Robust Optimization

by

Melvyn Sim

B.Eng, Electrical Engineering, National Univerity of Singapore (1995)
M.Eng, Electrical Engineering, National Univerity of Singapore (1996)
S.M, HPCES, Singapore-MIT-Alliance (2000)

Submitted to the Sloan School of Management
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Operations Research

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2004

Author ....
Sloan School of Management
May 14, 2004

Certified by ....
Dimitris J. Bertsimas
Boeing Professor of Operations Research
Sloan School of Management
Thesis Supervisor

Accepted by ....
James B. Orlin
Edward Pennell Brooks Professor of Operations Research
Co-Director, Operations Research Center

Robust Optimization

by

Melvyn Sim

Submitted to the Sloan School of Management
on May 14, 2004, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Operations Research

Abstract

We propose new methodologies in robust optimization that promise greater tractability, both theoretically and practically than the classical robust framework. We cover a broad range of mathematical optimization problems, including linear optimization (LP), quadratic constrained quadratic optimization (QCQP), general conic optimization including second order cone programming (SOCP) and semidefinite optimization (SDP), mixed integer optimization (MIP), network flows and $0-1$ discrete optimization. Our approach allows the modeler to vary the level of conservatism of the robust solutions in terms of probabilistic bounds of constraint violations, while keeping the problem tractable. Specifically, for LP, MIP, SOCP, SDP, our approaches retain the same complexity class as the original model. The robust QCQP becomes a SOCP, which is computationally as attractive as the nominal problem. In network flows, we propose an algorithm for solving the robust minimum cost flow problem in polynomial number of nominal minimum cost flow problems in a modified network. For $0-1$ discrete optimization problem with cost uncertainty, the robust counterpart of a polynomially solvable $0-1$ discrete optimization problem remains polynomially solvable and the robust counterpart of an NP-hard α-approximable $0-1$ discrete optimization problem, remains α-approximable. Under an ellipsoidal uncertainty set, we show that the robust problem retains the complexity of the nominal problem when the data is uncorrelated and identically distributed. For uncorrelated, but not identically distributed data, we propose an approximation method that solves the robust problem within arbitrary accuracy. We also propose a Frank-Wolfe type algorithm for this case, which we prove converges to a locally optimal solution, and in computational experiments is remarkably effective.

Thesis Supervisor: Dimitris J. Bertsimas
Title: Boeing Professor of Operations Research
Sloan School of Management

Acknowledgments

I would like to thank my thesis committee members, Dimitris Bertsimas, James Orlin and Georgia Perakis for their time, suggestions and valuable comments. I would also like to thank the National University of Singapore for the graduate scholarship.

This thesis is a celebration of my cordial relationship with Dimitris as my advisor and as a friend. I have benefited intellectually from his piercing insights and tastes for good research. His perpetual encouragement and motivation have illuminated my graduate life. When I was pursuing my master's program with the Singapore-MIT-Alliance, Dimitris encouraged me to pursue an academic career and was instrumental for admitting me to the ORC. At my crossroad, he wrote

> My only advice (on this but more generally in life): Do not let the financial situations dominate your decisions on what to do in life... Your main motivation should be to select what will give you the best chance to advance as a scientist and as a person. I believe very strongly in helping young talented people to achieve the best for themselves.

His message was comforting, far reaching and aptly exemplified his commitment to the welfare of his students.

I am fortunate to know Chung-Piaw Teo as my mentor, friend and colleague. Whenever we meet, either physically or in cyberspace, Chung-Piaw is always enthusiastic in sharing his knowledge and experiences. He has prepared me mentally for the challenges at MIT. Similarly, I am thankful to Robert Freund for sharing his mathematical insights and his teaching, research and peer review experiences. He is undoubtedly an excellent educator. With bittersweet memories, I am indebted to the late Kwok-Hong Lee for his guidance and encouragement. I will fondly remember his enthusiasm in teaching and passion for research.

The ORC is a great home. Friends such as Jose Correa, Michael Yee, David Craft, Ping Xu, Anshul Sood, Elodie Adida, Amr Farahat, Victor Martinez, Nicolas Stier, Michael Wagner, Adam Mesaurau, Alexandre Nogueira, Timothy Chan, Michele Aghassi, Margret Bjarnadottir, Agustin Bompadre, Pranava Goundan, Ekaterina Lesnaia, Stephen Shum, Dan Stratila, Dmitriy Rogozhnikov, Shobhit Gupta, Raghavendran Sivaraman, Romy Shoida, and many others have enriched my ORC experience. I am also privileged to work with Dessislava Pachamanova, who has contributed to Chapter 3 of this thesis.

I am grateful to Peng Sun, Huina Chen, Xin Chen, Yu Sheng for their warmth and hospitality, and to my Singaporean friends Harn-Wei Kua, Leonard Lee, Chiang-Juay Teo, Ruimin He, Pee-Ping Lai, Eng-Kiat Soh, Kwong-Meng Teo and Charlotte Yap for the fun, laughter and cordial friendship. We have forged a closer friendship after going through thick and thin.

My encounter with the Buddhist chaplain, Lama Tenzin Legphel Priyadarshi is one of my most gratifying experiences at MIT. I will fondly remember his teachings, meditative practices and penetrating insights of the Dharma for which I have a lifetime to discover and practice. I also appreciate the companionship of Chuan-Seng Tan, Irving Johnson, Scott Kennedy and many others in the Dharma group.

I am immensely indebted to my parents for their nurture, upbringing and unconditional love. Likewise, I am deeply grateful to my Guru, Lama Tan for his spiritual teachings and guidance. I am also thankful to my siblings Simon, Vincent, Steven, Boon-Khong, Yueh-Ling and my parents-in-law for their moral support and encouragement.

Finally, I would like to express my heartfelt gratitude, love and admiration to my wife, Yueh-Ping. Without her incredible support, encouragement, perseverance, patience, contentment, uplifting spirit and loving kindness, I would not have made it through the program so smoothly. We have made it together as a family.

Contents

1 Introduction **15**

1.1 Motivations and Philosophy 16

1.2 Literature Review 18

1.3 Structure of the Thesis 20

2 The Price of Robustness **23**

2.1 Robust Formulation of Linear Optimization Problems 24

2.1.1 Data Uncertainty in Linear Optimization 24

2.1.2 The Robust Formulation of Soyster 25

2.1.3 The Robust Formulation of Ben-Tal and Nemirovski 25

2.2 The New Robust Approach 26

2.3 Probability Bounds of Constraint Violation 29

2.3.1 On the Conservatism of Robust Solutions 40

2.3.2 Local Sensitivity Analysis of the Protection Level 45

2.4 Correlated Data 47

2.5 Experimental Results 48

2.5.1 A Simple Portfolio Problem 48

2.5.2 Robust Solutions of a Real-World Linear Optimization Problem 56

2.6 Conclusions 58

3 Robust Linear Optimization under General Norms **59**

3.1 Uncertainty Sets Defined by a Norm 61

3.2 The D-Norm and its Dual 64
3.2.1 Comparison with the Euclidean Norm 66
3.3 Probabilistic Guarantees 69
3.4 Conclusions 71

4 Robust Conic Optimization **73**
4.1 The Robust Model 76
4.1.1 Model for Parameter Uncertainty 76
4.1.2 Uncertainty Sets and Related Norms 76
4.1.3 The Class of Functions $f(\boldsymbol{x}, \boldsymbol{D})$ 78
4.2 The Proposed Robust Framework and its Tractability 79
4.2.1 Tractability of the Proposed Framework 82
4.2.2 Representation of $\max\{-f(\boldsymbol{x}, \Delta\boldsymbol{D}), -f(\boldsymbol{x}, -\Delta\boldsymbol{D})\}$ 84
4.2.3 The Nature and Size of the Robust Problem 88
4.3 Probabilistic Guarantees 94
4.4 General Cones 100
4.5 Conclusions 107

5 Robust Discrete Optimization and Network Flows **109**
5.1 Robust Formulation of Discrete Optimization Problems 110
5.1.1 Robust MIP Formulation 111
5.2 Robust Combinatorial Optimization 113
5.2.1 Algorithm for Robust Combinatorial Optimization Problems 115
5.3 Robust Approximation Algorithms 118
5.4 Robust Network Flows 120
5.5 Experimental Results 125
5.5.1 The Robust Knapsack Problem 125
5.5.2 Robust Sorting 128
5.5.3 The Robust Shortest Path Problem 131

5.6 Conclusions 134

6 Robust Discrete Optimization under an Ellipsoidal Uncertainty Set 137

6.1 Formulation of Robust Discrete Optimization Problems 139

6.2 Structural Results 141

6.3 Approximation via Piecewise Linear Functions 144

6.4 A Frank-Wolfe Type Algorithm 149

6.5 Generalized Discrete Robust Formulation 155

6.6 Experimental Results 159

6.6.1 The Robust Knapsack Problem 159

6.6.2 The Robust Minimum Cost over a Uniform Matroid 160

6.7 Conclusions 161

7 Conclusions 165

7.1 Future research 166

List of Figures

2-1 Comparison of probability bounds for $n = |J_i| = 10$. 40

2-2 Comparison of probability bounds for $n = |J_i| = 100$. 41

2-3 Comparison of probability bounds for $n = |J_i| = 2000$. 42

2-4 The return and the objective function value (risk adjusted return) as a function of the protection level Γ. 51

2-5 The solution of the portfolio for various protection levels. 52

2-6 Simulation study of the probability of underperforming the nominal return as a function of Γ. 53

2-7 Empirical Result of expected, maximum and minimum yield. 54

2-8 Tradeoffs between probability of underperforming and returns. 55

2-9 The tradeoff between cost and robustness. 57

5-1 Conversion of arcs with cost uncertainties. 122

5-2 The tradeoff between robustness and optimality in twenty instances of the 0-1 knapsack problem. 127

5-3 The cumulative distribution of cost (for $\rho = 0.2$) for various values of Γ for the robust sorting problem. 130

5-4 Randomly generated digraph and the set of robust shortest $\{s, t\}$ paths for various Γ values. 132

5-5 Influence of Γ on the distribution of path cost for $\rho = 0.1$. 133

5-6 The cumulative distribution of cost (for $\rho = 0.1$) for various values of Γ for the robust shortest path problem. 134

6-1 Illustration of the maximum gap between the function $f(w)$ and $g(w)$. 154

List of Tables

2.1 Choice of Γ_i as a function of $n = |J_i|$ so that the probability of constraint violation is less than 1%. 41

2.2 Simulation results given by the robust solution. 54

2.3 Distributions of $|J_i|$ in PILOT4. 56

2.4 The tradeoff between optimal cost and robustness. 57

4.1 The function $f(\boldsymbol{x}, \boldsymbol{D})$ for different conic optimization problems. . . . 79

4.2 Representation of the dual norm for $\boldsymbol{t} \geq \boldsymbol{0}$. 85

4.3 The function $\mathcal{H}(\boldsymbol{x}, \boldsymbol{\Delta D}^j)$ and the norm $\|\cdot\|_g$ for different conic optimization problems. 88

4.4 Size increase and nature of robust formulation when each data entry has independent uncertainty. 94

4.5 Probability bounds of $\mathrm{P}(f(\boldsymbol{x}, \tilde{\boldsymbol{D}}) < 0)$ for $\tilde{\boldsymbol{z}} \sim \mathcal{N}(0, \boldsymbol{I})$. 99

4.6 Sample calculations of Ω using Probability Bounds of Table 4.5 for $m = 100$, $n = 100$, $|N| = 495,000$. 100

5.1 Robust Knapsack Solutions. 126

5.2 Influence of Γ on $\overline{Z}(\Gamma)$ and $\sigma(\Gamma)$. 129

6.1 Number of subproblems, k as a function of the desired precision ϵ, size of the problem $\boldsymbol{d}'\boldsymbol{e}$ and $\Omega = 4$. 148

6.2 Robust Knapsack Solutions. 160

6.3 Performance of Algorithm 6.4 the Robust Minimum Cost problem over a Uniform Matroid. 162

Chapter 1

Introduction

In mathematical optimization models, we commonly assume that the data inputs are precisely known and ignore the influence of parameter uncertainties on the optimality and feasibility of the models. It is therefore conceivable that as the data differs from the assumed nominal values, the generated "optimal solution" may violate critical constraints and perform poorly from an objective function point of view. These observations motivate the need for methodologies in mathematical optimization models that solve for solutions that are immune to data uncertainty.

Robust optimization addresses the issue of data uncertainties from the perspective of computational tractability. In the past decade, there was considerable development in the theory of robust convex optimization. However, under the robust framework found in the literature, the robust models generally lead to an increase in computational complexity over the nominal problem, which is an issue when solving large problems. Moreover, there is often lack of probabilistic justification motivating the choice of parameters used in the robust framework. Furthermore, these results, though valid in convex optimization, do not necessarily carry forward in a tractable way to discrete optimization.

In this thesis, we propose new methodologies in robust optimization that lead to greater tractability, both theoretically and empirically, than those found in the literature. In particular, we cover a broad range of mathematical optimization problems, including linear optimization (LP), quadratic constrained quadratic optimization (QCQP), second order cone optimization (SOCP), semidefinite optimization (SDP), general conic optimization, mixed integer optimization (MIP), network flows and combinatorial optimization. To justify our robust framework, we focus on deriving probability bounds on the feasibility of the robust solution, which are generally lacking in related literature on robust optimization.

Structure of the chapter. In Section 1.1, we discuss the motivations and philosophy of robust optimizations. In Section 1.2, we review the works related to robust optimization. In Section 1.3, we outline the structure of this thesis.

1.1 Motivations and Philosophy

Data uncertainty is present in many real-world optimization problems. For example, in supply chain optimization, the actual demand for products, financial returns, actual material requirements and other resources are not precisely known when critical decisions need to be made. In engineering and science, data is subjected to measurement errors, which also constitute sources of data uncertainty in the optimization model.

In mathematical optimization, we generally assume that the data is precisely known. We then seek to minimize (or maximize) an objective function over a set of decision variables as follows:

$$\begin{array}{ll} \text{minimize} & f_0(\boldsymbol{x}, \boldsymbol{D}_0) \\ \text{subject to} & f_i(\boldsymbol{x}, \boldsymbol{D}_i) \geq 0 \quad \forall i \in I, \end{array} \tag{1.1}$$

where $\boldsymbol{x}$ is the vector of decision variables and $\boldsymbol{D}_i$, $i \in I \cup \{0\}$ are the data that is part of the inputs of the optimization problem.

When parameters in the objective function are uncertain, we are unlikely to achieve the desired "optimal value." However, the extent of adverse variation of the objective is often a cause for concern. Many modelers are willing to tradeoff optimality for a solution that has greater reliability in achieving their desired goal.

If parameter uncertainty arises at the constraints, when implementing a solution, it is likely that these constraints would be violated upon realization of the actual data. In many practical optimization problems, constraint violations can potentially influence the usability of the solution. We quote from the case study by Ben-Tal and Nemirovski [7] on linear optimization problems from the Net Lib library:

> In real-world applications of Linear Programming, one cannot ignore the possibility that a small uncertainty in the data can make the usual optimal solution completely meaningless from a practical viewpoint.

The classical methods of addressing parameter uncertainty includes sensitivity analysis and stochastic optimization. In the former approach, practitioners ignore the influence of data uncertainty in their models and subsequently perform sensitivity analysis to justify their solutions. However, sensitivity analysis is only a tool for analyzing the goodness of a solution. It is not particularly helpful for generating solutions that are robust to data changes. Furthermore, it is impractical to perform joint sensitivity analysis in models with large number of uncertain parameters.

In stochastic optimization, we express the feasibility of a solution using chance constraints. Assuming that we are given the distributions of the input parameters, the corresponding stochastic optimization problem to problem (1.1) is:

$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & \Pr\left(f_0(\boldsymbol{x}, \tilde{\boldsymbol{D}}_0) \leq t\right) \geq p_0 \\ & \Pr\left(f_i(\boldsymbol{x}, \tilde{\boldsymbol{D}}_i) \geq 0\right) \geq p_i \quad \forall i \in I, \end{array} \tag{1.2}$$

where $\tilde{\boldsymbol{D}}_i$, $i \in I \cup \{0\}$ are the random variables associated with the ith constraint. Although the model (1.2) is expressively rich, there are some fundamental difficulties. We can rarely obtain the actual distributions of the uncertain parameters. Moreover, even if we know the distributions, it is still computationally challenging to evaluate the chance constraints, let alone to optimize the model. Furthermore, the chance constraint can destroy the convexity properties and elevate significantly the complexity of the original problem .

In view of the difficulties, robust optimization presents a different approach to handling data uncertainty. In designing such an approach two criteria are important in our view:

- **Tractability:** Preserving the computational tractability both theoretically and most importantly practically of the nominal problem. From a theoretical perspective it is desirable that if the nominal problem is solvable in polynomial time, then the robust problem is also polynomially solvable.

- **Probability bounds:** Being able to find a guarantee on the probability that the robust solution is feasible, when the uncertain coefficients obey some natural probability distributions. This is important, since from these guarantees we can select parameters that allow to control the tradeoff between robustness and optimality.

In our literature review, we found that these criteria are not always fulfilled in classical models on robust optimization. Nevertheless, we feel that for robust optimization to have an impact, in theory and practice, we need to address these criteria, which is the primary focus of this thesis.

1.2 Literature Review

In recent years a body of literature is developing under the name of robust optimization, in which we optimize against the worst instances that might arise by using a min-max objective. Mulvey et al. [22] present an approach that integrates goal optimization formulations with scenario-based description of the problem data. Soyster, in the early 1970s, proposes a linear optimization model to construct a solution that is feasible for all data that belong in a convex set. The resulting model produces solutions that are too conservative in the sense that we give up too much of optimality for the nominal problem in order to ensure robustness (see the comments of Ben-Tal and Nemirovski [7]).

A significant step forward for developing a theory for robust optimization was taken independently by Ben-Tal and Nemirovski [7, 6, 4] and El-Ghaoui et al. [11, 12]. They proposed the following framework on robust optimization:

$$\begin{array}{lll} \text{minimize} & \max_{\boldsymbol{D}_0 \in \mathcal{U}_0} f_0(\boldsymbol{x}, \boldsymbol{D}_0) & \\ \text{subject to} & f(\boldsymbol{x}, \boldsymbol{D}_i) \geq 0 & \forall i \in I, \forall \boldsymbol{D}_i \in \mathcal{U}_i, \end{array} \tag{1.3}$$

where $\mathcal{U}_i$, $i \in I \cup \{0\}$, are the given uncertainty sets. They showed that under the assumption that the the set $\mathcal{U}_i$ are ellipsoids of the form

$$\mathcal{U} = \left\{ \boldsymbol{D} \mid \exists \boldsymbol{u} \in \Re^{|N|} : \boldsymbol{D} = \boldsymbol{D}^0 + \sum_{j \in N} \Delta \boldsymbol{D}^j u_j, \|\boldsymbol{u}\| \leq \Omega \right\},$$

the robust counterparts of some important generic convex optimization problems (linear optimization (LP), second order cone optimization problems (SOCP), semidefinite optimization (SDP) are either exactly, or approximately tractable problems that are efficiently solvable via interior point methods. However, under ellipsoidal uncertainty sets, the robust counterpart of an LP becomes an SOCP, of an SOCP becomes an SDP, while the robust counterpart of an SDP is NP-hard to solve. In other words, the difficulty of the robust problem increases, as SDPs are numerically harder to solve than SOCPs, which in turn are harder to solve than LPs. Hence, a practical drawback of such an approach, is that it leads to nonlinear, although convex, models, which are more demanding computationally than the earlier linear models by Soyster [26] (see also the discussion in Ben-Tal and Nemirovski [7]). Furthermore, except for the case of LP, these papers do not provide a guarantee on the probability that the robust solution is feasible, when the uncertain coefficients obey some natural probability distributions.

Another disadvantage of their nonlinear robust framework is the natural exclusion of discrete optimization models, which is predominantly LP based. Specifically for discrete optimization problems, Kouvelis and Yu [20] proposed a framework for robust discrete optimization, which seeks to find a solution that minimizes the worst case performance under a set of scenarios for the data. Unfortunately, under their approach, the robust counterpart of many polynomially solvable discrete optimization problems becomes NP-hard. A related objective is the minimax-regret approach, which seeks to minimize the worst case loss in objective value that may occur. Again, under the minimax-regret notion of robustness, many of the polynomially solvable discrete optimization problems become NP-hard. Under the minimax-regret robustness approach, Averbakh [3] showed that polynomial solvability is preserved for a specific discrete optimization problem (optimization over a uniform matroid) when each cost coefficient can vary within an interval (interval representation of uncertainty); however, the approach does not seem to generalize to other discrete optimization problems. There have also been research efforts to apply stochastic optimization methods to discrete optimization (see for example Schultz et al. [25]), but the computational requirements are even more severe in this case.

1.3 Structure of the Thesis

This thesis is organized as follows:

- **Chapter 2: The Price of Robustness.** Under robust optimization, we are willing to accept a suboptimal solution for the nominal values of the data, in order to ensure that the solution remains feasible and near optimal when the data changes. A concern with such an approach is that it might be too conservative. In Chapter 2, we propose an approach that attempts to make this tradeoff more attractive, that is we investigate ways to decrease what we call the price of robustness. In particular, we can flexibly adjust the level of conservatism of the robust solutions in terms of probabilistic bounds of constraint violations. An attractive aspect of our method is that the new robust formulation is also a linear optimization problem. We report numerical results for a portfolio optimization problem, and a problem from the Net Lib library.

- **Chapter 3: Robust Linear Optimization under General Norms.** We explicitly characterize the robust counterpart as a convex optimization problem that involves the dual norm of the given norm. Our approach encompasses several approaches from the literature and provide guarantees for constraints violation under probabilistic models that allow arbitrary dependencies in the distribution of the uncertain coefficients.

- **Chapter 4: Robust Conic Optimization.** In earlier proposals, the robust counterpart of conic optimization problems exhibits a lateral increase in complexity, i.e., robust LPs become SOCPs, robust SOCPs become SDPs, and robust SDPs become NP-hard. We propose a relaxed robust counterpart for general conic optimization problems that **(a)** preserves the computational tractability of the nominal problem; specifically the robust conic optimization problem retains its original structure, i.e., robust linear optimization problems (LPs) remain LPs, robust second order cone optimization problems (SOCPs) remain SOCPs and robust semidefinite optimization problems (SDPs) remain SDPs; moreover, when the data entries are independently distributed, the size of the proposed robust problem especially under the l_2 norm is practically the same as the nominal problem, and **(b)** allows us to provide a guarantee on the probability that the robust solution is feasible, when the uncertain coefficients obey independent and identically distributed normal distributions.

- **Chapter 5: Robust Discrete Optimization and Network Flows.** We propose an approach to address data uncertainty for discrete optimization and network flow problems that allows controlling the degree of conservatism of the solution, and is computationally tractable both practically and theoretically. In particular, when both the cost coefficients and the data in the constraints of an integer optimization problem are subject to uncertainty, we propose a robust integer optimization problem of moderately larger size that allows controlling the degree of conservatism of the solution in terms of probabilistic bounds on constraint violation. When only the cost coefficients are subject to uncertainty and the problem is a $0-1$ discrete optimization problem on n variables, then we solve the robust counterpart by solving at most $n+1$ instances of the original problem. Thus, the robust counterpart of a polynomially solvable $0-1$ discrete optimization problem remains polynomially solvable. In particular, robust matching, spanning tree, shortest path, matroid intersection, etc. are polynomially solvable. We also show that the robust counterpart of an NP-hard α-approximable $0-1$ discrete optimization problem, remains α-approximable. Finally, we propose an algorithm for robust network flows that solves the robust counterpart by solving a polynomial number of nominal minimum cost flow problems in a modified network.

- **Chapter 6: Robust Discrete Optimization under an Ellipsoidal Uncertainty Sets.** We address the complexity and practically efficient methods for robust discrete optimization under ellipsoidal uncertainty sets. Specifically, we show that the robust counterpart of a discrete optimization problem under ellipsoidal uncertainty is NP-hard even though the nominal problem is polynomially solvable. For uncorrelated and identically distributed data we show that the robust problem retains the complexity of the nominal problem. For uncorrelated, but not identically distributed data we propose an approximation method that solves the robust problem within arbitrary accuracy. We also propose a Frank-Wolfe type algorithm for this case, which we prove converges to a locally optimal solution, and in computational experiments is remarkably effective. Finally, we propose a generalization of the robust discrete optimization framework presented in Chapter 5 that allows the key parameter that controls the tradeoff between robustness and optimality to depend on the solution that results in increased flexibility and decreased conservatism, while maintaining the complexity of the nominal problem.

- **Chapter 7: Conclusions.** This chapter contains the concluding remarks of the thesis.

Chapter 2

The Price of Robustness

In this chapter, we propose a new approach for robust linear optimization that retains the advantages of the linear framework of Soyster [26]. More importantly, our approach offers full control on the degree of conservatism for every constraint. We protect against violation of constraint i deterministically, when only a prespecified number Γ_i of the coefficients changes, that is we guarantee that the solution is feasible if less than Γ_i uncertain coefficients change. Moreover, we provide a probabilistic guarantee that even if more than Γ_i change, then the robust solution will be feasible with high probability. In the process we prove a new, to the best of our knowledge, tight bound on sums of symmetrically distributed random variables. In this way, the proposed framework is at least as flexible as the one proposed by Ben-Tal and Nemirovski [7, 6, 4] and El-Ghaoui et al. [11, 12] and possibly more. Unlike these approaches, the robust counterparts we propose are linear optimization problems, and thus our approach readily generalizes to discrete optimization problems. To the best of our knowledge, there was no similar work done in the robust discrete optimization domain that involves deterministic and probabilistic guarantees of constraints against violation.

Structure of the chapter. In Section 2.1, we present the different approaches for robust linear optimization from the literature and discuss their merits. In Section 2.2 we propose the new approach and show that it can be solved as a linear optimization problem. In Section 2.3, we show that the proposed robust LP has attractive probabilistic and deterministic guarantees. Moreover, we perform sensitivity analysis of the degree of protection the proposed method offers. We provide extensions to our basic framework dealing with correlated uncertain data in Section 2.4. In Section 2.5, we apply the proposed approach to a portfolio problem and a problem from the Net Lib library. Finally, Section 2.6 contains some concluding remarks.

2.1 Robust Formulation of Linear Optimization Problems

2.1.1 Data Uncertainty in Linear Optimization

We consider the following nominal linear optimization problem:

$$\begin{aligned} \text{maximize} \quad & \boldsymbol{c}'\boldsymbol{x} \\ \text{subject to} \quad & \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b} \\ & \boldsymbol{l} \leq \boldsymbol{x} \leq \boldsymbol{u}. \end{aligned}$$

In the above formulation, we assume that data uncertainty only affects the elements in matrix $\boldsymbol{A}$. We assume without loss of generality that the objective function $\boldsymbol{c}$ is not subject to uncertainty, since we can use the objective maximize z, add the constraint $z - \boldsymbol{c}'\boldsymbol{x} \leq 0$, and thus include this constraint into $\boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b}$.

Model of Data Uncertainty U:

Consider a particular row i of the matrix $\boldsymbol{A}$ and let J_i the set of coefficients in row i that are subject to uncertainty. Each entry a_{ij}, $j \in J_i$ is modeled as a symmetric and bounded random variable $\tilde{a}_{ij}, j \in J_i$ (see Ben-Tal and Nemirovski [7]) that takes values in $[a_{ij} - \hat{a}_{ij}, a_{ij} + \hat{a}_{ij}]$. Associated with the uncertain data $\tilde{a}_{ij}$, we define the random variable $\eta_{ij} = (\tilde{a}_{ij} - a_{ij})/\hat{a}_{ij}$, which obeys an unknown, but symmetric distribution, and takes values in $[-1, 1]$.

2.1.2 The Robust Formulation of Soyster

As we have mentioned in the Introduction Soyster [26] considers column-wise uncertainty. Under the model of data uncertainty U, the robust formulation is as follows:

$$\begin{aligned} \text{maximize} \quad & \boldsymbol{c}'\boldsymbol{x} \\ \text{subject to} \quad & \sum_j a_{ij}x_j + \sum_{j\in J_i} \hat{a}_{ij}y_j \le b_i \quad \forall i \\ & -y_j \le x_j \le y_j \quad \forall j \\ & \boldsymbol{l} \le \boldsymbol{x} \le \boldsymbol{u} \\ & \boldsymbol{y} \ge \boldsymbol{0}. \end{aligned} \tag{2.1}$$

Let $\boldsymbol{x}^*$ be the optimal solution of Formulation (2.1). At optimality clearly, $y_j = |x_j^*|$, and thus

$$\sum_j a_{ij}x_j^* + \sum_{j\in J_i} \hat{a}_{ij}|x_j^*| \le b_i \quad \forall i.$$

We next show that for every possible realization $\tilde{a}_{ij}$ of the uncertain data, the solution remains feasible, that is the solution is "robust." We have

$$\sum_j \tilde{a}_{ij}x_j^* = \sum_j a_{ij}x_j^* + \sum_{j\in J_i} \eta_{ij}\hat{a}_{ij}x_j^* \le \sum_j a_{ij}x_j^* + \sum_{j\in J_i} \hat{a}_{ij}|x_j^*| \le b_i \quad \forall i$$

For every ith constraint, the term, $\sum_{j\in J_i} \hat{a}_{ij}|x_j|$ gives the necessary "protection" of the constraint by maintaining a gap between $\sum_j a_{ij}x_j^*$ and b_i.

2.1.3 The Robust Formulation of Ben-Tal and Nemirovski

Although the Soyster's method admits the highest protection, it is also the most conservative in practice in the sense that the robust solution has an objective function value much worse than the objective function value of the solution of the nominal linear optimization problem. To address this conservatism, Ben-Tal and Nemirovski

[7] propose the following robust problem:

$$\begin{aligned}
\text{maximize} \quad & \boldsymbol{c}'\boldsymbol{x} \\
\text{subject to} \quad & \sum_j a_{ij}x_j + \sum_{j \in J_i} \hat{a}_{ij}y_{ij} + \Omega_i \sqrt{\sum_{j \in J_i} \hat{a}_{ij}^2 z_{ij}^2} \leq b_i \quad \forall i \\
& -y_{ij} \leq x_j - z_{ij} \leq y_{ij} \qquad \forall i, j \in J_i \\
& \boldsymbol{l} \leq \boldsymbol{x} \leq \boldsymbol{u} \\
& \boldsymbol{y} \geq \boldsymbol{0}.
\end{aligned} \tag{2.2}$$

Under the model of data uncertainty U, the authors have shown that the probability that the i constraint is violated is at most $\exp(-\Omega_i^2/2)$. The robust Model (2.2) is less conservative than Model (2.1) as every feasible solution of the latter problem is a feasible solution to the former problem.

We next examine the sizes of Formulations (2.1) and (2.2). We assume that there are k coefficients of the $m \times n$ nominal matrix $\boldsymbol{A}$ that are subject to uncertainty. Given that the original nominal problem has n variables and m constraints (not counting the bound constraints), Model (2.1) is a linear optimization problem with $2n$ variables, and $m + 2n$ constraints. In contrast, Model (2.2) is a second order cone problem, with $n + 2k$ variables and $m + 2k$ constraints. Since Model (2.2) is a nonlinear one, it is particularly not attractive for solving robust discrete optimization models.

2.2 The New Robust Approach

In this section, we propose a robust formulation that is linear, is able to withstand parameter uncertainty under the model of data uncertainty U without excessively affecting the objective function, and readily extends to discrete optimization problems.

We motivate the formulation as follows. Consider the ith constraint of the nominal problem $\boldsymbol{a}_i'\boldsymbol{x} \leq b_i$. Let J_i be the set of coefficients a_{ij}, $j \in J_i$ that are subject to parameter uncertainty, i.e., $\tilde{a}_{ij}$, $j \in J_i$ takes values according to a symmetric distribution with mean equal to the nominal value a_{ij} in the interval $[a_{ij} - \hat{a}_{ij}, a_{ij} + \hat{a}_{ij}]$. For every i, we introduce a parameter Γ_i, not necessarily integer, that takes values in the interval $[0, |J_i|]$. As it would become clear below, the role of the parameter Γ_i is to adjust the robustness of the proposed method against the level of conservatism of the solution. Speaking intuitively, it is unlikely that all of the a_{ij}, $j \in J_i$ will change. Our goal is to be protected against all cases that up to $\lfloor \Gamma_i \rfloor$ of these coefficients are allowed to change, and one coefficient a_{it} changes by $(\Gamma_i - \lfloor \Gamma_i \rfloor)\hat{a}_{it}$. In other words, we stipulate that nature will be restricted in its behavior, in that only a subset of the coefficients will change in order to adversely affect the solution. We will develop an approach, that has the property that if nature behaves like this, then the robust solution will be feasible **deterministically**, and moreover, even if more than $\lfloor \Gamma_i \rfloor$ change, then the robust solution will be feasible **with very high probability**.

We consider the following (still nonlinear) formulation:

$$\begin{aligned}
\text{maximize} \quad & \boldsymbol{c}'\boldsymbol{x} \\
\text{subject to} \quad & \sum_j a_{ij}x_j + \max_{\{S_i \cup \{t_i\} | \ S_i \subseteq J_i, |S_i| = \lfloor \Gamma_i \rfloor, t_i \in J_i \setminus S_i\}} \\
& \qquad \left\{ \sum_{j \in S_i} \hat{a}_{ij} y_j + (\Gamma_i - \lfloor \Gamma_i \rfloor) \hat{a}_{it_i} y_t \right\} \leq b_i \quad \forall i \\
& -y_j \leq x_j \leq y_j \qquad \forall j \\
& \boldsymbol{l} \leq \boldsymbol{x} \leq \boldsymbol{u} \\
& \boldsymbol{y} \geq \boldsymbol{0}.
\end{aligned} \tag{2.3}$$

If Γ_i is chosen as an integer, the ith constraint is protected by

$$\beta_i(\boldsymbol{x}, \Gamma_i) = \max_{\{S_i | \ S_i \subseteq J_i, |S_i| = \Gamma_i\}} \{ \sum_{j \in S_i} \hat{a}_{ij} |x_j| \}.$$

Note that when $\Gamma_i = 0$, $\beta_i(\boldsymbol{x}, \Gamma_i) = 0$ the constraints are equivalent to that of the nominal problem. Likewise, if $\Gamma_i = |J_i|$, we have Soyster's method. Therefore, by varying $\Gamma_i \in [0, |J_i|]$, we have the flexibility of adjusting the robustness of the method against the level of conservatism of the solution.

In order to reformulate Model (2.3) as a linear optimization model we need the following proposition.

Proposition 1 *Given a vector $\boldsymbol{x}^*$, the protection function of the ith constraint,*

$$\beta_i(\boldsymbol{x}^*, \Gamma_i) = \max_{\{S_i \cup \{t_i\} |\ S_i \subseteq J_i, |S_i| = \lfloor \Gamma_i \rfloor, t_i \in J_i \setminus S_i\}} \left\{ \sum_{j \in S_i} \hat{a}_{ij} |x_j^*| + (\Gamma_i - \lfloor \Gamma_i \rfloor) \hat{a}_{it_i} |x_j^*| \right\} \quad (2.4)$$

equals to the objective function of the following linear optimization problem:

$$\begin{aligned} \beta_i(\boldsymbol{x}^*, \Gamma_i) = \text{maximize} \quad & \sum_{j \in J_i} \hat{a}_{ij} |x_j^*| z_{ij} \\ \text{subject to} \quad & \sum_{j \in J_i} z_{ij} \leq \Gamma_i \\ & 0 \leq z_{ij} \leq 1 \quad \forall j \in J_i. \end{aligned} \quad (2.5)$$

Proof : Clearly the optimal solution value of Problem (2.5) consists of $\lfloor \Gamma_i \rfloor$ variables at 1 and one variable at $\Gamma_i - \lfloor \Gamma_i \rfloor$. This is equivalent to the selection of subset $\{S_i \cup \{t_i\} |\ S_i \subseteq J_i, |S_i| = \lfloor \Gamma_i \rfloor, t_i \in J_i \setminus S_i\}$ with corresponding cost function $\sum_{j \in S_i} \hat{a}_{ij} |x_j^*| + (\Gamma_i - \lfloor \Gamma_i \rfloor) \hat{a}_{it_i} |x_j^*|$. ■

We next reformulate Model (2.3) as a linear optimization model.

Theorem 1 *Model (2.3) has an equivalent linear formulation as follows:*

$$\begin{aligned} \text{maximize} \quad & \boldsymbol{c}'\boldsymbol{x} \\ \text{subject to} \quad & \sum_j a_{ij} x_j + z_i \Gamma_i + \sum_{j \in J_i} p_{ij} \leq b_i \quad && \forall i \\ & z_i + p_{ij} \geq \hat{a}_{ij} y_j && \forall i, j \in J_i \\ & -y_j \leq x_j \leq y_j && \forall j \\ & l_j \leq x_j \leq u_j && \forall j \\ & p_{ij} \geq 0 && \forall i, j \in J_i \\ & y_j \geq 0 && \forall j \\ & z_i \geq 0 && \forall i. \end{aligned} \quad (2.6)$$

Proof : We first consider the dual of Problem (2.5):

$$\begin{array}{lll} \text{minimize} & \sum_{j \in J_i} p_{ij} + \Gamma_i z_i & \\ \text{subject to} & z_i + p_{ij} \geq \hat{a}_{ij} |x_j^*| & \forall i, j \in J_i \\ & p_{ij} \geq 0 & \forall j \in J_i \\ & z_i \geq 0 & \forall i. \end{array} \tag{2.7}$$

By strong duality, since Problem (2.5) is feasible and bounded for all $\Gamma_i \in [0, |J_i|]$, then the dual problem (2.7) is also feasible and bounded and their objective values coincide. Using Proposition 1, we have that $\beta_i(\boldsymbol{x}^*, \Gamma_i)$ is equal to the objective function value of Problem 2.7. Substituting to Problem (2.3) we obtain that Problem (2.3) is equivalent to the linear optimization problem (2.6). ■

Remark : The robust linear optimization Model (2.6) has $n + k + 1$ variables and $m + k + n$ constraints, where $k = \sum_i |J_i|$ the number of uncertain data, contrasted with $n + 2k$ variables and $m + 2k$ constraints for the nonlinear Formulation (2.2). In most real-world applications, the matrix $\boldsymbol{A}$ is sparse. An attractive characteristic of Formulation (2.6) is that it preserves the sparsity of the matrix $\boldsymbol{A}$.

2.3 Probability Bounds of Constraint Violation

It is clear by the construction of the robust formulation that if up to $\lfloor \Gamma_i \rfloor$ of the J_i coefficients a_{ij} change within their bounds, and up to one coefficient a_{it_i} changes by $(\Gamma_i - \lfloor \Gamma_i \rfloor)\hat{a}_{it}$, then the solution of Problem (2.6) will remain feasible. In this section, we show that under the Model of Data Uncertainty U, the robust solution is feasible with high probability. The parameter Γ_i controls the tradeoff between the probability of violation and the effect to the objective function of the nominal problem, which is what we call **the price of robustness**.

In preparation for our main result in this section, we prove the following proposition.

Proposition 2 *Let $\boldsymbol{x}^*$ be an optimal solution of Problem (2.6). Let S_i^* and t_i^* be the* *set and the index respectively that achieve the maximum for $\beta_i(\boldsymbol{x}^*, \Gamma_i)$ in Eq. (2.4). Suppose that the data in matrix $\boldsymbol{A}$ are subjected to the model of data uncertainty U.*

(a) *The probability that the ith constraint is violated satisfies:*

$$\mathrm{P}\left(\sum_j \tilde{a}_{ij}x_j^* > b_i\right) \leq \mathrm{P}\left(\sum_{j\in J_i} \gamma_{ij}\eta_{ij} \geq \Gamma_i\right)$$

where

$$\gamma_{ij} = \begin{cases} 1, & \text{if } j \in S_i^* \\ \dfrac{\hat{a}_{ij}|x_j^*|}{\hat{a}_{ir^*}|x_{r^*}^*|}, & \text{if } j \in J_i \backslash S_i^* \end{cases}$$

and

$$r^* = \arg\min_{r\in S_i^*\cup\{t_i^*\}} \hat{a}_{ir}|x_r^*|.$$

(b) *The quantities γ_{ij} satisfy $\gamma_{ij} \leq 1$ for all $j \in J_i\backslash S_i^*$.*

Proof : **(a)** Let x^*, S_i^* and t_i^* be the solution of Model (2.3). Then the probability of violation of the ith constraint is as follows:

$$\mathrm{P}\left(\sum_j \tilde{a}_{ij}x_j^* > b_i\right) \tag{2.8}$$

$$= \mathrm{P}\left(\sum_j a_{ij}x_j^* + \sum_{j\in J_i} \eta_{ij}\hat{a}_{ij}x_j^* > b_i\right)$$

$$\leq \mathrm{P}\left(\sum_{j\in J_i} \eta_{ij}\hat{a}_{ij}|x_j^*| > \sum_{j\in S_i^*} \hat{a}_{ij}|x_j^*| + (\Gamma_i - \lfloor\Gamma_i\rfloor)\hat{a}_{it_i^*}|x_{t_i^*}^*|\right) \tag{2.9}$$

$$= \mathrm{P}\left(\sum_{j\in J_i\backslash S_i^*} \eta_{ij}\hat{a}_{ij}|x_j^*| > \sum_{j\in S_i^*} \hat{a}_{ij}|x_j^*|(1-\eta_{ij}) + (\Gamma_i - \lfloor\Gamma_i\rfloor)\hat{a}_{it_i^*}|x_{t_i^*}^*|\right)$$

$$\leq \mathrm{P}\left(\sum_{j\in J_i\backslash S_i^*} \eta_{ij}\hat{a}_{ij}|x_j^*| > \hat{a}_{ir^*}|x_{r^*}^*|\left(\sum_{j\in S_i^*}(1-\eta_{ij}) + (\Gamma_i - \lfloor\Gamma_i\rfloor)\right)\right) \tag{2.10}$$

$$= \mathrm{P}\left(\sum_{j\in S_i^*} \eta_{ij} + \sum_{j\in J_i\backslash S_i^*} \frac{\hat{a}_{ij}|x_j^*|}{\hat{a}_{ir^*}|x_{r^*}^*|}\eta_{ij} > \Gamma_i\right)$$

$$= \mathrm{P}\left(\sum_{j\in J_i} \gamma_{ij}\eta_{ij} > \Gamma_i\right)$$

$$\leq \mathrm{P}\left(\sum_{j\in J_i} \gamma_{ij}\eta_{ij} \geq \Gamma_i\right).$$

Inequality (2.9) follows from Inequality (2.3), since $\boldsymbol{x}^*$ satisfies

$$\sum_j a_{ij}x_j^* + \sum_{j\in S_i^*} \hat{a}_{ij}y_j + (\Gamma_i - \lfloor\Gamma_i\rfloor)\hat{a}_{it_i^*}y_{t_i^*} \leq b_i.$$

Inequality (2.10) follows from $1 - \eta_{ij} \geq 0$ and $r^* = \arg\min_{r\in S_i^*\cup\{t_i^*\}} \hat{a}_{ir}|x_r^*|$.

(**b**) Suppose there exist $l \in J_i \backslash S_i^*$ such that $\hat{a}_{il}|x_l^*| > \hat{a}_{ir^*}|x_{r^*}^*|$. If $l \neq t_i^*$, then, since $r^* \in S_i^* \cup \{t^*\}$, we could increase the objective value of $\sum_{j\in S_i^*} \hat{a}_{ij}|x_j^*| + (\Gamma_i - \lfloor\Gamma_i\rfloor)\hat{a}_{it^*}|x_{t^*}^*|$ by exchanging l with r^* from the set $S_i^* \cup \{t_i^*\}$. Likewise, if $l = t_i^*$, $r^* \in S_i^*$, we could exchange t_i^* with r^* in the set S_i^* to increase the same objective function. In both cases, we arrive at a contradiction that $S_i^* \cup \{t_i^*\}$ is an optimum solution to this objective function. ■

We are naturally led to bound the probability $\mathrm{P}(\sum_{j\in J_i} \gamma_{ij}\eta_{ij} \geq \Gamma_i)$. The next result provides a bound that is independent of the solution $\boldsymbol{x}^*$.

Theorem 2 *If $\eta_{ij}, j \in J_i$ are independent and symmetrically distributed random variables in $[-1, 1]$, then*

$$\mathrm{P}\left(\sum_{j\in J_i} \gamma_{ij}\eta_{ij} \geq \Gamma_i\right) \leq \exp\left(-\frac{\Gamma_i^2}{2|J_i|}\right). \tag{2.11}$$

Proof : Let $\theta > 0$. Then

$$\begin{aligned}
\mathrm{P}\left(\sum_{j\in J_i} \gamma_{ij}\eta_{ij} \geq \Gamma_i\right) &\leq \frac{E[\exp(\theta\sum_{j\in J_i} \gamma_{ij}\eta_{ij})]}{\exp(\theta\Gamma_i)} &(2.12)\\
&= \frac{\prod_{j\in J_i} E[\exp(\theta\ \gamma_{ij}\eta_{ij})]}{\exp(\theta\Gamma_i)} &(2.13)\\
&= \frac{\prod_{j\in J_i} 2\int_0^1 \sum_{k=0}^{\infty} \frac{(\theta\gamma_{ij}\eta)^{2k}}{(2k)!} dF_{\eta_{ij}}(\eta)}{\exp(\theta\Gamma_i)} &(2.14)\\
&\leq \frac{\prod_{j\in J_i} \sum_{k=0}^{\infty} \frac{(\theta\gamma_{ij})^{2k}}{(2k)!}}{\exp(\theta\Gamma_i)}
\end{aligned}$$

$$
\begin{aligned}
&\leq \frac{\prod_{j \in J_i} \exp\left(\frac{\theta^2 \gamma_{ij}^2}{2}\right)}{\exp(\theta \Gamma_i)} \\
&\leq \exp\left(|J_i|\frac{\theta^2}{2} - \theta \Gamma_i\right). \qquad (2.15)
\end{aligned}
$$

Inequality (2.12) follows from Markov's inequality, Eqs. (2.13) and (2.14) follow from the independence and symmetric distribution assumption of the random variables η_{ij}. Inequality (2.15) follows from $\gamma_{ij} \leq 1$. Selecting $\theta = \Gamma_i/|J_i|$, we obtain (2.11). ■

Remark : While the bound we established has the attractive feature that is independent of the solution $\boldsymbol{x}^*$, it is not particularly attractive especially when $\frac{\Gamma_i^2}{2|J_i|}$ is small. We next derive the best possible bound, i.e., a bound that is achievable. We assume that $\Gamma_i \geq 1$.

Theorem 3 **(a)** *If $\eta_{ij}, j \in J_i$ are independent and symmetrically distributed random variables in $[-1, 1]$, then*

$$
\mathrm{P}\left(\sum_{j \in J_i} \gamma_{ij}\eta_{ij} \geq \Gamma_i\right) \leq B(n, \Gamma_i), \qquad (2.16)
$$

where

$$
\begin{aligned}
B(n, \Gamma_i) &= \frac{1}{2^n}\left\{(1-\mu)\sum_{l=\lfloor \nu \rfloor}^{n} \binom{n}{l} + \mu \sum_{l=\lfloor \nu \rfloor + 1}^{n} \binom{n}{l}\right\} \\
&= \frac{1}{2^n}\left\{(1-\mu)\binom{n}{\lfloor \nu \rfloor} + \sum_{l=\lfloor \nu \rfloor + 1}^{n} \binom{n}{l}\right\},
\end{aligned} \qquad (2.17)
$$

where $n = |J_i|$, $\nu = \frac{\Gamma_i + n}{2}$ and $\mu = \nu - \lfloor \nu \rfloor$.

(b) *The bound (2.16) is tight for η_{ij} having a discrete probability distribution:* $\mathrm{P}(\eta_{ij} = 1) = 1/2$ *and* $\mathrm{P}(\eta_{ij} = -1) = 1/2$, $\gamma_{ij} = 1$, *an integral value of* $\Gamma_i \geq 1$ *and* $\Gamma_i + n$ *being even.*

(c) *The bound (2.16) satisfies*

$$
B(n, \Gamma_i) \leq (1-\mu)C(n, \lfloor \nu \rfloor) + \sum_{l=\lfloor \nu \rfloor + 1}^{n} C(n, l), \qquad (2.18)
$$

where

$$C(n,l) = \begin{cases} \frac{1}{2^n}, & \text{if } l{=}0 \text{ or } l{=}n, \\ \frac{1}{\sqrt{2\pi}}\sqrt{\frac{n}{(n-l)l}} \exp\left(n \log\left(\frac{n}{2(n-l)}\right) + l \log\left(\frac{n-l}{l}\right)\right), & \text{otherwise.} \end{cases} \tag{2.19}$$

(d) *For* $\Gamma_i = \theta\sqrt{n}$,

$$\lim_{n\to\infty} B(n, \Gamma_i) = 1 - \Phi(\theta), \tag{2.20}$$

where

$$\Phi(\theta) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\theta} \exp\left(-\frac{y^2}{2}\right) dy$$

is the cumulative distribution function of a standard normal.

Proof : (a) The proof follows from Proposition 2 parts (a) and (b). To simplify the exposition we will drop the subscript i, which represents the index of the constraint. We prove the bound in (2.16) by induction on n. We define the auxiliary quantities:

$$\nu(\Gamma, n) = \frac{\Gamma + n}{2}, \qquad \mu(\Gamma, n) = \nu(\Gamma, n) - \lfloor \nu(\Gamma, n) \rfloor, \qquad \Upsilon(s, n) = \frac{1}{2^n} \sum_{l=s}^{n} \binom{n}{l}.$$

The induction hypothesis is formulated as follows:

$$\begin{aligned} &\mathrm{P}\left(\textstyle\sum_{j=1}^{n} \gamma_j \eta_j \geq \Gamma\right) \\ \leq &\begin{cases} (1 - \mu(\Gamma, n))\Upsilon(\lfloor \nu(\Gamma, n) \rfloor, n) + \mu(\Gamma, n)\Upsilon(\lfloor \nu(\Gamma, n) \rfloor + 1, n) & \text{if } \Gamma \in [1, n] \\ 0 & \text{if } \Gamma > n. \end{cases} \end{aligned}$$

For $n = 1$, then $\Gamma = 1$, and so $\nu(1,1) = 1, \mu(1,1) = 0$, $\Upsilon(1,1) = 1/2$ leading to:

$$\begin{aligned} \mathrm{P}(\eta_1 \geq \Gamma) &\leq \mathrm{P}(\eta_1 \geq 0) \\ &\leq \frac{1}{2} \\ &= (1 - \mu(1,1))\Upsilon(\lfloor \nu(1,1) \rfloor, 1) + \mu(1,1)\Upsilon(\lfloor \nu(1,1) \rfloor + 1, 1). \end{aligned}$$

Assuming the induction hypothesis holds for n, we have

$$
\begin{aligned}
& \mathrm{P}\left(\sum_{j=1}^{n+1} \gamma_j \eta_j \geq \Gamma\right) && (2.21)\\
= & \int_{-1}^{1} \mathrm{P}\left(\sum_{j=1}^{n} \gamma_j \eta_j \geq \Gamma - \gamma_{n+1}\eta_{n+1} | \eta_{n+1} = \eta\right) dF_{\eta_{n+1}}(\eta) \\
= & \int_{-1}^{1} \mathrm{P}\left(\sum_{j=1}^{n} \gamma_j \eta_j \geq \Gamma - \gamma_{n+1}\eta\right) dF_{\eta_{n+1}}(\eta) && (2.22)\\
= & \int_{0}^{1} \left[\mathrm{P}\left(\sum_{j=1}^{n} \gamma_j \eta_j \geq \Gamma - \gamma_{n+1}\eta\right)\right. \\
& \left. + \mathrm{P}\left(\sum_{j=1}^{n} \gamma_j \eta_j \geq \Gamma + \gamma_{n+1}\eta\right)\right] dF_{\eta_{n+1}}(\eta) && (2.23)\\
\leq & \max_{\phi \in [0, \gamma_{n+1}]} \left\{\mathrm{P}\left(\sum_{j=1}^{n} \gamma_j \eta_j \geq \Gamma - \phi\right)\right. \\
& \left. + \mathrm{P}\left(\sum_{j=1}^{n} \gamma_j \eta_j \geq \Gamma + \phi\right)\right\} \int_{0}^{1} dF_{\eta_{n+1}}(\eta) \\
= & \frac{1}{2} \max_{\phi \in [0, \gamma_{n+1}]} \left\{\mathrm{P}\left(\sum_{j=1}^{n} \gamma_j \eta_j \geq \Gamma - \phi\right) + \mathrm{P}\left(\sum_{j=1}^{n} \gamma_j \eta_j \geq \Gamma + \phi\right)\right\} \\
\leq & \frac{1}{2} \max_{\phi \in [0, \gamma_{n+1}]} \Psi_n(\phi) && (2.24)\\
\leq & \frac{1}{2} \Psi_n(1) && (2.25)\\
= & (1 - \mu(\Gamma, n+1)) \Upsilon(\lfloor \nu(\Gamma, n+1) \rfloor, n+1) + \\
& \mu(\Gamma, n+1) \Upsilon(\lfloor \nu(\Gamma, n+1) \rfloor + 1, n+1), && (2.26)
\end{aligned}
$$

where

$$
\Psi_n(\phi) = (1 - \mu(\Gamma - \phi, n)\Upsilon(\lfloor \nu(\Gamma - \phi, n) \rfloor, n) + \mu(\Gamma - \phi, n)\Upsilon(\lfloor \nu(\Gamma - \phi, n) \rfloor + 1, n)
$$

$$
+ (1 - \mu(\Gamma + \phi, n))\Upsilon(\lfloor \nu(\Gamma + \phi, n) \rfloor, n) + \mu(\Gamma + \phi, n)\Upsilon(\lfloor \nu(\Gamma + \phi, n) \rfloor + 1, n).
$$

Eqs. (2.22) and (2.23) follow from the assumption that η_j's are independent, symmetrically distributed random variables in $[-1, 1]$. Inequality (2.24) represents the induction hypothesis. Eq. (2.26) follows from:

$$
\begin{aligned}
\Psi_n(1) &= (1-\mu(\Gamma-1,n))\Upsilon(\lfloor\nu(\Gamma-1,n)\rfloor,n) \\
&\quad +\mu(\Gamma-1,n)\Upsilon(\lfloor\nu(\Gamma-1,n)\rfloor+1,n) \\
&\quad +(1-\mu(\Gamma+1,n))\Upsilon(\lfloor\nu(\Gamma+1,n)\rfloor,n) \\
&\quad +\mu(\Gamma+1,n)\Upsilon(\lfloor\nu(\Gamma+1,n)\rfloor+1,n) \qquad (2.27) \\
&= (1-\mu(\Gamma+1,n))(\Upsilon(\lfloor\nu(\Gamma+1,n)\rfloor-1,n) \\
&\quad +\Upsilon(\lfloor\nu(\Gamma+1,n)\rfloor,n))+ \\
&\quad \mu(\Gamma+1,n)(\Upsilon(\lfloor\nu(\Gamma+1,n)\rfloor,n) \\
&\quad +\Upsilon(\lfloor\nu(\Gamma+1,n)\rfloor+1,n)) \qquad (2.28) \\
&= 2\{(1-\mu(\Gamma+1,n))\Upsilon(\lfloor\nu(\Gamma+1,n)\rfloor,n+1)+ \\
&\quad \mu(\Gamma+1,n)\Upsilon(\lfloor\nu(\Gamma+1,n)\rfloor+1,n+1)\} \qquad (2.29) \\
&= 2\{(1-\mu(\Gamma,n+1))\Upsilon(\lfloor\nu(\Gamma,n+1)\rfloor,n+1)+ \\
&\quad \mu(\Gamma,n+1)\Upsilon(\lfloor\nu(\Gamma,n+1)\rfloor+1,n+1)\} \qquad (2.30)
\end{aligned}
$$

Eqs. (2.27) and (2.28) follow from noting that $\mu(\Gamma-1,n)=\mu(\Gamma+1,n)$ and $\lfloor\nu(\Gamma-1,n)\rfloor=\lfloor\nu(\Gamma+1,n)\rfloor-1$. Eq. (2.29) follows from the claim that $\Upsilon(s,n)+\Upsilon(s+1,n)=2\Upsilon(s+1,n+1)$, which is presented next:

$$
\begin{aligned}
\Upsilon(s,n)+\Upsilon(s+1,n) &= \frac{1}{2^n}\left\{\sum_{l=s}^{n}\binom{n}{l}+\sum_{l=s+1}^{n}\binom{n}{l}\right\} \\
&= \frac{1}{2^n}\left(\sum_{l=s}^{n-1}\left[\binom{n}{l}+\binom{n}{l+1}\right]+1\right) \\
&= \frac{1}{2^n}\left(\sum_{l=s}^{n-1}\binom{n+1}{l+1}+1\right) \\
&= \frac{1}{2^n}\sum_{l=s+1}^{n+1}\binom{n+1}{l} \\
&= 2\Upsilon(s+1,n+1),
\end{aligned}
$$

and Eq. (2.30) follows from $\mu(\Gamma+1,n)=\mu(\Gamma,n+1)=(\Gamma+n+1)/2$.

We are left to show that $\Psi_n(\phi)$ is a monotonically non-decreasing function in the domain $\phi \in [0,1]$, which implies that for any $\phi_1, \phi_2 \in [0,1]$ such that $\phi_1 > \phi_2$, $\Psi_n(\phi_1) - \Psi_n(\phi_2) \geq 0$. We fix Γ and n. To simplify the notation we use: $\mu(\phi) = \mu(\Gamma+\phi, n) = (\Gamma+\phi+n)/2$, $\nu(\phi) = \nu(\Gamma+\phi, n)$. For any choice of ϕ_1 and ϕ_2, we have $\rho = \lfloor \nu_{-\phi_1} \rfloor \leq \lfloor \nu_{-\phi_2} \rfloor \leq \lfloor \nu_{\phi_2} \rfloor \leq \lfloor \nu_{\phi_1} \rfloor \leq \rho+1$. Therefore, we consider the following cases:

For $\rho = \lfloor \nu_{-\phi_1} \rfloor = \lfloor \nu_{-\phi_2} \rfloor = \lfloor \nu_{\phi_2} \rfloor = \lfloor \nu_{\phi_1} \rfloor$,

$$\begin{aligned}
\mu_{-\phi_1} - \mu_{-\phi_2} &= -\frac{\phi_1 - \phi_2}{2} \\
\mu_{\phi_1} - \mu_{\phi_2} &= \frac{\phi_1 - \phi_2}{2} \\
\Psi_n(\phi_1) - \Psi_n(\phi_2) &= \frac{\phi_1 - \phi_2}{2}\{\Upsilon(\rho, n) - \Upsilon(\rho+1, n) - \Upsilon(\rho, n) + \Upsilon(\rho+1, n)\} \\
&= 0.
\end{aligned}$$

For $\rho = \lfloor \nu_{-\phi_1} \rfloor = \lfloor \nu_{-\phi_2} \rfloor = \lfloor \nu_{\phi_2} \rfloor - 1 = \lfloor \nu_{\phi_1} \rfloor - 1$,

$$\begin{aligned}
\mu_{-\phi_1} - \mu_{-\phi_2} &= -\frac{\phi_1 - \phi_2}{2} \\
\mu_{\phi_1} - \mu_{\phi_2} &= \frac{\phi_1 - \phi_2}{2} \\
\Psi_n(\phi_1) - \Psi_n(\phi_2) &= \frac{\phi_1 - \phi_2}{2}\{\Upsilon(\rho, n) - 2\Upsilon(\rho+1, n) + \Upsilon(\rho+2, n)\} \\
&= \frac{\phi_1 - \phi_2}{2^{n+1}(n+1)}\binom{n+1}{\rho+1}(1 + 2\rho - n) \\
&\geq \frac{\phi_1 - \phi_2}{2^{n+1}(n+1)}\binom{n+1}{\rho+1}\left(1 + 2\left\lfloor \frac{1+n-\phi_1}{2} \right\rfloor - n\right) \\
&\geq 0.
\end{aligned}$$

For $\rho = \lfloor \nu_{-\phi_1} \rfloor = \lfloor \nu_{-\phi_2} \rfloor - 1 = \lfloor \nu_{\phi_2} \rfloor - 1 = \lfloor \nu_{\phi_1} \rfloor - 1$,

$$\begin{aligned}
\mu_{-\phi_1} - \mu_{-\phi_2} &= -\frac{\phi_1 - \phi_2}{2} + 1 \\
\mu_{\phi_1} - \mu_{\phi_2} &= \frac{\phi_1 - \phi_2}{2} \\
\Psi_n(\phi_1) - \Psi_n(\phi_2) &= (1 - \mu_{-\phi_1})\{\Upsilon(\rho, n) - 2\Upsilon(\rho+1, n) + \Upsilon(\rho+2, n)\} \\
&\geq 0.
\end{aligned}$$

For $\rho = \lfloor \nu_{-\phi_1} \rfloor = \lfloor \nu_{-\phi_2} \rfloor = \lfloor \nu_{\phi_2} \rfloor = \lfloor \nu_{\phi_1} \rfloor - 1$,

$$\begin{aligned}
\mu_{-\phi_1} - \mu_{-\phi_2} &= -\frac{\phi_1 - \phi_2}{2} \\
\mu_{\phi_1} - \mu_{\phi_2} &= \frac{\phi_1 - \phi_2}{2} - 1 \\
\Psi_n(\phi_1) - \Psi_n(\phi_2) &= \mu_{\phi_1}\{\Upsilon(\rho, n) - 2\Upsilon(\rho + 1, n) + \Upsilon(\rho + 2, n)\} \\
&\geq 0.
\end{aligned}$$

(b) Let η_j obey a discrete probability distribution $\mathrm{P}(\eta_j = 1) = 1/2$ and $\mathrm{P}(\eta_j = -1) = 1/2$, $\gamma_j = 1$, $\Gamma \geq 1$ is integral and $\Gamma + n$ is even. Let S_n obeys a Binomial distribution with parameters n and $1/2$. Then,

$$\begin{aligned}
\mathrm{P}\left(\sum_{j=1}^{n} \eta_j \geq \Gamma\right) &= \mathrm{P}(S_n - (n - S_n) \geq \Gamma) \\
&= \mathrm{P}(2S_n - n \geq \Gamma) \\
&= \mathrm{P}(S_n \geq \frac{n + \Gamma}{2}) \\
&= \frac{1}{2^n} \sum_{l=\frac{n+\Gamma}{2}}^{n} \binom{n}{l}, \qquad (2.31)
\end{aligned}$$

which implies that the bound (2.16) is indeed tight.

(c) From Eq. (2.17), we need to find an upper bound for the function $\frac{1}{2^n}\binom{n}{l}$. From Stirling's formula (see Robbins [24]) we obtain for $n \geq 1$,

$$\sqrt{2\pi} n^{n+1/2} \exp(-n + 1/(12n + 1)) \leq n! \leq \sqrt{2\pi} n^{n+1/2} \exp(-n + 1/(12n)),$$

we can establish for $l \in \{1, \ldots, n - 1\}$,

$$\begin{aligned}
\frac{1}{2^n}\binom{n}{l} &= \frac{n!}{2^n (n-l)! l!} \\
&\leq \frac{1}{\sqrt{2\pi}} \sqrt{\frac{n}{(n-l)l}} \exp\left(\frac{1}{12n} - \frac{1}{12(n-l)+1} - \frac{1}{12l+1}\right) \times \\
&\quad \left(\frac{n}{2(n-l)}\right)^n \left(\frac{n-l}{l}\right)^l
\end{aligned}$$

$$\leq \frac{1}{\sqrt{2\pi}}\sqrt{\frac{n}{(n-l)l}}\left(\frac{n}{2(n-l)}\right)^n\left(\frac{n-l}{l}\right)^l \tag{2.32}$$
$$= \frac{1}{\sqrt{2\pi}}\sqrt{\frac{n}{(n-l)l}}\exp\left(n\log\left(\frac{n}{2(n-l)}\right)+l\log\left(\frac{n-l}{l}\right)\right),$$

where Eq. (2.32) follows from

$$\frac{1}{12(n-l)+1}+\frac{1}{12l+1}\geq\frac{2}{(12(n-l)+1)+(12l+1)}=\frac{2}{12n+2}>\frac{1}{12n}.$$

For $l=0$ and $l=n$ $\frac{1}{2^n}\binom{n}{l}=\frac{1}{2^n}$.

(d) Bound (2.16) can be written as

$$B(n,\Gamma_i)=(1-\mu)\mathrm{P}(S_n\geq\lfloor\nu\rfloor)+\mu\mathrm{P}(S_n\geq\lfloor\nu\rfloor+1),$$

where S_n represents a Binomial distribution with parameters n and $1/2$. Since $\mathrm{P}(S_n\geq\lfloor\nu\rfloor+1)\leq\mathrm{P}(S_n\geq\lfloor\nu\rfloor)$, we have

$$\mathrm{P}(S_n\geq\nu+1)=\mathrm{P}(S_n\geq\lfloor\nu\rfloor+1)\leq B(n,\Gamma_i)\leq\mathrm{P}(S_n\geq\lfloor\nu\rfloor)=\mathrm{P}(S_n\geq\nu),$$

since S_n is a discrete distribution. For $\Gamma_i=\theta\sqrt{n}$, where θ is a constant, we have

$$\mathrm{P}\left(\frac{S_n-n/2}{\sqrt{n}/2}\geq\theta+\frac{2}{\sqrt{n}}\right)\leq B(n,\Gamma_i)\leq\mathrm{P}\left(\frac{S_n-n/2}{\sqrt{n}/2}\geq\theta\right).$$

By the central limit theorem, we obtain that

$$\lim_{n\to\infty}\mathrm{P}\left(\frac{S_n-n/2}{\sqrt{n}/2}\geq\theta+\frac{2}{\sqrt{n}}\right)=\lim_{n\to\infty}\mathrm{P}\left(\frac{S_n-n/2}{\sqrt{n}/2}\geq\theta\right)=1-\Phi(\theta),$$

where $\Phi(\theta)$ is the cumulative distribution function of a standard normal. Thus, for $\Gamma_i=\theta\sqrt{n}$, we have

$$\lim_{n\to\infty}B(n,\Gamma_i)=1-\Phi(\theta).$$

■

Remarks:

(a) While Bound (2.16) is best possible (Theorem 3(b)), it poses computational difficulties in evaluating the sum of combination functions for large n. For this reason, we have calculated Bound (2.18), which is simple to compute and, as we will see, it is also very tight.

(b) Eq. (2.20) is a formal asymptotic theorem that applies when $\Gamma_i = \theta\sqrt{n}$. We can use the De Moivre-Laplace approximation of the Binomial distribution to obtain the approximation

$$B(n, \Gamma_i) \approx 1 - \Phi\left(\frac{\Gamma_i - 1}{\sqrt{n}}\right), \tag{2.33}$$

that applies, even when Γ_i does not scale as $\theta\sqrt{n}$.

(c) We next compare the bounds: (2.11) (Bound 1), (2.16) (Bound 2), (2.18) (Bound 3) and the approximate bound (2.33) for $n = |J_i| = 10, 100, 2000$. In Figure 2-1 we compare Bounds 1 and 2 for $n = 10$ that clearly show that Bound 2 dominates Bound 1 (in this case there is no need to calculate Bounds 3 and the approximate bound as n is small). In Figure 2-2 we compare all bounds for $n = 100$. It is clear that Bound 3, which is simple to compute, is identical to Bound 2, and both Bounds 2 and 3 dominate Bound 1 by an order of magnitude. The approximate bound provides a reasonable approximation to Bound 2. In Figure 2-3 we compare Bounds 1 and 3 and the approximate bound for $n = 2000$. Bound 3 is identical to the approximate bound, and both dominate Bound 1 by an order of magnitude. In summary, in the remainder of the chapter, we will use Bound 3, as it is simple to compute, it is a true bound (as opposed to the approximate bound), and dominates Bound 1. To amplify this point, Table 2.1 illustrates the choice of Γ_i as a function of $n = |J_i|$ so that the probability that a constraint is violated is less than 1%, where we used Bounds 1, 2, 3 and the approximate bound to evaluate the probability. It is clear that using Bounds 2,3 or the approximate bound gives essentially identical values of Γ_i, while using Bound 1 leads to unnecessarily higher values of Γ_i. For $|J_i| = 200$, we need to use $\Gamma = 33.9$, i.e., only 17% of the number of uncertain data, to guarantee violation probability of less than 1%. For constraints with fewer number of uncertain data such as $|J_i| = 5$, it is necessary to ensure full protection, which is equivalent to the Soyster's method.

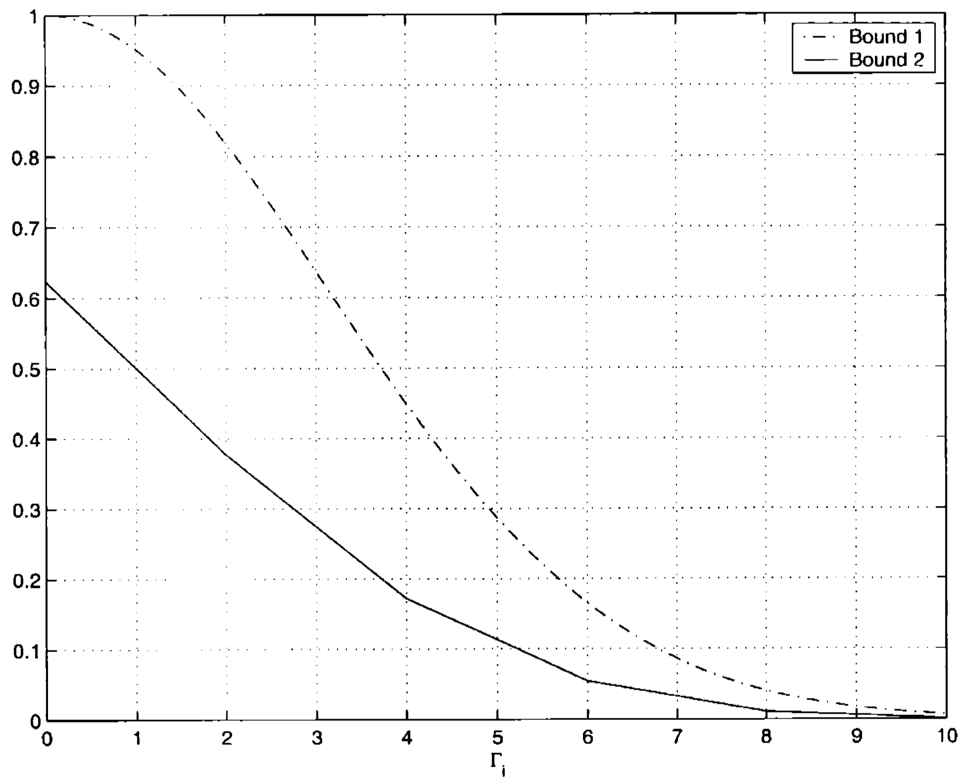

Figure 2-1: Comparison of probability bounds for $n = |J_i| = 10$.

Clearly, for constraints with large number of uncertain data, the proposed approach is capable of delivering less conservative solutions compared to the Soyster's method.

2.3.1 On the Conservatism of Robust Solutions

We have argued so far that the linear optimization framework of our approach has some computational advantages over the conic quadratic framework of Ben-Tal and Nemirovski [7, 6, 4] and El-Ghaoui et al. [11, 12] especially with respect to discrete optimization problems. Our objective in this section is to provide some insight, but not conclusive evidence, on the degree of conservatism for both approaches.

Given a constraint $\boldsymbol{a}'\boldsymbol{x} \leq b$, with $\boldsymbol{a} \in [\overline{\boldsymbol{a}} - \hat{\boldsymbol{a}}, \overline{\boldsymbol{a}} + \hat{\boldsymbol{a}}]$, the robust counterpart of Ben-Tal and Nemirovski [7, 6, 4] and El-Ghaoui et al. [11, 12] in its simplest form of ellipsoidal uncertainty (Formulation (2.2) includes combined interval and ellipsoidal

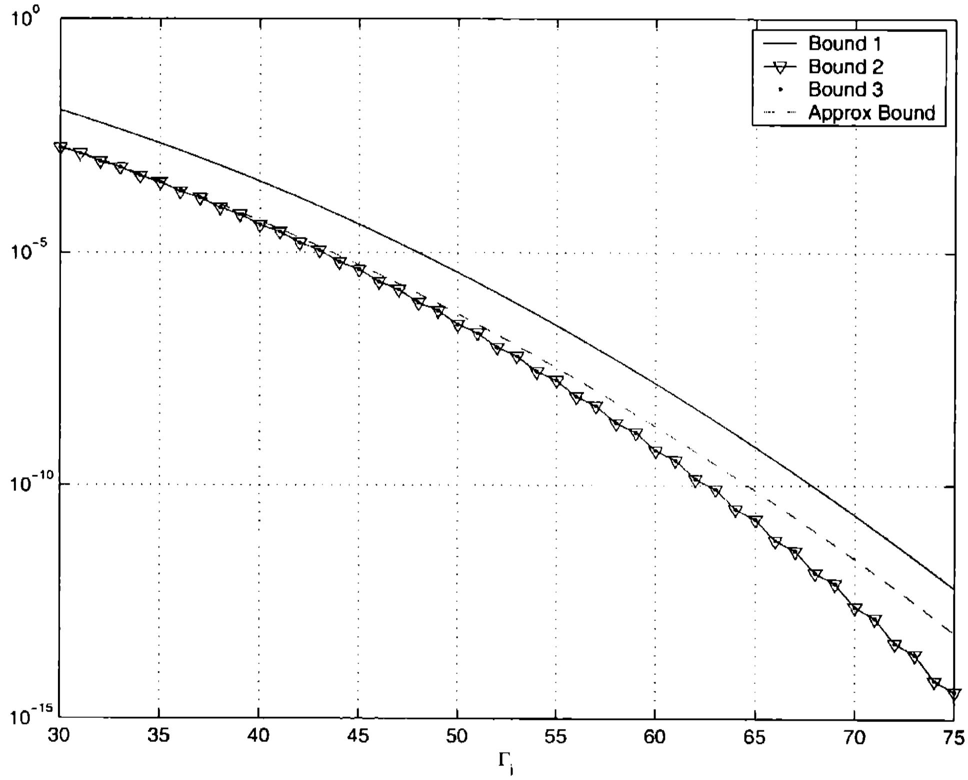

Figure 2-2: Comparison of probability bounds for $n = |J_i| = 100$.

$\|J_i\|$	Γ_i **from Bound 1**	Γ_i **from Bounds 2, 3**	Γ_i **from Approx.**
5	5	5	5
10	9.6	8.2	8.4
100	30.3	24.3	24.3
200	42.9	33.9	33.9
2000	135.7	105	105

Table 2.1: Choice of Γ_i as a function of $n = |J_i|$ so that the probability of constraint violation is less than 1%.

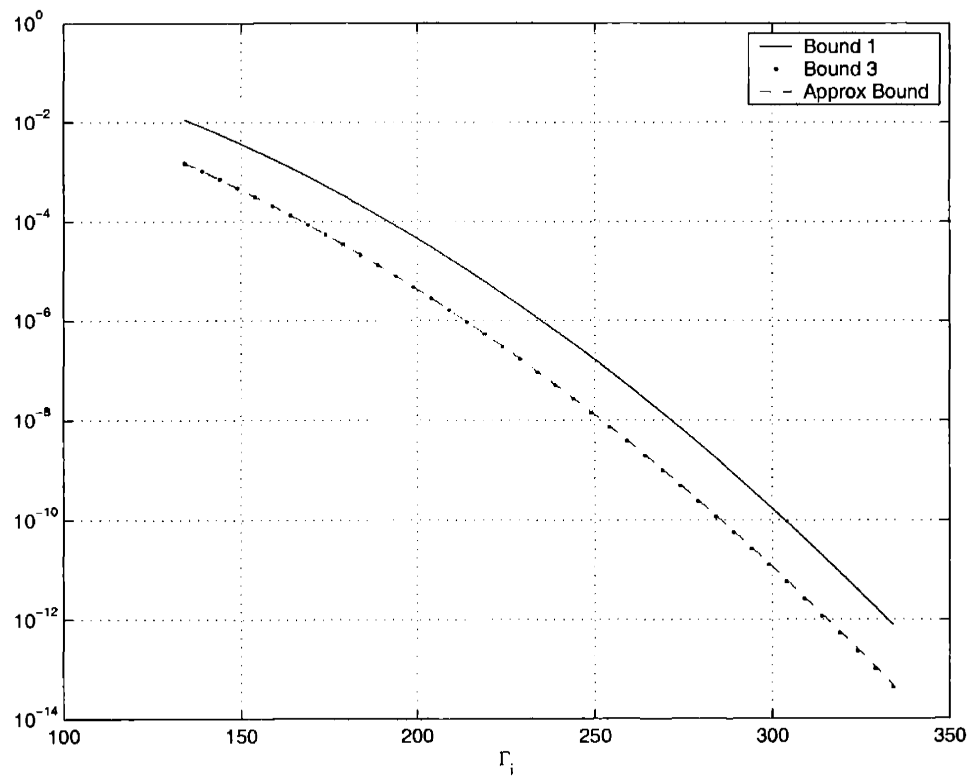

Figure 2-3: Comparison of probability bounds for $n = |J_i| = 2000$.

uncertainty) is:

$$\overline{\boldsymbol{a}}'\boldsymbol{x} + \Omega||\hat{\boldsymbol{A}}\boldsymbol{x}||_2 \leq b,$$

where $\hat{\boldsymbol{A}}$ is a diagonal matrix with elements $\hat{a}_i$ in the diagonal. Ben-Tal and Nemirovski [7] show that under the model of data uncertainty U for $\boldsymbol{a}$, the probability that the constraint is violated is bounded above by $\exp(-\Omega^2/2)$.

The robust counterpart of the current approach is

$$\overline{\boldsymbol{a}}'\boldsymbol{x} + \beta(\boldsymbol{x}, \Gamma) \leq b,$$

where we assumed that Γ is integral and

$$\beta(\boldsymbol{x}, \Gamma) = \max_{S,\ |S|=\Gamma} \sum_{i \in S} \hat{a}_i |x_i|.$$

From Eq. (2.11), the probability that the constraint is violated under the model of data uncertainty U for $\boldsymbol{a}$ is bounded above by $\exp(-\Gamma^2/(2n))$. Note that we do not use the stronger bound (2.16) for simplicity.

Let us select $\Gamma = \Omega\sqrt{n}$ so that the bounds for the probability of violation are the same for both approaches. The protection levels are $\Omega||\hat{\boldsymbol{A}}\boldsymbol{x}||_2$ and $\beta(\boldsymbol{x}, \Gamma)$. We will compare the protection levels both from a worst and an average case point of view in order to obtain some insight on the degree of conservatism. To simplify the exposition we define $y_i = \hat{a}_i|x_i|$. We also assume without loss of generality that $y_1 \geq y_2 \geq \ldots \geq y_n \geq 0$. Then the two protection levels become $\Omega||\boldsymbol{y}||_2$ and $\sum_{i=1}^{\Gamma} y_i$.

For $\Gamma = \theta\sqrt{n}$, and $y_1^0 = \ldots = y_\Gamma^0 = 1$, $y_k^0 = 0$ for $k \geq \Gamma + 1$, we have $\sum_{i=1}^{\Gamma} y_i^0 = \Gamma = \theta\sqrt{n}$, while $\Omega||\boldsymbol{y}||_2 = \theta^{3/2}n^{1/4}$, i.e., in this example the protection level of the conic quadratic framework is asymptotically smaller than our framework by a multiplicative factor of $n^{1/4}$. This order of the magnitude is in fact worst possible, since

$$\sum_{i=1}^{\Gamma} y_i \leq \sqrt{\Gamma}||\boldsymbol{y}||_2 = \sqrt{\frac{n}{\Gamma}}(\Omega||\boldsymbol{y}||_2),$$

which for $\Gamma = \theta\sqrt{n}$ leads to

$$\sum_{i=1}^{\Gamma} y_i \le \frac{n^{1/4}}{\theta^{1/2}} (\Omega ||\boldsymbol{y}||_2).$$

Moreover, we have

$$\begin{aligned}
\Omega ||\boldsymbol{y}||_2 &\le \Omega \sqrt{\sum_{i=1}^{\Gamma} y_i^2 + y_\Gamma^2 (n - \Gamma)} \\
&\le \Omega \sqrt{\sum_{i=1}^{\Gamma} y_i^2 + \left(\frac{\sum_{i=1}^{\Gamma} y_i}{\Gamma}\right)^2 (n - \Gamma)} \\
&\le \Omega \sqrt{\left(\sum_{i=1}^{\Gamma} y_i\right)^2 + \left(\sum_{i=1}^{\Gamma} y_i\right)^2 \left(\frac{n - \Gamma}{\Gamma^2}\right)} \\
&= \frac{\Gamma}{\sqrt{n}} \sum_{i=1}^{\Gamma} y_i \sqrt{1 + \frac{n - \Gamma}{\Gamma^2}} \\
&= \sqrt{\frac{\Gamma^2 + n - \Gamma}{n}} \sum_{i=1}^{\Gamma} y_i.
\end{aligned}$$

If we select $\Gamma = \theta\sqrt{n}$, which makes the probability of violation $\exp(-\theta^2/2)$, we obtain that

$$\Omega ||\boldsymbol{y}||_2 \le \sqrt{1 + \theta^2} \sum_{i=1}^{\Gamma} y_i.$$

Thus, in the worst case the protection level of our framework can only be smaller than the conic quadratic framework by a multiplicative factor of a constant. We conclude that in the worst case, the protection level for the conic quadratic framework can be smaller than our framework by a factor of $n^{1/4}$, while the protection of our framework can be smaller than the conic quadratic framework by at most a constant.

Let us compare the protection levels on average, however. In order to obtain some insight let us assume that y_i are independently and uniformly distributed in $[0, 1]$. Simple calculations show that for the case in question ($\Omega = \Gamma/\sqrt{n}$, $\Gamma = \theta\sqrt{n}$)

$$E[\Omega ||\boldsymbol{y}||_2] = \Theta(\sqrt{n}), \qquad E\left[\max_{S, |S| = \Gamma} \sum_{i \in S} y_i\right] = \Theta(\sqrt{n}),$$

which implies that on average the two protection levels are of the same order of magnitude.

It is admittedly unclear whether it is the worst or the average case we presented which is more relevant, and thus the previous discussion is inconclusive. It is fair to say, however, that both approaches allow control of the degree of conservatism by adjusting the parameters Γ and Ω. Moreover, we think that the ultimate criterion for comparing the degree of conservatism of these methods will be computation in real problems.

2.3.2 Local Sensitivity Analysis of the Protection Level

Given the solution of Problem (2.6), it is desirable to estimate the change in the objective function value with respect to the change of the protection level Γ_i. In this way we can assess the price of increasing or decreasing the protection level Γ_i of any constraint. Note that when Γ_i changes, only one parameter in the coefficient matrix in Problem (2.6) changes. Thus, we can use results from sensitivity analysis (see Freund [15] for a comprehensive analysis) to understand the effect of changing the protection level under nondegeneracy assumptions.

Theorem 4 *Let $\boldsymbol{z}^*$ and $\boldsymbol{q}^*$ be the optimal nondegenerate primal and dual solutions for the linear optimization problem (2.6) (under nondegeneracy, the primal and dual optimal solutions are unique). Then, the derivative of the objective function value with respect to protection level Γ_i of the ith constraint is*

$$-z_i^* q_i^* \tag{2.34}$$

where z_i^* is the optimal primal variable corresponding to the protection level Γ_i and q_i^* is the optimal dual variable of the ith constraint.

Proof : We transform Problem (2.6) in standard form,

$$\begin{aligned} G(\Gamma_i) = \text{maximize} \quad & \boldsymbol{c}'\boldsymbol{x} \\ \text{subject to} \quad & \boldsymbol{A}\boldsymbol{x} + \Gamma_i z_i \boldsymbol{e_i} = \boldsymbol{b} \\ & \boldsymbol{x} \geq \boldsymbol{0}, \end{aligned}$$

where e_i is a unit vector with an one in the ith position. Let B be the optimal basis, which is unique under primal and dual nondegeneracy. If the column $\Gamma_i e_i$ corresponding to the variable z_i is not in the basis, then $z_i^* = 0$. In this case, under dual nondegeneracy all reduced costs associated with the nonbasic variables are strictly negative, and thus a marginal change in the protection level does not affect the objective function value. Eq. (2.34) correctly indicates the zero variation.

If the column $\Gamma_i e_i$ corresponding to the variable z_i is in the basis, and the protection level Γ_i changes by $\Delta\Gamma_i$, then B becomes $B + \Delta\Gamma_i e_i e_i'$. By the *Matrix Inversion Lemma* we have:

$$(B + \Delta\Gamma_i e_i e_i')^{-1} = B^{-1} - \frac{\Delta\Gamma_i B^{-1} e_i e_i' B^{-1}}{1 + \Delta\Gamma_i e_i' B^{-1} e_i}.$$

Under primal and dual nondegeneracy, for small changes $\Delta\Gamma_i$, the new solutions preserve primal and dual feasibility. Therefore, the corresponding change in the objective function value is,

$$G(\Gamma_i + \Delta\Gamma_i) - G(\Gamma_i) = -\frac{\Delta\Gamma_i c_B' B^{-1} e_i e_i' B^{-1} b}{1 + \Delta\Gamma_i e_i' B^{-1} e_i} = -\frac{\Delta\Gamma_i z_i^* q_i^*}{1 + \Delta\Gamma_i e_i' B^{-1} e_i},$$

where c_B is the part of the vector $\mathbf{c}$ corresponding to the columns in B. Thus,

$$G'(\Gamma_i) = \lim_{\Delta\Gamma_i \to 0} \frac{G(\Gamma_i + \Delta\Gamma_i) - G(\Gamma_i)}{\Delta\Gamma_i} = -z_i^* q_i^*.$$

■

Remark : An attractive aspect of Eq. (2.34) is its simplicity as it only involves only the primal optimal solution corresponding to the protection level Γ_i and the dual optimal solution corresponding to the ith constraint.

2.4 Correlated Data

So far we assumed that the data are independently uncertain. It is possible, however, that the data are correlated. In particular, we envision that there are few sources of data uncertainty that affect all the data. More precisely, we assume that the model of data uncertainty is a follows.

Correlated Model of Data Uncertainty C:
Consider a particular row i of the matrix $\boldsymbol{A}$ and let J_i the set of coefficients in row i that are subject to uncertainty. Each entry $a_{ij},\ j \in J_i$ is modeled as

$$\tilde{a}_{ij} = a_{ij} + \sum_{k \in K_i} \tilde{\eta}_{ik} g_{kj}$$

and $\tilde{\eta}_{ik}$ are independent and symmetrically distributed random variables in $[-1, 1]$.

Note that under this model, there are only $|K_i|$ sources of data uncertainty that affect the data in row i. Note that these sources of uncertainty affect all the entries $a_{ij},\ j \in J_i$. For example if $|K_i| = 1$, then all data in a row are affected by a single random variable. For a concrete example, consider a portfolio construction problem, in which returns of various assets are predicted from a regression model. In this case, there are a few sources of uncertainty that affect globally all the assets classes.

Analogously to (2.3), we propose the following robust formulation:

$$\begin{aligned}
\text{maximize} \quad & \boldsymbol{c}'\boldsymbol{x} \\
\text{subject to} \quad & \sum_j a_{ij} x_j + \max_{\{S_i \cup \{t_i\} | S_i \subseteq K_i, |S_i| = \lfloor \Gamma_i \rfloor, t_i \in K_i \setminus S_i\}} \\
& \left\{ \sum_{k \in S_i} \Big| \sum_{j \in J_i} g_{kj} x_j \Big| + (\Gamma_i - \lfloor \Gamma_i \rfloor) \Big| \sum_{j \in J_i} g_{t_i j} x_j \Big| \right\} \le b_i \quad \forall i \qquad (2.35) \\
& \boldsymbol{l} \le \boldsymbol{x} \le \boldsymbol{u},
\end{aligned}$$

which can be written as a linear optimization problem as follows:

$$\begin{aligned}
\text{maximize} \quad & \boldsymbol{c}'\boldsymbol{x} & & \\
\text{subject to} \quad & \sum_j a_{ij}x_j + z_i\Gamma_i + \sum_{k\in K_i} p_{ik} \le b_i & & \forall i \\
& z_i + p_{ik} \ge y_{ik} & & \forall i,\ k \in K_i \\
& -y_{ik} \le \textstyle\sum_{j\in J_i} g_{kj}x_j \le y_{ik} & & \forall i,\ k \in K_i \\
& l_j \le x_j \le u_j & & \forall j \\
& p_{ik}, y_{ik} \ge 0 & & \forall i,\ k \in K_i \\
& z_i \ge 0 & & \forall i.
\end{aligned} \tag{2.36}$$

Analogously to Theorem 3, we can show that the probability that the ith constraint is violated is at most $B(|K_i|, \Gamma_i)$ defined in Eq. (2.16).

2.5 Experimental Results

In this section, we present three experiments illustrating our robust solution to problems with data uncertainty. The first example is a simple portfolio optimization problem from Ben-Tal and Nemirovski [6], which has data uncertainty in the objective function. In the last experiment we apply our method to a problem PILOT4 from the well known Net Lib collection to examine the effectiveness of our approach to real world problems.

2.5.1 A Simple Portfolio Problem

In this section we consider a portfolio construction problem consisting of a set of N stocks ($|N| = n$). Stock i has return $\tilde{p}_i$ which is of course uncertain. The objective is to determine the fraction x_i of wealth invested in stock i, so as to maximize the portfolio value $\sum_{i=1}^{n} \tilde{p}_i x_i$. We model the uncertain return $\tilde{p}_i$ as a random variable that has an arbitrary symmetric distribution in the interval $[p_i - \sigma_i, p_i + \sigma_i]$, where p_i is the expected return and σ_i is a measure of the uncertainty of the return of stock i. We further assume that the returns $\tilde{p}_i$ are independent.

The classical approach in portfolio construction is to use quadratic optimization and solve the following problem:

$$\begin{aligned} \text{maximize} \quad & \sum_{i=1}^{n} p_i x_i - \phi \sum_{i=1}^{n} \sigma_i^2 x_i^2 \\ \text{subject to} \quad & \sum_{i=1}^{n} x_i = 1 \\ & x_i \geq 0, \end{aligned}$$

where we interpret σ_i as the standard deviation of the return for stock i, and ϕ is a parameter that controls the tradeoff between risk and return. Applying our approach, we will solve instead the following problem (which can be reformulated as a linear optimization problem as in Theorem 2.6):

$$\begin{aligned} \text{maximize} \quad & z \\ \text{subject to} \quad & z \leq \sum_{i=1}^{n} p_i x_i - \beta(\boldsymbol{x}, \Gamma) \\ & \sum_{i=1}^{n} x_i = 1 \\ & x_i \geq 0. \end{aligned} \tag{2.37}$$

where

$$\beta(\boldsymbol{x}, \Gamma) = \max_{\{S \cup \{t\} |\ S \subseteq N, |S| = \lfloor \Gamma \rfloor, t \in N \setminus S\}} \left\{ \sum_{j \in S} \sigma_j x_j + (\Gamma - \lfloor \Gamma \rfloor) \sigma_t x_t \right\}$$

In this setting Γ is the protection level of the actual portfolio return in the following sense. Let $\boldsymbol{x}^*$ be an optimal solution of Problem (2.37) and let z^* be the optimal solution value of Problem (2.37). Then, $\boldsymbol{x}^*$ satisfies that $\mathrm{P}(\tilde{\boldsymbol{p}}'\boldsymbol{x}^* < z^*)$ is less than or equal to than the bound in Eq. (2.18). Ben-Tal and Nemirovski [6] consider the same portfolio problem using $n = 150$,

$$p_i = 1.15 + i \frac{0.05}{150}, \qquad \sigma_i = \frac{0.05}{450} \sqrt{2in(n+1)}.$$

Note that in this experiment, stocks with higher returns are also more risky.

Optimization Results

Let $\boldsymbol{x}^*(\Gamma)$ be an optimal solution to Problem (2.37) corresponding to the protection level Γ. A classical measure of risk is the *Standard Deviation*,

$$w(\Gamma) = \sqrt{\sum_{i \in N} \sigma_i^2 (x_i^*(\Gamma))^2}.$$

We first solved Problem (2.37) for various levels of Γ. Figure 2-4 illustrates the performance of the robust solution as a function of the protection level Γ, while Figure 2-5 shows the solution itself for various levels of the protection level. The solution exhibits some interesting "phase transitions" as the protection level increases:

1. For $\Gamma \leq 17$, both the expected return as well as the risk adjusted return (the objective function value) gradually decrease. Starting with $\Gamma = 0$, for which the solution consists of the stock 150 that has the highest expected return, the portfolio becomes gradually more diversified putting more weight on stocks with higher ordinal numbers. This can be seen for example for $\Gamma = 10$ in Figure 2-5.

2. For $17 < \Gamma \leq 41$, the risk adjusted return continues to gradually decrease as the protection level increases, while the expected return is insensitive to the protection level. In this range, $x_i* = \frac{\sum_i (1/\sigma_i)}{\sigma_i}$, i.e., the portfolio is fully diversified.

3. For $\Gamma \geq 41$, there is a sudden phase transition (see Figure 2-4). The portfolio consists of only stock 1, which is the one that has the largest risk adjusted return $p_i - \sigma_i$. This is exactly the solution given by the Soyster method as well. In this range both the expected and the risk adjusted returns are insensitive to Γ.

Simulation Results

To examine the quality of the robust solution, we run 10,000 simulations of random yields and compare robust solutions generated by varying the protection level Γ. As

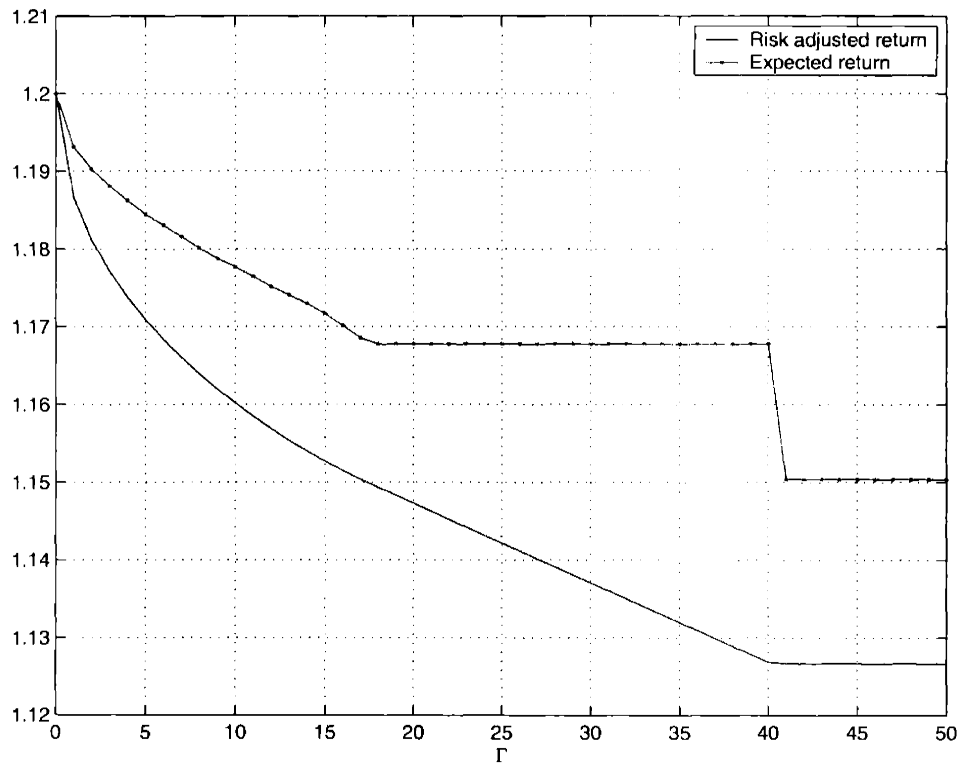

Figure 2-4: The return and the objective function value (risk adjusted return) as a function of the protection level Γ.

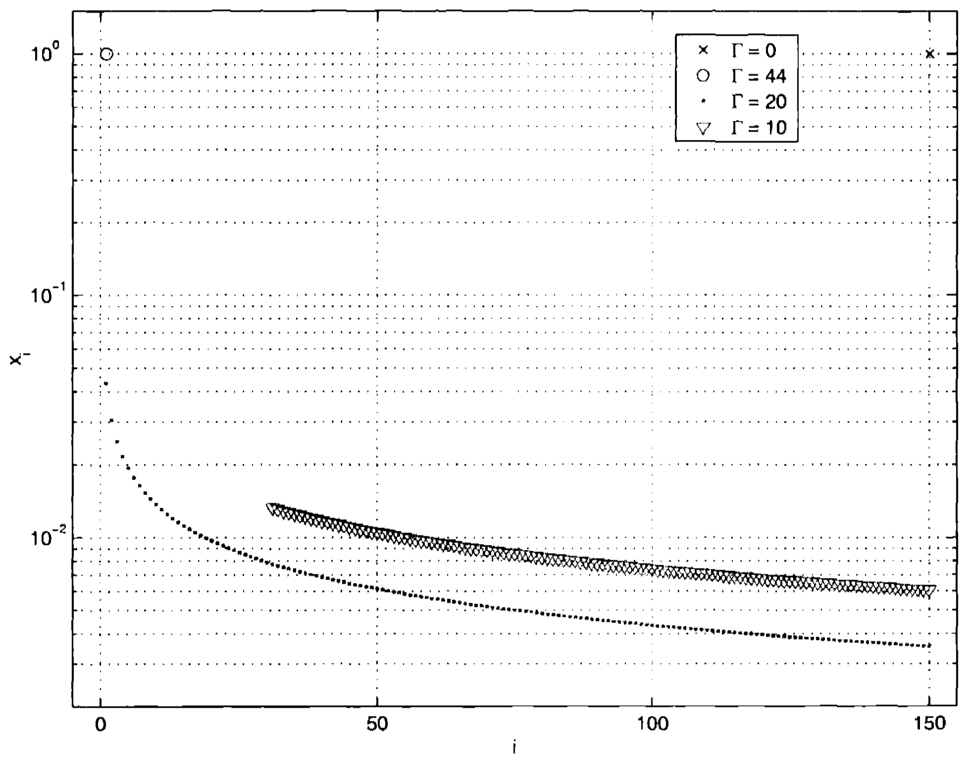

Figure 2-5: The solution of the portfolio for various protection levels.

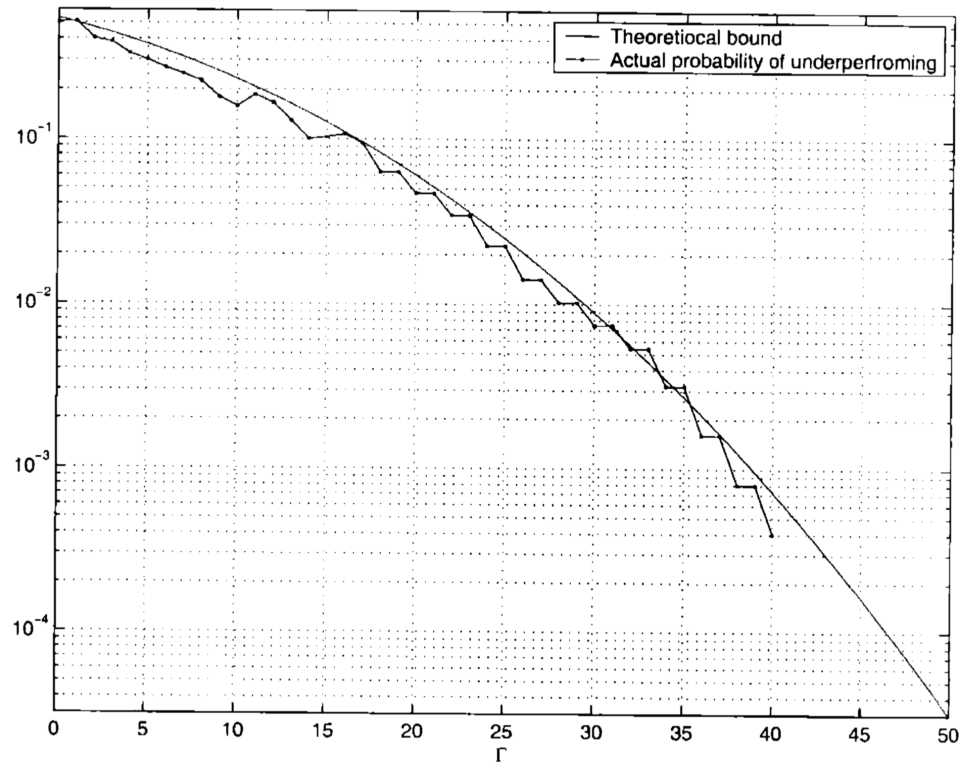

Figure 2-6: Simulation study of the probability of underperforming the nominal return as a function of Γ.

we have discussed, for the worst case simulation, we consider the distribution with $\tilde{p}_i$ taking with probability 1/2 the values at $p_i \pm \sigma_i$. In Figure 2-6, we compare the theoretical bound in Eq. (2.18) with the fraction of the simulated portfolio returns falling below the optimal solution, z^*. The empirical results suggest that the theoretical bound is close to the empirically observed values.

In Table 2.2, we present the results of the simulation indicating the tradeoff between risk and return. The corresponding plots are also presented in Figures 2-7 and 2-8. As expected as the protection level increases, the expected and maximum returns decrease, while the minimum returns increase. For instance, with $\Gamma \geq 15$, the minimum return is maintained above 12% for all simulated portfolios.

This example suggests that our approach captures the tradeoff between risk and return, very much like the mean variance approach, but does so in a linear framework. Additionally the robust approach provides both a deterministic guarantee about the

Γ	Prob. violation	Exp. Return	Min. Return	Max. Return	$w(\Gamma)$
0	0.5325	1.200	0.911	1.489	0.289
5	0.3720	1.184	1.093	1.287	0.025
10	0.2312	1.178	1.108	1.262	0.019
15	0.1265	1.172	1.121	1.238	0.015
20	0.0604	1.168	1.125	1.223	0.013
25	0.0250	1.168	1.125	1.223	0.013
30	0.0089	1.168	1.125	1.223	0.013
35	0.0028	1.168	1.125	1.223	0.013
40	0.0007	1.168	1.125	1.223	0.013
45	0.0002	1.150	1.127	1.174	0.024

Table 2.2: Simulation results given by the robust solution.

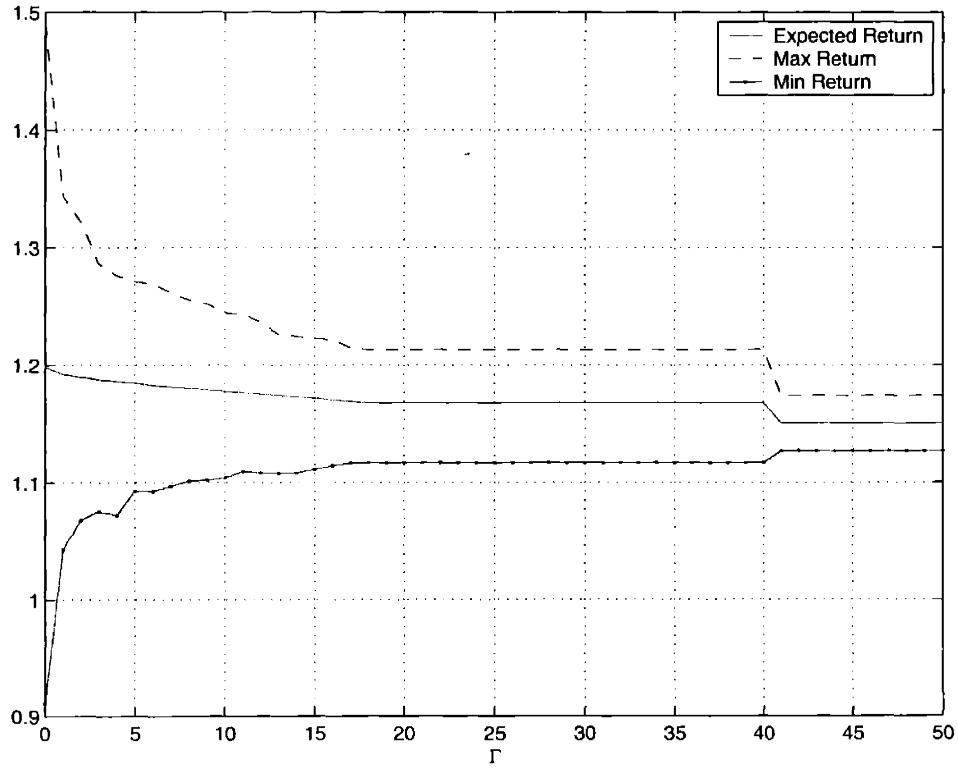

Figure 2-7: Empirical Result of expected, maximum and minimum yield.

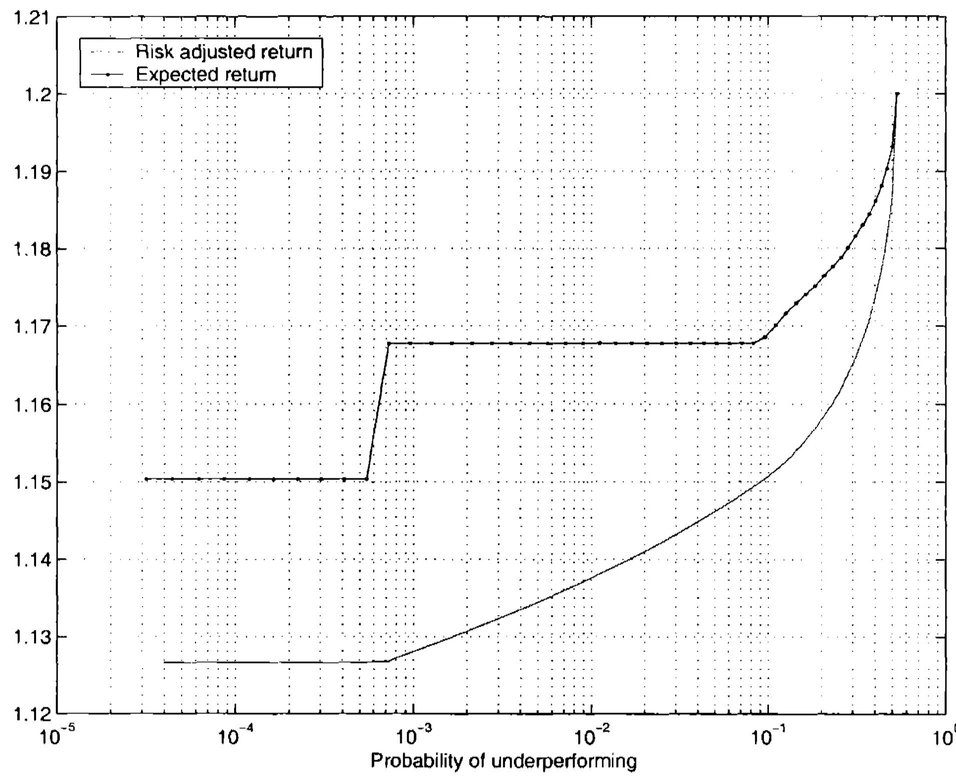

Figure 2-8: Tradeoffs between probability of underperforming and returns.

$\lvert J_i \rvert$	# constraints	$\lvert J_i \rvert$	# constraints
1	21	24	12
11	4	25	4
14	4	26	8
15	4	27	8
16	4	28	4
17	8	29	8
22	4		

Table 2.3: Distributions of $\lvert J_i \rvert$ in PILOT4.

return of the portfolio, as well as a probabilistic guarantee that is valid for all symmetric distributions.

2.5.2 Robust Solutions of a Real-World Linear Optimization Problem

As noted by Ben-Tal and Nemirovski [7], optimal solutions of linear optimization problems may become severely infeasible if the nominal data are slightly perturbed. In this experiment, we applied our method to the problem PILOT4 from the Net Lib library of problems. Problem PILOT4 is a linear optimization problem with 411 rows, 1000 columns, 5145 nonzero elements and optimum objective value, −2581.1392613. It contains coefficients such as 717.562256, -1.078783, -3.053161, -.549569, -22.634094, -39.874283, which seem unnecessarily precise. In our study, we assume that the coefficients of this type that participate in the inequalities of the formulation have a maximum 2% deviation from the corresponding nominal values. Table 2.3 presents the distributions of the number of uncertain data in the problem. We highlight that each of the constraints has at most 29 uncertain data.

We solve the robust problem (2.6) and report the results in Table 2.4. In Figure 2-9, we present the efficient frontier of the probability of constraint violation and cost.

We note that the cost of full protection (Soyster's method) is equal to −2397.5799. In this example, we observe that relaxing the need of full protection, still leads to a high increase in the cost unless one is willing to accept unrealistically high probabilities

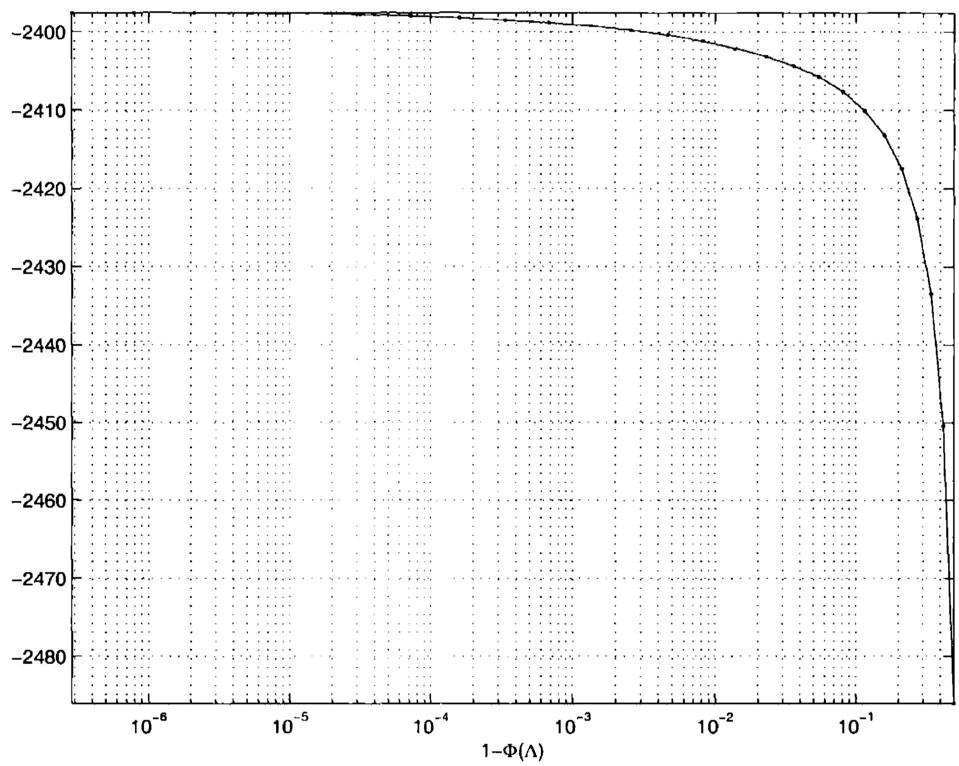

Figure 2-9: The tradeoff between cost and robustness.

Optimal Value	% Change	Prob. violation
-2486.03345	3.68%	0.5
-2450.4324	5.06%	0.421
-2433.4959	5.72%	0.345
-2413.2013	6.51%	0.159
-2403.1495	6.90%	0.0228
-2399.2592	7.05%	0.00135
-2397.8405	7.10%	3.17×10^{-5}
-2397.5799	7.11%	2.87×10^{-7}
-2397.5799	7.11%	9.96×10^{-8}

Table 2.4: The tradeoff between optimal cost and robustness.

for constraint violation. We attribute this to the fact that there are very few uncertain coefficients in each constraint (Table 2.3), and thus probabilistic protection is quite close to deterministic protection.

2.6 Conclusions

The major insights from our analysis are:

1. Our proposed robust methodology provides solutions that ensure deterministic and probabilistic guarantees that constraints will be satisfied as data change.

2. Under the proposed method, the protection level determines probability bounds of constraint violation, which do not depend on the solution of the robust model.

3. The method naturally applies to discrete optimization problems which we will illustrate in Chapter 4.

4. We feel that this is indicative of the fact that the attractiveness of the method increases as the number of uncertain data increases.

Chapter 3

Robust Linear Optimization under General Norms

In this chapter, we study the framework of robust optimization applied to the linear optimization problem

$$\max\left\{\boldsymbol{c}'\boldsymbol{x} \mid \tilde{\boldsymbol{A}}\boldsymbol{x} \leq \boldsymbol{b}, \boldsymbol{x} \in P^x\right\}, \tag{3.1}$$

with $\boldsymbol{x} \in \Re^{n\times 1}$, $\tilde{\boldsymbol{A}} \in \Re^{m\times n}$ a matrix of uncertain coefficients belonging to a known uncertainty set $\mathcal{U}$, $\boldsymbol{c} \in \Re^{n\times 1}$ and P^x a given set representing the constraints involving certain coefficients, the *robust counterpart* of Problem (3.1) is

$$\max\left\{\boldsymbol{c}'\boldsymbol{x} \mid \tilde{\boldsymbol{A}}\boldsymbol{x} \leq \boldsymbol{b}, \boldsymbol{x} \in P^x, \forall \tilde{\boldsymbol{A}} \in \mathcal{U}\right\}. \tag{3.2}$$

An optimal solution $\boldsymbol{x}^*$ is robust with respect to any realization of the data, that is, it satisfies the constraints for any $\tilde{\boldsymbol{A}} \in \mathcal{U}$.

In Chapter 2, we consider LPs such that each entry $\tilde{a}_{ij}$ of $\tilde{\boldsymbol{A}} \in \Re^{m\times n}$ is assumed to take values in the interval $[\bar{a}_{ij} - \Delta_{ij}, \bar{a}_{ij} + \Delta_{ij}]$ and protect for the case that up to Γ_i of the uncertain coefficients in constraint i, $i = 1, \ldots, m$, can take their worst-case values at the same time. The parameter Γ_i controls the tradeoff between robustness and optimality. The attractive aspect of the framework is that the robust counterpart is still an LP.

In this chapter, we propose a framework for robust modeling of linear optimization problems using uncertainty sets described by an arbitrary norm. We explicitly characterize the robust counterpart as a convex optimization problem that involves the dual norm of the given norm. Under a Euclidean norm we recover the second order cone formulation in Ben-Tal and Nemirovski [4, 6], El Ghaoui et al. [11, 12], while under a particular D-norm we introduce we recover the linear optimization formulation proposed in Chapter 1. In this way, we shed some new light to the nature and structure of the robust counterpart of an LP.

Structure of the chapter. In Section 3.1, we review from Ben-Tal and Nemirovski [8] the robust counterpart of a linear optimization problem when the deviations of the uncertain coefficients lie in a convex set and characterize the robust counterpart of an LP when the uncertainty set is described by a general norm, as a convex optimization problem that involves the dual norm of the given norm. In Section 3.2, we show that by varying the norm used to define the uncertainty set, we recover the second order cone formulation in Ben-Tal and Nemirovski [4, 6], El Ghaoui et al. [11, 12], while under a particular D-norm we introduce we recover the linear optimization formulation proposed in Chapter 1. In Section 3.3, we provide guarantees for constraint violation under general probabilistic models that allow arbitrary dependencies in the distribution of the uncertain coefficients. The final section contains some concluding remarks.

Notation

In this chapter, lowercase boldface will be used to denote vectors, while uppercase boldface will denote matrices. Tilde ($\tilde{\boldsymbol{a}}$) will denote uncertain coefficients, while over line ($\bar{\boldsymbol{a}}$) will be used for nominal values. $\tilde{\boldsymbol{A}} \in \Re^{m \times n}$ will usually be the matrix of uncertain coefficients in the linear optimization problem, and $\text{vec}(\tilde{\boldsymbol{A}}) \in \Re^{(m \cdot n) \times 1}$ will denote the vector obtained by stacking its rows on top of one another.

3.1 Uncertainty Sets Defined by a Norm

In this section, we review from Ben-Tal and Nemirovski [8] the structure of the robust counterpart for uncertainty sets defined by general norms. These characterizations are used to develop the new characterizations in Section 3.

Let S be a closed, bounded convex set and consider an uncertainty set in which the uncertain coefficients are allowed to vary in such a way that the deviations from their nominal values fall in a convex set S

$$\mathcal{U} = \left\{ \tilde{\boldsymbol{A}} \mid (\mathrm{vec}(\tilde{\boldsymbol{A}}) - \mathrm{vec}(\bar{\boldsymbol{A}})) \in S \right\}.$$

The next theorem characterizes the robust counterpart.

Theorem 5 *Problem*

$$\begin{aligned} \max \quad & \boldsymbol{c}'\boldsymbol{x} \\ \text{s.t.} \quad & \tilde{\boldsymbol{A}}\boldsymbol{x} \leq \boldsymbol{b} \\ & \boldsymbol{x} \in P^x \\ & \forall \tilde{\boldsymbol{A}} \in \Re^{m \times n} \text{ such that } (vec(\tilde{\boldsymbol{A}}) - vec(\bar{\boldsymbol{A}})) \in S \end{aligned} \tag{3.3}$$

can be formulated as

$$\begin{aligned} \max \quad & \boldsymbol{c}'\boldsymbol{x} \\ \text{s.t.} \quad & \bar{\boldsymbol{a}}_i\boldsymbol{x} + \max_{\boldsymbol{y} \in S}\{\boldsymbol{y}'\boldsymbol{x}\} \leq \boldsymbol{b}_i, \quad i = 1, \dots, m \\ & \boldsymbol{x} \in P^x. \end{aligned} \tag{3.4}$$

Proof : Clearly since S is compact, for each constraint i, $\tilde{\boldsymbol{a}}_i'\boldsymbol{x} \leq b_i$ for all $\mathrm{vec}(\tilde{\boldsymbol{A}}) - \mathrm{vec}(\bar{\boldsymbol{A}}) \in S$ if and only if

$$\max_{\left\{\mathrm{vec}(\tilde{\boldsymbol{A}}) - \mathrm{vec}(\bar{\boldsymbol{A}}) \in S\right\}} \{\tilde{\boldsymbol{a}}_i'\boldsymbol{x}\} \leq b_i.$$

Since

$$\max_{\{\text{vec}(\tilde{\boldsymbol{A}})-\text{vec}(\bar{\boldsymbol{A}})\in S\}} \{\tilde{\boldsymbol{a}}_i{}'\boldsymbol{x}\} = \max_{\{\text{vec}(\tilde{\boldsymbol{A}})-\text{vec}(\bar{\boldsymbol{A}})\in S\}} \left\{(\text{vec}(\tilde{\boldsymbol{A}}))'\boldsymbol{x}_i\right\}$$
$$= (\text{vec}(\bar{\boldsymbol{A}}))'\boldsymbol{x}_i + \max_{\{\boldsymbol{y}\in S\}} \{\boldsymbol{y}'\boldsymbol{x}\}$$

the theorem follows. ■

We next consider uncertainty sets that are arise from the requirement that the distance (as measured by an arbitrary norm) between uncertain coefficients and their nominal values is bounded. Specifically, we consider an uncertainty set $\mathcal{U}$ given by:

$$\mathcal{U} = \left\{\tilde{\boldsymbol{A}} \mid ||\boldsymbol{M}(\text{vec}(\tilde{\boldsymbol{A}}) - \text{vec}(\bar{\boldsymbol{A}}))|| \leq \Delta\right\}, \tag{3.5}$$

where $\boldsymbol{M}$ is an invertible matrix, $\Delta \geq 0$ and $||\cdot||$ a general norm.

Given a norm $||\boldsymbol{x}||$ for a real space of vectors $\boldsymbol{x}$, its *dual norm* induced over the dual space of linear functionals $\boldsymbol{s}$ is defined as follows:

Definition 1 (Dual Norm)

$$||\boldsymbol{s}||^* \doteq \max_{\{||\boldsymbol{x}||\leq 1\}} \boldsymbol{s}'\boldsymbol{x}. \tag{3.6}$$

The next result is well known (see, for example, Lax [21]).

Proposition 3 (a) *The dual norm of the dual norm is the original norm.*

(b) *The dual norm of the L_p norm*

$$||\boldsymbol{x}||_p \doteq \left(\sum_{j=1}^{n} |x_j|^p\right)^{1/p}, \tag{3.7}$$

is the L_q norm $||\boldsymbol{s}||_q$ with $q = 1 + \frac{1}{p-1}$.

The next theorem derives the form of the robust counterpart, when the uncertainty set is given by Eq. (3.5).

Theorem 6 *Problem*

$$
\begin{aligned}
\max \quad & \boldsymbol{c}'\boldsymbol{x} \\
\text{s.t.} \quad & \tilde{\boldsymbol{A}}\boldsymbol{x} \leq \boldsymbol{b} \\
& \boldsymbol{x} \in P^x \\
& \forall \tilde{\boldsymbol{A}} \in \mathcal{U} = \left\{ \tilde{\boldsymbol{A}} \mid ||\boldsymbol{M}(vec(\tilde{\boldsymbol{A}}) - vec(\bar{\boldsymbol{A}}))|| \leq \Delta \right\},
\end{aligned}
\tag{3.8}
$$

where $\boldsymbol{M}$ *is an invertible matrix, can be formulated as*

$$
\begin{aligned}
\max \quad & \boldsymbol{c}'\boldsymbol{x} \\
\text{s.t.} \quad & \bar{\boldsymbol{a}}_i\boldsymbol{x} + \Delta||\boldsymbol{M}'^{-1}\boldsymbol{x}_i||^* \leq \boldsymbol{b}_i, \quad i = 1, \ldots, m \\
& \boldsymbol{x} \in P^x,
\end{aligned}
\tag{3.9}
$$

where $\boldsymbol{x}_i \in \Re^{(m \cdot n) \times 1}$ *is a vector that contains* $\boldsymbol{x} \in \Re^{n \times 1}$ *in entries* $(i-1) \cdot n + 1$ *through* $i \cdot n$*, and* 0 *everywhere else.*

Proof : Introducing $\boldsymbol{y} = \dfrac{\boldsymbol{M}(\text{vec}(\tilde{\boldsymbol{A}}) - \text{vec}(\bar{\boldsymbol{A}}))}{\Delta}$, we can write $\mathcal{U} = \{\boldsymbol{y} \mid ||\boldsymbol{y}|| \leq 1\}$ and obtain

$$
\begin{aligned}
\max_{\{\text{vec}(\tilde{\boldsymbol{A}}) \in \mathcal{U}\}} \{\tilde{\boldsymbol{a}}_i{}'\boldsymbol{x}\} &= \max_{\{\text{vec}(\tilde{\boldsymbol{A}}) \in \mathcal{U}\}} \left\{ (\text{vec}(\tilde{\boldsymbol{A}}))'\boldsymbol{x}_i \right\} \\
&= \max_{\{\boldsymbol{y} \mid ||\boldsymbol{y}|| \leq 1\}} \left\{ (\text{vec}(\bar{\boldsymbol{A}}))'\boldsymbol{x}_i + \Delta(\boldsymbol{M}^{-1}\boldsymbol{y})'\boldsymbol{x} \right\} \\
&= \bar{\boldsymbol{a}}'\boldsymbol{x} + \Delta \max_{\{\boldsymbol{y} \mid ||\boldsymbol{y}|| \leq 1\}} \left\{ \boldsymbol{y}'(\boldsymbol{M}'^{-1}\boldsymbol{x}) \right\}.
\end{aligned}
$$

By Definition 1, the second term in the last expression is $\Delta||\boldsymbol{M}'^{-1}\boldsymbol{x}||^*$. The theorem follows by applying Theorem 5. ■

In applications, Ben-Tal and Nemirovski [4, 6] and El Ghaoui et al. [11, 12] work with uncertainty sets given by the Euclidean norm, i.e.,

$$
\mathcal{U} = \left\{ \tilde{\boldsymbol{A}} \mid ||\boldsymbol{M}(\text{vec}(\tilde{\boldsymbol{A}}) - \text{vec}(\bar{\boldsymbol{A}}))||_2 \leq \Delta \right\},
$$

where $||\cdot||_2$ denotes the Euclidean norm. Since the Euclidean norm is self dual, it follows that the robust counterpart is a second order cone problem. If the uncertainty set is described by either $||\cdot||_1$ or $||\cdot||_\infty$ (these norms are dual to each other), then the resulting robust counterpart can be formulated as an LP.

3.2 The D-Norm and its Dual

In this section, we show that the robust approach in Chapter 2 follows also from Theorem 6 by simply considering a different norm, called the D-norm and its dual as opposed to the Euclidean norm considered in Ben-Tal and Nemirovski and El Ghaoui et al.. Furthermore, we show worst case bounds on the proximity of the D-norm to the Euclidean norm.

In Chapter 2, we consider uncertainty sets given by

$$|||\boldsymbol{M}(\text{vec}(\tilde{\boldsymbol{A}}) - \text{vec}(\bar{\boldsymbol{A}}))|||_p \leq \Delta$$

with $p \in [1, n]$ and for $\boldsymbol{y} \in \Re^{n\times 1}$

$$|||\boldsymbol{y}|||_p = \max_{\{S\cup\{t\}|S\subseteq N,\ |S|\leq\lfloor p\rfloor,\ t\in N\setminus S\}} \left\{ \sum_{j\in S} |y_j| + (p - \lfloor p\rfloor)|y_t| \right\}.$$

The fact that $|||\boldsymbol{y}|||_p$ is indeed a norm, i.e., $|||\boldsymbol{y}|||_p \geq 0$, $|||c\boldsymbol{y}|||_p = |c|\cdot|||\boldsymbol{y}|||_p$, $|||\boldsymbol{y}|||_p = 0$ if and only $\boldsymbol{y} = \boldsymbol{0}$, and $|||\boldsymbol{x}+\boldsymbol{y}|||_p \leq |||\boldsymbol{x}|||_p + |||\boldsymbol{y}|||_p$ follows easily. Specifically, in the robust framework of Chapter 2, we consider constraint-wise uncertainty sets, $\boldsymbol{M}$ a diagonal matrix containing the inverses of the ranges of coefficient variation, and $\Delta = 1$. We next derive the dual norm.

Proposition 4 *The dual norm of the norm $|||\cdot|||_p$ is given by*

$$|||\boldsymbol{s}|||_p^* = \max(||\boldsymbol{s}||_\infty, ||\boldsymbol{s}||_1/p).$$

Proof : The norm $|||\boldsymbol{y}|||_p$ can be written as

$$
\begin{aligned}
|||\boldsymbol{y}|||_p = \max \quad & \sum_{j=1}^{n} u_j y_j & = \quad \min \quad & pr + \sum_{j=1}^{n} t_j \\
\text{s.t.} \quad & \sum_{j=1}^{n} u_j \leq p, \;\; 0 \leq u_j \leq 1 & \text{s.t.} \quad & r + t_j \geq |y_j|, t_j \geq 0, j = 1, \ldots, n \\
& & & r \geq 0,
\end{aligned}
$$

where the second equality follows by strong duality in linear optimization. Thus, $|||\boldsymbol{y}|||_p \leq 1$ if and only if

$$
pr + \sum_{j=1}^{n} t_j \leq 1, \quad r + t_j \geq |y_j|, t_j \geq 0, \;\; j = 1, \ldots, n, r \geq 0 \tag{3.10}
$$

is feasible. The dual norm $|||\boldsymbol{s}|||_p^*$ is given by

$$
|||\boldsymbol{s}|||_p^* = \max_{|||\boldsymbol{y}|||_p \leq 1} \boldsymbol{s}'\boldsymbol{y}.
$$

From Eq. (3.10) we have that

$$
\begin{aligned}
|||\boldsymbol{s}|||_p^* = \max \quad & \boldsymbol{s}'\boldsymbol{y} \\
\text{s.t.} \quad & pr + \sum_{j=1}^{n} t_j \leq 1, \\
& y_j - t_j - r \leq 0, \qquad j = 1, \ldots, n \\
& -y_j - t_j - r \leq 0, \quad j = 1, \ldots, n \\
& r \geq 0, t_j \geq 0, \qquad\;\; j = 1, \ldots, n.
\end{aligned}
$$

Applying LP duality again we obtain

$$
\begin{aligned}
|||\boldsymbol{s}|||_p^* = \min\ & \theta \\
\text{s.t.}\ & p\theta - \sum_{j=1}^{n} u_j - \sum_{j=1}^{n} v_j \geq 0, \\
& \theta - u_j - v_j \geq 0, && j = 1, \dots, n \\
& u_j - v_j = s_j, && j = 1, \dots, n \\
& \theta \geq 0, u_j, v_j \geq 0, && j = 1, \dots, n.
\end{aligned}
$$

Thus,

$$
\begin{aligned}
|||\boldsymbol{s}|||_p^* = \min\ & \theta \\
\text{s.t.}\ & \theta \geq |s_j|, && j = 1, \dots, n \\
& \theta \geq \sum_{j=1}^{n} |s_j|/p,
\end{aligned}
$$

and hence,

$$
|||\boldsymbol{s}|||_p^* = \max(||\boldsymbol{s}||_\infty, ||\boldsymbol{s}||_1/p).
$$

■

Combining Theorem 6 and Proposition 4 leads to an LP formulation of the robust counterpart of the uncertainty set proposed in Chapter 2. We thus observe that Theorem 6 provides a unified treatment of both the approach of Ben-Tal and Nemirovski [4, 6], El Ghaoui et al. [11, 12] and the approach in Chapter 2.

3.2.1 Comparison with the Euclidean Norm

Since uncertainty sets in the literature have been described using the Euclidean and the D-norm it is of interest to understand the proximity between these two norms.

Proposition 5 *For every $\boldsymbol{y} \in \Re^n$*

$$
\begin{aligned}
\min\left\{1, \frac{p}{\sqrt{n}}\right\} &\leq \frac{|||\boldsymbol{y}|||_p}{||\boldsymbol{y}||_2} \leq \sqrt{\lfloor p \rfloor + (p - \lfloor p \rfloor)^2} \\
\min\left\{\frac{1}{p}, \frac{1}{\sqrt{n}}\right\} &\leq \frac{|||\boldsymbol{y}|||_p^*}{||\boldsymbol{y}||_2} \leq \max\left\{\frac{\sqrt{n}}{p}, 1\right\}
\end{aligned}
$$

Proof : We find a lower bound on $|||y|||_p/\|y\|_2$ by solving the following nonlinear optimization problem:

$$\begin{aligned} \max \quad & \sum_{j \in N} x_j^2 \\ \text{s.t.} \quad & |||\boldsymbol{x}|||_p = 1. \end{aligned} \tag{3.11}$$

Let $S = \{1, \ldots, \lfloor p \rfloor\}$, $t = \lfloor p \rfloor + 1$. We can impose nonnegativity constraints on $\boldsymbol{x}$ and the constraints that $x_j \geq x_t$, $\forall j \in S$ and $x_j \leq x_t$, $\forall j \in N \backslash S$, without affecting the objective value of the Problem (3.11). Observing that the objective can never decrease if we let $x_j = x_t$, $\forall j \in N \backslash S$, we reduce (3.11) to the following problem:

$$\begin{aligned} \max \quad & \sum_{j \in S} x_j^2 + (n - \lfloor p \rfloor) x_t^2 \\ \text{s.t.} \quad & \sum_{j \in S} x_j + (p - \lfloor p \rfloor) x_t = 1 \\ & x_j \geq x_t \quad \forall j \in S \\ & x_t \geq 0. \end{aligned} \tag{3.12}$$

Since we are maximizing a convex function over a polytope, there exist an extreme point optimal solution to Problem (3.12). There are $|S| + 1$ inequality constraints. Out of those, $|S|$ need to be tight to establish an extreme point solution. The $|S| + 1$ extreme points can be found to be:

$$\boldsymbol{x}^k = \boldsymbol{e}_k \quad \forall k \in S \tag{3.13}$$

$$\boldsymbol{x}^{|S|+1} = \frac{1}{p}\boldsymbol{e}, \tag{3.14}$$

where $\boldsymbol{e}$ is the vector of ones and $\boldsymbol{e}_k$ is the unit vector with the kth element equals one, and the rest equal zero. Clearly, substituting all possible solutions, Problem (3.12) yields the optimum value of $\max\{1, n/p^2\}$. Taking the square root, the inequality $\|y\|_2 \leq \max\{1, \sqrt{n}/p\}\ |||y|||_p$ follows.

Similarly, in order to derive an upper bound of $|||y|||_p/\|y\|_2$, we solve the following problem:

$$
\begin{aligned}
\min \quad & \sum_{j \in N} x_j^2 \\
\text{s.t.} \quad & |||\boldsymbol{x}|||_p = 1.
\end{aligned}
\tag{3.15}
$$

Using the same partition of the solution an before, and observing that the objective can never increase with $x_j = 0, \ \forall j \in N \backslash S \ \backslash \{t\}$, we reduce Problem (3.15) to the following problem:

$$
\begin{aligned}
\min \quad & \sum_{j \in S} x_j^2 + x_t^2 \\
\text{s.t.} \quad & \sum_{j \in S} x_j + (p - \lfloor p \rfloor) x_t = 1 \\
& x_j \geq x_t \quad \forall j \in S \\
& x_t \geq 0.
\end{aligned}
\tag{3.16}
$$

An optimal solution of Problem (3.16) can be found using the KKT conditions:

$$
x_j = \begin{cases} \dfrac{1}{\lfloor p \rfloor + (p - \lfloor p \rfloor)^2} & \text{if } j \in S, \\ \dfrac{p - \lfloor p \rfloor}{\lfloor p \rfloor + (p - \lfloor p \rfloor)^2} & \text{if } j = t, \\ 0 & \text{otherwise.} \end{cases}
$$

Substituting, the optimal objective value of Problem (3.16) is $(\lfloor p \rfloor + (p - \lfloor p \rfloor)^2)^{-1}$. Hence, taking the square root, we establish the inequality

$$
|||\boldsymbol{y}|||_p \leq \sqrt{\lfloor p \rfloor + (p - \lfloor p \rfloor)^2} \|\boldsymbol{y}\|_2
$$

.

Since $1 \leq \frac{\|y\|_1}{\|y\|_2} \leq \sqrt{n}$, $\frac{1}{\sqrt{n}} \leq \frac{\|y\|_\infty}{\|y\|_2} \leq 1$, and $|||y|||_p^* = \max\left\{\frac{\|y\|_1}{p}, \|y\|_\infty\right\}$ the bounds follow. ■

3.3 Probabilistic Guarantees

In this section, we derive probabilistic guarantees on the feasibility of an optimal robust solution when the uncertainty set $\mathcal{U}$ is described by a general norm $||\cdot||$ with a dual norm $||\cdot||^*$.

We assume that the matrix $\tilde{\boldsymbol{A}}$ has an arbitrary (and unknown) distribution, but with known expected value $\bar{\boldsymbol{A}} \in \Re^{m\cdot n}$ and known covariance matrix $\Sigma \in \Re^{(m\cdot n)\times(m\cdot n)}$. Note that we allow arbitrary dependencies on $\tilde{\boldsymbol{A}}$. We define $\boldsymbol{M} = \boldsymbol{\Sigma}^{-\frac{1}{2}}$, which is a symmetric matrix.

Let $\boldsymbol{x}^* \in \Re^{n\times 1}$ be an optimal solution to Problem (3.9). Recall that $\boldsymbol{x}_i^* \in \Re^{(m\cdot n)\times 1}$ denotes the vector containing $\boldsymbol{x}^*$ in entries $(i-1)\cdot n$ through $i\cdot n$, and 0 everywhere else.

Theorem 7 **(a)** *The probability that $\boldsymbol{x}^*$ satisfies constraint i for any realization of the uncertain matrix $\tilde{\boldsymbol{A}}$ is*

$$\mathrm{P}\left(\tilde{\boldsymbol{a}}_i'\boldsymbol{x}^* \le b_i\right) = \mathrm{P}\left((vec(\tilde{\boldsymbol{A}}))'\boldsymbol{x}_i^* \le b_i\right) \ge 1 - \frac{1}{1+\Delta^2\left(\frac{||\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{x}_i^*||^*}{||\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{x}_i^*||_2}\right)^2}. \tag{3.17}$$

(b) *If the norm used in $\mathcal{U}$ is the D-norm $|||\cdot|||_p$, then*

$$\mathrm{P}\left(\tilde{\boldsymbol{a}}_i'\boldsymbol{x}^* \le b_i\right) \ge 1 - \frac{1}{1+\Delta^2\min\left\{\frac{1}{p^2}, \frac{1}{n}\right\}}. \tag{3.18}$$

(c) *If the norm used in $\mathcal{U}$ is the dual D-norm $|||\cdot|||_p^*$, then*

$$\mathrm{P}\left(\tilde{\boldsymbol{a}}_i'\boldsymbol{x}^* \le b_i\right) \ge 1 - \frac{1}{1+\Delta^2\min\left\{1, \frac{p^2}{n}\right\}}. \tag{3.19}$$

(d) *If the norm used in $\mathcal{U}$ is the Euclidean norm, then*

$$\mathrm{P}\left(\tilde{\boldsymbol{a}}_i'\boldsymbol{x}^* \le b_i\right) \ge 1 - \frac{1}{1+\Delta^2}. \tag{3.20}$$

Proof : Since an optimal robust solution satisfies

$$(\text{vec}(\bar{\boldsymbol{A}}))'\boldsymbol{x}_i^* + \Delta||\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{x}_i^*||^* \le b_i,$$

we obtain that

$$\text{P}\left((\text{vec}(\tilde{\boldsymbol{A}}))'\boldsymbol{x}_i^* > b_i\right) \le \text{P}\left((\text{vec}(\tilde{\boldsymbol{A}}))'\boldsymbol{x}_i^* \ge (\text{vec}(\bar{\boldsymbol{A}}))'\boldsymbol{x}_i^* + ||\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{x}_i^*||^*\right). \tag{3.21}$$

Bertsimas and Popescu [9] show that if S is a convex set, and $\tilde{\boldsymbol{X}}$ is a vector of random variables with mean $\bar{\boldsymbol{X}}$ and covariance matrix $\boldsymbol{\Sigma}$, then

$$\text{P}\left(\tilde{\boldsymbol{X}} \in S\right) \le \frac{1}{1+d^2}, \tag{3.22}$$

where

$$d^2 = \inf_{\tilde{\boldsymbol{X}} \in S} \left(\tilde{\boldsymbol{X}} - \bar{\boldsymbol{X}}\right)' \boldsymbol{\Sigma}^{-1} \left(\tilde{\boldsymbol{X}} - \bar{\boldsymbol{X}}\right).$$

We consider the convex set

$$S_i = \left\{\text{vec}(\tilde{\boldsymbol{A}}) \mid (\text{vec}(\tilde{\boldsymbol{A}}))'\boldsymbol{x}_i^* > (\text{vec}(\bar{\boldsymbol{A}}))'\boldsymbol{x}_i^* + \Delta||\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{x}_i^*||^*\right\}$$

. In our case,

$$d_i^2 = \inf_{\text{vec}(\tilde{\boldsymbol{A}}) \in S_i} \left(\text{vec}(\tilde{\boldsymbol{A}}) - \text{vec}(\bar{\boldsymbol{A}})\right)' \boldsymbol{\Sigma}^{-1} \left(\text{vec}(\tilde{\boldsymbol{A}}) - \text{vec}(\bar{\boldsymbol{A}})\right).$$

Applying the KKT conditions for this optimization problem we obtain an optimal solution

$$\text{vec}(\bar{\boldsymbol{A}}) + \Delta \left(\frac{||\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{x}_i^*||^*}{||\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{x}_i^*||_2}\right)^2 \boldsymbol{\Sigma}\boldsymbol{x}_i^*,$$

with

$$d^2 = \Delta^2 \left(\frac{||\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{x}_i^*||^*}{||\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{x}_i^*||_2}\right)^2.$$

Applying (3.22), Eq. (3.17) follows.

(b) If the norm used in the uncertainty set $\mathcal{U}$ is the D-norm, then by applying

Proposition 5, we obtain Eq. (3.18).

(c) If the norm used in the uncertainty set $\mathcal{U}$ is the dual D-norm, then by applying Proposition 5, we obtain Eq. (3.19).

(d) If the norm used in the uncertainty set $\mathcal{U}$ is the Euclidean norm, then Eq. (3.20) follows immediately from Eq. (3.17). ■

In Chapter 2, we prove stronger bounds under the stronger assumption that the data in each constraint are independently distributed according to a symmetric distribution. In contrast the bounds in Theorem 7 are weaker, but have wider applicability as they include arbitrary dependencies.

3.4 Conclusions

We have proposed a framework for robust modeling of linear optimization problems that unifies models in Ben-Tal and Nemirovski [4, 6], El Ghaoui et al. [11, 12] and our approach in Chapter 2. The use of the Euclidean norm in the uncertainty set leads to the formulation of the robust counterpart as an SOCP, while the use of the D-norm leads to the formulation of the robust counterpart as an LP. More general norms lead to more involved (although always convex) robust formulations.

Chapter 4

Robust Conic Optimization

The general optimization problem under parameter uncertainty is as follows:

$$\begin{aligned} \max \quad & f_0(\boldsymbol{x}, \tilde{\boldsymbol{D}}_0) \\ \text{subject to} \quad & f_i(\boldsymbol{x}, \tilde{\boldsymbol{D}}_i) \geq 0, \qquad i \in I, \\ & \boldsymbol{x} \in X, \end{aligned} \tag{4.1}$$

where $f_i(\boldsymbol{x}, \tilde{\boldsymbol{D}}_i)$, $i \in \{0\} \cup I$ are given functions, X is a given set and $\tilde{\boldsymbol{D}}_i$, $i \in \{0\} \cup I$ is the vector of uncertain coefficients. We define the nominal problem to be Problem (4.1) when the uncertain coefficients $\tilde{\boldsymbol{D}}_i$ take values equal to their expected values $\boldsymbol{D}_i^0$.

In order to address parameter uncertainty Problem (4.1) Ben-Tal and Nemirovski [4, 6] and independently by El Ghaoui et al. [11, 12] propose to solve the following robust optimization problem

$$\begin{aligned} \max \quad & \min_{\boldsymbol{D}_0 \in \mathcal{U}_0} f_0(\boldsymbol{x}, \boldsymbol{D}_0) \\ \text{s.t.} \quad & \min_{\boldsymbol{D}_i \in \mathcal{U}_i} f_i(\boldsymbol{x}, \boldsymbol{D}_i) \geq 0, i \in I \\ & \boldsymbol{x} \in X, \end{aligned} \tag{4.2}$$

where $\mathcal{U}_i$, $i \in \{0\} \cup I$ are given uncertainty sets. The motivation for solving Problem (4.2) is to find a solution $\boldsymbol{x}^* \in X$ that "immunizes" Problem (4.1) against parameter uncertainty. By selecting appropriate uncertainty sets $\mathcal{U}_i$, we can address the tradeoff between robustness and optimality. In designing such an approach two criteria are important in our view:

(a) Preserving the computational tractability both theoretically and most importantly practically of the nominal problem. From a theoretical perspective it is desirable that if the nominal problem is solvable in polynomial time, then the robust problem is also polynomially solvable. More specifically, it is desirable that robust conic optimization problems retain their original structure, i.e., robust linear optimization problems (LPs) remain LPs, robust second order cone optimization problems (SOCPs) remain SOCPs and robust semidefinite optimization problems (SDPs) remain SDPs.

(b) Being able to find a guarantee on the probability that the robust solution is feasible, when the uncertain coefficients obey some natural probability distributions. This is important, since from these guarantees we can select parameters that affect the uncertainty sets U_i that allows to control the tradeoff between robustness and optimality.

Let us examine whether the state of the art in robust optimization has the two properties mentioned above:

1. **Linear Optimization:** A uncertain LP constraint is of the form $\tilde{\boldsymbol{a}}'\boldsymbol{x} \geq \tilde{b}$, for which $\tilde{\boldsymbol{a}}$ and $\tilde{b}$ are subject to uncertainty. When the corresponding uncertainty set $\mathcal{U}$ is a polyhedron, then the robust counterpart is also an LP. When $\mathcal{U}$ is ellipsoidal, then the robust counterpart becomes an SOCP. For linear optimization there are probabilistic guarantees for feasibility available under reasonable probabilistic assumptions on data variation.

2. **Quadratic Constrained Quadratic Optimization (QCQP):** An uncertain QCQP constraint is of the form $\|\tilde{\boldsymbol{A}}\boldsymbol{x}\|_2^2 + \tilde{\boldsymbol{b}}'\boldsymbol{x} + \tilde{c} \leq 0$, where $\tilde{\boldsymbol{A}}$, $\tilde{\boldsymbol{b}}$ and $\tilde{c}$ are subject to data uncertainty. The robust counterpart is an SDP if the uncertainty set is a simple ellipsoid, and NP-hard if the set is polyhedral (Ben-Tal and Nemirovski [4, 6]. To the best of our knowledge, there are no available

probabilistic bounds.

3. **Second Order Cone Optimization (SOCP):** An uncertain SOCP constraint is of the form $\|\tilde{\boldsymbol{A}}\boldsymbol{x} + \tilde{\boldsymbol{b}}\|_2 \leq \tilde{\boldsymbol{c}}'\boldsymbol{x} + \tilde{d}$, where $\tilde{\boldsymbol{A}}$, $\tilde{\boldsymbol{b}}$, $\tilde{\boldsymbol{c}}$ and $\tilde{d}$ are subject to data uncertainty. The robust counterpart is an SDP if $\tilde{\boldsymbol{A}}$, $\tilde{\boldsymbol{b}}$ belong in an ellipsoidal uncertainty set $\mathcal{U}_1$ and $\tilde{\boldsymbol{c}}$, $\tilde{d}$ belong in another ellipsoidal set $\mathcal{U}_2$. The complexity of the problem is unknown, however, if $\tilde{\boldsymbol{A}}$, $\tilde{\boldsymbol{b}}$, $\tilde{\boldsymbol{c}}$, $\tilde{d}$ vary together in a common ellipsoidal set. To the best of our knowledge, there are no available probabilistic bounds.

4. **Semidefinite Optimization (SDP):** An uncertain SDP constraint of the form $\sum_{j=1}^{n} \tilde{\boldsymbol{A}}_j x_j \succeq \tilde{\boldsymbol{B}}$, where $\tilde{\boldsymbol{A}}_1, ..., \tilde{\boldsymbol{A}}_n$ and $\tilde{\boldsymbol{B}}$ are subject to data uncertainty. The robust counterpart is NP-hard for ellipsoidal uncertainty sets, while there are no available probabilistic bounds.

5. **Conic Optimization:** An uncertain conic optimization constraint of the form $\sum_{j=1}^{n} \tilde{\boldsymbol{A}}_j x_j \succeq_{\boldsymbol{K}} \tilde{\boldsymbol{B}}$, where $\tilde{\boldsymbol{A}}_1, ..., \tilde{\boldsymbol{A}}_n$ and $\tilde{\boldsymbol{B}}$ are subject to data uncertainty. The cone $\boldsymbol{K}$ is closed, pointed and with a nonempty interior. To the best of our knowledge, there are no results available regarding tractability and probabilistic guarantees in this case.

Our goal in this chapter is to address (a) and (b) above for robust conic optimization problems. Specifically, we propose a new robust counterpart of Problem (4.1) that has two properties: (a) It inherits the character of the nominal problem; for example, robust SOCPs remain SOCPs and robust SDPs remain SDPs; (b) under reasonable probabilistic assumptions on data variation we establish probabilistic guarantees for feasibility that lead to explicit ways for selecting parameters that control robustness.

The structure of the chapter is as follows. In Section 4.1, we describe the proposed robust model and in Section 4.2, we show that the robust model inherits the character of the nominal problem for LPs, QCQPs, SOCPs and SDPs. In Section 4.3, we prove probabilistic guarantees for feasibility for these classes of problems. In Section 4.4, we show tractability and give explicit probabilistic bounds for general conic problems. Section 4.5 concludes this chapter.

4.1 The Robust Model

In this section, we outline the ingredients of the proposed framework for robust conic optimization.

4.1.1 Model for Parameter Uncertainty

The model of data uncertainty we consider is

$$\tilde{\boldsymbol{D}} = \boldsymbol{D}^0 + \sum_{j \in N} \Delta \boldsymbol{D}^j \tilde{z}_j, \tag{4.3}$$

where $\boldsymbol{D}^0$ is the nominal value of the data, $\Delta\boldsymbol{D}^j$, $j \in N$ is a direction of data perturbation, and $\tilde{z}_j$, $j \in N$ are independent and identically distributed random variables with mean equal to zero, so that $E[\tilde{\boldsymbol{D}}] = \boldsymbol{D}^0$. The cardinality of N may be small, modeling situations involving a small collection of primitive independent uncertainties (for example a factor model in a finance context), or large, potentially as large as the number of entries in the data. In the former case, the elements of $\tilde{\boldsymbol{D}}$ are strongly dependent, while in the latter case the elements of $\tilde{\boldsymbol{D}}$ are weakly dependent or even independent (when $|N|$ is equal to the number of entries in the data).

4.1.2 Uncertainty Sets and Related Norms

In the robust optimization framework of (4.2), we consider the uncertainty set $\mathcal{U}$ as follows:

$$\mathcal{U} = \left\{ \boldsymbol{D} \mid \exists \boldsymbol{u} \in \Re^{|N|} : \boldsymbol{D} = \boldsymbol{D}^0 + \sum_{j \in N} \Delta \boldsymbol{D}^j u_j, \|\boldsymbol{u}\| \leq \Omega \right\}, \tag{4.4}$$

where Ω is a parameter controling the tradeoff between robustness and optimality (robustness increases as Ω increases). We restrict the vector norm $\|.\|$ we consider by imposing the condition:

$$\|\boldsymbol{u}\| = \|\boldsymbol{u}^+\|, \tag{4.5}$$

where $u_j^+ = |u_j| \ \forall j \in N$. The following norms commonly used in robust optimization satisfy Eq. (4.5):

- The polynomial norm l_k, $k = 1, \ldots, \infty$ (see Ben-Tal and Nemirovski [4, 7] and Soyster [26]).

- The $l_2 \cap l_\infty$ norm: $\max\{\|\boldsymbol{u}\|_2, \Omega\|\boldsymbol{u}\|_\infty\}$, $\Omega > 0$ (see Ben-Tal and Nemirovski [7]). This norm is used in modeling bounded and symmetrically distributed random data.

- The $l_1 \cap l_\infty$ norm: $\max\{\frac{1}{\Gamma}\|\boldsymbol{u}\|_1, \|\boldsymbol{u}\|_\infty\}$, $\Gamma > 0$. Note that this norm is equal to l_∞ if $\Gamma = |N|$, and l_1 if $\Gamma = 1$. This norm is used in modeling bounded and symmetrically distributed random data, and has the additional property that the robust counterpart of an LP is still an LP.

Given a norm $\|.\|$ we consider the dual norm $\|.\|^*$ defined as

$$\|\boldsymbol{s}\|^* = \max_{\|\boldsymbol{x}\| \leq 1} \boldsymbol{s}'\boldsymbol{x}.$$

We next show some basic properties of norms satisfying Eq. (4.5), which we will subsequently use in our development.

Proposition 6 *If the norm* $\|\cdot\|$ *satisfies Eq. (4.5), then we have*

(a) $\|\boldsymbol{w}\|^* = \|\boldsymbol{w}^+\|^*$.

(b) *For all* $\boldsymbol{v}, \boldsymbol{w}$ *such that* $\boldsymbol{v}^+ \leq \boldsymbol{w}^+$, $\|\boldsymbol{v}\|^* \leq \|\boldsymbol{w}\|^*$.

(c) *For all* $\boldsymbol{v}, \boldsymbol{w}$ *such that* $\boldsymbol{v}^+ \leq \boldsymbol{w}^+$, $\|\boldsymbol{v}\| \leq \|\boldsymbol{w}\|$.

Proof :

(a) Let $\boldsymbol{y} \in \arg\max_{\|\boldsymbol{x}\| \leq 1} \boldsymbol{w}'\boldsymbol{x}$, and for every $j \in N$, let $z_j = |y_j|$ if $w_j \geq 0$ and $z_j = -|y_j|$, otherwise. Clearly, $\boldsymbol{w}'\boldsymbol{z} = (\boldsymbol{w}^+)'\boldsymbol{y}^+ \geq \boldsymbol{w}'\boldsymbol{y}$. Since, $\|\boldsymbol{z}\| = \|\boldsymbol{z}^+\| =$

$\|y^+\| = \|y\| \leq 1$, and from the optimality of y, we have $w'z \leq w'y$, leading to $w'z = (w^+)'y^+ = w'y$. Since $\|w\| = \|w^+\|$, we obtain

$$\|w\|^* = \max_{\|x\|\leq 1} (w)'x = \max_{\|x\|\leq 1} (w^+)'x^+ = \max_{\|x\|\leq 1} (w^+)'x = \|w^+\|^*.$$

(b) Note that

$$\|w\|^* = \max_{\|x\|\leq 1} (w^+)'x^+ = \max_{\substack{\|x\|\leq 1\\ x\geq 0}} (w^+)'x.$$

If $v^+ \leq w^+$,

$$\|v\|^* = \max_{\substack{\|x\|\leq 1\\ x\geq 0}} (v^+)'x \leq \max_{\substack{\|x\|\leq 1\\ x\geq 0}} (w^+)'x = \|w\|^*.$$

(c) We apply part (b) to the norm $\|.\|^*$. From the self dual property of norms $\|.\|^{**} = \|.\|$, we obtain part (c). ■

4.1.3 The Class of Functions $f(x, D)$

We impose the following restrictions on the class of functions $f(x, D)$ in Problem (4.1) (we drop index j for clarity):

Assumption 1 *The function $f(x, D)$ satisfies:*

(a) *The function $f(x, D)$ is concave in D for all $x \in \Re^n$.*

(b) *$f(x, kD) = kf(x, D)$, for all $k \geq 0$, D, $x \in \Re^n$.*

Note that for functions $f(\cdot, \cdot)$ satisfying Assumption 1 we have:

$$f(x, A + B) \geq \frac{1}{2}f(x, 2A) + \frac{1}{2}f(x, 2B) = f(x, A) + f(x, B). \tag{4.6}$$

The restrictions implied by Assumption 1 still allow us to model LPs, QCQPs, SOCPs and SDPs. Table 4.1 shows the function $f(x, D)$ for these problems. Note that SOCP(1) models situations that only A and b vary, while SOCP(2) models situations that A, b, c and d vary. Note that for QCQP, the function, $-\|Ax\|_2^2 - b'x - c$ does not satisfy the second assumption. However, by extending the dimension of the problem, it is well-known that the QCQP constraint is SOCP constraint representable (see

Type	Constraint	$\boldsymbol{D}$	$f(\boldsymbol{x}, \boldsymbol{D})$
LP	$\boldsymbol{a}'\boldsymbol{x} \geq b$	$(\boldsymbol{a}, b)$	$\boldsymbol{a}'\boldsymbol{x} - b$
QCQP	$\|\boldsymbol{A}\boldsymbol{x}\|_2^2 + \boldsymbol{b}'\boldsymbol{x} + c \leq 0$	$(\boldsymbol{A}, \boldsymbol{b}, c, d)$ $d^0 = 1$, $\Delta d^j = 0\ \forall j \in N$	$\frac{d-(\boldsymbol{b}'\boldsymbol{x}+c)}{2} - \sqrt{\|\boldsymbol{A}\boldsymbol{x}\|_2^2 + \left(\frac{d+\boldsymbol{b}'\boldsymbol{x}+c}{2}\right)^2}$
SOCP(1)	$\|\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}\|_2 \leq \boldsymbol{c}'\boldsymbol{x} + d$	$(\boldsymbol{A}, \boldsymbol{b}, \boldsymbol{c}, d)$ $\boldsymbol{\Delta c}^j = \boldsymbol{0}$, $\Delta d^j = 0\ \forall j \in N$	$\boldsymbol{c}'\boldsymbol{x} + d - \|\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}\|_2$
SOCP(2)	$\|\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}\|_2 \leq \boldsymbol{c}'\boldsymbol{x} + d$	$(\boldsymbol{A}, \boldsymbol{b}, \boldsymbol{c}, d)$	$\boldsymbol{c}'\boldsymbol{x} + d - \|\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}\|_2$
SDP	$\sum_{j=1}^n \boldsymbol{A}_i x_i - \boldsymbol{B} \in \boldsymbol{S}_+^m$	$(\boldsymbol{A}_1, \ldots, \boldsymbol{A}_n, \boldsymbol{B})$	$\lambda_{min}(\sum_{j=1}^n \boldsymbol{A}_i x_i - \boldsymbol{B})$

Table 4.1: The function $f(\boldsymbol{x}, \boldsymbol{D})$ for different conic optimization problems.

Ben-Tal and Nemirovski [8]). Finally, the SDP constraint,

$$\sum_{j=1}^{n} \boldsymbol{A}_i x_i \succeq \boldsymbol{B},$$

is equivalent to

$$\lambda_{min}\left(\sum_{j=1}^{n} \boldsymbol{A}_i x_i - \boldsymbol{B}\right) \geq 0,$$

where $\lambda_{min}(\boldsymbol{M})$ is the function that returns the smallest eigenvalue of the symmetric matric $\boldsymbol{M}$.

4.2 The Proposed Robust Framework and its Tractability

The robust framework (4.2) leads to a significant increase in complexity for conic optimization problems. For this reason, we propose a more restricted robust problem, which, as we show in this section, retains the complexity of the nominal problem.

Specifically, under the model of data uncertainty in Eq. (4.3) we propose the following constraint for addressing the data uncertainty in the constraint $f(\boldsymbol{x}, \tilde{\boldsymbol{D}}) \geq 0$:

$$\min_{(\boldsymbol{v},\boldsymbol{w})\in\mathcal{V}} f(\boldsymbol{x}, \boldsymbol{D}^0) + \sum_{j\in N} \left\{ f(\boldsymbol{x}, \boldsymbol{\Delta D}^j)v_j + f(\boldsymbol{x}, -\boldsymbol{\Delta D}^j)w_j \right\} \geq 0, \tag{4.7}$$

where

$$\mathcal{V} = \left\{ (\boldsymbol{v}, \boldsymbol{w}) \in \Re_{+}^{|N| \times |N|} \mid \|\boldsymbol{v} + \boldsymbol{w}\| \leq \Omega \right\}, \tag{4.8}$$

and the norm $\|.\|$ satisfies Eq. (4.5). We next show that under Assumption 1, Eq. (4.7) implies the classical definition of robustness:

$$f(\boldsymbol{x}, \boldsymbol{D}) \geq 0, \qquad \forall \boldsymbol{D} \in \mathcal{U}, \tag{4.9}$$

where $\mathcal{U}$ is defined in Eq. (4.4). Moreover, if the function $f(\boldsymbol{x}, \boldsymbol{D})$ is linear in $\boldsymbol{D}$, then Eq. (4.7) is equivalent to Eq. (4.9).

Proposition 7 *Suppose the given norm $\|.\|$ satisfies Eq. (4.5).*

(a) *If $f(\boldsymbol{x}, \boldsymbol{A} + \boldsymbol{B}) = f(\boldsymbol{x}, \boldsymbol{A}) + f(\boldsymbol{x}, \boldsymbol{B})$, then $\boldsymbol{x}$ satisfies (4.7) if and only if $\boldsymbol{x}$ satisfies (4.9).*

(b) *Under Assumption 1, if $\boldsymbol{x}$ is feasible in Problem (4.7), then $\boldsymbol{x}$ is feasible in Problem (4.9).*

Proof :

(a) Under the linearity assumption, Eq. (4.7) is equivalent to:

$$f\left(\boldsymbol{x}, \boldsymbol{D}^0 + \sum_{j \in N} \boldsymbol{\Delta D}^j (v_j - w_j)\right) \geq 0 \qquad \forall \|\boldsymbol{v} + \boldsymbol{w}\| \leq \Omega, \quad \boldsymbol{v}, \boldsymbol{w} \geq \boldsymbol{0}, \tag{4.10}$$

while Eq. (4.9) can be written as:

$$f\left(\boldsymbol{x}, \boldsymbol{D}^0 + \sum_{j \in N} \boldsymbol{\Delta D}^j r_j\right) \geq 0 \qquad \forall \|\boldsymbol{r}\| \leq \Omega. \tag{4.11}$$

Suppose $\boldsymbol{x}$ is infeasible in (4.11), that is, there exists $\boldsymbol{r}$, $\|\boldsymbol{r}\| \leq \Omega$ such that

$$f\left(\boldsymbol{x}, \boldsymbol{D}^0 + \sum_{j \in N} \boldsymbol{\Delta D}^j r_j\right) < 0.$$

For all $j \in N$, let $v_j = \max\{r_j, 0\}$ and $w_j = -\min\{r_j, 0\}$. Clearly, $\boldsymbol{r} = \boldsymbol{v} - \boldsymbol{w}$ and since $v_j + w_j = |r_j|$, we have from Eq. (4.5) that $\|\boldsymbol{v} + \boldsymbol{w}\| = \|\boldsymbol{r}\| \leq \Omega$. Hence, $\boldsymbol{x}$ is infeasible in (4.10) as well.

Conversely, suppose $\boldsymbol{x}$ is infeasible in (4.10), then there exist $\boldsymbol{v}, \boldsymbol{w} \geq \mathbf{0}$ and $\|\boldsymbol{v}+\boldsymbol{w}\| \leq \Omega$ such that

$$f\left(\boldsymbol{x}, \boldsymbol{D}^0 + \sum_{j\in N} \Delta\boldsymbol{D}^j(v_j - w_j)\right) < 0.$$

For all $j \in N$, we let $r_j = v_j - w_j$ and we observe that $|r_j| \leq v_j + w_j$. Therefore, for norms satisfying Eq. (4.5) we have

$$\|\boldsymbol{r}\| = \|\boldsymbol{r}^+\| \leq \|\boldsymbol{v} + \boldsymbol{w}\| \leq \Omega,$$

and hence, $\boldsymbol{x}$ is infeasible in (4.11).

(b) Suppose $\boldsymbol{x}$ is feasible in Problem (4.7), i.e.,

$$f(\boldsymbol{x}, \boldsymbol{D}^0) + \sum_{j\in N} \left\{ f(\boldsymbol{x}, \Delta\boldsymbol{D}^j)v_j + f(\boldsymbol{x}, -\Delta\boldsymbol{D}^j)w_j \right\} \geq 0, \qquad \forall \|\boldsymbol{v}+\boldsymbol{w}\| \leq \Omega, \;\; \boldsymbol{v}, \boldsymbol{w} \geq \mathbf{0}.$$

From Eq. (4.6) and Assumption 1(b)

$$0 \leq f(\boldsymbol{x}, \boldsymbol{D}^0) + \sum_{j\in N} \left\{ f(\boldsymbol{x}, \Delta\boldsymbol{D}^j)v_j + f(\boldsymbol{x}, -\Delta\boldsymbol{D}^j)w_j \right\} \leq f(\boldsymbol{x}, \boldsymbol{D}^0 + \sum_{j\in N} \Delta\boldsymbol{D}^j(v_j - w_j))$$

for all $\|\boldsymbol{v} + \boldsymbol{w}\| \leq \Omega, \;\; \boldsymbol{v}, \boldsymbol{w} \geq \mathbf{0}$. In the proof of part (a) we established that

$$f(\boldsymbol{x}, \boldsymbol{D}^0 + \sum_{j\in N} \Delta\boldsymbol{D}^j r_j) \geq 0 \qquad \forall \|\boldsymbol{r}\| \leq \Omega$$

is equivalent to

$$f(\boldsymbol{x}, \boldsymbol{D}^0 + \sum_{j\in N} \Delta\boldsymbol{D}^j(v_j - w_j)) \geq 0 \qquad \forall \|\boldsymbol{v} + \boldsymbol{w}\| \leq \Omega, \;\; \boldsymbol{v}, \boldsymbol{w} \geq \mathbf{0},$$

and thus $\boldsymbol{x}$ satisfies (4.9). ■

Note that there are other proposals that relax the classical definition of robustness (4.9) (see for instance Ben-Tal and Nemirovski [5]) and lead to tractable solutions. However, our particular proposal in Eq. (4.7) combines tractability with the ability to derive probabilistic guarantees that the solution of Eq. (4.7) would remain feasible under reasonable assumptions on data variation.

4.2.1 Tractability of the Proposed Framework

Unlike the classical definition of robustness (4.9), which can not be represented in a tractable manner, we next show that Eq. (4.7) can be represented in a tractable manner.

Theorem 8 *For a norm satisfying Eq. (4.5) and a function* $f(\boldsymbol{x}, \boldsymbol{D})$ *satisfying Assumption 1*

(a) *Constraint (4.7) is equivalent to*

$$f(\boldsymbol{x}, \boldsymbol{D}^0) \geq \Omega \|\boldsymbol{s}\|^*, \tag{4.12}$$

where

$$s_j = \max\{-f(\boldsymbol{x}, \Delta\boldsymbol{D}^j), -f(\boldsymbol{x}, -\Delta\boldsymbol{D}^j)\}, \quad \forall j \in N.$$

(b) *Eq. (4.12) can be written as:*

$$\begin{aligned}
& f(\boldsymbol{x}, \boldsymbol{D}^0) \geq \Omega y \\
& f(\boldsymbol{x}, \Delta\boldsymbol{D}^j) + t_j \geq 0 \quad \forall j \in N \\
& f(\boldsymbol{x}, -\Delta\boldsymbol{D}^j) + t_j \geq 0 \quad \forall j \in N \\
& \|\boldsymbol{t}\|^* \leq y \\
& y \in \Re, \ \boldsymbol{t} \in \Re^{|N|}.
\end{aligned} \tag{4.13}$$

Proof :

(a) We introduce the following problems:

$$\begin{aligned}
z_1 = \max \quad & \boldsymbol{a}'\boldsymbol{v} + \boldsymbol{b}'\boldsymbol{w} \\
\text{s.t.} \quad & \|\boldsymbol{v} + \boldsymbol{w}\| \leq \Omega \\
& \boldsymbol{v}, \boldsymbol{w} \geq \boldsymbol{0},
\end{aligned} \tag{4.14}$$

and

$$
\begin{aligned}
z_2 = \max \quad & \sum_{j \in N} \max\{a_j, b_j, 0\} r_j \\
\text{s.t.} \quad & \|\boldsymbol{r}\| \leq \Omega,
\end{aligned}
\tag{4.15}
$$

and show that $z_1 = z_2$. Suppose $\boldsymbol{r}^*$ is an optimal solution to (4.15). For all $j \in N$, let

$$
\begin{aligned}
v_j = w_j = 0 \qquad & \text{if } \max\{a_j, b_j\} \leq 0 \\
v_j = |r_j^*|, w_j = 0 \quad & \text{if } a_j \geq b_j, a_j > 0 \\
w_j = |r_j^*|, v_j = 0 \quad & \text{if } b_j > a_j, b_j > 0.
\end{aligned}
$$

Observe that $a_j v_j + b_j w_j \geq \max\{a_j, b_j, 0\} r_j^*$ and $w_j + v_j \leq |r_j^*|$ $\forall j \in N$. From Proposition 6(c) we have $\|\boldsymbol{v} + \boldsymbol{w}\| \leq \|\boldsymbol{r}^*\| \leq \Omega$, and thus $\boldsymbol{v}, \boldsymbol{w}$ are feasible in Problem (4.14), leading to

$$
z_1 \geq \sum_{j \in N} (a_j v_j + b_j w_j) \geq \sum_{j \in N} \max\{a_j, b_j, 0\} r_j^* = z_2.
$$

Conversely, let $\boldsymbol{v}^*, \boldsymbol{w}^*$ be an optimal solution to Problem (4.14). Let $\boldsymbol{r} = \boldsymbol{v}^* + \boldsymbol{w}^*$. Clearly $\|\boldsymbol{r}\| \leq \Omega$ and observe that

$$
r_j \max\{a_j, b_j, 0\} \geq a_j v_j^* + b_j w_j^* \quad \forall j \in N.
$$

Therefore, we have

$$
z_2 \geq \sum_{j \in N} \max\{a_j, b_j, 0\} r_j \geq \sum_{j \in N} (a_j v_j^* + b_j w_j^*) = z_1,
$$

leading to $z_1 = z_2$. We next observe that

$$
\begin{aligned}
& \min_{(\boldsymbol{v}, \boldsymbol{w}) \in \mathcal{V}} \sum_{j \in N} \left\{ f(\boldsymbol{x}, \boldsymbol{\Delta D}^j) v_j + f(\boldsymbol{x}, -\boldsymbol{\Delta D}^j) w_j \right\} \\
= & - \max_{(\boldsymbol{v}, \boldsymbol{w}) \in \mathcal{V}} \sum_{j \in N} \left\{ -f(\boldsymbol{x}, \boldsymbol{\Delta D}^j) v_j - f(\boldsymbol{x}, -\boldsymbol{\Delta D}^j) w_j \right\} \\
= & - \max_{\{\|\boldsymbol{r}\| \leq \Omega\}} \sum_{j \in N} \left\{ \max\{-f(\boldsymbol{x}, \boldsymbol{\Delta D}^j), -f(\boldsymbol{x}, -\boldsymbol{\Delta D}^j), 0\} r_j \right\}
\end{aligned}
$$

and using the definition of dual norm, $\|\boldsymbol{s}\|^* = \max_{\|\boldsymbol{x}\| \leq 1} \boldsymbol{s}'\boldsymbol{x}$, we obtain $\Omega\|\boldsymbol{s}\|^* = \max_{\|\boldsymbol{x}\| \leq \Omega} \boldsymbol{s}'\boldsymbol{x}$, i.e., Eq. (4.12) follows. Note that

$$s_j = \max\{-f(\boldsymbol{x}, \Delta \boldsymbol{D}^j), -f(\boldsymbol{x}, -\Delta \boldsymbol{D}^j)\} \geq 0,$$

since otherwise there exists an $\boldsymbol{x}$ such that $s_j < 0$, i.e.,

$$f(\boldsymbol{x}, \Delta \boldsymbol{D}^j) > 0 \text{ and } f(\boldsymbol{x}, -\Delta \boldsymbol{D}^j) > 0.$$

From Assumption 1(b) $f(\boldsymbol{x}, \boldsymbol{0}) = 0$, contradicting the concavity of $f(\boldsymbol{x}, \boldsymbol{D})$ (Assumption 1(a)).

Suppose that $\boldsymbol{x}$ is feasible in Problem (4.12). Defining $\boldsymbol{t} = \boldsymbol{s}$ and $y = \|\boldsymbol{s}\|^*$, we can easily check that $(\boldsymbol{x}, \boldsymbol{t}, y)$ are feasible in Problem (4.13). Conversely, suppose, $\boldsymbol{x}$ is infeasible in (4.12), that is,

$$f(\boldsymbol{x}, \boldsymbol{D}^0) < \Omega\|\boldsymbol{s}\|^*.$$

Since, $t_j \geq s_j = \max\{-f(\boldsymbol{x}, \Delta \boldsymbol{D}^j), -f(\boldsymbol{x}, -\Delta \boldsymbol{D}^j)\} \geq 0$ we apply Proposition 6(b) to obtain $\|\boldsymbol{t}\|^* \geq \|\boldsymbol{s}\|^*$. Thus,

$$f(\boldsymbol{x}, \boldsymbol{D}^0) < \Omega\|\boldsymbol{s}\|^* \leq \Omega\|\boldsymbol{t}\|^* \leq \Omega y,$$

i.e., $\boldsymbol{x}$ is infeasible in (4.13).

(b) It is immediate that Eq. (4.12) can be written in the form of Eq. (4.13). ■

In Table 4.2, we list the common choices of norms, the representation of their dual norms and the corresponding references.

4.2.2 Representation of $\max\{-f(\boldsymbol{x}, \Delta \boldsymbol{D}), -f(\boldsymbol{x}, -\Delta \boldsymbol{D})\}$

The function $g(\boldsymbol{x}, \Delta \boldsymbol{D}^j) = \max\{-f(\boldsymbol{x}, \Delta \boldsymbol{D}^j), -f(\boldsymbol{x}, -\Delta \boldsymbol{D}^j)\}$ naturally arises in Theorem 8. Recall that a norm satisfies $\|\boldsymbol{A}\| \geq 0$, $\|k\boldsymbol{A}\| = |k| \cdot \|\boldsymbol{A}\|$, $\|\boldsymbol{A} + \boldsymbol{B}\| \leq \|\boldsymbol{A}\| + \|\boldsymbol{B}\|$, and $\|\boldsymbol{A}\| = 0$, implies that $\boldsymbol{A} = \boldsymbol{0}$. We show next that the function

Norms	$\|\boldsymbol{u}\|$	$\|\boldsymbol{t}\|^* \leq y$	Ref.
l_2	$\|\boldsymbol{u}\|_2$	$\|\boldsymbol{t}\|_2 \leq y$	[7]
l_1	$\|\boldsymbol{u}\|_1$	$t_j \leq y, \forall j \in N$	Chap. 3
l_∞	$\|\boldsymbol{u}\|_\infty$	$\sum_{j\in N} t_j \leq y$	Chap. 3
$l_p,\ p \geq 1$	$\|\boldsymbol{u}\|_p$	$\left(\sum_{j\in N} t_j^{\frac{q}{q-1}}\right)^{\frac{q-1}{q}} \leq y$	Chap. 3
$l_2 \cap l_\infty$ norm	$\max\{\|\boldsymbol{u}\|_2, \Omega\|\boldsymbol{u}\|_\infty\}$	$\exists \boldsymbol{s} \in \Re^{\|N\|}$: $\|\boldsymbol{s} - \boldsymbol{t}\|_2 + \frac{1}{\Omega}\sum_{j\in N} s_j \leq y$ $\boldsymbol{s} \geq \boldsymbol{0}$	[7]
$l_1 \cap l_\infty$ norm	$\max\{\frac{1}{\Gamma}\|\boldsymbol{u}\|_1, \|\boldsymbol{u}\|_\infty\}$	$\exists \boldsymbol{s}, \in \Re^{\|N\|}, p \in \Re$: $\Gamma p + \sum_{j\in N} s_j \leq y$ $s_j + p \geq t_j, \forall j \in N$ $p \geq 0, \boldsymbol{s} \geq \boldsymbol{0}$	Chap. 3

Table 4.2: Representation of the dual norm for $\boldsymbol{t} \geq \boldsymbol{0}$.

$g(\boldsymbol{x}, \boldsymbol{A})$ satisfies all these properties except the last one, i.e., it behaves almost like a norm.

Proposition 8 *Under Assumption 1, the function*

$$g(\boldsymbol{x}, \boldsymbol{A}) = \max\{-f(\boldsymbol{x}, \boldsymbol{A}), -f(\boldsymbol{x}, -\boldsymbol{A})\}$$

satisfies the following properties:

(a) $g(\boldsymbol{x}, \boldsymbol{A}) \geq 0$,

(b) $g(\boldsymbol{x}, k\boldsymbol{A}) = |k| g(\boldsymbol{x}, \boldsymbol{A})$,

(c) $g(\boldsymbol{x}, \boldsymbol{A} + \boldsymbol{B}) \leq g(\boldsymbol{x}, \boldsymbol{A}) + g(\boldsymbol{x}, \boldsymbol{B})$.

Proof :

(a) Suppose there exists $\boldsymbol{x}$ such that $g(\boldsymbol{x}, \boldsymbol{A}) < 0$, i.e., $f(\boldsymbol{x}, \boldsymbol{A}) > 0$ and $f(\boldsymbol{x}, -\boldsymbol{A}) > 0$. From Assumption 1(b) $f(\boldsymbol{x}, \boldsymbol{0}) = 0$, contradicting the concavity of $f(\boldsymbol{x}, \boldsymbol{A})$ (Assumption 1(a)).

(b) For $k \geq 0$, we apply Assumption 1(b) and obtain

$$\begin{aligned} g(\boldsymbol{x}, k\boldsymbol{A}) &= \max\{-f(\boldsymbol{x}, k\boldsymbol{A}), -f(\boldsymbol{x}, -k\boldsymbol{A})\} \\ &= k\max\{-f(\boldsymbol{x}, \boldsymbol{A}), -f(\boldsymbol{x}, -\boldsymbol{A})\} \\ &= kg(\boldsymbol{x}, \boldsymbol{A}). \end{aligned}$$

Similarly, if $k < 0$ we have

$$g(\boldsymbol{x}, k\boldsymbol{A}) = \max\{-f(\boldsymbol{x}, -k(-\boldsymbol{A})), -f(\boldsymbol{x}, -k(\boldsymbol{A}))\} = -kg(\boldsymbol{x}, \boldsymbol{A}).$$

(c) Using Eq. (4.6) we obtain

$$g(\boldsymbol{x}, \boldsymbol{A}+\boldsymbol{B}) = g(\boldsymbol{x}, \frac{1}{2}(2\boldsymbol{A}+2\boldsymbol{B})) \leq \frac{1}{2}g(\boldsymbol{x}, 2\boldsymbol{A}) + \frac{1}{2}g(\boldsymbol{x}, 2\boldsymbol{B}) = g(\boldsymbol{x}, \boldsymbol{A}) + g(\boldsymbol{x}, \boldsymbol{B}).$$

■

Note that the function $g(\boldsymbol{x}, \boldsymbol{A})$ does not necessarily define a norm for $\boldsymbol{A}$, since $g(\boldsymbol{x}, \boldsymbol{A}) = 0$ does not necessarily imply $\boldsymbol{A} = \boldsymbol{0}$. However, for LP, QCQP. SOCP(1), SOCP(2) and SDP, and specific direction of data perturbation, $\boldsymbol{\Delta D}^j$, we can map $g(\boldsymbol{x}, \boldsymbol{\Delta D}^j)$ to a function of a norm such that

$$g(\boldsymbol{x}, \boldsymbol{\Delta D}^j) = \|\mathcal{H}(\boldsymbol{x}, \boldsymbol{\Delta D}^j)\|_g,$$

where $\mathcal{H}(\boldsymbol{x}, \boldsymbol{\Delta D}^j)$ is linear in $\boldsymbol{\Delta D}^j$ and defined as follows (see also the summary in Table 4.3):

(a) LP:
$f(\boldsymbol{x}, \boldsymbol{D}) = \boldsymbol{a}'\boldsymbol{x} - b$, where $\boldsymbol{D} = (\boldsymbol{a}, b)$ and $\boldsymbol{\Delta D}^j = (\boldsymbol{\Delta a}^j, \Delta b^j)$. Hence,

$$g(\boldsymbol{x}, \boldsymbol{\Delta D}^j) = \max\{-(\boldsymbol{\Delta a}^j)'\boldsymbol{x} + \Delta b^j, (\boldsymbol{\Delta a}^j)'\boldsymbol{x} - \Delta b^j\} = |(\boldsymbol{\Delta a}^j)'\boldsymbol{x} - \Delta b^j|.$$

(b) QCQP:
$f(\boldsymbol{x}, \boldsymbol{D}) = (d - (\boldsymbol{b}'\boldsymbol{x} + c))/2 - \sqrt{\|\boldsymbol{A}\boldsymbol{x}\|_2^2 + ((d + \boldsymbol{b}'\boldsymbol{x} + c)/2)^2}$, where

$\boldsymbol{D} = (\boldsymbol{A}, \boldsymbol{b}, c, d)$ and $\boldsymbol{\Delta D}^j = (\boldsymbol{\Delta A}^j, \boldsymbol{\Delta b}^j, \Delta c^j, 0)$. Therefore,

$$\begin{aligned} g(\boldsymbol{x}, \boldsymbol{\Delta D}^j) &= \max\left\{\frac{(\boldsymbol{\Delta b}^j)'\boldsymbol{x}+\Delta c^j}{2} + \sqrt{\|\boldsymbol{\Delta A}^j\boldsymbol{x}\|_2^2 + \left(\frac{(\boldsymbol{\Delta b}^j)'\boldsymbol{x}+\Delta c^j}{2}\right)^2},\right. \\ &\qquad \left. -\frac{(\boldsymbol{\Delta b}^j)'\boldsymbol{x}+\Delta c^j}{2} + \sqrt{\|\boldsymbol{\Delta A}^j\boldsymbol{x}\|_2^2 + \left(\frac{(\boldsymbol{\Delta b}^j)'\boldsymbol{x}+\Delta c^j}{2}\right)^2}\right\} \\ &= \left|\frac{(\boldsymbol{\Delta b}^j)'\boldsymbol{x}+\Delta c^j}{2}\right| + \sqrt{\|\boldsymbol{\Delta A}^j\boldsymbol{x}\|_2^2 + \left(\frac{(\boldsymbol{\Delta b}^j)'\boldsymbol{x}+\Delta c^j}{2}\right)^2}. \end{aligned}$$

(c) SOCP(1):

$f(\boldsymbol{x}, \boldsymbol{D}) = \boldsymbol{c}'\boldsymbol{x} + d - \|\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}\|_2^2$, where $\boldsymbol{D} = (\boldsymbol{A}, \boldsymbol{b}, \boldsymbol{c}, d)$ and $\boldsymbol{\Delta D}^j = (\boldsymbol{\Delta A}^j, \boldsymbol{\Delta b}^j, \boldsymbol{0}, 0)$. Therefore,

$$g(\boldsymbol{x}, \boldsymbol{\Delta D}^j) = \|\boldsymbol{\Delta A}^j\boldsymbol{x} + \boldsymbol{\Delta b}^j\|_2.$$

(d) SOCP(2):

$f(\boldsymbol{x}, \boldsymbol{D}) = \boldsymbol{c}'\boldsymbol{x} + d - \|\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}\|_2^2$, where $\boldsymbol{D} = (\boldsymbol{A}, \boldsymbol{b}, \boldsymbol{c}, d)$ and $\boldsymbol{\Delta D}^j = (\boldsymbol{\Delta A}^j, \boldsymbol{\Delta b}^j, \boldsymbol{\Delta c}^j, d^j)$. Therefore,

$$\begin{aligned} g(\boldsymbol{x}, \boldsymbol{\Delta D}^j) &= \max\left\{-(\boldsymbol{\Delta c}^j)'\boldsymbol{x} - \Delta d^j + \|\boldsymbol{\Delta A}^j\boldsymbol{x} + \boldsymbol{\Delta b}^j\|_2,\right. \\ &\qquad \left.(\boldsymbol{\Delta c}^j)'\boldsymbol{x} + \Delta d^j + \|\boldsymbol{\Delta A}^j\boldsymbol{x} + \boldsymbol{\Delta b}^j\|_2\right\} \\ &= |(\boldsymbol{\Delta c}^j)'\boldsymbol{x} + \Delta d^j| + \|\boldsymbol{\Delta A}^j\boldsymbol{x} + \boldsymbol{\Delta b}^j\|_2. \end{aligned}$$

(e) SDP:

$f(\boldsymbol{x}, \boldsymbol{D}) = \lambda_{min}(\sum_{j=1}^n \boldsymbol{A}_i x_i - \boldsymbol{B})$, where $\boldsymbol{D} = (\boldsymbol{A}_1, ..., \boldsymbol{A}_n, \boldsymbol{B})$ and $\boldsymbol{\Delta D}^j = (\boldsymbol{\Delta A}_1^j, ..., \boldsymbol{\Delta A}_n^j, \boldsymbol{\Delta B}^j)$. Therefore,

$$\begin{aligned} g(\boldsymbol{x}, \boldsymbol{\Delta D}^j) &= \max\left\{-\lambda_{min}(\textstyle\sum_{j=1}^n \boldsymbol{\Delta A}_i^j x_i - \boldsymbol{\Delta B}^j),\right. \\ &\qquad \left.-\lambda_{min}\left(-\left(\textstyle\sum_{j=1}^n \boldsymbol{\Delta A}_i^j x_i - mb\Delta B^j\right)\right)\right\} \\ &= \max\left\{\lambda_{max}\left(-\left(\textstyle\sum_{j=1}^n \boldsymbol{\Delta A}_i^j x_i - \boldsymbol{\Delta B}^j\right)\right),\right. \\ &\qquad \left.\lambda_{max}(\textstyle\sum_{j=1}^n \boldsymbol{\Delta A}_i^j x_i - \boldsymbol{\Delta B}^j)\right\} \\ &= \left\|\sum_{j=1}^n \boldsymbol{\Delta A}_i^j x_i - \boldsymbol{\Delta B}^j\right\|_2. \end{aligned}$$

Type	$\boldsymbol{r} = \mathcal{H}(\boldsymbol{x}, \boldsymbol{\Delta D}^j)$	$g(\boldsymbol{x}, \boldsymbol{\Delta D}^j) = \lVert \boldsymbol{r} \rVert_g$
LP	$r = (\boldsymbol{\Delta a}^j)'\boldsymbol{x} - \Delta b^j$	$\lvert r \rvert$
QCQP	$\boldsymbol{r} = \begin{bmatrix} \boldsymbol{r}_1 \\ r_0 \end{bmatrix}, \boldsymbol{r}_1 = \begin{bmatrix} \boldsymbol{\Delta A}^j \boldsymbol{x} \\ ((\boldsymbol{\Delta b}^j)'\boldsymbol{x} + \Delta c^j)/2 \end{bmatrix},$ $r_0 = ((\boldsymbol{\Delta b}^j)'\boldsymbol{x} + \Delta c^j)/2$	$\lVert \boldsymbol{r}_1 \rVert_2 + \lvert r_0 \rvert$
SOCP(1)	$\boldsymbol{r} = \boldsymbol{\Delta A}^j \boldsymbol{x} + \boldsymbol{\Delta b}^j$	$\lVert \boldsymbol{r} \rVert_2$
SOCP(2)	$\boldsymbol{r} = \begin{bmatrix} \boldsymbol{r}_1 \\ r_0 \end{bmatrix}, \boldsymbol{r}_1 = \boldsymbol{\Delta A}^j \boldsymbol{x} + \boldsymbol{\Delta b}^j,$ $r_0 = (\boldsymbol{\Delta c}^j)'\boldsymbol{x} + \Delta d^j$	$\lVert \boldsymbol{r}_1 \rVert_2 + \lvert r_0 \rvert$
SDP	$\boldsymbol{R} = \sum_{i=1}^n \boldsymbol{\Delta A}_i^j x_i - \boldsymbol{\Delta B}^j$	$\lVert \boldsymbol{R} \rVert_2$

Table 4.3: The function $\mathcal{H}(\boldsymbol{x}, \boldsymbol{\Delta D}^j)$ and the norm $\lVert\cdot\rVert_g$ for different conic optimization problems.

4.2.3 The Nature and Size of the Robust Problem

In this section, we discuss the nature and size of the proposed robust conic problem. Note that in the proposed robust model (4.13) for every uncertain conic constraint $f(\boldsymbol{x}, \tilde{\boldsymbol{D}})$ we add at most $|N| + 1$ new variables, $2|N|$ conic constraints of the same nature as the nominal problem and an additional constraint involving the dual norm. The nature of this constraint depends on the norm we use to describe the uncertainty set $\mathcal{U}$ defined in Eq. (4.4).

When all the data entries of the problem have independent random perturbations, by exploiting sparsity of the additional conic constraints, we can further reduce the size of the robust model. Essentially, we can express the model of uncertainty in the form of Eq. (4.3), for which $\tilde{z}_j$ is the independent random variable associated with the jth data element, and $\boldsymbol{\Delta D}^j$ contains mostly zeros except at the entries corresponding to the data element. As an illustration, consider the following semidefinite constraint,

$$\begin{pmatrix} a_1 & a_2 \\ a_2 & a_3 \end{pmatrix} x_1 + \begin{pmatrix} a_4 & a_5 \\ a_5 & a_6 \end{pmatrix} x_2 \succeq \begin{pmatrix} a_7 & a_8 \\ a_8 & a_9 \end{pmatrix},$$

such that each element in the data $\boldsymbol{d} = (a_1, \ldots, a_9)'$ has an independent random perturbation, that is $\tilde{a}_i = a_i^0 + \Delta a_i \tilde{z}_i$ and $\tilde{z}_i$ are independently distributed. Equivalently, in Eq. (4.3) we have

$$\tilde{\boldsymbol{d}} = \boldsymbol{d}^0 + \sum_{i=1}^{9} \boldsymbol{\Delta d}^i \tilde{z}_i,$$

where $\boldsymbol{d}^0 = (a_1^0, \ldots, a_9^0)'$ and $\boldsymbol{\Delta d}^i$ is a vector with Δa_i at the ith entry and zero, otherwise. Hence, we can simplify the conic constraint in Eq. (4.13), $f(\boldsymbol{x}, \boldsymbol{\Delta d}^1)+t_1 \geq 0$ or

$$\lambda_{min}\left(\left(\begin{array}{cc} \Delta a_1 & 0 \\ 0 & 0 \end{array}\right)x_1 + \left(\begin{array}{cc} 0 & 0 \\ 0 & 0 \end{array}\right)x_2 - \left(\begin{array}{cc} 0 & 0 \\ 0 & 0 \end{array}\right)\right) + t_1 \geq 0,$$

as $t_1 \geq -\min\{\Delta a_1 x_1, 0\}$ or equivalently as linear constraints $t_1 \geq -\Delta a_1 x_1, t_1 \geq 0$. In this section, we show that if each data entry of the model has independent uncertainty, we can substantially reduce the size of the robust formulation (4.13). We focus on the equivalent representation (4.12),

$$f(\boldsymbol{x}, \boldsymbol{D}^0) \geq \Omega y, \|\boldsymbol{s}\|^* \leq y,$$

where, $s_j = \max\{-f(\boldsymbol{x}, \boldsymbol{\Delta D}^j), -f(\boldsymbol{x}, -\boldsymbol{\Delta D}^j)\} = g(\boldsymbol{x}, \boldsymbol{\Delta D}^j)$, for $j \in N$.

In a more general setting, we will show that if each data entry of the model has independent uncertainty, we can substantially reduce the size of the robust formulation of (4.13). We will focus on the equivalent representation (4.12),

$$f(\boldsymbol{x}, \boldsymbol{D}^0) \geq \Omega y, \|\boldsymbol{s}\|^* \leq y, y \in \Re$$

where, $s_j = \max\{-f(\boldsymbol{x}, \boldsymbol{\Delta D}^j), -f(\boldsymbol{x}, -\boldsymbol{\Delta D}^j)\} = g(\boldsymbol{x}, \boldsymbol{\Delta D}^j)$, for $j \in N$.

Proposition 9 *For LP, QCQP, SOCP(1), SOCP(2) and SDP, we can express $s_j = |\Delta d_j x_{i(j)}|$ for which Δd_j, $j \in N$ are constants and the function, $i : N \to \{0, \ldots, n\}$ maps $j \in N$ to the index of the corresponding variable. We define $x_0 = 1$ to handle the case when s_j is not variable dependent.*

Proof :

In the following exposition, we associate the jth data entry, $j \in N$ with an iid random variable $\tilde{z}_j$. For the function of interest, the corresponding expression of $g(\boldsymbol{x}, \boldsymbol{\Delta D}^j)$ is shown in Table (4.3).

(a) LP:

Uncertain LP data is represented as $\tilde{\boldsymbol{D}} = (\tilde{\boldsymbol{a}}, \tilde{b})$, where

$$\tilde{a}_j = a_j^0 + \Delta a_j \tilde{z}_j \ \forall j = \{1, \ldots, n\}$$
$$\tilde{b} = b^0 + \Delta b \tilde{z}_{n+1}.$$

We have $|N| = n + 1$ and

$$s_j = |\Delta a_j x_j| \ \forall j = \{1, \ldots, n\}$$
$$s_{n+1} = |\Delta b|$$

(b) QCQP:

Uncertain QCQP data is represented as $\tilde{\boldsymbol{D}} = (\tilde{\boldsymbol{A}}, \tilde{\boldsymbol{b}}, \tilde{c}, 1)$, where

$$\tilde{A}_{kj} = A_{kj}^0 + \Delta A_{kj} \tilde{z}_{n(k-1)+j} \quad \forall j \in \{1, \ldots, n\}, k = \{1, \ldots, l\},$$
$$\tilde{b}_j = b_j^0 + \Delta b_j \tilde{z}_{nl+j} \quad \forall j \in \{1, \ldots, n\},$$
$$\tilde{c} = c^0 + \Delta c \tilde{z}_{n(l+1)+1}.$$

We have $|N| = n(l + 1) + 1$ and

$$s_{n(k-1)+j} = |\Delta A_{kj} x_j| \quad \forall j \in \{1, \ldots, n\}, k = \{1, \ldots, l\},$$
$$s_{nl+j} = |\Delta b_j x_j| \quad \forall j \in \{1, \ldots, n\},$$
$$s_{n(l+1)+1} = |\Delta c|.$$

(c) SOCP(1)/SOCP(2):

Uncertain SOCP(2) data is represented as $\tilde{\boldsymbol{D}} = (\tilde{\boldsymbol{A}}, \tilde{\boldsymbol{b}}, \tilde{\boldsymbol{c}}, d)$, where

$$\tilde{A}_{kj} = A_{kj}^0 + \Delta A_{kj} \tilde{z}_{n(k-1)+j} \quad \forall j \in \{1, \ldots, n\}, k = \{1, \ldots, l\},$$
$$\tilde{b}_k = b_k^0 + \Delta b_k \tilde{z}_{nl+k} \quad \forall k \in \{1, \ldots, l\},$$
$$\tilde{c}_j = c_j^0 + \Delta c_j \tilde{z}_{(n+1)l+j} \quad \forall j \in \{1, \ldots, n\},$$
$$\tilde{d} = d^0 + \Delta d \tilde{z}_{(n+1)l+n+1}.$$

We have $|N| = (n+1)l + n + 1$ and

$$\begin{aligned}
s_{n(k-1)+j} &= |\Delta A_{kj} x_j| \ \ \forall j \in \{1, \ldots, n\}, k = \{1, \ldots, l\},\\
s_{nl+k} &= |\Delta b_k| \ \ \forall j \in \{1, \ldots, l\},\\
s_{(n+1)l+j} &= |\Delta c_j x_j| \ \ \forall j \in \{1, \ldots, n\},\\
s_{(n+1)l+n+1} &= |\Delta d|.
\end{aligned}$$

Observe that SOCP(1) is a special case of SOCP(2) for which $|N| = (n+1)l$, that is, $s_j = 0$ for all $j > (n+1)l$.

(d) SDP:

Uncertain SDP data for is represented as $\tilde{\boldsymbol{D}} = (\tilde{\boldsymbol{A}}_1, \ldots, \tilde{\boldsymbol{A}}_n, \tilde{\boldsymbol{B}})$, where

$$\begin{aligned}
\tilde{\boldsymbol{A}}_i &= \boldsymbol{A}_i^0 + \textstyle\sum_{k=1}^m \sum_{j=1}^k [\Delta \boldsymbol{A}_i]_{jk} \boldsymbol{I}_{jk} \tilde{z}_{p(i,j,k)} \ \ \forall i \in \{1, \ldots, n\},\\
\tilde{\boldsymbol{B}} &= \boldsymbol{B}^0 + \textstyle\sum_{k=1}^n \sum_{j=1}^k [\Delta \boldsymbol{B}]_{jk} \boldsymbol{I}_{jk} \tilde{z}_{p(n+1,j,k)},
\end{aligned}$$

where the index function $p(i,j,k) = (i-1)(m(m+1)/2) + k(k-1)/2 + j$, and the symmetric matrix $\boldsymbol{I}_{jk} \in \Re^{m \times m}$ satisfies,

$$\boldsymbol{I}_{jk} = \begin{cases} (\boldsymbol{e}_j \boldsymbol{e}_k' + \boldsymbol{e}_k \boldsymbol{e}_j') & \text{if } j \neq k \\ \boldsymbol{e}_k \boldsymbol{e}_k' & \text{if } k = j \end{cases},$$

$\boldsymbol{e}_k$ being a unit vector with zero entries and 1 at the kth entry. Hence, $|N| = (n+1)(m(m+1))/2$. Note that if $j = k$, it is obvious that $\|\boldsymbol{I}_{jk}\|_2 = 1$. Otherwise, observe that $\boldsymbol{I}_{jk}$ has rank 2 and $\frac{1}{\sqrt{2}}(\boldsymbol{e}_j + \boldsymbol{e}_k)$ and $\frac{1}{\sqrt{2}}(\boldsymbol{e}_j - \boldsymbol{e}_k)$ are two eigenvectors of $\boldsymbol{I}_{jk}$ with corresponding eigenvalues 1 and -1. Hence, $\|\boldsymbol{I}_{jk}\|_2 = 1$ for all valid indices j and k. Therefore, we have

$$\begin{aligned}
s_{p(i,j,k)} &= |[\Delta \boldsymbol{A}_i]_{jk} x_i| \ \ \forall i \in \{1, \ldots, n\}, j, k \in \{1, \ldots, m\}, j \leq k\\
s_{p(n+1,j,k)} &= |[\Delta \boldsymbol{B}]_{jk}| \ \ \forall j, k \in \{1, \ldots, m\}, j \leq k
\end{aligned}$$

.

■

We define the set $J(l) = \{j : i(j) = l, j \in N\}$ for $l \in \{0, \ldots, n\}$. Following from Table (4.2), we have the following robust formulations under the different norms in the restriction set $\mathcal{V}$ of (4.8).

(a) l_∞-norm

The constraint $\|\boldsymbol{s}\|^* \leq y$ for l_∞-norm is equivalent to

$$\sum_{j \in N} |\Delta d_j x_{i(j)}| \leq y \Leftrightarrow \sum_{l=0}^{n} \left(\sum_{j \in J(l)} |\Delta d_j| \right) |x_l| \leq y$$

or

$$\begin{aligned} &\textstyle\sum_{j \in J(0)} |\Delta d_j| + \sum_{l=1}^{n} \left(\sum_{j \in J(l)} |\Delta d_j| \right) t_l \leq y \\ &\boldsymbol{t} \geq \boldsymbol{x}, \boldsymbol{t} \geq -\boldsymbol{x} \\ &\boldsymbol{t} \in \Re^n. \end{aligned}$$

We introduce additional $n+1$ variables, including the variable y and $2n+1$ linear constraints to the nominal problem.

(b) l_1-norm

The constraint $\|\boldsymbol{s}\|^* \leq y$ for l_1-norm is equivalent to

$$\max_{j \in N} |\Delta d_j x_{i(j)}| \leq y \Leftrightarrow \max_{l \in \{0, \ldots, n\}} \left(\max_{j \in J(l)} |\Delta d_j| \right) |x_l| \leq y$$

or

$$\begin{aligned} \max\nolimits_{j \in J(0)} |\Delta d_j| &\leq y \\ \max\nolimits_{j \in J(l)} |\Delta d_j| x_l &\leq y \quad \forall l \in \{1, \ldots, n\} \\ -\max\nolimits_{j \in J(l)} |\Delta d_j| x_l &\leq y \quad \forall l \in \{1, \ldots, n\}. \end{aligned}$$

We introduce an additional variable and $2n+1$ linear constraints to the nominal problem.

(c) $l_1 \cap l_\infty$-norm

The constraint $\|s\|^* \leq y$ for $l_1 \cap l_\infty$-norm is equivalent to

$$\begin{aligned}
&t_j \geq |\Delta d_j| x_{i(j)} && \forall j \in N \\
&t_j \geq -|\Delta d_j| x_{i(j)} && \forall j \in N \\
&\Gamma p + \textstyle\sum_{j \in N} r_j \leq y \\
&r_j + p \geq t_j, \forall j \in N \\
&\boldsymbol{r} \in \Re_+^{|N|},\ \boldsymbol{t} \in \Re^{|N|},\ p \in \Re_+,
\end{aligned}$$

leading to an additional of $2|N|+2$ variables and $4|N|+2$ linear constraints, including non-negative constraints, to the nominal problem.

(d) l_2-norm

The constraint $\|s\|^* \leq y$ for l_2-norm is equivalent to

$$\sqrt{\sum_{j \in N} (\Delta d_j x_{i(j)})^2} \leq y \Leftrightarrow \sqrt{\sum_{j \in J(0)} |\Delta d_j| + \sum_{l=1}^{n} \left(\sum_{j \in J(l)} \Delta d_j^2 \right) x_l^2} \leq y.$$

We only introduce an additional variable, y and a second order cone (SOC) constraint to the nominal problem.

(e) $l_2 \cap l_\infty$-norm

The constraint $\|s\|^* \leq y$ for l_2-norm is equivalent to

$$\begin{aligned}
&t_j \geq |\Delta d_j| x_{i(j)} && \forall j \in N \\
&t_j \geq -|\Delta d_j| x_{i(j)} && \forall j \in N \\
&\|\boldsymbol{r} - \boldsymbol{t}\|_2 + \tfrac{1}{\Omega} \textstyle\sum_{j \in N} r_j \leq y \\
&\boldsymbol{t} \in \Re^{|N|},\ \boldsymbol{r} \in \Re_+^{|N|}.
\end{aligned}$$

We introduce additional $2|N| + 1$ variables, one SOCP constraint and $3|N|$ linear constraints, including non-negative constraints, to the nominal problem.

In Table 4.4 we summarize the number of variables and constraints and their nature when the nominal problem is an LP, QCQP, SOCP (1) (only $\boldsymbol{A}, \boldsymbol{b}$ vary), SOCP (2) ($\boldsymbol{A}, \boldsymbol{b}, \boldsymbol{c}, d$ vary) and SDP for various choices of norms. Note that for the

	l_∞	l_∞	$l_1 \cap l_\infty$	l_2	$l_2 \cap l_\infty$
Num. Variables	$n+1$	1	$2\|N\|+2$	1	$2\|N\|+1$
Num. Linear Const.	$2n+1$	$2n+1$	$4\|N\|+2$	0	$3\|N\|$
Num. SOC Const.	0	0	0	1	1
LP	LP	LP	LP	SOCP	SOCP
QCQP	SOCP	SOCP	SOCP	SOCP	SOCP
SOCP(1)	SOCP	SOCP	SOCP	SOCP	SOCP
SOCP(2)	SOCP	SOCP	SOCP	SOCP	SOCP
SDP	SDP	SDP	SDP	SDP	SDP

Table 4.4: Size increase and nature of robust formulation when each data entry has independent uncertainty.

cases of the l_1, l_∞ and l_2 norms, we are able to collate terms so that the number of variables and constraints introduced is minimal. Furthermore, using the l_2 norm results in only one additional variable, one additional SOCP type of constraint, while maintaining the nature of the original conic optimization problem of SOCP and SDP. The use of other norms comes at the expense of more variables and constraints of the order of $|N|$, which is not very appealing for large problems.

4.3 Probabilistic Guarantees

In this section, we derive a guarantee on the probability that the robust solution is feasible, when the uncertain coefficients obey some natural probability distributions. An important component of our analysis is the relation among different norms. We denote by $\langle \ , \ \rangle$ the inner product on a vector space, $\Re^m$ or the space of m by m symmetric matrices, $\boldsymbol{S}^{m \times m}$. The inner product induces a norm $\sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle}$. For a vector space, the natural inner product is the Euclidian inner product, $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \boldsymbol{x}'\boldsymbol{y}$, and the induced norm is the Euclidian norm $\|\boldsymbol{x}\|_2$. For the space of symmetric matrices, the natural inner product is the trace product or $\langle \boldsymbol{X}, \boldsymbol{Y} \rangle = \text{trace}(\boldsymbol{X}\boldsymbol{Y})$ and the corresponding induced norm is the Frobenius norm, $\|\boldsymbol{X}\|_F$ (see Renegar [23]).

We analyze the relation of the inner product norm $\sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle}$ with the norm $\|\boldsymbol{x}\|_g$ defined in Table 4.3 for the conic optimization problems we consider. Since $\|\boldsymbol{x}\|_g$ and $\sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle}$ are valid norms in a finite dimensional space, there exist finite $\alpha_1, \alpha_2 > 0$

such that

$$\frac{1}{\alpha_1}\|\boldsymbol{r}\|_g \le \sqrt{\langle \boldsymbol{r}, \boldsymbol{r}\rangle} \le \alpha_2 \|\boldsymbol{r}\|_g, \tag{4.16}$$

for all $\boldsymbol{r}$ in the relevant space.

Proposition 10 *For the norm $\|\cdot\|_g$ defined in Table 4.3 for the conic optimization problems we consider, Eq. (4.16) holds with the following parameters:*

(a) *LP: $\alpha_1 = \alpha_2 = 1$.*

(b) *QCQP, SOCP(2): $\alpha_1 = \sqrt{2}$ and $\alpha_2 = 1$.*

(c) *SOCP(1): $\alpha_1 = \alpha_2 = 1$.*

(d) *SDP: $\alpha_1 = 1$ and $\alpha_2 = \sqrt{m}$.*

Proof :

(a) LP: For $r \in \Re$ and $\|r\|_g = |r|$, leading to Eq. (4.16) with $\alpha_1 = \alpha_2 = 1$.

(b) QCQP, SOCP(2): For $\boldsymbol{r} = (\boldsymbol{r}_1, r_0)' \in \Re^{l+1}$, let $a = \|\boldsymbol{r}_1\|_2$ and $b = |r_0|$. Since $a, b > 0$, using the inequality $a + b \le \sqrt{2}\sqrt{a^2 + b^2}$ and $\sqrt{a^2+b^2} \le a + b$, we have

$$\frac{1}{\sqrt{2}}\left(\|\boldsymbol{r}_1\|_2 + |r_0|\right) \le \sqrt{\boldsymbol{r}'\boldsymbol{r}} = \|\boldsymbol{r}\|_2 \le \|\boldsymbol{r}_1\|_2 + |r_0|$$

leading to Eq. (4.16) with $\alpha_1 = \sqrt{2}$ and $\alpha_2 = 1$.

(c) SOCP(1): For all $\boldsymbol{r}$, Eq. (4.16) holds with $\alpha_1 = \alpha_2 = 1$.

(d) Let λ_j, $j = 1, \ldots, m$ be the eigenvalues of the matrix $\boldsymbol{A}$. Since $\|\boldsymbol{A}\|_F = \sqrt{\text{trace}(\boldsymbol{A}^2)} = \sqrt{\sum_j \lambda_j^2}$ and $\|\boldsymbol{A}\|_2 = \max_j |\lambda_j|$, we have

$$\|\boldsymbol{A}\|_2 \le \|\boldsymbol{A}\|_F \le \sqrt{m}\|\boldsymbol{A}\|_2,$$

leading to Eq. (4.16) with $\alpha_1 = 1$ and $\alpha_2 = \sqrt{m}$. ■

The central result of the section is as follows.

Theorem 9 **(a)** *Under the model of uncertainty in Eq. (4.3), and given a feasible* *solution $\boldsymbol{x}$ in Eq. (4.7), then*

$$\mathrm{P}(f(\boldsymbol{x}, \tilde{\boldsymbol{D}}) < 0) \le \mathrm{P}\left(\| \sum_{j \in N} \boldsymbol{r}_j \tilde{z}_j \|_g > \Omega \|\boldsymbol{s}\|^* \right),$$

where

$$\boldsymbol{r}_j = \mathcal{H}(\boldsymbol{x}, \boldsymbol{\Delta D}^j), \qquad s_j = \|\boldsymbol{r}_j\|_g, \quad j \in N.$$

(b) *When we use the l_2-norm in Eq. (4.8), i.e., $\|\boldsymbol{s}\|^* = \|\boldsymbol{s}\|_2$, and under the assumption that z_j are normally and independently distributed with mean zero and variance one, i.e., $\tilde{\boldsymbol{z}} \sim \mathcal{N}(0, \boldsymbol{I})$, then*

$$\mathrm{P}\left(\left\| \sum_{j \in N} \boldsymbol{r}_j \tilde{z}_j \right\|_g > \Omega \sqrt{\sum_{j \in N} \|\boldsymbol{r}_j\|_g^2}, \right) \le \frac{\sqrt{e}\Omega}{\alpha} \exp\left(-\frac{\Omega^2}{2\alpha^2} \right), \tag{4.17}$$

where $\alpha = \alpha_1 \alpha_2$, α_1, α_2 derived in Proposition 10 and $\Omega > \alpha$.

Proof :

(a) We have

$$\begin{aligned}
& \mathrm{P}(f(\boldsymbol{x}, \tilde{\boldsymbol{D}}) < 0) \\
\le\ & \mathrm{P}\left(f(\boldsymbol{x}, \boldsymbol{D}^0) + f(\boldsymbol{x}, \sum_{j \in N} \boldsymbol{\Delta D}^j \tilde{z}_j) < 0 \right) && (4.18) \\
\le\ & \mathrm{P}\left(f(\boldsymbol{x}, \sum_{j \in N} \boldsymbol{\Delta D}^j \tilde{z}_j) < -\Omega \|\boldsymbol{s}\|^* \right) && (4.19) \\
\le\ & \mathrm{P}\left(\min\left(f(\boldsymbol{x}, \sum_{j \in N} \boldsymbol{\Delta D}^j \tilde{z}_j), f(\boldsymbol{x}, -\sum_{j \in N} \boldsymbol{\Delta D}^j \tilde{z}_j) \right) < -\Omega \|\boldsymbol{s}\|^* \right) \\
=\ & \mathrm{P}\left(g(\boldsymbol{x}, \sum_{j \in N} \boldsymbol{\Delta D}^j \tilde{z}_j) > \Omega \|\boldsymbol{s}\|^* \right) \\
=\ & \mathrm{P}\left(\|\mathcal{H}(\boldsymbol{x}, \sum_{j \in N} \boldsymbol{\Delta D}^j \tilde{z}_j)\|_g > \Omega \|\boldsymbol{s}\|^* \right) \\
=\ & \mathrm{P}\left(\| \sum_{j \in N} \mathcal{H}(\boldsymbol{x}, \boldsymbol{\Delta D}^j) \tilde{z}_j \|_g > \Omega \|\boldsymbol{s}\|^* \right) && (4.20) \\
=\ & \mathrm{P}\left(\| \sum_{j \in N} \boldsymbol{r}_j \tilde{z}_j \|_g > \Omega \|\boldsymbol{s}\|^* \right),
\end{aligned}$$

where inequality (4.18) follows from (4.6), inequality (4.19) follows from (4.12), $s_j = \|\mathcal{H}(\boldsymbol{x}, \boldsymbol{\Delta D}^j)\|_g$ and Eq. (4.20) follows from (4.12) $\mathcal{H}(\boldsymbol{x}, \boldsymbol{D})$ being linear in $\boldsymbol{D}$.

(b) Using, the relations $\|\boldsymbol{r}\|_g \leq \alpha_1\sqrt{\langle \boldsymbol{r}, \boldsymbol{r}\rangle}$ and $\|\boldsymbol{r}\|_g \geq \frac{1}{\alpha_2}\sqrt{\langle \boldsymbol{r}, \boldsymbol{r}\rangle}$ from Proposition 10, we obtain

$$\begin{aligned}
& \mathrm{P}\left(\left\|\sum_{j\in N} \boldsymbol{r}_j\tilde{z}_j\right\|_g > \Omega\sqrt{\sum_{j\in N}\|\boldsymbol{r}_j\|_g^2}\right) \\
= \; & \mathrm{P}\left(\left\|\sum_{j\in N} \boldsymbol{r}_j\tilde{z}_j\right\|_g^2 > \Omega^2\sum_{j\in N}\|\boldsymbol{r}_j\|_g^2\right) \\
\leq \; & \mathrm{P}\left(\alpha_1^2\alpha_2^2\left\langle \sum_{j\in N} \boldsymbol{r}_j\tilde{z}_j, \sum_{k\in N} \boldsymbol{r}_k\tilde{z}_k\right\rangle > \Omega^2\sum_{j\in N}\langle \boldsymbol{r}_j, \boldsymbol{r}_j\rangle\right) \\
= \; & \mathrm{P}\left(\alpha^2\sum_{j\in N}\sum_{k\in N}\langle \boldsymbol{r}_j, \boldsymbol{r}_k\rangle\tilde{z}_j\tilde{z}_k > \Omega^2\sum_{j\in N}\langle \boldsymbol{r}_j, \boldsymbol{r}_j\rangle\right) \\
= \; & \mathrm{P}\left(\alpha^2\tilde{\boldsymbol{z}}'\boldsymbol{R}\tilde{\boldsymbol{z}} > \Omega^2\sum_{j\in N}\langle \boldsymbol{r}_j, \boldsymbol{r}_j\rangle\right),
\end{aligned}$$

where $R_{jk} = \langle \boldsymbol{r}_j, \boldsymbol{r}_k\rangle$. Clearly, $\boldsymbol{R}$ is a symmetric positive semidefinite matrix and can be spectrally decomposed such that $\boldsymbol{R} = \boldsymbol{Q}'\boldsymbol{\Lambda}\boldsymbol{Q}$, where $\boldsymbol{\Lambda}$ is the diagonal matrix of the eigenvalues and $\boldsymbol{Q}$ is the corresponding orthonormal matrix. Let $\tilde{\boldsymbol{y}} = \boldsymbol{Q}\tilde{\boldsymbol{z}}$ so that $\tilde{\boldsymbol{z}}'\boldsymbol{R}\tilde{\boldsymbol{z}} = \tilde{\boldsymbol{y}}'\boldsymbol{\Lambda}\tilde{\boldsymbol{y}} = \sum_{j\in N}\lambda_j\tilde{y}_j^2$. Since $\tilde{\boldsymbol{z}} \sim \mathcal{N}(0, \boldsymbol{I})$, we also have $\tilde{\boldsymbol{y}} \sim \mathcal{N}(0, \boldsymbol{I})$, that is, $\tilde{y}_j$, $j \in N$ are independent and normally distributed. Moreover,

$$\sum_{j\in N}\lambda_j = \mathrm{trace}(\boldsymbol{R}) = \sum_{j\in N}\langle \boldsymbol{r}_j, \boldsymbol{r}_j\rangle.$$

Therefore,

$$
\begin{aligned}
& \mathrm{P}\left(\alpha^2 \tilde{\boldsymbol{z}}' \boldsymbol{R} \tilde{\boldsymbol{z}} > \Omega^2 \sum_{j\in N} \langle \boldsymbol{r}_j, \boldsymbol{r}_j \rangle\right) \\
&= \mathrm{P}\left(\alpha^2 \sum_{j\in N} \lambda_j \tilde{y}_j^2 > \Omega^2 \sum_{j\in N} \lambda_j\right) \\
&\leq \frac{E\left(\exp\left(\theta\alpha^2 \sum_{j\in N} \lambda_j \tilde{y}_j^2\right)\right)}{\exp\left(\theta\Omega^2 \sum_{j\in N}\lambda_j\right)} && \text{(From Markov's inequality, } \theta > 0\text{)} \\
&= \frac{\prod_{j\in N} E\left(\exp\left(\theta\alpha^2 \lambda_j \tilde{y}_j^2\right)\right)}{\exp\left(\theta\Omega^2 \sum_{j\in N}\lambda_j\right)} && (\tilde{y}_j^2 \text{ are independent}) \\
&= \frac{\prod_{j\in N} E\left(\exp\left(\frac{\tilde{y}_j^2}{\beta}\right)^{\theta\alpha^2\lambda_j\beta}\right)}{\exp\left(\theta\Omega^2 \sum_{j\in N}\lambda_j\right)} && \text{for all } \beta > 2 \text{ and } \theta\alpha^2\lambda_j\beta \leq 1, \forall j \in N \\
&\leq \frac{\prod_{j\in N} \left(E\left(\exp\left(\frac{\tilde{y}_j^2}{\beta}\right)\right)^{\theta\alpha^2\lambda_j\beta}\right)}{\exp\left(\theta\Omega^2 \sum_{j\in N}\lambda_j\right)},
\end{aligned}
$$

where the last inequality follows from Jensen inequality, noting that $x^{\theta\alpha^2\lambda_j\beta}$ is a concave function of x if $\theta\alpha^2\lambda_j\beta \in [0,1]$. Since $\tilde{y}_j \sim \mathcal{N}(0,1)$,

$$
E\left(\exp\left(\frac{\tilde{y}_j^2}{\beta}\right)\right) = \frac{1}{\sqrt{2\pi}} \int_{\infty}^{\infty} \exp\left(-\frac{y^2}{2}\left(\frac{\beta-2}{\beta}\right)\right) dy = \sqrt{\frac{\beta}{\beta-2}}.
$$

Thus, we obtain

$$
\begin{aligned}
\frac{\prod_{j\in N} \left(E\left(\exp\left(\frac{\tilde{y}_j^2}{\beta}\right)\right)^{\theta\alpha^2\lambda_j\beta}\right)}{\exp\left(\theta\Omega^2 \sum_{j\in N}\lambda_j\right)} &= \frac{\prod_{j\in N}\left(\exp\left(\theta\alpha^2\lambda_j\beta\frac{1}{2}\ln\left(\frac{\beta}{\beta-2}\right)\right)\right)}{\exp\left(\theta\Omega^2 \sum_{j\in N}\lambda_j\right)} \\
&= \frac{\exp\left(\theta\alpha^2\beta\frac{1}{2}\ln\left(\frac{\beta}{\beta-2}\right)\sum_{j\in N}\lambda_j\right)}{\exp\left(\theta\Omega^2 \sum_{j\in N}\lambda_j\right)}.
\end{aligned}
$$

We select $\theta = 1/(\alpha^2\beta\lambda^*)$, where $\lambda^* = \max_{j\in N}\lambda_j$, and obtain

$$
\frac{\exp\left(\theta\alpha^2\beta\frac{1}{2}\ln\left(\frac{\beta}{\beta-2}\right)\sum_{j\in N}\lambda_j\right)}{\exp\left(\theta\Omega^2 \sum_{j\in N}\lambda_j\right)} = \exp\left(\rho\left(\frac{1}{2}\ln\left(\frac{\beta}{\beta-2}\right) - \frac{\Omega^2}{\alpha^2\beta}\right)\right),
$$

Type	Probability bound of infeasibility
LP	$\sqrt{e}\Omega\exp(-\frac{\Omega^2}{2})$
QCQP	$\sqrt{\frac{e}{2}}\Omega\exp(-\frac{\Omega^2}{4})$
SOCP(1)	$\sqrt{e}\Omega\exp(-\frac{\Omega^2}{2})$
SOCP(2)	$\sqrt{\frac{e}{2}}\Omega\exp(-\frac{\Omega^2}{4})$
SDP	$\sqrt{\frac{e}{m}}\Omega\exp(-\frac{\Omega^2}{2m})$

Table 4.5: Probability bounds of $\mathrm{P}(f(\boldsymbol{x},\tilde{\boldsymbol{D}})<0)$ for $\tilde{\boldsymbol{z}}\sim\mathcal{N}(0,\boldsymbol{I})$.

where $\rho=(\sum_{j\in N}\lambda_j)/\lambda^*$. Taking derivatives and choosing the best β, we have

$$\beta=\frac{2\Omega^2}{\Omega^2-\alpha^2},$$

for which $\Omega>\alpha$. Substituting and simplifying, we have

$$\exp\left(\rho\left(\frac{1}{2}\ln\left(\frac{\beta}{\beta-2}\right)-\frac{\Omega^2}{\alpha^2\beta}\right)\right)=\left(\frac{\sqrt{e}\Omega}{\alpha}\exp(-\frac{\Omega^2}{2\alpha^2})\right)^\rho\leq\frac{\sqrt{e}\Omega}{\alpha}\exp(-\frac{\Omega^2}{2\alpha^2}),$$

where the last inequality follows from $\rho\geq 1$, and from $\frac{\sqrt{e}\Omega}{\alpha}\exp(-\frac{\Omega^2}{2\alpha^2})<1$ for $\Omega>\alpha$.
■

Note that $f(\boldsymbol{x},\tilde{\boldsymbol{D}})<0$, implies that $\|\tilde{\boldsymbol{z}}\|>\Omega$. Thus, when $\tilde{\boldsymbol{z}}\sim\mathcal{N}(0,\boldsymbol{I})$

$$\mathrm{P}(f(\boldsymbol{x},\tilde{\boldsymbol{D}})<0)\leq\mathrm{P}(\|\tilde{\boldsymbol{z}}\|>\Omega)=1-\chi^2_{|N|}(\Omega^2), \tag{4.21}$$

where $\chi^2_{|N|}(\cdot)$ is the cdf of a χ-square distribution with $|N|$ degrees of freedom. Note that the bound (4.21) does not take into account the structure of $f(\boldsymbol{x},\tilde{\boldsymbol{D}})$ in contrast to bound (4.17) that depends on $f(\boldsymbol{x},\tilde{\boldsymbol{D}})$ via the parameter α. To illustrate this, we substitute the value of the parameter α from Proposition 10 in Eq. (4.17) and report in Table 4.6 the bound in Eq. (4.17).

To amplify the previous discussion, we show in Table 4.6 the value of Ω in order for the bound (4.17) to be less than or equal to ϵ. The last column shows the value of Ω using bound (4.21) that is independent of the structure of the problem. We choose $|N|=495000$ which is approximately the maximum number of data entries in

ϵ	LP	QCQP	SOCP(1)	SOCP(2)	SDP	Eq. (4.21)
10^{-1}	2.76	3.91	2.76	3.91	27.6	704.5
10^{-2}	3.57	5.05	3.57	5.05	35.7	705.2
10^{-3}	4.21	5.95	4.21	5.95	42.1	705.7
10^{-6}	5.68	7.99	5.68	7.99	56.8	706.9

Table 4.6: Sample calculations of Ω using Probability Bounds of Table 4.5 for $m = 100$, $n = 100$, $|N| = 495,000$.

a SDP constraint with $n = 100$ and $m = 100$. Although the size $|N|$ is unrealistic for constraints with less data entries such as LP, the derived probability bounds remain valid. Note that bound (4.21) leads to $\Omega = O(\sqrt{|N|}\ln(1/\epsilon))$.

For LP, SOCP, and QCQP, bound (4.17) leads to $\Omega = O(\ln(1/\epsilon))$, which is independent of the dimension of the problem. For SDP it leads to we have $\Omega = O(\sqrt{m}\ln(1/\epsilon))$. As a result, ignoring the structure of the problem and using bound (4.21) leads to very conservative solutions.

4.4 General Cones

In this section, we generalize the results in Sections 4.1-4.3 to arbitrary conic constraints of the form,

$$\sum_{j=1}^{n} \tilde{\boldsymbol{A}}_j x_j \succeq_{\boldsymbol{K}} \tilde{\boldsymbol{B}}, \tag{4.22}$$

where $\{\tilde{\boldsymbol{A}}_1, ..., \tilde{\boldsymbol{A}}_n, \tilde{\boldsymbol{B}}\} = \tilde{\boldsymbol{D}}$ constitutes the set of data that is subject to uncertainty, and $\boldsymbol{K}$ is a closed, convex, pointed cone with nonempty interior. For notational simplicity, we define

$$\mathcal{A}(\boldsymbol{x}, \tilde{\boldsymbol{D}}) = \sum_{j=1}^{n} \tilde{\boldsymbol{A}}_j x_j - \tilde{\boldsymbol{B}}$$

so that Eq. (4.22) is equivalent to

$$\mathcal{A}(\boldsymbol{x}, \tilde{\boldsymbol{D}}) \succeq_{\boldsymbol{K}} \boldsymbol{0}. \tag{4.23}$$

We assume that the model for data uncertainty is given in Eq. (4.3) with $\tilde{\boldsymbol{z}} \sim \mathcal{N}(0, \boldsymbol{I})$. The uncertainty set $\mathcal{U}$ satisfies Eq. (4.4) with the given norm satisfying $\|\boldsymbol{u}\| = \|\boldsymbol{u}^+\|$.

Paralleling the earlier development, starting with a cone $\boldsymbol{K}$ and constraint (4.23), we define the function $f(\cdot,\cdot)$ as follows so that $f(\boldsymbol{x}, \boldsymbol{D}) > 0$ if and only if $\mathcal{A}(\boldsymbol{x}, \boldsymbol{D}) \succ_{\boldsymbol{K}} \boldsymbol{0}$.

Proposition 11 *For any $\boldsymbol{V} \succ_{\boldsymbol{K}} 0$, the function*

$$\begin{aligned} f(\boldsymbol{x}, \boldsymbol{D}) = \quad & \max \quad \theta \\ & \text{s.t.} \quad \mathcal{A}(\boldsymbol{x}, \boldsymbol{D}) \succeq_{\boldsymbol{K}} \theta \boldsymbol{V}, \end{aligned} \tag{4.24}$$

satisfies the properties:

(a) *$f(\boldsymbol{x}, \boldsymbol{D})$ is bounded and concave in $\boldsymbol{x}$ and $\boldsymbol{D}$.*

(b) *$f(\boldsymbol{x}, k\boldsymbol{D}) = kf(\boldsymbol{x}, \boldsymbol{D}), \forall k \geq 0$.*

(c) *$f(\boldsymbol{x}, \boldsymbol{D}) \geq y$ if and only if $\mathcal{A}(\boldsymbol{x}, \boldsymbol{D}) \succeq_{\boldsymbol{K}} y\boldsymbol{V}$.*

(d) *$f(\boldsymbol{x}, \boldsymbol{D}) > y$ if an donly if $\mathcal{A}(\boldsymbol{x}, \boldsymbol{D}) \succ_{\boldsymbol{K}} y\boldsymbol{V}$.*

Proof :

(a) Consider the dual of Problem (4.24):

$$\begin{aligned} z^* = \quad & \min \quad \langle \boldsymbol{u}, \mathcal{A}(\boldsymbol{x}, \boldsymbol{D}) \rangle \\ & \text{s.t.} \quad \langle \boldsymbol{u}, \boldsymbol{V} \rangle = 1 \\ & \qquad \boldsymbol{u} \succeq_{\boldsymbol{K}^*} \boldsymbol{0}, \end{aligned}$$

where $\boldsymbol{K}^*$ is the dual cone of $\boldsymbol{K}$. Since $\boldsymbol{K}$ is a closed, convex, pointed cone with nonempty interior, so is $\boldsymbol{K}^*$ (see Ben-Tal and Nemirovski [8]). As $\boldsymbol{V} \succ_{\boldsymbol{K}} \boldsymbol{0}$, for all $\boldsymbol{u} \succeq_{\boldsymbol{K}^*} \boldsymbol{0}$ and $\boldsymbol{u} \neq \boldsymbol{0}$, we have $\langle \boldsymbol{u}, \boldsymbol{V} \rangle > 0$, hence, the dual problem is bounded. Furthermore, since $\boldsymbol{K}^*$ has a nonempty interior, the dual problem is strictly feasible, i.e., there exists $\boldsymbol{u} \succ_{\boldsymbol{K}^*} \boldsymbol{0}$, $\langle \boldsymbol{u}, \boldsymbol{V} \rangle = 1$. Therefore, by conic duality, the dual objective z^* has the same finite objective as the primal objective function $f(\boldsymbol{x}, \boldsymbol{D})$. Since $\mathcal{A}(\boldsymbol{x}, \boldsymbol{D})$ is a linear mapping of $\boldsymbol{D}$ and an affine mapping of $\boldsymbol{x}$, it follows that $f(\boldsymbol{x}, \boldsymbol{D})$ is concave in $\boldsymbol{x}$ and $\boldsymbol{D}$.

(b) Using the dual expression of $f(\boldsymbol{x}, \boldsymbol{D})$, and that $\mathcal{A}(\boldsymbol{x}, k\boldsymbol{D}) = k\mathcal{A}(\boldsymbol{x}, \boldsymbol{D})$, the result follows.

(c) If $\theta = y$ is feasible in Problem (4.24), we have $f(\boldsymbol{x}, \boldsymbol{D}) \geq \theta = y$. Conversely, if $f(\boldsymbol{x}, \boldsymbol{D}) \geq y$, then $\mathcal{A}(\boldsymbol{x}, \boldsymbol{D}) \succeq_{\boldsymbol{K}} f(\boldsymbol{x}, \boldsymbol{D})\boldsymbol{V} \succeq_{\boldsymbol{K}} y\boldsymbol{V}$.

(d) Suppose $\mathcal{A}(\boldsymbol{x}, \boldsymbol{D}) \succ_{\boldsymbol{K}} y\boldsymbol{V}$, then there exists $\epsilon > 0$ such that $\mathcal{A}(\boldsymbol{x}, \boldsymbol{D}) - y\boldsymbol{V} \succeq_{\boldsymbol{K}} \epsilon\boldsymbol{V}$ or $\mathcal{A}(\boldsymbol{x}, \boldsymbol{D}) \succeq_{\boldsymbol{K}} (\epsilon + y)\boldsymbol{V}$. Hence, $f(\boldsymbol{x}, \boldsymbol{D}) \geq \epsilon + y > y$. Conversely, since $\boldsymbol{V} \succ_{\boldsymbol{K}} \boldsymbol{0}$, if $f(\boldsymbol{x}, \boldsymbol{D}) > y$ then $(f(\boldsymbol{x}, \boldsymbol{D}) - y)\boldsymbol{V} \succ_{\boldsymbol{K}} \boldsymbol{0}$. Hence, $\mathcal{A}(\boldsymbol{x}, \boldsymbol{D}) \succeq_{\boldsymbol{K}} f(\boldsymbol{x}, \boldsymbol{D})\boldsymbol{V} \succ_{\boldsymbol{K}} y\boldsymbol{V}$. ■

Remark : With $y = 0$, (c) establishes that $\mathcal{A}(\boldsymbol{x}, \boldsymbol{D}) \succeq_{\boldsymbol{K}} \boldsymbol{0}$ if and only if $f(\boldsymbol{x}, \boldsymbol{D}) \geq 0$ and (d) establishes that $\mathcal{A}(\boldsymbol{x}, \boldsymbol{D}) \succ_{\boldsymbol{K}} \boldsymbol{0}$ if and only if $f(\boldsymbol{x}, \boldsymbol{D}) > 0$.

The proposed robust model is given in Eqs. (4.7) and (4.8). We next derive an expression for $g(\boldsymbol{x}, \Delta\boldsymbol{D}) = \max\{-f(\boldsymbol{x}, \Delta\boldsymbol{D}), -f(\boldsymbol{x}, -\Delta\boldsymbol{D})\}$.

Proposition 12 *Let* $g(\boldsymbol{x}, \Delta\boldsymbol{D}) = \max\{-f(\boldsymbol{x}, \Delta\boldsymbol{D}), -f(\boldsymbol{x}, -\Delta\boldsymbol{D})\}$. *Then*

$$g(\boldsymbol{x}, \Delta\boldsymbol{D}) = \|\mathcal{H}(\boldsymbol{x}, \Delta\boldsymbol{D})\|_g,$$

where $\mathcal{H}(\boldsymbol{x}, \Delta\boldsymbol{D}) = \mathcal{A}(\boldsymbol{x}, \Delta\boldsymbol{D})$ *and*

$$\|\boldsymbol{S}\|_g = \min\{y : y\boldsymbol{V} \succeq_{\boldsymbol{K}} \boldsymbol{S} \succeq_{\boldsymbol{K}} -y\boldsymbol{V}\}.$$

Proof :

We observe that

$$\begin{aligned} g(\boldsymbol{x}, \Delta\boldsymbol{D}) &= \max\{-f(\boldsymbol{x}, \Delta\boldsymbol{D}), -f(\boldsymbol{x}, -\Delta\boldsymbol{D})\} \\ &= \min\{y \mid -f(\boldsymbol{x}, \Delta\boldsymbol{D}) \leq y, -f(\boldsymbol{x}, -\Delta\boldsymbol{D}) \leq y\} \\ &= \min\{y \mid \mathcal{A}(\boldsymbol{x}, \Delta\boldsymbol{D}) \succeq_{\boldsymbol{K}} -y\boldsymbol{V}, -\mathcal{A}(\boldsymbol{x}, \Delta\boldsymbol{D}) \succeq_{\boldsymbol{K}} -y\boldsymbol{V}\}, \quad (4.25) \\ &= \|\mathcal{A}(\boldsymbol{x}, \Delta\boldsymbol{D})\|_g. \end{aligned}$$

We also need to show that $\|.\|_g$ is indeed a valid norm. Since $\boldsymbol{V} \succ_{\boldsymbol{K}} \boldsymbol{0}$, then $\|\boldsymbol{S}\|_g \geq 0$. Clearly, $\|\boldsymbol{0}\|_g = 0$ and if $\|\boldsymbol{S}\|_g = 0$, then $\boldsymbol{0} \succeq_{\boldsymbol{K}} \boldsymbol{S} \succeq_{\boldsymbol{K}} \boldsymbol{0}$, which implies that $\boldsymbol{S} = \boldsymbol{0}$.

To show that $\|k\boldsymbol{S}\|_g = |k|\|\boldsymbol{S}\|_g$, we observe that for $k > 0$,

$$\begin{aligned}\|k\boldsymbol{S}\|_g &= \min\{y \mid y\boldsymbol{V} \succeq_{\boldsymbol{K}} k\boldsymbol{S} \succeq_{\boldsymbol{K}} -y\boldsymbol{V}\} \\ &= k\min\left\{\frac{y}{k} \mid \frac{y}{k}\boldsymbol{V} \succeq_{\boldsymbol{K}} \boldsymbol{S} \succeq_{\boldsymbol{K}} -\frac{y}{k}\boldsymbol{V}\right\} \\ &= k\|\boldsymbol{S}\|_g.\end{aligned}$$

Likewise, if $k < 0$

$$\begin{aligned}\|k\boldsymbol{S}\|_g &= \min\{y \mid y\boldsymbol{V} \succeq_{\boldsymbol{K}} k\boldsymbol{S} \succeq_{\boldsymbol{K}} -y\boldsymbol{V}\} \\ &= \min\{y \mid y\boldsymbol{V} \succeq_{\boldsymbol{K}} -k\boldsymbol{S} \succeq_{\boldsymbol{K}} -y\boldsymbol{V}\} \\ &= \|-k\boldsymbol{S}\|_g \\ &= -k\|\boldsymbol{S}\|_g.\end{aligned}$$

Finally, to verify triangle inequality,

$$\begin{aligned}\|\boldsymbol{S}\|_g + \|\boldsymbol{T}\|_g &= \min\{y \mid y\boldsymbol{V} \succeq_{\boldsymbol{K}} \boldsymbol{S} \succeq_{\boldsymbol{K}} -y\boldsymbol{V}\} + \min\{z \mid z\boldsymbol{V} \succeq_{\boldsymbol{K}} \boldsymbol{T} \succeq_{\boldsymbol{K}} -z\boldsymbol{V}\} \\ &= \min\{y + z \mid y\boldsymbol{V} \succeq_{\boldsymbol{K}} \boldsymbol{S} \succeq_{\boldsymbol{K}} -y\boldsymbol{V},\ z\boldsymbol{V} \succeq_{\boldsymbol{K}} \boldsymbol{T} \succeq_{\boldsymbol{K}} -z\boldsymbol{V}\} \\ &\geq \min\{y + z \mid (y+z)\boldsymbol{V} \succeq_{\boldsymbol{K}} \boldsymbol{S} + \boldsymbol{T} \succeq_{\boldsymbol{K}} -(y+z)\boldsymbol{V}\} \\ &= \|\boldsymbol{S} + \boldsymbol{T}\|_g.\end{aligned}$$

■

For the general conic constraint, the norm, $\|\cdot\|_g$ is dependent on the cone $\boldsymbol{K}$ and a point in the interior of the cone $\boldsymbol{V}$. Hence, we define $\|\cdot\|_{\boldsymbol{K},\boldsymbol{V}} := \|\cdot\|_g$. Using Proposition 11 and Theorem 8 we next show that the robust counterpart for the conic constraint (4.23) is tractable and provide a bound on the probability that the constraint is feasible.

Theorem 10 *We have*

(a) **(Tractability)** *For a norm satisfying Eq. (4.5), constraint (4.7) for general* *cones is equivalent to*

$$\begin{aligned}
&\mathcal{A}(\boldsymbol{x}, \boldsymbol{D}^0) \succeq_{\boldsymbol{K}} \Omega y \boldsymbol{V}, \\
&t_j \boldsymbol{V} \succeq_{\boldsymbol{K}} \mathcal{A}(\boldsymbol{x}, \boldsymbol{\Delta D}^j) \succeq_{\boldsymbol{K}} -t_j \boldsymbol{V}, \quad j \in N, \\
&\|\boldsymbol{t}\|^* \le y, \\
&y \in \Re, \ \boldsymbol{t} \in \Re^{|N|},
\end{aligned}$$

(b) (Probabilistic guarantee) *When we use the l_2-norm in Eq. (4.8), i.e., $\|\boldsymbol{s}\|^* = \|\boldsymbol{s}\|_2$, and under the assumption that $\tilde{\boldsymbol{z}} \sim \mathcal{N}(0, \boldsymbol{I})$, then for all $\boldsymbol{V}$ we have*

$$\mathrm{P}(\mathcal{A}(\boldsymbol{x}, \tilde{\boldsymbol{D}}) \notin \boldsymbol{K}) \le \frac{\sqrt{e}\Omega}{\alpha_{\boldsymbol{K},\boldsymbol{V}}} \exp\left(-\frac{\Omega^2}{2\alpha^2_{\boldsymbol{K},\boldsymbol{V}}}\right),$$

where

$$\alpha_{\boldsymbol{K},\boldsymbol{V}} = \left(\max_{\sqrt{\langle \boldsymbol{S},\boldsymbol{S}\rangle}=1} \|\boldsymbol{S}\|_{\boldsymbol{K},\boldsymbol{V}}\right)\left(\max_{\|\boldsymbol{S}\|_{\boldsymbol{K},\boldsymbol{V}}=1} \sqrt{\langle \boldsymbol{s}, \boldsymbol{s}\rangle}\right)$$

and

$$\|\boldsymbol{S}\|_{\boldsymbol{K},\boldsymbol{V}} = \min\left\{y : y\boldsymbol{V} \succeq_{\boldsymbol{K}} \boldsymbol{S} \succeq_{\boldsymbol{K}} -y\boldsymbol{V}\right\}.$$

Proof :

The Theorem follows directly from Propositions 11, 12, Theorems 8, 9. ■

From Theorem 10, for any cone $\boldsymbol{K}$, we select $\boldsymbol{V}$ in order to minimize $\alpha_{\boldsymbol{K},\boldsymbol{V}}$, i.e.,

$$\alpha_{\boldsymbol{K}} = \min_{\boldsymbol{V} \succ \boldsymbol{0}} \alpha_{\boldsymbol{K},\boldsymbol{V}}.$$

We next show that the smallest parameter α is $\sqrt{2}$ and $\sqrt{m}$ for SOCP and SDP respectively. For the second order cone, $\boldsymbol{K} = \boldsymbol{L}^{n+1}$,

$$\boldsymbol{L}^{n+1} = \{\boldsymbol{x} \in \Re^{n+1} : \|\boldsymbol{x}_n\|_2 \le x_{n+1}\},$$

where $\boldsymbol{x}_n = (x_1, \ldots, x_n)'$. The induced norm is given by

$$\begin{aligned} \|\boldsymbol{x}\|_{\boldsymbol{L}^{n+1},\boldsymbol{v}} &= \min\{y : y\boldsymbol{v} \succeq_{\boldsymbol{L}^{n+1}} \boldsymbol{x} \succeq_{\boldsymbol{L}^{n+1}} -y\boldsymbol{v}\} \\ &= \min\{y : \|\boldsymbol{x}_n + \boldsymbol{v}_n y\|_2 \leq v_{n+1}y + x_{n+1}, \|\boldsymbol{x}_n - \boldsymbol{v}_n y\|_2 \leq v_{n+1}y - x_{n+1}, \} \end{aligned}$$

and

$$\alpha_{\boldsymbol{L}^{n+1},\boldsymbol{v}} = \left(\max_{\|\boldsymbol{x}\|_2=1} \|\boldsymbol{x}\|_{\boldsymbol{L}^{n+1},\boldsymbol{v}}\right)\left(\max_{\|\boldsymbol{x}\|_{\boldsymbol{L}^{n+1},\boldsymbol{v}}=1} \|\boldsymbol{x}\|_2\right).$$

For the symmetric positive semidefinite cone, $\boldsymbol{K} = \boldsymbol{S}_+^m$,

$$\|\boldsymbol{X}\|_{\boldsymbol{S}_+^m,\boldsymbol{V}} = \min\{y : y\boldsymbol{V} \succeq \boldsymbol{X} \succeq -y\boldsymbol{V}\},$$

$$\alpha_{\boldsymbol{S}_+^m,\boldsymbol{V}} = \left(\max_{\sqrt{\langle \boldsymbol{X},\boldsymbol{X}\rangle}=1} \|\boldsymbol{x}\|_{\boldsymbol{S}_+^m,\boldsymbol{V}}\right)\left(\max_{\|\boldsymbol{X}\|_{\boldsymbol{S}_+^m,\boldsymbol{V}}=1} \sqrt{\langle \boldsymbol{X},\boldsymbol{X}\rangle}\right).$$

Proposition 13 *We have*

(a) *For the second order cone, $\alpha_{\boldsymbol{L}^{n+1},\boldsymbol{v}} \geq \sqrt{2}$, for all $\boldsymbol{v} \succ_{\boldsymbol{L}^{n+1}} 0$ ($\|\boldsymbol{v}_n\|_2 < v_{n+1}$) with equality holding for $\boldsymbol{v} = (\boldsymbol{0}, 1)'$.*

(b) *For the symmetric positive semidefinite cone, $\alpha_{\boldsymbol{S}_+^m,\boldsymbol{V}} \geq \sqrt{m}$, for all $\boldsymbol{V} \succ 0$ with equality hoding for $\boldsymbol{V} = \boldsymbol{I}$.*

Proof :

For any $\boldsymbol{V} \succ_{\boldsymbol{K}} \boldsymbol{0}$, we observe that

$$\|\boldsymbol{V}\|_{\boldsymbol{K},\boldsymbol{V}} = \min\{y : y\boldsymbol{V} \succeq_{\boldsymbol{K}} \boldsymbol{V} \succeq_{\boldsymbol{K}} -y\boldsymbol{V}\} = 1.$$

Otherwise, if $\|\boldsymbol{V}\|_{\boldsymbol{K},\boldsymbol{V}} < 1$, there exist $y < 1$ such that $y\boldsymbol{V} \succeq_{\boldsymbol{K}} \boldsymbol{V}$, which implies that $-\boldsymbol{V} \succeq_{\boldsymbol{K}} \boldsymbol{0}$, contradicting $\boldsymbol{V} \succ_{\boldsymbol{K}} \boldsymbol{0}$. Hence, $\|\boldsymbol{v}\|_{\boldsymbol{L}^{n+1},\boldsymbol{v}} = 1$ and we obtain

$$\left(\max_{\|\boldsymbol{x}\|_{\boldsymbol{L}^{n+1},\boldsymbol{v}}=1} \|\boldsymbol{x}\|_2\right) \geq \|\boldsymbol{v}\|_2.$$

Likewise, when $\boldsymbol{x}_n = (\boldsymbol{v}_n)/(\sqrt{2}\|\boldsymbol{v}_n\|_2)$ and $x_{n+1} = -1/(\sqrt{2})$, so that $\|\boldsymbol{x}\|_2 = 1$, we can also verify that the inequalities

$$\begin{aligned}\|\frac{\boldsymbol{v}_n}{\sqrt{2}\|\boldsymbol{v}_n\|_2}+\boldsymbol{v}_n y\|_2 &\le v_{n+1}y-\frac{1}{\sqrt{2}}\\ \|\frac{\boldsymbol{v}_n}{\sqrt{2}\|\boldsymbol{v}_n\|_2}-\boldsymbol{v}_n y\|_2 &\le v_{n+1}y+\frac{1}{\sqrt{2}}\end{aligned}$$

hold if and only if $y \ge \sqrt{2}/(v_{n+1}-\|\boldsymbol{v}_n\|_2)$. Hence, $\|\boldsymbol{x}\|_{\boldsymbol{L}^{n+1},\boldsymbol{v}} = \sqrt{2}/(v_{n+1}-\|\boldsymbol{v}_n\|_2)$ and we obtain

$$\max_{\|\boldsymbol{x}\|_2=1}\|\boldsymbol{x}\|_{\boldsymbol{L}^{n+1},\boldsymbol{v}} \ge \frac{\sqrt{2}}{v_{n+1}-\|\boldsymbol{v}_n\|_2}.$$

Therefore, since $0 < v_{n+1}-\|\boldsymbol{v}_n\|_2 \le v_{n+1} \le \|\boldsymbol{v}\|$, we have

$$\alpha_{\boldsymbol{L}^{n+1},\boldsymbol{v}} = \left(\max_{\|\boldsymbol{x}\|_2=1}\|\boldsymbol{x}\|_{\boldsymbol{L}^{n+1},\boldsymbol{v}}\right)\left(\max_{\|\boldsymbol{x}\|_{\boldsymbol{L}^{n+1},\boldsymbol{v}}=1}\|\boldsymbol{x}\|_2\right) \ge \frac{\sqrt{2}\|\boldsymbol{v}\|_2}{v_{n+1}-\|\boldsymbol{v}_n\|_2} \ge \sqrt{2}.$$

When $\boldsymbol{v} = (\boldsymbol{0}, 1)'$, we have

$$\|\boldsymbol{x}\|_{\boldsymbol{L}^{n+1},\boldsymbol{v}} = \|\boldsymbol{x}_n\|_2 + |x_{n+1}|,$$

and from Proposition 10(b), the bound is achieved. Hence, $\alpha_{\boldsymbol{L}^{n+1}} = \sqrt{2}$.

(b) Since $\boldsymbol{V}$ is an invertible matrix, we observe that

$$\begin{aligned}\|\boldsymbol{X}\|_{S^m_+,\boldsymbol{V}} &= \min\{y : y\boldsymbol{V} \succeq \boldsymbol{X} \succeq -y\boldsymbol{V}\}\\ &= \min\left\{y : y\boldsymbol{I} \succeq \boldsymbol{V}^{-\frac{1}{2}}\boldsymbol{X}\boldsymbol{V}^{-\frac{1}{2}} \succeq -y\boldsymbol{I}\right\}\\ &= \|\boldsymbol{V}^{-\frac{1}{2}}\boldsymbol{X}\boldsymbol{V}^{-\frac{1}{2}}\|_2.\end{aligned}$$

For any $\boldsymbol{V} \succ \boldsymbol{0}$, let $\boldsymbol{X} = \boldsymbol{V}$, we have $\|\boldsymbol{X}\|_{S^m_+,\boldsymbol{V}} = 1$ and

$$\langle \boldsymbol{X}, \boldsymbol{X}\rangle = \text{trace}(\boldsymbol{V}\boldsymbol{V}) = \|\boldsymbol{\lambda}\|_2^2,$$

where $\boldsymbol{\lambda} \in \Re^m$ is a vector corresponding to all the eigenvalues of the matrix $\boldsymbol{V}$. Hence, we obtain

$$\left(\max_{\|\boldsymbol{X}\|_{S^m_+,\boldsymbol{V}}=1}\sqrt{\langle \boldsymbol{X}, \boldsymbol{X}\rangle}\right) \ge \|\boldsymbol{\lambda}\|_2.$$

Without loss of generality, let λ_1 be the smallest eigenvalue of $\boldsymbol{V}$ with corresponding normalized eigenvector, $\boldsymbol{q}_1$. Now, let $\boldsymbol{X} = \boldsymbol{q}_1\boldsymbol{q}_1'$. Observe that

$$\begin{aligned}
\langle \boldsymbol{X}, \boldsymbol{X} \rangle &= \text{trace}(\boldsymbol{X}\boldsymbol{X}) \\
&= \text{trace}(\boldsymbol{q}_1\boldsymbol{q}_1'\boldsymbol{q}_1\boldsymbol{q}_1') \\
&= \text{trace}(\boldsymbol{q}_1'\boldsymbol{q}_1\boldsymbol{q}_1'\boldsymbol{q}_1) \\
&= 1.
\end{aligned}$$

We can express the matrix, $\boldsymbol{V}$ in its spectral decomposition, so that $\boldsymbol{V} = \sum_j \boldsymbol{q}_j\boldsymbol{q}_j'\lambda_j$. Hence,

$$\begin{aligned}
\|\boldsymbol{X}\|_{S_+^m,\boldsymbol{V}} &= \|\boldsymbol{V}^{-\frac{1}{2}}\boldsymbol{X}\boldsymbol{V}^{-\frac{1}{2}}\|_2 \\
&= \|\textstyle\sum_j \boldsymbol{q}_j\boldsymbol{q}_j'\lambda_j^{-\frac{1}{2}}\boldsymbol{q}_1\boldsymbol{q}_1'\sum_j \boldsymbol{q}_j\boldsymbol{q}_j'\lambda_j^{-\frac{1}{2}}\|_2 \\
&= \|\lambda_1^{-1}\boldsymbol{q}_1\boldsymbol{q}_1'\|_2 \\
&= \lambda_1^{-1}.
\end{aligned}$$

Therefore, we establish that

$$\left(\max_{\sqrt{\langle \boldsymbol{X},\boldsymbol{X}\rangle}=1} \|\boldsymbol{X}\|_{S_+^m,\boldsymbol{V}}\right) \geq \lambda_1^{-1}.$$

Combining the results, we have

$$\alpha_{S^m,\boldsymbol{V}} = \left(\max_{\|\boldsymbol{X}\|_{S_+^m,\boldsymbol{V}}=1} \sqrt{\langle \boldsymbol{X}, \boldsymbol{X}\rangle}\right)\left(\max_{\sqrt{\langle \boldsymbol{X},\boldsymbol{X}\rangle}=1} \|\boldsymbol{X}\|_{S_+^m,\boldsymbol{V}}\right) \geq \frac{\|\boldsymbol{\lambda}\|_2}{\lambda_1} \geq \sqrt{m}.$$

When $\boldsymbol{V} = \boldsymbol{I}$, we have

$$\|\boldsymbol{X}\|_{S^m,\boldsymbol{V}} = \|\boldsymbol{X}\|_2,$$

and from Proposition 10(d), the bound is achieved. Hence, $\alpha_{S^m} = \sqrt{m}$. ■

4.5 Conclusions

We proposed a relaxed robust counterpart for general conic optimization problems that we believe achieves the objectives outlined in the introduction, namely:

(a) It preserves the computational tractability of the nominal problem. Specifically the robust conic optimization problem retains its original structure, i.e., robust linear optimization problems (LPs) remain LPs, robust second order cone optimization problems (SOCPs) remain SOCPs and robust semidefinite optimization problems (SDPs) remain SDPs. Moreover, the size of the proposed robust problem especially under the l_2 norm is practically the same as the nominal problem.

(b) It allows us to provide a guarantee on the probability that the robust solution is feasible, when the uncertain coefficients obey independent and identically distributed normal distributions.

Chapter 5

Robust Discrete Optimization and Network Flows

Our goal in this chapter is to propose an approach to address data uncertainty for discrete optimization and network flow problems that has the following features:

(a) It allows to control the degree of conservatism of the solution;

(b) It is computationally tractable both practically and theoretically.

Specifically, our contributions include:

(a) When both the cost coefficients and the data in the constraints of an integer optimization problem are subject to uncertainty, we propose, following the approach in Chapter 2, a robust integer optimization problem of moderately larger size that allows to control the degree of conservatism of the solution in terms of probabilistic bounds on constraint violation.

(b) When only the cost coefficients are subject to uncertainty and the problem is a $0-1$ discrete optimization problem on n variables, then we solve the robust counterpart by solving $n+1$ nominal problems. Thus, we show that the robust counterpart of a polynomially solvable $0-1$ discrete optimization problem remains polynomially solvable. In particular, robust matching, spanning tree, shortest path, matroid intersection, etc. are polynomially solvable. Moreover, we show that the robust counterpart of an NP-hard α-approximable $0-1$ discrete optimization problem, remains α-approximable.

(c) When only the cost coefficients are subject to uncertainty and the problem is a minimum cost flow problem, then we propose a polynomial time algorithm for the robust counterpart by solving a collection of minimum cost flow problems in a modified network.

Structure of the chapter. In Section 5.1, we present the general framework and formulation of robust discrete optimization problems. In Section 5.2, we propose an efficient algorithm for solving robust combinatorial optimization problems. In Section 5.3, we show that the robust counterpart of an NP-hard $0-1$ α-approximable discrete optimization problem remains α-approximable. In Section 5.4, we propose an efficient algorithm for robust network flows. In Section 5.5, we present some experimental findings relating to the computation speed and the quality of robust solutions. Finally, Section 5.6 contains some remarks with respect to the practical applicability of the proposed methods.

5.1 Robust Formulation of Discrete Optimization Problems

Let $\boldsymbol{c}$, $\boldsymbol{l}$, $\boldsymbol{u}$ be n-vectors, let $\boldsymbol{A}$ be an $m \times n$ matrix, and $\boldsymbol{b}$ be an m-vector. We consider the following nominal mixed integer optimization problem (MIP) on a set of n variables, the first k of which are integral:

$$
\begin{aligned}
\text{minimize} \quad & \boldsymbol{c}'\boldsymbol{x} \\
\text{subject to} \quad & \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b} \\
& \boldsymbol{l} \leq \boldsymbol{x} \leq \boldsymbol{u} \\
& x_i \in \mathcal{Z}, \qquad i = 1, \dots, k,
\end{aligned}
\tag{5.1}
$$

We assume without loss of generality that data uncertainty affects only the elements of the matrix $\boldsymbol{A}$ and $\boldsymbol{c}$, but not the vector $\boldsymbol{b}$, since in this case we can introduce a new variable x_{n+1}, and write $\boldsymbol{Ax} - \boldsymbol{b}x_{n+1} \leq \boldsymbol{0}$, $\boldsymbol{l} \leq \boldsymbol{x} \leq \boldsymbol{u}, 1 \leq x_{n+1} \leq 1$, thus augmenting $\boldsymbol{A}$ to include $\boldsymbol{b}$.

In typical applications, we have reasonable estimates for the mean value of the coefficients a_{ij} and its range $\hat{a}_{ij}$. We feel that it is unlikely that we know the exact distribution of these coefficients. Similarly, we have estimates for the cost coefficients c_j and an estimate of its range d_j. Specifically, the model of data uncertainty we consider is as follows:

Model of Data Uncertainty U:

(a) **(Uncertainty for matrix $\boldsymbol{A}$)**: Let $N = \{1, 2, \ldots, n\}$. Each entry $a_{ij},\ j \in N$ is modeled as independent, symmetric and bounded random variable (but with unknown distribution) $\tilde{a}_{ij}, j \in N$ that takes values in $[a_{ij} - \hat{a}_{ij}, a_{ij} + \hat{a}_{ij}]$.

(b) **(Uncertainty for cost vector $\boldsymbol{c}$)**: Each entry $c_j,\ j \in N$ takes values in $[c_j, c_j + d_j]$, where d_j represents the deviation from the nominal cost coefficient, c_j.

Note that we allow the possibility that $\hat{a}_{ij} = 0$ or $d_j = 0$. Note also that the only assumption that we place on the distribution of the coefficients a_{ij} is that it is symmetric.

5.1.1 Robust MIP Formulation

For robustness purposes, for every i, we introduce a number Γ_i, $i = 0, 1, \ldots, m$ that takes values in the interval $[0, |J_i|]$, where $J_i = \{j |\ \hat{a}_{ij} > 0\}$. Γ_0 is assumed to be integer, while Γ_i, $i = 1, \ldots, m$ are not necessarily integers.

The role of the parameter Γ_i in the constraints is to adjust the robustness of the proposed method against the level of conservatism of the solution. Consider the ith constraint of the nominal problem $\boldsymbol{a}_i'\boldsymbol{x} \leq b_i$. Let J_i be the set of coefficients a_{ij}, $j \in J_i$ that are subject to parameter uncertainty, i.e., $\tilde{a}_{ij},\ j \in J_i$ independently takes values according to a symmetric distribution with mean equal to the nominal value a_{ij} in the interval $[a_{ij} - \hat{a}_{ij}, a_{ij} + \hat{a}_{ij}]$. Speaking intuitively, it is unlikely that all of the a_{ij}, $j \in J_i$ will change. Our goal is to be protected against all cases in which up to $\lfloor \Gamma_i \rfloor$ of these coefficients are allowed to change, and one coefficient a_{it} changes by at most $(\Gamma_i - \lfloor \Gamma_i \rfloor)\hat{a}_{it}$. In other words, we stipulate that nature will be restricted in its behavior, in that only a subset of the coefficients will change in order to adversely affect the solution. We will then guarantee that if nature behaves like this then the robust solution will be feasible deterministically. We will also show that, essentially because the distributions we allow are symmetric, even if more than $\lfloor \Gamma_i \rfloor$ change, then the robust solution will be feasible with very high probability. Hence, we call Γ_i the protection level for the ith constraint.

The parameter Γ_0 controls the level of robustness in the objective. We are interested in finding an optimal solution that optimizes against all scenarios under which a number Γ_0 of the cost coefficients can vary in such a way as to maximally influence the objective. Let $J_0 = \{j \mid d_j > 0\}$. If $\Gamma_0 = 0$, we completely ignore the influence of the cost deviations, while if $\Gamma_0 = |J_0|$, we are considering all possible cost deviations, which is indeed most conservative. In general a higher value of Γ_0 increases the level of robustness at the expense of higher nominal cost.

Specifically, the proposed robust counterpart of Problem (5.1) is as follows:

$$\begin{aligned}
\text{minimize} \quad & \boldsymbol{c}'\boldsymbol{x} + \max_{\{S_0 \mid S_0 \subseteq J_0, |S_0| \leq \Gamma_0\}} \left\{ \sum_{j \in S_0} d_j |x_j| \right\} \\
\text{subject to} \quad & \sum_j a_{ij} x_j + \max_{\{S_i \cup \{t_i\} \mid S_i \subseteq J_i, |S_i| \leq \lfloor \Gamma_i \rfloor, t_i \in J_i \setminus S_i\}} \\
& \quad \left\{ \sum_{j \in S_i} \hat{a}_{ij} |x_j| + (\Gamma_i - \lfloor \Gamma_i \rfloor) \hat{a}_{it_i} |x_{t_i}| \right\} \leq b_i, \quad \forall i \\
& \boldsymbol{l} \leq \boldsymbol{x} \leq \boldsymbol{u} \\
& x_i \in \mathcal{Z}, \qquad \forall i = 1, \ldots, k.
\end{aligned} \tag{5.2}$$

Following Theorem 1 in Chapter 2, we have the following equivalent robust counterpart:

$$
\begin{aligned}
\text{minimize} \quad & \boldsymbol{c}'\boldsymbol{x} + z_0\Gamma_0 + \textstyle\sum_{j \in J_0} p_{0j} & & \\
\text{subject to} \quad & \sum_j a_{ij}x_j + z_i\Gamma_i + \sum_{j \in J_i} p_{ij} \leq b_i & & \forall i \\
& z_0 + p_{0j} \geq d_j y_j & & \forall j \in J_0 \\
& z_i + p_{ij} \geq \hat{a}_{ij} y_j & & \forall i \neq 0, j \in J_i \\
& p_{ij} \geq 0 & & \forall i, j \in J_i \\
& y_j \geq 0 & & \forall j \\
& z_i \geq 0 & & \forall i \\
& -y_j \leq x_j \leq y_j & & \forall j \\
& l_j \leq x_j \leq u_j & & \forall j \\
& x_i \in \mathcal{Z} & & i = 1, \ldots, k.
\end{aligned}
\tag{5.3}
$$

While the original Problem (5.1) involves n variables and m constraints, its robust counterpart Problem (5.3) has $2n+m+l$ variables, where $l = \sum_{i=0}^{m} |J_i|$ is the number of uncertain coefficients, and $2n + m + l$ constraints.

As we discussed, if less than $\lfloor \Gamma_i \rfloor$ coefficients a_{ij}, $j \in J_i$ participating in the ith constraint vary, then the robust solution will be feasible deterministically. Theorem 3 shows that that even if more than $\lfloor \Gamma_i \rfloor$ change, then the robust solution will be feasible with very high probability. We make no theoretical claims regarding suboptimality given that we made no probabilistic assumptions on the cost coefficients. In Section 5.5.1, we apply these bounds of Theorem 3 in the context of the zero-one knapsack problem.

5.2 Robust Combinatorial Optimization

Combinatorial optimization is an important class of discrete optimization whose decision variables are binary, that is $\boldsymbol{x} \in X \subseteq \{0,1\}^n$. In this section, the nominal combinatorial optimization problem we consider is:

$$
\begin{aligned}
\text{minimize} \quad & \boldsymbol{c}'\boldsymbol{x} \\
\text{subject to} \quad & \boldsymbol{x} \in X.
\end{aligned}
\tag{5.4}
$$

We are interested in the class of problems where each entry $\tilde{c}_j,\ j \in N = \{1, 2, \ldots, n\}$ takes values in $[c_j, c_j + d_j]$, $d_j \geq 0,\ j \in N$, but the set X is fixed. We would like to find a solution $\boldsymbol{x} \in X$ that minimizes the maximum cost $\boldsymbol{c}'\boldsymbol{x}$ such that at most Γ of the coefficients $\tilde{c}_j$ are allowed to change:

$$\begin{aligned} Z^* = \text{minimize} \quad & \boldsymbol{c}'\boldsymbol{x} + \max_{\{S \mid S \subseteq N, |S| \leq \Gamma\}} \sum_{j \in S} d_j x_j \\ \text{subject to} \quad & \boldsymbol{x} \in X. \end{aligned} \tag{5.5}$$

Without loss of generality, we assume that the indices are ordered in such that $d_1 \geq d_2 \geq \ldots \geq d_n$. We also define $d_{n+1} = 0$ for notational convenience. Examples of such problems include the shortest path, the minimum spanning tree, the minimum assignment, the traveling salesman, the vehicle routing and matroid intersection problems. Data uncertainty in the context of the vehicle routing problem for example, captures the variability of travel times in some of the links of the network.

In the context of scenario based uncertainty, finding an optimally robust solution involves solving the problem (for the case that only two scenarios for the cost vectors $\boldsymbol{c_1}$, $\boldsymbol{c_2}$ are known):

$$\begin{aligned} \text{minimize} \quad & \max(\boldsymbol{c}_1'\boldsymbol{x}, \boldsymbol{c}_2'\boldsymbol{x}) \\ \text{subject to} \quad & \boldsymbol{x} \in X. \end{aligned}$$

For many classical combinatorial problems (for example the shortest path problem), finding such a robust solution is NP-hard, even if minimizing $\boldsymbol{c}_i'\boldsymbol{x}$ subject to $\boldsymbol{x} \in X$ is polynomially solvable (Kouvelis and Yu [20]).

Clearly the robust counterpart of an NP-hard combinatorial optimization problem is NP-hard. We next show that surprisingly, the robust counterpart of a polynomially solvable combinatorial optimization problem is also polynomially solvable.

5.2.1 Algorithm for Robust Combinatorial Optimization Problems

In this section, we show that we can solve Problem (5.5) by solving at most $n+1$ nominal problems min $\boldsymbol{f}_i'\boldsymbol{x}$, subject to $\boldsymbol{x} \in X$, for $i = 1, \ldots, n+1$.

Theorem 11 *Problem (5.5) can be solved by solving the* $n+1$ *nominal problems:*

$$Z^* = \min_{l=1,\ldots,n+1} G^l, \tag{5.6}$$

where for $l = 1, \ldots, n+1$*:*

$$\begin{aligned} G^l = \Gamma d_l + \min \quad & \left(\boldsymbol{c}'\boldsymbol{x} + \sum_{j=1}^{l} (d_j - d_l)\, x_j \right) \\ \text{subject to} \quad & \boldsymbol{x} \in X. \end{aligned} \tag{5.7}$$

Proof : Problem (5.5) can be rewritten as follows:

$$\begin{aligned} Z^* = \min_{\mathbf{x} \in X} \left(\boldsymbol{c}'\boldsymbol{x} + \max \quad \right. & \left. \sum_{j \in N} d_j x_j u_j \right) \\ \text{subject to} \quad & 0 \le u_j \le 1, \qquad j \in N \\ & \sum_{j \in N} u_j \le \Gamma. \end{aligned}$$

Given a fixed $\boldsymbol{x} \in X$, we consider the inner maximization problem and formulate its dual. Applying strong duality to this problem we obtain:

$$\begin{aligned} Z^* = \min_{\mathbf{x} \in X} \boldsymbol{c}'\boldsymbol{x} + \min \quad & \left(\Gamma\theta + \sum_{j \in N} y_j \right) \\ \text{subject to} \quad & y_j + \theta \ge d_j x_j, \qquad j \in N \\ & y_j, \theta \ge 0, \end{aligned}$$

which can be rewritten as:

$$
\begin{aligned}
Z^* = \min \quad & \boldsymbol{c}'\boldsymbol{x} + \Gamma\theta + \sum_{j\in N} y_j \\
\text{subject to} \quad & y_j + \theta \geq d_j x_j, \qquad j \in N \\
& y_j, \theta \geq 0, \\
& \boldsymbol{x} \in X.
\end{aligned}
\tag{5.8}
$$

Clearly an optimal solution $(\boldsymbol{x}^*, \boldsymbol{y}^*, \theta^*)$ of Problem (5.8) satisfies:

$$y_j^* = \max(d_j x_j^* - \theta^*, 0),$$

and therefore,

$$Z^* = \min_{\mathbf{x}\in X, \theta\geq 0} \left(\Gamma\theta + \boldsymbol{c}'\boldsymbol{x} + \sum_{j\in N} \max(d_j x_j - \theta, 0)\right).$$

Since $X \subset \{0,1\}^n$,

$$\max(d_j x_j - \theta, 0) = \max(d_j - \theta, 0)\ x_j, \tag{5.9}$$

Hence, we obtain

$$Z^* = \min_{\mathbf{x}\in X, \theta\geq 0} \left(\Gamma\theta + \boldsymbol{c}'\boldsymbol{x} + \sum_{j\in N} \max(d_j - \theta, 0) x_j\right). \tag{5.10}$$

In order to find the optimal value for θ we decompose $\Re^+$ into the intervals $[0, d_n]$, $[d_n, d_{n-1}], \ldots, [d_2, d_1]$ and $[d_1, \infty)$. Then, recalling that $d_{n+1} = 0$, we obtain

$$\sum_{j\in N} \max(d_j - \theta, 0) x_j = \begin{cases} \sum_{j=1}^{l-1} (d_j - \theta) x_j, & \text{if } \theta \in [d_l, d_{l-1}], \quad l = n+1, \ldots, 2, \\ 0, & \text{if } \theta \in [d_1, \infty). \end{cases}$$

Therefore, $Z^* = \min_{l=1,\ldots,n+1} Z^l$, where for $l = 1, \ldots, n+1$:

$$Z^l = \min_{\mathbf{x}\in X, \theta\in[d_l, d_{l-1}]} \left(\Gamma\theta + \boldsymbol{c}'\boldsymbol{x} + \sum_{j=1}^{l-1} (d_j - \theta) x_j\right),$$

where the sum for $l = 1$ is equal to zero. Since we are optimizing a linear function of θ over the interval $[d_l, d_{l-1}]$, the optimal is obtained for $\theta = d_l$ or $\theta = d_{l-1}$, and thus for $l = 1, \ldots, n+1$:

$$\begin{aligned} Z^l &= \min\left(\Gamma d_l + \min_{\mathbf{x}\in X}\left(\boldsymbol{c}'\boldsymbol{x} + \sum_{j=1}^{l-1}(d_j - d_l)x_j\right),\right. \\ &\qquad \left.\Gamma d_{l-1} + \min_{\mathbf{x}\in X}\left(\boldsymbol{c}'\boldsymbol{x} + \sum_{j=1}^{l-1}(d_j - d_{l-1})x_j\right)\right) \\ &= \min\left(\Gamma d_l + \min_{\mathbf{x}\in X}\left(\boldsymbol{c}'\boldsymbol{x} + \sum_{j=1}^{l}(d_j - d_l)x_j\right),\right. \\ &\qquad \left.\Gamma d_{l-1} + \min_{\mathbf{x}\in X}\left(\boldsymbol{c}'\boldsymbol{x} + \sum_{j=1}^{l-1}(d_j - d_{l-1})x_j\right)\right). \end{aligned}$$

Thus,

$$\begin{aligned} Z^* = \min\Bigg(&\Gamma d_1 + \min_{\mathbf{x}\in X}\boldsymbol{c}'\boldsymbol{x}, \ldots, \Gamma d_l + \min_{\mathbf{x}\in X}\left(\boldsymbol{c}'\boldsymbol{x} + \sum_{j=1}^{l}(d_j - d_l)x_j\right), \\ &\ldots, \min_{\mathbf{x}\in X}\left(\boldsymbol{c}'\boldsymbol{x} + \sum_{j=1}^{n} d_j x_j\right)\Bigg). \end{aligned}$$

■

Remark: Note that we critically used the fact that the nominal problem is a 0-1 discrete optimization problem, i.e., $X \subseteq \{0,1\}^n$, in Eq. (5.9). For general integer optimization problems Eq. (5.9) does not apply.

Theorem 11 leads to the following algorithm.

Algorithm A

1. For $l = 1, \ldots, n+1$ solve the $n+1$ nominal problems Eqs. (5.7):

$$G^l = \Gamma d_l + \min_{\mathbf{x}\in X}\left(\boldsymbol{c}'\boldsymbol{x} + \sum_{j=1}^{l}(d_j - d_l)x_j\right),$$

   and let $\boldsymbol{x}^l$ be an optimal solution of the corresponding problem.

2. Let $l^* = \arg\min_{l=1,\ldots,n+1} G^l$.

3. $Z^* = G^{l^*}$; $\boldsymbol{x}^* = \boldsymbol{x}^{l^*}$.

Note that Z^l is not in general equal to G^l. If f is the number of distinct values among $d_1, \ldots, d_n$, then it is clear that Algorithm A solves $f+1$ nominal problems, since if $d_l = d_{l+1}$, then $G^l = G^{l+1}$. In particular, if all $d_j = d$ for all $j = 1, \ldots, n$, then Algorithm A solves only two nominal problems. Thus, if τ is the time to solve one nominal problem, Algorithm A solves the robust counterpart in $(f+1)\tau$ time, thus preserving the polynomial solvability of the nominal problem. In particular, Theorem 11 implies that the robust counterpart of many classical 0-1 combinatorial optimization problems like the minimum spanning tree, the minimum assignment, minimum matching, shortest path and matroid intersection, are polynomially solvable.

5.3 Robust Approximation Algorithms

In this section, we show that if the nominal combinatorial optimization problem (5.4) has an α-approximation polynomial time algorithm, then the robust counterpart Problem (5.5) with optimal solution value Z^* is also α-approximable. Specifically, we assume that there exists a polynomial time Algorithm H for the nominal problem (5.4), that returns a solution with an objective Z_H: $Z \leq Z_H \leq \alpha Z$, $\alpha \geq 1$.

The proposed algorithm for the robust Problem (5.5) is to utilize Algorithm H in Algorithm A, instead of solving the nominal instances exactly. The proposed algorithm is as follows:

Theorem 12 *Algorithm B yields a solution $\boldsymbol{x}^B$ with an objective value Z_B that satisfies:*

$$Z^* \leq \; Z_B \leq \; \alpha Z^*.$$

Proof : Since Z^* is the optimal objective function value of the robust problem, clearly $Z^* \leq \; Z_B$. Let l the index such that $Z^* = G^l$ in Theorem 11. Let $\boldsymbol{x}_H^l$ be an

Algorithm B

1. For $l = 1, \ldots, n+1$ find an α-approximate solution $\boldsymbol{x}_H^l$ using Algorithm H for the nominal problem:

$$G^l - \Gamma d_l = \min_{\mathbf{x} \in X} \left(\boldsymbol{c}'\boldsymbol{x} + \sum_{j=1}^{l} (d_j - d_l) x_j \right). \tag{5.11}$$

2. For $l = 1, \ldots, n+1$, let

$$Z_H^l = \boldsymbol{c}'\boldsymbol{x}_H^l + \max_{\{S \mid\, S \subseteq N, |S| \leq \Gamma\}} \sum_{j \in S} d_j (\boldsymbol{x}_H^l)_j.$$

3. Let $l^* = \arg \min_{l=1,\ldots,n+1} Z_H^l$.

4. $Z_B = Z_H^{l^*}$; $\boldsymbol{x}^B = \boldsymbol{x}_H^{l^*}$.

α-optimal solution for Problem (5.11). Then, we have

$$\begin{aligned}
Z_B &\leq Z_H^l \\
&= \boldsymbol{c}'\boldsymbol{x}_H^l + \max_{\{S \mid\, S \subseteq N, |S| \leq \Gamma\}} \sum_{j \in S} d_j (\boldsymbol{x}_H^l)_j \\
&= \min_{\theta \geq 0} \left\{ \boldsymbol{c}'\boldsymbol{x}_H^l + \sum_{j \in N} \max(d_j - \theta, 0)(\boldsymbol{x}_H^l)_j + \Gamma\theta \right\} && \text{(from Eq. (5.10))} \\
&\leq \boldsymbol{c}'\boldsymbol{x}_H^l + \sum_{j=1}^{l} (d_j - d_l)(\boldsymbol{x}_H^l)_j + \Gamma d_l \\
&\leq \alpha(G^l - \Gamma d_l) + \Gamma d_l && \text{(from Eq. (5.11))} \\
&\leq \alpha G^l && \text{(since } \alpha \geq 1\text{)} \\
&= \alpha Z^*.
\end{aligned}$$

■

Remark : Note that Algorithm A is a special case of Algorithm B for $\alpha = 1$. Note that it is critical to have an α-approximation algorithm for all nominal instances (5.11). In particular, if the nominal problem is the traveling salesman problem under triangle inequality, which can be approximated within $\alpha = 3/2$, Algorithm B is not an α-approximation algorithm for the robust counterpart, as the instances (5.11) may not satisfy the triangle inequality.

5.4 Robust Network Flows

In this section, we apply the methods of Section 5.2 to show that robust minimum cost flows can also be solved by solving a collection of modified nominal minimum cost flows. Given a directed graph $G = (\mathcal{N}, \mathcal{A})$, the minimum cost flow is defined as follows:

$$\begin{aligned}
\text{minimize} \quad & \sum_{(i,j)\in\mathcal{A}} c_{ij}x_{ij} \\
\text{subject to} \quad & \sum_{\{j:(i,j)\in\mathcal{A}\}} x_{ij} - \sum_{\{j:(j,i)\in\mathcal{A}\}} x_{ji} = b_i \quad \forall i \in \mathcal{N} \\
& 0 \le x_{ij} \le u_{ij} \qquad\qquad\qquad\qquad \forall (i,j) \in \mathcal{A}.
\end{aligned} \tag{5.12}$$

Let X be the set of feasible solutions of Problem (5.12).

We are interested in the class of problems in which each entry $\tilde{c}_{ij},\ (i,j) \in \mathcal{A}$ takes values in $[c_{ij}, c_{ij} + d_{ij}]$, $d_{ij}, c_{ij} \ge 0, (i,j) \in \mathcal{A}$. From Eq. (5.5) the robust minimum cost flow problem is:

$$\begin{aligned}
Z^* = \min \quad & \boldsymbol{c}'\boldsymbol{x} + \max_{\{S|\ S\subseteq\mathcal{A}, |S|\le\Gamma\}} \sum_{(i,j)\in S} d_{ij}x_{ij} \\
\text{subject to} \quad & \boldsymbol{x} \in X.
\end{aligned} \tag{5.13}$$

From Eq. (5.8) we obtain that Problem (5.13) is equivalent to solving the following problem:

$$Z^* = \min_{\theta\ge 0} Z(\theta), \tag{5.14}$$

where

$$\begin{aligned}
Z(\theta) = \Gamma\theta + \min \quad & \boldsymbol{c}'\boldsymbol{x} + \sum_{(i,j)\in\mathcal{A}} p_{ij} \\
\text{subject to} \quad & p_{ij} \ge d_{ij}x_{ij} - \theta \quad \forall (i,j) \in \mathcal{A} \\
& p_{ij} \ge 0 \qquad\qquad\quad \forall (i,j) \in \mathcal{A} \\
& \boldsymbol{x} \in X.
\end{aligned} \tag{5.15}$$

We next show that for a fixed $\theta \ge 0$, we can solve Problem (5.15) as a network flow problem.

Theorem 13 *For a fixed $\theta \geq 0$, Problem (5.15) can be solved as a network flow problem.*

Proof :

We eliminate the variables p_{ij} from Formulation (5.15) and obtain:

$$\begin{aligned} Z(\theta) = \Gamma\theta + \min \quad & \boldsymbol{c}'\boldsymbol{x} + \sum_{(i,j)\in\mathcal{A}} d_{ij} \max\left(x_{ij} - \frac{\theta}{d_{ij}}, 0\right) \\ \text{subject to} \quad & \mathbf{x} \in X. \end{aligned} \tag{5.16}$$

For every arc $(i,j) \in \mathcal{A}$, we introduce nodes i' and j' and replace the arc (i,j) with arcs (i,i'), (i',j'), (j',j) and (i',j) with the following costs and capacities (see also Figure 5-1):

$$\begin{aligned} c_{ii'} &= c_{ij} & u_{ii'} &= u_{ij} \\ c_{j'j} &= 0 & u_{j'j} &= \infty \\ c_{i'j} &= 0 & u_{i'j} &= \frac{\theta}{d_{ij}} \\ c_{i'j'} &= d_{ij} & u_{i'j'} &= \infty. \end{aligned}$$

Let $G' = (\mathcal{N}', \mathcal{A}')$ be the new direceted graph.

We show that solving a linear minimum cost flow problem with data as above, leads to the solution of Problem (5.16). Consider an optimal solution of Problem (5.16). If $x_{ij} \leq \theta/d_{ij}$ for a given arc $(i,j) \in \mathcal{A}$, then the flow x_{ij} will be routed along the arcs (i,i') and (i',j) an the total contribution to cost is

$$c_{ii'}x_{ij} + c_{i'j}x_{ij} = c_{ij}x_{ij}.$$

If, however, $x_{ij} \geq \theta/d_{ij}$, then the flow x_{ij} will be routed along the arcs (i,i'), then θ/d_{ij} will be routed along arc (i',j), and the excess $x_{ij} - (\theta/d_{ij})$ is routed through the arcs (i',j') and (j',j). The total contribution to cost is

$$c_{ii'}x_{ij} + c_{i'j}\frac{\theta}{d_{ij}} + c_{i'j'}\left(x_{ij} - \frac{\theta}{d_{ij}}\right) + c_{j'j}\left(x_{ij} - \frac{\theta}{d_{ij}}\right) =$$

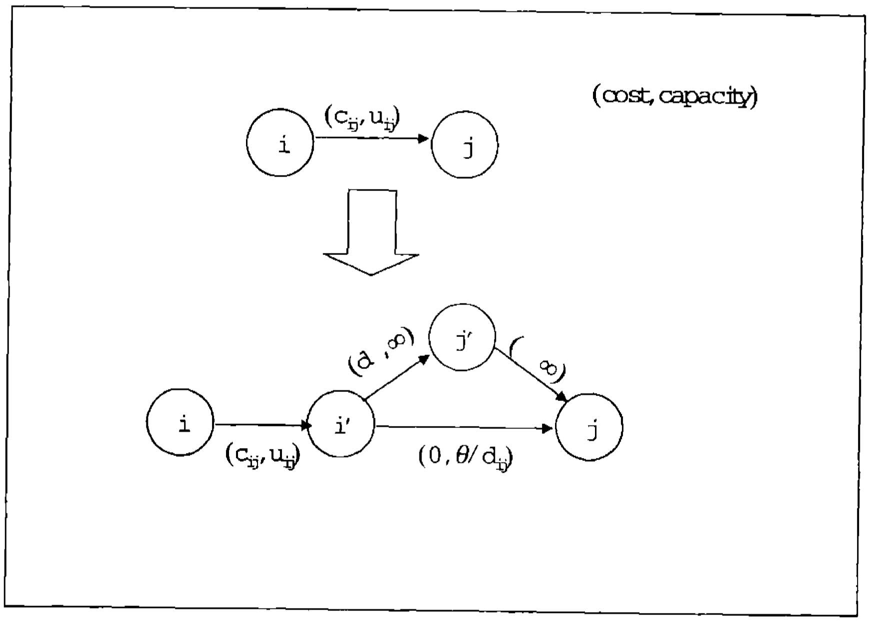

Figure 5-1: Conversion of arcs with cost uncertainties.

$$c_{ij}x_{ij} + d_{ij}\left(x_{ij} - \frac{\theta}{d_{ij}}\right).$$

In both cases the contribution to cost matches the objective function value in Eq. (5.16). ■

Without loss of generality, we can assume that all the capacities u_{ij}, $(i, j) \in \mathcal{A}$ are finitely bounded. Then, clearly $\theta \leq \overline{\theta} = \max\{u_{ij}d_{ij} \ : (i, j) \in \mathcal{A}\}$. Theorem 13 shows that the robust counterpart of the minimum cost flow problem can be converted to a minimum cost flow problem in which capacities on the arcs are linear functions of θ. Srinivasan and Thompsom [27] proposed a simplex based method for solving such parametric network flow problems for all values of the parameter $\theta \in [0, \overline{\theta}]$. Using this method, we can obtain the complete set of robust solutions for $\Gamma \in [0, |\mathcal{A}|]$. However, while the algorithm may be practical, it is not polynomial. We next provide a polynomial time algorithm. We first establish some properties of the function $Z(\theta)$.

Theorem 14 **(a)** *$Z(\theta)$ is a convex function.*

(b) *For all* $\theta_1, \theta_2 \geq 0$, *we have*

$$|Z(\theta_1) - Z(\theta_2)| \leq |\mathcal{A}||\theta_1 - \theta_2|. \tag{5.17}$$

Proof :

(a) Let $(\boldsymbol{x_1}, \boldsymbol{p_1})$ and $(\boldsymbol{x_2}, \boldsymbol{p_2})$ be optimal solutions to Problem (5.15) with $\theta = \theta_1$ and $\theta = \theta_2$ respectively. Clearly, since the feasible region is convex, for all $\lambda \in [0, 1]$, $(\lambda \boldsymbol{x_1} + (1-\lambda)\boldsymbol{x_2}, \lambda \boldsymbol{p_1} + (1-\lambda)\boldsymbol{p_2})$ is feasible to the problem with $\theta = \lambda\theta_1 + (1-\lambda)\theta_2$. Therefore,

$$\begin{aligned} & \lambda Z(\theta_1) + (1-\lambda) Z(\theta_2) \\ = \; & \boldsymbol{c}'(\lambda \boldsymbol{x_1} + (1-\lambda)\boldsymbol{x_2}) + \boldsymbol{e}'(\lambda \boldsymbol{p_1} + (1-\lambda)\boldsymbol{p_2}) \\ & \quad + \Gamma(\lambda\theta_1 + (1-\lambda)\theta_2) \geq Z(\lambda\theta_1 + (1-\lambda)\theta_2), \end{aligned}$$

where $\boldsymbol{e}$ is a vector of ones.

(b) By introducing Lagrange multiplies $\boldsymbol{r}$ to the first set of constraints of Problem (5.15), we obtain:

$$\begin{aligned} Z(\theta) &= \max_{\mathbf{r} \geq \mathbf{0}} \min_{\mathbf{x} \in X, \mathbf{p} \geq \mathbf{0}} \left\{ \Gamma\theta + \boldsymbol{c}'\boldsymbol{x} + \sum_{(i,j)\in\mathcal{A}} p_{ij} + \sum_{(i,j)\in\mathcal{A}} r_{ij}(d_{ij}x_{ij} - p_{ij} - \theta) \right\} \\ &= \max_{\mathbf{r} \geq \mathbf{0}} \min_{\mathbf{x} \in X, \mathbf{p} \geq \mathbf{0}} \left\{ (\Gamma - \sum_{(i,j)\in\mathcal{A}} r_{ij})\theta + \boldsymbol{c}'\boldsymbol{x} + \sum_{(i,j)\in\mathcal{A}} p_{ij}(1 - r_{ij}) + \sum_{(i,j)\in\mathcal{A}} r_{ij}d_{ij}x_{ij} \right\} \\ &= \max_{\mathbf{0} \leq \mathbf{r} \leq \mathbf{e}} \min_{\mathbf{x} \in X} \left\{ (\Gamma - \sum_{(i,j)\in\mathcal{A}} r_{ij})\theta + \boldsymbol{c}'\boldsymbol{x} + \sum_{(i,j)\in\mathcal{A}} r_{ij}d_{ij}x_{ij} \right\}, \end{aligned} \tag{5.18}$$

where Eq. (5.18) follows from the fact that $\min_{\mathbf{p} \geq \mathbf{0}} \left\{ \sum_{(i,j)\in\mathcal{A}} p_{ij}(1 - r_{ij}) \right\}$ is unbounded if any $r_{ij} > 1$ and equals to zero for $\mathbf{0} \leq \boldsymbol{r} \leq \mathbf{e}$. Without loss of generality, let $\theta_1 > \theta_2 \geq 0$. For $\mathbf{0} \leq \boldsymbol{r} \leq \mathbf{e}$, we have

$$-|\mathcal{A}| \leq \Gamma - \sum_{(i,j)\in\mathcal{A}} r_{ij} \leq |\mathcal{A}|.$$

Thus,

$$\begin{aligned}
Z(\theta_1) &= \max_{\mathbf{0}\le\mathbf{r}\le\mathbf{e}} \min_{\mathbf{x}\in X} \left\{ (\Gamma - \sum_{(i,j)\in\mathcal{A}} r_{ij})\theta_1 + \boldsymbol{c}'\boldsymbol{x} + \sum_{(i,j)\in\mathcal{A}} r_{ij} d_{ij} x_{ij} \right\} \\
&= \max_{\mathbf{0}\le\mathbf{r}\le\mathbf{e}} \min_{\mathbf{x}\in X} \left\{ (\Gamma - \sum_{(i,j)\in\mathcal{A}} r_{ij})(\theta_2 + (\theta_1 - \theta_2)) + \boldsymbol{c}'\boldsymbol{x} + \sum_{(i,j)\in\mathcal{A}} r_{ij} d_{ij} x_{ij} \right\} \\
&\le \max_{\mathbf{0}\le\mathbf{r}\le\mathbf{e}} \min_{\mathbf{x}\in X} \left\{ (\Gamma - \sum_{(i,j)\in\mathcal{A}} r_{ij})\theta_2 + |\mathcal{A}|(\theta_1 - \theta_2) + \boldsymbol{c}'\boldsymbol{x} + \sum_{(i,j)\in\mathcal{A}} r_{ij} d_{ij} x_{ij} \right\} \\
&= Z(\theta_2) + |\mathcal{A}|(\theta_1 - \theta_2).
\end{aligned}$$

Similarly,

$$\begin{aligned}
Z(\theta_1) &= \max_{\mathbf{0}\le\mathbf{r}\le\mathbf{e}} \min_{\mathbf{x}\in X} \left\{ (\Gamma - \sum_{(i,j)\in\mathcal{A}} r_{ij})(\theta_2 + (\theta_1 - \theta_2)) + \boldsymbol{c}'\boldsymbol{x} + \sum_{(i,j)\in\mathcal{A}} r_{ij} d_{ij} x_{ij} \right\} \\
&\ge \max_{\mathbf{0}\le\mathbf{r}\le\mathbf{e}} \min_{\mathbf{x}\in X} \left\{ (\Gamma - \sum_{(i,j)\in\mathcal{A}} r_{ij})\theta_2 - |\mathcal{A}|(\theta_1 - \theta_2) + \boldsymbol{c}'\boldsymbol{x} + \sum_{(i,j)\in\mathcal{A}} r_{ij} d_{ij} x_{ij} \right\} \\
&= Z(\theta_2) - |\mathcal{A}|(\theta_1 - \theta_2).
\end{aligned}$$

■

We next show that the robust minimum cost flow problem (5.13) can be solved by solving a polynomial number of network flow problems.

Theorem 15 *For any fixed* $\Gamma \le |\mathcal{A}|$ *and every* $\epsilon > 0$*, we can find a solution* $\hat{\mathbf{x}} \in X$ *with robust objective value*

$$\hat{Z} = \boldsymbol{c}'\hat{\mathbf{x}} + \max_{\{S|\ S\subseteq\mathcal{A}, |S|\le\Gamma\}} \sum_{(i,j)\in S} d_{ij}\hat{x}_{ij}$$

such that

$$Z^* \le \hat{Z} \le (1+\epsilon)Z^*$$

by solving $2\lceil \log_2(|\mathcal{A}|\overline{\theta}/\epsilon)\rceil + 3$ *network flow problems, where* $\overline{\theta} = \max\{u_{ij}d_{ij}\ :\ (i,j) \in \mathcal{A}\}$.

Proof : Let $\theta^* \ge 0$ be such that $Z^* = Z(\theta^*)$. Since $Z(\theta)$ is a convex function

(Theorem 14(a)), we use binary search to find a $\hat{\theta}$ such that

$$|\hat{\theta} - \theta^*| \le \frac{\overline{\theta}}{2^k},$$

by solving $2k + 3$ minimum cost flow problems of the type described in Theorem 13. We first need to evaluate $Z(0), Z(\overline{\theta}/2), Z(\overline{\theta})$, and then we need two extra points $Z(\overline{\theta}/4)$ and $Z(3\overline{\theta}/4)$ in order to decide whether θ^* belongs in the interval $[0, \overline{\theta}/2]$ or $[\overline{\theta}/2, \overline{\theta}]$ or $[\overline{\theta}/4, 3\overline{\theta}/4]$. From then on, we need two extra evaluations in order to halve the interval θ^* can belong to.

From Theorem 14(b)

$$|Z(\hat{\theta}) - Z(\theta^*)| \le |\mathcal{A}||\hat{\theta} - \theta^*| \le |\mathcal{A}|\frac{\overline{\theta}}{2^k} \le \epsilon,$$

for $k = \lceil \log_2(|\mathcal{A}|\overline{\theta}/\epsilon) \rceil$. Note that $\hat{\mathbf{x}}$ is the flow corresponding to the nominal network flow problem for $\theta = \hat{\theta}$.

■

5.5 Experimental Results

In this section we consider concrete discrete optimization problems and solve the robust counterparts.

5.5.1 The Robust Knapsack Problem

The zero-one nominal knapsack problem is:

$$\begin{aligned} \text{maximize} \quad & \sum_{i \in N} c_i x_i \\ \text{subject to} \quad & \sum_{i \in N} w_i x_i \le b \\ & \boldsymbol{x} \in \{0,1\}^n. \end{aligned}$$

We assume that the weights $\tilde{w}_i$ are uncertain, independently distributed and follow symmetric distributions in $[w_i - \delta_i, w_i + \delta_i]$. The objective value vector $\boldsymbol{c}$ is not subject

Γ	Violation Probability	Optimal Value	Reduction
0	1	5592	0%
2.8	4.49×10^{-1}	5585	0.13%
36.8	5.71×10^{-3}	5506	1.54%
82.0	5.04×10^{-9}	5408	3.29%
200	0	5283	5.50%

Table 5.1: Robust Knapsack Solutions.

to data uncertainty. An application of this problem is to maximize the total value of goods to be loaded on a cargo that has strict weight restrictions. The weight of the individual item is assumed to be uncertain, independent of other weights and follows a symmetric distribution. In our robust model, we want to maximize the total value of the goods but allowing a maximum of 1% chance of constraint violation.

The robust Problem (5.2) is as follows:

$$\begin{aligned} \text{maximize} \quad & \sum_{i \in N} c_i x_i \\ \text{subject to} \quad & \sum_{i \in N} w_i x_i + \max_{\{S \cup \{t\} | \ S \subseteq N, |S| = \lfloor \Gamma \rfloor, t \in N \setminus S\}} \left\{ \sum_{j \in S} \delta_j x_j + (\Gamma - \lfloor \Gamma \rfloor) \delta_t x_t \right\} \leq b \\ & \boldsymbol{x} \in \{0,1\}^n. \end{aligned}$$

For this experiment, we solve Problem (5.3) using CPLEX 7.0 for a random knapsack problem of size, $|N| = 200$. We set the capacity limit, b to 4000, the nominal weight, w_i being randomly chosen from the set $\{20, 21, \ldots, 29\}$ and the cost c_i randomly chosen from the set $\{16, 17, \ldots, 77\}$. We set the weight uncertainty δ_i to equal 10% of the nominal weight. The time to solve the robust discrete problems to optimality using CPLEX 7.0 on a Pentium II 400 PC ranges from 0.05 to 50 seconds.

Under zero protection level, $\Gamma = 0$, the optimal value is $5,592$. However, with full protection, $\Gamma = 200$, the optimal value is reduced by 5.5% to $5,283$. In Table 5.1, we present a sample of the objective function value and the probability bound of constraint violation computed from Eq. (2.16). It is interesting to note that the optimal value is marginally affected when we increase the protection level. For instance, to have a probability guarantee of at most 0.57% chance of constraint violation, we only

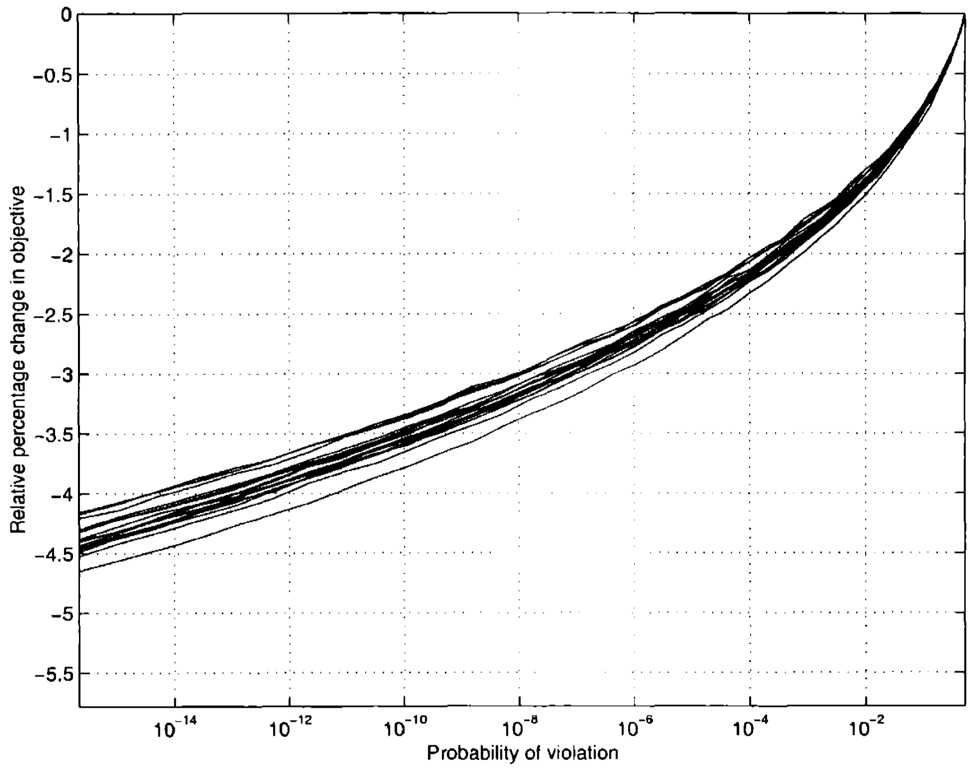

Figure 5-2: The tradeoff between robustness and optimality in twenty instances of the 0-1 knapsack problem.

reduce the objective by 1.54%. It appears that in this example we do not heavily penalize the objective function value in order to protect ourselves against constraint violation.

We repeated the experiment twenty times and in Figure 5-2 we report the tradeoff between robustness and optimality for all twenty problems. We observe that by allowing a profit reduction of 2%, we can make the probability of constraint violation smaller than 10^{-3}. Moroever, the conclusion did not seem to depend a lot on the specific instance we generated.

5.5.2 Robust Sorting

We consider the problem of minimizing the total cost of selecting k items out of a set of n items that can be expressed as the following integer optimization problem:

$$\begin{aligned} \text{minimize} \quad & \sum_{i \in N} c_i x_i \\ \text{subject to} \quad & \sum_{i \in N} x_i = k \\ & \boldsymbol{x} \in \{0,1\}^n. \end{aligned} \tag{5.19}$$

In this problem, the cost components are subjected to uncertainty. If the model is deterministic, we can easily solve the problem in $O(n \log n)$ by sorting the costs in ascending order and choosing the first k items. However, under the influence of data uncertainty, we will illustrate empirically that the deterministic model could lead to large deviations when the cost components are subject to uncertainty. Under our proposed Problem (5.5), we solve the following problem,

$$\begin{aligned} Z^*(\Gamma) = \quad \text{minimize} \quad & \boldsymbol{c}'\boldsymbol{x} + \max_{\{S \mid \; S \subseteq J, |S| \leq \Gamma\}} \sum_{j \in S} d_j x_j \\ \text{subject to} \quad & \sum_{i \in N} x_i = k \\ & \boldsymbol{x} \in \{0,1\}^n. \end{aligned} \tag{5.20}$$

We experiment with a problem of size $|N| = 200$ and $k = 100$. The cost and deviation components, c_j and d_j are uniformly distributed in $[50, 200]$ and $[20, 200]$ respectively. Since only k items will be selected, the robust solution for $\Gamma > k$ is the same as when $\Gamma = k$. Hence, Γ takes integral values from $[0, k]$. By varying Γ, we will illustrate empirically that we can control the deviation of the objective value under the influence of cost uncertainty.

We solve Problem (5.20) in two ways. First using Algorithm A, and second solving

Γ	$\overline{Z}(\Gamma)$	% Change in $\overline{Z}(\Gamma)$	$\sigma(\Gamma)$	% Change in $\sigma(\Gamma)$
0	8822	0 %	501.0	0.0 %
10	8827	0.056 %	493.1	-1.6 %
20	8923	1.145 %	471.9	-5.8 %
30	9059	2.686 %	454.3	-9.3 %
40	9627	9.125 %	396.3	-20.9 %
50	10049	13.91 %	371.6	-25.8 %
60	10146	15.00 %	365.7	-27.0 %
70	10355	17.38 %	352.9	-29.6 %
80	10619	20.37 %	342.5	-31.6 %
100	10619	20.37 %	340.1	-32.1 %

Table 5.2: Influence of Γ on $\overline{Z}(\Gamma)$ and $\sigma(\Gamma)$.

Problem (5.3):

$$\begin{aligned} \text{minimize} \quad & \boldsymbol{c}'\boldsymbol{x} + z\Gamma + \sum_{j \in N} p_j \\ \text{subject to} \quad & z + p_j \geq d_j x_j \qquad \forall j \in N \\ & \sum_{i \in N} x_i = k \\ & z \geq 0 \\ & p_j \geq 0 \\ & \boldsymbol{x} \in \{0,1\}^n. \end{aligned} \tag{5.21}$$

Algorithm A was able to find the robust solution for all $\Gamma \in \{0, \dots k\}$ in less than a second. The typical running time using CPLEX 7.0 to solve Problem (5.21) for only one of the Γ ranges from 30 to 80 minutes, which underscores the effectiveness of Algorithm A.

We let $\boldsymbol{x}(\Gamma)$ be an optimal solution to the robust model, with parameter Γ and define $\overline{Z}(\Gamma) = \boldsymbol{c}'\boldsymbol{x}(\Gamma)$ as the nominal cost in the absence of any cost deviations. To analyze the robustness of the solution, we simulate the distribution of the objective by subjecting the cost components to random perturbations. Under the simulation, each cost component independently deviates with probability p from the nominal value c_j to $c_j + d_j$. In Table 5.2, we report $\overline{Z}(\Gamma)$ and the standard deviation $\sigma(\Gamma)$ found in the simulation for $p = 0.2$ (we generated 20,000 instances to evaluate $\sigma(\Gamma)$).

Table 5.2 suggests that as we increase Γ, the standard deviation of the objective,

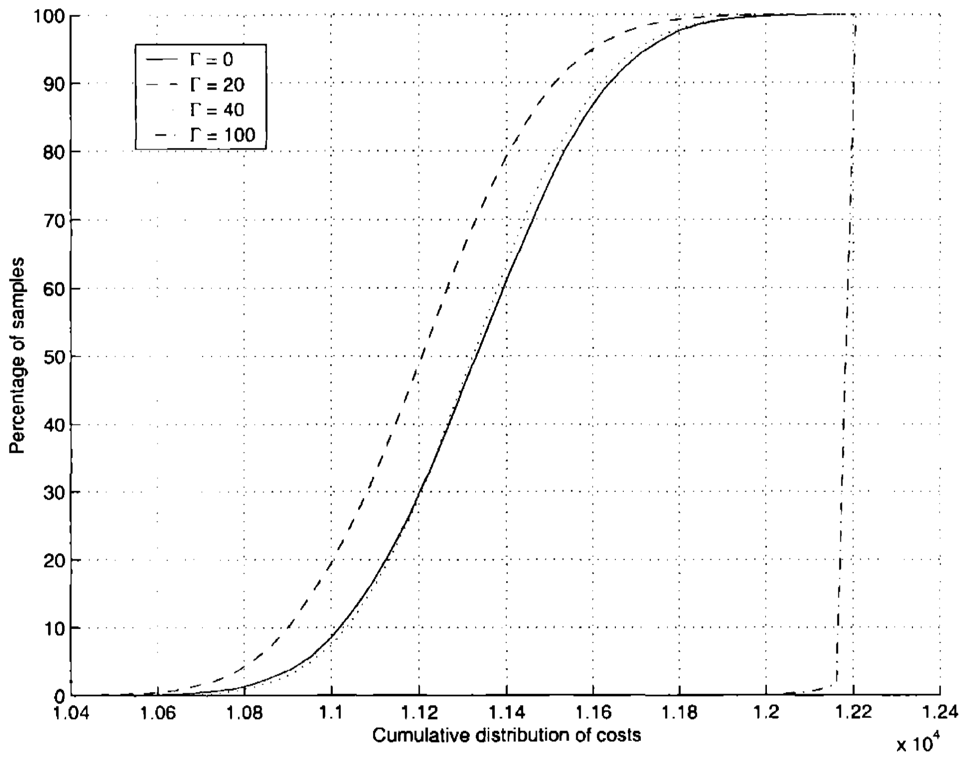

Figure 5-3: The cumulative distribution of cost (for $\rho = 0.2$) for various values of Γ for the robust sorting problem.

$\sigma(\Gamma)$ decreases, implying that the robustness of the solution increases, and $\overline{Z}(\Gamma)$ increases. Varying Γ we can find the tradeoff between the variability of the objective and the increase in nominal cost. Note that the robust formulation does not explicitly consider standard deviation. We chose to represent robustness in the numerical results with standard deviation of the objective, since standard deviation is the standard measure of variability and it has intuitive appeal.

In Figure 5-3 we report the cumulative distribution of cost (for $\rho = 0.2$) for various values of Γ for the robust sorting problem. We see that $\Gamma = 20$ dominates the nominal case $\Gamma = 0$, which in turn dominates $\Gamma = 100$ that appears over conservative. In particular, it is clear that not only the robust solution for $\Gamma = 20$ has lower variability than the nominal solution, it leads to a more favorable distribution of cost.

5.5.3 The Robust Shortest Path Problem

Given a directed graph $G = (\mathcal{N} \cup \{s,t\}, \mathcal{A})$, with non-negative arc cost c_{ij}, $(i,j) \in \mathcal{A}$, the shortest $\{s,t\}$ path problem seeks to find a path of minimum total arc cost from the source node s to the terminal node t. The problem can be modeled as a $0-1$ integer optimization problem:

$$
\begin{aligned}
\text{minimize} \quad & \sum_{(i,j)\in\mathcal{A}} c_{ij}x_{ij} \\
\text{subject to} \quad & \sum_{\{j:(i,j)\in\mathcal{A}\}} x_{ij} - \sum_{\{j:(j,i)\in\mathcal{A}\}} x_{ji} = \begin{cases} 1, & \text{if } i = s \\ \text{-1}, & \text{if } i = t \\ 0, & \text{otherwise,} \end{cases} \\
& \boldsymbol{x} \in \{0,1\}^{|\mathcal{A}|},
\end{aligned}
\tag{5.22}
$$

The shortest path problem surfaces in many important problems and has a wide range of applications from logistics planning to telecommunications (see for example, Ahuja et al. [1]. In these applications, the arc costs are estimated and subjected to uncertainty. The robust counterpart is then:

$$
\begin{aligned}
\text{minimize} \quad & \sum_{(i,j)\in\mathcal{A}} c_{ij}x_{ij} + \max_{\{S|\ S\subseteq\mathcal{A},|S|=\Gamma\}} \sum_{(i,j)\in S} d_{ij}x_{ij} \\
\text{subject to} \quad & \sum_{\{j:(i,j)\in\mathcal{A}\}} x_{ij} - \sum_{\{j:(j,i)\in\mathcal{A}\}} x_{ji} = \begin{cases} 1, & \text{if } i = s \\ \text{-1}, & \text{if } i = t \\ 0, & \text{otherwise,} \end{cases} \\
& \boldsymbol{x} \in \{0,1\}^{|\mathcal{A}|}.
\end{aligned}
\tag{5.23}
$$

Dijkstra's algorithm [10] solves the shortest path problem in $O(|\mathcal{N}|^2)$, while Algorithm A runs in $O(|\mathcal{A}||\mathcal{N}|^2)$. In order to test the performance of Algorithm A, we construct a randomly generated directed graph with $|\mathcal{N}| = 300$ and $|\mathcal{A}| = 1475$ as shown in Figure 5-4. The starting node, s is at the origin $(0,0)$ and the terminal node t is placed in coordinate $(1,1)$. The nominal arc cost, c_{ij} equals to the Euclidean distance between the adjacent nodes $\{i,j\}$ and the arc cost deviation, d_{ij} is set to γc_{ij}, where γ is uniformly distributed in $[0,8]$. Hence, some of the arcs have cost

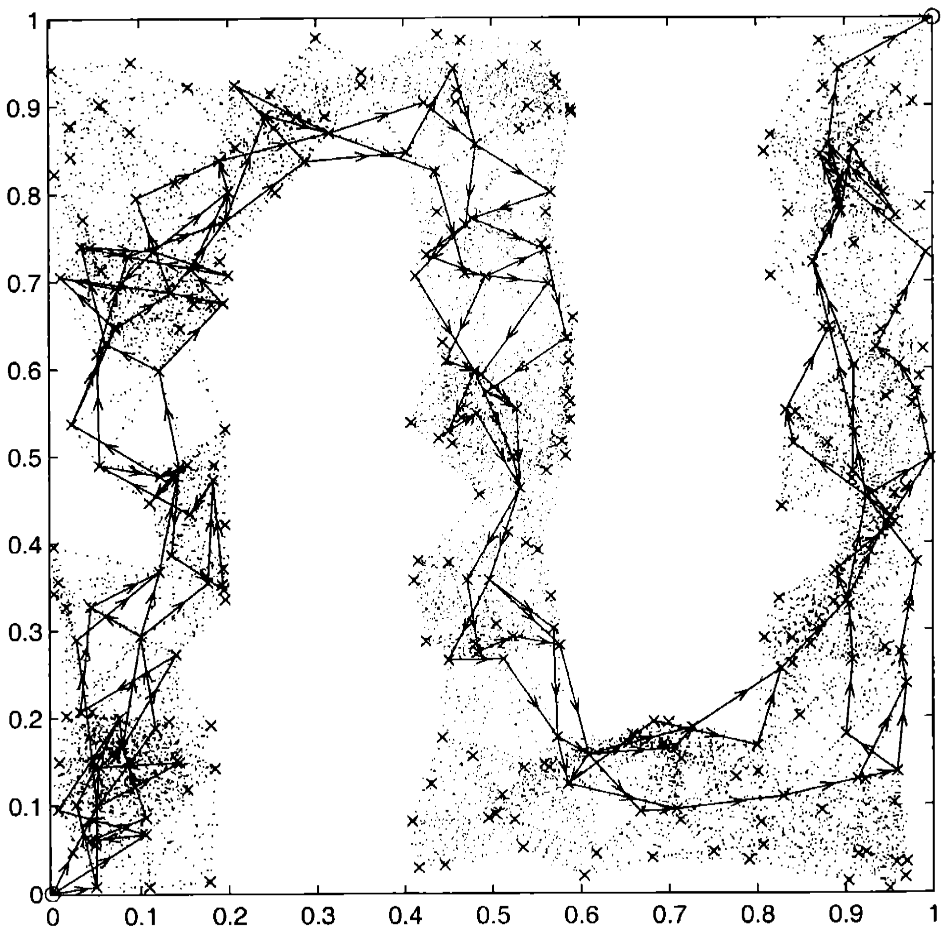

Figure 5-4: Randomly generated digraph and the set of robust shortest $\{s, t\}$ paths for various Γ values.

deviations of at most eight times of their nominal values. Using Algorithm A (calling Dijkstra's algorithm $|\mathcal{A}| + 1$ times), we solve for the complete set of robust shortest paths (for various Γ's), which are drawn in bold in Figure 5-4.

We simulate the distribution of the path cost by subjecting the arc cost to random perturbations. In each instance of the simulation, every arc (i, j) has cost that is independently perturbed, with probability ρ, from its nominal value c_{ij} to $c_{ij} + d_{ij}$. Setting $\rho = 0.1$, we generate 20,000 random scenarios and plot the distributions of the path cost for $\Gamma = 0, 3, 6$ and 10, which are shown in Figure 5-5. We observe that as Γ increases, the nominal path cost also increases, while cost variability decreases.

In Figure 5-6 we report the cumulative distribution of cost (for $\rho = 0.1$) for various

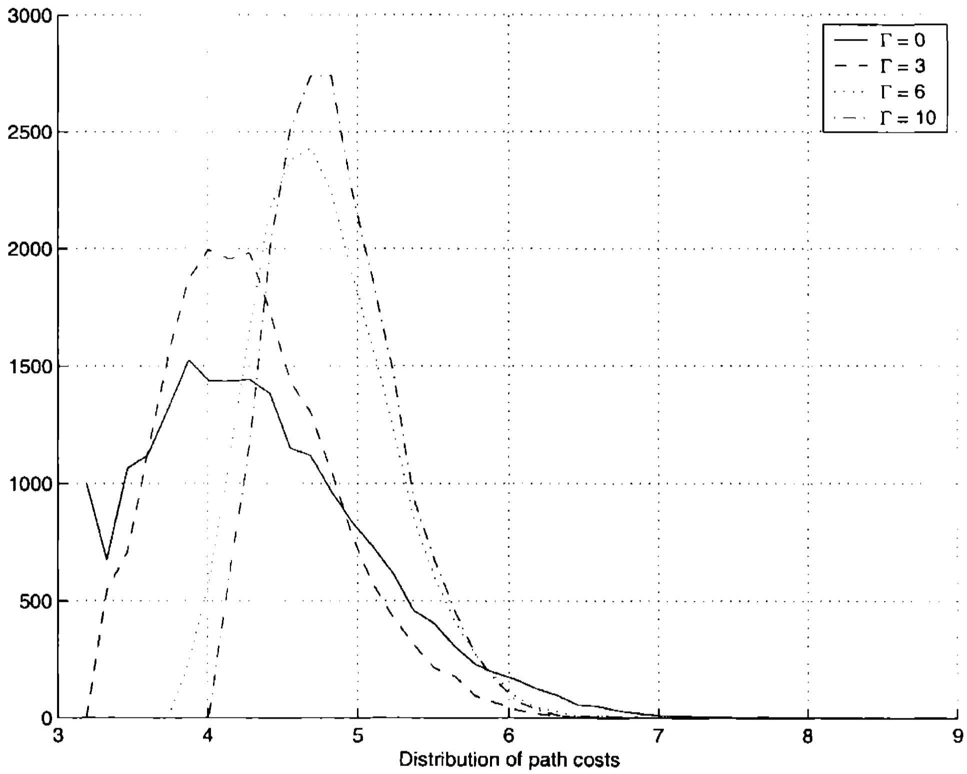

Figure 5-5: Influence of Γ on the distribution of path cost for $\rho = 0.1$.

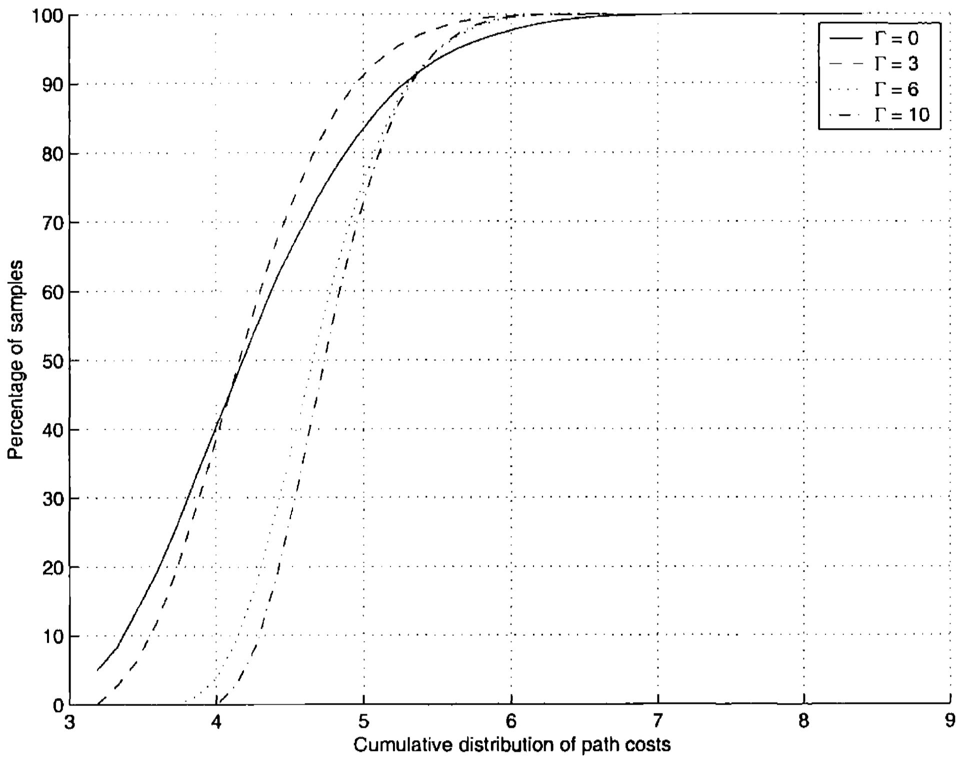

Figure 5-6: The cumulative distribution of cost (for $\rho = 0.1$) for various values of Γ for the robust shortest path problem.

values of Γ for the robust shortest path problem. Comparing the distributions for $\Gamma = 0$ (the nominal problem) and $\Gamma = 3$, we can see that none of the two distributions dominate each other. In other words, even if a decision maker is primarily cost conscious, he might still select to use a value of Γ that is different than zero, depending on his risk preference.

5.6 Conclusions

We feel that the proposed approach has the potential of being practically useful especially for combinatorial optimization and network flow problems that are subject to cost uncertainty. Unlike all other approaches that create robust solutions for combinatorial optimization problems, the proposed approach retains the complexity of the nominal problem or its approximability guarantee and offers the modeler the capability to control the tradeoff between cost and robustness by varying a single parameter Γ. For arbitrary discrete optimization problems, the increase in problem size is still moderate, and thus the proposed approach has the potential of being practically useful in this case as well.

Chapter 6

Robust Discrete Optimization under an Ellipsoidal Uncertainty Set

In Chapter 5, we propose an approach in solving robust discrete optimization problems that has the flexibility of adjusting the level of conservativeness of the solution while preserving the computational complexity of the nominal problem. This is attractive as it shows that adding robustness does not come at the price of a change in computational complexity. Ishii et. al. [18] consider solving a stochastic minimum spanning tree problem with costs that are independently and normally distributed leading to a similar framework as robust optimization with an ellipsoidal uncertainty set. However, to the best of our knowledge, there has not been any work or complexity results on extending this approach to solving general discrete optimization problems.

It is thus natural to ask whether adding robustness in discrete optimization problems under ellipsoidal sets leads to a change in computational complexity. In addition to the theoretical investigation, can we develop practically efficient methods to solve robust discrete optimization problems under ellipsoidal uncertainty sets?

Our objective in this chapter is to address these questions. Specifically our contributions include:

(a) Under an ellipsoidal uncertainty set, we show that the robust counterpart can be NP-hard even though the nominal problem is polynomially solvable in contrast with the uncertainty sets proposed in Chapter 5.

(b) Under an ellipsoidal uncertainty set with uncorrelated data, we show that the robust problem can be reduced to solving a collection of nominal problems with different linear objectives. If the distributions are identical, we show that we only require to solve $r + 1$ nominal problems, where r is the number of uncertain cost components, that is in this case the computational complexity is preserved. Under uncorrelated data, we propose an approximation method that solves the robust problem within an additive ϵ. The complexity of the method is $O((nd_{\max})^{1/4}\epsilon^{-1/2})$, where $d_{\max}$ is the largest number in the data describing the ellipsoidal set, that is the complexity is not polynomial as it depends on the data. We also propose a Frank-Wolfe type algorithm for this case, which we prove converges to a locally optimal solution, and in computational experiments is remarkably effective. We also link the robust problem with uncorrelated data to classical problems in parametric discrete optimization.

(c) We propose a generalization of the robust discrete optimization framework in Chapter 5 that allows the key parameter that controls the tradeoff between robustness and optimality to depend on the solution that results in increased flexibility and decreased conservatism, while maintaining the complexity of the nominal problem.

Structure of the chapter. In Section 6.1, we formulate robust discrete optimization problems under ellipsoidal uncertainty sets and show that the problem is NP-hard even for nominal problems that are polynomially solvable. In Section 6.2, we present structural results and establish that the robust problem under ball uncertainty (uncorrelated and identically distributed data) has the same complexity as the nominal problem. In Sections 6.3 and 6.4, we propose approximation methods for the robust problem under ellipsoidal uncertainty sets with uncorrelated but not identically distributed data. In Section 6.6, we present some experimental findings relating to the computation speed and the quality of robust solutions. The final section contains some concluding remarks.

6.1 Formulation of Robust Discrete Optimization Problems

A nominal discrete optimization problem is:

$$\begin{aligned} \text{minimize} \quad & \boldsymbol{c}'\boldsymbol{x} \\ \text{subject to} \quad & \boldsymbol{x} \in X, \end{aligned} \tag{6.1}$$

with $X \subseteq \{0,1\}^n$. We are interested in problems where each entry $\tilde{c}_j,\ j \in N = \{1,2,\ldots,n\}$ is uncertain and described by an uncertainty set C. Under the robust optimization paradigm, we solve

$$\begin{aligned} \text{minimize} \quad & \max_{\tilde{\boldsymbol{c}} \in C} \tilde{\boldsymbol{c}}'\boldsymbol{x} \\ \text{subject to} \quad & \boldsymbol{x} \in X. \end{aligned} \tag{6.2}$$

Writing $\tilde{\boldsymbol{c}} = \boldsymbol{c} + \tilde{\boldsymbol{s}}$, where $\boldsymbol{c}$ is the nominal value and the deviation $\tilde{\boldsymbol{s}}$ is restricted to the set $D = C - \boldsymbol{c}$, Problem (6.2) becomes:

$$\begin{aligned} \text{minimize} \quad & \boldsymbol{c}'\boldsymbol{x} + \xi(\boldsymbol{x}) \\ \text{subject to} \quad & \boldsymbol{x} \in X, \end{aligned} \tag{6.3}$$

where $\xi(\boldsymbol{x}) = \max_{\tilde{\boldsymbol{s}} \in D} \tilde{\boldsymbol{s}}'\boldsymbol{x}$. Special cases of Formulation (6.3) include:

(a) $D = \{\boldsymbol{s} :\ \tilde{s}_j \in [0, d_j]\}$, leading to $\xi(\boldsymbol{x}) = \boldsymbol{d}'\boldsymbol{x}$.

(b) $D = \{\boldsymbol{s} : \|\boldsymbol{\Sigma}^{-1/2}\boldsymbol{s}\|_2 \leq \Omega\}$ that models ellipsoidal uncertainty sets proposed by Ben-Tal and Nemirovski [7, 6, 4] and El-Ghaoui et al. [11, 12]. It easily follows that $\xi(\boldsymbol{x}) = \Omega\sqrt{\boldsymbol{x}'\boldsymbol{\Sigma}\boldsymbol{x}}$, where $\boldsymbol{\Sigma}$ is the covariance matrix of the random cost coefficients. For the special case that $\boldsymbol{\Sigma} = \text{diag}(d_1, \ldots, d_n)$, i.e., the random cost coefficients are uncorrelated, we obtain that $\xi(\boldsymbol{x}) = \Omega\sqrt{\sum_{j \in N} d_j x_j^2} = \Omega\sqrt{\boldsymbol{d}'\boldsymbol{x}}$.

(c) $D = \{s : 0 \leq s_j \leq d_j \ \forall j \in J, \sum_{k \in N} \frac{s_k}{d_k} \leq \Gamma\}$ proposed in Chapter 5. It follows that in this case $\xi(\boldsymbol{x}) = \max_{\{S: |S| = \Gamma, S \subseteq J\}} \sum_{j \in S} d_j x_j$, where J is the set of random cost components. We have also shown that Problem (6.3) reduces to solving at most $|J| + 1$ nominal problems for different cost vectors. In other words, the robust counterpart is polynomially solvable if the nominal problem is polynomially solvable.

Under models (a) and (c), robustness preserves the computational complexity of the nominal problem. Our objective in this chapter is to investigate the price (in increased complexity) of robustness under ellipsoidal uncertainty sets (model (b)) and propose effective algorithmic methods to tackle models (b), (c).

Our first result is unfortunately negative. Under ellipsoidal uncertainty sets with general covariance matrices, the price of robustness is high. The robust counterpart may become NP-hard even though the nominal problem is polynomially solvable.

Theorem 16 *The robust problem (6.3) with $\xi(\boldsymbol{x}) = \Omega\sqrt{\boldsymbol{x}'\Sigma\boldsymbol{x}}$ (Model (b)) is NP-hard, for the following classes of polynomially solvable nominal problems: shortest path, minimum cost assignment, resource scheduling, minimum spanning tree.*

Proof : Kouvelis and Yu [20] prove that the problem

$$\begin{aligned} \text{minimize} \quad & \max\{\boldsymbol{c}_1'\boldsymbol{x}, \boldsymbol{c}_2'\boldsymbol{x}\} \\ \text{subject to} \quad & \boldsymbol{x} \in X, \end{aligned} \tag{6.4}$$

is NP-hard for the polynomially solvable problems mentioned in the statement of the theorem. We show a simple transformation of Problem (6.4) to Problem (6.3) with $\xi(\boldsymbol{x}) = \Omega\sqrt{\boldsymbol{x}'\Sigma\boldsymbol{x}}$ as follows:

$$\begin{aligned} \max\{\boldsymbol{c}_1'\boldsymbol{x}, \boldsymbol{c}_2'\boldsymbol{x}\} &= \max\left\{\frac{\boldsymbol{c}_1'\boldsymbol{x} + \boldsymbol{c}_2'\boldsymbol{x}}{2} + \frac{\boldsymbol{c}_1'\boldsymbol{x} - \boldsymbol{c}_2'\boldsymbol{x}}{2}, \frac{\boldsymbol{c}_1'\boldsymbol{x} + \boldsymbol{c}_2'\boldsymbol{x}}{2} - \frac{\boldsymbol{c}_1'\boldsymbol{x} - \boldsymbol{c}_2'\boldsymbol{x}}{2}\right\} \\ &= \frac{\boldsymbol{c}_1'\boldsymbol{x} + \boldsymbol{c}_2'\boldsymbol{x}}{2} + \max\left\{\frac{\boldsymbol{c}_1'\boldsymbol{x} - \boldsymbol{c}_2'\boldsymbol{x}}{2}, -\frac{\boldsymbol{c}_1'\boldsymbol{x} - \boldsymbol{c}_2'\boldsymbol{x}}{2}\right\} \\ &= \frac{\boldsymbol{c}_1'\boldsymbol{x} + \boldsymbol{c}_2'\boldsymbol{x}}{2} + \left|\frac{\boldsymbol{c}_1'\boldsymbol{x} - \boldsymbol{c}_2'\boldsymbol{x}}{2}\right| \\ &= \frac{\boldsymbol{c}_1'\boldsymbol{x} + \boldsymbol{c}_2'\boldsymbol{x}}{2} + \frac{1}{2}\sqrt{\boldsymbol{x}'(\boldsymbol{c}_1 - \boldsymbol{c}_2)(\boldsymbol{c}_1 - \boldsymbol{c}_2)'\boldsymbol{x}}. \end{aligned}$$

The NP-hard Problem (6.4) is transformed to Problem (6.3) with $\xi(\boldsymbol{x}) = \Omega\sqrt{\boldsymbol{x}'\Sigma\boldsymbol{x}}$, $\boldsymbol{c} = (\boldsymbol{c_1} + \boldsymbol{c_2})/2$, $\Omega = 1/2$ and $\Sigma = (\boldsymbol{c_1} - \boldsymbol{c_2})(\boldsymbol{c_1} - \boldsymbol{c_2})'$. Thus, Problem (6.3) with $\xi(\boldsymbol{x}) = \Omega\sqrt{\boldsymbol{x}'\Sigma\boldsymbol{x}}$ is NP-hard. ■

We next would like to propose methods for model (b) with $\boldsymbol{\Sigma} = \text{diag}(d_1, \ldots, d_n)$. We are thus naturally led to consider the problem

$$\begin{aligned} G^* = \text{minimize} \quad & \boldsymbol{c}'\boldsymbol{x} + f(\boldsymbol{d}'\boldsymbol{x}) \\ \text{subject to} \quad & \boldsymbol{x} \in X, \end{aligned} \tag{6.5}$$

with $f(\cdot)$ a concave function. In particular, $f(x) = \Omega\sqrt{x}$ models ellipsoidal uncertainty sets with uncorrelated random cost coefficients (model (b)).

6.2 Structural Results

We first show that Problem (6.5) reduces to solving a number of nominal problems (6.1). Let $W = \{\boldsymbol{d}'\boldsymbol{x} \mid \boldsymbol{x} \in \{0,1\}^n\}$ and $\eta(w)$ be a subgradient of the concave function $f(\cdot)$ evaluated at w, that is, $f(u) - f(w) \leq \eta(w)(u - w)\ \forall u \in R$. If $f(w)$ is a differentiable function and $f'(0) = \infty$, we choose

$$\eta(w) = \begin{cases} f'(w) & \text{if } w \in W\backslash\{0\} \\ \frac{f(\underline{d}) - f(0)}{\underline{d}} & \text{if } w = 0 \end{cases},$$

where $\underline{d} = \min_{\{j:\ d_j > 0\}} d_j$.

Theorem 17 *Problem (6.5) is reducible to solving $|W|$ problems of the form:*

$$\begin{aligned} Z(w) = \text{minimize} \quad & (\boldsymbol{c} + \eta(w)\boldsymbol{d})'\boldsymbol{x} + f(w) - w\eta(w) \\ \textit{subject to} \quad & \boldsymbol{x} \in X. \end{aligned} \tag{6.6}$$

Moreover, $w^* = \arg\min_{\{w \in W\}} Z(w)$ yields the optimal solution to Problem (6.5) and $G^* = Z(w^*)$.

Proof : We first show that $G^* \geq \min_{w \in W} Z(w)$. Let $\boldsymbol{x}^*$ be an optimal solution to Problem (6.5) and $w^* = \boldsymbol{d}'\boldsymbol{x}^* \in W$. We have

$$\begin{aligned} G^* &= \boldsymbol{c}'\boldsymbol{x}^* + f(\boldsymbol{d}'\boldsymbol{x}^*) = \boldsymbol{c}'\boldsymbol{x}^* + f(w^*) = (\boldsymbol{c} + \eta(w^*)\boldsymbol{d})'\boldsymbol{x}^* + f(w^*) - w^*\eta(w^*) \\ &\geq \min_{\boldsymbol{x} \in X} (\boldsymbol{c} + \eta(w^*)\boldsymbol{d})'\boldsymbol{x} + f(w^*) - w^*\eta(w^*) = Z(w^*) \geq \min_{w \in W} Z(w). \end{aligned}$$

Conversely, for any $w \in W$, let y_w be an optimal solution to Problem (6.6). We have

$$\begin{aligned} Z(w) &= (\boldsymbol{c} + \eta(w)\boldsymbol{d})'\boldsymbol{y_w} + f(w) - w\eta(w) \\ &= \boldsymbol{c}'\boldsymbol{y_w} + f(\boldsymbol{d}'\boldsymbol{y_w}) + \eta(w)(\boldsymbol{d}'\boldsymbol{y_w} - w) - (f(\boldsymbol{d}'\boldsymbol{y_w}) - f(w)) \\ &\geq \boldsymbol{c}'\boldsymbol{y_w} + f(\boldsymbol{d}'\boldsymbol{y_w}) \qquad (6.7) \\ &\geq \min_{\boldsymbol{x} \in X} \boldsymbol{c}'\boldsymbol{x} + f(\boldsymbol{d}'\boldsymbol{x}) = G^*, \end{aligned}$$

where inequality (6.7) for $w \in W \backslash \{0\}$ follows, since $\eta(w)$ is a subgradient. To see that inequality (6.7) follows for $w = 0$ we argue as follows. Since $f(v)$ is concave and $v \geq \underline{d}\ \forall v \in W \backslash \{0\}$, we have

$$f(\underline{d}) \geq \frac{v - \underline{d}}{v} f(0) + \frac{\underline{d}}{v} f(v), \quad \forall v \in W \backslash \{0\}.$$

Rearranging, we have

$$\frac{f(v) - f(0)}{v} \leq \frac{f(\underline{d}) - f(0)}{\underline{d}} = \eta(0) \quad \forall v \in W \backslash \{0\},$$

leading to $\eta(0)(\boldsymbol{d}'\boldsymbol{y_w} - 0) - (f(\boldsymbol{d}'\boldsymbol{y_w}) - f(0)) \geq 0$. Therefore $G^* = \min_{w \in W} Z(w)$. ∎

Note that when $d_j = \sigma^2$, then $W = \{0, \sigma^2, \ldots, n\sigma^2\}$, and thus $|W| = n+1$, In this case, Problem (6.5) reduces to solving $n+1$ nominal problems (6.6), i.e., polynomial solvability is preserved. Specifically, for the case of an ellipsoidal uncertainty set $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{I}$, leading to $\xi(\boldsymbol{x}) = \Omega\sqrt{\sum_j \sigma^2 x_j^2} = \Omega\sigma\sqrt{\boldsymbol{e}'\boldsymbol{x}}$, we derive explicitly the subproblems involved.

Proposition 14 *Under an ellipsoidal uncertainty set with* $\xi(\boldsymbol{x}) = \Omega\sigma\sqrt{\boldsymbol{e}'\boldsymbol{x}}$,

$$G^* = \min_{w=0,1,\ldots,n} Z(w),$$

where

$$Z(w) = \begin{cases} \text{minimize}_{\boldsymbol{x}\in X} & \left(\boldsymbol{c} + \frac{\Omega\sigma}{2\sqrt{w}}\boldsymbol{e}\right)' \boldsymbol{x} + \frac{\Omega\sigma\sqrt{w}}{2} \quad w = 1,\ldots,n \\ \text{minimize}_{\boldsymbol{x}\in X} & (\boldsymbol{c} + \Omega\sigma\boldsymbol{e})'\boldsymbol{x} \quad w = 0. \end{cases} \tag{6.8}$$

Proof : With $\xi(\boldsymbol{x}) = \Omega\sigma\sqrt{\boldsymbol{e}'\boldsymbol{x}}$, we have $f(r) = \Omega\sigma\sqrt{r}$ and $f'(r) = \frac{\Omega\sigma}{2\sqrt{r}}$. Furthermore, $W = \{0,\ldots,n\}$, we choose $\eta(w) = f'(w)$, $\forall\, w \in W\backslash\{0\}$. Since $f'(0) = \infty$, and $\underline{d} = 1$, we obtain $\eta(0) = (f(\underline{d}) - f(0))/\underline{d} = f(1) - f(0) = \Omega\sigma$. ■

Proposition 14 suggests that for uncorrelated and identically distributed data, the computational complexity of the nominal problem is preserved.

An immediate corollary of Theorem 17 is to consider a parametric approach as follows:

Corollary 1 *An optimal solution to Problem (6.5) coincides with one of the optimal solutions to the parametric problem:*

$$\begin{aligned} \text{minimize} \quad & (\boldsymbol{c} + \theta\boldsymbol{d})'\boldsymbol{x} \\ \text{subject to} \quad & \boldsymbol{x} \in X, \end{aligned} \tag{6.9}$$

for $\theta \in [\eta(\boldsymbol{e}'\boldsymbol{d}), \eta(0)]$.

This establishes a connection of Problem (6.5) with parametric discrete optimization (see Gusfield [16] and Hassin and Tamir [17]). It turns out that if X is a matroid, the minimal set of optimal solutions to Problem (6.9) as θ varies is polynomial in size, see Eppstein [13] and Fern andez-Baca et al. [14]. For optimization over a matroid, the optimal solution depends on the ordering of the cost components. Since, as θ varies, it is easy to see that there are at most $\binom{n}{2} + 1$ different orderings, the corresponding robust problem is also polynomially solvable.

For the case of shortest paths, Karp and Orlin [19] provide a polynomial time algorithm using the parametric approach when all d_j's are equal. In contrast, the polynomial reduction in Proposition 14 applies to all discrete optimization problems.

More generally, $|W| \leq d_{\max} n$ with $d_{\max} = \max_j d_j$. If $d_{\max} \leq n^\alpha$, then Problem (6.5) reduces to solving $n^\alpha(n+1)$ nominal problems (6.6). However, when $d_{\max}$ is exponential in n, then the reduction does not preserve polynomiallity. For this reason, as well as deriving more practical algorithms even in the case that $|W|$ is polynomial in n we develop in the next section new algorithms.

6.3 Approximation via Piecewise Linear Functions

In this section, we develop a method for solving Problem (6.5) that is based on approximating the function $f(\cdot)$ with a piecewise linear concave function. We first show that if $f(\cdot)$ is a piecewise linear concave function with a polynomial number of segments, we can also reduce Problem (6.5) to solving a polynomial number of subproblems.

Proposition 15 *If $f(w), w \in [0, \boldsymbol{e}'\boldsymbol{d}]$ is a continuous piecewise linear concave function of k segments, Problem (6.5) can be reduced to solving k subproblems as follows:*

$$\begin{aligned} &\text{minimize} \quad (\boldsymbol{c} + \eta_j \boldsymbol{d})'\boldsymbol{x} \\ &\textit{subject to} \quad \boldsymbol{x} \in X, \end{aligned} \tag{6.10}$$

where η_j is the gradient of the jth linear piece of the function $f(\cdot)$.

Proof : The proof follows directly from Theorem 17 and the observations that if $f(w), w \in [0, \boldsymbol{e}'\boldsymbol{d}]$ is a continuous piecewise linear concave function of k linear pieces, the set of subgradients of each of the linear pieces constitutes the minimal set of subgradients for the function f. ■

We next show that approximating the function $f(\cdot)$ with a piecewise linear concave function leads to an approximate solution to Problem (6.5).

Theorem 18 *For $W = [\underline{w}, \bar{w}]$ such that $\boldsymbol{d}'\boldsymbol{x} \in W \ \forall \boldsymbol{x} \in X$, let $g(w)$, $w \in W$ be a piecewise linear concave function approximating the function $f(w)$ such that $-\epsilon_1 \leq f(w) - g(w) \leq \epsilon_2$ with $\epsilon_1, \epsilon_2 \geq 0$. Let $\boldsymbol{x}_{\mathrm{H}}$ be an optimal solution of the problem:*

$$\begin{aligned} \text{minimize} \quad & \boldsymbol{c}'\boldsymbol{x} + g(\boldsymbol{d}'\boldsymbol{x}) \\ \textit{subject to} \quad & \boldsymbol{x} \in X \end{aligned} \tag{6.11}$$

and let $G_{\mathrm{H}} = \boldsymbol{c}'\boldsymbol{x}_{\mathrm{H}} + f(\boldsymbol{d}'\boldsymbol{x}_{\mathrm{H}})$. Then,

$$G^* \leq G_{\mathrm{H}} \leq G^* + \epsilon_1 + \epsilon_2.$$

Proof : We have that

$$\begin{aligned} G^* &= \min_{\boldsymbol{x} \in X} \{\boldsymbol{c}'\boldsymbol{x} + f(\boldsymbol{d}'\boldsymbol{x})\} \\ &\leq G_{\mathrm{H}} = \boldsymbol{c}'\boldsymbol{x}_{\mathrm{H}} + f(\boldsymbol{d}'\boldsymbol{x}_{\mathrm{H}}) \\ &\leq \boldsymbol{c}'\boldsymbol{x}_{\mathrm{H}} + g(\boldsymbol{d}'\boldsymbol{x}_{\mathrm{H}}) + \epsilon_2 \qquad (6.12) \\ &= \min_{\boldsymbol{x} \in X} \{\boldsymbol{c}'\boldsymbol{x} + g(\boldsymbol{d}'\boldsymbol{x})\} + \epsilon_2 \\ &\leq \min_{\boldsymbol{x} \in X} \{\boldsymbol{c}'\boldsymbol{x} + f(\boldsymbol{d}'\boldsymbol{x})\} + \epsilon_1 + \epsilon_2 \qquad (6.13) \\ &= G^* + \epsilon_1 + \epsilon_2, \end{aligned}$$

where inequalities (6.12) and (6.13) follow from $-\epsilon_1 \leq f(w) - g(w) \leq \epsilon_2$. ■

We next apply the approximation idea to the case of ellipsoidal uncertainty sets. Specifically, we approximate the function $f(w) = \Omega\sqrt{w}$ in the domain $[\underline{w}, \bar{w}]$ with a piecewise linear concave function $g(w)$ satisfying $0 \leq f(w) - g(w) \leq \epsilon$ using the least number of linear pieces.

Proposition 16 *For $\epsilon > 0$, w_0 given, let $\phi = \epsilon/\Omega$ and for $i = 1, \ldots, k$ define*

$$w_i = \phi^2 \left\{ 2 \left(i + \sqrt{\frac{\sqrt{w_0}}{2\phi} + \frac{1}{4}} \right)^2 - \frac{1}{2} \right\}^2. \tag{6.14}$$

Let $g(w)$ be a piecewise linear concave function on the domain $w \in [w_0, w_k]$, with *breakpoints* $(w, g(w)) \in \{(w_0, \Omega\sqrt{w_0}), \ldots, (w_k, \Omega\sqrt{w_k})\}$. *Then, for all* $w \in [w_0, w_k]$

$$0 \leq \Omega\sqrt{w} - g(w) \leq \epsilon.$$

Proof : Since at the breakpoints w_i, $g(w_i) = \Omega\sqrt{w_i}$, $g(w)$ is a concave function with $g(w) \leq \Omega\sqrt{w}$, $\forall w \in [w_0, w_k]$. For $w \in [w_{i-1}, w_i]$, we have

$$\begin{aligned} \Omega\sqrt{w} - g(w) &= \Omega\sqrt{w} - \left\{\Omega\sqrt{w_{i-1}} + \frac{\Omega\sqrt{w_i} - \Omega\sqrt{w_{i-1}}}{w_i - w_{i-1}}(w - w_{i-1})\right\} \\ &= \Omega\left\{\sqrt{w} - \sqrt{w_{i-1}} - \frac{w - w_{i-1}}{\sqrt{w_i} + \sqrt{w_{i-1}}}\right\}. \end{aligned}$$

Clearly, the maximum value of $\Omega\sqrt{w} - g(w)$ is attained at $\sqrt{w^*} = \frac{\sqrt{w_i}+\sqrt{w_{i-1}}}{2}$. Therefore,

$$\begin{aligned} \Omega\sqrt{w} - g(w) &\leq \Omega\left\{\sqrt{w^*} - \sqrt{w_{i-1}} - \frac{w^* - w_{i-1}}{\sqrt{w_i} + \sqrt{w_{i-1}}}\right\} \\ &= \Omega\left\{\frac{\sqrt{w_i} - \sqrt{w_{i-1}}}{2} - \frac{\left(\frac{\sqrt{w_i}+\sqrt{w_{i-1}}}{2}\right)^2 - w_{i-1}}{\sqrt{w_i} + \sqrt{w_{i-1}}}\right\} \\ &= \Omega\left\{\frac{\sqrt{w_i} - \sqrt{w_{i-1}}}{2} - \frac{\left(\frac{\sqrt{w_i}+3\sqrt{w_{i-1}}}{2}\right)\left(\frac{\sqrt{w_i}-\sqrt{w_{i-1}}}{2}\right)}{\sqrt{w_i} + \sqrt{w_{i-1}}}\right\} \\ &= \frac{\Omega(\sqrt{w_i} - \sqrt{w_{i-1}})^2}{4(\sqrt{w_i} + \sqrt{w_{i-1}})} \\ &= \Omega\phi = \epsilon. \end{aligned} \tag{6.15}$$

The last Equality (6.15) by substituting Eq. (6.14). Since

$$\max_{w \in [w_{i-1}, w_i]} \left\{\Omega\sqrt{w} - g(w)\right\} = \epsilon,$$

the proposition follows. ■

Propositions 15, 16 and Theorem 18 lead to Algorithm 6.3.

Theorem 19 *Algorithm 6.3 is correct.*

Approximation by piecewise linear concave functions.
Input: $\boldsymbol{c}, \boldsymbol{d}, \underline{w}, \bar{w}, \Omega, \epsilon,\ f(x) = \Omega\sqrt{x}$ and a routine that optimizes a linear function over the set $X \subseteq \{0,1\}^n$.
Output: A solution $\boldsymbol{x}_{\mathrm{H}} \in X$ for which $G^* \leq \boldsymbol{c}'\boldsymbol{x}_{\mathrm{H}} + f(\boldsymbol{d}'\boldsymbol{x}_{\mathrm{H}}) \leq G^* + \epsilon$, where $G^* = \min_{\boldsymbol{x}\in X} \boldsymbol{c}'\boldsymbol{x} + f(\boldsymbol{d}'\boldsymbol{x})$.
Algorithm.

1. (Initialization) Let $\phi = \epsilon/\Omega$; Let $w_0 = \underline{w}$; Let

$$k = \left\lceil \sqrt{\frac{\Omega\sqrt{\bar{w}}}{2\epsilon} + \frac{1}{4}} - \sqrt{\frac{\Omega\sqrt{\underline{w}}}{2\epsilon} + \frac{1}{4}} \right\rceil = O\left(\sqrt{\frac{\Omega}{\epsilon}}(nd_{\max})^{\frac{1}{4}}\right)$$

where $d_{\max} = \max_j d_j$ and for $i = 1, \ldots, k$ let

$$w_i = \phi^2 \left\{ 2\left(i + \sqrt{\frac{\sqrt{\underline{w}}}{2\phi} + \frac{1}{4}}\right)^2 - \frac{1}{2} \right\}^2.$$

2. For $i = 1, \ldots, k$ solve the problem

$$\begin{aligned} Z_i = \text{minimize} \quad & \left(\boldsymbol{c} + \frac{\Omega}{\sqrt{w_i} + \sqrt{w_{i-1}}}\boldsymbol{d}\right)'\boldsymbol{x} \\ \text{subject to} \quad & \boldsymbol{x} \in X, \end{aligned} \tag{6.16}$$

Let $\boldsymbol{x_i}$ be an optimal solution to Problem (6.16).

3. Output $G^*_{\mathrm{H}} = Z_{i^*} = \min_{i=1,\ldots,k} Z_i$ and $\boldsymbol{x}_{\mathrm{H}} = \boldsymbol{x}_{i^*}$.

Proof : Using Proposition 16 we find a piecewise linear concave function $g(w)$ that approximates within a given tolerance $\epsilon > 0$ the function $\Omega\sqrt{w}$. From Proposition 15 and since the gradient of the ith segment of the function $g(w)$ for $w \in [w_{i-1}, w_i]$ is

$$\eta_i = \Omega\frac{\sqrt{w_i} - \sqrt{w_{i-1}}}{w_i - w_{i-1}} = \frac{\Omega}{\sqrt{w_i} + \sqrt{w_{i-1}}}.$$

we solve the Problems for $i = 1, \ldots, k$

$$\begin{aligned} Z_i = \text{minimize} \quad & \left(\boldsymbol{c} + \frac{\Omega}{\sqrt{w_i} + \sqrt{w_{i-1}}}\boldsymbol{d}\right)' \boldsymbol{x} \\ \text{subject to} \quad & \boldsymbol{x} \in X. \end{aligned}$$

Taking $G^*_{\mathrm{H}} = \min_i Z_i$ and using Theorem 18 it follows that Algorithm 6.3 produces a solution within ϵ. ■

Although the number of subproblems solved in Algorithm 6.3 is not polynomial with respect to the bit size of the input data, the computation involved is reasonable from a practical point of view. For example, in Table 6.1 we report the number of subproblems we need to solve for $\Omega = 4$, as a function of ϵ and $\boldsymbol{d}'\boldsymbol{e} = \sum_{j=1}^{n} d_j$.

ϵ	$\boldsymbol{d}'\boldsymbol{e}$	k
0.01	10	25
0.01	100	45
0.01	1000	80
0.01	10000	121
0.001	10	80
0.001	100	141
0.001	1000	251
0.001	10000	447

Table 6.1: Number of subproblems, k as a function of the desired precision ϵ, size of the problem $\boldsymbol{d}'\boldsymbol{e}$ and $\Omega = 4$.

6.4 A Frank-Wolfe Type Algorithm

A natural method to solve Problem (6.5) is to apply a Frank-Wolfe type algorithm, that is to successively linearize the function $f(\cdot)$.

The Frank-Wolfe type algorithm.
Input: $\boldsymbol{c}, \boldsymbol{d}, \Omega$, $\theta \in [\eta(\boldsymbol{d'e}), \eta(0)]$, $f(w)$, $\eta(w)$ and a routine that optimizes a linear function over the set $X \subseteq \{0,1\}^n$.
Output: A locally optimal solution to Problem (6.5).
Algorithm.

1. (Initialization) $k = 0; \boldsymbol{x_0} := \arg\min_{\boldsymbol{y} \in X}(\boldsymbol{c} + \theta \boldsymbol{d})'\boldsymbol{y})$
2. Until $\boldsymbol{d'x_{k+1}} = \boldsymbol{d'x_k}$, $\boldsymbol{x_{k+1}} := \arg\min_{\boldsymbol{y} \in X}(\boldsymbol{c} + \eta(\boldsymbol{d'x_k})\boldsymbol{d})'\boldsymbol{y})$.
3. Output $\boldsymbol{x_{k+1}}$.

We next show that Algorithm 6.4 converges to a locally optimal solution.

Theorem 20 *Let $\boldsymbol{x}$, $\boldsymbol{y}$ and $\boldsymbol{z}_\eta$ be optimal solutions to the following problems:*

$$\boldsymbol{x} = \arg\min_{\boldsymbol{u} \in X}(\boldsymbol{c} + \theta \boldsymbol{d})'\boldsymbol{u}, \tag{6.17}$$

$$\boldsymbol{y} = \arg\min_{\boldsymbol{u} \in X}(\boldsymbol{c} + \eta(\boldsymbol{d'x})\boldsymbol{d})'\boldsymbol{u} \tag{6.18}$$

$$\boldsymbol{z}_\eta = \arg\min_{\boldsymbol{u} \in X}(\boldsymbol{c} + \eta \boldsymbol{d})'\boldsymbol{u}, \tag{6.19}$$

for some η strictly between θ and $\eta(\boldsymbol{d'x}.)$

(a) *(Improvement)* $\boldsymbol{c'y} + f(\boldsymbol{d'y}) \leq \boldsymbol{c'x} + f(\boldsymbol{d'x})$.

(b) *(Monotonicity) If $\theta > \eta(\boldsymbol{d'x})$, then $\eta(\boldsymbol{d'x}) \geq \eta(\boldsymbol{d'y})$. Likewise, if $\theta < \eta(\boldsymbol{d'x})$, then $\eta(\boldsymbol{d'x}) \leq \eta(\boldsymbol{d'y})$. Hence, the sequence $\theta_k = \eta(\boldsymbol{d'x_k})$ for which*

$$\boldsymbol{x_k} = \arg\min_{x \in X}(\boldsymbol{c} + \eta(\boldsymbol{d'x_{k-1}})\boldsymbol{d})'\boldsymbol{x}$$

is either non-decreasing or non-increasing.

(c) *(Local optimality)*

$$\boldsymbol{c'y} + f(\boldsymbol{d'y}) \leq \boldsymbol{c'z}_\eta + f(\boldsymbol{d'z}_\eta),$$

for all η strictly between θ and $\eta(\boldsymbol{d'x})$. Moreover, if $\boldsymbol{d'y} = \boldsymbol{d'x}$, then the solution $\boldsymbol{y}$

is locally optimal, that is

$$\boldsymbol{y} = \arg\min_{\boldsymbol{u}\in X}(\boldsymbol{c} + \eta(\boldsymbol{d}'\boldsymbol{y})\boldsymbol{d})'\boldsymbol{u}$$

and

$$\boldsymbol{c}'\boldsymbol{y} + f(\boldsymbol{d}'\boldsymbol{y}) \leq \boldsymbol{c}'\boldsymbol{z}_\eta + f(\boldsymbol{d}'\boldsymbol{z}_\eta),$$

for all η between θ and $\eta(\boldsymbol{d}'\boldsymbol{y})$.

Proof : (a) We have

$$\begin{aligned}
\boldsymbol{c}'\boldsymbol{x} + f(\boldsymbol{d}'\boldsymbol{x}) &= (\boldsymbol{c} + \eta(\boldsymbol{d}'\boldsymbol{x})\boldsymbol{d})'\boldsymbol{x} - \eta(\boldsymbol{d}'\boldsymbol{x})\boldsymbol{d}'\boldsymbol{x} + f(\boldsymbol{d}'\boldsymbol{x}) \\
&\geq \boldsymbol{c}'\boldsymbol{y} + \eta(\boldsymbol{d}'\boldsymbol{x})\boldsymbol{d}'\boldsymbol{y} - \eta(\boldsymbol{d}'\boldsymbol{x})\boldsymbol{d}'\boldsymbol{x} + f(\boldsymbol{d}'\boldsymbol{x}) \\
&= \boldsymbol{c}'\boldsymbol{y} + f(\boldsymbol{d}'\boldsymbol{y}) + \{\eta(\boldsymbol{d}'\boldsymbol{x})(\boldsymbol{d}'\boldsymbol{y} - \boldsymbol{d}'\boldsymbol{x}) - (f(\boldsymbol{d}'\boldsymbol{y}) - f(\boldsymbol{d}'\boldsymbol{x})\} \\
&\geq \boldsymbol{c}'\boldsymbol{y} + f(\boldsymbol{d}'\boldsymbol{y}),
\end{aligned}$$

since $\eta(\cdot)$ is a subgradient.

(b) From the optimality of $\boldsymbol{x}$ and $\boldsymbol{y}$, we have

$$\begin{aligned}
\boldsymbol{c}'\boldsymbol{y} + \eta(\boldsymbol{d}'\boldsymbol{x})\boldsymbol{d}'\boldsymbol{y} &\leq \boldsymbol{c}'\boldsymbol{x} + \eta(\boldsymbol{d}'\boldsymbol{x})\boldsymbol{d}'\boldsymbol{x} \\
-(\boldsymbol{c}'\boldsymbol{y} + \theta\boldsymbol{d}'\boldsymbol{y}) &\leq -(\boldsymbol{c}'\boldsymbol{x} + \theta\boldsymbol{d}'\boldsymbol{x}).
\end{aligned}$$

Adding the two inequalities we obtain

$$(\boldsymbol{d}'\boldsymbol{x} - \boldsymbol{d}'\boldsymbol{y})(\eta(\boldsymbol{d}'\boldsymbol{x}) - \theta) \geq 0.$$

Therefore, if $\eta(\boldsymbol{d}'\boldsymbol{x}) > \theta$ then $\boldsymbol{d}'\boldsymbol{y} \leq \boldsymbol{d}'\boldsymbol{x}$ and since $f(w)$ is a concave function, i.e., $\eta(w)$ is non-increasing, $\eta(\boldsymbol{d}'\boldsymbol{y}) \geq \eta(\boldsymbol{d}'\boldsymbol{x})$. Likewise, if $\eta(\boldsymbol{d}'\boldsymbol{x}) < \theta$ then $\eta(\boldsymbol{d}'\boldsymbol{y}) \leq \eta(\boldsymbol{d}'\boldsymbol{x})$. Hence, the sequence $\theta_k = \eta(\boldsymbol{d}'\boldsymbol{x_k})$ is monotone.

(c) We first show that $\boldsymbol{d}'\boldsymbol{z}_\eta$ is in the convex hull of $\boldsymbol{d}'\boldsymbol{x}$ and $\boldsymbol{d}'\boldsymbol{y}$. From the optimality of $\boldsymbol{x}$, $\boldsymbol{y}$, and $\boldsymbol{z}_\eta$ we obtain

$$\begin{aligned} \boldsymbol{c}'\boldsymbol{x} + \theta \boldsymbol{d}'\boldsymbol{x} &\leq \boldsymbol{c}'\boldsymbol{z}_\eta + \theta \boldsymbol{d}'\boldsymbol{z}_\eta \\ \boldsymbol{c}'\boldsymbol{x} + \eta \boldsymbol{d}'\boldsymbol{x} &\geq \boldsymbol{c}'\boldsymbol{z}_\eta + \eta \boldsymbol{d}'\boldsymbol{z}_\eta \\ \boldsymbol{c}'\boldsymbol{y} + \eta(\boldsymbol{d}'\boldsymbol{x})\boldsymbol{d}'\boldsymbol{y} &\leq \boldsymbol{c}'\boldsymbol{z}_\eta + \eta(\boldsymbol{d}'\boldsymbol{x})\boldsymbol{d}'\boldsymbol{z}_\eta \\ \boldsymbol{c}'\boldsymbol{y} + \eta \boldsymbol{d}'\boldsymbol{y} &\geq \boldsymbol{c}'\boldsymbol{z}_\eta + \eta \boldsymbol{d}'\boldsymbol{z}_\eta \end{aligned}$$

From the first two inequalities we obtain

$$(\boldsymbol{d}'\boldsymbol{z}_\eta - \boldsymbol{d}'\boldsymbol{x})(\theta - \eta) \geq 0,$$

and from the last two we have

$$(\boldsymbol{d}'\boldsymbol{z}_\eta - \boldsymbol{d}'\boldsymbol{y})(\eta(\boldsymbol{d}'\boldsymbol{x}) - \eta) \geq 0.$$

If $\theta < \eta < \eta(\boldsymbol{d}'\boldsymbol{y})$, we conclude since $\eta(\cdot)$ is non-increasing that $\boldsymbol{d}'\boldsymbol{y} \leq \boldsymbol{d}'\boldsymbol{z}_\eta \leq \boldsymbol{d}'\boldsymbol{x}$. Likewise, if $\eta(\boldsymbol{d}'\boldsymbol{x}) < \eta < \theta$, we have $\boldsymbol{d}'\boldsymbol{x} \leq \boldsymbol{d}'\boldsymbol{z}_\eta \leq \boldsymbol{d}'\boldsymbol{y}$. Next, we have

$$\begin{aligned} \boldsymbol{c}'\boldsymbol{y} + f(\boldsymbol{d}'\boldsymbol{y}) &= (\boldsymbol{c} + \eta(\boldsymbol{d}'\boldsymbol{x})\boldsymbol{d})'\boldsymbol{y} - \eta(\boldsymbol{d}'\boldsymbol{x})\boldsymbol{d}'\boldsymbol{y} + f(\boldsymbol{d}'\boldsymbol{y}) \\ &\leq (\boldsymbol{c} + \eta(\boldsymbol{d}'\boldsymbol{x})\boldsymbol{d})'\boldsymbol{z}_\eta - \eta(\boldsymbol{d}'\boldsymbol{x})\boldsymbol{d}'\boldsymbol{y} + f(\boldsymbol{d}'\boldsymbol{y}) \\ &= \boldsymbol{c}'\boldsymbol{z}_\eta + f(\boldsymbol{d}'\boldsymbol{z}_\eta) + \{f(\boldsymbol{d}'\boldsymbol{y}) - f(\boldsymbol{d}'\boldsymbol{z}_\eta) - \eta(\boldsymbol{d}'\boldsymbol{x})(\boldsymbol{d}'\boldsymbol{y} - \boldsymbol{d}'\boldsymbol{z}_\eta)\} \\ &= \boldsymbol{c}'\boldsymbol{z}_\eta + f(\boldsymbol{d}'\boldsymbol{z}_\eta) + h(\boldsymbol{d}'\boldsymbol{z}_\eta) \\ &\leq \boldsymbol{c}'\boldsymbol{z}_\eta + f(\boldsymbol{d}'\boldsymbol{z}_\eta), \qquad (6.20) \end{aligned}$$

where inequality (6.20) follows from observing that the function $h(\alpha) = f(\boldsymbol{d}'\boldsymbol{y}) - f(\alpha) - \eta(\boldsymbol{d}'\boldsymbol{x})(\boldsymbol{d}'\boldsymbol{y} - \alpha)$ is a convex function with $h(\boldsymbol{d}'\boldsymbol{y}) = 0$ and $h(\boldsymbol{d}'\boldsymbol{x}) \leq 0$. Since $\boldsymbol{d}'\boldsymbol{z}_\eta$ is in the convex hull of $\boldsymbol{d}'\boldsymbol{x}$ and $\boldsymbol{d}'\boldsymbol{y}$, by convexity, $h(\boldsymbol{d}'\boldsymbol{z}_\eta) \leq \mu h(\boldsymbol{d}'\boldsymbol{y}) + (1 - \mu)h(\boldsymbol{d}'\boldsymbol{x}) \leq 0$, for some $\mu \in [0, 1]$. ■

Given a feasible solution, $\boldsymbol{x}$, Theorem 20(a) implies that we may improve the objective by solving a sequence of problems using Algorithm 6.4. Note that at each iteration, we are optimizing a linear function over X. Theorem 20(b) implies that the sequence of $\theta_k = \eta(\boldsymbol{d}'\boldsymbol{x_k})$ is monotone and since it is bounded it converges. Since X is finite, then the algorithm converges in a finite number of steps. Theorem 20(c) implies that at termination (recall that the termination condition is $\boldsymbol{d}'\boldsymbol{y} = \boldsymbol{d}'\boldsymbol{x}$) Algorithm 6.4 finds a locally optimal solution.

Suppose $\theta = \eta(\boldsymbol{e}'\boldsymbol{d})$ and $\{\boldsymbol{x_1}, \ldots, \boldsymbol{x_k}\}$ be the sequence of solutions of Algorithm 6.4. From Theorem 20(b), we have

$$\theta = \eta(\boldsymbol{e}'\boldsymbol{d}) \le \theta_1 = \eta(\boldsymbol{d}'\boldsymbol{x}_1) \le \ldots \le \theta_k = \eta(\boldsymbol{d}'\boldsymbol{x_k}).$$

When Algorithm 6.4 terminates at the solution $\boldsymbol{x_k}$, then from Theorem 20(c),

$$\boldsymbol{c}'\boldsymbol{x_k} + f(\boldsymbol{d}'\boldsymbol{x_k}) \le \boldsymbol{c}'\boldsymbol{z}_\eta + f(\boldsymbol{d}'\boldsymbol{z}_\eta), \tag{6.21}$$

where $\boldsymbol{z}_\eta$ is defined in Eq. (6.19) for all $\eta \in [\eta(\boldsymbol{e}'\boldsymbol{d}), \eta(\boldsymbol{d}'\boldsymbol{x_k})]$. Likewise, if $\bar{\theta} = \eta(0)$, and let $\{\boldsymbol{y_1}, \ldots, \boldsymbol{y_l}\}$ be the sequence of solutions of Algorithm 6.4, we have

$$\bar{\theta} = \eta(0) \ge \bar{\theta}_1 = \eta(\boldsymbol{d}'\boldsymbol{y}_1) \ge \ldots \ge \bar{\theta}_l = \eta(\boldsymbol{d}'\boldsymbol{y_l}),$$

and

$$\boldsymbol{c}'\boldsymbol{y_l} + f(\boldsymbol{d}'\boldsymbol{y_l}) \le \boldsymbol{c}'\boldsymbol{z}_\eta + f(\boldsymbol{d}'\boldsymbol{z}_\eta) \tag{6.22}$$

for all $\eta \in [\eta(\boldsymbol{d}'\boldsymbol{y_l}), \eta(0)]$. If $\eta(\boldsymbol{d}'\boldsymbol{x_k}) \ge \eta(\boldsymbol{d}'\boldsymbol{y_l})$, we have $\eta(\boldsymbol{d}'\boldsymbol{x_k}) \in [\eta(\boldsymbol{d}'\boldsymbol{y_l}), \eta(0)]$ and $\eta(\boldsymbol{d}'\boldsymbol{y_k}) \in [\eta(\boldsymbol{e}'\boldsymbol{d}), \eta(\boldsymbol{d}'\boldsymbol{x_l})]$. Hence, following from the inequalities (6.21) and (6.22), we conclude that

$$\boldsymbol{c}'\boldsymbol{y_l} + f(\boldsymbol{d}'\boldsymbol{y_l}) = \boldsymbol{c}'\boldsymbol{x_k} + f(\boldsymbol{d}'\boldsymbol{x_{k\eta}}) \le \boldsymbol{c}'\boldsymbol{z}_\eta + f(\boldsymbol{d}'\boldsymbol{z})$$

for all $\eta \in [\eta(\boldsymbol{e}'\boldsymbol{d}), \eta(\boldsymbol{d}'\boldsymbol{x_l})] \cup [\eta(\boldsymbol{d}'\boldsymbol{y_l}), \eta(0)] = [\eta(\boldsymbol{e}'\boldsymbol{d}), \eta(0)]$. Therefore, both $\boldsymbol{y_l}$ and $\boldsymbol{x_k}$ are globally optimal solutions. However, if $\eta(\boldsymbol{d}'\boldsymbol{y_l}) > \eta(\boldsymbol{d}'\boldsymbol{x_k})$, we are assured that the global optimal solution is $\boldsymbol{x_k}$, $\boldsymbol{y_l}$ or in $\{\boldsymbol{x} : \ \boldsymbol{x} = \arg\min_{\boldsymbol{u} \in X} (\boldsymbol{c} + \eta \boldsymbol{d})'\boldsymbol{u}$, $\eta \in (\eta(\boldsymbol{d}'\boldsymbol{x_k}), \eta(\boldsymbol{d}'\boldsymbol{y_l}))\}$. We next determine an error bound between the optimal objective and the objective of the best local solution, which is either $\boldsymbol{x_k}$ or $\boldsymbol{y_l}$.

Theorem 21 **(a)** *Let* $W = [\underline{w}, \overline{w}]$, $\eta(\underline{w}) > \eta(\overline{w})$, $X' = X \cap \{\boldsymbol{x} : \boldsymbol{d}'\boldsymbol{x} \in W\}$, $\boldsymbol{x_1}$ *and* $\boldsymbol{x_2}$ *being optimal solutions to the following problems,*

$$\boldsymbol{x_1} = \arg\min_{\boldsymbol{y}\in X'} (\boldsymbol{c} + \eta(\underline{w})\boldsymbol{d})'\boldsymbol{y}, \tag{6.23}$$

$$\boldsymbol{x_2} = \arg\min_{\boldsymbol{y}\in X'} (\boldsymbol{c} + \eta(\overline{w})\boldsymbol{d})'\boldsymbol{y}, \tag{6.24}$$

then

$$G'^* \le \min\{\boldsymbol{c}'\boldsymbol{x_1} + f(\boldsymbol{d}'\boldsymbol{x_1}), \boldsymbol{c}'\boldsymbol{x_2} + f(\boldsymbol{d}'\boldsymbol{x_2})\} \le G'^* + \varepsilon,$$

where

$$\begin{aligned} G'^* = \text{minimize} \quad & \boldsymbol{c}'\boldsymbol{y} + f(\boldsymbol{d}'\boldsymbol{y}) \\ \text{subject to} \quad & \boldsymbol{y} \in X', \end{aligned} \tag{6.25}$$

$$\varepsilon = \eta(\underline{w})(w^* - \underline{w}) + f(\underline{w}) - f(w^*),$$

and

$$w^* = \frac{f(\overline{w}) - f(\underline{w}) + \eta(\underline{w})\underline{w} - \eta(\overline{w})\overline{w}}{\eta(\underline{w}) - \eta(\overline{w})}$$

(b) *Suppose the feasible solutions* $\boldsymbol{x_1}$ *and* $\boldsymbol{x_2}$ *satisfy*

$$\boldsymbol{x_1} = \arg\min_{\boldsymbol{y}\in X} (\boldsymbol{c} + \eta(\boldsymbol{d}'\boldsymbol{x_1})\boldsymbol{d})'\boldsymbol{y}, \tag{6.26}$$

$$\boldsymbol{x_2} = \arg\min_{\boldsymbol{y}\in X} (\boldsymbol{c} + \eta(\boldsymbol{d}'\boldsymbol{x_2})\boldsymbol{d})'\boldsymbol{y}, \tag{6.27}$$

such that $\eta(\underline{w}) > \eta(\bar{w})$, *with* $\underline{w} = \boldsymbol{d}'\boldsymbol{x_1}$, $\bar{w} = \boldsymbol{d}'\boldsymbol{x_2}$ *and there exists an optimal solution* $\boldsymbol{x}^* = \arg\min_{\boldsymbol{y}\in X}(\boldsymbol{c} + \eta\boldsymbol{d})'\boldsymbol{y}$ *for some* $\eta \in (\eta(\bar{w}), \eta(\underline{w}))$, *then*

$$G^* \le \min\{\boldsymbol{c}'\boldsymbol{x_1} + f(\boldsymbol{d}'\boldsymbol{x_1}), \boldsymbol{c}'\boldsymbol{x_2} + f(\boldsymbol{d}'\boldsymbol{x_2})\} \le G^* + \varepsilon, \tag{6.28}$$

where $G^* = \boldsymbol{c}'\boldsymbol{x}^* + f(\boldsymbol{d}'\boldsymbol{x}^*)$.

Proof : **(a)** Let $g(w)$, $w \in W$ be a piecewise concave function comprising of two line segments through $(\underline{w}, f(\underline{w}))$, $(\overline{w}, f(\overline{w}))$ with respective subgradients $\eta(\underline{w})$ and $\eta(\overline{w})$.

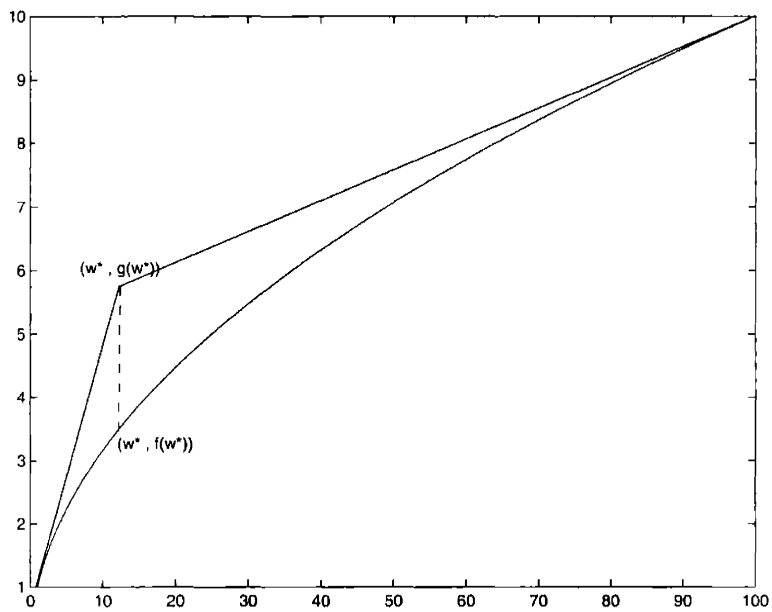

Figure 6-1: Illustration of the maximum gap between the function $f(w)$ and $g(w)$.

Clearly $f(w) \leq g(w)$ for $w \in W$, and hence, we have $-\varepsilon \leq f(w) - g(w) \leq 0$, where $\varepsilon = \max_{w \in W}(g(w) - f(w)) = g(w^*) - f(w^*)$, noting that the maximum difference occurs at the intersection (see Figure (6-1)). Therefore,

$$g(w^*) = \eta(\underline{w})(w^* - \underline{w}) + f(\underline{w}) = \eta(\overline{w})(w^* - \overline{w}) + f(\overline{w}).$$

Solving, we have

$$w^* = \frac{f(\overline{w}) - f(\underline{w}) + \eta(\underline{w})\underline{w} - \eta(\overline{w})\overline{w}}{\eta(\underline{w}) - \eta(\overline{w})}.$$

Applying Proposition 15 with X' instead of X and $k = 2$, we obtain

$$\min_{y \in X'} \boldsymbol{c'y} + g(\boldsymbol{d'y}) = \min \left\{ \boldsymbol{c'x_1} + g(\boldsymbol{d'x_1}), \boldsymbol{c'x_2} + g(\boldsymbol{d'x_2}) \right\}.$$

Finally, from Theorem 18, we have

$$G'^* \leq \min \left\{ \boldsymbol{c'x_1} + f(\boldsymbol{d'x_1}), \boldsymbol{c'x_2} + f(\boldsymbol{d'x_2}) \right\} \leq G'^* + \varepsilon.$$

(b) Under the stated conditions, observe that the optimal solutions of the problems

(6.26) and (6.27) are respectively the same as the optimal solutions of the problems (6.23) and (6.24). Let $\eta \in (\eta(\boldsymbol{d}'\boldsymbol{x_2}), \eta(\boldsymbol{d}'\boldsymbol{x_1}))$ such that $\boldsymbol{x}^* = \arg\min_{y \in X}(\boldsymbol{c} + \eta \boldsymbol{d})'\boldsymbol{y}$. We establish that

$$
\begin{aligned}
\boldsymbol{c}'\boldsymbol{x}^* + \eta \boldsymbol{d}'\boldsymbol{x}^* &\leq \boldsymbol{c}'\boldsymbol{x_1} + \eta \boldsymbol{d}'\boldsymbol{x_1} \\
\boldsymbol{c}'\boldsymbol{x}^* + \eta(\boldsymbol{d}'\boldsymbol{x_1})\boldsymbol{d}'\boldsymbol{x}^* &\geq \boldsymbol{c}'\boldsymbol{x_1} + \eta(\boldsymbol{d}'\boldsymbol{x_1})\boldsymbol{d}'\boldsymbol{x_1} \\
\boldsymbol{c}'\boldsymbol{x}^* + \eta \boldsymbol{d}'\boldsymbol{x}^* &\leq \boldsymbol{c}'\boldsymbol{x_2} + \eta \boldsymbol{d}'\boldsymbol{x_2} \\
\boldsymbol{c}'\boldsymbol{x}^* + \eta(\boldsymbol{d}'\boldsymbol{x_2})\boldsymbol{d}'\boldsymbol{x}^* &\geq \boldsymbol{c}'\boldsymbol{x_2} + \eta(\boldsymbol{d}'\boldsymbol{x_2})\boldsymbol{d}'\boldsymbol{x_2}.
\end{aligned}
$$

Since $\eta(\boldsymbol{d}'\boldsymbol{x_2}) < \eta < \eta(\boldsymbol{d}'\boldsymbol{x_1})$, it follows from the above that $\boldsymbol{d}'\boldsymbol{x}^* \in [\boldsymbol{d}'\boldsymbol{x_1}, \boldsymbol{d}'\boldsymbol{x_2}]$ and hence, $G^* = G'^*$ and the bounds of (6.25) follows from part (a). ■

If $\eta(\boldsymbol{d}'\boldsymbol{y_l}) > \eta(\boldsymbol{d}'\boldsymbol{x_k})$, Theorem 21(b) provides a guarantee on the quality of the best solution of the two locally optimal solution $\boldsymbol{x_k}$ and $\boldsymbol{y_l}$ relative to the global optimum. Moreover, we can improve the error bound by partitioning the interval $[\eta(\bar{w}), \eta(\underline{w})]$, with $\underline{w} = \boldsymbol{d}'\boldsymbol{y_l}$, $\bar{w} = \boldsymbol{d}'\boldsymbol{x_k}$ into two subintervals, $[\eta(\bar{w}), (\eta(\bar{w}) + \eta(\underline{w}))/2]$ and $[(\eta(\bar{w}) + \eta(\underline{w}))/2, \eta(\underline{w})]$ and applying Algorithm 2 in the intervals. Using Theorem 21(a), we can obtain improved bounds. Continuing this way, we can find the globally optimal solution.

6.5 Generalized Discrete Robust Formulation

In Chapter 5, we propose the following model for robust discrete optimization:

$$
\begin{aligned}
Z^* &= \min_{\boldsymbol{x} \in X} \boldsymbol{c}'\boldsymbol{x} + \max_{\{S \cup \{t\} |\ S \subseteq J, |S| = \lfloor \Gamma \rfloor, t \in J \setminus S\}} \left\{ \sum_{j \in S} d_j x_j + (\Gamma - \lfloor \Gamma \rfloor) d_t x_t \right\} \\
&= \min_{\boldsymbol{x} \in X} \boldsymbol{c}'\boldsymbol{x} + \max_{\{\boldsymbol{z} : \boldsymbol{e}'\boldsymbol{z} \leq \Gamma, \boldsymbol{0} \leq \boldsymbol{z} \leq \boldsymbol{e}\}} \left\{ \sum_{j \in J} d_j x_j z_j \right\}
\end{aligned}
\tag{6.29}
$$

Motivated from Theorem 3, if we select $\Gamma = \Lambda\sqrt{r}$, then the probability that the robust solution exceeds Z^* is approximately $1 - \Phi(\Lambda)$. Since in this case feasible solutions are restricted to binary values, we can achieve a less conservative solution by replacing r by $\sum_{j\in J} x_j^* = \boldsymbol{e'_J x}$, i.e., the parameter Γ in the robust problem depends on $\boldsymbol{e'_J x}$. We write $\Gamma = f(\boldsymbol{e'_J x})$, where $f(\cdot)$ is a concave function. Thus, we propose to solve the following problem:

$$\begin{aligned} Z^* = \quad & \text{minimize} \quad \boldsymbol{c'x} + \max_{\{\boldsymbol{z}:\boldsymbol{e'z}\le f(\boldsymbol{e'_J x}), \boldsymbol{0}\le\boldsymbol{z}\le\boldsymbol{e}\}} \left\{ \sum_{j\in J} d_j x_j z_j \right\} \\ & \text{subject to} \quad \boldsymbol{x} \in X. \end{aligned} \tag{6.30}$$

Without loss of generality, we assume that $d_1 \geq d_2 \geq \ldots \geq d_r$. We define $d_{r+1} = 0$ and let $S_l = \{1, \ldots, l\}$. For notational convenience, we also define $d_0 = 0$ and $S_0 = J$.

Theorem 22 *Let $\eta(w)$ be a subgradient of the concave function $f(\cdot)$ evaluated at w. Problem (6.30) satisfies $Z^* = \min_{(l,k):l,k\in J\cup\{0\}} Z_{lk}$, where*

$$\begin{aligned} Z_{lk} = \quad & \text{minimize} \quad \boldsymbol{c'x} + \sum_{j\in S_l} (d_j - d_l)x_j + \eta(k) d_l \boldsymbol{e'_J x} + d_l(f(k) - k\eta(k)) \\ & \text{subject to} \quad \boldsymbol{x} \in X. \end{aligned} \tag{6.31}$$

Proof : By strong duality of the inner maximization function with respect to $\boldsymbol{z}$, Problem (6.30) is equivalent to solving the following problem:

$$\begin{aligned} \text{minimize} \quad & \boldsymbol{c'x} + \sum_{j\in J} p_j + f(\boldsymbol{e'_J x})\theta \\ \text{subject to} \quad & p_j \geq d_j x_j - \theta \qquad \forall j \in J \\ & p_j \geq 0 \qquad \forall j \in J \\ & x \in X \\ & \theta \geq 0, \end{aligned} \tag{6.32}$$

We eliminate the variables p_j and express Problem (6.32) as follows:

$$\begin{aligned} \text{minimize} \quad & \boldsymbol{c'x} + \sum_{j\in J} \max\{d_j x_j - \theta, 0\} + f(\boldsymbol{e'_J x})\theta \\ \text{subject to} \quad & x \in X \\ & \theta \geq 0. \end{aligned} \tag{6.33}$$

Since $\boldsymbol{x} \in \{0,1\}^n$, we observe that

$$\max\{d_j x_j - \theta, 0\} = \begin{cases} d_j - \theta & \text{if } x_j = 1 \text{ and } d_j \geq \theta \\ 0 & \text{if } x_j = 0 \text{ or } d_j < \theta. \end{cases} \tag{6.34}$$

By restricting the interval of θ can vary we obtain that $Z^* = \min_{\theta \geq 0} \min_{l=0,\ldots,r} Z_l(\theta)$ where $Z_l(\theta)$, $l = 1, \ldots, r$, is defined for $\theta \in [d_l, d_{l+1}]$ is

$$\begin{aligned} Z_l(\theta) = \text{minimize} \quad & \boldsymbol{c'x} + \sum_{j \in S_l} (d_j - \theta) x_j + f(\boldsymbol{e'_J x})\theta \\ \text{subject to} \quad & \boldsymbol{x} \in X, \end{aligned} \tag{6.35}$$

and for $\theta \in [d_1, \infty)$:

$$\begin{aligned} Z_0(\theta) = \quad \text{minimize} \quad & \boldsymbol{c'x} + f(\boldsymbol{e'_J x})\theta \\ \text{subject to} \quad & x \in X. \end{aligned} \tag{6.36}$$

Since each function $Z_l(\theta)$ is optimized over the interval $[d_l, d_{l+1}]$, the optimal solution is realized in either d_l or d_{l+1}. Hence, we can restrict θ from the set $\{d_1, \ldots, d_r, 0\}$ and establish that

$$Z^* = \min_{l \in J \cup \{0\}} \boldsymbol{c'x} + \sum_{j \in S_l} (d_j - d_l) x_j + f(\boldsymbol{e'_J x}) d_l. \tag{6.37}$$

Since $\boldsymbol{e'_J x} \in \{0, 1, \ldots, r\}$, we apply Theorem 17 to obtain the subproblem decomposition of (6.31). ■

Theorem 22 suggests that the robust problem remains polynomially solvable if the nominal problem is polynomially solvable, but at the expense of higher computational complexity. We next explore faster algorithms that are only guarantee local optimality. In this spirit and analogously to Theorem 20, we provide a necessary condition for optimality, which can be exploited for local search algorithmic design.

Theorem 23 *An optimal solution, $\boldsymbol{x}$ to the Problem (6.30) is also an optimal solu-*

tion to the following problem:

$$\begin{aligned} &\text{minimize} \quad \boldsymbol{c}'\boldsymbol{y} + \sum_{j \in S_{l^*}} (d_j - d_{l^*}) y_j + \eta(\boldsymbol{e}_J'\boldsymbol{x}) d_{l^*} \boldsymbol{e}_J'\boldsymbol{y} \\ &\text{subject to} \quad \boldsymbol{y} \in X, \end{aligned} \tag{6.38}$$

where $l^* = \arg\min_{l \in J \cup \{0\}} \sum_{j \in S_l} (d_j - d_l) x_j + f(\boldsymbol{e}_J'\boldsymbol{x}) d_l$.

Proof : Suppose $\boldsymbol{x}$ is an optimal solution for Problem (6.30) but not for Problem (6.38). Let $\boldsymbol{y}$ be the optimal solution to Problem (6.38). Therefore,

$$\begin{aligned}
& \boldsymbol{c}'\boldsymbol{x} + \max_{\{\boldsymbol{z}: \boldsymbol{e}'\boldsymbol{z} \le f(\boldsymbol{e}_J'\boldsymbol{x}), \boldsymbol{0} \le \boldsymbol{z} \le \boldsymbol{e}\}} \left\{ \sum_{j \in J} d_j x_j z_j \right\} \\
= \; & \min_{l \in J \cup \{0\}} \boldsymbol{c}'\boldsymbol{x} + \sum_{j \in S_l} (d_j - d_l) x_j + f(\boldsymbol{e}_J'\boldsymbol{x}) d_l \qquad (6.39) \\
= \; & \boldsymbol{c}'\boldsymbol{x} + \sum_{j \in S_{l^*}} (d_j - d_{l^*}) x_j + f(\boldsymbol{e}_J'\boldsymbol{x}) d_{l^*} \\
= \; & \boldsymbol{c}'\boldsymbol{x} + \sum_{j \in S_{l^*}} (d_j - d_{l^*}) x_j + \eta(\boldsymbol{e}_J'\boldsymbol{x}) d_{l^*} \boldsymbol{e}_J'\boldsymbol{x} - \eta(\boldsymbol{e}_J'\boldsymbol{x}) d_{l^*} \boldsymbol{e}_J'\boldsymbol{x} + f(\boldsymbol{e}_J'\boldsymbol{x}) d_{l^*} \\
> \; & \boldsymbol{c}'\boldsymbol{y} + \sum_{j \in S_{l^*}} (d_j - d_{l^*}) y_j + \eta(\boldsymbol{e}_J'\boldsymbol{x}) d_{l^*} \boldsymbol{e}_J'\boldsymbol{y} - \eta(\boldsymbol{e}_J'\boldsymbol{x}) d_{l^*} \boldsymbol{e}_J'\boldsymbol{x} + f(\boldsymbol{e}_J'\boldsymbol{x}) d_{l^*} \\
= \; & \boldsymbol{c}'\boldsymbol{y} + \sum_{j \in S_{l^*}} (d_j - d_{l^*}) y_j + f(\boldsymbol{e}_J'\boldsymbol{y}) d_{l^*} + \\
& \quad \left(\eta(\boldsymbol{e}_J'\boldsymbol{x})(\boldsymbol{e}_J'\boldsymbol{y} - \boldsymbol{e}_J'\boldsymbol{x}) - \left(f(\boldsymbol{e}_J'\boldsymbol{y}) - f(\boldsymbol{e}_J'\boldsymbol{x}) \right) \right) d_{l^*} \\
\ge \; & \boldsymbol{c}'\boldsymbol{y} + \sum_{j \in S_{l^*}} (d_j - d_{l^*}) y_j + f(\boldsymbol{e}_J'\boldsymbol{y}) d_{l^*} \\
\ge \; & \min_{l \in J \cup \{0\}} \boldsymbol{c}'\boldsymbol{y} + \sum_{j \in S_l} (d_j - d_l) y_j + f(\boldsymbol{e}_J'\boldsymbol{y}) d_l \\
= \; & \boldsymbol{c}'\boldsymbol{y} + \max_{\{\boldsymbol{z}: \boldsymbol{e}'\boldsymbol{z} \le f(\boldsymbol{e}_J'\boldsymbol{y}), \boldsymbol{0} \le \boldsymbol{z} \le \boldsymbol{e}\}} \left\{ \sum_{j \in J} d_j y_j z_j \right\} \qquad (6.40)
\end{aligned}$$

where the Eqs. (6.39) and (6.40) follows from Eq. (6.37). This contradicts that $\boldsymbol{x}$ is optimal. ∎

6.6 Experimental Results

In this section we provide experimental evidence on the effectiveness of Algorithm 6.4. We apply Algorithm 6.4 as follows. We start with two initial solutions $\boldsymbol{x_1}$ and $\boldsymbol{x_2}$. Starting with $\boldsymbol{x_1}$ ($\boldsymbol{x_2}$) Algorithm 6.4 finds a locally optimal solution $\boldsymbol{y_1}$ ($\boldsymbol{y_2}$). If $\boldsymbol{y_1} = \boldsymbol{y_2}$, by Theorem 20, the optimum solution is found. Otherwise, we report the optimality gap ε derived from Theorem 21. If we want to find the optimal solution, we partition into smaller search regions using Theorem 19 and repeatedly apply Algorithm 6.4 until all regions are covered.

We apply the proposed approach to the binary knapsack and the uniform matroid problems.

6.6.1 The Robust Knapsack Problem

The binary knapsack problem is:

$$\begin{aligned} \text{maximize} \quad & \sum_{i \in N} \tilde{c}_i x_i \\ \text{subject to} \quad & \sum_{i \in N} w_i x_i \leq b \\ & \boldsymbol{x} \in \{0,1\}^n. \end{aligned}$$

We assume that the costs $\tilde{c}_i$ are random variables that are independently distributed with mean c_i and variance $d_i = \sigma_i^2$. Under the ellipsoidal uncertainty set, the robust model is:

$$\begin{aligned} \text{maximize} \quad & \sum_{i \in N} c_i x_i + \Omega\sqrt{\boldsymbol{d'x}} \\ \text{subject to} \quad & \sum_{i \in N} w_i x_i \leq b \\ & \boldsymbol{x} \in \{0,1\}^n. \end{aligned}$$

The instance of the robust knapsack problem is generated randomly with $|N| = 200$ and capacity limit, b equals $20,000$. The nominal weight w_i is randomly chosen from the set $\{100, \ldots, 1500\}$, the cost c_i is randomly chosen from the set $\{10,000, \ldots, 15,000\}$, and the standard deviation σ_j is dependent on c_j such that $\sigma_j = \delta_j c_j$, where

Ω	Z_H	Iterations	ε	$\frac{\varepsilon}{Z_H}$
1.0	1965421.36	4	0	0
2.0	2054638.82	6	0	0
2.5	2097656.46	6	0	0
3.0	2140207.75	6	3.1145	1.45523×10^{-6}
3.05	2144317.00	5	0	0
3.5	2182235.78	5	0	0
4.0	2224365.19	6	3.4046	1.53059×10^{-6}
4.5	2266054.21	7	0	0
5.0	2307475.12	8	0	0

Table 6.2: Robust Knapsack Solutions.

δ_j is uniformly distributed in $[0, 1]$. We vary the parameter Ω from 1 to 5 and report in Table 6.2 the best attainable objective, Z_H, the number of instance of nominal problem solved, as well as the optimality gap ε.

It is surprising that in all of the instances, we can obtain the optimal solution of the robust problem using a small number of iterations. Even for the cases, $\Omega = 3, 4$, where the Algorithm 6.4 terminates with more than one local minimum solutions, the resulting optimality gap is very small, which is usually acceptable in practical settings.

6.6.2 The Robust Minimum Cost over a Uniform Matroid

We consider the problem of minimizing the total cost of selecting k items out of a set of n items that can be expressed as the following integer optimization problem:

$$\begin{aligned} \text{minimize} \quad & \sum_{i \in N} \tilde{c}_i x_i \\ \text{subject to} \quad & \sum_{i \in N} x_i = k \\ & \boldsymbol{x} \in \{0, 1\}^n. \end{aligned} \tag{6.41}$$

In this problem, the cost components are subjected to uncertainty. If the model is deterministic, we can easily solve the problem in $O(n \log n)$ by sorting the costs in ascending order and choosing the first k items. In the robust framework under the ellipsoidal uncertainty set, we solve the following problem:

$$\begin{aligned} \text{minimize} \quad & \boldsymbol{c}'\boldsymbol{x} + \Omega\sqrt{\boldsymbol{d}'\boldsymbol{x}} \\ \text{subject to} \quad & \sum_{i \in N} x_i = k \\ & \boldsymbol{x} \in \{0,1\}^n. \end{aligned} \tag{6.42}$$

Since the underlying set is a matroid, it is well known that Problem (6.42) can be solved in strongly polynomial time using parametric optimization. Instead, we apply Algorithm 6.4 and observe the number of iterations needed before converging to a local minimum solution. Setting $|k| = |N|/2$, c_j and $\sigma_j = \sqrt{d_j}$ being uniformly distributed in $[5000, 20000]$ and $[500, 5000]$ respectively, we study the convergence properties as we vary $|N|$ from 200 to 20,000 and Ω from 1 to 3. For a given $|N|$ and Ω, we generate $\boldsymbol{c}$ and $\boldsymbol{d}$ randomly and solve 100 instances of the problem. Aggregating the results from solving the 100 instances, we report in Table 6.3 the average number of iterations before finding a local solution, the maximum relative optimality gap, ε/Z_H and the percentage of the local minimum solutions that are global, i.e $\varepsilon = 0$.

The overall performance of Algorithm 6.4 is surprisingly good. It also suggests scalability, as the number of iterations is marginally affected by an increase in $|N|$. In fact, in most of the problems tested, we obtain the optimal solution by solving less than 10 iterations of the nominal problem. Even in cases when local solutions are found, the corresponding optimality gap is negligible. In summary, Algorithm 6.4 seems practically promising.

6.7 Conclusions

A message of the present chapter is that the complexity of robust discrete optimization is affected by the choice of the uncertainty set. For ellipsoidal uncertainty sets, we have shown an increase in complexity for the robust counterpart of a discrete optimization problem for general covariance matrices $\boldsymbol{\Sigma}$, a preservation of complexity when $\boldsymbol{\Sigma} = \sigma \boldsymbol{I}$ (uncorrelated and identically distributed data), while we have left

Ω	$\|N\|$	Ave. Iter.	$\max \frac{\varepsilon}{Z_H}$	Opt. Sol. %
1	200	5.73	7.89×10^{-7}	98%
1	500	5.91	3.71×10^{-8}	99%
1	1000	6.18	5.80×10^{-9}	99%
1	2000	6.43	0	100%
1	5000	6.72	0	100%
1	10000	6.92	0	100%
1	20000	6.98	0	100%
2	200	6.24	0	100%
2	500	6.50	0	100%
2	1000	6.80	0	100%
2	2000	6.95	0	100%
2	5000	6.98	0	100%
2	10000	7.01	0	100%
2	20000	7.02	0	100%
3	200	6.55	1.62×10^{-6}	94%
3	500	6.85	7.95×10^{-8}	97%
3	1000	6.92	0	100%
3	2000	7.01	1.08×10^{-9}	99%
3	5000	7.06	5.13×10^{-10}	98%
3	10000	7.07	0	100%
3	20000	7.07	0	100%

Table 6.3: Performance of Algorithm 6.4 the Robust Minimum Cost problem over a Uniform Matroid.

open the complexity when the matrix $\mathbf{\Sigma}$ is diagonal (uncorrelated data). In the latter case, we proposed two algorithms that in computational experiments have excellent empirical performance.

Chapter 7

Conclusions

In order for robust optimization to have an impact in theory and practice of optimization, we feel that two criteria are important:

(a) Preserving the computational tractability of the nominal problem both theoretically and most importantly practically.

(b) Being able to find a guarantee on the probability that the robust solution is feasible, when the uncertain coefficients obey some natural probability distributions.

In this thesis we propose robust methodologies that meet the above requirements for a broad range of optimization problems. Specifically,

1. **Linear and Mixed Integer Optimization:** The robust counterpart of a LP (or MIP) remains a LP (or MIP) (Chapter 2). If each data has bounded, symmetric and independent distribution, we derive tight probability bound and show that our approach is far less conservative compared to the classical robust method of Soyster [26].

2. **Quadratic Constrained Quadratic Optimization:** The robust counterpart of a quadratic constraint becomes a collection of second order conic constraints (Chapter 4). Under normal distributions, the probability bound suggests that the robust solution remains feasible with high probability without being over-conservative in the choice of the protection level.

3. **Conic Optimization:** The robust SOCP becomes a SOCP and the robust SDP becomes a SDP (Chapter 4). Under normal distributions, we relate the probability bound of feasibility with the underlying cone. Likewise, the robust constraint can remain feasible with high probability without being over-conservative in the choice of the protection level.

4. **Discrete $0-1$ Optimization:** For $0-1$ discrete optimization problem with cost uncertainty, the robust counterpart of a polynomially solvable $0-1$ discrete optimization problem remains polynomially solvable and the robust counterpart of an NP-hard α-approximable $0-1$ discrete optimization problem, remains α-approximable (Chapter 5). Under an ellipsoidal uncertainty set, we show that the robust problem retains the complexity of the nominal problem when the data is uncorrelated and identically distributed. For uncorrelated, but not identically distributed data, we propose an approximation method that solves the robust problem within arbitrary accuracy. We also propose a Frank-Wolfe type algorithm for this case, which we prove converges to a locally optimal solution, and in computational experiments is remarkably effective (Chapter 6).

5. **Network Flows:** We propose an algorithm for solving the robust minimum cost flow problem in a polynomial number of nominal minimum cost flow problems in a modified network (Chapter 5).

7.1 Future research

Possible theoretical research in robust optimization include the followings:

- **Stochastic Models with Recourse:** In some stochastic optimization problems, there are exogenous parameters that influence subsequent stages of decision making, but whose value is uncertain, and only become known after the initial decision has been made. Unfortunately, the robust framework in this thesis does not lead to natural representations of such stochastic models. Due to the practical importance, it is therefore desirable to extend the attractive features of robust optimization to modeling and solving stochastic models with recourse.

- **Probability Bounds for General Distributions:** In Chapter 4, we derive the probability bounds on feasibility of the robust solution based on the assumption of normal distributions. It is desirable to derive probability bounds for more general distributions and establish the tightness of the these bounds.

- **Correlations in Uncertain Discrete $0-1$ Problems:** In the robust $0-1$ discrete models of Chapter 5 and 6, the nature of cost uncertainty is uncorrelated. It is worth having robust models that address correlated cost perturbation while keeping the model tractable.

- **Strong Formulations of Robust Discrete $0-1$ problems:** Recently, Atamtürk [2] provided stronger formulations of the robust discrete $0-1$ framework of Chapter 5 and showed empirically to significantly improve computation time. Likewise, it is beneficial to identify and study computationally the effect of strong formulations for ellipsoidal uncertainty set in the robust framework of Chapter 6.

Apart from the above theoretical research possibilities, it is worthwhile to apply these methodologies in applications and understand the merits or weaknesses of such approaches. Ultimately, we feel that the criterion for justifying robust optimization will be computation studies in real problems.

Bibliography

[1] Ahuja, R., Magnanti, T. and Orlin, J. (1993): Network flows: theory, algorithms, and applications, *Prentice Hall*, New York.

[2] Atamtürk, A. (2003): Strong formulations of robust mixed 0-1 programming. *Working Paper*, Department of Industrial Engineering and Operations Research, University of California, Berkeley.

[3] Averbakh, I. (2001): On the complexity of a class of combinatorial optimization problems with uncertainty, *Math. Prog. 90*, 263-272.

[4] Ben-Tal, A. and Nemirovski, A. (1998): Robust convex optimization, *Math. Oper. Res.*, 23, 769-805.

[5] Ben-Tal, A. and Nemirovski, A. (1998): On the quality of SDP approximations of uncertain SDP programs, *Research Report #4/98*, Optimization Laboratory, Faculty of Industrial Engineering and Management, Technion - Israel Institute of Technology, Israel.

[6] Ben-Tal, A. and Nemirovski, A. (1999): Robust solutions to uncertain programs, *Oper. Res. Letters*, 25, 1-13.

[7] Ben-Tal, A. and Nemirovski, A. (2000): Robust solutions of Linear Programming problems contaminated with uncertain data, *Math. Program.*, 88, 411-424.

[8] Ben-Tal, A. and Nemirovski, A. (2001): Lectures on modern convex optimization: Analysis, algorithms, and engineering applications, *MPR-SIAM Series on Optimization*, SIAM, Philadelphia.

[9] Bertsimas, D. and Popescu, I. (2003): Optimal Inequalities in Probability Theory: A Convex Optimization Approach, *Working Paper*, INSEAD TM62.

[10] Dijkstra, E. (1959): A note on two problems in connection with graphs, *Numerische Mathematik*, 269-271.

[11] El-Ghaoui, L. and Lebret, H. (1997): Robust solutions to least-square problems to uncertain data matrices, *SIAM J. Matrix Anal. Appl.*, 18, 1035-1064.

[12] El-Ghaoui, L., Oustry, F. and Lebret, H. (1998): Robust solutions to uncertain semidefinite programs, *SIAM J. Optim.*, 9, 33-52.

[13] Eppstein, D. (1995): Geometric lower bounds for parametric matroid optimization, *Proc. 27th ACM Sympos. Theory of Comput.*, 1995, 662-671.

[14] Fern andez-Baca, D., Slutzki, G. and Eppstein, D. (1996): Using sparsification for parametric minimum spanning tree problems, *Proc. 5th Scand. Worksh. Algorithm Theory. Springer-Verlag, Lecture Notes in Computer Science* 1097, July 1996.

[15] Freund, R. M. (1985): Postoptimal analysis of a linear program under simultaneous changes in matrix coefficients, *Math. Prog. Study*, 24, 1-13.

[16] Gusfield, D. (1980): Sensitivity analysis for combinatorial optimization., *Technical Report* UCB/ERL M80/22, University of Califonial, Berkeley.

[17] Hassin, R. and Tamir, A. (1989) Maximizing classes of two-parameter objectives over matriods, *Math. Oper. Res.* 14, No. 2, 362-375.

[18] Ishii, H., Shiode, S., Nishida, T. and Namasuya, Y. (1981) : Stochastic Spanning Tree Problem, *Discrete Appl. Math.* 3, 263-271.

[19] Karp, R. and Orlin, J. (1981): Parametric shortest path algorithms with an applications to cyclic stuffing, *Discrete Appl. Math.* 3, 37-45.

[20] Kouvelis, P. and Yu, G. (1997): Robust discrete optimization and its applications, *Kluwer Academic Publishers*, Norwell, MA.

[21] Lax P., (1997) Linear Algebra, *John Wiley& Sons*, New York.

[22] Mulvey, J., Vanderbei, R. and Zenios, S. (1995): Robust optimization of large-scale systems, *Oper. Res.* 43, 264-281.

[23] Renegar, J. (2001): A mathematical view of interior-point methods in convex optimization, *MPR-SIAM Series on Optimization*, SIAM, Philadelphia.

[24] Robbins, H. (1955): A Remark of Stirling's Formula, *Amer. Math. Monthly*, 62, 26-29.

[25] Schultz, R., Stougie L. and van der Vlerk, M. (1998): Solving stochastic programs with complete integer recourse by enumeration: A framework using Grobner basis reductions, *Math. Prog.* 83, 1998, 229-252.

[26] Soyster, A.L. (1973): Convex programming with set-inclusive constraints and applications to inexact linear programming, *Oper. Res.*, 21, 1154-1157.

[27] Srinivasan, V. and Thompson, G. (1972): An operator theory of parametric programming for the transportation problem, *Naval Research Logistics Quarterly* 19, 205-252.