Summer 2017

Optimization under uncertainty: conic programming representations, relaxations, and approximations

Guanglin Xu
University of Iowa

OPTIMIZATION UNDER UNCERTAINTY:

CONIC PROGRAMMING REPRESENTATIONS, RELAXATIONS,

AND APPROXIMATIONS

by

Guanglin Xu

A thesis submitted in partial fulfillment of the
requirements for the Doctor of Philosophy
degree in Business Administration (Management Sciences)
in the Graduate College of
The University of Iowa

August 2017

Thesis Supervisor: Professor Samuel A. Burer

Graduate College
The University of Iowa
Iowa City, Iowa

CERTIFICATE OF APPROVAL

PH.D. THESIS

This is to certify that the Ph.D. thesis of

Guanglin Xu

has been approved by the Examining Committee for the thesis requirement for the Doctor of Philosophy degree in Business Administration (Management Sciences) at the August 2017 graduation.

Thesis Committee: _____
Samuel A. Burer, Thesis Supervisor

_____
Kurt M. Anstreicher

_____
Qihang Lin

_____
Tianbao Yang

_____
Luis F. Zuluaga

To my family

ACKNOWLEDGEMENTS

I give my deepest gratitude to my advisor and mentor, Professor Samuel A. Burer. I cannot imagine that I could complete this thesis without his insightful discussions, continued encouragement, and unconditional support. I am extremely thankful for his incredible patience and extraordinary enthusiasm, which are contagious and motivational for me, even during the toughest times. He taught me how to think critically, write precisely, and present clearly. He encouraged me to explore research topics that I am interested in. He also showed me how to be a rigorous and effective teacher. *"More importantly",* he set up an excellent example as a cool professor, which is demanding in the research community. Thank you, Sam, for all your invaluable guidance. It has been a great honor and a wonderful journey to be your Ph.D. student. I owe you a debt of gratefulness that is beyond any word in any language.

I was also extremely fortunate to have Professor Kurt Anstreicher, Professor Qihang Lin, Professor Tianbao Yang, and Professor Luis Zuluaga as my committee members, who not only provide me with invaluable research suggestions but also serve as a continuous source of advice. In particular, I would like to thank Prof. Anstreicher for his inspirational suggestions to my research. I also appreciate his warm encouragement to my presentation. I would like to thank Prof. Lin for discussing and conducting research with me. Thank you for your kindness to answer all my questions. I would like to thank Prof. Yang for introducing me to big data techniques. I would

like to thank Prof. Zuluaga for inviting me to present my research at COR@L Seminar and providing insightful thoughts on my thesis.

My research and study at Tippie were facilitated with the help of many professors, administrative staff, and my colleagues. I would like to thank Professor Johannes Ledolter for being in my preliminary committee and providing me with insightful suggestions. I was also lucky to have excellent teachers including Professors Ann Campbell, Young Chen, Pavlo Krokhmal, Renato de Matta, Jeff Ohlmann, and Nick Street. Your teaching has significantly strengthened my research. I sincerely appreciate all the help from the departmental administrative staff including Renea Jay, Scott Hansen, and Barbara Carr. Your excellent assistance makes my work and life at Tippie much easier. I am thankful for my colleagues for creating an incredibly friendly office atmosphere. Thank you, Huan Jin, Amin Khezerlou, Xiexin Liu, Yuanyang Liu, Runchao Ma, Maryam Rahmani, Michael Redmond, John Rios, Venous Roshidibenam, Stacy Voccia, Wenjun Wang, Nicholas Woyak, Qiong Zhang, and Shu Zhang.

Wine confidant thousands of cups less. Life in Iowa City would have been less joyful if I had not met my closest friends at the University of Iowa. I will never forget all the wonderful moments I spent together with you all. Thank you, Maciej Rysz, Xiupeng Wei, Yaohui Zeng, Xiafei He, Anoop Verma, Bhupesh Shetty, Zhongyi Yuan, Xi Chen, and Bo Sun.

Last but not the least, I would like to thank my parents, my parents-in-law, my wife Shuai, and my two lovely sons Charlie and Mark. I sincerely thank them for

staying with me throughout the happiest and toughest times. To them I attribute this thesis.

ABSTRACT

In practice, the presence of uncertain parameters in optimization problems introduces new challenges in modeling and solvability to operations research. There are three main paradigms proposed for optimization problems under uncertainty. These include *stochastic programming, robust optimization*, and *sensitivity analysis*. In this thesis, we examine, improve, and combine the latter two paradigms in several relevant models and applications.

In the second chapter, we study a two-stage adjustable robust linear optimization problem in which the right-hand sides are uncertain and belong to a compact, convex, and tractable uncertainty set. Under standard and simple assumptions, we reformulate the two-stage problem as a copositive optimization program, which in turns leads to a class of tractable semidefinite-based approximations that are at least as strong as the affine policy, which is a well studied tractable approximation in the literature. We examine our approach over several examples from the literature and the results demonstrate that our tractable approximations significantly improve the affine policy. In particular, our approach recovers the optimal values of a class of instances of increasing size for which the affine policy admits an arbitrary large gap.

In the third chapter, we leverage the concept of robust optimization to conduct sensitivity analysis of the optimal value of linear programming (LP). In particular, we propose a framework for sensitivity analysis of LP problems, allowing for simultaneous perturbations in the objective coefficients and right-hand sides, where the

perturbations are modeled in a compact, convex, and tractable uncertainty set. This framework unifies and extends multiple approaches for LP sensitivity analysis in the literature and has close ties to worst-case LP and two-stage adjustable linear programming. We define the best-case and worst-case LP optimal values over the uncertainty set. As the concept aligns well with the general spirit of robust optimization, we denote our approach as *robust sensitivity analysis*. While the best-case and worst-case optimal values are difficult to compute in general, we prove that they equal the optimal values of two separate, but related, copositive programs. We then develop tight, tractable conic relaxations to provide bounds on the best-case and worst case optimal values, respectively. We also develop techniques to assess the quality of the bounds, and we validate our approach computationally on several examples fromand inspired bythe literature. We find that the bounds are very strong in practice and, in particular, are at least as strong as known results for specific cases from the literature.

In the fourth chapter of this thesis, we study the expected optimal value of a mixed 0-1 programming problem with uncertain objective coefficients following a joint distribution. We assume that the true distribution is not known exactly, but a set of independent samples can be observed. Using the Wasserstein metric, we construct an ambiguity set centered at the empirical distribution from the observed samples and containing all distributions that could have generated the observed samples with a high confidence. The problem of interest is to investigate the bound on the expected optimal value over the Wasserstein ambiguity set. Under standard assumptions, we reformulate the problem into a copositive programming problem, which naturally

leads to a tractable semidefinite-based approximation. We compare our approach with a moment-based approach from the literature for two applications. The numerical results illustrate the effectiveness of our approach.

Finally, we conclude the thesis with remarks on some interesting open questions in the field of optimization under uncertainty. In particular, we point out that some interesting topics that can be potentially studied by copositive programming techniques.

PUBLIC ABSTRACT

Decision-makers involved in business, economic, or industrial activities often encounter different uncertainties in the process of making decisions. For instance, an operations manager needs to decide how much items to purchase at the beginning of each planning period under a situation where customer demands are not known exactly; a portfolio manager needs to choose a set of assets to buy in before knowing the actual returns; or a wind farm operator needs to make a schedule for the next day without knowing the exact wind distributions and electricity demands. Therefore, it is necessary to develop effective tools to mitigate the negative impact of the uncertainties on these decision-making processes.

In this thesis, we apply techniques including mathematical programming, data analytics, statistics, and algorithmic computations to conduct two research topics in decision making under uncertainty. In the first topic, we study a two-stage adjustable robust optimization problem in which decision makers need to make the first-stage decision before the realizations of uncertainties and then determine the second-stage decision after observing the realizations. We propose an approach to solve the two-stage problem and show that our approach provides a decision that performs at least as well as the one provided by a state-of-the-art approach in the literature. In the second topic, we study the effect of the uncertain parameters on the objectives of the decision making problems. We propose two approaches to investigate the effect: a distribution-free approach and a data-driven distributionally robust approach. We

demonstrate the effectiveness of our approaches in this thesis.

TABLE OF CONTENTS

Page

LIST OF TABLES xiii

LIST OF FIGURES xiv

1 INTRODUCTION 1

1.1 Robust Optimization 2
1.2 Sensitivity Analysis 6
1.3 Distributionally Robust Optimization 10
1.4 Copositive Programming 12
1.5 Thesis Structure 15

2 TWO-STAGE ADJUSTABLE ROBUST LINEAR OPTIMIZATION . 16

2.1 Introduction 16
2.1.1 Notation, terminology, and background 20
2.2 A Copositive Reformulation 22
2.3 The Affine Policy 25
2.4 Improving the Affine Policy 30
2.5 Examples 32
2.5.1 A temporal network example 32
2.5.2 Multi-item newsvendor problem 45
2.5.3 Lot-sizing problem on a network 47
2.5.4 Randomly generated instances 49
2.6 Final Remarks 50

3 ROBUST SENSITIVITY ANALYSIS 52

3.1 Introduction 52
3.1.1 Notation, terminology, and copositive optimization 57
3.2 Robust Sensitivity Analysis 58
3.2.1 The best- and worst-case optimal values 58
3.2.2 Attainment and complexity 63
3.2.3 Finite variants of p^- and p^+ 64
3.3 Copositive Formulations and Relaxations 67
3.3.1 Copositive formulations 67
3.3.2 SDP-based conic relaxations 69
3.3.3 Bounds from feasible solutions 72
3.4 Computational Experiments 74
3.4.1 Examples from classical sensitivity analysis 75

3.4.2 An example from interval linear programming 78
3.4.3 Worst-case linear optimization 79
3.4.4 A network flow problem 81
3.4.5 Additional details 85
3.5 Conclusion 86

4 A DATA-DRIVEN DISTRIBUTIONALLY ROBUST BOUND ON THE EXPECTED OPTIMAL VALUE OF UNCERTAIN MIXED 0-1 LINEAR PROGRAMMING 89

4.1 Introduction 89
4.1.1 Notation, terminology, and basic techniques 94
4.2 A Wasserstein-Based Ambiguity Set 95
4.2.1 An empirical Wasserstein radius 97
4.3 Problem Reformulation and Tractable Bound 99
4.3.1 A copositive reformulation 101
4.3.2 A semidefinite-based relaxation 105
4.4 Numerical Experiments 107
4.4.1 Statistical sensitivity analysis of highest-order statistic . . 108
4.4.2 Project management problem 114
4.5 Concluding Remarks 118

5 CONCLUSIONS AND FUTURE RESEARCH 122

5.1 Conclusions 122
5.2 Future Research and Open Questions 124

REFERENCES 125

LIST OF TABLES

Table Page

2.1 Unit transportation costs t_{ij} associated with pairs of locations. 48

3.1 Costs for each period of an instance of the inventory management problem. 79

3.2 The unit transportation costs associated with the arcs of the network. . . 83

3.3 Results for the transportation network problem. 84

3.4 Statistics on the sizes of problems solved in the examples as reported by Mosek. 85

3.5 Effectiveness of the linearized complementarity constraint. 88

4.1 The percentage of the 100 trials where the optimal values from WB approach are greater or equal to the simulated values over the 8 cases. . . . 114

4.2 The percentage of the 100 trials where the optimal values from WB are greater than or equal to the corresponding simulated optimal values over the 7 cases. 118

LIST OF FIGURES

Figure Page

1.1 Illustration of the structure of a lot-sizing network with two stores and the associated decision variables and problem parameters. 4

3.1 Illustration of Example 3.1. Note that the dashed line in both (3.1a) and (3.1b) corresponds to the feasible region of the nominal problem. 60

3.2 The transportation network of the 5 suppliers and 10 customers. 82

4.1 Reliabilities of different Wasserstein radii for $N \in \{20, 80, 320, 1280\}$ respectively. 111

4.2 Illustration of the comparison of WB and MB over different sample sizes for a particular randomly generated underlying distribution. Note that the moment-based value and simulated value remains the same over all runs since this experiment is only for a particular distribution. 112

4.3 Illustration of the average relative gaps from both MB and WB in the case of $N \in \{10, 20, 40, 80, 160, 320, 640, 1280\}$. The blue line represents the average relative gap between the optimal values from WB and the simulated value; the red line represents the average relative gap between the optimal values from MB and the simulated value. 119

4.4 The structure of a project network. 120

4.5 Illustration of the comparison of WB and MB over different sample sizes for a particular randomly generated underlying distribution. Note that the moment-based value and simulated value remains the same over all runs since this experiment is only for a particular distribution. 120

4.6 Illustration of the average gaps from both MB and WB in the case of $N \in \{10, 20, 40, 80, 160, 320, 640\}$. The blue line represents the average gap between the optimal values from WB and the simulated values; the red line represents the average gap between the optimal values from MB and the simulated values. 121

CHAPTER 1
INTRODUCTION

Linear programming (LP) has become a very powerful decision-making tool in operations research; see for example [47]. However, the presence of parameter uncertainties introduces challenges for LP problems [46]. For example, even small data perturbations can significantly affect the feasibility or optimality of the nominal optimal solution [12]. Optimization problems under uncertainty also require efficient methods to solve the underlying large-scale optimization problems [14]. An uncertain LP problem is defined formally as a family of regular LP problems:

$$\left\{\min_{x}\left\{c^T x : Ax \geq b\right\}\right\}_{(A,b,c)\in\mathcal{U}} \tag{1.1}$$

where x denotes the decision variable, (A, b, c) are uncertain problem parameters, and $\mathcal{U}$ is the uncertainty set that contains all realizations of (A, b, c). In general, the uncertain LP problem can contain infinitely many regular linear programs.

Depending on the treatment of the uncertainties, there are several main approaches for optimization problems under uncertainty. One approach is *stochastic programming*, in which: the probability distributions of uncertain parameters are known; the objective and constraints are defined by averaging over possible outcomes; or the constraints are satisfied with a probabilistic guarantee; see [30, 88, 123]. Despite of its powerful modeling merits, stochastic programming suffers from the so-called curse of dimensionality, even for problems with only linear objective functions and constraints [124, 58]. Moreover, in many cases, there is not sufficient historical data to calibrate

or estimate the underlying probability distribution, especially for joint multivariate cases. In contrast, *robust optimization* (RO) can often lead to tractable optimization problems, which we introduce in the next section.

1.1 Robust Optimization

Robust optimization (RO) [64, 128] specifies a so-called *uncertainty set* which contains all realizations of the uncertain parameters. This might include bounds on the quantities or linear relations linking multiple unknowns. In this case, we do not make any distribution assumptions on the uncertain parameters. The goal is to optimize the objective value in the worst-case scenario, and the corresponding solution should be feasible for all realizations in the uncertainty set. In general, an optimal solution can be computed in polynomial time even for large classes of RO problems; see [12, 19]. Therefore, RO has become a powerful approach; see Ben-Tal and Nemirovski [15, 16, 17, 18], Bertsimas and Sim [28, 29], Ghaoui and Lebret [72], and Goldfarb and Iyengar [75].

At the early stage of RO research, the approach only considered "here-and-now" decision variables. That is, the decision variables are determined before the uncertainty is revealed. Therefore, the RO approach can be too conservative in this single-stage setting. As an extension, Ben-Tal et. al. [14] introduced two-stage *adjustable robust optimization* (ARO), which considers both first-stage ("here-and-now") and second-stage ("wait-and-see") variables. In this two-stage setting, the second-stage decision variables could be specified after the uncertainty is revealed.

Thus, ARO can be significantly less conservative than the regular RO. Furthermore, this two-stage setting can be more adequate to model many real-world applications, in which the decision-making processes have a two-stage nature.

We next tailor an example from [49] to illustrate two situations in which ARO is more adequate for modeling an operations management problem than the regular RO.

Example 1.1. Consider a lot-sizing problem on a network with two stores, in which a supplier needs to allocate stock at two stores and ship certain amounts of stock between the two stores in order to meet the demand from customers, while incurring the lowest cost, which is the sum of the ordering and shipping costs. In a deterministic setting, this could be done by solving the following linear program:

$$\begin{aligned}
\min_{x_1,x_2,y_{12},y_{21}} \quad & c_1x_1 + c_2x_2 + t_{12}y_{12} + t_{21}y_{21} \\
\text{s. t.} \quad & x_1 + y_{21} - y_{12} \geq d_1 \\
& x_2 + y_{12} - y_{21} \geq d_2 \\
& 0 \leq x_1 \leq V_1 \\
& 0 \leq x_2 \leq V_2 \\
& y_{12} \geq 0, y_{21} \geq 0,
\end{aligned}$$

where x_i denotes the stock allocation at store i, y_{ij}denotes the shipping amounts from store i to store j, c_i denotes the per-unit allocation cost at store i, t_{ij} denotes the per-unit shipping costs from store i to store j, d_i is the demand at store i, V_i is the capacity of store i, and $i, j = 1, 2$. Figure 1.1 illustrates the structure of the network and the associated decision variables and problem parameters.

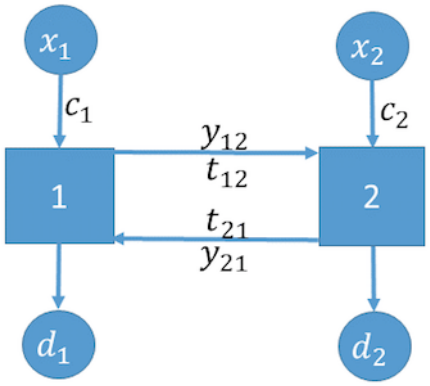


Figure 1.1: Illustration of the structure of a lot-sizing network with two stores and the associated decision variables and problem parameters.

Now consider a situation where the demand is uncertain. More precisely, let us assume that the demand (d_1, d_2) varies in an uncertainty set $\mathcal{U}$, which is convex, compact, and tractable. Then, the regular RO problem is formulated as the following:

$$
\begin{aligned}
\min_{x_1,x_2,y_{12},y_{21}} \quad & c_1x_1 + c_2x_2 + t_{12}y_{12} + t_{21}y_{21} \\
\text{s. t.} \quad & x_1 + y_{21} - y_{12} \geq d_1 \quad \forall\,(d_1, d_2) \in \mathcal{U} \\
& x_2 + y_{12} - y_{21} \geq d_2 \quad \forall\,(d_1, d_2) \in \mathcal{U} \\
& 0 \leq x_1 \leq V_1 \\
& 0 \leq x_2 \leq V_2 \\
& y_{12} \geq 0, y_{21} \geq 0.
\end{aligned}
$$

The above setting assumes that the allocation and shipping decisions have to be made before the uncertainty is revealed. In general, however, it might be more adequate to make the shipping decisions after the uncertain demand is revealed, which

leads to the following ARO problem:

$$
\begin{aligned}
\min_{x_1,x_2,y_{12},y_{21}} \quad & c_1x_1 + c_2x_2 + t_{12}y_{12}(d_1,d_2) + t_{21}y_{21}(d_1,d_2) \\
\text{s.t.} \quad & x_1 + y_{21}(d_1,d_2) - y_{12}(d_1,d_2) \geq d_1 \quad \forall\,(d_1,d_2) \in \mathcal{U} \\
& x_2 + y_{12}(d_1,d_2) - y_{21}(d_1,d_2) \geq d_2 \quad \forall\,(d_1,d_2) \in \mathcal{U} \\
& 0 \leq x_1 \leq V_1, 0 \leq x_2 \leq V_2 \\
& y_{12}(d_1,d_2) \geq 0, y_{21}(d_1,d_2) \geq 0,
\end{aligned}
$$

where $y_{12}(d_1,d_2)$ and $y_{21}(d_1,d_2)$ depend on the uncertainty parameters d_1 and d_2.

Under the adjustable robust setting, the optimal worst-case objective value is less than or equal to the one under the regular robust setting. This is due to the fact that $y_{12}(d_1,d_2)$ and $y_{21}(d_1,d_2)$ depend on the uncertainty parameters and thus can be adjusted after observing the realization of $(d_1,d_2) \in \mathcal{U}$. Furthermore, the adjustable robust setting in this application is more reasonable in practice.

Because of the immense modeling potential, real-world applications of ARO abound: unit commitment in renewable energy [24, 134, 145], facility location problems [6, 8, 68], emergency supply chain planning [11], and inventory management [9, 127]; see also [13, 66, 113]. We refer the reader to the excellent, recent tutorial [49] for background on ARO.

However, ARO is computationally intractable (NP-hard) in general; see [14, 49]. Therefore, different tractable approaches have been proposed to approximate the ARO. In some situations, an optimal solution of the regular RO problem can be used as an approximate solution, one that is in fact optimal in certain settings [16, 22]. On the other hand, the solution from the regular RO performs poorly in general [22]. The

affine policy [14], which forces the second-stage variables to be an affine function of the uncertainty parameters, is another common approximation for the ARO problem, but it is generally suboptimal. For example, Bertsimas and Goyal [21] show that the best affine policy can be $\Omega(m^{1/2-\delta})$ times the optimal cost for any $\delta > 0$, where m is the number of linear constraints.

With the above issues in mind, the first task we complete in the present thesis is improving the affine policy in a setting of two-stage adjustable robust linear optimization problem. Particularly, in Chapter 2, we study a two-stage ARO problem in which the right-hand sides are uncertain and belong to a convex, compact uncertainty set. This problem is NP-hard, and the affine policy is a popular, tractable approximation. We prove that under standard and simple conditions, the two-stage ARO can be reformulated as a copositive program [36, 37], which in turn leads to a class of tractable, semidefinite-based approximations that are at least as strong as the affine policy. We investigate several examples from the literature demonstrating that our tractable approximations significantly improve the affine policy. In particular, our approach solves exactly in polynomial time a class of instances of increasing size for which the affine policy admits an arbitrarily large gap.

1.2 Sensitivity Analysis

Along with stochastic programming and robust optimization, another paradigm for optimization under uncertainty is *sensitivity analysis* (SA), which examines how perturbations in the parameters affect the optimal value and solution. In particular,

let us consider the standard-form linear program:

$$\begin{aligned} \min \quad & \hat{c}^T x \\ \text{s.t.} \quad & \hat{A}x = \hat{b} \\ & x \geq 0 \end{aligned} \tag{1.2}$$

where $x \in \mathbb{R}^n$ is the variable and $(\hat{A}, \hat{b}, \hat{c}) \in \mathbb{R}^{m \times n} \times \mathbb{R}^m \times \mathbb{R}^n$ are the problem parameters. In practice, $(\hat{A}, \hat{b}, \hat{c})$ may not be known exactly or may be predicted to change within a certain region. Ordinary SA considers the change of a single element in $(\hat{A}, \hat{b}, \hat{c})$ and examines the corresponding effects on the optimal basis and tableau; see [47]. SA also extends to the addition of a new variable or constraint, although we do not consider such changes in the present thesis.

Beyond ordinary SA, more sophisticated approaches that allow simultaneous changes in the coefficients $\hat{c}$ or right-hand sides $\hat{b}$ have been proposed by numerous researchers. Bradley et al. [34] discuss the *100-percent rule* that requires specification of directions of increase or decrease from each $\hat{c}_j$ and then guarantees that the same basis remains optimal as long as the sum of fractions, corresponding to the percent of maximum change in each direction derived from ordinary SA, is less than or equal to 1. Wendell [136, 137, 138] develops the *tolerance approach* to find the so-called *maximum tolerance percentage* by which the objective coefficients can be simultaneously and independently perturbed within *a priori* bounds. The tolerance approach also handles perturbations in one row or column of the matrix coefficients [117] or even more general perturbations in all elements of the matrix coefficients under certain

assumptions [116]. Freund [67] investigates the sensitivity of an LP to simultaneous changes in matrix coefficients. In particular, he considers a linear program whose coefficient matrix depends linearly on a scalar parameter θ and studies the effect of small perturbations on the the optimal objective value and solution; see also [89, 91, 108]. Readers are referred to [136] for a survey of approaches for SA of problem (1.2).

To the best of our knowledge, in the context of LP, no authors have considered simultaneous LP parameter changes in a general way, i.e., perturbations in the objective coefficients $\hat{c}$, right-hand sides $\hat{b}$, and constraint coefficients $\hat{A}$ within a general region (not just intervals). The obstacle for doing so is clear: general perturbations lead to nonconvex quadratic programs (QPs), which are NP-hard to solve. The following example illustrates an operations management problem that requires more sophisticated sensitivity analysis.

Example 1.2. Let us consider a transportation network flow problem. Suppose there are m facilities including m_1 suppliers (origins) and m_2 customers (destinations). The transportation network of these facilities is built on n arcs connecting the m_1 suppliers and the m_2 customers. The estimated supply and demand units for each supplier and customer are given as b_i $(i = 1, \ldots, m)$; $b_i < 0$ if the facility i is a supplier and $b_i \geq 0$ if the facility i is a customer.

The estimated unit transportation costs associated with the arcs of the network are given as c_j $(j = 1, \ldots, n)$. However, suppose at the early stage of planning, the supply and demand units and the unit transportation costs are uncertain. Thus, the operations manager would like to quantify the resulting uncertainty in the optimal

transportation cost.

For example, let $\hat{b} \in \mathbb{R}^m$ denote the "best-guess" supply or demand units of the suppliers or customers; and let $\hat{c} \in \mathbb{R}^n_+$ denote the "best-guess" unit transportation costs associated with the arcs of the network. However, due to uncertainty, the supply and demand could vary. Denote by (b, c) the perturbations with respect to $(\hat{b}, \hat{c})$. Thus, the true data could be $(\hat{b} + b, \hat{c} + c)$. In particular, the supply and demand can vary within a certain range, e.g., $\|b\|_2 \leq 0.01\|\hat{b}\|_2$ where b denotes the perturbation in $\hat{b}$. Similarly, the unit transportation cost could actually vary due to the traffic or road conditions of the network, e.g., $\|c\|_2 \leq 0.01\|\hat{c}\|_2$ where c denotes the perturbation in $\hat{c}$. With this uncertainty, the operations manager would like to know the worst- and best-case for the optimal transportation costs on the following uncertain set

$$\mathcal{U} := \left\{ (b, c) \in \mathbb{R}^m \times \mathbb{R}^n : \begin{array}{r} \|b\|_2 \leq 0.01\|\hat{b}\|_2 \\ \|c\|_2 \leq 0.01\|\hat{c}\|_2 \end{array} \right\}.$$

Here, we assume that, before the final decision must be made, the uncertainties will be resolved. At this moment, the operations manager is monitoring what may or may not happen in the future. For example, what are the best-case and worst-case optimal values over all the possible parameter realizations?

In Chapter 3, motivated by Example 1.2, we will leverage the concept of robust optimization to develop a framework for sensitivity analysis of linear programs (LPs), which allows for simultaneous perturbations in the objective coefficients and right-hand sides, where the perturbations are modeled in a compact, convex, tractable

uncertainty set. In particular, we will investigate the best-case and worst-case LP optimal values over the parameter perturbations. We will also show that this framework can unify and extend multiple approaches for LP sensitivity analysis in the literature and has close ties to worst-case linear optimization and two-stage adaptive optimization. Note that here we do not make any distributional assumptions about the uncertainty of the parameters, nor about their independence or dependence. In this sense, our search for the worst- and best-case optimal LP values aligns well with typical approaches in the field of robust optimization. Thus, we denote our approach as *robust sensitivity analysis*.

1.3 Distributionally Robust Optimization

Although stochastic programming is criticized by the fact that it is computationally demanding in general, *distributionally robust optimization* (DRO) opens a new paradigm for optimization under uncertainty; see [41, 50, 74]. DRO stems from Scarf's research work in 1958 studying the ambiguity-averse newsvendor problem [119]. DRO problems assume that some limited information about the distribution can be obtained and that the decision maker can construct an *ambiguity set* that contains all distributions consistent with the available information. It is shown that DRO can provide high-quality decisions at a moderate computational cost [50, 74, 140].

A key ingredient in DRO problems is the ambiguity set containing all possible distributions. According to [63], a good ambiguity set should be rich enough to contain the true underlying distribution with high confidence while still being small

enough to avoid making overly conservative decisions. A simple yet popular way to characterize the ambiguity set is to use the moment information of the distributions [50, 74, 140]. Another interesting way is to build an ambiguity set that contains all unimodal distributions that satisfy some given support constraints [93, 122]. An alternative is to define the ambiguity set as a ball in the space of probability distributions by using a probability distance function such as Prohorov metric [62], Kullback-Leibler divergence [86, 87], or Wasserstein metric [63, 81].

We now point out an interesting connection between DRO and regular RO. Denoting the support of the uncertain parameters as $\mathcal{U}$, if one chooses the ambiguity set to be one that contains distributions that put all their weights at a single point anywhere in $\mathcal{U}$, then the DRO problem reduces to the regular RO problem. Many other ways of formulating ambiguity sets can be found in [56].

Leveraging the concept of DRO, a stream of research investigates the best-case (maximal) expectations of the maxima of mixed zero-one linear programs with objective uncertainty. Natarajan et. al. [107] studied mixed zero-one maximization linear programs with uncertain objective coefficients, in which the mean vector and the second-moment matrix of the nonnegative objective coefficients are assumed to be known, but the exact form of the distribution is unknown. The authors showed that computing the best-case (maximal) expectation of the optimal value (maxima) is equal to computing the optimal value of a copositive program. In some situations, computing the best-case expectation can be reduced to computing a semidefinite program [105]. We refer the reader to a recent excellent survey in [98].

In Chapter 4, motivated by the work in [107], we will investigate the best-case expected optimal value of mixed zero-one linear programs with objective coefficient uncertainty, in which the exact distribution of the cost coefficients is unknown. We construct the ambiguity set of possible distributions by using Wasserstein balls [81]. We will show that the problem can be reformulated as a copositive program. We then provide a tractable semidefinite programming relaxation to approximate the best-case expected optimal value.

1.4 Copositive Programming

Let $\mathcal{K} \subseteq \mathbb{R}^n$ be a closed convex cone. A *linear conic programming* (LCP) problem is defined as

$$p^* := \inf_x \left\{ c^T x : Ax = b, x \in \mathcal{K} \right\} \tag{1.3}$$

and its conic dual problem is

$$d^* := \sup_{y,s} \left\{ b^T y : A^T y + s = c,\ s \in \mathcal{K}^* \right\} \tag{1.4}$$

where $\mathcal{K}^*$ is the dual cone of $\mathcal{K}$. Immediately, we have $d^* \leq p^*$ by weak duality theory. Constraint qualifications can guarantee strong duality to hold, i.e., $p^* = d^*$. A common constraint qualification is the so-called Slater's condition, which we state in Lemma 1.1.

Lemma 1.1 (Theorem 2.1.2 in [97])**.** *Consider the pair of primal and dual problems (1.3) and (1.4).*

(i) Assume that the primal problem (1.3) is bounded from below, i.e., $p^* > -\infty$ and

that it is strictly feasible. Then the dual problem (1.4) attains its supremum d^* and there is no duality gap, i.e., $p^* = d^*$.

(ii) Assume that the dual problem (1.4) is bounded from above, i.e., $d^* < +\infty$ and that it is strictly feasible. Then the primal problem (1.3) attains its infimum p^* and there is no duality gap, i.e., $p^* = d^*$.

If $\mathcal{K} = \mathbb{R}^n_+$, then LCP is regular linear programming. If $\mathcal{K}$ is the *second-order cone*, then LCP is so-called *second-order cone programming*; see [2]. If $\mathcal{K}$ is the cone of positive semidefinite matrices, then LCP is the well studied semidefinite programming problem; see [132]. One of the important properties shared by the three cones is that they are self-dual cones, i.e., $\mathcal{K}^* = \mathcal{K}$.

We next introduce a LCP problem that is closely related to this thesis. For a closed convex cone $\mathcal{K} \subseteq \mathbb{R}^n$, we define the *copositive cone* as

$$\mathcal{COP}(\mathcal{K}) := \{M \in \mathcal{S}^n : x^T M x \geq 0 \ \forall \ x \in \mathcal{K}\},$$

and its dual cone, the *completely positive cone*, is

$$\mathcal{CPP}(\mathcal{K}) := \{X \in \mathcal{S}^n : X = \textstyle\sum_i x^i (x^i)^T, \ x^i \in \mathcal{K}\},$$

where the summation over i is finite but its cardinality is unspecified. Then, the copositive program is given as

$$p^* := \inf_{X \in \mathcal{COP}(\mathcal{K})} \{C \bullet X : A_i \bullet X = b_i \ \forall \, i = 1, \ldots, m\}$$

and its dual is

$$d^* := \sup_{y \in \mathbb{R}^m} \left\{ b^T y : C - \sum_{i=1}^m y_i A_i \in \mathcal{CPP}(\mathcal{K}) \right\}$$

The term *copositive programming* refers to linear programming over $\mathcal{COP}(\mathcal{K})$ or, via duality, linear programming over $\mathcal{CPP}(\mathcal{K})$. In fact, these problems are sometimes called *generalized copositive programming* or *set-semidefinite optimization* [40, 59] in contrast with the standard case $\mathcal{K} = \mathbb{R}^n_+$. In this thesis, we work with generalized copositive programming, although we use the shorter phrase for convenience.

In fact, many standard NP-hard optimization problems can be represented as copositive programs. These include the standard quadratic program [32], maximum stable set problem [48], min-cut tri-partitioning problem [114], and quadratic assignment problem [115]. More generally, Burer [36] showed that any nonconvex quadratic conic program

$$\min \left\{ x^T Q x + 2c^T x : Ax = b,\ x \in \mathcal{K} \right\} \tag{1.5}$$

can be modeled as an explicit copositive program over $\mathcal{CPP}(\mathbb{R}_+ \times \mathcal{K})$. Burer's result in [36] also holds for two specific types of quadratic constraints: binary condition constraints $x_i^2 = x_i$ and complementarity constraints $x_i x_j = 0$ with x_i, x_j bounded.

In general, copositive programming is computationally intractable as the separation of a copositive matrix with an arbitrary dimension is NP-hard. Naturally, different inner or outer approximations are studied. A series of inner approximation hierarchies of $\mathcal{COP}(\mathcal{K})$ for the case of $\mathcal{K} = \mathbb{R}^n_+$ have been proposed in [31, 48, 109, 110]. Lasserre [96] proposed an outer approximation hierarchy of $\mathcal{COP}(\mathcal{K})$ for the case of $\mathcal{K} = \mathbb{R}^n_+$, which complements previous inner approximation hierarchies. In practice, tractable semidefinite programming (SDP) relaxations are generally used to approximate copositive programs over the cone of completely positive matrices. These SDP

relaxations can be further tightened by adding the so-called *reformulation linearization technique* (RLT) constraints. We refer the reader to [4, 125] and Chapter 3 for the definition and concept of RLT constraints.

1.5 Thesis Structure

The rest of the thesis is structured as follows. Chapter 2 proposes a copsotive approach for a two-stage adjustable robust linear optimization problem in which the right-hand sides are uncertainty and belong to a convex, compact, and tractable set. Chapter 3 presents robust sensitivity analysis which is a framework to extend and unify the literature sensitivity analysis by leveraging modern tools for nonconvex quadratic programs. Chapter 4 proposes a distributionally robust approach to analyze the expectation of the optimal value of a mixed zero-one linear optimization problem and the ambiguity set is defined as a Wasserstein ball. Chapter 5 concludes the thesis and discusses several potential future directions of optimization under uncertainty.

CHAPTER 2
TWO-STAGE ADJUSTABLE ROBUST LINEAR OPTIMIZATION

2.1 Introduction

As mentioned in Chapter 1, ARO is intractable in general. Therefore, multiple tractable approximations have been proposed for it. In certain situations, a static, robust-optimization-based solution can be used to approximate ARO, and sometimes this static solution is optimal [16, 22]. The *affine policy* [14], which forces the second-stage variables to be an affine function of the uncertainty parameters, is another common approximation for ARO, but it is generally suboptimal. Several nonlinear policies have also been used to approximate ARO. Chen and Zhang [44] proposed the *extended affine policy* in which the primitive uncertainty set is reparameterized by introducing auxiliary variables after which the regular affine policy is applied. Bertsimas et. al. [23] introduced a more accurate, yet more complicated, approximation which forces the second-stage variables to depend polynomially (with a user-specified, fixed degree) on the uncertain parameters. Their approach yields a hierarchy of Lasserre-type semidefinite approximations and can be extended to multistage robust optimization.

The approaches just described provide upper bounds when ARO is stated as a minimization. On the other hand, a single lower bound can be calculated, for example, by fixing a specific value in the uncertainty set and solving the resulting LP (linear program), and Monte Carlo simulation over the uncertainty set can then

be used to compute a best lower bound. Finally, global approaches for solving ARO exactly include column and constraint generation [144] and Benders decomposition [24, 53].

In this research, we consider the following two-stage adjustable robust linear minimization problem with uncertain right-hand side:

$$\begin{aligned} v^*_{\mathrm{RLP}} := \quad & \min_{x, y(\cdot)} \quad c^T x + \max_{u \in \mathcal{U}} d^T y(u) \\ & \text{s.t.} \quad Ax + By(u) \geq Fu \quad \forall\, u \in \mathcal{U} \\ & \qquad\quad x \in \mathcal{X}, \end{aligned} \tag{RLP}$$

where $A \in \mathbb{R}^{m \times n_1}, B \in \mathbb{R}^{m \times n_2}, c \in \mathbb{R}^{n_1}$, $d \in \mathbb{R}^{n_2}$, $F \in \mathbb{R}^{m \times k}$ and $\mathcal{X} \subseteq \mathbb{R}^{n_1}$ is a closed convex set containing the first-stage decision x. The uncertainty set $\mathcal{U} \subseteq \mathbb{R}^k$ is compact, convex, and nonempty, and in particular we model it as a slice of a closed, convex, full-dimensional cone $\mathcal{K} \subseteq \mathbb{R}_+ \times \mathbb{R}^{k-1}$:

$$\mathcal{U} := \{u \in \mathcal{K} : e_1^T u = u_1 = 1\}, \tag{2.1}$$

where e_1 is the first canonical basic vector in $\mathbb{R}^k$. In other words, $\mathcal{K}$ is the homogenization of $\mathcal{U}$. We choose this homogenized version for notational convenience and note that it allows the modeling of affine effects of the uncertain parameters. The second-stage variable is $y(\cdot)$, which is formally defined as a mapping $y : \mathcal{U} \to \mathbb{R}^{n_2}$. It is clear that (RLP) is equivalent to

$$v^*_{\mathrm{RLP}} = \min_{x \in \mathcal{X}} \; c^T x + \max_{u \in \mathcal{U}} \min_{y(u) \in \mathbb{R}^{n_2}} \{d^T y(u) : By(u) \geq Fu - Ax\}, \tag{2.2}$$

where $y(u)$ is a vector variable specifying the value of $y(\cdot)$ at u.

Regarding (RLP), we make three standard assumptions.

Assumption 2.1. The closed, convex set $\mathcal{X}$ and the closed, convex cone $\mathcal{K}$ are both full-dimensional and computationally tractable.

For example, $\mathcal{X}$ and $\mathcal{K}$ could be represented using a polynomial number of linear, second-order-cone, and semidefinite inequalities, each of which possesses a polynomial-time separation oracle [77].

Assumption 2.2. Problem (RLP) is feasible, i.e., there exists a choice $x \in \mathcal{X}$ and $y(\cdot)$ such that $Ax + By(u) \geq Fu$ for all $u \in \mathcal{U}$.

The existence of an affine policy, which can be checked in polynomial time, is sufficient to establish that Assumption 2.2 holds.

Assumption 2.3. Problem (RLP) is bounded, i.e., v^*_{RLP} is finite.

Note that the negative directions of recession $\{\tau : d^T\tau < 0, B\tau \geq 0\}$ for the innermost LP in (2.2) do not depend on x and u. Hence, in light of Assumptions 2.2 and 2.3, there must exist no negative directions of recession; otherwise, v^*_{RLP} would clearly equal $-\infty$. So every innermost LP in (2.2) is either feasible with bounded value or infeasible. In particular, Assumption 2.2 implies that at least one such LP is feasible with bounded value. It follows that the specific associated dual LP $\max\{(Fu - Ax)^T w : B^T w = d, w \geq 0\}$ is also feasible with bounded value. In particular, the fixed set

$$\mathcal{W} := \{w \geq 0 : B^T w = d\}$$

is nonempty. For this study, we also make one additional assumption:

Assumption 2.4. Problem (RLP) possesses relatively complete recourse, i.e., for all

$x \in \mathcal{X}$ and $u \in \mathcal{U}$, the innermost LP in (2.2) is feasible.

By the above discussion, Assumption 2.4 guarantees that the innermost LP is feasible with bounded value, and hence every dual $\max\{(Fu - Ax)^T w : B^T w = d, w \geq 0\}$ attains its optimal value at an extreme point of $\mathcal{W}$.

In Section 4.3, under Assumptions 2.1–2.4, we reformulate (RLP) as an equivalent copositive program, which first and foremost enables a new perspective on two-stage robust optimization. Compared to many existing copositive approaches for difficult problems, ours exploits copositive duality; indeed, Assumption 2.4 is sufficient for establishing strong duality between the copositive primal and dual. In Section 2.3, we then apply a similar approach to derive a new formulation of the affine policy, which is then, in Section 2.4, directly related to the copositive version of (RLP). This establishes two extremes: on the one side is the copositive representation of (RLP), while on the other is the affine policy. Section 2.4 also proposes semidefinite-based approximations of (RLP) that interpolate between the full copositive program and the affine policy. Finally, in Section 2.5, we investigate several examples from the literature that demonstrate our bounds can significantly improve the affine-policy value. In particular, we prove that our semidefinite approach solves a class of instances of increasing size for which the affine policy admits arbitrarily large gaps. We end the chapter with a short discussion of future directions in Section 2.6.

It is important to note that, even if Assumption 2.4 does not hold, our copositive program still yields a valid upper bound on v^*_{RLP} that is at least as strong as the affine policy. More comments are provided at the end of Section 4.3; see also Section

2.3.

During the completion of this thesis, we became aware of a recent technical report by Ardestani-Jaafari and Delage [7], which introduces an approach for (RLP) that is very similar in spirit to ours in Section 4.3. Whereas we use copositive duality to reformulate (RLP) exactly and then approximate it using semidefinite programming, [7] uses semidefinite duality to approximate (RLP) in one step. We prefer our two-step approach because it clearly separates the use of conic duality from the choice of approximation. We also feel that our derivation is more compact. In addition, [7] focuses mainly on the case when $\mathcal{U}$ is polyhedral, whereas our approach builds semidefinite-based approximations for any $\mathcal{U}$ that can be represented, say, by linear, second-order-cone, and semidefinite inequalities. In Section 2.5.2, we also provide an example in which our semidefinite bound outperforms the semidefinite bound of [7].

2.1.1 Notation, terminology, and background

Let $\mathbb{R}^n$ denote n-dimensional Euclidean space represented as column vectors, and let $\mathbb{R}^n_+$ denote the nonnegative orthant in $\mathbb{R}^n$. For a scalar $p \geq 1$, the p-norm of $v \in \mathbb{R}^n$ is defined $\|v\|_p := (\sum_{i=1}^n |v_i|^p)^{1/p}$, e.g., $\|v\|_1 = \sum_{i=1}^n |v_i|$. We will drop the subscript for the 2-norm, i.e., $\|v\| := \|v\|_2$. For $v, w \in \mathbb{R}^n$, the inner product of v and w is $v^T w := \sum_{i=1}^n v_i w_i$. The symbol $\mathbb{1}_n$ denotes the all-ones vector in $\mathbb{R}^n$.

The space $\mathbb{R}^{m\times n}$ denotes the set of real $m \times n$ matrices, and the trace inner product of two matrices $A, B \in \mathbb{R}^{m\times n}$ is $A \bullet B := \text{trace}(A^T B)$. $\mathcal{S}^n$ denotes the space of $n \times n$ symmetric matrices, and for $X \in \mathcal{S}^n$, $X \succeq 0$ means that X is positive

semidefinite. In addition, diag(X) denotes the vector containing the diagonal entries of X, and Diag(v) is the diagonal matrix with vector v along its diagonal. We denote the null space of a matrix A as Null(A), i.e., Null(A) := $\{x : Ax = 0\}$. For $\mathcal{K} \subseteq \mathbb{R}^n$ a closed, convex cone, $\mathcal{K}^*$ denotes its dual cone. For a matrix A with n columns, the inclusion Rows(A) $\in \mathcal{K}$ indicates that the rows of A—considered as column vectors—are members of $\mathcal{K}$.

We next introduce some basics of *copositive programming* with respect to the cone $\mathcal{K} \subseteq \mathbb{R}^n$. The *copositive cone* is defined as

$$\mathcal{COP}(\mathcal{K}) := \{M \in \mathcal{S}^n : x^T M x \geq 0 \ \forall \ x \in \mathcal{K}\},$$

and its dual cone, the *completely positive cone*, is

$$\mathcal{CPP}(\mathcal{K}) := \{X \in \mathcal{S}^n : X = \textstyle\sum_i x^i (x^i)^T, \ x^i \in \mathcal{K}\},$$

where the summation over i is finite but its cardinality is unspecified. The term *copositive programming* refers to linear optimization over $\mathcal{COP}(\mathcal{K})$ or, via duality, linear optimization over $\mathcal{CPP}(\mathcal{K})$. In fact, these problems are sometimes called *generalized copositive programming* or *set-semidefinite optimization* [40, 59] in contrast with the standard case $\mathcal{K} = \mathbb{R}^n_+$. In this study, we work with generalized copositive programming, although we use the shorter phrase for convenience.

Finally, for the specific dimensions k and m of problem (RLP), we let e_i denote the i-th standard basis vector in $\mathbb{R}^k$, and similarly, f_j denotes the j-th standard basis vector in $\mathbb{R}^m$. We will also use $g_1 := \binom{e_1}{0} \in \mathbb{R}^{k+m}$.

2.2 A Copositive Reformulation

In this section, we construct a copositive representation of (RLP) under Assumptions 2.1–2.4 by first reformulating the inner maximization of (2.2) as a copositive problem and then employing copositive duality.

Within (2.2), define

$$\pi(x) := \max_{u\in\mathcal{U}} \min_{y(u)\in\mathbb{R}^{n_2}} \{d^T y(u) : By(u) \geq Fu - Ax\}.$$

The dual of the inner minimization is $\max_{w\in\mathcal{W}}(Fu - Ax)^T w$, which is feasible as discussed in Assumption 2.2. Hence, strong duality for LP implies

$$\pi(x) = \max_{u\in\mathcal{U}} \max_{w} \{(Fu - Ax)^T w : w \in \mathcal{W}\} = \max_{(u,w)\in\mathcal{U}\times\mathcal{W}} (Fu - Ax)^T w, \tag{2.3}$$

In other words, $\pi(x)$ equals the optimal value of a bilinear program over convex constraints, which is NP-hard in general [92].

It holds also that $\pi(x)$ equals the optimal value of an associated copositive program (see [35, 36] for example), which we now describe. Define

$$z := \begin{pmatrix} u \\ w \end{pmatrix} \in \mathbb{R}^{k+m}, \quad E := \begin{pmatrix} -de_1^T & B^T \end{pmatrix} \in \mathbb{R}^{n_2\times(k+m)}, \tag{2.4}$$

where $e_1 \in \mathbb{R}^k$ is the first coordinate vector, and homogenize via the relationship (2.1) and the definition of $\mathcal{W}$:

$$\begin{aligned} \pi(x) = \max \;\; & (F - Axe_1^T) \bullet wu^T \\ \text{s.t.} \;\; & Ez = 0 \\ & z \in \mathcal{K} \times \mathbb{R}^m_+, \quad g_1^T z = 1, \end{aligned}$$

where g_1 is the first coordinate vector in $\mathbb{R}^{k+m}$. The copositive representation is thus

$$\begin{aligned} \pi(x) = \max\ & (F - Axe_1^T) \bullet Z_{21} \\ \text{s.t.}\ & \operatorname{diag}(EZE^T) = 0 \\ & Z \in \mathcal{CPP}(\mathcal{K} \times \mathbb{R}^m_+), \quad g_1 g_1^T \bullet Z = 1, \end{aligned} \tag{2.5}$$

where Z has the block structure

$$Z = \begin{pmatrix} Z_{11} & Z_{21}^T \\ Z_{21} & Z_{21} \end{pmatrix} \in \mathcal{S}^{k+m}.$$

Note that under positive semidefiniteness, which is implied by the completely positive constraint, the constraint $\operatorname{diag}(EZE^T) = 0$ is equivalent to $ZE^T = 0$; see Proposition 1 of [36], for example. For the majority of this chapter, we will focus on this second version:

$$\begin{aligned} \pi(x) = \max\ & (F - Axe_1^T) \bullet Z_{21} \\ \text{s.t.}\ & ZE^T = 0 \\ & Z \in \mathcal{CPP}(\mathcal{K} \times \mathbb{R}^m_+), \quad g_1 g_1^T \bullet Z = 1. \end{aligned} \tag{2.6}$$

By standard theory [120, corollary 3.2d], the extreme points of $\mathcal{W}$ are contained in a ball $w^T w \leq r_w$, where $r_w > 0$ is a radius that is polynomially computable and representable in the encoding length of the entries of B and d (assuming those entries are rational). Hence, Assumption 2.4 guarantees that the optimal value of $\max\{(Fu - Ax)^T w : B^T w = d, w \geq 0\}$ does not change when $w^T w \leq r_w$ is enforced. In addition, because $\mathcal{U}$ is bounded by Assumption 2.1, there exists a sufficiently large

scalar $r_z > 0$ such that the constraint $z^T z \leq r_z$ is redundant. It follows from these observations that, in the preceding argument, we can enforce $z^T z = u^T u + w^T w \leq r := r_z + r_w$ without cutting off all optimal solutions of (2.3). Thus, the lifted and linearized constraint $I \bullet Z \leq r$ can be added to (2.6) without changing its optimal value, although some feasible directions of recession may be cut off. We arrive at

$$\begin{aligned} \pi(x) = \max\ & (F - Axe_1^T) \bullet Z_{21} \\ \text{s.t.}\ & ZE^T = 0, \quad I \bullet Z \leq r \\ & Z \in \mathcal{CPP}(\mathcal{K} \times \mathbb{R}^m_+), \quad g_1 g_1^T \bullet Z = 1. \end{aligned} \tag{2.7}$$

Letting $\Lambda \in \mathbb{R}^{(k+m)\times n_2}$, $\lambda \in \mathbb{R}$, and $\rho \in \mathbb{R}$ be the respective dual multipliers of $ZE^T = 0$, $g_1 g_1^T \bullet Z = 1$, and $I \bullet Z \leq r$, standard conic duality theory implies the dual of (2.7) is

$$\begin{aligned} \min_{\lambda,\Lambda,\rho}\ & \lambda + r\rho \\ \text{s.t.}\ & \lambda g_1 g_1^T - \tfrac{1}{2}G(x) + \tfrac{1}{2}(E^T\Lambda^T + \Lambda E) + \rho I \in \mathcal{COP}(\mathcal{K} \times \mathbb{R}^m_+) \\ & \rho \geq 0 \end{aligned} \tag{2.8}$$

where

$$G(x) := \begin{pmatrix} 0 & (F - Axe_1^T)^T \\ F - Axe_1^T & 0 \end{pmatrix} \in \mathcal{S}^{k+m}$$

is affine in x. Holding all other dual variables fixed, for $\rho > 0$ large, the left-hand-side matrix in (2.8) is strictly copositive—in fact, positive definite—which establishes that Slater's condition is satisfied, thus ensuring strong duality:

Proposition 2.1. Under Assumption 2.4, suppose $r > 0$ is a constant such that $z^T z \leq r$

is satisfied by all $u \in \mathcal{U}$ and all extreme points $w \in \mathcal{W}$, where $z = (u, w)$. Then the optimal value of (2.8) equals $\pi(x)$.

Now, with $\pi(x)$ expressed as a minimization that depends affinely on x, we can collapse (2.2) into a single minimization that is equivalent to (RLP):

$$\begin{array}{ll} \min\limits_{x,\lambda,\Lambda,\rho} & c^T x + \lambda + r\rho \\ \text{s.t.} & x \in \mathcal{X},\ \lambda g_1 g_1^T - \frac{1}{2}G(x) + \frac{1}{2}(E^T\Lambda^T + \Lambda E) + \rho I \in \mathcal{COP}(\mathcal{K} \times \mathbb{R}^m_+) \\ & \rho \geq 0. \end{array} \qquad (\overline{RLP})$$

Theorem 2.1. *The optimal value of ($\overline{RLP}$) equals v^*_{RLP}.*

An equivalent version of ($\overline{RLP}$) can be derived based on the representation of $\pi(x)$ in (2.5):

$$\begin{array}{ll} \min\limits_{x,\lambda,v,\rho} & c^T x + \lambda + r\rho \\ \text{s.t.} & x \in \mathcal{X},\ \lambda g_1 g_1^T - \frac{1}{2}G(x) + E^T \operatorname{Diag}(v) E + \rho I \in \mathcal{COP}(\mathcal{K} \times \mathbb{R}^m_+) \\ & \rho \geq 0. \end{array} \qquad (2.9)$$

Our example in Section 2.5.1 will be based on this version.

We remark that, even if Assumption 2.4 fails and strong duality between (2.7) and (2.8) cannot be established, it still holds that the optimal value of ($\overline{RLP}$) is an upper bound on v^*_{RLP}. Note that, in this case, (2.7) should be modified to exclude $I \bullet Z \leq r$, and ρ should be set to 0 in (2.8).

2.3 The Affine Policy

Under the affine policy, the second-stage decision variable $y(\cdot)$ in (RLP) is modeled as a linear function of u via a free variable $Y \in \mathbb{R}^{n_2 \times k}$:

$$
\begin{aligned}
v^*_{\text{Aff}} := \quad \min_{x,y(\cdot),Y} \quad & c^T x + \max_{u\in\mathcal{U}} d^T y(u) \\
\text{s.t.} \quad & Ax + By(u) \geq Fu \quad \forall\, u \in \mathcal{U} \\
& y(u) = Yu \qquad\qquad \forall\, u \in \mathcal{U} \\
& x \in \mathcal{X}.
\end{aligned}
\tag{\textit{Aff}}
$$

Here, Y acts as a "dummy" first-stage decision, and so (Aff) can be recast as a regular robust optimization problem over $\mathcal{U}$. Specifically, using standard techniques [14], (Aff) is equivalent to

$$
\begin{aligned}
\min_{x,Y,\lambda} \quad & c^T x + \lambda \\
\text{s.t.} \quad & \lambda e_1 - Y^T d \in \mathcal{K}^* \\
& \text{Rows}(Axe_1^T - F + BY) \in \mathcal{K}^* \\
& x \in \mathcal{X}.
\end{aligned}
\tag{2.10}
$$

Problem (2.10) is tractable, but in general, the affine policy is only an approximation of (RLP), i.e., $v^*_{\text{RLP}} < v^*_{\text{Aff}}$. In what follows, we provide a copositive representation ($\overline{\text{Aff}}$) of (*Aff*), which is then used to develop an alternative formulation (IA) of (2.10). Later, in Section 2.4, problem (IA) will be compared directly to ($\overline{\text{RLP}}$).

Following the approach of Section 4.3, we may express (Aff) as $\min_{x\in\mathcal{X},Y} c^T x + \Pi(x, Y)$ where

$$
\Pi(x, Y) := \max_{u\in\mathcal{U}} \min_{y(u)\in\mathbb{R}^{n_2}} \{d^T y(u) : By \geq Fu - Ax,\ y(u) = Yu\}.
$$

Note that we do not replace $y(u)$ everywhere by Yu in the definition of $\Pi(x, Y)$; this is a small but critical detail in the subsequent derivations. The inner minimization

has dual

$$\begin{aligned}&\max_{w\geq 0,v}\{(Fu-Ax)^Tw+(Yu)^Tv : B^Tw+v=d\}\\&=\max_{w\geq 0}\left((Fu-Ax)^Tw+(Yu)^T(d-B^Tw)\right).\end{aligned}$$

After collecting terms, homogenizing, and converting to copostive optimization, we have

$$\begin{aligned}\Pi(x,Y)= \quad &\max \quad \tfrac{1}{2}(G(x)-H(Y))\bullet Z\\ &\text{s.t.} \quad Z\in\mathcal{CPP}(\mathcal{K}\times\mathbb{R}^m_+),\quad g_1g_1^T\bullet Z=1\end{aligned}\tag{2.11}$$

with dual

$$\begin{aligned}&\min_{\lambda} \quad \lambda\\ &\text{s.t.} \quad \lambda g_1g_1^T-\tfrac{1}{2}G(x)+\tfrac{1}{2}H(Y)\in\mathcal{COP}(\mathcal{K}\times\mathbb{R}^m_+),\end{aligned}\tag{2.12}$$

where $G(x)$ is defined as in Section 4.3 and

$$H(Y):=\begin{pmatrix}-e_1d^TY-Y^Tde_1^T & (BY)^T\\ BY & 0\end{pmatrix}\in\mathcal{S}^{k+m}.$$

Since $\mathcal{K}$ has interior by Assumption 2.1, it follows that (2.11) also has interior, and so Slater's condition holds, implying strong duality between (2.11) and (2.12). Thus, repeating the logic of Section 4.3, (Aff) is equivalent to

$$\begin{aligned}&\min_{x,\lambda,Y} \quad c^Tx+\lambda\\ &\text{s.t.} \quad x\in\mathcal{X},\quad \lambda g_1g_1^T-\tfrac{1}{2}G(x)+\tfrac{1}{2}H(Y)\in\mathcal{COP}(\mathcal{K}\times\mathbb{R}^m_+).\end{aligned}\tag{$\overline{Aff}$}$$

Proposition 2.2. The optimal value of ($\overline{Aff}$) is v^*_{Aff}.

We now show that $\mathcal{COP}(\mathcal{K}\times\mathbb{R}^m_+)$ in ($\overline{\text{Aff}}$) can be replaced by a particular inner approximation without changing the optimal value. Moreover, this inner approxima-

tion is tractable, so that the resulting optimization problem serves as an alternative to the formulation (2.10) of (Aff).

Using the mnemonic "IA" for "inner approximation," we define

$$\mathrm{IA}(\mathcal{K}\times\mathbb{R}^m_+) := \left\{ S = \begin{pmatrix} S_{11} & S_{21}^T \\ S_{21} & S_{22} \end{pmatrix} : \begin{array}{l} S_{11} = e_1\alpha^T + \alpha e_1^T,\ \alpha\in\mathcal{K}^*, \\ \mathrm{Rows}(S_{21})\in\mathcal{K}^*,\ S_{22}\geq 0 \end{array} \right\}.$$

This set is tractable because it is defined by affine constraints in $\mathcal{K}^*$ as well as nonnegativity constraints. Moreover, $\mathrm{IA}(\mathcal{K}\times\mathbb{R}^m_+)$ is indeed a subset of $\mathcal{COP}(\mathcal{K}\times\mathbb{R}^m_+)$:

Lemma 2.2. $\mathrm{IA}(\mathcal{K}\times\mathbb{R}^m_+)\subseteq\mathcal{COP}(\mathcal{K}\times\mathbb{R}^m_+)$.

Proof. We first note that (2.1) implies that the first coordinate of every element of $\mathcal{K}$ is nonnegative; hence, $e_1\in\mathcal{K}^*$. Now, for arbitrary $\binom{p}{q}\in\mathcal{K}\times\mathbb{R}^m_+$ and $S\in\mathrm{IA}(\mathcal{K}\times\mathbb{R}^m_+)$, we prove $t := \binom{p}{q}^T S\binom{p}{q}\geq 0$. We have

$$t = \begin{pmatrix} p \\ q \end{pmatrix}^T \begin{pmatrix} S_{11} & S_{21}^T \\ S_{21} & S_{22} \end{pmatrix} \begin{pmatrix} p \\ q \end{pmatrix} = p^T S_{11} p + 2\, q^T S_{21} p + q^T S_{22} q.$$

Analyzing each of the three summands separately, we first have

$$e_1,\alpha\in\mathcal{K}^* \quad\Longrightarrow\quad p^T S_{11} p = p^T(e_1\alpha^T+\alpha e_1^T)p = 2(p^T e_1)(\alpha^T p)\geq 0.$$

Second, $p\in\mathcal{K}$ and $\mathrm{Rows}(S_{21})\in\mathcal{K}^*$ imply $S_{21}p\geq 0$, which in turn implies $q^T S_{21}p = q^T(S_{21}p)\geq 0$ because $q\geq 0$. Finally, it is clear that $q^T S_{22} q\geq 0$ as $S_{22}\geq 0$ and $q\geq 0$. Thus, $t\geq 0+0+0=0$, as desired. □

The following tightening of ($\overline{\mathrm{Aff}}$) simply replaces $\mathcal{COP}(\mathcal{K}\times\mathbb{R}^m_+)$ with its inner

approximation $\mathrm{IA}(\mathcal{K} \times \mathbb{R}^m_+)$:

$$\begin{array}{rcll} v^*_{\mathrm{IA}} & := & \min\limits_{x,\lambda,Y} & c^T x + \lambda \\ & & \text{s. t.} & x \in \mathcal{X}, \quad \lambda g_1 g_1^T - \frac{1}{2}G(x) + \frac{1}{2}H(Y) \in \mathrm{IA}(\mathcal{K} \times \mathbb{R}^m_+). \end{array} \tag{IA}$$

By construction, $v^*_{\mathrm{IA}} \geq v^*_{\mathrm{Aff}}$, but in fact these values are equal.

Theorem 2.3. $v^*_{\mathrm{IA}} = v^*_{\mathrm{Aff}}$.

Proof. We show $v^*_{\mathrm{IA}} \leq v^*_{\mathrm{Aff}}$ by demonstrating that every feasible solution of (2.10) yields a feasible solution of (IA) with the same objective value. Let (x, Y, λ) be feasible for (2.10); we prove

$$S := \lambda g_1 g_1^T - \tfrac{1}{2}G(x) + \tfrac{1}{2}H(Y) \in \mathrm{IA}(\mathcal{K} \times \mathbb{R}^m_+),$$

which suffices. Note that the block form of S is

$$S = \begin{pmatrix} \lambda e_1 e_1^T - \frac{1}{2}(e_1 d^T Y + Y^T d e_1^T) & \frac{1}{2}(Axe_1^T - F + BY)^T \\ \frac{1}{2}(Axe_1^T - F + BY) & 0 \end{pmatrix}.$$

The argument decomposes into three pieces. First, we define $\alpha := \frac{1}{2}(\lambda e_1 - Y^T d)$, which satisfies $\alpha \in \mathcal{K}^*$ due to (2.10). Then

$$\begin{aligned} S_{11} &= \lambda e_1 e_1^T - \tfrac{1}{2}(e_1 d^T Y + Y^T d e_1^T) \\ &= \left(\tfrac{1}{2}\lambda e_1 e_1^T - \tfrac{1}{2} e_1 d^T Y\right) + \left(\tfrac{1}{2}\lambda e_1 e_1^T - \tfrac{1}{2} Y^T d e_1^T\right) \\ &= e_1 \alpha^T + \alpha e_1^T \end{aligned}$$

as desired. Second, we have $2\,\mathrm{Rows}(S_{21}) = \mathrm{Rows}(Axe_1^T - F + BY) \in \mathcal{K}^*$ by (2.10). Finally, $S_{22} = 0 \geq 0$. □

2.4 Improving the Affine Policy

A direct relationship holds between ($\overline{\text{RLP}}$) and (IA):

Proposition 2.3. In problem ($\overline{\text{RLP}}$), write $\Lambda = \binom{\Lambda_1}{\Lambda_2}$, where $\Lambda_1 \in \mathbb{R}^{k\times n_2}$ and $\Lambda_2 \in \mathbb{R}^{m\times n_2}$. Problem (IA) is a restriction of ($\overline{RLP}$) in which $\Lambda_2 = 0$, Y is identified with Λ_1^T, $\rho = 0$, and $\mathcal{COP}(\mathcal{K}\times\mathbb{R}^m_+)$ is tightened to $\text{IA}(\mathcal{K}\times\mathbb{R}^m_+)$.

Proof. Examining the similar structure of ($\overline{\text{RLP}}$) and (IA), it suffices to equate the terms $E^T\Lambda^T + \Lambda E$ and $H(Y)$ in the respective problems under the stated restrictions. From (4.10),

$$E^T\Lambda^T + \Lambda E = \begin{pmatrix} -e_1d^T\Lambda_1^T - \Lambda_1 de_1^T & \Lambda_1 B^T - e_1 d^T\Lambda_2^T \\ B\Lambda_1^T - \Lambda_2 de_1^T & B\Lambda_2^T + \Lambda_2 B^T \end{pmatrix}.$$

Setting $\Lambda_2 = 0$ and identifying $Y = \Lambda_1^T$, we see

$$E^T\Lambda^T + \Lambda E = \begin{pmatrix} -e_1d^TY - Y^T de_1^T & Y^TB^T \\ BY & 0 \end{pmatrix} = H(Y),$$

as desired. □

Now let $\text{IB}(\mathcal{K}\times\mathbb{R}^m_+)$ be any closed convex cone satisfying

$$\text{IA}(\mathcal{K}\times\mathbb{R}^m_+) \subseteq \text{IB}(\mathcal{K}\times\mathbb{R}^m_+) \subseteq \mathcal{COP}(\mathcal{K}\times\mathbb{R}^m_+),$$

where the mnemonic "IB" stands for "in between", and consider the following problem gotten by replacing $\mathcal{COP}(\mathcal{K}\times\mathbb{R}^m_+)$ in ($\overline{\text{RLP}}$) with $\text{IB}(\mathcal{K}\times\mathbb{R}^m_+)$:

$$\begin{aligned} v^*_{\text{IB}} \;:=\; \min_{x,\lambda,\Lambda} \quad & c^Tx + \lambda \\ \text{s.t.} \quad & x\in\mathcal{X}, \quad \lambda g_1g_1^T - \tfrac{1}{2}G(x) + \tfrac{1}{2}(E^T\Lambda^T + \Lambda E) \in \text{IB}(\mathcal{K}\times\mathbb{R}^m_+). \end{aligned} \tag{IB}$$

Problem (IB) is clearly a restriction of ($\overline{\text{RLP}}$), and by Proposition 2.3, it is simultaneously no tighter than (IA). Combining this with Theorems 2.1 and 2.3, we thus have:

Theorem 2.4. $v^*_{\text{RLP}} \leq v^*_{\text{IB}} \leq v^*_{\text{Aff}}$.

We end this section with a short discussion of example approximations $\text{IB}(\mathcal{K} \times \mathbb{R}^m_+)$ for typical cases of $\mathcal{K}$. In fact, there are complete hierarchies of approximations of $\mathcal{COP}(\mathcal{K} \times \mathbb{R}^m_+)$ [146], but we present a relatively simple construction that starts from a given inner approximation $\text{IB}(\mathcal{K})$ of $\mathcal{COP}(\mathcal{K})$:

Proposition 2.4. Suppose $\text{IB}(\mathcal{K}) \subseteq \mathcal{COP}(\mathcal{K})$, and define

$$\text{IB}(\mathcal{K} \times \mathbb{R}^m_+) := \left\{ S + M + R \ : \begin{array}{c} S \in \text{IA}(\mathcal{K} \times \mathbb{R}^m_+),\ M \succeq 0 \\ R_{11} \in \text{IB}(\mathcal{K}),\ R_{21} = 0,\ R_{22} = 0 \end{array} \right\}.$$

Then $\text{IA}(\mathcal{K} \times \mathbb{R}^m_+) \subseteq \text{IB}(\mathcal{K} \times \mathbb{R}^m_+) \subseteq \mathcal{COP}(\mathcal{K} \times \mathbb{R}^m_+)$.

Proof. For the first inclusion, simply take $M = 0$ and $R_{11} = 0$. For the second inclusion, let arbitrary $\binom{p}{q} \in \mathcal{K} \times \mathbb{R}^m_+$ be given. We need to show

$$\binom{p}{q}^T (S + M + R) \binom{p}{q} = \tbinom{p}{q}^T S \tbinom{p}{q} + \tbinom{p}{q}^T M \tbinom{p}{q} + p^T R_{11} p \geq 0.$$

The first term is nonnegative because $S \in \text{IA}(\mathcal{K} \times \mathbb{R}^m_+)$; the second term is nonnegative because $M \succeq 0$; and the third is nonnegative because $R_{11} \in \mathcal{COP}(\mathcal{K})$. □

When $\mathcal{K} = \{u \in \mathbb{R}^k : \|(u_2, \ldots, u_k)^T\| \leq u_1\}$ is the second-order cone, it is known [130] that

$$\mathcal{COP}(\mathcal{K}) = \{R_{11} = \tau J + M_{11} : \tau \geq 0,\ M_{11} \succeq 0\},$$

where $J = \text{Diag}(1, -1, \ldots, -1)$. Because of this simple structure, it often makes sense to take $\text{IB}(\mathcal{K}) = \mathcal{COP}(\mathcal{K})$ in practice. Note also that $M_{11} \succeq 0$ can be absorbed into $M \succeq 0$ in the definition of $\text{IB}(\mathcal{K} \times \mathbb{R}^m_+)$ above. When $\mathcal{K} = \{u \in \mathbb{R}^k : Pu \geq 0\}$ is a polyhedral cone based on some matrix P, a typical inner approximation of $\mathcal{COP}(\mathcal{K})$ is

$$\text{IB}(\mathcal{K}) := \{R_{11} = P^T N P : N \geq 0\},$$

where N is a symmetric matrix variable of appropriate size. This corresponds to the RLT approach of [4, 38, 126].

2.5 Examples

In this section, we demonstrate our approximation v^*_{IB} satisfying $v^*_{\text{RLP}} \leq v^*_{\text{IB}} \leq v^*_{\text{Aff}}$ on several examples from the literature. The first example is treated analytically, while the remaining examples are verified numerically. All computations are conducted with Mosek version 8.0.0.28 beta [5] on an Intel Core i3 2.93 GHz Windows computer with 4GB of RAM and implemented using the modeling language YALMIP [99] in MATLAB (R2014a).

2.5.1 A temporal network example

The paper [139] studies a so-called *temporal network* application, which for any integer $s \geq 2$ leads to the following problem, which is based on an uncertainty set $\Xi \subseteq \mathbb{R}^s$ and in which the first-stage decision x is fixed, say, at 0 and $y(\cdot)$ maps

into $\mathbb{R}^s$:

$$\begin{array}{lll} \min\limits_{y(\cdot)} & \max\limits_{\xi\in\Xi} \; y(\xi)_s & \\ \text{s.t.} & y(\xi)_1 \geq \max\{\xi_1, 1-\xi_1\} & \forall\, \xi \in \Xi \\ & y(\xi)_2 \geq \max\{\xi_2, 1-\xi_2\} + y(\xi)_1 & \forall\, \xi \in \Xi \\ & \vdots & \\ & y(\xi)_s \geq \max\{\xi_s, 1-\xi_s\} + y(\xi)_{s-1} & \forall\, \xi \in \Xi. \end{array} \tag{2.13}$$

Note that each of the above linear constraints can be expressed as two separate linear constraints. The authors of [139] consider a polyhedral uncertainty set (based on the 1-norm). A related paper [78] considers a conic uncertainty set (based on the 2-norm) for $s = 2$; we will extend this to $s \geq 2$. In particular, we consider the following two uncertainty sets for general s:

$$\Xi_1 := \{\xi \in \mathbb{R}^s : \|\xi - \tfrac{1}{2}\mathbb{1}_s\|_1 \leq \tfrac{1}{2}\},$$

$$\Xi_2 := \{\xi \in \mathbb{R}^s : \|\xi - \tfrac{1}{2}\mathbb{1}_s\| \leq \tfrac{1}{2}\},$$

where $\mathbb{1}_s$ denotes the all-ones vector in $\mathbb{R}^s$. For $j = 1, 2$, let $v^*_{\text{RLP},j}$ and $v^*_{\text{Aff},j}$ be the robust and affine values associated with (2.13) for the uncertainty set Ξ_j. Note that $\Xi_1 \subseteq \Xi_2$, and hence $v^*_{\text{RLP},1} \leq v^*_{\text{RLP},2}$.

The papers [78, 139] show that $v^*_{\text{Aff},1} = v^*_{\text{Aff},2} = s$, and [139] establishes $v^*_{\text{RLP},1} = \frac{1}{2}(s+1)$. Moreover, we can calculate $v^*_{\text{RLP},2}$ in this chapter by the following analysis.

Any feasible $y(\xi)$ satisfies

$$\begin{aligned} y(\xi)_s &\geq \max\{\xi_s, 1-\xi_s\} + y(\xi)_{s-1} \\ &\geq \max\{\xi_s, 1-\xi_s\} + \max\{\xi_{s-1}, 1-\xi_{s-1}\} + y(\xi)_{s-2} \\ &\geq \cdots \geq \sum_{i=1}^{s} \max\{\xi_i, 1-\xi_i\} \end{aligned}$$

Hence, applying this inequality at an optimal $y(\cdot)$, it follows that

$$v^*_{\mathrm{RLP},2} \geq \max_{\xi \in \Xi_2} \sum_{i=1}^{s} \max\{\xi_i, 1-\xi_i\}.$$

Under the change of variables $\mu := 2\xi - \mathbb{1}_s$, we have

$$\begin{aligned} v^*_{\mathrm{RLP},2} &\geq \max_{\xi \in \Xi_2} \sum_{i=1}^{s} \max\{\xi_i, 1-\xi_i\} = \max_{\|\mu\| \leq 1} \sum_{i=1}^{s} \tfrac{1}{2} \max\{1+\mu_i, 1-\mu_i\} \\ &= \frac{1}{2} \max_{\|\mu\| \leq 1} \sum_{i=1}^{s} (1 + |\mu_i|) = \frac{1}{2} \left(s + \max_{\|\mu\| \leq 1} \|\mu\|_1 \right) = \tfrac{1}{2}(\sqrt{s} + s), \end{aligned}$$

where the last equality follows from the fact that the largest 1-norm over the Euclidean unit ball is $\sqrt{s}$. Moreover, one can check that the specific, sequentially defined mapping

$$\begin{aligned} y(\xi)_1 &:= \max\{\xi_1, 1-\xi_1\} \\ y(\xi)_2 &:= \max\{\xi_2, 1-\xi_2\} + y(\xi)_1 \\ &\vdots \\ y(\xi)_s &:= \max\{\xi_s, 1-\xi_s\} + y(\xi)_{s-1} \end{aligned}$$

is feasible with objective value $\frac{1}{2}(\sqrt{s}+s)$. So $v^*_{\mathrm{RLP},2} \leq \frac{1}{2}(\sqrt{s}+s)$, and this completes the argument that $v^*_{\mathrm{RLP},2} = \frac{1}{2}(\sqrt{s}+s)$. Overall, we find that each $j = 1, 2$ yields a

class of problems with arbitrarily large gaps between the true robust adjustable and affine-policy values.

Using the similar change of variables

$$u := (1, u_2, \ldots, u_{s+1})^T = (1, 2\xi_1 - 1, \ldots, 2\xi_s - 1)^T \in \mathbb{R}^{s+1},$$

for each Ξ_j, we may cast (2.13) in the form of (RLP) by setting $x = 0$, defining

$$m = 2s, \quad k = s + 1, \quad n_2 = s,$$

and taking $\widehat{\mathcal{U}}_j$ to be the k-dimensional cone associated with the j-norm. For convenience, we continue to use s in the following discussion, but we will remind the reader of the relationships between s, m, k, and n_2 as necessary (e.g., $s = m/2$). We also set

$$d = (0, \ldots, 0, 1)^T \in \mathbb{R}^s,$$

$$B = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & 0 & \cdots & 0 & 0 \\ -1 & 1 & 0 & \cdots & 0 & 0 \\ -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & 0 & \cdots & -1 & 1 \\ 0 & 0 & 0 & \cdots & -1 & 1 \end{pmatrix} \in \mathbb{R}^{2s \times s}, \quad F = \frac{1}{2} \begin{pmatrix} 1 & 1 & 0 & \cdots & 0 \\ 1 & -1 & 0 & \cdots & 0 \\ 1 & 0 & 1 & \cdots & 0 \\ 1 & 0 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \cdots & 1 \\ 1 & 0 & 0 & \cdots & -1 \end{pmatrix} \in \mathbb{R}^{2s \times (s+1)}.$$

Furthermore,

$$\widehat{\mathcal{U}}_2 := \{u \in \mathbb{R}^{s+1} : \|(u_2, \ldots, u_{s+1})^T\| \leq u_1\}$$

is the second-order cone, and

$$\widehat{\mathcal{U}}_1 := \{u \in \mathbb{R}^{s+1} : Pu \geq 0\},$$

where each row of $P \in \mathbb{R}^{2^s \times (s+1)}$ has the following form: $(1, \pm 1, \ldots, \pm 1)$. That is, each row is an $(s+1)$-length vector with a 1 in its first position and some combination of $+1$'s and -1's in the remaining s positions. Note that the size of P is exponential in s. Using extra nonnegative variables, we could also represent $\widehat{\mathcal{U}}_1$ as the projection of a cone with size polynomial in s, and all of the subsequent discussion would still apply. In other words, the exact representation of $\widehat{\mathcal{U}}_1$ is not so relevant to our discussion here; we choose the representation $Pu \geq 0$ in the original space of variables for convenience.

It is important to note that, besides $\mathcal{K}_1$ and $\mathcal{K}_2$, all other data required for representing (2.13) in the form of (RLP), such as the matrices B and F, do not depend on j. Assumptions 2.1–2.3 clearly hold, and the following proposition shows that (2.13) also satisfies Assumption 2.4:

Proposition 2.5. For (2.13) and its formulation as an instance of (RLP), $\mathcal{W}$ is nonempty and bounded.

Proof. The system $B^T w = d$ is equivalent to the $2s-1$ equations $w_1 + w_2 = 1$, $w_2 + w_3 = 1, \cdots, w_{2s-1} + w_{2s} = 1$. It is thus straightforward to check that $\mathcal{W}$ is nonempty and bounded. □

2.5.1.1 The case $j = 2$

Let us focus on the case $j = 2$; we continue to make use of the subscript 2. Recall $v^*_{\mathrm{RLP},2} = \frac{1}{2}(\sqrt{s} + s)$, and consider problem (IB_2) with $\mathrm{IB}(\widehat{\mathcal{U}}_2 \times \mathbb{R}^{2s}_+)$ built as described for the second-order cone at the end of Section 2.4. We employ the equivalent formulation (2.9) of $(\overline{\mathrm{RLP}})$, setting $x = 0$ and replacing $\mathcal{COP}(\widehat{\mathcal{U}}_2 \times \mathbb{R}^{2s}_+)$ by $\mathrm{IB}(\widehat{\mathcal{U}}_2 \times \mathbb{R}^{2s}_+)$:

$$\begin{array}{rll} v^*_{\mathrm{IB},2} = & \min & \lambda + r\rho \\ & \text{s.t.} & \lambda g_1 g_1^T - \frac{1}{2}G(0) + E^T \operatorname{Diag}(v) E + \rho I \in \mathrm{IB}(\widehat{\mathcal{U}}_2 \times \mathbb{R}^{2s}_+) \\ & & \rho \geq 0. \end{array} \tag{2.14}$$

Note that the dimension of g_1 is $k + m = (s+1) + 2s = 3s + 1$.

We know $v^*_{\mathrm{RLP},2} \leq v^*_{\mathrm{IB},2}$ by Theorem 2.4. Substituting the definition of $\mathrm{IB}(\widehat{\mathcal{U}}_2 \times$

$\mathbb{R}^{2s}_+$) from Section 2.4, using the fact that $\widehat{\mathcal{U}}_2^* = \widehat{\mathcal{U}}_2$, and simplifying, we have

$$\begin{aligned} v^*_{\mathrm{IB},2} = \quad & \min \quad \lambda + r\rho \\ & \text{s.t.} \quad \rho I + \lambda g_1 g_1^T - \tfrac{1}{2}G(0) + E^T \operatorname{Diag}(v) E - S - R \succeq 0 \\ & \qquad \rho \geq 0,\ S_{11} = e_1\alpha^T + \alpha e_1^T,\ \alpha \in \widehat{\mathcal{U}}_2,\ S_{22} \geq 0,\ \operatorname{Rows}(S_{21}) \in \widehat{\mathcal{U}}_2 \\ & \qquad R_{11} = \tau J,\ \tau \geq 0,\ R_{21} = 0,\ R_{22} = 0. \end{aligned} \tag{2.15}$$

We will show that, for every $\rho > 0$, (2.15) has a feasible solution with objective value $v^*_{\mathrm{RLP},2} + r\rho$. Then, by letting $\rho \to 0$, we conclude that $v^*_{\mathrm{IB},2} \leq v^*_{\mathrm{RLP},2}$, which in turn establishes that $v^*_{\mathrm{IB},2} = v^*_{\mathrm{RLP},2}$, i.e., that our relaxation is in fact exact.

For fixed $\rho > 0$, let us construct the claimed feasible solution. Set

$$\lambda = v^*_{\mathrm{RLP},2} = \tfrac{1}{2}(\sqrt{s} + s), \quad \alpha = 0, \quad \tau = \tfrac{1}{4}\sqrt{s}, \quad S_{21} = 0,$$

and

$$S_{22} = \frac{1}{2\sqrt{s}} \sum_{i=1}^{s} \left(f_{2i} f_{2i-1}^T + f_{2i-1} f_{2i}^T \right) \geq 0,$$

where f_j denotes the j-th coordinate vector in $\mathbb{R}^m = \mathbb{R}^{2s}$. Note that clearly $\alpha \in \widehat{\mathcal{U}}_2$ and $\operatorname{Rows}(S_{21}) \in \widehat{\mathcal{U}}_2$. Also forcing $v = \mu \mathbb{1}_k$ for a single scalar variable μ, where $\mathbb{1}_k$ is the all-ones vector of size $k = s + 1$, the feasibility constraints of (2.15) simplify further to

$$\rho I + \begin{pmatrix} \frac{1}{2}(s + \sqrt{s}) e_1 e_1^T - \frac{1}{4}\sqrt{s} J & -\frac{1}{2}F^T \\ -\frac{1}{2}F & -S_{22} \end{pmatrix} + \mu E^T E \succeq 0, \tag{2.16}$$

where $e_1 \in \mathbb{R}^k = \mathbb{R}^{s+1}$ is the first coordinate vector. For compactness, we write

$$V := \begin{pmatrix} \frac{1}{2}(s + \sqrt{s}) e_1 e_1^T - \frac{1}{4}\sqrt{s} J & -\frac{1}{2}F^T \\ -\frac{1}{2}F & -S_{22} \end{pmatrix} \tag{2.17}$$

so that (2.16) reads $\rho I + V + \mu E^T E \succeq 0$. We next claim that, given ρ, V, and E, μ can be chosen so that (2.16) is indeed satisfied, which we prove in the following.

First, the existence of μ requires the following lemma:

Lemma 2.5. *If V is positive semidefinite on the null space of E (that is, $z \in \text{Null}(E) \Rightarrow z^T V z \geq 0$), then there exists $\mu > 0$ such that $\rho I + V + \mu E^T E \succ 0$.*

Proof. We prove the contrapositive. Suppose $\rho I + V + \mu E^T E$ is not positive definite for all $\mu > 0$. In particular, there exists a sequence of vectors $\{z_\ell\}$ such that

$$z_\ell^T (\rho I + V + \ell E^T E) z_\ell \leq 0, \quad \|z_\ell\| = 1.$$

Since $\{z_\ell\}$ is bounded, there exists a limit point $\bar{z}$ such that

$$z_\ell^T (\tfrac{1}{\ell}(\rho I + V) + E^T E) z_\ell \leq 0 \quad \Rightarrow \quad \bar{z}^T E^T E \bar{z} = \|Ez\|^2 \leq 0 \quad \Leftrightarrow \quad \bar{z} \in \text{Null}(E).$$

Furthermore,

$$\begin{aligned} z_\ell^T (\rho I + V) z_\ell \leq -\ell z_\ell^T E^T E z_\ell = -\ell \|E z_\ell\|^2 \leq 0 \quad &\Rightarrow \quad \bar{z}^T (\rho I + V) \bar{z} \leq 0 \\ &\Leftrightarrow \quad \bar{z}^T V \bar{z} \leq -\rho \|\bar{z}\|^2 < 0. \end{aligned}$$

Thus, V is not positive semidefinite on $\text{Null}(E)$. □

Then, with the lemma in hand, it suffices to prove that V is positive semidefinite on $\text{Null}(E)$, a fact which we establish directly.

Recall that $E \in \mathbb{R}^{n_2 \times (k+m)} = \mathbb{R}^{s \times (3s+1)}$. For notational convenience, we partition any $z \in \mathbb{R}^{k+m}$ into $z = \binom{u}{w}$ with $u \in \mathbb{R}^k = \mathbb{R}^{s+1}$ and $w \in \mathbb{R}^m = \mathbb{R}^{2s}$. Then, from

the definition of E, we have

$$z = \begin{pmatrix} u \\ w \end{pmatrix} \in \mathrm{Null}(E) \iff \left\{ \begin{array}{l} w_1 + w_2 = w_3 + w_4 \\ \\ w_3 + w_4 = w_5 + w_6 \\ \\ \vdots \\ \\ w_{2s-3} + w_{2s-2} = w_{2s-1} + w_{2s} \\ \\ w_{2s-1} + w_{2s} = u_1 \end{array} \right\}$$

$$\iff \quad w_{2i-1} = u_1 - w_{2i} \ \ \forall\ i = 1, \ldots, s.$$

So, taking into account the definition (2.17) of V,

$$4\, z^T V z = 4 \begin{pmatrix} u \\ w \end{pmatrix}^T V \begin{pmatrix} u \\ w \end{pmatrix} = u^T \left(2(s + \sqrt{s}) e_1 e_1^T - \sqrt{s} J \right) u - 4\, w^T F u - 4\, w^T S_{22} w,$$

which breaks into the three summands, and we will simplify each one by one. First,

$$u^T \left(2(s + \sqrt{s}) e_1 e_1^T - \sqrt{s} J \right) u = 2(s + \sqrt{s}) u_1^2 - \sqrt{s} u_1^2 + \sqrt{s} \sum_{j=2}^{s+1} u_j^2$$

$$= 2\, s\, u_1^2 + \sqrt{s}\, u^T u.$$

Second,

$$\begin{aligned}
-4w^T Fu &= -4\sum_{j=1}^{2s} w_j [Fu]_j = -4\sum_{i=1}^{s} \left(w_{2i-1}[Fu]_{2i-1} + w_{2i}[Fu]_{2i}\right) \\
&= -2\sum_{i=1}^{s} \left(w_{2i-1}(u_1 + u_{i+1}) + w_{2i}(u_1 - u_{i+1})\right) \\
&= -2\sum_{i=1}^{s} \left((w_{2i-1} + w_{2i})u_1 + u_{i+1}(w_{2i-1} - w_{2i})\right) \\
&= -2\sum_{i=1}^{s} \left(u_1^2 + u_{i+1}(w_{2i-1} - w_{2i})\right) \\
&= -2\,s\,u_1^2 - 2\sum_{i=1}^{s} u_{i+1}(w_{2i-1} - w_{2i}) \\
&= -2\,s\,u_1^2 + 2\sum_{i=1}^{s} u_{i+1}(w_{2i} - w_{2i-1}) = -2\,s\,u_1^2 + 2\sum_{i=1}^{s} u_{i+1}(2w_{2i} - u_1).
\end{aligned}$$

Finally,

$$\begin{aligned}
-4w^T S_{22} w &= -4w^T \left(\frac{1}{2\sqrt{s}} \sum_{i=1}^{s} \left(f_{2i} f_{2i-1}^T + f_{2i-1} f_{2i}^T\right)\right) w \\
&= -\frac{4}{\sqrt{s}} \sum_{i=1}^{s} w_{2i-1} w_{2i} = -\frac{4}{\sqrt{s}} \sum_{i=1}^{s} (u_1 - w_{2i}) w_{2i}.
\end{aligned}$$

Combining the three summands, we have as desired

$$\begin{aligned}
4z^T Vz &= \left(2s\,u_1^2 + \sqrt{s}\,u^T u\right) + \left(-2s\,u_1^2 + 2\sum_{i=1}^{s} u_{i+1}(2w_{2i} - u_1)\right) + \left(-\frac{4}{\sqrt{s}} \sum_{i=1}^{s} (u_1 - w_{2i}) w_{2i}\right) \\
&= \sqrt{s}\,u^T u + 2\sum_{i=1}^{s} u_{i+1}(2w_{2i} - u_1) - \frac{4}{\sqrt{s}} \sum_{i=1}^{s} (u_1 - w_{2i}) w_{2i} \\
&= \sum_{i=1}^{s} \left(\frac{1}{\sqrt{s}}\,u_1^2 + \sqrt{s}\,u_{i+1}^2 + 2\,u_{i+1}(2w_{2i} - u_1) - \frac{4}{\sqrt{s}}(u_1 - w_{2i}) w_{2i}\right) \\
&= \sum_{i=1}^{s} \left(\frac{1}{\sqrt{s}}\,u_1^2 - 2\,u_1\,u_{i+1} - \frac{4}{\sqrt{s}}\,u_1\,w_{2i} + \sqrt{s}\,u_{i+1}^2 + 4\,u_{i+1}\,w_{2i} + \frac{4}{\sqrt{s}}\,w_{2i}^2\right) \\
&= \sum_{i=1}^{s} \left(-(s)^{-1/4}\,u_1 + (s)^{1/4}\,u_{i+1} + 2(s)^{-1/4}\,w_{2i}\right)^2 \\
&\geq 0.
\end{aligned}$$

By the discussion above, it follows that indeed $v^*_{\mathrm{IB},2} = v^*_{\mathrm{RLP},2}$ for the instance (2.13) of (RLP) based on Ξ_2.

For completeness—and also to facilitate Section 2.5.1.2 next—we construct the corresponding optimal solution of the dual of (2.14), which can be derived from (2.5) by setting $x = 0$, adding the redundant constraint $I \bullet Z \leq r$, and replacing $\mathcal{CPP}(\widehat{\mathcal{U}}_2 \times \mathbb{R}^{2s}_+)$ by its relaxation $\mathrm{IB}(\widehat{\mathcal{U}}_2 \times \mathbb{R}^{2s}_+)^*$, the dual cone of $\mathrm{IB}(\widehat{\mathcal{U}}_2 \times \mathbb{R}^{2s}_+)$. Specifically, the dual is

$$\begin{array}{rll} v^*_{\mathrm{IB},2} = & \max & F \bullet Z_{21} \\ & \text{s.t.} & \mathrm{diag}(EZE^T) = 0,\ I \bullet Z \leq r \\ & & J \bullet Z_{11} \geq 0,\ Z_{11}e_1 \in \widehat{\mathcal{U}}_2,\ Z_{22} \geq 0,\ \mathrm{Rows}(Z_{21}) \in \widehat{\mathcal{U}}_2 \\ & & Z \succeq 0,\ g_1 g_1^T \bullet Z = 1. \end{array} \tag{2.18}$$

We construct the optimal solution of (2.18) to be

$$Z = \frac{1}{4}\left[\begin{pmatrix} 2e_1 \\ \mathbb{1}_m \end{pmatrix}\begin{pmatrix} 2e_1 \\ \mathbb{1}_m \end{pmatrix}^T + \sum_{i=1}^{s} \begin{pmatrix} \frac{2}{\sqrt{s}} e_{i+1} \\ f_{2i-1} - f_{2i} \end{pmatrix}\begin{pmatrix} \frac{2}{\sqrt{s}} e_{i+1} \\ f_{2i-1} - f_{2i} \end{pmatrix}^T\right],$$

where each $e_\bullet$ is a coordinate vector in $\mathbb{R}^k = \mathbb{R}^{s+1}$, each $f_\bullet$ is a coordinate vector in $\mathbb{R}^m = \mathbb{R}^{2s}$, and $\mathbb{1}_m \in \mathbb{R}^m$ is the all-ones vector. By construction, Z is positive semidefinite, and one can argue in a straightforward manner that

$$Z_{11} = \mathrm{Diag}(1, \tfrac{1}{s}, \ldots, \tfrac{1}{s}), \quad Z_{22} = \frac{1}{4}\left(I + \mathbb{1}_m \mathbb{1}_m^T - \sum_{i=1}^{s}(f_{2i} f_{2i-1}^T + f_{2i-1} f_{2i}^T)\right),$$

and

$$Z_{21} = \frac{1}{2}\begin{pmatrix} 1 & \frac{1}{\sqrt{s}} & 0 & \cdots & 0 \\ 1 & -\frac{1}{\sqrt{s}} & 0 & \cdots & 0 \\ 1 & 0 & \frac{1}{\sqrt{s}} & \cdots & 0 \\ 1 & 0 & -\frac{1}{\sqrt{s}} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \cdots & \frac{1}{\sqrt{s}} \\ 1 & 0 & 0 & \cdots & -\frac{1}{\sqrt{s}} \end{pmatrix}.$$

Then Z clearly satisfies $g_1 g_1^T \bullet Z = 1$, $Z_{11} e_1 \in \widehat{\mathcal{U}}_2$, $J \bullet Z_{11} \geq 0$, $Z_{22} \geq 0$, and $\mathrm{Rows}(Z_{21}) \in \widehat{\mathcal{U}}_2$. Furthermore, the constraint $I \bullet Z \leq r$ is easily satisfied for sufficiently large r. To check the constraint $\mathrm{diag}(EZE^T) = 0$, it suffices to verify $EZ = 0$, which amounts to two equations. First,

$$0 = E\begin{pmatrix} 2e_1 \\ \mathbb{1}_m \end{pmatrix} = -2\, d e_1^T e_1 + B^T \mathbb{1}_m = -2d + 2d = 0,$$

and second, for each $i = 1, \ldots, s$,

$$0 = E\begin{pmatrix} \frac{2}{\sqrt{s}} e_{i+1} \\ f_{2i-1} - f_{2i} \end{pmatrix} = -\frac{2}{\sqrt{s}}\, d e_1^T e_{i+1} + B^T (f_{2i-1} - f_{2i}) = 0 + B^T f_{2i-1} - B^T f_{2i} = 0.$$

So the proposed Z is feasible. Finally, it is clear that the corresponding objective value is $F \bullet Z_{21} = \frac{1}{2}(\sqrt{s} + s)$. So Z is indeed optimal.

2.5.1.2 The case $j = 1$

Recall that Ξ_1 is properly contained in Ξ_2. So $v^*_{\mathrm{RLP},1}$ cannot exceed $v^*_{\mathrm{RLP},2}$ due to its smaller uncertainty set. In fact, as discussed above, we have $\frac{1}{2}(\sqrt{s} + 1) = v^*_{\mathrm{RLP},1} < v^*_{\mathrm{RLP},2} = \frac{1}{2}(\sqrt{s} + s)$ and $v^*_{\mathrm{Aff},1} = v^*_{\mathrm{Aff},2} = s$. In this subsection, we further

exploit the inclusion $\Xi_1 \subseteq \Xi_2$ and the results of the previous subsection (case $j = 2$) to prove that, for the particular tightening $\mathrm{IB}(\widehat{\mathcal{U}}_1 \times \mathbb{R}^{2s}_+)$ proposed at the end of Section 2.4, we have $v^*_{\mathrm{RLP},1} < v^*_{\mathrm{IB},1} = \frac{1}{2}(\sqrt{s} + s) < v^*_{\mathrm{Aff},1}$. In other words, the case $j = 1$ provides an example in which our approach improves the affine value but does not completely close the gap with the robust value.

The inclusion $\Xi_1 \subseteq \Xi_2$ implies $\widehat{\mathcal{U}}_1 \subseteq \widehat{\mathcal{U}}_2$ and $\mathcal{CPP}(\widehat{\mathcal{U}}_1 \times \mathbb{R}^{2s}_+) \subseteq \mathcal{CPP}(\widehat{\mathcal{U}}_2 \times \mathbb{R}^{2s}_+)$. Hence, $\mathcal{COP}(\widehat{\mathcal{U}}_1 \times \mathbb{R}^{2s}_+) \supseteq \mathcal{COP}(\widehat{\mathcal{U}}_2 \times \mathbb{R}^{2s}_+)$. Moreover, it is not difficult to see that the construction of $\mathrm{IB}(\widehat{\mathcal{U}}_1 \times \mathbb{R}^{2s}_+)$ introduced at the end of Section 2.4 for the polyhedral cone $\widehat{\mathcal{U}}_1$ satisfies $\mathrm{IB}(\widehat{\mathcal{U}}_1 \times \mathbb{R}^{2s}_+) \supseteq \mathrm{IB}(\widehat{\mathcal{U}}_2 \times \mathbb{R}^{2s}_+)$. Thus, we conclude $v^*_{\mathrm{IB},1} \leq v^*_{\mathrm{IB},2} = \frac{1}{2}(\sqrt{s} + s)$.

We finally show $v^*_{\mathrm{IB},1} \geq v^*_{\mathrm{IB},2}$. Based on the definition of $\widehat{\mathcal{U}}_1$ using the matrix P, similar to (2.18) the corresponding dual problem is

$$\begin{aligned} v^*_{\mathrm{IB},1} = \quad & \max \quad F \bullet Z_{21} \\ & \text{s.t.} \quad \mathrm{diag}(EZE^T) = 0,\ I \bullet Z \leq r \\ & \qquad PZ_{11}e_1 \geq 0,\ PZ_{11}P^T \geq 0,\ Z_{22} \geq 0,\ PZ_{21}^T \geq 0 \\ & \qquad Z \succeq 0,\ g_1 g_1^T \bullet Z = 1. \end{aligned} \tag{2.19}$$

To complete the proof, we claim that the specific Z detailed in the previous subsection is also feasible for (2.19). It remains to show that $PZ_{11}e_1 \geq 0$, $PZ_{11}P^T \geq 0$, and $PZ_{21}^T \geq 0$.

Recall that $Z_{11} = \mathrm{Diag}(1, \frac{1}{s}, \ldots, \frac{1}{s})$ and every row of P has the form $(1, \pm 1, \ldots, \pm 1)$. Clearly, we have $PZ_{11}e_1 \geq 0$. Moreover, each entry of $PZ_{11}P^T$ can be expressed as

$\binom{1}{\alpha}^T Z_{11} \binom{1}{\beta}$ for some $\alpha, \beta \in \mathbb{R}^s$ each of the form $(\pm 1, \ldots, \pm 1)$. We have

$$\binom{1}{\alpha}^T Z_{11} \binom{1}{\beta} = 1 + \tfrac{1}{s} \cdot \alpha^T \beta \geq 1 + \tfrac{1}{s}(-s) \geq 0.$$

So indeed $PZ_{11}P^T \geq 0$. To check $PZ_{21}^T \geq 0$, recall also that every column of Z_{21}^T has the form $\frac{1}{2}(e_1 \pm \frac{1}{\sqrt{s}} e_{i+1})$ for $i = 1, \ldots, s$, where $e_\bullet$ is a coordinate vector in $\mathbb{R}^k = \mathbb{R}^{s+1}$. Then each entry of $2PZ_{21}^T$ can be expressed as

$$\binom{1}{\alpha}^T e_1 \pm \tfrac{1}{\sqrt{s}} \binom{1}{\alpha}^T e_{i+1} \geq 1 - \tfrac{1}{\sqrt{s}} > 0.$$

So $PZ_{21}^T \geq 0$, as desired.

2.5.2 Multi-item newsvendor problem

In this example, we consider the same robust multi-item newsvendor problem discussed in [7]:

$$\max_{x \geq 0} \min_{\xi \in \Xi} \sum_{j \in \mathcal{J}} \big[r_j \min(x_j, \xi_j) - c_j x_j + s_j \max(x_j - \xi_j, 0) - p_j \max(\xi_j - x_j, 0) \big], \quad (2.20)$$

where $\mathcal{J}$ represents the set of products; x is the vector of nonnegative order quantities x_j for all $j \in \mathcal{J}$; $\xi \in \Xi$ is the vector of uncertain demands ξ_j for all $j \in \mathcal{J}$; r_j, c_j, s_j, and p_j denote the sale price, order cost, salvage price, and shortage cost of a unit of product j with $s_j \leq \min(r_j, c_j)$. Problem (2.20) is equivalent to

$$\begin{array}{lll} \max\limits_{x, y(\cdot)} & \min\limits_{\xi \in \Xi} \sum_{j \in \mathcal{J}} y_j(\xi) & \\ \text{s.t.} & y_j(\xi) \leq (r_j - c_j) x_j - (r_j - s_j)(x_j - \xi_j) & \forall\, j \in \mathcal{J},\ \xi \in \Xi \\ & y_j(\xi) \leq (r_j - c_j) x_j - p_j(\xi_j - x_j) & \forall\, j \in \mathcal{J},\ \xi \in \Xi \\ & x \geq 0. & \end{array} \quad (2.21)$$

We consider the same instance as in [7] with $\mathcal{J} = \{1, 2, 3\}$,

$$r = (80, 80, 80),\ c = (70, 50, 20),\ s = (20, 15, 10),\ p = (60, 60, 50),$$

and

$$\Xi := \left\{ \xi : \exists\, \zeta^+,\ \zeta^- \text{ s. t. } \begin{array}{l} \zeta^+ \geq 0,\ \zeta^- \geq 0 \\ \zeta_j^+ + \zeta_j^- \leq 1\ \forall\, j \in \mathcal{J} \\ \sum_{j\in\mathcal{J}}(\zeta_j^+ + \zeta_j^-) = 2 \\ \xi_1 = 80 + 30(\zeta_1^+ + \zeta_2^+ - \zeta_1^- - \zeta_2^-) \\ \xi_2 = 80 + 30(\zeta_2^+ + \zeta_3^+ - \zeta_2^- - \zeta_3^-) \\ \xi_3 = 60 + 20(\zeta_3^+ + \zeta_1^+ - \zeta_3^- - \zeta_1^-) \end{array} \right\}.$$

Omitting the details, we reformulate problem (2.21) as an instance of (RLP) in minimization form. Assumption 2.1 clearly holds, and by using a method called *enumeration of robust linear constraints* in [76], we have $v^*_{\text{RLP}} = -825.83$ (so Assumption 2.3 holds). Moreover, the affine-policy value is $v^*_{\text{Aff}} = -41.83$, and thus Assumption 2.2 holds. As mentioned at the end of Section 4.3, whether or not Assumption 2.4 holds, in practice our approach still provides an upper bound. Indeed, we solve (IB) with the approximating cone $\text{IB}(\mathcal{K} \times \mathbb{R}^m_+)$ defined in Section 2.4, where $\mathcal{K}$ is a polyhedral cone, and obtain $v^*_{\text{IB}} = -411.08$, which closes the gap significantly. The first-stage decisions given by the affine policy and our approach, respectively, are

$$x^*_{\text{Aff}} \approx (52.083, 104.400, 80.000), \quad x^*_{\text{IB}} \approx (57.118, 78.162, 78.473).$$

2.5.3 Lot-sizing problem on a network

We next consider a network lot-sizing problem derived from section 5 of [20] for which the mathematical formulation is:

$$\begin{aligned}
\min_{x,y(\cdot)} \quad & c^T x + \max_{\xi\in\Xi} \textstyle\sum_{i=1}^N \sum_{j=1}^N t_{ij} y(\xi)_{ij} \\
\text{s.t.} \quad & x_i + \textstyle\sum_{j=1}^N y(\xi)_{ji} - \sum_{j=1}^N y(\xi)_{ij} \geq \xi_i && \forall\ \xi \in \Xi,\ i = 1,\dots,N \\
& y(\xi)_{ij} \geq 0 && \forall\ \xi \in \Xi,\ i,j = 1,\dots,N \\
& 0 \leq x_i \leq V_i && \forall\ i = 1,\dots,N,
\end{aligned}$$

where N is the number of locations in the network, x denotes the first-stage stock allocations, $y(\xi)_{ij}$ denotes the second-stage shipping amounts from location i to location j, and the uncertainty set is the ball $\Xi := \{\xi : \|\xi\| \leq \Gamma\}$ for a given radius Γ. (The paper [20] uses a polyhedral uncertainty set, which we will also discuss below.) The vector c consists of the first-stage costs, the t_{ij} are the second-stage transportation costs for all location pairs, and V_i represents the capacity of store location i. We refer the reader to [20] for a full description.

Consistent with [20], we consider an instance with $N = 8$, $\Gamma = 10\sqrt{N}$, each $V_i = 20$, and each $c_i = 20$. We randomly generate the positions of the N locations from $[0, 10]^2$ in the plane. Then we set t_{ij} to be the (rounded) Euclidean distances between all pairs of locations; see Table 3.2.

Omitting the details, we reformulate this problem as an instance of (RLP), and we calculate $v^*_{\text{LB}} = 1573.8$ (using the Monte Carlo sampling procedure mentioned in the Introduction) and $v^*_{\text{Aff}} = 1950.8$. It is also easy to see that Assumption 2.1 holds, and the existence of an affine policy implies that Assumption 2.2 holds. Moreover,

	Location i							
Location j	1	2	3	4	5	6	7	8
1	0	4	3	2	2	2	3	5
2	4	0	6	5	4	4	2	8
3	3	6	0	1	5	2	6	2
4	2	5	1	0	4	1	4	3
5	2	4	5	4	0	4	2	7
6	2	4	2	1	4	0	4	4
7	3	2	6	4	2	4	0	7
8	5	8	2	3	7	4	7	0

Table 2.1: Unit transportation costs t_{ij} associated with pairs of locations.

Assumption 2.3 holds because the original objective value above is clearly bounded below by 0. Again, as mentioned at the end of Section 4.3, whether or not Assumption 2.4 holds, in practice we can still use our approach to calculate bounds. We solve (IB) with the approximating cone $\mathrm{IB}(\mathcal{K} \times \mathbb{R}^m_+)$ defined in Section 2.4, where $\mathcal{K}$ is the second-order cone, and obtain $v^*_{\mathrm{IB}} = 1794.0$, which closes the gap significantly. The first-stage allocations given by the affine policy and our approach, respectively, are

$$x^*_{\mathrm{Aff}} \approx (9.097, 11.246,\ 9.516, 8.320, 10.384, 9.493, 10.211, 12.316),$$

$$x^*_{\mathrm{IB}} \approx (0.269, 16.447, 15.328, 0.091, 18.124, 0.375,\ 9.951, 19.934).$$

Letting other data remain the same, we also ran tests on a budget uncertainty set $\Xi := \{\xi : 0 \leq \xi \leq \hat{\xi}e,\ e^T\xi \leq \Gamma\}$, where $\hat{\xi} = 20$ and $\Gamma = 20\sqrt{N}$, which is consistent with [20]. We found that, in this case, our method did not perform better than the affine policy.

2.5.4 Randomly generated instances

Finally, we used the same method presented in [78] to generate random instances of (RLP) with $(k, m, n_1, n_2) = (17, 16, 3, 5)$, $\mathcal{X} = \mathbb{R}^{n_1}$, $\mathcal{U}$ equal to the unit ball, and $\mathcal{K}$ equal to the second-order cone. Specifically, the instances are generated as follows: (i) the elements of A and B are independently and uniformly sampled in $[-5, 5]$; (ii) the rows of F are uniformly sampled in $[-5, 5]$ such that each row is in $-\mathcal{K}^* = -\mathcal{K}$ guaranteeing $Fu \leq 0$ for all $u \in \mathcal{U}$; and (iii) a random vector $\mu \in \mathbb{R}^m$ is repeatedly generated according to the uniform distribution on $[0, 1]^m$ until $c := A^T\mu \geq 0$ and $d := B^T\mu \geq 0$. Note that, by definition, $\mu \in \mathcal{W}$.

Clearly Assumption 2.1 is satisfied. In addition, we can see that Assumption 2.2 is true as follows. Consider $x = 0$ and set $y(\cdot)$ to be the zero map, i.e., $y(u) = 0$ for all $u \in \mathcal{U}$. Then $Ax + By(u) \geq Fu$ for all u if and only $0 \geq Fu$ for all u, which has been guaranteed by construction. Finally, Assumption 2.3 holds due to the following

chain, where $\pi(x)$ is defined as at the beginning of Section 4.3:

$$\begin{aligned} c^T x + \pi(x) &= c^T x + \max_{u \in \mathcal{U}} \max_{w \in \mathcal{W}} (Fu - Ax)^T w \\ &\geq c^T x + \max_{u \in \mathcal{U}} (Fu - Ax)^T \mu = c^T x - (Ax)^T \mu + \max_{u \in \mathcal{U}} (Fu)^T \mu \\ &= (c - A^T \mu)^T x + \max_{u \in \mathcal{U}} (Fu)^T \mu = 0^T x + \max_{u \in \mathcal{U}} (Fu)^T \mu \\ &> -\infty. \end{aligned}$$

We do not know if Assumption 2.4 necessarily holds for this construction, but as mentioned at the end of Section 4.3, our approximations still hold even if Assumption 2.4 does not hold.

For 1,000 generated instances, we computed v^*_{Aff}, the lower bound v^*_{LB} from the sampling procedure of the Introduction, and our bound v^*_{IB} using the the approximating cone $\text{IB}(\mathcal{K} \times \mathbb{R}^m_+)$ defined in Section 2.4, where $\mathcal{K}$ is the second-order cone. Of all 1,000 instances, 971 have $v^*_{\text{LB}} < v^*_{\text{IB}} = v^*_{\text{Aff}}$, while the remaining 29 have $v^*_{\text{LB}} < v^*_{\text{IB}} < v^*_{\text{Aff}}$. For those 29 instances with a positive gap, the average relative gap closed is 20.2%, where

$$\text{relative gap closed} := \frac{v^*_{\text{Aff}} - v^*_{\text{IB}}}{v^*_{\text{Aff}} - v^*_{\text{LB}}} \times 100\%.$$

2.6 Final Remarks

In this chapter, we have provided a new perspective on the two-stage problem (RLP). It would be interesting to study tighter inner approximations $\text{IB}(\mathcal{K} \times \mathbb{R}^m_+)$ of $\mathcal{COP}(\mathcal{K} \times \mathbb{R}^m_+)$ or to pursue other classes of problems, such as the one described in Section 2.5.1, for which our approach allows one to establish the tractability of (RLP).

A significant open question for our approach—one which we have not been able to resolve—is whether the copositive approach corresponds to enforcing a particular class of policies $y(\cdot)$. For example, the paper [23] solves (RLP) by employing polynomial policies, but the form of our "copositive policies" is unclear even though we have proven they are rich enough to solve (RLP). A related question is how to extract a specific policy $y(\cdot)$ from the solution of the approximation (IB).

CHAPTER 3
ROBUST SENSITIVITY ANALYSIS

3.1 Introduction

Consider the standard-form linear program (LP):

$$\begin{aligned} \min \quad & \hat{c}^T x \\ \text{s.t.} \quad & \hat{A}x = \hat{b} \\ & x \geq 0 \end{aligned} \tag{3.1}$$

where $x \in \mathbb{R}^n$ is the variable and $(\hat{A}, \hat{b}, \hat{c}) \in \mathbb{R}^{m \times n} \times \mathbb{R}^m \times \mathbb{R}^n$ are the problem parameters. In practice, $(\hat{A}, \hat{b}, \hat{c})$ may not be known exactly or may be predicted to change within a certain region. In such cases, *sensitivity analysis* (SA) examines how perturbations in the parameters affect the optimal value and solution of (3.1). We refer the reader to Section 1.2 for a review of classical and more sophisticated sensitivity analysis.

An area closely related to SA is *interval linear programming* (ILP), which can be viewed as *multi-parametric linear programming* with independent interval domains for the parameters [70, 71, 101]. Steuer [129] presents three algorithms for solving LPs in which the objective coefficients are specified by intervals, and Gabrel et al. [69] study LPs in which the right-hand sides vary within intervals and discuss the maximum and minimum optimal values. Mraz [103] considers a general situation in which the matrix coefficients and right-hand sides change within intervals and calculates upper and lower bounds for the associated optimal values. A comprehensive survey

of ILP has been given by Hladik [85].

To the best of our knowledge, in the context of LP, no authors have considered simultaneous LP parameter changes in a general way, i.e., perturbations in the objective coefficients $\hat{c}$, right-hand sides $\hat{b}$, and constraint coefficients $\hat{A}$ within a general region (not just intervals). The obstacle for doing so is clear: general perturbations lead to nonconvex quadratic programs (QPs), which are NP-hard to solve (as discussed below).

In this chapter, we extend—and in many cases unify—the SA literature by employing modern tools for nonconvex QPs. Specifically, we investigate SA for LPs in which $(\hat{b}, \hat{c})$ may change within a general compact, convex set $\mathcal{U}$, called the *uncertainty set*. Our goal is to calculate—or bound—the corresponding minimum (best-case) and maximum (worst-case) optimal values. Since these values involve the solution of nonconvex QPs, we use standard techniques from *copositive optimization* to reformulate these problems into convex *copositive programs* (COPs), which provide a theoretical grounding upon which to develop tight, tractable convex relaxations. We suggest the use of *semidefinite programming* (SDP) relaxations, which also incorporate valid conic inequalities that exploit the structure of the uncertainty set. We refer the reader to [57] for a survey on copositive optimization and its connections to semidefinite programming. We refer the reader to Section 1.4 for the relevant definitions and concepts of copositive programming.

Our approach is related to the recent work on *worst-case linear optimization* introduced by Peng and Zhu [111] in which: (i) only $\hat{b}$ is allowed to change within

an ellipsoidal region; and (ii) only the worst-case LP value is considered. (In fact, one can see easily that, in the setup of [111] based on (i), the best-case LP value can be computed in polynomial time via second-order-cone programming, making it less interesting to study in their setup.) The authors argue that the worst-case value is NP-hard to compute and use a specialized nonlinear semidefinite program (SDP) to bound it from above. They also develop feasible solutions to bound the worst-case value from below and show through a series of empirical examples that the resulting gaps are usually quite small. Furthermore, they also demonstrate that their SDP-based relaxation is better than the so-called *affine-rule approximation* (see [14]) and the Lasserre *linear matrix inequality* relaxation (see [95, 84]).

Our approach is more general than [111] because we allow both $\hat{b}$ and $\hat{c}$ to change, we consider more general uncertainty sets, and we study both the worst- and best-case values. In addition, instead of developing a specialized SDP approach, we make use of the machinery of copositive programming, which provides a theoretical grounding for the construction of tight, tractable conic relaxations using existing techniques. Nevertheless, we have been inspired by their approach in several ways. For example, their proof of NP-hardness also shows that our problem is NP-hard; we will borrow their idea of using primal solutions to estimate the quality of the relaxation bounds; and we test our approach in some of their examples.

We mention two additional connections of our approach with the literature. In [21], Bertsimas and Goyal consider a two-stage adaptive linear optimization problem under right-hand side uncertainty with a min-max objective. A simplified version

of this problem, in which the first-stage variables are non-existent, reduces to worst-case linear optimization; see the introduction of [21]. In fact, Bertsimas and Goyal use this fact to prove that their problem is NP-hard via the so-called max-min fractional set cover problem, which is a specific worst-case linear optimization problem studied by Feige et al. [65]. Our work is also related to the study of *adjustable robust optimization* [14, 131], which allows for two sets of decisions—one that must be made before the uncertain data is realized, and one after. In fact, our problem can viewed as a simplified case of adjustable robust optimization having no first-stage decisions. On the other hand, our approach is distinguished by its application to sensitivity analysis and its use of copositive and semidefinite optimization.

We organize this chapter as follows. In Section 3.2, we extend many of the existing approaches for SA by considering simultaneous, general changes in $(\hat{b}, \hat{c})$ and the corresponding effect on the LP optimal value. Precisely, we model general perturbations of $(\hat{b}, \hat{c})$ within a compact, convex set $\mathcal{U}$—the uncertainty set, borrowing terminology from the robust-optimization literature—and define the corresponding minimum and maximum optimal values p^- and p^+, respectively. We call our approach *robust sensitivity analysis*, or *RSA*. Then, continuing in Section 3.2, we formulate the calculation of p^- and p^+ as nonconvex bilinear QPs (or BQPs) and briefly discuss attainability and complexity issues. We also discuss how p^- and p^+ may be infinite and suggest alternative bounded variants, q^- and q^+, which have the property that, if p^- is already finite, then $q^- = p^-$ and similarly for q^+ and p^+. Compared to related approaches in the literature, our discussion of finiteness is novel. We then

discuss the addition of redundant constraints to the formulations of q^- and q^+, which will strengthen later relaxations. Section 3.3 then establishes COP reformulations of the nonconvex BQPs by directly applying existing reformulation techniques. Then, based on the COPs, we develop tractable SDP-based relaxations that incorporate the structure of the uncertainty set $\mathcal{U}$, and we also discuss procedures for generating feasible solutions of the BQPs, which can also be used to verify the quality of the relaxation bounds. In Section 3.4, we validate our approach on several examples, which demonstrate that the relaxations provide effective approximations of q^+ and q^-. In fact, we find that the relaxations admit no gap with q^+ and q^- for all tested examples.

We mention some caveats about the approach. First, we focus only on how the optimal value is affected by uncertainty, not the optimal solution. We do so because we believe this will be a more feasible first endeavor; determining how general perturbations affect the optimal solution can certainly be a task for future research. Second, as mentioned above, we believe we are the first to consider these types of general perturbations, and thus the literature with which to compare is somewhat limited. However, we connect with the literature whenever possible, e.g., in special cases such as interval perturbations and worst-case linear optimization. Third, since we do not make any distributional assumptions about the uncertainty of the parameters, nor about their independence or dependence, we believe our approach aligns well with the general sprit of *robust optimization*. It is important to note, however, that our interest is *not* robust optimization and is *not* directly comparable to robust

optimization. For example, while in robust optimization one wishes to find a single optimal solution that works well for all realizations of the uncertain parameters, here we are only concerned with how the optimal value changes as the parameters change. Finally, we note the existence of other relaxations for nonconvex QPs including LP relaxations (see [125]) and Lasserre-type SDP relaxations. Generally speaking, LP-based relaxations are relatively weak (see [4]); we do not consider them in this work. In addition, SDP approaches can often be tailored to outperform the more general Lasserre approach as has been demonstrated in [111]. Our copostive- and SDP-based approach is similar; see for example the valid inequalities discussed in Section 3.3.2.

3.1.1 Notation, terminology, and copositive optimization

Let $\mathbb{R}^n$ denote n-dimensional Euclidean space represented as column vectors, and let $\mathbb{R}^n_+$ denote the nonnegative orthant in $\mathbb{R}^n$. For a scalar $p \geq 1$, the p-norm of $v \in \mathbb{R}^n$ is defined $\|v\|_p := (\sum_{i=1}^n |v_i|^p)^{1/p}$, e.g., $\|v\|_1 = \sum_{i=1}^n |v_i|$. We will drop the subscript for the 2-norm, e.g., $\|v\| := \|v\|_2$. For $v, w \in \mathbb{R}^n$, the inner product of v and w is defined as $v^T w = \sum_{i=1}^n v_i w_i$ and the Hadamard product of v and w is defined by $v \circ w := (v_1 w_1, ..., v_n w_n)^T \in \mathbb{R}^n$. $\mathbb{R}^{m \times n}$ denotes the set of real $m \times n$ matrices, and the trace inner product of two matrices $A, B \in \mathbb{R}^{m \times n}$ is defined $A \bullet B := \text{trace}(A^T B)$. $\mathcal{S}^n$ denotes the space of $n \times n$ symmetric matrices, and for $X \in \mathcal{S}^n$, $X \succeq 0$ denotes that X is positive semidefinite. In addition, diag(X) denotes the vector containing the diagonal entries of X.

3.2 Robust Sensitivity Analysis

In this section, we introduce the concept of robust sensitivity analysis (RSA) of the optimal value of (3.1). In particular, we define the best-case optimal value p^- and the worst-case optimal value p^+ over the uncertainty set $\mathcal{U}$. Next, we clarify when p^- and p^+ could be infinite and propose finite, closely related alternatives q^+ and q^-.

3.2.1 The best- and worst-case optimal values

In the Introduction, we have described $\hat{b}$ and $\hat{c}$ as varying, a concept that we now formalize. Hereafter, $(\hat{b}, \hat{c})$ denotes the *nominal*, "best guess" parameter values, and we let (b, c) denote perturbations with respect to $(\hat{b}, \hat{c})$. In other words, the true data could be $(\hat{b} + b, \hat{c} + c)$, and we think of b and c as varying. We also denote the *uncertainty set* containing all possible perturbations (b, c) as $\mathcal{U} \subseteq \mathbb{R}^m \times \mathbb{R}^n$. Throughout this chapter, we assume the following:

Assumption 3.1. $\mathcal{U}$ is compact, convex, and computationally tractable. In addition, $\mathcal{U}$ contains $(0, 0)$.

For the purposes of this chapter, by *computationally tractable*, we mean that $\mathcal{U}$ can be represented as the intersection of a finite number of linear and convex quadratic (or second-order-cone) inequalities.

Given $(b, c) \in \mathcal{U}$, we define the perturbed optimal value function at (b, c) as

$$p(b, c) := \min\{(\hat{c} + c)^T x : \hat{A}x = \hat{b} + b,\ x \geq 0\}. \tag{3.2}$$

For example, $p(0, 0)$ is the *nominal optimal value* of the *nominal problem* based on the

nominal parameters. The main idea of robust sensitivity analysis is then to compute the infimum (best-case) and supremum (worst-case) of all optimal values $p(b, c)$ over the uncertainty set $\mathcal{U}$, i.e., to calculate

$$p^- := \inf\{p(b, c) : (b, c) \in \mathcal{U}\}, \tag{3.3}$$

$$p^+ := \sup\{p(b, c) : (b, c) \in \mathcal{U}\}. \tag{3.4}$$

We illustrate p^- and p^+ with a small example.

Example 3.1. Consider the nominal LP

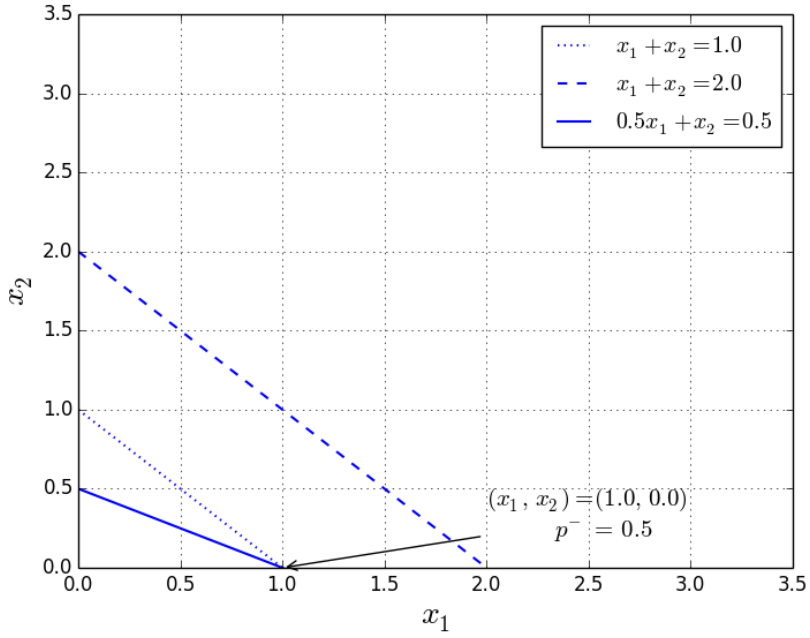
$$\begin{array}{ll} \min & x_1 + x_2 \\ \text{s.t.} & x_1 + x_2 = 2 \\ & x_1, x_2 \geq 0 \end{array} \tag{3.5}$$

and the uncertainty set

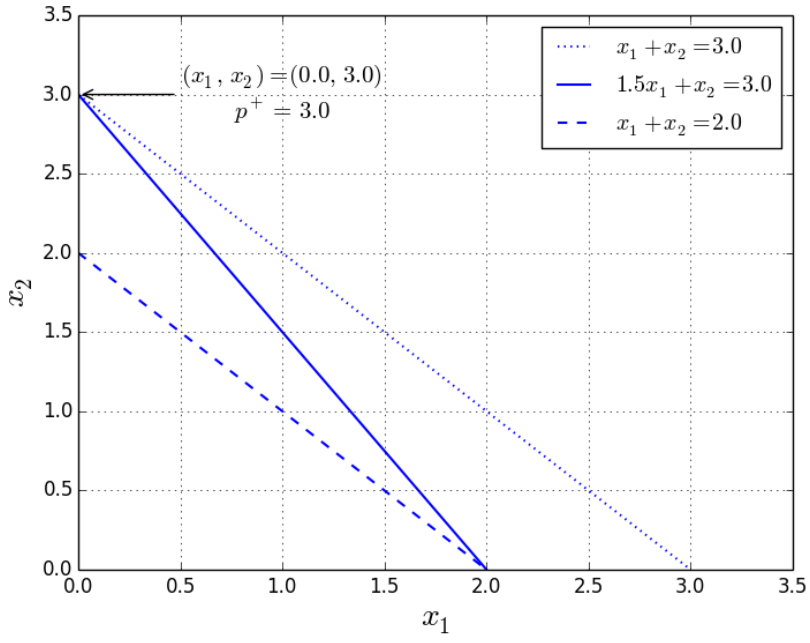
$$\mathcal{U} := \left\{ (b, c) : \begin{array}{c} b_1 \in [-1, 1] \\ c_1 \in [-0.5, 0.5],\ c_2 = 0 \end{array} \right\}.$$

Note that the perturbed data $\hat{b}_1 + b_1$ and $\hat{c}_1 + c_1$ remain positive, while $\hat{c}_2 + c_2$ is constant. Thus, the minimum optimal value p^- occurs when b_1 and c_1 are minimal, i.e., when $b_1 = -1$ and $c_1 = -0.5$. In this case, $p^- = 0.5$ at the solution $(x_1, x_2) = (1, 0)$. In a related manner, $p^+ = 3$ when $b_1 = 1$ and $c_1 = 0.5$ at the point $(x_1, x_2) = (0, 3)$. Actually, any perturbation with $c_1 \in [0, 0.5]$ and $b_1 = 1$ realizes the worst-case value $p^+ = 3$. Figure 3.1 illustrates this example.

We can obtain a direct formulation of p^- by simply collapsing the inner and outer minimizations of (3.3) into a single BLP:



(a) Illustration of the best-case optimal value.



(b) Illustration of the worst-case optimal value.

Figure 3.1: Illustration of Example 3.1. Note that the dashed line in both (3.1a) and (3.1b) corresponds to the feasible region of the nominal problem.

$$\begin{aligned} p^- = \quad & \inf_{b,c,x} \quad (\hat{c}+c)^T x \\ & \text{s. t.} \qquad \hat{A}x = \hat{b}+b,\ x \geq 0 \\ & \qquad\qquad (b,c) \in \mathcal{U}. \end{aligned} \tag{3.6}$$

The nonconvexity comes from the bilinear term $c^T x$ in the objective function. In the special case that $(b,c) \in \mathcal{U}$ implies $c = 0$, i.e., when there is no perturbation in the objective coefficients, we have the following:

Remark. If $c = 0$ for all $(b,c) \in \mathcal{U}$, then p^- can be computed in polynomial time as the optimal value of (3.6) with $c = 0$, which is a tractable convex program due to Assumption 3.1.

A direct formulation for p^+ can, under a fairly weak assumption, be gotten via duality. Define the perturbed primal and dual feasible sets for any $(b,c) \in \mathcal{U}$:

$$P(b) := \{x : \hat{A}x = \hat{b}+b, x \geq 0\},$$

$$D(c) := \{(y,s) : \hat{A}^T y + s = \hat{c}+c, s \geq 0\}.$$

For instance, $P(0)$ and $D(0)$ are the primal-dual feasible sets of the nominal problem. Next define the dual LP for (3.2) as

$$d(b,c) := \max\{(\hat{b}+b)^T y : (y,s) \in D(c)\}.$$

Considering the extended notion of strong duality, which handles the cases of infinite values, we have that $d(b,c) = p(b,c)$ when at least one of $P(b)$ and $D(c)$ is nonempty. Hence, under the assumption that every $(b,c) \in \mathcal{U}$ yields $P(b) \neq \emptyset$ or $D(c) \neq \emptyset$, a direct formulation for p^+ can be constructed by replacing $p(b,c)$ in (3.4) with $d(b,c)$

and then collapsing the subsequent inner and outer maximizations into the BLP

$$\begin{aligned} p^+ = \quad & \sup_{b,c,y,s} \quad (\hat{b}+b)^T y \\ & \text{s. t.} \qquad \hat{A}^T y + s = \hat{c} + c,\ s \geq 0 \\ & \qquad\qquad (b,c) \in \mathcal{U}. \end{aligned} \tag{3.7}$$

Here again, the nonconvexity arises due to the bilinear term $b^T y$ in the objective. If $(b,c) \in \mathcal{U}$ implies $b = 0$, then p^+ can be calculated in polynomial time:

Remark. If $b = 0$ for all $(b,c) \in \mathcal{U}$, then p^+ can be computed in polynomial time as the optimal value of (3.7) with $b = 0$, which is a tractable convex program due to Assumption 3.1.

We summarize the above discussion in the following proposition:

Proposition 3.1. The best-case value p^- equals the optimal value of (3.6). Moreover, if $P(b) \neq \emptyset$ or $D(c) \neq \emptyset$ for all $(b,c) \in \mathcal{U}$, then the worst-case value p^+ equals the optimal value of (3.7).

We view the condition in Proposition 3.1—that at least one of $P(b)$ and $D(c)$ is nonempty for each $(b,c) \in \mathcal{U}$—to be rather mild. Said differently, the case that $P(b) = D(c) = \emptyset$ for some $(b,c) \in \mathcal{U}$ appears somewhat pathological. For practical purposes, we hence consider (3.7) to be a valid formulation of p^+. Actually, in the next subsection, we will further restrict our attention to those $(b,c) \in \mathcal{U}$ for which both $P(b)$ and $D(c)$ are nonempty. In such cases, each $p(b,c)$ is guaranteed to be finite, which—as we will show—carefully handles the cases when p^+ and p^- are infinite.

Indeed, the worst-case value p^+ could equal $+\infty$ due to some perturbed $P(b)$ being empty as shown in the following example:

Example 3.2. In Example 3.1, change the uncertainty set to

$$\mathcal{U} := \left\{ (b,c) : \begin{array}{c} b_1 \in [-3,1] \\ c_1 \in [-0.5,0.5], c_2 = 0 \end{array} \right\}.$$

Then $p(b,c) = +\infty$ whenever $b_1 \in [-3,-2)$ since then the primal feasible set $P(b)$ is empty. Then $p^+ = +\infty$ overall. However, limiting b_1 to $[-2,1]$ yields a worst-case value of 3 as discussed in Example 3.1.

Similarly, p^- might equal $-\infty$ due to some perturbed LP having unbounded objective value, implying infeasibility of the corresponding dual feasible set $D(c)$.

3.2.2 Attainment and complexity

By an existing result concerning the attainment of the optimal value of BLPs, we have that p^- and p^+ are attainable when $\mathcal{U}$ has a relatively simple structure:

Proposition 3.2 (Theorem 2 of [100]). Suppose $\mathcal{U}$ is representable by a finite number of linear constraints and at most one convex quadratic constraint. Then, if the optimal value of (3.6) is finite, it is attained. A similar statement holds for (3.7).

In particular, attainment holds when $\mathcal{U}$ is polyhedral or second-order-cone representable with at most one second-order cone. Moreover, the bilinear nature of (3.6) implies that, if the optimal value is attained, then there exists an optimal solution (x^*, b^*, c^*) with (b^*, c^*) an extreme point of $\mathcal{U}$ and x^* an extreme point of $P(b^*)$. The same holds for (3.7) if its optimal value is attained.

As discussed in the Introduction, the worst-case value p^+ has been studied by Peng and Zhu [111] for the special case when $c = 0$ and b is contained in an ellipsoid.

The authors demonstrate (see their proposition 1.1) that calculating p^+ in this case is NP-hard. By the symmetry of duality, it thus also holds that p^- is NP-hard to compute in general.

3.2.3 Finite variants of p^- and p^+

We require the following feasibility and boundedness assumption:

Assumption 3.2. Both feasible sets $P(0)$ and $D(0)$ are nonempty, and one is bounded.

By standard theory, $P(0)$ and $D(0)$ cannot both be nonempty and bounded. Also define

$$\overline{\mathcal{U}} := \{(b,c) \in \mathcal{U} : P(b) \neq \emptyset, D(c) \neq \emptyset\}.$$

Note that $(0,0) \in \overline{\mathcal{U}}$ due to Assumption 3.2. In fact, $\overline{\mathcal{U}}$ can be captured with linear constraints that enforce primal-dual feasibility and hence is a compact, convex subset of $\mathcal{U}$:

$$\overline{\mathcal{U}} = \left\{ (b,c) \in \mathcal{U} : \exists\, (x,y,s) \text{ such that } \begin{array}{c} \hat{A}x = \hat{b} + b, x \geq 0 \\ \hat{A}^T y + s = \hat{c} + c, s \geq 0 \end{array} \right\}.$$

Analogous to p^+ and p^-, define

$$q^+ := \sup\{p(b,c) : (b,c) \in \overline{\mathcal{U}}\} \tag{3.8}$$

$$q^- := \inf\{p(b,c) : (b,c) \in \overline{\mathcal{U}}\}. \tag{3.9}$$

The following proposition establishes the finiteness of q^+ and q^-:

Proposition 3.3. Under Assumptions 3.1 and 3.2, both q^+ and q^- are finite.

Proof. We prove the contrapositive for q^-. (The argument for q^+ is similar.) Suppose $q^- = -\infty$. Then there exists a sequence $\{(b^k, c^k)\} \subseteq \overline{\mathcal{U}}$ with finite optimal values $p(b^k, c^k) \to -\infty$. By strong duality, there exists a primal-dual solution sequence $\{(x^k, y^k, s^k)\}$ with $(\hat{c} + c^k)^T x^k = (\hat{b} + b^k)^T y^k \to -\infty$. Since $\overline{\mathcal{U}}$ is bounded, it follows that $\|x^k\| \to \infty$ and $\|y^k\| \to \infty$.

There exists a $\hat{k}$ such that $\|x^k\| \neq 0$ for all $k \geq \hat{k}$ as $\|x^k\| \to \infty$. Consider the sequence $\{(z^k, d^k)\}$ with $(z^k, d^k) := (x^k, b^k)/\|x^k\|$ for all $k \geq \hat{k}$. We have $\hat{A}z^k = \hat{b}/\|x^k\| + d^k$, $z^k \geq 0$, and $\|z^k\| = 1$. Moreover, $\hat{b}/\|x^k\| + d^k \to 0$. Hence, there exists a subsequence converging to $(\bar{z}, 0)$ such that $\hat{A}\bar{z} = 0$, $\bar{z} \geq 0$, and $\|\bar{z}\| = 1$. This proves that the recession cone of $P(0)$ is nontrivial, and hence $P(0)$ is unbounded. In a similar manner, $D(0)$ is unbounded, which means Assumption 3.2 does not hold. □

Note that the proof of Proposition 3.3 only assumes that $\mathcal{U}$, and hence $\overline{\mathcal{U}}$, is bounded, which does not use the full power of Assumption 3.1.

We can also employ pertrubation theory for linear programming [118] to argue that, since every primal-dual pair of LPs implicit in the definition of q^+ and q^- are primal-dual feasible by construction, then the function $p(b, c)$ is continuous over $(b, c) \in \overline{\mathcal{U}}$. Since $\overline{\mathcal{U}}$ is compact by assumption, it follows that both q^+ and q^- are attained, allowing us to replace *sup* and *inf* by *max* and *min*.

Similar to p^-, a direct formulation of q^- can be constructed by employing the primal-dual formulation of $\overline{\mathcal{U}}$ and by collapsing the inner and outer minimizations of

(3.9) into a BLP:

$$\begin{array}{rcll} q^- & = & \min_{b,c,x,y,s} & (\hat{c}+c)^T x \\ & & \text{s. t.} & \hat{A}x = \hat{b}+b,\ x \geq 0 \\ & & & \hat{A}^T y + s = \hat{c}+c,\ s \geq 0 \\ & & & (b,c) \in \mathcal{U}. \end{array} \tag{3.10}$$

Similar to p^+, after replacing $p(b,c)$ in (3.8) by $d(b,c)$, we can collapse the inner and outer maximizations into the BLP:

$$\begin{array}{rcll} q^+ & = & \max_{b,c,x,y,s} & (\hat{b}+b)^T y \\ & & \text{s. t.} & \hat{A}x = \hat{b}+b,\ x \geq 0 \\ & & & \hat{A}^T y + s = \hat{c}+c,\ s \geq 0 \\ & & & (b,c) \in \mathcal{U}. \end{array} \tag{3.11}$$

The following proposition establishes $q^+ = p^+$ when p^+ is finite and, similarly, $q^- = p^-$ when p^- is finite.

Proposition 3.4. If p^+ is finite, then $q^+ = p^+$, and if p^- is finite, then $q^- = p^-$.

Proof. We prove the second statement only since the first is similar. Comparing the formulation (3.6) for p^- and the formulation (3.10) for q^-, it is clear that $p^- \leq q^-$. In addition, let (b,c,x) be any feasible solution of (3.6). Because p^- is finite, $p(b,c)$ is finite. Then the corresponding dual problem is feasible, which implies that we can extend (b,c,x) to a solution (b,c,x,y,s) of (3.10) with the same objective value. Hence, $p^- \geq q^-$. □

In the remaining sections of the chapter, we will focus on the finite variants q^- and q^+ given by the nonconvex QPs (3.10) and (3.11), which optimize the optimal

value function $p(b, c) = d(b, c)$ based on enforcing primal-dual feasibility. It is clear that we may also enforce the complementary slackness condition $x \circ s = 0$ without changing these problems. Although it might seem counterintuitive to add the redundant, nonconvex constraint $x \circ s = 0$ to an already difficult problem, in Section 3.3, we will propose convex relaxations to approximate q^- and q^+, in which case—as we will demonstrate—the relaxed versions of the redundant constraint can strengthen the relaxations.

3.3 Copositive Formulations and Relaxations

3.3.1 Copositive formulations

In order to formulate (3.10) and (3.11) as COPs, we apply a result of [36]; see also [35, 51, 60]. Consider the general nonconvex QP

$$\begin{aligned} \min \quad & z^T W z + 2\, w^T z \\ \text{s.t.} \quad & Ez = f,\ z \in \mathcal{K} \end{aligned} \tag{3.12}$$

where $\mathcal{K}$ is a closed, convex cone. Its copositive reformulation is

$$\begin{aligned} \min \quad & W \bullet Z + 2\, w^T z \\ \text{s.t.} \quad & Ez = f,\ \operatorname{diag}(EZE^T) = f \circ f \\ & \begin{pmatrix} 1 & z^T \\ z & Z \end{pmatrix} \in \mathcal{CPP}(\mathbb{R}_+ \times \mathcal{K}), \end{aligned} \tag{3.13}$$

as established by the following lemma:

Lemma 3.1 (Corollary 8.3 in [36])**.** *Problem (3.12) is equivalent to (3.13), i.e.: (i)*

both share the same optimal value; (ii) if (z^*, Z^*) *is optimal for (3.13), then* z^* *is in the convex hull of optimal solutions for (3.12).*

The following theorem establishes that problems (3.10) and (3.11) can be reformulated as copositive programs according to Lemma 3.1. The proof is based on describing how the two problems fit the form (3.12).

Theorem 3.2. *Problems (3.10) and (3.11) to compute* q^- *and* q^+ *are solvable as copositive programs of the form (3.13), where*

$$\mathcal{K} := \hom(\mathcal{U}) \times \mathbb{R}^n_+ \times \mathbb{R}^m \times \mathbb{R}^n_+$$

and

$$\hom(\mathcal{U}) := \{(t, b, c) \in \mathbb{R}_+ \times \mathbb{R}^m \times \mathbb{R}^n : t > 0,\ (b, c)/t \in \mathcal{U}\} \cup \{(0, 0, 0)\}$$

is the homogenization of $\mathcal{U}$*.*

Proof. We prove the result for just problem (3.10) since the argument for problem (3.11) is similar. First, we identify $z \in \mathcal{K}$ in (3.12) with $(t, b, c, x, y, s) \in \hom(\mathcal{U}) \times \mathbb{R}^n_+ \times \mathbb{R}^m \times \mathbb{R}^n_+$ in (3.10). In addition, in the constraints, we identify $Ez = f$ with the equations $\hat{A}x = t\hat{b} + b$, $\hat{A}^T y + s = t\hat{c} + c$, and $t = 1$. Note that the right-hand-side vector f is all zeros except for a single entry corresponding to the constraint $t = 1$. Moreover, in the objective, $z^T W z$ is identified with the bilinear term $c^T x$, and $2\, w^T z$ is identified with the linear term $\hat{c}^T x$. With this setup, it is clear that (3.10) is an instance of (3.12) and hence Lemma 3.1 applies to complete the proof. □

Under Assumption 3.1, $\mathcal{U}$ is tractable, and so $\mathcal{K}$ is as well. This is a key ingredient for devising tractable conic relaxations of (3.13); see Section 3.3.2 next.

3.3.2 SDP-based conic relaxations

We propose relaxations that are formed from (3.13) by relaxing the cone constraint

$$M := \begin{pmatrix} 1 & z^T \\ z & Z \end{pmatrix} \in \mathcal{CPP}(\mathbb{R}_+ \times \mathcal{K}).$$

As is well known—and direct from the definitions—cones of the form $\mathcal{CPP}(\cdot)$ are contained in the positive semidefinite cone. Hence, we will enforce $M \succeq 0$. It is also true that $M \in \mathcal{CPP}(\mathbb{R}_+ \times \mathcal{K})$ implies $z \in \mathcal{K}$, although $M \succeq 0$ does not necessarily imply this. So, in our relaxations, we will also enforce $z \in \mathcal{K}$. Including $z \in \mathcal{K}$ improves the relaxation and also helps in the calculation of bounds in Section 3.3.3.

Next, suppose that the description of $\mathbb{R}_+ \times \mathcal{K}$ contains at least two linear constraints, $a_1^T z \leq b_1$ and $a_2^T z \leq b_2$. By multiplying $b_1 - a_1^T z$ and $b_2 - a_2^T z$, we obtain a valid, yet redundant, quadratic constraint $b_1 b_2 - b_1 a_2^T z - b_2 a_1^T z + a_1^T z z^T a_2 \geq 0$ for (3.12). This quadratic inequality can in turn be linearized in terms of M as $b_1 b_2 - b_1 a_2^T z - b_2 a_1^T z + a_1^T Z a_2 \geq 0$, which is valid for $\mathcal{CPP}(\mathbb{R}_+ \times \mathcal{K})$. We add this linear inequality to our relaxation; it is called an *RLT constraint* [125]. In fact, we add all such RLT constraints arising from all pairs of linear constraints present in the description of $\mathbb{R}_+ \times \mathcal{K}$.

When the description of $\mathbb{R}_+ \times \mathcal{K}$ contains at least one linear constraint $a_1^T z \leq b_1$ and one second-order-cone constraint $\|d_2 - C_2^T z\| \leq b_2 - a_2^T z$, where d_2 is a vector and C_2 is a matrix, we will add a so-called *SOC-RLT constraint* to our relaxation [39]. The constraint is derived by multiplying the two constraints to obtain the valid

quadratic second-order-cone constraint

$$\|(b_1 - a_1^T z)(d_2 - C_2^T z)\| \leq (b_1 - a_1^T z)(b_2 - a_2^T z).$$

After linearization by M, we have the second-order-cone constraint

$$\|b_1 d_2 - d_2 a_1^T z - b_1 C_2^T z + C_2^T Z a_1\| \leq b_1 b_2 - b_1 a_2^T z - b_2 a_1^T z + a_1^T Z a_2.$$

Finally, recall the redundant complementarity constraint $x \circ s = 0$ described at the end of Section 3.2.3, which is valid for both (3.10) and (3.11). Decomposing it as $x_i s_i = 0$ for $i = 1, \ldots, n$, we may translate these n constraints to (3.13) as $z^T H_i z = 0$ for appropriately defined matrices matrices H_i. Then they may be linearized and added to our relaxation as $H_i \bullet Z = 0$.

The RLT and SOC-RLT constraints discussed here are not uniquely effective for BLPs such as (3.10) and (3.11). In fact, Anstreicher [4] provides a theoretical and geometric explanation of how the positive semidefinite constraint and RLT constraints together provide a tight relaxation for general nonconvex quadratically constrained quadratic programs. The power of RLT and SOC-RLT constraints in SDP relaxations is further discussed for extended trust-region subproblems [39]. In particular, in some cases, such relaxations are exact. More theoretical results concerning RLT and SOC-RLT constraints are discussed in [3, 143]. On the other hand, there is little to no literature concerning the effectiveness of the linearized complementarity constraints from a theoretical perspective. Computationally, Chen and Burer [43] demonstrate that the linearized complementarity constraints derived from the KKT conditions of nonconvex quadratic programming can strengthen corresponding SDP relaxations.

For our examples, we discuss the empirical strength of the linearized complementarity constraints in Section 3.4.5.

To summarize, let RLT denote the set of (z, Z) satisfying all the derived RLT constraints, and similarly, define SOCRLT as the set of (z, Z) satisfying all the derived SOC-RLT constraints. Then the SDP-based conic relaxation for (3.13) that we propose to solve is

$$\begin{aligned} \min \quad & W \bullet Z + 2\, w^T z \\ \text{s. t.} \quad & Ez = f,\ \operatorname{diag}(EZE^T) = f \circ f \\ & H_i \bullet Z = 0 \ \ \forall\, i = 1, \ldots, n \\ & (z, Z) \in \mathrm{RLT} \cap \mathrm{SOCRLT} \\ & \begin{pmatrix} 1 & z^T \\ z & Z \end{pmatrix} \succeq 0,\ z \in \mathcal{K}. \end{aligned} \tag{3.14}$$

It is worth mentioning that, in many cases, the RLT and SOC-RLT constraints will already imply $z \in \mathcal{K}$, but in such cases, we nevertheless write the constraint in (3.14) for emphasis; see also Section 3.3.3 below. Furthermore, we emphasize that (3.14) is computationally tractable as $\mathcal{K}$ is tractable under Assumption 3.1.

When translated to the problem (3.10) for calculating q^-, the relaxation (3.14) gives rise to a lower bound $q^-_{\mathrm{sdp}} \leq q^-$. Similarly, when applied to (3.11), we get an upper bound $q^+_{\mathrm{sdp}} \geq q^+$.

3.3.3 Bounds from feasible solutions

We now discuss two methods to approximate q^- from above and q^+ from below, i.e., to bound q^- and q^+ using feasible solutions of (3.10) and (3.11), respectively.

The first method, which has been inspired by [111], utilizes the optimal solution of the SDP relaxation (3.14). Let us discuss how to obtain such a bound for (3.10), as the discussion for (3.11) is similar. We first observe that any feasible solution (z, Z) of (3.14) satisfies $Ez = f$ and $z \in \mathcal{K}$, i.e., z satisfies all of the constraints of (3.12). Since (3.12) is equivalent to (3.10) under the translation discussed in the proof of Theorem 3.2, z gives rise to a feasible solution (x, y, s, b, c) of (3.10). From this feasible solution, we can calculate $(\hat{c} + c)^T x \geq q^-$. In practice, we will start from the optimal solution (z^-, Z^-) of (3.14). Note that computing (z^-, Z^-) is computationally tractable as (3.14) is tractable. We summarize this approach in the following remark.

Remark. Suppose that (z^-, Z^-) is an optimal solution of the relaxation (3.14) corresponding to (3.10), and let $(x^-, y^-, s^-, b^-, c^-)$ be the translation of z^- to a feasible point of (3.10). Then, $r^- := (\hat{c} + c^-)^T x^- \geq q^-$. Similarly, we define $r^+ := (\hat{b} + b^+)^T y^+ \leq q^+$ based on an optimal solution (z^+, Z^+) of (3.14) corresponding to (3.11).

Our second method for bounding q^- and q^+ using feasible solutions is a sampling procedure detailed in Algorithm 3.1. The main idea is to generate randomly a point $(b, c) \in \overline{\mathcal{U}}$ and then to calculate $p(b, c)$, which serves as an upper bound of q^- and a lower bound of q^+, i.e., $q^- \leq p(b, c) \leq q^+$. Multiple points (b^k, c^k) and values $p^k := p(b^k, c^k)$ are generated and the best bounds $q^- \leq v^- := \min_k\{p^k\}$ and

$\max_k\{p^k\} =: v^+ \leq q^+$ are saved. In fact, by the bilinearity of (3.10) and (3.11), we may restrict attention to the extreme points (b, c) of $\overline{\mathcal{U}}$ without reducing the quality of the resultant bounds; see also the discussion in Section 3.2.2. Hence, Algorithm 3.1 generates—with high probability—a random extreme point of $\overline{\mathcal{U}}$ by optimizing a random linear objective over $\overline{\mathcal{U}}$, and we generate the random linear objective as a vector uniform on the sphere, which is implemented by a well-known, quick procedure. Note that, even though the random objective is generated according to a specific distribution, we cannot predict the resulting distribution over the extreme points of $\overline{\mathcal{U}}$.

Algorithm 3.1 Sampling procedure to bound q^- from above and q^+ from below

Inputs: Instance with uncertainty set $\mathcal{U}$ and restricted uncertainty set $\overline{\mathcal{U}}$. Number of random trials T.

Outputs: Bounds $v^- := \min_k\{p^k\} \geq q^-$ and $v^+ := \max_k\{p^k\} \leq q^+$.

for $k = 1, \ldots, T$ **do**

Generate $(f, g) \in \mathbb{R}^m \times \mathbb{R}^n$ uniformly on the unit sphere.

Calculate $(b^k, c^k) \in \operatorname{Arg\,min}\{f^T b + g^T c : (b, c) \in \overline{\mathcal{U}}\}$.

Set $p^k := p(b^k, c^k)$.

end for

As all four of the bounds r^-, r^+, v^-, and v^+ are constructed from feasible solutions, we can further improve them heuristically by exploiting the bilinear objective

functions in (3.10) and (3.11). In particular, we employ the standard local improvement heuristic for programs with a bilinear objective and convex constraints (e.g., see [92]). Suppose, for example, that we have a feasible point $(x^-, y^-, s^-, b^-, c^-)$ for problem (3.10) as discussed in Remark 3.3.3. To attempt to improve the solution, we fix the variable c in (3.10) at the value c^-, and we solve the resulting convex problem for a new, hopefully better point $(x^1, y^1, s^1, b^1, c^1)$, where $c^1 = c^-$. Then, we fix x to x^1, resolve, and get a new point $(x^2, y^2, s^2, b^2, c^2)$, where $x^2 = x^1$. This alternating process is repeated until there is no further improvement in the objective of (3.10), and the final objective is our bound r^-.

In Section 3.4 below, we use the bounds r^-, r^+, v^-, and v^+ to verify the quality of our bounds q^-_{sdp} and q^+_{sdp}. Our tests indicate that neither bound, r^- nor v^-, dominates the other—and similarly for the bounds r^+ and v^+. Hence, we will actually report the better of each pair: $\min\{r^-, v^-\}$ and $\max\{r^+, v^+\}$. Also, for the calculations of v^- and v^+, we always take $T = 10,000$ in Algorithm 3.1.

3.4 Computational Experiments

We test our approach on six examples from the literature as well as an example of our own. The first three examples in Section 3.4.1 correspond to classical sensitivity analysis approaches for LP; the fourth example in Section 3.4.2 corresponds to an interval LP in inventory management; the fifth example in Section 3.4.3 corresponds to a systemic-risk calculation in financial systems; and the last example in Section 3.4.4 is a transportation network flow problem. We implement our tests in Python

(version 3.5.2) with Mosek (version 8.0.0.28 beta, limited to 1 thread) as our convex-optimization solver. All of Mosek's settings are set at their defaults, and computations are conducted on a Macintosh OS X El Capitan system with a quad-core 3.20GHz Intel Core i5 CPU and 8 GB RAM.

3.4.1 Examples from classical sensitivity analysis

Consider the following nominal problem from [135]:

$$\begin{aligned} \min \quad & -12x_1 - 18x_2 - 18x_3 - 40x_4 \\ \text{s. t.} \quad & 4x_1 + 9x_2 + 7x_3 + 10x_4 + x_5 = 6000 \\ & x_1 + x_2 + 3x_3 + 40x_4 + x_6 = 4000 \\ & x_1, \ldots, x_6 \geq 0. \end{aligned}$$

One of the optimal basis is $B = \{1, 4\}$ with optimal solution $\frac{1}{3}(4000, 0, 0, 200, 0, 0)$ and the optimal value $p(0, 0) = -18667$. We next study three examples that consider perturbations only in the cost coefficients of the nominal problem above.

According to standard, "textbook" sensitivity analysis, the optimal basis B persists when the coefficient of x_1 lies in the interval $[-16, -10]$ and other parameters remain the same. Along this interval, one can easily compute the best-case value -24000 and worst-case value -16000, and we attempt to reproduce this analysis with our approach. So let us choose the uncertainty set

$$\mathcal{U} = \left\{ (b, c) \in \mathbb{R}^2 \times \mathbb{R}^6 : \begin{array}{c} b_1 = b_2 = 0 \\ c_1 \in [-4, 2] \\ c_2 = \cdots = c_6 = 0 \end{array} \right\},$$

which corresponds precisely to the above allowable decrease and increase on the coefficient of x_1. Note that Assumptions 3.1 and 3.2 are satisfied. We thus know from above that $q^- = -24000$ and $q^+ = -16000$. Since $b = 0$ in $\mathcal{U}$, Remark 3.2.1 implies that q^+ is easy to calculate. So we apply our approach, i.e., solving the SDP-based relaxation, to approximate q^-. The relaxation value is $q^-_{\text{sdp}} = -24000$, which recovers q^- exactly. The CPU time for computing q^-_{sdp} is 0.10 seconds.

Our second example is based on the same nominal problem, but we consider the 100%-rule. The readers are refereed to [34] for the details of the mechanism. By the 100%-rule, the nominal optimal basis $B = \{1, 4\}$ persists for $(\tilde{c}_1, \tilde{c}_2)$ in the following simplex:

$$\left\{ (\tilde{c}_1, \tilde{c}_2) : \begin{array}{c} \tilde{c}_1 \in [-16, -12] \\ \tilde{c}_2 \in [-134/3, -18] \\ \frac{-12-\tilde{c}_1}{4} + \frac{-18-\tilde{c}_2}{80/3} \leq 1 \end{array} \right\}.$$

By evaluating the three extreme points $(-12, -18)$, $(-16, -18)$ and $(-12, -134/3)$ of this set with respect to the nominal optimal solution, one can calculate the best-case optimal value as $q^- = -24000$ and the worst-case optimal value as $q^+ = -18667$. We again apply our approach in an attempt to recover empirically the 100%-rule. Specifically, let

$$\mathcal{U} = \left\{ (b, c) : \begin{array}{c} b_1 = b_2 = 0 \\ c_1 \in [-4, 0],\ c_2 \in [-\frac{80}{3}, 0] \\ -\frac{c_1}{4} - \frac{c_2}{80/3} \leq 1 \\ c_3 = \cdots = c_6 = 0 \end{array} \right\}.$$

Note that Assumptions 3.1 and 3.2 are satisfied. Due to $b = 0$ and Remark 3.2.1, we

focus our attention on q^-. Calculating the SDP-based relaxation value, we see that $q^-_{\text{sdp}} = -24000$, which recovers q^- precisely. The CPU time is 0.15 seconds.

Our third example illustrates the tolerance approach, and we continue to use the same nominal problem from [135]. The tolerance approach considers simultaneous and independent perturbations in the objective coefficients by calculating a maximum tolerance percentage such that, as long as selected coefficients are accurate to within that percentage of their nominal values, the nominal optimal basis persists; see details in [138]. Let us consider perturbations in the coefficients of x_1 and x_2 with respect to the nominal problem. The tolerance approach ensures that the same nominal optimal basis $B = \{1, 4\}$ persists for $(\tilde{c}_1, \tilde{c}_2)$ in the following set

$$\left\{ (\tilde{c}_1, \tilde{c}_2) : \ \ c_1 \in [-2, 2], c_2 \in [-3, 3] \ \right\}.$$

By testing the four extreme points of the box of changes $[-14, -10] \times [-21, -15]$ with respect to the optimal nominal solution, one can calculate the best-case optimal value as $q^- = -21333$ and the worst-case optimal value as $q^+ = -16000$. To test our approach in this setting, we set

$$\mathcal{U} := \left\{ (b, c) : \begin{array}{c} b_1 = b_2 = 0, \ c_3 = \cdots = c_6 = 0 \\ \\ c_1 \in [-2, 2], c_2 \in [-3, 3] \end{array} \right\}$$

and, as in the previous two examples, we focus on q^-. Assumptions 3.1 and 3.2 are again satisfied, and we calculate the lower bound $q^-_{\text{sdp}} = -21333$, which recovers q^- precisely. The CPU time for computing q^-_{sdp} is 0.13 seconds.

3.4.2 An example from interval linear programming

We consider an optimization problem that is typical in inventory management, and this particular example originates from [69]. Let T be the number of periods; s_k and x_k be the available stock and the quantity ordered respectively at the beginning of period k; y_k be the holding cost or shortage cost at the period k. The order quantities x_k are nonnegative and further subject to uniform upper and lower bounds, u and l, and every stock level s_k is bounded above by U. At time k, the purchase cost is denoted as c_k, the holding cost is denoted as h_k, and the shortage cost is denoted g_k. The goal of this model is to satisfy exogenous demands d_k for each period k, while simultaneously minimizing the total of purchasing, holding, and shortage costs. Then, the problem can be formulated as the following linear programming problem (assuming that the initial inventory is 0):

$$\begin{aligned}
\min \quad & \textstyle\sum_{k=1}^{T}(c_k x_x + y_k) \\
\text{s.t.} \quad & s_0 = 0 \\
& s_{k-1} + x_k - s_k = d_k \qquad k = 1, \ldots, T \\
& y_k \geq h_k s_k \qquad\qquad k = 1, \ldots, T \\
& y_k \geq -g_k s_k \qquad\qquad k = 1, \ldots, T \\
& l \leq x_k \leq u \qquad\qquad k = 1, \ldots, T \\
& s_k \leq U \qquad\qquad k = 1, \ldots, T \\
& x_k, y_k \geq 0 \qquad\qquad k = 1, \ldots, T.
\end{aligned} \tag{3.15}$$

As in [69], consider an instance of (3.15) in which $T = 4$, $u = 1500, l = 1000$, $U = 600$, and all costs are as in Table 3.1. Moreover, suppose the demands

Period (k)	Purchasing cost (c_k)	Holding cost (h_k)	Shortage cost (g_k)
1	7	2	3
2	1	1	4
3	10	1	3
4	6	1	3

Table 3.1: Costs for each period of an instance of the inventory management problem.

d_k are each uncertain and may be estimated by the intervals $d_1 \in [700, 900], d_2 \in [1300, 1600], d_3 \in [900, 1100]$, and $d_4 \in [500, 700]$. From [69], the worst-case optimal value over this uncertainty set is $q^+ = 25600$. For our approach, it is easy to verify that Assumptions 3.1 and 3.2 are satisfied, and solving our SDP-based conic relaxation with an uncertainty set corresponding to the intervals on d_k, we recover q^+ exactly, i.e., we have $q^+_{\text{sdp}} = 25600$. The CPU time for computing our SDP optimal value is 1,224 seconds.

Since the uncertainties only involve the right-hand sides, Remark 3.2.1 implies that the best-case value q^- can be calculated in polynomial-time by solving an LP that directly incorporates the uncertainty.

3.4.3 Worst-case linear optimization

For an interbank market, systemic risk is used to evaluate the potential loss of the whole market as a response to the decisions made by the individual banks

[61, 111]. For a market consisting n banks, we use an $n \times n$ matrix $\hat{L}$ to denote the liability relationship between any two banks in the market. We use $\hat{b}_i$ to denote the exogenous operating cash flow received by bank i. Given the vector $\hat{b}$, we calculate the systemic loss $l(\hat{b})$ of the market, which measures the amount of overall failed liabilities [111]:

$$\begin{aligned} l(\hat{b}) = \quad & \min_{x} \quad \textstyle\sum_{i=1}^{n}(1 - x_i) \\ & \text{s.t.} \quad (\textstyle\sum_{j=1}^{n} \hat{L}_{ij})x_i - \sum_{j=1}^{n} \hat{L}_{ji}x_j \leq \hat{b}_i \quad \forall\, i = 1, \ldots, n \\ & \qquad\quad x_i \leq 1 \qquad\qquad\qquad\qquad\qquad\quad \forall\, i = 1, \ldots, n. \end{aligned}$$

Here the decision variables x_i represent the ratio of the total payment by bank i to the total obligation of bank i. These ratios are naturally less than or equal to 1 ($x_i \leq 1$) as the banks do not pay more than their obligations. In contrast, $1 - x_i$ denotes the ratio of bank i failing to fulfill its obligations.

In practice, however, there exist uncertainties in the exogenous operating cash flows. Allowing for uncertainties, the worst-case systemic risk problem [111] is given as

$$\begin{aligned} \max_{b \in \mathcal{V}} \quad \min_{x} \quad & \textstyle\sum_{i=1}^{n}(1 - x_i) \\ & \text{s.t.} \quad (\textstyle\sum_{j=1}^{n} \hat{L}_{ij})x_i - \sum_{j=1}^{n} \hat{L}_{ji}x_j \leq \hat{b}_i + Q_{i\cdot}b \quad \forall\, i = 1, \ldots, n \\ & \qquad\quad x_i \leq 1 \qquad\qquad\qquad\qquad\qquad\quad \forall\, i = 1, \ldots, n. \end{aligned}$$

where $\mathcal{V} := \{b \in \mathbb{R}^m : \|b\| \leq 1\}$ denotes the uncertainty set, $Q \in \mathbb{R}^{n \times m}$ for some $m \leq n$ corresponds to an affine scaling of $\mathcal{V}$, and $Q_{i\cdot}$ denotes the i-th row of Q. Readers are refereed to [61, 111] for the details of the model.

After converting the nominal LP to our standard form, we can easily put the systemic risk problem into our framework by defining $\mathcal{U} := \{(b, c) : b \in \mathcal{V},\ c = 0\}$

and slightly changing $\overline{\mathcal{U}}$ to reflect the dependence on the affine transformation as represented by the matrix Q.

Similar to section 5.2 in [111], we randomly generate 25 instances of size $m \times n = 5 \times 20$. In accordance with Remark 3.2.1, which states that q^- is easy to calculate in this case, we focus our attention on the worst-case value q^+. It is straightforward to verify Assumptions 3.1 and 3.2. To evaluate the quality of q^+_{sdp}, we also calculate $\max\{r^+, v^+\}$ for each instance and the associated relative gap:

$$\text{gap}^+ = \frac{q^+_{\text{sdp}} - \max\{r^+, v^+\}}{\max\{|\max\{r^+, v^+\}|,\ 1\}} \times 100\%.$$

The computational results indicate that our approach recovers q^+ exactly for all 25 instances, which matches the quality of results from [111]. Furthermore, the average computation time for solving each instance is about 915 seconds. The computation times for computing r^+ are trivial, while the average computation time for computing each v^+ is about 287 seconds.

3.4.4 A network flow problem

Consider a transportation network flow problem from [142], which has $m_1 = 5$ suppliers/origins and $m_2 = 10$ customers/destinations for a total of $m = 15$ facilities. The network is bipartite and consists of $n = 24$ arcs connecting suppliers and customers; see Figure 4.4. Also shown in Figure 4.4 are the (estimated) supply and demand numbers ($\hat{b}$) for each supplier and customer. In addition, the (estimated) unit transportation costs ($\hat{c}$) associated with the arcs of the network are given in Table 3.2. Suppose at the early stages of planning, the supply and demand units and the

unit transportation costs are uncertain. Thus, the manager would like to quantify the resulting uncertainty in the optimal transportation cost.

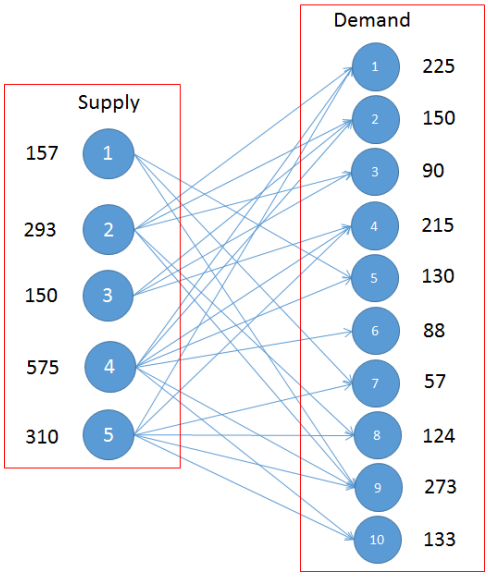


Figure 3.2: The transportation network of the 5 suppliers and 10 customers.

We consider three cases for the uncertainty set, each of which is also parameterized by a scalar $\gamma \in (0, 1)$. In the first case ("POLY"), we consider the polyhedral uncertainty set

$$\mathcal{U}_1(\gamma) = \{(b, c) : \ \|b\|_1 \leq \gamma\|\hat{b}\|_1, \ \|c\|_1 \leq \gamma\|\hat{c}\|_1 \ \};$$

in the second case ("SOC"), we consider the second-order-cone uncertainty set

$$\mathcal{U}_2(\gamma) := \{(b, c) : \ \|b\| \leq \gamma\|\hat{b}\|, \ \|c\| \leq \gamma\|\hat{c}\| \ \};$$

	Customer									
Supplier	1	2	3	4	5	6	7	8	9	10
1					2		3		2	
2	3	3	4					1	4	
3		4	5	3						
4	1	2		1	3	1			8	2
5	4			3			2	1	2	1

Table 3.2: The unit transportation costs associated with the arcs of the network.

and in the third case ("MIX"), we consider a mixture of the first two cases:

$$\mathcal{U}_3(\gamma) := \{(b,c) : \ \|b\|_1 \leq \gamma\|\hat{b}\|_1, \ \|c\| \leq \gamma\|\hat{c}\| \ \}.$$

For each, γ controls the perturbation magnitude in b and c relative to $\hat{b}$ and $\hat{c}$, respectively. In particular, we will consider three choices of γ: 0.01, 0.03, and 0.05. For example, $\gamma = 0.03$ roughly means that b can vary up to 3% of the magnitude of $\hat{b}$. In total, we have three cases with three choices for γ resulting in nine overall experiments.

Assumptions 3.1 and 3.2 are satisfied in this example, and so we apply our approach to bound q^- and q^+; see Table 3.3. Our 18 bounds (lower and upper bounds for each of the nine experiments) are listed in the two columns titled q^-_{sdp} and q^+_{sdp}, respectively. We also report the computation times (in seconds) for all 18 instances

under the two columns marked t^-_{sdp} and t^+_{sdp}. We also compute r^-, v^-, r^+, and v^+ and define the relative gaps

$$\text{gap}^- = \frac{\min\{r^-, v^-\} - q^-_{\text{sdp}}}{\max\{|\min\{r^-, v^-\}|,\ 1\}} \times 100\%,$$
$$\text{gap}^+ = \frac{q^+_{\text{sdp}} - \max\{r^+, v^+\}}{\max\{|\max\{r^+, v^+\}|,\ 1\}} \times 100\%.$$

Again, the computation times for r^- and r^+ are trivial. The average computation time for computing v^- and v^+ is about 216 seconds.

Table 3.3 shows that our relaxations capture q^- and q^+ in all cases. As ours is the first approach to study general perturbations in the literature, we are aware of no existing methods for this problem with which to compare our results.

Case	γ	q^-_{sdp}	gap^-	$\min\{r^-, v^-\}$	$\max\{r^+, v^+\}$	gap^+	q^+_{sdp}	$t^-_{\text{sdp}}(s)$	$t^+_{\text{sdp}}(s)$
	0.01	2638.4	0.0%	2638.4	3088.8	0.0%	3088.8	2592	2037
POLY	0.03	2139.6	0.0%	2139.6	3437.4	0.0%	3437.4	3421	2581
	0.05	1640.9	0.0%	1640.9	3769.8	0.0%	3769.8	3295	2713
	0.01	2745.6	0.0%	2745.6	2981.6	0.0%	2981.6	129	99
SOC	0.03	2498.9	0.2%	2504.3	3212.1	0.0%	3212.1	127	83
	0.05	2257.6	0.2%	2263.0	3442.7	0.0%	3442.7	113	85
	0.01	2724.1	0.0%	2724.1	3008.4	0.0%	3008.4	704	588
MIX	0.03	2429.2	0.0%	2429.2	3281.9	0.0%	3281.9	700	472
	0.05	2134.3	0.0%	2134.3	3560.7	0.0%	3560.7	716	493

Table 3.3: Results for the transportation network problem.

3.4.5 Additional details

Example	scalars	cones	matrices (size)	constraints
Sec. 3.4.1 (#1)	0	0	1 (23×23)	535
Sec. 3.4.1 (#2)	0	0	1 (23×23)	561
Sec. 3.4.1 (#3)	0	0	1 (23×23)	538
Sec. 3.4.2	0	0	1 (157×157)	14346
Sec. 3.4.3	486	81	1 (106×106)	13369
Sec. 3.4.4 (POLY)	0	0	1 (142×142)	21274
Sec. 3.4.4 (SOC)	2009	98	1 (103×103)	6300
Sec. 3.4.4 (MIX)	1975	79	1 (118×118)	11339

Table 3.4: Statistics on the sizes of problems solved in the examples as reported by Mosek.

Table 3.4 lists the statistics on the sizes of the conic programs solved in the above examples as reported by Mosek. The columns refer to the number of scalar variables (*scalars*), the number of second-order cones (*cones*), the number of positive semidefinite matrices along with their size (*matrices (size)*), and the number of linear constraints (*constraints*). Note that the number of linear constraints corresponds very closely to—indeed, is dominated by—the number of RLT constraints added, and the number of second-order cones equals the number of SOCRLT constraints. Note

also that, within Section 3.4.3, all SDP relaxations share the same size statistics. Similarly, the relaxations within the cases POLY, SOC, and MIX of Section 3.4.4 are all of the same size.

Finally, we investigate the effectiveness of the linearized complementarity constraints of (3.14) in the calculation of q^-_{sdp} and q^+_{sdp} by also solving the relaxations without the constraints. As it turns out, in all calculations of q^-_{sdp}, dropping those constraints does not change the relaxation value, i.e., those constraints are inactive in the optimal solution. However, in all calculations of q^+_{sdp} from Sections 3.4.1, 3.4.2, and 3.4.4, dropping the constraints has a significant negative effect as shown in Table 3.5. In the table, the gap is defined as

$$\text{gap} = \frac{(\text{value without constraint}) - q^+_{\text{sdp}}}{\max\{|q^+_{\text{sdp}}|,\ 1\}} \times 100\%.$$

We mention the effect of these constraints to highlight their practical importance—at least in certain situations such as ours. In the future, perhaps more theoretical light can be shed on these constraints as has already been done for the RLT and SOC-RLT constraints; see also the discussion in Section 3.3.2.

3.5 Conclusion

In this chapter, we have introduced the idea of robust sensitivity analysis for the optimal value of LP. In particular, we have discussed the best- and worst-case optimal values under general perturbations in the objective coefficients and right-hand sides. We have also presented finite variants that avoid cases of infeasibility and unboundedness. As the involved problems are nonconvex and very difficult to solve in

general, we have proposed copositive reformulations, which provide a theoretical basis for constructing tractable SDP-based relaxations that take into account the nature of the uncertainty set, e.g., through RLT and SOC-RLT constraints. Numerical experiments have indicated that our approach works very well on examples from, and inspired by, the literature. In future research, it would be interesting to improve the solution speed of the largest relaxations and to explore the possibility of also handling perturbations in the constraint matrix.

Example	q^+_{sdp}	gap	value without constraint
Sec. 3.4.1 (#1)	-16000	100%	0
Sec. 3.4.1 (#2)	-18667	100%	0
Sec. 3.4.1 (#3)	-16000	100%	0
Sec. 3.4.2	25600	1671%	453298
Sec. 3.4.4 (POLY 0.01)	3088.8	5.7%	3265.8
Sec. 3.4.4 (POLY 0.03)	3437.4	2.7%	3528.5
Sec. 3.4.4 (POLY 0.05)	3769.8	2.3%	3855.6
Sec. 3.4.4 (SOC 0.01)	2981.6	87.6%	5593.1
Sec. 3.4.4 (SOC 0.03)	3212.1	84.3%	5920.2
Sec. 3.4.4 (SOC 0.05)	3442.7	81.6%	6252.7
Sec. 3.4.4 (MIX 0.01)	3008.4	10.3%	3319.4
Sec. 3.4.4 (MIX 0.03)	3281.9	5.2%	3453.5
Sec. 3.4.4 (MIX 0.05)	3560.7	3.6%	3689.4

Table 3.5: Effectiveness of the linearized complementarity constraint.

CHAPTER 4
A DATA-DRIVEN DISTRIBUTIONALLY ROBUST BOUND ON THE EXPECTED OPTIMAL VALUE OF UNCERTAIN MIXED 0-1 LINEAR PROGRAMMING

4.1 Introduction

In this chapter, we consider the following uncertain mixed 0-1 linear programming problem:

$$v(\xi) := \max \left\{ (F\xi)^T x : \begin{array}{l} Ax = b,\ x \geq 0 \\ \\ x_j \in \{0,1\}\ \forall\, j \in \mathcal{B} \end{array} \right\} \tag{4.1}$$

where $A \in \mathbb{R}^{m \times n}$, $F \in \mathbb{R}^{n \times k}$, and $b \in \mathbb{R}^m$ are problem data, $x \in \mathbb{R}^n_+$ is the vector of decision variables, and $\mathcal{B} \subseteq \{1, \ldots, n\}$ is an index set. The objective coefficients are linear in the random vector $\xi \in \mathbb{R}^k$ via F. Problem (4.1) entails two extreme classes of programs: if $\mathcal{B} = \emptyset$, then (4.1) represents the regular linear program with uncertain objective coefficients; if $\mathcal{B} = \{1, \ldots, n\}$, then (4.1) represents the regular binary program with uncertain coefficients. In general, problem (4.1) is NP-hard [141].

The optimal value $v(\xi)$ is a random variable as ξ is a random vector in problem (4.1). We assume that ξ follows a multivariate distribution $\mathbb{P}$ supported on a nonempty set $\Xi \subseteq \mathbb{R}^k$, which is, in particular, defined as a slice of a closed, convex, full-dimensional cone $\widehat{\Xi} \subseteq \mathbb{R}_+ \times \mathbb{R}^{k-1}$:

$$\Xi := \left\{ \xi \in \widehat{\Xi} : e_1^T \xi = \xi_1 = 1 \right\},$$

where e_1 is the first standard basis vector in $\mathbb{R}^k$. In words, $\widehat{\Xi}$ is the homogenization

of Ξ. We choose this homogenized version for notational convenience. Note that it enables us to model affine effects of the uncertain parameters in (4.1).

The expected optimal value of (4.1), denoted by $v_{\mathbb{P}}$, is defined as

$$v_{\mathbb{P}} := \mathrm{E}_{\mathbb{P}}[v(\xi)] = \int_{\Xi} v(\xi)\, d\,\mathbb{P}(\xi).$$

The problem of computing $v_{\mathbb{P}}$ has been extensively studied in the literature. Hagstrom [79] showed that computing $v_{\mathbb{P}}$ for the longest path problem over a directed acyclic graph is $\#\mathcal{P}$-complete if the arc lengths are each independently distributed and restricted to taking two possible values. Aldous [1] studied a linear assignment problem with random cost coefficients following either an independent uniform distribution on $[0, 1]$ or an exponential distribution with parameter 1 and proved that the asymptotic value of $v_{\mathbb{P}}$ approaches $\frac{\pi^2}{6}$ as the number of assignments goes to infinity. For additional studies, see [33, 52, 102].

In practice, it is difficult or impossible to know $\mathbb{P}$ completely and computing $v_{\mathbb{P}}$ is thus not well defined in this situation. An alternative is to construct an ambiguity set, denoted by $\mathcal{D}$, that contains a family of distributions supported on Ξ consistent with any known properties of $\mathbb{P}$. Ideally, the ambiguity set will possess some statistical guarantee, e.g., the probability that $\mathbb{P} \in \mathcal{D}$ will be at least $1 - \beta$, where β is the significance level. In analogy with $v_{\mathbb{P}}$, we define $v_{\mathbb{Q}}$ for any $\mathbb{Q} \in \mathcal{D}$. Then, we are interested in computing the maximum expected optimal value $v_{\mathbb{Q}}$ over the ambiguity set $\mathcal{D}$:

$$v_{\mathcal{D}}^{+} := \sup_{\mathbb{Q} \in \mathcal{D}} v_{\mathbb{Q}}. \tag{4.2}$$

Note that, when the probability of $\mathbb{P} \in \mathcal{D}$ is at least $1-\beta$, the probability of $v_{\mathbb{P}} \leq v_{\mathcal{D}}^{+}$ is at least $1-\beta$.

There are three main issues (somehow conflicting) regarding the computation of $v_{\mathcal{D}}^{+}$. First, one would like an ambiguity set $\mathcal{D}$ with a high statistical guarantee to contain the true distribution $\mathbb{P}$. In this way, the computed $v_{\mathcal{D}}^{+}$ will be an upper bound on $v_{\mathbb{P}}$ with a high confidence level. (We will introduce several approaches in the following paragraph.) Second, one would like $v_{\mathcal{D}}^{+}$ to be tight in the sense that it is as close to $v_{\mathbb{P}}$ as possible. Generally, if $\mathcal{D}$ captures more information about $\mathbb{P}$, then $v_{\mathcal{D}}^{+}$ will be closer to $v_{\mathbb{P}}$. Finally, the third concern is the complexity of the resulting optimization problem, i.e., whether the problem can be solved in polynomial time.

Bertsimas et al. [25, 26] constructed ambiguity sets using the first two marginal moments of each ξ_i. Denote the first and second of each uncertain parameter by μ_i and σ_i respectively. They computed $v_{\mathcal{D}(\mu,\,\sigma)}^{+}$ over all joint distributions sharing the same first two marginal moments and proved polynomial time computability if the corresponding deterministic problem is solvable in polynomial time. However, the computed bound may not be tight with respect to $v_{\mathbb{P}}$ since the marginal-moment model does not capture the dependence of the random variables. In a closely related direction, Natarajan et al. [104] proposed an ambiguity set that was constructed from the known marginal distributions of each random variable ξ_i, and they computed $v_{\mathcal{D}(\mu,\,\sigma)}^{+}$ by solving a concave maximization problem. As an extension to the marginal moment-based approach, Natarajan et al. [107] proposed a cross-moment model that was based on an ambiguity set constructed using both marginal and cross

moments. Compared to the marginal-moment approach, the quality of the bound is tighter as the cross-moment model captures the dependence of the random variables. However, computing the bound requires solving a completely positive program, which itself can only be approximated in general. The authors then proposed semidefinite programming (SDP) relaxations to approximate $v^{+}_{\mathcal{D}(\mu,\,\Sigma)}$ over the cross-moment-based ambiguity set, where μ and Σ are the first and second moments respectively.

Moment-based ambiguity sets are also used prominently in a parallel vein of research, called distributionally robust optimization (DRO); see [27, 42, 45, 50, 55, 62, 73, 82, 106, 133, 147, 148]. The popularity of the moment-based approach is mainly due to the fact that it often leads to tractable optimization problems and relatively simple models. Its weakness, however, is that moment-based sets are not guaranteed to converge to the true distribution $\mathbb{P}$ when the sample size increases to infinity, even though the estimations of the first and second moments are themselves guaranteed to converge.

As an attractive alternative to moment-based ambiguity sets, distance-based ambiguity sets haven been proposed in recent years. This approach defines $\mathcal{D}$ as a ball in the space of probability distributions equipped with a distance measure, and the center of the ball is typically the empirical distribution derived from a series of independent realizations of the random vector ξ. The key ingredient of this approach is the distance function. Classical distance functions include the Kullback-Leibler divergence [86, 87], the ϕ-divergence [10, 54, 90], the Prohorov metric [62], empirical Burg-entropy divergence balls [94], and the Wasserstein metric [112, 140].

In this chapter, we apply the Wasserstein metric to construct a data-driven ambiguity set $\mathcal{D}$ centered at the empirical distribution $\widehat{\mathbb{P}}_N$ derived from N independent observations of ξ. This approach has several benefits. The conservativeness of the ambiguity set can be controlled by tuning a single parameter, the radius of the Wasserstein ball; we will discuss this parameter in detail in Section 4.2. Also, the Wasserstein ambiguity provides a natural confidence set for $\mathbb{P}$. Specifically, the Wasserstein ball around the empirical distribution on N independent identical samples contains $\mathbb{P}$ with confidence $1-\beta$ if its radius exceeds an explicit threshold $\epsilon_N(\beta)$ that can be computed via a closed form equation [59, 63]. We then formulate $v_{\mathcal{D}}^+$ in (4.2) over the constructed Wasserstein ambiguity set. That is, we model the maximum value of $v_{\mathbb{Q}}$ over the ambiguity set $\mathcal{D}$ constructed by the Wasserstein metric. In Section 4.3, we reformulate problem (4.2) into a copositive problem under some standard assumptions. As the copositive reformulation is computationally intractable, we apply a standard approach based on semidefinite programming techniques to approximate $v_{\mathcal{D}}^+$ from above. In Section 4.4, we numerically verify our approach on two applications from the literature. In particular, we compare our approach with the moment-based approach in [107], and we find that the bounds form our semidefinite-based programs approach the true expected optimal values as the number of observations increases, while the moment-based bounds remain constant. We conclude our research and discuss some future directions in Section 4.5.

We point out some similarities of our study to a recent technical report by Hanasusanto and Kuhn [81]. In their report, they proposed a Wasserstein-metric

ambiguity set for a two-stage DRO problem. In particular, they applied copositive programming techniques to reformulate the second-stage worst-case value function, which is essentially a max-min optimization problem, while we use copositive techniques to reformulate a max-max optimization problem; see (4.3). Furthermore, they directly used a hierarchy schema to approximate the copositive cones, while we derive natural SDP approximations based on the copositive reformulation. Note that their hierarchy approximations lead to SDP approximations as well. Finally, they developed an approach to derive an empirical Wasserstein radius, which is in spirit similar to our approach in this chapter.

4.1.1 Notation, terminology, and basic techniques

We denote by $\mathbb{R}^n$ the n-dimensional Euclidean space, and denote by $\mathbb{R}^n_+$ the nonnegative orthant in $\mathbb{R}^n$. For a scalar $p \geq 1$, the p-norm of $z \in \mathbb{R}^n$ is defined $\|z\|_p := (\sum_{i=1}^n |z_i|^p)^{1/p}$, e.g., $\|z\|_1 = \sum_{i=1}^n |z_i|$. We will drop the subscript for the 2-norm, i.e., $\|z\| := \|z\|_2$. For $v, w \in \mathbb{R}^n$, the inner product of v and w is denoted by $v^T w := \sum_{i=1}^n v_i w_i$. We denote by δ_ξ the Dirac distribution concentrating unit mass at $\xi \in \mathbb{R}^k$. For any $N \in \mathbb{N}$, we define $[N] := \{1, \ldots, N\}$.

The space $\mathbb{R}^{m\times n}$ denotes the set of real $m \times n$ matrices, and the trace inner product of two matrices $A, B \in \mathbb{R}^{m\times n}$ is $A \bullet B := \text{trace}(A^T B)$. $\mathcal{S}^n$ denotes the space of $n \times n$ symmetric matrices, and for $X \in \mathcal{S}^n$, $X \succeq 0$ means that X is positive semidefinite. In addition, $\text{diag}(X)$ denotes the vector containing the diagonal entries of X, and $\text{Diag}(v)$ is the diagonal matrix with vector v along its diagonal. We denote

by $I \in \mathcal{S}^n$ as the identity matrix.

For $\mathcal{K} \subseteq \mathbb{R}^n$ a closed, convex cone, $\mathcal{K}^*$ denotes its dual cone. We next introduce some basics of *copositive programming* with respect to the cone $\mathcal{K} \subseteq \mathbb{R}^n$. The *copositive cone* is defined as

$$\mathcal{COP}(\mathcal{K}) := \{M \in \mathcal{S}^n : x^T M x \geq 0 \ \forall \ x \in \mathcal{K}\},$$

and its dual cone, the *completely positive cone*, is

$$\mathcal{CPP}(\mathcal{K}) := \{X \in \mathcal{S}^n : X = \textstyle\sum_i x^i (x^i)^T, \ x^i \in \mathcal{K}\},$$

where the summation over i is finite but its cardinality is unspecified. The term *copositive programming* refers to linear optimization over $\mathcal{COP}(\mathcal{K})$ or, via duality, linear optimization over $\mathcal{CPP}(\mathcal{K})$. In fact, these problems are sometimes called *generalized copositive programming* or *set-semidefinite optimization* [40, 59] in contrast with the standard case $\mathcal{K} = \mathbb{R}^n_+$. In this chapter, we work with generalized copositive programming, although we use the shorter phrase for convenience.

For the specific dimensions k and n of the problem in this chapter, we let e_i denote the i-th standard basis vector in $\mathbb{R}^k$, and similarly, f_j denotes the j-th standard basis vector in $\mathbb{R}^n$. We will also use $g_1 := \binom{e_1}{0} \in \mathbb{R}^{k+n}$.

4.2 A Wasserstein-Based Ambiguity Set

In this section, we define the Wasserstein metric and discuss a standard method to construct a Wasserstein-based ambiguity set. Using this ambiguity set, we fully specify problem (4.2).

Denote by $\widehat{\Theta}_N := \{\hat{\xi}^1, \ldots, \hat{\xi}^N\}$ the data set that contains N independent samples of ξ governed by $\mathbb{P}$. The uniform empirical distribution based on $\widehat{\Theta}_N$ is $\widehat{\mathbb{P}}_N := \frac{1}{N}\sum_{i=1}^N \delta_{\hat{\xi}^i}$ where δ_ζ is the Dirac distribution concentrating unit mass at $\zeta \in \mathbb{R}^k$.

Definition 4.1 (Definition 3 in [81])**.** Let $\mathrm{IA}^2(\Xi)$ be the set of all probability distributions $\mathbb{Q}$ that are supported on Ξ and that satisfy $\mathrm{E}_{\mathbb{Q}}[\|\xi - \xi'\|^2] = \int_\Xi \|\xi - \xi'\|^2 d\,\mathbb{Q}(\xi) < \infty$ where $\xi' \in \Xi$ is some reference point, e.g., $\xi' = \hat{\xi}^i$ for some $i \in [N]$.

Definition 4.2 (Definition 3 in [81])**.** The 2-Wasserstein distance between any $\mathbb{Q}, \mathbb{Q}' \in \mathcal{M}^2(\Xi)$ is

$$W^2(\mathbb{Q}, \mathbb{Q}') := \inf \left\{ \left(\int_{\Xi^2} \|\xi - \xi'\|^2 \, \Pi(d\xi, d\xi') \right)^{1/2} : \begin{array}{l} \Pi \text{ is a joint distribution of } \xi \text{ and } \xi' \\ \text{with marginals } \mathbb{Q} \text{ and } \mathbb{Q}', \text{ respectively} \end{array} \right\}.$$

With this setting, our ambiguity set contains a family of distributions that are close to $\widehat{\mathbb{P}}_N$ with respect to the Wasserstein metric. In particular, we define our ambiguity set $\mathcal{D}$ as a 2-Wasserstein ball of radius ϵ that is centered at the uniform empirical distribution $\widehat{\mathbb{P}}_N$:

$$\mathcal{D}(\widehat{\mathbb{P}}_N, \epsilon) := \left\{ \mathbb{Q} \in \mathrm{IA}^2(\Xi) : W^2(\mathbb{Q}, \widehat{\mathbb{P}}_N) \leq \epsilon \right\}.$$

Note that $\widehat{\mathbb{P}}_N$ in $\mathcal{D}(\widehat{\mathbb{P}}_N, \epsilon)$ is defined on the N independent samples in the dataset $\widehat{\Theta}_N$. The reader is referred to [81] for the general case of $\mathrm{IA}^r(\Xi)$ and $W^r(\mathbb{Q}, \mathbb{Q}')$ for any $r \geq 1$. We use the 2-Wasserstein distance in this study for two reasons. First, the Euclidean distance is one of the most popular distances considered in the relevant literature; see [63, 81]. Second, we will find that problem (4.2) with an ambiguity set

based on the 2-Wasserstein distance can be reformulated into a copositive program; see Section 4.3.

Then, the problem of computing $v^+_{\mathcal{D}}$ based on the Wasserstein-based ambiguity set can be stated as

$$\begin{array}{rll} v^+_{\mathcal{D}(\widehat{\mathbb{P}}_N,\ \epsilon)} = & \sup\limits_{\Pi,\ \mathbb{Q}\in \mathrm{IA}^2(\Xi)} & \int_\Xi v(\xi)\ d\,\mathbb{Q}(\xi) \\ & \text{s. t.} & \int_{\Xi^2} \|\xi - \xi'\|^2\ \Pi(d\xi, d\xi') \le \epsilon^2 \\ & & \Pi \text{ is a joint distribution of } \xi \text{ and } \xi' \\ & & \text{with marginals } \mathbb{Q} \text{ and } \widehat{\mathbb{P}}_N, \text{ respectively.} \end{array} \tag{4.3}$$

The Wasserstein ball radius in problem (4.3) controls the conservatism of the optimal value. A larger radius is more likely to contain the true distribution and thus a more likely valid upper bound on $v_{\mathbb{P}}$, but even if it is valid, it could be a weaker upper bound. Therefore, it is crucial to choose an appropriate radius for the Wasserstein ball.

4.2.1 An empirical Wasserstein radius

The papers [59, 63] present a theoretical radius $\epsilon_N(\beta)$ for datasets of size N, which guarantees a desired confidence level $1-\beta$ for $\mathbb{P} \in \mathcal{D}(\mathbb{P}_N, \epsilon_N(\beta))$. Note that $\epsilon_N(\beta)$ depends on N and β. However, $\epsilon_N(\beta)$ is known to be conservative in practice; see [59] for example. In other words, $\mathcal{D}(\mathbb{P}_N, \epsilon_N(\beta))$ might contain significantly more irrelevant distributions so that the computed $v^+_{\mathcal{D}(\mathbb{P}_N,\ \epsilon_N(\beta))}$ is significantly larger than $v_{\mathbb{P}}$ and provides relatively less useful information. So, we propose a procedure to derive an empirical radius that provides a desired confidence level $1-\beta$ but is much smaller than $\epsilon_N(\beta)$. Our approach is based on the data set $\widehat{\Theta}_N$. In particular, we

apply a cross-validation procedure (see details in the next paragraphs) to compute an empirical confidence level (between 0 and 1) for a given radius ϵ. Our procedure guarantees that a larger radius leads to a higher confidence level. Therefore, by iteratively testing different ϵ, we can find a radius with a desired confidence level based on the data set $\widehat{\Theta}_N$. Although the derived $\epsilon(\widehat{\Theta}_N,\ \beta)$ depends on the data set $\widehat{\Theta}_N$, our experimental results indicate that it can be used for other datasets of the same sample size. We will show the numerical evidence in Section 4.4. Our approach is also similar in spirit to the one used in [63, 81].

Our procedure requires an oracle to compute (or approximate) $v^+_{\mathcal{D}(\mathbb{P}_N,\,\epsilon)}$. Later in Section 4.3, we will propose a specific approximation; see (4.20). Assume also that, in addition to the dataset $\widehat{\Theta}_N$, we predetermine a set $\mathcal{E}$ containing a large, yet finite, number of candidate radii ϵ. We randomly divide $\widehat{\Theta}_N$ into training and validation datasets K times. We enforce the same dataset size denoted by N_T on each of the K training datasets.

Next, for each $\epsilon \in \mathcal{E}$, we derive an empirical probability based on the following procedure: (i) we use each of the K training datasets to approximate $v^+_{\mathcal{D}(\widehat{\mathbb{P}}_{N_T}\,\epsilon)}$ with a value called $v_{\mathrm{WB}}(\epsilon)$ by calling the oracle; (ii) we then use the corresponding K validation datasets to simulate the expected optimal values denoted by $v_{\mathrm{SB}}(\epsilon)$; and (iii) we finally compute $\widehat{\mathbb{P}}_N[v_{\mathrm{WB}}(\epsilon) \geq v_{\mathrm{SB}}(\epsilon)]$, which is the percentage of the K instances where $v_{\mathrm{WB}}(\epsilon) \geq v_{\mathrm{SB}}(\epsilon)$. Let us call this empirical probability the *reliability index*. Thus, the reliability index can roughly approximate the confidence level that the underlying distribution is contained by the Wasserstein-based ambiguity set with

the radius ϵ. Note that $\widehat{\mathbb{P}}_N[\,\cdot\,]$ is non-decreasing in ϵ and equal to 1 for some large ϵ_0. Therefore, the set containing all the reliability indices with respect to the set $\mathcal{E}$ is essentially an empirical cumulative distribution. Then, given a desired confidence level, we can choose a corresponding empirical radius $\epsilon \in \mathcal{E}$ with a reliability index that is close to the confidence level. We specify the above procedure in Algorithm 4.1.

Algorithm 4.1 Procedure to compute a reliability index for any $\epsilon \in \mathcal{E}$

Inputs: A dataset $\widehat{\Theta}_N = \{\hat{\xi}^1, \ldots, \hat{\xi}^N\}$ and a radius $\epsilon \in \mathcal{E}$

Outputs: The reliability index

for $k = 1, \ldots, K$ **do**

Use the k^{th} training dataset to compute $v^k_{\text{WB}}(\epsilon)$

Use the k^{th} validation dataset to simulate $v^k_{\text{SB}}(\epsilon)$

end for

Calculate the reliability index for ϵ as the percentage of the K instances where $v^k_{\text{WB}}(\epsilon) \geq v^k_{\text{SB}}(\epsilon)$

4.3 Problem Reformulation and Tractable Bound

In this section, we propose a copositive programming reformulation for problem (4.3) under some mild assumptions. As copositive programs are computationally intractable, we then propose semidefinite-based relaxations for the purposes of com-

putation.

Let us first define the feasible set for $x \in \mathbb{R}^n$ in (4.1) as follows:

$$\mathcal{X} := \left\{ x \in \mathbb{R}^n : \begin{array}{l} Ax = b,\ x \geq 0 \\ \\ x_j \in \{0,1\}\ \forall\, j \in \mathcal{B} \end{array} \right\}.$$

We now introduce the following standard assumptions:

Assumption 4.1. The set $\mathcal{X} \subseteq \mathbb{R}^n$ is nonempty and bounded.

Assumption 4.2. $x \in \{x : Ax = b,\ x \geq 0\} \implies 0 \leq x_j \leq 1\ \forall\, j \in \mathcal{B}$.

Assumption 4.2 can be easily enforced based on [35]. For example, if $\mathcal{B} = \emptyset$, then the assumption is redundant. If problem (4.1) is derived from the network flow problems, for instance the longest path problem on a directed acyclic graph, then Assumption 4.2 is implied from the network flow constraints. When $\mathcal{B}$ is a nonempty set and the assumption is not implied by the constraints, we can add constraints $x_j + s_j = 1,\ \forall\, j \in \mathcal{B}$.

Assumption 4.3. The support set $\Xi \subseteq \mathbb{R}^k$ is convex, closed, and computationally tractable.

For example, Ξ could be represented using a polynomial number of linear, second-order-cone, and semidefinite inequalities. The set Ξ possesses a polynomial-time separation oracle [77].

Assumption 4.4. Ξ is bounded.

By Assumption 4.1, we know that $v(\xi)$ is finite and attainable for any $\xi \in \Xi$. Under Assumption 4.3, the distributions of ξ in the ambiguity set $\mathcal{D}$ are guaranteed

to be continuous over Ξ. Note that under Assumptions 4.1 - 4.4, $v_{\mathcal{D}}^{+}$ is finite and attainable and thus we can replace sup with max in (4.3) under these conditions. Assumption 4.4 could be merged with Assumption 4.3, but it is stated separately to highlight its role in proving the exactness of the copositive programming reformulation below.

4.3.1 A copositive reformulation

We reformulate problem (4.3) via conic programming duality theory and probability theory. We introduce a useful result from the literature as follows.

Lemma 4.1 (Theorem 1 in [81]). *$v^{+}_{\mathcal{D}(\widehat{\mathbb{P}}_N,\epsilon)}$ equals the optimal value of*

$$\begin{array}{ll} \sup & \frac{1}{N}\sum_{i=1}^{N}\int_{\Xi} v(\xi)\; d\,\mathbb{Q}_i(\xi) \\ \text{s.t.} & \frac{1}{N}\sum_{i=1}^{N}\int_{\Xi} \|\xi-\hat{\xi}_i\|^2\; d\,\mathbb{Q}_i(\xi) \le \epsilon^2 \\ & \mathbb{Q}_i \in \mathcal{M}^2(\Xi) \quad \forall\, i \in [N], \end{array} \tag{4.4}$$

where $\mathbb{Q}_i$ represents the distribution of ξ conditional on $\xi' = \hat{\xi}^i$ for all $i \in [N]$.

Proof. As $\mathbb{Q}_i$ represents the distribution of ξ conditional on $\xi' = \hat{\xi}^i$, the joint probability Π in problem (4.3) can be decomposed as $\Pi = \frac{1}{N}\sum\limits_{i\in[N]} \mathbb{Q}_i$ by the law of total probability. Thus, the optimal value of (4.4) coincides with $v^{+}_{\mathcal{D}(\widehat{\mathbb{P}}_N,\epsilon)}$, which completes the proof. □

We next provide a copositive programming reformulation for problem (4.4). As the first step, we use a standard duality argument to write the dual of (4.4) (see

also [63]):

$$v^+_{\mathcal{D}(\widehat{\mathbb{P}}_N,\epsilon)} = \sup_{\mathbb{Q}_i \in \mathrm{M}^2(\Xi)} \inf_{\lambda \geq 0} \frac{1}{N} \sum_{i=1}^{N} \int_{\Xi} v(\xi)\, d\,\mathbb{Q}_i(\xi) + \lambda \left(\epsilon^2 - \frac{1}{N} \sum_{i=1}^{N} \int_{\Xi} \|\xi - \hat{\xi}_i\|^2 \, d\,\mathbb{Q}_i(\xi) \right) \tag{4.5}$$

$$\leq \inf_{\lambda \geq 0} \sup_{\mathbb{Q}_i \in \mathrm{M}^2(\Xi)} \lambda\, \epsilon^2 + \frac{1}{N} \sum_{i=1}^{N} \int_{\Xi} (v(\xi) - \lambda \|\xi - \hat{\xi}_i\|^2)\, d\,\mathbb{Q}_i(\xi) \tag{4.6}$$

$$= \inf_{\lambda \geq 0} \lambda\, \epsilon^2 + \frac{1}{N} \sum_{i=1}^{N} \sup_{\xi \in \Xi} (v(\xi) - \lambda \|\xi - \hat{\xi}_i\|^2), \tag{4.7}$$

where (4.6) follows from the max-min inequality, while Equation (4.7) follows from the fact that $\mathcal{M}^2(\Xi)$ contains all the Dirac distributions supported on Ξ.

By Assumption 4.1, $v(\xi)$ is finite for all $\xi \in \Xi$. Then, the inequality in (4.6) becomes an equality for any $\epsilon > 0$ due to a straightforward generalization of a strong duality result for moment problems in Proposition 3.4 in [121]; see also Theorem 1 in [81] and Lemma 7 in [83]. By introducing auxiliary variables s_i, the minimization problem in (4.7) is equivalent to

$$\begin{array}{rll} v^+_{\mathcal{D}(\widehat{\mathbb{P}}_N,\epsilon)} = & \inf\limits_{\lambda,\, s_i} & \lambda\, \epsilon^2 + \frac{1}{N} \sum_{i=1}^{N} s_i \\ & \text{s.t.} & \sup\limits_{\xi \in \Xi} (v(\xi) - \lambda \|\xi - \hat{\xi}_i\|^2) \leq s_i \quad \forall\, i \in [N] \\ & & \lambda \geq 0. \end{array} \tag{4.8}$$

For each $i \in [N]$, consider the following maximization problem corresponding to the left-hand side of the constraints in (4.8):

$$\begin{array}{rll} h^i(\lambda) := & \sup & (F\xi)^T x - \lambda(\xi^T \xi - 2\hat{\xi}_i^T \xi + \|\hat{\xi}_i\|^2) \\ & \text{s.t.} & Ax = b,\ x \geq 0 \\ & & x_j \in \{0,1\} \quad \forall\, j \in \mathcal{B} \\ & & e_1^T \xi = 1,\ \xi \in \hat{\Xi}, \end{array} \tag{4.9}$$

which is a mixed 0-1 bilinear program. Under Assumption 4.2, it holds also that the optimal value of (4.9) equals the optimal value of an associated copositive program [35, 37], which we now describe.

Define

$$z := \begin{pmatrix} \xi \\ x \end{pmatrix} \in \mathbb{R}^{k+n}, \quad E := \begin{pmatrix} -be_1^T & A \end{pmatrix} \in \mathbb{R}^{m\times(k+n)}, \tag{4.10}$$

$$H^i(\lambda) := \begin{pmatrix} -\lambda(I - \hat{\xi}_i e_1^T - e_1\hat{\xi}_i^T + \|\hat{\xi}_i\|^2 e_1 e_1^T) & \frac{1}{2}F^T \\ \frac{1}{2}F & 0 \end{pmatrix} \in \mathcal{S}^{k+n}, \tag{4.11}$$

and for any $j \in \mathcal{B}$, define

$$Q_j := \begin{pmatrix} 0 \\ f_j \end{pmatrix}\begin{pmatrix} 0 \\ f_j \end{pmatrix}^T + \frac{1}{2}\begin{pmatrix} 0 \\ f_j \end{pmatrix}\begin{pmatrix} 0 \\ f_1 \end{pmatrix}^T + \frac{1}{2}\begin{pmatrix} 0 \\ f_1 \end{pmatrix}\begin{pmatrix} 0 \\ f_j \end{pmatrix}^T \in \mathcal{S}^{k+n}. \tag{4.12}$$

where f_j denotes the j-th standard basis vector in $\mathbb{R}^n$.

Because both $\mathcal{X}$ and Ξ are bounded by Assumptions 4.1 and 4.4, there exists a scalar $r > 0$ such that the constraint $z^T z = \xi^T\xi + x^T x \leq r$ is redundant for (4.9). Furthermore, it is well-known that we can use the following quadratic constraints to represent the binary variables in the description of $\mathcal{X}$:

$$x_j^2 - x_j = 0 \quad \Leftrightarrow \quad Q_j \bullet zz^T = 0.$$

After adding the redundant constraint and representing the binary variables, we ho-

mogenize problem (4.9) as follows:

$$\begin{aligned} \max \quad & H^i(\lambda) \bullet zz^T \\ \text{s.t.} \quad & Ez = 0,\ g_1^T z = 1 \\ & I \bullet zz^T \leq r \\ & Q_j \bullet zz^T = 0 \quad \forall\, j \in \mathcal{B} \\ & z \in \widehat{\Xi} \times \mathbb{R}^n_+, \end{aligned} \tag{4.13}$$

where $g_1 = \binom{e_1}{0} \in \mathbb{R}^{k+n}$ and e_1 denotes the standard basis vector in $\mathbb{R}^k$. The copositive representation is thus

$$\begin{aligned} \max \quad & H^i(\lambda) \bullet Z \\ \text{s.t.} \quad & \operatorname{diag}(EZE^T) = 0 \\ & g_1 g_1^T \bullet Z = 1 \\ & I \bullet Z \leq r \\ & Q_j \bullet Z = 0 \quad \forall\, j \in \mathcal{B} \\ & Z \in \mathcal{CPP}(\widehat{\Xi} \times \mathbb{R}^n_+). \end{aligned} \tag{4.14}$$

Letting $u^i \in \mathbb{R}^m$, $\rho^i \in \mathbb{R}_+$, $\alpha^i \in \mathbb{R}$, and $v^i \in \mathbb{R}^{|\mathcal{B}|}$ be the respective dual multipliers of $\operatorname{diag}(EZE^T) = 0$, $I \bullet Z \leq r$, $g_1 g_1^T \bullet Z = 1$, and $Q_j \bullet Z = 0$, standard conic duality theory implies the dual of (4.14) is

$$\begin{aligned} \min_{\alpha^i, \rho^i, u^i, v^i} \quad & \alpha^i + r\rho^i \\ \text{s.t.} \quad & \alpha^i g_1 g_1^T - H^i(\lambda) + E^T \operatorname{Diag}(u^i) E + \sum_{j \in \mathcal{B}} v^i_j Q_j + \rho^i I \in \mathcal{COP}(\widehat{\Xi} \times \mathbb{R}^n_+) \\ & \rho^i \geq 0. \end{aligned} \tag{4.15}$$

Holding all other dual variables fixed, for $\rho^i > 0$ large, the matrix variable in (4.15) is strictly copositive—in fact, positive definite—which establishes that Slater's condition

is satisfied, thus ensuring strong duality: the optimal value of (4.14) equals the optimal value of (4.15). Therefore, we can reformulate problem (4.8) as follows:

$$\begin{aligned} v^+_{\mathcal{D}(\widehat{\mathbb{P}}_N,\epsilon)} = \quad & \min \quad \lambda\epsilon^2 + \tfrac{1}{N}\sum_{i=1}^{N}(\alpha^i + r\rho^i) \\ & \text{s.t.} \quad \alpha^i g_1 g_1^T - H^i(\lambda) + E^T \operatorname{Diag}(u^i) E + \sum_{j\in\mathcal{B}} v^i_j Q_j + \rho^i I \in \mathcal{COP}(\widehat{\Xi}\times\mathbb{R}^n_+)\ \forall\, i \in [N] \\ & \qquad \rho^i \geq 0 \quad \forall\, i \in [N] \\ & \qquad \lambda \geq 0. \end{aligned} \tag{4.16}$$

Note that if Assumption 4.4 fails, the constraint $I \bullet Z \leq r$ should be excluded from (4.14) and thus the terms $r\rho^i$ and $\rho^i I$ in the objective function and the constraint respectively should be excluded in (4.15) as well. Therefore, strong duality between (4.14) and (4.15) cannot be established in this case. However, the modified (4.15) still provides an upper bound on $h^i(\lambda)$. Accordingly, the modified problem (4.16) still provides an upper bound on $v^+_{\mathcal{D}(\widehat{\mathbb{P}}_N,\epsilon)}$.

4.3.2 A semidefinite-based relaxation

As problem (4.16) is difficult to solve in general, we propose a tractable approximation based on semidefinite programming techniques. In particular, we propose an inner approximation of $\mathcal{COP}(\widehat{\Xi}\times\mathbb{R}^n_+)$ in (4.16) so that the resulting problem has an optimal value that is an upper bound on $v^+_{\mathcal{D}}$. Now, define

$$\operatorname{IA}(\widehat{\Xi}\times\mathbb{R}^n_+) := \left\{ S + M \ : \begin{array}{c} S_{11} \in \operatorname{IA}(\widehat{\Xi}), \operatorname{Rows}(S_{21}) \in \widehat{\Xi}^* \\ S_{22} \geq 0,\ M \succeq 0 \end{array} \right\},$$

where $\mathrm{IA}(\widehat{\Xi})$ is an inner approximation of $\mathcal{COP}(\widehat{\Xi})$, i.e., $\mathrm{IA}(\widehat{\Xi}) \subseteq \mathcal{COP}(\widehat{\Xi})$. Immediately, we have a relationship between $\mathrm{IA}(\widehat{\Xi} \times \mathbb{R}^n_+)$ and $\mathcal{COP}(\widehat{\Xi} \times \mathbb{R}^n_+)$:

Lemma 4.2. $\mathrm{IA}(\widehat{\Xi} \times \mathbb{R}^n_+) \subseteq \mathcal{COP}(\widehat{\Xi} \times \mathbb{R}^n_+)$.

Proof. Let arbitrary $\binom{p}{q} \in \widehat{\Xi} \times \mathbb{R}^n_+$ be given. We need to show

$$\begin{pmatrix} p \\ q \end{pmatrix}^T (S + M) \begin{pmatrix} p \\ q \end{pmatrix} = \tbinom{p}{q}^T S \tbinom{p}{q} + \tbinom{p}{q}^T M \tbinom{p}{q} \geq 0.$$

$$\begin{pmatrix} p \\ q \end{pmatrix}^T (S + M) \begin{pmatrix} p \\ q \end{pmatrix} = \tbinom{p}{q}^T S \tbinom{p}{q} + \tbinom{p}{q}^T M \tbinom{p}{q} \tag{4.17}$$

$$= p^T S_{11} p + 2q^T S_{21} p + q^T S_{22} q + \tbinom{p}{q}^T M \tbinom{p}{q} \tag{4.18}$$

$$\geq 0 \tag{4.19}$$

The first term is nonnegative because $p \in \widehat{\Xi}$ and $S_{11} \in \mathrm{IA}(\widehat{\Xi}) \subseteq \mathcal{COP}(\widehat{\Xi})$; the second term is nonnegative because $p \in \widehat{\Xi}, q \geq 0$, and $\mathrm{Rows}(S_{21}) \in \widehat{\Xi}^*$; the third term is nonnegative because $q \geq 0$ and $S_{22} \geq 0$; the last term is nonnegative because $M \succeq 0$. □

When $\widehat{\Xi} = \{\xi \in \mathbb{R}^k : P\xi \geq 0\}$ is a polyhedral cone based on some matrix $P \in \mathbb{R}^{p \times k}$, a typical inner approximation $\mathrm{IA}(\widehat{\Xi})$ of $\mathcal{COP}(\widehat{\Xi})$ is given by

$$\mathrm{IA}(\widehat{\Xi}) := \{S_{11} = P^T Y P : Y \geq 0\},$$

where $Y \in \mathcal{S}^p$ is a symmetric matrix variable. This corresponds to the RLT approach of [4, 38, 126]. When $\widehat{\Xi} = \{\xi \in \mathbb{R}^k : \|(\xi_2, \ldots, \xi_k)^T\| \leq \xi_1\}$ is the second-order cone, it is known [130] that

$$\mathcal{COP}(\widehat{\Xi}) = \{S_{11} = \tau J + M_{11} : \tau \geq 0,\ M_{11} \succeq 0\},$$

where $J = \text{Diag}(1, -1, \ldots, -1)$. Because of this simple structure, it often makes sense to take $\text{IA}(\widehat{\Xi}) = \mathcal{COP}(\widehat{\Xi})$ in practice.

Now consider the following problem by replacing $\mathcal{COP}(\widehat{\Xi}\times\mathbb{R}^n_+)$ with $\text{IA}(\widehat{\Xi}\times\mathbb{R}^n_+)$ in (4.16).

$$\begin{aligned}
\bar{v}^+_{\mathcal{D}(\widehat{\mathbb{P}},\epsilon)} = \quad & \min \quad \lambda\epsilon^2 + \tfrac{1}{N}\sum_{i=1}^{N}(\alpha^i + r\rho^i) \\
& \text{s.t.} \quad \alpha^i g_1 g_1^T - H^i(\lambda) + E^T \,\text{Diag}(u^i)E + \sum_{j\in\mathcal{B}} v^i_j Q_j + \rho^i I \in \text{IA}(\widehat{\Xi}\times\mathbb{R}^n_+)\ \forall\, i \in [N] \\
& \qquad \rho^i \geq 0 \quad \forall\, i \in [N] \\
& \qquad \lambda \geq 0.
\end{aligned} \tag{4.20}$$

Obviously, we have the following result:

Theorem 4.3. $v^+_{\mathcal{D}(\widehat{\mathbb{P}}_N,\epsilon)} \leq \bar{v}^+_{\mathcal{D}(\widehat{\mathbb{P}},\epsilon)}$.

4.4 Numerical Experiments

In this section, we validate our proposed Wasserstein-ball approach (WB) on two applications. As moment-based approaches are popular in the literature, we will compare WB with the moment-based approach (MB) proposed in [107] where the first two moments of the distributions are required. In practice, the moments of the distribution are often not known exactly. Delage and Ye [50] proposed a data-driven approach to handle this case. However, in this research, we assume that the moments are known exactly for MB. This setting actually favors MB. In particular, we compute these moments either from a closed-form formula or from a sufficiently large number of simulated samples. The goal of our experiments is to demonstrate that our approach can provide an upper bound, which gets closer to $v_{\mathbb{P}}$ as the size of

the data set increases, while the MB provides an upper bound, which is invariant to the size of the data set.

All computations are conducted with Mosek version 8.0.0.28 beta [5] on an Intel Core i3 2.93 GHz Windows computer with 4GB of RAM and are implemented using the modeling language YALMIP [99] in MATLAB (R2014a) version 8.3.0.532. In order to demonstrate the effectiveness of WB, we also implement a Monte Carlo simulation-based approach (SB) which requires a sufficiently large number of randomly generated samples. In the project management problem, we need to solve a linear program for each sample of the Monte Carlo simulation. We employ CPLEX 12.4 to solve these linear programs.

4.4.1 Statistical sensitivity analysis of highest-order statistic

The problem of finding the maximum value from a set $\zeta = (\zeta_1, \ldots, \zeta_n)$ of n numbers can be formulated as the optimization problem:

$$\max \left\{ \zeta^T x : e^T x = 1,\ x \geq 0 \right\}. \tag{4.21}$$

Suppose $\zeta_1 = \max\{\zeta_1, \ldots, \zeta_n\}$, then the optimal solution to (4.21) is $x_1^* = 1, x_2^* = \cdots = x_n^* = 0$. For the statistical sensitivity analysis problem, we consider a random vector ζ following a joint distribution $\mathbb{P}$. In the situation where the true distribution is not known exactly, our focus is to investigate the upper bound on the expected maximum value over an ambiguity set containing distributions that possess partial shared information.

We consider an instance with $n = 3$ and the true distribution $\mathbb{P}$ of ζ is assumed

to be jointly lognormal with first and second moments given by $\mu_{\log} \in \mathbb{R}^3$ and $\Sigma_{\log} \in \mathcal{S}^3$, respectively. In the experiments, we use the following procedure to randomly generate the first and second moments. We first sample $\mu \in \mathbb{R}^3$ from a uniform distribution $[0, 2]^3$. Then, we randomly generate a matrix $\Sigma \in \mathcal{S}^3$ as follows: we set the vector of standard deviations to $\sigma = \frac{1}{4}e \in \mathbb{R}^3$, sample a random correlation matrix $C \in \mathcal{S}^3$ using the MATLAB command 'gallery('randcorr',3)', and set $\Sigma = \mathrm{diag}(\sigma)C\,\mathrm{diag}(\sigma) + \mu\mu^T$. Then $\mu_{\log}$ and $\Sigma_{\log}$ can be computed based on the following formulae [80]:

$$\begin{aligned} (\mu_{\log})_i &= e^{\mu_i + 0.5\Sigma_{ii}}, \\ (\Sigma_{\log})_{ij} &= e^{\mu_i + \mu_j + 0.5(\Sigma_{ii} + \Sigma_{jj})}(e^{\Sigma_{ij}} - 1). \end{aligned} \tag{4.22}$$

We can cast this problem into our framework by setting $m = 1, k = n + 1$, $\xi = (1, \zeta_1, \ldots, \zeta_n)$, $F = (0, I)$, and $\mathcal{B} = \emptyset$. Obviously, Assumptions 4.1 and 4.3 are satisfied. Assumption 4.2 is vacuous. Although Assumption 4.4 does not hold, problem (4.20) can still provide a valid upper bound on the expected optimal value as discussed in Section 4.3.1.

4.4.1.1 The deviation of empirical Wasserstein radii

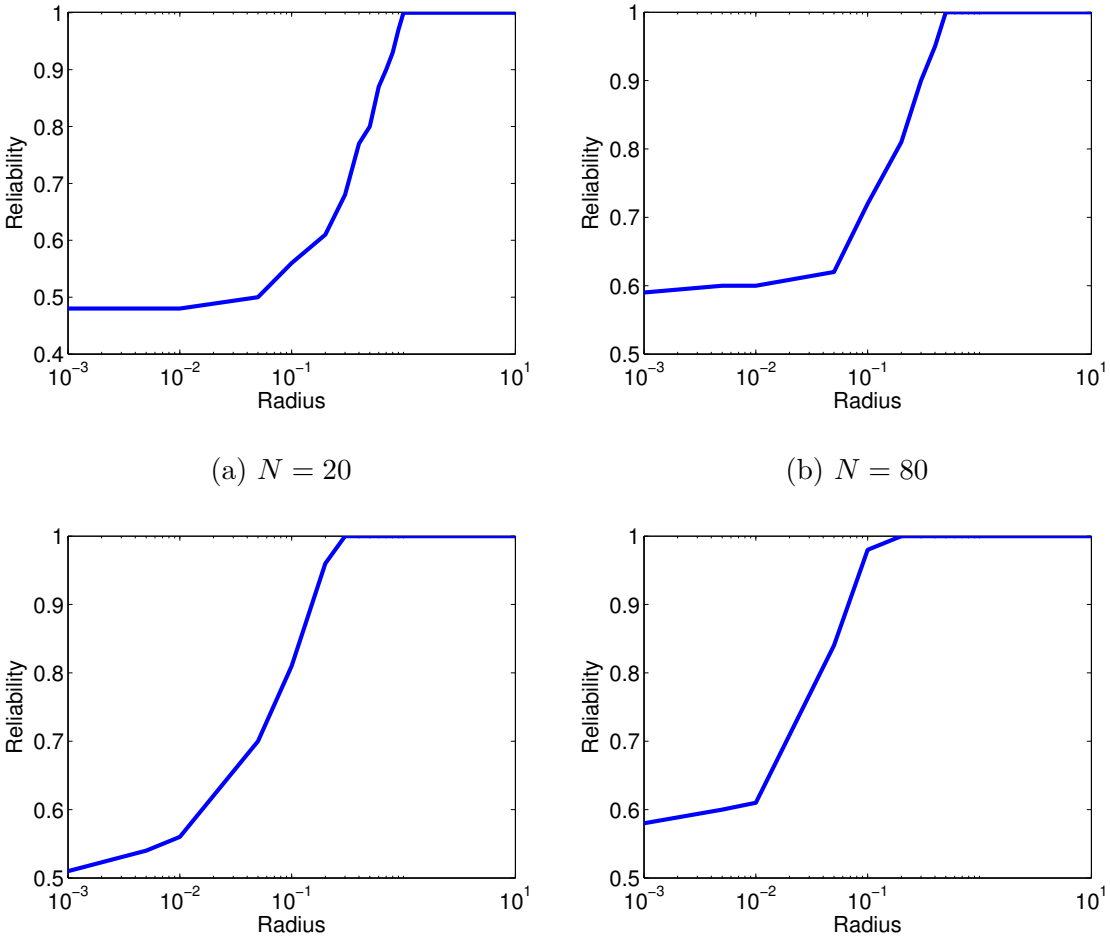
In this experiment, we consider a particular underlying distribution $\mathbb{P}$ that is generated by the procedure mentioned above. Also, we consider eight cases for the size of the dataset: $N \in \{10,\ 20,\ 40,\ 80,\ 160,\ 320,\ 640,\ 1280\}$. For each case, we randomly generate a dataset $\widehat{\Theta}_N$ containing N independent samples from $\mathbb{P}$ and use the procedure in Section 4.2 to determine a desired radius from a pre-specified

set $\mathcal{E} = \{0.001, 0.005, 0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0, 2.0\}$[1]. In particular, we set $K = 100$ in Algorithm 4.1. Figure 4.1 shows the trend of the reliabilities over different Wasserstein radii for $N \in \{20, 80, 320, 1280\}$. Clearly, smaller Wasserstein radii tend to have lower reliability indices. Furthermore, as the sample size increases, the reliability index increases as well for the same Wasserstein radius. The result of this experiment indicates that we can practically choose a Wasserstein radius with a desired statistical guarantee for each case of N.

4.4.1.2 Instances with the same underlying distribution

Our next experiment is to focus on a particular joint lognormal distribution $\mathbb{P}$ whose first and second moments are randomly generated based on the above procedure. We consider eight cases: $N \in \{10, 20, 40, 80, 160, 320, 640, 1280\}$. For each case, we test 100 trials and in each trial we randomly generate N independent samples from $\mathbb{P}$ and choose the Wasserstein radius with a reliability index of around 0.90, denoted by $\hat{\epsilon} := \epsilon(\widehat{\Theta}_N, 0.1)$. We compare our approach with MB with the first two moments of the distribution computed by (4.22). We also randomly generate 100000 independent samples from $\mathbb{P}$ to simulate the true expected optimal value. Figure 4.2 shows that our approach provides weaker bounds on the expected optimal value for smaller sample sizes. However, as the size of samples increases, our approach provides stronger bounds and the bounds get relatively close to the simulated value. In

[1]From preliminary experiments, the largest element 2.0 in set $\mathcal{E}$ returned 1 as the reliability index for all the experiments we conducted. Thus, it is sufficient to have 2.0 as the largest element here.



(a) $N = 20$

(b) $N = 80$

(c) $N = 320$

(d) $N = 1280$

Figure 4.1: Reliabilities of different Wasserstein radii for $N \in \{20, 80, 320, 1280\}$ respectively.

contrast, the value from MB remains the same regardless of the change of sample sizes.

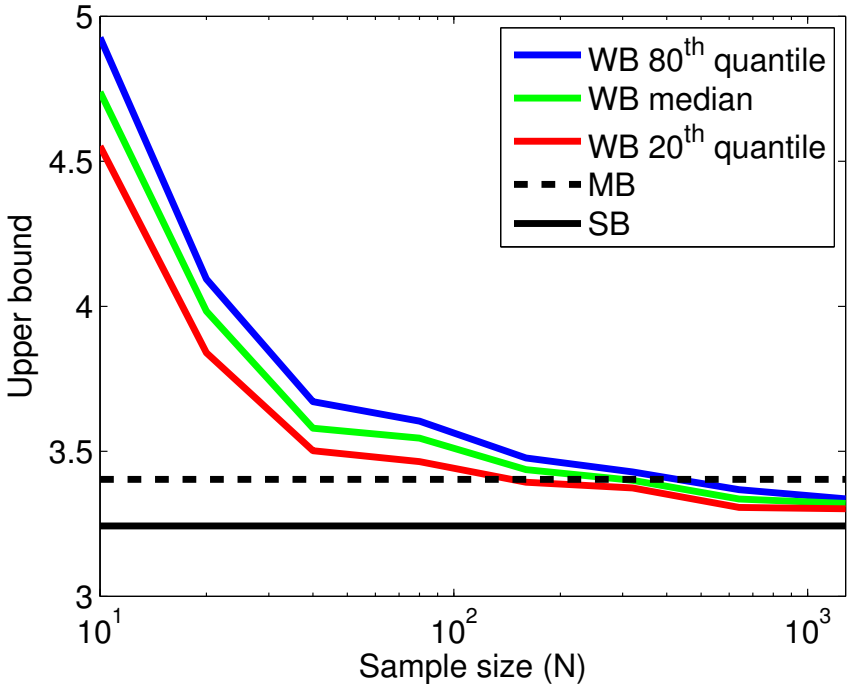


Figure 4.2: Illustration of the comparison of WB and MB over different sample sizes for a particular randomly generated underlying distribution. Note that the moment-based value and simulated value remains the same over all runs since this experiment is only for a particular distribution.

4.4.1.3 Instances with different underlying distributions

Again, in the third experiment, we consider eight cases $N \in \{10, 20, 40, 80, 160, 320, 640, 1280\}$. However, for each case, we randomly generate 100 trials and in each trial, we randomly generate N independent samples from a *randomly generated* joint lognormal distribution $\mathbb{P}$. For each trial in each case, we solve problem (4.20) with the Wasserstein radius corresponding to a reliability index of around 0.90 and simulate the expected optimal value over $100,000$ samples. Denote the optimal value from WB by $\bar{v}^+_{\mathcal{D}(\widehat{\mathbb{P}}_N,\hat{\epsilon})}$ and the simulated value by v_{SB}. Then, we calculate the relative gap between WB and SB as

$$\text{gap} := \frac{\bar{v}^+_{\mathcal{D}(\widehat{\mathbb{P}}_N,\hat{\epsilon})} - v_{\text{SB}}}{v_{\text{SB}}}.$$

For each case, we take the average of the relative gaps over the 100 trials. For each trial in each case, we also solve MB with the first two moments computed by (4.22). Denote the optimal value from MB by v^+_{MB}. Similarly, we calculate the relative gap between MB and SB as

$$\text{gap} := \frac{v^+_{\text{MB}} - v_{\text{SB}}}{v_{\text{SB}}}.$$

We then take the average of the relative gaps over the 100 trials in each case. Figure 4.3 illustrates the average relative gaps from both WB and MB over the eight cases. Clearly, the upper bound from WB approaches the simulated value along with the increase of the size of samples, while the average relative gap between the bound from MB and the simulated value does not.

Table 4.1 shows the percentage of the 100 trials where the optimal values from WB are greater than or equal to the corresponding simulated optimal values in the

eight cases. The result demonstrates that the derived empirical Wasserstein radii indeed provide desired statistical guarantees in practice.

Case number	1	2	3	4	5	6	7	8
Reliability index	1.00	1.00	1.00	1.00	0.97	0.98	0.97	0.98

Table 4.1: The percentage of the 100 trials where the optimal values from WB approach are greater or equal to the simulated values over the 8 cases.

4.4.2 Project management problem

In this example, we consider a project management problem, which can be formulated as a longest-path problem on a directed acyclic graph. The arcs denote activities and nodes denote completion of a set of activities. Arc lengths denote the time to complete the activities. Thus, the longest path from the starting node s to the ending node t gives the time needed to compete the whole project. Let ζ_{ij} be the length (time) of arc (activity) from node i to node j. The problem can be solved as a linear program due to the network flow structure as follows:

$$
\begin{aligned}
\max \quad & \sum_{(i,j)\in\mathcal{A}} \zeta_{ij} x_{ij} \\
\text{s. t.} \quad & \sum_{i:(i,j)\in\mathcal{A}} x_{ij} - \sum_{j:(i,j)\in\mathcal{A}} x_{ji} = \begin{cases} 1, & \text{if } i = s \\ 0, & \text{if } i \in \mathcal{N}, \text{ and } i \neq s,t \\ -1, & \text{if } i = t \end{cases} \\
& x_{ij} \geq 0, \ \forall\ (i,j) \in \mathcal{A},
\end{aligned}
\tag{4.23}
$$

where $\mathcal{A}$ denotes the set containing all the arcs and $\mathcal{N}$ denotes the set containing all nodes on the network. For the stochastic project management problem, the activity times are random. In such cases, due to the resource allocation and management constraints, the project manager would like to quantify the worst-case expected completion time of the project, which is corresponding to the worst-case longest path of the network.

We consider an instance with a network structure shown in Figure 4.4. This network consists of 7 arcs and 6 nodes. There are 3 paths from the starting node to the ending node on the network. In the experiments of this example, we consider truncated joint normal distributions. We use the following procedure to generate a truncated joint normal distribution $\mathbb{P}$: we generate $\zeta \geq 0$ from a jointly normal distribution with first and second moments given by $\mu \in \mathbb{R}^{|\mathcal{A}|}$ and $\Sigma \in \mathcal{S}^{|\mathcal{A}|}$, respectively. Specifically, we sample μ from a uniform distribution $[0,5]^{|\mathcal{A}|}$ while the matrix Σ is generated randomly using the following procedure: we set the vector of standard deviations to $\sigma = e$, sample a random correlation matrix $C \in \mathcal{S}^{|\mathcal{A}|}$ using the MATLAB command 'gallery('randcorr',$|\mathcal{A}|$)', and set $\Sigma = \operatorname{diag}(\sigma) C \operatorname{diag}(\sigma) + \mu\mu^T$. Skipping

the details, we can cast the network flow problem into our framework. It is straightforward to check that Assumptions 4.1, 4.3, and 4.4 are satisfied and Assumption 4.2 is vacuous.

4.4.2.1 Instances with the same underlying distribution

In the first experiment, we focus on a particular underlying distribution $\mathbb{P}$. We then consider seven cases where $N \in \{10, 20, 40, 80, 160, 320, 640\}$. For each case, we run 100 trials and in each trial we randomly generate a dataset $\widehat{\Theta}_N$ containing N independent samples $\{\hat{\zeta}^i\}_{i\in[N]}$ from $\mathbb{P}$. We use the procedure in Section 4.2 to compute a reliability set for each case. Then, we use computed reliability sets to derive empirical Wasserstein radii for the following computations. For each trial in each case, we solve problem (4.20) with a Wasserstein radius $\hat{\epsilon} = \hat{\epsilon}(\widehat{\Theta}_N, 0.1)$ corresponding to a reliability index of 0.90. We compare our WB to the literature MB required the first two moments. In this experiment, we approximate the moments based on a randomly generated dataset containing $100,000$ samples. The computed moments are close to their theoretical counterparts since the sample size is considerably large. We also simulate the expected optimal value over the $100,000$ samples. Figure 4.5 shows that WB provides weaker bounds on the expected optimal value for smaller sample sizes. However, as the size of samples increases, WB provides stronger bounds and the bounds get relatively close to the simulated value. In contrast, the bounds from MB remains the same regardless of the change of sample sizes.

4.4.2.2 Instances with different underlying distributions

The next experiment is also to consider the seven cases $N \in \{10, 20, 40, 80, 160, 320, 640\}$. Again, for each case, we randomly generate 100 trials in which N independent samples are drawn from a randomly generated truncated joint normal distribution using the above procedure. Then, for each trial in each case, we solve problem (4.20) with a Wasserstein radius $\hat{\epsilon}(\widehat{\Theta}_N, 0.1)$ corresponding to a 0.90 reliability index and we solve MB problem with moments approximated from a randomly generated dataset containing $100,000$ samples. We also simulate the expected optimal value over the $100,000$ samples for each trial in each case. Similarly, we compute the relative gap between the WB optimal value and the simulated optimal value and the relative gap between the MB optimal value and simulated optimal value. Then, for each case, we take the average of the relative gaps from both WB and MB over the 100 trials. Figure 4.3 illustrates the average relative gaps over the seven cases. Clearly, the upper bound from WB approaches to the simulated value along with the increase in the size of samples, while the gap between the bound from MB and the simulated value does not become narrow with the increase in the sample size. Table 4.2 shows the percentage of the 100 trials where the optimal values from WB are greater than or equal to the corresponding simulated optimal values over the seven cases.

Case number	1	2	3	4	5	6	7
Reliability index	1.00	1.00	1.00	1.00	0.97	0.98	0.97

Table 4.2: The percentage of the 100 trials where the optimal values from WB are greater than or equal to the corresponding simulated optimal values over the 7 cases.

4.5 Concluding Remarks

In this chapter, we have studied the expected optimal value of a mixed 0-1 programming problem with uncertain objective coefficients following a joint distribution whose information is not known exactly but a set of independent samples can be collected. Using the samples, we have constructed a Wasserstein-based ambiguity set that contains the true distribution with a desired confidence level. We proposed an approach to compute the upper bound on the expected optimal value. Then under mild assumption, the problem was reformulated to a copositive program, which leads to a semidefinite-based relaxation. We have validated the effectiveness of our approach over several applications.

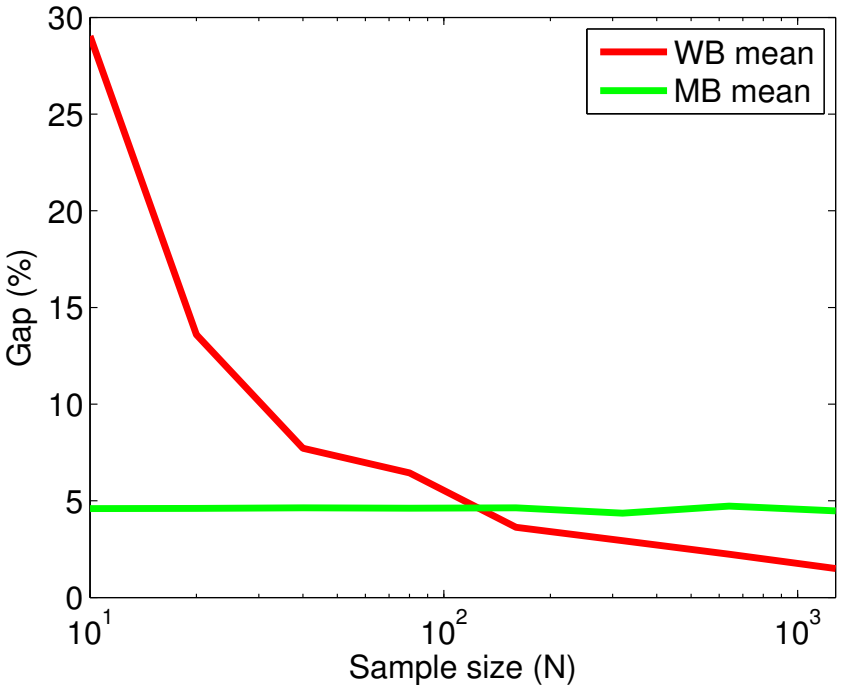


Figure 4.3: Illustration of the average relative gaps from both MB and WB in the case of $N \in \{10, 20, 40, 80, 160, 320, 640, 1280\}$. The blue line represents the average relative gap between the optimal values from WB and the simulated value; the red line represents the average relative gap between the optimal values from MB and the simulated value.

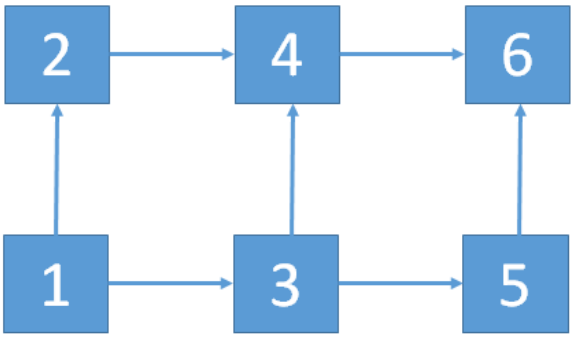


Figure 4.4: The structure of a project network.

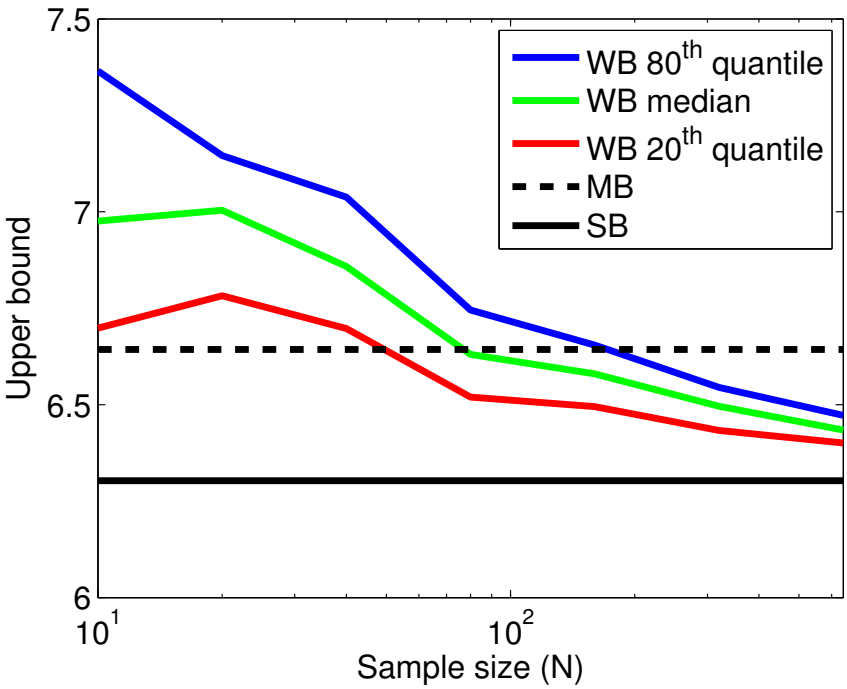


Figure 4.5: Illustration of the comparison of WB and MB over different sample sizes for a particular randomly generated underlying distribution. Note that the moment-based value and simulated value remains the same over all runs since this experiment is only for a particular distribution.

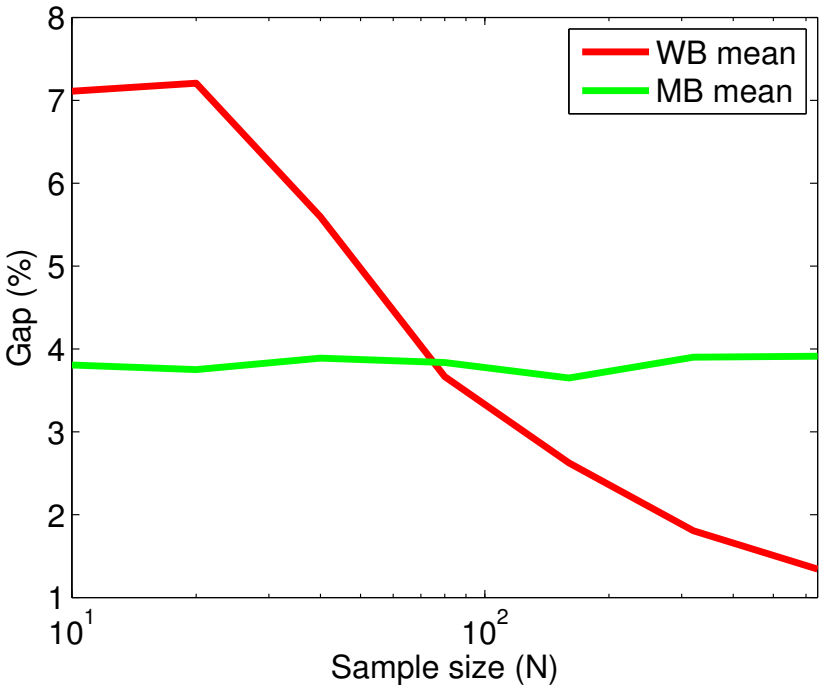


Figure 4.6: Illustration of the average gaps from both MB and WB in the case of $N \in \{10, 20, 40, 80, 160, 320, 640\}$. The blue line represents the average gap between the optimal values from WB and the simulated values; the red line represents the average gap between the optimal values from MB and the simulated values.

CHAPTER 5
CONCLUSIONS AND FUTURE RESEARCH

5.1 Conclusions

In this thesis, we apply techniques including mathematical programming, data analytics, and algorithmic computations to address several issues in optimization problems under uncertainty. In particular, we study a two-stage adjustable robust linear optimization problem with uncertain right-hand sides in which the first-stage decision is determined before the realizations of the uncertainties while the second-stage decision is determined after observing the realizations. We also study the effects of the uncertain parameters on the objective values of linear programming or mixed 0-1 linear programming from either distribution-free or distributionally robust perspectives. We highlight the main contributions of this thesis as follows.

In Chapter 2, we study a two-stage adjustable robust linear programming in which the right-hand sides are uncertain and belong to a convex, compact uncertainty set. The two-stage problem, in general, is computationally intractable. The affine policy is a popular, tractable approximation. Under some standard and simple conditions, we show that the two-stage problem can be reformulated as a copositive optimization problem, which in turn leads to a class of tractable, semidefinite-based approximations. We also show that the semidefinite-based approximation performs at least as well as the affine policy. We investigate several examples from the literature and the numerical results indicate that our tractable approximations significantly im-

prove the affine policy. In particular, our approach solves exactly in polynomial time a class of instances of increasing size for which the affine policy admits an arbitrarily large gap.

In Chapter 3, we leverage the concept of robust optimization to develop a framework for sensitivity analysis of linear programs (LPs) in minimization form. This framework allows for simultaneous perturbations in the objective coefficients and right-hand sides. We assume that the perturbations are modeled in a compact, convex, tractable uncertainty set. We investigate the best-case and worst-case LP optimal values over the parameter perturbations. We show that this framework can unify and extend multiple approaches for LP sensitivity analysis in the literature and has close ties to worst-case linear optimization and two-stage adaptive optimization. We reformulate the problems into copositive programs, which lead to semidefinite-based relaxations. We numerically show that the framework works effectively over several examples.

In Chapter 4, we investigate the distributionally robust bound on the expected optimal value of mixed 0-1 linear optimization problems with uncertain objective coefficients, in which the exact distribution of the cost coefficients is unknown and assumed to be in an ambiguity set. We construct the ambiguity set by using Wasserstein balls. We show that the problem can be reformulated as a copositive program under some standard and simple conditions. We then provide a tractable semidefinite programming relaxation to approximate the robust bound. We compare our approach with a well-known literature approach over several examples. The numerical results

indicate that our approach can be an effective alternative.

5.2 Future Research and Open Questions

Our result in Chapter 2 motivates us to extend the approach to two-stage robust optimization problems with binary recourse variables, which is known to be NP-hard in general. An interesting question is to see if there exists some conditions under which we can reformulate the two-stage binary recourse problem into a copositive program. A following question is if there exists efficient tight approximations to the problem.

The robust sensitivity analysis in Chapter 3 is conducted on linear programming problems, the semidefinite-based relaxations can provide tight bounds on the best- and worst-case optimal values. Inspired by this result, a natural question is if we can extend to approach to consider general perturbations in all problem data including the constraint matrix, cost coefficients, and right-hand sides. Another interesting question is if we can extend the approach to conduct robust sensitivity analysis in the optimal value of binary programming problems.

REFERENCES

[1] D. J. Aldous. The ζ (2) limit in the random assignment problem. *Random Structures & Algorithms*, 18(4):381–418, 2001.

[2] F. Alizadeh and D. Goldfarb. Second-order cone programming. *Mathematical programming*, 95(1):3–51, 2003.

[3] K. Anstreicher. Kronecker product constraints for semidefinite optimization. 2016.

[4] K. M. Anstreicher. Semidefinite programming versus the reformulation-linearization technique for nonconvex quadratically constrained quadratic programming. *Journal of Global Optimization*, 43(2-3):471–484, 2009.

[5] M. ApS. *The MOSEK optimization toolbox for MATLAB manual. Version 8.0.*, 2016.

[6] A. Ardestani-Jaafari and E. Delage. The value of flexibility in robust location-transportation problem. *Les Cahiers du GERAD G–2014–83, GERAD, HEC Montréal*, 2014.

[7] A. Ardestani-Jaafari and E. Delage. Linearized robust counterparts of two-stage robust optimization problem with applications in operations management. Manuscript, HEC Montreal, 2016.

[8] A. Atamtürk and M. Zhang. Two-stage robust network flow and design under demand uncertainty. *Operations Research*, 55(4):662–673, 2007.

[9] A. Ben-Tal, G. Boaz, and S. Shimrit. Robust multi-echelon multi-period inventory control. *European Journal of Operational Research*, 199(3):922–935, 2009.

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

[11] A. Ben-Tal, B. Do Chung, S. R. Mandala, and T. Yao. Robust optimization for emergency logistics planning: Risk mitigation in humanitarian relief supply chains. *Transportation research part B: methodological*, 45(8):1177–1189, 2011.

[12] A. Ben-Tal, L. E. Ghaoui, and A. Nemirovski. *Robust Optimization*. Princeton University Prss, Princeton, N.J.,, 2009.

[13] A. Ben-Tal, B. Golany, A. Nemirovski, and J.-P. Vial. Retailer-supplier flexible commitments contracts: a robust optimization approach. *Manufacturing & Service Operations Management*, 7(3):248–271, 2005.

[14] A. Ben-Tal, A. Goryashko, E. Guslitzer, and A. Nemirovski. Adjustable robust solutions of uncertain linear programs. *Mathematical Programming*, 99(2):351–376, 2004.

[15] A. Ben-Tal and A. Nemirovski. Robust convex optimization. *Mathematics of Operations Research*, 23(4):769–805, 1998.

[16] A. Ben-Tal and A. Nemirovski. Robust solutions of uncertain linear programs. *Operations Research Letters*, 25(1):1–14, 1999.

[17] A. Ben-Tal and A. Nemirovski. Robust solutions of linear programming problems contaminated with uncertain data. *Mathematical Programming Ser. A*, 88(3):411–424, 2000.

[18] A. Ben-Tal and A. Nemirovski. Robust optimization—methodology and applications. *Math. Program. Ser. A*, 92(3):453–480, 2002.

[19] D. Bertsimas, D. B. Brown, and C. Caramanis. Theory and applications of robust optimization. *SIAM Review*, 53(3):464–501, 2011.

[20] D. Bertsimas and F. J. de Ruiter. Duality in two-stage adaptive linear optimization: Faster computation and stronger bounds. *INFORMS Journal on Computing*, 28(3):500–511, 2016.

[21] D. Bertsimas and V. Goyal. On the power and limitations of affine policies in two-stage adaptive optimization. *Mathematical programming*, 134(2):491–531, 2012.

[22] D. Bertsimas, V. Goyal, and P. Y. Lu. A tight characterization of the performance of static solutions in two-stage adjustable robust linear optimization. *Mathematical Programming Ser. A*, 150(2):281–319, 2014.

[23] D. Bertsimas, D. A. Iancu, and P. A. Parrilo. A hierarchy of near-optimal policies for multistage adaptive optimization. *IEEE Transactions on Automatic Control*, 56(12):2809–2824, 2011.

[24] D. Bertsimas, E. Litvinov, X. A. Sun, J. Zhao, and T. Zheng. Adaptive robust optimization for the security constrained unit commitment problem. *IEEE transactions on power systems*, 28(1):52–63, 2013.

[25] D. Bertsimas, K. Natarajan, and C.-P. Teo. Probabilistic combinatorial optimization: Moments, semidefinite programming, and asymptotic bounds. *SIAM Journal on Optimization*, 15(1):185–209, 2004.

[26] D. Bertsimas, K. Natarajan, and C.-P. Teo. Persistence in discrete optimization under data uncertainty. *Mathematical programming*, 108(2):251–274, 2006.

[27] D. Bertsimas and I. Popescu. Optimal inequalities in probability theory: A convex optimization approach. *SIAM Journal on Optimization*, 15(3):780–804, 2005.

[28] D. Bertsimas and M. Sim. Robust discrete optimization and network flows. *Math. Program. Ser. B*, 98:49–71, 2003.

[29] D. Bertsimas and M. Sim. The price of robustness. *Operations Research*, 52(2):35–53, 2004.

[30] J. R. Birge and F. Louveaux. *Introduction to stochastic programming*. Springer-Verlag, New York, 1997.

[31] I. M. Bomze and E. De Klerk. Solving standard quadratic optimization problems via linear, semidefinite and copositive programming. *Journal of Global Optimization*, 24(2):163–185, 2002.

[32] I. M. Bomze, M. Dür, E. De Klerk, C. Roos, A. J. Quist, and T. Terlaky. On copositive programming and standard quadratic optimization problems. *Journal of Global Optimization*, 18(4):301–320, 2000.

[33] R. Bowman. Efficient estimation of arc criticalities in stochastic activity networks. *Management Science*, 41(1):58–67, 1995.

[34] S. Bradley, A. Hax, and T. Magnanti. Applied mathematical programming. 1977.

[35] S. Burer. On the copositive representation of binary and continuous nonconvex quadratic programs. *Mathematical Programming*, 120:479–495, 2009.

[36] S. Burer. Copositive programming. In M. Anjos and J. Lasserre, editors, *Handbook of Semidefinite, Cone and Polynomial Optimization: Theory, Algorithms, Software and Applications*, International Series in Operational Research and Management Science, pages 201–218. Springer, 2011.

[37] S. Burer. Copositive programming. In *Handbook on semidefinite, conic and polynomial optimization*, pages 201–218. Springer, 2012.

[38] S. Burer. A gentle, geometric introduction to copositive optimization. *Mathematical Programming*, 151(1):89–116, 2015.

[39] S. Burer and K. M. Anstreicher. Second-order-cone constraints for extended trust-region subproblems. *SIAM Journal on Optimization*, 23(1):432–451, 2013.

[40] S. Burer and H. Dong. Representing quadratically constrained quadratic programs as generalized copositive programs. *Operations Research Letters*, 40:203–206, 2012.

[41] G. C. Calafiore and L. El Ghaoui. On distributionally robust chance-constrained linear programs. *Journal of Optimization Theory and Applications*, 130(1):1–22, 2006.

[42] G. C. Calafiore and L. El Ghaoui. On distributionally robust chance-constrained linear programs. *Journal of Optimization Theory and Applications*, 130(1):1–22, 2006.

[43] J. Chen and S. Burer. Globally solving nonconvex quadratic programming problems via completely positive programming. *Mathematical Programming Computation*, 4(1):33–52, 2012.

[44] X. Chen and Y. Zhang. Uncertain linear programs: Extended affinely adjustable robust counterparts. *Operations Research*, 57(6):1469–1482, 2009.

[45] Z. Chen, M. Sim, and H. Xu. Distributionally robust optimization with infinitely constrained ambiguity sets. Working Paper, 2016.

[46] G. B. Dantzig. Linear programming under uncertainty. *Management Sci.*, 1:197–206, 1955.

[47] G. B. Dantzig. *Linear programming and extensions*. Princeton University Prss, Princeton, N.J.,, 1963.

[48] E. De Klerk and D. V. Pasechnik. Approximation of the stability number of a graph via copositive programming. *SIAM Journal on Optimization*, 12(4):875–892, 2002.

[49] E. Delage and D. A. Iancu. *Robust Multistage Decision Making*, chapter 2, pages 20–46. INFORMS, 2015.

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

[51] P. J. Dickinson, G. Eichfelder, and J. Povh. Erratum to: On the set-semidefinite representation of nonconvex quadratic programs over arbitrary feasible sets. *Optimization Letters*, 7(6):1387–1397, 2013.

[52] B. Dodin. Bounding the project completion time distribution in pert networks. *Operations Research*, 33(4):862–881, 1985.

[53] S. H. H. Doulabi, P. Jaillet, G. Pesant, and L.-M. Rousseau. Exploiting the structure of two-stage robust optimization models with integer adversarial variables. Manuscript, MIT, 2016.

[54] J. Duchi, P. Glynn, and H. Namkoong. Statistics of robust optimization: a generalized empirical likelihood approach. *arXiv preprint arXiv:1610.03425*, 2016.

[55] J. Dupačová. The minimax approach to stochastic programming and an illustrative application. *Stochastics: An International Journal of Probability and Stochastic Processes*, 20(1):73–88, 1987.

[56] J. Dupačová. Stochastic programming: minimax approach stochastic programming: Minimax approach. pages 3778–3782, 2008.

[57] M. Dür. Copositive Programming—A Survey. In M. Diehl, F. Glineur, E. Jarlebring, and W. Michiels, editors, *Recent Advances in Optimization and its Applications in Engineering*, pages 3–20. Springer, 2010.

[58] M. Dyer and L. Stougie. Computational complexity of stochastic programming problems. *Math. Program. Ser. A*, 116(3):423–432, 2006.

[59] G. Eichfelder and J. Jahn. Set-semidefinite optimization. *Journal of Convex Analysis*, 15:767–801, 2008.

[60] G. Eichfelder and J. Povh. On the set-semidefinite representation of non-convex quadratic programs over arbitrary feasible sets. *Optimization Letters*, 7(6):1373–1386, 2013.

[61] L. Eisenberg and T. H. Noe. Systemic risk in financial systems. *Management Science*, 47(2):236–249, 2001.

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

[63] P. M. Esfahani and D. Kuhn. Data-driven distributionally robust optimization using the wasserstein metric: Performance guarantees and tractable reformulations. *arXiv preprint arXiv:1505.05116*, 2015.

[64] J. E. Falk. Exact solutions of inexact linear programs. *Operations Research*, 24(4):783–787, 1976.

[65] U. Feige, K. Jain, M. Mahdian, and V. Mirrokni. Robust combinatorial optimization with exponential scenarios. In *Integer programming and combinatorial optimization*, volume 4513 of *Lecture Notes in Comput. Sci.*, pages 439–453. Springer, Berlin, 2007.

[66] R. J. Fonseca and B. Rustem. International portfolio management with affine policies. *European Journal of Operational Research*, 223(1):177–187, 2012.

[67] R. M. Freund. Postoptimal analysis of a linear program under simultaneous changes in matrix coefficients. *Mathematical Programming Study*, 24:1–13, 1985.

[68] V. Gabrel, M. Lacroix, C. Murat, and N. Remli. Robust location transportation problems under uncertain demands. *Discrete Applied Mathematics*, 164:100–111, 2014.

[69] V. Gabrel, C. Murat, and N. Remli. Linear programming with interval right hand sides. *International Transactions in Operational Research*, 17(3):397–408, 2010.

[70] T. Gal. Postoptimal analyses, parametric programming and related topics. 1995.

[71] T. Gal and J. Nedoma. Multiparametric linear programming. *Management Science*, 18(7):406–422, 1972.

[72] L. E. Ghaoui and H. Lebret. Robust solutions to least-squares problems with uncertain data. *SIAM Journal on Matrix analysis and applications*, 18:1035–1064, 1997.

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

[74] J. Goh and M. Sim. Distributionally robust optimization and its tractable approximations. *Operations research*, 58(4-part-1):902–917, 2010.

[75] D. Goldfarb and G. Iyengar. Robust convex quadratically constrained programs. *Mathematical Programming*, 97:495–515, 2003.

[76] B. L. Gorissen and D. Den Hertog. Robust counterparts of inequalities containing sums of maxima of linear functions. *European Journal of Operational Research*, 227(1):30–43, 2013.

[77] M. Grötschel, L. Lovász, and A. Schrijver. The ellipsoid method and its consequences in combinatorial optimization. *Combinatorica*, 1(2):169–197, 1981.

[78] M. J. Hadjiyiannis, P. J. Goulart, and D. Kuhn. A scenario approach for estimating the suboptimality of linear decision rules in two-stage robust optimization. In *50th IEEE conference on decision and control and European control conference (CDC-ECC)*, Orlando, FL, USA, December 12-15 2011.

[79] J. N. Hagstrom. Computational complexity of pert problems. *Networks*, 18(2):139–147, 1988.

[80] L. J. Halliwell. The lognormal random multivariate. In *Casualty Actuarial Society E-Forum, Spring 2015*.

[81] G. A. Hanasusanto and D. Kuhn. Conic programming reformulations of two-stage distributionally robust linear programs over wasserstein balls. *arXiv preprint arXiv:1609.07505*, 2016.

[82] G. A. Hanasusanto, D. Kuhn, S. W. Wallace, and S. Zymler. Distributionally robust multi-item newsvendor problems with multimodal demand distributions. *Mathematical Programming*, 152(1):1–32, 2015.

[83] G. A. Hanasusanto, V. Roitch, D. Kuhn, and W. Wiesemann. Ambiguous joint chance constraints under mean and dispersion information. *Operations Research*, 2017.

[84] D. Henrion, J.-B. Lasserre, and J. Löfberg. Gloptipoly 3: moments, optimization and semidefinite programming. *Optimization Methods & Software*, 24(4-5):761–779, 2009.

[85] M. Hladık. Interval linear programming: A survey. *Linear programming-new frontiers in theory and applications*, pages 85–120, 2012.

[86] Z. Hu and L. J. Hong. Kullback-leibler divergence constrained distributionally robust optimization. *Available at Optimization Online*, 2013.

[87] R. Jiang and Y. Guan. Data-driven chance constrained stochastic program. *Mathematical Programming*, pages 1–37, 2015.

[88] P. Kall and S. W. Wallace. *Stochastic programming*. John Wiley & Sons Ltd., Chichester, 1994.

[89] C. Kim. Parameterizing an activity vector in linear programming. *Operations Research*, 19:1632–1646, 1971.

[90] D. Klabjan, D. Simchi-Levi, and M. Song. Robust stochastic lot-sizing by means of histograms. *Production and Operations Management*, 22(3):691–710, 2013.

[91] D. Klatte. On the lower semicontinuity of the optimal sets in convex parametric optimization. *Mathematical Programming Study*, 10:104–109, 1979.

[92] H. Konno. A cutting plane algorithm for solving bilinear programs. *Mathematical Programming*, 11(1):14–27, 1976.

[93] C. M. Lagoa and B. R. Barmish. Distributionally robust monte carlo simulation: a tutorial survey. *IFAC Proceedings Volumes*, 35(1):151–162, 2002.

[94] H. Lam. Recovering best statistical guarantees via the empirical divergence-based distributionally robust optimization. *arXiv preprint arXiv:1605.09349*, 2016.

[95] J. B. Lasserre. Global optimization with polynomials and the problem of moments. *SIAM Journal on Optimization*, 11(3):796–817, 2001.

[96] J. B. Lasserre. New approximations for the cone of copositive matrices and its dual. *Mathematical Programming*, 144(1-2):265–276, 2014.

[97] M. Laurent and F. Vallentin. Semidefinite optimization. 2012.

[98] X. Li, K. Natarajan, C.-P. Teo, and Z. Zheng. Distributionally robust mixed integer linear programs: Persistency models with applications. *European Journal of Operational Research*, 233(3):459–473, 2014.

[99] J. Lofberg. Yalmip: A toolbox for modeling and optimization in matlab. In *Computer Aided Control Systems Design, 2004 IEEE International Symposium on*, pages 284–289. IEEE, 2004.

[100] Z.-Q. Luo and S. Zhang. On extensions of the frank-wolfe theorems. *Computational Optimization and Applications*, 13(1-3):87–110, 1999.

[101] P. G. McKeown and R. A. Minch. Notemultiplicative interval variation of objective function coefficients in linear programming. *Management Science*, 28(12):1462–1470, 1982.

[102] R. H. Möhring. Scheduling under uncertainty: Bounding the makespan distribution. In *Computational Discrete Mathematics*, pages 79–97. Springer, 2001.

[103] F. Mráz. Calculating the exact bounds of optimal valuesin lp with interval coefficients. *Annals of Operations Research*, 81:51–62, 1998.

[104] K. Natarajan, M. Song, and C.-P. Teo. Persistency model and its applications in choice modeling. *Management Science*, 55(3):453–469, 2009.

[105] K. Natarajan and C.-P. Teo. On reduced semidefinite programs for second order moment bounds with applications. *Mathematical Programming*, pages 1–32, 2016.

[106] K. Natarajan and C. P. Teo. On reduced semidefinite programs for second order moment bounds with applications. *Mathematical Programming*, 161(1):487–518, 2017.

[107] K. Natarajan, C. P. Teo, and Z. Zheng. Mixed 0-1 linear programs under objective uncertainty: A completely positive representation. *Operations research*, 59(3):713–728, 2011.

[108] W. Orchard-Hayes. *Advanced Linear-Programming Computing Techniques*. McGraw-Hill, New York, 1968.

[109] P. A. Parrilo. *Structured semidefinite programs and semialgebraic geometry methods in robustness and optimization*. PhD thesis, Citeseer, 2000.

[110] J. Pena, J. Vera, and L. F. Zuluaga. Computing the stability number of a graph via linear and semidefinite programming. *SIAM Journal on Optimization*, 18(1):87–105, 2007.

[111] J. Peng and T. Zhu. A nonlinear semidefinite optimization relaxation for the worst-case linear optimization under uncertainties. *Mathematical Programming*, 152(1-2):593–614, 2015.

[112] G. Pflug and D. Wozabal. Ambiguity in portfolio selection. *Quantitative Finance*, 7(4):435–442, 2007.

[113] M. Poss and C. Raack. Affine recourse for the robust network design problem: Between static and dynamic routing. *Networks*, 61(2):180–198, 2013.

[114] A. Pothen, H. D. Simon, and K.-P. Liou. Partitioning sparse matrices with eigenvectors of graphs. *SIAM journal on matrix analysis and applications*, 11(3):430–452, 1990.

[115] J. Povh and F. Rendl. Copositive and semidefinite relaxations of the quadratic assignment problem. *Discrete Optimization*, 6(3):231–241, 2009.

[116] N. Ravi and R. E. Wendell. The tolerance approach to sensitivity analysis of matrix coefficients in linear programming: General perturbations. *Journal of the Operational Research Society*, 36(10):943–950, 1985.

[117] N. Ravi and R. E. Wendell. The tolerance approach to sensitivity analysis of matrix coefficients in linear programming. *Management Science*, 35(9):1106–1119, 1989.

[118] J. Renegar. Some perturbation theory for linear programming. *Math. Programming*, 65(1, Ser. A):73–91, 1994.

[119] H. Scarf, K. Arrow, and S. Karlin. A min-max solution of an inventory problem. *Studies in the mathematical theory of inventory and production*, 10:201–209, 1958.

[120] A. Schrijver. *Theory of linear and integer programming*. Wiley-Interscience Series in Discrete Mathematics. John Wiley & Sons, Ltd., Chichester, 1986. A Wiley-Interscience Publication.

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

[122] A. Shapiro. Worst-case distribution analysis of stochastic programs. *Mathematical Programming*, 107(1):91–96, 2006.

[123] A. Shapiro. Stochastic programming approach to optimization under uncertainty. *Math. Program. Ser. B*, 112(1):183–220, 2008.

[124] A. Shapiro and A. Nemirovski. On complexity of stochastic programming problems. In V. Jeyakumar and A. M. Rubinov, editors, *Continuous Optimization: Current Trends and Applications*, pages 111–144. 2005.

[125] H. D. Sherali and W. P. Adams. *A Reformulation-Linearization Technique for Solving Discrete and Continuous Nonconvex Problems*. Kluwer, 1997.

[126] H. D. Sherali and W. P. Adams. *A reformulation-linearization technique for solving discrete and continuous nonconvex problems*, volume 31. Springer Science & Business Media, 2013.

[127] O. Solyalı, J.-F. Cordeau, and G. Laporte. The impact of modeling on robust inventory management under demand uncertainty. *Management Science*, 62(4):1188–1201, 2016.

[128] A. L. Soyster. Convex programming with set-inclusive constraints and applications to inexact linear programming. *Operations Research*, 21:1154–1157, 1973.

[129] R. E. Steuer. Algorithms for linear programming problems with interval objective function coefficients. *Mathematics of Operations Research*, 6(3):333–348, 1981.

[130] J. F. Sturm and S. Zhang. On cones of nonnegative quadratic functions. *Mathematics of Operations Research*, 28(2):246–267, 2003.

[131] A. Takeda, S. Taguchi, and R. H. Tütüncü. Adjustable robust optimization models for a nonlinear two-period system. *Journal of Optimization Theory and Applications*, 136(2):275–295, 2008.

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

[133] L. Vandenberghe, S. Boyd, and K. Comanor. Generalized chebyshev bounds via semidefinite programming. *SIAM review*, 49(1):52–64, 2007.

[134] Q. Wang, J. P. Watson, and Y. Guan. Two-stage robust optimization for nk contingency-constrained unit commitment. *Power Systems, IEEE Transactions on*, 28:2366–2375, 2013.

[135] J. E. Ward and R. E. Wendell. Approaches to sensitivity analysis in linear programming. *Annals of Operations Research*, 27(1):3–38, 1990.

[136] R. E. Wendell. A preview of a tolerance approach to sensitivity analysis in linear programming. *Discrete Mathematics*, 38(1):121–124, 1982.

[137] R. E. Wendell. Using bounds on the data in linear programming: The tolerance approach to sensitivity analysis. *Mathematical Programming*, 29(3):304–322, 1984.

[138] R. E. Wendell. The tolerance approach to sensitivity analysis in linear programming. *Management Science*, 31(5):564–578, 1985.

[139] W. Wiesemann, D. Kuhn, and B. Rustem. Robust resource allocations in temporal networks. *Math. Programm., Ser. A*, 135(1):437–471, 2011.

[140] W. Wiesemann, D. Kuhn, and M. Sim. Distributionally robust convex optimization. *Operations Research*, 62(6):1358–1376, 2014.

[141] L. A. Wolsey. *Integer programming*, volume 42. Wiley New York, 1998.

[142] F. Xie and R. Jia. Nonlinear fixed charge transportation problem by minimum cost flow-based genetic algorithm. *Computers & Industrial Engineering*, 63(4):763–778, 2012.

[143] B. Yang and S. Burer. A two-variable approach to the two-trust-region subproblem. *SIAM Journal on Optimization*, 26(1):661–680, 2016.

[144] B. Zeng and L. Zhao. Solving two-stage robust optimization problems using a column-and-constraint generation method. *Operations Research letters*, 41:457–461, 2013.

[145] L. Zhao and B. Zeng. Robust unit commitment problem with demand response and wind energy. In *Power and Energy Society General Meeting, 2012 IEEE*, pages 1–8. IEEE, 2012.

[146] L. F. Zuluaga, J. Vera, and J. Pea. LMI approximations for cones of positive semidefinite forms. *SIAM Journal on Optimization*, 16(4):1076–1091, 2006.

[147] S. Zymler, D. Kuhn, and B. Rustem. Distributionally robust joint chance constraints with second-order moment information. *Mathematical Programming*, 137(1-2):167–198, 2013.

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