

COMPUTATIONALLY EFFICIENT APPROACHES FOR  
OPTIMIZATION UNDER UNCERTAINTY AND ITS  
APPLICATIONS

by

Meysam Cheramin

---

Copyright © Meysam Cheramin 2021

A Dissertation Submitted to the Faculty of the  
DEPARTMENT OF SYSTEMS AND INDUSTRIAL ENGINEERING

In Partial Fulfillment of the Requirements  
For the Degree of

DOCTOR OF PHILOSOPHY

In the Graduate College

THE UNIVERSITY OF ARIZONA

2021

THE UNIVERSITY OF ARIZONA  
GRADUATE COLLEGE

As members of the Dissertation Committee, we certify that we have read the dissertation prepared by: Meysam Cheramin  
titled:

Computationally Efficient Approaches  
for Optimization under Uncertainty and its Applications

and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy.

*Cheng*

\_\_\_\_\_  
Jianqiang Cheng

Date: May 11, 2021

*Afroz Jalilzadeh*

\_\_\_\_\_  
Afroz Jalilzadeh

Date: May 11, 2021

*Hongyue Jin*

\_\_\_\_\_  
Hongyue Jin

Date: May 11, 2021

*Ruiwei Jiang*

\_\_\_\_\_  
Ruiwei Jiang

Date: May 11, 2021

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copies of the dissertation to the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

*Cheng*

\_\_\_\_\_  
Jianqiang Cheng  
Systems and Industrial Engineering

Date: May 11, 2021

## ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to those who supported me in making this journey possible. Without their help, this achievement would not have been possible.

First and foremost, I would like to express my deepest gratitude to my advisor, Dr. Jianqiang Cheng, for his endless support, patience, and encouragement. He not only mentored me through every stage of my Ph.D. studies but also provided a friendly and scientific atmosphere for my academic progress. This opportunity enabled me to learn many lessons from his positive attitude and valuable insights, which will be useful for both my academic and personal life.

I would like to express my appreciation to my committee members, Dr. Ruiwei Jiang, Dr. Hongyue Jin, and Dr. Afrooz Jalilzadeh for their supports, evaluations, and suggestions, which significantly improved the quality of this dissertation.

I would like to express my sincere thanks to the staff and faculty members of the Systems and Industrial Engineering department who made my Ph.D. studies a great experience by providing a pleasant educational environment. I am eternally grateful to Professor Young-Jun Son, the head of the Systems and Industrial Engineering department, for his continuous guidance, support, and positive energy.

Last but not least, my deepest debt of gratitude goes to my wife, parents, and sister for their endless love and support throughout my life.

## DEDICATION

I dedicate this dissertation to

*my wife, Elahe,*

*my parents, Hossein and Zahra,*

*and my sister, Nafiseh,*

for their endless **support, love and encouragement.**

## TABLE OF CONTENTS

LIST OF FIGURES . . . . .	8
LIST OF TABLES . . . . .	9
ABSTRACT . . . . .	11
CHAPTER 1 Introduction . . . . .	13
1.1 Motivation . . . . .	13
1.2 Contributions . . . . .	14
1.2.1 Computationally Efficient Approximations for DRO . . . . .	15
1.2.2 Resilient NdFeB Magnet Recycling . . . . .	16
1.2.3 Data-driven RO Using Scenario-Induced Uncertainty Sets . . . . .	17
1.3 Dissertation Organization . . . . .	18
CHAPTER 2 Computationally Efficient Approximations for Distributionally Robust Optimization under Moment and Wasserstein Ambiguity . . . . .	20
2.1 Introduction . . . . .	20
2.2 Moment-based Ambiguity Set . . . . .	24
2.2.1 Lower Bound . . . . .	26
2.2.2 Lower Bound Quality . . . . .	33
2.2.3 Upper Bounds . . . . .	37
2.2.3.1 PCA based Upper Bound . . . . .	37
2.2.3.2 Vector Splitting based Upper Bound . . . . .	40
2.2.4 Upper Bound Quality . . . . .	45
2.3 Combined Ambiguity Set . . . . .	54
2.3.1 Lower Bound . . . . .	56
2.3.2 Upper Bound . . . . .	62
2.4 Computational Experiments . . . . .	68
2.4.1 Computational Setup . . . . .	68
2.4.1.1 Production-Transportation Problem . . . . .	69
2.4.1.2 Multi-Product Newsvendor Problem . . . . .	70
2.4.2 Computational Results . . . . .	71
2.4.2.1 Instance Generation and Table Header Description . . . . .	71
2.4.2.2 Lower Bound Performance . . . . .	73
2.4.2.3 Upper Bound Performance . . . . .	75
2.4.2.4 Interval Performance . . . . .	77

TABLE OF CONTENTS – *Continued*

2.4.2.5	Benefits of Choosing Leading Components . . . . .	80
2.4.2.6	Sensitivity Analyses . . . . .	81
2.5	Discussions . . . . .	83
2.6	Supplement to Section 2.3: Further Including the First-Order Mo- ment Information . . . . .	84
2.6.1	Lower Bound . . . . .	88
2.6.2	Upper Bound . . . . .	89
2.6.2.1	Computational Experiments . . . . .	90
2.7	Supplement to Section 2.4: Applying the Proposed Approximations in the Context of Production-Transportation Problem . . . . .	94
CHAPTER 3 Resilient NdFeB Magnet Recycling under the Impacts of COVID- 19 Pandemic: Stochastic Programming and Benders Decomposition . . . . .		
19	Pandemic: Stochastic Programming and Benders Decomposition . . . . .	97
3.1	Introduction . . . . .	98
3.2	Literature Review . . . . .	101
3.2.1	Supply Chain and Logistics Network Design under Uncertainty	102
3.2.2	Supply Chain Risk Management during the COVID-19 Pan- demic . . . . .	103
3.2.3	Gaps and Contributions . . . . .	104
3.3	Problem Description . . . . .	105
3.4	Mathematical Model Formulation . . . . .	107
3.5	Solution Approach . . . . .	115
3.6	Case Study . . . . .	118
3.6.1	Scenario Generation Method . . . . .	121
3.6.1.1	Disruption Rates of Facilities . . . . .	122
3.6.1.2	Demand for Recycled NdFeB Magnets . . . . .	123
3.6.2	Case Study Results . . . . .	124
3.7	Computational Experiments . . . . .	126
3.7.1	Test Instance Generation . . . . .	127
3.7.2	Optimization and Robustness Testing . . . . .	127
3.7.3	Computational Performance of the Solution Approach . . . . .	130
3.8	Discussions . . . . .	133
3.8.1	Managerial Implications . . . . .	134
3.9	Supplement to Section 3.2: Classification of Relevant Studies . . . . .	136
3.10	Supplement to Section 3.6: Supplementary data . . . . .	137
CHAPTER 4 Data-Driven Robust Optimization Using Scenario-Induced Un- certainty Sets . . . . .		
4.1	Introduction . . . . .	141

TABLE OF CONTENTS – *Continued*

4.2	Polyhedral Uncertainty Sets . . . . .	145
4.3	Scenario-Induced Uncertainty Sets . . . . .	147
4.3.1	Low-rank Approximation with PCA . . . . .	148
4.3.2	PCA-based Polyhedral Uncertainty Sets . . . . .	149
4.4	Lower Bound Quality for Static RO . . . . .	151
4.5	Probabilistic Guarantees . . . . .	154
4.6	Computational Experiments . . . . .	157
4.6.1	Computational Setup . . . . .	157
4.6.1.1	Knapsack Problem . . . . .	158
4.6.1.2	Power Grid Problem . . . . .	160
4.6.2	Computational Results . . . . .	164
4.6.2.1	Instance Generation and Table Header Description . . . . .	164
4.6.2.2	Uncertainty Set Performance . . . . .	168
4.7	Discussions . . . . .	172
CHAPTER 5 Conclusions . . . . .		173
REFERENCES . . . . .		178

## LIST OF FIGURES

3.1	Schematic view of the designed supply chain and logistics network . . . .	106
3.2	Geographic representation of facilities and material flows in the optimal solution . . . . .	125
3.3	Geographic representation of the optimal solution of the deterministic problem . . . . .	127
4.1	$\mathcal{U}_{\text{conv}}(\mathcal{S})$ of uncertain parameters $u_1, u_2$ . . . . .	146
4.2	$\mathcal{U}_{\text{box}}, \mathcal{U}_{\text{budget}}$ with $\Gamma = 1$ , and $\mathcal{U}_{\text{conv}}(\mathcal{S})$ for uncertain parameters $u_1, u_2$ . .	147
4.3	$\mathcal{U}_{\text{PCA}}(\mathcal{S}, 2)$ VS. $\mathcal{U}_{\text{conv}}(\mathcal{S}), \mathcal{U}_{\text{budget}}$ with $\Gamma = 1$ , and $\mathcal{U}_{\text{box}}$ for uncertain parameters $u_1, u_2$ . . . . .	150
4.4	$\mathcal{U}_{\text{PCA}}(\mathcal{S}, 1)$ VS. $\mathcal{U}_{\text{PCA}}(\mathcal{S}, 2)$ for uncertain parameters $u_1, u_2$ . . . . .	150

## LIST OF TABLES

2.1	Lower bound (2.15) on the production-transportation problem . . . . .	73
2.2	Lower bound (2.15) on the newsvendor problem . . . . .	74
2.3	Lower bound (2.60) on the production-transportation problem . . . . .	74
2.4	Lower bound (2.60) on the newsvendor problem . . . . .	75
2.5	Upper bound (2.24) on the production-transportation problem . . . . .	75
2.6	Upper bound (2.24) on the newsvendor problem . . . . .	75
2.7	Upper bound (2.37) on the production-transportation problem . . . . .	76
2.8	Upper bound (2.37) on the newsvendor problem . . . . .	76
2.9	Upper bound (2.72) on the production-transportation problem . . . . .	77
2.10	Upper bound (2.72) on the newsvendor problem . . . . .	77
2.11	Intervals on the production-transportation problem . . . . .	78
2.12	Intervals on the newsvendor problem . . . . .	78
2.13	Intervals on the production-transportation problem . . . . .	79
2.14	Intervals on the newsvendor problem . . . . .	79
2.15	Arbitrary vs. leading components on the newsvendor problem . . . . .	81
2.16	Sensitivity analysis for lower bound (2.15) on the newsvendor problem with respect to $\gamma_1$ . . . . .	82
2.17	Sensitivity analysis for lower bound (2.15) on the newsvendor problem with respect to $\gamma_2$ . . . . .	82
2.18	Sensitivity analysis for lower bound (2.60) on the newsvendor problem with respect to $R_0$ . . . . .	83
2.19	Lower bound (2.95) on the production-transportation problem . . . . .	92
2.20	Upper bound (2.97) on the production-transportation problem . . . . .	92
2.21	Lower bounds (2.60) and (2.95) on the production-transportation problem	92
2.22	Upper bounds (2.72) and (2.97) on the production-transportation problem	93
2.23	Sensitivity analysis for lower bound (2.95) with respect to $(\gamma_1, R_0)$ . . . . .	93
2.24	Sensitivity analysis for upper bound (2.95) with respect to $(\gamma_1, R_0)$ . . . . .	93
3.1	Notations used in the CTSP model . . . . .	109
3.2	Benders decomposition algorithm with a multi-cut decomposition scheme .	119
3.3	Size of test problems . . . . .	128
3.4	The CTSP model performance versus the deterministic model and TSP model performances . . . . .	129
3.5	Impact of risk tolerance on the CTSP model performance . . . . .	130
3.6	The comparison of the decomposition schemes . . . . .	132

LIST OF TABLES – *Continued*

3.7	Benders decomposition algorithm’s performance versus the default CPLEX performance . . . . .	132
3.8	Lower bounds and upper bounds of the case study problem obtained by the Benders decomposition algorithm . . . . .	133
3.9	The description of abbreviations used in the literature review tables . . .	136
3.10	Classification of the studies considering both disruption and operational risks . . . . .	138
3.11	Classification of studies (Continued) . . . . .	139
3.12	Classification of studies (Continued) . . . . .	140
4.1	The Knapsack problem with $W = 3000$ . . . . .	168
4.2	The Knapsack problem with $W = 4000$ . . . . .	168
4.3	The Knapsack problem with $W = 5000$ . . . . .	169
4.4	The power grid problem with $\rho_1 = 0.5$ . . . . .	170
4.5	The power grid problem with $\rho_1 = 0.5$ . . . . .	170
4.6	The power grid problem with $\rho_1 = 0.9$ . . . . .	171
4.7	Lower bounds for the power grid problem with $\rho_1 = 0.5$ . . . . .	171
4.8	Lower bounds for power grid problem with $\rho_1 = 0.9$ . . . . .	171

## ABSTRACT

Uncertainty poses a significant challenge to decision making in many real-world problems, especially when it is high-dimensional. For example, disruptive events such as the COVID-19 pandemic have caused significant uncertainties in the supply and demand for many important products such as Neodymium-iron-boron (NdFeB) magnets. To overcome such challenge, advanced optimization approaches such as stochastic programming (SP), robust optimization (RO), and distributionally robust optimization (DRO) have been developed in order to enable decision makers to find an optimal trade-off between risk and reward by including some knowledge of the uncertainty into their decision-making process. In this dissertation, we study computationally efficient approaches for SP, RO, and DRO and their applications.

First, we propose computationally efficient inner and outer approximations for DRO problems with a moment-based ambiguity set and a combined ambiguity set including Wasserstein distance and moment information. In these approximations, we split a random vector into smaller pieces, leading to smaller matrix constraints. In addition, we use principal component analysis (PCA) to shrink uncertainty space dimensionality. We quantify the quality of the developed approximations by deriving theoretical bounds on their optimality gap. We display the practical applicability of the proposed approximations in production-transportation and multi-product newsvendor problems. The results demonstrate that these approximations dramatically reduce computational time while maintaining high solution quality. The approximations also help construct an interval that includes the (unknown) optimal value for a large-scale DRO problem.

Next, we propose a resilient reverse supply chain and logistics network design

for recycling NdFeB magnets as a means to diversify the supply and alleviate the risks. We develop scenarios to model the unique impact of the COVID-19 pandemic on the proposed business, incorporating both disruption intensity and recovery rate. We formulate a chance-constrained two-stage SP model to maximize the profit while guaranteeing the network resiliency against disruption risks. The decision variables include facility locations, processing capacities, inventory levels, and material flows. To solve the problem in large-scale instances, we develop an efficient Benders decomposition algorithm. Finally, we apply the model to the United States to demonstrate the practical applicability of the proposed model and algorithm.

Lastly, we propose a systematic approach to develop data-driven polyhedral uncertainty sets that mitigates these drawbacks. The proposed uncertainty sets are polytopes induced by a given set of scenarios, capture correlation information between uncertain parameters, and allows for direct trade-offs between tractability and conservativeness issue of conventional polyhedral uncertainty sets. To develop these uncertainty sets, we use principal component analysis (PCA) to transform the correlated scenarios into their uncorrelated principal components and to shrink the uncertainty space dimensionality. Thus, decision-makers can use the number of the leading principal components as a tool to trade-off tractability, conservativeness, and robustness of RO models. We quantify the quality of the lower bound of a static RO problem with a scenario-induced uncertainty set by deriving a theoretical bound on the gap between the optimal value of this problem and that of its lower bound. Additionally, we derive probabilistic guarantees for the performance of the proposed scenario-induced uncertainty sets by developing explicit lower bounds on the number of scenarios required to obtain the desired guarantees. Finally, we demonstrate the practical applicability of the proposed uncertainty sets to trade-off tractability, robustness, and conservativeness by examining a range of knapsack and power grid problems.

## CHAPTER 1

# Introduction

In this chapter, we first introduce the research motivation and contributions of this dissertation. Then, we outline the organization of the dissertation.

### 1.1 Motivation

Due to the high complexity of the uncertainty involved in real-world problems, decision making in such problems is very challenging. Therefore, different modeling techniques such as stochastic programming (SP), robust optimization (RO), and distributionally robust optimization (DRO) have been developed that enable decision makers to include some knowledge of the uncertainty into their decision-making process. A decision maker can adopt SP to model a problem if the complete knowledge about the probability distribution of the uncertainty is available. Otherwise, one can leverage either RO or DRO to model the problem. However, these approaches are usually computationally expensive or too conservative. This dissertation aims to alleviate such drawbacks by proposing computationally efficient approaches for these modeling frameworks that are less conservative than their conventional alternatives. In this dissertation, we propose computationally efficient approximations for DRO problems that are tractable and maintain high solution quality. Moreover, we develop data-driven polyhedral uncertainty sets that result in RO models that are less conservative and more tractable than those with conventional polyhedral uncertainty sets.

Neodymium-iron-boron (NdFeB) magnets are the most powerful magnets whose

supply has been dominated by a few countries. This permanent magnet is made of rare-earth elements (REEs) which are classified as critical materials to the United States and European Union due to their increasing importance to clean energy and persistent supply risk. The COVID-19 pandemic, in addition to other black swan events such as geopolitical conflicts and international trade wars, has caused significant uncertainties in the global supply and demand of REEs and NdFeB magnets. Therefore, it is crucial to develop domestic sources for NdFeB magnets, such as magnet-to-magnet recycling, that are resilient to such uncertainties. NdFeB magnet recycling can not only reduce new REE consumption but also lower the environmental footprint of NdFeB magnets. In this dissertation, we design a resilient reverse supply chain and logistics network for recycling NdFeB magnets by developing a chance-constrained two-stage stochastic programming (CTSP) model. This model ensures the network resiliency to cope with demand and disruption uncertainties as its scenarios in mimic different types of supply and demand disruptions caused by the COVID-19 pandemic. Since this problem is NP-hard, we develop a customized Benders decomposition algorithm to solve the large-scale instances to optimality within a short time.

## 1.2 Contributions

This dissertation aims to develop computationally efficient approaches for SP, RO, and DRO problems. Moreover, it applies these approaches to some real-world problems, including a supply chain and logistics problem, a power grid problem, and a production-transportation problem, to demonstrate their practical applicability. In what follows, we point out the main contributions of this dissertation.

### 1.2.1 Computationally Efficient Approximations for DRO

1. We derive computationally efficient inner and outer approximations of DRO problems with a moment-based ambiguity set accounting for the support, mean, and covariance of the uncertainty. The inner approximation is based on splitting a random vector into smaller sub-vectors and is parameterized by the number of split pieces. Such approximation appears to be new in the DRO literature. The outer approximation generalizes Cheng et al. (2018).
2. We quantify the quality of our inner and outer approximations by deriving theoretical bounds on the gap between the optimal value of the DRO problems and those of their approximations. These theoretical bounds guide us to select specific numbers of split pieces for reaching a predetermined error bound. They also allow us to trade-off between solution quality and the computational burden of solving DRO formulations.
3. We extend the inner and outer approximations, as well as their theoretical bounds, to a combined ambiguity set that contains covariance information and the Wasserstein distance between the true distribution and an empirical distribution. We also investigate the corresponding results by additionally including the first-order moment information in the combined ambiguity set.
4. Our proposed inner and outer approximations together, while quickly finding a feasible solution with a small optimality gap, enable us to construct a tight interval that includes the (unknown) optimal value of DRO formulations. Such an interval is very helpful for decision making in many real-world applications with large-scale instances and high-dimensional uncertainties (e.g., energy and transportation), where the corresponding DRO model cannot be solved to optimality (or even feasibility in most cases) by existing methods in a reasonable time.
5. We perform extensive computational experiments to demonstrate the effec-

tiveness of our approximations in solving DRO formulations. Notably, while commercial solvers were unable to even find a feasible solution to most large-size instances, our inner and outer approximations quickly found solutions with optimality guarantee.

### 1.2.2 Resilient NdFeB Magnet Recycling

1. We design an NdFeB magnet recycling supply chain and logistics network that is resilient to operational and disruption risks triggered by black swan events such as the COVID-19 pandemic. We consider various risk management strategies, including dynamic material flow adjustment, backup facilities, buffer inventories, and minimum service level enforcement to reinforce the resiliency of the reverse supply network under disruptive events.
2. We develop a CTSP model tailored to NdFeB magnet recycling to maximize the total profit while keeping the supply chain resilient enough to disruptive events on the same scale as the COVID-19 pandemic. To the best of our knowledge, this is the first study that develops a CTSP model for a supply chain and logistics network in the REE and NdFeB magnet industry.
3. We design disruption scenarios that mimic the real-world COVID-19 pandemic impacts. Our scenarios are unique in the literature from several aspects. First, we consider disruption and recovery durations, while earlier studies have often ignored when a disruption starts and ends. Second, we consider both demand and supply disruptions that have different timings, as was evidenced by the COVID-19 pandemic. The traditional assumption of simultaneous supply and demand disruptions could be optimistic due to synchronous recovery (Ivanov 2020). Thirdly, we consider the post-disruption recovery process and model it using piecewise functions, as opposed to the typical linear recovery assumption, to capture the real-world scenarios where the recovery rates are non-linear. As far as we are aware, no study has considered both factors in this way.

Fourthly, we examine the rebound effect, which increases demand after the end of disruption. If suppliers do not prepare for this sudden surge in demand, they may lose a large portion of the potential sales.

4. We develop an efficient customized Benders decomposition algorithm to solve the large-scale CTSP model tailored for the NdFeB magnet recycling problem. Moreover, we propose different decomposition schemes that can be adopted to decompose this problem and equip the Benders decomposition algorithm with each of them to identify the most efficient one. Although commercial solvers fail to find even a feasible solution for most of the test instances due to the computational complexity of the problem, the developed Benders decomposition algorithm equipped with the most efficient proposed decomposition scheme can solve all test instances to optimality within a short time.

### 1.2.3 Data-driven RO Using Scenario-Induced Uncertainty Sets

1. We use PCA to propose a systematic approach for developing data-driven polyhedral uncertainty sets that alleviate the disadvantages of conventional polyhedral uncertainty sets. Unlike the box and budget uncertainty sets, they can capture the correlation information of uncertainty. Moreover, they are less conservative than the box uncertainty set and computationally cheaper than the convex hull of the uncertainty data. Furthermore, the number of the leading principal components in these uncertainty sets can be used as a tool to trade-off tractability, conservativeness, and robustness of RO models.
2. We quantify the quality of the lower bound of a static RO problem with a scenario-induced uncertainty set by deriving a theoretical bound on the gap between the optimal value of this problem and that of its lower bound. This theoretical bound provides a rough approximation for the optimal value of the static RO Problem, which may not be solved efficiently in practice. Moreover, it determines how many principal components are needed to reach a preferred

gap, demonstrating a trade-off between computational burden and solution quality.

3. We provide probabilistic guarantees for the performance of the proposed uncertainty sets by deriving explicit lower bounds on the number of scenarios required to construct the uncertainty sets with desired probabilistic performance, which complements the existing work.

### 1.3 Dissertation Organization

This dissertation includes 5 chapters and its remaining chapters are organized as follows:

In Chapter 2, we develop computationally efficient inner and outer approximations for DRO problems with moment-based and combined ambiguity sets. First, we briefly review the literature related to DRO. Then, we study DRO with the moment-based ambiguity set, develop its approximations, and derive theoretical bounds of their optimality gaps. Similarly, we propose the same approximations and theoretical bounds for DRO with the combined ambiguity set. Next, we conduct comprehensive computational experiments on a couple of applications to evaluate the strength of the proposed approximations and validity of the theoretical bounds.

In Chapter 3, we mainly focus on designing a supply chain and logistics network for recycling NdFeB magnets that is resilient to operational and disruption risks caused by the COVID-19 pandemic. First, we provide a concise overview of the NdFeB magnet industry and the negative impacts of black swan events such as the COVID-19 pandemic on this industry. We also review the literature related to supply chain risk management in the COVID-19 pandemic and supply chain and logistics network design under uncertainty. Then, we state the problem, and its assumptions, and develop a chance-constrained two-stage stochastic programming model for this problem. Besides, we apply the proposed model to a case study of the United States

to obtain some managerial insights. Finally, we elaborate on the developed Benders decomposition algorithm and perform computational experiments to evaluate the performance of the model and algorithm.

The main focus of Chapter 4 is on proposing a systematic approach to develop data-driven polyhedral uncertainty sets that alleviate certain drawbacks of conventional polyhedral uncertainty sets. In this chapter, after providing a review of RO and a concise background on conventional polyhedral uncertainty sets, we introduce these scenario-induced uncertainty sets. Then, we derive a theoretical bound on the gap between the optimal value of a static RO problem with a scenario-induced uncertainty set and that of its lower bound. Next, we derive lower bounds on the number of scenario samples required to achieve the desired probabilistic performance guarantees for the developed uncertainty sets. Finally, we conduct comprehensive computational experiments on RO Knapsack and power grid problems with the proposed uncertainty sets to evaluate their performance.

Finally, Chapter 5 concludes the dissertation with a summary of contributions and the potential directions for further research.

## CHAPTER 2

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

In this chapter, we propose computationally efficient inner and outer approximations for distributionally robust optimization (DRO) problems with moment-based and combined ambiguity sets, and quantify the quality of the developed approximations by deriving theoretical bounds on their optimality gap. We provide a review of DRO in Section 2.1. In Section 2.2 (resp. Section 2.3), we study DRO with the moment-based ambiguity set (resp. the combined ambiguity set), propose its inner and outer approximations, and derive theoretical bounds of their optimality gaps. In Section 2.4, we perform extensive computational experiments on distributionally robust multiproduct newsvendor and production-transportation problems to evaluate the theoretical results and demonstrate the strength of the proposed approximations. Finally, Section 2.5 concludes this chapter.

## 2.1 Introduction

Uncertainty poses significant challenge to decision making in many real-world problems. To overcome such challenge, advanced optimization approaches have been developed to model uncertainty from various perspectives. Among them, stochastic programming (SP), robust optimization (RO), and DRO prevail nowadays. SP assumes that a decision maker has complete knowledge about the probability distribution of the uncertain parameters, whereas the distribution may not be precisely

estimated due to limited data availability (Shapiro et al. 2009). RO assumes the uncertain parameters run in a given set, and it hedges against the worst-case possible scenario within this set, leading to potentially conservative decisions (Ben-Tal and Nemirovski 1998, Bertsimas and Sim 2004). Scarf (1958) introduced the first DRO model by relaxing the complete-knowledge assumption in SP and reducing the conservativeness of RO. DRO models uncertainty through a distributional ambiguity set that specifies available information of the probability distribution of the uncertain parameters. In addition, DRO searches for an optimal solution that concerns the worst-case distribution in the ambiguity set. Thus, the performance of DRO is less conservative than RO; see Rahimian and Mehrotra (2019) for more details.

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

Due to high complexity of the uncertainty involved in real-world problems, none of the individual ambiguity sets can perfectly perform under all circumstances. For example, moment-based ambiguity sets do not guarantee asymptotic consistency, i.e., they do not converge to the true distribution of the uncertain parameters even if the number of historical data points increases to infinity (Chen et al. 2019, Liu et al. 2019). Meanwhile, decision-makers usually face the difficulty of exactly estimating an ambiguity set because high-dimensional and correlated uncertainties are involved in the real-world problems, where different decision-makers may also have different understanding and estimates of the ambiguity. Thus, to cope with such situation and further give the decision-makers more flexibility and freedom in selecting an appropriate ambiguity set from available alternatives, we may better consider different types of ambiguity sets. For instance, we can combine two different types of ambiguity sets to construct a better one that enjoys the advantages of both. In particular, we may consider a combined moment and Wasserstein ambiguity set. This combination can help exclude pathological distributions and result in a less conservative DRO model, which is also asymptotically consistent. Such benefits can be significant when the uncertainty is highly complex (Wang et al. 2018, Gao and Kleywegt 2017).

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

(2018), and more SOCP reformulations were proposed by Li et al. (2018) and Mieth and Dvorkin (2018). Moreover, Hanasusanto and Kuhn (2018) proposed CP and CPP reformulations and approximations of two-stage DRO linear programs over Wasserstein ambiguity sets.

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

1. We use random vector splitting and PCA techniques to derive computationally efficient inner and outer approximations of DRO problems with a moment-based ambiguity set accounting for the support, mean, and covariance of the uncertainty as well as a combined ambiguity set that contain covariance information and the Wasserstein distance between the true distribution and an empirical distribution.
2. We quantify the quality of our inner and outer approximations by deriving theoretical bounds on their optimality gaps that allow us to trade-off between

solution quality and computational burden.

3. Our proposed inner and outer approximations together enable us to construct a tight interval that includes the (unknown) optimal value of DRO formulations.
4. We perform extensive computational experiments in the context of production-transportation and multi-product newsvendor problems to demonstrate the effectiveness of our approximations in solving DRO formulations.

### Notation

In this chapter, scalar values are denoted by non-bold symbols, e.g.,  $s$  and  $\gamma_1$ , while vectors are denoted in the column form by bold symbols, e.g.,  $\mathbf{x} = (x_1, \dots, x_m)^\top$  and  $\mathbf{q}$ . Similarly, matrices are represented by bold capital symbols, e.g.,  $\mathbf{A}$  and  $\mathbf{\Sigma}$ , and the size of a matrix is indicated by  $r \times c$ , where  $r$  and  $c$  indicate the numbers of rows and columns, respectively. Italic subscripts indicate indices, e.g.,  $s_k$ , while non-italic ones represent simplified specifications, e.g.,  $\mathbf{Q}_r$ . We use  $\mathbb{E}_{\mathbb{P}}[\cdot]$  to represent expectation over distribution  $\mathbb{P}$  and use “ $\bullet$ ” to denote the inner product defined by  $\mathbf{A} \bullet \mathbf{B} = \sum_{i,j} A_{ij} B_{ij}$ , where  $\mathbf{A}$  and  $\mathbf{B}$  are two conformal matrices. If a matrix  $\mathbf{M}$  is positive semi-definite (PSD), it is indicated by  $\mathbf{M} \succeq 0$ . Symbols  $\|\cdot\|_1$  and  $\|\cdot\|_2$  denote L1-Norm and L2-Norm, respectively. Symbol  $\|\cdot\|_*$  represents the dual norm of  $\|\cdot\|_1$ . We reserve symbols  $\mathcal{D}$  and  $\mathcal{S}$  for ambiguity set and support, respectively. For any strictly positive integer number  $n$ , we use  $[n]$  to represent the set  $\{1, 2, \dots, n\}$ . The identity matrix of size  $m$  is denoted by  $\mathbf{I}_m$ . Symbols  $\mathbf{0}_m$  and  $\mathbf{0}_{r \times c}$  represent a zero vector of size  $m$  and a zero matrix of size  $r \times c$ , respectively.

## 2.2 Moment-based Ambiguity Set

In this section, we introduce a DRO problem with a moment-based ambiguity set. To solve it towards practical uses, we first recast it as an SDP formulation. In view of the computational challenge of solving the SDP problem in practice, we develop

its inner and outer approximations that can be solved more efficiently. Moreover, we derive theoretical bounds for their optimality gaps as compared to the original DRO problem, leading to a quality measurement.

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

$$\min_{\boldsymbol{x} \in \mathcal{X}} \mathbb{E}_{\mathbb{P}} [f(\boldsymbol{x}, \boldsymbol{\xi})]. \quad (2.1)$$

Here  $\mathbb{P}$  is assumed to be known exactly, which though in practice may not be precisely estimated due to limited data availability (e.g., missing data, lack of data, and expensive data acquirement). Nevertheless, some partial information of  $\boldsymbol{\xi}$  (e.g., mean and covariance) can be easily obtained from historical data. Therefore, instead of solving Problem (2.1) with a given distribution, we may seek a risk-averse solution that hedges against all the possible distributions that share such available information, leading to the following DRO model with the available information collected in a distributional ambiguity set  $\mathcal{D}_{\text{M1}}$ :

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

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

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

which specifies the support ( $\mathcal{S}$ ), mean ( $\boldsymbol{\mu}$ ), and covariance of random variable  $\boldsymbol{\xi}$  that could be derived using available historical data. We assume that  $\mathcal{S}$  is a convex set,  $\boldsymbol{\mu}$  lies in the strict interior of  $\mathcal{S}$ , and  $\boldsymbol{\Sigma}$  is a positive definite matrix. Parameters  $\gamma_1 \geq 0$  and  $\gamma_2 \geq 1$  are derived from historical data to control the size of the ambiguity set

and the conservatism of optimal solutions. The three constraints in  $\mathcal{D}_{M1}$  describe that (i) the support of  $\boldsymbol{\xi}$  is a subset of  $\mathcal{S}$ ; (ii) the mean of  $\boldsymbol{\xi}$  lies in an ellipsoid of size  $\gamma_1$  centered at  $\boldsymbol{\mu}$ ; and (iii) the centered second-order moment matrix is bounded by  $\gamma_2 \boldsymbol{\Sigma}$  in a PSD sense.

Although (DRO-M) admits a convex reformulation (e.g., SDP reformulation; see Delage and Ye (2010)), as discussed above, solving it in practice can be very challenging. Instead, we can solve good inner and outer approximations of (DRO-M) much more efficiently and obtain high-quality solutions, thereby complementing the existing studies such as Delage and Ye (2010) and Cheng et al. (2018). More importantly, the inner and outer approximations together can help characterize an interval that includes the unknown optimal value for a large-scale instance that may not be solved to feasibility. In the following, we derive an outer (resp. inner) approximation of (DRO-M), leading to a lower (resp. upper) bound, in Section 2.2.1 (resp. Section 2.2.3). We make the following assumption in this section for practical purpose.

**Assumption 2.1.** *Function  $f(\mathbf{x}, \boldsymbol{\xi})$  is piecewise linear convex in  $\boldsymbol{\xi}$ , i.e.,  $f(\mathbf{x}, \boldsymbol{\xi}) = \max_{k=1}^K \{y_k^0(\mathbf{x}) + y_k(\mathbf{x})^\top \boldsymbol{\xi}\}$  with both  $y_k(\mathbf{x}) = (y_k^1(\mathbf{x}), \dots, y_k^m(\mathbf{x}))^\top$  and  $y_k^0(\mathbf{x})$  affine in  $\mathbf{x}$  for any  $k \in [K]$ , and  $\mathcal{S}$  is polyhedral, i.e.,  $\mathcal{S} = \{\boldsymbol{\xi} | \mathbf{A}\boldsymbol{\xi} \leq \mathbf{b}\}$  with  $\mathbf{A} \in \mathbb{R}^{n \times m}$  and  $\mathbf{b} \in \mathbb{R}^n$ , with at least one interior point.*

### 2.2.1 Lower Bound

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

uncertainty onto a lower dimensional space by preserving the components with the highest variance and relaxing the rest of ones.

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

$$\min_{\mathbf{x} \in \mathcal{X}} \max_{\mathbb{P}_I \in \mathcal{D}_{M2}} \mathbb{E}_{\mathbb{P}_I} \left[ f \left( \mathbf{x}, \mathbf{U}\Lambda^{\frac{1}{2}}\boldsymbol{\xi}_I + \boldsymbol{\mu} \right) \right], \quad (2.2)$$

where

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

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

**Theorem 2.1.** *If  $f(\mathbf{x}, \mathbf{U}\Lambda^{\frac{1}{2}}\boldsymbol{\xi}_I + \boldsymbol{\mu})$  is  $\mathbb{P}_I$ -integrable for any  $\mathbb{P}_I \in \mathcal{D}_{M2}$ , then (DRO-M) has the same optimal value as the following problem:*

$$\min_{\mathbf{x}, s, \mathbf{q}, \mathbf{Q}} s + \gamma_2 \mathbf{I}_m \bullet \mathbf{Q} + \sqrt{\gamma_1} \|\mathbf{q}\|_2 \quad (2.3a)$$

$$\text{s.t. } s \geq f \left( \mathbf{x}, \mathbf{U}\Lambda^{\frac{1}{2}}\boldsymbol{\xi}_I + \boldsymbol{\mu} \right) - \boldsymbol{\xi}_I^\top \mathbf{q} - \boldsymbol{\xi}_I^\top \mathbf{Q} \boldsymbol{\xi}_I, \quad \forall \boldsymbol{\xi}_I \in \mathcal{S}_I, \quad (2.3b)$$

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

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

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

Problem (2.3) reduces to a SDP formulation with regard to a wide range of objective functions and support of uncertainty, which are specified in Assumption 2.1. Our approximation techniques may be applied to DRO problems with more general objective functions.

**Proposition 2.1.** *Under Assumption 2.1, (DRO-M) has the same optimal value as the following SDP formulation:*

$$Z_M^*(m) := \min_{\mathbf{x}, s, \hat{\boldsymbol{\lambda}}, \mathbf{q}, \mathbf{Q}} s + \gamma_2 \mathbf{I}_m \bullet \mathbf{Q} + \sqrt{\gamma_1} \|\mathbf{q}\|_2 \quad (2.4a)$$

$$\text{s.t.} \quad \begin{bmatrix} s - y_k^0(\mathbf{x}) - \boldsymbol{\lambda}_k^\top \mathbf{b} - y_k(\mathbf{x})^\top \boldsymbol{\mu} + \boldsymbol{\lambda}_k^\top \mathbf{A} \boldsymbol{\mu} & \frac{1}{2} \left( \mathbf{q} + (\mathbf{U} \boldsymbol{\Lambda}^{\frac{1}{2}})^\top (\mathbf{A}^\top \boldsymbol{\lambda}_k - y_k(\mathbf{x})) \right)^\top \\ \frac{1}{2} \left( \mathbf{q} + (\mathbf{U} \boldsymbol{\Lambda}^{\frac{1}{2}})^\top (\mathbf{A}^\top \boldsymbol{\lambda}_k - y_k(\mathbf{x})) \right) & \mathbf{Q} \end{bmatrix} \succeq 0, \quad \forall k \in [K], \quad (2.4b)$$

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

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

*Proof.* We apply the strong duality theorem to constraints (2.3b). As function  $f(\mathbf{x}, \boldsymbol{\xi})$  is piecewise linear convex, we reformulate constraints (2.3b) as:

$$s \geq y_k^0(\mathbf{x}) + y_k(\mathbf{x})^\top \boldsymbol{\xi} - \boldsymbol{\xi}^\top \mathbf{q} - \boldsymbol{\xi}^\top \mathbf{Q} \boldsymbol{\xi}, \quad \forall \boldsymbol{\xi} \in \mathcal{S}, \quad \forall k \in [K], \quad (2.5)$$

which are equivalent to  $\min_{\mathbf{A} \boldsymbol{\xi} \leq \mathbf{b}, \boldsymbol{\xi} \in \mathbb{R}^m} g_k(\boldsymbol{\xi}) \geq 0$ , where  $g_k(\boldsymbol{\xi}) = s + \boldsymbol{\xi}^\top \mathbf{q} + \boldsymbol{\xi}^\top \mathbf{Q} \boldsymbol{\xi} - y_k^0(\mathbf{x}) - y_k(\mathbf{x})^\top \boldsymbol{\xi}$ , for any  $k \in [K]$ . Moreover, we consider the Lagrangian dual problem of  $\min_{\mathbf{A} \boldsymbol{\xi} \leq \mathbf{b}, \boldsymbol{\xi} \in \mathbb{R}^m} g_k(\boldsymbol{\xi})$ , i.e.,  $\max_{\boldsymbol{\lambda}_k \geq 0} \inf_{\boldsymbol{\xi}} g_k(\boldsymbol{\xi}) + \boldsymbol{\lambda}_k^\top (\mathbf{A} \boldsymbol{\xi} - \mathbf{b})$ , where  $\boldsymbol{\lambda}_k \in \mathbb{R}^n$ . Note that function  $g_k(\boldsymbol{\xi})$  is convex in  $\boldsymbol{\xi}$  because  $\mathbf{Q} \succeq 0$ . Due to Assumption 2.1, there exists an interior point for the primal problem. It follows that constraints (2.5) are equivalent to the following ones:

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

which are further equivalent to the following constraints:

$$\exists \boldsymbol{\lambda}_k \geq 0, \quad s + \boldsymbol{\xi}^\top \mathbf{q} + \boldsymbol{\xi}^\top \mathbf{Q} \boldsymbol{\xi} - y_k^0(\mathbf{x}) - y_k(\mathbf{x})^\top \boldsymbol{\xi} + \boldsymbol{\lambda}_k^\top (\mathbf{A} \boldsymbol{\xi} - \mathbf{b}) \geq 0, \quad \forall \boldsymbol{\xi} \in \mathbb{R}^m, \quad (2.6) \\ \forall k \in [K].$$

As  $\boldsymbol{\xi} = \mathbf{U} \boldsymbol{\Lambda}^{\frac{1}{2}} \boldsymbol{\xi}_1 + \boldsymbol{\mu}$ , we replace  $\boldsymbol{\xi}$  with  $\mathbf{U} \boldsymbol{\Lambda}^{\frac{1}{2}} \boldsymbol{\xi}_1 + \boldsymbol{\mu}$  in (2.6). Thus, we have

$$(2.6) \Leftrightarrow \exists \boldsymbol{\lambda}_k \geq 0, \quad (1, \boldsymbol{\xi}_1^\top) \mathbf{Z}_k (1, \boldsymbol{\xi}_1^\top)^\top \geq 0, \quad \forall \boldsymbol{\xi}_1 \in \mathbb{R}^m, \quad \forall k \in [K], \quad (2.7)$$

$$\Leftrightarrow \exists \boldsymbol{\lambda}_k \geq 0, \mathbf{Z}_k \succeq 0, \forall k \in [K], \quad (2.8)$$

where

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

and the first equivalence holds due to the definition of  $\mathbf{Z}_k$ . For the second equivalence, clearly  $\Leftarrow$  follows from the definition of a PSD matrix. To prove  $\Rightarrow$ , we consider two possible cases for any  $(\eta_0 \in \mathbb{R}, \boldsymbol{\eta}^\top \in \mathbb{R}^m)^\top \in \mathbb{R}^{m+1}$ : (1) if  $\eta_0 = 0$ , then  $(\eta_0, \boldsymbol{\eta}^\top) \mathbf{Z}_k (\eta_0, \boldsymbol{\eta}^\top)^\top = \boldsymbol{\eta}^\top \mathbf{Q} \boldsymbol{\eta} \geq 0$  because  $\mathbf{Q}$  is PSD; (2) if  $\eta_0 \neq 0$ , then we have  $(\eta_0, \boldsymbol{\eta}^\top) \mathbf{Z}_k (\eta_0, \boldsymbol{\eta}^\top)^\top = \eta_0^2 (1, \frac{\boldsymbol{\eta}^\top}{\eta_0}) \mathbf{Z}_k (1, \frac{\boldsymbol{\eta}^\top}{\eta_0})^\top \geq 0$  according to (2.7). Therefore,  $\Rightarrow$  holds and we obtain Problem (2.4) by replacing constraints (2.3b) with (2.8).  $\square$

Next, to derive a lower bound, we approximate  $\boldsymbol{\xi}$  by capturing the dominant variability of  $\mathbf{U} \boldsymbol{\Lambda}^{\frac{1}{2}} \boldsymbol{\xi}_1$  through considering only the first  $m_1$  random variables of  $\boldsymbol{\xi}_1$ , i.e.,

$$\boldsymbol{\xi} \approx \mathbf{U} \boldsymbol{\Lambda}^{\frac{1}{2}} [\boldsymbol{\xi}_r; \mathbf{0}_{m-m_1}] + \boldsymbol{\mu} = \mathbf{U}_{m \times m_1} \boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}} \boldsymbol{\xi}_r + \boldsymbol{\mu}, \quad (2.9)$$

where  $\mathbf{U}_{m \times m_1} \in \mathbb{R}^{m \times m_1}$  and  $\boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}} \in \mathbb{R}^{m_1 \times m_1}$  are upper-left submatrices of  $\mathbf{U}$  and  $\boldsymbol{\Lambda}^{\frac{1}{2}}$ , respectively, and  $\boldsymbol{\xi}_r \in \mathbb{R}^{m_1}$  consists of the first  $m_1$  entries of  $\boldsymbol{\xi}_1$ . As the uncertainty of the last  $(m - m_1)$  entries of  $\boldsymbol{\xi}_1$  vanishes, this yields a relaxation of (DRO-M):

$$\min_{\mathbf{x} \in \mathcal{X}} \max_{\mathbb{P}_r \in \mathcal{D}_{M3}} \mathbb{E}_{\mathbb{P}_r} \left[ f \left( \mathbf{x}, \mathbf{U}_{m \times m_1} \boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}} \boldsymbol{\xi}_r + \boldsymbol{\mu} \right) \right], \quad (2.10a)$$

where

$$\mathcal{D}_{M3}(\mathcal{S}_r, \gamma_1, \gamma_2) = \left\{ \mathbb{P}_r \left| \begin{array}{l} \mathbb{P}_r(\boldsymbol{\xi}_r \in \mathcal{S}_r) = 1 \\ \mathbb{E}_{\mathbb{P}_r}[\boldsymbol{\xi}_r^\top] \mathbb{E}_{\mathbb{P}_r}[\boldsymbol{\xi}_r] \leq \gamma_1 \\ \mathbb{E}_{\mathbb{P}_r}[\boldsymbol{\xi}_r \boldsymbol{\xi}_r^\top] \preceq \gamma_2 \mathbf{I}_{m_1} \end{array} \right. \right\} \quad (2.10b)$$

with

$$\mathcal{S}_r := \left\{ \boldsymbol{\xi}_r \in \mathbb{R}^{m_1} : \mathbf{U}_{m \times m_1} \boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}} \boldsymbol{\xi}_r + \boldsymbol{\mu} \in \mathcal{S} \right\}. \quad (2.10c)$$

**Theorem 2.2.** *If  $f(\mathbf{x}, \mathbf{U}_{m \times m_1} \Lambda_{m_1}^{\frac{1}{2}} \boldsymbol{\xi}_r + \boldsymbol{\mu})$  is  $\mathbb{P}_r$ -integrable for any  $\mathbb{P}_r \in \mathcal{D}_{M3}$ , then Problem (2.10) has the same optimal value as the following problem:*

$$\begin{aligned} \min_{\mathbf{x}, s, \mathbf{q}_r, \mathbf{Q}_r} \quad & s + \gamma_2 \mathbf{I}_{m_1} \bullet \mathbf{Q}_r + \sqrt{\gamma_1} \|\mathbf{q}_r\|_2 \\ \text{s.t.} \quad & s \geq f\left(\mathbf{x}, \mathbf{U}_{m \times m_1} \Lambda_{m_1}^{\frac{1}{2}} \boldsymbol{\xi}_r + \boldsymbol{\mu}\right) - \boldsymbol{\xi}_r^\top \mathbf{q}_r - \boldsymbol{\xi}_r^\top \mathbf{Q}_r \boldsymbol{\xi}_r, \quad \forall \boldsymbol{\xi}_r \in \mathcal{S}_r, \\ & \mathbf{Q}_r \succeq 0, \quad \mathbf{x} \in \mathcal{X}, \end{aligned} \quad (2.11)$$

where  $\mathbf{q}_r \in \mathbb{R}^{m_1}$  and  $\mathbf{Q}_r \in \mathbb{R}^{m_1 \times m_1}$ . Furthermore, we have the following: (i) Problem (2.11) provides a lower bound for the optimal value of (DRO-M); (ii) the optimal value of Problem (2.11) is nondecreasing in  $m_1$ ; and (iii) if  $m_1 = m$ , then (DRO-M) and (2.11) have the same optimal value.

*Proof.* The proof of the deterministic reformulation (2.11) is the same as that of Theorem 2.1 and thus is omitted here. With (2.11), we define  $\boldsymbol{\zeta} = \mathbf{U}_{m \times m_1} \Lambda_{m_1}^{\frac{1}{2}} \boldsymbol{\xi}_r + \boldsymbol{\mu}$  and use  $\mathcal{S}_\zeta$  and  $\mathcal{D}_\zeta$  to denote its support and ambiguity set, respectively. As  $\mathcal{S}_r = \{\boldsymbol{\xi}_r \in \mathbb{R}^{m_1} : \mathbf{U}_{m \times m_1} \Lambda_{m_1}^{\frac{1}{2}} \boldsymbol{\xi}_r + \boldsymbol{\mu} \in \mathcal{S}\}$  and  $\mathcal{S}_\zeta = \{\boldsymbol{\zeta} \in \mathbb{R}^m : \boldsymbol{\zeta} = \mathbf{U}_{m \times m_1} \Lambda_{m_1}^{\frac{1}{2}} \boldsymbol{\xi}_r + \boldsymbol{\mu}, \boldsymbol{\xi}_r \in \mathcal{S}_r\}$ , we can deduce  $\mathcal{S}_\zeta \subset \mathcal{S}$ . We also have

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

where the last inequality is due to (2.10b). Note that

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

It follows that

$$(\mathbb{E}_{\mathbb{P}_\zeta} [\boldsymbol{\zeta}] - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbb{E}_{\mathbb{P}_\zeta} [\boldsymbol{\zeta}] - \boldsymbol{\mu}) \leq \gamma_1. \quad (2.12)$$

Meanwhile, we have

$$\begin{aligned} \mathbb{E}_{\mathbb{P}_\zeta} [(\zeta - \boldsymbol{\mu})(\zeta - \boldsymbol{\mu})^\top] &\preceq \gamma_2 \mathbf{U}_{m \times m_1} \boldsymbol{\Lambda}_{m_1} \mathbf{U}_{m \times m_1}^\top \\ &= \gamma_2 \mathbf{U} \begin{bmatrix} \boldsymbol{\Lambda}_{m_1} & \mathbf{0}_{m_1 \times (m-m_1)} \\ \mathbf{0}_{(m-m_1) \times m_1} & \mathbf{0}_{(m-m_1) \times (m-m_1)} \end{bmatrix} \mathbf{U}^\top \preceq \gamma_2 \mathbf{U} \boldsymbol{\Lambda} \mathbf{U}^\top = \gamma_2 \boldsymbol{\Sigma}. \end{aligned} \quad (2.13)$$

By  $\mathcal{S}_\zeta \subset \mathcal{S}$ , (2.12), and (2.13), it follows that  $\mathcal{D}_\zeta$  lies in  $\mathcal{D}_{M1}$ , i.e.,  $\mathcal{D}_\zeta \subset \mathcal{D}_{M1}$ , and accordingly

$$\max_{\mathbb{P}_\zeta \in \mathcal{D}_\zeta} \mathbb{E}_{\mathbb{P}_\zeta} [f(\mathbf{x}, \zeta)] \leq \max_{\mathbb{P} \in \mathcal{D}_{M1}} \mathbb{E}_{\mathbb{P}} [f(\mathbf{x}, \boldsymbol{\xi})]. \quad (2.14)$$

By the definition of  $\zeta$ , we have  $\max_{\mathbb{P}_\zeta \in \mathcal{D}_\zeta} \mathbb{E}_{\mathbb{P}_\zeta} [f(\mathbf{x}, \zeta)] = \max_{\mathbb{P}_r \in \mathcal{D}_{M3}} \mathbb{E}_{\mathbb{P}_r} [f(\mathbf{x}, \mathbf{U}_{m \times m_1} \boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}} \boldsymbol{\xi}_r + \boldsymbol{\mu})]$  due to change of variable. Then (2.14) implies the following inequality

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

which demonstrates that the optimal value of Problem (2.11) (i.e., Problem (2.10)) is a lower bound for that of Problem (2.3) (i.e., Problem (2.2)).

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

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

where  $\mathcal{D}_{r_i}$  (defined as (2.10b)) represents the ambiguity set of  $\boldsymbol{\xi}_{r_i}$  for any  $i \in [2]$ . For any  $\zeta_1 \sim \mathbb{P}_{\zeta_1} \in \mathcal{D}_{\zeta_1}$ , there exists a  $\boldsymbol{\xi}_{r_1} \sim \mathbb{P}_{r_1} \in \mathcal{D}_{r_1}$  such that  $\zeta_1 = \mathbf{U}_{m \times m_1} \boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}} \boldsymbol{\xi}_{r_1} + \boldsymbol{\mu} = \mathbf{U}_{m \times m_2} \boldsymbol{\Lambda}_{m_2}^{\frac{1}{2}} \bar{\boldsymbol{\xi}}_{r_2} + \boldsymbol{\mu}$ , where  $\bar{\boldsymbol{\xi}}_{r_2} = (\boldsymbol{\xi}_{r_1}^\top, \mathbf{0}_{m_2-m_1}^\top)^\top \in \mathbb{R}^{m_2}$ . By using  $\mathcal{S}_{r_i}$  (defined as (2.10c)) to denote the support of  $\boldsymbol{\xi}_{r_i}$  for any  $i \in [2]$ , we have

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

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

and

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

It follows that the distribution of  $\bar{\boldsymbol{\xi}}_{r_2}$  belongs to  $\mathcal{D}_{r_2}$  and thus  $\mathbb{P}_{\zeta_1} \in \mathcal{D}_{\zeta_2}$ . Therefore, we have  $\mathcal{D}_{\zeta_1} \subset \mathcal{D}_{\zeta_2}$  and

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

That is, the optimal value of Problem (2.11) is nondecreasing in  $m_1$ .

Finally, Problem (2.10) is equivalent to Problem (2.2) when  $m_1 = m$ . Then, Problem (2.11) results in an exact reformulation of Problem (2.2) by Theorem 2.1.  $\square$

**Proposition 2.2.** *Under Assumption 2.1, Problem (2.11) has the same optimal value as the following SDP formulation*

$$Z_M^*(m_1) := \min_{\substack{\mathbf{x}, s, \hat{\boldsymbol{\lambda}}, \\ \mathbf{q}_r, \mathbf{Q}_r}} s + \gamma_2 \mathbf{I}_{m_1} \bullet \mathbf{Q}_r + \sqrt{\gamma_1} \|\mathbf{q}_r\|_2 \quad (2.15a)$$

$$\text{s.t.} \quad \begin{bmatrix} s - y_k^0(\mathbf{x}) - \boldsymbol{\lambda}_k^\top b - y_k(\mathbf{x})^\top \boldsymbol{\mu} + \boldsymbol{\lambda}_k^\top \mathbf{A} \boldsymbol{\mu} & \frac{1}{2} \left( \mathbf{q}_r + \left( \mathbf{U}_{m \times m_1} \boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}} \right)^\top (\mathbf{A}^\top \boldsymbol{\lambda}_k - y_k(\mathbf{x})) \right)^\top \\ \frac{1}{2} \left( \mathbf{q}_r + \left( \mathbf{U}_{m \times m_1} \boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}} \right)^\top (\mathbf{A}^\top \boldsymbol{\lambda}_k - y_k(\mathbf{x})) \right) & \mathbf{Q}_r \end{bmatrix} \succeq \mathbf{0}, \quad \forall k \in [K], \quad (2.15b)$$

$$\mathbf{x} \in \mathcal{X}, \quad \boldsymbol{\lambda}_k \in \mathbb{R}_+^n, \quad \forall k \in [K], \quad (2.15c)$$

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

*Proof.* The proof is similar with that of Proposition 2.1 and thus is omitted here.  $\square$

Comparing Problems (2.4) and (2.15) in terms of size, one can observe that Problem (2.15) is significantly easier to solve than Problem (2.4) because: (i) Problem (2.15) includes fewer decision variables than Problem (2.4), i.e.,  $(m_1^2 + m_1 + 2n + 1)$  vs.  $(m^2 + m + 2n + 1)$ ; and (ii) the size of PSD matrices in Problem (2.15) is smaller than in Problem (2.4), i.e.,  $(m_1 + 1) \times (m_1 + 1)$  vs.  $(m + 1) \times (m + 1)$ .

## 2.2.2 Lower Bound Quality

To measure the quality of our derived lower bound, i.e.,  $Z_M^*(m_1)$  in (2.15), we develop a theoretical upper bound for the gap between the optimal values of Problems (2.4) and (2.15). This upper bound brings two benefits: (i) it provides a rough approximation for the optimal value of Problem (2.4), which may not be solved efficiently in practice; and (ii) it determines how many principal components are required to reach a preferred gap between the original and approximated optimal values, indicating a trade-off between solution quality and computational time.

**Proposition 2.3.** *It holds that*

$$0 \leq Z_M^*(m) - Z_M^*(m_1) \leq \sqrt{\gamma_2} \sum_{k=1}^K \sqrt{\sum_{i=m_1+1}^m \Lambda_{i,i} \left( (\mathbf{A}^\top \boldsymbol{\lambda}_k^* - y_k(\mathbf{x}^*))^\top \mathbf{U}_i \right)^2}, \quad (2.16)$$

where  $\mathbf{U}_i$  represents the  $i^{\text{th}}$  column of matrix  $\mathbf{U}$ , and  $\mathbf{x}^*$  and  $\boldsymbol{\lambda}_k^*$  ( $\forall k \in [K]$ ) are optimal solutions of Problem (2.15).

*Proof.* By Theorem 2.2, we have  $Z_M^*(m) - Z_M^*(m_1) \geq 0$ . Meanwhile, when  $m_1 = m$ , Problem (2.15) is equivalent to Problem (2.4). We use  $(\mathbf{x}^*, s^*, \boldsymbol{\lambda}_k^* \forall k \in [K], \mathbf{q}_r^*, \mathbf{Q}_r^*)$  to denote an optimal solution of Problem (2.15). Based on this optimal solution, we construct a feasible solution of Problem (2.4), represented by  $(\bar{\mathbf{x}}, \bar{s}, \bar{\boldsymbol{\lambda}}_k \forall k \in [K], \bar{\mathbf{q}}, \bar{\mathbf{Q}})$ . For clarity, we define

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

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

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

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

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

where  $s_1^k > 0$  and  $s_2^k > 0$  for any  $k \in [K]$ . As  $\bar{\mathbf{x}} = \mathbf{x}^* \in \mathcal{X}$  and  $\bar{\boldsymbol{\lambda}}_k = \boldsymbol{\lambda}_k^* \in \mathbb{R}_+^n$ , for any  $k \in [K]$ , due to constraint (2.15c), we only require  $(\bar{\mathbf{x}}, \bar{s}, \bar{\boldsymbol{\lambda}}_k \forall k \in [K], \bar{\mathbf{q}}, \bar{\mathbf{Q}})$  to satisfy (2.4b). Thus, we will find the values of  $s_1^k$  and  $s_2^k$  for any  $k \in [K]$  that enable this solution to satisfy (2.4b).

We plug  $(\bar{\mathbf{x}}, \bar{s}, \bar{\boldsymbol{\lambda}}_k \forall k \in [K], \bar{\mathbf{q}}, \bar{\mathbf{Q}})$  to (2.4b) and use  $\bar{\mathbf{Y}}^k$  for any  $k \in [K]$  to denote the corresponding matrix in (2.4b). For any given  $k \in [K]$ , we perform the following decomposition:

$$\begin{aligned} \bar{\mathbf{Y}}^k &= \begin{bmatrix} S^k & \frac{1}{2}(\mathbf{q}_r^* + \mathbf{q}_{m_1}^k)^\top & \mathbf{0}_{1 \times (m-m_1)} \\ \frac{1}{2}(\mathbf{q}_r^* + \mathbf{q}_{m_1}^k) & \mathbf{Q}_r^* & \mathbf{0}_{m_1 \times (m-m_1)} \\ \mathbf{0}_{(m-m_1) \times 1} & \mathbf{0}_{(m-m_1) \times m_1} & \mathbf{0}_{(m-m_1) \times (m-m_1)} \end{bmatrix} + \begin{bmatrix} \sum_{k=1}^K s_1^k & \mathbf{0}_{1 \times m_1} & \frac{1}{2} \mathbf{q}_{m-m_1}^k \top \\ \mathbf{0}_{m_1 \times 1} & \mathbf{0}_{m_1 \times m_1} & \mathbf{0}_{m_1 \times (m-m_1)} \\ \frac{1}{2} \mathbf{q}_{m-m_1}^k & \mathbf{0}_{(m-m_1) \times m_1} & \sum_{k=1}^K \frac{s_2^k}{4} \mathbf{q}_{m-m_1}^k (\mathbf{q}_{m-m_1}^k)^\top \end{bmatrix} \\ &\succeq \begin{bmatrix} S^k & \frac{1}{2}(\mathbf{q}_r^* + \mathbf{q}_{m_1}^k)^\top & \mathbf{0}_{1 \times (m-m_1)} \\ \frac{1}{2}(\mathbf{q}_r^* + \mathbf{q}_{m_1}^k) & \mathbf{Q}_r^* & \mathbf{0}_{m_1 \times (m-m_1)} \\ \mathbf{0}_{(m-m_1) \times 1} & \mathbf{0}_{(m-m_1) \times m_1} & \mathbf{0}_{(m-m_1) \times (m-m_1)} \end{bmatrix} + \begin{bmatrix} s_1^k & \mathbf{0}_{1 \times m_1} & \frac{1}{2} \mathbf{q}_{m-m_1}^k \top \\ \mathbf{0}_{m_1 \times 1} & \mathbf{0}_{m_1 \times m_1} & \mathbf{0}_{m_1 \times (m-m_1)} \\ \frac{1}{2} \mathbf{q}_{m-m_1}^k & \mathbf{0}_{(m-m_1) \times m_1} & \frac{s_2^k}{4} \mathbf{q}_{m-m_1}^k (\mathbf{q}_{m-m_1}^k)^\top \end{bmatrix} \end{aligned} \quad (2.17)$$

The first matrix in (2.17) is clearly PSD because the elimination of its zero components leads to a PSD matrix due to constraint (2.15b). Now we find the values of  $s_1^k$  and  $s_2^k$  to make the second matrix PSD as well, and then accordingly the constructed solution is feasible for (2.4).

Next, we use  $\begin{bmatrix} \bar{\mathbf{A}} & \bar{\mathbf{B}} \\ \bar{\mathbf{B}}^\top & \bar{\mathbf{C}} \end{bmatrix}$  to denote second matrix in (2.17) by letting  $\bar{\mathbf{A}} = s_1^k$ ,  $\bar{\mathbf{B}}^\top = (\mathbf{0}_{1 \times m_1} \frac{1}{2} \mathbf{q}_{m-m_1}^k \top)$ , and  $\bar{\mathbf{C}} = \begin{bmatrix} \mathbf{0}_{m_1 \times m_1} & \mathbf{0}_{m_1 \times (m-m_1)} \\ \mathbf{0}_{(m-m_1) \times m_1} & \frac{s_2^k}{4} \mathbf{q}_{m-m_1}^k (\mathbf{q}_{m-m_1}^k)^\top \end{bmatrix}$ . It follows that

$$\begin{aligned} \bar{\mathbf{C}} - \bar{\mathbf{B}} \bar{\mathbf{A}}^{-1} \bar{\mathbf{B}}^\top &= \begin{bmatrix} \mathbf{0}_{m_1 \times m_1} & \mathbf{0}_{m_1 \times (m-m_1)} \\ \mathbf{0}_{(m-m_1) \times m_1} & \frac{s_2^k}{4} \mathbf{q}_{m-m_1}^k (\mathbf{q}_{m-m_1}^k)^\top \end{bmatrix} - \frac{1}{s_1^k} (\mathbf{0}_{1 \times m_1} \frac{1}{2} \mathbf{q}_{m-m_1}^k \top)^\top (\mathbf{0}_{1 \times m_1} \frac{1}{2} \mathbf{q}_{m-m_1}^k \top) \\ &= \begin{bmatrix} \mathbf{0}_{m_1 \times m_1} & \mathbf{0}_{m_1 \times (m-m_1)} \\ \mathbf{0}_{(m-m_1) \times m_1} & \left( \frac{s_2^k}{4} - \frac{1}{4s_1^k} \right) \mathbf{q}_{m-m_1}^k (\mathbf{q}_{m-m_1}^k)^\top \end{bmatrix}, \end{aligned} \quad (2.18)$$

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

In addition, since Problem (2.4) is a minimization problem, its optimal value is no larger than the objective corresponding to the feasible solution

$(\bar{\mathbf{x}}, \bar{s}, \bar{\boldsymbol{\lambda}}_k \forall k \in [K], \bar{\mathbf{q}}, \bar{\mathbf{Q}})$ . That is,

$$\begin{aligned} Z^*_M(m) &\leq \bar{s} + \gamma_2 \mathbf{I}_m \bullet \bar{\mathbf{Q}} + \sqrt{\gamma_1} \|\bar{\mathbf{q}}\| \\ &= Z^*_M(m_1) + \sum_{k=1}^K s_1^k + \gamma_2 \sum_{k=1}^K \frac{s_2^k}{4} \text{trace}(\mathbf{q}_{m-m_1}^k (\mathbf{q}_{m-m_1}^k)^\top). \end{aligned} \quad (2.19)$$

Due to the condition  $s_1^k \times s_2^k \geq 1$ , we let  $s_1^k = \frac{\sqrt{\gamma_2 (\mathbf{q}_{m-m_1}^k)^\top \mathbf{q}_{m-m_1}^k}}{2}$  and  $s_2^k = \frac{2}{\sqrt{\gamma_2 (\mathbf{q}_{m-m_1}^k)^\top \mathbf{q}_{m-m_1}^k}}$ , which leads to the smallest possible value of the RHS of (2.19). Therefore, we have

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

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

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

□

**Remark 2.1.** Note that Cheng et al. (2018) derived a similar upper bound by specifically considering  $\gamma_1 = 0$  and  $\gamma_2 = 1$ , while (2.16) applies to general values of  $\gamma_1$  and  $\gamma_2$ .

Before closing this subsection, we can observe that theoretical optimality gap (2.16) does not explicitly depend on parameter  $\gamma_1$ . It is because we develop this gap based on a specific feasible solution of Problem (2.4), i.e.,  $(\bar{\mathbf{x}}, \bar{s}, \bar{\boldsymbol{\lambda}}_k \forall k \in [K], \bar{\mathbf{q}}, \bar{\mathbf{Q}})$  where  $\bar{\mathbf{q}} = (\mathbf{q}_r^{\ast \top}, \mathbf{0}_{m-m_1}^\top)^\top$ . By setting the last  $m - m_1$  elements of  $\bar{\mathbf{q}}$  to be 0, we eventually obtain the optimality gap that does not explicitly depend on parameter  $\gamma_1$ . Technically we can construct a different feasible solution and correspondingly

develop a possibly better optimality gap, which will explicitly depend on  $\gamma_1$  though. For instance, we can let  $\bar{\mathbf{x}} = \mathbf{x}^*$ ,  $\bar{\boldsymbol{\lambda}}_k = \boldsymbol{\lambda}_k^*$  for any  $k \in [K]$ ,  $\bar{\mathbf{q}} = (\mathbf{q}_r^{*\top}, \hat{\mathbf{q}}^\top)^\top$ ,  $\bar{s} = s^* + \sum_{k=1}^K s^k$ , and

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

where  $s^k > 0$  for any  $k \in [K]$ . Instead of fixing the values of  $\hat{\mathbf{q}}$  and  $\hat{\mathbf{Q}}$  as what do in the proof of Proposition 2.3, we can optimize the values of  $s^k, \forall k \in [K]$ ,  $\hat{\mathbf{q}}$ , and  $\hat{\mathbf{Q}}$  together towards minimizing the difference between the objective value corresponding to this feasible solution and the original optimal value  $Z^*_M(m)$ , leading to a better gap denoted by  $Z^1_{\text{gap}}(m_1)$ . That is, we solve the following optimization problem

$$\begin{aligned} Z^1_{\text{gap}}(m_1) := & \min_{\substack{s^k, \forall k \in [K], \\ \hat{\mathbf{q}}, \hat{\mathbf{Q}}}} \sum_{k=1}^K s^k + \gamma_2 \mathbf{I}_{m-m_1} \bullet \hat{\mathbf{Q}} + \sqrt{\gamma_1} \|\hat{\mathbf{q}}\|_2 & (2.20) \\ \text{s.t.} & \begin{bmatrix} s^k & \frac{1}{2} (\mathbf{q}_{m-m_1}^k - \hat{\mathbf{q}})^\top \\ \frac{1}{2} (\mathbf{q}_{m-m_1}^k - \hat{\mathbf{q}}) & \hat{\mathbf{Q}} \end{bmatrix} \succeq 0, \forall k \in [K], \\ & s^k \in \mathbb{R}_+, \forall k \in [K], \hat{\mathbf{q}} \in \mathbb{R}^{m-m_1}, \hat{\mathbf{Q}} \in \mathbb{R}^{(m-m_1) \times (m-m_1)}, \end{aligned}$$

where the SDP constraints are enforced to ensure the constructed solution to be feasible, similar to what we do in the proof of Proposition 2.3.

Solving Problem (2.20) clearly can give us a better theoretical error bound than (2.16). We also notice that (2.20) can be computationally intensive, particularly when  $m_1$  is large. Thus, we can further develop a more conservative yet computationally tractable error bound, based on Problem (2.20), by fixing  $\hat{\mathbf{Q}}$  at a feasible value. Specifically, the SDP constraints in (2.20) implies that  $\hat{\mathbf{Q}} \succeq \frac{(\mathbf{q}_{m-m_1}^k - \hat{\mathbf{q}})(\mathbf{q}_{m-m_1}^k - \hat{\mathbf{q}})^\top}{4s^k}$  for any  $k \in [K]$  by Schur's complement. Thus, we fix  $\hat{\mathbf{Q}}$  at the following feasible value by letting

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

It follows that we can reduce Problem (2.20) to the following problem where we only

optimize the values of  $s^k, \forall k \in [K]$ , and  $\hat{\mathbf{q}}$ :

$$Z_{\text{gap}}^2(m_1) := \min_{s^k, \forall k \in [K], \hat{\mathbf{q}}} \sum_{k=1}^K s^k + \gamma_2 \sum_{k=1}^K \frac{(\mathbf{q}_{m-m_1}^k - \hat{\mathbf{q}})^\top (\mathbf{q}_{m-m_1}^k - \hat{\mathbf{q}})}{4s^k} + \sqrt{\gamma_1} \|\hat{\mathbf{q}}\|_2 \quad (2.21)$$

s.t.  $s^k \in \mathbb{R}_+, \forall k \in [K], \hat{\mathbf{q}} \in \mathbb{R}^{m-m_1}.$

Problem (2.21) can be reformulated as a second-order conic program, which is significantly more tractable than (2.20). Meanwhile, based on the above descriptions on how to construct the required feasible solution for deriving the corresponding error bound, we have

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

### 2.2.3 Upper Bounds

We further develop computationally efficient inner approximations for Problem (2.2), leading to upper bounds of its optimal value. Specifically, we derive two inner approximations in Sections 2.2.3.1 and 2.2.3.2.

#### 2.2.3.1 PCA based Upper Bound

Similar to Section 2.2.1, we utilize PCA to consider only the first  $m_1$  entries of  $\boldsymbol{\xi}_{\text{I}}$  in the second-moment constraint in  $\mathcal{D}_{\text{M}2}$ . This is a relaxation of the second-moment constraint, leading to a larger ambiguity set and so an inner approximation of Problem (2.2):

$$\min_{\mathbf{x} \in \mathcal{X}} \max_{\mathbb{P}_{\text{I}} \in \mathcal{D}_{\text{M}4}} \mathbb{E}_{\mathbb{P}_{\text{I}}} \left[ f \left( \mathbf{x}, \mathbf{U} \boldsymbol{\Lambda}^{\frac{1}{2}} \boldsymbol{\xi}_{\text{I}} + \boldsymbol{\mu} \right) \right], \quad (2.22)$$

where

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

**Theorem 2.3.** *If  $f(x, \mathbf{U}\Lambda^{\frac{1}{2}}\boldsymbol{\xi}_I + \boldsymbol{\mu})$  is  $\mathbb{P}_I$ -integrable for any  $\mathbb{P}_I \in \mathcal{D}_{M4}$ , then Problem (2.22) has the same optimal value as the following problem:*

$$\min_{\mathbf{x}, s, \mathbf{q}, \mathbf{Q}_r} \quad s + \gamma_2 \mathbf{I}_{m_1} \bullet \mathbf{Q}_r + \sqrt{\gamma_1} \|\mathbf{q}\|_2 \quad (2.23a)$$

$$\text{s.t.} \quad s \geq f\left(\mathbf{x}, \mathbf{U}\Lambda^{\frac{1}{2}}\boldsymbol{\xi}_I + \boldsymbol{\mu}\right) - \boldsymbol{\xi}_r^\top \mathbf{Q}_r \boldsymbol{\xi}_r - \mathbf{q}^\top \boldsymbol{\xi}_I, \quad \forall \boldsymbol{\xi}_I \in \mathcal{S}_I, \quad (2.23b)$$

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

where  $\mathbf{q} \in \mathbb{R}^m$  and  $\mathbf{Q}_r \in \mathbb{R}^{m_1 \times m_1}$ . We also have the following: (i) Problem (2.23) provides an upper bound for the optimal value of Problem (2.3); (ii) the optimal value of Problem (2.23) is non-increasing in  $m_1$ ; and (iii) if  $m_1 = m$ , then Problems (2.3) and (2.23) have the same optimal value.

*Proof.* The proof is similar with that of Theorem 2.2 and thus is omitted here.  $\square$

**Proposition 2.4.** *Under Assumption 2.1, Problem (2.23) has the same optimal value as the following SDP formulation:*

$$\bar{Z}_M^*(m_1) := \min_{\substack{\mathbf{x}, s, \hat{\boldsymbol{\lambda}}, \\ \mathbf{q}, \mathbf{Q}_r}} \quad s + \gamma_2 \mathbf{I}_{m_1} \bullet \mathbf{Q}_r + \sqrt{\gamma_1} \|\mathbf{q}\|_2 \quad (2.24a)$$

$$\text{s.t.} \quad \left[ \begin{array}{cc} s - y_k^0(\mathbf{x}) - \boldsymbol{\lambda}_k^\top \mathbf{b} - y_k(\mathbf{x})^\top \boldsymbol{\mu} + \boldsymbol{\lambda}_k^\top \mathbf{A} \boldsymbol{\mu} & \frac{1}{2} \left( \mathbf{q}_1 + \left( \mathbf{U}_{m \times m_1} \Lambda_{m_1}^{\frac{1}{2}} \right)^\top (\mathbf{A}^\top \boldsymbol{\lambda}_k - y_k(\mathbf{x})) \right)^\top \\ \frac{1}{2} \left( \mathbf{q}_1 + \left( \mathbf{U}_{m \times m_1} \Lambda_{m_1}^{\frac{1}{2}} \right)^\top (\mathbf{A}^\top \boldsymbol{\lambda}_k - y_k(\mathbf{x})) \right) & \mathbf{Q}_r \end{array} \right] \succeq 0, \quad \forall k \in [K], \quad (2.24b)$$

$$\mathbf{q}_2 + \left( \mathbf{U}_{m \times (m-m_1)} \Lambda_{m-m_1}^{\frac{1}{2}} \right)^\top (\mathbf{A}^\top \boldsymbol{\lambda}_k - y_k(\mathbf{x})) = \mathbf{0}, \quad \forall k \in [K], \quad (2.24c)$$

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

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

*Proof.* We apply the strong duality theorem to constraints (2.23b). As function  $f(\mathbf{x}, \boldsymbol{\xi})$  is piecewise linear convex,  $\boldsymbol{\xi} = \mathbf{U}\Lambda^{\frac{1}{2}}\boldsymbol{\xi}_I + \boldsymbol{\mu}$ , and  $\boldsymbol{\xi}_I = (\boldsymbol{\xi}_r^\top \in \mathbb{R}^{m_1}, \boldsymbol{\xi}_{r2}^\top \in \mathbb{R}^{m-m_1})^\top$ , we reformulate (2.23b) as

$$s \geq y_k^0(\mathbf{x}) + y_k(\mathbf{x})^\top \left( \mathbf{U}\Lambda^{\frac{1}{2}} (\boldsymbol{\xi}_r^\top, \boldsymbol{\xi}_{r2}^\top)^\top + \boldsymbol{\mu} \right) - \boldsymbol{\xi}_r^\top \mathbf{Q}_r \boldsymbol{\xi}_r - \mathbf{q}^\top (\boldsymbol{\xi}_r^\top, \boldsymbol{\xi}_{r2}^\top)^\top, \quad \forall \boldsymbol{\xi}_I \in \mathcal{S}_I, \quad \forall k \in [K], \quad (2.25)$$

which are equivalent to  $\min_{\mathbf{A}(\mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_I + \boldsymbol{\mu}) \leq \mathbf{b}, \boldsymbol{\xi}_I \in \mathbb{R}^m} g_k(\boldsymbol{\xi}_I) \geq 0$ , where  $g_k(\boldsymbol{\xi}_I) = s + \mathbf{q}^\top (\boldsymbol{\xi}_r^\top, \boldsymbol{\xi}_{r2}^\top)^\top + \boldsymbol{\xi}_r^\top \mathbf{Q}_r \boldsymbol{\xi}_r - y_k^0(\mathbf{x}) - y_k(\mathbf{x})^\top (\mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}} (\boldsymbol{\xi}_r^\top, \boldsymbol{\xi}_{r2}^\top)^\top + \boldsymbol{\mu})$ , for any  $k \in [K]$ . Moreover, we consider the Lagrangian dual problem of  $\min_{\mathbf{A}(\mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_I + \boldsymbol{\mu}) \leq \mathbf{b}, \boldsymbol{\xi}_I \in \mathbb{R}^m} g_k(\boldsymbol{\xi}_I)$ , i.e.,  $\max_{\boldsymbol{\lambda}_k \geq 0} \inf_{\boldsymbol{\xi}_I} g_k(\boldsymbol{\xi}_I) + \boldsymbol{\lambda}_k^\top (\mathbf{A}(\mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_I + \boldsymbol{\mu}) - \mathbf{b})$ , where  $\boldsymbol{\lambda}_k \in \mathbb{R}^m$ . Note that function  $g_k(\boldsymbol{\xi}_I)$  is convex in  $\boldsymbol{\xi}_I$  because it is a quadratic function that can be written as the general form  $f(\mathbf{x}) = \mathbf{x}^\top \mathbf{M}\mathbf{x} + \mathbf{x}^\top \mathbf{b} + c$  where  $\mathbf{M}$  is PSD, i.e.,  $\mathbf{Q} \succeq 0$ . Due to Assumption 2.1, there exists an interior point for the primal problem. It follows that constraints (2.25) are equivalent to the following ones:

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

which are further equivalent to the following constraints:

$$\begin{aligned} \exists \boldsymbol{\lambda}_k \geq 0, s + \mathbf{q}^\top (\boldsymbol{\xi}_r^\top, \boldsymbol{\xi}_{r2}^\top)^\top + \boldsymbol{\xi}_r^\top \mathbf{Q}_r \boldsymbol{\xi}_r - y_k^0(\mathbf{x}) - y_k(\mathbf{x})^\top (\mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}} (\boldsymbol{\xi}_r^\top, \boldsymbol{\xi}_{r2}^\top)^\top + \boldsymbol{\mu}) \\ + \boldsymbol{\lambda}_k^\top (\mathbf{A}(\mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}} (\boldsymbol{\xi}_r^\top, \boldsymbol{\xi}_{r2}^\top)^\top + \boldsymbol{\mu}) - \mathbf{b}) \geq 0, \quad \forall \boldsymbol{\xi}_I \in \mathbb{R}^m, \forall k \in [K]. \end{aligned} \quad (2.26)$$

Then, we perform the following decomposition:

$$\mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}} (\boldsymbol{\xi}_r^\top, \boldsymbol{\xi}_{r2}^\top)^\top + \boldsymbol{\mu} = \mathbf{U}_{m \times m_1} \mathbf{\Lambda}_{m_1}^{\frac{1}{2}} \boldsymbol{\xi}_r + \mathbf{U}_{m \times (m-m_1)} \mathbf{\Lambda}_{m-m_1}^{\frac{1}{2}} \boldsymbol{\xi}_{r2} + \boldsymbol{\mu}, \quad (2.27)$$

where  $\mathbf{U}_{m \times m_1} \in \mathbb{R}^{m \times m_1}$  and  $\mathbf{\Lambda}_{m_1}^{\frac{1}{2}} \in \mathbb{R}^{m_1 \times m_1}$  are upper-left submatrices of  $\mathbf{U}$  and  $\mathbf{\Lambda}^{\frac{1}{2}}$ , respectively, and  $\mathbf{\Lambda}_{m-m_1}^{\frac{1}{2}} \in \mathbb{R}^{(m-m_1) \times (m-m_1)}$  and  $\mathbf{U}_{m \times (m-m_1)} \in \mathbb{R}^{m \times (m-m_1)}$  are their lower-right submatrices, respectively. By plugging (2.27) to (2.26) and defining  $\mathbf{q} = (\mathbf{q}_1^\top \in \mathbb{R}^{m_1}, \mathbf{q}_2^\top \in \mathbb{R}^{m-m_1})^\top$ , we have

$$\begin{aligned} (2.26) \Leftrightarrow \exists \boldsymbol{\lambda}_k \geq 0, s - y_k^0(\mathbf{x}) - \boldsymbol{\lambda}_k^\top \mathbf{b} - y_k(\mathbf{x})^\top \boldsymbol{\mu} + \boldsymbol{\lambda}_k^\top \mathbf{A}\boldsymbol{\mu} + \left( \mathbf{q}_1 + \left( \mathbf{U}_{m \times m_1} \mathbf{\Lambda}_{m_1}^{\frac{1}{2}} \right)^\top (\mathbf{A}^\top \boldsymbol{\lambda}_k - y_k(\mathbf{x})) \right)^\top \boldsymbol{\xi}_r \\ + \left( \mathbf{q}_2 + \left( \mathbf{U}_{m \times (m-m_1)} \mathbf{\Lambda}_{m-m_1}^{\frac{1}{2}} \right)^\top (\mathbf{A}^\top \boldsymbol{\lambda}_k - y_k(\mathbf{x})) \right)^\top \boldsymbol{\xi}_{r2} + \boldsymbol{\xi}_r^\top \mathbf{Q}_r \boldsymbol{\xi}_r \geq 0, \quad \forall \boldsymbol{\xi}_I \in \mathbb{R}^m, \forall k \in [K], \quad (2.28) \\ \Leftrightarrow (1, \boldsymbol{\xi}_r^\top) \mathbf{Z}_k (1, \boldsymbol{\xi}_r^\top)^\top + \mathbf{W}_k^\top \boldsymbol{\xi}_{r2} \geq 0, \quad \forall \boldsymbol{\xi}_I \in \mathbb{R}^m, \forall k \in [K], \end{aligned}$$

where

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

and

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

Since  $\mathbf{W}_k^\top \boldsymbol{\xi}_{r2}$  in (2.28) is affine and  $\boldsymbol{\xi}_{r2} \in \mathbb{R}^{m-m_1}$ , we set  $\mathbf{W}_k = \mathbf{0}$  for any  $k \in [K]$ , which prevents the objective value of the Lagrangian dual problem from going to infinity and accordingly leads to constraints (2.24c). Thus, we have

$$(2.26) \Leftrightarrow \exists \boldsymbol{\lambda}_k \geq 0, (\mathbf{1}, \boldsymbol{\xi}_r^\top) \mathbf{Z}_k (\mathbf{1}, \boldsymbol{\xi}_r^\top)^\top \geq 0, \forall \boldsymbol{\xi}_r \in \mathbb{R}^{m_1}, \forall k \in [K]; (2.24c), (2.29)$$

$$\Leftrightarrow \exists \boldsymbol{\lambda}_k \geq 0, \mathbf{Z}_k \succeq 0, \forall k \in [K]; (2.24c). \quad (2.30)$$

The first equivalence holds due to the definition of  $\mathbf{Z}_k$ . For the second equivalence, clearly  $\Leftarrow$  follows from the definition of a PSD matrix. To prove  $\Rightarrow$ , we consider two possible cases for any  $(\eta_0 \in \mathbb{R}, \boldsymbol{\eta}^\top \in \mathbb{R}^{m_1})^\top \in \mathbb{R}^{m_1+1}$ : (1) if  $\eta_0 = 0$ , then  $(\eta_0, \boldsymbol{\eta}^\top) \mathbf{Z}_k (\eta_0, \boldsymbol{\eta}^\top)^\top = \boldsymbol{\eta}^\top \mathbf{Q}_r \boldsymbol{\eta} \geq 0$  because  $\mathbf{Q}_r$  is PSD; (2) if  $\eta_0 \neq 0$ , then we have  $(\eta_0, \boldsymbol{\eta}^\top) \mathbf{Z}_k (\eta_0, \boldsymbol{\eta}^\top)^\top = \eta_0^2 (1, \frac{\boldsymbol{\eta}^\top}{\eta_0}) \mathbf{Z}_k (1, \frac{\boldsymbol{\eta}^\top}{\eta_0})^\top \geq 0$  according to (2.29). Therefore,  $\Rightarrow$  holds and we obtain Problem (2.24) by replacing constraints (2.23b) with (2.30).  $\square$

One can observe that Problem (2.24) is significantly easier to solve than Problem (2.4) due to fewer decision variables and lower-dimensional PSD matrices in Problem (2.24).

### 2.2.3.2 Vector Splitting based Upper Bound

We derive the second inner approximation by splitting the random vector  $\boldsymbol{\xi}_I$  into  $P$  pieces, i.e.,  $\boldsymbol{\xi}_I = (\boldsymbol{\xi}_{I_1}^\top, \boldsymbol{\xi}_{I_2}^\top, \dots, \boldsymbol{\xi}_{I_P}^\top)^\top$ , where  $\boldsymbol{\xi}_{I_i} \in \mathbb{R}^{m_i}$ ,  $\forall i \in [P]$ , and  $\sum_{i=1}^P m_i = m$ . Accordingly, we revise the second-moment constraint in  $\mathcal{D}_{M2}$  with respect to these smaller pieces, leading to the following ambiguity set:

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

Set  $\mathcal{D}_{M5}$  is a superset of  $\mathcal{D}_{M2}$  because we ignore the correlations among  $\boldsymbol{\xi}_{I_p}$  and  $\boldsymbol{\xi}_{I_q}$  for any  $p, q \in [P]$  with  $p \neq q$ . This leads to the following inner approximation of Problem (2.2):

$$\min_{\mathbf{x} \in \mathcal{X}} \max_{\mathbb{P}_I \in \mathcal{D}_{M5}} \mathbb{E}_{\mathbb{P}_I} \left[ f \left( \mathbf{x}, \mathbf{U} \boldsymbol{\Lambda}^{\frac{1}{2}} \boldsymbol{\xi}_I + \boldsymbol{\mu} \right) \right]. \quad (2.31)$$

**Theorem 2.4.** *If  $f(\mathbf{x}, \mathbf{U} \boldsymbol{\Lambda}^{\frac{1}{2}} \boldsymbol{\xi}_I + \boldsymbol{\mu})$  is  $\mathbb{P}_I$ -integrable for any  $\mathbb{P}_I \in \mathcal{D}_{M5}$ , then Problem (2.31) has the same optimal value as the following problem:*

$$\min_{\mathbf{x}, s, \mathbf{q}, \hat{\mathbf{Q}}} s + \gamma_2 \sum_{i=1}^P \mathbf{I}_{m_i} \bullet \mathbf{Q}_i + \sqrt{\gamma_1} \|\mathbf{q}\|_2 \quad (2.32a)$$

$$\begin{aligned} \text{s.t.} \quad & s \geq f \left( \mathbf{x}, \mathbf{U} \boldsymbol{\Lambda}^{\frac{1}{2}} \boldsymbol{\xi}_I + \boldsymbol{\mu} \right) - \sum_{i=1}^P \boldsymbol{\xi}_{I_i}^\top \mathbf{Q}_i \boldsymbol{\xi}_{I_i} - \mathbf{q}^\top \boldsymbol{\xi}_I, \quad \forall \boldsymbol{\xi}_I \in \mathcal{S}_I, \quad (2.32b) \\ & \mathbf{x} \in \mathcal{X}, \quad \mathbf{Q}_i \succeq 0, \quad \forall i \in [P], \end{aligned}$$

where  $\mathbf{q} \in \mathbb{R}^m$ ,  $\mathbf{Q}_i \in \mathbb{R}^{m_i \times m_i}$  for any  $i \in [P]$ , and  $\hat{\mathbf{Q}} = \{\mathbf{Q}_1, \dots, \mathbf{Q}_P\}$ . Furthermore, Problem (2.32) provides an upper bound for the optimal value of Problem (2.3).

*Proof.* As  $\mathbb{P}_I$  is a probability measure on  $(\mathbb{R}^m, \mathbb{B})$ , where  $\mathbb{B}$  denotes the Borel  $\sigma$ -algebra on  $\mathbb{R}^m$ , Problem (2.31) can be described as the following problem:

$$\min_{\mathbf{x} \in \mathcal{X}} \max_{\mathbb{P}_I \in \mathcal{D}_{M5}} \int_{\mathcal{S}_I} f \left( \mathbf{x}, \mathbf{U} \boldsymbol{\Lambda}^{\frac{1}{2}} \boldsymbol{\xi}_I + \boldsymbol{\mu} \right) d\mathbb{P}_I(\boldsymbol{\xi}_I) \quad (2.33a)$$

$$\text{s.t.} \quad \int_{\mathcal{S}_I} d\mathbb{P}_I(\boldsymbol{\xi}_I) = 1, \quad (2.33b)$$

$$\int_{\mathcal{S}_I} \begin{bmatrix} \mathbf{I}_m & \boldsymbol{\xi}_I \\ \boldsymbol{\xi}_I^\top & \gamma_1 \end{bmatrix} d\mathbb{P}_I(\boldsymbol{\xi}_I) \succeq 0, \quad (2.33c)$$

$$\int_{\mathcal{S}_I} \boldsymbol{\xi}_{I_i} \boldsymbol{\xi}_{I_i}^\top d\mathbb{P}_I(\boldsymbol{\xi}_I) \preceq \gamma_2 \mathbf{I}_{m_i}, \quad \forall i \in [P], \quad (2.33d)$$

where (2.33c) is derived due to Schur's complement. In the following, we first formulate the dual of Problem (2.33) and then we show that strong duality holds.

Considering  $s, \begin{bmatrix} \mathbf{W} & \mathbf{w} \\ \mathbf{w}^\top & r \end{bmatrix} \succeq 0$ , and  $\mathbf{Q}_i \succeq 0$  for any  $i \in [P]$  as Lagrangian multipliers of constraints (2.33b), (2.33c), and (2.33d), respectively, we formulate the following

problem as the Lagrangian dual problem of (2.33):

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

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

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

Accordingly, the dual problem of (2.33) can be described as follows:

$$\begin{aligned} \min_{\substack{\mathbf{x}, s, \mathbf{W} \\ \mathbf{w}, r, \hat{\mathbf{Q}}}} \quad & s + \mathbf{I}_m \bullet \mathbf{W} + \gamma_1 r + \gamma_2 \sum_{i=1}^P \mathbf{I}_{m_i} \bullet \mathbf{Q}_i \end{aligned} \quad (2.34a)$$

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

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

$$\begin{bmatrix} \mathbf{W} & \mathbf{w} \\ \mathbf{w}^\top & r \end{bmatrix} \succeq 0, \quad (2.34b)$$

where  $\hat{\mathbf{Q}} = \{\mathbf{Q}_1, \dots, \mathbf{Q}_P\}$ . We further simplify Problem (2.34) towards eliminating variables  $\mathbf{W}$  and  $r$ . To that end, we keep variables  $\mathbf{Q}_i$ , for any  $i \in [P]$ , and  $s$  fixed while solving Problem (2.34) analytically for variables  $\mathbf{W}$ ,  $\mathbf{w}$ , and  $r$ . It follows that we solve  $\min_{\mathbf{x}, \mathbf{W}, \mathbf{w}, r} \mathbf{I}_m \bullet \mathbf{W} + \gamma_1 r$  analytically for  $\mathbf{W}$ ,  $\mathbf{w}$ , and  $r$ . We consider two cases for the optimal solution of  $r$  (denoted by  $r^*$ ) due to constraint (2.34b), i.e.,  $r^* > 0$  and  $r^* = 0$ , as follows.

- If  $r^* > 0$ , then constraint (2.34b) can be reformulated as  $\mathbf{W} \succeq \frac{\mathbf{w}\mathbf{w}^\top}{r}$  by Schur's complement. As a result,  $\mathbf{W}^* = \frac{\mathbf{w}\mathbf{w}^\top}{r}$  is a valid optimal solution because  $\min_{\mathbf{x}, \mathbf{W}, \mathbf{w}, r} \mathbf{I}_m \bullet \mathbf{W} + \gamma_1 r$  is a minimization problem. Replacing  $\mathbf{W}^*$  by  $\frac{\mathbf{w}\mathbf{w}^\top}{r}$  leads to solve a one-dimensional convex optimization problem, i.e.,  $\min_{r>0} \frac{\mathbf{w}^\top \mathbf{w}}{r} + \gamma_1 r$ .

By applying the necessary first-order optimality condition to this problem, i.e., setting the derivative of the objective function over  $r$  to zero, we have  $r^* = \frac{\|\mathbf{w}\|_2}{\sqrt{\gamma_1}}$  as the optimal solution of  $r$ . If we plug  $\mathbf{W}^* = \frac{\mathbf{w}\mathbf{w}^\top}{r}$  and  $r^* = \frac{\|\mathbf{w}\|_2}{\sqrt{\gamma_1}}$  in (2.34a), we obtain the following problem:

$$\begin{aligned} \min_{\mathbf{x}, s, \mathbf{w}, \hat{\mathbf{Q}}} \quad & s + \gamma_2 \sum_{i=1}^P \mathbf{I}_{m_i} \bullet \mathbf{Q}_i + \sqrt{\gamma_1} \|2\mathbf{w}\|_2 \\ \text{s.t.} \quad & s - 2\mathbf{w}^\top \boldsymbol{\xi}_I + \sum_{i=1}^P \boldsymbol{\xi}_i^\top \mathbf{Q}_i \boldsymbol{\xi}_i - f\left(\mathbf{x}, \mathbf{U}\Lambda^{\frac{1}{2}}\boldsymbol{\xi}_I + \boldsymbol{\mu}\right) \geq 0, \quad \forall \boldsymbol{\xi}_I \in \mathcal{S}_I, \\ & \mathbf{x} \in \mathcal{X}, \quad \mathbf{Q}_i \succeq 0, \quad \forall i \in [P]. \end{aligned} \quad (2.35)$$

By introducing a new variable  $\mathbf{q} = -2\mathbf{w}$ , we obtain Problem (2.32).

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

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

which contradicts constraint (2.34b). Considering  $r^* = 0$  and  $\mathbf{w}^* = \mathbf{0}$ ,

$\min_{\mathbf{x}, \mathbf{W}, \mathbf{w}, r} \mathbf{I}_m \bullet \mathbf{W} + \gamma_1 r$  reduces to  $\min_{\mathbf{x}, \mathbf{W}} \mathbf{I}_m \bullet \mathbf{W}$  whose optimal solution is clearly  $\mathbf{W}^* = \mathbf{0}$  as it is a minimization problem. Here also by replacing  $\mathbf{q} = -2\mathbf{w}$ , we obtain Problem (2.32).

Note that our conditions on  $\gamma_1$ ,  $\gamma_2$ , and  $\mathbf{I}_{m_i}$  for any  $i \in [P]$  are sufficient to ensure that the Dirac measure lies in the relative interior of the feasible set of Problem (2.31). Therefore, we can conclude that there is no duality gap between Problems (2.31) and (2.32) according to the weaker version of Proposition 3.4 in Shapiro (2001).

Finally, to prove Problem (2.32) provides an upper bound for Problem (2.3), we can equivalently prove that Problem (2.31) is an upper bound of Problem (2.2) since Problems (2.32) and (2.3) are equivalent reformulations of Problems (2.31)

and (2.2), respectively. To that end, we only need to prove  $\mathcal{D}_{M2} \subset \mathcal{D}_{M5}$ , i.e., any distribution in  $\mathcal{D}_{M2}$  also belongs to  $\mathcal{D}_{M5}$ . As the first two constraints of  $\mathcal{D}_{M2}$  and  $\mathcal{D}_{M5}$  are the same, any distribution in  $\mathcal{D}_{M2}$  satisfies constraints  $\mathbb{P}(\boldsymbol{\xi}_I \in \mathcal{S}_I) = 1$  and  $\mathbb{E}_{\mathbb{P}_1}[\boldsymbol{\xi}_I^\top] \mathbb{E}_{\mathbb{P}_1}[\boldsymbol{\xi}_I] \leq \gamma_1$  in  $\mathcal{D}_{M5}$ . Thus, we only require to show any distribution in  $\mathcal{D}_{M2}$  satisfies constraint  $\mathbb{E}_{\mathbb{P}_1}[\boldsymbol{\xi}_{I_i} \boldsymbol{\xi}_{I_i}^\top] \preceq \gamma_2 \mathbf{I}_{m_i}$ ,  $\forall i \in [P]$ , in  $\mathcal{D}_{M5}$ . To that end, we let  $\boldsymbol{\xi}_I = (\boldsymbol{\xi}_{I_1}^\top, \boldsymbol{\xi}_{I_2}^\top, \dots, \boldsymbol{\xi}_{I_P}^\top)^\top$ ,  $\boldsymbol{\xi}_{I_i} \in \mathbb{R}^{m_i}$  for any  $i \in [P]$ , and reformulate the second-order moment constraint of  $\mathcal{D}_{M2}$  as the following equivalent constraint:

$$\mathbb{E}_{\mathbb{P}_1} \begin{bmatrix} \boldsymbol{\xi}_{I_1} \boldsymbol{\xi}_{I_1}^\top & \boldsymbol{\xi}_{I_1} \boldsymbol{\xi}_{I_2}^\top & \cdots & \boldsymbol{\xi}_{I_1} \boldsymbol{\xi}_{I_P}^\top \\ \boldsymbol{\xi}_{I_2} \boldsymbol{\xi}_{I_1}^\top & \boldsymbol{\xi}_{I_2} \boldsymbol{\xi}_{I_2}^\top & \cdots & \boldsymbol{\xi}_{I_2} \boldsymbol{\xi}_{I_P}^\top \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\xi}_{I_P} \boldsymbol{\xi}_{I_1}^\top & \boldsymbol{\xi}_{I_P} \boldsymbol{\xi}_{I_2}^\top & \cdots & \boldsymbol{\xi}_{I_P} \boldsymbol{\xi}_{I_P}^\top \end{bmatrix} \preceq \begin{bmatrix} \gamma_2 \mathbf{I}_{m_1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \gamma_2 \mathbf{I}_{m_2} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \gamma_2 \mathbf{I}_{m_P} \end{bmatrix}, \quad (2.36)$$

which implies  $\mathbb{E}_{\mathbb{P}_1}[\boldsymbol{\xi}_{I_i} \boldsymbol{\xi}_{I_i}^\top] \preceq \gamma_2 \mathbf{I}_{m_i}$ ,  $\forall i \in [P]$  by simply considering the diagonal components of the matrices on both sides of (2.36). That is, any distribution in  $\mathcal{D}_{M2}$ , which satisfies  $\mathbb{E}_{\mathbb{P}_1}[\boldsymbol{\xi}_I \boldsymbol{\xi}_I^\top] \preceq \gamma_2 \mathbf{I}_m$ , also satisfies  $\mathbb{E}_{\mathbb{P}_1}[\boldsymbol{\xi}_{I_i} \boldsymbol{\xi}_{I_i}^\top] \preceq \gamma_2 \mathbf{I}_{m_i}$ ,  $\forall i \in [P]$  in  $\mathcal{D}_{M5}$ , i.e.,  $\mathcal{D}_{M2} \subset \mathcal{D}_{M5}$ .  $\square$

**Proposition 2.5.** *Under Assumption 2.1, Problem (2.32) has the same optimal value as the following SDP formulation:*

$$\begin{aligned} UB_M^* := & \min_{\substack{\mathbf{x}, s, \mathbf{q}, \\ \hat{\mathbf{Q}}, \hat{\boldsymbol{\lambda}}, \hat{\mathbf{s}}}} s + \gamma_2 \sum_{i=1}^P \mathbf{I}_{m_i} \bullet \mathbf{Q}_i + \sqrt{\gamma_1} \|\mathbf{q}\|_2 & (2.37) \\ \text{s.t.} & \begin{bmatrix} s_{ik} & \frac{1}{2} \left( \mathbf{q}_i + \left( \mathbf{U}_{m \times m_i} \boldsymbol{\Lambda}_{m_i}^{\frac{1}{2}} \right)^\top (\mathbf{A}^\top \boldsymbol{\lambda}_k - y_k(\mathbf{x})) \right)^\top \\ \frac{1}{2} \left( \mathbf{q}_i + \left( \mathbf{U}_{m \times m_i} \boldsymbol{\Lambda}_{m_i}^{\frac{1}{2}} \right)^\top (\mathbf{A}^\top \boldsymbol{\lambda}_k - y_k(\mathbf{x})) \right) & \mathbf{Q}_i \end{bmatrix} \succeq \mathbf{0}, \\ & \forall i \in [P], \forall k \in [K], \\ & \sum_{i=1}^P s_{ik} = s - y_k^0(\mathbf{x}) - \boldsymbol{\lambda}_k^\top \mathbf{b} - y_k(\mathbf{x})^\top \boldsymbol{\mu} + \boldsymbol{\lambda}_k^\top \mathbf{A} \boldsymbol{\mu}, \forall k \in [K], \\ & \boldsymbol{\lambda}_k \in \mathbb{R}_+^n, \forall k \in [K], \mathbf{x} \in \mathcal{X}, \end{aligned}$$

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

*Proof.* The proof is similar with that of Proposition 2.4 and thus is omitted here.  $\square$

One can observe that Problem (2.37) is significantly easier to solve than Problem (2.4) because it has smaller-sized PSD matrices and matrix variables compared to Problem (2.4). We also note that the size of each piece (i.e.,  $m_i$  with  $i \in [P]$ ) does not necessarily equal to each other. Thus, our theoretical results, including theoretical reformulations in this section and optimal gap bounds in the following Section 2.2.4, are general with respect to the size of each piece. Nevertheless, different ways of assigning a random vector to pieces lead to different computational performance. We will explain how we split the random vector in our numerical experiments in Section 2.4.

### 2.2.4 Upper Bound Quality

To measure the quality of our derived upper bounds in Section 2.2.3, we derive a theoretical bound for the gap between the optimal values of Problem (2.4) and Problem (2.24) (resp. Problem (2.37)). Before that, we present the following lemma that will facilitate the proofs in this section.

**Lemma 2.1.** *Consider the following PSD matrix with dimension  $(m+1) \times (m+1)$ :*

$$\mathbf{Z} = \begin{bmatrix} s & \mathbf{q}_1^\top & \cdots & \mathbf{q}_K^\top \\ \mathbf{q}_1 & \mathbf{Q}_1 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{q}_K & \mathbf{0} & \cdots & \mathbf{Q}_K \end{bmatrix} \succeq 0, \quad (2.38)$$

where  $s \in \mathbb{R}$ ,  $\mathbf{q}_k \in \mathbb{R}^{m_k}$  for any  $k \in [K]$  with  $\sum_{k=1}^K m_k = m$ ,  $\mathbf{Q}_k \in \mathbb{R}^{m_k \times m_k}$  for any  $k \in [K]$ , and other components are zero. Inequality (2.38) holds if and only if there exist  $\{s_k\}_{k=1}^K$  with  $\sum_{k=1}^K s_k = s$  such that

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

*Proof.* We prove the following equivalence:

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

Matrix  $\mathbf{Z}$  is PSD if and only if  $\boldsymbol{\eta}^\top \mathbf{Z} \boldsymbol{\eta} \geq 0$  for any  $\boldsymbol{\eta} \in \mathbb{R}^m$ , where  $\boldsymbol{\eta} = (\eta_0, \boldsymbol{\eta}_1^\top, \boldsymbol{\eta}_2^\top, \dots, \boldsymbol{\eta}_K^\top)^\top$  and  $\boldsymbol{\eta}_k \in \mathbb{R}^{m_k}$  for any  $k \in [K]$  with  $\sum_{k=1}^K m_k = m$ . Similar to the proofs of Propositions 2.1 and 2.4, respectively, we assume that  $\eta_0 = 1$  without loss of generality. Thus, we have

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

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

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

In summary, we have

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

□

**Proposition 2.6.** *Suppose that  $\mathbf{x}^*$  is an optimal solution of Problem (2.24) and  $\mathbf{Y}_r^{k*} = \begin{bmatrix} Y_{11}^{k*} & \mathbf{Y}_{12r}^{k* \top} \\ \mathbf{Y}_{12r}^{k*} & \mathbf{Y}_{22r}^{k*} \end{bmatrix}$  and  $\boldsymbol{\theta}_k^*$  are the corresponding dual optimal solutions associated with constraints (2.24b) and (2.24c), respectively. Then, it holds that*

$$0 \leq \bar{Z}_M^*(m_1) - Z_M^*(m) \leq \sum_{k=1}^K \left( \left( (\mathbf{U} \boldsymbol{\Lambda}^{\frac{1}{2}})^\top \mathbf{y}_k(\mathbf{x}^*) \right)^\top \left( \mathbf{Y}_{12r}^{k* \top}, \boldsymbol{\theta}_k^{* \top} \right)^\top \left( \frac{c^k - 1}{c^k} \right) \right), \quad (2.39)$$

where

$$c^k = \sqrt{1 + \frac{1}{Y_{11}^{k*}} \sum_{i=1}^{m-m_1} \frac{K}{\gamma_2} \theta_{ki}^{*2}}, \quad \forall k \in [K]. \quad (2.40)$$

*Proof.* We prove the results by investigating the duals of Problems (2.4) and (2.24) where  $\mathbf{x}$  is fixed at  $\mathbf{x}^*$ . Given  $\mathbf{x} = \mathbf{x}^*$ , we use  $\mathbf{Y}^k$  to denote the dual variable of constraints (2.4b) for any  $k \in [K]$ , where the dual of Problem (2.4) can be described as follows:

$$Z_M^*(m) = \max_{\mathbf{Y}} \sum_{k=1}^K (y_k^0(\mathbf{x}) + y_k(\mathbf{x})^\top \boldsymbol{\mu}) Y_{11}^k + \sum_{k=1}^K \left( \left( \mathbf{U} \boldsymbol{\Lambda}^{\frac{1}{2}} \right)^\top y_k(\mathbf{x}) \right)^\top \mathbf{Y}_{12}^k \quad (2.41a)$$

$$\text{s.t.} \quad 1 - \sum_{k=1}^K Y_{11}^k = 0, \quad (2.41b)$$

$$\sqrt{\gamma_1} - \left\| \sum_{k=1}^K \mathbf{Y}_{12}^k \right\|_2 \geq 0, \quad (2.41c)$$

$$\gamma_2 \mathbf{I}_m - \sum_{k=1}^K \mathbf{Y}_{22}^k \succeq 0, \quad (2.41d)$$

$$\mathbf{b} Y_{11}^k - \mathbf{A} \boldsymbol{\mu} Y_{11}^k - \mathbf{A} \mathbf{U} \boldsymbol{\Lambda}^{\frac{1}{2}} \mathbf{Y}_{12}^k \geq 0, \quad \forall k \in [K], \quad (2.41e)$$

$$\mathbf{Y}^k : \begin{bmatrix} Y_{11}^k & \mathbf{Y}_{12}^{k \top} \\ \mathbf{Y}_{12}^k & Y_{22}^k \end{bmatrix} \succeq 0, \quad \forall k \in [K]. \quad (2.41f)$$

Similarly, given  $\mathbf{x} = \mathbf{x}^*$ , we use  $\mathbf{Y}_r^k$  and  $\boldsymbol{\theta}_k$  to denote the dual variables of constraints (2.24b) and (2.24c), respectively, for any  $k \in [K]$ , where the dual of Problem (2.24) can be described as follows:

$$\begin{aligned} \bar{Z}_M^*(m_1) = \max_{\mathbf{Y}, \boldsymbol{\theta}} & \sum_{k=1}^K (y_k^0(\mathbf{x}) + y_k(\mathbf{x})^\top \boldsymbol{\mu}) Y_{11}^k + \sum_{k=1}^K \left( \left( \mathbf{U}_{m \times m_1} \boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}} \right)^\top y_k(\mathbf{x}) \right)^\top \mathbf{Y}_{12r}^k \\ & + \sum_{k=1}^K \left( \left( \mathbf{U}_{m \times (m-m_1)} \boldsymbol{\Lambda}_{m-m_1}^{\frac{1}{2}} \right)^\top y_k(\mathbf{x}) \right)^\top \boldsymbol{\theta}_k \end{aligned} \quad (2.42a)$$

$$\text{s.t.} \quad 1 - \sum_{k=1}^K Y_{11}^k = 0, \quad (2.42b)$$

$$\sqrt{\gamma_1} - \left\| \sum_{k=1}^K \left( \mathbf{Y}_{12r}^{k \top}, \boldsymbol{\theta}_k^\top \right)^\top \right\|_2 \geq 0, \quad (2.42c)$$

$$\gamma_2 \mathbf{I}_{m_1} - \sum_{k=1}^K \mathbf{Y}_{22r}^k \succeq 0, \quad (2.42d)$$

$$\mathbf{b} Y_{11}^k - \mathbf{A} \boldsymbol{\mu} Y_{11}^k - \mathbf{A} \mathbf{U}_{m \times m_1} \boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}} \mathbf{Y}_{12r}^k - \mathbf{A} \mathbf{U}_{m \times (m-m_1)} \boldsymbol{\Lambda}_{m-m_1}^{\frac{1}{2}} \boldsymbol{\theta}_k \geq 0, \quad (2.42e)$$

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

$$\mathbf{Y}_r^k : \begin{bmatrix} Y_{11}^k & \mathbf{Y}_{12r}^{k \top} \\ \mathbf{Y}_{12r}^k & \mathbf{Y}_{22r}^k \end{bmatrix} \succeq 0, \quad \boldsymbol{\theta}_k \text{ free}, \quad \forall k \in [K]. \quad (2.42f)$$

Given an optimal solution of Problem (2.42), i.e.,  $\mathbf{Y}_r^{k*}$  and  $\boldsymbol{\theta}_k^*$  for any  $k \in [K]$ , we construct a feasible solution of Problem (2.41), represented by  $\bar{\mathbf{Y}}^k = \begin{bmatrix} \bar{Y}_{11}^k & \bar{Y}_{12}^k \\ \bar{\mathbf{Y}}_{12}^k & \bar{\mathbf{Y}}_{22}^k \end{bmatrix}$  for any  $k \in [K]$ . For any given  $k \in [K]$ , we let  $\bar{Y}_{11}^k = Y_{11}^{k*}$ ,  $\bar{\mathbf{Y}}_{12}^k = \frac{1}{c^k} (\mathbf{Y}_{12r}^{k* \top}, \boldsymbol{\theta}_k^{*\top})^\top$  with  $c^k \geq 1$ , and

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

where the value of  $c^k$  as well as the value of  $w_i^k$  for any  $i \in [m - m_1]$  will be determined later so that  $\bar{\mathbf{Y}}^k$  satisfies all the constraints in Problem (2.41).

First, for the solution  $(\bar{\mathbf{Y}}^k \forall k \in [K])$ , it satisfies constraint (2.41b) because  $1 - \sum_{k=1}^K Y_{11}^{k*} = 0$  due to constraint (2.42b). This solution satisfies constraint (2.41c) because

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

where the first equality is due to the definition of  $\bar{\mathbf{Y}}_{12}^k$ , the first inequality is because  $c^k \geq 1$ , and the second inequality is because of constraint (2.42c). This solution also satisfies constraints (2.41e) because for any  $k \in [K]$ , we have  $\mathbf{b}\bar{Y}_{11}^k - \mathbf{A}\boldsymbol{\mu}\bar{\mathbf{Y}}_{11}^k - \mathbf{A}\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\bar{\mathbf{Y}}_{12}^k = \mathbf{b}Y_{11}^{k*} - \mathbf{A}\boldsymbol{\mu}Y_{11}^{k*} - \mathbf{A}\mathbf{U}_{m \times m_1}\boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}}\mathbf{Y}_{12r}^{k*} - \mathbf{A}\mathbf{U}_{m \times (m-m_1)}\boldsymbol{\Lambda}_{m-m_1}^{\frac{1}{2}}\boldsymbol{\theta}_k^* \geq 0$ , where the first equality is due to the definition of  $\bar{\mathbf{Y}}_{12}^k$  and the first inequality is because of constraints (2.42e).

Next, in order for  $(\bar{\mathbf{Y}}^k \forall k \in [K])$  to satisfy constraint (2.41d), we require  $\gamma_2 \mathbf{I}_m - \sum_{k=1}^K \bar{\mathbf{Y}}_{22}^k \succeq 0$ , which is equivalent to

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

Since  $\sum_{k=1}^K \mathbf{Y}_{22r}^{k*} \preceq \gamma_2 \mathbf{I}_{m_1}$  due to constraint (2.42d), we require

$$\sum_{k=1}^K w_i^k \leq \gamma_2, \quad \forall i \in [m - m_1]. \quad (2.43)$$

In addition, in order for  $(\bar{\mathbf{Y}}^k \forall k \in [K])$  to satisfy constraints (2.41f), matrix  $\bar{\mathbf{Y}}^k$  must be PSD for any  $k \in [K]$ . Note that if  $c^k \bar{\mathbf{Y}}^k$  is PSD, then  $\bar{\mathbf{Y}}^k$  is PSD because  $c^k \geq 1$ . Therefore, we consider the following decomposition on  $c^k \bar{\mathbf{Y}}^k$ :

$$c^k \bar{\mathbf{Y}}^k = \begin{bmatrix} \frac{1}{c^k} \mathbf{Y}_{11}^{k*} & \mathbf{Y}_{12r}^{k* \top} & \mathbf{0}_{1 \times (m-m_1)} \\ \mathbf{Y}_{12r}^{k*} & c^k \mathbf{Y}_{22r}^{k*} & \\ \mathbf{0}_{(m-m_1) \times 1} & & \mathbf{0}_{(m-m_1) \times (m-m_1)} \end{bmatrix} + \begin{bmatrix} (c^k - \frac{1}{c^k}) \mathbf{Y}_{11}^{k*} & \mathbf{0}_{1 \times m_1} & \boldsymbol{\theta}_k^{* \top} \\ \mathbf{0}_{m_1 \times 1} & \mathbf{0}_{m_1 \times m_1} & \mathbf{0} \\ \boldsymbol{\theta}_k^* & \mathbf{0} & \begin{matrix} c^k w_1^k & & \\ & \ddots & \\ & & c^k w_{m-m_1}^k \end{matrix} \end{bmatrix}, \forall k \in [K], \quad (2.44)$$

where the first matrix on the RHS of (2.44) is clearly PSD because  $\mathbf{Y}_r^{k*}$  is PSD due to constraints (2.42f) and we require the second one to be PSD as well so that  $\bar{\mathbf{Y}}^k$  can be PSD. By Lemma 2.1, we equivalently require

$$\begin{bmatrix} y_i^k & \theta_{ki}^* \\ \theta_{ki}^* & c^k w_i^k \end{bmatrix} \succeq 0, \forall i \in [m-m_1], \forall k \in [K], \text{ with } \sum_{i=1}^{m-m_1} y_i^k = \left(c^k - \frac{1}{c^k}\right) Y_{11}^{k*}, \forall k \in [K]. \quad (2.45)$$

Now, for a given  $i \in [m-m_1]$ , we let  $w_i^k = w_i$  for any  $k \in [K]$ , and then we have  $w_i \leq \frac{\gamma_2}{K}$  due to (2.43). It follows that, from (2.45), we equivalently require  $c^k w_i y_i^k \geq (\theta_{ki}^*)^2$  for any  $k \in [K]$  and  $i \in [m-m_1]$ . Therefore, for any given  $k \in [K]$ , we have

$$c^k \geq \frac{(\theta_{ki}^*)^2}{w_i y_i^k} \geq \frac{K (\theta_{ki}^*)^2}{\gamma_2 y_i^k}, \quad (2.46)$$

where the first inequality is due to  $c^k w_i y_i^k \geq (\theta_{ki}^*)^2$  and the second inequality is because of  $w_i \leq \frac{\gamma_2}{K}$ . Since (2.46) is equivalent to  $y_i^k \geq \frac{K (\theta_{ki}^*)^2}{c^k \gamma_2}$  and we have  $\sum_{i=1}^{m-m_1} y_i^k = (c^k - \frac{1}{c^k}) Y_{11}^{k*}$  from (2.45), we can conclude that

$$\begin{aligned} \sum_{i=1}^{m-m_1} y_i^k &= \left(c^k - \frac{1}{c^k}\right) Y_{11}^{k*} \geq \sum_{i=1}^{m-m_1} \frac{K (\theta_{ki}^*)^2}{\gamma_2 c^k}, \forall k \in [K] \\ \Rightarrow c^k &\geq \sqrt{1 + \frac{1}{Y_{11}^{k*}} \sum_{i=1}^{m-m_1} \frac{K (\theta_{ki}^*)^2}{\gamma_2}}, \forall k \in [K]. \end{aligned} \quad (2.47)$$

Therefore, we choose the value of  $c^k$  such that (2.47) is satisfied at equality, leading to (2.40), while the constructed solution  $(\bar{\mathbf{Y}}^k \forall k \in [K])$  satisfies all the constraints in Problem (2.41).

Finally, by Theorem 2.3, we have  $\bar{Z}_M^*(m_1) - Z_M^*(m) \geq 0$ . Meanwhile, we have

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

where the inequality holds because the constructed solution  $(\bar{\mathbf{Y}}^k \forall k \in [K])$  is feasible for Problem (2.41), which is a maximization problem. Therefore, it follows that

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

□

**Remark 2.2.** *Optimality gap (2.39) does not explicitly depend on  $\gamma_1$  while  $\gamma_2$  is explicitly there because upper bound (2.24) captures the complete first-order moment information while dropping some information about the second-order moment.*

**Proposition 2.7.** *Suppose that Assumption 2.1 holds,  $\boldsymbol{\mu} \in \mathcal{S}$ , and  $\min_{k=1}^K \{y_k^0(\mathbf{x}^*) + y_k(\mathbf{x}^*)^\top \boldsymbol{\mu}\} \geq 0$ , where  $\mathbf{x}^*$  is an optimal solution of Problem (2.4), then the relative gap between the optimal values of Problems (2.37) and (2.4) is bounded from above by  $\sqrt{P} - 1$ , i.e.,  $0 \leq UB_M^* - Z_M^*(m) \leq (\sqrt{P} - 1)Z_M^*(m)$ .*

*Proof.* We reformulate Problem (2.37) as the following problem:

$$UB_M^* = \min_{\substack{\mathbf{x}, s, \mathbf{q}, \\ \hat{\mathbf{Q}}, \hat{\boldsymbol{\lambda}}}} s + \gamma_2 \sum_{i=1}^P \mathbf{I}_{m_i} \bullet \mathbf{Q}_i + \sqrt{\gamma_1} \|\mathbf{q}\|_2 \quad (2.48a)$$

$$\text{s.t.} \quad \left[ \begin{array}{cc} s - y_k^0(\mathbf{x}) - \boldsymbol{\lambda}_k^\top b - y_k(\mathbf{x})^\top \boldsymbol{\mu} + \boldsymbol{\lambda}_k^\top \mathbf{A} \boldsymbol{\mu} & \frac{1}{2} \left( \mathbf{q} + \left( \mathbf{U} \boldsymbol{\Lambda}^{\frac{1}{2}} \right)^\top (\mathbf{A}^\top \boldsymbol{\lambda}_k - y_k(\mathbf{x})) \right)^\top \\ \frac{1}{2} \left( \mathbf{q} + \left( \mathbf{U} \boldsymbol{\Lambda}^{\frac{1}{2}} \right)^\top (\mathbf{A}^\top \boldsymbol{\lambda}_k - y_k(\mathbf{x})) \right) & \mathbf{Q}' \end{array} \right] \succeq 0, \quad (2.48b)$$

$\forall k \in [K],$

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

where

$$\mathbf{Q}' = \begin{bmatrix} \mathbf{Q}_1 & \mathbf{0}_{m_1 \times m_2} & \cdots & \mathbf{0}_{m_1 \times m_{P-1}} & \mathbf{0}_{m_1 \times m_P} \\ \mathbf{0}_{m_2 \times m_1} & \mathbf{Q}_2 & \cdots & \mathbf{0}_{m_2 \times m_{P-1}} & \mathbf{0}_{m_2 \times m_P} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0}_{m_{P-1} \times m_1} & \mathbf{0}_{m_{P-1} \times m_2} & \cdots & \mathbf{Q}_{P-1} & \mathbf{0}_{m_{P-1} \times m_P} \\ \mathbf{0}_{m_P \times m_1} & \mathbf{0}_{m_P \times m_2} & \cdots & \mathbf{0}_{m_P \times m_{P-1}} & \mathbf{Q}_P \end{bmatrix}, \quad (2.49)$$

$\mathbf{Q}_i \in \mathbb{R}^{m_i \times m_i}$  for any  $i \in [P]$  with  $\sum_{i=1}^P m_i = m$ ,  $\hat{\mathbf{Q}} = \{\mathbf{Q}_1, \dots, \mathbf{Q}_P\}$ , and  $\hat{\boldsymbol{\lambda}} = \{\boldsymbol{\lambda}_1, \dots, \boldsymbol{\lambda}_K\}$ . Let  $(\mathbf{x}^*, s^*, \boldsymbol{\lambda}_k^* \forall k \in [K], \mathbf{q}^*, \mathbf{Q}^*)$  denote an optimal solution of Problem (2.4) with

$$\mathbf{Q}^* = \begin{bmatrix} \mathbf{Q}_1^* & \mathbf{Q}_{m_1 \times m_2}^* & \cdots & \mathbf{Q}_{m_1 \times m_{P-1}}^* & \mathbf{Q}_{m_1 \times m_P}^* \\ \mathbf{Q}_{m_2 \times m_1}^* & \mathbf{Q}_2^* & \cdots & \mathbf{Q}_{m_2 \times m_{P-1}}^* & \mathbf{Q}_{m_2 \times m_P}^* \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{Q}_{m_{P-1} \times m_1}^* & \mathbf{Q}_{m_{P-1} \times m_2}^* & \cdots & \mathbf{Q}_{P-1}^* & \mathbf{Q}_{m_{P-1} \times m_P}^* \\ \mathbf{Q}_{m_P \times m_1}^* & \mathbf{Q}_{m_P \times m_2}^* & \cdots & \mathbf{Q}_{m_P \times m_{P-1}}^* & \mathbf{Q}_P^* \end{bmatrix}. \quad (2.50)$$

Based on this optimal solution, in the following, we construct a feasible solution of Problem (2.48), denoted by  $(\bar{\mathbf{x}}, \bar{s}, \bar{\mathbf{q}}, \bar{\mathbf{Q}}', \bar{\boldsymbol{\lambda}}_k \forall k \in [K])$ .

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

$$\bar{\mathbf{Q}}' = \begin{bmatrix} k_1 \mathbf{Q}_1^* & \mathbf{0}_{m_1 \times m_2} & \cdots & \mathbf{0}_{m_1 \times m_{P-1}} & \mathbf{0}_{m_1 \times m_P} \\ \mathbf{0}_{m_2 \times m_1} & k_2 \mathbf{Q}_2^* & \cdots & \mathbf{0}_{m_2 \times m_{P-1}} & \mathbf{0}_{m_2 \times m_P} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0}_{m_{P-1} \times m_1} & \mathbf{0}_{m_{P-1} \times m_2} & \cdots & k_{P-1} \mathbf{Q}_{P-1}^* & \mathbf{0}_{m_{P-1} \times m_P} \\ \mathbf{0}_{m_P \times m_1} & \mathbf{0}_{m_P \times m_2} & \cdots & \mathbf{0}_{m_P \times m_{P-1}} & k_P \mathbf{Q}_P^* \end{bmatrix}, \quad (2.51)$$

with  $k_i \geq 1$  for any  $i \in \{0, 1, 2, \dots, P\}$ . In order for this solution to satisfy (2.48b), we require

$$\begin{bmatrix} k_0 s^* - y_k^0(\mathbf{x}^*) - \boldsymbol{\lambda}_k^{*\top} \mathbf{b} - y_k(\mathbf{x}^*)^\top \boldsymbol{\mu} + \boldsymbol{\lambda}_k^{*\top} \mathbf{A} \boldsymbol{\mu} & \frac{1}{2} \left( \mathbf{q}^* + (\mathbf{U} \boldsymbol{\Lambda}^{\frac{1}{2}})^\top (\mathbf{A}^\top \boldsymbol{\lambda}_k^* - y_k(\mathbf{x}^*)) \right)^\top \\ \frac{1}{2} \left( \mathbf{q}^* + (\mathbf{U} \boldsymbol{\Lambda}^{\frac{1}{2}})^\top (\mathbf{A}^\top \boldsymbol{\lambda}_k^* - y_k(\mathbf{x}^*)) \right) & \bar{\mathbf{Q}}' \end{bmatrix} \succeq 0, \quad \forall k \in [K]. \quad (2.52)$$

In the following, we find the values of  $k_i$  for any  $i \in \{0, 1, 2, \dots, P\}$  so that (2.52) holds. To that end, we construct the following matrix

$$\begin{bmatrix} k_0 (s^* - y_k^0(\mathbf{x}^*) - \boldsymbol{\lambda}_k^{*\top} \mathbf{b} - y_k(\mathbf{x}^*)^\top \boldsymbol{\mu} + \boldsymbol{\lambda}_k^{*\top} \mathbf{A} \boldsymbol{\mu}) & \frac{1}{2} \left( \mathbf{q}^* + (\mathbf{U} \boldsymbol{\Lambda}^{\frac{1}{2}})^\top (\mathbf{A}^\top \boldsymbol{\lambda}_k^* - y_k(\mathbf{x}^*)) \right)^\top \\ \frac{1}{2} \left( \mathbf{q}^* + (\mathbf{U} \boldsymbol{\Lambda}^{\frac{1}{2}})^\top (\mathbf{A}^\top \boldsymbol{\lambda}_k^* - y_k(\mathbf{x}^*)) \right) & \bar{\mathbf{Q}}' \end{bmatrix}, \quad \forall k \in [K]. \quad (2.53)$$

Note that subtracting (2.53) from (2.52) leads to the following matrix:

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

which is PSD because its eigenvalues are non-negative. In fact,  $(k_0 - 1)(y_k^0(\mathbf{x}^*) + y_k(\mathbf{x}^*)^\top \boldsymbol{\mu} - \boldsymbol{\lambda}_k^{*\top}(\mathbf{A}\boldsymbol{\mu} - \mathbf{b}))$  is the only non-zero eigenvalue of this matrix that is non-negative because  $k_0 \geq 1$ ,  $-\boldsymbol{\lambda}_k^{*\top}(\mathbf{A}\boldsymbol{\mu} - \mathbf{b}) \geq 0$  due to  $\mathbf{A}\boldsymbol{\mu} \leq \mathbf{b}$  and  $\boldsymbol{\lambda}_k^* \geq 0$ , and we have  $y_k^0(\mathbf{x}^*) + y_k(\mathbf{x}^*)^\top \boldsymbol{\mu} \geq 0$  according to the assumption  $\min_{k=1}^K \{y_k^0(\mathbf{x}^*) + y_k(\mathbf{x}^*)^\top \boldsymbol{\mu}\} \geq 0$ . Thus, we choose good values of  $k_i$  for any  $i \in \{0, 1, 2, \dots, P\}$  to ensure (2.53) to be a PSD matrix and accordingly will make (2.52) hold.

Next, by Lemma 2.1, in order for (2.53) to be a PSD, we equivalently require

$$\left[ \begin{array}{c} s_i(s^* - y_k^0(\mathbf{x}^*) - \boldsymbol{\lambda}_k^{*\top} \mathbf{b} - y_k(\mathbf{x}^*)^\top \boldsymbol{\mu} + \boldsymbol{\lambda}_k^{*\top} \mathbf{A}\boldsymbol{\mu}) \quad \frac{1}{2} \left( \mathbf{q}_i^* + \left( \mathbf{U}_{m \times m_i} \boldsymbol{\Lambda}_{m_i}^{\frac{1}{2}} \right)^\top (\mathbf{A}^\top \boldsymbol{\lambda}_k^* - y_k(\mathbf{x}^*)) \right)^\top \\ \frac{1}{2} \left( \mathbf{q}_i^* + \left( \mathbf{U}_{m \times m_i} \boldsymbol{\Lambda}_{m_i}^{\frac{1}{2}} \right)^\top (\mathbf{A}^\top \boldsymbol{\lambda}_k^* - y_k(\mathbf{x}^*)) \right) \quad k_i \mathbf{Q}_i^* \end{array} \right] \succeq 0, \quad (2.54)$$

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

with  $\sum_{i=1}^P s_i = k_0$ . Constraints (2.54) can be satisfied by allowing  $s_i \times k_i \geq 1$  for any  $i \in [P]$  due to (2.4b). Then, we let  $k_0 = k_1 = \dots = k_P$  and  $s_i \times k_i = 1$  for any  $i \in [P]$ , leading to  $k_0 = k_1 = \dots = k_P = \sqrt{P}$ .

Finally, we have  $UB_M^* \geq Z_M^*(m)$  by Theorem 2.4. Meanwhile, as Problem (2.37) is a minimization problem,  $UB_M^*$  is no larger than the objective value corresponding to our constructed feasible solution  $(\bar{\mathbf{x}}, \bar{s}, \bar{\mathbf{q}}, \bar{\mathbf{Q}}, \bar{\boldsymbol{\lambda}}_k \forall k \in [K])$ . That is, we have

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

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

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

□

We observe that the theoretical upper bound in Proposition 2.7 is achievable through the following example.

**Example 2.1.** Suppose that  $\boldsymbol{\mu} = \mathbf{0}$ ,  $\mathcal{S} = \mathbb{R}^m$ ,  $\gamma_1 = +\infty$ , and  $\gamma_2 = 1$ . With  $f(\mathbf{x}, \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_I + \boldsymbol{\mu}) = |\mathbf{x}^\top \boldsymbol{\xi}_I|$ , Problem (2.4) can be recast as the following SDP for-

mulation:

$$\begin{aligned} Z_M^*(m) = \min_{\mathbf{x}, s, \mathbf{Q}} \quad & s + \mathbf{I}_m \bullet \mathbf{Q} \\ \text{s.t.} \quad & \begin{bmatrix} s & \frac{\mathbf{x}^\top}{2} \\ \frac{\mathbf{x}}{2} & \mathbf{Q} \end{bmatrix} \succeq 0, \quad \begin{bmatrix} s & \frac{-\mathbf{x}^\top}{2} \\ \frac{-\mathbf{x}}{2} & \mathbf{Q} \end{bmatrix} \succeq 0, \quad \mathbf{x} \in \mathcal{X}. \end{aligned} \quad (2.55)$$

For fixed  $\mathbf{x} \in \mathcal{X}$ , optimizing over the remaining decision variables in (2.55) yields  $\mathbf{Q}^* = \frac{\mathbf{x}\mathbf{x}^\top}{4s}$ , and  $s^* = \frac{\sqrt{\mathbf{x}^\top \mathbf{x}}}{2}$  with objective value  $\sqrt{\mathbf{x}^\top \mathbf{x}}$ . By Proposition 2.5, we obtain an upper bound by considering

$$\begin{aligned} UB_M^* = \min_{\mathbf{x}, s, \hat{\mathbf{Q}}} \quad & s + \sum_{i=1}^P \mathbf{I}_i \bullet \mathbf{Q}_i \\ \text{s.t.} \quad & \begin{bmatrix} s & \frac{\mathbf{x}^\top}{2} \\ \frac{\mathbf{x}}{2} & \text{diag}(\mathbf{Q}_i) \end{bmatrix} \succeq 0, \quad \begin{bmatrix} s & \frac{-\mathbf{x}^\top}{2} \\ \frac{-\mathbf{x}}{2} & \text{diag}(\mathbf{Q}_i) \end{bmatrix} \succeq 0, \quad \mathbf{x} \in \mathcal{X}, \end{aligned} \quad (2.56)$$

where  $\hat{\mathbf{Q}} = \{\mathbf{Q}_1, \dots, \mathbf{Q}_P\}$  and  $\text{diag}(\mathbf{Q}_i)$  is a block diagonal matrix consisting of  $\mathbf{Q}_1, \dots, \mathbf{Q}_P$ . For fixed  $\mathbf{x} \in \mathcal{X}$ , where  $\mathbf{x} = (\mathbf{x}_1^\top, \dots, \mathbf{x}_P^\top)^\top$  and  $\mathbf{x}_i \in \mathbb{R}^{m_i}$  for all  $i \in [P]$ , optimizing over the remaining decision variables in Problem (2.56) yields  $\mathbf{Q}_i^* = \frac{\mathbf{x}_i \mathbf{x}_i^\top}{4s_i}$  for all  $i \in [P]$ , and  $s^* = \sum_{i=1}^P \frac{\sqrt{\mathbf{x}_i^\top \mathbf{x}_i}}{2}$  with objective value  $\sum_{i=1}^P \sqrt{\mathbf{x}_i^\top \mathbf{x}_i}$ . Now we let  $m_i = \frac{m}{P}$  for any  $i \in [P]$  and  $\mathcal{X} = \{\mathbf{x} \in \mathbb{R}^m | x_i \geq 1, \forall i \in [m]\}$ . It follows that  $Z_M^*(m) = \sqrt{m}$  and  $UB_M^* = P\sqrt{\frac{m}{P}}$ , and so the relative gap between  $Z_M^*(m)$  and  $UB_M^*$  is

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

attaining the theoretical upper bound in Proposition 2.7.

**Remark 2.3.** The theoretical error bound  $\sqrt{P} - 1$  in Proposition 2.7, albeit achievable in the above example, is usually conservative because it needs to hold valid for arbitrary (including many pathetic and worst-case) DRO problem instances. For instance, if  $P > 1$ ,  $\gamma_1 > 0$ , and  $\|q^*\|_2 > 0$ , then the theoretical bound cannot be attained. The above example is comparable to the worst-case distribution induced by solving a DRO problem. Nevertheless, the inner approximation (2.37) leads to much

*better computational optimality gaps under various instances, which will be shown in our numerical experiments in Section 2.4.*

### 2.3 Combined Ambiguity Set

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

Formally, we consider DRO problem

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

where

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

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

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

where  $\pi$  denotes a joint distribution of  $\boldsymbol{\xi}$  and  $\hat{\boldsymbol{\xi}}$  with marginals  $\mathbb{P}$  and  $\mathbb{P}_0$ , respectively. Intuitively,  $W(\mathbb{P}, \mathbb{P}_0)$  represents the minimum expected distance between  $\boldsymbol{\xi}$  and  $\hat{\boldsymbol{\xi}}$  over all possible joint distributions  $\pi$ . It has been shown that, as  $N \rightarrow \infty$ ,  $\mathbb{P}_0$  converges to the true distribution of  $\boldsymbol{\xi}$  almost surely (Van der Vaart 2000). As a result, if we select the value of  $R_0 > 0$  appropriately, then the Wasserstein ball centered at  $\mathbb{P}_0$  with radius  $R_0$  will include such true distribution with high

confidence. Besides the Wasserstein ball,  $\mathcal{D}_{C_1}$  designates that the centered second-order moment matrix of  $\boldsymbol{\xi}$  is bounded by  $\gamma_2 \boldsymbol{\Sigma}$ . We notice that  $R_0$  controls the conservatism degree of  $\mathcal{D}_{C_1}$ . The larger radius  $R_0$  is,  $\mathcal{D}_{C_1}$  has higher confidence to contain the true distribution of  $\boldsymbol{\xi}$ , while it leads to a more conservative optimal solution to (DRO-C). In contrast, when  $R_0$  decreases to zero, (DRO-C) reduces to an ambiguity-free stochastic program with regard to  $\mathbb{P}_0$ . For (DRO-C), we consider a setting slightly stronger than that in Assumption 2.1.

**Assumption 2.2.** *Function  $f(\mathbf{x}, \boldsymbol{\xi})$  is piecewise linear convex in  $\boldsymbol{\xi}$ , i.e.,  $f(\mathbf{x}, \boldsymbol{\xi}) = \max_{k=1}^K \{y_k^0(\mathbf{x}) + y_k(\mathbf{x})^\top \boldsymbol{\xi}\}$  with both  $y_k(\mathbf{x}) = (y_k^1(\mathbf{x}), \dots, y_k^m(\mathbf{x}))^\top$  and  $y_k^0(\mathbf{x})$  affine in  $\mathbf{x}$  for any  $k \in [K]$ . Additionally,  $\mathcal{S} = \mathbb{R}^m$ .*

**Proposition 2.8.** *Under Assumption 2.2, (DRO-C) can be recast as the following SDP formulation:*

$$Z_C^*(m) := \min_{\mathbf{x}, \lambda, \mathbf{Q}, \hat{\boldsymbol{\zeta}}, \hat{y}} \lambda R_0 + \gamma_2 \boldsymbol{\Sigma} \bullet \mathbf{Q} + \frac{1}{N} \sum_{i=1}^N y_i \quad (2.57a)$$

$$\text{s.t.} \quad \begin{bmatrix} \mathbf{Q} & \frac{1}{2}(-y_k(\mathbf{x}) + \boldsymbol{\zeta}^i - 2\mathbf{Q}\boldsymbol{\mu}) \\ \frac{1}{2}(-y_k(\mathbf{x}) + \boldsymbol{\zeta}^i - 2\mathbf{Q}\boldsymbol{\mu})^\top & y_i - y_k^0(\mathbf{x}) - \boldsymbol{\zeta}^{i\top} \hat{\boldsymbol{\xi}}^i + \boldsymbol{\mu}^\top \mathbf{Q}\boldsymbol{\mu} \end{bmatrix} \succeq 0, \quad (2.57b)$$

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

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

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

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

As discussed in Section 2.2, Problem (2.57) can be computationally difficult when  $\boldsymbol{\xi}$  is high-dimensional and/or correlated, leading to many large-scale PSD constraints. We derive more efficiently solvable outer and inner approximations of Problem (2.57) (i.e., (DRO-C)) in Sections 2.3.1 and 2.3.2, leading to lower and upper bounds, respectively, while theoretically showing their quality.

### 2.3.1 Lower Bound

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

$$\min_{\mathbf{x} \in \mathcal{X}} \max_{\mathbb{P}_I \in \mathcal{D}_{C2}} \mathbb{E}_{\mathbb{P}_I} \left[ f \left( \mathbf{x}, \mathbf{U} \Lambda^{\frac{1}{2}} \boldsymbol{\xi}_I + \boldsymbol{\mu} \right) \right], \quad (2.58)$$

where

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

under the condition that  $f(\mathbf{x}, \mathbf{U} \Lambda^{\frac{1}{2}} \boldsymbol{\xi}_I + \boldsymbol{\mu})$  is  $\mathbb{P}_I$ -integrable for any  $\mathbb{P}_I \in \mathcal{D}_{C2}$  and  $\mathcal{S}_I := \{\boldsymbol{\xi}_I \in \mathbb{R}^m : \mathbf{U} \Lambda^{\frac{1}{2}} \boldsymbol{\xi}_I + \boldsymbol{\mu} \in \mathcal{S}\}$ . Next, by the approximation of  $\boldsymbol{\xi}$  in (2.9) due to PCA, we outer approximate (2.58) as the following problem:

$$\min_{\mathbf{x} \in \mathcal{X}} \max_{\mathbb{P}_r \in \mathcal{D}_{C3}} \mathbb{E}_{\mathbb{P}_r} \left[ f \left( \mathbf{x}, \mathbf{U}_{m \times m_1} \Lambda_{m_1}^{\frac{1}{2}} \boldsymbol{\xi}_r + \boldsymbol{\mu} \right) \right], \quad (2.59a)$$

where

$$\mathcal{D}_{C3}(\mathcal{S}_r, \boldsymbol{\mu}, \gamma_2, \mathbb{P}_0, R_0) = \left\{ \mathbb{P}_r \left| \begin{array}{l} \mathbb{E}_{\mathbb{P}_r} [\boldsymbol{\xi}_r \boldsymbol{\xi}_r^\top] \preceq \gamma_2 \mathbf{I}_{m_1} \\ \exists \pi : \int_{\mathcal{S}_r^2} \left\| \mathbf{U}_{m \times m_1} \Lambda_{m_1}^{\frac{1}{2}} \boldsymbol{\xi}_r + \boldsymbol{\mu} - \hat{\boldsymbol{\xi}} \right\|_1 \pi \left( \mathbf{U}_{m \times m_1} \Lambda_{m_1}^{\frac{1}{2}} \boldsymbol{\xi}_r + \boldsymbol{\mu}, \hat{\boldsymbol{\xi}} \right) \leq R_0 \end{array} \right. \right\} \quad (2.59b)$$

with

$$\mathcal{S}_r := \left\{ \boldsymbol{\xi}_r \in \mathbb{R}^{m_1} : \mathbf{U}_{m \times m_1} \Lambda_{m_1}^{\frac{1}{2}} \boldsymbol{\xi}_r + \boldsymbol{\mu} \in \mathcal{S} \right\}. \quad (2.59c)$$

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

**Theorem 2.5.** *Under Assumption 2.2, Problem (2.59) has the same optimal value*

as the following SDP formulation:

$$Z_C^*(m_1) := \min_{\mathbf{x}, \lambda, \mathbf{Q}_r, \hat{\boldsymbol{\zeta}}, \hat{y}} \lambda R_0 + \gamma_2 \mathbf{I}_{m_1} \bullet \mathbf{Q}_r + \frac{1}{N} \sum_{i=1}^N y_i \quad (2.60a)$$

$$\text{s.t.} \quad \begin{bmatrix} \mathbf{Q}_r & \frac{1}{2} \left( (-y_k(\mathbf{x}) + \boldsymbol{\zeta}^i)^\top \mathbf{U}_{m \times m_1} \boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}} \right)^\top \\ \frac{1}{2} (-y_k(\mathbf{x}) + \boldsymbol{\zeta}^i)^\top \mathbf{U}_{m \times m_1} \boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}} & y_i - y_k(\mathbf{x})^\top \boldsymbol{\mu} - y_k^0(\mathbf{x}) + (\boldsymbol{\mu} - \hat{\boldsymbol{\zeta}}^i)^\top \boldsymbol{\zeta}^i \end{bmatrix} \succeq 0, \quad \forall i \in [N], \forall k \in [K], \quad (2.60b)$$

$$\lambda \in \mathbb{R}_+, \mathbf{x} \in \mathcal{X}, \|\boldsymbol{\zeta}^i\|_* \leq \lambda, \forall i \in [N], \quad (2.60c)$$

where  $\mathbf{Q}_r \in \mathbb{R}^{m_1 \times m_1}$ ,  $\boldsymbol{\zeta}^i \in \mathbb{R}^m$  for any  $i \in [N]$ ,  $\hat{\boldsymbol{\zeta}} = \{\boldsymbol{\zeta}^1, \dots, \boldsymbol{\zeta}^N\}$ , and  $\hat{y} = \{y_1, \dots, y_N\}$ . Furthermore, we have the following: (i) Problem (2.60) provides a lower bound for the optimal value of (DRO-C); (ii) the optimal value of Problem (2.60) is nondecreasing in  $m_1$ ; and (iii) if  $m_1 = m$ , then (DRO-C) and (2.60) have the same optimal value.

*Proof.* By Theorem 1 in Gao and Kleywegt (2017), Problem (2.59) has the following strong dual problem:

$$\min_{\mathbf{x} \in \mathcal{X}, \mathbf{Q}_r, \lambda} \left\{ \lambda R_0 + \gamma_2 \mathbf{I}_{m_1} \bullet \mathbf{Q}_r + \int_{\mathbb{R}^{m_1}} \sup_{\boldsymbol{\xi}_r} g(\boldsymbol{\xi}_r, \hat{\boldsymbol{\xi}}) \mathbb{P}_0(d\hat{\boldsymbol{\xi}}) \right\}, \quad (2.61)$$

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

$$\int_{\mathbb{R}^{m_1}} \sup_{\boldsymbol{\xi}_r} g(\boldsymbol{\xi}_r, \hat{\boldsymbol{\xi}}) \mathbb{P}_0(d\hat{\boldsymbol{\xi}}) = \frac{1}{N} \sum_{i=1}^N \sup_{\boldsymbol{\xi}_r} g(\boldsymbol{\xi}_r, \hat{\boldsymbol{\xi}}^i). \quad (2.62)$$

Thus, by plugging (2.62) into (2.61), (2.61) can be reformulated as

$$\min_{\mathbf{x} \in \mathcal{X}, \mathbf{Q}_r, \lambda} \left\{ \lambda R_0 + \gamma_2 \mathbf{I}_{m_1} \bullet \mathbf{Q}_r + \frac{1}{N} \sum_{i=1}^N y_i \right\} \quad (2.63a)$$

$$\text{s.t.} \quad y_i = \sup_{\boldsymbol{\xi}_r} g(\boldsymbol{\xi}_r, \hat{\boldsymbol{\xi}}^i), \quad \forall i \in [N]. \quad (2.63b)$$

Since Problem (2.63) is a minimization problem, constraints (2.63b) can be relaxed to  $y_i \geq \sup_{\xi_r} g(\xi_r, \hat{\xi}^i)$  for any  $i \in [N]$ . Thus, we have

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

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

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

and accordingly we have

$$\begin{aligned} y_i &\geq \sup_{\xi_r} \inf_{\|\hat{\zeta}\|_* \leq 1} \left\{ y_k(\mathbf{x})^\top \left( \mathbf{U}_{m \times m_1} \Lambda_{m_1}^{\frac{1}{2}} \xi_r + \boldsymbol{\mu} \right) + y_k^0(\mathbf{x}) - \xi_r^\top \mathbf{Q}_r \xi_r - \lambda \hat{\zeta}^\top \left( \mathbf{U}_{m \times m_1} \Lambda_{m_1}^{\frac{1}{2}} \xi_r + \boldsymbol{\mu} - \hat{\xi}^i \right) \right\}, \forall k \in [K] \\ y_i &\geq \inf_{\|\hat{\zeta}\|_* \leq 1} \sup_{\xi_r} \left\{ y_k(\mathbf{x})^\top \left( \mathbf{U}_{m \times m_1} \Lambda_{m_1}^{\frac{1}{2}} \xi_r + \boldsymbol{\mu} \right) + y_k^0(\mathbf{x}) - \xi_r^\top \mathbf{Q}_r \xi_r - \lambda \hat{\zeta}^\top \left( \mathbf{U}_{m \times m_1} \Lambda_{m_1}^{\frac{1}{2}} \xi_r + \boldsymbol{\mu} - \hat{\xi}^i \right) \right\}, \forall k \in [K] \\ \exists \hat{\zeta} \text{ s.t. } \|\hat{\zeta}\|_* &\leq 1, y_i \geq \sup_{\xi_r} \left\{ y_k(\mathbf{x})^\top \left( \mathbf{U}_{m \times m_1} \Lambda_{m_1}^{\frac{1}{2}} \xi_r + \boldsymbol{\mu} \right) + y_k^0(\mathbf{x}) - \xi_r^\top \mathbf{Q}_r \xi_r - \lambda \hat{\zeta}^\top \left( \mathbf{U}_{m \times m_1} \Lambda_{m_1}^{\frac{1}{2}} \xi_r + \boldsymbol{\mu} - \hat{\xi}^i \right) \right\}, \\ &\quad \forall k \in [K] \\ \Leftrightarrow \exists \hat{\zeta} \text{ s.t. } \|\hat{\zeta}\|_* &\leq 1, y_i \geq y_k(\mathbf{x})^\top \left( \mathbf{U}_{m \times m_1} \Lambda_{m_1}^{\frac{1}{2}} \xi_r + \boldsymbol{\mu} \right) + y_k^0(\mathbf{x}) - \xi_r^\top \mathbf{Q}_r \xi_r - \lambda \hat{\zeta}^\top \left( \mathbf{U}_{m \times m_1} \Lambda_{m_1}^{\frac{1}{2}} \xi_r + \boldsymbol{\mu} - \hat{\xi}^i \right), \\ &\quad \forall \xi_r \in \mathbb{R}^{m_1}, \forall k \in [K] \\ \Leftrightarrow \exists \hat{\zeta} \text{ s.t. } \|\hat{\zeta}\|_* &\leq 1, \begin{bmatrix} \mathbf{Q}_r & \frac{1}{2} \left( (-y_k(\mathbf{x})^\top + \lambda \hat{\zeta}^\top) \mathbf{U}_{m \times m_1} \Lambda_{m_1}^{\frac{1}{2}} \right)^\top \\ \frac{1}{2} (-y_k(\mathbf{x})^\top + \lambda \hat{\zeta}^\top) \mathbf{U}_{m \times m_1} \Lambda_{m_1}^{\frac{1}{2}} & y_i - y_k(\mathbf{x})^\top \boldsymbol{\mu} - y_k^0(\mathbf{x}) + \lambda \hat{\zeta}^\top (\boldsymbol{\mu} - \hat{\xi}^i) \end{bmatrix} \succeq 0, \forall k \in [K], \quad (2.64) \end{aligned}$$

where the first equivalence is due to the convexity of  $g(\xi_r, \hat{\xi}^i)$ ,  $\mathcal{S}_r$ , and the feasible region defined by  $\|\hat{\zeta}\|_* \leq 1$ . For any given  $i \in [N]$ , we replace  $\lambda \hat{\zeta}$  by  $\zeta^i$ , and then we can obtain Problem (2.60) by further replacing (2.63b) by (2.64) for any  $i \in [N]$ .

To prove Problem (2.60) provides a lower bound for Problem (2.57), we consider Problem (2.59) and define  $\zeta = \mathbf{U}_{m \times m_1} \Lambda_{m_1}^{\frac{1}{2}} \xi_r + \boldsymbol{\mu}$ , denoting its support and ambiguity set by  $\mathcal{S}_\zeta$  and  $\mathcal{D}_\zeta$ , respectively. As  $\mathcal{S}_r = \{\xi_r \in \mathbb{R}^{m_1} : \mathbf{U}_{m \times m_1} \Lambda_{m_1}^{\frac{1}{2}} \xi_r + \boldsymbol{\mu} \in \mathcal{S}\}$  and  $\mathcal{S}_\zeta = \{\zeta \in \mathbb{R}^m : \zeta = \mathbf{U}_{m \times m_1} \Lambda_{m_1}^{\frac{1}{2}} \xi_r + \boldsymbol{\mu}, \xi_r \in \mathcal{S}_r\}$ , we can deduce  $\mathcal{S}_\zeta \subset \mathcal{S}$ . We also have

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

Moreover, we have

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

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

$$\max_{\mathbb{P}_{\zeta} \in \mathcal{D}_{\zeta}} \mathbb{E}_{\mathbb{P}_{\zeta}} [f(\mathbf{x}, \zeta)] \leq \max_{\mathbb{P} \in \mathcal{D}_{C_1}} \mathbb{E}_{\mathbb{P}} [f(\mathbf{x}, \boldsymbol{\xi})]. \quad (2.65)$$

By the definition of  $\zeta$ , we have  $\max_{\mathbb{P}_{\zeta} \in \mathcal{D}_{\zeta}} \mathbb{E}_{\mathbb{P}_{\zeta}} [f(\mathbf{x}, \zeta)] = \max_{\mathbb{P}_r \in \mathcal{D}_{C_3}} \mathbb{E}_{\mathbb{P}_r} [f(\mathbf{x}, \mathbf{U}_{m \times m_1} \Lambda_{m_1}^{\frac{1}{2}} \boldsymbol{\xi}_r + \boldsymbol{\mu})]$  due to change of variable. Then (2.65) implies the following inequality

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

which demonstrates that the optimal value of Problem (2.59) (i.e., Problem (2.60)) is a lower bound for that of Problem (DRO-C) (i.e., Problem (2.57)).

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

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

where  $\mathcal{D}_{r_i}$  (defined as (2.59b)) represents the ambiguity set of  $\boldsymbol{\xi}_{r_i}$  for any  $i \in [2]$ . For any  $\zeta_1 \sim \mathbb{P}_{\zeta_1} \in \mathcal{D}_{\zeta_1}$ , there exists a  $\boldsymbol{\xi}_{r_1} \sim \mathbb{P}_{r_1} \in \mathcal{D}_{r_1}$  such that  $\zeta_1 = \mathbf{U}_{m \times m_1} \Lambda_{m_1}^{\frac{1}{2}} \boldsymbol{\xi}_{r_1} + \boldsymbol{\mu} = \mathbf{U}_{m \times m_2} \Lambda_{m_2}^{\frac{1}{2}} \bar{\boldsymbol{\xi}}_{r_2} + \boldsymbol{\mu}$ , where  $\bar{\boldsymbol{\xi}}_{r_2} = (\boldsymbol{\xi}_{r_1}^{\top}, \mathbf{0}_{m_2 - m_1}^{\top})^{\top} \in \mathbb{R}^{m_2}$ . By using  $\mathcal{S}_{r_i}$  (defined as (2.59c)) to denote the support of  $\boldsymbol{\xi}_{r_i}$  for any  $i \in [2]$ , we have

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

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

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

It follows that the distribution of  $\bar{\boldsymbol{\xi}}_{r_2}$  belongs to  $\mathcal{D}_{r_2}$  and thus  $\mathbb{P}_{\zeta_1} \in \mathcal{D}_{\zeta_2}$ . Therefore, we have  $\mathcal{D}_{\zeta_1} \subset \mathcal{D}_{\zeta_2}$  and

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

That is, the optimal value of Problem (2.59) (i.e., Problem (2.60)) is nondecreasing in  $m_1$ .

Finally, Problem (2.59) is equivalent to Problem (2.58) when  $m_1 = m$ . Therefore, Problem (2.60) results in an exact reformulation of Problem (2.57).  $\square$

We show the quality of the outer approximation (2.60) in the following proposition.

**Proposition 2.9.** *It holds that*

$$0 \leq Z^*_C(m) - Z^*_C(m_1) \leq \sqrt{\frac{\gamma_2}{N}} \sum_{i=1}^N \sum_{k=1}^K \sqrt{\mathbf{L}_{m-m_1}^{ik} (\mathbf{L}_{m-m_1}^{ik})^\top}, \quad (2.66)$$

where  $\mathbf{L}_{m-m_1}^{ik} = \left(-y_k(\mathbf{x}^*) + \zeta^{i*}\right)^\top \mathbf{U}_{m \times (m-m_1)} (\boldsymbol{\Lambda}^{m-m_1})^{\frac{1}{2}}$ ,  $\mathbf{x}^*$  and  $\zeta^{i*}$  ( $\forall i \in [N]$ ) denote an optimal solution of Problem (2.60), and  $\boldsymbol{\Lambda}^{m-m_1} \in \mathbb{R}^{(m-m_1) \times (m-m_1)}$  denotes the lower-right submatrix of  $\boldsymbol{\Lambda}$ .

*Proof.* By Theorem 2.5, we have  $Z^*_C(m) - Z^*_C(m_1) \geq 0$ . Moreover, according to this theorem, Problem (2.57) and the following problem, i.e., Problem (2.60) with  $m_1 = m$ , have the same optimal value.

$$\min_{\mathbf{x}, \lambda, \mathbf{Q}, \hat{\zeta}, \hat{y}} \lambda R_0 + \gamma_2 \mathbf{I}_m \bullet \mathbf{Q} + \frac{1}{N} \sum_{i=1}^N y_i \quad (2.67a)$$

$$\text{s.t.} \quad \begin{bmatrix} \mathbf{Q} & \frac{1}{2} \left( (-y_k(\mathbf{x}) + \zeta^i)^\top \mathbf{U} \boldsymbol{\Lambda}^{\frac{1}{2}} \right)^\top \\ \frac{1}{2} (-y_k(\mathbf{x}) + \zeta^i)^\top \mathbf{U} \boldsymbol{\Lambda}^{\frac{1}{2}} & y_i - y_k(\mathbf{x})^\top \boldsymbol{\mu} - y_k^0(\mathbf{x}) + \zeta^{i\top} (\boldsymbol{\mu} - \hat{\boldsymbol{\xi}}^i) \end{bmatrix} \succeq 0, \quad \forall i \in [N], \forall k \in [K], \quad (2.67b)$$

$$\lambda \in \mathbb{R}_+, \mathbf{x} \in \mathcal{X}, \|\zeta^i\|_* \leq \lambda, \forall i \in [N], \quad (2.67c)$$

We use  $(\mathbf{x}^*, \lambda^*, \mathbf{Q}_r^*, \zeta^{i*} \forall i \in [N], y_i^* \forall i \in [N])$  to denote an optimal solution of Problem (2.60). Based on this optimal solution, we construct a feasible solution of Problem (2.67), represented by  $(\bar{\mathbf{x}}, \bar{\lambda}, \bar{\mathbf{Q}}, \bar{\zeta}^i \forall i \in [N], \bar{y}_i \forall i \in [N])$ . For clarity, we define

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

$\mathbf{L}_c^{ik} = \left(-y_k(\mathbf{x}^*) + \boldsymbol{\zeta}^{i*}\right)^\top \mathbf{U}_{m \times c} (\boldsymbol{\Lambda}^c)^{\frac{1}{2}}, \forall i \in [N], \forall k \in [K], \forall c \in \{m_1, m - m_1, m\}$ , where  $\boldsymbol{\Lambda}^{m_1} \in \mathbb{R}^{m_1 \times m_1}$  and  $\boldsymbol{\Lambda}^{m-m_1} \in \mathbb{R}^{(m-m_1) \times (m-m_1)}$  represent the upper-left and lower-right submatrices of  $\boldsymbol{\Lambda}$ , respectively.

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

$$\bar{\mathbf{Q}} = \begin{bmatrix} \mathbf{Q}_r^* & \mathbf{0}_{m_1 \times (m-m_1)} \\ \mathbf{0}_{(m-m_1) \times m_1} & \sum_{i=1}^N \sum_{k=1}^K \frac{s_1^{ik}}{4} (\mathbf{L}_{m-m_1}^{ik})^\top \mathbf{L}_{m-m_1}^{ik} \end{bmatrix}, \text{ and} \quad (2.68)$$

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

where  $s_1^{ik} > 0$  and  $s_2^{ik} > 0$  for any  $i \in [N]$  and  $k \in [K]$ . As  $\bar{\lambda} = \lambda^* \geq 0$ ,  $\bar{\mathbf{x}} = \mathbf{x}^* \in \mathcal{X}$ , and  $\|\bar{\boldsymbol{\zeta}}^i\|_* = \|\boldsymbol{\zeta}^{i*}\|_* \leq \lambda$ ,  $\forall i \in [N]$  due to constraint (2.60c), we only require  $(\bar{\mathbf{x}}, \bar{\lambda}, \bar{\mathbf{Q}}, \bar{\boldsymbol{\zeta}}^i \forall i \in [N], \bar{y}_i \forall i \in [N])$  to satisfy (2.67b). Thus, we will find the values of  $s_1^{ik}$  and  $s_2^{ik}$  for any  $i \in [N]$  and  $k \in [K]$  that enable this solution to satisfy (2.67b).

We plug  $(\bar{\mathbf{x}}, \bar{\lambda}, \bar{\mathbf{Q}}, \bar{\boldsymbol{\zeta}}^i \forall i \in [N], \bar{y}_i \forall i \in [N])$  to (2.67b) and use  $\bar{\mathbf{Y}}^{ik}$  for any  $i \in [N]$  and  $k \in [K]$  to denote the corresponding matrix in (2.67b). For any given  $i \in [N]$  and  $k \in [K]$ , we perform the following decomposition:

$$\begin{aligned} \bar{\mathbf{Y}}^{ik} &= \begin{bmatrix} \mathbf{Q}_r^* & \mathbf{0}_{m_1 \times (m-m_1)} & \frac{1}{2} (\mathbf{L}_{m_1}^{ik})^\top \\ \mathbf{0}_{(m-m_1) \times m_1} & \mathbf{0}_{(m-m_1) \times (m-m_1)} & \mathbf{0}_{(m-m_1) \times 1} \\ \frac{1}{2} \mathbf{L}_{m_1}^{ik} & \mathbf{0}_{1 \times (m-m_1)} & s^{ik} \end{bmatrix} + \begin{bmatrix} \mathbf{0}_{m_1 \times m_1} & \mathbf{0}_{m_1 \times (m-m_1)} & \mathbf{0}_{m_1 \times 1} \\ \mathbf{0}_{(m-m_1) \times m_1} & \sum_{i=1}^N \sum_{k=1}^K \frac{s^{ik}}{4} (\mathbf{L}_{m-m_1}^{ik})^\top \mathbf{L}_{m-m_1}^{ik} & \frac{1}{2} (\mathbf{L}_{m-m_1}^{ik})^\top \\ \mathbf{0}_{1 \times m_1} & \frac{1}{2} \mathbf{L}_{m-m_1}^{ik} & \sum_{k=1}^K s_2^{ik} \end{bmatrix} \\ &\succeq \begin{bmatrix} \mathbf{Q}_r^* & \mathbf{0}_{m_1 \times (m-m_1)} & \frac{1}{2} (\mathbf{L}_{m_1}^{ik})^\top \\ \mathbf{0}_{(m-m_1) \times m_1} & \mathbf{0}_{(m-m_1) \times (m-m_1)} & \mathbf{0}_{(m-m_1) \times 1} \\ \frac{1}{2} \mathbf{L}_{m_1}^{ik} & \mathbf{0}_{1 \times (m-m_1)} & s^{ik} \end{bmatrix} + \begin{bmatrix} \mathbf{0}_{m_1 \times m_1} & \mathbf{0}_{m_1 \times (m-m_1)} & \mathbf{0}_{m_1 \times 1} \\ \mathbf{0}_{(m-m_1) \times m_1} & \frac{s^{ik}}{4} (\mathbf{L}_{m-m_1}^{ik})^\top \mathbf{L}_{m-m_1}^{ik} & \frac{1}{2} (\mathbf{L}_{m-m_1}^{ik})^\top \\ \mathbf{0}_{1 \times m_1} & \frac{1}{2} \mathbf{L}_{m-m_1}^{ik} & s_2^{ik} \end{bmatrix}. \quad (2.69) \end{aligned}$$

The first matrix in (2.69) is clearly PSD because the elimination of its zero components leads to a PSD matrix due to constraints (2.60b). Now we find the values of  $s_1^{ik}$  and  $s_2^{ik}$  to make the second matrix PSD as well, and then accordingly the constructed solution is feasible for (2.67).

Next, we use  $\begin{bmatrix} \bar{\mathbf{A}} & \bar{\mathbf{B}} \\ \bar{\mathbf{B}}^\top & \bar{\mathbf{C}} \end{bmatrix}$  to denote the second matrix in (2.69) by letting  $\bar{\mathbf{A}} = \begin{bmatrix} \mathbf{0}_{m_1 \times m_1} & \mathbf{0}_{m_1 \times (m-m_1)} \\ \mathbf{0}_{(m-m_1) \times m_1} & \frac{s^{ik}}{4} (\mathbf{L}_{m-m_1}^{ik})^\top \mathbf{L}_{m-m_1}^{ik} \end{bmatrix}$ ,  $\bar{\mathbf{B}}^\top = (\mathbf{0}_{1 \times m_1} \frac{1}{2} \mathbf{L}_{m-m_1}^{ik})$ , and  $\bar{\mathbf{C}} = s_2^{ik}$ . It follows that

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

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

In addition, since Problem (2.67) is a minimization problem, its optimal value is no larger than the objective corresponding to the feasible solution  $(\bar{\mathbf{x}}, \bar{\lambda}, \bar{\mathbf{Q}}, \bar{\boldsymbol{\zeta}}^i \forall i \in [N], \bar{y}_i \forall i \in [N])$ . That is,

$$\begin{aligned} Z^*_{\text{C}}(m) &\leq \bar{\lambda}R_0 + \gamma_2 \mathbf{I}_m \bullet \bar{\mathbf{Q}} + \frac{1}{N} \sum_{i=1}^N \bar{y}_i \\ &= Z^*_{\text{C}}(m_1) + \gamma_2 \sum_{i=1}^N \sum_{k=1}^K \frac{s_1^{ik}}{4} \text{trace} \left( \left( \mathbf{L}_{m-m_1}^{ik} \right)^\top \mathbf{L}_{m-m_1}^{ik} \right) + \frac{1}{N} \sum_{i=1}^N \sum_{k=1}^K s_2^{ik}. \end{aligned} \quad (2.70)$$

Due to the condition  $s_1^{ik} \times s_2^{ik} \geq 1$ , we let  $s_1^{ik} = \frac{2}{\sqrt{\gamma_2 N \mathbf{L}_{m-m_1}^{ik} (\mathbf{L}_{m-m_1}^{ik})^\top}}$  and  $s_2^{ik} = \frac{\sqrt{\gamma_2 N \mathbf{L}_{m-m_1}^{ik} (\mathbf{L}_{m-m_1}^{ik})^\top}}{2}$  for any  $i \in [N]$  and  $k \in [K]$ , which leads to the smallest possible value of the RHS of (2.70). Therefore, we have

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

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

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

□

### 2.3.2 Upper Bound

We further inner approximate Problem (2.57), leading to an upper bound, by splitting random vector  $\boldsymbol{\xi}_1$  into  $P$  pieces in the second-moment constraint in  $\mathcal{D}_{\text{C}2}$  so that

$\boldsymbol{\xi}_I = (\boldsymbol{\xi}_{I_1}^\top, \boldsymbol{\xi}_{I_2}^\top, \dots, \boldsymbol{\xi}_{I_P}^\top)^\top$ , where  $\boldsymbol{\xi}_{I_j} \in \mathbb{R}^{m_j}$ ,  $\forall j \in [P]$ , and  $\sum_{j=1}^P m_j = m$ . This gives rise to an inner approximation

$$\min_{\mathbf{x} \in \mathcal{X}} \max_{\mathbb{P}_I \in \mathcal{D}_{C4}} \mathbb{E}_{\mathbb{P}_I} \left[ f \left( \mathbf{x}, \mathbf{U} \boldsymbol{\Lambda}^{\frac{1}{2}} \boldsymbol{\xi}_I + \boldsymbol{\mu} \right) \right], \quad (2.71)$$

where

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

**Theorem 2.6.** *Under Assumption 2.2, Problem (2.71) has the same optimal value as the following SDP formulation:*

$$\begin{aligned} UB_C^* := & \min_{\substack{\mathbf{x}, \lambda, \hat{\mathbf{Q}}, \\ \zeta, \hat{y}, \hat{s}}} \lambda R_0 + \gamma_2 \sum_{j=1}^P \mathbf{I}_{m_j} \bullet \mathbf{Q}_j + \frac{1}{N} \sum_{i=1}^N y_i & (2.72) \\ \text{s.t.} & \begin{bmatrix} \mathbf{Q}_j & \frac{1}{2} \left( (-y_k(\mathbf{x}) + \zeta^i)^\top \mathbf{U}_{m \times m_j} \boldsymbol{\Lambda}_{m_j}^{\frac{1}{2}} \right)^\top \\ \frac{1}{2} (-y_k(\mathbf{x}) + \zeta^i)^\top \mathbf{U}_{m \times m_j} \boldsymbol{\Lambda}_{m_j}^{\frac{1}{2}} & s_{jik} \end{bmatrix} \succeq 0, \\ & \forall j \in [P], \forall i \in [N], \forall k \in [K], \\ & \sum_{j=1}^P s_{jik} = y_i - y_k(\mathbf{x})^\top \boldsymbol{\mu} - y_k^0(\mathbf{x}) + \left( \boldsymbol{\mu} - \hat{\boldsymbol{\xi}}^i \right)^\top \boldsymbol{\zeta}^i, \forall i \in [N], \forall k \in [K], \\ & \lambda \in \mathbb{R}_+, \mathbf{x} \in \mathcal{X}, \|\boldsymbol{\zeta}^i\|_* \leq \lambda, \forall i \in [N], \end{aligned}$$

where  $\mathbf{Q}_j \in \mathbb{R}^{m_j \times m_j}$ ,  $\boldsymbol{\zeta}^i \in \mathbb{R}^m$  for any  $i \in [N]$ ,  $\hat{\mathbf{Q}} = \{\mathbf{Q}_1, \dots, \mathbf{Q}_P\}$ ,  $\hat{\boldsymbol{\zeta}} = \{\boldsymbol{\zeta}^1, \dots, \boldsymbol{\zeta}^N\}$ ,  $\hat{y} = \{y_1, \dots, y_N\}$ , and  $\hat{s} = \{s_{jik}, \forall i \in [N], \forall j \in [P], \forall k \in [K]\}$ . Furthermore, Problem (2.72) provides an upper bound for the optimal value of (DRO-C).

*Proof.* By Theorem 1 in Gao and Kleywegt (2017), Problem (2.71) has the following strong dual problem:

$$\min_{\mathbf{x} \in \mathcal{X}, \mathbf{Q}_j \forall j, \lambda} \left\{ \lambda R_0 + \gamma_2 \sum_{j=1}^P \mathbf{I}_{m_j} \bullet \mathbf{Q}_j + \int_{\mathbb{R}^m} \sup_{\boldsymbol{\xi}_I} g(\boldsymbol{\xi}_I, \hat{\boldsymbol{\xi}}) \mathbb{P}_0(d\hat{\boldsymbol{\xi}}) \right\}, \quad (2.73)$$

where  $g(\boldsymbol{\xi}_I, \hat{\boldsymbol{\xi}}) = \max_{k=1}^K \{y_k(\mathbf{x})^\top (\mathbf{U} \boldsymbol{\Lambda}^{\frac{1}{2}} \boldsymbol{\xi}_I + \boldsymbol{\mu}) + y_k^0(\mathbf{x})\} - \sum_{j=1}^P \boldsymbol{\xi}_{I_j}^\top \mathbf{Q}_j \boldsymbol{\xi}_{I_j} - \lambda \|\mathbf{U} \boldsymbol{\Lambda}^{\frac{1}{2}} \boldsymbol{\xi}_I + \boldsymbol{\mu} - \hat{\boldsymbol{\xi}}\|_1$ , and  $\mathbf{Q}_j$  for any  $j \in [P]$  and  $\lambda$  are the Lagrangian multipliers of the primal

second-order moment and Wasserstein constraints, respectively. As  $\mathbb{P}_0$  denotes an empirical distribution of  $\boldsymbol{\xi}$  generated by i.i.d. samples  $\{\hat{\boldsymbol{\xi}}^i : i \in N\} \subseteq \mathcal{S}$  from the  $\mathbb{P}$ , i.e.,  $\mathbb{P}\{\boldsymbol{\xi} = \hat{\boldsymbol{\xi}}^i\} = \frac{1}{N}$ , we have

$$\int_{\mathbb{R}^m} \sup_{\boldsymbol{\xi}_I} g(\boldsymbol{\xi}_I, \hat{\boldsymbol{\xi}}) \mathbb{P}_0(d\hat{\boldsymbol{\xi}}) = \frac{1}{N} \sum_{i=1}^N \sup_{\boldsymbol{\xi}_I} g(\boldsymbol{\xi}_I, \hat{\boldsymbol{\xi}}^i). \quad (2.74)$$

because  $g(\boldsymbol{\xi}_I, \hat{\boldsymbol{\xi}})$  is a convex function and  $\mathcal{S}_I = \mathbb{R}^m$ , which is convex. Thus, by plugging (2.74) into (2.73), (2.73) can be reformulated as

$$\min_{\mathbf{x} \in \mathcal{X}, \mathbf{Q}_j \forall j, \lambda} \left\{ \lambda R_0 + \gamma_2 \sum_{j=1}^P \mathbf{I}_{m_j} \bullet \mathbf{Q}_j + \frac{1}{N} \sum_{i=1}^N y_i \right\} \quad (2.75a)$$

$$\text{s.t. } y_i = \sup_{\boldsymbol{\xi}_I} g(\boldsymbol{\xi}_I, \hat{\boldsymbol{\xi}}^i), \quad \forall i \in [N]. \quad (2.75b)$$

Since Problem (2.75) is a minimization problem, constraints (2.75b) can be relaxed to  $y_i \geq \sup_{\boldsymbol{\xi}_I} g(\boldsymbol{\xi}_I, \hat{\boldsymbol{\xi}}^i)$  for any  $i \in [N]$ . Thus, we have

$$\begin{aligned} y_i &\geq \sup_{\boldsymbol{\xi}_I} \left\{ \max_{k=1}^K \left\{ y_k(\mathbf{x})^\top (\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_I + \boldsymbol{\mu}) + y_k^0(\mathbf{x}) \right\} - \sum_{j=1}^P \boldsymbol{\xi}_{I_j}^\top \mathbf{Q}_j \boldsymbol{\xi}_{I_j} - \lambda \left\| \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_I + \boldsymbol{\mu} - \hat{\boldsymbol{\xi}}^i \right\|_1 \right\}, \quad \forall i \in [N] \\ \Leftrightarrow y_i &\geq \max_{k=1}^K \sup_{\boldsymbol{\xi}_I} \left\{ y_k(\mathbf{x})^\top (\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_I + \boldsymbol{\mu}) + y_k^0(\mathbf{x}) - \sum_{j=1}^P \boldsymbol{\xi}_{I_j}^\top \mathbf{Q}_j \boldsymbol{\xi}_{I_j} - \lambda \left\| \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_I + \boldsymbol{\mu} - \hat{\boldsymbol{\xi}}^i \right\|_1 \right\}, \quad \forall i \in [N] \\ \Leftrightarrow y_i &\geq \sup_{\boldsymbol{\xi}_I} \left\{ y_k(\mathbf{x})^\top (\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_I + \boldsymbol{\mu}) + y_k^0(\mathbf{x}) - \sum_{j=1}^P \boldsymbol{\xi}_{I_j}^\top \mathbf{Q}_j \boldsymbol{\xi}_{I_j} - \lambda \left\| \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_I + \boldsymbol{\mu} - \hat{\boldsymbol{\xi}}^i \right\|_1 \right\}, \quad \forall i \in [N], \forall k \in [K]. \end{aligned} \quad (2.76)$$

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

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

and accordingly we have

$$\begin{aligned}
& y_i \geq \sup_{\xi_{\mathbf{I}}} \inf_{\|\hat{\zeta}\|_* \leq 1} \left\{ y_k(\mathbf{x})^\top \left( \mathbf{U} \boldsymbol{\Lambda}^{\frac{1}{2}} \xi_{\mathbf{I}} + \boldsymbol{\mu} \right) + y_k^0(\mathbf{x}) - \sum_{j=1}^P \xi_{\mathbf{I}_j}^\top \mathbf{Q}_j \xi_{\mathbf{I}_j} - \lambda \hat{\zeta}^\top \left( \mathbf{U} \boldsymbol{\Lambda}^{\frac{1}{2}} \xi_{\mathbf{I}} + \boldsymbol{\mu} - \hat{\xi}^i \right) \right\}, \forall k \in [K] \\
\Leftrightarrow & y_i \geq \inf_{\|\hat{\zeta}\|_* \leq 1} \sup_{\xi_{\mathbf{I}}} \left\{ y_k(\mathbf{x})^\top \left( \mathbf{U} \boldsymbol{\Lambda}^{\frac{1}{2}} \xi_{\mathbf{I}} + \boldsymbol{\mu} \right) + y_k^0(\mathbf{x}) - \sum_{j=1}^P \xi_{\mathbf{I}_j}^\top \mathbf{Q}_j \xi_{\mathbf{I}_j} - \lambda \hat{\zeta}^\top \left( \mathbf{U} \boldsymbol{\Lambda}^{\frac{1}{2}} \xi_{\mathbf{I}} + \boldsymbol{\mu} - \hat{\xi}^i \right) \right\}, \forall k \in [K] \\
\Leftrightarrow & \exists \hat{\zeta} \text{ s.t. } \|\hat{\zeta}\|_* \leq 1, y_i \geq \sup_{\xi_{\mathbf{I}}} \left\{ y_k(\mathbf{x})^\top \left( \mathbf{U} \boldsymbol{\Lambda}^{\frac{1}{2}} \xi_{\mathbf{I}} + \boldsymbol{\mu} \right) + y_k^0(\mathbf{x}) - \sum_{j=1}^P \xi_{\mathbf{I}_j}^\top \mathbf{Q}_j \xi_{\mathbf{I}_j} - \lambda \hat{\zeta}^\top \left( \mathbf{U} \boldsymbol{\Lambda}^{\frac{1}{2}} \xi_{\mathbf{I}} + \boldsymbol{\mu} - \hat{\xi}^i \right) \right\}, \\
& \hspace{25em} \forall k \in [K] \\
\Leftrightarrow & \exists \hat{\zeta} \text{ s.t. } \|\hat{\zeta}\|_* \leq 1, y_i \geq y_k(\mathbf{x})^\top \left( \mathbf{U} \boldsymbol{\Lambda}^{\frac{1}{2}} \xi_{\mathbf{I}} + \boldsymbol{\mu} \right) + y_k^0(\mathbf{x}) - \sum_{j=1}^P \xi_{\mathbf{I}_j}^\top \mathbf{Q}_j \xi_{\mathbf{I}_j} - \lambda \hat{\zeta}^\top \left( \mathbf{U} \boldsymbol{\Lambda}^{\frac{1}{2}} \xi_{\mathbf{I}} + \boldsymbol{\mu} - \hat{\xi}^i \right), \\
& \hspace{25em} \forall \xi_{\mathbf{I}} \in \mathbb{R}^m, \forall k \in [K] \\
\Leftrightarrow & \exists \hat{\zeta} \text{ s.t. } \|\hat{\zeta}\|_* \leq 1, \begin{bmatrix} Q' & \frac{1}{2} \left( (-y_k(\mathbf{x})^\top + \lambda \hat{\zeta}^\top) \mathbf{U} \boldsymbol{\Lambda}^{\frac{1}{2}} \right)^\top \\ \frac{1}{2} (-y_k(\mathbf{x})^\top + \lambda \hat{\zeta}^\top) \mathbf{U} \boldsymbol{\Lambda}^{\frac{1}{2}} & y_i - y_k(\mathbf{x})^\top \boldsymbol{\mu} - y_k^0(\mathbf{x}) + \lambda \hat{\zeta}^\top (\boldsymbol{\mu} - \hat{\xi}^i) \end{bmatrix} \succeq 0, \forall k \in [K], \\
& \hspace{25em} (2.77)
\end{aligned}$$

where decision variable  $Q'$  is described as (2.49) and the first equivalence is due to the convexity of  $g(\xi_{\mathbf{I}}, \hat{\xi})$ ,  $\mathcal{S}_{\mathbf{I}}$ , and the feasible region defined by  $\|\hat{\zeta}\|_* \leq 1$ . For any given  $i \in [N]$ , we replace  $\lambda \hat{\zeta}$  by  $\zeta^i$ , and then we can reduce Problem (2.71) to the following problem by further replacing (2.75b) by (2.77) for any  $i \in [N]$ :

$$\min_{\substack{\mathbf{x}, \lambda, \mathbf{Q}_j \forall j, \\ \zeta^i \forall i, y_i \forall i}} \lambda R_0 + \gamma_2 \sum_{j=1}^P \mathbf{I}_{m_j} \bullet \mathbf{Q}_j + \frac{1}{N} \sum_{i=1}^N y_i \quad (2.78a)$$

$$\text{s.t. } \begin{bmatrix} Q' & \frac{1}{2} \left( (-y_k(\mathbf{x}) + \zeta^i)^\top \mathbf{U} \boldsymbol{\Lambda}^{\frac{1}{2}} \right)^\top \\ \frac{1}{2} (-y_k(\mathbf{x}) + \zeta^i)^\top \mathbf{U} \boldsymbol{\Lambda}^{\frac{1}{2}} & y_i - y_k(\mathbf{x})^\top \boldsymbol{\mu} - y_k^0(\mathbf{x}) + \zeta^{i\top} (\boldsymbol{\mu} - \hat{\xi}^i) \end{bmatrix} \succeq 0, \forall i \in [N], \quad (2.78b)$$

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

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

Finally, by Lemma 2.1, we reformulate Problem (2.78) as Problem (2.72) by decomposing the PSD matrix in (2.78b) equivalently to  $K$  PSD matrices. The proof of the claim that Problem (2.72) provides an upper bound for Problem (2.57) is the same as that of Theorem 2.4 and thus is omitted here.  $\square$

We show the quality of our derived inner approximation (2.72) in the following proposition.

**Proposition 2.10.** *If  $\min_{k=1}^K \{y_k^0(\mathbf{x}^*) + y_k(\mathbf{x}^*)^\top \boldsymbol{\mu}\} \geq 0$  and  $\max_{i=1}^N \{(\boldsymbol{\mu} - \hat{\boldsymbol{\xi}}^i)^\top \boldsymbol{\zeta}^{i*}\} \leq 0$ , where  $\mathbf{x}^*$  and  $\boldsymbol{\zeta}^{i*}$  are optimal solutions of Problem (2.57), then the relative gap between the optimal values of Problems (2.72) and (2.57) is bounded from above by  $\sqrt{P} - 1$ , i.e.,  $0 \leq UB_C^* - Z_C^*(m) \leq (\sqrt{P} - 1)Z_C^*(m)$ .*

*Proof.* We reformulate Problem (2.72) as Problem (2.78). Let  $(\mathbf{x}^*, \lambda^*, \mathbf{Q}^*, \boldsymbol{\zeta}^{i*} \forall i \in [N], y_i^* \forall i \in [N])$  denote an optimal solution of Problem (2.57) with  $\mathbf{Q}^*$  represented by (2.50). Based on this optimal solution, in the following, we construct a feasible solution of Problem (2.78), denoted by  $(\bar{\mathbf{x}}, \bar{\lambda}, \bar{\mathbf{Q}}_j \forall j \in [P], \bar{\boldsymbol{\zeta}}^i \forall i \in [N], \bar{y}_i \forall i \in [N])$ .

First, we let  $\bar{\mathbf{x}} = \mathbf{x}^*$ ,  $\bar{\lambda} = \lambda^*$ ,  $\bar{\boldsymbol{\zeta}}^i = \boldsymbol{\zeta}^{i*}$  for any  $i \in [N]$ ,  $\bar{y}_i = k_0 y_i^*$  for any  $i \in [N]$ , and  $\bar{\mathbf{Q}}_j = \bar{\mathbf{Q}}'$  (as described in (2.51)), with  $k_j \geq 1$  for any  $j \in \{0, 1, 2, \dots, P\}$ . In order for this solution to satisfy (2.78b), we require

$$\begin{bmatrix} \bar{\mathbf{Q}}' & \frac{1}{2} \left( (-y_k(\mathbf{x}^*) + \boldsymbol{\zeta}^{i*})^\top \mathbf{U} \boldsymbol{\Lambda}^{\frac{1}{2}} \right)^\top \\ \frac{1}{2} (-y_k(\mathbf{x}^*) + \boldsymbol{\zeta}^{i*})^\top \mathbf{U} \boldsymbol{\Lambda}^{\frac{1}{2}} & k_0 y_i^* - y_k(\mathbf{x}^*)^\top \boldsymbol{\mu} - y_k^0(\mathbf{x}^*) + \boldsymbol{\zeta}^{i* \top} (\boldsymbol{\mu} - \hat{\boldsymbol{\xi}}^i) \end{bmatrix} \succeq 0, \quad \forall i \in [N], \quad \forall k \in [K]. \quad (2.79)$$

In the following, we find the values of  $k_j$  for any  $j \in \{0, 1, 2, \dots, P\}$  so that (2.79) holds. To that end, we construct the following matrix

$$\begin{bmatrix} \bar{\mathbf{Q}}' & \frac{1}{2} \left( (-y_k(\mathbf{x}^*) + \boldsymbol{\zeta}^{i*})^\top \mathbf{U} \boldsymbol{\Lambda}^{\frac{1}{2}} \right)^\top \\ \frac{1}{2} (-y_k(\mathbf{x}^*) + \boldsymbol{\zeta}^{i*})^\top \mathbf{U} \boldsymbol{\Lambda}^{\frac{1}{2}} & k_0 \left( y_i^* - y_k(\mathbf{x}^*)^\top \boldsymbol{\mu} - y_k^0(\mathbf{x}^*) + \boldsymbol{\zeta}^{i* \top} (\boldsymbol{\mu} - \hat{\boldsymbol{\xi}}^i) \right) \end{bmatrix}, \quad \forall i \in [N], \quad \forall k \in [K]. \quad (2.80)$$

Note that subtracting (2.80) from (2.79) leads to the following matrix:

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

which is PSD because its eigenvalues are non-negative. In fact,  $(k_0 - 1)(y_k^0(\mathbf{x}^*) + y_k(\mathbf{x}^*)^\top \boldsymbol{\mu} - \boldsymbol{\zeta}^{i* \top} (\boldsymbol{\mu} - \hat{\boldsymbol{\xi}}^i))$  is the only non-zero eigenvalue of this matrix that is non-negative because  $k_0 \geq 1$ ,  $-\boldsymbol{\zeta}^{i* \top} (\boldsymbol{\mu} - \hat{\boldsymbol{\xi}}^i) \geq 0$  due to the assumption  $\max_{i=1}^N \{ \boldsymbol{\zeta}^{i* \top} (\boldsymbol{\mu} - \hat{\boldsymbol{\xi}}^i) \} \leq 0$ , and we have  $y_k^0(\mathbf{x}^*) + y_k(\mathbf{x}^*)^\top \boldsymbol{\mu} \geq 0$  according to the assumption  $\min_{k=1}^K \{ y_k^0(\mathbf{x}^*) + y_k(\mathbf{x}^*)^\top \boldsymbol{\mu} \} \geq 0$ . Thus, we choose good values of  $k_j$  for any  $j \in \{0, 1, 2, \dots, P\}$  to ensure (2.80) to be a PSD matrix and accordingly will make (2.79) hold.

Next, by Lemma 2.1, in order for (2.80) to be a PSD, we equivalently require

$$\left[ \begin{array}{cc} k_j \mathbf{Q}_j^* & \frac{1}{2} \left( (-y_k(\mathbf{x}^*) + \boldsymbol{\zeta}^{i*})^\top \mathbf{U}_{m \times m_j} \boldsymbol{\Lambda}_{m_j}^{\frac{1}{2}} \right)^\top \\ \frac{1}{2} (-y_k(\mathbf{x}^*) + \boldsymbol{\zeta}^{i*})^\top \mathbf{U}_{m \times m_j} \boldsymbol{\Lambda}_{m_j}^{\frac{1}{2}} & s_j \left( \mathbf{y}_i^* - y_k(\mathbf{x}^*)^\top \boldsymbol{\mu} - y_k^0(\mathbf{x}^*) + \boldsymbol{\zeta}^{i* \top} (\boldsymbol{\mu} - \boldsymbol{\xi}^i) \right) \end{array} \right] \succeq 0, \quad \forall k \in [K], \quad (2.81)$$

$$\forall i \in [N], \quad \forall j \in [P],$$

with  $\sum_{j=1}^P s_j = k_0$ . Constraints (2.81) can be satisfied by allowing  $s_j \times k_j \geq 1$  for any  $j \in [P]$  due to (2.57b). Then, we let  $k_0 = k_1 = \dots = k_P$  and  $s_j \times k_j = 1$  for any  $j \in [P]$ , leading to  $k_0 = k_1 = \dots = k_P = \sqrt{P}$ .

Finally, we have  $UB_C^* \geq Z_C^*(m)$  by Theorem 2.6. Meanwhile, as Problem (2.72) is a minimization problem,  $UB_C^*$  is no larger than the objective value corresponding to our constructed feasible solution  $(\bar{\mathbf{x}}, \bar{\lambda}, \bar{\mathbf{Q}}_j \ \forall j \in [P], \bar{\boldsymbol{\zeta}}^i \ \forall i \in [N], \bar{y}_i \ \forall i \in [N])$ . That is, we have

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

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

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

□

**Remark 2.4.** *The theoretical upper bound in Proposition 2.10 is achievable. Indeed, when we enlarge  $R_0$  enough such that  $\mathcal{D}_{C_4}$  degenerates to a moment-based ambiguity set, we can follow the same setting of Example 2.1 to achieve the theoretical upper bound in Proposition 2.10. Meanwhile, similar to our notes in Remark 2.3, the theoretical bound in Proposition 2.10 is also conservative and only achievable in pathetic problem instances. The quality of the inner approximation (2.72) is usually much more optimistic due to the corresponding numerical results in Section 2.4.*

Before closing this section, we note that the first-order moment constraint, albeit important to reduce solution conservatism, is not included in the combined ambiguity set of (DRO-C). We consider additionally including the first-order moment

constraint, leading to a DRO problem with a combined ambiguity set consisting of Wasserstein distance and first- and second-order moment information; see Problem (DRO-C2) in Section 2.6. We correspondingly derive the inner and outer approximations of (DRO-C2), and conduct computational experiments to compare their performance with the approximations of (DRO-C). Based on the results, we observe that adding the first-order moment information does further reduce the conservatism of optimal solutions for both (DRO-C) and its approximations, though such reduction is not significant. This is because the second-order moment information together with the Wasserstein information has already (partially) implied the first-order moment information. Such insight is also explained in Gao and Kleywegt (2017). Due to this, we only keep (DRO-C) in the main body.

## 2.4 Computational Experiments

We perform extensive computational experiments to demonstrate the effectiveness of our proposed inner and outer approximations in two applications: production-transportation and multi-product newsvendor problems. The mathematical models are implemented in MATLAB R2017a (ver. 9.2) by the modeling language CVX (ver. 2.1) (Grant and Boyd 2008, 2014) with the Mosek solver (8.0.0.60) on a PC with 64-bit Windows Operating System, an Intel(R) Core(TM) i7-7700 CPU @ 3.60 GHz processor, and a 16 GB RAM. The time limit for each run is set at 36 hours. In Section 2.4.1, we specify the proposed lower and upper bounds in the context of the two aforementioned applications. In Section 2.4.2, we explain how to randomly generate test instances and report the numerical results together with analyses.

### 2.4.1 Computational Setup

In this section, we specify the proposed lower and upper bounds, as well as the theoretical upper bounds for their gaps with the original DRO model, in the context

of production-transportation and multi-product newsvendor problems.

#### 2.4.1.1 Production-Transportation Problem

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

$$\min_{\mathbf{x}, \mathbf{z}} \quad \sum_{i=1}^m c_i x_i + \sum_{i=1}^m \sum_{j=1}^n \xi_{ij} z_{ij} \quad (2.82a)$$

$$\text{s.t.} \quad \sum_{i=1}^m z_{ij} = d_j, \quad \forall j \in [n], \quad (2.82b)$$

$$\sum_{j=1}^n z_{ij} = x_i, \quad \forall i \in [m], \quad (2.82c)$$

$$0 \leq x_i \leq 1, \quad \forall i \in [m], \quad (2.82d)$$

$$z_{ij} \geq 0, \quad \forall i \in [m], \quad \forall j \in [n]. \quad (2.82e)$$

Now we derive the DRO counterpart of Problem (2.82). Specifically, we assume that  $\boldsymbol{\xi}$  is random and its probability distribution  $\mathbb{P}$  is unknown but it belongs to a predefined distributional ambiguity set  $\mathcal{D}$ . The decision  $\mathbf{x}$  is decided before the realization of randomness and  $\mathbf{z}$  is made as a recourse to specific realizations (Bertsimas et al. 2010). This leads to a two-stage DRO counterpart

$$\min_{\mathbf{x}} \left\{ \mathbf{c}^\top \mathbf{x} + \max_{\mathbb{P} \in \mathcal{D}} \mathbb{E}_{\mathbb{P}} [\mathcal{U}(\mathcal{Q}(\mathbf{z}, \boldsymbol{\xi}))] : (2.82d) \right\}, \quad (2.83)$$

where  $\mathcal{U}(\cdot)$  is a convex nondecreasing disutility function used to incorporate risk

considerations into the second-stage cost. In particular, we define

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

where  $\mathcal{Q}(z, \boldsymbol{\xi}) = \min_z \{z^\top \boldsymbol{\xi} : (2.82b), (2.82c), (2.82e)\}$ . We can apply the proposed inner and outer approximations (i.e., Problems (2.15), (2.60), (2.24), (2.37), and (2.72)) to approximate Problem (2.83) in the context of production-transportation problem, with the details provided in Section 2.7.

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

#### 2.4.1.2 Multi-Product Newsvendor Problem

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

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

where  $\mathbf{c}$  represents the wholesale price,  $\mathbf{v}$  represents the retail price,  $\mathbf{g}$  represents the salvage price, and the minimum and nonnegativity operator are applied componentwise. Now we consider that demand  $\boldsymbol{\xi}$  is uncertain and its probability distribution belongs to a distributional ambiguity set  $\mathcal{D}$ . The DRO counterpart of the

multi-product newsvendor problem to minimize the expected total loss against the worst-case distribution in  $\mathcal{D}$  can be described as follows:

$$\min_{\mathbf{x} \geq 0} \max_{\mathbb{P} \in \mathcal{D}} \mathbb{E}_{\mathbb{P}} \left[ \max \left\{ (\mathbf{c} - \mathbf{v})^\top \mathbf{x}, (\mathbf{c} - \mathbf{g})^\top \mathbf{x} + (\mathbf{g} - \mathbf{v})^\top \boldsymbol{\xi} \right\} \right]. \quad (2.84)$$

Note that the procedure of applying the proposed inner and outer approximations and the theoretical bounds to Problem (2.84) is similar to that for Problem (2.83) and thus is omitted here.

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

## 2.4.2 Computational Results

We first evaluate the performance of our proposed lower and upper bounds and then show how they can help construct a tight interval, which includes unknown optimal solutions of large-sized DRO problems that cannot be solved to optimality (or even feasibility in most cases) by existing methods in reasonable time. We further investigate the benefits of choosing the components with the largest variability (i.e., leading components) in the PCA approach (as compared to randomly choosing components) and also perform sensitivity analyses with respect to several given parameters.

### 2.4.2.1 Instance Generation and Table Header Description

We perform our experiments to solve various instances. First, we consider different levels of problem size, namely small, medium and large, by varying  $m$

and  $n$  in the production-transportation problem and varying  $n$  in the newsvendor problem. Second, we consider different levels of approximation. In particular, for PCA-based lower and upper bounds, we consider different values of  $\frac{m_1}{m}$  in  $\{100\%, 75\%, 50\%, 25\%, 10\%\}$ . For vector splitting based upper bounds, we consider  $P \in \{2, 4, 5\}$ , by which the random vector  $\boldsymbol{\xi}$  is equally split with  $m_1 = m_2 = \dots = m_P$ ; that is, here we use a homogeneous division of the random vector. This is because such homogeneous division eventually leads to a set of SDP constraints, with all of them having the smallest possible matrices simultaneously, in our proposed formulations. It helps improve the computational performance of the corresponding inner approximation. In addition, when assigning all the components of  $\boldsymbol{\xi}$  to pieces, we consider the non-increasing order of the components based on their variance. Specifically, the first piece includes the first  $m_1$  components with the largest variance, the next  $m_2$  components are assigned to the second pieces, and so on. The reason is that in our approximation, the correlations among different pieces are dropped, but we would like to capture the correlations among those important components within each piece. Third, we consider different supports, i.e.,  $\mathcal{S} \in \{[-2\boldsymbol{\sigma}, 2\boldsymbol{\sigma}], [-3\boldsymbol{\sigma}, 3\boldsymbol{\sigma}], [-4\boldsymbol{\sigma}, 4\boldsymbol{\sigma}]\}$ , where  $\boldsymbol{\sigma}$  represents the sample standard deviation of random vector  $\boldsymbol{\xi}$ . In addition,  $\gamma_1$  and  $\gamma_2$  are set as 1 and 2, respectively, for the moment-based ambiguity set, the number of i.i.d data samples  $N = 10$  for the combined ambiguity set, the Wasserstein radius  $R_0$  is set as 30 for the production-transportation problem and as 700 for the newsvendor problem. We will perform sensitivity analyses with respect to other different values of  $\gamma_1$ ,  $\gamma_2$ , and  $R_0$ . For each combination of the above three variants, we randomly generate five instances and report the average results over them.

In the following Sections 2.4.2.2 - 2.4.2.6, we will use tables to report our results and here we describe several table headers that are shared by most of the tables. Column “Size” reports the values of  $m$  and  $n$  in the production-transportation problem and  $n$  in the newsvendor problem, indicating different levels of the problem size. Column “Orig.” represents the computational time in seconds required to

solve the original DRO problem and column “Time” represents the computational time in seconds required to solve the corresponding inner or outer approximations. Column “Gap” represents the relative gap in percentage between a lower or upper bound and the original optimal value. Here, the relative gap between two values is defined as their difference divided by the maximum. As such relative gaps are theoretically bounded from above, e.g., (2.16) and (2.39), we use column “Gap2” in percentage to represent the value of theoretical bound. Whenever needed, we use “LB” and “UB” to denote the lower and upper bounds, respectively. Note that the percentage of  $\xi$ 's components utilized to construct lower and upper bounds, i.e.,  $\frac{m_1}{m} \times 100\%$ , is represented by  $\frac{m_1}{m}(\%)$  and  $P$  represents the number of split pieces of  $\xi$ , with each piece having the same size.

#### 2.4.2.2 Lower Bound Performance

We summarize the lower bounds of the DRO problem with the moment-based ambiguity set on both applications in Tables 2.1 and 2.2, while Tables 2.3 and 2.4 report the results for the DRO problem with the combined ambiguity set.

Table 2.1: Lower bound (2.15) on the production-transportation problem

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

From Tables 2.1 and 2.2, where the column “Gap2” represents the relative gap induced by the theoretical upper bound in (2.16), we have the following observations. First, when the number of principal components  $m_1$  increases, both Gap and Gap2 decrease and the computational time increases. When  $m_1$  increases to  $m$ , we obtain the lower bound equivalent to the original optimal value but in a large computational

Table 2.2: Lower bound (2.15) on the newsvendor problem

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

time. In practice, thus we can leverage the number of principal components as a tool to trade-off between solution quality and computational time. Second, when the problem size increases, the original problem becomes more difficult to solve, while our approximations reduce the computational time significantly and maintain very high solution quality. For instance, even with a 90% reduction in the dimension of uncertainty space, our approximation solution can give an objective value at around 2% of the original optimal value, as shown in the column “Gap”. Third, when comparing the values of Gap and Gap2, we can observe that the latter is always larger than the former, demonstrating that theoretical bound (2.16) is valid. Meanwhile, the quality of theoretical bound (2.16) is sensitive to different problems and datasets, as Gap2 is closer to Gap in the newsvendor problem as compared to in the production-transportation problem. Similarly, the lower bound (2.15) performs better when applied to the newsvendor application, as the relative gap, i.e., Gap, is generally smaller than that in the production-transportation problem.

Table 2.3: Lower bound (2.60) on the production-transportation problem

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

From Tables 2.3 and 2.4, where the column “Gap2” represents the relative gap induced by the theoretical upper bound in (2.66), we have the similar observations as from Tables 2.1 and 2.2. In addition, a comparison among Tables 2.1 - 2.4

Table 2.4: Lower bound (2.60) on the newsvendor problem

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

shows that: (i) solving DRO problems with the combined ambiguity set and their outer approximations takes more computational time than solving those with the moment-based ambiguity set; and (ii) theoretical bound (2.66) is more conservative than theoretical bound (2.16) and can be improved in our future studies.

### 2.4.2.3 Upper Bound Performance

We report performance of the upper bound (2.24) in Tables 2.5 - 2.6 and report that of (2.37) and (2.72) in Tables 2.7 - 2.10.

Table 2.5: Upper bound (2.24) on the production-transportation problem

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

Table 2.6: Upper bound (2.24) on the newsvendor problem

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

From Tables 2.5 - 2.6, where the column “Gap2” represents the relative gap induced by the theoretical upper bound in (2.39), we have similar observations as from Tables 2.1 and 2.2. In addition, our approximation (2.24) performs better when solving the production-transportation problem as compared to solving the newsvendor problem. Meanwhile, the values of Gap2 are always larger than those of Gap, implying that theoretical bound (2.39) is valid.

Table 2.7: Upper bound (2.37) on the production-transportation problem

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

Table 2.8: Upper bound (2.37) on the newsvendor problem

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

From Tables 2.7 - 2.10, we have the following observations. First, when  $P$  increases, the computational time decreases and the Gap increases. In practice, thus we can leverage the number of split pieces as a tool to balance between solution quality and computational time. Second, the quality of upper bound (2.37) is sensitive to different problems and datasets because it performs better (i.e., with smaller Gap) on the production-transportation problem than on the newsvendor problem. Third, as mentioned in Remarks 2.3 and 2.4, although the theoretical error bound

$\sqrt{P} - 1$  is very conservative, as represented in the column “Gap2”, the computational optimality gap perform very well, as represented in the column “Gap”. In addition, by comparing Tables 2.5 - 2.6 and Tables 2.7 - 2.10, we can observe that the vector splitting based upper bounds are much tighter than the PCA based upper bounds.

Table 2.9: Upper bound (2.72) on the production-transportation problem

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

Table 2.10: Upper bound (2.72) on the newsvendor problem

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

#### 2.4.2.4 Interval Performance

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

The interval results of the DRO problem with the moment-based ambiguity set are summarized in Tables 2.11 and 2.12, while Tables 2.13 and 2.14 report the results for the DRO problem with the combined ambiguity set. The first row of each table,

Table 2.11: Intervals on the production-transportation problem

[LB, UB]		[(2.15), (2.24)]		[(2.15), (2.37)]				
Size ( $m, n$ )	Support	$(\frac{m_1}{m}\%, P)$	Orig. (secs)	Itv-Time (secs)	Itv-Gap (%)	Itv-Time (secs)	Itv-Gap (%)	
(6, 50)	$[-2\sigma, 2\sigma]$	(25%, 5)	-	100.0	7.98	203.5	3.53	
		(50%, 4)	-	1017.9	5.74	730.5	2.32	
		(75%, 2)	-	5626.6	2.56	3706.4	0.93	
	$[-3\sigma, 3\sigma]$	(25%, 5)	-	106.2	13.60	219.6	3.31	
		(50%, 4)	-	1071.3	10.75	813.3	2.42	
		(75%, 2)	-	5955.1	6.39	3913.1	1.23	
	$[-4\sigma, 4\sigma]$	(25%, 5)	-	102.7	20.37	233.2	3.54	
		(50%, 4)	-	1093.9	16.40	780.7	2.87	
		(75%, 2)	-	5564.6	6.82	4124.3	1.09	
	(8, 50)	$[-2\sigma, 2\sigma]$	(25%, 5)	-	265.7	4.94	604.7	2.78
			(50%, 4)	-	3375.2	4.53	2181.1	2.47
			(75%, 2)	-	-	-	-	-
$[-3\sigma, 3\sigma]$		(25%, 5)	-	302.6	10.28	629.1	2.80	
		(50%, 4)	-	3697.8	8.21	2667.7	2.21	
		(75%, 2)	-	-	-	-	-	
$[-4\sigma, 4\sigma]$		(25%, 5)	-	299.4	15.86	650	2.86	
		(50%, 4)	-	4129.7	14.08	2596.4	2.41	
		(75%, 2)	-	-	-	-	-	

Table 2.12: Intervals on the newsvendor problem

[LB, UB]		[(2.15), (2.24)]		[(2.15), (2.37)]				
Size ( $n$ )	Support	$(\frac{m_1}{m}\%, P)$	Orig. (secs)	Itv-Time (secs)	Itv-Gap (%)	Itv-Time (secs)	Itv-Gap (%)	
300	$[-2\sigma, 2\sigma]$	(25%, 5)	-	10.3	33.06	11.1	2.99	
		(50%, 4)	-	227.0	14.31	124.5	1.98	
		(75%, 2)	-	1703.2	3.24	982.4	0.71	
	$[-3\sigma, 3\sigma]$	(25%, 5)	-	11.1	51.19	13.3	2.62	
		(50%, 4)	-	244.7	17.90	144.2	1.76	
		(75%, 2)	-	1693.3	3.79	782.1	0.60	
	$[-4\sigma, 4\sigma]$	(25%, 5)	-	11.1	84.47	11.9	2.72	
		(50%, 4)	-	251.9	26.62	126.4	1.85	
		(75%, 2)	-	1814.3	7.00	950.8	0.63	
	400	$[-2\sigma, 2\sigma]$	(25%, 5)	-	36	28.82	35.7	2.38
			(50%, 4)	-	958.2	15.86	423.5	1.82
			(75%, 2)	-	-	-	-	-
$[-3\sigma, 3\sigma]$		(25%, 5)	-	34.7	58.55	35.4	2.48	
		(50%, 4)	-	1093.7	20.06	525.8	1.74	
		(75%, 2)	-	-	-	-	-	
$[-4\sigma, 4\sigma]$		(25%, 5)	-	38.5	80.60	39.1	2.38	
		(50%, 4)	-	990.7	23.66	430.4	1.55	
		(75%, 2)	-	-	-	-	-	

indicated by [LB,UB], shows that each interval is constructed by which lower and upper bounds. Column “Itv-Time” reports the time spent to construct each interval, which equals to the summation of the computational times needed to find the lower and upper bounds. Column “Itv-Gap”, calculated by  $\frac{UB-LB}{UB} \times 100\%$ , demonstrates how tight the interval [LB, UB] is. Symbol “-” indicates that no optimal solution of the original DRO problem or its approximations can be found within the time limit. In Column “ $(\frac{m_1}{m}\%, P)$ ”,  $\frac{m_1}{m}\%$  is used to define the lower bound problem and the PCA based upper bound problems, i.e., Problems (2.15), (2.24), and (2.60), while  $P$  is used to define the vector splitting based upper bound problems, i.e., Problems (2.37) and (2.72).

Table 2.13: Intervals on the production-transportation problem

		[LB,UB]	[(2.60), (2.72)]	
Size ( $m, n$ )	$(\frac{m_1}{m}\%, P)$	Orig. (secs)	Itv-Time (secs)	Itv-Gap (%)
(8, 30)	(25%, 5)	-	2144.6	7.08
	(50%, 4)	-	4204.3	6.32
	(75%, 2)	-	14043.1	3.14
(8, 40)	(25%, 5)	-	5153.1	7.38
	(50%, 4)	-	10290.3	6.98
	(75%, 2)	-	40784.1	2.73

Table 2.14: Intervals on the newsvendor problem

		[LB, UB]	[(2.60), (2.72)]	
Size ( $n$ )	$(\frac{m_1}{m}\%, P)$	Orig. (secs)	Itv-Time (secs)	Itv-Gap (%)
240	(25%, 5)	-	503.7	3.34
	(50%, 4)	-	891.1	2.44
	(75%, 2)	-	2853.4	0.88
320	(25%, 5)	-	1018.8	3.36
	(50%, 4)	-	2275.7	2.28
	(75%, 2)	-	7946.2	0.77
400	(25%, 5)	-	3078.8	3.24
	(50%, 4)	-	7780.3	2.19
	(75%, 2)	-	-	-

From Tables 2.11 - 2.14, we have the following observations. First, when the optimal solution cannot be found for most original cases, an interval that includes

the unknown optimal value can be constructed fast by using our proposed inner and outer approximations. Such interval is relatively tight for most cases, as demonstrated by the interval [(2.15),(2.37)] in Tables 2.11 - 2.12 and the interval [(2.60), (2.72)] in Tables 2.13 - 2.14. Second, when  $m_1$  increases and  $P$  decreases, such interval can be tighter but it costs more computational time. In practice, thus we can leverage the number of principal components and split pieces as a tool to balance between the interval tightness and computational time. Third, from Tables 2.11 and 2.12, we can observe that the vector splitting based upper bounds due to (2.37) are much tighter than the PCA based upper bounds due to (2.24) and take less computational time.

#### 2.4.2.5 Benefits of Choosing Leading Components

When using the PCA approach to derive our inner and outer approximations, we choose the leading components and relax the rest of ones. To show the benefits of choosing the leading components in guaranteeing solution quality, we take our derived lower bound (2.15) together with its corresponding theoretical error bound (2.16) as an example and compare its performance of solving the newsvendor problem under two cases: (a) choose  $m_1$  arbitrary components; (b) choose  $m_1$  leading components.

If we choose  $m_1$  arbitrary components, then we can follow the same process in Section 2.2.1 to derive the corresponding outer approximation (denoted by (8') here for short) and theoretical error bound (denoted by (9') here for short). Following the table header of Tables 2.1 and 2.2, which show the performance of lower bound (2.15), we introduce subscripts a and b to use "Gap<sub>a</sub>" to denote the relative gap in percentage between the lower bound (8') and the original optimal value, "Gap2<sub>a</sub>" to denote the value of theoretical error bound (9') in percentage, use "Gap<sub>b</sub>" to denote the relative gap in percentage between the lower bound (8) and the original optimal value, and "Gap2<sub>b</sub>" to denote the value of theoretical error bound (2.16) in

percentage. We further use “Det<sub>c</sub>” to denote the relative gap in percentage between “Gap<sub>a</sub>” and “Gap<sub>b</sub>”, i.e.,  $(\text{Gap}_a - \text{Gap}_b) / \text{Gap}_a \times 100\%$ , and use “Det<sub>t</sub>” to denote the relative gap in percentage between “Gap2<sub>a</sub>” and “Gap2<sub>b</sub>”, i.e.,  $(\text{Gap2}_a - \text{Gap2}_b) / \text{Gap2}_a \times 100\%$ . We show the corresponding results in Table 2.15.

Table 2.15: Arbitrary vs. leading components on the newsvendor problem

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

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

#### 2.4.2.6 Sensitivity Analyses

To better analyze the performance of our derived approximations more thoroughly, we conduct sensitivity analyses with respect to parameters  $\gamma_1$ ,  $\gamma_2$ , and  $R_0$ . We take lower bound (2.15) with support  $[-3\sigma, 3\sigma]$  on the newsvendor problem as an example to conduct the sensitivity analyses with respect to  $\gamma_1$  and  $\gamma_2$  and summarize the results in Tables 2.16 and 2.17, respectively. Similarly, we conduct the sensitivity analysis with respect to  $R_0$  for lower bound (2.60) on the newsvendor problem and summarize the results in Table 2.18.

Table 2.16: Sensitivity analysis for lower bound (2.15) on the newsvendor problem with respect to  $\gamma_1$ 

$\frac{m_1}{m}(\%)$		100%			75%			50%			25%			10%			
Size (n)	$\gamma_1$	Orig. (secs)	Time (secs)	Gap (%)	Gap2 (%)	Time (secs)	Gap (%)	Gap2 (%)	Time (secs)	Gap (%)	Gap2 (%)	Time (secs)	Gap (%)	Gap2 (%)	Time (secs)	Gap (%)	Gap2 (%)
100	0.4	13.2	13.6	0.00	0.00	4.1	0.03	0.78	1.18	0.24	2.32	0.38	1.07	4.68	0.3	2.16	6.24
	0.6	13.7	13.5	0.00	0.00	4.0	0.03	0.78	1.1	0.25	2.33	0.4	1.10	4.69	0.3	2.23	6.25
	0.8	12.8	13.0	0.00	0.00	4.3	0.03	0.78	1.1	0.25	2.33	0.4	1.12	4.69	0.3	2.27	6.26
	1	18.8	19.5	0.00	0.00	5.7	0.03	0.78	1.4	0.25	2.33	0.4	1.13	4.70	0.3	2.29	6.26
	1.2	13.9	14.1	0.00	0.00	4.5	0.03	0.78	1.2	0.25	2.33	0.4	1.13	4.70	0.3	2.29	6.26
160	0.4	129.2	130.0	0.00	0.00	34.4	0.02	0.55	6.2	0.24	1.97	0.9	0.94	3.83	0.5	2.00	5.17
	0.6	136.7	138.5	0.00	0.00	34.8	0.02	0.55	6.4	0.25	1.97	0.9	0.97	3.83	0.4	2.07	5.18
	0.8	145.0	143.4	0.00	0.00	33.7	0.02	0.55	6.2	0.25	1.97	0.9	0.99	3.83	0.5	2.11	5.18
	1	163.6	164.5	0.00	0.00	46.4	0.02	0.55	9.1	0.25	1.98	1.0	0.99	3.83	0.4	2.12	5.18
	1.2	132.1	129.0	0.00	0.00	31.7	0.02	0.55	6.1	0.25	1.98	0.8	0.99	3.83	0.4	2.12	5.18
200	0.4	402.2	402.8	0.00	0.00	88.4	0.02	0.53	16.1	0.18	1.67	1.4	0.77	3.30	0.4	1.57	4.40
	0.6	435.1	434.9	0.00	0.00	90.6	0.02	0.53	15.2	0.19	1.67	1.3	0.80	3.30	0.4	1.62	4.40
	0.8	425.3	425.1	0.00	0.00	103.0	0.02	0.53	14.4	0.19	1.67	1.3	0.81	3.31	0.4	1.65	4.41
	1	483.0	483.5	0.00	0.00	113.9	0.02	0.53	21.5	0.19	1.67	1.8	0.81	3.31	0.4	1.65	4.41
	1.2	371.3	371.6	0.00	0.00	100.6	0.02	0.53	15.2	0.19	1.67	1.3	0.81	3.31	0.4	1.65	4.41

Table 2.17: Sensitivity analysis for lower bound (2.15) on the newsvendor problem with respect to  $\gamma_2$ 

$\frac{m_1}{m}(\%)$		100%			75%			50%			25%			10%			
Size (n)	$\gamma_2$	Orig. (secs)	Time (secs)	Gap (%)	Gap2 (%)	Time (secs)	Gap (%)	Gap2 (%)	Time (secs)	Gap (%)	Gap2 (%)	Time (secs)	Gap (%)	Gap2 (%)	Time (secs)	Gap (%)	Gap2 (%)
100	1	12.2	12.0	0.00	0.00	3.9	0.01	0.42	1.1	0.17	1.54	0.4	0.83	3.25	0.3	1.72	4.36
	2	17.8	17.5	0.00	0.00	5.7	0.02	0.60	1.4	0.24	2.21	0.4	1.19	4.66	0.3	2.46	6.26
	3	13.3	13.4	0.00	0.00	4.04	0.02	0.74	1.1	0.29	2.73	0.4	1.45	5.77	0.3	3.01	7.76
	4	14.1	14.1	0.00	0.00	4.44	0.02	0.86	1.2	0.33	3.18	0.4	1.65	6.72	0.3	3.43	9.03
	5	13.8	13.7	0.00	0.00	4.4	0.03	0.97	1.2	0.37	3.58	0.4	1.83	7.57	0.3	3.80	10.17
160	1	114.0	113.3	0.00	0.00	28.8	0.01	0.38	5.4	0.14	1.29	0.7	0.64	2.66	0.3	1.34	3.61
	2	158.0	159.8	0.00	0.00	45.2	0.02	0.54	8.3	0.20	1.86	0.9	0.91	3.82	0.4	1.92	5.18
	3	140.3	140.5	0.00	0.00	32.7	0.02	0.67	6.0	0.24	2.29	0.8	1.11	4.72	0.4	2.34	6.40
	4	152.7	152.8	0.00	0.00	33.5	0.02	0.78	6.6	0.27	2.67	0.7	1.27	5.49	0.3	2.67	7.45
	5	142.2	142.8	0.00	0.00	38.8	0.03	0.88	7.4	0.30	3.00	0.7	1.40	6.18	0.4	2.95	8.38
200	1	396.4	398.2	0.00	0.00	103.2	0.01	0.36	15.5	0.11	1.06	1.4	0.52	2.26	0.4	1.12	3.09
	2	473.8	473.3	0.00	0.00	125.4	0.02	0.52	22.1	0.15	1.52	1.9	0.74	3.23	0.5	1.60	4.42
	3	452.03	451.8	0.00	0.00	86.7	0.02	0.64	19.1	0.19	1.87	1.3	0.91	3.99	0.4	1.95	5.45
	4	429.3	430.0	0.00	0.00	106.6	0.02	0.74	15.9	0.21	2.18	1.4	1.03	4.63	0.4	2.22	6.34
	5	466.9	465.9	0.00	0.00	98.6	0.03	0.84	14.8	0.23	2.45	1.4	1.14	5.20	0.4	2.45	7.12

When parameter  $\gamma_1$  increases, the values of Gap and Gap2 increase very slightly, as shown in Table 2.16. In contrast, when parameter  $\gamma_2$  increases, the values of Gap and Gap2 also increase, though at higher rates. It implies that the value of  $\gamma_2$  and accordingly the second-order moment constraint are more important than the value of  $\gamma_1$  and the first-order moment constraint, respectively, in terms of affecting the computational performance of solving the corresponding DRO problems. It also further supports our focus on approximating the covariance matrix in an ambiguity

Table 2.18: Sensitivity analysis for lower bound (2.60) on the newsvendor problem with respect to  $R_0$

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

set through either the PCA approach or the vector splitting approach in this chapter. In addition, we can observe that an increase in  $R_0$  will also lead to slightly increasing Gap and Gap2, as shown in Table 2.18. Meanwhile, the values of computational time are close for the changes of values in  $\gamma_1$ ,  $\gamma_2$ , and  $R_0$ , respectively.

## 2.5 Discussions

In this chapter, we proposed computationally efficient inner and outer approximations for DRO problems with two types of ambiguity sets: the moment-based ambiguity set and combined ambiguity set. We approximated the original DRO problems mainly through two approaches: (i) use PCA to shrink the dimensionality of the uncertainty space; and (ii) split the random parameter vector into smaller pieces, both of which lead to smaller PSD matrix constraints. Furthermore, we derived theoretical bounds on the gap between the optimal values of DRO problems and their approximations. Such bounds help determine the required numbers of split pieces and principal components to reach a predetermined error bound. Our proposed approximations enable decision-makers to better balance the trade-off between solution quality and computational time by leveraging the appropriate numbers of split pieces and principal components. Meanwhile, the inner and outer approximations

together enable us to construct an interval that contains the (unknown) optimal solution for a large-scale DRO problem, which cannot be solved to optimality (or even feasibility in most cases) by existing methods in a reasonable time. Such interval is tight for most cases, as demonstrated by our numerical experiments. Finally, we demonstrated the significant effectiveness of the proposed approximations in solving the distributionally robust production-transportation and multi-product newsvendor problems. The results showed that our approximations significantly reduce the computational time while maintaining high solution quality, with the strengths of our derived theoretical bounds well justified.

## 2.6 Supplement to Section 2.3: Further Including the First-Order Moment Information

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

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

where

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

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

**Proposition 2.11.** *Under Assumption 2.2, (DRO-C2) can be recast as the following*

SDP formulation:

$$Z_{C2}^*(m) := \min_{\substack{\mathbf{x}, \lambda, \mathbf{Q} \\ \hat{\boldsymbol{\zeta}}, \hat{\mathbf{y}}, \mathbf{w}}} \lambda R_0 + \gamma_2 \boldsymbol{\Sigma} \bullet \mathbf{Q} + 2\sqrt{\gamma_1 \mathbf{w}^\top \boldsymbol{\Sigma} \mathbf{w}} + \frac{1}{N} \sum_{i=1}^N y_i \quad (2.85a)$$

$$\text{s.t.} \quad \begin{bmatrix} \mathbf{Q} & \frac{1}{2}(-y_k(\mathbf{x}) + \boldsymbol{\zeta}^i - 2\mathbf{Q}\boldsymbol{\mu} - 2\mathbf{w}) \\ \frac{1}{2}(-y_k(\mathbf{x}) + \boldsymbol{\zeta}^i - 2\mathbf{Q}\boldsymbol{\mu} - 2\mathbf{w})^\top & y_i - y_k^0(\mathbf{x}) - \boldsymbol{\zeta}^{i\top} \hat{\boldsymbol{\xi}}^i + \boldsymbol{\mu}^\top \mathbf{Q}\boldsymbol{\mu} + 2\mathbf{w}^\top \boldsymbol{\mu} \end{bmatrix} \succeq 0, \quad (2.85b)$$

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

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

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

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

$$\min_{\mathbf{x} \in \mathcal{X}} \max_{\mathbb{P}, \pi} \int_{\mathbb{R}^m} f(\mathbf{x}, \boldsymbol{\xi}) d\mathbb{P}(\boldsymbol{\xi}) \quad (2.86a)$$

$$\text{s.t.} \quad \int_{\mathbb{R}^m} \begin{bmatrix} \boldsymbol{\Sigma} & (\boldsymbol{\xi} - \boldsymbol{\mu}) \\ (\boldsymbol{\xi} - \boldsymbol{\mu})^\top & \gamma_1 \end{bmatrix} d\mathbb{P}(\boldsymbol{\xi}) \succeq 0, \quad (2.86b)$$

$$\int_{\mathbb{R}^m} (\boldsymbol{\xi} - \boldsymbol{\mu})(\boldsymbol{\xi} - \boldsymbol{\mu})^\top d\mathbb{P}(\boldsymbol{\xi}) \preceq \gamma_2 \boldsymbol{\Sigma}, \quad (2.86c)$$

$$\int_{(\mathbb{R}^m)^2} \|\boldsymbol{\xi} - \hat{\boldsymbol{\xi}}\|_1 \pi(\boldsymbol{\xi}, \hat{\boldsymbol{\xi}}) \leq R_0, \quad (2.86d)$$

where  $\mathbb{R}^m = \mathcal{S}$  and (2.86b) is derived due to Schur's complement.

We let  $\begin{bmatrix} \mathbf{W} & \mathbf{w} \\ \mathbf{w}^\top & r \end{bmatrix} \succeq 0$ ,  $\mathbf{Q} \succeq 0$ , and  $\lambda \in \mathbb{R}_+$  denote the Lagrangian multipliers of constraints (2.86b), (2.86c), and (2.86d), respectively, and thus derive the Lagrangian dual problem of (2.86) as follows:

$$\min_{\substack{\mathbf{x}, \mathbf{W}, \mathbf{w} \\ r, \mathbf{Q}, \lambda}} \lambda R_0 + \gamma_2 \boldsymbol{\Sigma} \bullet \mathbf{Q} + \boldsymbol{\Sigma} \bullet \mathbf{W} + \gamma_1 r + \int_{\mathbb{R}^m} \sup_{\boldsymbol{\xi}} g(\boldsymbol{\xi}, \hat{\boldsymbol{\xi}}) \mathbb{P}_0(d\hat{\boldsymbol{\xi}}) \quad (2.87a)$$

$$\text{s.t.} \quad \lambda \in \mathbb{R}_+, \mathbf{x} \in \mathcal{X}, \mathbf{Q} \succeq 0,$$

$$\begin{bmatrix} \mathbf{W} & \mathbf{w} \\ \mathbf{w}^\top & r \end{bmatrix} \succeq 0, \quad (2.87b)$$

where  $g(\boldsymbol{\xi}, \hat{\boldsymbol{\xi}}) = f(\mathbf{x}, \boldsymbol{\xi}) + 2\mathbf{w}^\top (\boldsymbol{\xi} - \boldsymbol{\mu}) - (\boldsymbol{\xi} - \boldsymbol{\mu})^\top \mathbf{Q} (\boldsymbol{\xi} - \boldsymbol{\mu}) - \lambda \|\boldsymbol{\xi} - \hat{\boldsymbol{\xi}}\|_1$ . By following the similar steps in the proof of Lemma 1 in Gao and Kleywegt (2017), we can prove that

here the strong duality holds for Problem (2.87). We then further simplify (2.87) towards eliminating variables  $\mathbf{W}$  and  $r$  in the outer minimization problem. To that end, we keep variables  $\mathbf{x}$ ,  $\mathbf{Q}$ ,  $\lambda$ , and  $\mathbf{w}$  fixed while solving Problem (2.87) analytically for variables  $\mathbf{W}$  and  $r$ . It follows that we solve  $\min_{\mathbf{W}, r} \Sigma \bullet \mathbf{W} + \gamma_1 r$  analytically for  $\mathbf{W}$  and  $r$ . We consider two cases for the optimal solution of  $r$  (denoted by  $r^*$ ) due to constraint (2.87b), i.e.,  $r^* > 0$  and  $r^* = 0$ , as follows.

- If  $r^* > 0$ , then constraint (2.87b) can be reformulated as  $\mathbf{W} \succeq \frac{\mathbf{w}\mathbf{w}^\top}{r}$  by Schur's complement. As a result,  $\mathbf{W}^* = \frac{\mathbf{w}\mathbf{w}^\top}{r}$  is a valid optimal solution because  $\min_{\mathbf{W}, r} \Sigma \bullet \mathbf{W} + \gamma_1 r$  is a minimization problem. Replacing  $\mathbf{W}^*$  by  $\frac{\mathbf{w}\mathbf{w}^\top}{r}$  leads to solve a one-dimensional convex optimization problem, i.e.,  $\min_{r>0} \left(\frac{1}{r}\right) \mathbf{w}^\top \Sigma \mathbf{w} + \gamma_1 r$ . By applying the necessary first-order optimality condition to this problem, i.e., setting the derivative of the objective function over  $r$  to zero, we have  $r^* = \sqrt{\left(\frac{1}{\gamma_1}\right) \mathbf{w}^\top \Sigma \mathbf{w}}$  as the optimal solution of  $r$ , resulting in  $\min_{\mathbf{W}, r} \Sigma \bullet \mathbf{W} + \gamma_1 r = 2\sqrt{\gamma_1 \mathbf{w}^\top \Sigma \mathbf{w}}$ .
- If  $r^* = 0$ , then we let  $\mathbf{w}^*$  denote the optimal solution of  $\mathbf{w}$  and we must have  $\mathbf{w}^* = \mathbf{0}$ . Otherwise, we have  $\mathbf{w}^{*\top} \mathbf{w}^* > 0$ , and by defining  $\mathbf{Z} = (\mathbf{w}^{*\top}, \eta)^\top$  with  $\eta < \frac{-\mathbf{w}^{*\top} \mathbf{W}^* \mathbf{w}^*}{2\mathbf{w}^{*\top} \mathbf{w}^*}$ , we further have

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

which contradicts constraint (2.87b). Considering  $r^* = 0$  and  $\mathbf{w}^* = \mathbf{0}$ ,  $\min_{\mathbf{x}, \mathbf{W}, \mathbf{w}, r} \Sigma \bullet \mathbf{W} + \gamma_1 r$  reduces to  $\min_{\mathbf{W}} \Sigma \bullet \mathbf{W}$  whose optimal solution is clearly  $\mathbf{W}^* = \mathbf{0}$  as it is a minimization problem. As a result,  $\min_{\mathbf{W}, r} \Sigma \bullet \mathbf{W} + \gamma_1 r = 2\sqrt{\gamma_1 \mathbf{w}^\top \Sigma \mathbf{w}}$ .

Therefore, Problem (2.87) can be recast as follows:

$$\begin{aligned} \min_{\substack{\mathbf{x}, \mathbf{w} \\ \mathbf{Q}, \lambda}} \quad & \lambda R_0 + \gamma_2 \Sigma \bullet \mathbf{Q} + 2\sqrt{\gamma_1 \mathbf{w}^\top \Sigma \mathbf{w}} + \int_{\mathbb{R}^m} \sup_{\xi} g(\xi, \hat{\xi}) \mathbb{P}_0(d\hat{\xi}) \quad (2.88) \\ \text{s.t.} \quad & \lambda \in \mathbb{R}_+, \quad \mathbf{x} \in \mathcal{X}, \quad \mathbf{Q} \succeq 0. \end{aligned}$$

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

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

As  $\mathbb{P}_0$  denotes an empirical distribution of  $\boldsymbol{\xi}$  generated by i.i.d. samples  $\{\hat{\boldsymbol{\xi}}^i : i \in [N]\} \subseteq \mathcal{S}$  from the  $\mathbb{P}$ , i.e.,  $\mathbb{P}\{\boldsymbol{\xi} = \hat{\boldsymbol{\xi}}^i\} = \frac{1}{N}$  for any  $i \in [N]$ , we have

$$\int_{\mathbb{R}^m} \sup_{\boldsymbol{\xi}} g(\boldsymbol{\xi}, \hat{\boldsymbol{\xi}}) \mathbb{P}_0(d\hat{\boldsymbol{\xi}}) = \frac{1}{N} \sum_{i=1}^N \sup_{\boldsymbol{\xi}} g(\boldsymbol{\xi}, \hat{\boldsymbol{\xi}}^i). \quad (2.89)$$

Thus, by plugging (2.89) into (2.88), (2.88) can be reformulated as

$$\min_{\substack{\mathbf{x}, \mathbf{w} \\ \mathbf{Q}, \lambda}} \left\{ \lambda R_0 + \gamma_2 \boldsymbol{\Sigma} \bullet \mathbf{Q} + 2\sqrt{\gamma_1 \mathbf{w}^\top \boldsymbol{\Sigma} \mathbf{w}} + \frac{1}{N} \sum_{i=1}^N y_i \right\} \quad (2.90a)$$

$$\begin{aligned} \text{s.t. } & y_i = \sup_{\boldsymbol{\xi}} g(\boldsymbol{\xi}, \hat{\boldsymbol{\xi}}^i), \quad \forall i \in [N], \\ & \lambda \in \mathbb{R}_+, \quad \mathbf{x} \in \mathcal{X}, \quad \mathbf{Q} \succeq 0. \end{aligned} \quad (2.90b)$$

Since Problem (2.90) is a minimization problem, constraints (2.90b) can be relaxed to  $y_i \geq \sup_{\boldsymbol{\xi}} g(\boldsymbol{\xi}, \hat{\boldsymbol{\xi}}^i)$  for any  $i \in [N]$ . Thus, we have

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

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

$$\begin{aligned} y_i &\geq \sup_{\boldsymbol{\xi}} \inf_{\|\hat{\boldsymbol{\zeta}}\|_* \leq 1} \left\{ y_k^0(\mathbf{x}) + y_k(\mathbf{x})^\top \boldsymbol{\xi} + 2\mathbf{w}^\top (\boldsymbol{\xi} - \boldsymbol{\mu}) - (\boldsymbol{\xi} - \boldsymbol{\mu})^\top \mathbf{Q} (\boldsymbol{\xi} - \boldsymbol{\mu}) - \lambda \hat{\boldsymbol{\zeta}}^\top (\boldsymbol{\xi} - \hat{\boldsymbol{\xi}}^i) \right\}, \quad \forall k \in [K] \\ \Leftrightarrow y_i &\geq \inf_{\|\hat{\boldsymbol{\zeta}}\|_* \leq 1} \sup_{\boldsymbol{\xi}} \left\{ y_k^0(\mathbf{x}) + y_k(\mathbf{x})^\top \boldsymbol{\xi} + 2\mathbf{w}^\top (\boldsymbol{\xi} - \boldsymbol{\mu}) - (\boldsymbol{\xi} - \boldsymbol{\mu})^\top \mathbf{Q} (\boldsymbol{\xi} - \boldsymbol{\mu}) - \lambda \hat{\boldsymbol{\zeta}}^\top (\boldsymbol{\xi} - \hat{\boldsymbol{\xi}}^i) \right\}, \quad \forall k \in [K] \\ \Leftrightarrow \exists \hat{\boldsymbol{\zeta}} \text{ s.t. } & \|\hat{\boldsymbol{\zeta}}\|_* \leq 1, y_i \geq \sup_{\boldsymbol{\xi}} \left\{ y_k^0(\mathbf{x}) + y_k(\mathbf{x})^\top \boldsymbol{\xi} + 2\mathbf{w}^\top (\boldsymbol{\xi} - \boldsymbol{\mu}) - (\boldsymbol{\xi} - \boldsymbol{\mu})^\top \mathbf{Q} (\boldsymbol{\xi} - \boldsymbol{\mu}) - \lambda \hat{\boldsymbol{\zeta}}^\top (\boldsymbol{\xi} - \hat{\boldsymbol{\xi}}^i) \right\}, \\ & \forall k \in [K] \\ \Leftrightarrow \exists \hat{\boldsymbol{\zeta}} \text{ s.t. } & \|\hat{\boldsymbol{\zeta}}\|_* \leq 1, y_i \geq y_k^0(\mathbf{x}) + y_k(\mathbf{x})^\top \boldsymbol{\xi} + 2\mathbf{w}^\top (\boldsymbol{\xi} - \boldsymbol{\mu}) - (\boldsymbol{\xi} - \boldsymbol{\mu})^\top \mathbf{Q} (\boldsymbol{\xi} - \boldsymbol{\mu}) - \lambda \hat{\boldsymbol{\zeta}}^\top (\boldsymbol{\xi} - \hat{\boldsymbol{\xi}}^i), \\ & \forall \boldsymbol{\xi} \in \mathbb{R}^m, \forall k \in [K] \\ \Leftrightarrow \exists \hat{\boldsymbol{\zeta}} \text{ s.t. } & \|\hat{\boldsymbol{\zeta}}\|_* \leq 1, \left[ \begin{array}{c} \mathbf{Q} \\ \frac{1}{2}(-y_k(\mathbf{x}) + \lambda \hat{\boldsymbol{\zeta}} - 2\mathbf{Q}\boldsymbol{\mu} - 2\mathbf{w}) \\ y_i - y_k^0(\mathbf{x}) - \lambda \hat{\boldsymbol{\zeta}}^\top \hat{\boldsymbol{\xi}}^i + \boldsymbol{\mu}^\top \mathbf{Q}\boldsymbol{\mu} + 2\mathbf{w}^\top \boldsymbol{\mu} \end{array} \right] \succeq 0, \quad \forall k \in [K], \end{aligned} \quad (2.91)$$

where the first equivalence is due to the convexity of  $g(\boldsymbol{\xi}, \hat{\boldsymbol{\xi}})$ ,  $\mathcal{S}$ , and the feasible region defined by  $\|\hat{\boldsymbol{\zeta}}\|_* \leq 1$ . For any given  $i \in [N]$ , we replace  $\lambda \hat{\boldsymbol{\zeta}}$  by  $\boldsymbol{\zeta}^i$ , and then we can obtain Problem (2.85) by further replacing (2.90b) by (2.91) for any  $i \in [N]$ .  $\square$

In the following Sections 2.6.1 and 2.6.2, we derive computationally efficient outer and inner approximations, leading to lower and upper bounds, for Problem (2.85), respectively. We first reformulate (DRO-C2) as the following problem by performing the eigenvalue decomposition on matrix  $\boldsymbol{\Sigma}$ :

$$\min_{\boldsymbol{x} \in \mathcal{X}} \max_{\mathbb{P}_I \in \mathcal{D}_{C6}} \mathbb{E}_{\mathbb{P}_I} \left[ f \left( \boldsymbol{x}, \boldsymbol{U} \boldsymbol{\Lambda}^{\frac{1}{2}} \boldsymbol{\xi}_I + \boldsymbol{\mu} \right) \right], \quad (2.92)$$

where

$$\mathcal{D}_{C6}(\mathcal{S}_I, \boldsymbol{\mu}, \gamma_1, \gamma_2, \mathbb{P}_0, R_0) = \left\{ \mathbb{P}_I \left| \begin{array}{l} \mathbb{E}_{\mathbb{P}_I}[\boldsymbol{\xi}_I^\top] \mathbb{E}_{\mathbb{P}_I}[\boldsymbol{\xi}_I] \leq \gamma_1 \\ \mathbb{E}_{\mathbb{P}_I}[\boldsymbol{\xi}_I \boldsymbol{\xi}_I^\top] \preceq \gamma_2 \boldsymbol{I}_m \\ \exists \pi : \int_{\mathcal{S}^2} \left\| \boldsymbol{U} \boldsymbol{\Lambda}^{\frac{1}{2}} \boldsymbol{\xi}_I + \boldsymbol{\mu} - \hat{\boldsymbol{\xi}} \right\|_1 \pi \left( \boldsymbol{U} \boldsymbol{\Lambda}^{\frac{1}{2}} \boldsymbol{\xi}_I + \boldsymbol{\mu}, \hat{\boldsymbol{\xi}} \right) \leq R_0 \end{array} \right. \right\}$$

under the condition that  $f(\boldsymbol{x}, \boldsymbol{U} \boldsymbol{\Lambda}^{\frac{1}{2}} \boldsymbol{\xi}_I + \boldsymbol{\mu})$  is  $\mathbb{P}_I$ -integrable for any  $\mathbb{P}_I \in \mathcal{D}_{C6}$  and  $\mathcal{S}_I := \{\boldsymbol{\xi}_I \in \mathbb{R}^m : \boldsymbol{U} \boldsymbol{\Lambda}^{\frac{1}{2}} \boldsymbol{\xi}_I + \boldsymbol{\mu} \in \mathcal{S}\}$ .

### 2.6.1 Lower Bound

By the approximation of  $\boldsymbol{\xi}$  in (2.9) due to PCA, we outer approximate (2.92) as the following problem:

$$\min_{\boldsymbol{x} \in \mathcal{X}} \max_{\mathbb{P}_r \in \mathcal{D}_{C7}} \mathbb{E}_{\mathbb{P}_r} \left[ f \left( \boldsymbol{x}, \boldsymbol{U}_{m \times m_1} \boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}} \boldsymbol{\xi}_r + \boldsymbol{\mu} \right) \right], \quad (2.93)$$

where

$$\mathcal{D}_{C7}(\mathcal{S}_r, \boldsymbol{\mu}, \gamma_1, \gamma_2, \mathbb{P}_0, R_0) = \left\{ \mathbb{P}_r \left| \begin{array}{l} \mathbb{E}_{\mathbb{P}_r}[\boldsymbol{\xi}_r^\top] \mathbb{E}_{\mathbb{P}_r}[\boldsymbol{\xi}_r] \leq \gamma_1 \\ \mathbb{E}_{\mathbb{P}_r}[\boldsymbol{\xi}_r \boldsymbol{\xi}_r^\top] \preceq \gamma_2 \boldsymbol{I}_{m_1} \\ \exists \pi : \int_{\mathcal{S}^2} \left\| \boldsymbol{U}_{m \times m_1} \boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}} \boldsymbol{\xi}_r + \boldsymbol{\mu} - \hat{\boldsymbol{\xi}} \right\|_1 \pi \left( \boldsymbol{U}_{m \times m_1} \boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}} \boldsymbol{\xi}_r + \boldsymbol{\mu}, \hat{\boldsymbol{\xi}} \right) \leq R_0 \end{array} \right. \right\}$$

with  $\mathcal{S}_r := \{\boldsymbol{\xi}_r \in \mathbb{R}^{m_1} : \boldsymbol{U}_{m \times m_1} \boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}} \boldsymbol{\xi}_r + \boldsymbol{\mu} \in \mathcal{S}\}$ .

**Theorem 2.7.** *Under Assumption 2.2, Problem (2.93) has the same optimal value as the following SDP formulation:*

$$Z_{C2}^*(m_1) := \min_{\substack{\mathbf{x}, \lambda, \mathbf{q}_r \\ \mathbf{Q}_r, \hat{\boldsymbol{\zeta}}, \hat{\mathbf{y}}}} \lambda R_0 + \gamma_2 \mathbf{I}_{m_1} \bullet \mathbf{Q}_r + \sqrt{\gamma_1} \|\mathbf{q}_r\|_2 + \frac{1}{N} \sum_{i=1}^N y_i \quad (2.95a)$$

$$\text{s.t.} \quad \begin{bmatrix} \mathbf{Q}_r & \frac{1}{2} \left( \mathbf{q}_r + (-y_k(\mathbf{x}) + \boldsymbol{\zeta}^i)^\top \mathbf{U}_{m \times m_1} \boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}} \right)^\top \\ \frac{1}{2} \left( \mathbf{q}_r + (-y_k(\mathbf{x}) + \boldsymbol{\zeta}^i)^\top \mathbf{U}_{m \times m_1} \boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}} \right) & y_i - y_k(\mathbf{x})^\top \boldsymbol{\mu} - y_k^0(\mathbf{x}) + (\boldsymbol{\mu} - \hat{\boldsymbol{\zeta}}^i)^\top \boldsymbol{\zeta}^i \end{bmatrix} \succeq 0, \quad \forall i \in [N], \forall k \in [K], \quad (2.95b)$$

$$\lambda \in \mathbb{R}_+, \mathbf{x} \in \mathcal{X}, \|\boldsymbol{\zeta}^i\|_* \leq \lambda, \forall i \in [N], \quad (2.95c)$$

where  $\mathbf{q}_r \in \mathbb{R}^{m_1}$ ,  $\mathbf{Q}_r \in \mathbb{R}^{m_1 \times m_1}$ ,  $\boldsymbol{\zeta}^i \in \mathbb{R}^m$ ,  $\hat{\boldsymbol{\zeta}} = \{\boldsymbol{\zeta}^1, \dots, \boldsymbol{\zeta}^N\}$ , and  $\hat{\mathbf{y}} = \{y_1, \dots, y_N\}$ . Furthermore, (i) Problem (2.95) provides a lower bound for the optimal value of (DRO-C2); (ii) the optimal value of Problem (2.95) is nondecreasing in  $m_1$ ; and (iii) if  $m_1 = m$ , then (DRO-C2) and (2.95) have the same optimal value.

*Proof.* The reformulation proof is similar to that in Proposition 2.11, while the proofs for claims (i), (ii), and (iii) are similar to those in Theorems 2.2 and 2.5. Thus we omit them here for brevity.  $\square$

## 2.6.2 Upper Bound

We approximate Problem (2.85) by splitting  $\boldsymbol{\xi}_I$  in  $\mathcal{D}_{C6}$  into  $P$  pieces so that  $\boldsymbol{\xi}_I = (\boldsymbol{\xi}_{I_1}^\top, \boldsymbol{\xi}_{I_2}^\top, \dots, \boldsymbol{\xi}_{I_P}^\top)^\top$ , where  $\boldsymbol{\xi}_{I_j} \in \mathbb{R}^{m_j}$ ,  $\forall j \in [P]$ , and  $\sum_{j=1}^P m_j = m$ . This gives rise to an inner approximation

$$\min_{\mathbf{x} \in \mathcal{X}} \max_{\mathbb{P}_I \in \mathcal{D}_{C8}} \mathbb{E}_{\mathbb{P}_I} \left[ f \left( \mathbf{x}, \mathbf{U} \boldsymbol{\Lambda}^{\frac{1}{2}} \boldsymbol{\xi}_I + \boldsymbol{\mu} \right) \right], \quad (2.96)$$

where

$$\mathcal{D}_{C8}(\mathcal{S}_I, \boldsymbol{\mu}, \gamma_1, \gamma_2, \mathbb{P}_0, R_0) = \left\{ \mathbb{P}_I \left| \begin{array}{l} \mathbb{E}_{\mathbb{P}_I} [\boldsymbol{\xi}_I^\top] \mathbb{E}_{\mathbb{P}_I} [\boldsymbol{\xi}_I] \leq \gamma_1 \\ \mathbb{E}_{\mathbb{P}_I} [\boldsymbol{\xi}_{I_j} \boldsymbol{\xi}_{I_j}^\top] \preceq \gamma_2 \mathbf{I}_{m_j}, \forall j \in [P] \\ \exists \pi : \int_{\mathcal{S}_2} \left\| \mathbf{U} \boldsymbol{\Lambda}^{\frac{1}{2}} \boldsymbol{\xi}_I + \boldsymbol{\mu} - \hat{\boldsymbol{\xi}} \right\|_1 \pi \left( \mathbf{U} \boldsymbol{\Lambda}^{\frac{1}{2}} \boldsymbol{\xi}_I + \boldsymbol{\mu}, \hat{\boldsymbol{\xi}} \right) \leq R_0 \end{array} \right. \right\}.$$

**Theorem 2.8.** *Under Assumption 2.2, Problem (2.96) has the same optimal value as the following SDP formulation:*

$$\begin{aligned}
UB_{C2}^* := & \min_{\substack{\mathbf{x}, \lambda, \mathbf{q} \\ \hat{\mathbf{Q}}, \hat{\boldsymbol{\zeta}}, \hat{\mathbf{y}}, \hat{\mathbf{s}}}} \lambda R_0 + \gamma_2 \sum_{j=1}^P \mathbf{I}_{m_j} \bullet \mathbf{Q}_j + \sqrt{\gamma_1} \|\mathbf{q}\|_2 + \frac{1}{N} \sum_{i=1}^N y_i & (2.97) \\
\text{s.t.} & \left[ \begin{array}{cc} \mathbf{Q}_j & \frac{1}{2} \left( \mathbf{q}_j + (-y_k(\mathbf{x}) + \boldsymbol{\zeta}^i)^\top \mathbf{U}_{m \times m_j} \boldsymbol{\Lambda}_{m_j}^{\frac{1}{2}} \right)^\top \\ \frac{1}{2} \left( \mathbf{q}_j + (-y_k(\mathbf{x}) + \boldsymbol{\zeta}^i)^\top \mathbf{U}_{m \times m_j} \boldsymbol{\Lambda}_{m_j}^{\frac{1}{2}} \right) & s_{jik} \end{array} \right] \succeq 0, \\
& \forall j \in [P], \forall i \in [N], \forall k \in [K], \\
& \sum_{j=1}^P s_{jik} = y_i - y_k(\mathbf{x})^\top \boldsymbol{\mu} - y_k^0(\mathbf{x}) + \left( \boldsymbol{\mu} - \hat{\boldsymbol{\xi}}^i \right)^\top \boldsymbol{\zeta}^i, \forall i \in [N], \forall k \in [K], \\
& \lambda \in \mathbb{R}_+, \mathbf{x} \in \mathcal{X}, \|\boldsymbol{\zeta}^i\|_* \leq \lambda, \forall i \in [N],
\end{aligned}$$

where  $\mathbf{q} = (\mathbf{q}_1^\top \in \mathbb{R}^{m_1}, \dots, \mathbf{q}_P^\top \in \mathbb{R}^{m_P})^\top$ ,  $\mathbf{Q}_j \in \mathbb{R}^{m_j \times m_j}$  for any  $j \in [P]$ ,  $\hat{\mathbf{Q}} = \{\mathbf{Q}_1, \dots, \mathbf{Q}_P\}$ ,  $\boldsymbol{\zeta}^i \in \mathbb{R}^m$  for any  $i \in [N]$ ,  $\hat{\boldsymbol{\zeta}} = \{\boldsymbol{\zeta}^1, \dots, \boldsymbol{\zeta}^N\}$ ,  $\hat{\mathbf{y}} = \{y_1, \dots, y_N\}$ , and  $\hat{\mathbf{s}} = \{s_{ijk}, \forall i \in [N], \forall j \in [P], \forall k \in [K]\}$ . Furthermore, Problem (2.97) provides an upper bound for the optimal value of (DRO-C2).

*Proof.* The reformulation proof is similar to that in Proposition 2.11 and Theorem 2.6, while the proof of the claim that Problem (2.97) provides an upper bound for Problem (2.85) is the same as that of Theorem 2.4. Thus we omit them here for brevity.  $\square$

### 2.6.2.1 Computational Experiments

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



Table 2.19: Lower bound (2.95) on the production-transportation problem

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

Table 2.20: Upper bound (2.97) on the production-transportation problem

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

(DRO-C2), as reported in Tables 2.21 and 2.22. In Tables 2.21 and 2.22, the columns “Orig1” and “Orig2” represent the optimal values of Problems (2.57) (i.e., original Problem (DRO-C)) and (2.85) (i.e., original Problem (DRO-C2)), respectively. The column “Obj1” (resp. “Obj2”) represents the objective value of the approximation of Problem (DRO-C) (resp. Problem (DRO-C2)). From Tables 2.21 and 2.22, we can observe that by including the first-order moment information in the combined ambiguity set, the conservatism of the optimal solution can be reduced (leading to a smaller objective value), though very slightly.

Table 2.21: Lower bounds (2.60) and (2.95) on the production-transportation problem

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

Third, we perform sensitivity analyses with respect to parameters  $\gamma_1$  and  $R_0$ , where we consider  $(m, n) = (5, 20)$ . The results are reported in Tables 2.23 and 2.24. When either  $\gamma_1$  or  $R_0$  increases, both the computational gap induced by lower bound (2.95) increases slightly, while the computational time change is not

Table 2.22: Upper bounds (2.72) and (2.97) on the production-transportation problem

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

significant. From Table 2.24 we can observe that an increase in  $R_0$  slightly increases the computational gap, while an increase in  $\gamma_1$  slightly decreases it.

Table 2.23: Sensitivity analysis for lower bound (2.95) with respect to  $(\gamma_1, R_0)$ 

$\frac{m_1}{m}$ (%)	100%			75%		50%		25%		10%	
$(\gamma_1, R_0)$	Orig. (secs)	Time (secs)	Gap (%)	Time (secs)	Gap (%)	Time (secs)	Gap (%)	Time (secs)	Gap (%)	Time (secs)	Gap (%)
(0.4,30)	802.7	759.1	0.00	287.7	0.87	100.7	3.07	39.7	3.87	19.1	3.96
(0.6,30)	776.8	760.8	0.00	295.6	1.02	113.1	3.62	40.3	4.57	18.3	4.68
(0.8,30)	861.0	845.5	0.00	282.8	1.15	109.9	4.07	39.5	5.15	17.0	5.27
(0.4,40)	842.7	728.9	0.00	273.7	0.97	101.8	3.64	38.3	4.47	19.53	4.72
(0.6,40)	769.2	728.9	0.00	281.4	1.25	94.3	4.03	37.5	5.05	17.0	5.26
(0.8,40)	799.9	731.9	0.00	282.8	1.40	94.5	4.53	37.1	5.78	17.2	6.26
(0.4,50)	760.0	703.6	0.00	276.6	1.18	98.6	4.05	40.08	4.93	17.9	5.11
(0.6,50)	725.3	731.0	0.00	277.2	1.39	101.2	4.52	38.7	5.47	17.7	5.94
(0.8,50)	747.7	703.4	0.00	286.4	1.57	105.1	5.12	38.2	6.43	17.1	7.63

Table 2.24: Sensitivity analysis for upper bound (2.95) with respect to  $(\gamma_1, R_0)$ 

$P$		2		4		5	
$(\gamma_1, R_0)$	Orig. (secs)	Time (secs)	Gap (%)	Time (secs)	Gap (%)	Time (secs)	Gap (%)
(0.4,30)	823.3	202.8	0.25	127.2	0.74	119.4	0.77
(0.6,30)	852.6	205.4	0.24	126.2	0.70	117.4	0.74
(0.8,30)	712.3	189.1	0.23	128.1	0.68	120.7	0.72
(0.4,40)	795.8	199.4	0.26	133.1	0.75	121.9	0.77
(0.6,40)	827.6	202.6	0.25	135.3	0.71	126.4	0.75
(0.8,40)	775.5	200.7	0.23	134.4	0.69	122.6	0.73
(0.4,50)	747.0	198.0	0.26	130.4	0.76	124.7	0.78
(0.6,50)	749.2	206.3	0.26	130.3	0.71	125.0	0.75
(0.8,50)	773.5	205.4	0.24	131.7	0.70	124.9	0.74

## 2.7 Supplement to Section 2.4: Applying the Proposed Approximations in the Context of Production-Transportation Problem

First, the outer approximation (2.15) leads to the following problem:

$$\min_{\substack{\mathbf{x}, \mathbf{z}, s, \\ \lambda, \mathbf{q}_r, \mathbf{Q}_r}} \mathbf{c}^\top \mathbf{x} + s + \gamma_2 \mathbf{I}_{m_1} \bullet \mathbf{Q}_r + \sqrt{\gamma_1} \|\mathbf{q}_r\|_2 \quad (2.98a)$$

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

$$\boldsymbol{\lambda}_k \in \mathbb{R}_+^n, \quad \forall k \in [K], \quad (2.82d),$$

$$\sum_{i=1}^m z_{ijk} = d_j, \quad \forall j \in [n], \quad \forall k \in [K], \quad (2.98b)$$

$$\sum_{j=1}^n z_{ijk} = x_i, \quad \forall i \in [m], \quad \forall k \in [K], \quad (2.98c)$$

$$z_{ijk} \geq 0, \quad \forall i \in [m], \quad \forall j \in [n], \quad \forall k \in [K], \quad (2.98d)$$

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

Second, the outer approximation (2.60) leads to the following problem:

$$\min_{\substack{\mathbf{x}, \mathbf{z}, \lambda, \mathbf{Q}_r, \hat{\boldsymbol{\zeta}}, \hat{\boldsymbol{\gamma}}} \mathbf{c}^\top \mathbf{x} + \lambda R_0 + \gamma_2 \mathbf{I}_{m_1} \bullet \mathbf{Q}_r + \frac{1}{N} \sum_{i=1}^N y_i \quad (2.99)$$

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

$$\lambda \in \mathbb{R}_+, \quad \|\boldsymbol{\zeta}^i\|_* \leq \lambda, \quad \forall i \in [N], \quad (2.82d), \quad (2.98b) - (2.98d),$$

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

Third, the inner approximation (2.24) leads to the following problem:

$$\begin{aligned}
& \min_{\mathbf{x}, \mathbf{z}, s, \hat{\boldsymbol{\lambda}}, \mathbf{q}, \mathbf{Q}_r} \quad \mathbf{c}^\top \mathbf{x} + s + \gamma_2 \mathbf{I}_{m_1} \bullet \mathbf{Q}_r + \sqrt{\gamma_1} \|\mathbf{q}\|_2 \\
& \text{s.t.} \quad \left[ \begin{array}{cc} s - \beta_k - \boldsymbol{\lambda}_k^\top \mathbf{b} - \alpha_k \mathbf{z}_k^\top \boldsymbol{\mu} + \boldsymbol{\lambda}_k^\top \mathbf{A} \boldsymbol{\mu} & \frac{1}{2} \left( \mathbf{q}_1 + \left( \mathbf{U}_{mn \times m_1} \boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}} \right)^\top (\mathbf{A}^\top \boldsymbol{\lambda}_k - \alpha_k \mathbf{z}_k) \right)^\top \\ \frac{1}{2} \left( \mathbf{q}_1 + \left( \mathbf{U}_{mn \times m_1} \boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}} \right)^\top (\mathbf{A}^\top \boldsymbol{\lambda}_k - \alpha_k \mathbf{z}_k) \right) & \mathbf{Q}_r \end{array} \right] \succeq 0, \\
& \hspace{15em} \forall k \in [K], \\
& \mathbf{q}_2 + \left( \mathbf{U}_{mn \times (mn - m_1)} \boldsymbol{\Lambda}_{mn - m_1}^{\frac{1}{2}} \right)^\top (\mathbf{A}^\top \boldsymbol{\lambda}_k - \alpha_k \mathbf{z}_k) = 0, \quad \forall k \in [K], \\
& \boldsymbol{\lambda}_k \in \mathbb{R}_+^n, \quad \forall k \in [K], \quad (2.82d), \quad (2.98b) - (2.98d),
\end{aligned}$$

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

Fourth, the inner approximation (2.37) leads to the following problem:

$$\begin{aligned}
& \min_{\mathbf{x}, \mathbf{z}, s, \mathbf{q}, \hat{\mathbf{Q}}, \hat{\boldsymbol{\lambda}}} \quad \mathbf{c}^\top \mathbf{x} + s + \gamma_2 \sum_{i=1}^P \mathbf{I}_{m_i} \bullet \mathbf{Q}_i + \sqrt{\gamma_1} \|\mathbf{q}\|_2 \\
& \text{s.t.} \quad \left[ \begin{array}{cc} s_{ik} & \frac{1}{2} \left( \mathbf{q}_i + \left( \mathbf{U}_{mn \times m_i} \boldsymbol{\Lambda}_{m_i}^{\frac{1}{2}} \right)^\top (\mathbf{A}^\top \boldsymbol{\lambda}_k - \alpha_k \mathbf{z}_k) \right)^\top \\ \frac{1}{2} \left( \mathbf{q}_i + \left( \mathbf{U}_{mn \times m_i} \boldsymbol{\Lambda}_{m_i}^{\frac{1}{2}} \right)^\top (\mathbf{A}^\top \boldsymbol{\lambda}_k - \alpha_k \mathbf{z}_k) \right) & \mathbf{Q}_i \end{array} \right] \succeq 0, \\
& \hspace{15em} \forall i \in [P], \quad \forall k \in [K], \\
& \sum_{i=1}^P s_{ik} = s - \beta_k - \boldsymbol{\lambda}_k^\top \mathbf{b} - \alpha_k \mathbf{z}_k^\top \boldsymbol{\mu} + \boldsymbol{\lambda}_k^\top \mathbf{A} \boldsymbol{\mu}, \quad \forall k \in [K], \\
& \boldsymbol{\lambda}_k \in \mathbb{R}_+^n, \quad \forall k \in [K], \quad (2.82d), \quad (2.98b) - (2.98d),
\end{aligned}$$

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



## CHAPTER 3

# Resilient NdFeB Magnet Recycling under the Impacts of COVID-19 Pandemic: Stochastic Programming and Benders Decomposition

In this chapter, we propose a chance-constrained two-stage stochastic programming (CTSP) model for a reverse supply chain network of Neodymium-iron-boron (Nd-FeB) magnets to maximize the profit while guaranteeing network resiliency. Furthermore, we develop an efficient Benders decomposition algorithm to solve the problem in large-scale instances. In Section 3.1, a brief background of the NdFeB magnet industry and the impacts of some disruptive events, including the COVID-19 pandemic, on this industry is provided. Section 3.2 provides a brief review of the literature about supply chain and logistics network design under uncertainty as well as supply chain risk management in the COVID-19 pandemic. Section 3.3 defines the studied problem in this chapter and Section 3.4 presents the proposed mathematical model for this problem. The developed Benders decomposition algorithm is introduced in Section 3.5 and the proposed model is applied to the United States in Section 3.6. In Section 3.7, comprehensive computational experiments are conducted to evaluate the performance of the proposed model and algorithm. Finally, Section 3.8 concludes this chapter by providing some managerial implications.

### 3.1 Introduction

Today's global supply chains are like a double-edged sword. Although they are efficient and productive in normal conditions, they are incredibly vulnerable to uncertain events that can disrupt the continuous flow of materials through the supply chain. The source of these uncertainties can be categorized into two groups: disruption risks and operational risks (Tang 2006, Tomlin 2006). While operational risks represent disturbances in daily operations like lead time or demand fluctuation, disruption risks are related to low-frequency but high-impact events (also called Black Swan events) like natural disasters (e.g., tsunami, earthquake, flood) and human-made disasters (e.g., fire, terrorism, utility failure, labor strikes, cyber-attack). Such risks can entirely or partially disrupt one or more supply chain and logistics network components such as suppliers, manufacturers, distribution centers, and transportation links. The consequences of these disruptions range from productivity and revenue reduction to loss of goodwill (Alcantara 2017).

The COVID-19 pandemic is the most recent example of such disruptive events. The long-term supply and demand disruptions caused by this pandemic distinguish it from other small-scale and medium-scale disruptive events (Ivanov 2020). In the first nine months of the pandemic, the COVID-19 pandemic infected about 35 million people, affected 216 countries (Worldometer 2020), and created a severe supply shortage in the global market. According to Sherman (2020), 94% of Fortune 1000 companies experienced supply chain disruptions from the COVID-19 pandemic. The most vulnerable companies had concentrated their production facilities geographically to save costs. For example, Bradstreet (2020) found that at least 51,000 companies around the world have Tier 1 suppliers in Wuhan, where the COVID-19 was observed for the first time, while at least five million companies have one or more Tier 2 suppliers in Wuhan. Moreover, the fast spread of COVID-19 in the United States and Europe obstructed the movement of the products and materials worldwide. Consequently, the COVID-19 pandemic created substantial demand and

supply disruptions globally.

One of the essential products, whose supply chain has been significantly affected by the COVID-19 pandemic, is the NdFeB magnet. More than 20 industrial sectors, including hard disk drives (HDDs) and electric vehicles (EVs), are dependent on NdFeB magnet supplies (Shaw and Constantinides 2012). This permanent magnet is made of rare-earth elements (REEs), which are classified as critical materials to the United States and European Union due to their increasing importance to clean energy and persistent supply risk. Although NdFeB magnet demand is distributed worldwide, China is the leading producer of this magnet, with approximately 80% of the global market share (Eggert et al. 2016). Therefore, the companies that rely heavily on China for NdFeB magnets or REEs face severe supply disruption risks during the COVID-19 pandemic. The facility shutdown and temporary production loss triggered by this pandemic have amplified the global REE and NdFeB magnet supply chain risk.

Besides the COVID-19 pandemic, other black swan events such as geopolitical conflicts and international trade wars have made an unpredictable, widespread, and severe impact on the global supply chains of REEs and NdFeB magnets. For example, the risk of near-monopolistic supply became evident in 2010 when China blocked the export of REEs to Japan over a geopolitical conflict. The supply disruption eventually led to an unstable market and a tenfold increase in REE prices (Golev et al. 2014). In 2019, the US increased the tariff on Chinese products from 10% to 25% as a part of the ongoing trade war. This decision raised a fear that China, as a retaliation, could ban the export of rare-earth magnets to the US since 80% of the US rare-earth element imports come from China (Schmid 2019). Therefore, several countries, including the US, encourage domestic companies to develop alternative sources to produce NdFeB magnets, such as recycling NdFeB magnets. Building a new REE supply chain with both upstream (e.g., REE mining) and downstream (e.g., NdFeB magnet production) processing capabilities is both expensive and time-consuming. However, direct magnet-to-magnet recycling can be a viable

option for producing NdFeB magnets from the end-of-life (EOL) products.

NdFeB magnet recycling could have significant benefits over new production. For example, new NdFeB magnet production is heavily dependent on the extraction of neodymium (Nd) and dysprosium (Dy) from REE bearing ores and clays, which is expensive and environmentally disastrous (Tharumarajah and Koltun 2011). Moreover, the growth of the existing industrial sectors and the development of new technologies that are heavily dependent on REEs have significantly increased the consumption of these critical resources (Nlebedim and King 2018). According to Habib et al. (2020), the annual worldwide demands for REEs are expected to increase significantly from 2020 to 2035, causing a depletion of REE geological reserves. Therefore, the primary supply of Nd and Dy may fail to meet the future demand by 2050 (Habib and Wenzel 2014).

NdFeB magnet recycling has a promising potential because the worldwide in-use stock of NdFeB magnets is about 400% larger than the annual extraction rate of REEs (Du and Graedel 2011). Companies are developing NdFeB magnet recycling technologies that utilize EOL products either for reusing the magnets or recovering REEs (Nlebedim and King 2018). According to Habib and Wenzel (2014), NdFeB magnet recycling can meet almost 50% of the neodymium (Nd) and dysprosium (Dy) demand by 2100. NdFeB magnet recycling can not only reduce new REE consumption but also lower the environmental footprint of NdFeB magnets (Jin et al. 2018). The economic viability and environmental benefits of NdFeB magnet recycling are highly sensitive to the processing volume (for economies of scale) and transportation distance among various supply chain members, highlighting the importance of optimizing the reverse logistics network.

To the best of our knowledge, there exists no quantitative decision-making model in the literature that develops a resilient NdFeB magnet recycling supply chain. Our goal is to fill this gap in the literature by designing a reverse logistics network for NdFeB magnet recycling that is resilient to both disruption and operational risks.

In this study, we aim to answer the following questions:

1. How should we strategically locate the facilities and optimize the processing capacities, inventories, and transportation flows to maximize the profit?
2. How can we make the supply chain and logistics network resilient enough to large-scale disruptions caused by a black swan event like the COVID-19 pandemic?

To answer these questions, we develop a CTSP model that not only maximizes the total profit of the recycling supply chain and logistics network but also ensures the network resiliency to cope with demand and disruption uncertainties. We also develop scenarios to model different types of supply and demand disruptions, rebound effect on NdFeB magnet demand, and the non-linear recovery process caused by large-scale disruptive events like the COVID-19 pandemic. As the CTSP model is NP-hard, we develop a customized Benders decomposition algorithm to find the optimal solution for large-scale instances and perform experiments to demonstrate its computational efficiency. The proposed model was applied to a case study of Nd-FeB magnet recycling in the US to draw managerial insights that could be helpful for supply chain and logistics operations managers.

## **3.2 Literature Review**

Shekarian and Mellat Parast (2020) provided a comprehensive review of different strategies for supply chain disruption risk management, and Chowdhury et al. (2021) reviewed COVID-19 related supply chain studies. To make the literature review more streamlined, in this section, we provide a brief review of the literature in two streams of research: (1) supply chain and logistics network design under uncertainty and (2) supply chain risk management in the COVID-19 pandemic.

### 3.2.1 Supply Chain and Logistics Network Design under Uncertainty

The increasing complexity of the global supply chain and logistics network has made it more vulnerable to operational and disruption risks than ever before. As strategic decisions like facility location are long-term decisions, it is essential to consider uncertainties while designing a supply chain and logistics network. In the literature, numerous quantitative models have been developed, focusing on either disruption or operational risks. Interested readers may refer to Golan et al. (2020), Govindan et al. (2017), Hosseini et al. (2019), Paul et al. (2016), and Snyder et al. (2016) for a more comprehensive review. However, considering both operational and disruption risks will result in more robust and reliable networks (Heckmann et al. 2015, Shukla et al. 2011). In Section 3.9, we summarize the studies that considered both operational and disruption risks in their quantitative models.

To manage the supply chain risks, various risk mitigation strategies have been proposed, which include alternative sourcing (Zhalechian et al. 2018), backup suppliers (Paul et al. 2017), backorder (Hishamuddin et al. 2013), spare capacity (Paul et al. 2014), buffer inventory (Lücker et al. 2019), facility fortification (Li et al. 2013), substitute products (Saha et al. 2020), and relationship management with suppliers (DuHadway et al. 2019). Tables 3.10 - 3.12 show that a combination of these common strategies was used to enhance supply chain and logistics network resilience.

The majority of the studies reported in Tables 3.10 - 3.12 used discrete scenarios to model disruption risks caused by natural or human-made disasters, while the stochastic programming approach has been the most popular method in this area (Ahmadi-Javid and Seddighi 2013, Mak and Shen 2012). On the other hand, operational risks were modeled using two-stage stochastic programming (Listeş 2007), robust optimization (Pishvae et al. 2011), possibilistic programming (Pishvae and Torabi 2010) or their combinations. While most of the studies adopted commercial optimization solvers to solve their mathematical models, a few studies developed

approximate or exact solution methods to solve their large-scale models (Fattahi and Govindan 2018, Hasani and Khosrojerdi 2016, Zahiri et al. 2017).

### 3.2.2 Supply Chain Risk Management during the COVID-19 Pandemic

The traditional risk management strategies may not be sufficient for large-scale disruptions like the COVID-19 pandemic due to its unique features. In preparation for the next pandemic, Choi et al. (2020), Ivanov and Das (2020), and Ivanov and Dolgui (2020) suggested supply chain practitioners exploring the ideas of Industry 4.0 and digital manufacturing for higher transparency and better control. Dolgui et al. (2020) proposed a framework for supply chain models that can survive disruptions through reconfiguring its structure and replanning economic performance with long-term impacts. During the COVID-19 pandemic, many customers inclined toward online shopping due to government-issued “stay at home” orders. Therefore, Hobbs (2020) encouraged the retail supply chain designers to develop a multi-channel distribution network. However, Chopra (2020), Gurvich and Hussain (2020), and Shih (2020) promoted the localization of production and supplier bases as an effective strategy to deal with disruptions resulted from globalization. These studies employed qualitative modeling approaches to advance their risk management strategies.

On the other hand, some studies benefited from quantitative modeling approaches to develop risk management strategies. For example, Mehrotra et al. (2020) developed a stochastic programming model that considered storing buffer inventories at each supply chain network node. They showed that sharing idle inventories between different supply chain network nodes can effectively reduce supply shortage during disruptive events. Paul and Chowdhury (2020b) proposed four strategies, i.e., resource sharing, collective sourcing, manufacturing of basic quality items, and small-sized packing, to manage the increased demand for essential items during the pandemic. They also developed an analytical model to evaluate the improvement

in service level after implementing these strategies. Paul and Chowdhury (2020a) also formulated a mathematical recovery model using production capacity expansion and emergency sourcing as the strategies against supply and demand uncertainties. Their model can be used in the production plan of high-demand, essential products.

### 3.2.3 Gaps and Contributions

Although significant contributions have been made in the literature of supply chain risk management, no articles have focused on the resiliency of REE and rare-earth magnet industries. Jin et al. (2019) formulated a deterministic model that maximizes the economic and environmental benefits of NdFeB magnet recycling. With the advent of the COVID-19 pandemic, a resilient supply chain has become more critical. As a result, we propose new models and solution approaches to design a resilient supply chain and logistics network for NdFeB magnet recycling.

Our contributions to the literature are summarized as follows:

1. We design an NdFeB magnet recycling supply chain and logistics network that is resilient to COVID-19-related operational and disruption risks by considering various risk management strategies.
2. We develop a CTSP model tailored to NdFeB magnet recycling for the first time to maximize the total profit while keeping the supply chain resilient enough to disruptive events on the same scale as the COVID-19 pandemic.
3. We design disruption scenarios that mimic the real-world COVID-19 pandemic impacts by considering disruption and recovery durations, both demand and supply disruptions with different timings, a non-linear post-disruption recovery process, and the rebound effect.
4. We develop an efficient Benders decomposition algorithm to solve the large-scale CTSP model while proposing different decomposition schemes that can

be adopted to decompose this problem.

### 3.3 Problem Description

This section describes an NdFeB magnet recycling supply chain and logistics network, which is highly complex due to its geographically scattered, multi-product, multi-period, and multi-echelon nature. Figure 3.1 shows the structure of the proposed supply chain and logistics network and its material flows. Collection centers collect EOL HDDs and EV motors from all over the US, which are transported to dismantling centers. At dismantling centers, the EOL products are disassembled to obtain the valuable NdFeB magnets and printed circuit boards (PCBs), and the rest of the materials are shredded for their scrap value. Then, the EOL magnets are transported to the recycling centers for value recovery, and PCBs and metal scraps are sold to other recyclers outside of our supply chain and logistics network. Finally, the recycled magnets are transported from recycling centers to sales points. It should be noted that the demand for recycled NdFeB magnets and disruption rates of facilities are the uncertain parameters considered in this study under the COVID-19 pandemic.

In this multi-echelon problem, the decisions are made for facilities in each echelon, including collection centers, dismantling centers, recycling centers, and sales points. As we consider HDDs and EV motors as the feed materials, the collection centers are already established in the US market: EOL HDDs are collected primarily due to data security reasons, and EV motors are collected for their economic value through the existing vehicle recovery infrastructure. Therefore, the remaining strategic decisions are where to establish, how many dismantling and recycling centers at what capacities. Dismantling and recycling centers are capital-intensive to construct, so we limit the number of new facilities.

We assume each dismantling and recycling center has a warehouse with a limited

capacity for holding buffer inventories. Therefore, tactical decisions need to be made on (1) the material flows between facilities and (2) the inventory levels of dismantling and recycling centers on a seasonal basis (i.e., three months). Again, due to the large capital investment, multi-period planning is necessary to better utilize the facilities. A five-year horizon is assumed for this purpose.

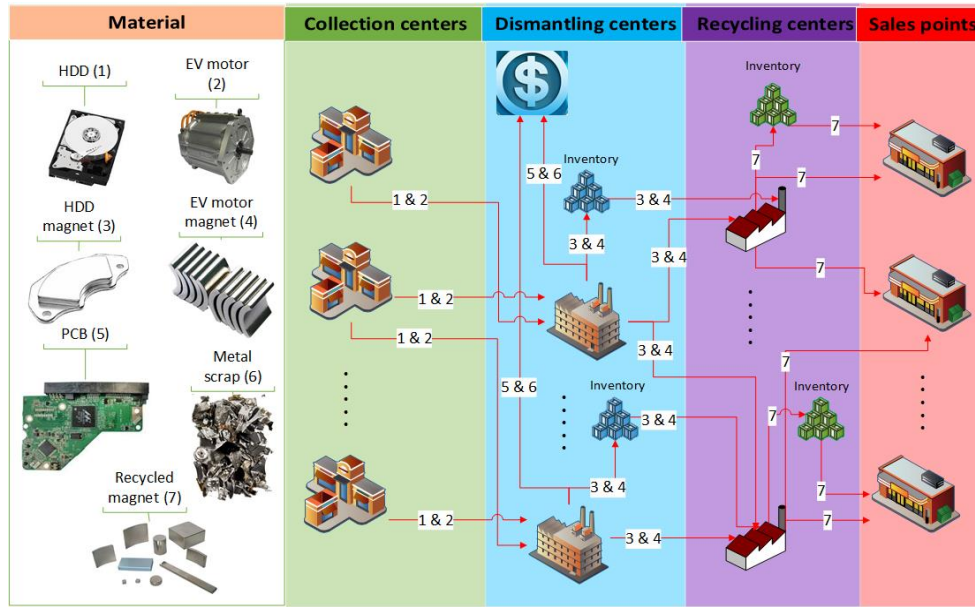


Figure 3.1: Schematic view of the designed supply chain and logistics network

We assume that the disrupted facilities have lower processing capacities to a fraction of their nominal capacities. Consequently, the customers of the disrupted facilities may face supply shortages. The recovery of the disrupted facilities is supposed to begin right after the disruption. We assume that a small fraction of the disrupted capacity can be restored at the initial stage, and the recovery rate gradually increases. Besides, we assume that demand disruption can start either simultaneously with supply disruption or at a different period. After a demand disruption, sales points may face demand shock due to the rebound effect. To deal with these uncertainties, we apply risk management strategies that increase the resilience of the supply chain and logistics network. Finally, we assume that the uncertain parameters are discretely distributed, i.e., they are represented by a set of finite scenarios

where each scenario has a specified probability of occurrence.

The resilience of this supply chain and logistics network is defined as its capacity to supply enough recycled NdFeB magnets to satisfy the demands of sales points. To ensure that the proposed supply chain and logistics network is resilient to disruptive events on the same scale as the COVID-19 pandemic, we propose several strategies. First, we use a dynamic material flow adjustment plan where material flow decisions can be revised after a disruption. Second, we allow each facility to have more than one supplier. Therefore, even if a supplier is partially or fully disrupted, other active suppliers act as a backup to ensure the continuous flow of materials through the network. Third, we diversify the dismantling and recycling center locations to minimize the risk of losing multiple facilities simultaneously due to a disruptive event in a specific location. Fourth, we assume that dismantling and recycling centers can process more materials than their customer needs and store the extra processed materials as buffer inventories. Therefore, even if the production is disrupted, they can use the buffer inventories to satisfy customer demand.

We consider a penalty cost for unsatisfied demand to monetize goodwill loss. In addition, we enforce that all the sales point demand must be satisfied with a high probability (e.g., 93%). On the other hand, inventory holding costs penalize supply overages to make balanced decisions for the supply chain and logistics network.

### **3.4 Mathematical Model Formulation**

In this section, we develop a CTSP model for the problem described in Section 3.3. This model aims to maximize the total profit of the supply chain while keeping it resilient enough to disruptive events on the same scale as the COVID-19 pandemic. At the beginning of the planning horizon, strategic or first-stage decisions are made before any disruptive event occurs. The decision-makers determine the optimal locations of dismantling and recycling centers from a set of candidate locations,

along with their optimal processing capacities. Next, the operational or second-stage decisions, which depend on the first-stage ones, are made once the actual demand and disruption rates are revealed. At this stage, the decision-makers determine the optimal volume of material flows between facilities and the optimal inventory levels of dismantling and recycling centers. The CTSP model also contains a chance constraint that guarantees meeting the demand with a specified probability. We also embed a set of discrete and independent disruption scenarios with probabilities of occurrence in the CSTP model that mimics the real disruption and operational risks. We discuss these randomly generated scenarios in more detail in Section 3.6.1. The notations used in the CTSP model are summarized in Table 3.1. Constant parameters are denoted by capital letters, and decision variables are denoted by lower cases.

The CTSP model aims to maximize the profit (before taxes), i.e., the difference between total revenue and total cost. The revenue comes from the sales of PCBs and metal scraps ( $f_{2\omega}$ ) and recycled NdFeB magnets ( $f_{3\omega}$ ). On the other hand, the total cost consists of the facility setup cost ( $f_1$ ), EOL product acquisition cost ( $f_{4\omega}$ ), transportation cost ( $f_{5\omega}$ ), other operating costs ( $f_{6\omega}$ ), penalty cost for unsatisfied demand of NdFeB magnets ( $f_{7\omega}$ ), and inventory cost ( $f_{8\omega}$ ). Therefore, the objective function of the CTSP model contains the following revenue and cost components:

Facility set-up cost:

$$f_1 = \sum_j (y_j^{Dis} \cdot FC_j^{Dis} + VC_j^{Dis} \cdot q_j^{Dis}) + \sum_r (y_r^{Rec} \cdot FC_r^{Rec} + VC_r^{Rec} \cdot q_r^{Rec}).$$

Revenue from materials other than NdFeB magnets:

$$f_{2\omega} = \sum_t \sum_j \sum_i \sum_k \sum_l P_{lt} q_{kijt\omega}^{Col} W_{lk}.$$

Revenue from recycled NdFeB magnets:

$$f_{3\omega} = \sum_t \sum_s \sum_r P_t^{Nd} \cdot q_{rst\omega}^{Rec}.$$

Table 3.1: Notations used in the CTSP model

Notation	Description
<b>Sets</b>	
$K$	Set of EOL product types
$L$	Set of components in EOL products other than NdFeB magnet
$I$	Set of collection centers
$J$	Set of dismantling centers
$R$	Set of recycling centers
$S$	Set of sales points
$T$	Set of seasons $T = \{t   t \geq 1, t \in Z^+\}$
$\Omega$	Set of scenarios
<b>Parameters</b>	
$AC_{kt}$	The acquisition cost of EOL product $k$ in season $t$ (\$/kg)
$C_{ijt}^{Col}$	Transportation cost from collection center $i$ to dismantling center $j$ in season $t$ (\$/kg)
$C_{jrt}^{Dis}$	Transportation cost from dismantling center $j$ to recycling center $r$ in season $t$ (\$/kg)
$C_{rst}^{Rec}$	Transportation cost from recycling center $r$ to sales point $s$ in season $t$ (\$/kg)
$FC_j^{Dis}$	The fixed setup cost of dismantling center $j$ (\$)
$FC_r^{Rec}$	The fixed setup cost of recycling center $r$ (\$)
$VC_j^{Dis}$	Variable set-up cost of dismantling center $j$ (\$)
$VC_r^{Rec}$	Variable set-up cost of recycling center $r$ in season $t$ (\$)
$OC_{kjt}^{Dis}$	Operating cost of EOL product $k$ in dismantling center $j$ in season $t$ (\$/kg)
$OC_r^{Rec}$	Operating cost of NdFeB magnets in recycling center $r$ in season $t$ (\$/kg)
$N^{Dis}$	Maximum number of constructed dismantling centers
$N^{Rec}$	Maximum number of constructed recycling centers
$Q_j^{Dis}$	Maximum processing capacity of dismantling center $j$ per season (kg)
$Q_r^{Rec}$	Maximum processing capacity of recycling center $r$ per season (kg)
$P_t^{Nd}$	The sales price of recycled NdFeB magnet in season $t$ (\$/kg)
$P_{lt}$	The sales price of component $l$ in season $t$ (\$/kg)
$W_{lk}$	The weight percentage of component $l$ in EOL product $k$ (%)
$W_k^{Nd}$	The weight percentage of NdFeB magnet in EOL product $k$ (%)
$S_{kit}$	Supply of EOL product $k$ in collection center $i$ in season $t$ (kg)
$\Theta$	NdFeB magnet recycling efficiency (%)
$WA_j^{Dis}$	Warehouse capacity of dismantling center $j$ per season (kg)
$WA_r^{Rec}$	Warehouse capacity of recycling center $r$ per season (kg)
$HC_t^{Dis}$	Inventory holding cost of each dismantling center in season $t$ (\$/kg)
$HC_t^{Rec}$	Inventory holding cost of each recycling center in season $t$ (\$/kg)
$PC_t$	Penalty cost of unsatisfied demand in season $t$ (\$/kg)
$\epsilon$	Risk tolerance
$P_\omega$	Probability of scenario $\omega \in \Omega$
$WP$	Warm-up period length
<b>Variables</b>	
$D_{st\omega}$	Recycled NdFeB magnet demand in sales point $s$ in season $t$ under scenario $\omega$ (kg)
$Z_{i\omega}^{Col}$	Disruption rate of collection center $i$ in season $t$ under scenario $\omega$ (%)
$Z_{j\omega}^{Dis}$	Disruption rate of dismantling center $j$ in season $t$ under scenario $\omega$ (%)
$Z_{r\omega}^{Rec}$	Disruption rate of recycling center $r$ in season $t$ under scenario $\omega$ (%)
<b>First-stage decision variables</b>	
$y_j^{Dis}$	Indicates if dismantling center $j$ is constructed (i.e., $y_j^{Dis} = 1$ ) or not (i.e., $y_j^{Dis} = 0$ )
$y_r^{Rec}$	Indicates if recycling center $r$ is constructed (i.e., $y_r^{Rec} = 1$ ) or not (i.e., $y_r^{Rec} = 0$ )
$q_j^{Dis}$	Processing capacity of dismantling center $j$ per season (kg)
$q_r^{Rec}$	Processing capacity of recycling center $r$ per season (kg)
<b>Second-stage decision variables</b>	
$q_{kijl\omega}^{Col}$	Transported product $k$ from collection center $i$ to dismantling center $j$ in season $t$ under scenario $\omega$ (kg)
$q_{jrt\omega}^{Dis}$	Transported NdFeB magnet from dismantling center $j$ to recycling center $r$ in season $t$ under scenario $\omega$ (kg)
$q_{rst\omega}^{Rec}$	Transported recycled magnet from recycling center $r$ to sales point $s$ in season $t$ under scenario $\omega$ (kg)
$inv_{j\omega}^{Dis}$	Inventory level of dismantling center $j$ in season $t$ under scenario $\omega$ (kg)
$inv_{r\omega}^{Rec}$	Inventory level of recycling center $r$ in season $t$ under scenario $\omega$ (kg)
$z_\omega$	Indicates if there is any unsatisfied demand after the warm-up period under scenario $\omega$ (i.e., $z_\omega = 1$ ) or not (i.e., $z_\omega = 0$ )

Feedstock acquisition cost:

$$f_{4\omega} = \sum_t \sum_j \sum_i \sum_k AC_{kt} q_{kijt\omega}^{Col}.$$

Transportation cost:

$$f_{5\omega} = \sum_t \left( \sum_j \sum_i \sum_k C_{ijt}^{Col} \cdot q_{kijt\omega}^{Col} + \sum_r \sum_j C_{jrt}^{Dis} \cdot q_{jrt\omega}^{Dis} + \sum_s \sum_r C_{rst}^{Rec} \cdot q_{rst\omega}^{Rec} \right).$$

Other operating costs:

$$f_{6\omega} = \sum_t \left( \sum_j \sum_i \sum_k OC_{kjt}^{Dis} q_{kijt\omega}^{col} + \sum_r \sum_j OC_{rt}^{Rec} q_{jrt\omega}^{Dis} \right).$$

Penalty cost for unsatisfied demand of NdFeB magnets:

$$f_{7\omega} = \sum_t \sum_s PC_t \cdot \left( D_{st\omega} - \sum_r q_{rst\omega}^{Rec} \right).$$

Inventory holding cost:

$$f_{8\omega} = \sum_t \left( \sum_j HC_t^{Dis} \cdot inv_{jt\omega}^{Dis} + \sum_r HC_t^{Rec} \cdot inv_{rt\omega}^{Rec} \right).$$

where  $f_{7\omega}$  is non-negative due to Constraint (3.15) shown below that ensures the sales of recycled magnets ( $q_{rst\omega}^{Rec}$ ) not exceeding the demand  $D_{st\omega}$  (i.e., we cannot sell more than what is requested by the buyers). We use  $\mathbf{x}$  to denote the first-stage decision variables, i.e.,  $y_j^{Dis}, y_r^{Rec}, q_j^{Dis}$ , and  $q_r^{Rec}$ , and  $\xi(\omega)$  to denote uncertain parameters, i.e.,  $D_{st\omega}, Z_{it\omega}^{Col}, Z_{jt\omega}^{Dis}$ , and  $Z_{rt\omega}^{Rec}$ . Then the CTSP model can be expressed in the following formulation:

First-stage objective function:

$$\max -f_1 + \mathbb{E}_\xi [Q(\mathbf{x}, \xi(\omega))], \quad (3.1)$$

where  $Q(\mathbf{x}, \xi(\omega))$  represents the optimal objective value of the second-stage model, i.e., Equation (3.7). Equation (3.1), which includes setup cost at the first stage and the expected value of the revenues and costs from the second stage decisions, maximizes the total expected profit (before taxes).

Facility location constraints:

*s.t.*

$$\sum_j y_j^{Dis} \leq N^{Dis} \quad (3.2)$$

$$\sum_r y_r^{Rec} \leq N^{Rec} \quad (3.3)$$

Constraints (3.2) - (3.3) limit the number of newly constructed dismantling and recycling centers to a maximum allowed number.

Processing capacity constraints:

$$q_j^{Dis} \leq Q_j^{Dis} \cdot y_j^{Dis}, \quad \forall j \in J, \quad (3.4)$$

$$q_r^{Rec} \leq Q_r^{Rec} \cdot y_r^{Rec}, \quad \forall r \in R, \quad (3.5)$$

Constraints (3.4) - (3.5) prevent dismantling and recycling centers from exceeding the maximum allowed processing capacity.

Domain of the first-stage decision variables:

$$y_j^{Dis}, y_r^{Rec} \in \{0, 1\}, \quad q_j^{Dis}, q_r^{Rec} \geq 0, \quad \forall j \in J, \forall r \in R. \quad (3.6)$$

Second-stage objective function:

$$Q(\mathbf{x}, \xi(\omega)) = \max f_{2\omega} + f_{3\omega} - f_{4\omega} - f_{5\omega} - f_{6\omega} - f_{7\omega} - f_{8\omega} \quad (3.7)$$

Equation (3.7) maximizes the potential profit, considering the revenues obtained from selling PCBs, metal scraps, and recycled NdFeB magnets and the costs of feedstock acquisition, transportation, operation, unsatisfied demand penalty, and inventory.

Material flow constraints:

*s.t.*

$$\sum_j q_{kijtw}^{Col} \leq Z_{itw}^{Col} \cdot S_{kit}, \quad \forall k \in K, \forall i \in I, \forall t \in T, \quad (3.8)$$

$$\sum_k q_{kijtw}^{Col} \leq Q_j^{Dis} \cdot y_j^{Dis}, \quad \forall i \in I, \forall j \in J, \forall t \in T, \quad (3.9)$$

$$\sum_i \sum_k q_{kijtw}^{Col} \leq Z_{jtw}^{Dis} \cdot q_j^{Dis}, \quad \forall j \in J, \forall t \in T, \quad (3.10)$$

$$q_{jrtw}^{Dis} \leq Q_r^{Rec} \cdot y_j^{Dis}, \quad \forall j \in J, \forall r \in R, \forall t \in T, \quad (3.11)$$

$$q_{jrtw}^{Dis} \leq Q_r^{Rec} \cdot y_r^{Rec}, \quad \forall j \in J, \forall r \in R, \forall t \in T, \quad (3.12)$$

$$\sum_j q_{jrtw}^{Dis} \leq Z_{rtw}^{Rec} \cdot q_r^{Rec}, \quad \forall r \in R, \forall t \in T, \quad (3.13)$$

$$q_{rstw}^{Rec} \leq \min\{D_{stw}, Q_r^{Rec}\} \cdot y_r^{Rec}, \quad \forall r \in R, \forall s \in S, \forall t \in T, \quad (3.14)$$

$$\sum_r q_{rstw}^{Rec} \leq D_{stw}, \quad \forall s \in S, \forall t \in T. \quad (3.15)$$

Constraints (3.8) – (3.15) determine the quantity of material flows between facilities. More specifically, Constraints (3.8) - (3.10) assure the constructed dismantling centers receive EOL HDDs and EV motors up to their processing capacities and the total supply availability, considering disruptions in collection and dismantling centers. Constraints (3.11) – (3.13) ensure recycling centers receive used NdFeB magnets from eligible dismantling centers up to their processing capacities under disruption. Similarly, Constraints (3.14) – (3.15) limit the transportation volume of recycled magnets between recycling centers and sales points up to the recyclers' processing capacities and sales points' demand.

Inventory constraints:

$$inv_{jt\omega}^{Dis} \leq WA_j^{Dis} \cdot y_j^{Dis}, \quad \forall j \in J, \forall t \in T, \quad (3.16)$$

$$inv_{rt\omega}^{Rec} \leq WA_r^{Rec} \cdot y_r^{Rec}, \quad \forall r \in R, \forall t \in T, \quad (3.17)$$

$$inv_{jt\omega}^{Dis} = inv_{j(t-1)\omega}^{Dis} + \sum_i \sum_k W_k^{Nd, Col} q_{kij t\omega} - \sum_r q_{jrt\omega}^{Dis}, \quad \forall j \in J, \forall t \in T, \quad (3.18)$$

$$inv_{rt\omega}^{Rec} = inv_{r(t-1)\omega}^{Rec} + \Theta \sum_j q_{jrt\omega}^{Dis} - \sum_s q_{rst\omega}^{Rec}, \quad \forall r \in R, \forall t \in T, \quad (3.19)$$

$$\sum_r q_{jrt\omega}^{Dis} \leq inv_{jt\omega}^{Dis} + \sum_i \sum_k W_k^{Nd, Col} q_{kij t\omega}, \quad \forall j \in J, \forall t \in T, \quad (3.20)$$

$$\sum_s q_{rst\omega}^{Rec} \leq inv_{rt\omega}^{Rec} + \Theta \sum_j q_{jrt\omega}^{Dis}, \quad \forall r \in R, \forall t \in T. \quad (3.21)$$

Constraints (3.16) – (3.17) allow facilities to hold inventory up to the warehouse capacities. Constraints (3.18) – (3.19) define a facility's inventory level based on its incoming and outgoing material flows and its inventory level in the previous season. Finally, constraints (3.20) – (3.21) imply that the outgoing flow of each dismantling and recycling center must be up to its processed material and inventory level.

Chance constraint:

$$Pr \left( \sum_r q_{rst\omega}^{Rec} \geq D_{st\omega}, \quad \forall s \in S, \forall t \in T, t > WP \right) \geq 1 - \epsilon \quad (3.22)$$

Constraint (3.22) ensures the demand at each sales point is completely met after the warm-up period with the probability of at least  $1 - \epsilon$ , where  $\epsilon$  denotes a predefined tolerable risk with a value between 0 and 1. Note that Constraint (3.22) does not contradict with Constraint (3.15) because they have equality sign in common. In other words, they together imply that recycling centers must transport recycled magnets to the sales points as much as the sales points need but they can transport less recycled magnets as long as the total probability of scenarios with unsatisfied demand is less than or equal to  $\epsilon$ . Constraint (3.22) is then replaced with the

following sets of deterministic constraints:

$$D_{stw} - \sum_r q_{rstw}^{Rec} \leq D_{stw} \cdot z_\omega, \quad \forall s \in S, \forall t \in T, t > WP, \quad (3.23a)$$

$$\sum_\omega P_\omega \cdot z_\omega \leq \epsilon, \quad (3.23b)$$

where  $z_\omega$  indicates if there is any unsatisfied demand after the warm-up period under scenario  $\omega$  (i.e.,  $z_\omega = 1$ ) or not (i.e.,  $z_\omega = 0$ ).

We assume no initial inventories for the dismantling and recycling centers at the beginning of the planning horizon. Therefore, we may not meet the desired service level in the first few periods of facility set-up. Thus, we consider a warm-up period (WP) that is necessary for the system to reach the working status of a supply chain and logistics network. During this warm-up period, we relax the stringent requirement of meeting all the demands to help mimic the real-world system. Moreover, since the dismantling and recycling centers are being established during the warm-up period, they can operate at a processing capacity that is smaller than their nominal processing capacities, i.e.,  $q_j^{Dis}$  and  $q_r^{Rec}$ . As the construction of the facilities progresses during the warm-up period, their processing capacity increases gradually so that they reach their nominal processing capacities by the end of this period. Therefore, after the warm-up period, all the newly established dismantling and recycling centers will be able to operate at their full processing capacities given that they are not impacted by a disruption.

Domain of the second-stage decision variables:

$$q_{kijt\omega}^{Col}, q_{jrt\omega}^{Dis}, q_{rst\omega}^{Rec} \geq 0, \quad \forall i \in I, \forall j \in J, \forall r \in R, \forall s \in S, \forall t \in T, \quad (3.24)$$

$$inv_{j0\omega}^{Dis}, inv_{r0\omega}^{Rec} = 0, \quad inv_{rt\omega}^{Rec}, inv_{jt\omega}^{Dis} \geq 0, \quad \forall j \in J, \forall r \in R, \forall t \in T, \quad (3.25)$$

$$z_\omega \in \{0, 1\}. \quad (3.26)$$

Considering the discrete distribution of the random variables, the CTSP model can be expressed as the following deterministic mixed-integer linear programming

(MILP) problem:

$$\begin{aligned} \max \quad & -f_1 + \sum_{\omega} P_{\omega} \cdot Q(\mathbf{x}, \xi(\omega)) \\ \text{s.t.} \quad & \\ & (3.2) - (3.6) \\ & (3.8) - (3.26), \quad \forall \omega \in \Omega. \end{aligned}$$

### 3.5 Solution Approach

Although optimization solvers can directly solve the deterministic MILP problem, they are not computationally efficient in solving large-scale instances. A computationally efficient method to solve two-stage stochastic programs is Benders decomposition, or so-called L-shaped algorithm (Van Slyke and Wets 1969).

Benders (1962) developed the Benders decomposition algorithm for the first time to solve mixed-integer programming problems to optimality. This solution method splits the original problem between a master problem and a subproblem. The master problem contains integer (and possibly some continuous) decision variables of the original problem, while the subproblem consists of the remaining decision variables. In other words, the master problem includes complicating decision variables. The complicating decision variables are referred to as the decision variables that when they are fixed as constants, the remaining mathematical model can be solved relatively easily. The subproblem can often be further decomposed into smaller problems due to the block diagonal structure of its constraint matrix. At each iteration of this algorithm, the master problem is solved to optimality. Then, the subproblem(s) is constructed based on the optimal solution of the master problem. In the next step, the constructed subproblem(s) is solved to optimality, and the Benders optimality and feasibility cuts are generated based on this optimal solution. In the next iteration, the updated master problem, which includes the newly generated Benders cuts, is solved again and the algorithm continues as before until no further cuts can

be found by the subproblem(s). A comprehensive literature review of the Benders decomposition algorithm can be found in (Rahmaniani et al. 2017).

Different decomposition schemes can be adopted to decompose the deterministic MILP problem into a master problem and one or more subproblems. Each decomposition scheme may differ from others in terms of computational time. Therefore, we propose different decomposition schemes to investigate which one is the most efficient. In the first decomposition scheme, used by CPLEX’s Benders decomposition algorithm and called “CPLEX’s default decomposition scheme” in this chapter, the master problem includes only binary decision variables, while continuous decision variables are included in the subproblem(s). In addition, we propose two more decomposition schemes (i.e., single-cut and multi-cut) that exploit the model’s two-stage structure. In these schemes, the first-stage decision variables along with the auxiliary decision variable  $z_\omega$  are assigned to the master problem, while the other second-stage decision variables are included in the subproblem(s). When the first-stage decision variables are fixed, the second stage problem can be decomposed to  $|\Omega|$  independent linear subproblems that can be solved more quickly. The number of the Benders cuts added to the master problem at each iteration is equal to the number of subproblems, i.e., each subproblem adds one cut to the master problem. Therefore, various decomposition schemes can be proposed by considering different levels of aggregation for the second-stage problems. In this case, we use the complete aggregation of the second-stage problems which leads to one subproblem (single-cut decomposition scheme). We also use completely disaggregated second-stage problems which lead to one subproblem for each scenario (multi-cut decomposition scheme) (Khasiba et al. 2020).

Each of the single-cut and multi-cut decomposition schemes has its own advantages. The single-cut decomposition scheme includes fewer decision variables and Benders cuts. Thus, the master problem is typically solved faster. On the other hand, in the multi-cut decomposition scheme, many Benders cuts are added per iteration. So, it typically converges in fewer iterations. Also, the master prob-

lem might become large and slow to solve (Luedtke 2016). Therefore, inspired by (Khassiba et al. 2020), we also consider a hybrid decomposition scheme called a “partially-aggregated-cut decomposition scheme”. In this scheme, the second-stage problems are partially aggregated into different clusters where each cluster contains a few scenarios. Therefore, the number of subproblems in this scheme equals the number of clusters.

In this study, we develop three partially aggregated cut decomposition schemes with three, five, and seven clusters and we distribute the scenarios over the clusters uniformly. Let the following problem represent a two-stage stochastic programming (TSP) problem, which is a CTSP model with no chance constraint, i.e.,  $\epsilon = 1$ :

$$\begin{aligned} \min \quad & c^\top \mathbf{x} + \sum_{\omega} P_{\omega} \cdot Q(\mathbf{x}, \xi(\omega)) & (3.27) \\ \text{s.t.} \quad & \\ & A\mathbf{x} \geq b, \\ & \mathbf{x} \in \mathcal{X}, \end{aligned}$$

where for all  $\omega \in \Omega$  we have  $Q(\mathbf{x}, \xi(\omega)) = \min_{\mathbf{y}} \{u_{\omega}^\top \mathbf{y} : D_{\omega} \mathbf{y} = g_{\omega} - B_{\omega} \mathbf{x}, \mathbf{y} \in \mathcal{Y}\}$  whose dual problem is  $\max_{\alpha} \{\alpha^\top (g_{\omega} - B_{\omega} \mathbf{x}) : \alpha^\top D_{\omega} \leq u_{\omega}^\top\}$ . Let  $\Gamma_{\omega}$ ,  $\text{EP}(\Gamma_{\omega})$ , and  $\text{ER}(\Gamma_{\omega})$  denote the feasible region of the dual problem and its sets of the extreme points and extreme rays, respectively. We also define  $V^{\omega} \subseteq \text{EP}(\Gamma_{\omega})$  and  $R^{\omega} \subseteq \text{ER}(\Gamma_{\omega})$  where  $\alpha^{\omega} \in V^{\omega}$  and  $r^{\omega} \in R^{\omega}$  indicate an extreme point and extreme ray of the  $\Gamma_{\omega}$ , respectively. Under the multi-cut decomposition scheme, Problem (3.27) is decomposed to the following master problem (MP) and subproblems ( $\text{SP}_{\omega}$ ):

$$\begin{aligned} \text{MP} : z = \min_{\mathbf{x}, \theta_{\omega}} \quad & c^\top \mathbf{x} + \sum_{\omega} P_{\omega} \theta_{\omega} \\ \text{s.t.} \quad & \\ & A\mathbf{x} \geq b, \quad \mathbf{x} \in \mathcal{X}, \\ & \theta_{\omega} \geq (\alpha^{\omega})^\top (g_{\omega} - B_{\omega} \mathbf{x}), \quad \forall \omega \in \Omega, \quad \forall \alpha^{\omega} \in \text{EP}(\Gamma_{\omega}), \\ & (r^{\omega})^\top (g_{\omega} - B_{\omega} \mathbf{x}) \leq 0, \quad \forall \omega \in \Omega, \quad \forall r^{\omega} \in \text{ER}(\Gamma_{\omega}), \end{aligned}$$

where the second and third constraints are optimality and feasibility cuts, respectively.

$$\begin{aligned} \text{SP}_\omega : Q_\omega(\hat{\mathbf{x}}^t, \xi(\omega)) &= \max_{\alpha} \{ \alpha^\top (g_\omega - B_\omega \hat{\mathbf{x}}^t) : \alpha^\top D_\omega = u_\omega^\top \} \\ &= \min_{\mathbf{y}} \{ u_\omega^\top \mathbf{y} : D_\omega \mathbf{y} = g_\omega - B_\omega \mathbf{x}, \mathbf{y} \in \mathcal{Y} \} \end{aligned}$$

Similarly, the following problem represents the master problem under the single-cut decomposition scheme:

$$\text{MP: } z = \min_{\mathbf{x}, \Phi} c^\top \mathbf{x} + \Phi$$

*s.t.*

$$A\mathbf{x} \geq b, \mathbf{x} \in \mathcal{X},$$

$$\Phi \geq \sum_{\omega} P_\omega (\alpha^\omega)^\top (g_\omega - B_\omega \mathbf{x}), (\alpha^1, \dots, \alpha^\omega) \in \text{EP}(\Gamma_1) \times \dots \times \text{EP}(\Gamma_\omega),$$

$$(r^\omega)^\top (g_\omega - B_\omega \mathbf{x}) \leq 0, \forall \omega \in \Omega, r^\omega \in \text{ER}(\Gamma_\omega),$$

where the second and third constraints are optimality and feasibility cuts, respectively. With this background, we provide the pseudocode of the developed Benders decomposition algorithm with a multi-cut decomposition scheme in Table 3.2. In Section 3.7, we conduct comprehensive computational experiments to compare the performance of different decomposition schemes.

### 3.6 Case Study

Since NdFeB magnet recycling is a new business in the US, we apply the developed CTSP model and solution approach to set up an optimal reverse supply chain and logistics network in terms of profitability and resilience against a black swan event like the COVID-19 pandemic. We use a part of the data reported in (Jin et al. 2019) for this case study. The following assumptions are made on the parameters that affect the objective function:

Table 3.2: Benders decomposition algorithm with a multi-cut decomposition scheme

---

```

1: Set  $t = 0$ ;
2: Solve  $\min_{\mathbf{x}} \{c^\top \mathbf{x} : A\mathbf{x} \geq b, \mathbf{x} \in \mathcal{X}\}$  to get the optimal solution  $\hat{\mathbf{x}}^0$ ;
3: for  $\omega = 1, 2, \dots, |\Omega|$ 
4:   Solve subproblem  $Q(\hat{\mathbf{x}}^0, \xi(\omega))$ ;
5:   if subproblem  $\omega$  is feasible, i.e.,  $\hat{\theta}_\omega^0 < Q(\hat{\mathbf{x}}^0, \xi(\omega))$ 
6:     let  $\alpha^\omega$  be an optimal dual solution with  $\hat{\theta}_\omega < (\hat{\alpha}^\omega)^\top (g_\omega - B_\omega \hat{\mathbf{x}}^0)$ ;
7:     Add optimality cut to the master problem:  $V^\omega = \{\hat{\alpha}^\omega\}$ ;
8:   else
9:     Let  $\hat{r}^\omega$  be a dual extreme ray with  $(\hat{r}^\omega)^\top (g_\omega - B_\omega \hat{\mathbf{x}}^0) > 0$ ;
10:    Add feasibility cut to the master problem:  $R^\omega = \{\hat{r}^\omega\}$ ;
11:   end if
12: end for
13: while  $\hat{z}_t < c^\top \hat{\mathbf{x}}^t + \sum_\omega P_\omega Q(\hat{\mathbf{x}}^t, \xi(\omega)) - \epsilon$  do
14:   Set  $t = t + 1$ ;
15:   Solve master problem to get solution  $(\hat{\mathbf{x}}^t, \hat{\theta}^t)$  with objective value  $\hat{z}_t$ ;
16:   for  $\omega = 1, 2, \dots, |\Omega|$ 
17:     Solve subproblem  $Q(\hat{\mathbf{x}}^t, \xi(\omega))$ ;
18:     if subproblem  $\omega$  is feasible, i.e.,  $\hat{\theta}_\omega^t < Q(\hat{\mathbf{x}}^t, \xi(\omega))$ 
19:       let  $\alpha^\omega$  be an optimal dual solution with  $\hat{\theta}_\omega < (\hat{\alpha}^\omega)^\top (g_\omega - B_\omega \hat{\mathbf{x}}^t)$ ;
20:       Add optimality cut to the master problem:  $V^\omega \leftarrow V^\omega \cup \{\hat{\alpha}^\omega\}$ ;
21:     else
22:       Let  $\hat{r}^\omega$  be a dual extreme ray with  $(\hat{r}^\omega)^\top (g_\omega - B_\omega \hat{\mathbf{x}}^t) > 0$ ;
23:       Add feasibility cut to the master problem:  $R^\omega \leftarrow R^\omega \cup \{\hat{r}^\omega\}$ ;
24:     end if
25:   end for
26: end while

```

---

1. The planning horizon is five years, i.e., 20 seasons.
2. A total of 48 collection centers are considered, one per state except for Hawaii and Alaska. The assumed coordinate of each collection center, along with the weight and cost of the collected HDDs and EV motors at each facility in the first season, can be found in Section 3.10.
3. Recycled magnets are assumed to be sold to six facilities of an automotive motor manufacturer located in Michigan, Ohio, Oklahoma, Alabama, and New York (2 facilities). The coordinates of each sales point and their assumed demand are shown in Section 3.10.

4. We consider five candidate locations for constructing dismantling centers, and three candidate locations for constructing recycling centers. The input data of this case study considers only one location in each state that leads to facility diversification because at most one dismantling and one recycling center can be constructed in each state. The assumed coordinates of these locations, costs, and maximum processing capacities are available in Section 3.10. Since dismantling and recycling centers are capital-intensive facilities, we assume at most half of the candidate sites for each facility type (or the ceiling of this number if it is not an integer) to be constructed.
5. The price of the recycled NdFeB magnet is assumed to be \$60/kg in the first season (Handwerker 2019).
6. The inventory holding costs of used and recycled NdFeB magnets are assumed to be 10% of the processing costs of EOL products and used magnets in the same season, respectively.
7. The penalty cost of unsatisfied demand is set to 15% of the recycled NdFeB magnet price in the same season. Moreover, 7% of risk tolerance, i.e., the service level of 93%, is considered for this supply chain and logistics network. In other words, the total probabilities of the scenarios under which demand is not completely satisfied after the warm-up period must be less than or equal to 7%.
8. The warehouse capacities of dismantling and recycling centers equal to their maximum processing capacities in a season.
9. Considering the average inflation rate of the last five years, i.e., from 2016 to 2020, all monetary values are assumed to increase by 0.5% per season (Calculator 2020). The values of other parameters can be found in Section 3.10.

### 3.6.1 Scenario Generation Method

In our study, we consider the following facility disruption cases:

1. Disruption in only collection centers
2. Disruption in only dismantle centers
3. Disruption in only recycling centers
4. Disruption in both collection and dismantling centers
5. Disruption in both collection and recycling centers
6. Disruption in both dismantling and recycling centers
7. Disruption in collection, dismantling centers, and recycling centers

The disruption probability of each facility is chosen based on expert opinion. Facilities located at more disaster prone areas are assigned higher disruption probabilities. It is assumed that the disruptions of facilities are independent of each other. In this study, only high-frequency scenarios that significantly affect the supply chain and logistics network are selected so that less likely events are ignored. Typically, complete loss of capacity after disruption is a less likely event and thus it is not explored. Therefore, we assumed a facility may lose 30% to 80% of its capacity after a disruption. Also, in all seven cases, recovery actions are taken to restore the disrupted facilities to their normal status. The number of seasons required for the full recovery may vary depending on the severity of disruption and resource availability. In this study, we consider three types of recovery for each of the seven facility disruption cases: short-term (up to two seasons), medium-term (up to six seasons), and long-term (up to ten seasons). Therefore, in total, we have 21 groups for disruption scenarios. According to Olson and Wu (2013), the number of scenarios in a stochastic programming problem should be kept limited and manageable. However, our assumption about the partial disruptions of facilities can easily lead to an exponential growth of scenarios within each group. To resolve this

issue, we decided to choose a representative scenario from each group. To this end, we first randomly generated a set of sample scenarios for each group and ordered them based on the probability of occurrence. After that, we chose the most probable scenarios as a representative of that group. Including the ideal scenario when no disruption occurs, we have 22 scenarios along with their probability of occurrence. Finally, we normalized the probabilities so that we can use these scenarios to evaluate the expected value in (3.1) using the Sample Average Approximation (SAA) method (Shapiro and Homem-de Mello 1998, Kleywegt et al. 2002).

### 3.6.1.1 Disruption Rates of Facilities

In each scenario, the disruption may start in any season of the planning horizon. The binary parameter  $\beta_\omega^t$  is set to 1 if a disruptive event occurs in season  $t$  under scenario  $\omega$ ; otherwise, it is set to 0. Similar to Azad et al. (2014), Fattahi et al. (2017), and Teimuory et al. (2013), we also assume that the lost capacity of disrupted facility  $f$  under scenario  $\omega$ , denoted by  $\gamma_{f\omega}$  in percentages, follows a normal distribution. The working capacity fraction of facility  $f$  in season  $t$  under scenario  $\omega$ , denoted by  $CP_{f\omega}^t$ , is calculated by the following equation:

$$CP_{f\omega}^t = (CP_{f\omega}^{t-1} + R_{f\omega}^t) \times \max\left(0, 1 - \beta_\omega^t \times \frac{\max(0, \gamma_{f\omega})}{100}\right), \forall t \in T,$$

where

$$R_{f\omega}^t = \begin{cases} 0, & \text{if } CP_{f\omega}^{t-1} = 1 \\ \min(1 - CP_{f\omega}^{t-1}, f(\nabla t)), & \text{if } 0 \leq CP_{f\omega}^{t-1} \leq 1 \end{cases}$$

Here,  $R_{f\omega}^t$  denotes the recovered capacity of facility  $f$  in season  $t$  under scenario  $\omega$ , and  $f(\nabla t)$  denotes the recovery rate at time  $\nabla t$  which is the time difference between the current season and the season in which disruption occurs.

Very few studies have considered the post-disruption recovery process of disrupted facilities. They have typically assumed that the recovery rate of dis-

rupted facilities is linear. However, the assumption is often not the case, as the recovered capacity in the first few periods after a disruption may be very low in practice, and then the recovery rate may increase gradually. Hence, we use piecewise linear functions to capture the desired characteristics of the recovery process. We develop the following recovery functions,  $f(\nabla t)$ , for short-term, medium-term, and long-term recoveries:

Short-term recovery:

$$f(\nabla t) = \begin{cases} 0.25\nabla t, & \text{if } 0 \leq \nabla t \leq 1 \\ 0.3\nabla t - 0.05, & \text{if } 1 \leq \nabla t \leq 2 \end{cases}$$

Medium-term recovery:

$$f(\nabla t) = \begin{cases} 0.067\nabla t, & \text{if } 0 \leq \nabla t \leq 2 \\ -0.006\nabla t + 0.146, & \text{if } 2 \leq \nabla t \leq 4 \\ 0.0705\nabla t - 0.16, & \text{if } 4 \leq \nabla t \leq 6 \end{cases}$$

Long-term recovery:

$$f(\nabla t) = \begin{cases} 0.00965\nabla t, & \text{if } 0 \leq \nabla t \leq 2 \\ 0.013\nabla t - 0.0067, & \text{if } 2 \leq \nabla t \leq 4 \\ 0.03308\nabla t - 0.08702, & \text{if } 4 \leq \nabla t \leq 6 \\ 0.01441\nabla t + 0.025, & \text{if } 6 \leq \nabla t \leq 8 \\ 0.01951\nabla t - 0.0158, & \text{if } 8 \leq \nabla t \leq 10 \end{cases}$$

### 3.6.1.2 Demand for Recycled NdFeB Magnets

A disruptive event may also cause demand disruption that could start simultaneously or a few seasons after supply disruption. Without considering the disruption effect, historical data projects an increasing NdFeB magnet demand in the US. To capture this trend, we use the seasonal autoregressive integrated moving average (SARIMA) method to forecast the expected demand at sales point  $s$  in season  $t$  under scenario  $\omega$  ( $\bar{D}_{s\omega}^t$ ). However, the actual demand may

be different from the expected value because of monthly fluctuations. To simulate this uncertainty, we use a triangular distribution to generate random demand within the 95% confidence interval provided by the SARIMA model. Besides operational risks, demand disruption and the rebound effect in the post-disruption period can unexpectedly shift from the average. To represent this variation, we introduce a new parameter  $\epsilon_{s\omega}^t$  for use in the following equation to simulate the demand for the NdFeB magnet in the US:

$$D_{s\omega}^t = \text{triag}(LB, \bar{D}_{s\omega}^t, UB) \times \epsilon_{s\omega}^t,$$

where

$$\epsilon_{s\omega}^t = \begin{cases} 1, & \text{no demand disruption} \\ < 1, & \text{demand disruption} \\ > 1, & \text{a surge in demand} \end{cases}$$

### 3.6.2 Case Study Results

We implement the CTSP model and solve it to optimality using the solution method proposed in Section 3.5. The first-stage decisions, including facility locations and their processing capacities, are made at the beginning of the planning horizon and do not change over time. However, the second-stage decisions, including inventory and material flow decisions, are made after the realization of uncertainties and vary by season and scenario. Due to the chapter length limitation, we elaborate on the major optimal decisions of only one scenario to showcase some important insights. Figure 3.2 presents optimal material flows of a scenario in season 1 and optimal facility locations where the states with brighter colors have lower operating costs. The optimal solutions of other decision variables (e.g., processing capacity) are summarized in Section 3.10.

- About 13% of EOL HDDs and EV motors are collected from California, highlighting the importance of including California in this recycling business. The

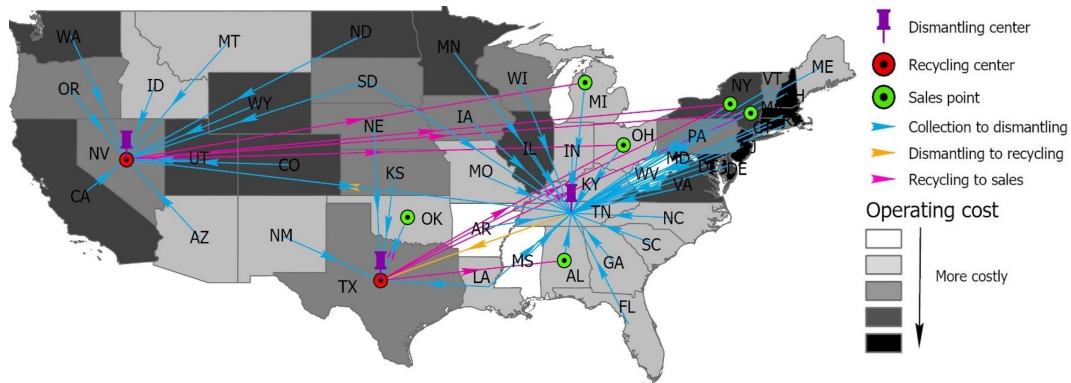


Figure 3.2: Geographic representation of facilities and material flows in the optimal solution

other top suppliers are in Texas, Florida, New York, Pennsylvania, Illinois, Ohio, Georgia, North Carolina, and Michigan (ranked in the descending order), contributing about 42% of the total EOL supply.

- Three dismantling centers are suggested for construction. The largest dismantling center shall be located in Tennessee primarily due to its proximity to the sales points and low operating cost. Its processing capacity is 1,250 tons of EOL products per season, followed by Nevada with a processing capacity of 997 tons/season, and Texas with a processing capacity of 920 tons/season.
- Two recycling centers are suggested for construction: one in Texas (with a processing capacity of 62.5 tons/season) and the other in Nevada (with a processing capacity of 56.3 tons/season). These states benefit from the collocation of dismantling and recycling centers to minimize transportation costs.
- The total profit is projected to be \$165 million (before taxes). Approximately 59% of the revenue is obtained by selling PCBs and metal scraps, while the remaining 41% is from selling the recycled NdFeB magnets, highlighting the importance of dismantling that yields multiple components and revenue streams.
- The optimal plan leads to a resilient supply chain and logistics network. The expected value of the average unsatisfied demand over all scenarios is 0.28%,

meaning only 0.28% of the total demand after the warm-up period is not met. Notably, in the worst-case scenario where a significant number of facilities are heavily disrupted, only 8.08% of the total demand is not satisfied.

- Buffer inventories significantly increase the resilience of the network at a low cost: Inventory holding cost constitutes only 0.31% of the total cost, but this small investment leads to satisfying 99.72% of the total demand. The best warehouse size for a dismantling center is 25% of its maximum processing capacity and that for a recycling center is 80% of its maximum processing capacity. The optimal inventory levels are relatively low (i.e.,  $\sim 12\%$  and  $\sim 10\%$  of warehouse capacities of dismantlers and recyclers, respectively, on average).
- If we assume the exact future demand and disruption rates are known ahead, the resulting optimal supply chain and logistics network (shown in Figure 3.3) is much different from the proposed CTSP results. The optimal plan of the CTSP model locates one more dismantling center in Texas to improve the resilience of the supply chain. Note that this deterministic model can be considered as a CTSP model whose risk tolerance is one, i.e., it has no chance constraint, and it has only one scenario, called the average scenario. The demand and disruption rates of the average scenario are set to be the expected values of the demand and disruption rates of all scenarios.

### 3.7 Computational Experiments

This section first explains how we generate test instances whose scenarios mimic the disruption and operational risks caused by the COVID-19 pandemic. Using the generated test instances, we compare the CTSP, TSP, and deterministic models in terms of profit, resilience, and computation burden and show the impact of the risk tolerance  $\epsilon$  on these metrics. Finally, we conduct a comprehensive experiment to

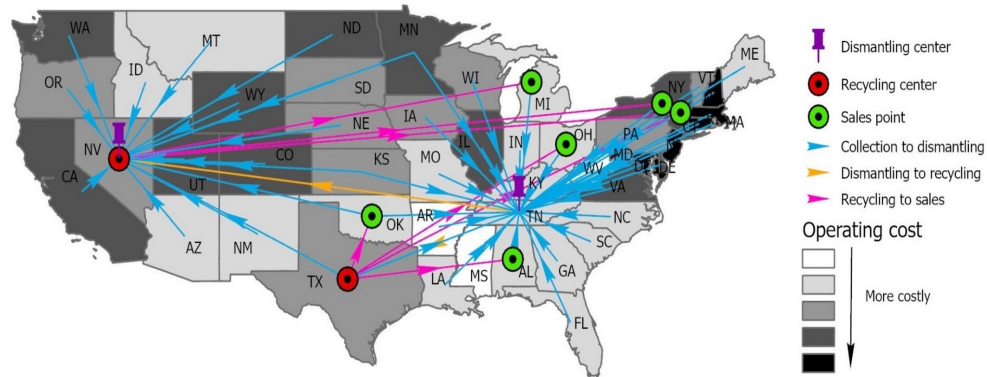


Figure 3.3: Geographic representation of the optimal solution of the deterministic problem

find the best Benders decomposition method among those proposed in Section 3.5. The mathematical model and solution algorithms are implemented using C++ API and IBM CPLEX optimizer (version 12.10.0.0) on a PC with a 64-bit Windows Operating System, a 16 GB RAM, and an Intel(R) Core(TM) i7-7700 CPU @3.60 GHz processor. The time limit for each test problem is set to 40 hours.

### 3.7.1 Test Instance Generation

To evaluate the performance of the mathematical models and solution methods, we generate nine test instances. Table 3.3 reports the size of the test instances where the first instance is the case study described in Section 3.6. More details about the test instances, including the coordinates of candidate sites and scenarios, are available in Section 3.10.

### 3.7.2 Optimization and Robustness Testing

To evaluate how robust our findings are, we conduct three types of robustness tests on the CTSP model. First, we assess the robustness of this model in terms of its performance in providing optimal plans for small-sized, medium-sized, and large-sized supply chain and logistics networks. To that end, all experiments of this

Table 3.3: Size of test problems

Test No.	$ I $	$ J $	$ R $	$ S $	$ T $	$ \Omega $	Binary decision variables	Continues decision variables	Total decision variables	Constraints
1	48	5	3	6	20	22	30	229248	229278	187924
2	48	5	4	6	20	22	31	234529	234560	196725
3	48	7	6	6	20	22	35	335733	335768	270649
4	48	9	8	6	20	22	39	440457	440496	351613
5	48	11	10	6	20	22	43	548701	548744	439617
6	48	13	12	6	20	22	47	660465	660512	534661
7	48	15	14	6	20	22	51	775749	775800	636745
8	48	17	16	6	20	22	55	894553	894608	745869
9	48	19	18	6	20	22	59	1016877	1016936	862033

section are conducted on nine supply chain and logistics networks, as described in Table 3.3, ranging in size from small to large. Second, we evaluate the robustness of the CTSP model in terms of its capability of satisfying the demand compared to other alternative models. Finally, we assess how robust the CTSP model is by evaluating its optimal plans for different risk levels.

For the first and second robustness tests, we compare the optimal plan of the CTSP model with those of TSP and the deterministic model in terms of profitability, resiliency, and computational burden using the test instances of Section 3.7.1 and scenario generation method described in Section 3.6.1. The deterministic model is a TSP model with only one scenario, called the average scenario, whose demand and disruption rate values are equal to the expected value of the demand and disruption rate values of all scenarios. In this section, the risk tolerance of the CTSP model is 0.07, unless specified otherwise, and the models are solved with the best solution approach identified in Section 3.7.3. Moreover, the unsatisfied demand excludes that of the warm-up period.

We summarize the results of this comparison in Table 3.4 where columns “Test No.,” “OPT,” and “CPU” represent the test instance number, the expected total profit, and the computation time, respectively. Column “Exp.” indicates the expected value of the average unsatisfied demand of all scenarios and column “Prob.”

denotes the total probability of encountering scenarios with unsatisfied demand.

From Table 3.4, we have the following observations. First, the optimal plan of the CTSP model is less profitable but more resilient than that of the TSP model. More specifically, although the CTSP model yields 0.73% lower profit and 15.53% more CPU time than the TSP model on average, the unsatisfied demand is significantly lower (i.e., 84.81%) than the TSP model on average. Second, the total probability of encountering scenarios with unsatisfied demand equals its specified risk tolerance of each model. This observation clearly shows that as more resilience leads to less profit, the models try to satisfy just the minimum enforced service level to maximize the profit. We discuss the impact of risk tolerance on profit and resilience in more detail in Table 3.5. Third, the deterministic model is 99.95% faster and 85.22% less resilient than the CTSP model on average.

Table 3.4: The CTSP model performance versus the deterministic model and TSP model performances

Test No.	Deterministic				TSP				CTSP			
	OPT ( $\times 10^8$ )	CPU (Sec.)	Shortage		OPT ( $\times 10^8$ )	CPU (Sec.)	Shortage		OPT ( $\times 10^8$ )	CPU (Sec.)	Shortage	
			Exp. (%)	Prob.			Exp. (%)	Prob.			Exp. (%)	Prob.
1	1.67011	0.1	1.75	1	1.67552	92.3	1.69	1	1.65828	111.5	0.28	0.07
2	1.68248	0.1	2.96	1	1.68632	96.2	2.86	1	1.67314	168.4	0.51	0.07
3	1.68654	0.1	1.65	1	1.69037	205.7	1.63	1	1.68684	210.6	0.29	0.07
4	1.6871	0.2	1.88	1	1.69192	321.5	1.83	1	1.68533	341.1	0.35	0.07
5	1.70188	0.2	3.03	1	1.70684	384.6	2.9	1	1.69815	427.1	0.31	0.07
6	1.7764	0.2	5.52	1	1.77887	459.7	5.47	1	1.7681	588.2	0.54	0.07
7	1.69524	0.2	1.9	1	1.70795	757.2	1.8	1	1.69615	888.3	0.29	0.07
8	1.69844	0.3	3.33	1	1.70657	954.2	3.29	1	1.69489	1148.9	0.61	0.07
9	1.70818	0.3	2.29	1	1.7125	1198.7	2.25	1	1.68395	1304.1	0.23	0.07

For the first and third robustness tests, we evaluate the optimal plans provided by the CTSP model for different risk levels in supply chain and logistics networks with different sizes. Table 3.5 displays how the risk tolerance value,  $\epsilon$ , affects the profit, resilience, and computation time of the CTSP model. First, increasing  $\epsilon$  leads to a non-decreasing trend in profit. This observation reveals that profit and resilience are contradictory goals. Second, increasing  $\epsilon$  leads to a non-decreasing trend in unsatisfied demand. Third, an increase in the value of  $\epsilon$  can either increase or decrease the computation time.

Table 3.5: Impact of risk tolerance on the CTSP model performance

Test No.	Risk level (%)	OPT	Shortage Exp. (%)	Shortage Prob.	CPU (Sec.)
1	25	167313000	0.86	0.25	119.5
	50	167341000	0.95	0.3	62.1
	75	167341000	0.95	0.3	115.3
2	25	167557000	1.11	0.25	122.6
	50	167587000	1.25	0.3	63
	75	167745000	1.93	0.75	144.4
3	25	168866000	0.71	0.25	205.2
	50	168882000	0.79	0.3	117.9
	75	168898000	0.81	0.31	316.3
4	25	168673000	0.75	0.25	238.9
	50	168689000	0.82	0.3	209.9
	75	168987000	1.3	0.75	638.2
5	25	169962000	0.7	0.25	359.9
	50	169978000	0.77	0.3	288.3
	75	169978000	0.77	0.3	956.4
6	25	177035000	1.48	0.25	500.8
	50	177067000	1.69	0.3	378.9
	75	177561000	4.07	0.75	1114.6
7	25	170575000	0.81	0.25	669
	50	170599000	0.87	0.3	532.6
	75	170599000	0.87	0.3	853.7
8	25	169681000	1.03	0.25	893.2
	50	169702000	1.12	0.3	683
	75	170405000	2.71	0.73	1598.4
9	25	170340000	1.02	0.25	988.5
	50	170354000	1.07	0.3	752.5
	75	170354000	1.07	0.3	2202.5

### 3.7.3 Computational Performance of the Solution Approach

The decomposition scheme applied to the deterministic MILP problem plays a vital role in the performance of the developed Benders decomposition algorithm. Therefore, we experiment to identify the decomposition scheme that imposes the least computational burden to solve the proposed CTSP problem. For this purpose, we solve the deterministic MILP problem for each test instance by the Benders decomposition algorithm with the following decomposition schemes respectively:

1. CPLEX's default decomposition scheme
2. Single-cut decomposition scheme
3. Partially-aggregated-cut decomposition scheme with three clusters of scenarios

4. Partially-aggregated-cut decomposition scheme with five clusters of scenarios
5. Partially-aggregated-cut decomposition scheme with seven clusters of scenarios
6. Multi-cut decomposition scheme

The corresponding computation times are summarized in Table 3.6.

According to Table 3.6 the multi-cut decomposition scheme leads to the fastest Benders decomposition algorithm. More precisely, the average computation time of the algorithm equipped with the multi-cut decomposition scheme is 70.7%, 77.9%, 54.4%, 37.9%, and 29.6% less than that of the algorithm equipped with decomposition schemes 1 to 5, respectively. This is because the multi-cut decomposition scheme results in the smallest linear subproblems that can be solved efficiently in parallel. It should be noted that the CPLEX's default decomposition scheme is the second worst decomposition scheme in terms of CPU time. This is because it fails to decompose the CTSP model according to its two-stage structure, i.e., it cannot include all first-stage decision variables in the master problem since the model contains both first-stage binary and continuous decision variables. As a result, the subproblem may not be further decomposed under the CPLEX's default decomposition scheme.

Finally, we compare the performance of the Benders decomposition algorithm, equipped with the multi-cut decomposition scheme, with the performance of the CPLEX's default algorithm. in Table 3.7. In this table, column "Gap" reports the difference between two computation times divided by the maximum computation time, and NA indicates that the CPLEX's default algorithm fails to find an optimal solution within the 40-hour limit. According to the Gap values, the developed Benders decomposition algorithm performs significantly better than the CPLEX counterpart. The average computation time is 98.5% less for the Benders decomposition algorithm, and it also finds the optimal solution when the CPLEX's default algorithm fails to do so within a 40-hour time limit.

Table 3.6: The comparison of the decomposition schemes

Test No.	CPU (Sec.)					
	1	2	3	4	5	6
1	281.2	317.6	179.3	148.6	151.2	111.5
2	279.1	547	247.3	184.6	168.4	100.8
3	678	890.6	466.6	334.4	290.9	210.6
4	989.6	1653.6	786.2	535.4	469	341.1
5	1529.1	2066.3	815.3	642.5	521.6	427.1
6	3964.9	4613.1	1831.6	1290.3	1042.5	588.2
7	3708.6	4326.6	1848.6	1396.5	1195.8	888.3
8	3067.2	4149.2	2093.8	1501.5	1346.1	1148.9
9	5468.5	6665.2	4171.6	2597.6	2300	1304.1
<b>Average</b>	<b>2218.5</b>	<b>2803.2</b>	<b>1382.3</b>	<b>959</b>	<b>831.7</b>	<b>569</b>

Table 3.7: Benders decomposition algorithm's performance versus the default CPLEX performance

Test No.	CPLEX (Sec.)	Benders (Sec.)	Gap (%)
1	3300.1	111.5	96.62
2	11514.7	100.8	99.12
3	68889.7	210.6	99.69
4	NA	341.1	NA
5	NA	427.1	NA
6	NA	588.2	NA
7	NA	888.3	NA
8	NA	1148.9	NA
9	NA	1304.1	NA

For our case study, the Benders decomposition algorithm converges pretty fast. At each iteration of this algorithm, the objective value of the master problem provides an upper bound on the CTSP model while its lower bound is obtained after having solved the scenario subproblems. This algorithm is terminated when the difference between the upper and lower bounds is less than or equal to a predefined tolerance parameter. For example, the Benders decomposition algorithm solved the case study problem to optimality after 1653 iterations. Table 3.8 shows a few lower and upper bounds of the case study problem obtained by this algorithm. In Table 3.8, the first column titled "Iteration No." shows the algorithm iteration number

in which the bounds are obtained and the last column “Gap” reports the relative gap between the lower bound and upper bound at each iteration. Note that Gap in iteration 1653, which is the last iteration, is almost zero.

Table 3.8: Lower bounds and upper bounds of the case study problem obtained by the Benders decomposition algorithm

Iteration No.	Lower bound	Upper bound	Gap (%)
684	165669000	166491000	0.49
1130	165761000	166491000	0.44
1240	165761000	166433000	0.4
1269	165761000	166409999	0.39
1275	165761000	166393295	0.38
1299	165761000	166376593	0.37
1302	165761000	166359896	0.36
1313	165761000	166343201	0.35
1328	165828000	166339000	0.31

### 3.8 Discussions

Black swan events such as the COVID-19 pandemic have amplified the supply chain risk of NdFeB magnets whose supply has been dominated by a few countries. Therefore, it is crucial to develop domestic sources for NdFeB magnet that is resilient to such uncertainties. We design a profitable and resilient reverse supply chain and logistics network that is multi-product, multi-period, and multi-echelon through developing a CTSP model. This model optimally locates the facilities and determines their optimal processing capacities as first-stage decisions. Then, it determines optimal material flows and inventory levels for each period of the planning horizon. The CTSP model uses multiple strategies, such as buffer inventories, dynamic material and inventory adjustments, facility diversification, penalization of unsatisfied demand, and enforcement of minimum service level to guarantee the network resilience against potential risks. We also design scenarios for the CTSP model that mimic the unique features of large-scale disruptions, such as different disruptions rates, non-linear recovery rates, and the rebound effect on NdFeB magnet demand.

Since CPLEX cannot solve the model to optimality in large-scale instances, we develop a computationally efficient Benders decomposition algorithm, equipped with a multi-cut decomposition scheme, that reduces the CPLEX's computation time by  $\sim 98.5\%$ .

### 3.8.1 Managerial Implications

By applying the CTSP model to the US case study, we gain valuable insights about implementing a NdFeB magnet recycling business in this country. For example, the optimal solution is directed at constructing three dismantling centers, each in Nevada, Texas, and Tennessee, and two recycling centers, each in Nevada and Texas. As the business is at an early stage of development, there are ample opportunities to configure the supply chain and logistics network that is resilient to disruptive events including the COVID-19 pandemic. Based on the research findings, we gain the following managerial insights:

- It is crucial to strategically plan the reverse supply chain and logistics network, as the capital investment constitutes a major cost in the initial phase of the business development. Supply chain and logistics operations managers shall select candidate facility locations based on multiple criteria (e.g., proximity to demand and supply, resource availability, and costs) and apply the proposed methods to determine the optimal solutions. The case study revealed that investing  $\sim 1\%$  of the total cost in warehousing for inventory could satisfy over 99% of the total demand. As such, the proposed method will save costs and improve business resiliency over the long term.
- There is a trade off between profit and resiliency: the smaller risk tolerance, the higher cost to maintain the higher service level. Supply chain and logistics operations managers shall engage with multiple stakeholders in each echelon of the supply chain to assess their risk tolerance level and determine the optimal value through simulation with the proposed methods.

- Unsatisfied NdFeB magnet demand is mainly caused by insufficient supplies of EOL HDDs and EV motors. As the demand for clean energy is projected to increase substantially in the next few decades, supply chain and logistics operations decision makers shall delve into collecting more EOL products from diverse sources (e.g., wind turbines) in the future.

### 3.9 Supplement to Section 3.2: Classification of Relevant Studies

Tables 3.10 - 3.12 summarize the studies that considered both operational and disruption risks in their quantitative models. Table 3.9 describes the abbreviations used in these tables. Moreover, symbol “-” in column “Solution approach” means the original mathematical model is solved by just an optimization solver.

Table 3.9: The description of abbreviations used in the literature review tables

Abbreviation	Description
CTSP	Chance-constrained two-stage stochastic programming
MSP	Multi-stage stochastic programming
TSP	Two-stage stochastic programming
SRO	Scenario-based robust optimization
IRO	Interval robust optimization
RPP	Robust possibilistic programming
PP	Possibilistic programming
L	Linear
PL	Piece-wise linear
C	Complete
P	Partial
F	Forward
R	Reverse
D	Demand
CT	Transportation cost
AF	Availability of facilities
AT	Availability of transportation link
SC	Supply capacity
RQ	Return quantity
QP	Quality of products
PC	Production capacity
TC	Transmission capacity
EP	Environment parameters
SP	Social parameters
HC	Holding capacity
SQ	Supply quantity

### **3.10 Supplement to Section 3.6: Supplementary data**

Supplementary data associated with this study is available online at <https://doi.org/>.

Table 3.10: Classification of the studies considering both disruption and operational risks

Reference	Multi-period	Multi-product	Risk management strategy	Disruption type	Network flow	Scenario Modeling			Uncertainty modeling*	Uncertain parameters	Solution approach**
						Period	Recovery process	Demand rebound			
Lin and Wang (2011)	×	✓	postponement with downward substitution, centralized stocking, and supplier sourcing base	C	F	-	-	-	TSP	D, SC	Exact
Baghalian et al. (2013)	×	✓	-	C	F	-	-	-	SRO	AF, AT	Heuristics
Hatefi and Jolai (2014)	×	×	-	C	F, R	-	-	-	IRO	D, QP, RQ	-
Hatefi et al. (2015)	×	×	Goods-sharing between reliable and unreliable facilities	P	F, R	-	-	-	PP	D, CT	-
Sadghiani et al. (2015)	×	×	Multiple sourcing	C, P	F	-	-	-	PP	D, SC, AT	-
Torabi et al. (2015)	×	×	Multiple sourcing, Prepositioning emergency inventory, Backup supplier, Business continuity plan	P	F	-	-	-	PP	D, SC, All cost & price	Meta-heuristics
Torabi et al. (2016)	×	×	-	C, P	F, R	-	-	-	PP	D, SC, Cost, RQ	-

Table 3.11: Classification of studies (Continued)

Reference	Multi-period		Risk management strategy	Disruption type	Network flow	Scenario Modeling			Solution approach**		
	Multi-product	period				Disruption Period	Recovery process	Demand rebound		Uncertainty modeling*	Uncertain parameters
Hasani and Khosrojerdi (2016)	✓	✓	Fortification of suppliers, Using multiple sourcing, Pre-positioning emergency inventory	P	F	-	-	-	IRO	D, SC, Procurement Cost	Meta-heuristics
Jabbarzadeh et al. (2016)	×	×	Fortification of facilities, goods sharing between reliable and unreliable facilities	C,P	F	-	-	-	IRO	D, SC	Exact
Fattahi et al. (2017)	✓	✓	Customer reassignment, Facility fortification, Extra capacity, Backup	C,P	F	-	L	-	MSP	D, SC, Recovery Cost	-
Khalili et al. (2017)	×	×	supplier, Prepositioning emergency inventory	C,P	F,R	-	-	-	PP	D, PC, TC	-
Hamidieh and Fazli-Khalaf (2017)	×	×	-	C,P	F	-	-	-	PP	SC	-
Zahiri et al. (2017)	✓	✓	Multiple sourcing, Managing flow/node complexity	P	F	-	-	-	PP	EP, SP, D, SC All cost & price.	Meta-heuristics
Fattahi and Govindan (2018)	✓	✓	-	C,P	F	-	L	-	MSP	HC, SQ, Recovery cost	Exact

Table 3.12: Classification of studies (Continued)

Reference	Scenario						Modeling				Solution approach**
	Multi-period	Multi-product	Risk management strategy	Disruption type	Network flow	Disruption Period	Recovery process	Demand rebound	Uncertainty modeling*	Uncertainty parameters	
Namdar et al. (2018)	×	×	Multiple sourcing, Contracting with backup suppliers	P	F	-	-	-	TSP	SC, Lost-sale cost, Market price	-
Sabouhi et al. (2018)	×	×	Multiple sourcing, Prepositioning emergency inventory	P	F	-	-	-	PP	SC	-
Ahranjani et al. (2020)	✓	✓	-	C,P	F	-	-	-	RPP	SQ, All cost & price	-
Hosseini-Motlagh et al. (2020)	✓	✓	-	C,P	F	-	-	-	SRO	D, SC	Exact
This study	✓	✓	Dynamic materialflow, Dynamic Buffer Inventory, Dynamic facility-customer allocation	C,P	R	✓	PL	✓	CTSP	D, SC, Recovery time	Exact

## CHAPTER 4

Data-Driven Robust Optimization Using Scenario-Induced  
Uncertainty Sets

In this chapter, we propose a systematic approach to develop data-driven polyhedral uncertainty sets that alleviate certain drawbacks of conventional polyhedral uncertainty sets. We provide a review of robust optimization (RO) in Section 4.1. In Section 4.2, we present a concise background on conventional polyhedral uncertainty sets. In Section 4.3, after introducing the principal component analysis (PCA) technique, we propose an efficient approach to construct polyhedral scenario-induced uncertainty sets. In section 4.4, we derive a theoretical bound on the gap between the optimal value of a static RO problem with a scenario-induced uncertainty set and that of its lower bound. In Section 4.5, we elaborate on deriving lower bounds on the number of scenario samples required to achieve the desired probabilistic performance guarantees for the developed uncertainty sets. In Section 4.6, we conduct extensive computational experiments on RO Knapsack and power grid problems with the proposed uncertainty sets to evaluate their performance. Finally, Section 4.7 contains concluding remarks.

## 4.1 Introduction

Decision-making in real-world problems is challenging due to the uncertainty involved in them. The challenge is even more significant when the uncertainty is high-dimensional. To overcome this challenge, researchers have proposed various optimization techniques that enable decision-makers to include some knowledge of

the uncertainty into their decision-making process to optimize the trade-off between risk and reward. RO is one of these techniques and seeks an optimal solution that is feasible for all realizations within an uncertainty set. RO assumes that all realizations of uncertainty is prescribed by given uncertainty set and hedges against the worst-case scenario in the set (Ben-Tal and Nemirovski 1998, Bertsimas and Sim 2004).

RO has gained increasing popularity over the last two decades because: (i) it considers uncertainties in the absence of explicit knowledge about their probability distributions; and (ii) its models are usually more tractable than other optimization under uncertainty techniques. Indeed, RO is commonly used in various areas, including but not limited to inventory management, energy management, revenue management, network design, and finance (Bertsimas and Thiele 2006). For a detailed review of RO, we refer interested readers to Ben-Tal and Nemirovski (2008), Ben-Tal et al. (2009), Bertsimas et al. (2011), Gabrel et al. (2014), and Sözüer and Thiele (2016), which provide comprehensive surveys of the RO-related studies.

Uncertainty sets are the core of RO models and play a key role in their performance, greatly impacting solution quality and computational tractability. A well-constructed uncertainty set typically should: (i) capture the most significant aspects of the underlying uncertainty; (ii) be computationally tractable; and (iii) balance robustness and conservativeness of the solution (Lorca and Sun 2014). In other words, the uncertainty set should be large enough to include any true realization of uncertainty with high confidence and small enough to exclude pathological scenarios. Since the introduction of RO by Soyster (1973), several popular uncertainty sets have been proposed and analyzed. Among them, polyhedral uncertainty sets are the most widely used uncertainty sets due to their computational advantages in deriving linear robust counterparts (Lappas and Gounaris 2016). Moreover, certain polyhedral uncertainty sets can capture key features of uncertainty, such as asymmetry and correlation, due to their flexibility in including uncertainty data by adjusting their hyperplanes (Ning and You 2018a). The box and budget uncertainty

sets are two popular types of polyhedral uncertainty sets. Soyster (1973) proposed the box uncertainty set in which each uncertain parameter belongs to a range, while Bertsimas and Sim (2004) introduced the budget uncertainty set where the number of the uncertain parameters that are allowed to vary from their nominal values is limited to a pre-specified budget.

Data-driven RO has provided an efficient alternative to traditional decision-making under uncertainty techniques. As a combination of robust and data-driven frameworks, data-driven RO injects a given set of historical data or scenarios into the model through different methods such as constructing a data-driven uncertainty set (Bertsimas and Thiele 2006).

In more recent literature, machine learning techniques have been adopted to develop data-driven uncertainty sets. For instance, Ning and You (2018a,b) and Dai et al. (2020) proposed hybrid methods to construct data-driven uncertainty sets by combining the robust kernel density estimation and PCA methods. In other examples, Shang et al. (2017), Zhao et al. (2019), Qiu et al. (2019), Shen et al. (2020), and Mohseni and Pishvaei (2020) developed data-driven uncertainty sets using the support vector clustering (SVC) method. Despite the growing popularity of these approaches, there are still many practical limitations. For example, SVC suffers from the curse of dimensionality when uncertainty is high-dimensional (Scott 2015).

In this study, we develop data-driven polyhedral uncertainty sets using PCA. The proposed scenario-induced uncertainty sets have computational benefits for static RO and adaptive robust optimization (ARO) problems by leveraging only a small number of the principal components of uncertainty data while maintaining high solution quality. The contributions of this chapter can be summarized as follows.

1. We use PCA to propose a systematic approach for developing data-driven polyhedral uncertainty sets that alleviate the disadvantages of conventional polyhedral uncertainty sets. Our proposed uncertainty sets efficiently capture

the correlations among uncertainties.

2. We quantify the quality of the lower bound of a static RO problem with a scenario-induced uncertainty set by deriving a theoretical bound on the gap between the optimal value of this problem and that of its lower bound.
3. We provide probabilistic guarantees for the performance of the proposed uncertainty sets by deriving explicit lower bounds on the number of scenarios.

### Notation

In this chapter, we denote scalar values by non-bold symbols, e.g.,  $m_1$ , while we represent vectors by bold symbols in the column form (e.g.,  $\mathbf{u} = (u_1, \dots, u_m)^\top$  and  $\mathbf{w}$ ). Similarly, we denote a matrix by bold capital symbols (e.g.,  $\mathbf{X}$ ) and indicate its size by  $r \times c$ , where  $r$  and  $c$  demonstrate the numbers of rows and columns, respectively. Italic subscripts represent indices, e.g.,  $c_g$ , while non-italic subscripts indicate simplified specifications, e.g.,  $\mathcal{U}_{\text{box}}$ . Symbol  $\|\cdot\|$  denotes the Euclidean norm and  $|\cdot|$  indicates absolute value. We use  $[G]$  to represent the set  $\{1, 2, \dots, G\}$  for any positive integer number  $G$ . We reserve symbol  $[a, b]$  to represent a range whose minimum and maximum values are  $a$  and  $b$ , respectively. The Euler number is indicated by  $e$  while  $\mathbf{e}_i$  represents a vector with all zero elements, except for the  $i^{\text{th}}$  element. Symbol  $F(\cdot)$  indicates the cumulative distribution function of a variable and  $\prod$  represents the operator for the product of a sequence. The number of uncertain parameters, i.e., the size of random variable vector, is denoted by  $m$  and  $\mathbf{u} = (u_1, \dots, u_m)^\top \in \mathbb{R}^m$  represents the random variable vector. We adopt  $N$  to denote the number of available scenarios for  $\mathbf{u}$ . We reserve symbol  $\mathcal{S}$  to represent the set of the  $N$  scenarios, where each scenario is denoted by  $\mathbf{s}_j \in \mathbb{R}^m$ , i.e.,  $\mathbf{s}_j \in \mathcal{S}, \forall j \in [N]$ . The number of utilized principal components in the scenario-induced uncertainty sets is indicated by  $m_1$ . Symbol  $\lceil x \rceil$  represents the smallest integer that is not smaller than  $x$ . Symbol  $\text{unif}(0, 1)$  stands for uniform distribution over the interval  $[0, 1]$ .

## 4.2 Polyhedral Uncertainty Sets

Polyhedral uncertainty sets are widely used in RO because they have a flexible structure to model uncertainty. A general polyhedron uncertainty set is defined as the intersection of closed half-spaces that are represented by linear inequalities of uncertain parameters. More specifically,  $\mathcal{U}_{\text{poly}} = \{\mathbf{u} : \mathbf{A}_i \mathbf{u} \leq b_i, \forall i \in [I]\}$  represents the general formulation of the polyhedral uncertainty set, where  $\mathbf{A}_i$  and  $b_i$  are the coefficients of its  $i^{\text{th}}$  linear inequality.

The box and budget uncertainty sets are two special cases of  $\mathcal{U}_{\text{poly}}$ . Soyster (1973) introduced  $\mathcal{U}_{\text{gen}} = \{\mathbf{u} : a_i \leq u_i \leq b_i, \forall i \in [m]\}$  as a general box uncertainty set. Alternatively, box uncertainty set can also be defined as follows:

$$\mathcal{U}_{\text{box}} = \{\mathbf{u} : u_i = \bar{u}_i + z_i \hat{u}_i, -1 \leq z_i \leq 1, \forall i \in [m]\},$$

where  $\bar{u}_i$  represents the nominal value of  $u_i$  and  $\hat{u}_i$  denotes the largest possible deviation of  $u_i$ , i.e.,  $u_i$  belongs to range  $[\bar{u}_i - \hat{u}_i, \bar{u}_i + \hat{u}_i]$ . Bertsimas and Sim (2004) introduced a budget uncertainty set defined as follows:

$$\mathcal{U}_{\text{budget}} = \{\mathbf{u} : u_i = \bar{u}_i + z_i \hat{u}_i, -1 \leq z_i \leq 1, \sum_{i=1}^m |z_i| \leq \Gamma, \forall i \in [m]\},$$

where parameter  $\Gamma \in [0, m]$  can be used as a tool to trade-off the conservativeness and robustness of RO models with  $\mathcal{U}_{\text{budget}}$ . Indeed,  $\Gamma = 0$  yields the nominal problem, which does not incorporate any uncertainty, while  $\Gamma = m$  results in the most conservative problem, in which  $u_i$  is allowed to deviate between its maximum and minimum value. Uncertainty set  $\mathcal{U}_{\text{box}}$  is a special case of  $\mathcal{U}_{\text{budget}}$  because  $\mathcal{U}_{\text{budget}}$  is equivalent to  $\mathcal{U}_{\text{box}}$  if  $\Gamma = m$ .

Polyhedral uncertainty sets can be constructed based on historical uncertainty data. These uncertainty sets are referred to as data-driven polyhedral uncertainty sets. For example, the following convex hull of  $\mathcal{S}$  (i.e., the smallest convex set that includes all  $N$  scenarios in  $\mathcal{S}$ ) can be considered as a scenario-induced polyhedral

uncertainty set:

$$\mathcal{U}_{\text{conv}}(\mathcal{S}) = \left\{ \mathbf{u} : \mathbf{u} = \sum_{j=1}^N \alpha_j \mathbf{s}_j, \sum_{j=1}^N \alpha_j = 1, 0 \leq \alpha_j \leq 1, \forall j \in [N] \right\}.$$

Figure 4.1 illustrates the  $\mathcal{U}_{\text{conv}}(\mathcal{S})$  constructed by positively correlated, negatively correlated, and uncorrelated scenarios of  $\mathbf{u} = (u_1, u_2)^\top \in \mathbb{R}^2$ . Figure 4.2 shows  $\mathcal{U}_{\text{box}}$ ,  $\mathcal{U}_{\text{budget}}$  with  $\Gamma = 1$ , and  $\mathcal{U}_{\text{conv}}(\mathcal{S})$  together for this random variable vector. In these figures, each blue point indicates a scenario and the gray rectangle, green lozenge, and red polygon represent  $\mathcal{U}_{\text{box}}$ ,  $\mathcal{U}_{\text{budget}}$ , and  $\mathcal{U}_{\text{conv}}(\mathcal{S})$ , respectively.

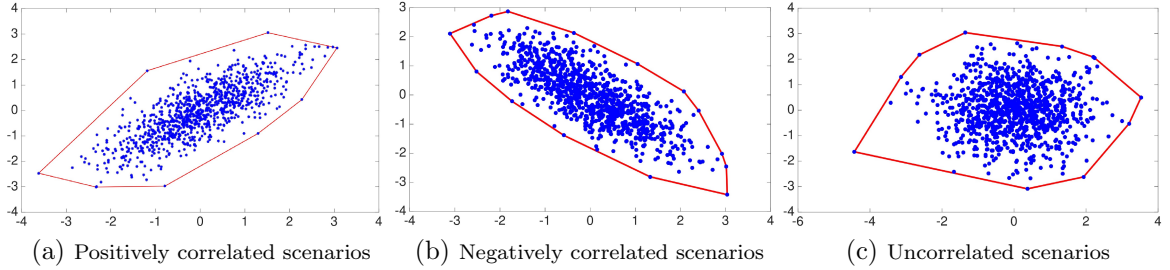


Figure 4.1:  $\mathcal{U}_{\text{conv}}(\mathcal{S})$  of uncertain parameters  $u_1, u_2$

From Figures 4.1 and 4.2, we can observe: (i) Uncertainty set  $\mathcal{U}_{\text{box}}$  is the most conservative among them; (ii) Uncertainty sets  $\mathcal{U}_{\text{box}}$  and  $\mathcal{U}_{\text{budget}}$  cannot capture the correlation information of uncertainty; and (iii) Uncertainty set  $\mathcal{U}_{\text{conv}}(\mathcal{S})$  is the most computationally expensive because it involves many more decision variables than the other two uncertainty sets ( $N$  vs.  $m$ ), leading to larger-size RO formulations. Given these observations, it would be highly desirable to develop data-driven polyhedral uncertainty sets that can capture dependent information of uncertainty, alleviate conservatism, and result in more computationally tractable RO models compared to  $\mathcal{U}_{\text{conv}}(\mathcal{S})$ . To this end, we propose such data-driven polyhedral uncertainty sets in section 4.3.

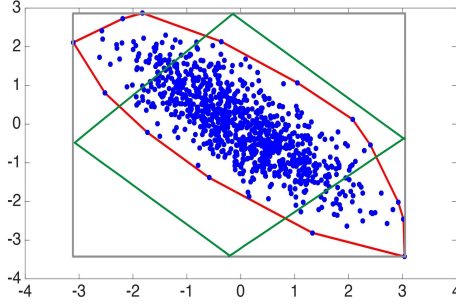


Figure 4.2:  $\mathcal{U}_{\text{box}}$ ,  $\mathcal{U}_{\text{budget}}$  with  $\Gamma = 1$ , and  $\mathcal{U}_{\text{conv}}(\mathcal{S})$  for uncertain parameters  $u_1, u_2$

### 4.3 Scenario-Induced Uncertainty Sets

In this section, we develop scenario-induced polyhedral uncertainty sets by leveraging PCA to alleviate the drawbacks of  $\mathcal{U}_{\text{box}}$ ,  $\mathcal{U}_{\text{budget}}$ , and  $\mathcal{U}_{\text{conv}}(\mathcal{S})$ . The merits of the developed uncertainty sets are as follows. First, they explicitly capture the correlation information of uncertainty. Second, they yield more tractable RO models compared to  $\mathcal{U}_{\text{conv}}(\mathcal{S})$  because their RO formulations are more computationally efficient in comparison with  $\mathcal{U}_{\text{conv}}(\mathcal{S})$ , due to fewer decision variables. Third, they are less conservative than  $\mathcal{U}_{\text{box}}$ . Fourth, a portion of the principal components of data can be used to improve the tractability and conservativeness of RO models at the expense of robustness reduction.

In brief, the proposed data-driven approach used to construct scenario-induced uncertainty sets includes the following steps: (i) Calculating the sample mean vector and sample covariance matrix of  $\mathbf{u}$  based on  $N$  scenarios; (ii) Obtaining the principal directions of uncertainty data by performing the eigenvalue decomposition on the sample covariance matrix; (iii) Projecting centered scenarios onto each principal direction. In the remainder of this section, we elaborate on these steps in more detail.

### 4.3.1 Low-rank Approximation with PCA

The PCA technique enables us to project high-dimensional uncertainty onto a lower-dimensional space by preserving the components with the highest variance. Moreover, it transforms the correlated uncertain parameters into their uncorrelated principal components (Wold et al. 1987). We refer interested readers to Wold et al. (1987) and Reris and Brooks (2015) for more information about PCA.

Let  $\bar{\mathbf{s}} = \frac{1}{N} \sum_{j=1}^N \mathbf{s}_j$  be the sample mean of the uncertainty and  $\mathbf{X} = [s_{ij}]_{N \times m}$  be the uncertainty data matrix, where the  $j^{\text{th}}$  row represents  $\mathbf{s}_j^\top \in \mathbb{R}^{1 \times m}$ ,  $\forall j \in [N]$ . Without loss of generality, we center  $\mathbf{X}$  at the sample mean by subtracting  $\bar{\mathbf{s}}$  from each scenario (row), i.e.,  $\mathbf{s}_{j0} = \mathbf{s}_j - \bar{\mathbf{s}}$ ,  $\forall j \in [N]$ . Therefore, the centered data matrix, denoted by  $\mathbf{X}_0$ , enables us to approximate the covariance matrix of  $\mathbf{u}$  by the sample covariance matrix  $\mathbf{C}$  given by  $\mathbf{C} = \frac{1}{N-1} \mathbf{X}_0^\top \mathbf{X}_0$ .

The PCA technique can be performed by conducting the eigenvalue decomposition (EVD) on  $\mathbf{C}$ . With  $\mathbf{X}_0^\top = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$  as the singular value decomposition of  $\mathbf{X}_0^\top$ , the EAD of  $\mathbf{C}$  is as follows:

$$\mathbf{C} = \frac{1}{N-1} (\mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top) (\mathbf{V}\mathbf{\Sigma}^\top\mathbf{U}^\top) = \mathbf{U} \left( \frac{\mathbf{\Sigma}\mathbf{\Sigma}^\top}{N-1} \right) \mathbf{U}^\top = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top,$$

where  $\mathbf{U} \in \mathbb{R}^{m \times m}$ ,  $\mathbf{V} \in \mathbb{R}^{N \times N}$ , and  $\mathbf{\Sigma} \in \mathbb{R}^{m \times N}$ . The columns of  $\mathbf{U}$  and the diagonal entries of  $\mathbf{\Lambda}$  represent the eigenvectors and eigenvalues of  $\mathbf{C}$ , respectively. The eigenvectors are the principal directions of the centered data, denoted by  $\mathbf{d}_i$ ,  $\forall i \in [m]$ . The eigenvalue related to each eigenvector represents the variance of the centered data along the corresponding principal direction. Without loss of generality, we assume that the eigenvalues are in non-increasing order. Therefore, the first principal directions characterize most of the variance. The projections of the centered data on the principal directions are called principal components, given by  $\mathbf{s}'_{ji} = \frac{\mathbf{s}_{j0} \cdot \mathbf{d}_i}{\|\mathbf{d}_i\|^2} \mathbf{d}_i$ ,  $\forall j \in [N], \forall i \in [m]$ .

To reduce the dimensionality of the centered uncertainty data from  $m$  to  $m_1$ , we preserve only the first  $m_1$  columns of  $\mathbf{U}$  and  $m_1 \times m_1$  upper-left entries of  $\mathbf{\Lambda}$ , which

are related to the principal directions with the largest variance. Since the dropped components play the least important role in characterizing the uncertainty, PCA projects the  $m$ -dimensional uncertainty space onto an  $m_1$ -dimensional space with the least information loss.

### 4.3.2 PCA-based Polyhedral Uncertainty Sets

By applying PCA to  $\mathcal{S}$  according to the steps discussed in subsection 4.3.1, we propose the following scenario-induced uncertainty set:

$$\mathcal{U}_{\text{PCA}}(\mathcal{S}, m_1) = \left\{ \mathbf{u} : \mathbf{u} = \bar{\mathbf{s}} + \sum_{i=1}^{m_1} \left( \alpha_i \left( \frac{\bar{\omega}_i}{\|\mathbf{d}_i\|} \mathbf{d}_i \right) + (1 - \alpha_i) \left( \frac{\omega_i}{\|\mathbf{d}_i\|} \mathbf{d}_i \right) \right) + \sum_{i=m_1+1}^m \frac{\bar{\omega}_i + \omega_i}{2\|\mathbf{d}_i\|} \mathbf{d}_i, 0 \leq \alpha_i \leq 1, \forall i \in [m_1] \right\},$$

where

$$\bar{\omega}_i = \max_{j=1}^N \left\{ \frac{\mathbf{s}_{j0} \cdot \mathbf{d}_i}{\|\mathbf{d}_i\|} \right\} \in \mathbb{R}, \quad \text{and} \quad \omega_i = \min_{j=1}^N \left\{ \frac{\mathbf{s}_{j0} \cdot \mathbf{d}_i}{\|\mathbf{d}_i\|} \right\} \in \mathbb{R},$$

meaning  $\left( \frac{\bar{\omega}_i}{\|\mathbf{d}_i\|} \mathbf{d}_i \right)$  and  $\left( \frac{\omega_i}{\|\mathbf{d}_i\|} \mathbf{d}_i \right)$  are the largest and smallest projected centered scenarios onto the principal direction  $\mathbf{d}_i$ , respectively. The sample mean  $\bar{\mathbf{s}}$  is added to  $\mathcal{U}_{\text{PCA}}(\mathcal{S}, m_1)$  because the scenarios have already been centered at  $\bar{\mathbf{s}}$ .

In Figure 4.3, the blue rectangle, red polygon, green lozenge, and gray rectangle respectively represent  $\mathcal{U}_{\text{PCA}}(\mathcal{S}, 2)$ ,  $\mathcal{U}_{\text{conv}}(\mathcal{S})$ ,  $\mathcal{U}_{\text{budget}}$  with  $\Gamma = 1$ , and  $\mathcal{U}_{\text{box}}$  of  $\mathbf{u} = (u_1, u_2)^\top \in \mathbb{R}^2$  for a set of positively correlated scenarios. According to this figure, we have  $\mathcal{U}_{\text{conv}}(\mathcal{S}) \subseteq \mathcal{U}_{\text{PCA}}(\mathcal{S}, 2)$ , therefore, RO models with  $\mathcal{U}_{\text{PCA}}(\mathcal{S}, m_1 = m)$  are more conservative (robust) than those with  $\mathcal{U}_{\text{conv}}(\mathcal{S})$ . On the other hand,  $\mathcal{U}_{\text{PCA}}(\mathcal{S}, m_1 = m)$  results in more tractable RO models compared to  $\mathcal{U}_{\text{conv}}(\mathcal{S})$  because it involves fewer decision variables in RO models in comparison with  $\mathcal{U}_{\text{conv}}(\mathcal{S})$ .

Figure 4.4 shows  $\mathcal{U}_{\text{PCA}}(\mathcal{S}, 2)$  and  $\mathcal{U}_{\text{PCA}}(\mathcal{S}, 1)$  for the same set of scenarios. In this example,  $\mathbf{d}_1$  and  $\mathbf{d}_2$ , indicated by the green and dashed line respectively, are the principal directions where the most variance of data exists along  $\mathbf{d}_1$ . The uncertainty set  $\mathcal{U}_{\text{PCA}}(\mathcal{S}, 1)$ , which considers  $\mathbf{d}_1$  as the only leading principal direction, is the green

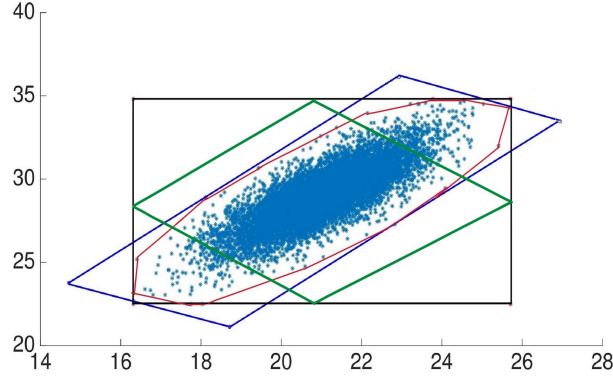


Figure 4.3:  $\mathcal{U}_{\text{PCA}}(\mathcal{S}, 2)$  VS.  $\mathcal{U}_{\text{conv}}(\mathcal{S})$ ,  $\mathcal{U}_{\text{budget}}$  with  $\Gamma = 1$ , and  $\mathcal{U}_{\text{box}}$  for uncertain parameters  $u_1, u_2$

line segment whose endpoints are generated by  $\alpha_1 = 0$  and  $\alpha_1 = 1$  and any value of  $\alpha_1$  between 0 and 1 generates a unique point on this line. As  $\mathcal{U}_{\text{PCA}}(\mathcal{S}, 1)$  considers  $\mathbf{d}_2$  as the non-leading principal direction, it sets  $\alpha_2 = \frac{1}{2}$  to keep only the middle value of  $\mathbf{d}_2$  that is located on the green line. Thus, when  $m_1$  reduces from 2 to 1, the blue rectangle shrinks to the green line.

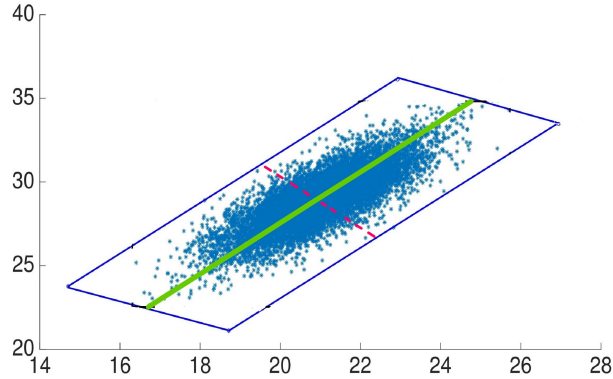


Figure 4.4:  $\mathcal{U}_{\text{PCA}}(\mathcal{S}, 1)$  VS.  $\mathcal{U}_{\text{PCA}}(\mathcal{S}, 2)$  for uncertain parameters  $u_1, u_2$

**Remark 4.1.** A smaller  $m_1$  yields a lower dimensional uncertainty set  $\mathcal{U}_{\text{PCA}}(\mathcal{S}, m_1)$ , which leads to a more tractable and less conservative (robust) RO model. Therefore,  $m_1$  can be used as tool to trade-off tractability and conservativeness of RO models directly.

**Remark 4.2.** If we choose  $\mathbf{d}_i = \mathbf{e}_i$ ,  $\forall i \in [m]$ , then the uncertainty set  $\hat{\mathcal{U}}_{\text{PCA}}(\mathcal{S}, m) = \{\mathbf{u} : \mathbf{u} \in \prod_{i=1}^m [\bar{\mathbf{s}} + \underline{\omega}_i, \bar{\mathbf{s}} + \bar{\omega}_i]\}$  is a box uncertainty set and a special case of

$\mathcal{U}_{\text{PCA}}(\mathcal{S}, m)$ .

**Remark 4.3.** *The intersection of  $\mathcal{U}_{\text{PCA}}(\mathcal{S}, m_1)$  and  $\hat{\mathcal{U}}_{\text{PCA}}(\mathcal{S}, m_1)$ ,  $\mathcal{U}_{\cap}(\mathcal{S}, m_1) = \mathcal{U}_{\text{PCA}}(\mathcal{S}, m_1) \cap \hat{\mathcal{U}}_{\text{PCA}}(\mathcal{S}, m_1)$ , is another scenario induced uncertainty set.*

#### 4.4 Lower Bound Quality for Static RO

For any  $m_1$  smaller than  $m$ , a minimization RO problem with  $\mathcal{U}_{\text{PCA}}(\mathcal{S}, m_1)$  leads to a lower bound for the same problem with  $\mathcal{U}_{\text{PCA}}(\mathcal{S}, m)$ . Similarly, a maximization RO problem with  $\mathcal{U}_{\text{PCA}}(\mathcal{S}, m_1)$  results in an upper bound for the same problem with  $\mathcal{U}_{\text{PCA}}(\mathcal{S}, m)$ . The smaller  $m_1$  is, the RO problem with  $\mathcal{U}_{\text{PCA}}(\mathcal{S}, m_1)$  yields a looser bound that is more computationally tractable. In this section, we quantify the quality of the lower bound of a static minimization RO problem with  $\mathcal{U}_{\text{PCA}}(\mathcal{S}, m)$  by deriving a theoretical bound on the gap between its optimal value and the optimal value of its lower bound. Consider the following static RO problem

$$Z^*(m) := \min_{\mathbf{x} \in \mathcal{X}} \max_{\substack{\mathbf{u} \in \mathcal{U}_{\text{PCA}} \\ (\mathcal{S}, m)}} f(\mathbf{x}, \mathbf{u}), \quad (4.1)$$

whose lower bound is

$$Z^*(m_1) := \min_{\mathbf{x} \in \mathcal{X}} \max_{\substack{\mathbf{u} \in \mathcal{U}_{\text{PCA}} \\ (\mathcal{S}, m_1)}} f(\mathbf{x}, \mathbf{u}). \quad (4.2)$$

The following theorem provides a theoretical bound on the solution quality of problem (4.2).

**Theorem 4.1.** *When  $f(\mathbf{x}, \mathbf{u})$  is piecewise linear convex in  $\mathbf{u}$ , i.e.,  $f(\mathbf{x}, \mathbf{u}) = \max_{k=1}^K \{y_k^0(\mathbf{x}) + y_k(\mathbf{x})^\top \mathbf{u}\}$  with both  $y_k(\mathbf{x}) = (y_k^1(\mathbf{x}), \dots, y_k^m(\mathbf{x}))^\top$  and  $y_k^0(\mathbf{x})$  affine in  $\mathbf{x}$  for any  $k \in [K]$ , it holds that*

$$0 \leq Z^*(m) - Z^*(m_1) \leq \max_{k=1}^K \sum_{i=m_1+1}^m |y_k(\hat{\mathbf{x}})^\top \bar{\mathbf{d}}_i| \left( \frac{\bar{\omega}_i - \omega_i}{2} \right),$$

where  $\hat{\mathbf{x}}$  is an optimal solution of the RO problem with  $\mathcal{U}_{\text{PCA}}(\mathcal{S}, m_1)$ , i.e., Problem (4.2), and  $\bar{\mathbf{d}}_i = \frac{\mathbf{d}_i}{\|\mathbf{d}_i\|}$ ,  $\forall i \in \{m_1 + 1, \dots, m\}$ .

*Proof.* Since Problem (4.2) is a lower bound of Problem (4.1), it is trivial that  $Z^*_M(m) - Z^*_M(m_1) \geq 0$ . In what follows, we derive the upper bound of the gap. Problem (4.2) can be rewritten as

$$\min_{\mathbf{x} \in \mathcal{X}} \max_{\substack{\mathbf{u} \in \mathcal{U}_{\text{PCA}} \\ (\mathcal{S}, m_1)}} \max_{k=1}^K y_k^0(\mathbf{x}) + y_k(\mathbf{x})^\top \mathbf{u},$$

which is equivalent to

$$\min_{\mathbf{x} \in \mathcal{X}} \max_{k=1}^K \max_{\substack{\mathbf{u} \in \mathcal{U}_{\text{PCA}} \\ (\mathcal{S}, m_1)}} y_k^0(\mathbf{x}) + y_k(\mathbf{x})^\top \mathbf{u}. \quad (4.3)$$

For clarity, we define  $\bar{\mathbf{d}}_i = \frac{\mathbf{d}_i}{\|\mathbf{d}_i\|}$ ,  $\forall i \in [m]$ . With this definition,  $\mathbf{u}$  in  $\mathcal{U}_{\text{PCA}}(\mathcal{S}, m_1)$  is rewritten as

$$\mathbf{u} = \bar{\mathbf{s}} + \sum_{i=1}^{m_1} (\alpha_i \bar{\omega}_i \bar{\mathbf{d}}_i + (1 - \alpha_i) \underline{\omega}_i \bar{\mathbf{d}}_i) + \sum_{i=m_1+1}^m \frac{\bar{\omega}_i + \underline{\omega}_i}{2} \bar{\mathbf{d}}_i, \quad 0 \leq \alpha_i \leq 1, \quad \forall i \in [m_1]. \quad (4.4)$$

By plugging (4.4) to (4.3), Problem (4.3) is reformulated as the following problem:

$$\min_{\mathbf{x} \in \mathcal{X}} \max_{k=1}^K \max_{\substack{0 \leq \alpha_i \leq 1 \\ \forall i \in [m_1]}} y_k^0(\mathbf{x}) + y_k(\mathbf{x})^\top \left[ \bar{\mathbf{s}} + \sum_{i=1}^{m_1} (\alpha_i \bar{\omega}_i \bar{\mathbf{d}}_i + (1 - \alpha_i) \underline{\omega}_i \bar{\mathbf{d}}_i) + \sum_{i=m_1+1}^m \frac{\bar{\omega}_i + \underline{\omega}_i}{2} \bar{\mathbf{d}}_i \right].$$

Similarly, Problem (4.1) is formulated as the following problem:

$$\min_{\mathbf{x} \in \mathcal{X}} \max_{k=1}^K \max_{\substack{0 \leq \alpha_i \leq 1 \\ \forall i \in [m]}} y_k^0(\mathbf{x}) + y_k(\mathbf{x})^\top \left[ \bar{\mathbf{s}} + \sum_{i=1}^m (\alpha_i \bar{\omega}_i \bar{\mathbf{d}}_i + (1 - \alpha_i) \underline{\omega}_i \bar{\mathbf{d}}_i) \right].$$

Let  $\mathbf{x}^*$  and  $\hat{\mathbf{x}}$  be an optimal solution of Problems (4.1) and (4.2), respectively. For clarity, we define

$$h(\mathbf{x}^*, \hat{\alpha}_m) = y_k^0(\mathbf{x}^*) + y_k(\mathbf{x}^*)^\top \left[ \bar{\mathbf{s}} + \sum_{i=1}^m (\alpha_i \bar{\omega}_i \bar{\mathbf{d}}_i + (1 - \alpha_i) \underline{\omega}_i \bar{\mathbf{d}}_i) \right],$$

$$g(\hat{\mathbf{x}}, \hat{\alpha}_{m_1}) = y_k^0(\hat{\mathbf{x}}) + y_k(\hat{\mathbf{x}})^\top \left[ \bar{\mathbf{s}} + \sum_{i=1}^{m_1} (\alpha_i \bar{\omega}_i \bar{\mathbf{d}}_i + (1 - \alpha_i) \underline{\omega}_i \bar{\mathbf{d}}_i) + \sum_{i=m_1+1}^m \frac{\bar{\omega}_i + \underline{\omega}_i}{2} \bar{\mathbf{d}}_i \right],$$

where  $\hat{\alpha}_m = \{\alpha_1, \dots, \alpha_m\}$  and  $\hat{\alpha}_{m_1} = \{\alpha_1, \dots, \alpha_{m_1}\}$ . With the definitions, we have

$$Z^*(m) - Z^*(m_1) = \max_{k=1}^K \max_{\substack{0 \leq \alpha_i \leq 1 \\ \forall i \in [m]}} h(\mathbf{x}^*, \hat{\alpha}_m) - \max_{k=1}^K \max_{\substack{0 \leq \alpha_i \leq 1 \\ \forall i \in [m_1]}} g(\hat{\mathbf{x}}, \hat{\alpha}_{m_1}).$$

Since  $\hat{\mathbf{x}}$  is a feasible solution of Problem (4.1) as well, we have

$$\begin{aligned}
\Leftrightarrow Z^*(m) - Z^*(m_1) &\leq \max_{k=1}^K \max_{\substack{0 \leq \alpha_i \leq 1 \\ \forall i \in [m]}} h(\hat{\mathbf{x}}, \hat{\alpha}_m) - \max_{k=1}^K \max_{\substack{0 \leq \alpha_i \leq 1 \\ \forall i \in [m_1]}} g(\hat{\mathbf{x}}, \hat{\alpha}_{m_1}), \\
\Leftrightarrow Z^*(m) - Z^*(m_1) &\leq \max_{k=1}^K \max_{\substack{0 \leq \alpha_i \leq 1 \\ \forall i \in [m]}} y_k^0(\hat{\mathbf{x}}) + y_k(\hat{\mathbf{x}})^\top \left[ \bar{\mathbf{s}} + \sum_{i=1}^m (\alpha_i \bar{\omega}_i \bar{\mathbf{d}}_i + (1 - \alpha_i) \underline{\omega}_i \underline{\mathbf{d}}_i) \right] \\
&\quad - \max_{k=1}^K \max_{\substack{0 \leq \alpha_i \leq 1 \\ \forall i \in [m_1]}} \left( y_k^0(\hat{\mathbf{x}}) + y_k(\hat{\mathbf{x}})^\top \left[ \bar{\mathbf{s}} + \sum_{i=1}^{m_1} (\alpha_i \bar{\omega}_i \bar{\mathbf{d}}_i + (1 - \alpha_i) \underline{\omega}_i \underline{\mathbf{d}}_i) + \sum_{i=m_1+1}^m \frac{\bar{\omega}_i + \underline{\omega}_i}{2} \bar{\mathbf{d}}_i \right] \right) \\
\Leftrightarrow Z^*(m) - Z^*(m_1) &\leq \max_{k=1}^K \max_{\substack{0 \leq \alpha_i \leq 1 \\ \forall i \in \{m_1+1, \dots, m\}}} y_k(\hat{\mathbf{x}})^\top \bar{\mathbf{d}}_i \sum_{i=m_1+1}^m \left( \alpha_i \bar{\omega}_i + (1 - \alpha_i) \underline{\omega}_i - \frac{\bar{\omega}_i + \underline{\omega}_i}{2} \right) \\
\Leftrightarrow Z^*(m) - Z^*(m_1) &\leq \max_{k=1}^K \max_{\substack{0 \leq \alpha_i \leq 1 \\ \forall i \in \{m_1+1, \dots, m\}}} y_k(\hat{\mathbf{x}})^\top \bar{\mathbf{d}}_i \sum_{i=m_1+1}^m \left( (\alpha_i - \frac{1}{2}) \bar{\omega}_i + (\frac{1}{2} - \alpha_i) \underline{\omega}_i \right)
\end{aligned} \tag{4.5}$$

$$\begin{aligned}
\Leftrightarrow Z^*(m) - Z^*(m_1) &\leq \max_{k=1}^K \sum_{i=m_1+1}^m \max \left\{ y_k(\hat{\mathbf{x}})^\top \bar{\mathbf{d}}_i \left( -\frac{1}{2} \bar{\omega}_i + \frac{1}{2} \underline{\omega}_i \right), y_k(\hat{\mathbf{x}})^\top \bar{\mathbf{d}}_i \left( \frac{1}{2} \bar{\omega}_i - \frac{1}{2} \underline{\omega}_i \right) \right\}
\end{aligned} \tag{4.6}$$

$$\begin{aligned}
\Leftrightarrow Z^*(m) - Z^*(m_1) &\leq \max_{k=1}^K \sum_{i=m_1+1}^m |y_k(\hat{\mathbf{x}})^\top \bar{\mathbf{d}}_i| \left( \frac{\bar{\omega}_i - \underline{\omega}_i}{2} \right).
\end{aligned}$$

Note that (4.6) is equivalent to (4.5) because the inner maximization problem in (4.5) is linear. Therefore, its optimal solution is one of the extreme points of  $0 \leq \alpha_i \leq 1$ , i.e., either  $\alpha_i = 0$  or  $\alpha_i = 1$ .  $\square$

The theoretical upper bound developed in Theorem 4.1 brings two benefits: (i) it provides a rough approximation for the optimal value of Problem (4.1), which may not be solved efficiently in practice; and (ii) it determines how many principal components are required to reach a preferred gap, demonstrating a trade-off between computational burden and solution quality.

## 4.5 Probabilistic Guarantees

In this section, we derive probabilistic guarantees for the performance of  $\mathcal{U}_{\text{PCA}}(\mathcal{S}, m_1)$  when all the principal components are utilized to construct these uncertainty sets, i.e.,  $m_1 = m$ . To that end, we develop explicit lower bounds on the number of scenario samples required to construct these sets with desired probabilistic performance. To derive the probabilistic guarantees, we consider no assumptions on the probability distribution of uncertainty data.

**Theorem 4.2.** *If  $N \geq N^*(m) = \lceil \frac{1}{\epsilon} \frac{e}{e-1} (2m-1 + \ln \frac{1}{\beta}) \rceil$ , then we have  $1-\beta$  confidence that any realization  $\mathbf{s}$  belongs to the uncertainty set  $\mathcal{U}_{\text{PCA}}(\mathcal{S}, m)$  with the probability of at least  $1-\epsilon$ , i.e.,*

$$\mathbb{P}_{\hat{\mathbf{s}}}\{\mathbb{P}_{\mathbf{s}}\{\mathbf{s} \in \mathcal{U}_{\text{PCA}}(\mathcal{S}, m)\} \geq 1-\epsilon\} \geq 1-\beta,$$

where  $0 < \epsilon < 1$ ,  $0 < \beta < 1$ , and  $\hat{\mathbf{s}} = \{\mathbf{s}_1, \dots, \mathbf{s}_N\}$ .

*Proof.* The result is deduced from Margellos et al. (2014) and thus we omit the proof.  $\square$

From Theorem 4.2, when  $m = 1$ ,  $N^*(m)$  becomes  $N^*(1) = \lceil \frac{1}{\epsilon} \frac{e}{e-1} (1 + \ln \frac{1}{\beta}) \rceil$ .

**Theorem 4.3.** *When  $m = 1$ . If  $N \geq \frac{\ln \beta - \ln(1-\epsilon + N^*(1)\epsilon)}{\ln(1-\epsilon)} + 1$ , then we have  $1-\beta$  confidence that any realization  $\mathbf{s}$  belongs to the uncertainty set  $\mathcal{U}_{\text{PCA}}(\mathcal{S}, m)$  with the probability of at least  $1-\epsilon$ , i.e.,*

$$\mathbb{P}_{\hat{\mathbf{s}}}\{\mathbb{P}_{\mathbf{s}}\{\mathbf{s} \in \mathcal{U}_{\text{PCA}}(\mathcal{S}, m)\} \geq 1-\epsilon\} \geq 1-\beta, \quad (4.7)$$

where  $0 < \epsilon < 1$ ,  $0 < \beta < 1$ , and  $\hat{\mathbf{s}} = \{\mathbf{s}_1, \dots, \mathbf{s}_N\}$ .

*Proof.* With  $m = 1$ , we have

$$\begin{aligned} \mathbb{P}_{\mathbf{s}}\{\mathbf{s} \in \mathcal{U}_{\text{PCA}}(\mathcal{S}, 1)\} &= \mathbb{P}_{\mathbf{s}}\{\bar{\mathbf{s}}_j \leq \bar{\mathbf{s}} \leq \bar{\mathbf{s}}_j, \quad \forall j \in [N]\} \\ &= \mathbb{P}_{\mathbf{s}}\{\min_{j=1}^N \bar{\mathbf{s}}_j \leq \bar{\mathbf{s}} \leq \max_{j=1}^N \bar{\mathbf{s}}_j\} \end{aligned}$$

Thus we have a sufficient and necessary condition for  $\mathbb{P}_{\mathbf{s}}\{\mathbf{s} \in \mathcal{U}_{\text{PCA}}(\mathcal{S}, 1)\} \geq 1 - \epsilon$ :

$$\mathbb{P}_{\mathbf{s}}\left\{\min_{j=1}^N \bar{\mathbf{s}}_j \leq \bar{\mathbf{s}} \leq \max_{j=1}^N \bar{\mathbf{s}}_j\right\} \geq 1 - \epsilon.$$

Similarly, we have

$$\begin{aligned} \mathbb{P}_{\mathbf{s}}\left\{\mathbb{P}_{\mathbf{s}}\left\{\min_{j=1}^N \bar{\mathbf{s}}_j \leq \bar{\mathbf{s}} \leq \max_{j=1}^N \bar{\mathbf{s}}_j\right\} \geq 1 - \epsilon\right\} &= \mathbb{P}_{\mathbf{s}}\left\{F_1\left(\max_{j=1}^N \bar{\mathbf{s}}_j\right) - F_1\left(\min_{j=1}^N \bar{\mathbf{s}}_j\right) \geq 1 - \epsilon\right\} \\ &= \mathbb{P}_{\mathbf{s}}\left\{\max_{j=1}^N F_1(\bar{\mathbf{s}}_j) - \min_{j=1}^N F_1(\bar{\mathbf{s}}_j) \geq 1 - \epsilon\right\}, \end{aligned}$$

where the second equality is due to the fact that  $F_1$  is non-decreasing. Thus, to make inequality (4.7) hold, it is equivalent to have

$$\mathbb{P}_{\mathbf{s}}\left\{\max_{j=1}^N F_1(\bar{\mathbf{s}}_j) - \min_{j=1}^N F_1(\bar{\mathbf{s}}_j) \geq 1 - \epsilon\right\} \geq 1 - \beta.$$

As  $F_1(\bar{\mathbf{s}}_j)$  is a random variable with the probability distribution of  $\text{unif}(0, 1)$ , by defining  $\xi_1 = \max_{j=1}^N F_1(\bar{\mathbf{s}}_j)$  and  $\xi_2 = \min_{j=1}^N F_1(\bar{\mathbf{s}}_j)$ , the joint probability density function of order statistics  $\xi_1$  and  $\xi_2$  is  $N(N-1)(\xi_1 - \xi_2)^{N-2}$  when  $\xi_1 \geq \xi_2$  and zero otherwise (Casella and Berger 2021). Consequently, we have

$$\begin{aligned} \mathbb{P}_{\mathbf{s}}\left\{\max_{j=1}^N F_1(\bar{\mathbf{s}}_j) - \min_{j=1}^N F_1(\bar{\mathbf{s}}_j) \geq 1 - \epsilon\right\} &= \int_0^\epsilon \int_{1-\epsilon+\xi_2}^1 N(N-1)(\xi_1 - \xi_2)^{N-2} d\xi_1 d\xi_2 \\ &= 1 - (1 - \epsilon)^N - N(1 - \epsilon)^{N-1} \epsilon. \end{aligned}$$

Moreover,  $1 - (1 - \epsilon)^N - N(1 - \epsilon)^{N-1} \epsilon \geq 1 - \beta$  is equivalent to

$$(1 - \epsilon)^{N-1} (1 - \epsilon + N\epsilon) \leq \beta. \quad (4.8)$$

Thus, to complete the proof, it is sufficient to show  $N = \frac{\ln \beta - \ln(1 - \epsilon + N^*(1)\epsilon)}{\ln(1 - \epsilon)} + 1$  satisfies the inequality (4.8). Based on the results of Theorem 4.2, we know that  $N^*(1)$  satisfies the inequality (4.8). That means there exists an  $\bar{N} \leq N^*(1)$ , such that the inequality (4.8) holds with  $N = \bar{N}$ , i.e.,

$$(1 - \epsilon)^{\bar{N}-1} (1 - \epsilon + \bar{N}\epsilon) \leq \beta, \quad (4.9)$$

The above inequality (4.9) can be implied by the condition  $(1 - \epsilon)^{\bar{N}-1} (1 - \epsilon + N^*(1)\epsilon) \leq \beta$ , which is equivalent to  $\bar{N} \geq \frac{\ln \beta - \ln(1 - \epsilon + N^*(1)\epsilon)}{\ln(1 - \epsilon)} + 1$ . Therefore the proof is complete.  $\square$

It is worth noting that the derived lower bound  $N_1^* = \lceil \frac{\ln \beta - \ln(1 - \epsilon + N^*(1)\epsilon)}{\ln(1 - \epsilon)} + 1 \rceil$  for  $N$  in Theorem 4.3 is always smaller than the existing lower bound  $N^*(1) = \lceil \frac{1}{\epsilon} \frac{e}{e-1} (1 + \ln \frac{1}{\beta}) \rceil$  in Theorem 4.2 when  $m = 1$ . For instance, when  $\alpha = 0.1$ ,  $\beta = 0.1$  and  $m = 1$ ,  $N_1^* = 41$  is while  $N^*(1)$  is 53. Therefore, the developed lower bound improves the existing work. We next extend the results of Theorem 4.3 to the general case of  $m$ .

**Corollary 4.1.** *Let  $N^* = \lceil \frac{m}{\epsilon} \frac{e}{e-1} (1 + \ln \frac{m}{\beta}) \rceil$ . If  $N \geq \frac{\ln \beta - \ln(m - \epsilon + N^*\epsilon)}{\ln(1 - \frac{\epsilon}{m})} + 1$ , then we have  $1 - \beta$  confidence that any realization  $\mathbf{s}$  belongs to the uncertainty set  $\mathcal{U}_{\text{PCA}}(\mathcal{S}, m)$  with the probability of at least  $1 - \epsilon$ , i.e.,*

$$\mathbb{P}_{\hat{\mathbf{s}}}\{\mathbb{P}_{\mathbf{s}}\{\mathbf{s} \in \mathcal{U}_{\text{PCA}}(\mathcal{S}, m)\} \geq 1 - \epsilon\} \geq 1 - \beta, \quad (4.10)$$

where  $0 < \epsilon < 1$ ,  $0 < \beta < 1$ , and  $\hat{\mathbf{s}} = \{\mathbf{s}_1, \dots, \mathbf{s}_N\}$ .

*Proof.* We have

$$\begin{aligned} \mathbb{P}_{\mathbf{s}}\{\mathbf{s} \in \mathcal{U}_{\text{PCA}}(\mathcal{S}, m)\} &= \mathbb{P}_{\mathbf{s}}\{\bar{\mathbf{s}}_j^i \leq \bar{\mathbf{s}}^i \leq \bar{\mathbf{s}}_j^i, \quad \forall i \in [m], \forall j \in [N]\} \\ &= \mathbb{P}_{\mathbf{s}}\{\min_{j=1}^N \bar{\mathbf{s}}_j^i \leq \bar{\mathbf{s}}^i \leq \max_{j=1}^N \bar{\mathbf{s}}_j^i, \quad \forall i \in [m]\} \\ &\geq \sum_{i=1}^m \mathbb{P}_{\mathbf{s}}\{\min_{j=1}^N \bar{\mathbf{s}}_j^i \leq \bar{\mathbf{s}}^i \leq \max_{j=1}^N \bar{\mathbf{s}}_j^i\} - m + 1 \end{aligned}$$

where the last inequality is due to the Bonferroni Inequalities. Thus we have a sufficient condition for  $\mathbb{P}_{\mathbf{s}}\{\mathbf{s} \in \mathcal{U}_{\text{PCA}}(\mathcal{S}, m)\} \geq 1 - \epsilon$ :

$$\sum_{i=1}^m \mathbb{P}_{\mathbf{s}}\{\min_{j=1}^N \bar{\mathbf{s}}_j^i \leq \bar{\mathbf{s}}^i \leq \max_{j=1}^N \bar{\mathbf{s}}_j^i\} - m + 1 \geq 1 - \epsilon,$$

which can be further implied by:

$$\mathbb{P}_{\mathbf{s}}\{\min_{j=1}^N \bar{\mathbf{s}}_j^i \leq \bar{\mathbf{s}}^i \leq \max_{j=1}^N \bar{\mathbf{s}}_j^i\} \geq 1 - \frac{\epsilon}{m}, \quad \forall i \in [m].$$

Similarly, we have

$$\begin{aligned} \mathbb{P}_{\hat{\mathbf{s}}}\left\{\mathbb{P}_{\mathbf{s}}\left\{\min_{j=1}^N \bar{\mathbf{s}}_j^i \leq \bar{\mathbf{s}}^i \leq \max_{j=1}^N \bar{\mathbf{s}}_j^i\right\} \geq 1 - \frac{\epsilon}{m}, \quad \forall i \in [m]\right\} &= \mathbb{P}_{\hat{\mathbf{s}}}\left\{F_i(\max_{j=1}^N \bar{\mathbf{s}}_j^i) - F_i(\min_{j=1}^N \bar{\mathbf{s}}_j^i) \geq 1 - \frac{\epsilon}{m}, \quad \forall i \in [m]\right\} \\ &\geq \sum_{i=1}^m \mathbb{P}_{\hat{\mathbf{s}}}\left\{\max_{j=1}^N F_i(\bar{\mathbf{s}}_j^i) - \min_{j=1}^N F_i(\bar{\mathbf{s}}_j^i) \geq 1 - \frac{\epsilon}{m}\right\} - m + 1. \end{aligned}$$

Thus, to make inequality (4.10) hold, it is sufficient to have

$$\mathbb{P}_{\mathfrak{s}} \left\{ \max_{j=1}^N F_i(\bar{\mathfrak{s}}_j^i) - \min_{j=1}^N F_i(\bar{\mathfrak{s}}_j^i) \geq 1 - \frac{\epsilon}{m} \right\} \geq 1 - \frac{\beta}{m}.$$

Then by the results of Theorem 4.3, the conclusion follows.  $\square$

It should be noted that, in contrast to the case  $m = 1$ , the derived lower bound  $N_1^{**} = \lceil \frac{\ln \beta - \ln(m - \epsilon + N^* \epsilon)}{\ln(1 - \frac{\epsilon}{m})} + 1 \rceil$  for  $N$  in Corollary 4.1 is not always smaller than the existing lower bound  $N_2^{**} = \lceil \frac{1}{\epsilon} \frac{e}{e-1} (2m - 1 + \ln \frac{1}{\beta}) \rceil$  in Theorem 4.2 when  $m \geq 2$ . Therefore, the developed lower bound complements the existing work.

## 4.6 Computational Experiments

We conduct comprehensive computational experiments to show the effectiveness of the proposed scenario-induced uncertainty sets using two applications: Knapsack and power grid problems. The mathematical models are implemented by MATLAB R2021a (ver. 9.10) API of Gurobi (ver. 9.1) on a PC with a 64-bit Windows Operating System, an Intel(R) Core(TM) i7-7700 CPU @ 3.60 GHz processor, and 16 GB RAM. In Section 4.6.1, we specify the proposed uncertainty sets in the context of the Knapsack and power grid problems. In Section 4.6.2, we address how to randomly generate test instances of these applications and report the numerical results along with their analyses.

### 4.6.1 Computational Setup

In this section, we specify the proposed  $\mathcal{U}_{\text{PCA}}(\mathcal{S}, m_1)$ , and  $\mathcal{U}_{\cap}(\mathcal{S}, m_1)$  uncertainty sets in the context of the knapsack and power grid problems.

#### 4.6.1.1 Knapsack Problem

We are given a set of items, each with a given value and uncertain weight, that we wish to pack into a container with a maximum capacity limit. The goal is to maximize the total value of the packed items by choosing a subset of the items that fit into the container.

$$\begin{aligned} \max \quad & \mathbf{v}^\top \mathbf{x} & (4.11) \\ \text{s.t.} \quad & \mathbf{w}^\top \mathbf{x} \leq W, \quad \forall \mathbf{w} \in \mathcal{U}, \\ & x_z \in \{0, 1\}, \quad \forall z \in [n]. \end{aligned}$$

This problem is a static RO problem. In Problem (4.11), parameter  $n$  represents the number of items and  $\mathbf{v} \in \mathbb{R}^n$  and  $\mathbf{w} \in \mathbb{R}^n$  denote the values and weights of the items, respectively. Parameter  $W$  indicates the maximum capacity of the container and  $\mathcal{U}$  represents the uncertainty set of the uncertain weights. Decision variable  $x_z$ ,  $\forall z \in [n]$ , indicates if item  $z$  is packed into the container (i.e.,  $x_z = 1$ ) or not (i.e.,  $x_z = 0$ ). The objective is to maximize the total value of the packed items subject to the constraint that the total weight of the packed items does not exceed the maximum capacity of the container. Problem (4.11) can be reformulated as the following bi-level problem:

$$\max \quad \mathbf{v}^\top \mathbf{x} \quad (4.12a)$$

$$\text{s.t.} \quad \max_{\mathbf{w} \in \mathcal{U}} \mathbf{w}^\top \mathbf{x} \leq W, \quad (4.12b)$$

$$x_z \in \{0, 1\}, \quad \forall z \in [n].$$

After applying  $\mathcal{U}_{\text{PCA}}(\mathcal{S}, m_1)$  set to Problem (4.12) and replacing the inner optimization problem (4.12b) with its dual formulation, Problem (4.12) is equivalent to the

following problem:

$$\begin{aligned}
& \max \quad \mathbf{v}^\top \mathbf{x} \\
& \text{s.t.} \quad \sum_{i=1}^{m_1} \beta_i + \sum_{i=1}^{m_1} \frac{\underline{\omega}_i}{\|\mathbf{d}_i\|} \mathbf{d}_i^\top \mathbf{x} + \sum_{i=m_1+1}^m \frac{\bar{\omega}_i + \underline{\omega}_i}{2\|\mathbf{d}_i\|} \mathbf{d}_i^\top \mathbf{x} \leq W \\
& \quad \beta_i \geq \frac{\bar{\omega}_i - \underline{\omega}_i}{\|\mathbf{d}_i\|} \mathbf{d}_i^\top \mathbf{x}, \quad \forall i \in [m_1], \\
& \quad \beta_i \in \mathbb{R}_+, \quad \forall i \in [m_1], \\
& \quad x_z \in \{0, 1\}, \quad \forall z \in [n],
\end{aligned}$$

where  $\beta_i$ ,  $\forall i \in [m_1]$  are dual decision variables. Moreover, after applying uncertainty set  $\mathcal{U}_\cap(\mathcal{S}, m_1)$  to Problem (4.12) and replacing the inner optimization problem (4.12b) with its dual formulation, Problem (4.12) is reformulated as the following problem:

$$\begin{aligned}
& \max \quad \mathbf{v}^\top \mathbf{x} \\
& \text{s.t.} \quad \sum_{i=1}^{m_1} \beta_i + (\mathbf{U} - \mathbf{c})^\top \boldsymbol{\gamma} + (\mathbf{c} - \mathbf{L})^\top \boldsymbol{\zeta} + \mathbf{c}_0^\top \mathbf{x} \leq W, \\
& \quad \beta_i + \frac{\bar{\omega}_i - \underline{\omega}_i}{\|\mathbf{d}_i\|} \mathbf{d}_i^\top (\boldsymbol{\gamma} - \boldsymbol{\zeta}) \geq \frac{\bar{\omega}_i - \underline{\omega}_i}{\|\mathbf{d}_i\|} \mathbf{d}_i^\top \mathbf{x}, \quad \forall i \in [m_1], \\
& \quad \boldsymbol{\gamma}, \boldsymbol{\zeta} \in \mathbb{R}_+^n, \\
& \quad \beta_i \in \mathbb{R}_+, \quad \forall i \in [m_1], \\
& \quad x_z \in \{0, 1\}, \quad \forall z \in [n],
\end{aligned}$$

where  $\mathbf{U}$  and  $\mathbf{L}$  are respectively the upper bound and lower bound of the box uncertainty set,

$$\mathbf{c} = \sum_{i=1}^{m_1} \frac{\underline{\omega}_i}{\|\mathbf{d}_i\|} \mathbf{d}_i + \sum_{i=m_1+1}^m \frac{\bar{\omega}_i + \underline{\omega}_i}{2\|\mathbf{d}_i\|} \mathbf{d}_i,$$

$\beta_i$ ,  $\forall i \in [m_1]$ ,  $\boldsymbol{\gamma}$ , and  $\boldsymbol{\zeta}$  are dual decision variables.

#### 4.6.1.2 Power Grid Problem

In this problem, we consider a dispatchable power grid. This power grid is a network of generator stations, transmission systems, and consumers that delivers power from generators to consumers. The power generated by generators is referred to as output while consumers' power demand is referred to as load. Load is also considered as any component of the power grid that consumes power. A bus is defined as a vertical line at which several components of a power grid such as loads or generators are connected. This power grid includes load buses and generator buses. The generators of this power grid are dispatchable, i.e., they can be dispatched on demand by adjusting their output according to power orders. Moreover, the dispatchable generators are subject to ramping constraints. A ramp event is defined as a power increase or decrease event that happens in a time unit. More specifically, a ramp-up event occurs when power increases while a ramp-down event occurs when the power decreases.

To balance output and load, the load shedding and output curtailment procedures are performed on this power grid. Load shedding is the act of switching off output to some consumers when the load is more than output to prevent the power grid from collapsing. Output curtailment is the act of deliberately reducing output to below what could have been generated due to the maintenance of the transmission system or the overloaded transmission system when output is more than load. We assume there is no power flow limitation for this power grid. With this background,

the power grid problem with uncertain load (demand) is defined as follows:

$$\max_{\hat{d} \in \mathcal{U}} \min_{\hat{p}, \hat{q}, \hat{\underline{q}}} \sum_{t=1}^T \sum_{g=1}^G c_g p_g^t + \overline{M} \sum_{t=1}^T \overline{q}^t + \underline{M} \sum_{t=1}^T \underline{q}^t \quad (4.13a)$$

$$\text{s.t.} \quad \sum_{g=1}^G p_g^t + \overline{q}^t - \underline{q}^t = \sum_{l=1}^L d_l^t, \quad \forall t \in [T], \quad (4.13b)$$

$$- \underline{R}_g \leq p_g^t - p_g^{t-1} \leq \overline{R}_g, \quad \forall g \in [G], \forall t \in [T], \quad (4.13c)$$

$$p_g^t \leq \overline{P}_g, \quad \forall g \in [G], \forall t \in [T], \quad (4.13d)$$

$$p_g^t, \overline{q}^t, \underline{q}^t \in \mathbb{R}_+, \quad \forall g \in [G], \forall t \in [T].$$

where  $\hat{d} = \{d_l^t, \forall l \in [L], \forall t \in [T]\}$ ,  $\hat{p} = \{p_g^t, \forall g \in [G], \forall t \in [T]\}$ ,  $\hat{q} = \{\overline{q}^t, \forall t \in [T]\}$ , and  $\hat{\underline{q}} = \{\underline{q}^t, \forall t \in [T]\}$ . This problem is a special case of an ARO problem. In Problem (4.13), parameters  $T$ ,  $G$ ,  $L$  indicate the total number of time units, generator buses, and load buses while each time unit, generator bus, and load bus is identified by indices  $t$ ,  $g$ , and  $l$ , respectively. The cost of generating one megawatt of output by generator bus  $g$  in one time unit is denoted by  $c_g$  and parameters  $\overline{M}$  and  $\underline{M}$  indicate the penalty costs for one megawatt of load shedding and output curtailment performed in one time unit, respectively. The uncertain load (demand) of load bus  $l$  at time  $t$  is represented by  $d_l^t$ . Parameters  $\underline{R}_g$  and  $\overline{R}_g$  denote the maximum allowed ramp-down and ramp-up in megawatt between two consecutive time units for generator bus  $g$  and parameter  $\overline{P}_g$  represents the maximum output capacity of generator bus  $g$  in each time unit. Decision variable  $p_g^t$  denotes the generated output by generator bus  $g$  at time  $t$  in megawatt. Decision variables  $\overline{q}^t$  and  $\underline{q}^t$  represent the performed load shedding and output curtailment on the power grid at time  $t$  in megawatt.

Problem (4.13) minimizes the worst-case total economic dispatch cost, including output generation cost, load shedding penalty cost, and output curtailment penalty cost, by determining the optimal generated output in megawatt by each generator bus in each time unit and optimal performed output curtailment and load shedding in megawatt in each time unit. Constraint (4.13b) balances the total output of generator buses and the total load of load buses considering performed output

curtailment and load shedding in each time unit. Constraint (4.13c) limits the ramp of each generator bus in each two consecutive time units to its maximum allowed ramp-up and ramp-down. Constraint (4.13d) guarantees that the generated output by each generator bus in each time unit does not exceed its maximum output generation capacity.

After replacing the inner optimization problem with its dual formulation, Problem (4.13) is equivalent to the following problem:

$$\begin{aligned}
\max_{\hat{d} \in \mathcal{U}} \max_{\hat{y}, \hat{y}_g, \hat{z}, \hat{x}} & \sum_{t=1}^T \left( \sum_{l=1}^L d_l^t \right) x^t - \sum_{t=1}^T \sum_{g=1}^G (\bar{R}_g \bar{y}_g^t + \underline{R}_g \underline{y}_g^t + \bar{P}_g z_g^t) & (4.14) \\
\text{s.t.} & x^t - \bar{y}_g^t + \bar{y}_g^{t+1} + \underline{y}_g^t - \underline{y}_g^{t+1} - z_g^t \leq c_g, \quad \forall g \in [G], \forall t \in [T-1], \\
& x^T - \bar{y}_g^T + \underline{y}_g^T - z_g^T \leq c_g, \quad \forall g \in [G], \\
& -\underline{M} \leq x^t \leq \bar{M}, \quad \forall t \in [T], \\
& \bar{y}_g^t, \underline{y}_g^t, z_g^t \in \mathbb{R}_+, \quad \forall g \in [G], \forall t \in [T],
\end{aligned}$$

where  $\hat{y} = \{\bar{y}_g^t, \forall g \in [G], t \in [T]\}$ ,  $\hat{y}_g = \{\underline{y}_g^t, \forall g \in [G], t \in [T]\}$ ,  $\hat{z} = \{z_g^t, \forall g \in [G], t \in [T]\}$ , and  $\hat{x} = \{x^t, \forall t \in [T]\}$  are dual decision variables. Then, applying  $\mathcal{U}_{\text{PCA}}(\mathcal{S}, m_1)$  set to Problem (4.14) leads to the following problem:

$$\begin{aligned}
\max_{\hat{y}, \hat{y}_g, \hat{z}, \hat{x}, \hat{\alpha}} & \sum_{t=1}^T \left( \sum_{l=1}^L \mathbf{u}_{(t-1)L+l} \right) x^t - \sum_{t=1}^T \sum_{g=1}^G (\bar{R}_g \bar{y}_g^t + \underline{R}_g \underline{y}_g^t + \bar{P}_g z_g^t) & (4.15) \\
\text{s.t.} & x^t - \bar{y}_g^t + \bar{y}_g^{t+1} + \underline{y}_g^t - \underline{y}_g^{t+1} - z_g^t \leq c_g, \quad \forall g \in [G], \forall t \in [T-1], \\
& x^T - \bar{y}_g^T + \underline{y}_g^T - z_g^T \leq c_g, \quad \forall g \in [G], \\
& -\underline{M} \leq x^t \leq \bar{M}, \quad \forall t \in [T], \\
& \bar{y}_g^t, \underline{y}_g^t, z_g^t \in \mathbb{R}_+, \quad \forall g \in [G], \forall t \in [T], \\
& 0 \leq \alpha_i \leq 1, \quad \forall i \in [m_1],
\end{aligned}$$

where  $\hat{\alpha} = \{\alpha_i, \forall i \in [m_1]\}$  and

$$\mathbf{u} = \sum_{i=1}^{m_1} \left( \alpha_i \left( \frac{\bar{\omega}_i}{\|\mathbf{d}_i\|} \mathbf{d}_i \right) + (1 - \alpha_i) \left( \frac{\underline{\omega}_i}{\|\mathbf{d}_i\|} \mathbf{d}_i \right) \right) + \sum_{i=m_1+1}^m \frac{\bar{\omega}_i + \underline{\omega}_i}{2\|\mathbf{d}_i\|} \mathbf{d}_i.$$

Uncertain parameters  $d_i^t$ ,  $\forall l \in [L], \forall t \in [T]$ , can be represented alternatively as a single uncertain parameter (random variable) vector  $\mathbf{u} \in \mathbb{R}^m$ , where  $m = TL$ . Accordingly, uncertain parameter  $d_i^t$  for a specific  $l$  and  $t$  is located in the  $((t-1)L+l)^{th}$  element of vector  $\mathbf{u} \in \mathbb{R}^{TL}$ . Therefore,  $\mathbf{u}_{(t-1)L+l}$  in the objective function denotes the  $((t-1)L+l)^{th}$  element of  $\mathbf{u}$ , which is  $d_i^t$  for a given  $t$  and  $l$ .

When  $\hat{y}, \hat{y}, \hat{z}$ , and  $\hat{x}$  are fixed, Problem (4.15) is equivalent to the following problem:

$$\begin{aligned} \max_{\hat{\alpha}} \quad & \sum_{t=1}^T \left( \sum_{l=1}^L \mathbf{u}_{(t-1)L+l} \right) x^t - \sum_{t=1}^T \sum_{g=1}^G (\bar{R}_g \bar{y}_g^t + \underline{R}_g \underline{y}_g^t + \bar{P}_g z_g^t) \\ \text{s.t.} \quad & x^t - \bar{y}_g^t + \bar{y}_g^{t+1} + \underline{y}_g^t - \underline{y}_g^{t+1} - z_g^t \leq c_g, \quad \forall g \in [G], \forall t \in [T-1], \\ & x^T - \bar{y}_g^T + \underline{y}_g^T - z_g^T \leq c_g, \quad \forall g \in [G], \\ & -\underline{M} \leq x^t \leq \bar{M}, \quad \forall t \in [T], \\ & \bar{y}_g^t, \underline{y}_g^t, z_g^t \in \mathbb{R}_+, \quad \forall g \in [G], \forall t \in [T], \\ & \alpha_i \in \{0, 1\}, \quad \forall i \in [m_1], \end{aligned}$$

where  $\hat{\alpha}$  is the only decision variable of this linear problem and its feasible region is a box. Therefore, Problem (4.15), which is bi-level, is equivalent to the following single-level mixed integer linear problem:

$$\begin{aligned} \max_{\hat{y}, \hat{y}, \hat{z}, \hat{v}, \hat{x}, \hat{\alpha}} \quad & \sum_{t=1}^T \sum_{l=1}^L \mathbf{q}_{(t-1)L+l} - \sum_{t=1}^T \sum_{g=1}^G (\bar{R}_g \bar{y}_g^t + \underline{R}_g \underline{y}_g^t + \bar{P}_g z_g^t) \\ \text{s.t.} \quad & x^t - \bar{y}_g^t + \bar{y}_g^{t+1} + \underline{y}_g^t - \underline{y}_g^{t+1} - z_g^t \leq c_g, \quad \forall g \in [G], \forall t \in [T-1], \\ & x^T - \bar{y}_g^T + \underline{y}_g^T - z_g^T \leq c_g, \quad \forall g \in [G], \\ & -\underline{M} \leq x^t \leq \bar{M}, \quad \forall t \in [T], \\ & -\underline{M} \alpha_i \leq v_i^t \leq \bar{M} \alpha_i, \quad \forall i \in [m_1], \forall t \in [T], \\ & x^t - (1 - \alpha_i) \bar{M} \leq v_i^t \leq x^t + (1 - \alpha_i) \underline{M}, \quad \forall i \in [m_1], \forall t \in [T], \\ & \bar{y}_g^t, \underline{y}_g^t, z_g^t, v_i^t \in \mathbb{R}_+, \quad \forall g \in [G], \forall t \in [T], \forall i \in [m_1], \\ & \alpha_i \in \{0, 1\}, \quad \forall i \in [m_1], \end{aligned}$$

where  $\hat{v} = \{v_i^t, \forall i \in [m_1], \forall t \in [T]\}$  and

$$\mathbf{q} = \sum_{i=1}^{m_1} \left( v_i^t \left( \frac{\bar{\omega}_i}{\|\mathbf{d}_i\|} \mathbf{d}_i \right) + (x^t - v_i^t) \left( \frac{\omega_i}{\|\mathbf{d}_i\|} \mathbf{d}_i \right) \right) + x^t \left( \sum_{i=m_1+1}^m \frac{\bar{\omega}_i + \omega_i}{2\|\mathbf{d}_i\|} \mathbf{d}_i \right).$$

## 4.6.2 Computational Results

We first explain how we generated random test instances for the Knapsack and power grid problems used to evaluate the performance of the proposed SIU-based RO models. Then, we compare the robust counterparts using these uncertainty sets in terms of the conservativeness of their solutions and the computational time needed to solve them to optimality. Finally, we further investigate the performance of the developed uncertainty sets by performing sensitivity analysis concerning the parameters of the uncertainty sets and the parameters of the Knapsack and power grid problems.

### 4.6.2.1 Instance Generation and Table Header Description

We conduct our computational experiments to solve various instances of the robust knapsack and power grid problems with the proposed uncertainty sets. To generate test instances of the knapsack problem, we follow the same experimental setup proposed by Bertsimas and Sim (2004). As this problem is NP-hard, we generate random Knapsack problems of size  $n = 200$ , which can be solved to optimality by off-the-shelf optimization solvers. The value of each item, i.e.,  $v_z, \forall z \in [200]$ , is randomly selected from the set  $\{16, 17, \dots, 77\}$ . The weight of each item, i.e.,  $w_z, \forall z \in [200]$ , is assumed to be uncertain, dependent on the weights of other items, and follows a Normal distribution, i.e.,  $w_z \sim N(\mu_z, \sigma_z^2), \forall z \in [200]$ , where  $\mu_z$  and  $\sigma_z^2$  denote its mean and variance, respectively. Parameter  $\mu_z$  is randomly chosen from the set  $\{20, 21, \dots, 29\}$  and  $\sigma_z^2$  is assumed to be  $\sigma_z^2 = \frac{\mu_z^2}{300}$ . There exists the same correlation between any two dependent weights, which is denoted by  $\rho$  and implies their dependency. More specifically, the weight of item  $z$  is assumed

to be correlated with the weight of item  $z + 1$  for any odd values of  $z$ . With this assumption and the definition of correlation, i.e.,  $\rho = \frac{\text{Cov}(w_z, w_{z+1})}{\sigma_z \sigma_{z+1}}$ , the covariance matrix of the weights is as follows:

$$\text{Cov}(\mathbf{w}) = \begin{bmatrix} \frac{\mu_1^2}{300} & \frac{\rho\mu_1\mu_2}{300} & 0 & \dots & 0 & 0 \\ \frac{\rho\mu_1\mu_2}{300} & \frac{\mu_2^2}{300} & 0 & \dots & 0 & 0 \\ 0 & 0 & \ddots & \ddots & 0 & 0 \\ \vdots & \vdots & & & \vdots & \vdots \\ 0 & 0 & & & 0 & 0 \\ 0 & 0 & \dots & 0 & \frac{\mu_{199}^2}{300} & \frac{\rho\mu_{199}\mu_{200}}{300} \\ 0 & 0 & \dots & 0 & \frac{\rho\mu_{199}\mu_{200}}{300} & \frac{\mu_{200}^2}{300} \end{bmatrix}.$$

To randomly generate correlated normally distributed scenarios, we first generate uncorrelated scenarios, denoted by  $\mathbf{s}_j'' \in \mathbb{R}^{200}, \forall j \in [N]$ , using  $w_z \sim N(\mu_z, \sigma_z^2), \forall z \in [200]$ . Then, we obtain matrix  $\mathbf{M}$  by the Cholesky decomposition of  $\text{Cov}(\mathbf{w})$  so that  $\mathbf{M}\mathbf{M}^\top = \text{Cov}(\mathbf{w})$ . Finally, we generate the correlated scenarios by  $\mathbf{s}_j = \boldsymbol{\mu} + \mathbf{M}\mathbf{s}_j'', \forall j \in [N]$ . In the next step, we perform EVD on the sample covariance matrix  $\mathbf{C} = \frac{1}{N-1}\mathbf{X}_0^\top\mathbf{X}_0$ , which is an approximation of  $\text{Cov}(\mathbf{w})$ , to develop the proposed uncertainty sets using the PCA technique.

The solutions of the RO Knapsack problems with each proposed uncertainty set are functions of input scenarios. Accordingly, conducting computational experiments based on only one set of scenarios might lead to a biased analysis of the uncertainty set performance. Therefore, we create 10 sets of scenarios, i.e.,  $\mathcal{S}_k, \forall k \in [10]$ , so that each  $\mathcal{S}_k$  includes  $N = 10,000$  scenarios of  $w_z, \forall z \in [200]$ , randomly generated by  $N(\mu_z, \sigma_z^2)$ . We construct each proposed uncertainty set based on all  $\mathcal{S}_k$  sets, which results in 10 different uncertainty sets of the same type. Then, these uncertainty sets are applied to each Knapsack test problem, leading to 10 problems with the same type of uncertainty set. Finally, the 10 robust knapsack problems are solved to optimality and the average of their optimal objective values and computational times is reported as the performance of the proposed uncertainty set in the context of the given Knapsack test problem.

To study how the maximum capacity of the container and the correlation between the weights of items affect the performance of the proposed uncertainty

sets, we perform sensitivity analysis with respect to parameters  $W$  and  $\rho$ . To that end, we conduct our experiments based on three values of the maximum capacity and seven values of the correlation, i.e.,  $W \in \{3000, 4000, 5000\}$  and  $\rho \in \{-0.8, -0.5, -0.2, 0, 0.2, 0.5, 0.8\}$ .

To generate the test instances of the power grid problem, we consider an IEEE 24-bus system [TK - citation] that consists of 32 generator buses and 17 load buses, planning for a 24-hour horizon, i.e.,  $T = 24$ . For simplicity, the 17 load buses are assumed to be grouped into two load buses. Accordingly, the number of generator buses and load buses are set to be 32 and 2, i.e.,  $G = 32$  and  $L = 2$ . For each generator bus, the output generation cost  $c_g$  is randomly generated by discrete uniform distribution  $\text{unif}(10, 150)$  and the maximum output capacity  $\bar{P}_g$  is randomly generated by continuous uniform distribution  $\text{unif}(5, 245)$ . Similarly, the maximum allowed ramp-down and ramp-up for each generator bus, i.e.,  $\underline{R}_g$  and  $\bar{R}_g$ , are randomly generated by continuous uniform distribution  $\text{unif}(5, 105)$ . The penalty costs of the load shedding and output curtailment are considered \$500 and \$50 per megawatt, i.e.,  $\bar{M} = 500$  and  $\underline{M} = 50$ .

The two grouped loads in each time, denoted by  $d_1^t$  and  $d_2^t$ ,  $\forall t \in [24]$ , are assumed to be uncertain, dependent on each other, and follow a Normal distribution, i.e.,  $d_l^t \sim N(\mu_l^t, \sigma_l^{t2})$ ,  $\forall l \in [2], \forall t \in [24]$ . We assume  $\mu_1^t = \frac{2}{5}Q^t$  and  $\mu_2^t = \frac{3}{5}Q^t$ , where  $Q^t$  represents the total loads of 17 load buses in hour  $t$  and is randomly generated by continuous uniform distribution  $\text{unif}(2100, 2900)$ . Moreover, the variances of the two grouped loads in each hour are set to be  $\sigma_1^{t2} = \frac{\mu_1^{t2}}{100}$  and  $\sigma_2^{t2} = \frac{\mu_2^{t2}}{100}$ .

As  $d_1^t$  and  $d_2^t$ ,  $\forall t \in [24]$ , form a collection of random variables or events indexed by different instants of time, we consider them as a stochastic process. This stochastic process is assumed to be a Markov chain, where each event depends only on the state attained in the previous event. In other words, random variables  $(d_1^{t+1}, d_2^{t+1})$  only depend on  $(d_1^t, d_2^t)$  for all  $t \in [23]$ . In a Markov chain, there are two types of relationships between random variables, referred to as temporal relationship and

spatial relationship. We define the temporal relationship as the correlation between  $d_l^t$  and  $d_l^{t+1}$ , which is quantified by the temporal correlation coefficient  $\rho_1$  so that  $\rho_1 = \frac{\text{Cov}(d_l^t, d_l^{t+1})}{\sigma_l^t \sigma_l^{t+1}}$ ,  $\forall l \in [2], \forall t \in [23]$ . On the other hand, the spatial relationship is defined as the correlation between  $d_1^t$  and  $d_2^t$ , which is quantified by the spatial correlation coefficient  $\rho_2$  so that  $\rho_2 = \frac{\text{Cov}(d_1^t, d_2^t)}{\sigma_1^t \sigma_2^t}$ ,  $\forall t \in [24]$ . Uncertain parameters  $d_1^t$  and  $d_2^t$ ,  $\forall t \in [24]$ , form vector  $\mathbf{u} \in \mathbb{R}^{48}$  whose  $(2(t-1) + l)^{th}$  element is  $d_l^t$ . Similar to the Knapsack problem, we randomly generate correlated normally distributed scenarios by the randomly generated mean vector, randomly generated uncorrelated scenarios, and the Cholesky decomposition of  $\text{Cov}(\mathbf{u})$ . Then, we perform EVD on the sample covariance matrix  $\mathbf{C}$ .

To study the effect of temporal and spatial correlations on the performance of the proposed uncertainty sets, we perform sensitivity analysis with respect to parameters  $\rho_1$  and  $\rho_2$ . For this purpose, we consider two and seven settings for the temporal and spatial correlations, respectively, i.e.,  $\rho_1 \in \{0.5, 0.9\}$  and  $\rho_2 \in \{-0.8, -0.5, -0.2, 0, 0.2, 0.5, 0.8\}$ . We further investigate the impact of the utilized principal components on the conservativeness of solutions and computational time by setting  $m_1$  to 42 and 36, i.e., 87.5% and 75% of the  $m = 48$  principal components. Like the Knapsack problem, we create 10 sets of 10,000 scenarios of  $d_l^t$ ,  $\forall l \in [2], t \in [24]$ , randomly generated by  $d_l^t \sim N(\mu_l^t, \sigma_l^{t2})$ .

In Section 4.6.2.2, we will summarize the results of our computational experiments in Tables 4.1 - 4.8. Columns “ $\rho$ ”, “ $\rho_1$ ”, “ $\rho_2$ ” report the values of conventional, temporal, and spatial correlations. Column “Value” represents the optimal objective value of the corresponding RO problem. Column “Time” reports the computational time of solving the corresponding RO problem to optimality in seconds. Symbol “Gap” represents the relative gap in percentage between “Value” of the first column and “Value” of the corresponding uncertainty set. We define the relative gap between two values as their difference divided by the maximum absolute value.

#### 4.6.2.2 Uncertainty Set Performance

We summarize the results for robust problems with  $\mathcal{U}_{\text{conv}}(\mathcal{S})$ ,  $\mathcal{U}_{\text{box}}$ ,  $\mathcal{U}_{\text{PCA}}(\mathcal{S}, m)$ , and  $\mathcal{U}_{\cap}(\mathcal{S}, m)$  uncertainty sets in the context of the Knapsack problem in Tables 4.1 - 4.3, while Tables 4.4 - 4.8 report the results for RO problems with  $\mathcal{U}_{\text{conv}}(\mathcal{S})$ ,  $\mathcal{U}_{\text{box}}$ ,  $\mathcal{U}_{\text{PCA}}(\mathcal{S}, m)$ , and  $\mathcal{U}_{\text{PCA}}(\mathcal{S}, m_1)$  uncertainty sets in the context of the power grid problem. Sensitivity analysis with respect to the maximum container capacity  $W$  and varying values of correlation  $\rho$  are reported in Tables 4.1 - 4.3. Tables 4.5 and 4.8 show how the number of utilized principal components affects the performance of  $\mathcal{U}_{\text{PCA}}(\mathcal{S}, m_1)$ , reporting the sensitivity analysis results with respect to  $m_1$ .

Sensitivity analyses with respect to the three values of temporal correlation  $\rho_1$  are reported through Tables 4.5 - 4.8. Within each of these tables, we hold  $\rho_1$  constant and present sensitivity analysis results for varying values of spatial correlation  $\rho_2$ . In general, shorter computational times imply greater tractability, and better objective values imply less conservativeness (i.e., a larger value for the Knapsack problem (maximization) and a smaller value for the power grid problem (minimization)).

Table 4.1: The Knapsack problem with  $W = 3000$

$\rho$	$\mathcal{U}_{\text{conv}}(\mathcal{S})$		$\mathcal{U}_{\text{box}}$			$\mathcal{U}_{\text{PCA}}(\mathcal{S}, m)$			$\mathcal{U}_{\cap}(\mathcal{S}, m)$		
	Value	Time (secs)	Value	Time (secs)	Gap (%)	Value	Time (secs)	Gap (%)	Value	Time (secs)	Gap (%)
-0.8	7134	747.1	6249	0.1	12.41	6468	547.5	9.34	6570	85.7	7.91
-0.5	7082	739.6	6210	0.1	12.31	6296	457.5	11.10	6392	144.1	9.74
-0.2	7111	742.6	6246	0.1	12.16	6222	83.0	12.50	6338	70.0	10.87
0	7040	742.0	6200	0.1	11.93	6133	54.3	12.88	6256	37.9	11.14
0.2	7158	741.4	6311	0.1	11.83	6214	50.4	13.19	6350	24.4	11.29
0.5	7017	740.9	6196	0.1	11.70	6098	55.3	13.10	6235	27.9	11.14
0.8	7109	742.8	6251	0.1	12.07	6175	32.5	13.14	6308	20.2	11.27

Table 4.2: The Knapsack problem with  $W = 4000$

$\rho$	$\mathcal{U}_{\text{conv}}(\mathcal{S})$		$\mathcal{U}_{\text{box}}$			$\mathcal{U}_{\text{PCA}}(\mathcal{S}, m)$			$\mathcal{U}_{\cap}(\mathcal{S}, m)$		
	Value	Time (secs)	Value	Time (secs)	Gap (%)	Value	Time (secs)	Gap (%)	Value	Time (secs)	Gap (%)
-0.8	8585	725.2	7686	0.1	10.47	8071	52.57	5.99	8147	12.5	5.10
-0.5	8484	728.4	7599	0.1	10.43	7852	30.2	7.45	7921	18.2	6.64
-0.2	8443	743.0	7585	0.1	10.16	7692	5.3	8.89	7780	19.5	7.85
0	8443	730.3	7609	0.1	9.88	7648	3.5	9.42	7746	11.5	8.26
0.2	8497	729.9	7646	0.1	10.02	7691	7.2	9.49	7795	12.5	8.26
0.5	8264	728.3	7438	0.1	10.00	7483	5.0	9.45	7581	12.9	8.26
0.8	8406	729.8	7575	0.1	9.89	7615	2.7	9.41	7715	10.8	8.22

Table 4.3: The Knapsack problem with  $W = 5000$ 

$\rho$	$\mathcal{U}_{\text{conv}}(\mathcal{S})$		$\mathcal{U}_{\text{box}}$			$\mathcal{U}_{\text{PCA}}(\mathcal{S}, m)$			$\mathcal{U}_{\cap}(\mathcal{S}, m)$		
	Value	Time (secs)	Value	Time (secs)	Gap (%)	Value	Time (secs)	Gap (%)	Value	Time (secs)	Gap (%)
-0.8	9291	678.9	8610	0.1	7.33	9086	2.2	2.21	9118	5.8	1.86
-0.5	9442	666.8	8756	0.1	7.27	9083	2.4	3.80	9128	6.0	3.33
-0.2	9275	662.4	8602	0.1	7.26	8787	1.7	5.26	8847	5.3	4.61
0	9352	666.2	8694	0.1	7.04	8813	1.0	5.76	8886	6.1	4.98
0.2	9088	664.8	8449	0.1	7.03	8564	1.4	5.77	8635	6.7	4.98
0.5	9263	664.2	8588	0.1	7.29	8716	1.7	5.91	8788	8.0	5.13
0.8	9296	663.1	8631	0.1	7.15	8740	1.3	5.98	8814	5.8	5.19

From Tables 4.1 - 4.3, we have the following observations. First, as expected  $\mathcal{U}_{\text{box}}$  results in the most tractable RO Knapsack problems while  $\mathcal{U}_{\text{conv}}(\mathcal{S})$  leads to the least tractable ones. This is directly related to the high dimensionality (i.e. more decision variables) of  $\mathcal{U}_{\text{conv}}(\mathcal{S})$  compared with  $\mathcal{U}_{\text{box}}$ . Second,  $\mathcal{U}_{\text{conv}}(\mathcal{S})$  results in the least conservative RO problems while  $\mathcal{U}_{\text{box}}$  leads to the most conservative ones. This is because  $\mathcal{U}_{\text{conv}}(\mathcal{S})$  defines the smallest uncertainty set while  $\mathcal{U}_{\text{box}}$  defines the largest uncertainty set. Similarly,  $\mathcal{U}_{\cap}(\mathcal{S}, m)$  leads to less conservative RO problems in comparison with  $\mathcal{U}_{\text{PCA}}(\mathcal{S}, m)$ . Third, RO problems using  $\mathcal{U}_{\cap}(\mathcal{S}, m)$  and  $\mathcal{U}_{\text{box}}$ , respectively, have the smallest and largest Gap values, meaning they have the most and least similar performance compared to  $\mathcal{U}_{\text{conv}}(\mathcal{S})$  in terms of conservativeness. Fourth, RO Knapsack problems with any uncertainty sets are more tractable when  $W$  is larger. Fifth,  $\mathcal{U}_{\cap}(\mathcal{S}, m)$  leads to more tractable RO problems compared to  $\mathcal{U}_{\text{PCA}}(\mathcal{S}, m)$  when  $W$  is smaller while  $\mathcal{U}_{\text{PCA}}(\mathcal{S}, m)$  outperforms  $\mathcal{U}_{\cap}(\mathcal{S}, m)$  in this regard for larger values of  $W$ . Sixth, when  $W$  is larger, RO problems with either  $\mathcal{U}_{\text{box}}$ ,  $\mathcal{U}_{\text{PCA}}(\mathcal{S}, m)$ , or  $\mathcal{U}_{\cap}(\mathcal{S}, m)$  have smaller Gap values. In other words,  $\mathcal{U}_{\text{conv}}(\mathcal{S})$  has a less significant benefit over other uncertainty sets in terms of conservativeness when  $W$  is larger. Seventh, when scenarios are more negatively correlated,  $\mathcal{U}_{\cap}(\mathcal{S}, m)$  and  $\mathcal{U}_{\text{PCA}}(\mathcal{S}, m)$  result in less conservative RO problems and, moreover, they have smaller Gap values.

We reached similar conclusions for results presented in Tables 4.4 - 4.6. RO problems with  $\mathcal{U}_{\text{PCA}}(\mathcal{S}, m)$  are more tractable than those with  $\mathcal{U}_{\text{conv}}(\mathcal{S})$  and less tractable than those with  $\mathcal{U}_{\text{box}}$ . While, RO problems with  $\mathcal{U}_{\text{PCA}}(\mathcal{S}, m)$  are less conservative

than RO problems with  $\mathcal{U}_{\text{box}}$  and more conservative than those with  $\mathcal{U}_{\text{conv}}(\mathcal{S})$ . Lastly, similar to RO problems with  $\mathcal{U}_{\text{conv}}(\mathcal{S})$ , RO problems with  $\mathcal{U}_{\text{PCA}}(\mathcal{S}, m)$  are less conservative when scenarios are more negatively correlated.

Table 4.4: The power grid problem with  $\rho_1 = 0.5$

$\rho_2$	$\mathcal{U}_{\text{conv}}(\mathcal{S})$		$\mathcal{U}_{\text{box}}$			$\mathcal{U}_{\text{PCA}}(\mathcal{S}, m)$		
	Value ( $\times 10^7$ )	Time (secs)	Value ( $\times 10^7$ )	Time (secs)	Gap (%)	Value ( $\times 10^7$ )	Time (secs)	Gap (%)
-0.8	0.79	960.7	1.77	0.3	55.37	1.37	73.2	42.34
-0.5	0.82	955.6	1.76	0.3	53.41	1.38	939.2	40.58
-0.2	0.85	954.5	1.77	0.3	51.98	1.48	816.5	42.57
0	0.88	956.3	1.77	0.3	50.28	1.53	344.1	42.48
0.2	0.88	961.2	1.77	0.3	50.28	1.50	509.3	41.33
0.5	0.94	962.9	1.75	0.3	46.29	1.50	236.6	37.33
0.8	0.93	1005.1	1.79	0.4	48.04	1.54	57.1	39.61

Based on results from Tables 4.5 and 4.6, we observed the following. First, when  $\rho_1$  is larger,  $\mathcal{U}_{\text{PCA}}(\mathcal{S}, m_1)$  results in more tractable RO problems while  $\mathcal{U}_{\text{box}}$  leads to less tractable RO problems. Second, RO problems with  $\mathcal{U}_{\text{box}}$  are more conservative when  $\rho_1$  is larger. Third, when  $\rho_2$  is smaller, RO problems with  $\mathcal{U}_{\text{PCA}}(\mathcal{S}, m_1)$  have larger Gap values. In other words, RO problems with  $\mathcal{U}_{\text{PCA}}(\mathcal{S}, m_1)$  have a more remarkable benefit over those with  $\mathcal{U}_{\text{box}}$  in terms of conservativeness when  $\rho_2$  is smaller. Fourth, a smaller  $m_1$  results in more tractable and less conservative (i.e. robust) RO problems compared with  $\mathcal{U}_{\text{PCA}}(\mathcal{S}, m_1)$  because it yields a more smaller uncertainty set. Therefore,  $m_1$  can be used as a tool to trade-off tractability, conservativeness, and robustness of RO models with  $\mathcal{U}_{\text{PCA}}(\mathcal{S}, m_1)$ .

Table 4.5: The power grid problem with  $\rho_1 = 0.5$

$\rho_2$	$\mathcal{U}_{\text{box}}$		$\mathcal{U}_{\text{PCA}}(\mathcal{S}, m_1 = m = 48)$			$\mathcal{U}_{\text{PCA}}(\mathcal{S}, m_1 = 42)$			$\mathcal{U}_{\text{PCA}}(\mathcal{S}, m_1 = 36)$		
	Value ( $\times 10^7$ )	Time (secs)	Value ( $\times 10^7$ )	Time (secs)	Gap (%)	Value ( $\times 10^7$ )	Time (secs)	Gap (%)	Value ( $\times 10^7$ )	Time (secs)	Gap (%)
-0.8	1.77	0.3	1.37	73.2	22.60	1.34	15.2	24.29	1.28	2.9	27.68
-0.5	1.76	0.3	1.38	938.2	21.59	1.35	63.2	23.30	1.30	10.6	26.14
-0.2	1.77	0.3	1.48	816.5	16.38	1.46	61.3	17.51	1.42	9.8	19.77
0	1.77	0.3	1.53	344.1	13.56	1.50	62.0	15.25	1.46	9.5	17.51
0.2	1.77	0.3	1.50	509.3	15.25	1.47	167.9	16.95	1.43	34.0	19.21
0.5	1.75	0.3	1.50	236.6	14.29	1.48	123.4	15.43	1.43	41.8	18.29
0.8	1.79	0.4	1.54	57.1	13.97	1.52	47.7	15.08	1.49	12.9	16.76

Base on to numbers in the ‘‘Value’’ columns of Tables 4.7 - 4.8, the power grid problem with  $\mathcal{U}_{\text{PCA}}(\mathcal{S}, m_1)$  leads to a lower bound for the power grid problem with

Table 4.6: The power grid problem with  $\rho_1 = 0.9$

$\rho_2$	$\mathcal{U}_{\text{box}}$		$\mathcal{U}_{\text{PCA}}(\mathcal{S}, m_1 = m = 48)$			$\mathcal{U}_{\text{PCA}}(\mathcal{S}, m_1 = 42)$			$\mathcal{U}_{\text{PCA}}(\mathcal{S}, m_1 = 36)$		
	Value ( $\times 10^7$ )	Time (secs)	Value ( $\times 10^7$ )	Time (secs)	Gap (%)	Value ( $\times 10^7$ )	Time (secs)	Gap (%)	Value ( $\times 10^7$ )	Time (secs)	Gap (%)
-0.8	1.86	0.4	1.31	7.7	29.57	1.30	2.1	30.11	1.29	1.2	30.65
-0.5	1.87	0.4	1.45	6.0	22.46	1.44	1.7	22.99	1.43	1.4	23.53
-0.2	1.87	0.4	1.47	17.1	21.39	1.45	3.6	22.46	1.44	1.5	22.99
0	1.88	0.4	1.59	6.1	15.43	1.58	2.6	15.96	1.56	1.4	17.02
0.2	1.91	0.4	1.62	6.9	15.18	1.61	2.9	15.71	1.60	1.5	16.23
0.5	1.91	0.4	1.63	2.8	14.66	1.62	2.5	15.18	1.61	1.4	15.71
0.8	1.86	0.4	1.64	4.1	11.83	1.64	9.2	11.83	1.63	2.1	12.37

$\mathcal{U}_{\text{PCA}}(\mathcal{S}, m)$ . Time and gap results show that for smaller  $m_1$ , the power grid problem with  $\mathcal{U}_{\text{PCA}}(\mathcal{S}, m_1)$  results in a looser lower bound and is more computationally tractable. Moreover, larger values of  $\rho_1$  lead to tighter and more tractable lower bounds. Again, the results justify that  $m_1$  can be used as a tool to trade-off tractability, conservativeness, and robustness of RO models with  $\mathcal{U}_{\text{PCA}}(\mathcal{S}, m_1)$ .

Table 4.7: Lower bounds for the power grid problem with  $\rho_1 = 0.5$

$\rho_2$	$\mathcal{U}_{\text{PCA}}(\mathcal{S}, m_1 = m)$		$\mathcal{U}_{\text{PCA}}(\mathcal{S}, m_1 = 42)$			$\mathcal{U}_{\text{PCA}}(\mathcal{S}, m_1 = 36)$		
	Value ( $\times 10^7$ )	Time (secs)	Value ( $\times 10^7$ )	Time (secs)	Gap (%)	Value ( $\times 10^7$ )	Time (secs)	Gap (%)
-0.8	1.37	73.2	1.34	15.2	2.19	1.28	2.9	6.57
-0.5	1.38	938.2	1.35	63.2	2.17	1.30	10.6	5.80
-0.2	1.48	816.5	1.46	61.3	1.35	1.42	9.8	4.05
0	1.53	344.1	1.50	62.0	1.96	1.46	9.5	4.58
0.2	1.50	509.3	1.47	167.9	2.00	1.43	34.0	4.67
0.5	1.50	236.6	1.48	123.4	1.33	1.43	41.8	4.67
0.8	1.54	57.1	1.52	47.7	1.30	1.49	12.9	3.25

Table 4.8: Lower bounds for power grid problem with  $\rho_1 = 0.9$

$\rho_2$	$\mathcal{U}_{\text{PCA}}(\mathcal{S}, m_1 = m)$		$\mathcal{U}_{\text{PCA}}(\mathcal{S}, m_1 = 42)$			$\mathcal{U}_{\text{PCA}}(\mathcal{S}, m_1 = 36)$		
	Value ( $\times 10^7$ )	Time (secs)	Value ( $\times 10^7$ )	Time (secs)	Gap (%)	Value ( $\times 10^7$ )	Time (secs)	Gap (%)
-0.8	1.31	7.7	1.30	2.1	0.76	1.29	1.2	1.53
-0.5	1.45	6.0	1.44	1.7	0.69	1.43	1.4	1.38
-0.2	1.47	17.1	1.45	3.6	1.36	1.44	1.5	2.04
0	1.59	6.1	1.58	2.6	0.63	1.56	1.4	1.89
0.2	1.62	6.9	1.61	2.9	0.62	1.60	1.5	1.23
0.5	1.63	2.8	1.62	2.5	0.61	1.61	1.4	1.23
0.8	1.64	14.1	1.64	9.2	0.00	1.63	2.1	0.61

## 4.7 Discussions

In this chapter, we proposed a systematic approach to develop data-driven polyhedral uncertainty sets using PCA. These uncertainty sets alleviate some of the drawbacks of conventional polyhedral uncertainty sets. Primarily, the proposed uncertainty sets capture the correlation information between uncertain parameters and are less conservative than the box uncertainty sets. Moreover, they lead to more computationally tractable RO models compared to the convex hull of uncertainty data. The number of the leading principal components in these uncertainty sets can be used as a tool to trade-off tractability, conservativeness, and robustness of RO models. Additionally, we developed a theoretical bound on the gap between the optimal value of a static RO problem under a piece-wise linear objective function with a scenario-induced uncertainty set and that of its lower bound to quantify the quality of the lower bound. We also derived probabilistic guarantees for the performance of the proposed uncertainty sets by developing explicit lower bounds on the number of scenarios required to construct uncertainty sets.

## CHAPTER 5

# Conclusions

In this dissertation, we focused on developing computationally efficient approaches for SP, RO, and DRO problems and applying them to some real-world applications. This chapter first summarizes the contributions of the dissertation and then proposes some future research directions to extend our study.

### Summary

In Chapter 2, we developed computationally efficient inner and outer approximations for DRO problems with the moment-based and combined ambiguity sets. We approximated the original DRO problems by PCA and uncertainty vector splitting techniques to make PSD matrix constraints smaller and to shrink the uncertainty space dimensionality. Moreover, we derived theoretical bounds on the optimality gaps of the approximations, which help determine the required numbers of split pieces and principal components to reach a predetermined error bound and a trade-off between solution quality and computational time. Meanwhile, the inner and outer approximations together construct an interval that contains the unknown optimal solution for a large-scale DRO problem, which cannot be solved to optimality or even feasibility. Finally, the results of the conducted experiments on distributionally robust production-transportation and multi-product newsvendor problems showed that our approximations significantly reduce the computational time while maintaining high solution quality.

In Chapter 3, we designed a resilient reverse supply chain and logistics network for recycling NdFeB magnets through developing a CTSP model. This model optimally locates the facilities and determines their optimal processing capacities at

the beginning of the planning horizon and then determines optimal material flows and inventory levels for each period of the planning horizon. To guarantee the resilience of this multi-product, multi-period, and multi-echelon network against potential risks, the proposed model uses multiple strategies, such as buffer inventories, dynamic material and inventory adjustments, and enforcement of minimum service level. Furthermore, we generated scenarios that mimic the unique features of COVID-19-related disruptions, such as different disruptions rates, non-linear recovery rates, and the rebound effect on NdFeB magnet demand. Finally, we developed a computationally efficient Benders decomposition algorithm that reduces the computation time of CPLEX's default algorithm by  $\sim 98.5\%$ .

In Chapter 4, we developed data-driven polyhedral uncertainty sets that capture the correlation information between uncertain parameters. These uncertainty sets are less conservative than the box uncertainty sets and lead to more computationally tractable RO models compared to the convex hull of uncertainty data. Besides, we developed a theoretical bound on the gap between the optimal value of a static RO problem with a scenario-induced uncertainty set and that of its lower bound to quantify the quality of the lower bound. We also derived probabilistic guarantees for the performance of the proposed uncertainty sets by developing explicit lower bounds on the number of scenarios.

### **Future Research Directions**

In future research, we can extend our work in Chapter 2 from the following perspectives by addressing the limitations of this study:

1. It would be nice to consider a more general objective function and further develop inner and outer approximations for DRO problems. Currently in this study, we only consider piece-wise linear objective functions.
2. It would be appealing to derive tighter and thus less conservative theoretical error bounds for our derived approximations, given that error bounds in

Propositions 2.7 and 2.10 are not tight in general.

3. By observing our derived approximations perform differently in the context of different applications, it would be very interesting to investigate how the problem structure and the technical insights of deriving the approximations can affect the computational performance of our approximations in solving various application problems.

Moreover, future studies may focus on the following extensions of the research presented in Chapter 3:

1. We may expand the current model to a multi-objective optimization problem by maximizing the environmental and social benefits in addition to the economic and resilience aspects to address the triple bottom line of sustainability. For example, the carbon footprint reduction from recycling compared to virgin magnet production can be used to measure environmental benefits, which may be negated by the additional transportation impacts from the reverse logistics, requiring a balanced decision. In another example, candidate facility locations can be selected based on social indices such as job creation opportunities and quality of life to enhance social sustainability.
2. We assumed that the disruption time and rate in one location are independent of those in another location, but the impact of large-scale disruptive events like COVID-19 can easily propagate through the supply chain sequentially or interactively among the stakeholders. Thus, future studies may consider the propagation speed and time of disruption while designing a resilient reverse logistics network.
3. We proposed the development of a national logistics network to shield the domestic supply chain from global disruptions, which could be expanded to a global network utilizing the existing supply chain of NdFeB magnets.

4. Future research may incorporate more uncertainties, such as NdFeB magnet price volatility and geopolitical conflicts that could act as another black swan event as was evidenced from the 2011 REE crisis. If there is no sufficient data about these uncertainties, other modeling approaches such as robust optimization and distributionally robust optimization can be utilized instead of stochastic programming that requires the complete knowledge of probability distribution of the uncertainties. For instance, as the COVID-19 pandemic is the most recent global black swan event, the volatile price of NdFeB magnets may not be captured sufficiently to allow estimation of its probability distribution function. Therefore, future research may use partial knowledge of the unknown probability distribution, e.g., its moment information, to develop distributionally robust optimization models. Alternatively, the available data can be utilized to develop robust optimization problems with data-driven uncertainty sets.
5. Dynamic pricing strategies could be adopted to increase the EOL product supply. As the unsatisfied NdFeB magnet demand is mainly caused by insufficient supplies of EOL HDDs and EV motors, a time-based pricing strategy can help collect more EOL products. More specifically, a time-based pricing strategy can consider higher prices for EOL products when the supply is lower, e.g., during the warm-up period, to encourage more end users to sell their EOL products. These approaches could increase the resiliency of the proposed supply chain and logistics network.

Finally, there are multiple important extensions of the study presented in Chapter 4 as follows:

1. It would be worthwhile to leverage other machine learning techniques to improve the proposed scenario-induced uncertainty sets.
2. It would be useful to derive a theoretical bound on the gaps between the

optimal value of an ARO problem, under a more general objective function, with a scenario-induced uncertainty set and that of its lower bound.

3. In this study, we used the first  $m_1$  principal directions with the largest variance to develop approximate scenario-induced uncertainty sets. However, these principal directions may not always lead to the best results. Future studies may focus on developing a systematic approach to finding the  $m_1$  directions that result in the best performance, with respect to computational tractability and robustness of the solution.
4. We would like to improve the lower bound on the number of scenario samples required to achieve the desired probabilistic performance guarantees.

## REFERENCES

- Ahmadi-Javid, A. and A. H. Seddighi (2013). A location-routing problem with disruption risk. *Transportation Research Part E: Logistics and Transportation Review*, **53**, pp. 63–82.
- Ahranjani, P. M., S. F. Ghaderi, A. Azadeh, and R. Babazadeh (2020). Robust design of a sustainable and resilient bioethanol supply chain under operational and disruption risks. *Clean Technologies and Environmental Policy*, **22**(1), pp. 119–151.
- Alcantara, R. G. A. L., P. (2017). Bci supply chain resilience report. Business Continuity Institute.
- Azad, N., H. Davoudpour, G. K. Saharidis, and M. Shiripour (2014). A new model to mitigating random disruption risks of facility and transportation in supply chain network design. *The International Journal of Advanced Manufacturing Technology*, **70**(9-12), pp. 1757–1774.
- Baghalian, A., S. Rezapour, and R. Z. Farahani (2013). Robust supply chain network design with service level against disruptions and demand uncertainties: A real-life case. *European journal of operational research*, **227**(1), pp. 199–215.
- Ben-Tal, A., D. Den Hertog, A. De Waegenaere, B. Melenberg, and G. Rennen (2013). Robust solutions of optimization problems affected by uncertain probabilities. *Management Sci.*, **59**(2), pp. 341–357.
- Ben-Tal, A., L. El Ghaoui, and A. Nemirovski (2009). *Robust optimization*, volume 28. Princeton University Press.
- Ben-Tal, A. and A. Nemirovski (1998). Robust convex optimization. *Math. Oper. Res.*, **23**(4), pp. 769–805.
- Ben-Tal, A. and A. Nemirovski (2008). Selected topics in robust convex optimization. *Mathematical Programming*, **112**(1), pp. 125–158.
- Benders, J. (1962). Partitioning procedures for solving mixed-variables programming problems ‘. *Numerische mathematik*, **4**(1), pp. 238–252.
- Bertsimas, D., D. B. Brown, and C. Caramanis (2011). Theory and applications of robust optimization. *SIAM Rev.*, **53**(3), pp. 464–501.

- Bertsimas, D., X. V. Doan, K. Natarajan, and C.-P. Teo (2010). Models for minimax stochastic linear optimization problems with risk aversion. *Math. Oper. Res.*, **35**(3), pp. 580–602.
- Bertsimas, D., V. Gupta, and N. Kallus (2018a). Data-driven robust optimization. *Math. Programming*, **167**(2), pp. 235–292.
- Bertsimas, D., V. Gupta, and N. Kallus (2018b). Robust sample average approximation. *Math. Programming*, **171**(1-2), pp. 217–282.
- Bertsimas, D. and M. Sim (2004). The price of robustness. *Oper. Res.*, **52**(1), pp. 35–53.
- Bertsimas, D. and A. Thiele (2006). Robust and data-driven optimization: modern decision making under uncertainty. In *Models, methods, and applications for innovative decision making*, pp. 95–122. INFORMS.
- Bradstreet, D. (2020). Business Impact of the Coronavirus.
- Calculator, U. I. (2020). Historical Inflation Rates: 1914-2020. <https://www.usinflationcalculator.com/inflation/historical-inflation-rates/>. Accessed: October 5 2020.
- Casella, G. and R. L. Berger (2021). *Statistical inference*. Cengage Learning.
- Chen, X., Q. Lin, and G. Xu (2019). Distributionally Robust Optimization with Confidence Bands for Probability Density Functions. *arXiv preprint arXiv:1901.02169*.
- Cheng, J., R. L.-Y. Chen, H. N. Najm, A. Pinar, C. Safta, and J.-P. Watson (2018). Distributionally Robust Optimization with Principal Component Analysis. *SIAM J. Optim.*, **28**(2), pp. 1817–1841.
- Cheng, J., E. Delage, and A. Lisser (2014). Distributionally robust stochastic knapsack problem. *SIAM J. Optim.*, **24**(3), pp. 1485–1506.
- Cheng, J., J. Leung, and A. Lisser (2016). New reformulations of distributionally robust shortest path problem. *Comput. & Oper. Res.*, **74**, pp. 196–204.
- Choi, T. Y., D. Rogers, and B. Vakil (2020). Coronavirus is a wake-up call for supply chain management. *Harvard Business Review*, **27**, pp. 364–398.
- Chopra, S. (2020). The Coronavirus Has Upended Supply Chains. Here’s How Companies Can Prepare for the Next Disrup-

tion. <https://insight.kellogg.northwestern.edu/article/coronavirus-upended-supply-chains-how-companies-can-prepare-disruption>. Accessed: October 5 2020.

- Chowdhury, P., S. K. Paul, S. Kaisar, and M. A. Muktadir (2021). COVID-19 pandemic related supply chain studies: a systematic review. *Transportation Research Part E: Logistics and Transportation Review*, p. 102271.
- Dai, X., X. Wang, R. He, W. Du, W. Zhong, L. Zhao, and F. Qian (2020). Data-driven robust optimization for crude oil blending under uncertainty. *Computers & Chemical Engineering*, **136**, p. 106595.
- Delage, E. and Y. Ye (2010). Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Oper. Res.*, **58**(3), pp. 595–612.
- Dolgui, A., D. Ivanov, and B. Sokolov (2020). Reconfigurable supply chain: the X-network. *International Journal of Production Research*, **58**(13), pp. 4138–4163.
- Du, X. and T. E. Graedel (2011). Global rare earth in-use stocks in NdFeB permanent magnets. *Journal of Industrial Ecology*, **15**(6), pp. 836–843.
- DuHadway, S., S. Carnovale, and B. Hazen (2019). Understanding risk management for intentional supply chain disruptions: Risk detection, risk mitigation, and risk recovery. *Annals of Operations Research*, **283**(1), pp. 179–198.
- Eggert, R., C. Wadia, C. Anderson, D. Bauer, F. Fields, L. Meinert, and P. Taylor (2016). Rare earths: market disruption, innovation, and global supply chains. *Annual Review of Environment and Resources*, **41**, pp. 199–222.
- El Ghaoui, L. E., M. Oks, and F. Oustry (2003). Worst-case value-at-risk and robust portfolio optimization: A conic programming approach. *Oper. Res.*, **51**(4), pp. 543–556.
- Erdođan, E. and G. Iyengar (2006). Ambiguous chance constrained problems and robust optimization. *Math. Programming*, **107**(1-2), pp. 37–61.
- Esfahani, P. M. and D. Kuhn (2018). Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations. *Math. Programming*, **171**(1-2), pp. 115–166.
- Fattahi, M. and K. Govindan (2018). A multi-stage stochastic program for the sustainable design of biofuel supply chain networks under biomass supply uncertainty and disruption risk: A real-life case study. *Transportation Research*

*Part E: Logistics and Transportation Review*, **118**, pp. 534–567.

- Fattahi, M., K. Govindan, and E. Keyvanshokoo (2017). Responsive and resilient supply chain network design under operational and disruption risks with delivery lead-time sensitive customers. *Transportation Research Part E: Logistics and Transportation Review*, **101**, pp. 176–200.
- Gabrel, V., C. Murat, and A. Thiele (2014). Recent advances in robust optimization: An overview. *European journal of operational research*, **235**(3), pp. 471–483.
- Gao, R. and A. J. Kleywegt (2017). Distributionally robust stochastic optimization with dependence structure. *arXiv preprint arXiv:1701.04200*.
- Golan, M. S., L. H. Jernegan, and I. Linkov (2020). Trends and applications of resilience analytics in supply chain modeling: systematic literature review in the context of the COVID-19 pandemic. *Environment Systems and Decisions*, **40**, pp. 222–243.
- Golev, A., M. Scott, P. D. Erskine, S. H. Ali, and G. R. Ballantyne (2014). Rare earths supply chains: Current status, constraints and opportunities. *Resources Policy*, **41**, pp. 52–59.
- Gotoh, J.-y., M. J. Kim, and A. E. Lim (2018). Robust empirical optimization is almost the same as mean–variance optimization. *Oper. Res. Lett.*, **46**(4), pp. 448–452.
- Govindan, K., M. Fattahi, and E. Keyvanshokoo (2017). Supply chain network design under uncertainty: A comprehensive review and future research directions. *European Journal of Operational Research*, **263**(1), pp. 108–141.
- Grant, M. and S. Boyd (2014). CVX: Matlab software for disciplined convex programming, version 2.1.
- Grant, M. C. and S. P. Boyd (2008). Graph implementations for nonsmooth convex programs. In *Recent Advances in Learning and Control*, pp. 95–110. Springer.
- Gurvich, V. J. and A. S. Hussain (2020). In and beyond COVID-19: US academic pharmaceutical science and engineering community must engage to meet critical National Needs. *AAPS PharmSciTech*, **21**, pp. 1–4.
- Habib, K., S. I. T. Hansd ttir, and H. Habib (2020). Critical metals for electromobility: Global demand scenarios for passenger vehicles, 2015–2050. *Resources, Conservation and Recycling*, **154**, p. 104603.

- Habib, K. and H. Wenzel (2014). Exploring rare earths supply constraints for the emerging clean energy technologies and the role of recycling. *Journal of Cleaner Production*, **84**, pp. 348–359.
- Hamidieh, A. and M. Fazli-Khalaf (2017). A possibilistic reliable and responsive closed loop supply chain network design model under uncertainty. *Journal of Advanced Manufacturing Systems*, **16**(04), pp. 317–338.
- Hanasusanto, G. A. and D. Kuhn (2018). Conic programming reformulations of two-stage distributionally robust linear programs over wasserstein balls. *Oper. Res.*, **66**(3), pp. 849–869.
- Handwerker, B., C. & Olson (2019). Value Recovery from Used Electronics, Phase 2.
- Hasani, A. and A. Khosrojerdi (2016). Robust global supply chain network design under disruption and uncertainty considering resilience strategies: A parallel memetic algorithm for a real-life case study. *Transportation Research Part E: Logistics and Transportation Review*, **87**, pp. 20–52.
- Hatefi, S. M. and F. Jolai (2014). Robust and reliable forward–reverse logistics network design under demand uncertainty and facility disruptions. *Applied Mathematical Modelling*, **38**(9-10), pp. 2630–2647.
- Hatefi, S. M., F. Jolai, S. A. Torabi, and R. Tavakkoli-Moghaddam (2015). A credibility-constrained programming for reliable forward–reverse logistics network design under uncertainty and facility disruptions. *International Journal of Computer Integrated Manufacturing*, **28**(6), pp. 664–678.
- Heckmann, I., T. Comes, and S. Nickel (2015). A critical review on supply chain risk–Definition, measure and modeling. *Omega*, **52**, pp. 119–132.
- Hishamuddin, H., R. A. Sarker, and D. Essam (2013). A recovery model for a two-echelon serial supply chain with consideration of transportation disruption. *Computers & Industrial Engineering*, **64**(2), pp. 552–561.
- Hobbs, J. E. (2020). Food supply chains during the COVID-19 pandemic. *Canadian Journal of Agricultural Economics/Revue canadienne d'agroeconomie*, **68**(2), pp. 171–176.
- Hosseini, S., D. Ivanov, and A. Dolgui (2019). Review of quantitative methods for supply chain resilience analysis. *Transportation Research Part E: Logistics and Transportation Review*, **125**, pp. 285–307.

- Hosseini-Motlagh, S.-M., M. R. G. Samani, and S. Homaei (2020). Blood supply chain management: robust optimization, disruption risk, and blood group compatibility (a real-life case). *Journal of Ambient Intelligence and Humanized Computing*, **11**(3), pp. 1085–1104.
- Hu, Z. and L. J. Hong (2013). Kullback-Leibler divergence constrained distributionally robust optimization. *Available at Optimization Online*.
- Ivanov, D. (2020). Viable supply chain model: integrating agility, resilience and sustainability perspectives—lessons from and thinking beyond the COVID-19 pandemic. *Annals of Operations Research*, pp. 1–21.
- Ivanov, D. and A. Dolgui (2020). A digital supply chain twin for managing the disruption risks and resilience in the era of Industry 4.0. *Production Planning & Control*, pp. 1–14.
- Jabbarzadeh, A., B. Fahimnia, J.-B. Sheu, and H. S. Moghadam (2016). Designing a supply chain resilient to major disruptions and supply/demand interruptions. *Transportation Research Part B: Methodological*, **94**, pp. 121–149.
- Jin, H., P. Afiuny, S. Dove, G. Furlan, M. Zakotnik, Y. Yih, and J. W. Sutherland (2018). Life cycle assessment of neodymium-iron-boron magnet-to-magnet recycling for electric vehicle motors. *Environmental science & technology*, **52**(6), pp. 3796–3802.
- Jin, H., B. D. Song, Y. Yih, and J. W. Sutherland (2019). A bi-objective network design for value recovery of neodymium-iron-boron magnets: A case study of the United States. *Journal of Cleaner Production*, **211**, pp. 257–269.
- Khalili, S. M., F. Jolai, and S. A. Torabi (2017). Integrated production–distribution planning in two-echelon systems: a resilience view. *International Journal of Production Research*, **55**(4), pp. 1040–1064.
- Khassiba, A., F. Bastin, S. Cafieri, B. Gendron, and M. Mongeau (2020). Two-stage stochastic mixed-integer programming with chance constraints for extended aircraft arrival management. *Transportation Science*, **54**(4), pp. 897–919.
- Kleywegt, A. J., A. Shapiro, and T. Homem-de Mello (2002). The sample average approximation method for stochastic discrete optimization. *SIAM Journal on Optimization*, **12**(2), pp. 479–502.
- Lappas, N. H. and C. E. Gounaris (2016). Multi-stage adjustable robust optimization for process scheduling under uncertainty. *AIChE Journal*, **62**(5), pp.

1646–1667.

- Li, B., R. Jiang, and J. L. Mathieu (2019). Ambiguous risk constraints with moment and unimodality information. *Math. Programming*, **173**(1-2), pp. 151–192.
- Li, Q., B. Zeng, and A. Savachkin (2013). Reliable facility location design under disruptions. *Computers & Operations Research*, **40**(4), pp. 901–909.
- Li, Z., W. Wu, B. Zhang, and X. Tai (2018). Kullback–Leibler divergence-based distributionally robust optimisation model for heat pump day-ahead operational schedule to improve PV integration. *IET Gener., Trans. & Distrib.*, **12**(13), pp. 3136–3144.
- Lin, C.-C. and T.-H. Wang (2011). Build-to-order supply chain network design under supply and demand uncertainties. *Transportation Research Part B: Methodological*, **45**(8), pp. 1162–1176.
- Listeş, O. (2007). A generic stochastic model for supply-and-return network design. *Computers & Operations Research*, **34**(2), pp. 417–442.
- Liu, Y., X. Yuan, and J. Zhang (2019). Discrete Approximation Scheme in Distributionally Robust Optimization. [http://www.optimization-online.org/DB\\_FILE/2019/01/7040.pdf](http://www.optimization-online.org/DB_FILE/2019/01/7040.pdf).
- Lorca, A. and X. A. Sun (2014). Adaptive robust optimization with dynamic uncertainty sets for multi-period economic dispatch under significant wind. *IEEE Transactions on Power Systems*, **30**(4), pp. 1702–1713.
- Lücker, F., R. W. Seifert, and I. Biçer (2019). Roles of inventory and reserve capacity in mitigating supply chain disruption risk. *International Journal of Production Research*, **57**(4), pp. 1238–1249.
- Luedtke, J. (2016). Benders Decomposition for Solving Two-stage Stochastic Optimization Models. Institute for Mathematics and its Applications.
- Mak, H.-Y. and Z.-J. Shen (2012). Risk diversification and risk pooling in supply chain design. *IIE transactions*, **44**(8), pp. 603–621.
- Margellos, K., P. Goulart, and J. Lygeros (2014). On the road between robust optimization and the scenario approach for chance constrained optimization problems. *IEEE Transactions on Automatic Control*, **59**(8), pp. 2258–2263.
- Mehrotra, S., H. Rahimian, M. Barah, F. Luo, and K. Schantz (2020). A model of supply-chain decisions for resource sharing with an application to ventilator

- allocation to combat COVID-19. *Naval Research Logistics (NRL)*, **67**(5), pp. 303–320.
- Mieth, R. and Y. Dvorkin (2018). Data-driven distributionally robust optimal power flow for distribution systems. *IEEE Control Syst. Lett.*, **2**(3), pp. 363–368.
- Mohseni, S. and M. S. Pishvaei (2020). Data-driven robust optimization for wastewater sludge-to-biodiesel supply chain design. *Computers & Industrial Engineering*, **139**, p. 105944.
- Namdar, J., X. Li, R. Sawhney, and N. Pradhan (2018). Supply chain resilience for single and multiple sourcing in the presence of disruption risks. *International Journal of Production Research*, **56**(6), pp. 2339–2360.
- Natarajan, K., M. Sim, and J. Uichanco (2010). Tractable robust expected utility and risk models for portfolio optimization. *Math. Finance*, **20**(4), pp. 695–731.
- Ning, C. and F. You (2018a). Data-driven adaptive robust optimization framework based on principal component analysis. In *2018 Annual American Control Conference (ACC)*, pp. 3020–3025. IEEE.
- Ning, C. and F. You (2018b). Data-driven decision making under uncertainty integrating robust optimization with principal component analysis and kernel smoothing methods. *Computers & Chemical Engineering*, **112**, pp. 190–210.
- Niu, S., S. Song, J.-Y. Ding, Y. Zhang, and R. Chiong (2019). Distributionally robust single machine scheduling with the total tardiness criterion. *Comput. & Oper. Res.*, **101**, pp. 13–28.
- Nlebedim, I. and A. H. King (2018). Addressing criticality in rare earth elements via permanent magnets recycling. *Jom*, **70**(2), pp. 115–123.
- Olson, D. L. and D. D. Wu (2013). Extreme-event risk management: a review of “Lee, B., Preston, F. 2012. Preparing for High-impact, Low-probability Events: Lessons from Eyjafjallajökull. London: Chatham House.”. *Journal of cleaner production*, **53**, pp. 67–68.
- Paul, S. K. and P. Chowdhury (2020a). A production recovery plan in manufacturing supply chains for a high-demand item during COVID-19. *International Journal of Physical Distribution & Logistics Management*.
- Paul, S. K. and P. Chowdhury (2020b). Strategies for managing the impacts of disruptions during COVID-19: an example of toilet paper. *Global Journal of Flexible Systems Management*, **21**(3), pp. 283–293.

- Paul, S. K., R. Sarker, and D. Essam (2014). Real time disruption management for a two-stage batch production–inventory system with reliability considerations. *European Journal of Operational Research*, **237**(1), pp. 113–128.
- Paul, S. K., R. Sarker, and D. Essam (2016). Managing risk and disruption in production-inventory and supply chain systems: A review. *Journal of Industrial and Management Optimization*.
- Paul, S. K., R. Sarker, and D. Essam (2017). A quantitative model for disruption mitigation in a supply chain. *European Journal of Operational Research*, **257**(3), pp. 881–895.
- Pishvae, M. S., M. Rabbani, and S. A. Torabi (2011). A robust optimization approach to closed-loop supply chain network design under uncertainty. *Applied Mathematical Modelling*, **35**(2), pp. 637–649.
- Pishvae, M. S. and S. A. Torabi (2010). A possibilistic programming approach for closed-loop supply chain network design under uncertainty. *Fuzzy sets and systems*, **161**(20), pp. 2668–2683.
- Qiu, R., Y. Sun, Z.-P. Fan, and M. Sun (2019). Robust multi-product inventory optimization under support vector clustering-based data-driven demand uncertainty set. *Soft Computing*, pp. 1–17.
- Rahimian, H. and S. Mehrotra (2019). Distributionally robust optimization: A review. *arXiv preprint arXiv:1908.05659*.
- Rahmaniani, R., T. G. Crainic, M. Gendreau, and W. Rei (2017). The Benders decomposition algorithm: A literature review. *European Journal of Operational Research*, **259**(3), pp. 801–817.
- Reris, R. and J. P. Brooks (2015). Principal component analysis and optimization: A tutorial.
- Sabouhi, F., M. S. Pishvae, and M. S. Jabalameli (2018). Resilient supply chain design under operational and disruption risks considering quantity discount: A case study of pharmaceutical supply chain. *Computers & Industrial Engineering*, **126**, pp. 657–672.
- Sadghiani, N. S., S. Torabi, and N. Sahebjamnia (2015). Retail supply chain network design under operational and disruption risks. *Transportation Research Part E: Logistics and Transportation Review*, **75**, pp. 95–114.
- Saha, A. K., A. Paul, A. Azeem, and S. K. Paul (2020). Mitigating partial-disruption

- risk: A joint facility location and inventory model considering customers' preferences and the role of substitute products and backorder offers. *Computers & Operations Research*, **117**, p. 104884.
- Scarf, H. (1958). A min-max solution of an inventory problem. *Studies in the Math. Theory of Inventory and Production*.
- Schmid, M. (2019). Rare earths in the trade dispute between the US and China: A déjà vu. *Intereconomics*, **54**(6), pp. 378–384.
- Scott, D. W. (2015). The curse of dimensionality and dimension reduction. *Multivariate Density Estimation: Theory, Practice, and Visualization*, pp. 195–217.
- Shang, C., X. Huang, and F. You (2017). Data-driven robust optimization based on kernel learning. *Computers & Chemical Engineering*, **106**, pp. 464–479.
- Shapiro, A. (2001). On duality theory of conic linear problems. In *Semi-infinite programming*, pp. 135–165. Springer.
- Shapiro, A., D. Dentcheva, and A. Ruszczyński (2009). *Lectures on Stochastic Programming: Modeling and Theory*. SIAM.
- Shapiro, A. and T. Homem-de Mello (1998). A simulation-based approach to two-stage stochastic programming with recourse. *Mathematical Programming*, **81**(3), pp. 301–325.
- Shaw, S. and S. Constantinides (2012). Permanent magnets: the demand for rare earths. In *8th International Rare Earths Conference*.
- Shekarian, M. and M. Mellat Parast (2020). An Integrative approach to supply chain disruption risk and resilience management: a literature review. *International Journal of Logistics Research and Applications*, pp. 1–29.
- Shen, F., L. Zhao, W. Du, W. Zhong, and F. Qian (2020). Large-scale industrial energy systems optimization under uncertainty: A data-driven robust optimization approach. *Applied Energy*, **259**, p. 114199.
- Sherman, E. (2020). 94% of the Fortune 1000 are seeing coronavirus supply chain disruptions: Report. <https://fortune.com/2020/02/21/fortune-1000-coronavirus-china-supply-chain-impact/>. Accessed: October 5 2020.
- Shih, W. (2020). Is it time to rethink globalized supply chains? <https://sloanreview.mit.edu/article/>

- [is-it-time-to-rethink-globalized-supply-chains/](#). Accessed: April 21 2021.
- Shukla, A., V. A. Lalit, and V. Venkatasubramanian (2011). Optimizing efficiency-robustness trade-offs in supply chain design under uncertainty due to disruptions. *International Journal of Physical Distribution & Logistics Management*.
- Snyder, L. V., Z. Atan, P. Peng, Y. Rong, A. J. Schmitt, and B. Sinsoysal (2016). OR/MS models for supply chain disruptions: A review. *Iie Transactions*, **48**(2), pp. 89–109.
- Soyster, A. L. (1973). Convex programming with set-inclusive constraints and applications to inexact linear programming. *Operations research*, **21**(5), pp. 1154–1157.
- Sözüer, S. and A. C. Thiele (2016). The state of robust optimization. In *Robustness analysis in decision aiding, optimization, and analytics*, pp. 89–112. Springer.
- Tang, C. S. (2006). Perspectives in supply chain risk management. *International journal of production economics*, **103**(2), pp. 451–488.
- Teimuory, E., F. Atoei, E. Mohammadi, and A. Amiri (2013). A multi-objective reliable programming model for disruption in supply chain. *Management Science Letters*, **3**(5), pp. 1467–1478.
- Tharumarajah, R. and P. Koltun (2011). Cradle to gate assessment of environmental impact of rare earth metals. In *Proceedings of the 7th Australian Conference on Life Cycle Assessment, Melbourne, Australia*, pp. 9–10.
- Tomlin, B. (2006). On the value of mitigation and contingency strategies for managing supply chain disruption risks. *Management science*, **52**(5), pp. 639–657.
- Torabi, S., M. Baghersad, and S. Mansouri (2015). Resilient supplier selection and order allocation under operational and disruption risks. *Transportation Research Part E: Logistics and Transportation Review*, **79**, pp. 22–48.
- Torabi, S., J. Namdar, S. Hatefi, and F. Jolai (2016). An enhanced possibilistic programming approach for reliable closed-loop supply chain network design. *International Journal of Production Research*, **54**(5), pp. 1358–1387.
- Van der Vaart, A. W. (2000). *Asymptotic Statistics*, volume 3. Cambridge university press.
- Van Slyke, R. M. and R. Wets (1969). L-shaped linear programs with applications

- to optimal control and stochastic programming. *SIAM Journal on Applied Mathematics*, **17**(4), pp. 638–663.
- Vandenberghe, L. and S. Boyd (1996). Semidefinite programming. *SIAM Rev.*, **38**(1), pp. 49–95.
- Wang, C., R. Gao, F. Qiu, J. Wang, and L. Xin (2018). Risk-based distributionally robust optimal power flow with dynamic line rating. *IEEE Trans. on Power Systems*, **33**(6), pp. 6074–6086.
- Wang, Z., P. W. Glynn, and Y. Ye (2016). Likelihood robust optimization for data-driven problems. *Comput. Management Sci.*, **13**(2), pp. 241–261.
- Wold, S., K. Esbensen, and P. Geladi (1987). Principal Component Analysis. *Chemom. Intell. Lab. Syst.*, **2**(1-3), pp. 37–52.
- Worldometer (2020). COVID-19 Coronavirus Pandemic. <https://www.worldometers.info/coronavirus/>. Accessed: October 5 2020.
- Xu, H., Y. Liu, and H. Sun (2018). Distributionally robust optimization with matrix moment constraints: Lagrange duality and cutting plane methods. *Math. Programming*, **169**(2), pp. 489–529.
- Yang, Y. and W. Wu (2019). A Distributionally Robust Optimization Model for Real-time Power Dispatch in Distribution Networks. *IEEE Trans. on Smart Grid*, **10**(4), pp. 3743–3752.
- Zahiri, B., J. Zhuang, and M. Mohammadi (2017). Toward an integrated sustainable-resilient supply chain: A pharmaceutical case study. *Transportation Research Part E: Logistics and Transportation Review*, **103**, pp. 109–142.
- Zhalechian, M., S. A. Torabi, and M. Mohammadi (2018). Hub-and-spoke network design under operational and disruption risks. *Transportation research part E: logistics and transportation review*, **109**, pp. 20–43.
- Zhang, Y., S. Shen, and S. A. Erdogan (2018). Solving 0–1 semidefinite programs for distributionally robust allocation of surgery blocks. *Optim. Lett.*, **12**(7), pp. 1503–1521.
- Zhao, L., C. Ning, and F. You (2019). A Data-Driven Robust Optimization Approach to Operational Optimization of Industrial Steam Systems under Uncertainty. In *Computer Aided Chemical Engineering*, volume 46, pp. 1399–1404. Elsevier.