2})^i\right); \tag{29}$$

$$\boldsymbol{\beta}_k^{i+1} = \boldsymbol{\beta}_k^i + \rho\left(\tilde{\mathbf{u}}_k^{i+1} - \mathbf{B}^{i+1}\mathbf{u}_k^{i+1}\right), \ \forall k \in [K],$$

$$\boldsymbol{\Lambda}_{\mathrm{U}1}^{i+1} = \boldsymbol{\Lambda}_{\mathrm{U}1}^i + \rho\left(\tilde{\mathbf{B}}^{i+1} - (\mathbf{C}^{i+1})^\top \mathbf{B}^{i+1}\right), \ \boldsymbol{\Lambda}_{\mathrm{U}2}^{i+1} = \boldsymbol{\Lambda}_{\mathrm{U}2}^i + \rho\left(\mathbf{C}^{i+1} - \mathbf{B}^{i+1}\right)$$

Until Convergence.

Following the framework of Algorithm 1, we design Algorithm 2 to solve Problem (27). We initialize $\mathbf{B}^0 = \begin{bmatrix} \mathbf{I}_{m_1} \\ \mathbf{0}_{(m-m_1)\times m_1} \end{bmatrix}$ based on the PCA approximation in Cheramin et al. (2022). In this algorithm, because of the newly introduced variables $\tilde{\mathbf{u}}_k, \forall k \in [K]$, $\mathbf{C}$, and $\tilde{\mathbf{B}}$, in each iteration, we have Problem (28) is a low-dimensional SDP problem and Problem (29) is a convex quadratic program. Thus, Algorithm 3 can be performed efficiently. Meanwhile, similar to Algorithm 1, the following convergence property holds.

PROPOSITION 6. *Consider Algorithm 2 for solving Problem (27). We have that any limit point of the sequence generated by Algorithm 2 is a stationary solution for Problem (27).*

7.2. A New ADMM for Problem (26)

Although Algorithm 2 has an appealing convergence property, it may generate a solution stuck in a local optimum because we may quickly obtain a local solution $\mathbf{C} = \mathbf{B}$ without further improvements. Here we would like not to rewrite the nonconvex constraint $\mathbf{B}^\top \mathbf{B} = \mathbf{I}_{m_1}$ and decide to solve a nonconvex optimization problem in each iteration of a new ADMM algorithm. Specifically, we only introduce $\tilde{\mathbf{u}}_k = \mathbf{B}\mathbf{u}_k$ for any $k \in [K]$ and rewrite Problem (26) as follows:

$$\Theta_{\mathrm{M}}(m_1) = \min_{\substack{\mathbf{B},\mathbf{x},s,\mathbf{q},\mathbf{Q}_{\mathrm{r}},\\ \boldsymbol{\lambda}_k,\tilde{\mathbf{u}}_k,\mathbf{u}_k,\forall k\in[K]}} \quad s + \gamma_2 \mathbf{I}_{m_1} \bullet \mathbf{Q}_{\mathrm{r}} + \sqrt{\gamma_1}\,\|\mathbf{q}\|_2 \tag{30a}$$

$$\text{s.t.} \quad \begin{bmatrix} s - y_k^0(\mathbf{x}) - \boldsymbol{\lambda}_k^\top \mathbf{b} - y_k(\mathbf{x})^\top \boldsymbol{\mu} + \boldsymbol{\lambda}_k^\top \mathbf{A}\boldsymbol{\mu} & \frac{1}{2}\mathbf{u}_k^\top \\ \frac{1}{2}\mathbf{u}_k & \mathbf{Q}_{\mathrm{r}} \end{bmatrix} \succeq 0,\ \forall k \in [K], \tag{30b}$$

$$\mathbf{q} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top \left(\mathbf{A}^\top \boldsymbol{\lambda}_k - y_k(\mathbf{x})\right) = \tilde{\mathbf{u}}_k,\ \forall k \in [K], \tag{30c}$$

$$\mathbf{x} \in \mathcal{X},\ \mathbf{B}^\top \mathbf{B} = \mathbf{I}_{m_1},\ \mathbf{B} \in \mathbb{R}^{m \times m_1},\ \mathbf{q} \in \mathbb{R}^m,\ \mathbf{Q}_{\mathrm{r}} \in \mathbb{R}^{m_1 \times m_1}, \tag{30d}$$

$$\boldsymbol{\lambda}_k \in \mathbb{R}_+^l,\ \mathbf{u}_k \in \mathbb{R}^{m_1},\ \tilde{\mathbf{u}}_k \in \mathbb{R}^m,\ \forall k \in [K], \tag{30e}$$

$$\tilde{\mathbf{u}}_k = \mathbf{B}\mathbf{u}_k,\ \forall k \in [K]. \tag{30f}$$

We consider the following augmented Lagrangian function for Problem (30):

$$\mathcal{L}_{\mathrm{U2}}\left(\mathbf{B};\ \mathbf{x}, s, \mathbf{q}, \mathbf{Q}_{\mathrm{r}}, \boldsymbol{\lambda}_k, \tilde{\mathbf{u}}_k, \mathbf{u}_k, \forall k \in [K];\ \boldsymbol{\beta}_k, \forall k \in [K]\right) = s + \gamma_2 \mathbf{I}_{m_1} \bullet \mathbf{Q}_{\mathrm{r}} + \sqrt{\gamma_1}\,\|\mathbf{q}\|_2$$
$$+ \sum_{k=1}^K \left(\boldsymbol{\beta}_k^\top \left(\tilde{\mathbf{u}}_k - \mathbf{B}\mathbf{u}_k\right)\right) + \sum_{k=1}^K \frac{\rho}{2} \left(\tilde{\mathbf{u}}_k - \mathbf{B}\mathbf{u}_k\right)^\top \left(\tilde{\mathbf{u}}_k - \mathbf{B}\mathbf{u}_k\right) + \mathbb{1}\left((30\text{b}) - (30\text{e})\right),$$

where $\boldsymbol{\beta}_k \in \mathbb{R}^m$ $(\forall k \in [K])$ are Lagrangian multipliers, $\rho > 0$ is the penalty parameter, and $\mathbb{1}((30\text{b}) - (30\text{e}))$ is an indicator function that takes 0 if $(\mathbf{B}, \mathbf{x}, s, \mathbf{q}, \mathbf{Q}_{\mathrm{r}}, \boldsymbol{\lambda}_k, \tilde{\mathbf{u}}_k, \mathbf{u}_k, \forall k \in [K])$ satisfies constraints (30b)–(30e) and takes $+\infty$ otherwise.

Algorithm 3 ADMM for Problem (30)

Initialize: $\mathbf{B}^0, \boldsymbol{\beta}_k^0, \forall k \in [K]$
Repeat: update $(\mathbf{x}, s, \mathbf{q}, \mathbf{Q}_{\mathrm{r}}, \boldsymbol{\lambda}_k, \tilde{\mathbf{u}}_k, \mathbf{u}_k, \forall k \in [K])$, $\mathbf{B}$ and $\boldsymbol{\beta}_k (\forall k \in [K])$ alternatingly by

$$\left(\mathbf{x}, s, \mathbf{q}, \mathbf{Q}_{\mathrm{r}}, \boldsymbol{\lambda}_k, \tilde{\mathbf{u}}_k, \mathbf{u}_k, \forall k \in [K]\right)^{i+1} = \operatorname*{arg\,min}_{(\mathbf{x},s,\mathbf{q},\mathbf{Q}_{\mathrm{r}},\boldsymbol{\lambda}_k,\tilde{\mathbf{u}}_k,\mathbf{u}_k,\forall k\in[K])} \mathcal{L}_{\mathrm{U2}}\left(\mathbf{x}, s, \mathbf{q}, \mathbf{Q}_{\mathrm{r}}, \boldsymbol{\lambda}_k, \tilde{\mathbf{u}}_k, \mathbf{u}_k, \forall k \in [K];\ \mathbf{B}^i,\ \boldsymbol{\beta}_k^i, \forall k \in [K]\right); \tag{31}$$

$$\mathbf{B}^{i+1} = \operatorname*{arg\,min}_{\mathbf{B}} \mathcal{L}_{\mathrm{U2}}\left(\mathbf{B};\ \left(\mathbf{x}, s, \mathbf{q}, \mathbf{Q}_{\mathrm{r}}, \boldsymbol{\lambda}_k, \tilde{\mathbf{u}}_k, \mathbf{u}_k, \forall k \in [K]\right)^{i+1},\ \boldsymbol{\beta}_k^i, \forall k \in [K]\right); \tag{32}$$

$$\boldsymbol{\beta}_k^{i+1} = \boldsymbol{\beta}_k^i + \rho\left(\tilde{\mathbf{u}}_k^{i+1} - \mathbf{B}^{i+1}\mathbf{u}_k^{i+1}\right),\ \forall k \in [K]$$

Until Convergence.

Now we design Algorithm 3 to solve Problem (30). We initialize $\mathbf{B}^0 = \begin{bmatrix} \mathbf{I}_{m_1} \\ \mathbf{0}_{(m-m_1)\times m_1} \end{bmatrix}$ based on the PCA approximation. In this algorithm, Problem (31) is a low-dimensional SDP problem, while, different from Problem (29) in Algorithm 2, Problem (32) is a nonconvex optimization problem. We note that Problem (32) has the same optimal solution with the following problem:

$$\max_{\mathbf{B}^\top \mathbf{B}=\mathbf{I}_{m_1}} \sum_{k=1}^{K} \left(\boldsymbol{\beta}_k \mathbf{u}_k^\top + \rho \tilde{\mathbf{u}}_k \mathbf{u}_k^\top\right) \bullet \mathbf{B}. \tag{33}$$

Considering a similar problem $\max_{\mathbf{B}^\top \mathbf{B}=\mathbf{I}_{m_1}} \mathbf{S} \bullet \mathbf{B}$ where $\mathbf{S} \in \mathbb{R}^{m \times m_1}$ and $\mathbf{S} = \mathbf{U}_\mathrm{S} \boldsymbol{\Sigma}_\mathrm{S} \mathbf{V}_\mathrm{S}^\top$ by the singular value decomposition (SVD), Eldén and Park (1999) show that the optimal solution $\mathbf{B}^* = \mathbf{U}_\mathrm{S} \mathbf{V}_\mathrm{S}^\top$. Now by letting $\sum_{k=1}^{K} (\boldsymbol{\beta}_k \mathbf{u}_k^\top + \rho \tilde{\mathbf{u}}_k \mathbf{u}_k^\top) = \mathbf{U}_\mathrm{U} \boldsymbol{\Sigma}_\mathrm{U} \mathbf{V}_\mathrm{U}^\top$, we can easily obtain the optimal solution of Problem (33) (i.e., Problem (32)) as $\mathbf{B}^* = \mathbf{U}_\mathrm{U} \mathbf{V}_\mathrm{U}^\top$. Therefore, although Problem (32) is nonconvex, it can be solved extremely easily, by which Algorithm 3 can be performed efficiently.

8. Numerical Experiments

We perform extensive numerical experiments to demonstrate the effectiveness of our ODR approach in solving two moment-based DRO problems: multiproduct newsvendor and CVaR problems. The mathematical models are implemented in MATLAB R2022a (ver. 9.12) by the modeling language CVX (ver. 2.2) and solved by the Mosek solver (ver. 9.3.20) on a PC with 64-bit Windows Operating System, an Intel(R) Xeon(R) W-2195 CPU @ 2.30GHz processor, and a 128 GB of memory. The time limit for each run is set at 2 hours. In Section 8.1, we specify the proposed inner and outer approximations under the ODR approach in the context of the multiproduct newsvendor and CVaR problems. In Section 8.2, we report and analyze all the numerical results.

8.1. Numerical Setup

In the deterministic multiproduct newsvendor problem, we consider m products and the demand for each product $i \in [m]$ is ξ_i. Given the wholesale, retail, and salvage prices: $\mathbf{c} \in \mathbb{R}_+^m$, $\mathbf{v} \in \mathbb{R}_+^m$, and $\mathbf{g} \in \mathbb{R}_+^m$, respectively, we decide an ordering amount $\mathbf{x} \in \mathbb{R}_+^m$ to minimize the total cost

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

Note that this piecewise linear function $f(\mathbf{x}, \boldsymbol{\xi})$ has only two pieces, i.e., $K = 2$. When the demand $\boldsymbol{\xi}$ is uncertain and its probability distribution belongs to a distributional ambiguity set $\mathcal{D}_{\mathrm{M0}}$ as defined in Section 3, we obtain the following DRO counterpart:

$$\min_{\mathbf{x} \geq 0} \max_{\mathbb{P} \in \mathcal{D}_{\mathrm{M0}}} \mathbb{E}_\mathbb{P} \left[\max\left\{(\mathbf{c} - \mathbf{v})^\top \mathbf{x}, (\mathbf{c} - \mathbf{g})^\top \mathbf{x} + (\mathbf{g} - \mathbf{v})^\top \boldsymbol{\xi}\right\}\right]. \tag{34}$$

By Proposition 1, Problem (34) has the same optimal value as the following SDP formulation:

$$\min_{\substack{\mathbf{x}, s, \boldsymbol{\lambda}_1, \\ \boldsymbol{\lambda}_2, \mathbf{q}, \mathbf{Q}}} \ s + \gamma_2 \mathbf{I}_m \bullet \mathbf{Q} + \sqrt{\gamma_1} \|\mathbf{q}\|_2 \tag{35a}$$

$$\text{s.t.} \quad \begin{bmatrix} s - (\mathbf{c}-\mathbf{v})^\top \mathbf{x} - \boldsymbol{\lambda}_1^\top (\mathbf{b} - \mathbf{A}\boldsymbol{\mu}) & \frac{1}{2}\left(\mathbf{q} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top \mathbf{A}^\top \boldsymbol{\lambda}_1\right)^\top \\ \frac{1}{2}\left(\mathbf{q} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top \mathbf{A}^\top \boldsymbol{\lambda}_1\right) & \mathbf{Q} \end{bmatrix} \succeq 0, \tag{35b}$$

$$\begin{bmatrix} s - (\mathbf{c}-\mathbf{g})^\top \mathbf{x} - \boldsymbol{\lambda}_2^\top (\mathbf{b} - \mathbf{A}\boldsymbol{\mu}) + (\mathbf{v}-\mathbf{g})^\top \boldsymbol{\mu} & \frac{1}{2}\left(\mathbf{q} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top \left(\mathbf{A}^\top \boldsymbol{\lambda}_2 + \mathbf{v} - \mathbf{g}\right)\right)^\top \\ \frac{1}{2}\left(\mathbf{q} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top \left(\mathbf{A}^\top \boldsymbol{\lambda}_2 + \mathbf{v} - \mathbf{g}\right)\right) & \mathbf{Q} \end{bmatrix} \succeq 0, \tag{35c}$$

$$\mathbf{x} \in \mathbb{R}_+^m,\ \boldsymbol{\lambda}_1 \in \mathbb{R}_+^l,\ \boldsymbol{\lambda}_2 \in \mathbb{R}_+^l,\ \mathbf{q} \in \mathbb{R}^m,\ \mathbf{Q} \in \mathbb{R}^{m\times m}. \tag{35d}$$

By the first outer approximation (18), the following problem provides a lower bound for the optimal value of Problem (35):

$$\max_{\substack{\mathbf{B}, t_1, \mathbf{p}_1, \mathbf{P}_1, \\ t_2, \mathbf{p}_2, \mathbf{P}_2}} \left(t_2 \boldsymbol{\mu}^\top + \mathbf{p}_2^\top \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\right)^\top\right)(\mathbf{g} - \mathbf{v}) \tag{36a}$$

$$\text{s.t.} \quad 1 - t_1 - t_2 = 0,\ \sqrt{\gamma_1} - \|\mathbf{p}_1 + \mathbf{p}_2\|_2 \geq 0, \tag{36b}$$

$$t_1(\mathbf{A}\boldsymbol{\mu} - \mathbf{b})^\top + \mathbf{p}_1^\top \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\right)^\top \mathbf{A}^\top \leq 0, \tag{36c}$$

$$t_2(\mathbf{A}\boldsymbol{\mu} - \mathbf{b})^\top + \mathbf{p}_2^\top \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\right)^\top \mathbf{A}^\top \leq 0, \tag{36d}$$

$$\gamma_2 \mathbf{I}_{m_1} - \mathbf{P}_1 - \mathbf{P}_2 \succeq 0,\ t_1(\mathbf{c} - \mathbf{v}) + t_2(\mathbf{c} - \mathbf{g}) \geq 0, \tag{36e}$$

$$\begin{bmatrix} t_1 & \mathbf{p}_1^\top \\ \mathbf{p}_1 & \mathbf{P}_1 \end{bmatrix} \succeq 0,\ \begin{bmatrix} t_2 & \mathbf{p}_2^\top \\ \mathbf{p}_2 & \mathbf{P}_2 \end{bmatrix} \succeq 0,\ \mathbf{B}^\top \mathbf{B} = \mathbf{I}_{m_1}, \tag{36f}$$

$$\mathbf{B} \in \mathbb{R}^{m\times m_1},\ \mathbf{p}_1 \in \mathbb{R}^{m_1},\ \mathbf{p}_2 \in \mathbb{R}^{m_1},\ \mathbf{P}_1 \in \mathbb{R}^{m_1\times m_1},\ \mathbf{P}_2 \in \mathbb{R}^{m_1\times m_1}. \tag{36g}$$

By the inner approximation (19), the following problem provides an upper bound for the optimal value of Problem (35) and reaches the optimal value of Problem (35) when $m_1 \geq 2$:

$$\min_{\substack{\mathbf{B}, \mathbf{x}, s, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \\ \mathbf{q}, \mathbf{Q}_\mathrm{r}, \mathbf{u}_1, \mathbf{u}_2}} s + \gamma_2 \mathbf{I}_{m_1} \bullet \mathbf{Q}_\mathrm{r} + \sqrt{\gamma_1}\|\mathbf{q}\|_2 \tag{37a}$$

$$\text{s.t.} \quad \begin{bmatrix} s - (\mathbf{c}-\mathbf{v})^\top \mathbf{x} - \boldsymbol{\lambda}_1^\top (\mathbf{b} - \mathbf{A}\boldsymbol{\mu}) & \frac{1}{2}\mathbf{u}_1^\top \\ \frac{1}{2}\mathbf{u}_1 & \mathbf{Q}_\mathrm{r} \end{bmatrix} \succeq 0, \tag{37b}$$

$$\begin{bmatrix} s - (\mathbf{c}-\mathbf{g})^\top \mathbf{x} - \boldsymbol{\lambda}_2^\top (\mathbf{b} - \mathbf{A}\boldsymbol{\mu}) + (\mathbf{v}-\mathbf{g})^\top \boldsymbol{\mu} & \frac{1}{2}\mathbf{u}_2^\top \\ \frac{1}{2}\mathbf{u}_2 & \mathbf{Q}_\mathrm{r} \end{bmatrix} \succeq 0, \tag{37c}$$

$$\mathbf{q} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top \mathbf{A}^\top \boldsymbol{\lambda}_1 = \mathbf{B}\mathbf{u}_1, \tag{37d}$$

$$\mathbf{q} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top \left(\mathbf{A}^\top \boldsymbol{\lambda}_2 + \mathbf{v} - \mathbf{g}\right) = \mathbf{B}\mathbf{u}_2, \tag{37e}$$

$$\mathbf{x} \in \mathbb{R}_+^m,\ \boldsymbol{\lambda}_1 \in \mathbb{R}_+^l,\ \boldsymbol{\lambda}_2 \in \mathbb{R}_+^l,\ \mathbf{B}^\top \mathbf{B} = \mathbf{I}_{m_1}, \tag{37f}$$

$$\mathbf{q} \in \mathbb{R}^m,\ \mathbf{Q}_\mathrm{r} \in \mathbb{R}^{m_1\times m_1},\ \mathbf{B} \in \mathbb{R}^{m\times m_1},\ \mathbf{u}_1 \in \mathbb{R}^{m_1},\ \mathbf{u}_2 \in \mathbb{R}^{m_1}. \tag{37g}$$

By the second outer approximation (24), the following problem with $m_1 \leq 2$ provides another lower bound for the optimal value of Problem (35) and reaches the optimal value of Problem (35) when $m_1 = 2$:

$$\min_{\substack{\mathbf{B}, \bar{\mathbf{B}}, \mathbf{x}, s, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \\ \mathbf{q}, \mathbf{Q}_\mathrm{r}, \mathbf{u}_1, \mathbf{u}_2, \mathbf{h}_1, \mathbf{h}_2}} s + \gamma_2 \mathbf{I}_{m_1} \bullet \mathbf{Q}_\mathrm{r} + \sqrt{\gamma_1}\|\mathbf{q}\|_2 \tag{38a}$$

$$\text{s.t.} \quad \begin{bmatrix} s-(\mathbf{c}-\mathbf{v})^\top\mathbf{x}-\boldsymbol{\lambda}_1^\top(\mathbf{b}-\mathbf{A}\boldsymbol{\mu}) & \frac{1}{2}\mathbf{u}_1^\top \\ \frac{1}{2}\mathbf{u}_1 & \mathbf{Q}_\mathrm{r} \end{bmatrix} \succeq 0, \tag{38b}$$

$$\begin{bmatrix} s-(\mathbf{c}-\mathbf{g})^\top\mathbf{x}-\boldsymbol{\lambda}_2^\top(\mathbf{b}-\mathbf{A}\boldsymbol{\mu})+(\mathbf{v}-\mathbf{g})^\top\boldsymbol{\mu} & \frac{1}{2}\mathbf{u}_2^\top \\ \frac{1}{2}\mathbf{u}_2 & \mathbf{Q}_\mathrm{r} \end{bmatrix} \succeq 0, \tag{38c}$$

$$\mathbf{q}+\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top \mathbf{A}^\top\boldsymbol{\lambda}_1 = \mathbf{B}\mathbf{u}_1+\bar{\mathbf{B}}\mathbf{h}_1, \tag{38d}$$

$$\mathbf{q}+\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top \left(\mathbf{A}^\top\boldsymbol{\lambda}_2+\mathbf{v}-\mathbf{g}\right) = \mathbf{B}\mathbf{u}_2+\bar{\mathbf{B}}\mathbf{h}_2, \tag{38e}$$

$$\mathbf{x}\in\mathbb{R}^m_+,\ \boldsymbol{\lambda}_1\in\mathbb{R}^l_+,\ \boldsymbol{\lambda}_2\in\mathbb{R}^l_+,\ [\mathbf{B},\bar{\mathbf{B}}]^\top[\mathbf{B},\bar{\mathbf{B}}]=\mathbf{I}_K, \tag{38f}$$

$$\mathbf{q}\in\mathbb{R}^m,\ \mathbf{Q}_\mathrm{r}\in\mathbb{R}^{m_1\times m_1},\ \mathbf{B}\in\mathbb{R}^{m\times m_1},\ \bar{\mathbf{B}}\in\mathbb{R}^{m\times(K-m_1)}, \tag{38g}$$

$$\mathbf{u}_1\in\mathbb{R}^{m_1},\ \mathbf{u}_2\in\mathbb{R}^{m_1},\ \mathbf{h}_1\in\mathbb{R}^{K-m_1},\ \mathbf{h}_2\in\mathbb{R}^{K-m_1}. \tag{38h}$$

When the dimension m of $\boldsymbol{\xi}$ is large, the original Problem (35) becomes very difficult to solve because of the large-scale SDP constraints. Nevertheless, with our ODR approach, Problems (36)–(38) have SDP matrices with very small sizes (e.g., $K+1=3$), largely reducing the computational burden while maintaining the original problem's optimal value and optimal solution $\mathbf{x}$. Note that the DRO counterpart of the CVaR problem has been introduced and reformulated in Example 1. The inner and outer approximations can be derived similarly and thus are omitted here.

8.2. Numerical Results

We compare the performance of our ODR approach (that provides two lower bounds and one upper bound) with three benchmark approaches: (i) the Mosek solver with default settings, which can provide the optimal value of the original problem; (ii) The low-rank algorithm proposed by Burer and Monteiro (2003) to solve the SDP reformulation of the original problem, i.e., Problem (4), generating a lower bound for the optimal value of Problem (4); and (iii) The existing PCA approximation proposed by Cheramin et al. (2022). For the third benchmark, we consider the reduced dimension $m_1 \in \{100\%\times m, 80\%\times m, 60\%\times m, 40\%\times m, 20\%\times m, K=2\}$, where this approach generates PCA-based lower and upper bounds for the original problem. Our proposed inner and outer approximations under the ODR approach are solved using Algorithm 3.

8.2.1. Instance Generation and Table Header Description We consider various instances of the multiproduct newsvendor and CVaR problems. In the former problem, the mean and standard deviation of $\boldsymbol{\xi}$ are randomly generated from the intervals $[0,10]$ and $[1,2]$, respectively. We further generate a correlation matrix randomly using the MATLAB function "gallery('randcorr',n)" and then convert it to a covariance matrix. We follow Xu et al. (2018) to set the wholesale, retail, and savage prices as $c_i = 0.1(5+i-1)$, $v_i = 0.15(5+i-1)$, and $g_i = 0.05(5+i-1)$ for any $i\in[m]$, respectively. Meanwhile, we consider $m \in \{100, 200, 400, 800, 1200, 1600, 2000\}$ in this problem.

In the CVaR problem, we set $\alpha = 0.05$, and the mean and standard deviation of $\boldsymbol{\xi}$, i.e., $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$, are randomly generated from the intervals $[-5,5]$ and $[1,2]$, respectively, where $\boldsymbol{\xi}$ is supported on

$[\boldsymbol{\mu}-2\boldsymbol{\sigma},\boldsymbol{\mu}+2\boldsymbol{\sigma}]$. The covariance matrix is generated similarly as above. Meanwhile, we consider $m \in \{100, 200, 400, 800, 1200, 1600, 2000\}$ in this problem.

For each of the above two problems with a given value of m, we randomly generate five instances. We report both the average and instance-level performance in six tables, with the average performance in Tables 1–2 and instance-level in Tables C2–C9 (see Appendix C). Here we describe several table headers that are shared by these tables. We use "Mosek" and "Low-rank" to represent the performance of the Mosek solver and the low-rank algorithm, respectively. The abbreviations "LB," "UB," and "RLB" represent lower, upper, and revisited lower bounds, respectively. Specifically, the labels "ODR-LB," "ODR-UB," and "ODR-RLB" denote the lower-bound performance after solving the first outer approximation (18) with $m_1 = K$, the upper-bound performance after solving the inner approximation (19) with $m_1 = K$, and the other lower-bound performance after solving the second outer approximation (24) with $m_1 = K$, respectively (see Tables 1–2 and C2–C3). The labels "PCA-LB" and "PCA-UB" denote the PCA-based lower and upper bounds provided by the PCA approximation in Cheramin et al. (2022), respectively (see Tables C4–C9). In Tables 1–2, we use "PCA-100%," "PCA-80%," "PCA-60%," "PCA-40%," "PCA-20%," and "PCA-$\frac{200}{m}$%" to denote the performance of the PCA approximation when the reduced dimension m_1 equals $100\% \times m$, $80\% \times m$, $60\% \times m$, $40\% \times m$, $20\% \times m$, and $K = 2$, respectively.

In all the tables, we use "Size" to represent the value of m and "Time" to represent the computational time in seconds required to solve each instance. We use "Gap1" (resp. "Gap2") to represent the relative gap in percentage between a lower (resp. an upper) bound and the optimal value provided by the Mosek solver. That is,

$$\text{Gap1} = \frac{\text{optimal value} - \text{lower bound}}{|\text{optimal value}|} \times 100\%, \ \text{Gap2} = \frac{\text{upper bound} - \text{optimal value}}{|\text{optimal value}|} \times 100\%.$$

We further use "Interval Gap" to represent the relative gap in percentage between a lower bound and an upper bound, i.e.,

$$\text{Interval Gap} = \frac{\text{upper bound} - \text{lower bound}}{|\text{upper bound}|} \times 100\%. \tag{39}$$

Specifically, for both the ODR approach and the low-rank algorithm, we take the objective value of "ODR-UB" as the value of "upper bound" in (39). For the PCA approximation approach, the value of "upper bound" in (39) is provided by this approach itself. The objective values of each instance provided by all the approaches, i.e., "Mosek," "Low-rank," "ODR-LB," "ODR-UB," "ODR-RLB," "PCA-LB," and "PCA-UB," are provided in Tables C2–C9.

Finally, we use "-" to represent that no result can be obtained within the time limit (i.e., two hours). For instance, the Mosek solver cannot solve the original problem to the optimality within two hours when $m \geq 400$. Hence, we cannot obtain the value of "Gap1" for the "Mosek," "ODR-LB," and "ODR-RLB" approaches.

Table 1 Average Performance on the Newsvendor Problem

	Size (m)	100	200	400	800	1200	1600	2000
Mosek	Time (secs)	13.02	363.54	-	-	-	-	-
Low-rank	Gap1 (%)	2.52	1.79	-	-	-	-	-
	Time (secs)	0.26	0.80	5.46	47.34	110.33	309.00	825.62
	Interval Gap (%)	4.27	3.66	2.67	2.32	2.28	2.36	2.52
ODR-LB	Gap1 (%)	0.09	0.00	-	-	-	-	-
	Time (secs)	0.77	0.78	0.83	0.85	1.13	2.01	2.54
	Interval Gap (%)	1.81	1.83	1.44	1.44	1.56	1.73	1.96
ODR-RLB	Gap1 (%)	0.03	0.03	-	-	-	-	-
	Time (secs)	1.95	2.60	4.33	9.75	20.83	38.36	56.68
	Interval Gap (%)	1.74	1.86	1.46	1.46	1.56	1.78	1.98
ODR-UB	Gap2 (%)	1.68	1.80	-	-	-	-	-
	Time (secs)	1.95	2.60	4.33	9.75	20.83	38.36	56.68
PCA-100%	Gap1 (%)	0.00	0.00	-	-	-	-	-
	Time (secs)	13.04	361.54	-	-	-	-	-
	Gap2 (%)	0.00	0.00	-	-	-	-	-
	Time (secs)	12.99	361.91	-	-	-	-	-
	Interval Gap (%)	0.00	0.00	-	-	-	-	-
PCA-80%	Gap1 (%)	0.50	0.31	-	-	-	-	-
	Time (secs)	5.05	120.72	3348.00	-	-	-	-
	Gap2 (%)	12.23	11.13	-	-	-	-	-
	Time (secs)	7.77	155.39	4793.72	-	-	-	-
	Interval Gap (%)	14.54	12.90	13.57	-	-	-	-
PCA-60%	Gap1 (%)	0.98	0.73	-	-	-	-	-
	Time (secs)	1.44	28.73	785.63	-	-	-	-
	Gap2 (%)	23.33	24.07	-	-	-	-	-
	Time (secs)	2.29	44.28	1196.98	-	-	-	-
	Interval Gap (%)	31.76	32.79	31.87	-	-	-	-
PCA-40%	Gap1 (%)	1.69	1.20	-	-	-	-	-
	Time (secs)	0.39	5.21	125.43	3351.00	-	-	-
	Gap2 (%)	35.79	35.94	-	-	-	-	-
	Time (secs)	0.57	8.03	177.96	5237.40	-	-	-
	Interval Gap (%)	58.50	58.26	56.45	57.18	-	-	-
PCA-20%	Gap1 (%)	2.71	1.90	-	-	-	-	-
	Time (secs)	0.15	0.43	6.49	136.97	971.60	3546.30	-
	Gap2 (%)	47.92	48.19	-	-	-	-	-
	Time (secs)	0.17	0.60	9.25	203.97	1340.46	4940.28	-
	Interval Gap (%)	97.74	84.05	90.65	93.17	92.38	94.27	-
PCA-$\frac{200}{m}$%	Gap1 (%)	4.26	3.24	-	-	-	-	-
	Time (secs)	0.11	0.12	0.13	0.20	0.26	0.36	0.50
	Gap2 (%)	57.60	59.25	-	-	-	-	-
	Time (secs)	0.13	0.14	0.16	0.22	0.32	0.46	0.60
	Interval Gap (%)	147.12	154.40	141.39	149.18	146.92	150.96	153.57

Table 2 Average Performance on the CVaR Problem

	Size (m)	100	200	400	800	1200	1600	2000
Mosek	Time (secs)	16.00	451.34	-	-	-	-	-
Low-rank	Gap1 (%)	2.81	3.34	-	-	-	-	-
	Time (secs)	2.79	6.91	20.78	75.84	171.48	635.15	1630.04
	Interval Gap (%)	3.64	3.92	4.54	4.81	7.14	4.80	4.73
ODR-LB	Gap1 (%)	0.03	0.03	-	-	-	-	-
	Time (secs)	0.97	1.31	5.03	14.85	28.85	48.23	71.18
	Interval Gap (%)	0.89	0.62	1.83	1.65	3.76	1.78	1.52
ODR-RLB	Gap1 (%)	8.29	6.65	-	-	-	-	-
	Time (secs)	2.29	4.77	9.44	39.87	84.58	113.93	172.53
	Interval Gap (%)	9.07	7.19	9.36	11.76	8.98	8.71	8.38
ODR-UB	Gap2 (%)	0.87	0.60	-	-	-	-	-
	Time (secs)	2.29	4.77	9.44	39.87	84.58	113.93	172.53
PCA-100%	Gap1 (%)	0.00	0.00	-	-	-	-	-
	Time (secs)	16.05	452.56	-	-	-	-	-
	Gap2 (%)	0.00	0.00	-	-	-	-	-
	Time (secs)	15.87	451.05	-	-	-	-	-
	Interval Gap (%)	0.00	0.00	-	-	-	-	-
PCA-80%	Gap1 (%)	34.06	33.79	-	-	-	-	-
	Time (secs)	6.19	154.19	3830.44	-	-	-	-
	Gap2 (%)	74.44	79.31	-	-	-	-	-
	Time (secs)	6.35	149.15	4172.78	-	-	-	-
	Interval Gap (%)	60.13	62.89	70.26	-	-	-	-
PCA-60%	Gap1 (%)	92.94	84.33	-	-	-	-	-
	Time (secs)	1.97	42.75	1103.60	-	-	-	-
	Gap2 (%)	170.40	157.30	-	-	-	-	-
	Time (secs)	2.12	38.33	1036.52	-	-	-	-
	Interval Gap (%)	96.94	93.32	102.80	-	-	-	-
PCA-40%	Gap1 (%)	155.36	150.26	-	-	-	-	-
	Time (secs)	0.49	6.26	152.85	4236.06	-	-	-
	Gap2 (%)	310.48	301.20	-	-	-	-	-
	Time (secs)	0.70	8.03	170.81	4185.96	-	-	-
	Interval Gap (%)	113.70	112.28	119.30	124.90	-	-	-
PCA-20%	Gap1 (%)	185.66	210.06	-	-	-	-	-
	Time (secs)	0.18	0.56	6.76	148.34	1032.09	3887.06	-
	Gap2 (%)	457.20	628.70	-	-	-	-	-
	Time (secs)	0.40	1.71	15.38	235.30	1351.00	4534.18	-
	Interval Gap (%)	115.47	119.49	119.97	122.84	127.39	127.89	-
PCA-$\frac{200}{m}$%	Gap1 (%)	208.99	260.15	-	-	-	-	-
	Time (secs)	0.11	0.12	0.14	0.19	0.28	0.43	0.53
	Gap2 (%)	490.51	745.23	-	-	-	-	-
	Time (secs)	0.35	1.34	2.32	4.47	5.73	8.24	12.65
	Interval Gap (%)	118.45	118.99	119.19	119.52	119.61	119.64	119.74

8.2.2. Numerical Performance From Tables 1–2 and C2–C9, we have the following observations. First, when $m \in \{100, 200\}$, where the Mosek solver solves each instance of the original problem to the optimality, our ODR approach performs much better than the three benchmark approaches. Both the "ODR-LB" and "ODR-RLB" provide a smaller value of "Gap1" than the low-rank algorithm and the PCA approximation, and require a shorter computational time than the three benchmark approaches. The "ODR-UB" also provides a smaller value of "Gap2" than the PCA approximation if $m_1 \neq 100\% \times m$ therein and requires shorter computational time.

Specifically, Tables C2–C3 show that the objective value of "ODR-LB" reaches the optimal value of the original problem for most of the instances, while the "ODR-LB" reduces the computational time by up to three orders of magnitude compared to the Mosek solver. The "ODR-UB" and "ODR-RLB" also provide objective values that are near-optimal for each instance and reduce the computational time significantly. Tables C4–C9 show that our ODR approach (including "ODR-LB," "ODR-UB," and "ODR-RLB") provides a better solution in terms of the objective value than the PCA approximation if the reduced dimension $m_1 \leq 80\% \times m$ in the latter approach. That is, even if we maintain 80% of the random parameters corresponding to the largest eigenvalues to be uncertain in the PCA approximation by focusing on only their statistical information, the performance is worse than our ODR approach, where we optimize the dimensionality reduction from m to $K = 2$ (i.e., maintaining only 1% of the original dimensionality size when $m = 200$). More importantly, the inner and outer approximations of our ODR approach can be solved efficiently.

Second, when $m \geq 400$, where the Mosek solver cannot solve any instance of the original problem to the optimality, our ODR approach also performs better than the benchmark approaches. The "ODR-LB" provides a smaller value of "Interval Gap" (within 2%) and requires a much shorter computational time than both the low-rank algorithm and the PCA approximation. For instance, when $m = 1600$, the low-rank algorithm and "ODR-LB" take 370.02 and 2.03 seconds to solve an instance of the multiproduct newsvendor problem and provide the value of "Interval Gap" at 1.97% and 1.36%, respectively. The PCA approximation solves this instance only when the reduced dimension m_1 is not larger than $20\% \times m$, by which it takes 3548.4 seconds while the solution quality is very poor, providing the value of "Interval Gap" at 90.8%. More importantly, our ODR approach is not sensitive to the value of m, while the benchmark approaches perform much worse when m is larger. Thus, when we cannot obtain the optimal value of the original problem, the "ODR-LB" and "ODR-UB" can be efficiently solved to provide a shorter interval that includes the optimal value than the benchmark approaches. That is, our ODR approach can provide a near-optimal solution very efficiently for the moment-DRO problems where other best possible benchmark approaches are struggling.

Third, the performance of "ODR-RLB" is similar to that of "ODR-LB," while the latter is more computationally efficient because the former has more equality constraints than the latter. Compared to "Low-rank," the "ODR-RLB" performs better in the multiproduct newsvendor problem while it provides a larger value of "Interval Gap" in the CVaR problem, and the "ODR-RLB" requires a shorter computational time in both problems. Compared to the PCA approximation, the "ODR-RLB" is better with respect to both the solution quality and computational time.

8.2.3. Numerical Insights Tables 1–2 and C2–C9 show that our ODR approach performs better than the PCA approximation with respect to the objective values for all the cases except that the "PCA-100%" (i.e., the original problem) provides the optimal value when the problem size is small, i.e., $m \in \{100, 200\}$. Note that the PCA approximation reduces the dimensionality of the random vector $\boldsymbol{\xi}$ by focusing on only the statistical information of $\boldsymbol{\xi}$, while the ODR approach integrates the dimensionality reduction with the optimization of the original problem. Here we would like to further demonstrate the benefits of our approach, thereby providing insights into how we can choose the value of $\mathbf{B}$ without solving the models in our ODR approach.

Consider the multiproduct newsvendor problem. The PCA approximation chooses the random parameters corresponding to the largest eigenvalues by maximizing the expectation of $\boldsymbol{\xi}^\top \boldsymbol{\xi}$, i.e., the variability of $\boldsymbol{\xi}$. Adopting the idea of our ODR approach to integrate the dimensionality reduction with the subsequent optimization problem, we can consider the objective function $f(\mathbf{x}, \boldsymbol{\xi})$ when choosing the random parameters in $\boldsymbol{\xi}$. Specifically, we can maximize the variability of $(\mathbf{g} - \mathbf{v})^\top \boldsymbol{\xi}$, which is the only random component in $f(\mathbf{x}, \boldsymbol{\xi})$. By (10), we solve the following problem to reduce the dimension from m to m_1:

$$
\begin{aligned}
\max_{\mathbf{B}^\top \mathbf{B} = \mathbf{I}_{m_1}} \mathbb{E}_{\mathbb{P}} \left[(\mathbf{g} - \mathbf{v})^\top \boldsymbol{\xi} \boldsymbol{\xi}^\top (\mathbf{g} - \mathbf{v}) \right] &\approx \mathbb{E}_{\mathbb{P}} \left[(\mathbf{g} - \mathbf{v})^\top \left(\mathbf{U} \boldsymbol{\Lambda}^{\frac{1}{2}} \mathbf{B} \boldsymbol{\xi}_{\mathrm{r}} + \boldsymbol{\mu} \right) \left(\mathbf{U} \boldsymbol{\Lambda}^{\frac{1}{2}} \mathbf{B} \boldsymbol{\xi}_{\mathrm{r}} + \boldsymbol{\mu} \right)^\top (\mathbf{g} - \mathbf{v}) \right] \\
&= \mathbb{E}_{\mathbb{P}} \left[(\mathbf{g} - \mathbf{v})^\top \left(\left(\mathbf{U} \boldsymbol{\Lambda}^{\frac{1}{2}} \mathbf{B} \boldsymbol{\xi}_{\mathrm{r}} \right) \left(\mathbf{U} \boldsymbol{\Lambda}^{\frac{1}{2}} \mathbf{B} \boldsymbol{\xi}_{\mathrm{r}} \right)^\top + 2 \mathbf{U} \boldsymbol{\Lambda}^{\frac{1}{2}} \mathbf{B} \boldsymbol{\xi}_{\mathrm{r}} \boldsymbol{\mu}^\top + \boldsymbol{\mu} \boldsymbol{\mu}^\top \right) (\mathbf{g} - \mathbf{v}) \right] \\
&= \mathbb{E}_{\mathbb{P}} \left[(\mathbf{g} - \mathbf{v})^\top \left(\left(\mathbf{U} \boldsymbol{\Lambda}^{\frac{1}{2}} \mathbf{B} \boldsymbol{\xi}_{\mathrm{r}} \right) \left(\mathbf{U} \boldsymbol{\Lambda}^{\frac{1}{2}} \mathbf{B} \boldsymbol{\xi}_{\mathrm{r}} \right)^\top + \boldsymbol{\mu} \boldsymbol{\mu}^\top \right) (\mathbf{g} - \mathbf{v}) \right]. \qquad (40)
\end{aligned}
$$

By introducing $\mathbf{r} = (\boldsymbol{\Lambda}^{\frac{1}{2}} \mathbf{U}^\top)(\mathbf{g} - \mathbf{v})$, Problem (40) clearly has the same optimal solution as

$$
\max_{\mathbf{B}^\top \mathbf{B} = \mathbf{I}_{m_1}} \mathbf{r}^\top \mathbf{B} \mathbf{B}^\top \mathbf{r}. \qquad (41)
$$

PROPOSITION 7. *We have* $\mathbf{B}^* = [\mathbf{r}/\|\mathbf{r}\|_2, \mathbf{0}_{m \times (m_1 - 1)}]$ *is an optimal solution of Problem* (41).

By considering the partial feature of the original optimization problem, the optimal $\mathbf{B}^*$ of Problem (40) by Proposition 7 performs better than the PCA approximation that only considers statistical information of random parameters. Note that our proposed inner and outer approximations of the ODR approach consider the complete feature of the original optimization problem and can provide an even better choice of $\mathbf{B}$. In the multiproduct newsvendor problem with

$K = 2$, we can compare the $\mathbf{B}^*$ of Problem (40) with the optimal $\mathbf{B}$ provided by our proposed outer approximation (18) with $m_1 = K$. Specifically, letting $m = 10$, we have (i) the optimal value given by the PCA approximation (lower bound) with $m_1 = K$ is -18.62; (ii) the optimal value given by (40) is -17.53 with $\mathbf{B} = \begin{bmatrix} -0.8696 & -0.0478 & 0.3285 & -0.0930 & -0.2762 & 0.2126 & -0.0456 & -0.0034 & 0.0361 & 0.0097 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}^\top$; (iii) the optimal value given by (18) (lower bound) with $m_1 = K$ is -17.38 with $\mathbf{B} = \begin{bmatrix} -0.8964 & -0.1886 & 0.2094 & -0.0327 & -0.2497 & 0.2215 & -0.0548 & -0.0216 & 0.0289 & 0.0104 \\ 0.0143 & 0.0052 & -0.0014 & -0.0004 & 0.0034 & -0.0035 & 0.0010 & 0.0006 & -0.0003 & -0.0002 \end{bmatrix}^\top$. Clearly, our ODR approach performs the best and the value of $\mathbf{B}$ from solving (40) is close to that from our ODR approach (the Frobenius norm of the difference between the two matrices is less than 0.1). That is, if a decision-maker does not have enough capacity to solve the approximations of our ODR approach, the decision-maker may consider partial feature of the optimization problem when reducing the dimensionality.

9. Conclusion

Moment-based DRO provides a theoretical framework to integrate moment-based information from available data with optimal decision-making. Extensive studies have demonstrated the effectiveness of this framework in solving various industrial applications under uncertainties. Although moment-based DRO problems can be reformulated as SDPs that can be solved in polynomial time, solving high-dimensional SDPs is significantly challenging. More importantly, high-dimensional random parameters are generally involved in industrial applications, demanding efficient approaches to solve the high-dimensional SDPs in the context of moment-based DRO.

Current approaches adopt the PCA to first reduce the dimensionality of random parameters using only the statistical information and then solve the subsequent low-dimensional approximation (SDPs). We show that performing dimensionality reduction using the components with the largest variability may not produce a good optimal value from the subsequent PCA approximation and it can be even worse than using the components with the least variability (Example 1). Thus, we integrate the dimensionality reduction with subsequent SDP problems and hence propose an optimized dimensionality reduction (ODR) approach for the moment-based DRO (Sections 4–6), aiming to drastically reduce the computational time of solving the SDP reformulations while maintaining the optimal solution of the original problem.

We first derive an outer approximation under the ODR approach to provide a lower bound for the optimal value of the original problem (Theorem 1), where the lower bound is nondecreasing in the reduced dimension m_1. We expect to choose a small m_1 to close the gap between the derived lower bound and the original optimal value. To that end, we show that the rank of each SDP matrix with respect to an optimal solution of the original high-dimensional SDP reformulation is small, guiding us on how to optimize the dimensionality reduction (Theorem 2). With this low-rank property, we observe that the derived lower bound can be close to the original optimal value (Theorem 3) but may not reach it (Example 2). Nevertheless, we show that the gap between this lower bound and the original optimal value is bounded by a constant (Proposition 3). Furthermore, we derive an inner approximation to provide an upper bound for the optimal value of the original problem (Theorem 4). More importantly, this upper bound reaches the original optimal value when the reduced dimension m_1 is small (Theorem 5). Building on this significant result, we further derive an outer approximation to provide the second lower bound for the optimal value of the original problem, where the gap between the new lower bound and the original optimal value can be closed when the reduced dimension m_1 is small (Theorem 6).

The two outer and one inner approximations derived for the original problem are all low-dimensional SDPs and nonconvex with bilinear terms (Propositions 2 and 4 and Theorem 6). We accordingly develop modified ADMM algorithms to solve them efficiently (Section 7). Finally, we demonstrate the effectiveness of our ODR approach in solving multiproduct newsvendor and CVaR problems. We compare our ODR approach and algorithms with three benchmark approaches: the Mosek solver, the low-rank algorithm by Burer and Monteiro (2003), and existing PCA approximations by Cheramin et al. (2022). Numerical results show that our ODR approach significantly outperforms these benchmark approaches in both computational time and solution quality. Our approach can obtain an optimal or near-optimal (mostly within 0.1%) solution and reduce the computational time by up to three orders of magnitude. More importantly, unlike the existing approaches that become more computationally challenging when the dimension m of random parameters increases, our approach is not sensitive to m, demonstrating significant strength in solving large-scale practical problems (Section 8.2.2). In addition, we provide insights into why our ODR approach performs better than the existing PCA approximations (Section 8.2.3).

Our research can be further extended in various directions. First, this paper considers a piecewise linear cost function in the original problem. Thus, it would be attractive to consider a more general objective function. Second, it would be interesting to apply our approach to more application problems to generate practical insights. Third, our ODR approach can be potentially generalized to solve general SDPs with certain structures. Thus, it would be appealing to exploit the structures of SDP constraints and apply the ODR approach to solve more general SDPs. We leave the above extensions for future research.

References

Abdi, H. and Williams, L. J. (2010). Principal component analysis. *Wiley interdisciplinary reviews: computational statistics*, 2(4):433–459.

Ban, G.-Y. and Rudin, C. (2019). The big data newsvendor: Practical insights from machine learning. *Operations Research*, 67(1):90–108.

Bertsimas, D. and Kallus, N. (2020). From predictive to prescriptive analytics. *Management Science*, 66(3):1025–1044.

Bertsimas, D., Sim, M., and Zhang, M. (2019). Adaptive distributionally robust optimization. *Management Science*, 65(2):604–618.

Burer, S. and Monteiro, R. D. (2003). A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization. *Mathematical Programming*, 95(2):329–357.

Candès, E. J., Li, X., Ma, Y., and Wright, J. (2011). Robust principal component analysis? *Journal of the ACM (JACM)*, 58(3):1–37.

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

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

Cheramin, M., Cheng, J., Jiang, R., and Pan, K. (2022). Computationally efficient approximations for distributionally robust optimization under moment and wasserstein ambiguity. *INFORMS Journal on Computing*, 34(3):1768–1794.

Comon, P. (1994). Independent component analysis, a new concept? *Signal Processing*, 36(3):287–314.

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

Eldén, L. and Park, H. (1999). A procrustes problem on the stiefel manifold. *Numerische Mathematik*, 82(4):599–619.

Elmachtoub, A. N. and Grigas, P. (2022). Smart "predict, then optimize". *Management Science*, 68(1):9–26.

Farnia, F. and Tse, D. (2016). A minimax approach to supervised learning. *Advances in Neural Information Processing Systems*, 29.

Gallego, G. and Moon, I. (1993). The distribution free newsboy problem: review and extensions. *Journal of the Operational Research Society*, 44(8):825–834.

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

Ghosal, S. and Wiesemann, W. (2020). The distributionally robust chance-constrained vehicle routing problem. *Operations Research*, 68(3):716–732.

Goldfarb, D. and Iyengar, G. (2003). Robust portfolio selection problems. *Mathematics of Operations Research*, 28(1):1–38.

Gotoh, J.-y. and Konno, H. (2002). Bounding option prices by semidefinite programming: A cutting plane algorithm. *Management Science*, 48(5):665–678.

Hajinezhad, D. and Shi, Q. (2018). Alternating direction method of multipliers for a class of nonconvex bilinear optimization: convergence analysis and applications. *Journal of Global Optimization*, 70(1):261–288.

He, X. and Niyogi, P. (2003). Locality preserving projections. *Advances in Neural Information Processing Systems*, 16.

Helmberg, C. (2002). Semidefinite programming. *European Journal of Operational Research*, 137(3):461–482.

Ho-Nguyen, N. and Kılınç-Karzan, F. (2022). Risk guarantees for end-to-end prediction and optimization processes. *Management Science*, 68(12):8680–8698.

Huang, D., Jiang, F., Li, K., Tong, G., and Zhou, G. (2022). Scaled PCA: A new approach to dimension reduction. *Management Science*, 68(3):1678–1695.

Kallus, N. and Mao, X. (2023). Stochastic optimization forests. *Management Science*, 69(4):1935–2545.

Lanckriet, G. R., Ghaoui, L. E., Bhattacharyya, C., and Jordan, M. I. (2002). A robust minimax approach to classification. *Journal of Machine Learning Research*, 3(Dec):555–582.

Landauer, T. K., Foltz, P. W., and Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes*, 25(2-3):259–284.

Lemon, A., So, A. M.-C., Ye, Y., et al. (2016). Low-rank semidefinite programming: Theory and applications. *Foundations and Trends® in Optimization*, 2(1-2):1–156.

Li, J. Y.-M. (2018). Closed-form solutions for worst-case law invariant risk measures with application to robust portfolio optimization. *Operations Research*, 66(6):1533–1541.

Li, K.-C. (1991). Sliced inverse regression for dimension reduction. *Journal of the American Statistical Association*, 86(414):316–327.

Li, K.-C. (1992). On principal hessian directions for data visualization and dimension reduction: Another application of stein's lemma. *Journal of the American Statistical Association*, 87(420):1025–1039.

Li, Y. and Zhu, L.-X. (2007). Asymptotics for sliced average variance estimation. *The Annals of Statistics*, 35(1):41–69.

Lin, F., Fang, X., and Gao, Z. (2022). Distributionally robust optimization: A review on theory and applications. *Numerical Algebra, Control & Optimization*, 12(1):159–212.

Liu, S., He, L., and Shen, Z.-J. M. (2021). On-time last-mile delivery: Order assignment with travel-time predictors. *Management Science*, 67(7):4095–4119.

Liu, Y., Meskarian, R., and Xu, H. (2017). Distributionally robust reward-risk ratio optimization with moment constraints. *SIAM Journal on Optimization*, 27(2):957–985.

Liyanage, L. H. and Shanthikumar, J. G. (2005). A practical inventory control policy using operational statistics. *Operations Research Letters*, 33(4):341–348.

Lotfi, S. and Zenios, S. A. (2018). Robust VaR and CVaR optimization under joint ambiguity in distributions, means, and covariances. *European Journal of Operational Research*, 269(2):556–576.

Mehrotra, S. and Papp, D. (2014). A cutting surface algorithm for semi-infinite convex programming with an application to moment robust optimization. *SIAM Journal on Optimization*, 24(4):1670–1697.

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

Natarajan, K., Sim, M., and Uichanco, J. (2018). Asymmetry and ambiguity in newsvendor models. *Management Science*, 64(7):3146–3167.

Popescu, I. (2007). Robust mean-covariance solutions for stochastic optimization. *Operations Research*, 55(1):98–112.

Qi, M., Shi, Y., Qi, Y., Ma, C., Yuan, R., Wu, D., and Shen, Z.-J. M. (2023). A practical end-to-end inventory management model with deep learning. *Management Science*, 69(2):759–773.

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

Rockafellar, R. T. and Uryasev, S. (2000). Optimization of conditional value-at-risk. *Journal of Risk*, 2:21–42.

Rujeerapaiboon, N., Kuhn, D., and Wiesemann, W. (2016). Robust growth-optimal portfolios. *Management Science*, 62(7):2090–2109.

Scarf, H. (1958). A min-max solution of an inventory problem. *Studies in the mathematical theory of inventory and production*.

Schölkopf, B., Smola, A., and Müller, K.-R. (1997). Kernel principal component analysis. In *International Conference on Artificial Neural Networks*, pages 583–588. Springer.

See, C.-T. and Sim, M. (2010). Robust approximation to multiperiod inventory management. *Operations Research*, 58(3):583–594.

Shehadeh, K. S., Cohn, A. E., and Jiang, R. (2020). A distributionally robust optimization approach for outpatient colonoscopy scheduling. *European Journal of Operational Research*, 283(2):549–561.

Sion, M. (1958). On general minimax theorems. *Pacific Journal of Mathematics*, 8(1):171–176.

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

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

Yin, X. and Li, B. (2011). Sufficient dimension reduction based on an ensemble of minimum average variance estimators. *The Annals of Statistics*, 39(6):3392–3416.

Yue, J., Chen, B., and Wang, M.-C. (2006). Expected value of distribution information for the newsvendor problem. *Operations Research*, 54(6):1128–1136.

Yurtsever, A., Tropp, J. A., Fercoq, O., Udell, M., and Cevher, V. (2021). Scalable semidefinite programming. *SIAM Journal on Mathematics of Data Science*, 3(1):171–200.

Zhang, L., Yang, J., and Gao, R. (2023). Optimal robust policy for feature-based newsvendor. *Management Science*, Forthcoming.

Zhang, Y., Song, S., Shen, Z.-J. M., and Wu, C. (2017). Robust shortest path problem with distributional uncertainty. *IEEE Transactions on Intelligent Transportation Systems*, 19(4):1080–1090.

Zhu, T., Xie, J., and Sim, M. (2022). Joint estimation and robustness optimization. *Management Science*, 68(3):1659–1677.

Zymler, S., Kuhn, D., and Rustem, B. (2013). Worst-case value at risk of nonlinear portfolios. *Management Science*, 59(1):172–188.

Appendix A: Table of Notations

Table A1 Summary of Notations

Notation	Description
Random Variables:	
$\boldsymbol{\xi}$	The random vector $\boldsymbol{\xi} \in \mathbb{R}^m$
$\boldsymbol{\xi}_{\mathrm{I}}$	The random vector $\boldsymbol{\xi}_{\mathrm{I}} \in \mathbb{R}^m$ obtained by the linearly transformation of $\boldsymbol{\xi}$
$\boldsymbol{\xi}_{\mathrm{r}}$	The random vector $\boldsymbol{\xi}_{\mathrm{r}} \in \mathbb{R}^{m_1}$ obtained by reducing the dimension of $\boldsymbol{\xi}_{\mathrm{I}}$
Distributions:	
$\mathbb{P}$	The probability distribution of the random vector $\boldsymbol{\xi}$
$\mathbb{P}_{\mathrm{I}}$	The probability distribution of the random vector $\boldsymbol{\xi}_{\mathrm{I}}$
$\mathbb{P}_{\mathrm{r}}$	The probability distribution of the random vector $\boldsymbol{\xi}_{\mathrm{r}}$
Decision Variables:	
$\mathbf{x}$	The decision variable $\mathbf{x} \in \mathbb{R}^n$
s, λ_k, $\mathbf{q}$, $\mathbf{Q}$	Decision variables in original SDP problem
$\hat{\boldsymbol{\lambda}}$	$\hat{\boldsymbol{\lambda}} := \{\lambda_1, \ldots, \lambda_K\}$
$\mathbf{q}_{\mathrm{r}}$, $\mathbf{Q}_{\mathrm{r}}$	Decision variables in PCA approximation
$\mathbf{B}$	The decision variable used in the optimized dimensionality reduction
$t_k, \mathbf{p}_k, \mathbf{P}_k, \mathbf{Z}$	Decision variables used in the lower bound
$\mathbf{Q}_{\mathrm{r}}', \hat{\mathbf{u}}', \hat{\mathbf{u}}'', \mathbf{B}_1, \mathbf{B}_2$	Decision variables used in the revisited lower bound
Parameters and Sets:	
$\mathcal{X}$	The feasible set of decision variable $\mathbf{x}$
$\mathcal{D}_{\mathrm{M0}}$	The distributional ambiguity set constructed by statistical information
$\mathcal{D}_{\mathrm{M}}$	The distributional ambiguity set corresponding to $\boldsymbol{\xi}_{\mathrm{I}}$
$\mathcal{S}$	The support of $\boldsymbol{\xi}$
γ_1	A scalar $\gamma_1 \geq 0$
γ_2	A scalar $\gamma_2 \geq 1$
$\boldsymbol{\mu}$	The estimated mean of $\boldsymbol{\xi}$
$\boldsymbol{\Sigma}$	The estimated covariance matrix of $\boldsymbol{\xi}$
$\mathbf{U}$, $\boldsymbol{\Lambda}$	Two matrices produced by the eigenvalue decomposition on the covariance matrix $\boldsymbol{\Sigma}$
$\mathbf{A}, \mathbf{b}$	$\mathcal{S} := \{\boldsymbol{\xi} \mid \mathbf{A}\boldsymbol{\xi} \leq \mathbf{b}\}$
$\mathcal{S}_{\mathrm{I}}$	The support of $\boldsymbol{\xi}_{\mathrm{I}}$
$\mathcal{S}_{\mathrm{r}}$	The support of $\boldsymbol{\xi}_{\mathrm{r}}$
$\mathcal{D}_{\mathrm{L}}$	The distributional ambiguity set corresponding to $\boldsymbol{\xi}_{\mathrm{r}}$
$\mathcal{B}_{m_1}$	The feasible set of decision variable $\mathbf{B} \in \mathbb{R}^{m \times m_1}$
$\mathcal{D}_{\mathrm{U}}$	The distributional ambiguity set by relaxing the second-moment constraint in $\mathcal{D}_{\mathrm{M}}$
Optimal Value Functions:	
$\Theta_{\mathrm{M}}(m)$	The optimal value of the original problem
$\Theta_{\mathrm{M}}(m_1)$	The optimal value of the PCA approximation
$\Theta_{\mathrm{L}}(m_1)$	The optimal value of the first outer approximation
$\underline{\Theta}(m_1, \mathbf{B})$	The optimal value of the subproblem of the first outer approximation
$\Theta_{\mathrm{U}}(m_1)$	The optimal value of the inner approximation
$\overline{\Theta}(m_1, \mathbf{B})$	The optimal value of the subproblem of the inner approximation
$\Theta_{\mathrm{L2}}(m_1)$	The optimal value of the second outer approximation

Appendix B: Technical Proofs

B.1. Proof of Lemma 1

First, we have

$$\begin{bmatrix} \mathbf{I}_m & \mathbf{B} \\ \mathbf{B}^\top & \mathbf{I}_{m_1} \end{bmatrix} \succeq 0 \iff \mathbf{I}_m - \mathbf{B}\mathbf{I}_{m_1}^{-1}\mathbf{B}^\top \succeq 0 \iff \mathbf{B}\mathbf{B}^\top \preceq \mathbf{I}_m,$$

where the first equivalence is by Schur complement and the second is because $\mathbf{I}_{m_1}^{-1} = \mathbf{I}_{m_1}$.

Second, we have

$$\begin{bmatrix} \mathbf{I}_m & \mathbf{B} \\ \mathbf{B}^\top & \mathbf{I}_{m_1} \end{bmatrix} \succeq 0 \iff \mathbf{I}_{m_1} - \mathbf{B}^\top\mathbf{I}_{m}^{-1}\mathbf{B} \succeq 0 \iff \mathbf{B}^\top\mathbf{B} \preceq \mathbf{I}_{m_1},$$

where the first equivalence is by Schur complement and the second is because $\mathbf{I}_m^{-1} = \mathbf{I}_m$. Thus, the lemma is proved. □

B.2. Proof of Lemma 2

(i) Suppose $\mathbf{X} \succeq \mathbf{Y}$. For any $\mathbf{a} \in \mathbb{R}^n$, we have $\mathbf{V}\mathbf{a} \in \mathbb{R}^m$. It follows that

$$\begin{aligned} \mathbf{X} \succeq \mathbf{Y} &\Longrightarrow (\mathbf{V}\mathbf{a})^\top(\mathbf{X}-\mathbf{Y})(\mathbf{V}\mathbf{a}) \geq 0,\ \forall \mathbf{a} \in \mathbb{R}^n \\ &\iff \mathbf{a}^\top \left(\mathbf{V}^\top(\mathbf{X}-\mathbf{Y})\mathbf{V}\right)\mathbf{a} \geq 0,\ \forall \mathbf{a} \in \mathbb{R}^n \\ &\iff \mathbf{V}^\top(\mathbf{X}-\mathbf{Y})\mathbf{V} \succeq 0 \iff \mathbf{V}^\top\mathbf{X}\mathbf{V} \succeq \mathbf{V}^\top\mathbf{Y}\mathbf{V}. \end{aligned}$$

(ii) First, for any $\mathbf{V} \in \mathbb{R}^{m\times m}$, we have

$$\mathbf{X} \succeq \mathbf{Y} \Longrightarrow \mathbf{V}^\top\mathbf{X}\mathbf{V} \succeq \mathbf{V}^\top\mathbf{Y}\mathbf{V}$$

by (i). Second, suppose $\mathbf{V}^\top\mathbf{X}\mathbf{V} \succeq \mathbf{V}^\top\mathbf{Y}\mathbf{V}$. Note that $\mathbf{V}^{-1} \in \mathbb{R}^{m\times m}$. According to (i), the matrix $\mathbf{V}^\top\mathbf{X}\mathbf{V} - \mathbf{V}^\top\mathbf{Y}\mathbf{V}$ remains as PSD if it multiplies $(\mathbf{V}^{-1})^\top$ before it and $\mathbf{V}^{-1}$ after it, i.e.,

$$\left(\mathbf{V}^{-1}\right)^\top \mathbf{V}^\top\mathbf{X}\mathbf{V}\mathbf{V}^{-1} \succeq \left(\mathbf{V}^{-1}\right)^\top \mathbf{V}^\top\mathbf{Y}\mathbf{V}\mathbf{V}^{-1}.$$

It follows that $\mathbf{X} \succeq \mathbf{Y}$ because $(\mathbf{V}^{-1})^\top\mathbf{V}^\top = \mathbf{I}_m$ and $\mathbf{V}\mathbf{V}^{-1} = \mathbf{I}_m$. Thus, $\mathbf{X} \succeq \mathbf{Y}$ is equivalent to $\mathbf{V}^\top\mathbf{X}\mathbf{V} \succeq \mathbf{V}^\top\mathbf{Y}\mathbf{V}$ if $\mathbf{V} \in \mathbb{R}^{m\times m}$ is invertible. □

B.3. Proof of Theorem 1

(i) Given any $\mathbf{x} \in \mathcal{X}$ and $\mathbf{B} \in \mathcal{B}_{m_1}$, i.e., $\mathbf{B}^\top\mathbf{B} = \mathbf{I}_{m_1}$, we define $\boldsymbol{\zeta} = \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\boldsymbol{\xi}_{\mathrm{r}} + \boldsymbol{\mu}$ and use $\mathcal{S}_\zeta$ and $\mathcal{D}_\zeta$ to denote its support and ambiguity set, respectively. As $\mathcal{S}_{\mathrm{r}} = \{\boldsymbol{\xi}_{\mathrm{r}} \in \mathbb{R}^{m_1} \mid \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\boldsymbol{\xi}_{\mathrm{r}} + \boldsymbol{\mu} \in \mathcal{S}\}$ and $\mathcal{S}_\zeta = \{\boldsymbol{\zeta} \in \mathbb{R}^m \mid \boldsymbol{\zeta} = \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\boldsymbol{\xi}_{\mathrm{r}} + \boldsymbol{\mu},\ \boldsymbol{\xi}_{\mathrm{r}} \in \mathcal{S}_{\mathrm{r}}\}$, we can deduce $\mathcal{S}_\zeta \subseteq \mathcal{S}$. We also have

$$\begin{aligned} &\left(\mathbb{E}_{\mathbb{P}_\zeta}[\boldsymbol{\zeta}] - \boldsymbol{\mu}\right)^\top \boldsymbol{\Sigma}^{-1}\left(\mathbb{E}_{\mathbb{P}_\zeta}[\boldsymbol{\zeta}] - \boldsymbol{\mu}\right) \qquad (42) \\ =& \left(\mathbb{E}_{\mathbb{P}_{\mathrm{r}}}\left[\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\boldsymbol{\xi}_{\mathrm{r}}\right]\right)^\top \boldsymbol{\Sigma}^{-1}\mathbb{E}_{\mathbb{P}_{\mathrm{r}}}\left[\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\boldsymbol{\xi}_{\mathrm{r}}\right] \end{aligned}$$

$$
\begin{aligned}
=&\mathbb{E}_{\mathbb{P}_{\mathrm{r}}}\left[\boldsymbol{\xi}_{\mathrm{r}}^{\top}\right]\mathbf{B}^{\top}\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^{\top}\boldsymbol{\Sigma}^{-1}\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)\mathbf{B}\mathbb{E}_{\mathbb{P}_{\mathrm{r}}}\left[\boldsymbol{\xi}_{\mathrm{r}}\right]\\
=&\mathbb{E}_{\mathbb{P}_{\mathrm{r}}}\left[\boldsymbol{\xi}_{\mathrm{r}}^{\top}\right]\mathbf{B}^{\top}\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^{\top}\left(\mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^{\top}\right)^{-1}\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)\mathbf{B}\mathbb{E}_{\mathbb{P}_{\mathrm{r}}}\left[\boldsymbol{\xi}_{\mathrm{r}}\right]\\
=&\mathbb{E}_{\mathbb{P}_{\mathrm{r}}}\left[\boldsymbol{\xi}_{\mathrm{r}}^{\top}\right]\mathbf{B}^{\top}\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^{\top}\left(\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^{\top}\right)^{-1}\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)\mathbf{B}\mathbb{E}_{\mathbb{P}_{\mathrm{r}}}\left[\boldsymbol{\xi}_{\mathrm{r}}\right]\\
=&\mathbb{E}_{\mathbb{P}_{\mathrm{r}}}\left[\boldsymbol{\xi}_{\mathrm{r}}^{\top}\right]\mathbf{B}^{\top}\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^{\top}\left(\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^{\top}\right)^{-1}\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^{-1}\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)\mathbf{B}\mathbb{E}_{\mathbb{P}_{\mathrm{r}}}\left[\boldsymbol{\xi}_{\mathrm{r}}\right]\\
=&\mathbb{E}_{\mathbb{P}_{\mathrm{r}}}\left[\boldsymbol{\xi}_{\mathrm{r}}^{\top}\right]\mathbf{B}^{\top}\mathbf{B}\mathbb{E}_{\mathbb{P}_{\mathrm{r}}}\left[\boldsymbol{\xi}_{\mathrm{r}}\right]=\mathbb{E}_{\mathbb{P}_{\mathrm{r}}}\left[\boldsymbol{\xi}_{\mathrm{r}}^{\top}\right]\mathbb{E}_{\mathbb{P}_{\mathrm{r}}}\left[\boldsymbol{\xi}_{\mathrm{r}}\right]\leq\gamma_1,
\end{aligned}
\tag{43}
$$

where the inequality holds because of (6b). Meanwhile, we have

$$
\begin{aligned}
&\mathbb{E}_{\mathbb{P}_{\zeta}}\left[(\boldsymbol{\zeta}-\boldsymbol{\mu})(\boldsymbol{\zeta}-\boldsymbol{\mu})^{\top}\right]=\mathbb{E}_{\mathbb{P}_{\mathrm{r}}}\left[\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\boldsymbol{\xi}_{\mathrm{r}}\boldsymbol{\xi}_{\mathrm{r}}^{\top}\mathbf{B}^{\top}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{U}^{\top}\right]\\
\preceq&\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\gamma_2\mathbf{I}_{m_1}\mathbf{B}^{\top}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{U}^{\top}=\gamma_2\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\mathbf{B}^{\top}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{U}^{\top}\preceq\gamma_2\mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^{\top}=\gamma_2\boldsymbol{\Sigma},
\end{aligned}
\tag{44}
$$

where the first inequality holds because of (6b) and the second inequality holds because $\mathbf{B}^{\top}\mathbf{B}=\mathbf{I}_{m_1}$, leading to $\mathbf{B}^{\top}\mathbf{B}\preceq\mathbf{I}_{m_1}$, which is further equivalent to $\mathbf{B}\mathbf{B}^{\top}\preceq\mathbf{I}_m$ by Lemma 1. By $\mathcal{S}_{\zeta}\subseteq\mathcal{S}$, (43), and (44), it follows that $\mathcal{D}_{\zeta}$ lies in $\mathcal{D}_{\mathrm{M0}}$, i.e., $\mathcal{D}_{\zeta}\subseteq\mathcal{D}_{\mathrm{M0}}$.

Therefore, given any $\mathbf{x}\in\mathcal{X}$ and $\mathbf{B}\in\mathcal{B}_{m_1}$, we have

$$
\max_{\mathbb{P}_{\mathrm{r}}\in\mathcal{D}_{\mathrm{L}}}\ \mathbb{E}_{\mathbb{P}_{\mathrm{r}}}\left[f\left(\mathbf{x},\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\boldsymbol{\xi}_{\mathrm{r}}+\boldsymbol{\mu}\right)\right]=\max_{\mathbb{P}_{\zeta}\in\mathcal{D}_{\zeta}}\ \mathbb{E}_{\mathbb{P}_{\zeta}}\left[f\left(\mathbf{x},\boldsymbol{\zeta}\right)\right]\leq\max_{\mathbb{P}\in\mathcal{D}_{\mathrm{M0}}}\ \mathbb{E}_{\mathbb{P}}\left[f\left(\mathbf{x},\boldsymbol{\xi}\right)\right],
$$

where the equality holds by change of variables and the inequality holds because $\mathcal{D}_{\zeta}\subseteq\mathcal{D}_{\mathrm{M0}}$. It follows that

$$
\max_{\mathbf{B}\in\mathcal{B}_{m_1}}\ \min_{\mathbf{x}\in\mathcal{X}}\ \max_{\mathbb{P}_{\mathrm{r}}\in\mathcal{D}_{\mathrm{L}}}\ \mathbb{E}_{\mathbb{P}_{\mathrm{r}}}\left[f\left(\mathbf{x},\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\boldsymbol{\xi}_{\mathrm{r}}+\boldsymbol{\mu}\right)\right]\leq\min_{\mathbf{x}\in\mathcal{X}}\ \max_{\mathbb{P}\in\mathcal{D}_{\mathrm{M0}}}\ \mathbb{E}_{\mathbb{P}}\left[f\left(\mathbf{x},\boldsymbol{\xi}\right)\right],
$$

which demonstrates that the optimal value of Problem (11) is a lower bound for that of Problem (3) (i.e., Problem (2)).

(ii) For any $m_1<m_2\leq m$, $\mathbf{B}_1\in\mathbb{R}^{m\times m_1}$, and $\mathbf{C}\in\mathbb{R}^{m\times(m_2-m_1)}$ such that $\mathbf{B}_1^{\top}\mathbf{B}_1=\mathbf{I}_{m_1}$ and $[\mathbf{B}_1,\mathbf{C}]^{\top}[\mathbf{B}_1,\mathbf{C}]=\mathbf{I}_{m_2}$, we have $\mathbf{B}_2=[\mathbf{B}_1,\mathbf{C}]\in\mathbb{R}^{m\times m_2}$. Meanwhile, we have $\mathcal{B}_{m_2}=\{\mathbf{B}\in\mathbb{R}^{m\times m_2}\mid\mathbf{B}^{\top}\mathbf{B}=\mathbf{I}_{m_2}\}$ and define $\boldsymbol{\zeta}_i=\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}_i\boldsymbol{\xi}_{\mathrm{r}_i}+\boldsymbol{\mu}\in\mathbb{R}^m$ for any $i\in[2]$, where $\boldsymbol{\xi}_{\mathrm{r}_i}\in\mathbb{R}^{m_i}$. Clearly, $\mathbf{B}_2\in\mathcal{B}_{m_2}$ because $\mathbf{B}_2^{\top}\mathbf{B}_2=\mathbf{I}_{m_2}$. We further define the ambiguity set of $\boldsymbol{\zeta}_i$ as

$$
\mathcal{D}_{\zeta_i}=\left\{\mathbb{P}_{\zeta_i}\ \middle|\ \boldsymbol{\zeta}_i\sim\mathbb{P}_{\zeta_i},\ \boldsymbol{\zeta}_i=\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}_i\boldsymbol{\xi}_{\mathrm{r}_i}+\boldsymbol{\mu},\ \boldsymbol{\xi}_{\mathrm{r}_i}\sim\mathbb{P}_{\mathrm{r}_i}\in\mathcal{D}_{\mathrm{r}_i}\right\},\ \forall i\in[2],
\tag{45}
$$

where $\mathcal{D}_{\mathrm{r}_i}$ represents the ambiguity set of $\boldsymbol{\xi}_{\mathrm{r}_i}$ for any $i\in[2]$. Given $\boldsymbol{\zeta}_1\sim\mathbb{P}_{\zeta_1}\in\mathcal{D}_{\zeta_1}$, there exists a $\boldsymbol{\xi}_{\mathrm{r}_1}\sim\mathbb{P}_{\mathrm{r}_1}\in\mathcal{D}_{\mathrm{r}_1}$ such that $\boldsymbol{\zeta}_1=\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}_1\boldsymbol{\xi}_{\mathrm{r}_1}+\boldsymbol{\mu}=\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}_2\bar{\boldsymbol{\xi}}_{\mathrm{r}_2}+\boldsymbol{\mu}$, where $\bar{\boldsymbol{\xi}}_{\mathrm{r}_2}=(\boldsymbol{\xi}_{\mathrm{r}_1}^{\top},\mathbf{0}_{m_2-m_1}^{\top})^{\top}\in\mathbb{R}^{m_2}$.

By using $\mathcal{S}_{\mathrm{r}_i}$ (see definition in (12)) to denote the support of $\boldsymbol{\xi}_{\mathrm{r}_i}$ for any $i\in[2]$, we have

$$
\mathbb{P}\left\{\boldsymbol{\xi}_{\mathrm{r}_1}\in\mathcal{S}_{\mathrm{r}_1}\right\}=\mathbb{P}\left\{\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}_1\boldsymbol{\xi}_{\mathrm{r}_1}+\boldsymbol{\mu}\in\mathcal{S}\right\}=\mathbb{P}\left\{\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}_2\bar{\boldsymbol{\xi}}_{\mathrm{r}_2}+\boldsymbol{\mu}\in\mathcal{S}\right\}=1,
$$

where the second equality holds because $\mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{B}_1\boldsymbol{\zeta}_{\mathrm{r}_1} = \mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{B}_2\bar{\boldsymbol{\xi}}_{\mathrm{r}_2}$. It follows that $\mathbb{P}\{\bar{\boldsymbol{\xi}}_{\mathrm{r}_2} \in \mathcal{S}_{\mathrm{r}_2}\} = 1$ by the definition of $\mathcal{S}_{\mathrm{r}_2}$. In addition, we have $\mathbb{E}[\bar{\boldsymbol{\xi}}_{\mathrm{r}_2}^\top]\mathbb{E}[\bar{\boldsymbol{\xi}}_{\mathrm{r}_2}] = \mathbb{E}[\boldsymbol{\xi}_{\mathrm{r}_1}^\top]\mathbb{E}[\boldsymbol{\xi}_{\mathrm{r}_1}] \leq \gamma_1$ and

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

Thus, the probability distribution of $\bar{\boldsymbol{\xi}}_{\mathrm{r}_2}$ belongs to $\mathcal{D}_{\mathrm{r}_2}$. Meanwhile, by the definition of $\mathcal{D}_{\zeta_i}$ for any $i \in [2]$ in (45) and $\boldsymbol{\zeta}_1 = \mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{B}_2\bar{\boldsymbol{\xi}}_{\mathrm{r}_2} + \boldsymbol{\mu}$, we have $\mathbb{P}_{\zeta_1} \in \mathcal{D}_{\zeta_2}$ and further $\mathcal{D}_{\zeta_1} \subseteq \mathcal{D}_{\zeta_2}$. Therefore, for any $\mathbf{x} \in \mathcal{X}$, $\mathbf{B}_1 \in \mathbb{R}^{m\times m_1}$, and $\mathbf{C} \in \mathbb{R}^{m\times(m_2-m_1)}$ such that $\mathbf{B}_1^\top\mathbf{B}_1 = \mathbf{I}_{m_1}$ and $[\mathbf{B}_1, \mathbf{C}]^\top[\mathbf{B}_1, \mathbf{C}] = \mathbf{I}_{m_2}$, we have

$$\max_{\mathbb{P}_{\zeta_1}\in\mathcal{D}_{\zeta_1}} \mathbb{E}_{\mathbb{P}_{\zeta_1}}\left[f\left(\mathbf{x}, \boldsymbol{\zeta}_1\right)\right] \leq \max_{\mathbb{P}_{\zeta_2}\in\mathcal{D}_{\zeta_2}} \mathbb{E}_{\mathbb{P}_{\zeta_2}}\left[f\left(\mathbf{x}, \boldsymbol{\zeta}_2\right)\right]. \tag{46}$$

Together with the definitions of $\boldsymbol{\zeta}_i$ ($\forall i \in [2]$) and $\mathbf{B}_2$, inequality (46) leads to

$$\max_{\mathbb{P}_{\mathrm{r}_1}\in\mathcal{D}_{\mathrm{r}_1}} \mathbb{E}_{\mathbb{P}_{\zeta_1}}\left[f\left(\mathbf{x}, \mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{B}_1\boldsymbol{\xi}_{\mathrm{r}_1} + \boldsymbol{\mu}\right)\right] \leq \max_{\mathbb{P}_{\mathrm{r}_2}\in\mathcal{D}_{\mathrm{r}_2}} \mathbb{E}_{\mathbb{P}_{\zeta_2}}\left[f\left(\mathbf{x}, \mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}[\mathbf{B}_1, \mathbf{C}]\boldsymbol{\xi}_{\mathrm{r}_2} + \boldsymbol{\mu}\right)\right].$$

Considering an optimal solution $\mathbf{B}_1^* \in \mathbb{R}^{m\times m_1}$ of Problem (11), for any $\mathbf{x} \in \mathcal{X}$ and $\mathbf{C} \in \mathbb{R}^{m\times(m_2-m_1)}$ such that $[\mathbf{B}_1^*, \mathbf{C}]^\top[\mathbf{B}_1^*, \mathbf{C}] = \mathbf{I}_{m_2}$, we have

$$\max_{\mathbb{P}_{\mathrm{r}_1}\in\mathcal{D}_{\mathrm{r}_1}} \mathbb{E}_{\mathbb{P}_{\zeta_1}}\left[f\left(\mathbf{x}, \mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{B}_1^*\boldsymbol{\xi}_{\mathrm{r}_1} + \boldsymbol{\mu}\right)\right] \leq \max_{\mathbb{P}_{\mathrm{r}_2}\in\mathcal{D}_{\mathrm{r}_2}} \mathbb{E}_{\mathbb{P}_{\zeta_2}}\left[f\left(\mathbf{x}, \mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}[\mathbf{B}_1^*, \mathbf{C}]\boldsymbol{\xi}_{\mathrm{r}_2} + \boldsymbol{\mu}\right)\right].$$

For any $\mathbf{C} \in \mathbb{R}^{m\times(m_2-m_1)}$ such that $[\mathbf{B}_1^*, \mathbf{C}]^\top[\mathbf{B}_1^*, \mathbf{C}] = \mathbf{I}_{m_2}$, we have

$$\min_{\mathbf{x}\in\mathcal{X}} \max_{\mathbb{P}_{\mathrm{r}_1}\in\mathcal{D}_{\mathrm{r}_1}} \mathbb{E}_{\mathbb{P}_{\zeta_1}}\left[f\left(\mathbf{x}, \mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{B}_1^*\boldsymbol{\xi}_{\mathrm{r}_1} + \boldsymbol{\mu}\right)\right] \leq \min_{\mathbf{x}\in\mathcal{X}} \max_{\mathbb{P}_{\mathrm{r}_2}\in\mathcal{D}_{\mathrm{r}_2}} \mathbb{E}_{\mathbb{P}_{\zeta_2}}\left[f\left(\mathbf{x}, \mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}[\mathbf{B}_1^*, \mathbf{C}]\boldsymbol{\xi}_{\mathrm{r}_2} + \boldsymbol{\mu}\right)\right]. \tag{47}$$

It follows that

$$\begin{aligned}
&\max_{\mathbf{B}_1^\top\mathbf{B}_1=\mathbf{I}_{m_1}} \min_{\mathbf{x}\in\mathcal{X}} \max_{\mathbb{P}_{\mathrm{r}_1}\in\mathcal{D}_{\mathrm{r}_1}} \mathbb{E}_{\mathbb{P}_{\zeta_1}}\left[f\left(\mathbf{x}, \mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{B}_1\boldsymbol{\xi}_{\mathrm{r}_1} + \boldsymbol{\mu}\right)\right] \\
&\quad = \min_{\mathbf{x}\in\mathcal{X}} \max_{\mathbb{P}_{\mathrm{r}_1}\in\mathcal{D}_{\mathrm{r}_1}} \mathbb{E}_{\mathbb{P}_{\zeta_1}}\left[f\left(\mathbf{x}, \mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{B}_1^*\boldsymbol{\xi}_{\mathrm{r}_1} + \boldsymbol{\mu}\right)\right] \\
&\quad \leq \min_{\mathbf{x}\in\mathcal{X}} \max_{\mathbb{P}_{\mathrm{r}_2}\in\mathcal{D}_{\mathrm{r}_2}} \mathbb{E}_{\mathbb{P}_{\zeta_2}}\left[f\left(\mathbf{x}, \mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}[\mathbf{B}_1^*, \mathbf{C}]\boldsymbol{\xi}_{\mathrm{r}_2} + \boldsymbol{\mu}\right)\right] \\
&\quad \leq \max_{\mathbf{B}_2\in\mathcal{B}_{m_2}} \min_{\mathbf{x}\in\mathcal{X}} \max_{\mathbb{P}_{\mathrm{r}_2}\in\mathcal{D}_{\mathrm{r}_2}} \mathbb{E}_{\mathbb{P}_{\zeta_2}}\left[f\left(\mathbf{x}, \mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{B}_2\boldsymbol{\xi}_{\mathrm{r}_2} + \boldsymbol{\mu}\right)\right],
\end{aligned}$$

where the first inequality holds by (47) and the second inequality holds because $[\mathbf{B}_1^*, \mathbf{C}] \in \mathcal{B}_{m_2}$. That is, the optimal value of Problem (11) is nondecreasing in m_1.

(iii) When $m_1 = m$, we have $\mathbf{B} \in \mathcal{B}_m \subseteq \mathbb{R}^{m\times m}$, i.e., $\mathbf{B}^\top\mathbf{B} = \mathbf{I}_m$. First, we have $\Theta_{\mathrm{L}}(m) \leq \Theta_{\mathrm{M}}(m)$ by the conclusion (i). Second, when $\mathbf{B} = \mathbf{I}_m$, Problem (11) becomes Problem (3). Because $\mathbf{B} = \mathbf{I}_m$ is a feasible solution of Problem (11), it follows that $\Theta_{\mathrm{L}}(m) \geq \Theta_{\mathrm{M}}(m)$. Therefore, we have $\Theta_{\mathrm{L}}(m) = \Theta_{\mathrm{M}}(m)$. □

B.4. Proof of Theorem 2

Note that the optimal solution $(\mathbf{x}^*, s^*, \hat{\boldsymbol{\lambda}}^*, \mathbf{q}^*, \mathbf{Q}^*)$ of Problem (4) leads to the optimal value $s^* + \gamma_2 \mathbf{I}_m \bullet \mathbf{Q}^* + \sqrt{\gamma_1}\, \|\mathbf{q}^*\|_2$. Based on this optimal solution, we construct a feasible solution of Problem (4), denoted by $(\mathbf{x}', s', \hat{\boldsymbol{\lambda}}', \mathbf{q}', \mathbf{Q}')$ such that $\mathbf{x}' = \mathbf{x}^*$, $s' = s^*$, and $\hat{\boldsymbol{\lambda}}' = \hat{\boldsymbol{\lambda}}^*$.

Now we construct the values of $\mathbf{q}'$ and $\mathbf{Q}'$. By constraints (4b), we have

$$\begin{bmatrix} S_k & \frac{1}{2}\left(\mathbf{q}^* + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top \left(\mathbf{A}^\top \boldsymbol{\lambda}_k^* - y_k(\mathbf{x}^*)\right)\right)^\top \\ \frac{1}{2}\left(\mathbf{q}^* + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top \left(\mathbf{A}^\top \boldsymbol{\lambda}_k^* - y_k(\mathbf{x}^*)\right)\right) & \mathbf{Q}^* \end{bmatrix} \succeq 0, \ \forall k \in [K]. \tag{48}$$

By Schur complement, we can equivalently rewrite (48) as

$$4S_k\mathbf{Q}^* \succeq \left(\mathbf{q}^* + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top \left(\mathbf{A}^\top \boldsymbol{\lambda}_k^* - y_k(\mathbf{x}^*)\right)\right)\left(\mathbf{q}^* + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top \left(\mathbf{A}^\top \boldsymbol{\lambda}_k^* - y_k(\mathbf{x}^*)\right)\right)^\top, \forall k \in [K]. \tag{49}$$

Note that $K < m$. Thus, through the Gram–Schmidt process, we can always construct K orthonormal vectors $\mathbf{v}_k \in \mathbb{R}^m$, $\forall k \in [K]$, and K real vectors $\mathbf{a}_k \in \mathbb{R}^K$, $\forall k \in [K]$, such that

$$\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top \left(\mathbf{A}^\top \boldsymbol{\lambda}_k^* - y_k(\mathbf{x}^*)\right) = \mathbf{V}\mathbf{a}_k, \ \forall k \in [K], \tag{50}$$

$\mathbf{V} = [\mathbf{v}_k, \ \forall k \in [K]] \in \mathbb{R}^{m \times K}$. We further extend $\mathbf{V}$ to $[\mathbf{V}, \bar{\mathbf{V}}] \in \mathbb{R}^{m \times m}$ with $\bar{\mathbf{V}} \in \mathbb{R}^{m \times (m-K)}$ such that all the column vectors of $[\mathbf{V}, \bar{\mathbf{V}}]$ can span the space of $\mathbb{R}^m$. As $\mathbf{q}^* \in \mathbb{R}^m$, we can find $\mathbf{a}_0 \in \mathbf{R}^K$ and $\bar{\mathbf{a}}_0 \in \mathbf{R}^{m-K}$ such that

$$\mathbf{q}^* = \mathbf{V}\mathbf{a}_0 + \bar{\mathbf{V}}\bar{\mathbf{a}}_0. \tag{51}$$

As $\mathbf{Q}^* \in \mathbb{R}^{m \times m}$, we can then decompose $\mathbf{Q}^*$ as

$$\begin{aligned} \mathbf{Q}^* &= [\mathbf{V} \ \bar{\mathbf{V}}] \begin{bmatrix} \mathbf{Y}_{11} & \mathbf{Y}_{12} \\ \mathbf{Y}_{21} & \mathbf{Y}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{V}^\top \\ \bar{\mathbf{V}}^\top \end{bmatrix} \\ &= \mathbf{V}\mathbf{Y}_{11}\mathbf{V}^\top + \bar{\mathbf{V}}\mathbf{Y}_{21}\mathbf{V}^\top + \mathbf{V}\mathbf{Y}_{12}\bar{\mathbf{V}}^\top + \bar{\mathbf{V}}\mathbf{Y}_{22}\bar{\mathbf{V}}^\top, \end{aligned} \tag{52}$$

where $\mathbf{Y}_{11} \in \mathbb{R}^{K \times K}$, $\mathbf{Y}_{12} \in \mathbb{R}^{K \times (m-K)}$, $\mathbf{Y}_{21} \in \mathbb{R}^{(m-K) \times K}$, and $\mathbf{Y}_{22} \in \mathbb{R}^{(m-K) \times (m-K)}$. As $\mathbf{Q}^* \succeq 0$, we have $\begin{bmatrix} \mathbf{Y}_{11} & \mathbf{Y}_{12} \\ \mathbf{Y}_{21} & \mathbf{Y}_{22} \end{bmatrix} = [\mathbf{v} \ \bar{\mathbf{v}}]^{-1} \mathbf{Q}^* \begin{bmatrix} \mathbf{V}^\top \\ \bar{\mathbf{V}}^\top \end{bmatrix}^{-1} \succeq 0$ by Lemma 2. By (49), (50), and (51), we have

$$4S_k\mathbf{Q}^* \succeq \left(\mathbf{V}\mathbf{a}_0 + \bar{\mathbf{V}}\bar{\mathbf{a}}_0 + \mathbf{V}\mathbf{a}_k\right)\left(\mathbf{V}\mathbf{a}_0 + \bar{\mathbf{V}}\bar{\mathbf{a}}_0 + \mathbf{V}\mathbf{a}_k\right)^\top, \ \forall k \in [K]. \tag{53}$$

By (52) and (53), we have

$$4S_k\left(\mathbf{V}\mathbf{Y}_{11}\mathbf{V}^\top + \bar{\mathbf{V}}\mathbf{Y}_{21}\mathbf{V}^\top + \mathbf{V}\mathbf{Y}_{12}\bar{\mathbf{V}}^\top + \bar{\mathbf{V}}\mathbf{Y}_{22}\bar{\mathbf{V}}^\top\right) \succeq \left(\mathbf{V}\mathbf{a}_0 + \bar{\mathbf{V}}\bar{\mathbf{a}}_0 + \mathbf{V}\mathbf{a}_k\right)\left(\mathbf{V}\mathbf{a}_0 + \bar{\mathbf{V}}\bar{\mathbf{a}}_0 + \mathbf{V}\mathbf{a}_k\right)^\top, \forall k \in [K].$$

By Lemma 2, we further have

$$4S_k\mathbf{V}^\top(\mathbf{V}\mathbf{Y}_{11}\mathbf{V}^\top + \bar{\mathbf{V}}\mathbf{Y}_{21}\mathbf{V}^\top + \mathbf{V}\mathbf{Y}_{12}\bar{\mathbf{V}}^\top + \bar{\mathbf{V}}\mathbf{Y}_{22}\bar{\mathbf{V}}^\top)\mathbf{V}$$

$$\succeq \mathbf{V}^\top(\mathbf{V}\mathbf{a}_0 + \bar{\mathbf{V}}\bar{\mathbf{a}}_0 + \mathbf{V}\mathbf{a}_k)(\mathbf{V}\mathbf{a}_0 + \bar{\mathbf{V}}\bar{\mathbf{a}}_0 + \mathbf{V}\mathbf{a}_k)^\top\mathbf{V},\ \forall k \in [K]. \tag{54}$$

Because $\mathbf{V}^\top\bar{\mathbf{V}} = \mathbf{0}$, $\bar{\mathbf{V}}^\top\mathbf{V} = \mathbf{0}$, and $\mathbf{V}^\top\mathbf{V} = \mathbf{I}_K$, constraints (54) become

$$4S_k\mathbf{Y}_{11} \succeq (\mathbf{a}_0 + \mathbf{a}_k)(\mathbf{a}_0 + \mathbf{a}_k)^\top,\ \forall k \in [K]. \tag{55}$$

Now we let $\mathbf{q}' = \mathbf{V}\mathbf{a}_0$ and $\mathbf{Q}' = \mathbf{V}\mathbf{Y}_{11}\mathbf{V}^\top$. By (55) and Lemma 2, we have

$$\begin{aligned}4S_k\mathbf{Q}' =& 4S_k\mathbf{V}\mathbf{Y}_{11}\mathbf{V}^\top \succeq (\mathbf{V}\mathbf{a}_0 + \mathbf{V}\mathbf{a}_k)(\mathbf{V}\mathbf{a}_0 + \mathbf{V}\mathbf{a}_k)^\top \\ =& \left(\mathbf{q}' + \left(\mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}\right)^\top \left(\mathbf{A}^\top\boldsymbol{\lambda}_k^* - y_k(\mathbf{x}^*)\right)\right)\left(\mathbf{q}' + \left(\mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}\right)^\top \left(\mathbf{A}^\top\boldsymbol{\lambda}_k^* - y_k(\mathbf{x}^*)\right)\right)^\top,\ \forall k \in [K].\end{aligned} \tag{56}$$

Comparing (4b) and (56), we have $(\mathbf{x}', s', \hat{\boldsymbol{\lambda}}', \mathbf{q}', \mathbf{Q}')$ is a feasible solution of Problem (4) and the corresponding objective value is

$$s' + \gamma_2\mathbf{I}_m \bullet \mathbf{Q}' + \sqrt{\gamma_1}\,\|\mathbf{q}'\|_2 \geq s^* + \gamma_2\mathbf{I}_m \bullet \mathbf{Q}^* + \sqrt{\gamma_1}\,\|\mathbf{q}^*\|_2, \tag{57}$$

where the inequality holds because $(\mathbf{x}', s', \hat{\boldsymbol{\lambda}}', \mathbf{q}', \mathbf{Q}')$ is a feasible solution of Problem (4) and Problem (4) is a minimization problem. Note that

$$\begin{aligned}\mathbf{I}_m \bullet \mathbf{Q}^* &= tr(\mathbf{Q}^*) = tr\left([\mathbf{V}\ \bar{\mathbf{V}}]\begin{bmatrix}\mathbf{Y}_{11} & \mathbf{Y}_{12}\\ \mathbf{Y}_{21} & \mathbf{Y}_{22}\end{bmatrix}\begin{bmatrix}\mathbf{V}^\top\\ \bar{\mathbf{V}}^\top\end{bmatrix}\right) = tr\left(\begin{bmatrix}\mathbf{Y}_{11} & \mathbf{Y}_{12}\\ \mathbf{Y}_{21} & \mathbf{Y}_{22}\end{bmatrix}\begin{bmatrix}\mathbf{V}^\top\\ \bar{\mathbf{V}}^\top\end{bmatrix}[\mathbf{V}\ \bar{\mathbf{V}}]\right) \\ &= tr\left(\begin{bmatrix}\mathbf{Y}_{11} & \mathbf{Y}_{12}\\ \mathbf{Y}_{21} & \mathbf{Y}_{22}\end{bmatrix}\right) = \mathbf{I}_K \bullet \mathbf{Y}_{11} + \mathbf{I}_{m-K} \bullet \mathbf{Y}_{22} \\ &\geq \mathbf{I}_K \bullet \mathbf{Y}_{11} = tr(\mathbf{Y}_{11}) = tr(\mathbf{Y}_{11}\mathbf{V}^\top\mathbf{V}) = tr(\mathbf{V}\mathbf{Y}_{11}\mathbf{V}^\top) = tr(\mathbf{Q}') \\ &= \mathbf{I}_m \bullet \mathbf{Q}',\end{aligned}$$

where the first equality holds by the definition of a matrix's trace, the second equality holds by (52), the third equality holds by the cyclic property of a matrix's trace, the fourth equality holds because $\begin{bmatrix}\mathbf{V}^\top\\ \bar{\mathbf{V}}^\top\end{bmatrix}[\mathbf{V}\ \bar{\mathbf{V}}] = \mathbf{I}_m$, and the first inequality holds because $\begin{bmatrix}\mathbf{Y}_{11} & \mathbf{Y}_{12}\\ \mathbf{Y}_{21} & \mathbf{Y}_{22}\end{bmatrix} \succeq 0$ and accordingly $\mathbf{I}_{m-K} \bullet \mathbf{Y}_{22} \geq 0$. Meanwhile,

$$\begin{aligned}\|\mathbf{q}^*\|_2^2 = (\mathbf{q}^*)^\top\mathbf{q}^* = (\mathbf{V}\mathbf{a}_0 + \bar{\mathbf{V}}\bar{\mathbf{a}}_0)^\top(\mathbf{V}\mathbf{a}_0 + \bar{\mathbf{V}}\bar{\mathbf{a}}_0) = \left(\mathbf{a}_0^\top\mathbf{a}_0 + \bar{\mathbf{a}}_0^\top\bar{\mathbf{a}}_0\right) \\ \geq \mathbf{a}_0^\top\mathbf{a}_0 = (\mathbf{V}\mathbf{a}_0)^\top(\mathbf{V}\mathbf{a}_0) = (\mathbf{q}')^\top\mathbf{q}' = \|\mathbf{q}'\|_2^2,\end{aligned}$$

where the second equality holds by (51), the third equality holds because $\mathbf{V}^\top\bar{\mathbf{V}} = \mathbf{0}$, $\bar{\mathbf{V}}^\top\mathbf{V} = \mathbf{0}$, $\mathbf{V}^\top\mathbf{V} = \mathbf{I}_K$, $\bar{\mathbf{V}}^\top\bar{\mathbf{V}} = \mathbf{I}_{m-K}$, and the first inequality holds because $\bar{\mathbf{a}}_0^\top\bar{\mathbf{a}}_0 \geq 0$. Thus, we have

$$s' + \gamma_2\mathbf{I}_m \bullet \mathbf{Q}' + \sqrt{\gamma_1}\,\|\mathbf{q}'\|_2 \leq s^* + \gamma_2\mathbf{I}_m \bullet \mathbf{Q}^* + \sqrt{\gamma_1}\,\|\mathbf{q}^*\|_2. \tag{58}$$

Combining (57) and (58) leads to

$$s' + \gamma_2\mathbf{I}_m \bullet \mathbf{Q}' + \sqrt{\gamma_1}\,\|\mathbf{q}'\|_2 = s^* + \gamma_2\mathbf{I}_m \bullet \mathbf{Q}^* + \sqrt{\gamma_1}\,\|\mathbf{q}^*\|_2,$$

which indicates that $(\mathbf{x}', s', \hat{\boldsymbol{\lambda}}', \mathbf{q}', \mathbf{Q}')$ is also an optimal solution of Problem (4). Meanwhile, note that $\text{rank}(\mathbf{Q}') = \text{rank}(\mathbf{V}\mathbf{Y}_{11}\mathbf{V}^\top) \leq \min\{\text{rank}(\mathbf{V}), \text{rank}(\mathbf{Y}_{11})\} \leq K$, $\boldsymbol{\delta} = \mathbf{a}_0$, and $\boldsymbol{\nu}_k = \mathbf{a}_k$ for any $k \in [K]$. Thus, the proof is complete. □

B.5. Proof of Theorem 3

We construct a solution $(\mathbf{x}^\dagger, s^\dagger, \hat{\boldsymbol{\lambda}}^\dagger, \mathbf{q}_\mathrm{r}^\dagger, \mathbf{Q}_\mathrm{r}^\dagger, \mathbf{B}^\dagger)$ of Problems (14) and (15) by setting $\mathbf{x}^\dagger = \mathbf{x}^*$, $s^\dagger = s^*$, $\hat{\boldsymbol{\lambda}}^\dagger = \hat{\boldsymbol{\lambda}}^*$, $\mathbf{q}_\mathrm{r}^\dagger = (\delta^\top, \mathbf{0}_{m_1-K}^\top)^\top$, $\mathbf{Q}_\mathrm{r}^\dagger = \begin{bmatrix} \mathbf{Y}_{11} & \mathbf{0}_{K\times(m_1-K)} \\ \mathbf{0}_{(m_1-K)\times K} & \mathbf{0}_{(m_1-K)\times(m_1-K)} \end{bmatrix}$, and $\mathbf{B}^\dagger = [\mathbf{V}, \mathbf{C}]$, where $\mathbf{C} \in \mathbb{R}^{m\times(m_1-K)}$ and $[\mathbf{V}, \mathbf{C}]^\top [\mathbf{V}, \mathbf{C}] = \mathbf{I}_{m_1}$. First, we show this constructed solution is feasible to Problems (14) and (15). Clearly, this solution satisfies constraints (15c). In addition, from Problem (4), as $\mathbf{q}' = \mathbf{V}\delta$ and $\mathbf{Q}' = \mathbf{V}\mathbf{Y}_{11}\mathbf{V}^\top$, for any $k \in [K]$, we have

$$\begin{bmatrix} S_k & \frac{1}{2}\left(\mathbf{V}\delta + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top (\mathbf{A}^\top \boldsymbol{\lambda}_k^* - y_k(\mathbf{x}^*))\right)^\top \\ \frac{1}{2}\left(\mathbf{V}\delta + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top (\mathbf{A}^\top \boldsymbol{\lambda}_k^* - y_k(\mathbf{x}^*))\right) & \mathbf{V}\mathbf{Y}_{11}\mathbf{V}^\top \end{bmatrix} \succeq 0,$$

which, by Schur complement, is equivalent to

$$S_k\left(\mathbf{V}\mathbf{Y}_{11}\mathbf{V}^\top\right) \succeq \frac{1}{4}\left(\mathbf{V}\delta + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top \left(\mathbf{A}^\top \boldsymbol{\lambda}_k^* - y_k(\mathbf{x}^*)\right)\right)\left(\mathbf{V}\delta + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top \left(\mathbf{A}^\top \boldsymbol{\lambda}_k^* - y_k(\mathbf{x}^*)\right)\right)^\top. \tag{59}$$

From (59), for any $k \in [K]$, we have the following inequality holds by Lemma 2:

$$\begin{aligned} &S_k\left([\mathbf{V}, \mathbf{C}]^\top \mathbf{V}\mathbf{Y}_{11}\mathbf{V}^\top [\mathbf{V}, \mathbf{C}]\right) \\ &\quad \succeq \frac{1}{4}[\mathbf{V}, \mathbf{C}]^\top \left(\mathbf{V}\delta + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top (\mathbf{A}^\top \boldsymbol{\lambda}_k^* - y_k(\mathbf{x}^*))\right)\left(\mathbf{V}\delta + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top (\mathbf{A}^\top \boldsymbol{\lambda}_k^* - y_k(\mathbf{x}^*))\right)^\top [\mathbf{V}, \mathbf{C}], \end{aligned}$$

which is equivalent to

$$S_k\mathbf{Q}_\mathrm{r}^\dagger \succeq \frac{1}{4}\left(\mathbf{q}_\mathrm{r}^\dagger + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}^\dagger\right)^\top (\mathbf{A}^\top \boldsymbol{\lambda}_k^* - y_k(\mathbf{x}^*))\right)\left(\mathbf{q}_\mathrm{r}^\dagger + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}^\dagger\right)^\top (\mathbf{A}^\top \boldsymbol{\lambda}_k^* - y_k(\mathbf{x}^*))\right)^\top \tag{60}$$

by the construction of the solution $\mathbf{q}_\mathrm{r}^\dagger, \mathbf{Q}_\mathrm{r}^\dagger, \mathbf{B}^\dagger$ and $[\mathbf{V}, \mathbf{C}]^\top \mathbf{V} = [\mathbf{I}_K, \mathbf{0}_{K\times(m_1-K)}]^\top$. By Schur complement, (60) indicates that the constructed solution $(\mathbf{x}^\dagger, s^\dagger, \hat{\boldsymbol{\lambda}}^\dagger, \mathbf{q}_\mathrm{r}^\dagger, \mathbf{Q}_\mathrm{r}^\dagger, \mathbf{B}^\dagger)$ also satisfies constraints (15b) and thus it is a feasible solution of Problems (14) and (15).

Second, we show the objective value of this feasible solution $(\mathbf{x}^\dagger, s^\dagger, \hat{\boldsymbol{\lambda}}^\dagger, \mathbf{q}_\mathrm{r}^\dagger, \mathbf{Q}_\mathrm{r}^\dagger, \mathbf{B}^\dagger)$ is equal to the optimal value of Problem (4). The objective value corresponding to this solution is

$$\begin{aligned} s^\dagger + \gamma_2\mathbf{I}_{m_1} \bullet \mathbf{Q}_\mathrm{r}^\dagger + \sqrt{\gamma_1}\left\|\mathbf{q}_\mathrm{r}^\dagger\right\|_2 &= s^* + \gamma_2\mathbf{I}_K \bullet \mathbf{Y}_{11} + \sqrt{\gamma_1}\|\delta\|_2 \\ &= s^* + \gamma_2\mathbf{I}_K \bullet (\mathbf{Y}_{11}\mathbf{V}^\top\mathbf{V}) + \sqrt{\gamma_1}\|\delta\|_2 \\ &= s^* + \gamma_2\mathbf{I}_m \bullet (\mathbf{V}\mathbf{Y}_{11}\mathbf{V}^\top) + \sqrt{\gamma_1}\|\delta\|_2 \\ &= s^* + \gamma_2\mathbf{I}_m \bullet \mathbf{Q}' + \sqrt{\gamma_1}\|\delta\|_2 \\ &= s^* + \gamma_2\mathbf{I}_m \bullet \mathbf{Q}' + \sqrt{\gamma_1}\sqrt{\delta^\top\delta} \\ &= s^* + \gamma_2\mathbf{I}_m \bullet \mathbf{Q}' + \sqrt{\gamma_1}\sqrt{\delta^\top\mathbf{V}^\top\mathbf{V}\delta} \\ &= s^* + \gamma_2\mathbf{I}_m \bullet \mathbf{Q}' + \sqrt{\gamma_1}\sqrt{\mathbf{q}'^\top\mathbf{q}'} \\ &= s^* + \gamma_2\mathbf{I}_m \bullet \mathbf{Q}' + \sqrt{\gamma_1}\|\mathbf{q}'\|_2 \end{aligned}$$

$$= \Theta_{\rm M}(m), \tag{61}$$

where the first equality holds by the construction of $(\mathbf{x}^\dagger, s^\dagger, \hat{\boldsymbol{\lambda}}^\dagger, \mathbf{q}_{\rm r}^\dagger, \mathbf{Q}_{\rm r}^\dagger, \mathbf{B}^\dagger)$, the second equality holds because $\mathbf{V}^\top \mathbf{V} = \mathbf{I}_K$, the third equality holds by the cyclic property of a matrix's trace, the fourth equality holds by the definition of $\mathbf{Q}'$ in Theorem 2, and the seventh equality holds because $\mathbf{q}' = \mathbf{V}\delta$. $\square$

B.6. Proof of Proposition 2

We consider the Lagrangian dual of the inner minimization part (i.e., Problem (15)) of Problem (14) as follows:

$$\max_{\substack{\begin{bmatrix} t_k & \mathbf{p}_k^\top \\ \mathbf{p}_k & \mathbf{P}_k \end{bmatrix} \succeq 0, \\ \forall k \in [K], \\ \mathbf{Z} \succeq 0}} \; \min_{\substack{\mathbf{x}, s, \hat{\lambda} \geq 0, \\ \mathbf{q}_{\rm r}, \mathbf{Q}_{\rm r} \succeq 0}} \mathcal{L}\left(\mathbf{x}, s, \hat{\lambda}, \mathbf{q}_{\rm r}, \mathbf{Q}_{\rm r}; \mathbf{Z}, t_k, \mathbf{p}_k, \mathbf{P}_k, \forall k \in [K]\right), \tag{62}$$

where the Lagrangian function

$$\begin{aligned}
&\mathcal{L}\left(\mathbf{x}, s, \hat{\lambda}, \mathbf{q}_{\rm r}, \mathbf{Q}_{\rm r}; \mathbf{Z}, t_k, \mathbf{p}_k, \mathbf{P}_k, \forall k \in [K]\right) \\
=& s + \gamma_2 \mathbf{I}_{m_1} \bullet \mathbf{Q}_{\rm r} + \sqrt{\gamma_1}\, \|\mathbf{q}_{\rm r}\|_2 - \mathbf{Z} \bullet \left(\sum_{i=1}^{n} (\boldsymbol{\Delta}_i x_i) + \boldsymbol{\Delta}_0 \right) - \sum_{k=1}^{K} \begin{bmatrix} t_k & \mathbf{p}_k^\top \\ \mathbf{p}_k & \mathbf{P}_k \end{bmatrix} \bullet \\
&\begin{bmatrix} s - y_k^0(\mathbf{x}) - \boldsymbol{\lambda}_k^\top b - y_k(\mathbf{x})^\top \boldsymbol{\mu} + \boldsymbol{\lambda}_k^\top \mathbf{A} \boldsymbol{\mu} & \frac{1}{2}\left(\mathbf{q}_{\rm r} + \left(\mathbf{U} \boldsymbol{\Lambda}^{\frac{1}{2}} \mathbf{B} \right)^\top (\mathbf{A}^\top \boldsymbol{\lambda}_k - y_k(\mathbf{x})) \right)^\top \\ \frac{1}{2}\left(\mathbf{q}_{\rm r} + \left(\mathbf{U} \boldsymbol{\Lambda}^{\frac{1}{2}} \mathbf{B} \right)^\top (\mathbf{A}^\top \boldsymbol{\lambda}_k - y_k(\mathbf{x})) \right) & \mathbf{Q}_{\rm r} \end{bmatrix} \\
=& \left(1 - \sum_{k=1}^{K} t_k\right) s - \sum_{k=1}^{K} \left(t_k (\mathbf{A}\boldsymbol{\mu} - \mathbf{b})^\top + \mathbf{p}_k^\top \left(\mathbf{U} \boldsymbol{\Lambda}^{\frac{1}{2}} \mathbf{B} \right)^\top \mathbf{A}^\top \right) \boldsymbol{\lambda}_k + \sqrt{\gamma_1}\, \|\mathbf{q}_{\rm r}\|_2 - \sum_{k=1}^{K} \mathbf{p}_k^\top \mathbf{q}_{\rm r} \\
&+ \left(\gamma_2 \mathbf{I}_{m_1} - \sum_{k=1}^{K} \mathbf{P}_k \right) \bullet \mathbf{Q}_{\rm r} - \sum_{i=1}^{\tau} \sum_{j=1}^{\tau} z_{ij} \left(\mathbf{a}_{ij} \mathbf{x} + a_{ij}^0 \right) + \sum_{k=1}^{K} \left(t_k y_k^0(x) + \left(t_k \boldsymbol{\mu}^\top + \mathbf{p}_k^\top \left(\mathbf{U} \boldsymbol{\Lambda}^{\frac{1}{2}} \mathbf{B} \right)^\top \right) y_k(x) \right) \\
=& \left(1 - \sum_{k=1}^{K} t_k\right) s - \sum_{k=1}^{K} \left(t_k (\mathbf{A}\boldsymbol{\mu} - \mathbf{b})^\top + \mathbf{p}_k^\top \left(\mathbf{U} \boldsymbol{\Lambda}^{\frac{1}{2}} \mathbf{B} \right)^\top \mathbf{A}^\top \right) \boldsymbol{\lambda}_k + \sqrt{\gamma_1}\, \|\mathbf{q}_{\rm r}\|_2 - \sum_{k=1}^{K} \mathbf{p}_k^\top \mathbf{q}_{\rm r} \\
&+ \left(\gamma_2 \mathbf{I}_{m_1} - \sum_{k=1}^{K} \mathbf{P}_k \right) \bullet \mathbf{Q}_{\rm r} - \sum_{i=1}^{\tau} \sum_{j=1}^{\tau} z_{ij} \mathbf{a}_{ij} \mathbf{x} - \sum_{i=1}^{\tau} \sum_{j=1}^{\tau} z_{ij} a_{ij}^0 + \sum_{k=1}^{K} \left(t_k \mathbf{w}_k^0 + \left(t_k \boldsymbol{\mu}^\top + \mathbf{p}_k^\top \left(\mathbf{U} \boldsymbol{\Lambda}^{\frac{1}{2}} \mathbf{B} \right)^\top \right) \mathbf{W}_k \right) \mathbf{x} \\
&+ \sum_{k=1}^{K} \left(t_k d_k^0 + \left(t_k \boldsymbol{\mu}^\top + \mathbf{p}_k^\top \left(\mathbf{U} \boldsymbol{\Lambda}^{\frac{1}{2}} \mathbf{B} \right)^\top \right) \mathbf{d}_k \right) \\
=& \left(1 - \sum_{k=1}^{K} t_k\right) s - \sum_{k=1}^{K} \left(t_k (\mathbf{A}\boldsymbol{\mu} - \mathbf{b})^\top + \mathbf{p}_k^\top \left(\mathbf{U} \boldsymbol{\Lambda}^{\frac{1}{2}} \mathbf{B} \right)^\top \mathbf{A}^\top \right) \boldsymbol{\lambda}_k + \sqrt{\gamma_1}\, \|\mathbf{q}_{\rm r}\|_2 - \sum_{k=1}^{K} \mathbf{p}_k^\top \mathbf{q}_{\rm r} \\
&+ \left(\gamma_2 \mathbf{I}_{m_1} - \sum_{k=1}^{K} \mathbf{P}_k \right) \bullet \mathbf{Q}_{\rm r} + \left(\sum_{k=1}^{K} \left(t_k \mathbf{w}_k^0 + \left(t_k \boldsymbol{\mu}^\top + \mathbf{p}_k^\top \left(\mathbf{U} \boldsymbol{\Lambda}^{\frac{1}{2}} \mathbf{B} \right)^\top \right) \mathbf{W}_k \right) - \sum_{i=1}^{\tau} \sum_{j=1}^{\tau} z_{ij} \mathbf{a}_{ij} \right) \mathbf{x} \\
&+ \sum_{k=1}^{K} \left(t_k d_k^0 + \left(t_k \boldsymbol{\mu}^\top + \mathbf{p}_k^\top \left(\mathbf{U} \boldsymbol{\Lambda}^{\frac{1}{2}} \mathbf{B} \right)^\top \right) \mathbf{d}_k \right) - \sum_{i=1}^{\tau} \sum_{j=1}^{\tau} z_{ij} a_{ij}^0.
\end{aligned}$$

To present the objective value of the inner minimization problem of (62) from going to negative infinity, we require

$$1-\sum_{k=1}^{K}t_k=0,\ \sqrt{\gamma_1}-\left\|\sum_{k=1}^{K}\mathbf{p}_k\right\|_2\geq 0,\ \gamma_2\mathbf{I}_{m_1}-\sum_{k=1}^{K}\mathbf{P}_k\succeq 0, \tag{63a}$$

$$t_k\left(\mathbf{A}\boldsymbol{\mu}-\mathbf{b}\right)^\top+\mathbf{p}_k^\top\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\right)^\top\mathbf{A}^\top\leq 0,\ \forall k\in[K], \tag{63b}$$

$$\sum_{k=1}^{K}\left(t_k\mathbf{w}_k^0+\left(t_k\boldsymbol{\mu}^\top+\mathbf{p}_k^\top\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\right)^\top\right)\mathbf{W}_k\right)-\sum_{i=1}^{\tau}\sum_{j=1}^{\tau}z_{ij}\mathbf{a}_{ij}=0, \tag{63c}$$

$$\begin{bmatrix} t_k & \mathbf{p}_k^\top \\ \mathbf{p}_k & \mathbf{P}_k \end{bmatrix}\succeq 0,\ \forall k\in[K],\ \mathbf{Z}\succeq 0. \tag{63d}$$

Then, the dual problem of Problem (15) can be described as follows:

$$\begin{aligned}\max_{t_k,\mathbf{p}_k,\mathbf{P}_k,\forall k\in[K],\mathbf{Z}}\ & \sum_{k=1}^{K}\left(t_k d_k^0+\left(t_k\boldsymbol{\mu}^\top+\mathbf{p}_k^\top\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\right)^\top\right)\mathbf{d}_k\right)-\sum_{i=1}^{\tau}\sum_{j=1}^{\tau}z_{ij}a_{ij}^0 \\ \text{s.t.}\quad & (63a)-(63d).\end{aligned} \tag{64}$$

By integrating the outer maximization part of Problem (14) and Problem (64), we obtain the bilinear SDP problem (18). Now we would like to prove the strong duality between Problem (15) and Problem (64); that is, these two problems share the same optimal value, which further shows that Problem (14) has the same optimal value as Problem (18). To that end, we find an interior point of Problem (15).

Let $\mathbf{x}'$ be an interior point in $\mathcal{X}$, we can construct an interior point by setting $\hat{\boldsymbol{\lambda}}'=\{\mathbf{1}_l,\ldots,\mathbf{1}_l\}$, $s'=\sum_{k=1}^{K}\left|y_k^0(\mathbf{x}')+\mathbf{1}_l^\top\mathbf{b}+y_k(\mathbf{x}')^\top\boldsymbol{\mu}-\mathbf{1}_l^\top\mathbf{A}\boldsymbol{\mu}\right|+1$, $\mathbf{q}_\mathrm{r}'=0$, and $\mathbf{Q}_\mathrm{r}'=\sum_{k=1}^{K}1/(4(s'-y_k^0(\mathbf{x}')-\mathbf{1}_l^\top b-y_k(\mathbf{x}')^\top\boldsymbol{\mu}+\mathbf{1}_l^\top\mathbf{A}\boldsymbol{\mu}))(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B})^\top(\mathbf{A}^\top\mathbf{1}_l-y_k(\mathbf{x}'))(\mathbf{A}^\top\mathbf{1}_l-y_k(\mathbf{x}'))^\top(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B})+\mathbf{I}_{m_1}$. Clearly, $(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B})^\top(\mathbf{A}^\top\mathbf{1}_l-y_k(\mathbf{x}'))(\mathbf{A}^\top\mathbf{1}_l-y_k(\mathbf{x}'))^\top(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B})\succeq 0$. Thus, $\mathbf{Q}_\mathrm{r}'\succ 0$. Now we only need to show that constraints (15b) hold in the positive-definite sense with respect to this constructed solution.

By the construction of $\mathbf{Q}_\mathrm{r}'$, for any $k\in[K]$, we have

$$\begin{aligned}&\mathbf{Q}_\mathrm{r}'-\frac{\left(\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\right)^\top\left(\mathbf{A}^\top\mathbf{1}_l-y_k(\mathbf{x}')\right)\right)\left(\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\right)^\top\left(\mathbf{A}^\top\mathbf{1}_l-y_k(\mathbf{x}')\right)\right)^\top}{4\left(s'-y_k^0(\mathbf{x}')-\mathbf{1}_l^\top b-y_k(\mathbf{x}')^\top\boldsymbol{\mu}+\mathbf{1}_l^\top\mathbf{A}\boldsymbol{\mu}\right)}\\ &=\sum_{\forall k'\in[K]:k'\neq k}\frac{\left(\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\right)^\top\left(\mathbf{A}^\top\mathbf{1}_l-y_{k'}(\mathbf{x}')\right)\right)\left(\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\right)^\top\left(\mathbf{A}^\top\mathbf{1}_l-y_{k'}(\mathbf{x}')\right)\right)^\top}{4\left(s'-y_{k'}^0(\mathbf{x}')-\mathbf{1}_l^\top b-y_{k'}(\mathbf{x}')^\top\boldsymbol{\mu}+\mathbf{1}_l^\top\mathbf{A}\boldsymbol{\mu}\right)}+\mathbf{I}_{m_1}\succ 0,\end{aligned} \tag{65}$$

where $s'-y_{k'}^0(\mathbf{x}')-\mathbf{1}_l^\top b-y_{k'}(\mathbf{x}')^\top\boldsymbol{\mu}+\mathbf{1}_l^\top\mathbf{A}\boldsymbol{\mu}>0$ by the construction of s'. By Schur complement, (65) is equivalent to

$$\begin{bmatrix} s'-y_k^0(\mathbf{x}')-\mathbf{1}_l^\top b-y_k(\mathbf{x}')^\top\boldsymbol{\mu}+\mathbf{1}_l^\top\mathbf{A}\boldsymbol{\mu} & \frac{1}{2}\left(\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\right)^\top\left(\mathbf{A}^\top\mathbf{1}_l-y_k(\mathbf{x}')\right)\right)^\top \\ \frac{1}{2}\left(\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\right)^\top\left(\mathbf{A}^\top\mathbf{1}_l-y_k(\mathbf{x}')\right)\right) & \mathbf{Q}_\mathrm{r}'\end{bmatrix}\succ 0,\ \forall k\in[K].$$

Thus, $(\mathbf{x}', s', \hat{\boldsymbol{\lambda}}', \mathbf{q}_{\mathrm{r}}', \mathbf{Q}_{\mathrm{r}}')$ is an interior point of Problem (15) and the strong duality between Problem (15) and Problem (64) holds. □

B.7. Proof of proposition 3

First, we have $\Theta_{\mathrm{M}}(m) \geq \Theta_{\mathrm{L}}(m_1) \geq \underline{\Theta}(m_1, \mathbf{B}') = s^* + \gamma_2 \mathbf{I}_{m_1} \bullet \mathbf{Q}_{\mathrm{r}}^* + \sqrt{\gamma_1} \|\mathbf{q}_{\mathrm{r}}^*\|_2$, where the first inequality holds by conclusion (i) of Theorem 1 and the second inequality holds because $\mathbf{B}'$ is a feasible solution of Problem (14) and this problem is a maximization problem.

Next, we would like to construct a feasible solution $(\mathbf{x}', s', \hat{\boldsymbol{\lambda}}', \mathbf{q}', \mathbf{Q}')$ of Problem (4). We set $\mathbf{x}' = \mathbf{x}^*$, $\hat{\boldsymbol{\lambda}}' = \hat{\boldsymbol{\lambda}}^*$, $s' = s^* + s_0$, $\mathbf{q}' = \mathbf{B}'\mathbf{q}_{\mathrm{r}}^*$, and $\mathbf{Q}' = \mathbf{B}'\mathbf{Q}_{\mathrm{r}}^*(\mathbf{B}')^\top + \mathbf{Q}_0$, where $s_0 \geq 0$ and $\mathbf{Q}_0 \succeq 0$ and their values will be decided later. Clearly, this solution satisfies constraints (4c). For this solution to satisfy constraints (4b), the values s_0 and $\mathbf{Q}_0$ should satisfy

$$(S_k + s_0)\left(\mathbf{B}'\mathbf{Q}_{\mathrm{r}}^*(\mathbf{B}')^\top + \mathbf{Q}_0\right) \succeq \frac{1}{4}\left(\mathbf{B}'\mathbf{q}_{\mathrm{r}}^* + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top \left(\mathbf{A}^\top \boldsymbol{\lambda}_k^* - y_k(\mathbf{x}^*)\right)\right) \times \left(\mathbf{B}'\mathbf{q}_{\mathrm{r}}^* + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top \left(\mathbf{A}^\top \boldsymbol{\lambda}_k^* - y_k(\mathbf{x}^*)\right)\right)^\top = \frac{1}{4}\mathbf{M}_k, \ \forall k \in [K]. \quad (66)$$

Note that, if $(S + s_0)\mathbf{Q}_0 \succeq (1/4)\mathbf{M}_k$ for any $k \in [K]$, then (66) holds. By the definition of $\mathbf{M}_k$, we have $\mathbf{M}_k \succeq 0$ for any $k \in [K]$. Therefore, for any $s_0 \geq 0$, we can construct

$$\mathbf{Q}_0 = \sum_{k=1}^{K} \frac{1}{4(S + s_0)} \mathbf{M}_k$$

such that (66) holds and hence $(\mathbf{x}', s', \hat{\boldsymbol{\lambda}}', \mathbf{q}', \mathbf{Q}')$ is a feasible solution of Problem (4). The objective value (denoted by Θ_{M}') with respect to this constructed solution is

$$\begin{aligned} s' + \gamma_2 \mathbf{I}_m \bullet \mathbf{Q}' + \sqrt{\gamma_1}\,\|\mathbf{q}'\|_2 &= s^* + s_0 + \gamma_2 \mathbf{I}_m \bullet \mathbf{B}'\mathbf{Q}_{\mathrm{r}}^*(\mathbf{B}')^\top + \gamma_2 \mathbf{I}_m \bullet \mathbf{Q}_0 + \sqrt{\gamma_1}\,\|\mathbf{B}'\mathbf{q}_{\mathrm{r}}^*\|_2 \\ &= s^* + s_0 + \gamma_2 \mathbf{I}_{m_1} \bullet \mathbf{Q}_{\mathrm{r}}^* + \gamma_2 \mathbf{I}_m \bullet \mathbf{Q}_0 + \sqrt{\gamma_1}\,\|\mathbf{q}_{\mathrm{r}}^*\|_2 \\ &= \underline{\Theta}(m_1, \mathbf{B}') + s_0 + \sum_{k=1}^{K} \frac{\gamma_2}{4(S + s_0)} \mathbf{I}_m \bullet \mathbf{M}_k, \end{aligned}$$

where the second equality holds because $\mathbf{I}_m \bullet \mathbf{B}'\mathbf{Q}_{\mathrm{r}}^*(\mathbf{B}')^\top = \mathbf{I}_{m_1} \bullet \mathbf{Q}_{\mathrm{r}}^*(\mathbf{B}')^\top\mathbf{B}' = \mathbf{I}_{m_1} \bullet \mathbf{Q}_{\mathrm{r}}^*$ and $(\mathbf{q}_{\mathrm{r}}^*)^\top(\mathbf{B}')^\top\mathbf{B}'\mathbf{q}_{\mathrm{r}}^* = (\mathbf{q}_{\mathrm{r}}^*)^\top\mathbf{q}_{\mathrm{r}}^*$. As this constructed solution is a feasible solution of Problem (4), which is a minimization problem, we have $\Theta_{\mathrm{M}}(m) \leq \Theta_{\mathrm{M}}'$ It follows that

$$\Theta_{\mathrm{M}}(m) - \Theta_{\mathrm{L}}(m_1) \leq \Theta_{\mathrm{M}}' - \underline{\Theta}(m_1, \mathbf{B}') = s_0 + \sum_{k=1}^{K} \frac{\gamma_2}{4(S + s_0)} \mathbf{I}_m \bullet \mathbf{M}_k. \quad (67)$$

We further choose a value of s_0 to minimize the the right-hand side (RHS) of (67). Note that (i) If $\sqrt{P} - S < 0$, then the RHS of (67) is minimized at P/S with $s_0 = 0$; (ii) If $\sqrt{P} - S \geq 0$, then the RHS of (67) is minimized at $2\sqrt{P} - S$ with $s_0 = \sqrt{P} - S$. Therefore, we conclude that the proposition holds. □

B.8. Proof of Theorem 4

(i) For any $\boldsymbol{\xi}_{\mathrm{I}} \sim \mathbb{P}_{\mathrm{I}} \in \mathcal{D}_{\mathrm{M}}$, we have $\mathbb{E}_{\mathbb{P}_{\mathrm{I}}}[\boldsymbol{\xi}_{\mathrm{I}}\boldsymbol{\xi}_{\mathrm{I}}^\top] \preceq \gamma_2 \mathbf{I}_{m_1}$. Then, by Lemma 2, for any given $\mathbf{x} \in \mathcal{X}$ and $\mathbf{B} \in \mathcal{B}_{m_1}$, i.e., $\mathbf{B}^\top\mathbf{B} = \mathbf{I}_{m_1}$, we further have $\mathbf{B}^\top(\mathbb{E}_{\mathbb{P}_{\mathrm{I}}}[\boldsymbol{\xi}_{\mathrm{I}}\boldsymbol{\xi}_{\mathrm{I}}^\top])\mathbf{B} \preceq \mathbf{B}^\top(\gamma_2\mathbf{I}_{m_1})\mathbf{B}$, i.e., $\mathbb{E}_{\mathbb{P}_{\mathrm{I}}}[\mathbf{B}^\top\boldsymbol{\xi}_{\mathrm{I}}\boldsymbol{\xi}_{\mathrm{I}}^\top\mathbf{B}] \preceq \gamma_2\mathbf{B}^\top\mathbf{I}_{m_1}\mathbf{B} = \gamma_2\mathbf{I}_{m_1}$. It follows that $\mathcal{D}_{\mathrm{M}} \subseteq \mathcal{D}_{\mathrm{U}}$. Thus, given any $\mathbf{x} \in \mathcal{X}$ and $\mathbf{B} \in \mathcal{B}_{m_1}$, we have

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

It follows that

$$\min_{\mathbf{B} \in \mathcal{B}_{m_1}} \min_{\mathbf{x} \in \mathcal{X}} \max_{\mathbb{P}_{\mathrm{I}} \in \mathcal{D}_{\mathrm{U}}} \mathbb{E}_{\mathbb{P}_{\mathrm{I}}}\left[f\left(\mathbf{x}, \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}} + \boldsymbol{\mu}\right)\right] \geq \min_{\mathbf{x} \in \mathcal{X}} \max_{\mathbb{P}_{\mathrm{I}} \in \mathcal{D}_{\mathrm{M}}} \mathbb{E}_{\mathbb{P}_{\mathrm{I}}}\left[f\left(\mathbf{x}, \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}} + \boldsymbol{\mu}\right)\right],$$

which demonstrates that the optimal value of Problem (19) is an upper bound for that of Problem (3) (i.e., Problem (2)).

(ii) Consider any $m_1 < m_2 \leq m$. We have $\mathcal{B}_{m_2} := \{\mathbf{B} \in \mathbb{R}^{m \times m_2} \mid \mathbf{B}^\top\mathbf{B} = \mathbf{I}_{m_2}\}$ and consider an optimal solution $(\mathbf{B}^*, \mathbf{x}^*)$ of Problem (19), i.e., $\min_{\mathbf{B} \in \mathcal{B}_{m_1}} \min_{\mathbf{x} \in \mathcal{X}} \max_{\mathbb{P}_{\mathrm{I}} \in \mathcal{D}_{\mathrm{U}}} \mathbb{E}_{\mathbb{P}_{\mathrm{I}}}[f(\mathbf{x}, \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}} + \boldsymbol{\mu})]$.

Note that $(\mathbf{B}^*)^\top\mathbf{B}^* = \mathbf{I}_{m_1}$. We can then construct $\mathbf{B}' = [\mathbf{B}^*, \mathbf{C}] \in \mathbb{R}^{m \times m_2}$ such that $\mathbf{C} \in \mathbb{R}^{m \times (m_2 - m_1)}$ and $\mathbf{B}' \in \mathcal{B}_{m_2}$, i.e., $(\mathbf{B}')^\top\mathbf{B}' = \mathbf{I}_{m_2}$. With $\mathbf{B}'$, we use $\mathcal{D}'_{\mathrm{U}}$ to denote the corresponding ambiguity set defined in (21). By the second-moment constraint in $\mathcal{D}'_{\mathrm{U}}$, we have

$$\begin{aligned}
&\mathbb{E}_{\mathbb{P}_{\mathrm{I}}}\left[(\mathbf{B}')^\top\boldsymbol{\xi}_{\mathrm{I}}\boldsymbol{\xi}_{\mathrm{I}}^\top\mathbf{B}'\right] \\
=&\mathbb{E}_{\mathbb{P}_{\mathrm{I}}}\left[[\mathbf{B}^*, \mathbf{C}]^\top\boldsymbol{\xi}_{\mathrm{I}}\boldsymbol{\xi}_{\mathrm{I}}^\top[\mathbf{B}^*, \mathbf{C}]\right] \\
=&\mathbb{E}_{\mathbb{P}_{\mathrm{I}}}\begin{bmatrix} (\mathbf{B}^*)^\top\boldsymbol{\xi}_{\mathrm{I}}\boldsymbol{\xi}_{\mathrm{I}}^\top\mathbf{B}^* & (\mathbf{B}^*)^\top\boldsymbol{\xi}_{\mathrm{I}}\boldsymbol{\xi}_{\mathrm{I}}^\top\mathbf{C} \\ \mathbf{C}^\top\boldsymbol{\xi}_{\mathrm{I}}\boldsymbol{\xi}_{\mathrm{I}}^\top\mathbf{B}^* & \mathbf{C}^\top\boldsymbol{\xi}_{\mathrm{I}}\boldsymbol{\xi}_{\mathrm{I}}^\top\mathbf{C} \end{bmatrix} \\
\preceq&\gamma_2\mathbf{I}_{m_2},
\end{aligned}$$

which implies that $\mathbb{E}_{\mathbb{P}_{\mathrm{I}}}[(\mathbf{B}^*)^\top\boldsymbol{\xi}_{\mathrm{I}}\boldsymbol{\xi}_{\mathrm{I}}^\top\mathbf{B}^*] \preceq \gamma_2\mathbf{I}_{m_1}$. It follows that $\mathcal{D}'_{\mathrm{U}} \subseteq \mathcal{D}_{\mathrm{U}}$. Therefore, we have

$$\max_{\mathbb{P}_{\mathrm{I}} \in \mathcal{D}_{\mathrm{U}}} \mathbb{E}_{\mathbb{P}_{\mathrm{I}}}\left[f\left(\mathbf{x}^*, \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}} + \boldsymbol{\mu}\right)\right] \geq \max_{\mathbb{P}_{\mathrm{I}} \in \mathcal{D}'_{\mathrm{U}}} \mathbb{E}_{\mathbb{P}_{\mathrm{I}}}\left[f\left(\mathbf{x}^*, \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}} + \boldsymbol{\mu}\right)\right]. \tag{68}$$

Because $\mathbf{B}' \in \mathcal{B}_{m_2}$ and $\mathbf{x}^* \in \mathcal{X}$, the constructed solution $(\mathbf{B}', \mathbf{x}^*)$ is feasible to the problem $\min_{\mathbf{B} \in \mathcal{B}_{m_2}} \min_{\mathbf{x} \in \mathcal{X}} \max_{\mathbb{P}_{\mathrm{I}} \in \mathcal{D}'_{\mathrm{U}}} \mathbb{E}_{\mathbb{P}_{\mathrm{I}}}[f(\mathbf{x}, \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}} + \boldsymbol{\mu})]$. Then, we have

$$\begin{aligned}
\Theta_{\mathrm{U}}(m_2) &= \min_{\mathbf{B} \in \mathcal{B}_{m_2}} \min_{\mathbf{x} \in \mathcal{X}} \max_{\mathbb{P}_{\mathrm{I}} \in \mathcal{D}'_{\mathrm{U}}} \mathbb{E}_{\mathbb{P}_{\mathrm{I}}}\left[f\left(\mathbf{x}, \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}} + \boldsymbol{\mu}\right)\right] \\
&\leq \max_{\mathbb{P}_{\mathrm{I}} \in \mathcal{D}'_{\mathrm{U}}} \mathbb{E}_{\mathbb{P}_{\mathrm{I}}}\left[f\left(\mathbf{x}^*, \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}} + \boldsymbol{\mu}\right)\right] \\
&\leq \max_{\mathbb{P}_{\mathrm{I}} \in \mathcal{D}_{\mathrm{U}}} \mathbb{E}_{\mathbb{P}_{\mathrm{I}}}\left[f\left(\mathbf{x}^*, \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}} + \boldsymbol{\mu}\right)\right] \\
&= \min_{\mathbf{B} \in \mathcal{B}_{m_1}} \min_{\mathbf{x} \in \mathcal{X}} \max_{\mathbb{P}_{\mathrm{I}} \in \mathcal{D}_{\mathrm{U}}} \mathbb{E}_{\mathbb{P}_{\mathrm{I}}}\left[f\left(\mathbf{x}, \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}} + \boldsymbol{\mu}\right)\right]
\end{aligned}$$

$$= \Theta_{\mathrm{U}}(m_1),$$

where the first inequality holds because $(\mathbf{B}', \mathbf{x}^*)$ is a feasible solution of the problem $\min_{\mathbf{B}\in\mathcal{B}_{m_2}} \min_{\mathbf{x}\in\mathcal{X}} \max_{\mathbb{P}_{\mathrm{I}}\in\mathcal{D}'_{\mathrm{U}}} \mathbb{E}_{\mathbb{P}_{\mathrm{I}}}[f(\mathbf{x}, \mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}} + \boldsymbol{\mu})]$, the second inequality holds by (68), and the second equality holds because $(\mathbf{B}^*, \mathbf{x}^*)$ is an optimal solution of Problem (19). That is, the optimal value of Problem (19) is nonincreasing in m_1.

(iii) When $m_1 = m$, we have $\mathbf{B} \in \mathcal{B}_m \subseteq \mathbb{R}^{m\times m}$, i.e., $\mathbf{B}^\top\mathbf{B} = \mathbf{I}_m$. First, we have $\Theta_{\mathrm{U}}(m) \geq \Theta_{\mathrm{M}}(m)$ by the conclusion (i). Second, when $\mathbf{B} = \mathbf{I}_m$, Problem (19) becomes Problem (3). Because $\mathbf{B} = \mathbf{I}_m$ is a feasible solution of Problem (19), it follows that $\Theta_{\mathrm{U}}(m) \leq \Theta_{\mathrm{M}}(m)$. Therefore, we have $\Theta_{\mathrm{U}}(m) = \Theta_{\mathrm{M}}(m)$. □

B.9. Proof of Proposition 4

First, by Theorem 3 in Cheramin et al. (2022), Problem (20) has the same optimal value as the following problem:

$$\min_{\mathbf{x},s,\mathbf{q},\mathbf{Q}_{\mathrm{r}}} \quad s + \gamma_2\mathbf{I}_{m_1} \bullet \mathbf{Q}_{\mathrm{r}} + \sqrt{\gamma_1}\,\|\mathbf{q}\|_2 \tag{69a}$$

$$\text{s.t.} \quad s \geq f\left(\mathbf{x}, \mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}} + \boldsymbol{\mu}\right) - \boldsymbol{\xi}_{\mathrm{I}}^\top\mathbf{B}\mathbf{Q}_{\mathrm{r}}\mathbf{B}^\top\boldsymbol{\xi}_{\mathrm{I}} - \mathbf{q}^\top\boldsymbol{\xi}_{\mathrm{I}},\ \forall\boldsymbol{\xi}_{\mathrm{I}} \in \mathcal{S}_{\mathrm{I}}, \tag{69b}$$

$$\mathbf{Q}_{\mathrm{r}} \succeq 0,\ \mathbf{x} \in \mathcal{X},\ \mathbf{Q}_{\mathrm{r}} \in \mathbb{R}^{m_1\times m_1},\ \mathbf{q} \in \mathbb{R}^m. \tag{69c}$$

Next, we apply the strong duality theorem to constraints (69b). We define

$$g_k(\boldsymbol{\xi}_{\mathrm{I}}) = s + \boldsymbol{\xi}_{\mathrm{I}}^\top\mathbf{B}\mathbf{Q}_{\mathrm{r}}\mathbf{B}^\top\boldsymbol{\xi}_{\mathrm{I}} + \mathbf{q}^\top\boldsymbol{\xi}_{\mathrm{I}} - y_k^0(\mathbf{x}) - y_k(\mathbf{x})^\top\left(\mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}} + \boldsymbol{\mu}\right),\ \forall k \in [K].$$

As function $f(\mathbf{x}, \boldsymbol{\xi})$ is piecewise linear convex, we can reformulate (69b) as

$$g_k(\boldsymbol{\xi}_{\mathrm{I}}) \geq 0,\ \forall\boldsymbol{\xi}_{\mathrm{I}} \in \mathcal{S}_{\mathrm{I}},\ \forall k \in [K],$$

which is equivalent to

$$\min_{\mathbf{A}\left(\mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}}+\boldsymbol{\mu}\right)\leq\mathbf{b},\ \boldsymbol{\xi}_{\mathrm{I}}\in\mathbb{R}^m} g_k(\boldsymbol{\xi}_{\mathrm{I}}) \geq 0,\ \forall k \in [K]. \tag{70}$$

For any $k \in [K]$, the Lagrangian dual problem of $\min_{\mathbf{A}(\mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}}+\boldsymbol{\mu})\leq\mathbf{b},\ \boldsymbol{\xi}_{\mathrm{I}}\in\mathbb{R}^m} g_k(\boldsymbol{\xi}_{\mathrm{I}})$ is

$$\max_{\boldsymbol{\lambda}_k\geq 0} \min_{\boldsymbol{\xi}_{\mathrm{I}}\in\mathbb{R}^m} g_k(\boldsymbol{\xi}_{\mathrm{I}}) + \boldsymbol{\lambda}_k^\top\left(\mathbf{A}\left(\mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}} + \boldsymbol{\mu}\right) - \mathbf{b}\right),$$

where $\boldsymbol{\lambda}_k \in \mathbb{R}^l$. Because there exists an interior point for the primal problem, the strong duality holds. Thus, constraints (70) are equivalent to

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

which are further equivalent to

$$\exists \boldsymbol{\lambda}_k \geq 0: s + \boldsymbol{\xi}_{\mathrm{I}}^\top \mathbf{B}\mathbf{Q}_{\mathrm{r}}\mathbf{B}^\top \boldsymbol{\xi}_{\mathrm{I}} + \mathbf{q}^\top \boldsymbol{\xi}_{\mathrm{I}} - y_k^0(\mathbf{x}) - y_k(\mathbf{x})^\top \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}} + \boldsymbol{\mu}\right) + \boldsymbol{\lambda}_k^\top \left(\mathbf{A}\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}} + \boldsymbol{\mu}\right) - \mathbf{b}\right) \geq 0,\ \forall \boldsymbol{\xi}_{\mathrm{I}} \in \mathbb{R}^m,\ \forall k \in [K]. \tag{71}$$

Note that $\mathbf{B}^\top\mathbf{B} = \mathbf{I}_{m_1}$; that is, all the column vectors of $\mathbf{B}$ are orthogonal. We can then extend $\mathbf{B}$ to $[\mathbf{B}, \bar{\mathbf{B}}] \in \mathbb{R}^{m\times m}$ with $\bar{\mathbf{B}} \in \mathbb{R}^{m\times(m-m_1)}$ such that all the column vectors of $[\mathbf{B}, \bar{\mathbf{B}}]$ span the space of $\mathbb{R}^m$. Thus, we can always find $\boldsymbol{\xi}_1 \in \mathbb{R}^{m_1}$ and $\boldsymbol{\xi}_2 \in \mathbb{R}^{m-m_1}$ such that

$$\boldsymbol{\xi}_{\mathrm{I}} = \mathbf{B}\boldsymbol{\xi}_1 + \bar{\mathbf{B}}\boldsymbol{\xi}_2.$$

It follows that constraints (71) become

$$\exists \boldsymbol{\lambda}_k \geq 0: s + \boldsymbol{\xi}_1^\top \mathbf{Q}_{\mathrm{r}}\boldsymbol{\xi}_1 + \mathbf{q}^\top (\mathbf{B}\boldsymbol{\xi}_1 + \bar{\mathbf{B}}\boldsymbol{\xi}_2) - y_k^0(\mathbf{x}) - y_k(\mathbf{x})^\top \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}(\mathbf{B}\boldsymbol{\xi}_1 + \bar{\mathbf{B}}\boldsymbol{\xi}_2) + \boldsymbol{\mu}\right) + \boldsymbol{\lambda}_k^\top \left(\mathbf{A}\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}(\mathbf{B}\boldsymbol{\xi}_1 + \bar{\mathbf{B}}\boldsymbol{\xi}_2) + \boldsymbol{\mu}\right) - \mathbf{b}\right) \geq 0,\ \forall \boldsymbol{\xi}_1 \in \mathbb{R}^{m_1}, \boldsymbol{\xi}_2 \in \mathbb{R}^{m-m_1},\ \forall k \in [K]. \tag{72}$$

We further define

$$\mathbf{Z}_k = \begin{bmatrix} s - y_k^0(\mathbf{x}) - \boldsymbol{\lambda}_k^\top \mathbf{b} - y_k(\mathbf{x})^\top \boldsymbol{\mu} + \boldsymbol{\lambda}_k^\top \mathbf{A}\boldsymbol{\mu} & \frac{1}{2}\left(\mathbf{B}^\top \mathbf{q} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\right)^\top (\mathbf{A}^\top\boldsymbol{\lambda}_k - y_k(\mathbf{x}))\right)^\top \\ \frac{1}{2}\left(\mathbf{B}^\top \mathbf{q} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\right)^\top (\mathbf{A}^\top\boldsymbol{\lambda}_k - y_k(\mathbf{x}))\right) & \mathbf{Q}_{\mathrm{r}} \end{bmatrix},\ \forall k \in [K].$$

Thus, we have

$$\begin{aligned}
(72) \iff & \exists \boldsymbol{\lambda}_k \geq 0: \left(1, \boldsymbol{\xi}_1^\top\right) \mathbf{Z}_k \left(1, \boldsymbol{\xi}_1^\top\right)^\top + \boldsymbol{\xi}_2^\top \left(\bar{\mathbf{B}}^\top \mathbf{q} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\bar{\mathbf{B}}\right)^\top (\mathbf{A}^\top\boldsymbol{\lambda}_k - y_k(\mathbf{x}))\right) \geq 0, \\
& \qquad \forall \boldsymbol{\xi}_1 \in \mathbb{R}^{m_1}, \boldsymbol{\xi}_2 \in \mathbb{R}^{m-m_1},\ \forall k \in [K]. \\
\iff & \exists \boldsymbol{\lambda}_k \geq 0: \left(1, \boldsymbol{\xi}_1^\top\right) \mathbf{Z}_k \left(1, \boldsymbol{\xi}_1^\top\right)^\top \geq 0,\ \forall \boldsymbol{\xi}_1 \in \mathbb{R}^{m_1},\ \forall k \in [K]; \qquad (73) \\
& \bar{\mathbf{B}}^\top \mathbf{q} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\bar{\mathbf{B}}\right)^\top (\mathbf{A}^\top\boldsymbol{\lambda}_k - y_k(\mathbf{x})) = 0,\ \forall k \in [K]. \\
\iff & \exists \boldsymbol{\lambda}_k \geq 0: \mathbf{Z}_k \succeq 0,\ \bar{\mathbf{B}}^\top \mathbf{q} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\bar{\mathbf{B}}\right)^\top (\mathbf{A}^\top\boldsymbol{\lambda}_k - y_k(\mathbf{x})) = 0,\ \forall k \in [K]. \\
\iff & \exists \boldsymbol{\lambda}_k \geq 0: \mathbf{Z}_k \succeq 0,\ \bar{\mathbf{B}}^\top \left(\mathbf{q} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top (\mathbf{A}^\top\boldsymbol{\lambda}_k - y_k(\mathbf{x}))\right) = 0,\ \forall k \in [K]. \qquad (74) \\
\iff & \exists \boldsymbol{\lambda}_k \geq 0, \mathbf{u}_k \in \mathbb{R}^{m_1}: \mathbf{Z}_k \succeq 0,\ \mathbf{q} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top (\mathbf{A}^\top\boldsymbol{\lambda}_k - y_k(\mathbf{x})) = \mathbf{B}\mathbf{u}_k,\ \forall k \in [K]. \qquad (75)
\end{aligned}$$

The first equivalence holds due to the definition of $\mathbf{Z}_k$. For the third equivalence, clearly $\Longleftarrow$ follows from the definition of a PSD matrix. To prove $\Longrightarrow$, we consider two possible cases for any $(\eta_0 \in \mathbb{R}, \boldsymbol{\eta}^\top \in \mathbb{R}^{m_1})^\top \in \mathbb{R}^{m_1+1}$: (i) if $\eta_0 = 0$, then $(\eta_0, \boldsymbol{\eta}^\top)\mathbf{Z}_k(\eta_0, \boldsymbol{\eta}^\top)^\top = \boldsymbol{\eta}^\top \mathbf{Q}_{\mathrm{r}}\boldsymbol{\eta} \geq 0$ because $\mathbf{Q}_{\mathrm{r}}$ is PSD; (ii) 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 (73). Therefore, $\Longrightarrow$ holds. For the fifth equivalence, (74) shows that $\mathbf{q} + (\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}})^\top(\mathbf{A}^\top\boldsymbol{\lambda}_k - y_k(\mathbf{x}))$ is in the null space of $\bar{\mathbf{B}}$ and thus cannot be represented by basis vectors in the space of $\bar{\mathbf{B}}$. Because $[\mathbf{B}, \bar{\mathbf{B}}]$ span the space of $\mathbb{R}^m$, we have $\mathbf{q} + (\mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}})^\top(\mathbf{A}^\top\boldsymbol{\lambda}_k - y_k(\mathbf{x}))$ should be in the space of $\mathbf{B}$. That is, there exists $\mathbf{u}_k \in \mathbb{R}^{m_1}$ such that $\mathbf{q} + (\mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}})^\top(\mathbf{A}^\top\boldsymbol{\lambda}_k - y_k(\mathbf{x})) = \mathbf{B}\mathbf{u}_k$ for any $k \in [K]$. Meanwhile, because $\mathbf{B}^\top\mathbf{B} = \mathbf{I}_{m_1}$, we have

$$\mathbf{B}^\top\mathbf{q} + \left(\mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{B}\right)^\top \left(\mathbf{A}^\top\lambda_k - y_k(\mathbf{x})\right) = \mathbf{B}^\top\mathbf{B}\mathbf{u}_k = \mathbf{u}_k, \ \forall k \in [K].$$

Finally, we obtain Problem (22) by replacing constraints (69b) with (75) and replacing $\mathbf{B}^\top\mathbf{q} + (\mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{B})^\top(\mathbf{A}^\top\boldsymbol{\lambda}_k - y_k(\mathbf{x}))$ with $\mathbf{u}_k$. $\square$

B.10. Proof of Theorem 5

Consider $m_1 = K$. We construct a solution $(\mathbf{x}^\dagger, s^\dagger, \hat{\boldsymbol{\lambda}}^\dagger, \mathbf{q}^\dagger, \mathbf{Q}_\mathrm{r}^\dagger, \hat{\mathbf{u}}^\dagger, \mathbf{B}^\dagger)$ of Problem (19) by setting $\mathbf{x}^\dagger = \mathbf{x}^*$, $s^\dagger = s^*$, $\hat{\boldsymbol{\lambda}}^\dagger = \hat{\boldsymbol{\lambda}}^*$, $\mathbf{q}^\dagger = \mathbf{q}' = \mathbf{V}\delta$, $\mathbf{Q}_\mathrm{r}^\dagger = \mathbf{Y}_{11}$, $\mathbf{B}^\dagger = \mathbf{V}$, and $\hat{\mathbf{u}}_k^\dagger = \delta + \boldsymbol{\nu}_k$ $(k \in [K])$.

First, we show this constructed solution is feasible to Problem (19). Clearly, this solution satisfies constraints (22d)–(22e). By the construction of the solution, for any $k \in [K]$, we further have

$$\begin{aligned}\mathbf{q}^\dagger + \left(\mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}\right)^\top \left(\mathbf{A}^\top\boldsymbol{\lambda}_k^\dagger - y_k(\mathbf{x}^\dagger)\right) &= \mathbf{V}\delta + \left(\mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}\right)^\top \left(\mathbf{A}^\top\boldsymbol{\lambda}_k^* - y_k(\mathbf{x}^*)\right) \\ &= \mathbf{V}\delta + \mathbf{V}\boldsymbol{\nu}_k = \mathbf{B}^\dagger\hat{\mathbf{u}}_k^\dagger,\end{aligned}$$

where the first equality holds by the construction of $\mathbf{q}^\dagger$, the second equality holds by (50), and the third equality holds by the construction of $\hat{\mathbf{u}}_k^\dagger$. Thus, this solution satisfies constraints (22c). Meanwhile, $\mathbf{V}^\top\mathbf{V} = \mathbf{I}_K = \mathbf{I}_{m_1}$. It follows that $(\mathbf{B}^\dagger)^\top\mathbf{B}^\dagger = \mathbf{I}_{m_1}$.

In addition, from Problem (4), as $\mathbf{q}' = \mathbf{V}\delta$ and $\mathbf{Q}' = \mathbf{V}\mathbf{Y}_{11}\mathbf{V}^\top$, for any $k \in [K]$, we have

$$\begin{bmatrix} S_k & \frac{1}{2}\left(\mathbf{V}\delta + \left(\mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}\right)^\top \left(\mathbf{A}^\top\boldsymbol{\lambda}_k^* - y_k(\mathbf{x}^*)\right)\right)^\top \\ \frac{1}{2}\left(\mathbf{V}\delta + \left(\mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}\right)^\top \left(\mathbf{A}^\top\boldsymbol{\lambda}_k^* - y_k(\mathbf{x}^*)\right)\right) & \mathbf{V}\mathbf{Y}_{11}\mathbf{V}^\top \end{bmatrix} \succeq 0,$$

which, by Schur complement, is equivalent to

$$\begin{aligned}4S_k\left(\mathbf{V}\mathbf{Y}_{11}\mathbf{V}^\top\right) &\succeq \left(\mathbf{V}\delta + \left(\mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}\right)^\top \left(\mathbf{A}^\top\boldsymbol{\lambda}_k^* - y_k(\mathbf{x}^*)\right)\right)\left(\mathbf{V}\delta + \left(\mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}\right)^\top \left(\mathbf{A}^\top\boldsymbol{\lambda}_k^* - y_k(\mathbf{x}^*)\right)\right)^\top \\ &= \left(\mathbf{V}\mathbf{u}_k^\dagger\right)\left(\mathbf{V}\mathbf{u}_k^\dagger\right)^\top, \end{aligned} \tag{76}$$

where the equality holds by (50) and the construction of $\hat{\mathbf{u}}_k^\dagger$. From (76), for any $k \in [K]$, we have the following inequality holds by Lemma 2:

$$4S_k\left(\mathbf{V}^\top\mathbf{V}\mathbf{Y}_{11}\mathbf{V}^\top\mathbf{V}\right) \succeq \mathbf{V}^\top\left(\mathbf{V}\mathbf{u}_k^\dagger\right)\left(\mathbf{V}\mathbf{u}_k^\dagger\right)^\top\mathbf{V},$$

which is equivalent to

$$4S_k\mathbf{Y}_{11} \succeq \mathbf{u}_k^\dagger\mathbf{u}_k^{\dagger\top} \tag{77}$$

because $\mathbf{V}^\top \mathbf{V} = \mathbf{I}_K$. By Schur complement, for any $k \in [K]$, (77) further becomes

$$\begin{bmatrix} S_k & \frac{1}{2}\mathbf{u}_k^{\dagger\top} \\ \frac{1}{2}\mathbf{u}_k^\dagger & \mathbf{Y}_{11} \end{bmatrix} \succeq 0,$$

which indicates that the constructed solution $(\mathbf{x}^\dagger, s^\dagger, \hat{\boldsymbol{\lambda}}^\dagger, \mathbf{q}^\dagger, \mathbf{Q}_\mathrm{r}^\dagger, \hat{\mathbf{u}}^\dagger, \mathbf{B}^\dagger)$ also satisfies constraints (22b) and thus it is a feasible solution of Problem (19).

Second, we show this feasible solution $(\mathbf{x}^\dagger, s^\dagger, \hat{\boldsymbol{\lambda}}^\dagger, \mathbf{q}^\dagger, \mathbf{Q}_\mathrm{r}^\dagger, \hat{\mathbf{u}}^\dagger, \mathbf{B}^\dagger)$ is an optimal solution of Problem (19). The objective value corresponding to this solution is

$$\begin{aligned} s^\dagger + \gamma_2 \mathbf{I}_{m_1} \bullet \mathbf{Q}_\mathrm{r}^\dagger + \sqrt{\gamma_1}\left\|\mathbf{q}^\dagger\right\|_2 &= s^* + \gamma_2 \mathbf{I}_{m_1} \bullet \mathbf{Y}_{11} + \sqrt{\gamma_1}\left\|\mathbf{q}'\right\|_2 \\ &= s^* + \gamma_2 \mathbf{I}_{m_1} \bullet (\mathbf{Y}_{11}\mathbf{V}^\top\mathbf{V}) + \sqrt{\gamma_1}\left\|\mathbf{q}'\right\|_2 \\ &= s^* + \gamma_2 \mathbf{I}_{m} \bullet (\mathbf{V}\mathbf{Y}_{11}\mathbf{V}^\top) + \sqrt{\gamma_1}\left\|\mathbf{q}'\right\|_2 \\ &= s^* + \gamma_2 \mathbf{I}_{m} \bullet \mathbf{Q}' + \sqrt{\gamma_1}\left\|\mathbf{q}'\right\|_2 \\ &= \Theta_\mathrm{M}(m), \end{aligned}$$

where the first equality holds by the construction of $(\mathbf{x}^\dagger, s^\dagger, \hat{\boldsymbol{\lambda}}^\dagger, \mathbf{q}^\dagger, \mathbf{Q}_\mathrm{r}^\dagger, \hat{\mathbf{u}}^\dagger, \mathbf{B}^\dagger)$, the second equality holds because $\mathbf{V}^\top\mathbf{V} = \mathbf{I}_K$, the third equality holds by the cyclic property of a matrix's trace, and the fourth equality holds by the definition of $\mathbf{Q}'$ in Theorem 2. Therefore, the solution $(\mathbf{x}^\dagger, s^\dagger, \hat{\boldsymbol{\lambda}}^\dagger, \mathbf{q}^\dagger, \mathbf{Q}_\mathrm{r}^\dagger, \hat{\mathbf{u}}^\dagger, \mathbf{B}^\dagger)$ is an optimal solution of Problem (19).

Finally, when $m_1 > K$ and $m_1 \le m$, we have $\Theta_\mathrm{U}(m_1) \ge \Theta_\mathrm{M}(m)$ by the conclusion (i) in Theorem 4 and $\Theta_\mathrm{U}(m_1) \le \Theta_\mathrm{U}(K) = \Theta_\mathrm{M}(m)$ by the conclusion (ii) in Theorem 4. It follows that $\Theta_\mathrm{U}(m_1) = \Theta_\mathrm{M}(m)$. □

B.11. Proof of Proposition 5

First, by Lemma 2, for any $\mathbf{B} \in \mathcal{B}_{m_1}$, we have

$$\mathbf{X} \preceq \mathbf{I}_m \Longrightarrow \mathbf{B}^\top\mathbf{X}\mathbf{B} \preceq \mathbf{B}^\top\mathbf{I}_m\mathbf{B} = \mathbf{I}_{m_1}.$$

Second, we perform eigenvalue decomposition on $\mathbf{X}$, i.e., $\mathbf{X} = \mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}^\top$, where $\mathbf{Q} \in \mathbb{R}^{m\times m}$ is a matrix with orthonormal column vectors and $\boldsymbol{\Lambda} \in \mathbb{R}^{m\times m}$ is a diagonal matrix. Without loss of generality, we assume that the diagonal elements of $\boldsymbol{\Lambda}$ are arranged in a nonincreasing order and let $\boldsymbol{\Lambda}_{m_1\times m_1}$ represent the upper-left submatrix of $\boldsymbol{\Lambda}$.

Now we let $\mathbf{B} = \mathbf{Q}_{m\times m_1}$, where $\mathbf{Q}_{m\times m_1}$ is the left submatrix of $\mathbf{Q}$. Then we have $\mathbf{B} \in \mathcal{B}_{m_1}$ and

$$\begin{aligned} \mathbf{B}^\top\mathbf{X}\mathbf{B} \preceq \mathbf{I}_{m_1} \Longrightarrow \mathbf{B}^\top\mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}^\top\mathbf{B} \preceq \mathbf{I}_{m_1} &\Longrightarrow \mathbf{Q}_{m\times m_1}^\top\mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}^\top\mathbf{Q}_{m\times m_1} \preceq \mathbf{I}_{m_1} \\ &\Longrightarrow [\mathbf{I}_{m_1}, \mathbf{0}_{m_1\times(m-m_1)}]\boldsymbol{\Lambda}[\mathbf{I}_{m_1}, \mathbf{0}_{m_1\times(m-m_1)}]^\top \preceq \mathbf{I}_{m_1} \\ &\Longrightarrow \boldsymbol{\Lambda}_{m_1\times m_1} \preceq \mathbf{I}_{m_1} \Longrightarrow \boldsymbol{\Lambda} \preceq \mathbf{I}_m \end{aligned}$$

$$\Longrightarrow \mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}^\top \preceq \mathbf{Q}\mathbf{I}_m\mathbf{Q}^\top \Longrightarrow \mathbf{X} \preceq \mathbf{I}_m$$

where the first deduction holds by the eigenvalue decomposition of $\mathbf{X}$, the second deduction holds by the construction of $\mathbf{B}$, the third deduction holds because all the column vectors in $\mathbf{Q}$ are orthonormal, the fourth deduction holds by the definition of $\boldsymbol{\Lambda}_{m_1\times m_1}$, the fifth deduction holds because $\text{rank}(\mathbf{X}) \leq m_1$, the sixth deduction holds by Lemma 2. Thus, if $\mathbf{B}^\top\mathbf{X}\mathbf{B} \preceq \mathbf{I}_{m_1}$ for any $\mathbf{B} \in \mathcal{B}_{m_1}$, then we have $\mathbf{X} \preceq \mathbf{I}_m$. The proof is complete. □

B.12. Proof of Theorem 6

By dualizing the inner maximization problem of Problem (23) and integrating it with the outer minimization operators, we first obtain the following formulation:

$$\min_{\substack{\mathbf{x},s,\mathbf{q},\mathbf{Q}'_\mathrm{r},\mathbf{Q}''_\mathrm{r}\\ \mathbf{B}_1,\mathbf{B}_2}} \quad s + \gamma_2\mathbf{I}_{m_1} \bullet \mathbf{Q}'_\mathrm{r} + \sqrt{\gamma_1}\,\|\mathbf{q}\|_2 \tag{78a}$$

$$\text{s.t.} \quad s \geq f\left(\mathbf{x}, \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_\mathrm{I} + \boldsymbol{\mu}\right) - \boldsymbol{\xi}_\mathrm{I}^\top\mathbf{B}_1\mathbf{Q}'_\mathrm{r}\mathbf{B}_1^\top\boldsymbol{\xi}_\mathrm{I} - \boldsymbol{\xi}_\mathrm{I}^\top\mathbf{B}_2\mathbf{Q}''_\mathrm{r}\mathbf{B}_2^\top\boldsymbol{\xi}_\mathrm{I} - \mathbf{q}^\top\boldsymbol{\xi}_\mathrm{I},\ \forall\boldsymbol{\xi}_\mathrm{I} \in \mathcal{S}_\mathrm{I}, \tag{78b}$$

$$\mathbf{Q}'_\mathrm{r} \succeq 0,\ \mathbf{Q}''_\mathrm{r} \succeq 0,\ \mathbf{x} \in \mathcal{X},\ \mathbf{Q}'_\mathrm{r} \in \mathbb{R}^{m_1\times m_1},\ \mathbf{Q}''_\mathrm{r} \in \mathbb{R}^{(K-m_1)\times(K-m_1)},\ \mathbf{q} \in \mathbb{R}^m, \tag{78c}$$

$$\mathbf{B}_1 \in \mathbb{R}^{m\times m_1}, \mathbf{B}_2 \in \mathbb{R}^{m\times(K-m_1)},\ [\mathbf{B}_1,\mathbf{B}_2]^\top[\mathbf{B}_1,\mathbf{B}_2] = \mathbf{I}_K. \tag{78d}$$

Next, we apply the strong duality theorem to constraints (78b). We define

$$g_k(\boldsymbol{\xi}_\mathrm{I}) = s + \boldsymbol{\xi}_\mathrm{I}^\top\mathbf{B}_1\mathbf{Q}'_\mathrm{r}\mathbf{B}_1^\top\boldsymbol{\xi}_\mathrm{I} + \boldsymbol{\xi}_\mathrm{I}^\top\mathbf{B}_2\mathbf{Q}''_\mathrm{r}\mathbf{B}_2^\top\boldsymbol{\xi}_\mathrm{I} + \mathbf{q}^\top\boldsymbol{\xi}_\mathrm{I} - y_k^0(\mathbf{x}) - y_k(\mathbf{x})^\top\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_\mathrm{I} + \boldsymbol{\mu}\right),\ \forall k \in [K].$$

As function $f(\mathbf{x},\boldsymbol{\xi})$ is piecewise linear convex, we can reformulate (78b) as

$$g_k(\boldsymbol{\xi}_\mathrm{I}) \geq 0,\ \forall\boldsymbol{\xi}_\mathrm{I} \in \mathcal{S}_\mathrm{I},\ \forall k \in [K],$$

which is equivalent to

$$\min_{\mathbf{A}\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_\mathrm{I}+\boldsymbol{\mu}\right)\leq\mathbf{b},\ \boldsymbol{\xi}_\mathrm{I}\in\mathbb{R}^m} g_k(\boldsymbol{\xi}_\mathrm{I}) \geq 0,\ \forall k \in [K]. \tag{79}$$

For any $k \in [K]$, the Lagrangian dual problem of $\min_{\mathbf{A}(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_\mathrm{I}+\boldsymbol{\mu})\leq\mathbf{b},\ \boldsymbol{\xi}_\mathrm{I}\in\mathbb{R}^m} g_k(\boldsymbol{\xi}_\mathrm{I})$ is

$$\max_{\boldsymbol{\lambda}_k\geq 0}\ \min_{\boldsymbol{\xi}_\mathrm{I}\in\mathbb{R}^m} g_k(\boldsymbol{\xi}_\mathrm{I}) + \boldsymbol{\lambda}_k^\top\left(\mathbf{A}\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_\mathrm{I} + \boldsymbol{\mu}\right) - \mathbf{b}\right),$$

where $\boldsymbol{\lambda}_k \in \mathbb{R}^l$. Because there exists an interior point for the primal problem, the strong duality holds. Thus, constraints (79) are equivalent to

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

which are further equivalent to

$$\exists \boldsymbol{\lambda}_k \geq 0: s + \boldsymbol{\xi}_{\mathrm{I}}^\top \mathbf{B}_1 \mathbf{Q}_{\mathrm{r}}' \mathbf{B}_1^\top \boldsymbol{\xi}_{\mathrm{I}} + \boldsymbol{\xi}_{\mathrm{I}}^\top \mathbf{B}_2 \mathbf{Q}_{\mathrm{r}}'' \mathbf{B}_2^\top \boldsymbol{\xi}_{\mathrm{I}} + \mathbf{q}^\top \boldsymbol{\xi}_{\mathrm{I}} - y_k^0(\mathbf{x}) - y_k(\mathbf{x})^\top \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}} \boldsymbol{\xi}_{\mathrm{I}} + \boldsymbol{\mu} \right) + \boldsymbol{\lambda}_k^\top \left(\mathbf{A} \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}} \boldsymbol{\xi}_{\mathrm{I}} + \boldsymbol{\mu} \right) - \mathbf{b} \right) \geq 0, \ \forall \boldsymbol{\xi}_{\mathrm{I}} \in \mathbb{R}^m, \ \forall k \in [K]. \tag{80}$$

Note that $\mathbf{B}^\top \mathbf{B} = \mathbf{I}_{m_1}$; that is, all the column vectors of $\mathbf{B}$ are orthogonal. We can then extend $\mathbf{B}$ to $[\mathbf{B}, \bar{\mathbf{B}}] \in \mathbb{R}^{m \times m}$ with $\bar{\mathbf{B}} \in \mathbb{R}^{m \times (m-K)}$ such that all the column vectors of $[\mathbf{B}, \bar{\mathbf{B}}]$ span the space of $\mathbb{R}^m$. Thus, we can always find $\boldsymbol{\xi}_1 \in \mathbb{R}^{m_1}$, $\boldsymbol{\xi}_2 \in \mathbb{R}^{K-m_1}$, and $\boldsymbol{\xi}_3 \in \mathbb{R}^{m-K}$ such that

$$\boldsymbol{\xi}_{\mathrm{I}} = \mathbf{B}_1 \boldsymbol{\xi}_1 + \mathbf{B}_2 \boldsymbol{\xi}_2 + \bar{\mathbf{B}} \boldsymbol{\xi}_3.$$

It follows that constraints (80) become

$$\begin{aligned} \exists \boldsymbol{\lambda}_k \geq 0: s &+ \boldsymbol{\xi}_1^\top \mathbf{Q}_{\mathrm{r}}' \boldsymbol{\xi}_1 + \boldsymbol{\xi}_2^\top \mathbf{Q}_{\mathrm{r}}'' \boldsymbol{\xi}_2 + \mathbf{q}^\top \left(\mathbf{B}_1 \boldsymbol{\xi}_1 + \mathbf{B}_2 \boldsymbol{\xi}_2 + \bar{\mathbf{B}} \boldsymbol{\xi}_3 \right) - y_k^0(\mathbf{x}) \\ &- y_k(\mathbf{x})^\top \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}} \left(\mathbf{B}_1 \boldsymbol{\xi}_1 + \mathbf{B}_2 \boldsymbol{\xi}_2 + \bar{\mathbf{B}} \boldsymbol{\xi}_3 \right) + \boldsymbol{\mu} \right) \\ &+ \boldsymbol{\lambda}_k^\top \left(\mathbf{A} \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}} \left(\mathbf{B}_1 \boldsymbol{\xi}_1 + \mathbf{B}_2 \boldsymbol{\xi}_2 + \bar{\mathbf{B}} \boldsymbol{\xi}_3 \right) + \boldsymbol{\mu} \right) - \mathbf{b} \right) \geq 0, \\ &\forall \boldsymbol{\xi}_1 \in \mathbb{R}^{m_1}, \boldsymbol{\xi}_2 \in \mathbb{R}^{K-m_1}, \boldsymbol{\xi}_3 \in \mathbb{R}^{m-K}, \ \forall k \in [K]. \end{aligned} \tag{81}$$

We further define

$$\mathbf{Z}_k = \begin{bmatrix} s - y_k^0(\mathbf{x}) - \boldsymbol{\lambda}_k^\top \mathbf{b} - y_k(\mathbf{x})^\top \boldsymbol{\mu} + \boldsymbol{\lambda}_k^\top \mathbf{A} \boldsymbol{\mu} & \frac{1}{2}(\mathbf{h}_k')^\top & \frac{1}{2}(\mathbf{h}_k'')^\top \\ \frac{1}{2}\mathbf{h}_k' & \mathbf{Q}_{\mathrm{r}}' & \mathbf{0}_{m_1 \times (K-m_1)} \\ \frac{1}{2}\mathbf{h}_k'' & \mathbf{0}_{(K-m_1) \times m_1} & \mathbf{Q}_{\mathrm{r}}'' \end{bmatrix}, \ \forall k \in [K],$$

where $\mathbf{h}_k' = \mathbf{B}_1^\top \mathbf{q} + (\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}} \mathbf{B}_1)^\top (\mathbf{A}^\top \boldsymbol{\lambda}_k - y_k(\mathbf{x}))$ and $\mathbf{h}_k'' = \mathbf{B}_2^\top \mathbf{q} + (\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}} \mathbf{B}_2)^\top (\mathbf{A}^\top \boldsymbol{\lambda}_k - y_k(\mathbf{x}))$ for any $k \in [K]$. It follows that

$$\begin{aligned} (81) \iff & \exists \boldsymbol{\lambda}_k \geq 0: \left(1, \boldsymbol{\xi}_1^\top, \boldsymbol{\xi}_2^\top\right) \mathbf{Z}_k \left(1, \boldsymbol{\xi}_1^\top, \boldsymbol{\xi}_2^\top\right)^\top + \boldsymbol{\xi}_3^\top \left(\bar{\mathbf{B}}^\top \mathbf{q} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}} \bar{\mathbf{B}} \right)^\top \left(\mathbf{A}^\top \boldsymbol{\lambda}_k - y_k(\mathbf{x}) \right) \right) \geq 0, \\ & \qquad \forall \boldsymbol{\xi}_1 \in \mathbb{R}^{m_1}, \boldsymbol{\xi}_2 \in \mathbb{R}^{K-m_1}, \boldsymbol{\xi}_3 \in \mathbb{R}^{m-K}, \ \forall k \in [K]. \\ \iff & \exists \boldsymbol{\lambda}_k \geq 0: \left(1, \boldsymbol{\xi}_1^\top, \boldsymbol{\xi}_2^\top\right) \mathbf{Z}_k \left(1, \boldsymbol{\xi}_1^\top, \boldsymbol{\xi}_2^\top\right)^\top \geq 0, \ \forall \boldsymbol{\xi}_1 \in \mathbb{R}^{m_1}, \boldsymbol{\xi}_2 \in \mathbb{R}^{K-m_1}, \ \forall k \in [K]; \qquad (82) \\ & \qquad \bar{\mathbf{B}}^\top \mathbf{q} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}} \bar{\mathbf{B}} \right)^\top \left(\mathbf{A}^\top \boldsymbol{\lambda}_k - y_k(\mathbf{x}) \right) = 0, \ \forall k \in [K]. \\ \iff & \exists \boldsymbol{\lambda}_k \geq 0: \ \mathbf{Z}_k \succeq 0, \ \bar{\mathbf{B}}^\top \mathbf{q} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}} \bar{\mathbf{B}} \right)^\top \left(\mathbf{A}^\top \boldsymbol{\lambda}_k - y_k(\mathbf{x}) \right) = 0, \ \forall k \in [K]. \\ \iff & \exists \boldsymbol{\lambda}_k \geq 0: \ \mathbf{Z}_k \succeq 0, \ \bar{\mathbf{B}}^\top \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) = 0, \ \forall k \in [K]. \qquad (83) \\ \iff & \exists \boldsymbol{\lambda}_k \geq 0, \mathbf{u}_k' \in \mathbb{R}^{m_1}, \mathbf{u}_k'' \in \mathbb{R}^{K-m_1}: \\ & \qquad \mathbf{Z}_k \succeq 0, \ \mathbf{q} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}} \right)^\top \left(\mathbf{A}^\top \boldsymbol{\lambda}_k - y_k(\mathbf{x}) \right) = \mathbf{B}_1 \mathbf{u}_k' + \mathbf{B}_2 \mathbf{u}_k'', \ \forall k \in [K]. \qquad (84) \end{aligned}$$

The first equivalence holds due to the definition of $\mathbf{Z}_k$. For the third equivalence, clearly $\Longleftarrow$ follows from the definition of a PSD matrix. To prove $\Longrightarrow$, we consider two possible cases for any

$(\eta_0 \in \mathbb{R}, \boldsymbol{\eta}_1^\top \in \mathbb{R}^{m_1}, \boldsymbol{\eta}_2^\top \in \mathbb{R}^{K-m_1})^\top \in \mathbb{R}^{K+1}$: (i) if $\eta_0 = 0$, then $(\eta_0, \boldsymbol{\eta}_1^\top, \boldsymbol{\eta}_2^\top)\mathbf{Z}_k(\eta_0, \boldsymbol{\eta}_1^\top, \boldsymbol{\eta}_2^\top)^\top = \boldsymbol{\eta}_1^\top \mathbf{Q}_\mathrm{r}' \boldsymbol{\eta}_1 + \boldsymbol{\eta}_2^\top \mathbf{Q}_\mathrm{r}'' \boldsymbol{\eta}_2 \geq 0$ because $\mathbf{Q}_\mathrm{r}'$ and $\mathbf{Q}_\mathrm{r}''$ are PSD; (ii) if $\eta_0 \neq 0$, then we have $(\eta_0, \boldsymbol{\eta}_1^\top, \boldsymbol{\eta}_2^\top)\mathbf{Z}_k(\eta_0, \boldsymbol{\eta}_1^\top, \boldsymbol{\eta}_2^\top)^\top = \eta_0^2(1, \frac{\boldsymbol{\eta}_1^\top}{\eta_0}, \frac{\boldsymbol{\eta}_2^\top}{\eta_0})\mathbf{Z}_k(1, \frac{\boldsymbol{\eta}_1^\top}{\eta_0}, \frac{\boldsymbol{\eta}_2^\top}{\eta_0})^\top \geq 0$ according to (82). Therefore, $\Longrightarrow$ holds. For the fifth equivalence, (83) shows that $\mathbf{q} + (\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}})^\top(\mathbf{A}^\top \boldsymbol{\lambda}_k - y_k(\mathbf{x}))$ is in the null space of $\bar{\mathbf{B}}$ and thus cannot be represented by basis vectors in the space of $\bar{\mathbf{B}}$. Because $[\mathbf{B}, \bar{\mathbf{B}}]$ span the space of $\mathbb{R}^m$, we have $\mathbf{q} + (\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}})^\top(\mathbf{A}^\top \boldsymbol{\lambda}_k - y_k(\mathbf{x}))$ should be in the space of $\mathbf{B}$. That is, there exists $\mathbf{u}_k' \in \mathbb{R}^{m_1}$ and $\mathbf{u}_k'' \in \mathbb{R}^{K-m_1}$ such that $\mathbf{q} + (\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}})^\top(\mathbf{A}^\top \boldsymbol{\lambda}_k - y_k(\mathbf{x})) = \mathbf{B}_1\mathbf{u}_k' + \mathbf{B}_2\mathbf{u}_k''$ for any $k \in [K]$. Meanwhile, because $\mathbf{B}^\top\mathbf{B} = \mathbf{I}_K$, we have

$$\mathbf{h}_k' = \mathbf{B}_1^\top \mathbf{q} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}_1\right)^\top \left(\mathbf{A}^\top \boldsymbol{\lambda}_k - y_k(\mathbf{x})\right) = \mathbf{B}_1^\top \mathbf{B}_1 \mathbf{u}_k' = \mathbf{u}_k',\ \forall k \in [K],$$
$$\mathbf{h}_k'' = \mathbf{B}_2^\top \mathbf{q} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}_2\right)^\top \left(\mathbf{A}^\top \boldsymbol{\lambda}_k - y_k(\mathbf{x})\right) = \mathbf{B}_2^\top \mathbf{B}_2 \mathbf{u}_k'' = \mathbf{u}_k'',\ \forall k \in [K].$$

By replacing constraints (78b) with (84), we obtain the following problem:

$$\min_{\substack{\mathbf{x}, s, \hat{\boldsymbol{\lambda}}, \mathbf{q}, \\ \mathbf{Q}_\mathrm{r}', \mathbf{Q}_\mathrm{r}'', \hat{\mathbf{u}}', \hat{\mathbf{u}}'', \\ \mathbf{B}_1, \mathbf{B}_2}} \quad s + \gamma_2 \mathbf{I}_{m_1} \bullet \mathbf{Q}_\mathrm{r}' + \sqrt{\gamma_1}\, \|\mathbf{q}\|_2 \tag{85a}$$

$$\text{s.t.} \quad \begin{bmatrix} s - y_k^0(\mathbf{x}) - \boldsymbol{\lambda}_k^\top \mathbf{b} - y_k(\mathbf{x})^\top \boldsymbol{\mu} + \boldsymbol{\lambda}_k^\top \mathbf{A}\boldsymbol{\mu} & \frac{1}{2}(\mathbf{u}_k')^\top & \frac{1}{2}(\mathbf{u}_k'')^\top \\ \frac{1}{2}\mathbf{u}_k' & \mathbf{Q}_\mathrm{r}' & \mathbf{0}_{m_1 \times (K-m_1)} \\ \frac{1}{2}\mathbf{u}_k'' & \mathbf{0}_{(K-m_1)\times m_1} & \mathbf{Q}_\mathrm{r}'' \end{bmatrix} \succeq 0,\ \forall k \in [K], \tag{85b}$$

$$\mathbf{q} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top \left(\mathbf{A}^\top \boldsymbol{\lambda}_k - y_k(\mathbf{x})\right) = \mathbf{B}_1\mathbf{u}_k' + \mathbf{B}_2\mathbf{u}_k'',\ \forall k \in [K], \tag{85c}$$

$$\mathbf{x} \in \mathcal{X},\ \mathbf{q} \in \mathbb{R}^m,\ \mathbf{Q}_\mathrm{r}' \in \mathbb{R}^{m_1 \times m_1},\ \mathbf{Q}_\mathrm{r}'' \in \mathbb{R}^{(K-m_1)\times(K-m_1)}, \tag{85d}$$

$$\mathbf{B}_1 \in \mathbb{R}^{m \times m_1}, \mathbf{B}_2 \in \mathbb{R}^{m \times (K-m_1)},\ [\mathbf{B}_1, \mathbf{B}_2]^\top [\mathbf{B}_1, \mathbf{B}_2] = \mathbf{I}_K, \tag{85e}$$

$$\hat{\boldsymbol{\lambda}} = \{\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_K\},\ \boldsymbol{\lambda}_k \in \mathbb{R}_+^l,\ \forall k \in [K], \tag{85f}$$

$$\hat{\mathbf{u}}' = \{\mathbf{u}_1', \ldots, \mathbf{u}_K'\},\ \mathbf{u}_k' \in \mathbb{R}^{m_1},\ \forall k \in [K], \tag{85g}$$

$$\hat{\mathbf{u}}'' = \{\mathbf{u}_1'', \ldots, \mathbf{u}_K''\},\ \mathbf{u}_k'' \in \mathbb{R}^{K-m_1},\ \forall k \in [K]. \tag{85h}$$

Note that the value of $\mathbf{Q}_\mathrm{r}''$ does not contribute to the objective function (85a). We can then let M be an arbitrarily large positive number and $\mathbf{Q}_\mathrm{r}'' = M\mathbf{I}_{(K-m_1)\times(K-m_1)}$ be an optimal solution, by which constraints (85b) become

$$\begin{bmatrix} s - y_k^0(\mathbf{x}) - \boldsymbol{\lambda}_k^\top \mathbf{b} - y_k(\mathbf{x})^\top \boldsymbol{\mu} + \boldsymbol{\lambda}_k^\top \mathbf{A}\boldsymbol{\mu} & \frac{1}{2}(\mathbf{u}_k')^\top \\ \frac{1}{2}\mathbf{u}_k' & \mathbf{Q}_\mathrm{r}' \end{bmatrix} \succeq 0,\ \forall k \in [K]. \tag{86}$$

By replacing (85b) with (86), we obtain the formulation of Problem (24).

Based on the formulation of Problem (24), now we show that the three conclusions hold. Note that for any $\mathbf{B} \in \mathcal{B}_K = \{\mathbf{B} \in \mathbb{R}^{m \times K} \mid \mathbf{B}^\top\mathbf{B} = \mathbf{I}_K\}$, the optimal value of Problem (19), i.e., $\Theta_\mathrm{U}(K)$, reaches the optimal value of the original Problem (4), i.e., $\Theta_\mathrm{M}(m)$. We would like to show that by relaxing the constraints in Problem (19), we can obtain the exact formulation of Problem (24), thereby the three conclusions hold.

First, we rewrite constraints (22b)–(22c) in Problem (19) with $m_1 = K$ by dividing $\mathbf{B}$ into $[\mathbf{B}_1, \mathbf{B}_2]$ and $\mathbf{u}_k$ into $((\mathbf{u}_k')^\top, (\mathbf{u}_k'')^\top)^\top$. Thus, we obtain the following formulation:

$$\min_{\substack{\mathbf{x}, s, \hat{\boldsymbol{\lambda}}, \\ \mathbf{q}, \mathbf{Q}_\mathrm{r}, \hat{\mathbf{u}}', \hat{\mathbf{u}}'', \\ \mathbf{B}_1, \mathbf{B}_2}} \quad s + \gamma_2 \mathbf{I}_K \bullet \mathbf{Q}_\mathrm{r} + \sqrt{\gamma_1} \|\mathbf{q}\|_2 \tag{87a}$$

$$\text{s.t.} \quad \begin{bmatrix} s - y_k^0(\mathbf{x}) - \boldsymbol{\lambda}_k^\top \mathbf{b} - y_k(\mathbf{x})^\top \boldsymbol{\mu} + \boldsymbol{\lambda}_k^\top \mathbf{A} \boldsymbol{\mu} & \frac{1}{2}\left((\mathbf{u}_k')^\top, (\mathbf{u}_k'')^\top\right) \\ \frac{1}{2}\left((\mathbf{u}_k')^\top, (\mathbf{u}_k'')^\top\right)^\top & \mathbf{Q}_\mathrm{r} \end{bmatrix} \succeq 0, \ \forall k \in [K], \tag{87b}$$

$$\mathbf{q} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top \left(\mathbf{A}^\top \boldsymbol{\lambda}_k - y_k(\mathbf{x})\right) = \mathbf{B}_1 \mathbf{u}_k' + \mathbf{B}_2 \mathbf{u}_k'', \ \forall k \in [K], \tag{87c}$$

$$\mathbf{x} \in \mathcal{X}, \ [\mathbf{B}_1, \mathbf{B}_2]^\top [\mathbf{B}_1, \mathbf{B}_2] = \mathbf{I}_K, \tag{87d}$$

$$\mathbf{q} \in \mathbb{R}^m, \ \mathbf{Q}_\mathrm{r} \in \mathbb{R}^{K \times K}, \ \mathbf{B}_1 \in \mathbb{R}^{m \times m_1}, \mathbf{B}_2 \in \mathbb{R}^{m \times (K - m_1)}, \tag{87e}$$

$$\hat{\boldsymbol{\lambda}} = \{\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_K\}, \ \boldsymbol{\lambda}_k \in \mathbb{R}_+^l, \ \forall k \in [K], \tag{87f}$$

$$\hat{\mathbf{u}}' = \{\mathbf{u}_1', \ldots, \mathbf{u}_K'\}, \ \mathbf{u}_k' \in \mathbb{R}^{m_1}, \ \forall k \in [K], \tag{87g}$$

$$\hat{\mathbf{u}}'' = \{\mathbf{u}_1'', \ldots, \mathbf{u}_K''\}, \ \mathbf{u}_k'' \in \mathbb{R}^{K - m_1}, \ \forall k \in [K]. \tag{87h}$$

Second, we relax constraints (87b) into

$$\begin{bmatrix} s - y_k^0(\mathbf{x}) - \boldsymbol{\lambda}_k^\top \mathbf{b} - y_k(\mathbf{x})^\top \boldsymbol{\mu} + \boldsymbol{\lambda}_k^\top \mathbf{A} \boldsymbol{\mu} & \frac{1}{2}(\mathbf{u}_k')^\top \\ \frac{1}{2}\mathbf{u}_k' & \mathbf{Q}_\mathrm{r}' \end{bmatrix} \succeq 0, \ \forall k \in [K], \tag{88}$$

where $\mathbf{Q}_\mathrm{r}' \in \mathbb{R}^{m_1 \times m_1}$ is the upper-left submatrix of $\mathbf{Q}_\mathrm{r}$. Note that if we use (88) to replace (87b), we obtain a relaxation and accordingly lower bound for Problem (87). In addition, we further reduce the optimal value of the relaxation by replacing $\mathbf{Q}_\mathrm{r}$ in the objective function (87a) with $\mathbf{Q}_\mathrm{r}'$. That is, we obtain a lower bound for the optimal value of Problem (19) with $m_1 = K$ (i.e., Problem (4)). After these two steps of relaxations, we obtain the exact formulation of Problem (24). Thus, we can conclude that Problem (24) is a relaxation of Problem (19) with $m_1 = K$. Therefore, by the conclusion in Theorem 5, we have

$$\Theta_{\mathrm{L2}}(m_1) \leq \Theta_{\mathrm{U}}(K) = \Theta_{\mathrm{M}}(m).$$

That is, the conclusion (i) holds.

For the conclusion (ii): For any $0 \leq m_1 < m_2 \leq K$, we can follow the above two steps of relaxations to relax Problem (87) to the problem with the optimal value $\Theta_{\mathrm{L2}}(m_2)$, and based on this relaxed problem, we can further relax it to the problem with the optimal value $\Theta_{\mathrm{L2}}(m_1)$. Because all these problems are minimization problems, we have $\Theta_{\mathrm{L2}}(m_1) \leq \Theta_{\mathrm{L2}}(m_2)$.

For the conclusion (iii): When $m_1 = K$, Problem (24) becomes Problem (19) with $m_1 = K$. Thus, by the conclusion in Theorem 5, we have

$$\Theta_{\mathrm{L2}}(K) = \Theta_{\mathrm{U}}(K) = \Theta_{\mathrm{M}}(m). \quad \square$$

B.13. Proof of Proposition 7

Because $\mathbf{B}^\top\mathbf{B} = \mathbf{I}_{m_1}$, we have $\mathbf{B}^\top\mathbf{B} \preceq \mathbf{I}_{m_1}$, which implies $\mathbf{B}\mathbf{B}^\top \preceq \mathbf{I}_m$ by Lemma 1. It follows that $\mathbf{r}^\top\mathbf{B}\mathbf{B}^\top\mathbf{r} \leq \mathbf{r}^\top\mathbf{r}$. Meanwhile, we have

$$\mathbf{r}^\top\mathbf{B}^*\mathbf{B}^{*\top}\mathbf{r} = \begin{bmatrix} \frac{\mathbf{r}^\top\mathbf{r}}{\|\mathbf{r}\|_2} & \mathbf{0}_{1\times(m_1-1)} \end{bmatrix} \begin{bmatrix} \frac{\mathbf{r}^\top\mathbf{r}}{\|\mathbf{r}\|_2} \\ \mathbf{0}_{(m_1-1)\times 1} \end{bmatrix} = \mathbf{r}^\top\mathbf{r},$$

indicating that $\mathbf{B}^* = [\mathbf{r}/\|\mathbf{r}\|_2, \mathbf{0}_{m\times(m_1-1)}]$ is an optimal solution of Problem (41). $\square$

Appendix C: Instance-level Performance

Table C2 Instance-level Performance of Mosek, Low-rank, and ODR approaches on the Newsvendor Problem

Size (m)	Inst No.	Mosek	Time (secs)	Low-rank	Time (secs)	Gap1 (%)	ODR-LB	Time (secs)	Gap1 (%)	ODR-UB	Time (secs)	Gap2 (%)	Interval Gap (%)	ODR-RLB	Gap1 (%)	Interval Gap (%)
	1	-1286.49	12.32	-1318.79	0.26	2.51	-1286.55	0.92	0.00	-1272.91	2.00	1.06	1.07	-1286.67	0.01	1.08
	2	-1273.87	11.34	-1306.17	0.26	2.54	-1273.95	0.69	0.01	-1247.32	2.00	2.08	2.13	-1274.26	0.03	2.16
100	3	-1332.03	13.36	-1364.44	0.24	2.43	-1337.56	0.52	0.42	-1305.45	1.97	2.00	2.46	-1332.36	0.02	2.06
	4	-1165.51	14.41	-1198.11	0.24	2.80	-1165.77	0.86	0.02	-1147.27	1.97	1.57	1.61	-1165.85	0.03	1.62
	5	-1404.18	13.67	-1436.49	0.28	2.30	-1404.44	0.87	0.02	-1380.13	1.79	1.71	1.76	-1404.71	0.04	1.78
	1	-5024.99	342.00	-5113.92	0.77	1.77	-5025.01	0.79	0.00	-4942.44	2.59	1.64	1.67	-5025.71	0.01	1.68
	2	-4905.57	432.57	-4994.05	0.73	1.80	-4905.69	0.77	0.00	-4826.53	2.60	1.61	1.64	-4906.38	0.02	1.65
200	3	-5224.18	368.04	-5312.91	0.74	1.70	-5224.19	0.77	0.00	-5143.56	2.60	1.54	1.57	-5224.81	0.01	1.58
	4	-4544.90	338.68	-4633.02	0.86	1.94	-4544.92	0.78	0.00	-4420.02	2.61	2.75	2.83	-4549.09	0.09	2.92
	5	-5082.08	336.39	-5170.61	0.93	1.74	-5082.11	0.78	0.00	-5008.97	2.62	1.44	1.46	-5082.67	0.01	1.47
	1	-	-	-21319.00	4.75	-	-21073.11	0.84	-	-20609.54	4.63	-	2.25	-21085.65	-	2.31
	2	-	-	-19704.00	4.18	-	-19456.89	0.83	-	-19154.47	4.11	-	1.58	-19458.73	-	1.59
400	3	-	-	-20768.00	9.18	-	-20522.01	0.83	-	-20329.67	4.15	-	0.95	-20523.16	-	0.95
	4	-	-	-20233.00	4.28	-	-19987.14	0.83	-	-19720.16	4.61	-	1.35	-19988.39	-	1.36
	5	-	-	-21098.00	4.92	-	-20851.12	0.83	-	-20625.09	4.17	-	1.10	-20852.53	-	1.10
	1	-	-	-79859.00	20.42	-	-79169.65	0.87	-	-78077.61	9.75	-	1.40	-79175.14	-	1.41
	2	-	-	-77257.00	27.47	-	-76565.62	0.85	-	-74916.99	10.33	-	2.20	-76599.05	-	2.25
800	3	-	-	-81070.00	35.62	-	-80377.76	0.84	-	-79315.51	9.53	-	1.34	-80382.74	-	1.35
	4	-	-	-81457.00	76.50	-	-80764.52	0.84	-	-79962.65	9.58	-	1.00	-80768.75	-	1.01
	5	-	-	-81496.00	76.69	-	-80805.26	0.85	-	-79792.38	9.55	-	1.27	-80813.71	-	1.28
	1	-	-	-183630.00	109.14	-	-182370.00	1.14	-	-180050.00	19.91	-	1.29	-182370.00	-	1.29
	2	-	-	-173670.00	117.37	-	-172400.00	1.12	-	-169680.00	19.78	-	1.60	-172410.00	-	1.61
1200	3	-	-	-182520.00	140.65	-	-181250.00	1.13	-	-177660.00	22.55	-	2.02	-181270.00	-	2.03
	4	-	-	-178900.00	74.59	-	-177640.00	1.13	-	-175060.00	20.25	-	1.47	-177640.00	-	1.47
	5	-	-	-185620.00	109.90	-	-184350.00	1.13	-	-181780.00	21.68	-	1.41	-184360.00	-	1.42
	1	-	-	-315650.00	400.93	-	-313700.00	2.01	-	-305200.00	41.39	-	2.79	-313730.00	-	2.79
	2	-	-	-309790.00	372.20	-	-307850.00	2.00	-	-302760.00	37.50	-	1.68	-308420.00	-	1.87
1600	3	-	-	-324440.00	149.56	-	-322490.00	2.01	-	-317910.00	38.19	-	1.44	-322510.00	-	1.45
	4	-	-	-317670.00	252.29	-	-315730.00	2.00	-	-311360.00	37.82	-	1.40	-315750.00	-	1.41
	5	-	-	-327380.00	370.02	-	-325430.00	2.03	-	-321070.00	36.89	-	1.36	-325450.00	-	1.36
	1	-	-	-498890.00	375.27	-	-496170.00	2.56	-	-486710.00	65.33	-	1.94	-496420.00	-	2.00
	2	-	-	-503680.00	456.47	-	-500960.00	2.53	-	-493000.00	59.05	-	1.61	-500980.00	-	1.62
2000	3	-	-	-474830.00	652.28	-	-472110.00	2.50	-	-464030.00	59.88	-	1.74	-472230.00	-	1.77
	4	-	-	-500770.00	1958.60	-	-498060.00	2.54	-	-483530.00	41.23	-	3.00	-498090.00	-	3.01
	5	-	-	-499180.00	685.50	-	-496460.00	2.54	-	-489140.00	57.90	-	1.50	-496490.00	-	1.50

Table C3 Instance-level Performance of Mosek, Low-rank, and ODR approaches on the CVaR Problem

Size (m)	Inst No.	Mosek	Time (secs)	Low-rank	Time (secs)	Gap1 (%)	ODR-LB	Time (secs)	Gap1 (%)	ODR-UB	Time (secs)	Gap2 (%)	Interval Gap (%)	ODR-RLB	Gap1 (%)	Interval Gap (%)
100	1	4.53	17.18	4.41	2.43	2.68	4.53	1.11	0.03	4.56	2.53	0.51	0.54	4.23	6.75	7.22
	2	4.80	17.12	4.63	3.83	3.51	4.79	0.88	0.03	4.84	2.57	0.92	0.94	4.38	8.74	9.57
	3	4.95	14.86	4.85	2.19	1.89	4.94	1.09	0.04	5.03	1.27	1.66	1.67	4.41	10.84	12.30
	4	4.43	14.71	4.30	3.05	2.91	4.43	0.88	0.05	4.47	1.92	0.86	0.91	4.04	8.77	9.56
	5	3.55	16.14	3.45	2.43	3.05	3.55	0.88	0.00	3.57	3.18	0.37	0.37	3.33	6.35	6.69
200	1	3.48	435.49	3.39	6.25	2.74	3.48	1.54	0.00	3.49	1.87	0.35	0.35	3.36	3.58	3.92
	2	3.08	438.93	2.96	8.91	3.81	3.08	1.52	0.01	3.10	5.08	0.66	0.66	2.99	2.84	3.47
	3	2.96	472.15	2.80	7.03	5.2	2.96	1.18	0.01	2.97	3.75	0.32	0.32	2.76	6.86	7.16
	4	2.61	469.65	2.56	5.83	1.92	2.61	1.15	0.12	2.64	5.64	1.29	1.40	2.30	11.93	13.06
	5	3.29	440.47	3.19	6.55	3.04	3.29	1.17	0.01	3.30	7.48	0.35	0.36	3.03	8.03	8.35
400	1	-	-	1.84	21.65	-	1.87	5.09	-	1.89	7.13	-	0.58	1.77	-	6.01
	2	-	-	1.37	19.35	-	1.42	5.06	-	1.45	7.85	-	2.36	1.33	-	8.58
	3	-	-	1.63	18.83	-	1.68	4.16	-	1.75	11.30	-	3.78	1.48	-	15.14
	4	-	-	1.31	26.35	-	1.35	5.79	-	1.37	10.19	-	1.50	1.26	-	8.00
	5	-	-	1.72	17.72	-	1.77	5.03	-	1.78	10.72	-	0.91	1.62	-	9.08
800	1	-	-	0.38	68.43	-	0.40	13.27	-	0.41	30.50	-	2.55	0.34	-	16.17
	2	-	-	0.43	84.54	-	0.44	17.25	-	0.44	53.51	-	0.41	0.40	-	8.89
	3	-	-	0.65	72.48	-	0.67	11.32	-	0.68	28.40	-	1.37	0.62	-	8.75
	4	-	-	0.32	84.69	-	0.33	20.70	-	0.34	35.89	-	3.20	0.28	-	18.17
	5	-	-	0.54	69.04	-	0.57	11.70	-	0.57	51.05	-	0.70	0.53	-	6.81
1200	1	-	-	-0.22	159.23	-	-0.21	26.73	-	-0.20	66.29	-	6.16	-0.22	-	9.70
	2	-	-	0.05	184.83	-	0.05	29.53	-	0.05	96.83	-	4.99	0.05	-	8.97
	3	-	-	-0.24	172.05	-	-0.23	27.23	-	-0.23	74.17	-	2.57	-0.25	-	8.32
	4	-	-	-0.05	193.84	-	-0.04	30.20	-	-0.04	96.19	-	3.49	-0.05	-	10.39
	5	-	-	-0.23	147.44	-	-0.22	30.58	-	-0.22	89.40	-	1.58	-0.24	-	7.51
1600	1	-	-	-0.49	634.88	-	-0.47	52.30	-	-0.47	108.48	-	1.37	-0.50	-	6.46
	2	-	-	-0.63	724.57	-	-0.60	47.77	-	-0.59	112.56	-	1.49	-0.66	-	11.10
	3	-	-	-0.65	548.77	-	-0.63	39.49	-	-0.61	107.97	-	2.31	-0.68	-	10.98
	4	-	-	-0.54	628.65	-	-0.53	56.91	-	-0.52	116.45	-	2.08	-0.56	-	8.11
	5	-	-	-0.67	638.91	-	-0.65	44.68	-	-0.64	124.21	-	1.66	-0.69	-	6.91
2000	1	-	-	-0.95	1654.35	-	-0.92	69.48	-	-0.91	168.84	-	0.82	-0.98	-	8.04
	2	-	-	-0.87	1259.47	-	-0.83	74.56	-	-0.82	176.96	-	1.42	-0.89	-	8.27
	3	-	-	-0.88	1622.80	-	-0.86	72.74	-	-0.85	173.70	-	1.81	-0.92	-	8.67
	4	-	-	-0.86	1558.60	-	-0.83	74.70	-	-0.82	171.74	-	2.06	-0.89	-	9.56
	5	-	-	-0.91	2055.00	-	-0.89	64.40	-	-0.88	171.40	-	1.47	-0.94	-	7.35

Table C4 Instance-level Performance of PCA Approximation on the Newsvendor Problem (Part 1)

$\frac{m_1}{m}$(%)				100%							80%						
Size (m)	Inst No.	Mosek	Time (secs)	PCA -LB	Time (secs)	Gap1 (%)	PCA -UB	Time (secs)	Gap2 (%)	Interval Gap (%)	PCA -LB	Time (secs)	Gap1 (%)	PCA -UB	Time (secs)	Gap2 (%)	Interval Gap (%)
	1	-1286.50	12.32	-1286.50	12.25	0.00	-1286.50	12.26	0.00	0.00	-1294.60	4.81	0.63	-1095.30	8.32	14.86	18.20
	2	-1273.90	11.34	-1273.90	11.34	0.00	-1273.90	11.31	0.00	0.00	-1280.30	6.26	0.50	-1125.30	9.26	11.66	13.77
100	3	-1332.00	13.36	-1332.00	13.41	0.00	-1332.00	13.45	0.00	0.00	-1338.40	5.00	0.48	-1177.70	5.81	11.58	13.65
	4	-1165.50	14.41	-1165.50	14.31	0.00	-1165.50	14.32	0.00	0.00	-1170.80	4.74	0.45	-1027.10	6.25	11.87	13.99
	5	-1404.20	13.67	-1404.20	13.90	0.00	-1404.20	13.59	0.00	0.00	-1410.20	4.43	0.43	-1247.20	9.20	11.18	13.07
	1	-5025.00	342.00	-5025.00	341.02	0.00	-5025.00	340.13	0.00	0.00	-5039.50	108.66	0.29	-4489.90	154.84	10.65	12.24
	2	-4905.60	432.57	-4905.60	428.21	0.00	-4905.60	428.05	0.00	0.00	-4923.40	159.12	0.36	-4291.10	145.74	12.53	14.74
200	3	-5224.20	368.04	-5224.20	367.90	0.00	-5224.20	365.32	0.00	0.00	-5236.10	129.56	0.23	-4735.40	165.19	9.36	10.57
	4	-4544.90	338.68	-4544.90	337.53	0.00	-4544.90	338.40	0.00	0.00	-4562.50	99.16	0.39	-3978.80	156.46	12.46	14.67
	5	-5082.10	336.39	-5082.10	333.06	0.00	-5082.10	337.63	0.00	0.00	-5096.90	107.12	0.29	-4539.00	154.74	10.69	12.29
	1	-	-	-	-	-	-	-	-	-	-21115.00	3571.70	-	-18777.00	4340.70	-	12.45
	2	-	-	-	-	-	-	-	-	-	-19502.00	3118.80	-	-17147.00	4737.10	-	13.73
400	3	-	-	-	-	-	-	-	-	-	-20576.00	3563.40	-	-17939.00	4335.40	-	14.70
	4	-	-	-	-	-	-	-	-	-	-20037.00	3130.60	-	-17530.00	5597.80	-	14.30
	5	-	-	-	-	-	-	-	-	-	-20896.00	3355.50	-	-18545.00	4957.60	-	12.68
	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
800	3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1200	3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1600	3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2000	3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Table C5 Instance-level Performance of PCA Approximation on the Newsvendor Problem (Part 2)

$\frac{m_1}{m}$ (%)				60%							40%						
Size (m)	Inst No.	Mosek	Time (secs)	PCA -LB	Time (secs)	Gap1 (%)	PCA -UB	Time (secs)	Gap2 (%)	Interval Gap (%)	PCA -LB	Time (secs)	Gap1 (%)	PCA -UB	Time (secs)	Gap2 (%)	Interval Gap (%)
100	1	-1286.50	12.32	-1302.20	1.34	1.22	-947.57	2.27	26.35	37.43	-1311.00	0.37	1.90	-798.73	0.61	37.91	64.14
	2	-1273.90	11.34	-1287.10	1.47	1.04	-977.48	2.41	23.27	31.68	-1295.20	0.37	1.67	-827.38	0.59	35.05	56.54
	3	-1332.00	13.36	-1343.80	1.37	0.89	-1034.90	2.30	22.30	29.85	-1352.70	0.44	1.55	-871.01	0.53	34.61	55.30
	4	-1165.50	14.41	-1175.80	1.49	0.88	-899.35	2.20	22.84	30.74	-1186.10	0.38	1.77	-726.28	0.56	37.69	63.31
	5	-1404.20	13.67	-1416.30	1.51	0.86	-1097.00	2.27	21.88	29.11	-1426.00	0.38	1.55	-930.82	0.53	33.71	53.20
200	1	-5025.00	342.00	-5060.30	27.90	0.70	-3850.80	45.91	23.37	31.41	-5082.70	5.46	1.15	-3280.60	7.94	34.71	54.93
	2	-4905.60	432.57	-4942.40	28.01	0.75	-3681.80	44.88	24.95	34.24	-4961.20	5.57	1.13	-3160.60	8.04	35.57	56.97
	3	-5224.20	368.04	-5254.40	28.38	0.58	-4159.80	38.24	20.37	26.31	-5279.80	4.30	1.06	-3500.30	8.22	33.00	50.84
	4	-4544.90	338.68	-4586.40	27.91	0.91	-3294.90	44.76	27.50	39.20	-4612.30	5.50	1.48	-2691.30	8.05	40.78	71.38
	5	-5082.10	336.39	-5117.90	31.46	0.70	-3853.90	47.60	24.17	32.80	-5141.40	5.26	1.17	-3271.50	7.87	35.63	57.16
400	1	-	-	-21169.00	821.43	-	-16441.00	1190.50	-	28.76	-21245.00	135.16	-	-13992.00	172.88	-	51.84
	2	-	-	-19565.00	806.29	-	-14518.00	1242.40	-	34.76	-19627.00	118.76	-	-12219.00	190.53	-	60.63
	3	-	-	-20633.00	746.19	-	-15382.00	1185.90	-	34.14	-20701.00	126.90	-	-13075.00	163.08	-	58.33
	4	-	-	-20103.00	745.90	-	-15039.00	1060.30	-	33.67	-20166.00	126.71	-	-12822.00	172.02	-	57.28
	5	-	-	-20948.00	808.32	-	-16365.00	1305.80	-	28.00	-21030.00	119.62	-	-13638.00	191.30	-	54.20
800	1	-	-	-	-	-	-	-	-	-	-79668.00	3315.70	-	-49697.00	5561.00	-	60.31
	2	-	-	-	-	-	-	-	-	-	-77065.00	3325.90	-	-47841.00	5349.40	-	61.09
	3	-	-	-	-	-	-	-	-	-	-80845.00	3064.50	-	-52462.00	4773.00	-	54.10
	4	-	-	-	-	-	-	-	-	-	-81234.00	3738.30	-	-52648.00	5373.90	-	54.30
	5	-	-	-	-	-	-	-	-	-	-81299.00	3310.60	-	-52083.00	5129.70	-	56.10
1200	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1600	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2000	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Table C6 Instance-level Performance of PCA Approximation on the Newsvendor Problem (Part 3)

$\frac{m_1}{m}(\%)$				20%							$m_1=2$						
Size (m)	Inst No.	Mosek	Time (secs)	PCA -LB	Time (secs)	Gap1 (%)	PCA -UB	Time (secs)	Gap2 (%)	Interval Gap (%)	PCA -LB	Time (secs)	Gap1 (%)	PCA -UB	Time (secs)	Gap2 (%)	Interval Gap (%)
100	1	-1286.50	12.32	-1326.80	0.15	3.13	-635.71	0.17	50.59	108.71	-1340.40	0.11	4.19	-543.04	0.13	57.79	146.83
	2	-1273.90	11.34	-1304.70	0.15	2.42	-687.64	0.17	46.02	89.74	-1327.60	0.11	4.22	-542.22	0.13	57.44	144.85
	3	-1332.00	13.36	-1365.40	0.15	2.51	-712.10	0.17	46.54	91.74	-1383.50	0.11	3.87	-578.03	0.12	56.60	139.35
	4	-1165.50	14.41	-1200.50	0.15	3.00	-566.47	0.17	51.40	111.93	-1226.10	0.11	5.20	-437.58	0.13	62.46	180.20
	5	-1404.20	13.67	-1439.40	0.15	2.51	-771.50	0.17	45.06	86.57	-1457.90	0.12	3.82	-649.78	0.13	53.73	124.37
200	1	-5025.00	342.00	-5119.80	0.44	1.89	-2596.60	0.60	48.33	97.17	-5186.10	0.12	3.21	-2060.80	0.14	58.99	151.65
	2	-4905.60	432.57	-4994.40	0.41	1.81	-2562.50	0.63	47.76	94.90	-5075.90	0.12	3.47	-1946.50	0.14	60.32	160.77
	3	-5224.20	368.04	-5315.90	0.46	1.76	-2866.90	0.61	45.12	85.42	-5381.30	0.12	3.01	-2311.10	0.14	55.76	132.85
	4	-4544.90	338.68	-4647.50	0.43	2.26	-2126.90	0.54	53.20	118.51	-4698.30	0.12	3.38	-1665.00	0.13	63.37	182.18
	5	-5082.10	336.39	-5173.10	0.43	1.79	-2715.90	0.64	46.56	90.47	-5242.30	0.12	3.15	-2143.50	0.14	57.82	144.57
400	1	-	-	-21339.00	6.70	-	-11594.00	9.11	-	84.05	-21522.00	0.13	-	-9371.90	0.16	-	129.64
	2	-	-	-19708.00	6.24	-	-9958.20	9.87	-	97.91	-19902.00	0.13	-	-7542.00	0.16	-	163.88
	3	-	-	-20799.00	6.73	-	-10727.00	8.69	-	93.89	-20966.00	0.13	-	-8775.80	0.16	-	138.91
	4	-	-	-20248.00	6.92	-	-10572.00	8.59	-	91.52	-20433.00	0.13	-	-8308.70	0.16	-	145.92
	5	-	-	-21117.00	5.85	-	-11362.00	9.99	-	85.86	-21303.00	0.13	-	-9320.10	0.16	-	128.57
800	1	-	-	-79911.00	154.48	-	-40500.00	200.80	-	97.31	-80463.00	0.31	-	-32126.00	0.24	-	150.46
	2	-	-	-77307.00	144.71	-	-38892.00	199.38	-	98.77	-77887.00	0.18	-	-29920.00	0.22	-	160.32
	3	-	-	-81103.00	124.46	-	-42637.00	188.42	-	90.22	-81643.00	0.18	-	-33281.00	0.22	-	145.31
	4	-	-	-81485.00	126.10	-	-43148.00	210.77	-	88.85	-82068.00	0.18	-	-33560.00	0.22	-	144.54
	5	-	-	-81558.00	135.08	-	-42763.00	220.47	-	90.72	-82118.00	0.18	-	-33479.00	0.23	-	145.28
1200	1	-	-	-183710.00	985.14	-	-96727.00	1366.70	-	89.93	-184810.00	0.26	-	-76036.00	0.33	-	143.06
	2	-	-	-173790.00	985.53	-	-86447.00	1300.90	-	101.04	-174850.00	0.26	-	-66649.00	0.32	-	162.34
	3	-	-	-182600.00	979.76	-	-96613.00	1363.80	-	89.00	-183690.00	0.26	-	-75896.00	0.33	-	142.03
	4	-	-	-179020.00	921.56	-	-91610.00	1247.70	-	95.42	-180010.00	0.26	-	-71834.00	0.32	-	150.59
	5	-	-	-185720.00	986.00	-	-99561.00	1423.20	-	86.54	-186830.00	0.26	-	-78973.00	0.32	-	136.57
1600	1	-	-	-315750.00	3582.30	-	-162510.00	4530.10	-	94.30	-317380.00	0.36	-	-125440.00	0.50	-	153.01
	2	-	-	-309960.00	3567.80	-	-153820.00	4507.80	-	101.51	-311670.00	0.36	-	-118400.00	0.44	-	163.23
	3	-	-	-324640.00	3695.80	-	-168020.00	5346.20	-	93.22	-326210.00	0.36	-	-132570.00	0.43	-	146.07
	4	-	-	-317780.00	3337.20	-	-165280.00	4731.50	-	92.27	-319500.00	0.38	-	-127390.00	0.47	-	150.80
	5	-	-	-327540.00	3548.40	-	-172320.00	5585.80	-	90.08	-329150.00	0.36	-	-136190.00	0.49	-	141.68
2000	1	-	-	-	-	-	-	-	-	-	-501430.00	0.50	-	-200430.00	0.60	-	150.18
	2	-	-	-	-	-	-	-	-	-	-506210.00	0.51	-	-206390.00	0.59	-	145.27
	3	-	-	-	-	-	-	-	-	-	-477400.00	0.51	-	-174350.00	0.61	-	173.82
	4	-	-	-	-	-	-	-	-	-	-503420.00	0.50	-	-201130.00	0.59	-	150.30
	5	-	-	-	-	-	-	-	-	-	-501710.00	0.51	-	-202070.00	0.61	-	148.29

Table C7 Instance-level Performance of PCA Approximation on the CVaR Problem (Part 1)

$\frac{m_1}{m}$ (%)				100%							80%						
Size (m)	Inst No.	Mosek	Time (secs)	PCA -LB	Time (secs)	Gap1 (%)	PCA -UB	Time (secs)	Gap2 (%)	Interval Gap (%)	PCA -LB	Time (secs)	Gap1 (%)	PCA -UB	Time (secs)	Gap2 (%)	Interval Gap (%)
100	1	4.53	17.18	4.53	17.01	0.00	4.53	16.89	0.00	0.00	2.97	5.91	34.50	7.28	6.01	60.62	59.22
	2	4.80	17.12	4.80	17.11	0.00	4.80	16.84	0.00	0.00	3.56	6.82	25.79	7.33	5.94	52.90	51.47
	3	4.95	14.86	4.95	15.03	0.00	4.95	14.80	0.00	0.00	3.78	6.79	23.55	6.75	6.41	36.54	44.01
	4	4.43	14.71	4.43	14.78	0.00	4.43	14.80	0.00	0.00	2.42	5.51	45.35	8.48	7.32	91.61	71.48
	5	3.55	16.14	3.55	16.31	0.00	3.55	16.00	0.00	0.00	2.09	5.94	41.13	8.19	6.07	130.54	74.46
200	1	3.48	435.49	3.48	439.62	0.00	3.48	439.48	0.00	0.00	2.29	165.75	34.16	6.02	167.46	72.80	61.90
	2	3.08	438.93	3.08	440.76	0.00	3.08	437.38	0.00	0.00	1.92	146.85	37.73	5.63	128.67	82.82	65.94
	3	2.96	472.15	2.96	470.65	0.00	2.96	469.18	0.00	0.00	2.08	136.76	29.78	5.20	147.38	75.64	60.02
	4	2.61	469.65	2.61	470.19	0.00	2.61	468.57	0.00	0.00	1.66	174.35	36.24	5.16	146.45	97.68	67.75
	5	3.29	440.47	3.29	441.60	0.00	3.29	440.62	0.00	0.00	2.27	147.26	31.04	5.52	155.78	67.59	58.86
400	1	-	-	-	-	-	-	-	-	-	1.36	3377.40	-	3.30	4254.30	-	58.86
	2	-	-	-	-	-	-	-	-	-	0.74	4014.60	-	3.34	4464.40	-	77.93
	3	-	-	-	-	-	-	-	-	-	1.06	3517.60	-	3.43	3920.40	-	69.01
	4	-	-	-	-	-	-	-	-	-	0.68	4446.20	-	3.83	4014.90	-	82.22
	5	-	-	-	-	-	-	-	-	-	1.30	3796.40	-	3.53	4209.90	-	63.28
800	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1200	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1600	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2000	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Table C8 Instance-level Performance of PCA Approximation on the CVaR Problem (Part 2)

$\frac{m_1}{m}$(%)				60%							40%						
Size (m)	Inst No.	Mosek	Time (secs)	PCA -LB	Time (secs)	Gap1 (%)	PCA -UB	Time (secs)	Gap2 (%)	Interval Gap (%)	PCA -LB	Time (secs)	Gap1 (%)	PCA -UB	Time (secs)	Gap2 (%)	Interval Gap (%)
	1	4.53	17.18	0.33	1.90	92.81	11.67	2.02	157.49	97.21	-2.76	0.50	160.98	16.82	0.68	270.99	116.44
	2	4.80	17.12	0.76	1.91	84.11	10.65	2.14	122.01	92.84	-2.18	0.49	145.43	14.35	0.71	199.20	115.19
100	3	4.95	14.86	0.80	2.14	83.82	11.67	2.14	135.91	93.14	-1.92	0.47	138.74	20.42	0.68	312.79	109.38
	4	4.43	14.71	-0.21	2.01	104.64	12.17	2.14	174.81	101.69	-2.55	0.49	157.66	17.42	0.72	293.34	114.66
	5	3.55	16.14	0.02	1.90	99.31	12.86	2.19	261.77	99.81	-2.63	0.50	174.02	20.47	0.72	476.10	112.85
	1	3.48	435.49	0.98	43.50	71.82	7.70	39.11	120.99	87.25	-1.03	7.37	129.57	11.53	7.88	230.93	108.94
	2	3.08	438.93	0.59	42.58	80.72	7.00	39.09	127.11	91.51	-1.80	5.84	158.52	11.43	8.01	270.79	115.78
200	3	2.96	472.15	0.37	44.19	87.41	8.05	38.07	172.10	95.37	-1.22	5.86	141.24	12.10	8.10	309.15	110.08
	4	2.61	469.65	-0.05	40.15	101.77	8.29	38.96	217.33	100.56	-2.07	6.34	179.30	13.44	8.12	414.57	115.41
	5	3.29	440.47	0.66	43.35	79.92	8.19	36.41	148.97	91.94	-1.40	5.88	142.67	12.53	8.06	280.57	111.21
	1	-	-	0.15	1173.60	-	4.99	1060.60	-	96.94	-1.71	152.21	-	9.13	154.43	-	118.75
	2	-	-	-0.35	1058.30	-	5.14	1004.10	-	106.90	-2.09	148.95	-	9.73	182.07	-	121.46
400	3	-	-	-0.21	1162.90	-	5.53	1057.90	-	103.81	-1.97	151.57	-	9.78	172.16	-	120.16
	4	-	-	-0.41	935.48	-	5.77	1000.00	-	107.11	-2.10	159.27	-	10.56	162.28	-	119.86
	5	-	-	0.04	1187.70	-	5.51	1060.00	-	99.24	-1.53	152.23	-	9.41	183.13	-	116.29
	1	-	-	-	-	-	-	-	-	-	-2.06	4063.20	-	7.78	4084.70	-	126.55
	2	-	-	-	-	-	-	-	-	-	-1.98	4058.60	-	8.43	4313.80	-	123.45
800	3	-	-	-	-	-	-	-	-	-	-1.71	4316.80	-	7.79	4092.10	-	122.00
	4	-	-	-	-	-	-	-	-	-	-2.09	4275.80	-	7.25	4089.90	-	128.80
	5	-	-	-	-	-	-	-	-	-	-1.80	4465.90	-	7.60	4349.30	-	123.70
	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1200	3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1600	3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2000	3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Table C9 Instance-level Performance of PCA Approximation on the CVaR Problem (Part 3)

$\frac{m_1}{m}$(%)				20%							$m_1 = 2$						
Size (m)	Inst No.	Mosek	Time (secs)	PCA -LB	Time (secs)	Gap1 (%)	PCA -UB	Time (secs)	Gap2 (%)	Interval Gap (%)	PCA -LB	Time (secs)	Gap1 (%)	PCA -UB	Time (secs)	Gap2 (%)	Interval Gap (%)
	1	4.53	17.18	-3.90	0.17	186.08	23.52	0.41	418.78	116.59	-4.75	0.11	204.81	26.01	0.35	473.84	118.26
	2	4.80	17.12	-3.50	0.18	172.99	20.50	0.41	327.37	117.08	-4.71	0.11	198.31	25.60	0.34	433.80	118.42
100	3	4.95	14.86	-3.56	0.18	172.05	25.95	0.39	424.76	113.73	-4.77	0.11	196.53	25.95	0.34	424.76	118.39
	4	4.43	14.71	-3.90	0.17	188.06	26.11	0.40	489.61	114.94	-4.89	0.11	210.47	26.33	0.34	494.64	118.58
	5	3.55	16.14	-3.88	0.18	209.10	25.78	0.39	625.50	115.04	-4.79	0.11	234.83	25.78	0.35	625.50	118.58
	1	3.48	435.49	-3.21	0.55	192.20	22.73	1.65	552.79	114.12	-4.92	0.12	241.30	25.95	1.29	645.08	118.96
	2	3.08	438.93	-3.48	0.58	212.88	23.36	1.77	657.93	114.89	-4.91	0.12	259.40	25.49	1.36	727.15	119.27
200	3	2.96	472.15	-3.41	0.56	215.42	19.86	1.75	571.47	117.19	-4.95	0.12	267.31	26.21	1.34	786.05	118.88
	4	2.61	469.65	-3.32	0.54	227.31	21.18	1.69	711.04	115.70	-4.74	0.12	281.72	26.17	1.35	902.11	118.13
	5	3.29	440.47	-3.37	0.56	202.50	24.69	1.71	650.30	113.66	-4.97	0.12	251.00	25.20	1.33	665.76	119.72
	1	-	-	-3.30	6.67	-	16.91	15.91	-	119.49	-4.90	0.13	-	25.50	8.32	-	119.21
	2	-	-	-3.37	6.62	-	15.66	14.98	-	121.55	-4.92	0.13	-	25.72	8.30	-	119.14
400	3	-	-	-3.59	7.17	-	16.55	15.43	-	121.69	-4.94	0.14	-	25.80	8.32	-	119.16
	4	-	-	-3.54	6.70	-	20.87	15.48	-	116.97	-4.94	0.14	-	25.49	8.29	-	119.39
	5	-	-	-3.53	6.65	-	17.51	15.08	-	120.15	-4.99	0.14	-	26.18	8.35	-	119.07
	1	-	-	-3.47	143.12	-	16.12	218.65	-	121.50	-4.93	0.19	-	25.38	12.28	-	119.43
	2	-	-	-3.50	144.60	-	15.70	216.28	-	122.28	-4.95	0.19	-	25.63	14.29	-	119.30
800	3	-	-	-3.34	155.60	-	15.46	245.10	-	121.62	-4.90	0.19	-	25.49	13.65	-	119.22
	4	-	-	-3.45	143.77	-	13.99	242.98	-	124.68	-4.97	0.19	-	25.04	11.44	-	119.85
	5	-	-	-3.51	154.61	-	14.56	253.47	-	124.13	-4.98	0.19	-	25.15	11.68	-	119.79
	1	-	-	-3.58	1080.30	-	12.77	1396.30	-	128.06	-4.97	0.28	-	25.21	20.99	-	119.70
	2	-	-	-3.46	956.30	-	14.20	1340.40	-	124.40	-4.98	0.29	-	25.52	23.64	-	119.53
1200	3	-	-	-3.51	1080.90	-	12.90	1337.00	-	127.22	-4.99	0.28	-	25.15	20.69	-	119.83
	4	-	-	-3.47	958.77	-	12.84	1347.00	-	127.04	-4.96	0.28	-	25.51	22.54	-	119.44
	5	-	-	-3.48	1084.20	-	11.51	1334.30	-	130.23	-4.98	0.29	-	25.45	19.78	-	119.57
	1	-	-	-3.46	3648.80	-	12.85	4063.90	-	126.94	-4.98	0.40	-	25.46	48.60	-	119.57
	2	-	-	-3.43	4048.20	-	11.87	4949.60	-	128.92	-4.99	0.40	-	25.53	41.03	-	119.56
1600	3	-	-	-3.53	3827.30	-	12.17	4732.50	-	129.00	-4.99	0.42	-	25.35	45.68	-	119.68
	4	-	-	-3.48	3652.40	-	13.09	4280.40	-	126.58	-4.98	0.43	-	25.47	46.43	-	119.57
	5	-	-	-3.56	4258.60	-	12.70	4644.50	-	128.02	-4.99	0.48	-	25.19	46.48	-	119.81
	1	-	-	-	-	-	-	-	-	-	-4.99	0.54	-	25.37	78.45	-	119.67
	2	-	-	-	-	-	-	-	-	-	-4.99	0.54	-	25.27	84.03	-	119.76
2000	3	-	-	-	-	-	-	-	-	-	-5.00	0.54	-	25.26	79.27	-	119.78
	4	-	-	-	-	-	-	-	-	-	-5.00	0.53	-	25.30	81.26	-	119.76
	5	-	-	-	-	-	-	-	-	-	-4.98	0.53	-	25.24	80.23	-	119.73