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

[*]Corresponding authors

Contact: shiyi-phd.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 by integrating 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 production-transportation problems. Numerical results show significant advantages of our approach regarding 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 (see Rahimian and Mehrotra 2019 and Lin et al. 2022 for detailed review).

To solve different applications, researchers study the DRO under various ambiguity sets. The ambiguity set plays a crucial role in connecting statistical information and optimization modeling, providing a flexible framework to model uncertainties and incorporate partial information of random parameters (e.g., from historical data and prior belief) into the model. Moreover, the performance of DRO depends significantly on the distributional ambiguity set (Mohajerin Esfahani and Kuhn 2018, Chen et al. 2023). This paper focuses on moment-based ambiguity sets, which include all distributions satisfying certain moment constraints. Examples of such constraints include restricting the exact mean and covariance matrix (Scarf 1958) and bounding the first and second moments (Ghaoui et al. 2003, Delage and Ye 2010). Moment-based DRO has been extensively studied 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 (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).

As a moment-based DRO model can be reformulated as a semi-infinite program (Xu et al. 2018), 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 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, theoretically efficient algorithms (e.g., the interior-point methods) to solve SDPs impose substantial demands on computational time and memory resources (Vandenberghe and Boyd 1996, Helmberg 2002), particularly when addressing high-dimensional SDPs. Widely adopted commercial solvers (e.g., Mosek) exhibit prohibitively long computational times when solving high-dimensional SDPs, and the computational burden escalates considerably even as the problem dimension increases gradually (see our numerical results in Section 7). Thus, it is of great interest to study efficient algorithms for solving SDPs in the context of moment-based DRO. Besides the generic methods (e.g., the interior point methods), two types of algorithms can speed up solving SDP reformulations of moment-based DRO: low-rank SDP algorithms and dimensionality reduction methods. First, some studies develop efficient algorithms by exploiting the low-rank properties of SDP constraints (Burer and Monteiro 2003, Yurtsever et al. 2021). These algorithms rarely have theoretical guarantees but are practically efficient. Specifically, the existing studies may reformulate convex SDPs as non-convex problems and subsequently develop efficient algorithms to deliver high-quality solutions within reduced time frames (Lemon et al. 2016). Second, dimensionality reduction techniques represent data with important statistical 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 computational time significantly.

However, both the general SDP algorithms and existing dimensionality reduction methods may not perform well for moment-based DRO. The general SDP algorithm aims to solve 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, leading to an *optimized dimensionality reduction (ODR) approach for moment-based DRO*. We summarize our contributions as follows:

1. Unlike the PCA approximation approaches (Cheng et al. 2018, Cheramin et al. 2022) that first reduce dimensionality and then solve approximation problems, we integrate the dimensionality reduction with the subsequent optimization problems and provide an ODR approach. 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.
2. We prove the low-rank property of the original high-dimensional SDP reformulations of moment-based DRO problems. Specifically, we show that there exists an optimal solution such that the ranks of matrices in SDP reformulations are less than the number of SDP constraints plus one. Such a property helps our low-dimensional approximations achieve the original optimal value, closing the approximation gap.

3. 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 prove that any accumulation point of the sequence produced by the ADMM algorithm satisfies the first-order stationary conditions of the low-dimensional bilinear SDP problem. We apply the ODR approach and ADMM algorithms to solve multiproduct newsvendor and production-transportation 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.

Note that the ODR approach echoes the recently emerging framework that integrates machine learning (e.g., parameter estimation) with decision-making (Bertsimas and Kallus 2020, Bertsimas and Koduri 2022, Elmachtoub and Grigas 2022). More relevant applications of such a framework 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. Unlike the above applications, we integrate dimensionality reduction with optimization in this paper (Jiang et al. 2023), which is recently followed by He and Mak (2023). He and Mak (2023) integrate the PCA with a subsequent stochastic program and provide a distributionally robust bound for the error between the objective values of the original and integrated problems. The integrated approach in He and Mak (2023) involves solving nonconvex and high-dimensional SDPs and may not reduce the error to zero, while our approach solves low-dimensional SDPs and can achieve the optimal value of the original moment-based DRO problem, thereby offering guidance on selecting the reduced dimension for practical applicability.

The remainder of this paper is organized as follows. Section 2 provides the SDP reformulation of moment-based DRO problems and illustrates the disadvantages of the PCA approximation approaches (Cheng et al. 2018, Cheramin et al. 2022). In Section 3, we propose the first outer approximation under the ODR approach, leading to a lower bound for the original problem, and are then motivated to develop the low-rank property of the original high-dimensional SDP reformulation, aiming to find a small reduced dimension to close the approximation gap. In Sections 4 and 5, motivated by the results in Section 3, we provide an inner approximation and a second outer approximation for the original problem, respectively, and both of them can achieve the original optimal value. Section 6 develops efficient algorithms to solve the above three approximations, all of which are low-dimensional bilinear SDP problems. In Section 7, we conduct extensive numerical experiments on multiproduct newsvendor and production-transportation problems. Section 8 concludes the paper. All proofs are presented in the Appendix if not specified.

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 $\boldsymbol{\Sigma}$, and the size of a matrix is indicated by $r \times c$, where r and c indicate the numbers of rows and columns, respectively. Italic subscripts indicate indices, e.g., S_k, while non-italic ones represent simplified specifications, e.g., $\mathbf{Q}_\mathrm{r}$. We use $\mathbb{E}_\mathbb{P}[\cdot]$ to represent the 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. We use $\mathbb{1}(\cdot)$ to denote the indicator function, which takes 1 if all the conditions encompassed in $(\cdot)$ are satisfied and takes 0 otherwise.

2. 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 information (Delage and Ye 2010) is included in $\mathcal{D}_{\mathrm{M0}}$ as follows:

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

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 $\gamma_1 \geq 0$, $\gamma_2 \geq 1$, 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

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

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., $\mathcal{S} = \{\boldsymbol{\xi} \mid \mathbf{A}\boldsymbol{\xi} \leq \mathbf{b}\}$ 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:*

$$\Theta_{\mathrm{M}}(m) = \min_{\mathbf{x}, s, \hat{\boldsymbol{\lambda}}, \mathbf{q}, \mathbf{Q}} \phi(m, s, \mathbf{q}, \mathbf{Q}) \tag{4a}$$

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

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

where $\hat{\boldsymbol{\lambda}} = \{\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_K\}$, $\mathbf{q} \in \mathbb{R}^m$, $\mathbf{Q} \in \mathbb{R}^{m\times m}$, $\phi(m, s, \mathbf{q}, \mathbf{Q}) := s + \gamma_2\mathbf{I}_m \bullet \mathbf{Q} + \sqrt{\gamma_1}\|\mathbf{q}\|_2$, $\chi(k, \mathbf{x}, s, \boldsymbol{\lambda}) := s - y_k^0(\mathbf{x}) - \boldsymbol{\lambda}^\top\mathbf{b} - y_k(\mathbf{x})^\top\boldsymbol{\mu} + \boldsymbol{\lambda}^\top\mathbf{A}\boldsymbol{\mu}$, and $\boldsymbol{\psi}(k, \mathbf{x}, \boldsymbol{\lambda}) := \mathbf{A}^\top\boldsymbol{\lambda} - y_k(\mathbf{x})$.

Note that the functions $\phi(m, s, \mathbf{q}, \mathbf{Q})$, $\chi(k, \mathbf{x}, s, \boldsymbol{\lambda})$, and $\boldsymbol{\psi}(k, \mathbf{x}, \boldsymbol{\lambda})$ will also be used in the remainder of this paper to simplify other SDP formulations. 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 $\mathbf{\Sigma}$ 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 $\mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}}$ by maintaining the first $m_1(\leq m)$ components of $\boldsymbol{\xi}_{\mathrm{I}}$ and fixing its other components at 0; that is,

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

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 $\mathbf{\Lambda}_{m_1}^{\frac{1}{2}} \in \mathbb{R}^{m_1\times m_1}$ are upper-left submatrices of $\mathbf{U}$ and $\mathbf{\Lambda}$, 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):

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

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

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

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

$$\Theta_{\mathrm{M}}(m_1) = \min_{\substack{\mathbf{x},s,\hat{\boldsymbol{\lambda}},\\ \mathbf{q}_{\mathrm{r}},\mathbf{Q}_{\mathrm{r}}}} \phi(m_1, s, \mathbf{q}_{\mathrm{r}}, \mathbf{Q}_{\mathrm{r}}) \tag{7a}$$

$$\text{s.t.} \begin{bmatrix} \chi(k,\mathbf{x},s,\boldsymbol{\lambda}_k) & \frac{1}{2}\left(\mathbf{q}_{\mathrm{r}} + \left(\mathbf{U}_{m\times m_1}\mathbf{\Lambda}_{m_1}^{\frac{1}{2}}\right)^{\top}\boldsymbol{\psi}(k,\mathbf{x},\boldsymbol{\lambda}_k)\right)^{\top} \\ \frac{1}{2}\left(\mathbf{q}_{\mathrm{r}} + \left(\mathbf{U}_{m\times m_1}\mathbf{\Lambda}_{m_1}^{\frac{1}{2}}\right)^{\top}\boldsymbol{\psi}(k,\mathbf{x},\boldsymbol{\lambda}_k)\right) & \mathbf{Q}_{\mathrm{r}} \end{bmatrix} \succeq 0, \quad \forall k \in [K], \tag{7b}$$

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

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):

$$\min_{t \in \mathbb{R}} \ t + \frac{1}{\alpha} \mathbb{E}_{\mathbb{P}} \left[g(\mathbf{x}, \boldsymbol{\xi}) - t \right]^{+}, \tag{8}$$

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 $\mathcal{S}$. We reformulate the distributionally robust counterpart of Problem (8) in Appendix B.1 and obtain Problem (46). Let $\alpha = 0.05$, $\mathcal{S} = \{\boldsymbol{\xi} \in \mathbb{R}^3 \mid 0 \leq \xi_1 \leq 8, 1 \leq \xi_2 \leq 12, 2 \leq \xi_3 \leq 16\}$, $\boldsymbol{\mu} = [1,2,3]$, $\boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0.2 & 0.1 \\ 0.2 & 3 & 0.15 \\ 0.1 & 0.15 & 2 \end{bmatrix}$ with eigenvalues 3.044, 1.983, and 0.973. Solving Problem (46) gives the optimal value 5.021 with $x_1 = 0.719$, $x_2 = 0.135$, $x_3 = 0.145$, and $t = 3.129$. Following Cheng et al. (2018) and Cheramin et al. (2022) to perform PCA approximation over Problem (46) by capturing only one of the three components in $\boldsymbol{\xi}$, we observe the following:

- Choosing the component corresponding to the largest eigenvalue 3.044, the PCA approximation gives the optimal value at 1.788 with $x_1 = 1$, $x_2 = 0$, $x_3 = 0$, and $t = 1.373$.
- Choosing the component corresponding to the second largest eigenvalue 1.983, the PCA approximation gives the optimal value at 1.3 with $x_1 = 0.7$, $x_2 = 0.3$, $x_3 = 0$, and $t = 1.3$.
- Choosing the component corresponding to the smallest eigenvalue 0.973, the PCA approximation gives the optimal value at 1.915 with $x_1 = 0.085$, $x_2 = 0.915$, $x_3 = 0$, and $t = 1.915$.

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 component with the smallest 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 3–5 and more importantly, the bounds can achieve the optimal value of the original Problem (3).

3. Lower Bound

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

$$\boldsymbol{\xi} = \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}} + \boldsymbol{\mu} \approx \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\boldsymbol{\xi}_{\mathrm{r}} + \boldsymbol{\mu}, \tag{9}$$

where $\mathbf{B}$ will be optimized in the subsequent PCA approximation, i.e., optimized dimensionality reduction. By (9), we project $\boldsymbol{\xi}_{\mathrm{I}}$ onto a subspace of $\mathbb{R}^{m \times m}$ spanned by the columns of $\mathbf{B} \in \mathcal{B}_{m_1}$ and approximate $\boldsymbol{\xi}_{\mathrm{I}}$ by the projection $\mathbf{B}\boldsymbol{\xi}_{\mathrm{r}}$, instead of considering only the random variables corresponding to the largest eigenvalues. When $\mathbf{B} = \begin{bmatrix} \mathbf{I}_{m_1} \\ \mathbf{0}_{(m-m_1)\times m_1} \end{bmatrix}$, (9) reduces to (5). 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). Unlike the existing PCA approach that first reduces the dimension of the uncertainty space and then provides approximations, our ODR method innovatively integrates dimensionality reduction with the subsequent optimization problems. Such an integrated framework deviates from the traditional dimensionality reduction method like PCA because we do not predetermine a low-dimensional space to consider in the subsequent optimization problem. Instead, we linearly map the high-dimensional uncertainty space to a low-dimensional space while such a mapping relationship (represented by the decision $\mathbf{B}$) is carefully optimized together with the subsequent optimization problems.

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:

$$\Theta_{\mathrm{L}}(m_1) = \max_{\mathbf{B} \in \mathcal{B}_{m_1}} \min_{\mathbf{x} \in \mathcal{X}} \max_{\mathbb{P}_{\mathrm{r}} \in \mathcal{D}_{\mathrm{L}}} \mathbb{E}_{\mathbb{P}_{\mathrm{r}}} \left[f\left(\mathbf{x}, \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\boldsymbol{\xi}_{\mathrm{r}} + \boldsymbol{\mu} \right) \right], \tag{10}$$

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

$$\mathcal{S}_{\mathrm{r}} := \left\{ \boldsymbol{\xi}_{\mathrm{r}} \in \mathbb{R}^{m_1} \;\middle|\; \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\boldsymbol{\xi}_{\mathrm{r}} + \boldsymbol{\mu} \in \mathcal{S} \right\}. \tag{11}$$

We will show that Problem (10) provides a lower bound for Problem (3) (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* (10) *provides a lower bound for the optimal value of Problem* (3)*, i.e.,* $\Theta_{\mathrm{L}}(m_1) \le \Theta_{\mathrm{M}}(m)$ *for any* $m_1 \le m$*; (ii) the optimal value of Problem* (10) *is nondecreasing in* m_1*, i.e.,* $\Theta_{\mathrm{L}}(m_1) \le \Theta_{\mathrm{L}}(m_2)$ *for any* $m_1 < m_2 \le m$*; and (iii) when* $m_1 = m$*, Problem* (3) *and Problem* (10) *have the same optimal value, i.e.,* $\Theta_{\mathrm{L}}(m) = \Theta_{\mathrm{M}}(m)$.

Theorem 1 shows that we obtain a lower bound for the optimal value of Problem (3) when reducing the dimensionality space of $\boldsymbol{\xi}_{\mathrm{I}}$ *while optimizing the choice of* $\mathbf{B} \in \mathcal{B}_{m_1}$ in Problem (10). 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 those in 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 (10) 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 $\Theta_{\mathrm{L}}(m_1)$ and $\Theta_{\mathrm{M}}(m)$ when we choose a small m_1. To that end, we follow the PCA approximation (7) to reformulate Problem (10) as the following SDP formulation:

$$\Theta_{\mathrm{L}}(m_1) = \max_{\mathbf{B}\in\mathcal{B}_{m_1}} \underline{\Theta}(m_1, \mathbf{B}), \tag{12}$$

where

$$\underline{\Theta}(m_1, \mathbf{B}) := \min_{\substack{\mathbf{x}, s, \hat{\boldsymbol{\lambda}}, \\ \mathbf{q}_{\mathrm{r}}, \mathbf{Q}_{\mathrm{r}}}} \phi(m_1, s, \mathbf{q}_{\mathrm{r}}, \mathbf{Q}_{\mathrm{r}}) \tag{13a}$$

$$\text{s.t.} \begin{bmatrix} \chi(k, \mathbf{x}, s, \boldsymbol{\lambda}_k) & \frac{1}{2}\left(\mathbf{q}_{\mathrm{r}} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\right)^\top \boldsymbol{\psi}(k, \mathbf{x}, \boldsymbol{\lambda}_k)\right)^\top \\ \frac{1}{2}\left(\mathbf{q}_{\mathrm{r}} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\right)^\top \boldsymbol{\psi}(k, \mathbf{x}, \boldsymbol{\lambda}_k)\right) & \mathbf{Q}_{\mathrm{r}} \end{bmatrix} \succeq 0, \quad \forall k \in [K], \tag{13b}$$

$$\mathbf{x} \in \mathcal{X};\ \hat{\boldsymbol{\lambda}} = \{\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_K\},\ \boldsymbol{\lambda}_k \in \mathbb{R}^l_+,\ \forall k \in [K];\ \mathbf{q}_{\mathrm{r}} \in \mathbb{R}^{m_1};\ \mathbf{Q}_{\mathrm{r}} \in \mathbb{R}^{m_1\times m_1}. \tag{13c}$$

Now we would like to find an $m_1 < m$ such that $\Theta_{\mathrm{L}}(m_1)$ in Problem (12) is close (even equal) to $\Theta_{\mathrm{M}}(m)$ in Problem (4). Note that if $\Theta_{\mathrm{L}}(m_1) = \Theta_{\mathrm{M}}(m)$, then comparing the SDP constraints between (4) and (12) shows that the rank of $\mathbf{Q}$ in the optimal solution of Problem (4) can be smaller than m. Specifically, we are motivated to explore the low-rank property of Problem (4) and obtain the following significant conclusion.

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 $k \in [K]$ depend on the optimal solution $(\mathbf{x}^*, s^*, \hat{\boldsymbol{\lambda}}^*, \mathbf{q}^*, \mathbf{Q}^*)$.*

Proof. 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

$$\begin{bmatrix} S_k & \frac{1}{2}\left(\mathbf{q}^* + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top(\mathbf{A}^\top\boldsymbol{\lambda}_k^* - y_k(\mathbf{x}^*))\right)^\top \\ \frac{1}{2}\left(\mathbf{q}^* + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top(\mathbf{A}^\top\boldsymbol{\lambda}_k^* - y_k(\mathbf{x}^*))\right) & \mathbf{Q}^* \end{bmatrix} \succeq 0,\ \ \forall k \in [K]. \tag{14}$$

We can equivalently rewrite (14) as

$$4S_k\mathbf{Q}^* \succeq \left(\mathbf{q}^* + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top\left(\mathbf{A}^\top\boldsymbol{\lambda}_k^* - y_k(\mathbf{x}^*)\right)\right)\left(\mathbf{q}^* + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top\left(\mathbf{A}^\top\boldsymbol{\lambda}_k^* - y_k(\mathbf{x}^*)\right)\right)^\top,\ \forall k \in [K];\ \ \mathbf{Q}^* \succeq 0. \tag{15}$$

Note that, if $S_k > 0$ for any $k \in [K]$, then (14) is equivalent to (15) by Schur complement; otherwise, when $S_k = 0$ for some $k \in [K]$, we have $(1/2)(\mathbf{q}^* + (\mathbf{U}\boldsymbol{\Lambda}^{1/2})^\top(\mathbf{A}^\top\boldsymbol{\lambda}_k^* - y_k(\mathbf{x}^*))) = \mathbf{0}_m$ by the definition of a PSD matrix. Thus, (14) is equivalent to $\mathbf{Q}^* \succeq 0$, i.e., (15).

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 $\boldsymbol{\kappa}_k \in \mathbb{R}^K$, $\forall k \in [K]$, such that

$$\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top\left(\mathbf{A}^\top\boldsymbol{\lambda}_k^* - y_k(\mathbf{x}^*)\right) = \mathbf{V}\boldsymbol{\kappa}_k,\ \forall k \in [K], \tag{16}$$

$\mathbf{V} = [\mathbf{v}_k,\ \forall k \in [K]] \in \mathbb{R}^{m\times K}$. 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 $\boldsymbol{\kappa}_0 \in \mathbb{R}^K$ and $\bar{\boldsymbol{\kappa}}_0 \in \mathbb{R}^{m-K}$ such that

$$\mathbf{q}^* = \mathbf{V}\boldsymbol{\kappa}_0 + \bar{\mathbf{V}}\bar{\boldsymbol{\kappa}}_0. \tag{17}$$

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

$$\mathbf{Q}^* = [\mathbf{V}\ \bar{\mathbf{V}}] \begin{bmatrix} \mathbf{Y}_{11} & \mathbf{Y}_{12} \\ \mathbf{Y}_{21} & \mathbf{Y}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{V}^\top \\ \bar{\mathbf{V}}^\top \end{bmatrix} = \mathbf{V}\mathbf{Y}_{11}\mathbf{V}^\top + \bar{\mathbf{V}}\mathbf{Y}_{21}\mathbf{V}^\top + \mathbf{V}\mathbf{Y}_{12}\bar{\mathbf{V}}^\top + \bar{\mathbf{V}}\mathbf{Y}_{22}\bar{\mathbf{V}}^\top, \tag{18}$$

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 (15), (16), and (17), we have

$$4S_k\mathbf{Q}^* \succeq \left(\mathbf{V}\boldsymbol{\kappa}_0 + \bar{\mathbf{V}}\bar{\boldsymbol{\kappa}}_0 + \mathbf{V}\boldsymbol{\kappa}_k\right)\left(\mathbf{V}\boldsymbol{\kappa}_0 + \bar{\mathbf{V}}\bar{\boldsymbol{\kappa}}_0 + \mathbf{V}\boldsymbol{\kappa}_k\right)^\top, \ \forall k \in [K]. \tag{19}$$

By (18) and (19), we have

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

By Lemma 2, we further have

$$\begin{aligned} &4S_k\mathbf{V}^\top(\mathbf{V}\mathbf{Y}_{11}\mathbf{V}^\top + \bar{\mathbf{V}}\mathbf{Y}_{21}\mathbf{V}^\top + \mathbf{V}\mathbf{Y}_{12}\bar{\mathbf{V}}^\top + \bar{\mathbf{V}}\mathbf{Y}_{22}\bar{\mathbf{V}}^\top)\mathbf{V} \\ &\qquad \succeq \mathbf{V}^\top(\mathbf{V}\boldsymbol{\kappa}_0 + \bar{\mathbf{V}}\bar{\boldsymbol{\kappa}}_0 + \mathbf{V}\boldsymbol{\kappa}_k)(\mathbf{V}\boldsymbol{\kappa}_0 + \bar{\mathbf{V}}\bar{\boldsymbol{\kappa}}_0 + \mathbf{V}\boldsymbol{\kappa}_k)^\top\mathbf{V}, \ \forall k \in [K]. \end{aligned} \tag{20}$$

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 (20) become

$$4S_k\mathbf{Y}_{11} \succeq (\boldsymbol{\kappa}_0 + \boldsymbol{\kappa}_k)(\boldsymbol{\kappa}_0 + \boldsymbol{\kappa}_k)^\top, \ \forall k \in [K]. \tag{21}$$

Now we let $\mathbf{q}' = \mathbf{V}\boldsymbol{\kappa}_0$ and $\mathbf{Q}' = \mathbf{V}\mathbf{Y}_{11}\mathbf{V}^\top$. By (21) and Lemma 2, we have

$$\begin{aligned} 4S_k\mathbf{Q}' =& 4S_k\mathbf{V}\mathbf{Y}_{11}\mathbf{V}^\top \succeq (\mathbf{V}\boldsymbol{\kappa}_0 + \mathbf{V}\boldsymbol{\kappa}_k)(\mathbf{V}\boldsymbol{\kappa}_0 + \mathbf{V}\boldsymbol{\kappa}_k)^\top \\ =& \left(\mathbf{q}' + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top\left(\mathbf{A}^\top\boldsymbol{\lambda}_k^* - y_k(\mathbf{x}^*)\right)\right)\left(\mathbf{q}' + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top\left(\mathbf{A}^\top\boldsymbol{\lambda}_k^* - y_k(\mathbf{x}^*)\right)\right)^\top, \ \forall k \in [K]. \end{aligned} \tag{22}$$

Comparing (4b) and (22), 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

$$s' + \gamma_2\mathbf{I}_m \bullet \mathbf{Q}' + \sqrt{\gamma_1}\,\|\mathbf{q}'\|_2 \geq s^* + \gamma_2\mathbf{I}_m \bullet \mathbf{Q}^* + \sqrt{\gamma_1}\,\|\mathbf{q}^*\|_2, \tag{23}$$

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

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

where the first equality holds by the definition of a matrix's trace, the second equality holds by (18), 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,

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

where the second equality holds by (17), 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 $\bar{\boldsymbol{\kappa}}_0^\top \bar{\boldsymbol{\kappa}}_0 \geq 0$. Thus, we have

$$s' + \gamma_2 \mathbf{I}_m \bullet \mathbf{Q}' + \sqrt{\gamma_1}\, \|\mathbf{q}'\|_2 \leq s^* + \gamma_2 \mathbf{I}_m \bullet \mathbf{Q}^* + \sqrt{\gamma_1}\, \|\mathbf{q}^*\|_2 . \tag{24}$$

Combining (23) and (24) 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 $\text{rank}(\mathbf{Q}') = \text{rank}(\mathbf{V}\mathbf{Y}_{11}\mathbf{V}^\top) \leq \min\{\text{rank}(\mathbf{V}), \text{rank}(\mathbf{Y}_{11})\} \leq K$, $\boldsymbol{\delta} = \boldsymbol{\kappa}_0$, and $\boldsymbol{\nu}_k = \boldsymbol{\kappa}_k$ for any $k \in [K]$. Thus, the proof is complete. □

When $K \geq m$, we have $\text{rank}(\mathbf{Q}') \leq m \leq K$, thereby no need to consider this case in Theorem 2. Note that K is the number of pieces formulating the piecewise linear function $f(\mathbf{x}, \boldsymbol{\xi})$ and it is usually small for practical problems. For instance, in the CVaR and newsvendor problems, we have $K = 2$ (see Example 1 and Section 7.1.1, respectively). Thus, Theorem 2 shows that the rank of $\mathbf{Q}'$ that optimizes Problem (4), i.e., K, is usually small. We then expect that for any $m_1 \in [m]$ and $\mathbf{B} \in \mathcal{B}_{m_1}$, the rank of the optimal $\mathbf{Q}_\mathrm{r}$ in Problem (13) might also be no greater than K and hence small for practical problems. With Problem (12), we then would like to choose a small $m_1 \geq K$ and find a $\mathbf{B} \in \mathcal{B}_{m_1}$ such that $\Theta_\mathrm{L}(m_1)$ can be close to $\Theta_\mathrm{M}(m)$.

We used to conjecture that the optimal value of Problem (12) equals that of Problem (4) when $m_1 \geq K$. Most numerical experiments (see Section 7) show this conjecture may be correct, while we also find a feasible solution of Problems (12) and (13) such that the corresponding objective value is equal to the optimal value of the original Problem (4) (see Theorem 7 in Appendix C.4). Nevertheless, we find an example to illustrate that the optimal value of Problem (12) with $m_1 = K$ can be strictly less than the optimal value of Problem (4) (see Example 2 in Appendix C.5). Thus, while the optimized dimensionality reduction maintains very high-quality solutions (mostly the optimal solutions as shown in our later numerical experiments in Section 7), 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 (12) 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 (12), we first reformulate it into a bilinear SDP problem (see Proposition 2) under the following assumption and then propose efficient algorithms (see Section 6) 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 $\mathbf{\Delta}_i \in \mathbb{R}^{\tau\times\tau}$ is symmetric 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 (12) has the same optimal value as the following bilinear SDP formulation:*

$$\Theta_{\mathrm{L}}(m_1) = \max_{\substack{t_k, \mathbf{p}_k, \mathbf{P}_k, \forall k \in [K], \\ \mathbf{Z}, \mathbf{B}}} \sum_{k=1}^K \left(t_k d_k^0 + \left(t_k \boldsymbol{\mu}^\top + \mathbf{p}_k^\top \left(\mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{B} \right)^\top \right) \mathbf{d}_k \right) - \sum_{i=1}^\tau \sum_{j=1}^\tau z_{ij} a_{ij}^0 \tag{25a}$$

$$\text{s.t.} \quad 1 - \sum_{k=1}^K t_k = 0, \ \sqrt{\gamma_1} - \left\| \sum_{k=1}^K \mathbf{p}_k \right\|_2 \geq 0, \ \gamma_2 \mathbf{I}_{m_1} - \sum_{k=1}^K \mathbf{P}_k = 0, \tag{25b}$$

$$t_k (\mathbf{A}\boldsymbol{\mu} - \mathbf{b})^\top + \mathbf{p}_k^\top \left(\mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{B} \right)^\top \mathbf{A}^\top \leq 0, \ \forall k \in [K], \tag{25c}$$

$$\sum_{k=1}^K \left(t_k \mathbf{w}_k^0 + \left(t_k \boldsymbol{\mu}^\top + \mathbf{p}_k^\top \left(\mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{B} \right)^\top \right) \mathbf{W}_k \right) - \sum_{i=1}^\tau \sum_{j=1}^\tau z_{ij} \mathbf{a}_{ij} = 0, \tag{25d}$$

$$\begin{bmatrix} t_k & \mathbf{p}_k^\top \\ \mathbf{p}_k & \mathbf{P}_k \end{bmatrix} \succeq 0, \ \forall k \in [K], \ \mathbf{B}^\top \mathbf{B} = \mathbf{I}_{m_1}, \ \mathbf{Z} \succeq 0, \tag{25e}$$

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 $\sum_{i=1}^n (\mathbf{\Delta}_i x_i) + \mathbf{\Delta}_0 \succeq 0$ 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 (13b).

Although solving Problem (25) may not achieve the optimal value of Problem (4), Theorem 7 demonstrates that we are not far from our target to close the approximation gap and motivates us to further develop an upper bound for the optimal value of Problem (4) while *closing the gap* between them in the next section.

4. Upper Bound

Motivated by the benefits of introducing a decision variable $\mathbf{B} \in \mathcal{B}_{m_1}$ in (9) to develop an outer approximation, we develop an inner approximation for Problem (3) in this section 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):

$$\Theta_\mathrm{U}(m_1) := \min_{\mathbf{B} \in \mathcal{B}_{m_1}} \overline{\Theta}(m_1, \mathbf{B}), \tag{26}$$

where

$$\overline{\Theta}(m_1, \mathbf{B}) = \min_{\mathbf{x} \in \mathcal{X}} \max_{\mathbb{P}_\mathrm{I} \in \mathcal{D}_\mathrm{U}} \mathbb{E}_{\mathbb{P}_\mathrm{I}}\left[f\left(\mathbf{x}, \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_\mathrm{I} + \boldsymbol{\mu}\right)\right] \tag{27}$$

with

$$\mathcal{D}_\mathrm{U}(\mathcal{S}_\mathrm{I}, \gamma_1, \gamma_2) = \left\{ \mathbb{P}_\mathrm{I} \;\middle|\; \begin{array}{l} \mathbb{P}_\mathrm{I}(\boldsymbol{\xi}_\mathrm{I} \in \mathcal{S}_\mathrm{I}) = 1, \;\; \mathbb{E}_{\mathbb{P}_\mathrm{I}}\left[\boldsymbol{\xi}_\mathrm{I}^\top\right] \mathbb{E}_{\mathbb{P}_\mathrm{I}}\left[\boldsymbol{\xi}_\mathrm{I}\right] \le \gamma_1 \\ \mathbb{E}_{\mathbb{P}_\mathrm{I}}\left[\mathbf{B}^\top \boldsymbol{\xi}_\mathrm{I}\boldsymbol{\xi}_\mathrm{I}^\top \mathbf{B}\right] \preceq \gamma_2 \mathbf{I}_{m_1} \end{array} \right\}. \tag{28}$$

The second-moment constraint in $\mathcal{D}_\mathrm{U}$ is relaxed from $\mathbb{E}_{\mathbb{P}_\mathrm{I}}[\boldsymbol{\xi}_\mathrm{I}\boldsymbol{\xi}_\mathrm{I}^\top] \preceq \gamma_2 \mathbf{I}_m$, where we introduce a decision variable $\mathbf{B} \in \mathcal{B}_{m_1}$ to *optimize such a relaxation* of the ambiguity set, leading to optimized dimensionality reduction over this second-moment constraint. 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 3. *The following three conclusions hold: (i) Problem* (26) *provides an upper bound for the optimal value of Problem* (3)*, i.e.,* $\Theta_\mathrm{U}(m_1) \ge \Theta_\mathrm{M}(m)$ *for any* $m_1 \le m$*; (ii) the optimal value of Problem* (26) *is nonincreasing in* m_1*, i.e.,* $\Theta_\mathrm{U}(m_1) \ge \Theta_\mathrm{U}(m_2)$ *for any* $m_1 < m_2 \le m$*; and (iii) when* $m_1 = m$*, Problem* (26) *and Problem* (3) *have the same optimal value, i.e.,* $\Theta_\mathrm{U}(m) = \Theta_\mathrm{M}(m)$*.*

Theorem 3 shows that Problem (26) 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 3. *Under Assumption 1, Problem* (27) *has the same optimal value as the following SDP formulation:*

$$\overline{\Theta}(m_1, \mathbf{B}) = \min_{\substack{\mathbf{x}, s, \hat{\boldsymbol{\lambda}}, \\ \mathbf{q}, \mathbf{Q}_\mathrm{r}, \hat{\mathbf{u}}}} \phi(m_1, s, \mathbf{q}, \mathbf{Q}_\mathrm{r}) \tag{29a}$$

$$\text{s.t.} \quad \begin{bmatrix} \chi(k, \mathbf{x}, s, \boldsymbol{\lambda}_k) & \frac{1}{2}\mathbf{u}_k^\top \\ \frac{1}{2}\mathbf{u}_k & \mathbf{Q}_\mathrm{r} \end{bmatrix} \succeq 0, \; \forall k \in [K], \tag{29b}$$

$$\mathbf{q} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top \boldsymbol{\psi}(k, \mathbf{x}, \boldsymbol{\lambda}_k) = \mathbf{B}\mathbf{u}_k, \; \forall k \in [K], \tag{29c}$$

$$\mathbf{x} \in \mathcal{X}, \; \mathbf{q} \in \mathbb{R}^m, \; \mathbf{Q}_\mathrm{r} \in \mathbb{R}^{m_1 \times m_1}, \tag{29d}$$

$$\hat{\boldsymbol{\lambda}} = \{\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_K\}, \; \boldsymbol{\lambda}_k \in \mathbb{R}_+^l, \; \hat{\mathbf{u}} = \{\mathbf{u}_1, \ldots, \mathbf{u}_K\}, \; \mathbf{u}_k \in \mathbb{R}^{m_1}, \; \forall k \in [K]. \tag{29e}$$

Proposition 3 shows that Problem (26) can be updated by replacing its inner optimization Problem (27) with Problem (29). More importantly, Problem (26) becomes a low-dimensional SDP formulation, which can be solved more efficiently than solving the original high-dimensional formulation (4). With the updated Problem (26), we have the following conclusion.

THEOREM 4. *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, $\mathbf{Y}_{11}$, *and* $\boldsymbol{\nu}_k$ $(\forall k \in [K])$ *that are defined in Theorem 2. If* $m_1 \geq K$, *then* $\Theta_{\mathrm{U}}(m_1) = \Theta_{\mathrm{M}}(m)$.

Theorem 4 shows that when $m_1 \geq K$, the optimal value of Problem (26) is always equal to the optimal value of the original problem, $\Theta_{\mathrm{M}}(m)$. Specifically, when $m_1 = K$, there exist optimal $\mathbf{B} = \mathbf{V}$ and $\mathbf{Q}_{\mathrm{r}} = \mathbf{Y}_{11}$ in Problem (26) such that $\Theta_{\mathrm{U}}(m_1) = \overline{\Theta}(m_1, \mathbf{V})$. Recall that K represents the number of pieces in the piecewise linear objective function $f(\mathbf{x}, \boldsymbol{\xi})$ and also determines the rank of the optimal $\mathbf{Q}$ in the original Problem (4) (see Theorem 2). Thus, we can decompose an optimal $\mathbf{Q}$ with a rank of K into $\mathbf{V}\mathbf{Y}_{11}\mathbf{V}^\top$, where $\mathbf{V} \in \mathbb{R}^{m \times K}$ such that $\mathbf{V}^\top\mathbf{V} = \mathbf{I}_K$ and $\mathbf{Y}_{11} \in \mathbb{R}^{K \times K}$. In Problem (26), when $m_1 = K$, we have $\mathbf{B} \in \mathbb{R}^{m \times K}$ and $\mathbf{Q}_{\mathrm{r}} \in \mathbb{R}^{K \times K}$ in the subproblem (29), by which constructing a solution with $\mathbf{B} = \mathbf{V}$ and $\mathbf{Q}_{\mathrm{r}} = \mathbf{Y}_{11}$ shows that $\Theta_{\mathrm{U}}(K) = \Theta_{\mathrm{M}}(m)$. Clearly, such a tightness result also holds when $m_1 > K$ by the monotonicity result in Theorem 3.

We may further interpret the insights as follows. Comparing the inner-approximation Problem (26) 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, when such a relaxation is jointly optimized via the decision $\mathbf{B}$ with the subsequent SDP reformulation, it 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 low-rank 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. More specifically, we have the following proposition holds.

PROPOSITION 4. *For any PSD matrix* $\mathbf{X} \in \mathbb{R}^{m \times m}$ *such that* $\operatorname{rank}(\mathbf{X}) \leq m_1 \leq m$, *we have the following equivalence holds:*

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

COROLLARY 1. *For any PSD matrix* $\mathbf{X} \in \mathbb{R}^{m \times m}$ *and* $\operatorname{rank}(\mathbf{X}) \leq m_1 \leq m$, *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 (26), Proposition 4 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 (26) with $m_1 \geq K$ does not lose the optimality.

5. Lower Bound Revisited

Given that Problem (26) with $m_1 = K$ maintains the optimal value of the original Problem (3), we can further perform dimensionality reduction based on Problem (26) as we do in Section 3, thereby obtaining a new lower bound for the optimal value of Problem (3). Specifically, we consider $K \leq m$ 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

$$\min_{\mathbf{B}\in\mathcal{B}_K} \min_{\mathbf{x}\in\mathcal{X}} \max_{\mathbb{P}_{\mathrm{I}}\in\mathcal{D}_{\mathrm{L2}}} \mathbb{E}_{\mathbb{P}_{\mathrm{I}}}\left[f\left(\mathbf{x}, \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}} + \boldsymbol{\mu}\right)\right] \tag{30}$$

with

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

$\mathbf{B} = [\mathbf{B}_1, \mathbf{B}_2]$, $\mathbf{B}_1 \in \mathbb{R}^{m\times m_1}$, 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 (26) 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 the columns of $\mathbf{B}_2$ 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 (26) and maintain high-quality solutions. Specifically, the following theorem holds.

THEOREM 5. *Under Assumption 1, by dualizing the inner maximization problem of Problem (30), we obtain the following SDP formulation:*

$$\Theta_{\mathrm{L2}}(m_1) = \min_{\substack{\mathbf{x},s,\hat{\boldsymbol{\lambda}},\\ \mathbf{q},\mathbf{Q}_{\mathrm{r}}',\hat{\mathbf{u}}',\hat{\mathbf{u}}'',\\ \mathbf{B}_1,\mathbf{B}_2}} \phi(m_1, s, \mathbf{q}, \mathbf{Q}_{\mathrm{r}}) \tag{31a}$$

$$\text{s.t.} \quad \begin{bmatrix} \chi(k, \mathbf{x}, s, \boldsymbol{\lambda}_k) & \frac{1}{2}(\mathbf{u}_k')^\top \\ \frac{1}{2}\mathbf{u}_k' & \mathbf{Q}_{\mathrm{r}}' \end{bmatrix} \succeq 0,\ \forall k \in [K], \tag{31b}$$

$$\mathbf{q} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top \boldsymbol{\psi}(k, \mathbf{x}, \boldsymbol{\lambda}_k) = \mathbf{B}_1\mathbf{u}_k' + \mathbf{B}_2\mathbf{u}_k'',\ \forall k \in [K], \tag{31c}$$

$$\mathbf{x} \in \mathcal{X},\ \mathbf{q} \in \mathbb{R}^m,\ \mathbf{Q}_{\mathrm{r}}' \in \mathbb{R}^{m_1\times m_1}, \tag{31d}$$

$$\mathbf{B}_1 \in \mathbb{R}^{m\times m_1}, \mathbf{B}_2 \in \mathbb{R}^{m\times(K-m_1)},\ [\mathbf{B}_1, \mathbf{B}_2]^\top[\mathbf{B}_1, \mathbf{B}_2] = \mathbf{I}_K, \tag{31e}$$

$$\hat{\boldsymbol{\lambda}} = \{\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_K\}, \ \boldsymbol{\lambda}_k \in \mathbb{R}^l_+, \ \forall k \in [K], \tag{31f}$$

$$\hat{\mathbf{u}}' = \{\mathbf{u}'_1, \ldots, \mathbf{u}'_K\}, \ \mathbf{u}'_k \in \mathbb{R}^{m_1}, \ \forall k \in [K], \tag{31g}$$

$$\hat{\mathbf{u}}'' = \{\mathbf{u}''_1, \ldots, \mathbf{u}''_K\}, \ \mathbf{u}''_k \in \mathbb{R}^{K-m_1}, \ \forall k \in [K]. \tag{31h}$$

In addition, the following three conclusions hold: (i) Problem (31) 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 (31) 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 (31) 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 (12) 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 (31) achieves the optimal value of the original problem when $m_1 = K$.

6. Efficient Algorithm

In Sections 3–5, we provide two outer approximations (i.e., Problems (25) and (31)) leading to lower bounds for the optimal value of Problem (3) and an inner approximation (i.e., Problem (26)) leading to an upper bound. Both Problems (26) and (31) can achieve the same optimal value as the original problem when $m_1 = K$. Note that the three approximations are low-dimensional bilinear SDP problems, which are nonconvex. It is hard to obtain the optimal solution of a bilinear SDP problem. To develop computationally efficient algorithms for bilinear SDPs, we derive Alternating Direction Method of Multipliers (ADMM) algorithms (Hajinezhad and Shi 2018, Wang et al. 2019, Themelis and Patrinos 2020) to solve the three approximations (see Appendix F.1 for more detailed reasons for choosing ADMM rather than other techniques). Meanwhile, it is important to note that our proposed algorithms are used to obtain a near-optimal dimensionality reduction matrix $\mathbf{B}$. Other solutions (e.g., $\mathbf{x}$) returned by the ADMM algorithm are not used here. Given this near-optimal $\mathbf{B}$, we can solve the low-dimensional SDP problems to retrieve the efficient lower and upper bounds for the original optimal value (see details in Section 6.2).

Note that Problems (25), (26), and (31) share the following generic formulation:

$$\min_{\mathbf{B},\mathbf{x},\tilde{\mathbf{u}}_k,\mathbf{u}_k,\forall k\in[K]} \ g\left(\mathbf{x}, \mathbf{u}_k, \tilde{\mathbf{u}}_k, \forall k \in [K]\right) \tag{32a}$$

$$\text{s.t.} \quad \left(\mathbf{x}, \mathbf{u}_k, \tilde{\mathbf{u}}_k, \ \forall k \in [K]\right) \in \mathcal{U}, \tag{32b}$$

$$\mathbf{B}^\top \mathbf{B} = \mathbf{I}_{m_1}, \tag{32c}$$

$$\tilde{\mathbf{u}}_k = \mathbf{B}\mathbf{u}_k, \ \forall k \in [K], \tag{32d}$$

where $\mathbf{B} \in \mathbb{R}^{m\times m_1}$, $\mathbf{u}_k \in \mathbb{R}^{m_1}$, $\tilde{\mathbf{u}}_k \in \mathbb{R}^{m}$, $\forall k \in [K]$, $\mathcal{U}$ is a convex set with at least one interior point, and $g(\cdot)$ is a differentiable convex function. Note that constraints (32c) and (32d) are bilinear constraints. We consider the following augmented Lagrangian problem for Problem (32):

$$\min_{\substack{\mathbf{x},\tilde{\mathbf{u}}_k,\mathbf{u}_k,\forall k\in[K];\\ \mathbf{B};\ \boldsymbol{\beta}_k,\forall k\in[K]}} \left\{ g\left(\mathbf{x},\mathbf{u}_k,\tilde{\mathbf{u}}_k,\forall k\in[K]\right) + \sum_{k=1}^{K}\boldsymbol{\beta}_k^\top\left(\tilde{\mathbf{u}}_k - \mathbf{B}\mathbf{u}_k\right) + \sum_{k=1}^{K}\frac{\rho_k}{2}\left\|\tilde{\mathbf{u}}_k - \mathbf{B}\mathbf{u}_k\right\|_2^2 \;\middle|\; (32b)-(32c)\right\}, \quad (33)$$

where $\boldsymbol{\beta}_k \in \mathbb{R}^m$ $(\forall k \in [K])$ are Lagrangian multipliers and $\rho_k > 0$ $(\forall k \in [K])$ are the penalty parameters. Thus, we design Algorithm 1 to solve Problem (33). Thereafter, we use the superscript i to denote the iteration step of Algorithm 1.

Algorithm 1 ADMM for Problem (32)

Initialize: $\mathbf{B}^0, \boldsymbol{\beta}_k^0, \forall k \in [K]$
Repeat: update $\left(\mathbf{x},\tilde{\mathbf{u}}_k,\mathbf{u}_k,\forall k\in[K]\right)$, $\mathbf{B}$ and $\boldsymbol{\beta}_k(\forall k\in[K])$ alternatingly by

Given $\mathbf{B}^i$ and $\boldsymbol{\beta}_k^i$ for any $k \in [K]$, solve Problem (33) to obtain the optimal solution $\left(\mathbf{x},\tilde{\mathbf{u}}_k,\mathbf{u}_k,\forall k\in[K]\right)^{i+1}$;

Given $\left(\mathbf{x},\tilde{\mathbf{u}}_k,\mathbf{u}_k,\forall k\in[K]\right)^{i+1}$ and $\boldsymbol{\beta}_k^i$ for any $k \in [K]$, solve Problem (33) to obtain the optimal solution $\mathbf{B}^{i+1}$;

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

In Algorithm 1, 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) and set $\boldsymbol{\beta}_k^0 = \mathbf{0}$ for any $k \in [K]$. We terminate the iteration when the improvement (regarding the relative gap) of the objective value is less than 10^{-4}. In this algorithm, given $\mathbf{B}$ and $\boldsymbol{\beta}_k$ for any $k \in [K]$, Problem (33) becomes a low-dimensional (i.e., m_1+1) SDP problem. Given $(\mathbf{x},\tilde{\mathbf{u}}_k,\mathbf{u}_k,\boldsymbol{\beta}_k,\forall k\in[K])$, Problem (33) becomes a nonconvex optimization problem, while the following proposition shows that it has an analytical optimal solution. Thus, Algorithm 1 can be performed efficiently.

PROPOSITION 5. *Given $(\mathbf{x},\tilde{\mathbf{u}}_k,\mathbf{u}_k,\boldsymbol{\beta}_k,\forall k\in[K])$, Problem (33) has an optimal solution $\mathbf{B}^* = \tilde{\mathbf{U}}\tilde{\mathbf{V}}^\top$, where $\sum_{k=1}^K(\boldsymbol{\beta}_k\mathbf{u}_k^\top + \rho_k\tilde{\mathbf{u}}_k\mathbf{u}_k^\top) = \tilde{\mathbf{U}}\tilde{\boldsymbol{\Sigma}}\tilde{\mathbf{V}}^\top$ for $\tilde{\mathbf{U}} \in \mathbb{R}^{m\times m_1}$, $\tilde{\boldsymbol{\Sigma}} \in \mathbb{R}^{m_1\times m_1}$, and $\tilde{\mathbf{V}} \in \mathbb{R}^{m_1\times m_1}$ by the singular value decomposition (SVD).*

6.1. Convergence Analyses

We further analyze the convergence property of Algorithm 1 to ensure the dimensionality reduction solution $\mathbf{B}$ returned by this algorithm is near-optimal, i.e., a theoretical guarantee. First, the following lemma holds.

LEMMA 3. *Given* $(\mathbf{x}, \tilde{\mathbf{u}}_k, \mathbf{u}_k, \boldsymbol{\beta}_k, \forall k \in [K])$, *we have* $\mathbf{B}^* = \tilde{\mathbf{U}}\tilde{\mathbf{V}}^\top$ *(an optimal solution of Problem (33)) is also an optimal solution of the following convex optimization problem:*

$$\max_{\mathbf{B} \in \mathbb{R}^{m \times m_1}} \left\{ \sum_{k=1}^{K} \left(\boldsymbol{\beta}_k \mathbf{u}_k^\top + \rho \tilde{\mathbf{u}}_k \mathbf{u}_k^\top \right) \bullet \mathbf{B} \;\middle|\; \mathbf{B}^\top \mathbf{B} \preceq \mathbf{I}_{m_1} \right\}. \tag{34}$$

To the best of our knowledge, no proof of convergence exists for the proposed ADMM algorithm in our specific setting. Existing convergence results focus on different problem settings. For instance, Hajinezhad and Shi (2018) employ the ADMM algorithm to solve a general bilinear optimization problem. Similar to our approach, they penalize the linking constraint and incorporate other constraints using an indicator function. However, in their setting, all the constraints except the linking constraint are convex, whereas we need to further deal with the nonconvex constraint $\mathbf{B}^\top \mathbf{B} = \mathbf{I}$. In addition, Chen and Goulart (2023) design an ADMM algorithm for diagonally constrained SDPs. Due to such a specific structure, they can derive the analytical optimal solution at each step. In contrast, the first step of our ADMM algorithm involves solving a general SDP problem for which no analytical optimal solution can be obtained.

Next, we present the convergence properties of our proposed ADMM algorithm. We let

$$\begin{aligned} &\mathcal{L}\left(\mathbf{B}, \left(\mathbf{x}, \tilde{\mathbf{u}}_k, \mathbf{u}_k, \forall k \in [K]\right), \left(\boldsymbol{\beta}_k, \forall k \in [K]\right)\right) = \\ &\qquad g\left(\mathbf{x}, \tilde{\mathbf{u}}_k, \mathbf{u}_k, \forall k \in [K]\right) + \sum_{k=1}^{K} \boldsymbol{\beta}_k^\top \left(\tilde{\mathbf{u}}_k - \mathbf{B}\mathbf{u}_k\right) + \sum_{k=1}^{K} \frac{\rho_k}{2} \left\|\tilde{\mathbf{u}}_k - \mathbf{B}\mathbf{u}_k\right\|_2^2, \end{aligned}$$

and introduce the following assumption commonly used in the convergence analysis of ADMM (Luo et al. 2008, Shen et al. 2014, Bertsekas 2014, Bai et al. 2021).

ASSUMPTION 3. *Given any* $k \in [K]$, *the sequence* $\{\boldsymbol{\beta}_k^i\}$ *is bounded and* $\sum_{i=0}^{\infty} \|\boldsymbol{\beta}_k^{i+1} - \boldsymbol{\beta}_k^i\|_2^2 < \infty$.

Because every bounded sequence has a convergent subsequence, we have

$$\boldsymbol{\beta}_k^i \to \boldsymbol{\beta}_k^*,\ i \to \infty, i \in \mathcal{I},\ \forall k \in [K], \tag{35}$$

where $\{\beta_k^i\}_{i \in \mathcal{I}}$ is a subsequence of $\{\beta_k^i\}$. Meanwhile, under Assumption 3, by the update rule $\boldsymbol{\beta}_k^{i+1} = \boldsymbol{\beta}_k^i + \rho_k^i(\tilde{\mathbf{u}}_k^{i+1} - \mathbf{B}^{i+1}\mathbf{u}_k^{i+1})$, we have

$$\boldsymbol{\beta}_k^{i+1} - \boldsymbol{\beta}_k^i = \rho_k^i(\tilde{\mathbf{u}}_k^{i+1} - \mathbf{B}^{i+1}\mathbf{u}_k^{i+1}) \to 0,\ i \to \infty,\ \forall k \in [K].$$

Because $\rho_k^i > 0$ for any $k \in [K]$, we have

$$\tilde{\mathbf{u}}_k^i - \mathbf{B}^i \mathbf{u}_k^i \to 0,\ i \to \infty,\ \forall k \in [K]. \tag{36}$$

We are now ready to state the convergence theorem of Algorithm 1 as follows.

THEOREM 6. *Let* $(\mathbf{B}^*, \mathbf{x}^*, \tilde{\mathbf{u}}_k^*, \mathbf{u}_k^*, \forall k \in [K])$ *be any accumulation point of the sequence* $\{\mathbf{B}^i, \mathbf{x}^i, \tilde{\mathbf{u}}_k^i, \mathbf{u}_k^i, \forall k \in [K]\}$ *generated by Algorithm 1. Then,* $(\mathbf{B}^*, \mathbf{x}^*, \tilde{\mathbf{u}}_k^*, \mathbf{u}_k^*, \forall k \in [K])$ *satisfies the first-order stationary conditions of Problem* (32).

Note that one sufficient condition to ensure the existence of the accumulation point is that, at each iteration step $i+1$, given $\mathbf{B}^i$ and $(\boldsymbol{\beta}_k^i, \forall k \in [K])$ obtained from the previous iteration i, the optimal solution of the following problem is bounded:

$$\min_{\mathbf{x}, \tilde{\mathbf{u}}_k, \mathbf{u}_k, \forall k \in [K]} \left\{ \mathcal{L}\left(\mathbf{B}^i, (\mathbf{x}, \tilde{\mathbf{u}}_k, \mathbf{u}_k, \forall k \in [K]), \left(\boldsymbol{\beta}_k^i, \forall k \in [K]\right)\right) \,\middle|\, (32b) \right\}. \tag{37}$$

This implies that, at iteration step $i+1$, the optimal solution of Problem (37), i.e., $(\mathbf{x}, \tilde{\mathbf{u}}_k, \mathbf{u}_k, \forall k \in [K])^{i+1}$, is bounded. By Assumption 3, the sequence $\{\boldsymbol{\beta}_k^i\}$ is also bounded. It follows that $\mathbf{B}^{i+1}$ is also bounded because $\mathbf{B}^{i+1}$ depends on $\boldsymbol{\beta}_k^i$, $\mathbf{u}_k^{i+1}$, and $\tilde{\mathbf{u}}_k^{i+1}$ by Proposition 5. Therefore, the bounded sequence $\{\mathbf{B}^i, \mathbf{x}^i, \tilde{\mathbf{u}}_k^i, \mathbf{u}_k^i, \forall k \in [K]\}$ must have accumulation points.

6.2. Quality of the Dimensionality Reduction Solution B

Although the optimal value of our proposed three approximations can provide efficient lower or upper bounds for the original optimal value, our derived ADMM algorithms may not produce an optimal solution for each corresponding approximation problem because all three approximations are nonconvex problems. Therefore, it is important to recover the efficient lower or upper bounds using the dimensionality reduction solution $\mathbf{B}^{\text{ADMM}}$ returned by the ADMM algorithms. Clearly, a better dimensionality reduction solution $\mathbf{B}^{\text{ADMM}}$ leads to a better lower or upper bound. Thus, we focus on the near-optimal dimensionality reduction solution $\mathbf{B}^{\text{ADMM}}$ that solving these outer and inner approximations can produce, rather than the solution of $\mathbf{x}$ that the ADMM algorithm returns. Specifically, we use this $\mathbf{B}^{\text{ADMM}}$ to recover the lower or upper bound for the optimal value of Problem (3) as follows.

(i) For Problem (25) (the first outer approximation), given the $\mathbf{B}^{\text{ADMM}}$ returned by the ADMM algorithm, we solve the low-dimensional SDP problem (13) to retrieve the lower bound.

(ii) For Problem (26) (the inner approximation), given the $\mathbf{B}^{\text{ADMM}}$ returned by the ADMM algorithm, we solve the low-dimensional SDP problem (29) to retrieve the upper bound.

(iii) For Problem (31) (the second outer approximation), given the $[\mathbf{B}_1, \mathbf{B}_2]^{\text{ADMM}}$ returned by the ADMM algorithm, we solve the low-dimensional SDP problem (13) to retrieve the lower bound.

Note that for Problem (31), given the $[\mathbf{B}_1, \mathbf{B}_2]^{\text{ADMM}}$ returned by the ADMM algorithm, solving the subproblem of Problem (31) may not provide a lower bound for the original optimal value. Therefore, to recover a lower bound, we solve the low-dimensional SDP problem (13) to retrieve the lower bound.

In addition, we can measure the quality of the $\mathbf{B}^{\text{ADMM}}$ by deriving a theoretical interval in which the optimal value of Problem (3) is located via the following proposition.

PROPOSITION 6. *Given any $m_1 \in [m]$ and $\mathbf{B}' \in \mathcal{B}_{m_1}$, we use $(\mathbf{x}^*, s^*, \hat{\boldsymbol{\lambda}}^*, \mathbf{q}_\mathrm{r}^*, \mathbf{Q}_\mathrm{r}^*)$ to denote an optimal solution of Problem (13). 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*

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

We have

$$\Theta_\mathrm{M}(m) - \underline{\Theta}(m_1, \mathbf{B}') \le \frac{P}{S}\mathbb{1}(\sqrt{P} - S < 0) + (2\sqrt{P} - S)\mathbb{1}(\sqrt{P} - S \ge 0).$$

Given any feasible $\mathbf{B}' \in \mathcal{B}_{m_1}$, solving the low-dimensional SDP problem (13) to obtain $(\mathbf{x}^*, s^*, \hat{\boldsymbol{\lambda}}^*, \mathbf{q}_\mathrm{r}^*, \mathbf{Q}_\mathrm{r}^*)$ is efficient. Thus, by solving an easy problem, we can obtain a theoretical lower bound $\underline{\Theta}(m_1, \mathbf{B}')$ and upper bound $\underline{\Theta}(m_1, \mathbf{B}') + (P/S)\mathbb{1}(\sqrt{P} - S < 0) + (2\sqrt{P} - S)\mathbb{1}(\sqrt{P} - S \ge 0)$ for the original Problem (3) and quantify the gap between them in Proposition 6. More specifically, we can obtain their theoretical relative gap (denoted by "Theoretical Gap") as follows:

$$\text{Theoretical Gap} = \frac{\frac{P}{S}\mathbb{1}(\sqrt{P} - S < 0) + (2\sqrt{P} - S)\mathbb{1}(\sqrt{P} - S \ge 0)}{|\underline{\Theta}(m_1, \mathbf{B}')|} \times 100\%. \tag{38}$$

Note that this theoretical gap can be applied to all the three proposed approximations (i.e., Problems (25), (26), and (31)) because the theoretical gap can be calculated for any $\mathbf{B}' \in \mathcal{B}_{m_1}$ and all three approximations can yield a high-quality dimensionality reduction matrix $\mathbf{B}^{\text{ADMM}}$ satisfying $(\mathbf{B}^{\text{ADMM}})^\top \mathbf{B}^{\text{ADMM}} = \mathbf{I}_{m_1}$ by using our proposed ADMM algorithm.

Proposition 6 implies that a better $\mathbf{B}^{\text{ADMM}}$ returned by our proposed ADMM algorithms leads to a tighter theoretical gap. Note that our proposed three approximations provide efficient lower and upper bounds for the original Problem (3) and two of them are exact when we reduce the dimensionality of the uncertain space from m to $m_1 = K$ (see Theorems 4 and 5). Thus, we expect to have a very high-quality dimensionality reduction solution $\mathbf{B}^{\text{ADMM}}$ returned by our ADMM algorithms, thereby implying a high-quality theoretical gap (38). Our numerical results in Section 7.2 show that such a theoretical gap is usually less than 5%.

7. 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 production-transportation 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 two hours. In Section 7.1, we specify the proposed inner and outer approximations under the ODR approach in the context of the multiproduct newsvendor and production-transportation problems. In Section 7.2, we report and analyze all the numerical results. All the data instances and source code can be found in Jiang et al. (2025).

7.1. Numerical Setup

7.1.1. Multiproduct Newsvendor Problem In the deterministic multiproduct newsvendor problem (Cheramin et al. 2022), 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

$$f(\mathbf{x}, \boldsymbol{\xi}) = \max\left\{(\mathbf{c}-\mathbf{v})^\top \mathbf{x}, (\mathbf{c}-\mathbf{g})^\top \mathbf{x} + (\mathbf{g}-\mathbf{v})^\top \boldsymbol{\xi}\right\}.$$

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 2, we obtain the following DRO counterpart:

$$\min_{\mathbf{x}\geq 0} \max_{\mathbb{P}\in\mathcal{D}_{\mathrm{M0}}} \mathbb{E}_{\mathbb{P}}\left[\max\left\{(\mathbf{c}-\mathbf{v})^\top \mathbf{x}, (\mathbf{c}-\mathbf{g})^\top \mathbf{x} + (\mathbf{g}-\mathbf{v})^\top \boldsymbol{\xi}\right\}\right]. \tag{39}$$

We can apply Proposition 1 to reformulate Problem (39) and the proposed inner and outer approximations (i.e., Problems (25), (26), and (31)) to approximate it (see details in Appendix G.1).

When the dimension m of $\boldsymbol{\xi}$ is large, the original SDP reformulation of Problem (39) (i.e., Problem (108)) becomes very difficult to solve because of the large-scale SDP constraints. Nevertheless, our approximations under the ODR approach (i.e., Problems (109)–(111)) have SDP matrices with very small sizes (e.g., $K+1=3$), largely reducing the computational burden while maintaining the solution quality.

7.1.2. Production-Transportation Problem (Bertsimas et al. 2010) The deterministic production-transportation problem considers G suppliers, each with normalized capacity one, and H customers, each with demand d_j ($\forall j \in [H]$) such that $\sum_{j=1}^{H} d_j \leq G$. Supplier $i \in [G]$ produces x_i goods with unit production cost c_i and delivers z_{ij} goods to customer $j \in [H]$ with unit transportation cost ξ_{ij}, thereby satisfying all customer demands and minimizing the total cost. We can formulate this problem as follows:

$$\min_{\mathbf{x},\mathbf{z}} \ \mathbf{c}^\top \mathbf{x} + \boldsymbol{\xi}^\top \mathbf{z} \tag{40a}$$

$$\text{s.t.} \quad 0 \le x_i \le 1, \ \forall i \in [G], \tag{40b}$$

$$\sum_{i=1}^{G} z_{ij} = d_j, \ \forall j \in [H], \quad \sum_{j=1}^{H} z_{ij} = x_i, \ \forall i \in [G], \quad z_{ij} \ge 0, \ \forall i \in [G], \ j \in [H], \tag{40c}$$

where $\boldsymbol{\xi} = (\xi_{ij}, \ \forall i \in [G], \ j \in [H])^\top$ and $\mathbf{z}_{ij} = (z_{ij}, \ \forall i \in [G], \ j \in [H])^\top$. Following the same setting in Bertsimas et al. (2010), we consider $\boldsymbol{\xi}$ is random and its probability distribution $\mathbb{P}$ belongs to $\mathcal{D}_{\mathrm{M0}}$. Thus, we can derive the two-stage DRO counterpart:

$$\min_{\mathbf{x}} \left\{ \mathbf{c}^\top \mathbf{x} + \max_{\mathbb{P} \in \mathcal{D}_{\mathrm{M0}}} \mathbb{E}_{\mathbb{P}} \left[\mathcal{U} \left(\mathcal{Q} \left(\mathbf{x}, \boldsymbol{\xi} \right) \right) \right] \, \middle| \, (40b) \right\}, \tag{41}$$

where $\mathcal{U}(\mathcal{Q}(\mathbf{x}, \boldsymbol{\xi})) = \max_{k \in [K]} \{\alpha_k \mathcal{Q}(\mathbf{x}, \boldsymbol{\xi}) + \beta_k\}$ is a convex nondecreasing disutility function that incorporates risk attitudes into the second-stage cost and $\mathcal{Q}(\mathbf{x}, \boldsymbol{\xi}) = \min_{\mathbf{z}} \{\boldsymbol{\xi}^\top \mathbf{z} \mid (40c)\}$. We can reformulate Problem (41) into an SDP problem and apply our inner and outer approximations to approximate it (see details in Appendix G.2).

When the dimension GH of $\boldsymbol{\xi}$ is large, the original SDP reformulation of Problem (41) (i.e., Problem (112)) becomes very difficult to solve due to the high-dimensional SDP constraints, while our approximations (i.e., Problems (114)–(116)) have SDP matrices with very small sizes (e.g., $K+1$) and can be solved efficiently.

7.2. 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), generating PCA-based lower and upper bounds. For the third benchmark, we consider the reduced dimension $m_1 \in \{100\% \times \dim(\boldsymbol{\xi}), 80\% \times \dim(\boldsymbol{\xi}), 60\% \times \dim(\boldsymbol{\xi}), 40\% \times \dim(\boldsymbol{\xi}), 20\% \times \dim(\boldsymbol{\xi}), K\}$, where $\dim(\boldsymbol{\xi})$ represents the dimension of the random vector $\boldsymbol{\xi}$. Note that $K = 2$ in the multiproduct newsvendor problem. For the production-transportation problem, we consider $K \in \{5, 10, 15\}$. Our proposed inner and outer approximations under the ODR approach are solved using Algorithm 1, providing near-optimal $\mathbf{B}^*$ and recovering valid lower and upper bounds (see Section 6.2). For all instances, through tuning a range of values, we set the initial Lagrangian multipliers $(\boldsymbol{\beta}_k, \forall k \in [K])$ to $\mathbf{0}$ and the penalty parameters $(\rho_k, \forall k \in [K])$ to 200 in Algorithm 1.

7.2.1. Instance Generation and Table Header Description We consider various instances of the multiproduct newsvendor and production-transportation problems. In the multiproduct newsvendor 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 production-transportation problem, we follow Bertsimas et al. (2010) to randomly generate the locations of G suppliers and H customers from a unit square and use ξ_{ij} to measure the distance between supplier $i \in [G]$ and customer $j \in [H]$. For any $i \in [G]$ and $j \in [H]$, we generate 10,000 samples of ξ_{ij} from the interval $[0.5\bar{\xi}_{ij}, 1.5\bar{\xi}_{ij}]$ and use them to estimate the mean, standard deviation, and covariance matrix of $\boldsymbol{\xi}$. Using γ_0 denoting the average transportation cost, we then generate production cost c_i and demand d_j uniformly on the intervals $[0.5\gamma_0, 1.5\gamma_0]$ and $[0.5G/H, G/H]$, respectively, for any $i \in [G]$ and $j \in [H]$. The disutility function $\mathcal{U}(x) = 0.25(e^{2x} - 1)$ and is approximated by a piecewise-linear function with K equidistant segments on the interval $[0,1]$. Meanwhile, we consider $(G,H) \in \{(4,25), (5,20), (5,40), (8,25), (10,40), (20,30), (20,40)\}$.

For any given value of m in the multiproduct newsvendor problem or (G,H) in the production-transportation problem, we randomly generate five instances and report the average performance in Tables 1–4. 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 (Burer and Monteiro 2003), 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 (25) with $m_1 = K$, the upper-bound performance after solving the inner approximation (26) with $m_1 = K$, and the other lower-bound performance after solving the second outer approximation (31) with $m_1 = K$, respectively. Note that using the dimensionality reduction solution $\mathbf{B}$ given by our proposed ADMM algorithms after solving each approximation, we retrieve the corresponding lower or upper bound following the recovery process described in Section 6.2. We use "PCA-100%," "PCA-80%," "PCA-60%," "PCA-40%," "PCA-20%," and "PCA-K" to denote the performance of the PCA approximation by Cheramin et al. (2022) when the reduced dimension m_1 equals $100\% \times \dim(\boldsymbol{\xi})$, $80\% \times \dim(\boldsymbol{\xi})$, $60\% \times \dim(\boldsymbol{\xi})$, $40\% \times \dim(\boldsymbol{\xi})$, $20\% \times \dim(\boldsymbol{\xi})$, and K, respectively. For instance, "PCA-20%" and "PCA-2" in Table 1 denote the performance of the case when $m_1 = 20\% \times \dim(\boldsymbol{\xi})$ and $m_1 = K = 2$, respectively.

In all the tables, we use "Size" to represent the value of m in the multiproduct newsvendor problem or (G, H) in the production-transportation problem and "Time" to represent the computational time in seconds required to solve each instance. Note that following the recovery process described in Section 6.2, the computational time includes two parts. The first part is the time spent using the ADMM algorithm to solve an approximation problem. The second part is the time spent solving the lower- or upper-bound subproblem. The lower-bound subproblem (13) and the upper-bound subproblem (29) are low-dimensional SDP problems. In all our computational experiments, the time spent on the second part is less than 1% of the time spent on the first part. Therefore, we report the total time without separately distinguishing between the two parts. 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.,

$$\text{Interval Gap} = \frac{\text{upper bound} - \text{lower bound}}{|\text{upper bound}|} \times 100\%. \tag{42}$$

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 (42). For the PCA approximation approach, the value of "upper bound" in (42) is provided by this approach itself. The value of the "Theoretical Gap" for each instance is defined in (38).

To further illustrate how to calculate the "Interval Gap," we consider the "Interval Gap" in the row "ODR-LB" of Table 1 as an example. We first use our ADMM algorithm to solve the first outer approximation, i.e., Problem (25), and record the obtained dimensionality reduction solution $\mathbf{B}^{\text{LB}}$. Given this $\mathbf{B}^{\text{LB}}$, we solve Problem (13) and obtain $\underline{\Theta}(m_1, \mathbf{B}^{\text{LB}})$, which is a lower bound. We then use the ADMM algorithm to solve the inner approximation, i.e., Problem (26), and record $\mathbf{B}^{\text{UB}}$. Given this $\mathbf{B}^{\text{UB}}$, we solve Problem (29) and obtain $\overline{\Theta}(m_1, \mathbf{B}^{\text{UB}})$, which is an upper bound. Thus, we calculate the "Interval Gap" by $(\overline{\Theta}(m_1, \mathbf{B}^{\text{UB}}) - \underline{\Theta}(m_1, \mathbf{B}^{\text{LB}})) / |\overline{\Theta}(m_1, \mathbf{B}^{\text{UB}})| \times 100\%$.

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 for the newsvendor problem with $m \geq 400$. Hence, we cannot obtain the value of "Gap1" for the "Mosek," "ODR-LB," and "ODR-RLB" approaches.

7.2.2. Numerical Performance From Tables 1–4, we have the following observations. First, in the newsvendor problem with $m \in \{100, 200\}$ and the production-transportation problem with $(G, H) \in \{(4, 25), (5, 20), (5, 40), (8, 25)\}$, the Mosek solver solves each instance of the original problem to the optimality. Our ODR approach performs much better than the three benchmark

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
	Theoretical Gap (%)	2.73	3.27	2.47	3.15	4.28	2.91	3.58
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
	Theoretical Gap (%)	1.92	2.36	1.27	2.44	1.90	3.01	2.58
ODR-UB	Gap2 (%)	1.68	1.80	-	-	-	-	-
	Time (secs)	1.95	2.60	4.33	9.75	20.83	38.36	56.68
	Theoretical Gap (%)	1.92	2.36	1.27	2.44	1.90	3.01	2.58
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-2	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 Production-Transportation Problem with $K = 5$

	Size $((G,H))$	(4,25)	(5,20)	(5,40)	(8,25)	(10,40)	(20,30)	(20,40)
Mosek	Time (secs)	56.04	70.10	1524.45	1612.68	-	-	-
Low-rank	Gap1 (%)	3.69	4.17	2.94	3.42	-	-	-
	Time (secs)	14.87	15.28	44.82	41.34	192.21	710.28	1678.21
	Interval Gap (%)	3.69	4.17	2.94	3.42	5.44	4.12	5.93
ODR-LB	Gap1 (%)	0.14	0.34	0.22	0.29	-	-	-
	Time (secs)	4.03	3.63	6.01	4.82	12.41	25.03	57.79
	Interval Gap (%)	0.15	0.34	0.43	0.29	0.52	0.55	0.51
	Theoretical Gap (%)	1.24	1.07	4.91	0.77	2.71	1.25	1.09
ODR-RLB	Gap1 (%)	0.01	0.02	0.01	0.00	-	-	-
	Time (secs)	5.42	5.36	22.40	20.86	123.64	330.29	665.43
	Interval Gap (%)	0.02	0.02	0.01	0.01	0.00	0.00	0.00
	Theoretical Gap (%)	1.13	1.17	1.80	0.81	1.22	1.92	1.32
ODR-UB	Gap2 (%)	0.01	0.01	0.00	0.00	-	-	-
	Time (secs)	5.48	5.32	22.41	20.86	123.48	329.95	665.35
	Theoretical Gap (%)	1.13	1.17	1.80	0.81	1.22	1.92	1.32
PCA-100%	Gap1 (%)	0.00	0.00	0.00	0.00	-	-	-
	Time (secs)	56.74	68.76	1521.32	1612.94	-	-	-
	Gap2 (%)	0.00	0.00	0.00	0.00	-	-	-
	Time (secs)	55.49	68.49	1521.04	1611.67	-	-	-
	Interval Gap (%)	0.00	0.00	0.00	0.00	-	-	-
PCA-80%	Gap1 (%)	0.52	0.99	0.33	1.50	-	-	-
	Time (secs)	23.41	27.68	589.74	625.65	-	-	-
	Gap2 (%)	3.31	2.18	1.15	2.76	-	-	-
	Time (secs)	29.33	31.20	657.78	639.09	-	-	-
	Interval Gap (%)	3.63	3.08	1.44	4.12	-	-	-
PCA-60%	Gap1 (%)	1.60	2.61	1.26	3.41	-	-	-
	Time (secs)	8.54	8.41	143.60	115.54	3087.98	-	-
	Gap2 (%)	9.02	7.33	3.45	8.62	-	-	-
	Time (secs)	9.04	8.88	158.95	152.21	3484.45	-	-
	Interval Gap (%)	9.50	9.17	4.36	11.04	4.85	-	-
PCA-40%	Gap1 (%)	3.75	4.23	2.33	4.00	-	-	-
	Time (secs)	2.18	2.11	30.80	23.44	512.53	3004.06	-
	Gap2 (%)	14.05	11.58	6.57	9.96	-	-	-
	Time (secs)	2.61	2.37	37.08	30.60	633.00	3149.11	-
	Interval Gap (%)	15.42	14.04	7.94	12.66	8.52	5.05	-
PCA-20%	Gap1 (%)	4.45	5.41	3.59	4.16	-	-	-
	Time (secs)	0.74	0.69	6.21	5.59	38.03	205.25	741.71
	Gap2 (%)	17.07	12.66	9.25	10.08	-	-	-
	Time (secs)	1.11	1.01	7.31	6.26	64.54	276.63	883.55
	Interval Gap (%)	18.35	15.95	11.32	12.91	9.39	5.22	4.85
PCA-5	Gap1 (%)	5.35	5.73	4.55	4.39	-	-	-
	Time (secs)	0.38	0.36	0.52	0.45	0.66	0.79	1.00
	Gap2 (%)	17.99	15.35	12.71	10.08	-	-	-
	Time (secs)	0.78	0.72	2.35	2.08	10.41	27.98	68.39
	Interval Gap (%)	19.76	18.26	15.15	13.12	9.69	5.48	5.14

Table 3 Average Performance on the Production-Transportation Problem with $K = 10$

	Size $((G,H))$	(4,25)	(5,20)	(5,40)	(8,25)	(10,40)	(20,30)	(20,40)
Mosek	Time (secs)	115.52	103.52	2866.54	2857.47	-	-	-
Low-rank	Gap1 (%)	3.41	3.92	3.21	4.08	-	-	-
	Time (secs)	17.67	18.41	56.22	54.17	227.71	929.00	1907.28
	Interval Gap (%)	3.41	3.92	3.21	4.08	5.27	6.2	4.62
ODR-LB	Gap1 (%)	0.12	0.19	0.13	0.25	-	-	-
	Time (secs)	6.65	6.41	11.11	10.31	31.42	61.41	131.10
	Interval Gap (%)	0.12	0.19	0.13	0.25	0.22	0.19	0.25
	Theoretical Gap (%)	1.78	1.97	2.96	1.19	2.10	1.49	1.22
ODR-RLB	Gap1 (%)	0.00	0.00	0.00	0.00	-	-	-
	Time (secs)	11.22	10.65	40.34	32.95	122.54	342.80	748.11
	Interval Gap (%)	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Theoretical Gap (%)	1.86	2.27	1.74	1.22	2.10	1.39	2.30
ODR-UB	Gap2 (%)	0.00	0.00	0.00	0.00	-	-	-
	Time (secs)	11.14	10.69	40.28	32.92	122.48	344.10	747.90
	Theoretical Gap (%)	1.86	2.27	1.74	1.22	2.10	1.39	2.30
PCA-100%	Gap1 (%)	0.00	0.00	0.00	0.00	-	-	-
	Time (secs)	115.74	102.90	2852.43	2851.09	-	-	-
	Gap2 (%)	0.00	0.00	0.00	0.00	-	-	-
	Time (secs)	116.32	103.70	2853.10	2859.01	-	-	-
	Interval Gap (%)	0.00	0.00	0.00	0.00	-	-	-
PCA-80%	Gap1 (%)	0.47	0.39	0.38	0.58	-	-	-
	Time (secs)	52.31	50.48	1101.61	979.90	-	-	-
	Gap2 (%)	0.04	0.04	0.02	0.02	-	-	-
	Time (secs)	58.12	58.88	1392.67	1364.82	-	-	-
	Interval Gap (%)	0.51	0.42	0.40	0.61	-	-	-
PCA-60%	Gap1 (%)	1.54	1.39	0.98	1.57	-	-	-
	Time (secs)	14.48	15.54	270.80	240.40	4182.33	-	-
	Gap2 (%)	0.96	0.34	0.34	0.40	-	-	-
	Time (secs)	19.64	20.83	455.00	440.12	4582.31	-	-
	Interval Gap (%)	2.46	1.72	1.32	1.95	2.61	-	-
PCA-40%	Gap1 (%)	2.30	2.31	1.76	2.18	-	-	-
	Time (secs)	4.19	4.04	53.67	45.22	831.61	4312.63	-
	Gap2 (%)	1.85	1.63	1.10	0.77	-	-	-
	Time (secs)	6.19	5.96	94.74	97.70	931.48	4792.10	-
	Interval Gap (%)	4.05	3.87	2.81	2.92	5.01	4.95	-
PCA-20%	Gap1 (%)	2.98	3.20	2.26	2.39	-	-	-
	Time (secs)	2.26	1.91	7.72	7.02	52.34	361.81	1038.72
	Gap2 (%)	2.24	2.44	1.71	1.07	-	-	-
	Time (secs)	3.83	3.52	18.70	18.07	71.38	428.93	1391.25
	Interval Gap (%)	5.09	5.49	3.90	3.42	6.21	5.48	4.29
PCA-10	Gap1 (%)	3.36	3.26	2.40	2.53	-	-	-
	Time (secs)	0.83	0.76	1.19	1.00	2.01	1.95	2.54
	Gap2 (%)	2.51	2.48	2.09	1.11	-	-	-
	Time (secs)	1.93	1.81	7.12	5.97	21.15	43.62	79.15
	Interval Gap (%)	5.72	5.59	4.40	3.60	6.36	7.28	6.04

Table 4 Average Performance on the Production-Transportation Problem with $K = 15$

	Size ((G,H))	(4,25)	(5,20)	(5,40)	(8,25)	(10,40)	(20,30)	(20,40)
Mosek	Time (secs)	154.22	159.12	4004.18	4099.14	-	-	-
Low-rank	Gap1 (%)	4.51	3.52	4.06	5.91	-	-	-
	Time (secs)	21.73	28.14	74.20	62.18	331.01	1124.41	2271.82
	Interval Gap (%)	4.51	3.52	4.06	5.91	4.22	5.11	4.83
ODR-LB	Gap1 (%)	0.05	0.10	0.12	0.15	-	-	-
	Time (secs)	13.92	16.56	22.11	26.16	43.02	80.51	168.90
	Interval Gap (%)	0.05	0.10	0.12	0.15	0.11	0.13	0.09
	Theoretical Gap (%)	4.69	5.22	5.57	6.38	3.29	4.41	3.95
ODR-RLB	Gap1 (%)	0.00	0.00	0.00	0.00	-	-	-
	Time (secs)	22.60	21.45	77.18	63.99	241.03	689.24	1550.21
	Interval Gap (%)	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Theoretical Gap (%)	1.38	2.61	1.79	1.73	2.14	1.94	2.25
ODR-UB	Gap2 (%)	0.00	0.00	0.00	0.00	-	-	-
	Time (secs)	22.61	21.44	76.99	64.08	149.21	401.63	878.68
	Theoretical Gap (%)	1.38	2.61	1.79	1.73	2.14	1.94	2.25
PCA-100%	Gap1 (%)	0.00	0.00	0.00	0.00	-	-	-
	Time (secs)	154.64	158.67	4024.62	4095.91	-	-	-
	Gap2 (%)	0.00	0.00	0.00	0.00	-	-	-
	Time (secs)	156.32	155.64	4050.04	4059.01	-	-	-
	Interval Gap (%)	0.00	0.00	0.00	0.00	-	-	-
PCA-80%	Gap1 (%)	0.47	0.53	0.62	0.58	-	-	-
	Time (secs)	68.31	69.22	1439.48	1428.65	-	-	-
	Gap2 (%)	0.04	0.06	0.15	0.08	-	-	-
	Time (secs)	79.55	80.20	2120.31	2146.05	-	-	-
	Interval Gap (%)	0.51	0.57	0.75	0.62	-	-	-
PCA-60%	Gap1 (%)	0.95	1.13	1.28	1.25	-	-	-
	Time (secs)	22.50	25.11	411.27	398.64	-	-	-
	Gap2 (%)	0.22	0.31	0.47	0.62	-	-	-
	Time (secs)	33.28	38.11	658.16	624.21	-	-	-
	Interval Gap (%)	1.17	1.40	1.74	1.85	-	-	-
PCA-40%	Gap1 (%)	1.68	1.67	1.78	1.81	-	-	-
	Time (secs)	6.61	8.53	81.20	80.88	1553.12	-	-
	Gap2 (%)	1.65	2.58	2.54	1.96	-	-	-
	Time (secs)	10.21	12.37	137.98	155.84	2296.17	-	-
	Interval Gap (%)	3.23	4.27	4.31	3.76	4.53	-	-
PCA-20%	Gap1 (%)	2.22	3.01	2.09	1.91	-	-	-
	Time (secs)	4.74	4.98	13.21	14.20	81.28	375.81	1671.21
	Gap2 (%)	3.07	3.61	2.98	2.48	-	-	-
	Time (secs)	9.24	8.19	27.26	28.79	124.04	496.38	2517.39
	Interval Gap (%)	5.27	6.59	5.06	4.22	5.78	6.29	5.88
PCA-15	Gap1 (%)	2.24	3.32	2.94	3.09	-	-	-
	Time (secs)	1.91	2.07	2.52	2.73	4.25	4.87	6.27
	Gap2 (%)	3.10	4.12	3.57	4.01	-	-	-
	Time (secs)	4.74	5.12	12.86	12.92	23.16	48.21	95.72
	Interval Gap (%)	5.31	7.40	6.41	7.03	6.42	7.19	6.83

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 $m_1 \neq 100\% \times \dim(\boldsymbol{\xi})$ therein and requires shorter computational time.

Specifically, the objective value of our "ODR-LB" reaches the optimal value of the original problem for some instances of the multiproduct newsvendor problem (see Table 1) and provides near-optimal solutions for the production-transportation problem with "Gap1" less than 0.34% (see Tables 2–4). More importantly, the "ODR-LB" reduces the computational time by up to three orders of magnitude compared to the Mosek solver. In addition, the "ODR-RLB" and "ODR-UB" *reach the optimal value of the original problem for all instances* in Tables 3–4 and provide objective values that are near-optimal (within 1.8% for all instances and 0.03% for most instances) in Tables 1–2, while reducing the computational time significantly. The results imply that our ADMM algorithms return the optimal $\mathbf{B}^*$ for most instances and the numerical gap between our derived lower and upper bounds (i.e., "Interval Gap") can be tight.

In addition, Tables 1–4 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 $m_1 \leq 80\% \times \dim(\boldsymbol{\xi})$ 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 $\dim(\boldsymbol{\xi})$ to K (i.e., maintaining only 1% of the original dimensionality size when $m = 200$ and $K = 2$ for the multiproduct newsvendor problem). More importantly, the inner and outer approximations of our ODR approach can be solved efficiently.

Second, when the problem size is large, i.e., $m \geq 400$ in the newsvendor problem and $(G, H) \in \{(10,40),(20,30),(20,40)\}$ in the production-transportation problem, the Mosek solver cannot solve any instance of the original problem to the optimality. Our ODR approach also performs better than the three benchmark approaches. Tables 1–4 show that "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 $m = 1600$ in the multiproduct newsvendor problem, the low-rank algorithm and "ODR-LB" take 309 and 2.01 seconds to solve an instance of the multiproduct newsvendor problem and provide the value of "Interval Gap" at 2.36% and 1.73%, 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 3546.3 seconds while the solution quality is very poor, providing the value of "Interval Gap" at 94.27%. Tables 2–4 show that the "ODR-RLB" and "ODR-UB" *reach the optimal value of the original problem for all instances* (with "Interval Gap" at 0); that is, our ADMM algorithms return the optimal $\mathbf{B}^*$. Meanwhile, the value of the "Theoretical Gap" is mostly within 5%. More importantly, our ODR approach is not sensitive to the value of $\dim(\boldsymbol{\xi})$, while the benchmark approaches perform much worse when $\dim(\boldsymbol{\xi})$ is larger. Thus, when we cannot obtain the optimal value of the original problem, the "ODR-LB", "ODR-RLB", and "ODR-UB" can be efficiently solved to provide a narrower 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-based DRO problems where other best possible benchmark approaches are struggling.

Third, "ODR-RLB" leads to better solution quality than "ODR-LB". From Tables 1-4, we observe that (i) there is no significant difference between the theoretical gaps of two outer approximations, (ii) the "Gap1" and "Interval Gap" of "ODR-RLB" are smaller than those of "ODR-LB" for most instances. Thus, we recommend that practitioners solve the second outer approximation to obtain a lower bound, where this approximation also provides a theoretical optimality guarantee when $m_1 = K$ (see Theorem 5).

Furthermore, although we obtain near-optimal solutions by setting $m_1 = K$ in the ODR approach, the sensitivity analyses of our ODR approach in Tables G3–G5 (see details in Appendix G.4) with respect to m_1 also show valuable results. Specifically, we consider the production-transportation problem (41), where m_1 takes values from $\{3,5,7\}$ when $K = 5$, $\{8,10,12\}$ when $K = 10$, and $\{13,15,17\}$ when $K = 15$. Concerning our ODR approach (i.e., "ODR-LB," "ODR-UB," and "ODR-RLB"), we observe a general trend where the values of "Gap1," "Gap2," and "Interval Gap" all tend to decrease as m_1 increases. This trend aligns with the theoretical results in Theorems 1, 3, and 5. Meanwhile, we consider the same problem with $G = 5$, $H = 20$, and $K = 10$ and provide a line chart in Figure 1 to demonstrate the trend concerning the value of m_1 in a higher granularity. These results demonstrate that (i) the computational times of our three approximations increase when we choose a larger m_1, (ii) for the first outer approximation, the gap between the lower bound and the optimal value decreases from 0.28% to 0.01% when we increase m_1 from 2 to 40, (iii) for the second outer approximation and the inner approximation, a similar conclusion holds, and they will achieve the original optimal value when we choose $m_1 \geq 6$. These results also demonstrate that $m_1 = K$ is a good choice, and increasing m_1 further may not bring significant improvement in terms of solution quality. Combining the tightness results from Theorems 4 and 5 and the sensitivity analyses here, we recommend that practitioners choose $m_1 = K$ to ensure that the proposed approximations yield high-quality results.

7.2.3. Numerical Insights In the multiproduct newsvendor problem, Tables 1 show that our ODR approach performs better than the PCA approximation with respect to the objective values

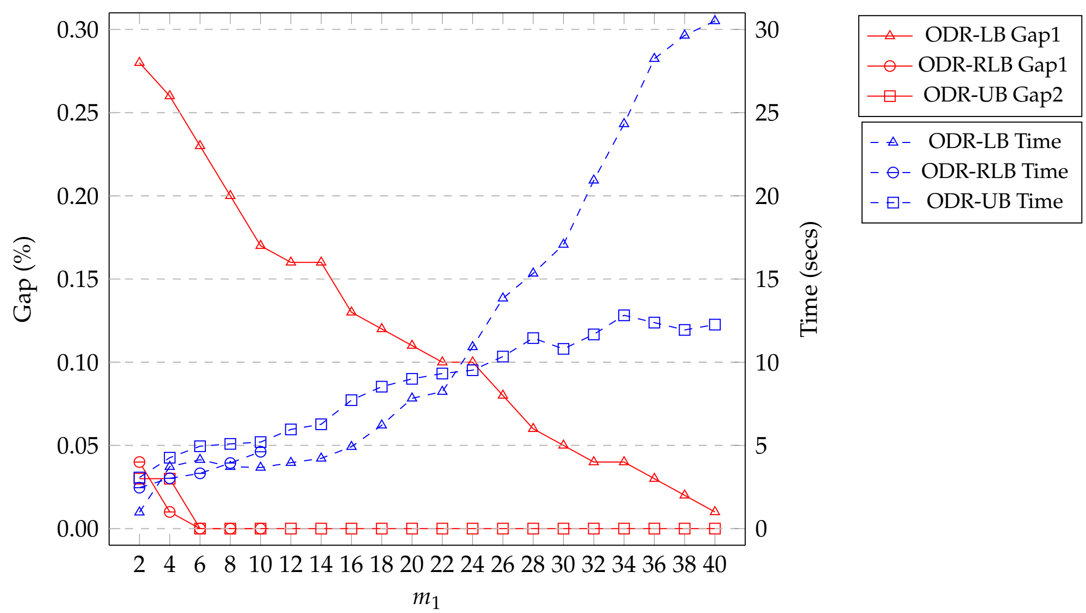

Figure 1 Sensitivity Analyses on the Production-Transportation Problem

for all the cases except that the "PCA-100%" (i.e., the original problem) provides the optimal value when the problem size is small. 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 (9), we solve the following problem to reduce the dimension from m to m_1:

$$\begin{aligned}\max_{\mathbf{B}^\top\mathbf{B}=\mathbf{I}_{m_1}} \mathbb{E}_{\mathbb{P}}\left[(\mathbf{g}-\mathbf{v})^\top\boldsymbol{\xi}\boldsymbol{\xi}^\top(\mathbf{g}-\mathbf{v})\right] &\approx \mathbb{E}_{\mathbb{P}}\left[(\mathbf{g}-\mathbf{v})^\top\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\boldsymbol{\xi}_{\mathrm{r}}+\boldsymbol{\mu}\right)\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\boldsymbol{\xi}_{\mathrm{r}}+\boldsymbol{\mu}\right)^\top(\mathbf{g}-\mathbf{v})\right]\\
&=\mathbb{E}_{\mathbb{P}}\left[(\mathbf{g}-\mathbf{v})^\top\left(\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\boldsymbol{\xi}_{\mathrm{r}}\right)\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\boldsymbol{\xi}_{\mathrm{r}}\right)^\top+2\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\boldsymbol{\xi}_{\mathrm{r}}\boldsymbol{\mu}^\top+\boldsymbol{\mu}\boldsymbol{\mu}^\top\right)(\mathbf{g}-\mathbf{v})\right]\\
&=\mathbb{E}_{\mathbb{P}}\left[(\mathbf{g}-\mathbf{v})^\top\left(\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\boldsymbol{\xi}_{\mathrm{r}}\right)\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\boldsymbol{\xi}_{\mathrm{r}}\right)^\top+\boldsymbol{\mu}\boldsymbol{\mu}^\top\right)(\mathbf{g}-\mathbf{v})\right].\end{aligned} \tag{43}$$

By introducing $\mathbf{r}=(\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{U}^\top)(\mathbf{g}-\mathbf{v})$, Problem (43) clearly has the same optimal solution as

$$\max_{\mathbf{B}^\top\mathbf{B}=\mathbf{I}_{m_1}} \mathbf{r}^\top\mathbf{B}\mathbf{B}^\top\mathbf{r}. \tag{44}$$

PROPOSITION 7. *We have* $\mathbf{B}^* = [\mathbf{r}/\|\mathbf{r}\|_2, \mathbf{0}_{m\times(m_1-1)}]$ *is an optimal solution of Problem* (44)*.*

By considering the partial feature of the original optimization problem, the optimal $\mathbf{B}^*$ of Problem (43) 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 (43) with the optimal $\mathbf{B}$ provided by our proposed outer approximation (25) 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 (43) 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 (25) (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 (43) 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 the partial feature of the optimization problem when reducing the dimensionality.

8. 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 3–5), 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 aim to choose a small m_1 to close the approximation 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 7) but may not reach it (Example 2). Nevertheless, Theorem 7 demonstrates that we are not far from closing the approximation gap and motivates us to derive an inner approximation to provide an upper bound for the optimal value of the original problem (Theorem 3). More importantly, this upper bound reaches the original optimal value when the reduced dimension m_1 is small (Theorem 4). 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 5).

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 3 and Theorem 5). We accordingly develop modified ADMM algorithms to solve them efficiently (Section 6) and analyze the convergence property of the ADMM algorithms (see Section 6.1). Based on the near-optimal dimensionality reduction solution $\mathbf{B}^{\text{ADMM}}$ returned by the ADMM algorithms, we also explain how to recover the corresponding lower and upper bounds for the original optimal value (see Section 6.2). Finally, we demonstrate the effectiveness of our ODR approach in solving multiproduct newsvendor and production-transportation 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). The numerical results show that our ODR approach significantly outperforms these benchmarks in 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 7.2.2). In addition, we provide insights into why our ODR approach performs better than the existing PCA approximations (Section 7.2.3).

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, it is also of great interest to integrate the idea of dimensionality reduction into the Wasserstein DRO or two-stage stochastic programming. Fourth, 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. Fifth, it would be an interesting extension to consider cases with $K \geq m$. Note that for these cases, our proposed ODR approach can still be used to improve the traditional PCA approximation by choosing an appropriate m_1 such that the solution quality is good and the computational time is short, and also provide theoretical guarantees that bound the gap between the proposed upper bound and the original optimal value, as well as the gap between the original optimal value and the proposed lower bound. Nevertheless, we may not provide clear guidance on choosing m_1 so that our proposed approximations achieve the optimal value of the original problem. We leave the above extensions for future research.

Acknowledgments

The authors thank the Area Editor, Associate Editor, and three anonymous reviewers very much for their valuable comments and suggestions, which have significantly improved the quality of this paper. Kai Pan and Zuo-Jun Max Shen are co-corresponding authors. Non-student authors (i.e., non-first authors) are listed in alphabetical order. This research was supported by the Research Grants Council of Hong Kong [Grant 15503723].

References

Bai, X., Sun, J., and Zheng, X. (2021). An augmented lagrangian decomposition method for chance-constrained optimization problems. *INFORMS Journal on Computing*, 33(3):1056–1069.

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

Bertsekas, D. P. (2014). *Constrained optimization and Lagrange multiplier methods*. Academic press.

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

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

Bertsimas, D. and Koduri, N. (2022). Data-driven optimization: A reproducing kernel Hilbert space approach. *Operations Research*, 70(1):454–471.

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.

Chen, L., Fu, C., Si, F., Sim, M., and Xiong, P. (2023). Robust optimization with moment-dispersion ambiguity. *Available at SSRN 4525224*.

Chen, Y. and Goulart, P. (2023). Burer-Monteiro ADMM for large-scale SDPs. *arXiv preprint arXiv:2302.04016*.

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.

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, L. and Mak, H.-Y. (2023). Prescriptive PCA: Dimensionality reduction for two-stage stochastic optimization. *arXiv preprint arXiv:2306.02223*, Jun 4, 2023.

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.

Jiang, S., Cheng, J., Pan, K., and Shen, Z.-J. M. (2023). Optimized dimensionality reduction for moment-based distributionally robust optimization. *arXiv preprint arXiv:2305.03996*, May 6, 2023.

Jiang, S., Cheng, J., Pan, K., and Shen, Z.-J. M. (2025). Optimized dimensionality reduction for moment-based distributionally robust optimization. Available for download at https://github.com/jsy1164014200/ODR-MDRO.

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.

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.

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.

Löfberg, J. (2004). Yalmip: A toolbox for modeling and optimization in matlab. In *In Proceedings of the CACSD Conference*, Taipei, Taiwan.

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.

Luo, H., Sun, X., and Li, D. (2008). On the convergence of augmented lagrangian methods for constrained global optimization. *SIAM Journal on Optimization*, 18(4):1209–1230.

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*, Aug 13, 2019.

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*.

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.

Shen, Y., Wen, Z., and Zhang, Y. (2014). Augmented lagrangian alternating direction method for matrix separation based on low-rank factorization. *Optimization Methods and Software*, 29(2):239–263.

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

Themelis, A. and Patrinos, P. (2020). Douglas–rachford splitting and ADMM for nonconvex optimization: Tight convergence results. *SIAM Journal on Optimization*, 30(1):149–181.

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

Wang, Y., Yin, W., and Zeng, J. (2019). Global convergence of ADMM in nonconvex nonsmooth optimization. *Journal of Scientific Computing*, 78:29–63.

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.

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.

Shiyi Jiang is a Ph.D. student at the Faculty of Business of The Hong Kong Polytechnic University. He received his Bachelor's degree in Information Management and Information Systems from Huazhong University of Science and Technology, China, in 2021. His research interests include stochastic programming, distributionally robust optimization, and their applications.

Jianqiang Cheng is an Associate Professor in the Department of Systems and Industrial Engineering at the University of Arizona. He received his B.S. degree in Mathematics and Applied Mathematics from Shanghai University, China, and his Ph.D. degree in Informatique from the University of Paris-Sud, France. His research interests include stochastic programming, robust optimization, conic optimization, and their applications.

Kai Pan is an Associate Professor at the Faculty of Business of The Hong Kong Polytechnic University. He received his Ph.D. degree from the University of Florida, USA, in 2016 and his Bachelor's degree from Zhejiang University, China, in 2010. His research interests include stochastic and discrete optimization, robust and data-driven optimization, and dynamic programming and their applications in the energy market, smart cities, supply chain, shared mobility, marketing, and transportation.

Zuo-Jun Max Shen is a Chair Professor, Vice-President, and Pro-Vice-Chancellor (Research) at the University of Hong Kong. He received his Ph.D. degree from Northwestern University, USA, in 2000. That same year, he began his academic career as an Assistant Professor at the University of Florida and joined the University of California, Berkeley in 2004, where he rose through the ranks to become Chancellor's Professor and Chair of the Department of Industrial Engineering and Operations Research and Professor of the Department of Civil and Environmental Engineering. His research concerns logistics and supply chain management, data-driven decision-making, and system optimization and their applications in businesses, energy systems, transportation systems, smart cities, healthcare management, and environmental protection.

Online Supplement for "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
[*]Corresponding authors
Contact: shiyi-phd.jiang@connect.polyu.hk (SJ), jqcheng@arizona.edu (JC),
kai.pan@polyu.edu.hk (KP), maxshen@berkeley.edu (Z-JMS)

This Online Supplement expands upon the main body of the paper by providing: (i) a list of notation in a table for quick reference, (ii) detailed reformulations in Example 1, (iii) complete proofs of lemmas, propositions, and theorems in the main body, (iv) Theorem 7 and its explanation, (v) a counter-example to illustrate that the optimal value of Problem (13) with $m_1 = K$ and $\mathbf{B} = \mathbf{V}$ is strictly less than the optimal value of Problem (4), (vi) motivation to derive ADMM algorithms and the detailed ADMM algorithm for Problem (26), (vii) detailed formulations of the multi-product newsvendor problem and the production-transportation problem, and (viii) sensitivity analyses for our proposed ODR approach.

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, $\boldsymbol{\lambda}_k$, $\mathbf{q}$, $\mathbf{Q}$	Decision variables in original SDP problem
$\hat{\boldsymbol{\lambda}}$	$\hat{\boldsymbol{\lambda}} := \{\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_K\}$
$\mathbf{q}_{\mathrm{r}}$, $\mathbf{Q}_{\mathrm{r}}$	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
$\overline{\Theta}_{\mathrm{U}}(m_1)$	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: Supplement to Section 2

B.1. Reformulations in Example 1

The distributionally robust counterpart of the CVaR problem (8) can be formulated as

$$\begin{aligned}&\min_{\mathbf{x}\in\mathcal{X}}\max_{\mathbb{P}\in\mathcal{D}}\min_{t\in\mathbb{R}} t+\frac{1}{\alpha}\mathbb{E}_{\mathbb{P}}\left[g\left(\mathbf{x},\boldsymbol{\xi}\right)-t\right]^{+}\\ =&\min_{\mathbf{x}\in\mathcal{X},\, t\in\mathbb{R}}\max_{\mathbb{P}\in\mathcal{D}} t+\frac{1}{\alpha}\mathbb{E}_{\mathbb{P}}\left[g\left(\mathbf{x},\boldsymbol{\xi}\right)-t\right]^{+},\end{aligned}\tag{45}$$

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 (45) has the same optimal value with the following SDP formulation:

$$\min_{\substack{\mathbf{x},s,t,\boldsymbol{\lambda}_1,\\ \boldsymbol{\lambda}_2,\mathbf{q},\mathbf{Q}}} s+\mathbf{I}_m\bullet\mathbf{Q}\tag{46a}$$

$$\text{s.t.}\quad\begin{bmatrix} s-t-\boldsymbol{\lambda}_1^\top\mathbf{b}+\boldsymbol{\lambda}_1^\top\mathbf{A}\boldsymbol{\mu} & \frac{1}{2}\left(\mathbf{q}+\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top\mathbf{A}^\top\boldsymbol{\lambda}_1\right)^\top\\ \frac{1}{2}\left(\mathbf{q}+\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top\mathbf{A}^\top\boldsymbol{\lambda}_1\right) & \mathbf{Q}\end{bmatrix}\succeq 0,\tag{46b}$$

$$\begin{bmatrix} s-\left(1-\frac{1}{\alpha}\right)t-\boldsymbol{\lambda}_2^\top\mathbf{b}-\left(\frac{1}{\alpha}\mathbf{x}\right)^\top\boldsymbol{\mu}+\boldsymbol{\lambda}_2^\top\mathbf{A}\boldsymbol{\mu} & \frac{1}{2}\left(\mathbf{q}+\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top\left(\mathbf{A}^\top\boldsymbol{\lambda}_2-\frac{1}{\alpha}\mathbf{x}\right)\right)^\top\\ \frac{1}{2}\left(\mathbf{q}+\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top\left(\mathbf{A}^\top\boldsymbol{\lambda}_2-\frac{1}{\alpha}\mathbf{x}\right)\right) & \mathbf{Q}\end{bmatrix}\succeq 0,\tag{46c}$$

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

Appendix C: Supplement to Section 3

C.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. □

C.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

$$\mathbf{X}\succeq\mathbf{Y}\Longrightarrow(\mathbf{V}\mathbf{a})^\top(\mathbf{X}-\mathbf{Y})(\mathbf{V}\mathbf{a})\geq 0,\ \forall\mathbf{a}\in\mathbb{R}^n$$

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

(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. □

C.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) = \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] \\
&= \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}}}[\boldsymbol{\xi}_{\mathrm{r}}] = \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}}}[\boldsymbol{\xi}_{\mathrm{r}}] \\
&= \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}}}[\boldsymbol{\xi}_{\mathrm{r}}] \\
&= \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}}}[\boldsymbol{\xi}_{\mathrm{r}}] \\
&= \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{47}$$

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

$$\begin{aligned}
&\mathbb{E}_{\mathbb{P}_\zeta}\left[(\boldsymbol{\zeta}-\boldsymbol{\mu})(\boldsymbol{\zeta}-\boldsymbol{\mu})^\top\right] = \mathbb{E}_{\mathbb{P}_{\mathrm{r}}}\left[\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\boldsymbol{\xi}_{\mathrm{r}}\boldsymbol{\xi}_{\mathrm{r}}^\top\mathbf{B}^\top\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{U}^\top\right] \\
\preceq &\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\gamma_2\mathbf{I}_{m_1}\mathbf{B}^\top\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{U}^\top = \gamma_2\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\mathbf{B}^\top\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{U}^\top \preceq \gamma_2\mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^\top = \gamma_2\boldsymbol{\Sigma},
\end{aligned} \tag{48}$$

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 $\mathcal{S}_\zeta \subseteq \mathcal{S}$, (47), and (48), 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

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

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

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

which demonstrates that the optimal value of Problem (10) is a lower bound for that of Problem (3) (i.e., Problem (2)).

(ii) For any $m_1 < m_2 \le m$, $\mathbf{B}_1 \in \mathbb{R}^{m\times m_1}$, 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

$$\mathcal{D}_{\zeta_i} = \left\{\mathbb{P}_{\zeta_i} \;\middle|\; \boldsymbol{\zeta}_i \sim \mathbb{P}_{\zeta_i},\ \boldsymbol{\zeta}_i = \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}_i\boldsymbol{\xi}_{\mathrm{r}_i} + \boldsymbol{\mu},\ \boldsymbol{\xi}_{\mathrm{r}_i} \sim \mathbb{P}_{\mathrm{r}_i} \in \mathcal{D}_{\mathrm{r}_i}\right\},\ \forall i \in [2], \tag{49}$$

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 (11)) 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 $\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}_1\boldsymbol{\xi}_{\mathrm{r}_1} = \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}_2\bar{\boldsymbol{\xi}}_{\mathrm{r}_2}$. 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 $\mathbb{E}[\bar{\boldsymbol{\xi}}_{\mathrm{r}_2}^\top]\mathbb{E}[\bar{\boldsymbol{\xi}}_{\mathrm{r}_2}] = \mathbb{E}[\boldsymbol{\xi}_{\mathrm{r}_1}^\top]\mathbb{E}[\boldsymbol{\xi}_{\mathrm{r}_1}] \le \gamma_1$ 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 (49) and $\boldsymbol{\zeta}_1 = \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}_2\bar{\boldsymbol{\xi}}_{\mathrm{r}_2} + \boldsymbol{\mu}$, 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

$$\max_{\mathbb{P}_{\zeta_1}\in\mathcal{D}_{\zeta_1}} \mathbb{E}_{\mathbb{P}_{\zeta_1}}\left[f\left(\mathbf{x},\boldsymbol{\zeta}_1\right)\right] \le \max_{\mathbb{P}_{\zeta_2}\in\mathcal{D}_{\zeta_2}} \mathbb{E}_{\mathbb{P}_{\zeta_2}}\left[f\left(\mathbf{x},\boldsymbol{\zeta}_2\right)\right]. \tag{50}$$

Together with the definitions of $\boldsymbol{\zeta}_i$ ($\forall i \in [2]$) and $\mathbf{B}_2$, inequality (50) leads to

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

Considering an optimal solution $\mathbf{B}_1^* \in \mathbb{R}^{m\times m_1}$ of Problem (10), 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

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

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

$$\min_{\mathbf{x}\in\mathcal{X}} \max_{\mathbb{P}_{\mathrm{r}_1}\in\mathcal{D}_{\mathrm{r}_1}} \mathbb{E}_{\mathbb{P}_{\zeta_1}} \left[f\left(\mathbf{x}, \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}_1^*\boldsymbol{\xi}_{\mathrm{r}_1} + \boldsymbol{\mu}\right)\right] \le \min_{\mathbf{x}\in\mathcal{X}} \max_{\mathbb{P}_{\mathrm{r}_2}\in\mathcal{D}_{\mathrm{r}_2}} \mathbb{E}_{\mathbb{P}_{\zeta_2}} \left[f\left(\mathbf{x}, \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}[\mathbf{B}_1^*, \mathbf{C}]\boldsymbol{\xi}_{\mathrm{r}_2} + \boldsymbol{\mu}\right)\right]. \quad (51)$$

It follows that

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

where the first inequality holds by (51) and the second inequality holds because $[\mathbf{B}_1^*, \mathbf{C}] \in \mathcal{B}_{m_2}$. That is, the optimal value of Problem (10) 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 $\Theta_{\mathrm{L}}(m) \le \Theta_{\mathrm{M}}(m)$ by the conclusion (i). Second, when $\mathbf{B} = \mathbf{I}_m$, Problem (10) becomes Problem (3). Because $\mathbf{B} = \mathbf{I}_m$ is a feasible solution of Problem (10), it follows that $\Theta_{\mathrm{L}}(m) \ge \Theta_{\mathrm{M}}(m)$. Therefore, we have $\Theta_{\mathrm{L}}(m) = \Theta_{\mathrm{M}}(m)$. $\square$

C.4. Theorem 7

THEOREM 7. *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* $m_1 \ge K$, *there exists a feasible solution* $\mathbf{B}^\dagger = [\mathbf{V}, \mathbf{C}]$ *in Problem* (12) *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* (13) *such that the corresponding objective value equals the optimal value of Problem* (4), $\Theta_{\mathrm{M}}(m)$.

Proof. 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 (12) and (13) 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 (12) and (13). Clearly, this solution satisfies constraints (13c). 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

$$S_k\left(\mathbf{V}\mathbf{Y}_{11}\mathbf{V}^\top\right) \succeq \frac{1}{4}\left(\mathbf{V}\delta + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top\left(\mathbf{A}^\top\boldsymbol{\lambda}_k^* - y_k(\mathbf{x}^*)\right)\right)\left(\mathbf{V}\delta + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top\left(\mathbf{A}^\top\boldsymbol{\lambda}_k^* - y_k(\mathbf{x}^*)\right)\right)^\top. \quad (52)$$

From (52), for any $k \in [K]$, we have the following inequality holds by Lemma 2:

$$S_k\left([\mathbf{V}, \mathbf{C}]^\top\mathbf{V}\mathbf{Y}_{11}\mathbf{V}^\top[\mathbf{V}, \mathbf{C}]\right)$$

$$\succeq \frac{1}{4}[\mathbf{V},\mathbf{C}]^\top \left(\mathbf{V}\delta + \left(\mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}\right)^\top (\mathbf{A}^\top \boldsymbol{\lambda}_k^* - y_k(\mathbf{x}^*))\right) \left(\mathbf{V}\delta + \left(\mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}\right)^\top (\mathbf{A}^\top \boldsymbol{\lambda}_k^* - y_k(\mathbf{x}^*))\right)^\top [\mathbf{V},\mathbf{C}],$$

which is equivalent to

$$S_k\mathbf{Q}_\mathrm{r}^\dagger \succeq \frac{1}{4}\left(\mathbf{q}_\mathrm{r}^\dagger + \left(\mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{B}^\dagger\right)^\top (\mathbf{A}^\top \boldsymbol{\lambda}_k^* - y_k(\mathbf{x}^*))\right) \left(\mathbf{q}_\mathrm{r}^\dagger + \left(\mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{B}^\dagger\right)^\top (\mathbf{A}^\top \boldsymbol{\lambda}_k^* - y_k(\mathbf{x}^*))\right)^\top \tag{53}$$

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, (53) 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 (13b) and thus it is a feasible solution of Problems (12) and (13).

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

$$\begin{aligned}
s^\dagger + \gamma_2\mathbf{I}_{m_1} \bullet \mathbf{Q}_\mathrm{r}^\dagger + \sqrt{\gamma_1}\left\|\mathbf{q}_\mathrm{r}^\dagger\right\|_2 &= s^* + \gamma_2\mathbf{I}_K \bullet \mathbf{Y}_{11} + \sqrt{\gamma_1}\|\delta\|_2 \\
&= s^* + \gamma_2\mathbf{I}_K \bullet (\mathbf{Y}_{11}\mathbf{V}^\top\mathbf{V}) + \sqrt{\gamma_1}\|\delta\|_2 \\
&= s^* + \gamma_2\mathbf{I}_m \bullet (\mathbf{V}\mathbf{Y}_{11}\mathbf{V}^\top) + \sqrt{\gamma_1}\|\delta\|_2 \\
&= s^* + \gamma_2\mathbf{I}_m \bullet \mathbf{Q}' + \sqrt{\gamma_1}\|\delta\|_2 \\
&= s^* + \gamma_2\mathbf{I}_m \bullet \mathbf{Q}' + \sqrt{\gamma_1}\sqrt{\delta^\top\delta} \\
&= s^* + \gamma_2\mathbf{I}_m \bullet \mathbf{Q}' + \sqrt{\gamma_1}\sqrt{\delta^\top\mathbf{V}^\top\mathbf{V}\delta} \\
&= s^* + \gamma_2\mathbf{I}_m \bullet \mathbf{Q}' + \sqrt{\gamma_1}\sqrt{\mathbf{q}'^\top\mathbf{q}'} \\
&= s^* + \gamma_2\mathbf{I}_m \bullet \mathbf{Q}' + \sqrt{\gamma_1}\|\mathbf{q}'\|_2 \\
&= \Theta_\mathrm{M}(m),
\end{aligned} \tag{54}$$

where the first equality holds by the construction of $(\mathbf{x}^\dagger, s^\dagger, \hat{\boldsymbol{\lambda}}^\dagger, \mathbf{q}_\mathrm{r}^\dagger, \mathbf{Q}_\mathrm{r}^\dagger, \mathbf{B}^\dagger)$, 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 $\mathbf{q}' = \mathbf{V}\delta$. □

Theorem 7 shows that, when $m_1 \geq K$, we can always find a feasible solution of Problems (12) and (13) such that the corresponding objective value is equal to the optimal value of the original Problem (4). More importantly, the SDP constraints in Problem (12) 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).

It is important to note that, although the constructed feasible solution of Problem (12) has an objective value of $\Theta_\mathrm{M}(m)$ and Problem (12) serves as an outer approximation for Problem (4), this does not imply that the constructed feasible solution is optimal for Problem (12). The primary reason is that Problem (12) is a max-min problem. For any $m_1 \leq m$, we would like to have an optimal $\mathbf{B}^* \in \mathcal{B}_{m_1}$ such that the optimal value of the inner minimization problem given this $\mathbf{B}^*$ can be maximized. By Theorem 1, we have $\underline{\Theta}(m_1, \mathbf{B}^*) = \Theta_{\mathrm{L}}(m_1) \leq \Theta_{\mathrm{M}}(m)$ for any $m_1 \leq m$.

According to Theorem 7, when $m_1 \geq K$, we can construct a feasible solution $\mathbf{B}^\dagger = [\mathbf{V}, \mathbf{C}]$ of the outer maximization problem. Note that $\mathbf{B}^\dagger$ is not necessarily optimal for Problem (12). Because the outer problem is a maximization problem, we have $\underline{\Theta}(m_1, \mathbf{B}^\dagger) \leq \underline{\Theta}(m_1, \mathbf{B}^*) = \Theta_{\mathrm{L}}(m_1) \leq \Theta_{\mathrm{M}}(m)$.

Given this feasible $\mathbf{B}^\dagger$, we can construct a feasible solution of the inner minimization problem (13) such that the corresponding objective value (here denoted by ϕ) is equal to $\Theta_{\mathrm{M}}(m)$. Note that this constructed feasible solution is not necessarily optimal for Problem (12) or Problem (13). Because the inner problem is a minimization problem, we have $\underline{\Theta}(m_1, \mathbf{B}^\dagger) \leq \phi = \Theta_{\mathrm{M}}(m)$.

Now, we construct a feasible solution $\mathbf{B}^\dagger$ for the outer maximization problem and, given this $\mathbf{B}^\dagger$, we also construct a feasible solution for the inner minimization problem. Although the corresponding objective value ϕ is equal to $\Theta_{\mathrm{M}}(m)$ and Problem (12) is an outer approximation for the original problem, we cannot claim that this constructed feasible solution is optimal due to the max-min natrue. One key reason is that the constructed solution of the inner minimization problem (13) may not be optimal. Thus, $\underline{\Theta}(m_1, \mathbf{B}^\dagger) < \phi = \Theta_{\mathrm{M}}(m)$ may hold.

We used to conjecture that this constructed feasible solution is an optimal solution of Problems (12) and (13) such that the optimal value of Problem (12) equals that of Problem (4) when $m_1 \geq K$. Most numerical experiments (see Section 7) show this conjecture may be correct, but we find a counter-example (see Example 2 in Appendix C.5). Example 2 illustrates that the optimal value of Problem (13) with $m_1 = K$ and $\mathbf{B} = \mathbf{V}$ is strictly less than the optimal value of Problem (4), which means that the constructed feasible solution ($\mathbf{B} = \mathbf{V}$) is not optimal.

C.5. Counter-Example

Now we provide an example as follows to illustrate that the optimal value of Problem (13) with $m_1 = K$ and $\mathbf{B} = \mathbf{V}$ 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

$$\min_{\mathbf{x}\in\mathcal{X}, s, \mathbf{q}, \mathbf{Q}} \left\{ s + 2\mathbf{I}_m \bullet \mathbf{Q} + \|\mathbf{q}\|_2 \;\middle|\; \begin{bmatrix} s - \mathbf{x}^\top \mathbf{W}_k^\top \mathbf{1}_m & \frac{1}{2}(\mathbf{q} - \mathbf{W}_k\mathbf{x})^\top \\ \frac{1}{2}(\mathbf{q} - \mathbf{W}_k\mathbf{x}) & \mathbf{Q} \end{bmatrix} \succeq 0,\ \forall k \in [K] \right\}. \tag{55}$$

Solving Problem (55) 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 (13) becomes

$$\min_{\mathbf{x}\in\mathcal{X},s,\mathbf{q}_\mathrm{r},\mathbf{Q}_\mathrm{r}} \left\{ s + 2\mathbf{I}_{m_1} \bullet \mathbf{Q}_\mathrm{r} + \|\mathbf{q}_\mathrm{r}\|_2 \;\middle|\; \begin{bmatrix} s - \mathbf{x}^\top \mathbf{W}_k^\top \mathbf{1}_m & \frac{1}{2}\left(\mathbf{q}_\mathrm{r} - \mathbf{V}^\top \mathbf{W}_k \mathbf{x}\right)^\top \\ \frac{1}{2}\left(\mathbf{q}_\mathrm{r} - \mathbf{V}^\top \mathbf{W}_k \mathbf{x}\right) & \mathbf{Q}_\mathrm{r} \end{bmatrix} \succeq 0,\ \forall k \in [K] \right\}. \tag{56}$$

Solving Problem (56) gives the optimal value 5.1139 with $\mathbf{x} = [1,-7,1,1]^\top$. That is, the optimal value of Problem (13) with $\mathbf{B} = \mathbf{V}$ is strictly less than the optimal value of Problem (4).

C.6. Proof of Proposition 2

We consider the Lagrangian dual of the inner minimization part (i.e., Problem (13)) of Problem (12) as follows:

$$\max_{\substack{\begin{bmatrix} t_k & \mathbf{p}_k^\top \\ \mathbf{p}_k & \mathbf{P}_k \end{bmatrix} \succeq 0, \\ \forall k\in[K], \\ \mathbf{Z}\succeq 0}} \ \min_{\substack{\mathbf{x},s,\hat{\lambda}\geq 0, \\ \mathbf{q}_\mathrm{r},\mathbf{Q}_\mathrm{r}}} \mathcal{L}\left(\mathbf{x}, s, \hat{\lambda}, \mathbf{q}_\mathrm{r}, \mathbf{Q}_\mathrm{r}; \mathbf{Z}, t_k, \mathbf{p}_k, \mathbf{P}_k, \forall k \in [K]\right), \tag{57}$$

where the Lagrangian function

$$\begin{aligned}
&\mathcal{L}\left(\mathbf{x}, s, \hat{\lambda}, \mathbf{q}_\mathrm{r}, \mathbf{Q}_\mathrm{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}_\mathrm{r} + \sqrt{\gamma_1}\|\mathbf{q}_\mathrm{r}\|_2 - \mathbf{Z} \bullet \left(\sum_{i=1}^{n}(\mathbf{\Delta}_i x_i) + \mathbf{\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}_\mathrm{r} + \left(\mathbf{U}\mathbf{\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}_\mathrm{r} + \left(\mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{B}\right)^\top \left(\mathbf{A}^\top \boldsymbol{\lambda}_k - y_k(\mathbf{x})\right)\right) & \mathbf{Q}_\mathrm{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}\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{B}\right)^\top \mathbf{A}^\top\right) \boldsymbol{\lambda}_k + \sqrt{\gamma_1}\|\mathbf{q}_\mathrm{r}\|_2 - \sum_{k=1}^{K} \mathbf{p}_k^\top \mathbf{q}_\mathrm{r} \\
&+ \left(\gamma_2 \mathbf{I}_{m_1} - \sum_{k=1}^{K} \mathbf{P}_k\right) \bullet \mathbf{Q}_\mathrm{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}\mathbf{\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}\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{B}\right)^\top \mathbf{A}^\top\right) \boldsymbol{\lambda}_k + \sqrt{\gamma_1}\|\mathbf{q}_\mathrm{r}\|_2 - \sum_{k=1}^{K} \mathbf{p}_k^\top \mathbf{q}_\mathrm{r} \\
&+ \left(\gamma_2 \mathbf{I}_{m_1} - \sum_{k=1}^{K} \mathbf{P}_k\right) \bullet \mathbf{Q}_\mathrm{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}\mathbf{\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}\mathbf{\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}\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{B}\right)^\top \mathbf{A}^\top\right) \boldsymbol{\lambda}_k + \sqrt{\gamma_1}\|\mathbf{q}_\mathrm{r}\|_2 - \sum_{k=1}^{K} \mathbf{p}_k^\top \mathbf{q}_\mathrm{r}
\end{aligned}$$

$$+\left(\gamma_2\mathbf{I}_{m_1}-\sum_{k=1}^{K}\mathbf{P}_k\right)\bullet\mathbf{Q}_{\mathrm{r}}+\left(\sum_{k=1}^{K}\left(t_k\mathbf{w}_k^0+\left(t_k\boldsymbol{\mu}^\top+\mathbf{p}_k^\top\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\right)^\top\right)\mathbf{W}_k\right)-\sum_{i=1}^{\tau}\sum_{j=1}^{\tau}z_{ij}\mathbf{a}_{ij}\right)\mathbf{x}$$
$$+\sum_{k=1}^{K}\left(t_kd_k^0+\left(t_k\boldsymbol{\mu}^\top+\mathbf{p}_k^\top\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\right)^\top\right)\mathbf{d}_k\right)-\sum_{i=1}^{\tau}\sum_{j=1}^{\tau}z_{ij}a_{ij}^0.$$

To present the objective value of the inner minimization problem of (57) from going to negative infinity, we require

$$1-\sum_{k=1}^{K}t_k=0,\ \sqrt{\gamma_1}-\left\|\sum_{k=1}^{K}\mathbf{p}_k\right\|_2\geq 0,\ \gamma_2\mathbf{I}_{m_1}-\sum_{k=1}^{K}\mathbf{P}_k=0, \tag{58a}$$
$$t_k\left(\mathbf{A}\boldsymbol{\mu}-\mathbf{b}\right)^\top+\mathbf{p}_k^\top\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\right)^\top\mathbf{A}^\top\leq 0,\ \forall k\in[K], \tag{58b}$$
$$\sum_{k=1}^{K}\left(t_k\mathbf{w}_k^0+\left(t_k\boldsymbol{\mu}^\top+\mathbf{p}_k^\top\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\right)^\top\right)\mathbf{W}_k\right)-\sum_{i=1}^{\tau}\sum_{j=1}^{\tau}z_{ij}\mathbf{a}_{ij}=0, \tag{58c}$$
$$\begin{bmatrix} t_k & \mathbf{p}_k^\top\\ \mathbf{p}_k & \mathbf{P}_k\end{bmatrix}\succeq 0,\ \forall k\in[K],\ \mathbf{Z}\succeq 0. \tag{58d}$$

Then, the dual problem of Problem (13) can be described as follows:

$$\begin{aligned}\max_{t_k,\mathbf{p}_k,\mathbf{P}_k,\forall k\in[K],\mathbf{Z}}\ & \sum_{k=1}^{K}\left(t_kd_k^0+\left(t_k\boldsymbol{\mu}^\top+\mathbf{p}_k^\top\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\right)^\top\right)\mathbf{d}_k\right)-\sum_{i=1}^{\tau}\sum_{j=1}^{\tau}z_{ij}a_{ij}^0\\ \text{s.t.}\quad & (58a)-(58d).\end{aligned} \tag{59}$$

By integrating the outer maximization part of Problem (12) and Problem (59), we obtain the bilinear SDP problem (25). Now we would like to prove the strong duality between Problem (13) and Problem (59); that is, these two problems share the same optimal value, which further shows that Problem (12) has the same optimal value as Problem (25). To that end, we find an interior point of Problem (13).

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 (13b) 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

$$\begin{aligned}&\mathbf{Q}_{\mathrm{r}}'-\frac{\left(\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\right)^\top\left(\mathbf{A}^\top\mathbf{1}_l-y_k(\mathbf{x}')\right)\right)\left(\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\right)^\top\left(\mathbf{A}^\top\mathbf{1}_l-y_k(\mathbf{x}')\right)\right)^\top}{4\left(s'-y_k^0(\mathbf{x}')-\mathbf{1}_l^\top b-y_k(\mathbf{x}')^\top\boldsymbol{\mu}+\mathbf{1}_l^\top\mathbf{A}\boldsymbol{\mu}\right)}\\ &=\sum_{\forall k'\in[K]:k'\neq k}\frac{\left(\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\right)^\top\left(\mathbf{A}^\top\mathbf{1}_l-y_{k'}(\mathbf{x}')\right)\right)\left(\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\right)^\top\left(\mathbf{A}^\top\mathbf{1}_l-y_{k'}(\mathbf{x}')\right)\right)^\top}{4\left(s'-y_{k'}^0(\mathbf{x}')-\mathbf{1}_l^\top b-y_{k'}(\mathbf{x}')^\top\boldsymbol{\mu}+\mathbf{1}_l^\top\mathbf{A}\boldsymbol{\mu}\right)}+\mathbf{I}_{m_1}\succ 0,\end{aligned} \tag{60}$$

where $s' - y^0_{k'}(\mathbf{x}') - \mathbf{1}_l^\top b - y_{k'}(\mathbf{x}')^\top \boldsymbol{\mu} + \mathbf{1}_l^\top \mathbf{A}\boldsymbol{\mu} > 0$ by the construction of s'. By Schur complement, (60) is equivalent to

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

Thus, $(\mathbf{x}', s', \hat{\boldsymbol{\lambda}}', \mathbf{q}'_\mathrm{r}, \mathbf{Q}'_\mathrm{r})$ is an interior point of Problem (13) and the strong duality between Problem (13) and Problem (59) holds. □

Appendix D: Supplement to Section 4

D.1. Proof of Theorem 3

(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 (26) 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 (26), 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 (28). 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

$$\max_{\mathbb{P}_\mathrm{I} \in \mathcal{D}_\mathrm{U}} \mathbb{E}_{\mathbb{P}_\mathrm{I}}\left[f\left(\mathbf{x}^*, \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_\mathrm{I} + \boldsymbol{\mu}\right)\right] \geq \max_{\mathbb{P}_\mathrm{I} \in \mathcal{D}'_\mathrm{U}} \mathbb{E}_{\mathbb{P}_\mathrm{I}}\left[f\left(\mathbf{x}^*, \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_\mathrm{I} + \boldsymbol{\mu}\right)\right]. \tag{61}$$

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 $\min\limits_{\mathbf{B}\in\mathcal{B}_{m_2}} \min\limits_{\mathbf{x}\in\mathcal{X}} \max\limits_{\mathbb{P}_{\mathrm{I}}\in\mathcal{D}'_{\mathrm{U}}} \mathbb{E}_{\mathbb{P}_{\mathrm{I}}}[f(\mathbf{x}, \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}} + \boldsymbol{\mu})]$. Then, we have

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

where the first inequality holds because $(\mathbf{B}', \mathbf{x}^*)$ is a feasible solution of the problem $\min\limits_{\mathbf{B}\in\mathcal{B}_{m_2}} \min\limits_{\mathbf{x}\in\mathcal{X}} \max\limits_{\mathbb{P}_{\mathrm{I}}\in\mathcal{D}'_{\mathrm{U}}} \mathbb{E}_{\mathbb{P}_{\mathrm{I}}}[f(\mathbf{x}, \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}} + \boldsymbol{\mu})]$, the second inequality holds by (61), and the second equality holds because $(\mathbf{B}^*, \mathbf{x}^*)$ is an optimal solution of Problem (26). That is, the optimal value of Problem (26) 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 (26) becomes Problem (3). Because $\mathbf{B} = \mathbf{I}_m$ is a feasible solution of Problem (26), it follows that $\Theta_{\mathrm{U}}(m) \leq \Theta_{\mathrm{M}}(m)$. Therefore, we have $\Theta_{\mathrm{U}}(m) = \Theta_{\mathrm{M}}(m)$. $\square$

D.2. Proof of Proposition 3

First, by Theorem 3 in Cheramin et al. (2022), Problem (27) has the same optimal value as the following problem:

$$\min_{\mathbf{x},s,\mathbf{q},\mathbf{Q}_{\mathrm{r}}} \quad s + \gamma_2 \mathbf{I}_{m_1} \bullet \mathbf{Q}_{\mathrm{r}} + \sqrt{\gamma_1}\,\|\mathbf{q}\|_2 \tag{62a}$$

$$\text{s.t.} \quad s \geq f\left(\mathbf{x}, \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}} + \boldsymbol{\mu}\right) - \boldsymbol{\xi}_{\mathrm{I}}^\top \mathbf{B}\mathbf{Q}_{\mathrm{r}}\mathbf{B}^\top\boldsymbol{\xi}_{\mathrm{I}} - \mathbf{q}^\top\boldsymbol{\xi}_{\mathrm{I}},\ \forall \boldsymbol{\xi}_{\mathrm{I}} \in \mathcal{S}_{\mathrm{I}}, \tag{62b}$$

$$\mathbf{Q}_{\mathrm{r}} \succeq 0,\ \mathbf{x} \in \mathcal{X},\ \mathbf{Q}_{\mathrm{r}} \in \mathbb{R}^{m_1\times m_1},\ \mathbf{q} \in \mathbb{R}^m. \tag{62c}$$

Next, we apply the strong duality theorem to constraints (62b). We define

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

As function $f(\mathbf{x}, \boldsymbol{\xi})$ is piecewise linear convex, we can reformulate (62b) 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

$$\min_{\mathbf{A}\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}}+\boldsymbol{\mu}\right)\leq\mathbf{b},\ \boldsymbol{\xi}_{\mathrm{I}}\in\mathbb{R}^m} \quad g_k(\boldsymbol{\xi}_{\mathrm{I}}) \geq 0,\ \forall k \in [K]. \tag{63}$$

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 (63) 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

$$\begin{aligned}\exists \boldsymbol{\lambda}_k \geq 0: s + \boldsymbol{\xi}_{\mathrm{I}}^\top \mathbf{B}\mathbf{Q}_{\mathrm{r}}\mathbf{B}^\top \boldsymbol{\xi}_{\mathrm{I}} + \mathbf{q}^\top \boldsymbol{\xi}_{\mathrm{I}} - y_k^0(\mathbf{x}) - y_k(\mathbf{x})^\top \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}} + \boldsymbol{\mu}\right) \\ + \boldsymbol{\lambda}_k^\top \left(\mathbf{A}\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}} + \boldsymbol{\mu}\right) - \mathbf{b}\right) \geq 0,\ \forall \boldsymbol{\xi}_{\mathrm{I}} \in \mathbb{R}^m,\ \forall k \in [K].\end{aligned} \tag{64}$$

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 (64) become

$$\begin{aligned}\exists \boldsymbol{\lambda}_k \geq 0: s + \boldsymbol{\xi}_1^\top \mathbf{Q}_{\mathrm{r}}\boldsymbol{\xi}_1 + \mathbf{q}^\top\left(\mathbf{B}\boldsymbol{\xi}_1 + \bar{\mathbf{B}}\boldsymbol{\xi}_2\right) - y_k^0(\mathbf{x}) - y_k(\mathbf{x})^\top \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\left(\mathbf{B}\boldsymbol{\xi}_1 + \bar{\mathbf{B}}\boldsymbol{\xi}_2\right) + \boldsymbol{\mu}\right) \\ + \boldsymbol{\lambda}_k^\top \left(\mathbf{A}\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\left(\mathbf{B}\boldsymbol{\xi}_1 + \bar{\mathbf{B}}\boldsymbol{\xi}_2\right) + \boldsymbol{\mu}\right) - \mathbf{b}\right) \geq 0,\ \forall \boldsymbol{\xi}_1 \in \mathbb{R}^{m_1}, \boldsymbol{\xi}_2 \in \mathbb{R}^{m-m_1},\ \forall k \in [K].\end{aligned} \tag{65}$$

We further define

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

Thus, we have

$$\begin{aligned}(65) \iff & \exists \boldsymbol{\lambda}_k \geq 0:\ \left(1, \boldsymbol{\xi}_1^\top\right)\mathbf{Z}_k\left(1, \boldsymbol{\xi}_1^\top\right)^\top + \boldsymbol{\xi}_2^\top\left(\bar{\mathbf{B}}^\top\mathbf{q} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\bar{\mathbf{B}}\right)^\top\left(\mathbf{A}^\top\boldsymbol{\lambda}_k - y_k(\mathbf{x})\right)\right) \geq 0, \\ & \qquad \forall \boldsymbol{\xi}_1 \in \mathbb{R}^{m_1}, \boldsymbol{\xi}_2 \in \mathbb{R}^{m-m_1},\ \forall k \in [K]. \\ \iff & \exists \boldsymbol{\lambda}_k \geq 0:\ \left(1, \boldsymbol{\xi}_1^\top\right)\mathbf{Z}_k\left(1, \boldsymbol{\xi}_1^\top\right)^\top \geq 0,\ \forall \boldsymbol{\xi}_1 \in \mathbb{R}^{m_1},\ \forall k \in [K]; \qquad (66) \\ & \bar{\mathbf{B}}^\top\mathbf{q} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\bar{\mathbf{B}}\right)^\top\left(\mathbf{A}^\top\boldsymbol{\lambda}_k - y_k(\mathbf{x})\right) = 0,\ \forall k \in [K]. \\ \iff & \exists \boldsymbol{\lambda}_k \geq 0:\ \mathbf{Z}_k \succeq 0,\ \bar{\mathbf{B}}^\top\mathbf{q} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\bar{\mathbf{B}}\right)^\top\left(\mathbf{A}^\top\boldsymbol{\lambda}_k - y_k(\mathbf{x})\right) = 0,\ \forall k \in [K]. \\ \iff & \exists \boldsymbol{\lambda}_k \geq 0:\ \mathbf{Z}_k \succeq 0,\ \bar{\mathbf{B}}^\top\left(\mathbf{q} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top\left(\mathbf{A}^\top\boldsymbol{\lambda}_k - y_k(\mathbf{x})\right)\right) = 0,\ \forall k \in [K]. \qquad (67)\end{aligned}$$

$$\iff \exists \boldsymbol{\lambda}_k \geq 0, \mathbf{u}_k \in \mathbb{R}^{m_1}: \ \mathbf{Z}_k \succeq 0, \ \mathbf{q} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top \left(\mathbf{A}^\top \boldsymbol{\lambda}_k - y_k(\mathbf{x})\right) = \mathbf{B}\mathbf{u}_k, \ \forall k \in [K]. \tag{68}$$

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 (66). Therefore, $\Longrightarrow$ holds. For the fifth equivalence, (67) 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 $\mathbf{u}_k \in \mathbb{R}^{m_1}$ such that $\mathbf{q} + (\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}})^\top(\mathbf{A}^\top \boldsymbol{\lambda}_k - y_k(\mathbf{x})) = \mathbf{B}\mathbf{u}_k$ for any $k \in [K]$. Meanwhile, because $\mathbf{B}^\top \mathbf{B} = \mathbf{I}_{m_1}$, we have

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

Finally, we obtain Problem (29) by replacing constraints (62b) with (68) and replacing $\mathbf{B}^\top \mathbf{q} + (\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B})^\top(\mathbf{A}^\top \boldsymbol{\lambda}_k - y_k(\mathbf{x}))$ with $\mathbf{u}_k$. □

D.3. Proof of Theorem 4

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 (26) 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 (26). Clearly, this solution satisfies constraints (29d)–(29e). By the construction of the solution, for any $k \in [K]$, we further have

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

where the first equality holds by the construction of $\mathbf{q}^\dagger$, the second equality holds by (16), and the third equality holds by the construction of $\hat{\mathbf{u}}_k^\dagger$. Thus, this solution satisfies constraints (29c). 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

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

which, by Schur complement, is equivalent to

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

$$= \left(\mathbf{V}\mathbf{u}_k^{\dagger}\right)\left(\mathbf{V}\mathbf{u}_k^{\dagger}\right)^{\top}, \tag{69}$$

where the equality holds by (16) and the construction of $\hat{\mathbf{u}}_k^{\dagger}$. From (69), 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

$$4S_k\mathbf{Y}_{11} \succeq \mathbf{u}_k^{\dagger}\mathbf{u}_k^{\dagger\top} \tag{70}$$

because $\mathbf{V}^{\top}\mathbf{V} = \mathbf{I}_K$. By Schur complement, for any $k \in [K]$, (70) 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 (29b) and thus it is a feasible solution of Problem (26).

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 (26). The objective value corresponding to this solution is

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

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 (26).

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 3 and $\Theta_{\mathrm{U}}(m_1) \le \Theta_{\mathrm{U}}(K) = \Theta_{\mathrm{M}}(m)$ by the conclusion (ii) in Theorem 3. It follows that $\Theta_{\mathrm{U}}(m_1) = \Theta_{\mathrm{M}}(m)$. □

D.4. Proof of Proposition 4

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

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

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 $\operatorname{rank}(\mathbf{X}) \le m_1$, 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. □

Appendix E: Supplement to Section 5

E.1. Proof of Theorem 5

By dualizing the inner maximization problem of Problem (30) and integrating it with the outer minimization operators, we first obtain the following formulation:

$$\min_{\substack{\mathbf{x},s,\mathbf{q},\mathbf{Q}_r',\mathbf{Q}_r''\\ \mathbf{B}_1,\mathbf{B}_2}} \quad s + \gamma_2\mathbf{I}_{m_1} \bullet \mathbf{Q}_r' + \sqrt{\gamma_1}\,\|\mathbf{q}\|_2 \tag{71a}$$

$$\text{s.t.} \quad s \ge f\left(\mathbf{x}, \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}} + \boldsymbol{\mu}\right) - \boldsymbol{\xi}_{\mathrm{I}}^\top\mathbf{B}_1\mathbf{Q}_r'\mathbf{B}_1^\top\boldsymbol{\xi}_{\mathrm{I}} - \boldsymbol{\xi}_{\mathrm{I}}^\top\mathbf{B}_2\mathbf{Q}_r''\mathbf{B}_2^\top\boldsymbol{\xi}_{\mathrm{I}} - \mathbf{q}^\top\boldsymbol{\xi}_{\mathrm{I}},\ \forall\boldsymbol{\xi}_{\mathrm{I}} \in \mathcal{S}_{\mathrm{I}}, \tag{71b}$$

$$\mathbf{Q}_r' \succeq 0,\ \mathbf{Q}_r'' \succeq 0,\ \mathbf{x} \in \mathcal{X},\ \mathbf{Q}_r' \in \mathbb{R}^{m_1\times m_1},\ \mathbf{Q}_r'' \in \mathbb{R}^{(K-m_1)\times(K-m_1)},\ \mathbf{q} \in \mathbb{R}^m, \tag{71c}$$

$$\mathbf{B}_1 \in \mathbb{R}^{m\times m_1}, \mathbf{B}_2 \in \mathbb{R}^{m\times(K-m_1)},\ [\mathbf{B}_1, \mathbf{B}_2]^\top[\mathbf{B}_1, \mathbf{B}_2] = \mathbf{I}_K. \tag{71d}$$

Next, we apply the strong duality theorem to constraints (71b). We define

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

As function $f(\mathbf{x}, \boldsymbol{\xi})$ is piecewise linear convex, we can reformulate (71b) as

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

which is equivalent to

$$\min_{\mathbf{A}\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}}+\boldsymbol{\mu}\right)\le\mathbf{b},\ \boldsymbol{\xi}_{\mathrm{I}}\in\mathbb{R}^m} \quad g_k(\boldsymbol{\xi}_{\mathrm{I}}) \ge 0,\ \forall k \in [K]. \tag{72}$$

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

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

where $\boldsymbol{\lambda}_k \in \mathbb{R}^l$. Because there exists an interior point for the primal problem, the strong duality holds. Thus, constraints (72) 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

$$\begin{aligned}\exists \boldsymbol{\lambda}_k \geq 0: s + \boldsymbol{\xi}_{\mathrm{I}}^\top \mathbf{B}_1 \mathbf{Q}_{\mathrm{r}}' \mathbf{B}_1^\top \boldsymbol{\xi}_{\mathrm{I}} + \boldsymbol{\xi}_{\mathrm{I}}^\top \mathbf{B}_2 \mathbf{Q}_{\mathrm{r}}'' \mathbf{B}_2^\top \boldsymbol{\xi}_{\mathrm{I}} + \mathbf{q}^\top \boldsymbol{\xi}_{\mathrm{I}} - y_k^0(\mathbf{x}) - y_k(\mathbf{x})^\top \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}} + \boldsymbol{\mu}\right) \\ + \boldsymbol{\lambda}_k^\top \left(\mathbf{A}\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}} + \boldsymbol{\mu}\right) - \mathbf{b}\right) \geq 0,\ \forall \boldsymbol{\xi}_{\mathrm{I}} \in \mathbb{R}^m,\ \forall k \in [K].\end{aligned} \tag{73}$$

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 (73) become

$$\begin{aligned}\exists \boldsymbol{\lambda}_k \geq 0: s + \boldsymbol{\xi}_1^\top \mathbf{Q}_{\mathrm{r}}' \boldsymbol{\xi}_1 + \boldsymbol{\xi}_2^\top \mathbf{Q}_{\mathrm{r}}'' \boldsymbol{\xi}_2 + \mathbf{q}^\top (\mathbf{B}_1 \boldsymbol{\xi}_1 + \mathbf{B}_2 \boldsymbol{\xi}_2 + \bar{\mathbf{B}} \boldsymbol{\xi}_3) - y_k^0(\mathbf{x}) \\ - y_k(\mathbf{x})^\top \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}} (\mathbf{B}_1 \boldsymbol{\xi}_1 + \mathbf{B}_2 \boldsymbol{\xi}_2 + \bar{\mathbf{B}} \boldsymbol{\xi}_3) + \boldsymbol{\mu}\right) \\ + \boldsymbol{\lambda}_k^\top \left(\mathbf{A}\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}} (\mathbf{B}_1 \boldsymbol{\xi}_1 + \mathbf{B}_2 \boldsymbol{\xi}_2 + \bar{\mathbf{B}} \boldsymbol{\xi}_3) + \boldsymbol{\mu}\right) - \mathbf{b}\right) \geq 0, \\ \forall \boldsymbol{\xi}_1 \in \mathbb{R}^{m_1}, \boldsymbol{\xi}_2 \in \mathbb{R}^{K-m_1}, \boldsymbol{\xi}_3 \in \mathbb{R}^{m-K},\ \forall k \in [K].\end{aligned} \tag{74}$$

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

$$\begin{aligned}(74) \iff & \exists \boldsymbol{\lambda}_k \geq 0:\ \left(1, \boldsymbol{\xi}_1^\top, \boldsymbol{\xi}_2^\top\right) \mathbf{Z}_k \left(1, \boldsymbol{\xi}_1^\top, \boldsymbol{\xi}_2^\top\right)^\top + \boldsymbol{\xi}_3^\top \left(\bar{\mathbf{B}}^\top \mathbf{q} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}} \bar{\mathbf{B}}\right)^\top (\mathbf{A}^\top \boldsymbol{\lambda}_k - y_k(\mathbf{x}))\right) \geq 0, \\ & \qquad \forall \boldsymbol{\xi}_1 \in \mathbb{R}^{m_1}, \boldsymbol{\xi}_2 \in \mathbb{R}^{K-m_1}, \boldsymbol{\xi}_3 \in \mathbb{R}^{m-K},\ \forall k \in [K]. \\ \iff & \exists \boldsymbol{\lambda}_k \geq 0:\ \left(1, \boldsymbol{\xi}_1^\top, \boldsymbol{\xi}_2^\top\right) \mathbf{Z}_k \left(1, \boldsymbol{\xi}_1^\top, \boldsymbol{\xi}_2^\top\right)^\top \geq 0,\ \forall \boldsymbol{\xi}_1 \in \mathbb{R}^{m_1}, \boldsymbol{\xi}_2 \in \mathbb{R}^{K-m_1},\ \forall k \in [K]; \qquad (75) \\ & \qquad \bar{\mathbf{B}}^\top \mathbf{q} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}} \bar{\mathbf{B}}\right)^\top (\mathbf{A}^\top \boldsymbol{\lambda}_k - y_k(\mathbf{x})) = 0,\ \forall k \in [K]. \\ \iff & \exists \boldsymbol{\lambda}_k \geq 0:\ \mathbf{Z}_k \succeq 0,\ \bar{\mathbf{B}}^\top \mathbf{q} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}} \bar{\mathbf{B}}\right)^\top (\mathbf{A}^\top \boldsymbol{\lambda}_k - y_k(\mathbf{x})) = 0,\ \forall k \in [K]. \\ \iff & \exists \boldsymbol{\lambda}_k \geq 0:\ \mathbf{Z}_k \succeq 0,\ \bar{\mathbf{B}}^\top \left(\mathbf{q} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top (\mathbf{A}^\top \boldsymbol{\lambda}_k - y_k(\mathbf{x}))\right) = 0,\ \forall k \in [K]. \qquad (76)\end{aligned}$$

$$\iff \exists \boldsymbol{\lambda}_k \geq 0, \mathbf{u}_k' \in \mathbb{R}^{m_1}, \mathbf{u}_k'' \in \mathbb{R}^{K-m_1}:$$
$$\mathbf{Z}_k \succeq 0,\ \mathbf{q} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top \left(\mathbf{A}^\top \boldsymbol{\lambda}_k - y_k(\mathbf{x})\right) = \mathbf{B}_1\mathbf{u}_k' + \mathbf{B}_2\mathbf{u}_k'',\ \forall k \in [K]. \tag{77}$$

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 (75). Therefore, $\Longrightarrow$ holds. For the fifth equivalence, (76) 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 (71b) with (77), we obtain the following problem:

$$\min_{\substack{\mathbf{x}, s, \hat{\boldsymbol{\lambda}}, \mathbf{q}, \\ \mathbf{Q}_\mathrm{r}', \mathbf{Q}_\mathrm{r}'', \hat{\mathbf{u}}', \hat{\mathbf{u}}'', \\ \mathbf{B}_1, \mathbf{B}_2}} \quad s + \gamma_2 \mathbf{I}_{m_1} \bullet \mathbf{Q}_\mathrm{r}' + \sqrt{\gamma_1}\,\|\mathbf{q}\|_2 \tag{78a}$$
$$\text{s.t.} \quad \begin{bmatrix} s - y_k^0(\mathbf{x}) - \boldsymbol{\lambda}_k^\top\mathbf{b} - y_k(\mathbf{x})^\top\boldsymbol{\mu} + \boldsymbol{\lambda}_k^\top\mathbf{A}\boldsymbol{\mu} & \frac{1}{2}(\mathbf{u}_k')^\top & \frac{1}{2}(\mathbf{u}_k'')^\top \\ \frac{1}{2}\mathbf{u}_k' & \mathbf{Q}_\mathrm{r}' & \mathbf{0}_{m_1\times(K-m_1)} \\ \frac{1}{2}\mathbf{u}_k'' & \mathbf{0}_{(K-m_1)\times m_1} & \mathbf{Q}_\mathrm{r}'' \end{bmatrix} \succeq 0,\ \forall k \in [K], \tag{78b}$$
$$\mathbf{q} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top \left(\mathbf{A}^\top\boldsymbol{\lambda}_k - y_k(\mathbf{x})\right) = \mathbf{B}_1\mathbf{u}_k' + \mathbf{B}_2\mathbf{u}_k'',\ \forall k \in [K], \tag{78c}$$
$$\mathbf{x} \in \mathcal{X},\ \mathbf{q} \in \mathbb{R}^m,\ \mathbf{Q}_\mathrm{r}' \in \mathbb{R}^{m_1\times m_1},\ \mathbf{Q}_\mathrm{r}'' \in \mathbb{R}^{(K-m_1)\times(K-m_1)}, \tag{78d}$$
$$\mathbf{B}_1 \in \mathbb{R}^{m\times m_1}, \mathbf{B}_2 \in \mathbb{R}^{m\times(K-m_1)},\ [\mathbf{B}_1, \mathbf{B}_2]^\top[\mathbf{B}_1, \mathbf{B}_2] = \mathbf{I}_K, \tag{78e}$$
$$\hat{\boldsymbol{\lambda}} = \{\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_K\},\ \boldsymbol{\lambda}_k \in \mathbb{R}_+^l,\ \forall k \in [K], \tag{78f}$$
$$\hat{\mathbf{u}}' = \{\mathbf{u}_1', \ldots, \mathbf{u}_K'\},\ \mathbf{u}_k' \in \mathbb{R}^{m_1},\ \forall k \in [K], \tag{78g}$$
$$\hat{\mathbf{u}}'' = \{\mathbf{u}_1'', \ldots, \mathbf{u}_K''\},\ \mathbf{u}_k'' \in \mathbb{R}^{K-m_1},\ \forall k \in [K]. \tag{78h}$$

Note that the value of $\mathbf{Q}_\mathrm{r}''$ does not contribute to the objective function (78a). 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 (78b) become

$$\begin{bmatrix} s - y_k^0(\mathbf{x}) - \boldsymbol{\lambda}_k^\top\mathbf{b} - y_k(\mathbf{x})^\top\boldsymbol{\mu} + \boldsymbol{\lambda}_k^\top\mathbf{A}\boldsymbol{\mu} & \frac{1}{2}(\mathbf{u}_k')^\top \\ \frac{1}{2}\mathbf{u}_k' & \mathbf{Q}_\mathrm{r}' \end{bmatrix} \succeq 0,\ \forall k \in [K]. \tag{79}$$

By replacing (78b) with (79), we obtain the formulation of Problem (31).

Based on the formulation of Problem (31), 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 (26), 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 (26), we can obtain the exact formulation of Problem (31), thereby the three conclusions hold.

First, we rewrite constraints (29b)–(29c) in Problem (26) 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:

$$
\begin{aligned}
\min_{\substack{\mathbf{x}, s, \hat{\boldsymbol{\lambda}},\\ \mathbf{q}, \mathbf{Q}_{\mathrm{r}}, \hat{\mathbf{u}}', \hat{\mathbf{u}}'',\\ \mathbf{B}_1, \mathbf{B}_2}} \quad & s + \gamma_2 \mathbf{I}_K \bullet \mathbf{Q}_{\mathrm{r}} + \sqrt{\gamma_1}\,\|\mathbf{q}\|_2 && \text{(80a)}\\
\text{s.t.} \quad & \begin{bmatrix} s - y_k^0(\mathbf{x}) - \boldsymbol{\lambda}_k^\top \mathbf{b} - y_k(\mathbf{x})^\top \boldsymbol{\mu} + \boldsymbol{\lambda}_k^\top \mathbf{A}\boldsymbol{\mu} & \frac{1}{2}\left((\mathbf{u}_k')^\top, (\mathbf{u}_k'')^\top\right) \\ \frac{1}{2}\left((\mathbf{u}_k')^\top, (\mathbf{u}_k'')^\top\right)^\top & \mathbf{Q}_{\mathrm{r}} \end{bmatrix} \succeq 0,\ \forall k \in [K], && \text{(80b)}\\
& \mathbf{q} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top \left(\mathbf{A}^\top \boldsymbol{\lambda}_k - y_k(\mathbf{x})\right) = \mathbf{B}_1 \mathbf{u}_k' + \mathbf{B}_2 \mathbf{u}_k'',\ \forall k \in [K], && \text{(80c)}\\
& \mathbf{x} \in \mathcal{X},\ [\mathbf{B}_1, \mathbf{B}_2]^\top [\mathbf{B}_1, \mathbf{B}_2] = \mathbf{I}_K, && \text{(80d)}\\
& \mathbf{q} \in \mathbb{R}^m,\ \mathbf{Q}_{\mathrm{r}} \in \mathbb{R}^{K\times K},\ \mathbf{B}_1 \in \mathbb{R}^{m\times m_1}, \mathbf{B}_2 \in \mathbb{R}^{m\times (K-m_1)}, && \text{(80e)}\\
& \hat{\boldsymbol{\lambda}} = \{\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_K\},\ \boldsymbol{\lambda}_k \in \mathbb{R}_+^l,\ \forall k \in [K], && \text{(80f)}\\
& \hat{\mathbf{u}}' = \{\mathbf{u}_1', \ldots, \mathbf{u}_K'\},\ \mathbf{u}_k' \in \mathbb{R}^{m_1},\ \forall k \in [K], && \text{(80g)}\\
& \hat{\mathbf{u}}'' = \{\mathbf{u}_1'', \ldots, \mathbf{u}_K''\},\ \mathbf{u}_k'' \in \mathbb{R}^{K-m_1},\ \forall k \in [K]. && \text{(80h)}
\end{aligned}
$$

Second, we relax constraints (80b) into

$$
\begin{bmatrix} s - y_k^0(\mathbf{x}) - \boldsymbol{\lambda}_k^\top \mathbf{b} - y_k(\mathbf{x})^\top \boldsymbol{\mu} + \boldsymbol{\lambda}_k^\top \mathbf{A}\boldsymbol{\mu} & \frac{1}{2}(\mathbf{u}_k')^\top \\ \frac{1}{2}\mathbf{u}_k' & \mathbf{Q}_{\mathrm{r}}' \end{bmatrix} \succeq 0,\ \forall k \in [K], \tag{81}
$$

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 (81) to replace (80b), we obtain a relaxation and accordingly lower bound for Problem (80). In addition, we further reduce the optimal value of the relaxation by replacing $\mathbf{Q}_{\mathrm{r}}$ in the objective function (80a) with $\mathbf{Q}_{\mathrm{r}}'$. That is, we obtain a lower bound for the optimal value of Problem (26) with $m_1 = K$ (i.e., Problem (4)). After these two steps of relaxations, we obtain the exact formulation of Problem (31). Thus, we can conclude that Problem (31) is a relaxation of Problem (26) with $m_1 = K$. Therefore, by the conclusion in Theorem 4, we have

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

That is, the conclusion (i) holds.

For the conclusion (ii): For any $0 \le m_1 < m_2 \le K$, we can follow the above two steps of relaxations to relax Problem (80) 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 $\Theta_{L2}(m_1)$. Because all these problems are minimization problems, we have $\Theta_{L2}(m_1) \le \Theta_{L2}(m_2)$.

For the conclusion (iii): When $m_1 = K$, Problem (31) becomes Problem (26) with $m_1 = K$. Thus, by the conclusion in Theorem 4, we have

$$\Theta_{L2}(K) = \Theta_U(K) = \Theta_M(m). \quad \square$$

Appendix F: Supplement to Section 6

F.1. Motivation to Derive ADMM Algorithms

Several techniques, including the McCormick envelopes and spatial branch-and-bound, can handle bilinear constraints and solve bilinear SDP problems like Problems (25), (26), and (31). However, to the best of our knowledge, the currently available bilinear SDP solvers, such as PENLAB and BMIBNB (Löfberg 2004), are not yet fully mature and only succeed on relatively small and simple problems. Therefore, we derive ADMM algorithms to solve the three approximations efficiently. To illustrate this, we consider the multiproduct newsvendor problem (39) in Section 7 as an example to compare our ADMM algorithms with benchmark solvers with default settings in solving this problem, where no time limit is given.

First, we use the Mosek solver to solve the original high-dimensional SDP problem (4) and report the optimality gap and computational time. Note that the Mosek solver uses the interior-point algorithm to solve an SDP problem and terminates when the relative gap between the primal and dual objective values is no greater than 10^{-9}. Second, we use the BMIBNB solver (the state-of-the-art bilinear SDP solver, to the best of our knowledge) to solve the low-dimensional bilinear SDP problem (26) with $m_1 = K$. Note that the BMIBNB solver uses the spatial branch-and-bound algorithm to solve a bilinear SDP problem. In the BMIBNB solver, we set "Mosek" as the lower bound solver and "fmincon" as the upper bound solver. The default termination condition of the BMIBNB solver is that the relative gap between the lower and upper bounds is no greater than 0.01 or the maximum number of nodes in the branch-and-bound tree is greater than 100. Third, we use our derived ADMM algorithms to solve Problems (25), (26), and (31), respectively. For each of the three problems, we follow the instructions in Section 6.2 to recover the lower or upper bounds in the following two steps: (i) We use the ADMM algorithm to solve the approximation problem and obtain a near-optimal dimensionality reduction matrix $\mathbf{B}^{\text{ADMM}}$; (ii) Given this $\mathbf{B}^{\text{ADMM}}$, we solve a low-dimensional SDP problem to recover the lower or upper bound for the original optimal value. Kindly note that the reported time for our lower or upper bounds reflects the combined time for both steps. Other numerical setups are detailed in Section 7.

Table F2 Performance of Bilinear SDP Solver on the Newsvendor Problem

	Size (m)-Instance Id	100-1	100-2	100-3	200-1	200-2	200-3
Mosek	Optimality Gap (%)	0.00	0.00	0.00	0.00	0.00	0.00
	Time (secs)	12.32	11.34	13.36	342.00	432.57	368.04
BMIBNB	Optimality Gap (%)	4.31	3.89	3.65	4.25	3.83	3.85
	Time (secs)	3840.00	5552.10	6412.50	9885.50	9942.50	10643.00
ODR-LB	Gap1 (%)	0.07	0.07	0.09	0.00	0.00	0.00
	Time (secs)	0.58	0.72	0.83	0.42	0.91	0.88
ODR-RLB	Gap1 (%)	0.03	0.04	0.03	0.06	0.02	0.02
	Time (secs)	2.45	1.74	1.98	2.57	2.42	2.76
ODR-UB	Gap2 (%)	1.56	1.81	1.65	1.73	1.92	2.03
	Time (secs)	2.45	1.74	1.98	2.57	2.42	2.76

Table F2 shows the numerical results, where "Optimality Gap (%)" and "Time (secs)" are reported by the corresponding solver, and "Gap1 (%)" (resp. "Gap2 (%)") represents the relative gap in percentage between a lower (resp. an upper) bound and the optimal value provided by the Mosek solver. Note that when $m > 200$, the BMIBNB solver fails to obtain any feasible solution within three hours, and thus, these cases are not reported in Table F2. For the reported six instances, the Mosek solver can solve them to optimality, and the BMIBNB solver terminates when reaching the limit of the maximum number of nodes. These numerical results demonstrate that using the bilinear SDP solver to solve the low-dimensional bilinear SDP reformulation of the high-dimensional SDP leads to poor performance. It is not surprising. The Mosek solver employs the interior-point method to solve the high-dimensional SDP problem, leading to a polynomial-time algorithm. In contrast, the BMIBNB solver uses spatial branch-and-bound to solve the low-dimensional bilinear SDP problem, leading to an exponential-time algorithm. More importantly, the bilinear SDP problem is nonconvex, significantly increasing the computational challenge. Finally, these numerical results demonstrate that the ADMM algorithms bring significant improvement in terms of optimality gap and computational time.

F.2. The ADMM Algorithm for Problem (26)

Note that Algorithm 1 serves as a unified ADMM algorithm for all the three proposed approximations, i.e., Problems (25), (26), and (31). Here, we take Problem (26) as an example to demonstrate corresponding specific details for this problem. Recall that Problem (26) can be formulated as follows:

$$\Theta_{\mathrm{U}}(m_1) = \min_{\substack{\mathbf{B},\mathbf{x},s,\mathbf{q},\mathbf{Q}_{\mathrm{r}},\\ \boldsymbol{\lambda}_k,\tilde{\mathbf{u}}_k,\mathbf{u}_k,\forall k\in[K]}} \quad \phi(m_1, s, \mathbf{q}, \mathbf{Q}_{\mathrm{r}}) \tag{82a}$$

$$\text{s.t.} \quad \begin{bmatrix} \chi(k,\mathbf{x},s,\boldsymbol{\lambda}_k) & \frac{1}{2}\mathbf{u}_k^\top \\ \frac{1}{2}\mathbf{u}_k & \mathbf{Q}_{\mathrm{r}} \end{bmatrix} \succeq 0, \ \forall k \in [K], \tag{82b}$$

$$\mathbf{q}+\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^{\top}\boldsymbol{\psi}(k,\mathbf{x},\boldsymbol{\lambda}_k)=\tilde{\mathbf{u}}_k,\ \forall k\in[K], \tag{82c}$$

$$\mathbf{x}\in\mathcal{X},\ \mathbf{B}^{\top}\mathbf{B}=\mathbf{I}_{m_1},\ \mathbf{B}\in\mathbb{R}^{m\times m_1},\ \mathbf{q}\in\mathbb{R}^{m},\ \mathbf{Q}_{\mathrm{r}}\in\mathbb{R}^{m_1\times m_1}, \tag{82d}$$

$$\boldsymbol{\lambda}_k\in\mathbb{R}^{l}_{+},\ \mathbf{u}_k\in\mathbb{R}^{m_1},\ \tilde{\mathbf{u}}_k\in\mathbb{R}^{m},\ \forall k\in[K], \tag{82e}$$

$$\tilde{\mathbf{u}}_k=\mathbf{B}\mathbf{u}_k,\ \forall k\in[K]. \tag{82f}$$

where constraints (82f) and $\mathbf{B}^{\top}\mathbf{B}=\mathbf{I}_{m_1}$ are bilinear constraints. We consider the following augmented Lagrangian problem for Problem (82):

$$\min_{\substack{\mathbf{x},s,\mathbf{q},\mathbf{Q}_{\mathrm{r}},\\ \boldsymbol{\lambda}_k,\tilde{\mathbf{u}}_k,\mathbf{u}_k,\forall k\in[K];\\ \mathbf{B};\ \boldsymbol{\beta}_k,\forall k\in[K]}}\left\{s+\gamma_2\mathbf{I}_{m_1}\bullet\mathbf{Q}_{\mathrm{r}}+\sqrt{\gamma_1}\,\|\mathbf{q}\|_2+\sum_{k=1}^{K}\boldsymbol{\beta}_k^{\top}\left(\tilde{\mathbf{u}}_k-\mathbf{B}\mathbf{u}_k\right)+\sum_{k=1}^{K}\frac{\rho_k}{2}\left\|\tilde{\mathbf{u}}_k-\mathbf{B}\mathbf{u}_k\right\|_2^2\ \middle|\ (82b)-(82e)\right\}, \tag{83}$$

where $\boldsymbol{\beta}_k\in\mathbb{R}^{m}$ $(\forall k\in[K])$ are Lagrangian multipliers and $\rho_k>0$ $(\forall k\in[K])$ are the penalty parameters. Thus, we design Algorithm 2 to solve Problem (82).

Algorithm 2 ADMM for Problem (82)

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

Given $\mathbf{B}^i$ and $\boldsymbol{\beta}_k^i$ for any $k\in[K]$, solve Problem (83) to obtain the optimal solution $\left(\mathbf{x},s,\mathbf{q},\mathbf{Q}_{\mathrm{r}},\boldsymbol{\lambda}_k,\tilde{\mathbf{u}}_k,\mathbf{u}_k,\forall k\in[K]\right)^{i+1}$;

Given $\left(\mathbf{x},s,\mathbf{q},\mathbf{Q}_{\mathrm{r}},\boldsymbol{\lambda}_k,\tilde{\mathbf{u}}_k,\mathbf{u}_k,\forall k\in[K]\right)^{i+1}$ and $\boldsymbol{\beta}_k^i$ for any $k\in[K]$, solve Problem (83) to obtain the optimal solution $\mathbf{B}^{i+1}$;

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

In this algorithm, given $\mathbf{B}$ and $\boldsymbol{\beta}_k$ for any $k\in[K]$, Problem (83) becomes a low-dimensional (i.e., m_1+1) SDP problem. Given $(\mathbf{x},s,\mathbf{q},\mathbf{Q}_{\mathrm{r}},\boldsymbol{\lambda}_k,\tilde{\mathbf{u}}_k,\mathbf{u}_k,\boldsymbol{\beta}_k,\forall k\in[K])$, Problem (83) becomes a nonconvex optimization problem, while the following proposition shows that it has an analytical optimal solution. Thus, Algorithm 2 can be performed efficiently.

PROPOSITION 8. *Given* $(\mathbf{x},s,\mathbf{q},\mathbf{Q}_{\mathrm{r}},\boldsymbol{\lambda}_k,\tilde{\mathbf{u}}_k,\mathbf{u}_k,\boldsymbol{\beta}_k,\forall k\in[K])$, *Problem* (83) *has an optimal solution* $\mathbf{B}^*=\tilde{\mathbf{U}}\tilde{\mathbf{V}}^{\top}$, *where* $\sum_{k=1}^{K}(\boldsymbol{\beta}_k\mathbf{u}_k^{\top}+\rho_k\tilde{\mathbf{u}}_k\mathbf{u}_k^{\top})=\tilde{\mathbf{U}}\tilde{\boldsymbol{\Sigma}}\tilde{\mathbf{V}}^{\top}$ *for* $\tilde{\mathbf{U}}\in\mathbb{R}^{m\times m_1}$, $\tilde{\boldsymbol{\Sigma}}\in\mathbb{R}^{m_1\times m_1}$, *and* $\tilde{\mathbf{V}}\in\mathbb{R}^{m_1\times m_1}$ *by the singular value decomposition (SVD).*

We further analyze the convergence property of Algorithm 2 to ensure the dimensionality reduction solution $\mathbf{B}$ returned by this algorithm is near-optimal, i.e., a theoretical guarantee. First, the following lemma holds.

LEMMA 4. *Given* $(\mathbf{x}, s, \mathbf{q}, \mathbf{Q}_\mathrm{r}, \boldsymbol{\lambda}_k, \tilde{\mathbf{u}}_k, \mathbf{u}_k, \boldsymbol{\beta}_k, \forall k \in [K])$, *we have* $\mathbf{B}^* = \tilde{\mathbf{U}}\tilde{\mathbf{V}}^\top$ *(an optimal solution of Problem (83)) is also an optimal solution of the following convex optimization problem:*

$$\max_{\mathbf{B} \in \mathbb{R}^{m \times m_1}} \left\{ \sum_{k=1}^{K} \left(\boldsymbol{\beta}_k \mathbf{u}_k^\top + \rho \tilde{\mathbf{u}}_k \mathbf{u}_k^\top \right) \bullet \mathbf{B} \;\middle|\; \mathbf{B}^\top \mathbf{B} \preceq \mathbf{I}_{m_1} \right\}. \tag{84}$$

We then present the convergence properties of our proposed ADMM algorithm. We let

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

and continue to adopt Assumption 3. We state the convergence theorem of Algorithm 2 as follows.

THEOREM 8. *Let* $(\mathbf{B}^*, \mathbf{x}^*, s^*, \mathbf{q}^*, \mathbf{Q}_\mathrm{r}^*, \boldsymbol{\lambda}_k^*, \tilde{\mathbf{u}}_k^*, \mathbf{u}_k^*, \forall k \in [K])$ *be any accumulation point of the sequence* $\{\mathbf{B}^i, \mathbf{x}^i, s^i, \mathbf{q}^i, \mathbf{Q}_\mathrm{r}^i, \boldsymbol{\lambda}_k^i, \tilde{\mathbf{u}}_k^i, \mathbf{u}_k^i, \forall k \in [K]\}$ *generated by Algorithm 2. Then,* $(\mathbf{B}^*, \mathbf{x}^*, s^*, \mathbf{q}^*, \mathbf{Q}_\mathrm{r}^*, \boldsymbol{\lambda}_k^*, \tilde{\mathbf{u}}_k^*, \mathbf{u}_k^*, \forall k \in [K])$ *satisfies the first-order stationary conditions of Problem (82).*

Proof. Let $(\mathbf{B}^*, \mathbf{x}^*, s^*, \mathbf{q}^*, \mathbf{Q}_\mathrm{r}^*, \boldsymbol{\lambda}_k^*, \tilde{\mathbf{u}}_k^*, \mathbf{u}_k^*, \forall k \in [K])$ be an accumulation point of the sequence $\{\mathbf{B}^i, \mathbf{x}^i, s^i, \mathbf{q}^i, \mathbf{Q}_\mathrm{r}^i, \boldsymbol{\lambda}_k^i, \tilde{\mathbf{u}}_k^i, \mathbf{u}_k^i, \forall k \in [K]\}$. Then, there exists a subsequence $\{\mathbf{B}^i, \mathbf{x}^i, s^i, \mathbf{q}^i, \mathbf{Q}_\mathrm{r}^i, \boldsymbol{\lambda}_k^i, \tilde{\mathbf{u}}_k^i, \mathbf{u}_k^i, \forall k \in [K]\}_{i \in \mathcal{I}}$ that converges to $(\mathbf{B}^*, \mathbf{x}^*, s^*, \mathbf{q}^*, \mathbf{Q}_\mathrm{r}^*, \boldsymbol{\lambda}_k^*, \tilde{\mathbf{u}}_k^*, \mathbf{u}_k^*, \forall k \in [K])$.

First, note that $(\mathbf{x}, s, \mathbf{q}, \mathbf{Q}_\mathrm{r}, \boldsymbol{\lambda}_k, \tilde{\mathbf{u}}_k, \mathbf{u}_k, \forall k \in [K])^{i+1}$ is the optimal solution of the convex problem

$$\min_{\substack{\mathbf{x}, s, \mathbf{q}, \mathbf{Q}_\mathrm{r}, \\ \boldsymbol{\lambda}_k, \tilde{\mathbf{u}}_k, \mathbf{u}_k, \forall k \in [K]}} \left\{ \mathcal{L}\left(\mathbf{B}^i, (\mathbf{x}, s, \mathbf{q}, \mathbf{Q}_\mathrm{r}, \boldsymbol{\lambda}_k, \tilde{\mathbf{u}}_k, \mathbf{u}_k, \forall k \in [K]), \left(\boldsymbol{\beta}_k^i, \forall k \in [K]\right)\right) \;\middle|\; (82\text{b}) - (82\text{e}) \right\}. \tag{85}$$

We show that there exists an interior point in the feasible region of Problem (85), by which the KKT conditions are first-order necessary conditions for the optimal solution of Problem (85). Specifically, by letting $\mathbf{x}'$ be an interior point in $\mathcal{X}$ (by Assumption 2), we can set $\hat{\boldsymbol{\lambda}}' = \{\mathbf{1}_l, \ldots, \mathbf{1}_l\}$, $s' = \sum_{k=1}^{K} \left| y_k^0(\mathbf{x}') + \mathbf{1}_l^\top \mathbf{b} + y_k(\mathbf{x}')^\top \boldsymbol{\mu} - \mathbf{1}_l^\top \mathbf{A} \boldsymbol{\mu} \right| + 1$, $\mathbf{q}' = \mathbf{0}_m$, $\tilde{\mathbf{u}}_k = (\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}})^\top (\mathbf{A}^\top \mathbf{1}_l - y_k(\mathbf{x}'))$, $\mathbf{u}_k = \mathbf{0}_{m_1}$, and $\mathbf{Q}_\mathrm{r}' = \mathbf{I}_{m_1}$. Clearly, $(\mathbf{x}', s', \hat{\boldsymbol{\lambda}}', \mathbf{q}', \mathbf{Q}_\mathrm{r}', \tilde{\mathbf{u}}_k, \mathbf{u}_k, \forall k \in [K])$ is an interior point in the feasible region of Problem (85). Therefore, the following first-order stationary conditions hold; that is, there exists $\begin{bmatrix} t_k^i & (\mathbf{p}_k^i)^\top \\ \mathbf{p}_k^i & \mathbf{P}_k^i \end{bmatrix} \succeq 0$, $\boldsymbol{\pi}_k^i \geq 0$, $\boldsymbol{\eta}_k^i \in \mathbb{R}^m$, and $\mathbf{Z}^i \succeq 0$ such that

$$1 - \sum_{k=1}^{K} t_k^i = 0, \tag{86a}$$

$$\gamma_2 \mathbf{I}_{m_1} - \sum_{k=1}^{K} \mathbf{P}_k^i = \mathbf{0}_{m_1 \times m_1}, \tag{86b}$$

$$\frac{\sqrt{\gamma_1} (\mathbf{q}^{i+1})^\top}{\|\mathbf{q}^{i+1}\|_2} + \sum_{k=1}^{K} (\boldsymbol{\eta}_k^i)^\top = 0, \tag{86c}$$

$$t_k^i (\mathbf{b} - \mathbf{A}\boldsymbol{\mu})^\top + (\boldsymbol{\eta}_k^i)^\top \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top \mathbf{A}^\top - (\boldsymbol{\pi}_k^i)^\top = 0, \; \forall k \in [K], \tag{86d}$$

$$-(\mathbf{p}_k^i)^\top - (\boldsymbol{\beta}_k^i)^\top \mathbf{B}^i - \rho_k^i(\tilde{\mathbf{u}}_k^{i+1} - \mathbf{B}^i\mathbf{u}_k^{i+1})\mathbf{B}^i = 0,\ \forall k \in [K], \tag{86e}$$

$$-(\boldsymbol{\eta}_k^i)^\top + (\boldsymbol{\beta}_k^i)^\top + \rho_k^i(\tilde{\mathbf{u}}_k^{i+1} - \mathbf{B}^i\mathbf{u}_k^{i+1}) = 0,\ \forall k \in [K], \tag{86f}$$

$$\sum_{k=1}^{K}\left(t_k^i\mathbf{w}_k^0 + \left(t_k^i\boldsymbol{\mu}^\top + (\boldsymbol{\eta}_k^i)^\top\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top\right)\mathbf{W}_k\right) - \sum_{i'=1}^{\tau}\sum_{j'=1}^{\tau} z_{i'j'}^i \mathbf{a}_{i'j'} = 0. \tag{86g}$$

By (86a) and $t_k^i \geq 0$ ($\forall k \in [K]$), we have that $\{t_k^i\}_{i\in\mathcal{I}}$ ($\forall k \in [K]$) are bounded. Because every bounded sequence has a convergent subsequence, without loss of generality, we can assume that $t_k^i \to t_k^*$ ($i \to \infty, i \in \mathcal{I}, \forall k \in [K]$). Taking limits in (86a) for $i \in \mathcal{I}$, we have

$$1 - \sum_{k=1}^{K} t_k^* = 0. \tag{87}$$

Similarly, by (86b), we have $\mathbf{P}_k^i \to \mathbf{P}_k^*$ ($i \to \infty, i \in \mathcal{I}, \forall k \in [K]$) and

$$\gamma_2\mathbf{I}_{m_1} - \sum_{k=1}^{K}\mathbf{P}_k^* = \mathbf{0}_{m_1\times m_1}. \tag{88}$$

Taking limits in (86e) and (86f) for $i \in \mathcal{I}$, by (35)and (36), we have that $\mathbf{p}_k^i \to \mathbf{p}_k^*$, $\boldsymbol{\eta}_k^i \to \boldsymbol{\eta}_k^*$ ($i \to \infty, i \in \mathcal{I}, \forall k \in [K]$), and

$$-\ (\mathbf{p}_k^*)^\top - (\boldsymbol{\beta}_k^*)^\top\mathbf{B}^* = 0,\ \forall k \in [K], \tag{89}$$

$$-\ (\boldsymbol{\eta}_k^*)^\top + (\boldsymbol{\beta}_k^*)^\top = 0,\ \forall k \in [K]. \tag{90}$$

Taking limits in (86d) and (86g) for $i \in \mathcal{I}$, we have that $\mathbf{Z}^i \to \mathbf{Z}^*$, $\boldsymbol{\pi}_k^i \to \boldsymbol{\pi}_k^*$ ($\forall k \in [K]$), and

$$t_k^*\left(\mathbf{b} - \mathbf{A}\boldsymbol{\mu}\right)^\top + (\boldsymbol{\eta}_k^*)^\top\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top\mathbf{A}^\top - (\boldsymbol{\pi}_k^*)^\top = 0,\ \forall k \in [K], \tag{91}$$

$$\sum_{k=1}^{K}\left(t_k^*\mathbf{w}_k^0 + \left(t_k^*\boldsymbol{\mu}^\top + (\boldsymbol{\eta}_k^*)^\top\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top\right)\mathbf{W}_k\right) - \sum_{i'=1}^{\tau}\sum_{j'=1}^{\tau} z_{i'j'}^* \mathbf{a}_{i'j'} = 0. \tag{92}$$

Taking limits in (86c) for $i \in \mathcal{I}$, we have

$$\frac{\sqrt{\gamma_1}(\mathbf{q}^*)^\top}{\|\mathbf{q}^*\|_2} + \sum_{k=1}^{K}(\boldsymbol{\eta}_k^*)^\top = 0. \tag{93}$$

Next, note that $\mathbf{B}^{i+1}$ is the optimal solution of the nonconvex problem

$$\min_{\mathbf{B}}\left\{\mathcal{L}\left(\mathbf{B}, (\mathbf{x}, s, \mathbf{q}, \mathbf{Q}_{\mathrm{r}}, \boldsymbol{\lambda}_k, \tilde{\mathbf{u}}_k, \mathbf{u}_k, \forall k \in [K])^{i+1}, \left(\boldsymbol{\beta}_k^i, \forall k \in [K]\right)\right)\ \middle|\ (82b) - (82e)\right\}.$$

In Proposition 8, we give an analytical optimal solution of $\mathbf{B}^{i+1}$. However, there is no optimality condition for this $\mathbf{B}^{i+1}$. By Lemma 4, we have that $\mathbf{B}^{i+1}$ is an optimal solution of the convex problem (84). Clearly, $\mathbf{B} = \mathbf{0}_{m\times m_1}$ is an interior point of Problem (84). Therefore, $\mathbf{B}^{i+1}$ satisfies the first-order stationary conditions of Problem (84); that is, there exists $\mathbf{C}^{i+1} \succeq 0$ such that

$$-\sum_{k=1}^{K}\boldsymbol{\beta}_k^i(\mathbf{u}_k^{i+1})^\top + \mathbf{B}^{i+1}(\mathbf{C}^{i+1})^\top - \rho_k^i\sum_{k=1}^{K}\left(\tilde{\mathbf{u}}_k^{i+1} - \mathbf{B}^{i+1}\mathbf{u}_k^i\right)(\mathbf{u}_k^i)^\top = 0. \tag{94}$$

Taking limits in (94) for $i \in \mathcal{I}$, we have that $\mathbf{B}^i(\mathbf{C}^i)^\top \to \sum_{k=1}^K \boldsymbol{\beta}_k^*(\mathbf{u}_k^*)^\top$ $(i \to \infty, \forall i \in \mathcal{I})$. Because $(\mathbf{B}^i)^\top \mathbf{B}^i = \mathbf{I}_{m_1}$ and $\mathbf{B}^i \to \mathbf{B}^*$ $(i \to \infty, \forall i \in \mathcal{I})$, it follows that

$$
\begin{aligned}
&(\mathbf{B}^i)^\top \mathbf{B}^i (\mathbf{C}^i)^\top \to (\mathbf{B}^i)^\top \sum_{k=1}^K \boldsymbol{\beta}_k^*(\mathbf{u}_k^*)^\top,\ i \to \infty, \forall i \in \mathcal{I},\\
\Longrightarrow\ &(\mathbf{C}^i)^\top \to (\mathbf{B}^i)^\top \sum_{k=1}^K \boldsymbol{\beta}_k^*(\mathbf{u}_k^*)^\top,\ i \to \infty, \forall i \in \mathcal{I},\\
\Longrightarrow\ &(\mathbf{C}^i)^\top \to (\mathbf{B}^*)^\top \sum_{k=1}^K \boldsymbol{\beta}_k^*(\mathbf{u}_k^*)^\top,\ i \to \infty, \forall i \in \mathcal{I},\\
\Longrightarrow\ &\mathbf{C}^i \to \mathbf{C}^* := \left(\sum_{k=1}^K \mathbf{u}_k^*(\boldsymbol{\beta}_k^*)^\top\right) \mathbf{B}^*,\ i \to \infty, \forall i \in \mathcal{I}.
\end{aligned}
$$

Thus, taking limits in (94) for $i \in \mathcal{I}$, we have

$$
-\sum_{k=1}^K \boldsymbol{\beta}_k^*(\mathbf{u}_k^*)^\top + \mathbf{B}^*(\mathbf{C}^*)^\top = 0. \tag{95}
$$

Finally, combining (87)–(93) and (95), we have that $(\mathbf{B}^*, \mathbf{x}^*, s^*, \mathbf{q}^*, \mathbf{Q}_r^*, \boldsymbol{\lambda}_k^*, \tilde{\mathbf{u}}_k^*, \mathbf{u}_k^*, \forall k \in [K])$ satisfies the stationary conditions of Problem (82); that is, there exists $\begin{bmatrix} t_k & \mathbf{p}_k^\top \\ \mathbf{p}_k & \mathbf{P}_k \end{bmatrix} \succeq 0$, $\boldsymbol{\eta}_k$, $\boldsymbol{\omega}_k$, $\boldsymbol{\pi}_k \geq 0$, $\mathbf{C}$, and $\mathbf{Z} \succeq 0$ such that

$$
1 - \sum_{k=1}^K t_k = 0, \tag{96a}
$$

$$
\gamma_2 \mathbf{I}_{m_1} - \sum_{k=1}^K \mathbf{P}_k = \mathbf{0}_{m_1 \times m_1}, \tag{96b}
$$

$$
\frac{\sqrt{\gamma_1}(\mathbf{q}^*)^\top}{\|\mathbf{q}^*\|_2} + \sum_{k=1}^K \boldsymbol{\eta}_k^\top = 0, \tag{96c}
$$

$$
t_k(\mathbf{b} - \mathbf{A}\boldsymbol{\mu})^\top + \boldsymbol{\eta}_k^\top \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top \mathbf{A}^\top - \boldsymbol{\pi}_k^\top = 0,\ \forall k \in [K], \tag{96d}
$$

$$
-\mathbf{p}_k^\top - \boldsymbol{\omega}_k^\top \mathbf{B}^* = 0,\ \forall k \in [K], \tag{96e}
$$

$$
-\boldsymbol{\eta}_k^\top + \boldsymbol{\omega}_k^\top = 0,\ \forall k \in [K], \tag{96f}
$$

$$
-\sum_{k=1}^K \boldsymbol{\omega}_k (\mathbf{u}_k^*)^\top + \mathbf{B}^* \mathbf{C}^\top = 0, \tag{96g}
$$

$$
\sum_{k=1}^K \left(t_k \mathbf{w}_k^0 + \left(t_k \boldsymbol{\mu}^\top + \boldsymbol{\eta}_k^\top \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top \right) \mathbf{W}_k \right) - \sum_{i=1}^\tau \sum_{j=1}^\tau z_{ij} \mathbf{a}_{ij} = 0. \tag{96h}
$$

This completes the proof. □

F.3. Proof of Propositions 5 and 8

In Proposition 5, given $(\mathbf{x}, \tilde{\mathbf{u}}_k, \mathbf{u}_k, \boldsymbol{\beta}_k, \forall k \in [K])$, we can omit the constant in the objective function of Problem (33) and rewrite this problem as follows:

$$
\min_{\mathbf{B} \in \mathbb{R}^{m \times m_1}} \left\{ \sum_{k=1}^K -\boldsymbol{\beta}_k^\top \mathbf{B} \mathbf{u}_k + \sum_{k=1}^K \left(-\rho_k \tilde{\mathbf{u}}_k^\top \mathbf{B} \mathbf{u}_k + \frac{\rho_k}{2} \mathbf{u}_k^\top \mathbf{B}^\top \mathbf{B} \mathbf{u}_k \right) \,\middle|\, \mathbf{B}^\top \mathbf{B} = \mathbf{I}_{m_1} \right\}. \tag{97}
$$

In Proposition 8, given $(\mathbf{x}, s, \mathbf{q}, \mathbf{Q}_r, \boldsymbol{\lambda}_k, \tilde{\mathbf{u}}_k, \mathbf{u}_k, \boldsymbol{\beta}_k, \forall k \in [K])$, we can also omit the constant in the objective function of Problem (83) and rewrite this problem as Problem (97).

By $\mathbf{B}^\top\mathbf{B} = \mathbf{I}_{m_1}$, the term $\mathbf{u}_k^\top\mathbf{B}^\top\mathbf{B}\mathbf{u}_k$ is also a constant. Because $\boldsymbol{\beta}_k^\top\mathbf{B}\mathbf{u}_k = (\boldsymbol{\beta}_k\mathbf{u}_k^\top) \bullet \mathbf{B}$ and $\tilde{\mathbf{u}}_k^\top\mathbf{B}\mathbf{u}_k = (\tilde{\mathbf{u}}_k\mathbf{u}_k^\top) \bullet \mathbf{B}$, we can further rewrite Problem (97) as follows:

$$\max_{\mathbf{B}\in\mathbb{R}^{m\times m_1}} \left\{ \sum_{k=1}^{K} \left(\boldsymbol{\beta}_k\mathbf{u}_k^\top + \rho_k\tilde{\mathbf{u}}_k\mathbf{u}_k^\top\right) \bullet \mathbf{B} \;\middle|\; \mathbf{B}^\top\mathbf{B} = \mathbf{I}_{m_1} \right\}.$$

By the SVD, i.e., $\sum_{k=1}^{K}(\boldsymbol{\beta}_k\mathbf{u}_k^\top + \rho_k\tilde{\mathbf{u}}_k\mathbf{u}_k^\top) = \tilde{\mathbf{U}}\tilde{\boldsymbol{\Sigma}}\tilde{\mathbf{V}}^\top$, we have

$$\begin{aligned}\mathbf{B}^* &= \arg\max_{\mathbf{B}^\top\mathbf{B}=\mathbf{I}_{m_1}} \left(\tilde{\mathbf{U}}\tilde{\boldsymbol{\Sigma}}\tilde{\mathbf{V}}^\top\right) \bullet \mathbf{B} = \arg\max_{\mathbf{B}^\top\mathbf{B}=\mathbf{I}_{m_1}} tr\left(\tilde{\mathbf{U}}\tilde{\boldsymbol{\Sigma}}\tilde{\mathbf{V}}^\top\mathbf{B}^\top\right)\\ &= \arg\max_{\mathbf{B}^\top\mathbf{B}=\mathbf{I}_{m_1}} tr\left(\tilde{\boldsymbol{\Sigma}}\tilde{\mathbf{V}}^\top\mathbf{B}^\top\tilde{\mathbf{U}}\right) = \arg\max_{\mathbf{B}^\top\mathbf{B}=\mathbf{I}_{m_1}} \tilde{\boldsymbol{\Sigma}} \bullet \left(\tilde{\mathbf{U}}^\top\mathbf{B}\tilde{\mathbf{V}}\right),\end{aligned}$$

where the second and fourth equalities hold by the definition of a matrix's trace and the third equality holds by the cyclic property of a matrix's trace. Eldén and Park (1999) show that $\mathbf{B}^* = \tilde{\mathbf{U}}\tilde{\mathbf{V}}^\top$ is an optimal solution. □

F.4. Proof of Lemmas 3 and 4

By the SVD in Propositions 5 and 8, we have

$$\sum_{k=1}^{K}(\boldsymbol{\beta}_k\mathbf{u}_k^\top + \rho_k\tilde{\mathbf{u}}_k\mathbf{u}_k^\top) = \tilde{\mathbf{U}}\tilde{\boldsymbol{\Sigma}}\tilde{\mathbf{V}}^\top.$$

We construct $\bar{\mathbf{U}} \in \mathbb{R}^{m\times m}$ by adding $m - m_1$ orthonormal columns to $\tilde{\mathbf{U}} \in \mathbb{R}^{m\times m_1}$ such that $\bar{\mathbf{U}}^\top\bar{\mathbf{U}} = \mathbf{I}_m$, and add $m - m_1$ zero rows to $\tilde{\boldsymbol{\Sigma}} \in \mathbb{R}^{m_1\times m_1}$ to construct $\bar{\boldsymbol{\Sigma}} = [\tilde{\boldsymbol{\Sigma}}; \mathbf{0}_{(m-m_1)\times m_1}] \in \mathbb{R}^{m\times m_1}$. It follows that $\tilde{\mathbf{U}}\tilde{\boldsymbol{\Sigma}}\tilde{\mathbf{V}}^\top = \bar{\mathbf{U}}\bar{\boldsymbol{\Sigma}}\tilde{\mathbf{V}}^\top$.

Meanwhile, by the cyclic property of a matrix's trace, we have

$$\left(\bar{\mathbf{U}}\bar{\boldsymbol{\Sigma}}\tilde{\mathbf{V}}^\top\right) \bullet \mathbf{B} = tr\left(\bar{\mathbf{U}}\bar{\boldsymbol{\Sigma}}\tilde{\mathbf{V}}^\top\mathbf{B}^\top\right) = tr\left(\bar{\boldsymbol{\Sigma}}\tilde{\mathbf{V}}^\top\mathbf{B}^\top\bar{\mathbf{U}}\right) = \bar{\boldsymbol{\Sigma}} \bullet \left(\bar{\mathbf{U}}^\top\mathbf{B}\tilde{\mathbf{V}}\right).$$

It follows that Problem (84) is equivalent to

$$\max_{\mathbf{B}\in\mathbb{R}^{m\times m_1}} \left\{\bar{\boldsymbol{\Sigma}} \bullet \left(\bar{\mathbf{U}}^\top\mathbf{B}\tilde{\mathbf{V}}\right) \;\middle|\; \mathbf{B}^\top\mathbf{B} \preceq \mathbf{I}_{m_1}\right\}. \tag{98}$$

Note that we have $\bar{\mathbf{U}}\bar{\mathbf{U}}^\top = \mathbf{I}_m$ and $\tilde{\mathbf{V}}\tilde{\mathbf{V}}^\top = \mathbf{I}_{m_1}$ by the SVD and the construction of $\bar{\mathbf{U}}$. We then have

$$\mathbf{B}^\top\mathbf{B} \preceq \mathbf{I}_{m_1} \iff \mathbf{B}^\top\bar{\mathbf{U}}\bar{\mathbf{U}}^\top\mathbf{B} \preceq \mathbf{I}_{m_1} \iff \tilde{\mathbf{V}}^\top\mathbf{B}^\top\bar{\mathbf{U}}\bar{\mathbf{U}}^\top\mathbf{B}\tilde{\mathbf{V}} \preceq \mathbf{I}_{m_1},$$

where the first equivalence holds by $\bar{\mathbf{U}}\bar{\mathbf{U}}^\top = \mathbf{I}_m$ and the second equivalence holds by $\tilde{\mathbf{V}}\tilde{\mathbf{V}}^\top = \mathbf{I}_{m_1}$ and Lemma 2.

Therefore, Problem (98) is equivalent to

$$\max_{\mathbf{B}\in\mathbb{R}^{m\times m_1}} \left\{\bar{\boldsymbol{\Sigma}} \bullet \left(\bar{\mathbf{U}}^\top \mathbf{B}\tilde{\mathbf{V}}\right) \;\middle|\; \left(\tilde{\mathbf{V}}^\top \mathbf{B}^\top \bar{\mathbf{U}}\right)\left(\bar{\mathbf{U}}^\top \mathbf{B}\tilde{\mathbf{V}}\right) \preceq \mathbf{I}_{m_1}\right\}. \tag{99}$$

Because $\bar{\boldsymbol{\Sigma}} \in \mathbb{R}^{m\times m_1}$ is a rectangular diagonal matrix with non-negative numbers on the diagonal, we have that $\bar{\mathbf{U}}^\top \mathbf{B}\tilde{\mathbf{V}} = [\mathbf{I}_{m_1}; \mathbf{0}_{(m-m_1)\times m_1}]$ will lead to the optimal value. Note that when we choose $\mathbf{B} = \tilde{\mathbf{U}}\tilde{\mathbf{V}}^\top$, we have

$$\bar{\mathbf{U}}^\top \mathbf{B}\tilde{\mathbf{V}} = \bar{\mathbf{U}}^\top \tilde{\mathbf{U}}\tilde{\mathbf{V}}^\top\tilde{\mathbf{V}} = [\mathbf{I}_{m_1}; \mathbf{0}_{(m-m_1)\times m_1}]\mathbf{I}_{m_1} = [\mathbf{I}_{m_1}; \mathbf{0}_{(m-m_1)\times m_1}].$$

Therefore, $\mathbf{B}^* = \tilde{\mathbf{U}}\tilde{\mathbf{V}}^\top$ is an optimal solution of Problem (99), i.e., Problem (84). ☐

F.5. Proof of Theorem 6

Let $(\mathbf{B}^*, \mathbf{x}^*, \tilde{\mathbf{u}}_k^*, \mathbf{u}_k^*, \forall k \in [K])$ be an accumulation point of the sequence $\{\mathbf{B}^i, \mathbf{x}^i, \tilde{\mathbf{u}}_k^i, \mathbf{u}_k^i, \forall k \in [K]\}$. Then, there exists a subsequence $\{\mathbf{B}^i, \mathbf{x}^i, \tilde{\mathbf{u}}_k^i, \mathbf{u}_k^i, \forall k \in [K]\}_{i\in\mathcal{I}}$ that converges to $(\mathbf{B}^*, \tilde{\mathbf{u}}_k^*, \mathbf{u}_k^*, \forall k \in [K])$.

First, note that $(\mathbf{x}, \tilde{\mathbf{u}}_k, \mathbf{u}_k, \forall k \in [K])^{i+1}$ is the optimal solution of the convex problem

$$\min_{\mathbf{x},\tilde{\mathbf{u}}_k,\mathbf{u}_k,\forall k\in[K]} \left\{\mathcal{L}\left(\mathbf{B}^i, (\mathbf{x}, \tilde{\mathbf{u}}_k, \mathbf{u}_k, \forall k \in [K]), \left(\boldsymbol{\beta}_k^i, \forall k \in [K]\right)\right) \;\middle|\; (32b) - (32c)\right\}. \tag{100}$$

Because there exists an interior point in the feasible region of Problem (100), by which the KKT conditions are first-order necessary conditions for the optimal solution of Problem (100). Specifically, the following first-order stationary conditions hold; that is,

$$-\nabla g(\mathbf{x}^{i+1}, \mathbf{u}_k^{i+1}, \tilde{\mathbf{u}}_k^{i+1}, \forall k \in [K]) + \begin{bmatrix} \mathbf{0} \\ (\mathbf{B}^i)^\top \boldsymbol{\beta}_1^i + \rho_1^i (\mathbf{B}^i)^\top \left(\tilde{\mathbf{u}}_1^{i+1} - \mathbf{B}^i \mathbf{u}_1^{i+1}\right) \\ \vdots \\ (\mathbf{B}^i)^\top \boldsymbol{\beta}_K^i + \rho_K^i (\mathbf{B}^i)^\top \left(\tilde{\mathbf{u}}_K^{i+1} - \mathbf{B}^i \mathbf{u}_K^{i+1}\right) \\ -\boldsymbol{\beta}_1^i - \rho_1^i \left(\tilde{\mathbf{u}}_1^{i+1} - \mathbf{B}^i \mathbf{u}_1^{i+1}\right) \\ \vdots \\ -\boldsymbol{\beta}_K^i - \rho_K^i \left(\tilde{\mathbf{u}}_K^{i+1} - \mathbf{B}^i \mathbf{u}_K^{i+1}\right) \end{bmatrix} \in \mathcal{N}_{\mathcal{U}}(\mathbf{x}^{i+1}, \mathbf{u}_k^{i+1}, \tilde{\mathbf{u}}_k^{i+1}, \forall k \in [K]), \tag{101}$$

where $\mathcal{N}_{\mathcal{U}}(\mathbf{x}^{i+1}, \mathbf{u}_k^{i+1}, \tilde{\mathbf{u}}_k^{i+1}, \forall k \in [K])$ is the normal cone of $\mathcal{U}$ at $(\mathbf{x}^{i+1}, \mathbf{u}_k^{i+1}, \tilde{\mathbf{u}}_k^{i+1}, \forall k \in [K])$. Taking limits in (101) for $i \in \mathcal{I}$, we have

$$-\nabla g(\mathbf{x}^*, \mathbf{u}_k^*, \tilde{\mathbf{u}}_k^*, \forall k \in [K]) + \begin{bmatrix} \mathbf{0} \\ (\mathbf{B}^*)^\top \boldsymbol{\beta}_1^* \\ \vdots \\ (\mathbf{B}^*)^\top \boldsymbol{\beta}_K^* \\ -\boldsymbol{\beta}_1^* \\ \vdots \\ -\boldsymbol{\beta}_K^* \end{bmatrix} \in \mathcal{N}_{\mathcal{U}}(\mathbf{x}^*, \mathbf{u}_k^*, \tilde{\mathbf{u}}_k^*, \forall k \in [K]). \tag{102}$$

Next, note that $\mathbf{B}^{i+1}$ is the optimal solution of the nonconvex problem

$$\min_{\mathbf{B}} \left\{ \mathcal{L}\left(\mathbf{B}, (\mathbf{x}, \tilde{\mathbf{u}}_k, \mathbf{u}_k, \forall k \in [K])^{i+1}, \left(\boldsymbol{\beta}_k^i, \forall k \in [K]\right)\right) \,\Big|\, (32b) - (32c) \right\}.$$

In Proposition 5, we give an analytical optimal solution of $\mathbf{B}^{i+1}$. However, there is no optimality condition for this $\mathbf{B}^{i+1}$. By Lemma 3, we have that $\mathbf{B}^{i+1}$ is an optimal solution of the convex problem (34). Clearly, $\mathbf{B} = \mathbf{0}_{m \times m_1}$ is an interior point of Problem (34). Therefore, $\mathbf{B}^{i+1}$ satisfies the first-order stationary conditions of Problem (34); that is, there exists $\mathbf{C}^{i+1} \succeq 0$ such that

$$-\sum_{k=1}^{K} \boldsymbol{\beta}_k^i (\mathbf{u}_k^{i+1})^\top + \mathbf{B}^{i+1}(\mathbf{C}^{i+1})^\top - \rho_k^i \sum_{k=1}^{K} \left(\tilde{\mathbf{u}}_k^{i+1} - \mathbf{B}^{i+1}\mathbf{u}_k^i\right) (\mathbf{u}_k^i)^\top = 0. \tag{103}$$

Taking limits in (103) for $i \in \mathcal{I}$, we have that $\mathbf{B}^i(\mathbf{C}^i)^\top \to \sum_{k=1}^{K} \boldsymbol{\beta}_k^* (\mathbf{u}_k^*)^\top$ $(i \to \infty, \forall i \in \mathcal{I})$. Because $(\mathbf{B}^i)^\top \mathbf{B}^i = \mathbf{I}_{m_1}$ and $\mathbf{B}^i \to \mathbf{B}^*$ $(i \to \infty, \forall i \in \mathcal{I})$, it follows that

$$\begin{aligned}
&(\mathbf{B}^i)^\top \mathbf{B}^i (\mathbf{C}^i)^\top \to (\mathbf{B}^i)^\top \sum_{k=1}^{K} \boldsymbol{\beta}_k^* (\mathbf{u}_k^*)^\top,\ i \to \infty, \forall i \in \mathcal{I},\\
\Longrightarrow\ &(\mathbf{C}^i)^\top \to (\mathbf{B}^i)^\top \sum_{k=1}^{K} \boldsymbol{\beta}_k^* (\mathbf{u}_k^*)^\top,\ i \to \infty, \forall i \in \mathcal{I},\\
\Longrightarrow\ &(\mathbf{C}^i)^\top \to (\mathbf{B}^*)^\top \sum_{k=1}^{K} \boldsymbol{\beta}_k^* (\mathbf{u}_k^*)^\top,\ i \to \infty, \forall i \in \mathcal{I},\\
\Longrightarrow\ &\mathbf{C}^i \to \mathbf{C}^* := \left(\sum_{k=1}^{K} \mathbf{u}_k^* (\boldsymbol{\beta}_k^*)^\top\right) \mathbf{B}^*,\ i \to \infty, \forall i \in \mathcal{I}.
\end{aligned}$$

Thus, taking limits in (103) for $i \in \mathcal{I}$, we have

$$-\sum_{k=1}^{K} \boldsymbol{\beta}_k^* (\mathbf{u}_k^*)^\top + \mathbf{B}^*(\mathbf{C}^*)^\top = 0. \tag{104}$$

Finally, combining (102) and (104), we have that $(\mathbf{B}^*, \mathbf{x}^*, \tilde{\mathbf{u}}_k^*, \mathbf{u}_k^*, \forall k \in [K])$ satisfies the stationary conditions of Problem (33); that is, there exist $\boldsymbol{\omega}_k$ and $\mathbf{C}$ such that

$$\begin{aligned}
&-\nabla g(\mathbf{x}^*, \mathbf{u}_k^*, \tilde{\mathbf{u}}_k^*, \forall k \in [K]) + \begin{bmatrix} \mathbf{0} \\ (\mathbf{B}^*)^\top \boldsymbol{\omega}_1 \\ \vdots \\ (\mathbf{B}^*)^\top \boldsymbol{\omega}_K \\ -\boldsymbol{\omega}_1 \\ \vdots \\ -\boldsymbol{\omega}_K \end{bmatrix} \in \mathcal{N}_{\mathcal{U}}(\mathbf{x}^*, \mathbf{u}_k^*, \tilde{\mathbf{u}}_k^*, \forall k \in [K])\\
&-\sum_{k=1}^{K} \boldsymbol{\omega}_k (\mathbf{u}_k^*)^\top + \mathbf{B}^* \mathbf{C}^\top = 0
\end{aligned} \tag{105}$$

This completes the proof. □

F.6. Proof of Proposition 6

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 (12) 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

$$(S_k + s_0)\left(\mathbf{B}'\mathbf{Q}_{\mathrm{r}}^*(\mathbf{B}')^\top + \mathbf{Q}_0\right) \succeq \frac{1}{4}\left(\mathbf{B}'\mathbf{q}_{\mathrm{r}}^* + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top \left(\mathbf{A}^\top\boldsymbol{\lambda}_k^* - y_k(\mathbf{x}^*)\right)\right) \times \left(\mathbf{B}'\mathbf{q}_{\mathrm{r}}^* + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top \left(\mathbf{A}^\top\boldsymbol{\lambda}_k^* - y_k(\mathbf{x}^*)\right)\right)^\top = \frac{1}{4}\mathbf{M}_k, \ \forall k \in [K]. \tag{106}$$

Note that, if $(S + s_0)\mathbf{Q}_0 \succeq (1/4)\mathbf{M}_k$ for any $k \in [K]$, then (106) 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 (106) 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

$$\Theta_{\mathrm{M}}(m) - \Theta_{\mathrm{L}}(m_1) \leq \Theta_{\mathrm{M}}' - \underline{\Theta}(m_1, \mathbf{B}') = s_0 + \sum_{k=1}^{K} \frac{\gamma_2}{4(S + s_0)}\mathbf{I}_m \bullet \mathbf{M}_k. \tag{107}$$

We further choose a value of s_0 to minimize the right-hand side (RHS) of (107). Note that (i) If $\sqrt{P} - S < 0$, then the RHS of (107) is minimized at P/S with $s_0 = 0$; (ii) If $\sqrt{P} - S \geq 0$, then the RHS of (107) is minimized at $2\sqrt{P} - S$ with $s_0 = \sqrt{P} - S$. Therefore, we conclude that the proposition holds. □

Appendix G: Supplement to Section 7

G.1. Multiproduct Newsvendor Problem

By Proposition 1, Problem (39) has the same optimal value as the following SDP formulation:

$$\min_{\substack{\mathbf{x},s,\boldsymbol{\lambda}_1,\\ \boldsymbol{\lambda}_2,\mathbf{q},\mathbf{Q}}} \; s+\gamma_2\mathbf{I}_m\bullet\mathbf{Q}+\sqrt{\gamma_1}\,\|\mathbf{q}\|_2 \tag{108a}$$

$$\text{s.t.}\quad \begin{bmatrix} s-(\mathbf{c}-\mathbf{v})^\top\mathbf{x}-\boldsymbol{\lambda}_1^\top(\mathbf{b}-\mathbf{A}\boldsymbol{\mu}) & \frac{1}{2}\left(\mathbf{q}+\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top\mathbf{A}^\top\boldsymbol{\lambda}_1\right)^\top \\ \frac{1}{2}\left(\mathbf{q}+\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top\mathbf{A}^\top\boldsymbol{\lambda}_1\right) & \mathbf{Q}\end{bmatrix}\succeq 0, \tag{108b}$$

$$\begin{bmatrix} s-(\mathbf{c}-\mathbf{g})^\top\mathbf{x}-\boldsymbol{\lambda}_2^\top(\mathbf{b}-\mathbf{A}\boldsymbol{\mu})+(\mathbf{v}-\mathbf{g})^\top\boldsymbol{\mu} & \frac{1}{2}\left(\mathbf{q}+\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top(\mathbf{A}^\top\boldsymbol{\lambda}_2+\mathbf{v}-\mathbf{g})\right)^\top \\ \frac{1}{2}\left(\mathbf{q}+\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top(\mathbf{A}^\top\boldsymbol{\lambda}_2+\mathbf{v}-\mathbf{g})\right) & \mathbf{Q}\end{bmatrix}\succeq 0, \tag{108c}$$

$$\mathbf{x}\in\mathbb{R}^m_+,\ \boldsymbol{\lambda}_1\in\mathbb{R}^l_+,\ \boldsymbol{\lambda}_2\in\mathbb{R}^l_+,\ \mathbf{q}\in\mathbb{R}^m,\ \mathbf{Q}\in\mathbb{R}^{m\times m}. \tag{108d}$$

By the first outer approximation (25), the following problem provides a lower bound for the optimal value of Problem (108):

$$\max_{\substack{\mathbf{B},t_1,\mathbf{p}_1,\mathbf{P}_1,\\ t_2,\mathbf{p}_2,\mathbf{P}_2}} \; \left(t_2\boldsymbol{\mu}^\top+\mathbf{p}_2^\top\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\right)^\top\right)(\mathbf{g}-\mathbf{v}) \tag{109a}$$

$$\text{s.t.}\quad 1-t_1-t_2=0,\ \sqrt{\gamma_1}-\|\mathbf{p}_1+\mathbf{p}_2\|_2\geq 0, \tag{109b}$$

$$t_1(\mathbf{A}\boldsymbol{\mu}-\mathbf{b})^\top+\mathbf{p}_1^\top\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\right)^\top\mathbf{A}^\top\leq 0, \tag{109c}$$

$$t_2(\mathbf{A}\boldsymbol{\mu}-\mathbf{b})^\top+\mathbf{p}_2^\top\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\right)^\top\mathbf{A}^\top\leq 0, \tag{109d}$$

$$\gamma_2\mathbf{I}_{m_1}-\mathbf{P}_1-\mathbf{P}_2\succeq 0,\ t_1(\mathbf{c}-\mathbf{v})+t_2(\mathbf{c}-\mathbf{g})\geq 0, \tag{109e}$$

$$\begin{bmatrix} t_1 & \mathbf{p}_1^\top \\ \mathbf{p}_1 & \mathbf{P}_1\end{bmatrix}\succeq 0,\ \begin{bmatrix} t_2 & \mathbf{p}_2^\top \\ \mathbf{p}_2 & \mathbf{P}_2\end{bmatrix}\succeq 0,\ \mathbf{B}^\top\mathbf{B}=\mathbf{I}_{m_1}, \tag{109f}$$

$$\mathbf{B}\in\mathbb{R}^{m\times m_1},\ \mathbf{p}_1\in\mathbb{R}^{m_1},\ \mathbf{p}_2\in\mathbb{R}^{m_1},\ \mathbf{P}_1\in\mathbb{R}^{m_1\times m_1},\ \mathbf{P}_2\in\mathbb{R}^{m_1\times m_1}. \tag{109g}$$

By the inner approximation (26), the following problem provides an upper bound for the optimal value of Problem (108) and achieves the optimal value of Problem (108) when $m_1\geq 2$:

$$\min_{\substack{\mathbf{B},\mathbf{x},s,\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_2,\\ \mathbf{q},\mathbf{Q}_r,\mathbf{u}_1,\mathbf{u}_2}} \; s+\gamma_2\mathbf{I}_{m_1}\bullet\mathbf{Q}_r+\sqrt{\gamma_1}\,\|\mathbf{q}\|_2 \tag{110a}$$

$$\text{s.t.}\quad \begin{bmatrix} s-(\mathbf{c}-\mathbf{v})^\top\mathbf{x}-\boldsymbol{\lambda}_1^\top(\mathbf{b}-\mathbf{A}\boldsymbol{\mu}) & \frac{1}{2}\mathbf{u}_1^\top \\ \frac{1}{2}\mathbf{u}_1 & \mathbf{Q}_r\end{bmatrix}\succeq 0, \tag{110b}$$

$$\begin{bmatrix} s-(\mathbf{c}-\mathbf{g})^\top\mathbf{x}-\boldsymbol{\lambda}_2^\top(\mathbf{b}-\mathbf{A}\boldsymbol{\mu})+(\mathbf{v}-\mathbf{g})^\top\boldsymbol{\mu} & \frac{1}{2}\mathbf{u}_2^\top \\ \frac{1}{2}\mathbf{u}_2 & \mathbf{Q}_r\end{bmatrix}\succeq 0, \tag{110c}$$

$$\mathbf{q}+\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top\mathbf{A}^\top\boldsymbol{\lambda}_1=\mathbf{B}\mathbf{u}_1, \tag{110d}$$

$$\mathbf{q}+\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top\left(\mathbf{A}^\top\boldsymbol{\lambda}_2+\mathbf{v}-\mathbf{g}\right)=\mathbf{B}\mathbf{u}_2, \tag{110e}$$

$$\mathbf{x} \in \mathbb{R}_+^m,\ \boldsymbol{\lambda}_1 \in \mathbb{R}_+^l,\ \boldsymbol{\lambda}_2 \in \mathbb{R}_+^l,\ \mathbf{B}^\top \mathbf{B} = \mathbf{I}_{m_1}, \tag{110f}$$

$$\mathbf{q} \in \mathbb{R}^m,\ \mathbf{Q}_\mathrm{r} \in \mathbb{R}^{m_1 \times m_1},\ \mathbf{B} \in \mathbb{R}^{m \times m_1},\ \mathbf{u}_1 \in \mathbb{R}^{m_1},\ \mathbf{u}_2 \in \mathbb{R}^{m_1}. \tag{110g}$$

By the second outer approximation (31), the following problem with $m_1 \leq 2$ provides another lower bound for the optimal value of Problem (108) and achieves the optimal value of Problem (108) when $m_1 = 2$:

$$\min_{\substack{\mathbf{B},\bar{\mathbf{B}},\mathbf{x},s,\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_2,\\ \mathbf{q},\mathbf{Q}_\mathrm{r},\mathbf{u}_1,\mathbf{u}_2,\mathbf{h}_1,\mathbf{h}_2}} \quad s + \gamma_2 \mathbf{I}_{m_1} \bullet \mathbf{Q}_\mathrm{r} + \sqrt{\gamma_1}\,\|\mathbf{q}\|_2 \tag{111a}$$

$$\text{s.t.} \quad \begin{bmatrix} s - (\mathbf{c} - \mathbf{v})^\top \mathbf{x} - \boldsymbol{\lambda}_1^\top (\mathbf{b} - \mathbf{A}\boldsymbol{\mu}) & \frac{1}{2}\mathbf{u}_1^\top \\ \frac{1}{2}\mathbf{u}_1 & \mathbf{Q}_\mathrm{r} \end{bmatrix} \succeq 0, \tag{111b}$$

$$\begin{bmatrix} s - (\mathbf{c} - \mathbf{g})^\top \mathbf{x} - \boldsymbol{\lambda}_2^\top (\mathbf{b} - \mathbf{A}\boldsymbol{\mu}) + (\mathbf{v} - \mathbf{g})^\top \boldsymbol{\mu} & \frac{1}{2}\mathbf{u}_2^\top \\ \frac{1}{2}\mathbf{u}_2 & \mathbf{Q}_\mathrm{r} \end{bmatrix} \succeq 0, \tag{111c}$$

$$\mathbf{q} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top \mathbf{A}^\top \boldsymbol{\lambda}_1 = \mathbf{B}\mathbf{u}_1 + \bar{\mathbf{B}}\mathbf{h}_1, \tag{111d}$$

$$\mathbf{q} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top \left(\mathbf{A}^\top \boldsymbol{\lambda}_2 + \mathbf{v} - \mathbf{g}\right) = \mathbf{B}\mathbf{u}_2 + \bar{\mathbf{B}}\mathbf{h}_2, \tag{111e}$$

$$\mathbf{x} \in \mathbb{R}_+^m,\ \boldsymbol{\lambda}_1 \in \mathbb{R}_+^l,\ \boldsymbol{\lambda}_2 \in \mathbb{R}_+^l,\ [\mathbf{B}, \bar{\mathbf{B}}]^\top [\mathbf{B}, \bar{\mathbf{B}}] = \mathbf{I}_K, \tag{111f}$$

$$\mathbf{q} \in \mathbb{R}^m,\ \mathbf{Q}_\mathrm{r} \in \mathbb{R}^{m_1 \times m_1},\ \mathbf{B} \in \mathbb{R}^{m \times m_1},\ \bar{\mathbf{B}} \in \mathbb{R}^{m \times (K - m_1)}, \tag{111g}$$

$$\mathbf{u}_1 \in \mathbb{R}^{m_1},\ \mathbf{u}_2 \in \mathbb{R}^{m_1},\ \mathbf{h}_1 \in \mathbb{R}^{K - m_1},\ \mathbf{h}_2 \in \mathbb{R}^{K - m_1}. \tag{111h}$$

G.2. Production-Transportation Problem

By the reformulation results of Section 4.1 of Bertsimas et al. (2010) and Appendix C of Cheramin et al. (2022), we have the following SDP reformulation for Problem (41):

$$\min_{\substack{\mathbf{x},\mathbf{z}_k(\forall k \in [K]),\\ s,\boldsymbol{\lambda}_k(\forall k \in [K]),\mathbf{q},\mathbf{Q}}} \quad s + \gamma_2 \mathbf{I}_{GH} \bullet \mathbf{Q} + \sqrt{\gamma_1}\,\|\mathbf{q}\|_2 \tag{112a}$$

$$\text{s.t.} \quad \begin{bmatrix} s - \mathbf{c}^\top \mathbf{x} - \beta_k - \boldsymbol{\lambda}_k^\top \mathbf{b} - \alpha_k \mathbf{z}_k^\top \boldsymbol{\mu} + \boldsymbol{\lambda}_k^\top \mathbf{A}\boldsymbol{\mu} & \frac{1}{2}\left(\mathbf{q} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top (\mathbf{A}^\top \boldsymbol{\lambda}_k - \alpha_k \mathbf{z}_k)\right)^\top \\ \frac{1}{2}\left(\mathbf{q} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top (\mathbf{A}^\top \boldsymbol{\lambda}_k - \alpha_k \mathbf{z}_k)\right) & \mathbf{Q} \end{bmatrix} \succeq 0, \quad \forall k \in [K], \tag{112b}$$

$$\boldsymbol{\lambda}_k \in \mathbb{R}_+^l,\ \forall k \in [K], \mathbf{q} \in \mathbb{R}^{GH}, \mathbf{Q} \in \mathbb{R}^{GH \times GH}, \tag{112c}$$

$$\mathbf{0} \leq \mathbf{x} \leq \mathbf{1}, \tag{112d}$$

$$\sum_{i=1}^{G} z_{ijk} = d_j,\ \forall j \in [H],\ k \in [K], \tag{112e}$$

$$\sum_{j=1}^{H} z_{ijk} = x_i,\ \forall i \in [G],\ k \in [K], \tag{112f}$$

$$z_{ijk} \geq 0,\ \forall i \in [G],\ j \in [H],\ k \in [K], \tag{112g}$$

where $\mathbf{z}_k$ is a vector whose $((i-1)G+j)$-th element is z_{ijk}. Note that, by Theorem 1 in Cheramin et al. (2022), the following problem can be reformulated as Problem (112):

$$\min_{\mathbf{x},\mathbf{z}_k,\forall k\in[K]} \max_{\mathbb{P}_{\rm I}\in\mathcal{D}_{\rm M}} \mathbb{E}_{\mathbb{P}}\left[\max_{k\in[K]}\left\{\mathbf{c}^\top\mathbf{x}+\alpha_k\mathbf{z}_k^\top\left(\mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\rm I}+\boldsymbol{\mu}\right)+\beta_k\right\}\right] \tag{113a}$$

$$\text{s.t.}\quad (112d)-(112g). \tag{113b}$$

By the first outer approximation (25), the following problem provides a lower bound for the optimal value of Problem (112):

$$\max_{\substack{t_k,\mathbf{p}_k,\mathbf{P}_k,\forall k\in[K],\\ \mathbf{w}_k,\mathbf{u}_k,\forall k\in[K],\mathbf{v},\mathbf{B}}} \sum_{k=1}^K t_k\beta_k-\sum_{i=i}^G v_i+\sum_{k=1}^K\sum_{j=1}^H w_{jk}d_j \tag{114a}$$

$$\text{s.t.}\quad 1-\sum_{k=1}^K t_k=0,\ \sqrt{\gamma_1}-\left\|\sum_{k=1}^K\mathbf{p}_k\right\|_2\geq 0, \tag{114b}$$

$$t_k(\mathbf{A}\boldsymbol{\mu}-\mathbf{b})^\top+\mathbf{p}_k^\top\left(\mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{B}\right)^\top\mathbf{A}^\top\leq 0,\ \forall k\in[K], \tag{114c}$$

$$\gamma_2\mathbf{I}_{m_1}-\sum_{k=1}^K\mathbf{P}_k\succeq 0,\ \sum_{k=1}^K(t_k\mathbf{c}+\mathbf{u}_k)+\mathbf{v}\geq 0, \tag{114d}$$

$$\alpha_k t_k\boldsymbol{\mu}^\top+\alpha_k\mathbf{p}_k^\top\left(\mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{B}\right)^\top-(\underbrace{\mathbf{w}_k^\top,\ldots,\mathbf{w}_k^\top}_{\text{repeat G times}})-(\underbrace{u_{1k},\ldots,u_{1k}}_{\text{repeat H times}},\ldots,u_{Gk},\ldots,u_{Gk})\geq 0,\ \forall k\in[K], \tag{114e}$$

$$\begin{bmatrix} t_k & \mathbf{p}_k^\top\\ \mathbf{p}_k & \mathbf{P}_k\end{bmatrix}\succeq 0,\ \forall k\in[K],\ \mathbf{B}^\top\mathbf{B}=\mathbf{I}_{m_1},\ \mathbf{v}\in\mathbb{R}_+^G,\ \mathbf{B}\in\mathbb{R}^{GH\times m_1}, \tag{114f}$$

$$\mathbf{p}_k\in\mathbb{R}^{m_1},\ \mathbf{P}_k\in\mathbb{R}^{m_1\times m_1},\ \mathbf{w}_k\in\mathbb{R}^H,\ \mathbf{u}_k\in\mathbf{R}^G,\ \forall k\in[K]. \tag{114g}$$

By the inner approximation (26), the following problem provides an upper bound for the optimal value of Problem (112) and achieves the optimal value of Problem (112) when $m_1\geq K$:

$$\min_{\substack{\mathbf{B},\mathbf{x},\mathbf{z}_k(\forall k\in[K]),\\ \mathbf{u}_k,\boldsymbol{\lambda}_k,\forall k\in[K],s,\mathbf{q},\mathbf{Q}_{\rm r}}} s+\gamma_2\mathbf{I}_{m_1}\bullet\mathbf{Q}_{\rm r}+\sqrt{\gamma_1}\|\mathbf{q}\|_2 \tag{115a}$$

$$\text{s.t.}\quad \begin{bmatrix} s-\mathbf{c}^\top\mathbf{x}-\beta_k-\boldsymbol{\lambda}_k^\top\mathbf{b}-\alpha_k\mathbf{z}_k^\top\boldsymbol{\mu}+\boldsymbol{\lambda}_k^\top\mathbf{A}\boldsymbol{\mu} & \frac{1}{2}\mathbf{u}_k^\top\\ \frac{1}{2}\mathbf{u}_k & \mathbf{Q}_{\rm r}\end{bmatrix}\succeq 0,\ \forall k\in[K], \tag{115b}$$

$$\mathbf{q}+\left(\mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}\right)^\top\left(\mathbf{A}^\top\boldsymbol{\lambda}_k-\alpha_k\mathbf{z}_k)\right)=\mathbf{B}\mathbf{u}_k,\ \forall k\in[K], \tag{115c}$$

$$\boldsymbol{\lambda}_k\in\mathbb{R}_+^l,\ \mathbf{u}_k\in\mathbb{R}^{m_1},\ \forall k\in[K],\mathbf{q}\in\mathbb{R}^{GH},\mathbf{Q}_{\rm r}\in\mathbb{R}^{m_1\times m_1}, \tag{115d}$$

$$(112d)-(112g),\ \mathbf{B}^\top\mathbf{B}=\mathbf{I}_{m_1},\ \mathbf{B}\in\mathbb{R}^{GH\times m_1}. \tag{115e}$$

By the second outer approximation (31), the following problem with $m_1\leq K$ provides another lower bound for the optimal value of Problem (112) and achieves the optimal value of Problem (112) when $m_1=K$:

$$\min_{\substack{\mathbf{B},\bar{\mathbf{B}},\mathbf{x},\mathbf{z}_k(\forall k\in[K]),\\ \mathbf{u}_k,\mathbf{h}_k,\boldsymbol{\lambda}_k,\forall k\in[K],s,\mathbf{q},\mathbf{Q}_{\rm r}}} s+\gamma_2\mathbf{I}_{m_1}\bullet\mathbf{Q}_{\rm r}+\sqrt{\gamma_1}\|\mathbf{q}\|_2 \tag{116a}$$

$$\text{s.t.} \quad \begin{bmatrix} s - \mathbf{c}^\top \mathbf{x} - \beta_k - \boldsymbol{\lambda}_k^\top \mathbf{b} - \alpha_k \mathbf{z}_k^\top \boldsymbol{\mu} + \boldsymbol{\lambda}_k^\top \mathbf{A}\boldsymbol{\mu} & \frac{1}{2}\mathbf{u}_k^\top \\ \frac{1}{2}\mathbf{u}_k & \mathbf{Q}_{\mathrm{r}} \end{bmatrix} \succeq 0,\ \forall k \in [K], \tag{116b}$$

$$\mathbf{q} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top \left(\mathbf{A}^\top \boldsymbol{\lambda}_k - \alpha_k \mathbf{z}_k)\right) = \mathbf{B}\mathbf{u}_k + \bar{\mathbf{B}}\mathbf{h}_k,\ \forall k \in [K], \tag{116c}$$

$$\boldsymbol{\lambda}_k \in \mathbb{R}_+^l,\ \mathbf{u}_k \in \mathbb{R}^{m_1},\ \mathbf{h}_k \in \mathbb{R}^{K-m_1},\ \forall k \in [K], \mathbf{q} \in \mathbb{R}^{GH}, \mathbf{Q}_{\mathrm{r}} \in \mathbb{R}^{m_1 \times m_1}, \tag{116d}$$

$$(112d) - (112g),\ [\mathbf{B}, \bar{\mathbf{B}}]^\top [\mathbf{B}, \bar{\mathbf{B}}] = \mathbf{I}_{m_1},\ \mathbf{B} \in \mathbb{R}^{GH \times m_1},\ \bar{\mathbf{B}} \in \mathbb{R}^{GH \times (K-m_1)}. \tag{116e}$$

G.3. 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 (44). □

G.4. Sensitivity Analyses

Table G3 Sensitivity Analyses on the Production-Transportation Problem with $K = 5$

		Size $((G,H))$	(4,25)	(5,20)	(5,40)	(8,25)	(10,40)	(20,30)	(20,40)
$m_1 = 3$	ODR-LB	Gap1 (%)	0.37	0.71	0.67	0.47	-	-	-
		Time (secs)	2.86	2.81	4.87	4.32	15.16	27.51	57.45
		Interval Gap (%)	0.69	0.79	0.73	0.51	0.49	0.51	0.57
		Theoretical Gap (%)	1.27	0.62	2.20	1.36	-	-	-
	ODR-RLB	Gap1 (%)	0.07	0.17	0.02	0.02	-	-	-
		Time (secs)	13.53	14.19	43.93	57.79	205.99	442.26	1092.43
		Interval Gap (%)	0.39	0.25	0.08	0.06	0.03	0.02	0.02
		Theoretical Gap (%)	1.08	3.19	1.96	0.96	-	-	-
	ODR-UB	Gap2 (%)	0.32	0.08	0.06	0.04	-	-	-
		Time (secs)	6.03	5.60	22.31	20.53	122.04	332.10	660.08
		Theoretical Gap (%)	1.28	1.03	3.20	2.05	-	-	-
$m_1 = 5$	ODR-LB	Gap1 (%)	0.14	0.34	0.22	0.29	-	-	-
		Time (secs)	4.03	3.63	6.01	4.82	12.41	25.03	57.79
		Interval Gap (%)	0.15	0.34	0.43	0.29	0.52	0.55	0.51
		Theoretical Gap (%)	1.24	1.07	4.91	0.77	2.71	1.25	1.09
	ODR-RLB	Gap1 (%)	0.01	0.02	0.01	0.00	-	-	-
		Time (secs)	5.42	5.36	22.40	20.86	123.64	330.29	665.43
		Interval Gap (%)	0.02	0.02	0.01	0.01	0.00	0.00	0.00
		Theoretical Gap (%)	1.13	1.17	1.80	0.81	1.22	1.92	1.32
	ODR-UB	Gap2 (%)	0.01	0.01	0.00	0.00	-	-	-
		Time (secs)	5.48	5.32	22.41	20.86	123.48	329.95	665.35
		Theoretical Gap (%)	1.13	1.17	1.80	0.81	1.22	1.92	1.32
$m_1 = 7$	ODR-LB	Gap1 (%)	0.16	0.21	0.53	0.23	-	-	-
		Time (secs)	3.56	4.29	5.39	5.70	19.27	22.89	58.38
		Interval Gap (%)	0.17	0.22	0.53	0.24	0.31	0.56	0.50
		Theoretical Gap (%)	1.22	0.88	1.67	0.79	-	-	-
	ODR-UB	Gap2 (%)	0.02	0.01	0.00	0.00	-	-	-
		Time (secs)	5.62	5.48	22.42	21.24	121.61	330.94	660.85
		Theoretical Gap (%)	1.15	1.04	1.28	0.59	-	-	-

Table G4 Sensitivity Analyses on the Production-Transportation Problem with $K = 10$

		Size $((G, H))$	(4,25)	(5,20)	(5,40)	(8,25)	(10,40)	(20,30)	(20,40)
$m_1 = 8$	ODR-LB	Gap1 (%)	0.17	0.15	0.16	0.28	-	-	-
		Time (secs)	6.26	5.75	9.80	8.87	29.94	54.71	133.58
		Interval Gap (%)	0.17	0.15	0.16	0.28	0.23	0.22	0.27
		Theoretical Gap (%)	4.32	4.53	6.22	6.09	-	-	-
	ODR-RLB	Gap1 (%)	0.01	0.00	0.01	0.00	-	-	-
		Time (secs)	13.03	12.91	53.84	35.92	205.07	593.38	1334.02
		Interval Gap (%)	0.01	0.00	0.01	0.01	0.00	0.00	0.00
		Theoretical Gap (%)	3.53	3.02	4.24	4.86	-	-	-
	ODR-UB	Gap2 (%)	0.00	0.00	0.00	0.00	-	-	-
		Time (secs)	10.31	10.23	39.26	31.59	119.58	339.84	755.18
		Theoretical Gap (%)	2.21	2.40	3.98	3.04	-	-	-
$m_1 = 10$	ODR-LB	Gap1 (%)	0.12	0.19	0.13	0.25	-	-	-
		Time (secs)	6.65	6.41	11.11	10.31	31.42	61.41	131.10
		Interval Gap (%)	0.12	0.19	0.13	0.25	0.22	0.19	0.25
		Theoretical Gap (%)	1.78	1.97	2.96	1.19	2.10	1.49	1.22
	ODR-RLB	Gap1 (%)	0.00	0.00	0.00	0.00	-	-	-
		Time (secs)	11.22	10.65	40.34	32.95	122.54	342.80	748.11
		Interval Gap (%)	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		Theoretical Gap (%)	1.86	2.27	1.74	1.22	2.10	1.39	2.30
	ODR-UB	Gap2 (%)	0.00	0.00	0.00	0.00	-	-	-
		Time (secs)	11.14	10.69	40.28	32.92	122.48	344.10	747.90
		Theoretical Gap (%)	1.86	2.27	1.74	1.22	2.10	1.39	2.30
$m_1 = 12$	ODR-LB	Gap1 (%)	0.11	0.14	0.15	0.23	-	-	-
		Time (secs)	8.89	7.90	11.78	13.44	32.16	63.96	122.91
		Interval Gap (%)	0.11	0.14	0.15	0.23	0.19	0.19	0.27
		Theoretical Gap (%)	3.64	5.01	7.41	7.85	-	-	-
	ODR-UB	Gap2 (%)	0.00	0.00	0.00	0.00	-	-	-
		Time (secs)	11.71	11.55	38.59	29.68	123.40	346.68	761.21
		Theoretical Gap (%)	0.81	1.32	2.20	2.75	-	-	-

Table G5 Sensitivity Analyses on the Production-Transportation Problem with $K = 15$

		Size $((G,H))$	(4,25)	(5,20)	(5,40)	(8,25)	(10,40)	(20,30)	(20,40)
$m_1 = 13$	ODR-LB	Gap1 (%)	0.08	0.21	0.15	0.18	-	-	-
		Time (secs)	11.12	11.10	18.00	24.05	31.32	71.04	142.88
		Interval Gap (%)	0.08	0.21	0.15	0.18	0.25	0.19	0.23
		Theoretical Gap (%)	4.56	7.01	5.79	6.95	-	-	-
	ODR-RLB	Gap1 (%)	0.00	0.03	0.00	0.00	-	-	-
		Time (secs)	26.19	25.91	97.35	68.99	213.44	641.74	1451.80
		Interval Gap (%)	0.00	0.03	0.00	0.00	0.00	0.01	0.00
		Theoretical Gap (%)	6.09	5.83	5.88	6.12	-	-	-
	ODR-UB	Gap2 (%)	0.00	0.00	0.00	0.00	-	-	-
		Time (secs)	19.46	19.15	71.71	50.62	128.92	367.33	833.81
		Theoretical Gap (%)	2.44	3.00	2.52	3.07	-	-	-
$m_1 = 15$	ODR-LB	Gap1 (%)	0.05	0.10	0.12	0.15	-	-	-
		Time (secs)	13.92	16.56	22.11	26.16	43.02	80.51	168.90
		Interval Gap (%)	0.05	0.10	0.12	0.15	0.11	0.13	0.09
		Theoretical Gap (%)	4.69	5.22	5.57	6.38	3.29	4.41	3.95
	ODR-RLB	Gap1 (%)	0.00	0.00	0.00	0.00	-	-	-
		Time (secs)	22.60	21.45	77.18	63.99	241.03	689.24	1550.21
		Interval Gap (%)	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		Theoretical Gap (%)	1.38	2.61	1.79	1.73	2.14	1.94	2.25
	ODR-UB	Gap2 (%)	0.00	0.00	0.00	0.00	-	-	-
		Time (secs)	22.61	21.44	76.99	64.08	149.21	401.63	878.68
		Theoretical Gap (%)	1.38	2.61	1.79	1.73	2.14	1.94	2.25
$m_1 = 17$	ODR-LB	Gap1 (%)	0.10	0.09	0.13	0.16	-	-	-
		Time (secs)	18.10	21.04	28.81	32.61	52.71	84.10	188.21
		Interval Gap (%)	0.10	0.09	0.13	0.16	0.10	0.08	0.13
		Theoretical Gap (%)	4.75	4.13	5.50	6.16	-	-	-
	ODR-UB	Gap2 (%)	0.00	0.00	0.00	0.00	-	-	-
		Time (secs)	26.83	27.84	84.70	63.17	161.79	447.03	926.33
		Theoretical Gap (%)	2.20	3.61	2.58	2.20	-	-	-