Spring 2015

Mathematical programming techniques for solving stochastic optimization problems with certainty equivalent measures of risk

Alexander Vinel
University of Iowa

MATHEMATICAL PROGRAMMING TECHNIQUES FOR SOLVING STOCHASTIC OPTIMIZATION PROBLEMS WITH CERTAINTY EQUIVALENT MEASURES OF RISK

by

Alexander Vinel

A thesis submitted in partial fulfillment of the
requirements for the Doctor of Philosophy
degree in Industrial Engineering
in the Graduate College of
The University of Iowa

May 2015

Thesis Supervisor: Associate Professor Pavlo Krokhmal

Graduate College
The University of Iowa
Iowa City, Iowa

CERTIFICATE OF APPROVAL

PH.D. THESIS

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

Alexander Vinel

has been approved by the Examining Committee for the thesis requirement for the Doctor of Philosophy degree in Industrial Engineering at the May 2015 graduation.

Thesis Committee: ___________________________
Pavlo Krokhmal, Thesis Supervisor

____
Andrew Kusiak

____
Yong Chen

____
Samuel Burer

____
Barrett Thomas

To my dear wife

ACKNOWLEDGEMENTS

I would like to express my special appreciation and thanks to my advisor Professor Pavlo Krokhmal. I would also like to thank my dissertation committee members Prof. Kusiak, Prof. Chen, Prof. Burer and Prof. Thomas, who provided invaluable feedback and advice that significantly improved this manuscript, as well as made the process of writing and defending much more enjoyable.

It is also my pleasure to thank Maciej Rysz and Yana Morenko for fruitful collaborations. My great thanks go to my fellow students Dmitri Chernikov, Nathaniel Richmond and Bo Sun for helping to create and maintain a positive learning and research atmosphere. I especially appreciate the help of Mohammad Mirghorbani who was one of the first to welcome me in the new country after I moved to the United States.

This research has been supported in part by the Air Force Office of Scientific Research and National Science Foundation grants. In addition, support by the AFRL Mathematical Modeling and Optimization Institute is gratefully acknowledged with special thanks to Dr. Eduardo Pasiliao.

Lastly, I would like thank department administrative staff: Andrea Flaherty, Mandy Kurtenbach, Karla Stout and Jennifer Rumping who were extremely helpful and were always there to solve any problems.

ABSTRACT

The problem of risk-averse decision making under uncertainties is studied from both modeling and computational perspectives. First, we consider a framework for constructing coherent and convex measures of risk which is inspired by infimal convolution operator, and prove that the proposed approach constitutes a new general representation of these classes. We then discuss how this scheme may be effectively employed to obtain a class of certainty equivalent measures of risk that can directly incorporate decision maker's preferences as expressed by utility functions. This approach is consequently utilized to introduce a new family of measures, the log-exponential convex measures of risk. Conducted numerical experiments show that this family can be a useful tool when modeling risk-averse decision preferences under heavy-tailed distributions of uncertainties. Next, numerical methods for solving the arising optimization problems are developed. A special attention is devoted to the class p-order cone programming problems and mixed-integer models. Solution approaches proposed include approximation schemes for p-order cone and more general nonlinear programming problems, lifted conic and nonlinear valid inequalities, mixed-integer rounding conic cuts and new linear disjunctive cuts.

PUBLIC ABSTRACT

Recently, stochastic programming and decision making under conditions of uncertainty have been receiving an increasing amount of attention in the literature. With the ongoing advances in the amount of computational power, it is now possible to successfully solve optimization problems in the presence of random parameters for many practical applications. In the present work two challenges associated with the introduction of randomness into optimization are discussed: how these uncertainties can be modeled, and then how the resulting problems can be solved numerically. Efforts in designing appropriate ”measures of risk” are outlined in the first chapter, with special consideration given to the phenomena of heavy-tailed distributions of losses and catastrophic risk. This leads to the introduction of a new general modeling framework that have not been considered in the literature before. Next, the mathematical programming consequences of the proposed modeling approaches are considered. This work includes design of novel solution procedures for both convex and mixed-integer programming problems of a special kind.

TABLE OF CONTENTS

LIST OF TABLES

Table

LIST OF FIGURES

Figure

CHAPTER 1
INTRODUCTION

1.1 Decision Making Under Uncertainties

A decision making problem under uncertainties can be stated as the problem of selecting a decision $\boldsymbol{x} \in C \subset \mathbb{R}^n$, given that the cost X of this decision depends not only on $\boldsymbol{x}$, but also on a random event $\omega \in \Omega$: $X = X(\boldsymbol{x}, \omega)$. A principal modeling challenge that one faces in this setting is to select an appropriate ordering of random outcomes X, or, in other words, define a way to choose one uncertain outcome, $X_1 = X(\boldsymbol{x}_1, \omega)$, over another, $X_2 = X(\boldsymbol{x}_2, \omega)$. A fundamental contribution in this context is represented by the expected utility theory of von Neumann and Morgenstern (1944), which argues that if the preferences of a decision maker are *rational*, i.e., they satisfy a specific system of properties (axioms), then there exists a utility function $u : \mathbb{R} \mapsto \mathbb{R}$, such that a decision under uncertainty is optimal if it maximizes the expected utility of the payoff. Equivalently, the random elements representing payoffs under uncertainty can be ordered based on the corresponding values of expected utility of these payoffs. Closely connected to the expected utility theory is the subject of stochastic orderings (see, for example, Levy (1998)), and particularly stochastic dominance relations, which have found applications in economics, decision theory, game theory, and so on.

An alternative approach to introducing preference relations over random outcomes $X(\boldsymbol{x}, \omega)$, which has traditionally been employed in optimization and operations research literature, and which is followed in the present work, is to introduce a function $\rho\colon \mathcal{X} \mapsto \overline{\mathbb{R}}$, where $\mathcal{X}$ is an appropriately defined space containing X, such that X_1 is preferred to X_2 whenever $\rho(X_1) < \rho(X_2)$. The decision making problem in the presence of uncertainties can then be expressed as a mathematical program

$$\min\{\rho(X) : X = X(\boldsymbol{x}, \omega) \in \mathcal{X},\ \boldsymbol{x} \in C\}, \tag{1.1}$$

where function ρ is usually referred to as a *risk measure*. In stochastic programming literature, the objective of a minimization problem like (1.1) has traditionally been chosen in the form of the expected cost, $\rho(X) = \mathsf{E}X$ (Prékopa (1995); Birge and Louveaux (1997)), which is commonly regarded as a representation of risk-neutral preferences. In the finance domain, a pioneering work of Markowitz (1952) has introduced a *risk-reward* paradigm for decision making under uncertainty, and variance was proposed as a measure of risk, $\rho(X) = \sigma^2(X)$. Since then, the problem of devising risk criteria suitable for quantification of specific risk-averse preferences has received significant attention (see a survey in Krokhmal et al. (2011)). It was noticed, however, that the "ad-hoc" approach of constructing ρ may yield risk functionals that, while serving well in a specific application, are faulty in the general methodological sense. Artzner et al. (1999) suggested an axiomatic approach, similar to that of von Neumann and Morgenstern (1944), to defining a well-behaved risk measure ρ in (1.1), and introduced the concept of *coherent measures of risk*. Subsequently, a range of variations and extensions of the axiomatic framework for designing risk functionals have been proposed in the literature, such as *convex* and *spectral* measures of risk, *deviation measures*, and so on (see Föllmer and Schied (2002); Rockafellar et al. (2006); Rockafellar and Uryasev (2013)). Since many classes of axiomatically defined risk measures represent risk preferences that are not fully compatible with the rational risk-averse preferences of utility theory, of additional interest are risk measures that possess such a compatibility in a certain sense.

1.2 Aim of the Study

In this context, the goal of the present study is to explore new approaches to risk-averse stochastic programming. We aim at pursuing two objectives: construct a new methodology for generating measures of risk and then develop mathematical programming techniques in order to ensure that the arising optimization problems can be efficiently solved.

In this study we propose a new representation for the classes of coherent and convex measures of risk, which builds upon a previous work of Krokhmal (2007). This representation is then used to introduce a class of coherent or convex measures of risk that can directly incorporate *rational* risk preferences as prescribed by the corresponding utility function, through the concept of certainty equivalent. This class of certainty equivalent measures of risk contains some of the existing risk measures, such as the popular Conditional Value-at-Risk (Rockafellar and Uryasev (2000, 2002)) as special cases. As an application of the general approach, we introduce a two-parameter family of log-exponential convex risk measures, which quantify risk by emphasizing extreme losses in the tail of the loss distribution. Two case studies illustrate the practical merits of the log-exponential risk measures; in particular, it is shown that these nonlinear measures of risk can be preferable to more traditional measures, such as Conditional Value-at-Risk, if the loss distribution is heavy-tailed and contains catastrophic losses.

We show that a decision making problem based on the proposed preference relation can be formulated as a convex nonlinear programming problem. Our next goal is then to explore computational approaches to effectively solve such problems in the case of both convex and mixed-integer models. A notable example of the optimization problems under consideration is constituted by the class of p-order cone programming problems. In our view, it is of particular interest to study solution methods targeted specifically at this class. First, such problems result from the decision making models based on evaluation of risk in terms of higher moments of the loss distributions, which have been shown in the literature to lead to promising practical results (see among others Krokhmal (2007); Malevergne and Sornette (2005)). Second, while such problems generalize the well-studied class of second-order cone programming, they have received significantly less attention in the literature. Finally, due to the conic property, such problems can be easier to solve compared to the general counterparts, thus they can be used as a test-ground for evaluating applicability of the proposed methods in more general settings.

The rest of the manuscript is organized as follows. In Chapter 2 we present our work on devising a new risk-averse stochastic programming framework: a new representation for convex and coherent risk measures, the class of certainty equivalent risk measures, and the family of log-exponential convex risk measures. The chapter concludes with a discussion of the conducted case studies. In Chapters 3 and 4 we discuss solution methods for p-order cone programming problems. Specifically, in Chapter 3 we explore approximation schemes for p-order cone programming and in Chapter 4 we propose two families of valid inequalities for the mixed-integer model. In the next two chapters we tackle the general stochastic programming problems with certainty equivalent constraints. Namely, we show how two main computational challenges associated with our problem formulation can be addressed. In Chapter 5 we develop a targeted scenario decomposition method in order to deal with stochastic constraints. Finally, in Chapter 6 we consider nonlinear constraints present in mixed-integer nonlinear programming problems that arise from risk-averse stochastic programming with certainty equivalent measures of risk. We have previously published some of the results presented in the following chapters in Vinel and Krokhmal (2014a,b, 2015); Rysz et al. (2014)

CHAPTER 2
CERTAINTY EQUIVALENT RISK MEASURES

2.1 Risk Measures Based on Infimal Convolution

2.1.1 Coherent and Convex Measures of Risk

Consider a random outcome $X \in \mathcal{X}$ defined on an appropriate probability space $(\Omega, \mathcal{F}, \mathsf{P})$, where $\mathcal{X}$ is a linear space of $\mathcal{F}$-measurable functions $X : \Omega \mapsto \mathbb{R}$. A function $\rho\colon \mathcal{X} \mapsto \overline{\mathbb{R}} = \mathbb{R} \cup \{+\infty\}$ is said to be a *convex measure of risk* if it satisfies the following axioms:

(A0) *lower semicontinuity (l.s.c.)*;

(A1) *monotonicity*: $\rho(X) \leq \rho(Y)$ for all $X \leq Y$;

(A2) *convexity*: $\rho\big(\lambda X + (1-\lambda)Y\big) \leq \lambda\rho(X) + (1-\lambda)\rho(Y), \quad \lambda \in [0,1]$;

(A3) *translation invariance*: $\rho(X + a) = \rho(X) + a, \quad a \in \mathbb{R}$.

Similarly, a function $\rho : \mathcal{X} \to \overline{\mathbb{R}}$ is said to be a *coherent measure of risk* if it satisfies (A0) – (A3), and, additionally,

(A4) *positive homogeneity*: $\rho(\lambda X) = \lambda\rho(X), \quad \lambda > 0$.

Remark 1. We assume that linear space $\mathcal{X}$ is endowed with necessary properties in such a way that risk measures that we consider are well defined (do not assume the value of $-\infty$). Specifically, it can be assumed that $\mathsf{E}|X| < +\infty$. Similarly, $\mathcal{X}$ is equipped with appropriate topology. Unless stated otherwise we would assume the topology induced by convergence in probability. We also assume throughout this chapter that all considered functions are proper. Recall that a function $f : \mathcal{X} \mapsto \overline{\mathbb{R}}$ is proper if $f(X) > -\infty$ for all $X \in \mathcal{X}$, and

$\operatorname{dom} f = \{X \in \mathcal{X} \mid f(X) < +\infty\} \neq \emptyset$.

Remark 2. In this work we adopt the traditional viewpoint of engineering literature that a random element X represents a cost or a loss, in the sense that smaller realizations of X are preferred. In economics literature it is customary to consider X as wealth or payoff variable, whose larger realizations are desirable. In most cases, these two approaches can be reconciled by inverting the sign of X, which may require some modifications to the properties discussed above. For example, the translation invariance axiom (A3) will have the form $\rho(X + a) = \rho(X) - a$ in the case when X is a payoff quantity.

Remark 3. Without loss of generality we also assume that a convex measure of risk satisfies normalization property: $\rho(0) = 0$ (observe that coherent measures necessarily satisfy this property). First, such a normalization requirement is natural from methodological and practical viewpoints, since there is usually no risk associated with zero costs or losses. Second, due to translation invariance any convex ρ can be normalized by setting $\tilde{\rho}(X) = \rho(X) - \rho(0)$.

Remark 4. It is worth noting that normalized convex measures of risk satisfy the so-called *subhomogeneity* property:

(A4′) *subhomoheneity:* $\rho(\lambda X) \leq \lambda\rho(X)$ for $\lambda \in (0, 1)$ and $\rho(\lambda X) \geq \lambda\rho(X)$ for $\lambda > 1$.

Indeed, in order to see that the first inequality in (A4′) holds, observe that $\lambda\rho(X) = \lambda\rho(X) + (1 - \lambda)\rho(0) \geq \rho(\lambda X + (1 - \lambda 0)) = \rho(\lambda X)$ for $\lambda \in (0, 1)$. Similarly, if $\lambda > 1$, then $\frac{1}{\lambda}\rho(\lambda X) = \frac{1}{\lambda}\rho(\lambda X) + \big(1 - \frac{1}{\lambda}\big)\rho(0) \geq \rho(X)$.

Artzner et al. (1999) and Delbaen (2002) have proposed a general representation for the class of coherent measures by showing that a mapping $\rho : \mathcal{X} \mapsto \mathbb{R}$ is a coherent risk measure if and only if $\rho(X) = \sup_{Q\in\mathcal{Q}} \mathsf{E}_Q X$, where $\mathcal{Q}$ is a closed convex subset of P-absolutely continuous probability measures. Föllmer and Schied (2002) have generalized this result to convex measures of risk. Since then, other representations have been proposed, (see Kusuoka (2001, 2012); Frittelli and Rosazza Gianin (2005); Dana (2005); Acerbi (2002)). For example, Acerbi (2002) has suggested a spectral representation: $\rho(X) = \int_0^1 \mathrm{VaR}_\lambda(X)\psi(\lambda)\mathrm{d}\lambda$, where $\psi \in \mathcal{L}^1([0,1])$. While many of these results led to important methodological conclusions, most of them did not provide a readily available, practical way to construct new risk measures in accordance with specified risk preferences. Below we discuss a representation that can be better suited for this purpose.

2.1.2 An Infimal Convolution Representation for Coherent and Convex Measures of Risk

An approach to constructing coherent measures of risk that was based on the operation of infimal convolution was proposed in Krokhmal (2007). Given a function $\phi\colon \mathcal{X} \mapsto \overline{\mathbb{R}}$, consider a risk measure ρ, which we will call a *convolution-based measure of risk*, in the form

$$\rho(X) = \inf_{\eta}\ \eta + \phi(X - \eta). \tag{2.1}$$

Then, the following claim has been shown to hold.

Proposition 2.1 (Krokhmal, 2007, Theorem 1). *Suppose that function ϕ satisfies axioms (A0)–(A2) and (A4), and, additionally, is such that $\phi(\eta) > \eta$ for all constant $\eta \neq 0$. Then* *the infimal convolution in* (2.1) *is a proper coherent risk measure. Moreover, the infimum in* (2.1) *is attained for all* X*, and can be replaced with a minimization operator.*

In this section we show that this approach can be substantially generalized which leads us to formulate Theorem 2.5 below. Before moving to this general result we establish a few subsidiary lemmas. First, we demonstrate that expression (2.1) is a *representation*, i.e., any coherent measure of risk can be expressed in the form of (2.1).

Lemma 2.2. *Let* ρ *be a coherent risk measure. Then, there exists a proper function* $\phi : \mathcal{X} \mapsto \overline{\mathbb{R}}$ *that satisfies axioms (A0)–(A2) and (A4),* $\phi(\eta) > \eta$ *for all constant* $\eta \neq 0$*, and is such that* $\rho(X) = \min_\eta \eta + \phi(X - \eta)$.

Proof: For a given proper and coherent ρ consider $\phi_\rho(X) = 2[\rho(X)]_+$, where $[X]_+ = \max\{X, 0\}$, and observe that ϕ_ρ is proper and satisfies (A0)–(A2) and (A4) if ρ is coherent, and, moreover, $\phi_\rho(\eta) = 2[\eta]_+ > \eta$ for all real $\eta \neq 0$. Finally, $\min_\eta \eta + \phi_\rho(X - \eta) = \min_\eta \eta + 2[\rho(X - \eta)]_+ = \min_\eta \eta + 2[\rho(X) - \eta]_+ = \rho(X)$, i.e., any coherent ρ can be represented in the form of (2.1). □

Remark 5. It is easy to see from the proof of Lemma 2.2 that the function ϕ in representation (2.1) is not determined uniquely for any given coherent measure ρ. Indeed, one can choose (among possibly others) $\phi(X) = \alpha[\rho(X)]_+$ for any $\alpha > 1$.

Next, we show that the infimal convolution representation (2.1) can be generalized to convex measures of risk. Technically, proof of Proposition 2.1 in Krokhmal (2007) relies heavily on the positive homogeneity property (A3) of coherent risk measures, but as we demonstrate below, it can be amended in order to circumvent this issue. Recall that, given a proper, l.s.c., convex function f on $\mathbb{R}^n$ and $\boldsymbol{x} \in \operatorname{dom} f$, its *recession function* $(f0^+)(\boldsymbol{y})$ can be defined as

$$(f0^+)(\boldsymbol{y}) = \lim_{\tau \to \infty} \frac{f(\boldsymbol{x} + \tau \boldsymbol{y}) - f(\boldsymbol{x})}{\tau}.$$

Note that in Rockafellar (1997), Theorem 8.5 it is shown that the expression above does not depend on $\boldsymbol{x} \in \operatorname{dom} f$, i.e., $(f0^+)(\boldsymbol{y})$ is well-defined. The result established below mirrors that of Proposition 2.1 in the case of convex measures.

Lemma 2.3. *Suppose that a proper function ϕ satisfies axioms (A0)–(A2), and, additionally, is such that $\phi(\eta) > \eta$ for all constant $\eta \neq 0$ and $\phi(0) = 0$. Then the infimal convolution $\rho(X) = \inf_\eta \ \eta + \phi(X - \eta)$ is a proper convex risk measure. Moreover, the infimum is attained for all X, and can be replaced with $\min_\eta$.*

Proof: For any fixed $X \in \mathcal{X}$ consider function $\phi_X(\eta) = \eta + \phi(X - \eta)$. Clearly, since ϕ is proper, l.s.c. and convex, ϕ_X is l.s.c., convex in η and $\phi_X > -\infty$ for all η. Next we will show that the infimum in the definition of ρ is attained for any X. First, suppose that $\operatorname{dom} \phi_X = \emptyset$, hence $\rho(X) = +\infty$, and the infimum in the definition is attained for any $\eta \in \mathbb{R}$. Now, assume that there exists $\tilde{\eta} \in \operatorname{dom} \phi_X$, and consequently both $\phi(X - \tilde{\eta}) < +\infty$ and $\rho(X) < +\infty$. Recall that a proper, l.s.c. function ϕ_X attains its infimum if it has no directions of recession (see, Theorem 27.2 in Rockafellar (1997)), or in other words, if $\phi_X 0^+(\xi) > 0$ for all $\xi \neq 0$. Observe that

$$\begin{aligned}(\phi_X 0^+)(\xi) &= \lim_{\tau \to \infty} \frac{\tilde{\eta} + \tau\xi + \phi(X - \tilde{\eta} - \tau\xi) - \tilde{\eta} - \phi(X - \tilde{\eta})}{\tau} \\ &= \xi + \lim_{\tau \to \infty} \frac{\phi(X - \tilde{\eta} - \tau\xi)}{\tau} \geq \xi + \lim_{\tau \to \infty} \phi\Big(\frac{X - \tilde{\eta}}{\tau} - \xi\Big),\end{aligned}$$

where the last inequality follows from Remark 4 for sufficiently large τ. Since ϕ is l.s.c. and $\phi(\xi) > \xi$ for all $\xi \neq 0$, we can conclude that $\lim_{\tau \to \infty} \phi\Big(\frac{X - \tilde{\eta}}{\tau} - \xi\Big) \geq \phi(-\xi) > -\xi$, whereby $(\phi_X 0^+)(\xi) > 0$ for all $\xi \neq 0$, which guarantees that the infimum in the definition is attained, and $\rho(X) = \min_\eta \ \eta + \phi(X - \eta)$. Next, we will verify that axiom (A0) holds. As shown above, for any $X \in \mathcal{X}$ there exists η_X such that $\rho(X) = \eta_X + \phi(X - \eta_X)$. Consequently,

$$\begin{aligned}\liminf_{Y\to X} \rho(Y) &= \liminf_{Y\to X} \Big(\eta_Y + \phi(Y - \eta_Y)\Big) \geq \liminf_{Y\to X} \Big(\eta_X + \phi(Y - \eta_X)\Big) \\ &= \eta_X + \liminf_{Y\to X} \phi(Y - \eta_X) \geq \eta_X + \phi(X - \eta_X) = \rho(X),\end{aligned}$$

where the last inequality holds due to lower semicontinuity of ϕ. Whence, by definition, ρ is l.s.c. Verification of properties (A1)–(A3) is analogous to that presented in Krokhmal (2007), Theorem 1. □

Lemma 2.4. *Let ρ be a convex risk measure and let $\rho(0) = 0$. Then there exists a proper function $\phi(X)$ that satisfies axioms of monotonicity and convexity, is lower semicontinuous, $\phi(\eta) > \eta$ for all $\eta \neq 0$, and such that $\rho(X) = \min_\eta \eta + \phi(X - \eta)$.*

Proof: Analogous to Lemma 2.2 we can take $\phi_\rho(X) = 2[\rho(X)]_+$. □

Combining the above results, we obtain a general conclusion.

Theorem 2.5. *A proper, l.s.c. function $\rho : \mathcal{X} \mapsto \overline{\mathbb{R}}$ is a convex (respectively, coherent) measure of risk if and only if there exists a proper, l.s.c. function $\phi : \mathcal{X} \mapsto \overline{\mathbb{R}}$, which satisfies the axioms of monotonicity and convexity (and, respectively, positive homogeneity), $\phi(\eta) > \eta$ for all $\eta \neq 0$, $\phi(0) = 0$, and such that $\rho(X) = \min_\eta \eta + \phi(X - \eta)$.*

The importance of infimal convolution representation (2.1) for convex/coherent risk measures lies in the fact that it is amenable for use in stochastic programming problems with risk constraints or risk objectives (note that the problem does not necessarily have to be convex).

Lemma 2.6. *Let ρ be a coherent measure of risk, and for some $F, H : \mathcal{X} \mapsto \mathbb{R}$ and $C(\mathcal{X}) \subset \mathcal{X}$ consider the following risk-constrained stochastic programming problem:*

$$\min\{F(X) : \rho(X) \leq H(X),\ X \in C(\mathcal{X})\}. \tag{2.2}$$

Then, for a given convolution representation (2.1) *of ρ, problem* (2.2) *is equivalent to a problem of the form*

$$\min\{F(X) : \eta + \phi(X - \eta) \leq H(X),\ X \in C(\mathcal{X}),\ \eta \in \mathbb{R}\}, \tag{2.3}$$

in the sense that if (2.2) *is feasible, they achieve minima at the same values of the decision variable X and their optimal objective values coincide. Moreover, if risk constraint is binding at optimality in* (2.2)*, then (X^*, η^*) delivers a minimum to* (2.3) *if and only if X^* is an optimal solution of* (2.2) *and $\eta^* \in \arg\min\{\eta + \phi(X^* - \eta)\}$.*

Proof: Analogous to that in Krokhmal (2007), Theorem 3. □

Additionally, representation (2.1) conveys the idea that *a risk measure represents an optimal value or optimal solution of a stochastic programming problem of special form.*

2.1.3 Convolution Representation and Certainty Equivalents

The infimal convolution representation (2.1) allows for construction of convex or coherent measures of risk that directly employ risk preferences of a decision maker through a connection to the expected utility theory of von Neumann and Morgenstern (1944). Assuming without loss of generality that the loss/cost elements $X \in \mathcal{X}$ are such that $-X$ represents *wealth* or *reward*, consider a non-decreasing, convex *deutility* function $v : \mathbb{R} \mapsto \mathbb{R}$ that quantifies dissatisfaction of a risk-averse rational decision maker with a loss or cost. Obviously, this is equivalent to having a non-decreasing concave utility function $u(t) = -v(-t)$. By the inverse of v we will understand function $v^{-1}(a) = \sup\{t \in \mathbb{R} : v(t) = a\}$.

Remark 6. Note that if a non-decreasing, convex $v(t) \not\equiv \text{const}$ then, according to the definition above, the inverse is finite, and moreover, if there exists t, such that $v(t) = a < +\infty$, then $v^{-1}(a) = \max\{t \in \mathbb{R} \,|\, v(t) = a\}$. Additionally, let $v^{-1}(+\infty) = +\infty$.

Then, for any given $\alpha \in (0, 1)$, consider function ϕ in the form

$$\phi(X) = \frac{1}{1-\alpha} v^{-1} \mathsf{E} v(X), \tag{2.4}$$

where we use an operator-like notation for v^{-1}, i.e., $v^{-1}\mathsf{E}v(X) = v^{-1}(\mathsf{E}v(X))$.

Expression $\text{CE}(X) = v^{-1}\mathsf{E}v(X)$ represents the *certainty equivalent* of an uncertain loss X, a deterministic loss/cost such that a rational decision maker would be indifferent between accepting $\text{CE}(X)$ or an uncertain X; it is also known as *quasi-arithmetic mean*, *Kolmogorov mean*, or *Kolmogorov-Nagumo mean* (see, among others, Bullen et al. (1988); Hardy et al. (1952)). Certainty equivalents play an important role in the decision making literature (see, for example, Wilson (1979); McCord and Neufville (1986)); in the context of modern risk theory, certainty equivalents were considered in the work of Ben-Tal and Teboulle (2007).

In order for function ϕ as defined by (2.4) to comply with the conditions of Theorem 2.5, the deutility function should be such that $\phi(\eta) = \frac{1}{1-\alpha} v^{-1} v(\eta) > \eta$ for $\eta \neq 0$. This necessarily implies that $v(\eta) = v(0)$ for all $\eta \leq 0$, provided that v is proper, non-decreasing and convex. Indeed, if $v(\eta^*) < v(0)$ for some $\eta^* < 0$, then according to the above remarks $v^{-1}v(\eta^*) = \max\{\eta : v(\eta) = v(\eta^*)\} = \eta^{**}$, where η^{**} is such that $\eta^* \leq \eta^{**} < 0$ and $v^{-1}v(\eta^{**}) = \eta^{**}$, and so $\phi(\eta^{**}) = (1-\alpha)^{-1}\eta^{**} < \eta^{**}$.

Additionally, without loss of generality it can be postulated that $v(0) = 0$, i.e., zero loss means zero dissatisfaction. Indeed, $\tilde{v}^{-1}\mathsf{E}\tilde{v}(X) = v^{-1}\mathsf{E}v(X)$ for $\tilde{v}(t) = v(t) - v(0)$, i.e., such a transformation of the deutility function does not change the value of the certainty equivalent. Similarly, it is assumed that $v(t) > 0$ for all $t > 0$. Unless $v(t) \equiv 0$ there exists a $\tilde{t}$, such that $v(t) = 0$ for all $t \leq \tilde{t}$ and $v(t) > 0$ for all $t > \tilde{t}$. Now, consider $\tilde{v}(t) = v(t - \tilde{t})$, in which case $\tilde{v}^{-1}(a) = v^{-1}(a) + \tilde{t}$, and $\tilde{v}^{-1}\mathsf{E}\tilde{v}(X) = v^{-1}\mathsf{E}v(X - \tilde{t}) + \tilde{t}$. Consequently, if we denote by ρ and $\tilde{\rho}$ the corresponding risk measures, then due to translation invariance,

$$\tilde{\rho}(X) = \min \eta + \frac{1}{1-\alpha}\tilde{v}^{-1}\mathsf{E}\tilde{v}(X - \eta) = \min \eta + \frac{1}{1-\alpha}v^{-1}\mathsf{E}v(X - \eta - \tilde{t}) + \tilde{t} = \rho(X) + C,$$

where C is a constant i.e., both ρ and $\tilde{\rho}$ correspond to the same decision preferences. In other words, unless $v(t) \equiv 0$, we can assume that $v(t) > 0$ for all $t > 0$. It also represents a practical consideration that positive losses entail positive deutility/dissatisfaction.

To sum up, we consider non-decreasing, convex deutility function $v : \mathbb{R} \mapsto \mathbb{R}$ such that

$$v(t) = v\big([t]_+\big) = \begin{cases} v(t) > 0, & t > 0, \\ 0, & t \leq 0. \end{cases}$$

We will refer to such a function as a *one-sided deutility*. The corresponding expression for risk measure (2.1) becomes

$$\begin{aligned}\rho(X) &= \min_{\eta}\ \eta + \frac{1}{1-\alpha} v^{-1}\mathsf{E}v(X-\eta) \\ &= \min_{\eta}\ \eta + \frac{1}{1-\alpha} v^{-1}\mathsf{E}v\big([X-\eta]_+\big).\end{aligned} \tag{2.5}$$

Next we analyze the conditions under which formulation (2.5) yields a coherent or convex measure of risk. We will assume that the space $\mathcal{X}$ is such that certainty equivalent above is well-defined, particularly, integrability condition is satisfied.

Proposition 2.7. *If v is a one-sided deutility function, then $\phi(X) = \frac{1}{1-\alpha} v^{-1}\mathsf{E}v(X)$ is proper, l.s.c., satisfies the axiom of monotonicity and $\phi(\eta) > \eta$ for all $\eta \neq 0$.*

Proof: Clearly, such a ϕ is proper and l.s.c. The monotonicity property of ϕ defined by (2.4), $\phi(X) \leq \phi(Y)$ for all $X \leq Y$, follows from both v and v^{-1} being non-decreasing. Finally, note that

$$\begin{aligned}\phi(\eta) =& \frac{1}{1-\alpha} v^{-1} v(\eta) = \frac{1}{1-\alpha} v^{-1} v([\eta]_+) \\ =& \frac{1}{1-\alpha} \sup\big\{t : v(t) = v([\eta]_+)\big\} \geq \frac{1}{1-\alpha}[\eta]_+ > \eta\end{aligned}$$

for all $\eta \neq 0$. □

From Proposition 2.7 we can conclude that in order for the conditions of Theorem 2.5 to be satisfied we only need to guarantee convexity of the certainty equivalent (2.4) (note that axiom (A4) is satisfied if certainty equivalent itself is positive homogeneous). A sufficient condition of this type has been established in Ben-Tal and Teboulle (2007).

Proposition 2.8 (Ben-Tal and Teboulle, 2007). *If $v \in \mathcal{C}^3(\mathbb{R})$ is strictly convex and $\dfrac{v'}{v''}$ is convex, then certainty equivalent $v^{-1}\mathsf{E}v$ is also convex.*

The following observation shows how this result can be adapted to establish convexity of certainty equivalents in the case of one-sided deutility functions.

Corollary 2.9. *If* $v \in \mathcal{C}^3[0,\infty)$ *is strictly convex and* $\dfrac{v'}{v''}$ *is convex on* $[0,+\infty)$*, then certainty equivalent* $v_+^{-1}\mathsf{E}v_+$ *is convex, where* $v_+(t) = v([t]_+)$*.*

Proof: Indeed, note that $v_+^{-1}\mathsf{E}v_+(X) = v_+^{-1}\mathsf{E}v([X]_+) = v^{-1}\mathsf{E}v([X]_+)$, which is convex as a superposition of a convex (Proposition 2.8 for function v) and a non-decreasing convex functions. □

Remark 7. Conditions of Proposition 2.8 are only sufficient, i.e., it is possible for a certainty equivalent to be convex, when Proposition 2.8 does not apply (as shown in Corollary 2.9). Moreover, these conditions are rather restrictive, specifically, requirement $v \in \mathcal{C}^3(\mathbb{R})$. Thus, it is worth noting that if v is a one-sided deutility function and its certainty equivalent is convex, then ρ defined in (2.5) is a convex (or coherent) measure of risk, regardless of whether Corollary 2.9 holds. At the same time, this result can be useful, as shown in Proposition 2.16.

Observe that if function ϕ is taken in the form of (2.4), where v is a one-sided deutility, the structure of the resulting risk measure (2.5) allows for an intuitive interpretation, similar to that proposed by Ben-Tal and Teboulle (2007). Consider, for instance, a problem of resource allocation for a hazardous mission planning, where X represents the unknown in advance cost of resources necessary to cover losses and damages. Assume that it is possible to allocate amount η worth of resources in advance, whereby the remaining part of costs, $[X-\eta]_+$, will have to be covered after the actual realization of X is observed. To a decision maker with deutility v, the uncertain cost remainder $[X-\eta]_+$ is equivalent to the deterministic amount of certainty equivalent $v^{-1}\mathsf{E}v([X-\eta]_+)$. Since this portion of resource allocation is "unplanned" additional penalty is imposed. If this penalty is modeled using a multiplier $\frac{1}{1-\alpha}$, then the expected additional cost of the resource is $\frac{1}{1-\alpha}v^{-1}\mathsf{E}v([X-\eta]_+)$. Thus, the risk associated with the mission amounts to $\eta + (1-\alpha)^{-1}v^{-1}\mathsf{E}v([X-\eta]_+)$, and can be minimized over all possible values of η, leading to definition (2.5). Moreover, when applied to the general definition (2.1), this argument provides an intuition behind the condition $\phi(\eta) > \eta$ above. Indeed, the positive difference $\phi(\eta) - \eta$ can be seen as a penalty for an unplanned loss.

We also note that certainty equivalent representation (2.5) for coherent or convex measures of risk is related to the *optimized certainty equivalents* (OCEs) due to Ben-Tal and Teboulle (2007),

$$\text{OCE}(X) = \sup_{\eta} \; \eta + \mathsf{E}u(X-\eta). \tag{2.6}$$

While interpretations of formulas (2.5) and (2.6) are similar, and moreover, it can be shown that, under certain conditions on the utility function, $\rho(X) = -\text{OCE}(X)$ is a convex measure of risk, there are important differences between these representations. In (2.6), the quantity being maximized is technically not a certainty equivalent, while the authors have argued that specific conditions on utility function u allowed them to consider it as one. In addition, representation (2.6) entails addition of values with generally inconsistent units, e.g., dollars and utility. Finally, as shown above, representation (2.5) allows for constructing both coherent and convex measures of risk, while the OCE approach yields a coherent risk measure if and only if the utility function is piecewise linear.

Remark 8. It is straightforward to observe that by choosing the one-sided deutility function in (2.5) in the form $v(t) = [t]_+$ one obtains the well-known Conditional-Value-at-Risk (CVaR) measure (Rockafellar and Uryasev (2002)), while one-sided deutility $v(t) = [t]_+^p$ yields the Higher-Moment Coherent Risk (HMCR) measures (Krokhmal (2007)).

Remark 9. In general, risk measure ρ is called a *tail measure of risk* if it quantifies the risk of X through its right-hand tail, $[X - c]_+$, where the tail cutoff point c can be adjusted according to risk preferences Krokhmal et al. (2011). Observe that the above analysis implies that coherent or convex risk measures based on certainty equivalents (2.5) are necessarily tail measures of risk (see also Propositions 2.14 and 2.15 below).

Another key property of the risk measures that admit a certainty equivalent representation (2.5) is that they "naturally" preserve stochastic orderings induced on the space $\mathcal{X}$ of random outcomes by the utility function u or, equivalently, deutility v. Assuming again that $\mathcal{X}$ is endowed with necessary properties, e.g., integrability, of particular interest are the properties of *isotonicity with respect to second order stochastic dominance (SSD)* (see, e.g., De Giorgi (2005); Pflug (2006); Krokhmal (2007)),

(A1′) *SSD isotonicity:* $\rho(X) \leq \rho(Y)$ for all $X, Y \in \mathcal{X}$ such that $-X \succcurlyeq_{\text{SSD}} -Y$,

and, more generally, *isotonicity with respect to k-th order stochastic dominance (kSD)*,

(A1″) *kSD isotonicity:* $\rho(X) \leq \rho(Y)$ for all $X, Y \in \mathcal{X}$ such that $-X \succcurlyeq_{k\text{SD}} -Y$,

for a given $k \geq 1$.

Recall that random outcome X is said to dominate outcome Y with respect to second-order stochastic dominance (SSD), $\mathcal{X} \succcurlyeq_{\rm SSD} Y$, if

$$\int_{-\infty}^{t} F_X(\xi)\,\mathrm{d}\xi \leq \int_{-\infty}^{t} F_Y(\xi)\,\mathrm{d}\xi \quad \text{for all} \quad t \in \mathbb{R},$$

where $F_Z(t) = \mathsf{P}\{Z \leq t\}$ is the c.d.f. of a random element $Z \in \mathcal{X}$. Similarly, outcome X dominates outcome Y with respect to kth-order stochastic dominance (kSD), $\mathcal{X} \succcurlyeq_{k\rm SD} Y$, if

$$F_X^{(k)}(t) \leq F_Y^{(k)}(t), \quad \text{for all} \quad t \in \mathbb{R},$$

where $F_X^{(k)}(t) = \int_{-\infty}^{t} F_X^{(k-1)}(\xi)\mathrm{d}\xi, \;\; F_X^{(1)}(t) = \mathsf{P}\{X \leq t\}$. Stochastic dominance relations in general, and SSD in particular have occupied a prominent place in decision making literature (see, for a example, Levy (1998) for an extensive account), in particular due to a direct connection to the expected utility theory. Namely, it is well known Rothschild and Stiglitz (1970) that $\mathcal{X} \succcurlyeq_{\rm SSD} Y$ if and only if $\mathsf{E}u(X) \geq \mathsf{E}u(Y)$ for all non-decreasing and concave utility functions u, i.e., if and only if Y is never preferred over X by any rational risk-averse decision maker. In general, it can be shown that $\mathcal{X} \succcurlyeq_{k\rm SD} Y$ if and only if $\mathsf{E}u(X) \geq \mathsf{E}u(Y)$ for all $u \in U^{(k)}$, where $U^{(k)}$ is a specific class of real-valued utility functions; particularly, $U^{(1)}$ consists of all non-decreasing functions, $U^{(2)}$ contains all non-decreasing and concave functions, $U^{(3)}$ amounts to all non-decreasing, concave functions with convex derivative, and so on (see, for example, Fishburn (1977) and references therein). This characterization of kSD dominance relation naturally implies that the proposed certainty equivalent representation yields risk measures that are necessarily kSD-isotonic, given that the set of considered deutility functions is appropriately restricted.

Proposition 2.10. *If deutility function v is such that $-v(-t) \in U^{(k)}$, then risk measure ρ given by the certainty equivalent representation* (2.5) *is kSD-isotonic, i.e., satisfies* (A1$''$).

Proof: Follows immediately from the definition of kSD dominance, kSD isotonicity and the discussion above. □

Corollary 2.11. *If a real-valued function v is a one-sided deutility, then* (2.5) *defines a risk measure that is isotonic with respect to second order stochastic dominance.*

The importance of SSD- and, generally, kSD isotonicity of risk measures (2.5) in the sense of Proposition 2.10 attributes to the fact that coherent or convex measures of risk in general are not kSD-isotonic for $k \geq 2$ (see De Giorgi (2005) for an explicit example). Note, that it has been established in the literature that a law-invariat convex measure of risk defined on an atomeless probability space is SSD isotonic (see, Föllmer and Schied (2004), Corollary 4.59). At the same time, Proposition 2.10 and Corollary 2.11 directly show that the proposed certainty equivalent measures are natirally connected to kSD orderings. In this context, the certainty equivalent representation (2.5) ensures that risk-averse preferences expressed by the utility (equivalently, deutility) function are "transparently" inherited by the corresponding certainty equivalent-based risk measure; indeed, note that Proposition 2.10 does not require that the certainty equivalent in (2.5) is convex.

2.1.4 Optimality Conditions and Some Properties of Optimal $\boldsymbol{\eta}$

Consider the definition of CVaR, $\mathsf{CVaR}_\alpha(X) = \min_\eta \eta + \frac{1}{1-\alpha}\mathsf{E}[X-\eta]_+$. The lowest value of η that delivers minimum in this definition is know in the literature as Value-at-Risk (VaR) at confidence level α, and while VaR in general is not convex, it is widely used as a measure of risk in practice, especially in financial applications (Jorion (1997); Duffie and Pan (1997)). Thus, it is of interest to investigate some properties of $\eta^*(X) \in \arg\min \eta + \frac{1}{1-\alpha} v^{-1}\mathsf{E}v(X-\eta)$. First, we can formulate necessary and sufficient optimality conditions.

Proposition 2.12. *Suppose that v is a non-decreasing and convex function, certainty equivalent $v^{-1}\mathsf{E}v$ is convex and $\mathsf{E}\partial_\pm v(X-\eta^*)$ is well defined, then*

$\eta^* \in \arg\min \eta + \frac{1}{1-\alpha} v^{-1}\mathsf{E}v(X-\eta)$ *if and only if*

$$\partial_- v^{-1}\big(\mathsf{E}v(X-\eta^*)\big) \cdot \mathsf{E}\partial_- v(X-\eta^*) \leq 1-\alpha \leq$$
$$\partial_+ v^{-1}\big(\mathsf{E}v(X-\eta^*)\big) \cdot \mathsf{E}\partial_+ v(X-\eta^*),$$

where $\partial_\pm v$ denote one-sided derivatives of v with respect to the argument.

Proof: Let us denote $\phi_X(\eta) = \eta + \frac{1}{1-\alpha} v^{-1}\mathsf{E}v(X-\eta)$. Since certainty equivalent $v^{-1}\mathsf{E}v$ is convex, ϕ_X is also convex, and thus, it has left and right derivatives everywhere on $\operatorname{dom}\phi_X \neq \emptyset$, and η delivers a minimum to ϕ_X if and only if $\partial_-\phi_X(\eta) \leq 0 \leq \partial_+\phi_X(\eta)$. In what follows, we determine closed form expressions for left and right derivatives of ϕ_X. By definition, if $\eta \in \operatorname{dom}\phi_X$ then

$$\begin{aligned}\partial_+\phi_X(\eta) =& \lim_{\varepsilon\downarrow 0} \frac{\phi_X(\eta+\varepsilon)-\phi_X(\eta)}{\varepsilon}\\ =& 1 + \frac{1}{1-\alpha}\lim_{\varepsilon\downarrow 0}\frac{v^{-1}\mathsf{E}v(X-\eta-\varepsilon) - v^{-1}\mathsf{E}v(X-\eta)}{\varepsilon}.\end{aligned}$$

Repeating a usual argument used to prove the chain rule of differentiation (see, e.g., Randolph (1952)), we can define

$$Q(y) = \begin{cases} \dfrac{v^{-1}(y) - v^{-1}\mathsf{E}v(X-\eta)}{y - \mathsf{E}v(X-\eta)}, & y < \mathsf{E}v(X-\eta),\\ \partial_- v^{-1}\big(\mathsf{E}v(X-\eta)\big), & \text{otherwise,}\end{cases}$$

in which case

$$\partial_+\phi_X(\eta) = 1 + \frac{1}{1-\alpha}\lim_{\varepsilon\downarrow 0}\left\{Q\big(\mathsf{E}v(X-\eta-\varepsilon)\big)\frac{\mathsf{E}v(X-\eta-\varepsilon)-\mathsf{E}v(X-\eta)}{\varepsilon}\right\}.$$

Clearly, $\lim_{\varepsilon\downarrow 0} Q\big(\mathsf{E}v(X-\eta-\varepsilon)\big) = \partial\_v^{-1}\big(\mathsf{E}v(X-\eta)\big)$ by monotone convergence theorem, and the only part left to find is

$$\lim_{\varepsilon\downarrow 0}\frac{\mathsf{E}v(X-\eta-\varepsilon)-\mathsf{E}v(X-\eta)}{\varepsilon} = -\lim_{\varepsilon\downarrow 0}\frac{\mathsf{E}v(X-\eta)-\mathsf{E}v(X-\eta-\varepsilon)}{\varepsilon}.$$

Observe that $\lim_{\varepsilon\downarrow 0}\dfrac{v(x-\eta)-v(x-\eta-\varepsilon)}{\varepsilon} = \partial\_v(x-\eta)$ for any fixed $x\in\mathbb{R}$ (note that $\partial\_v(x-\eta)$ exists since v is convex). Moreover,

$$\frac{v(x-\eta)-v(x-\eta-\varepsilon)}{\varepsilon} \nearrow \partial\_v(x-\eta) \quad \text{as} \quad \varepsilon \searrow 0,$$

where $\nearrow$ denotes monotonic convergence from below (Rockafellar, 1997, Theorem 23.1). Thus, by monotone convergence theorem, we can interchange the limit and expectation:

$$\begin{aligned}\lim_{\varepsilon\downarrow 0}\frac{\mathsf{E}v(X-\eta)-\mathsf{E}v(X-\eta-\varepsilon)}{\varepsilon} &= \mathsf{E}\lim_{\varepsilon\downarrow 0}\frac{v(X-\eta)-v(X-\eta-\varepsilon)}{\varepsilon}\\ &= \mathsf{E}\partial\_v(X-\eta),\end{aligned}$$

i.e., $\partial_+\phi_X(\eta) = 1 - \frac{1}{1-\alpha}\partial\_v^{-1}\big(\mathsf{E}v(X-\eta)\big)\cdot\mathsf{E}\partial\_v(X-\eta)$. Similar argument can be applied to evaluate $\partial_-\phi_X(\eta)$ in order to complete the proof. □

Corollary 2.13. *Condition*

$$(v^{-1})'(\mathsf{E}v(X-\eta))\mathsf{E}v'(X-\eta) = 1-\alpha$$

is sufficient for η *to deliver the minimum in* (2.4)*, given that* $(v^{-1})'$ *and* v' *are well-defined.*

Conditions established above show that for a fixed X, the location of $\eta^*(X)$ is determined by the parameter α. Two propositions below illustrate this observation.

Proposition 2.14. *Given an* $X \in \operatorname{dom}\rho$ *for all* $\alpha \in (0,1)$*, if* $\eta^*_\alpha(X) \in \arg\min \eta + \frac{1}{1-\alpha} v^{-1}\mathsf{E}v(X-\eta)$*, where* v *is a one-sided deutility function, and certainty equivalent* $v^{-1}\mathsf{E}v$ *exists for any* X*, and is convex, then* $\eta^*_{\alpha_1}(X) \leq \eta^*_{\alpha_2}(X)$ *for any* $\alpha_1 < \alpha_2$*.*

Proof: Below we will use $\eta^*_\alpha(X)$ and η^*_α interchangeably in order to simplify the notation. Let $\alpha_1 < \alpha_2$. Since v is a one-sided deutility, then $v(X-\eta) = v([X-\eta]_+)$, and by the definition of $\eta^*_\alpha(X)$,

$$\eta^*_{\alpha_1} + \frac{1}{1-\alpha_1} v^{-1}\mathsf{E}v\big([X-\eta^*_{\alpha_1}]_+\big) \leq \eta^*_{\alpha_2} + \frac{1}{1-\alpha_1} v^{-1}\mathsf{E}v\big([X-\eta^*_{\alpha_2}]_+\big).$$

Suppose that $\eta^*_{\alpha_1} > \eta^*_{\alpha_2}$, then one has

$$\begin{aligned} 0 < \eta^*_{\alpha_1} - \eta^*_{\alpha_2} &\leq \frac{1}{1-\alpha_1}\Big(v^{-1}\mathsf{E}v\big([X-\eta^*_{\alpha_2}]_+\big) - v^{-1}\mathsf{E}v\big([X-\eta^*_{\alpha_1}]_+\big)\Big) \\ &< \frac{1}{1-\alpha_2}\Big(v^{-1}\mathsf{E}v\big([X-\eta^*_{\alpha_2}]_+\big) - v^{-1}\mathsf{E}v([X-\eta^*_{\alpha_1}]_+)\Big). \end{aligned}$$

This immediately leads to

$$\eta^*_{\alpha_1} + \frac{1}{1-\alpha_2} v^{-1}\mathsf{E}v\big([X-\eta^*_{\alpha_1}]_+\big) < \eta^*_{\alpha_2} + \frac{1}{1-\alpha_2} v^{-1}\mathsf{E}v\big([X-\eta^*_{\alpha_2}]_+\big),$$

which contradicts the definition of $\eta^*_{\alpha_2}$, thus furnishing the statement of the proposition. □

Proposition 2.15. *Given an* $X \in \operatorname{dom}\rho$ *for all* $\alpha \in (0,1)$*, if* $\eta^*_\alpha(X) \in \arg\min \eta + \frac{1}{1-\alpha} v^{-1}\mathsf{E}v(X-\eta)$*, where* v *is a one-sided deutility function, and certainty equivalent* $v^{-1}\mathsf{E}v$ *exists for any* X*, and is convex, then*

$$\lim_{\alpha\to 1} \eta^*_\alpha(X) = \text{ess.sup}(X).$$

Proof: Again, let us consider function $\phi_X(\eta) = \eta + \frac{1}{1-\alpha} v^{-1} \mathsf{E} v(X-\eta)$, and since v is a one-sided deutility, $\phi_X(\eta) = \eta + \frac{1}{1-\alpha} v^{-1} \int_{X \geq \eta} v(X-\eta) \mathrm{d}\mathsf{P}$. Suppose that $\mathrm{ess.sup}(X) = A < +\infty$, consequently $\mathsf{P}(X \geq A - \varepsilon) > 0$ for any $\varepsilon > 0$. Note that $\phi_X(A) = A$. Now,

$$\begin{aligned} \phi_X(A-\varepsilon) =& A - \varepsilon + \frac{1}{1-\alpha} v^{-1} \int_{X \geq A-\varepsilon} v(X - A + \varepsilon) \mathrm{d}\mathsf{P} \geq \\ & A - \varepsilon + \frac{1}{1-\alpha} v^{-1} \int_{X \geq A - \frac{\varepsilon}{2}} v(X - A + \varepsilon) \mathrm{d}\mathsf{P} \geq \\ & A - \varepsilon + \frac{1}{1-\alpha} v^{-1} \Big(v\Big(\frac{\varepsilon}{2}\Big) \mathsf{P}\Big(X \geq A - \frac{\varepsilon}{2}\Big)\Big) = A - \varepsilon + \frac{1}{1-\alpha} M_\varepsilon, \end{aligned}$$

where $M_\varepsilon = v^{-1}\Big(v\big(\frac{\varepsilon}{2}\big)\mathsf{P}\big(X \geq A - \frac{\varepsilon}{2}\big)\Big) > 0$. Hence, $\phi_X(A-\varepsilon) > \phi_X(A)$ for any sufficiently large values of α, which means that in this case any $\eta^*_\alpha(X) \in \arg\min \eta + \frac{1}{1-\alpha} v^{-1} \mathsf{E} v(X-\eta)$ has to satisfy $\eta^*_\alpha(X) \in (A-\varepsilon, A]$, and thus $\lim_{\alpha \to 1} \eta^*_\alpha(X) = A = \mathrm{ess.sup}(X)$.

Now, let $\mathrm{ess.sup}(X) = +\infty$. Note that $\int_{X \geq \eta} v(X-\eta) \mathrm{d}\mathsf{P}$ is a non-increasing function of η. Let $A \in \mathbb{R}$ and $\phi_X(A) = A + \frac{1}{1-\alpha} v^{-1} \int_{X \geq A} v(X-A) \mathrm{d}\mathsf{P}$. Since $\mathrm{ess.sup}(X) = +\infty$, there exists $\tilde{A} > A$ such that $0 < \int_{X \geq \tilde{A}} v(X-\tilde{A}) \mathrm{d}\mathsf{P} < \int_{X \geq A} v(X-A) \mathrm{d}\mathsf{P}$. Thus, $\phi_X(\tilde{A}) = \tilde{A} + \frac{1}{1-\alpha} v^{-1} \int_{X \geq \tilde{A}} v(X-\tilde{A}) \mathrm{d}\mathsf{P} < \phi_X(A)$ for any sufficiently large α, which yields $\eta^*_\alpha(X) > A$. Since the value of A has been selected arbitrarily, $\lim_{\alpha \to 1} \eta^*_\alpha(X) = +\infty = \mathrm{ess.sup}(X)$. □

2.2 Application: Log-Exponential Convex Measures of Risk

As it was already mentioned above, CVaR and HMCR measures can be defined in terms of the proposed certainty equivalent-based representation (2.5). Note that both cases correspond to positively homogeneous functions ϕ, and, therefore, are coherent measures of risk. Next we consider a convex measure of risk resulting from the certainty equivalent representation (2.5) with an exponential one-sided deutility function $v(t) = -1 + \lambda^{[t]_+}$:

$$\rho_\alpha^{(\lambda)}(X) = \min_\eta \ \eta + \frac{1}{1-\alpha} \log_\lambda \mathsf{E}\lambda^{[X-\eta]_+}, \quad \text{where} \quad \lambda > 1 \quad \text{and} \quad \alpha \in (0,1). \tag{2.7}$$

We refer to such $\rho_\alpha^{(\lambda)}$ as the family of *log-exponential convex risk (LogExpCR) measures*. First, using the general framework developed above, it can be readily seen that LogExpCR family are convex measures of risk.

Proposition 2.16. *Functions $\rho_\alpha^{(\lambda)}(X)$ defined by* (2.7) *are proper convex measures of risk.*

Proof: Follows immediately from Theorem 2.5, Proposition 2.7 and Corollary 2.9. □

A particular member of the family of LogExpCR measures is determined by the values of two parameters, α and λ. Recall that in Section 2.1.4 we have established that parameter α plays a key role in determining the position of $\eta_\alpha^*(X) \in \arg\min \eta + \frac{1}{1-\alpha} v^{-1}\mathsf{E}v(X - \eta)$, particularly, $\alpha_1 < \alpha_2$ leads to $\eta^*_{\alpha_1}(X) \le \eta^*_{\alpha_2}(X)$, and $\lim_{\alpha\to 1} \eta^*_\alpha(X) = \text{ess.sup}(X)$. These two properties allow us to conclude that α determines the "length" of the tail of distribution of X, or, in other words, determines which part of the distribution should be considered "risky". This is in accordance with a similar property of the CVaR measure, which, in the case of a continuous loss distribution, quantifies the risk as the expected loss in the worst $1 - \alpha$ percent of the cases. See Krokhmal (2007) for a similar argument for HMCR measures.

Furthermore, one has

$$\begin{aligned}
\rho_\alpha^{(\lambda)}(X) =& \min_\eta \eta + \frac{1}{1-\alpha} \log_\lambda \mathsf{E}\lambda^{[X-\eta]_+} = \min_\eta \eta + \frac{1}{1-\alpha}\frac{1}{\ln\lambda} \ln \mathsf{E}e^{\ln\lambda [X-\eta]_+} \\
=& \frac{1}{\ln\lambda} \min_\eta \eta\ln\lambda + \frac{1}{1-\alpha} \mathsf{E}e^{[X\ln\lambda - \eta\ln\lambda]_+} = \\
& \frac{1}{\ln\lambda} \min_{\eta'} \eta' + \frac{1}{1-\alpha} \mathsf{E}e^{[X\ln\lambda - \eta']_+} = \frac{1}{\ln\lambda}\rho_\alpha^{(e)}(X\ln\lambda).
\end{aligned}$$

This implies that LogExpCR measures satisfy a "quasi positive homogeneity" property:

$$\rho_\alpha^{(\lambda)}(X)\ln\lambda = \rho_\alpha^{(e)}(X\ln\lambda),$$

where parameter $\ln\lambda$ plays the role of a scaling factor. Thus, in the case of log-exponential convex measures of risk (2.7), scaling can be seen as a way to designate the total range of the loss variable. Consequently, a combination of the parameters α and λ determines both the region of the loss distribution that should be considered "risky", and the emphasis that should be put on the larger losses. Note, that the specific choice of the parameter values should be determined by the decision-maker's preferences and attitude towards risk. A preliminary computational study may be required to calibrate this values.

It is of interest to note that LogExpCR measures are isotonic with respect to *any* order $k \geq 1$ of stochastic dominance:

Proposition 2.17. *The family of log-exponential convex measures of risk* (2.7) *are kSD-isotonic for any* $k \geq 1$*, i.e.,* $\rho_\alpha^{(\lambda)}(X) \leq \rho_\alpha^{(\lambda)}(Y)$ *for all* $X, Y \in \mathcal{X}$ *such that* $-X \succcurlyeq_{kSD} -Y$.

Proof: Follows immediately from Proposition 2.10 for v defined above. □

Based on these observations and the preceding discussion, we can conclude that the introduced family of LogExpCR measures possesses a number of desirable properties from both optimization and methodological perspectives. It is widely acknowledged in the literature that risk is associated with "heavy" tails of the loss distribution; for example, in Krokhmal (2007) it has been illustrated that evaluating risk exposure in terms of higher tail moments can lead to improved decision making in financial applications with heavy-tailed distributions of asset returns. Furthermore, there are many real-life applications where risk exposure is associated with catastrophic events of very low probability and extreme magnitude, such as natural disasters, which often turn out to be challenging for traditional analytic tools (see, for example, Kousky and Cooke (2009); Cooke and Nieboer (2011) and references therein, Iaquinta et al. (2009), or Kreinovich et al. (2012)). By construction, LogExpCR measures quantify risk by putting extra emphasis on the tail of the distribution, which allows us to hypothesize that they could perform favorably compared to conventional approaches in situations that involve heavy-tailed distributions of losses and catastrophic risks. This conjecture has been tested in two numerical case studies that are presented next. The idea is to evaluate the quality of solutions based on the risk estimates due to nonlinear LogExpCR measure with those obtained using linear CVaR measure, which can now be considered as a standard approach in risk-averse applications. Particularly, we were interested in assessing the influence that the behavior of the tails of the underlying losses distributions has in this comparison.

2.2.1 Case Study 1: Flood Insurance Claims Model

2.2.1.1 Dataset description

For the first part of the case study we used a dataset managed by a non-profit research organization *Resources for the Future* (Cooke and Nieboer (2011)). It contains flood insurance claims, filed through National Flood Insurance Program (NFIP), aggregated by county and year for the State of Florida from 1980 to 2006. The data is in 2000 US dollars divided by personal income estimates per county per year from the Bureau of Economic

Accounts (BEA), in order take into account substantial growth in exposure to flood risk. The dataset has 67 counties, and spans for 355 months.

2.2.1.2 Model formulation

Let random vector $\boldsymbol{\ell}$ represent the dollar values of insurance claims (individual elements of this vector correspond to individual counties), and consider the following stochastic programming problem, where ρ is a risk measure:

$$\min \quad \rho(\boldsymbol{\ell}^\top \boldsymbol{x}) \tag{2.8a}$$

$$\text{s. t.} \quad \sum_i x_i = K \tag{2.8b}$$

$$x_i \in \{0, 1\}. \tag{2.8c}$$

Such a formulation allows for a straightforward interpretation, namely, the goal here is to identify K counties with a minimal common insurance risk due to flood as estimated by ρ. Clearly, such a simplified model does not reflect the complexities of real-life insurance operations. At the same time, since the purpose of this case study is to analyze the properties of risk measures themselves, a deliberately simple formulation was chosen so as to highlight the differences between solutions of (2.8) due to different choices of the risk measure ρ in (2.8a).

Given that the distribution of $\boldsymbol{\ell}$ is represented by equiprobable scenario realizations $\ell_{i,j}$, and m is the number of scenarios (time periods), model (2.8) with risk measure chosen as the Conditional Value-at-Risk, $\rho(X) = \mathsf{CVaR}_\alpha(X)$, can be expressed as

$$\min \quad \eta + \frac{1}{1-\alpha_{\text{CVaR}}} \sum_j \frac{1}{m} \Big[\sum_i x_i \ell_{i,j} - \eta \Big]_+ \tag{2.9a}$$

$$\text{s. t.} \quad \sum_i x_i = K \tag{2.9b}$$

$$x_i \in \{0, 1\}. \tag{2.9c}$$

Similarly, if a LogExpCR measure is used, $\rho(X) = \rho_\alpha^{(e)}(X)$, then (2.8) can be formulated as

$$\min \quad \eta + \frac{1}{1-\alpha_{\text{LogExpCR}}} \log \sum_j \frac{1}{m} e^{[\sum_i x_i \ell_{i,j} - \eta]_+} \tag{2.10a}$$

$$\text{s. t.} \quad \sum_i x_i = K \tag{2.10b}$$

$$x_i \in \{0, 1\}. \tag{2.10c}$$

2.2.1.3 Normal data

In order to be able to evaluate the effect of the tail behavior of the loss distribution on the obtained solutions of decision making problems, we additionally generated a similar dataset based on normal distribution. Particularly, we draw 355 realizations from 67-dimensional normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, where $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are mean and covariance estimates of NFIP data respectively. Our goal here is to make sure that the main difference between the datasets lays in the tails (normal distribution is a well-known example of a light-tailed distribution), and by preserving mean vector and covariance matrix we secure that this dataset captures the leading trends present in the original data. Now, by comparing the decisions due to CVaR and LogExpCR measures for these two datasets we can make conclusions on the effects that the tails of the distributions have on the quality of subsequent decisions.

2.2.1.4 Implementation details

Problems (2.9) and (2.10) represent a mixed-integer linear programming (MIP) and a mixed-integer non-linear programming (MINLP) problems respectively. MIP problems were solved using IBM ILOG CPLEX 12.5 solver accessed through C++ API. For the MINLPs of the form (2.10) we implemented a custom branch-and-bound algorithm based on outer polyhedral approximation approach, which utilized CPLEX 12.5 MIP solver and MOSEK 6.0 for NLP subproblems.

In order to evaluate the quality of the decisions we employed a usual training-testing framework. Given a preselected value m, the first m scenarios were used to solve problems (2.9) and (2.10), then for the remaining $N - m$ scenarios the total loss was calculated as $L^\rho = \sum_{j=m}^{N} \sum_i \ell_{i,j} x_i^\rho$, where $\boldsymbol{x}^\rho$ represents an optimal solution of either problem (2.9) or problem (2.10), and N is the total number of scenarios in the dataset. In other words, the decision vector $\boldsymbol{x}^\rho$ is selected based on the first m observations of the historical data (training), and the quality of this solution is estimated based on the "future" realizations (testing).

For this experiments we have set the parameter α_{CVaR} set to 0.9, which is a usual practical choice and can be interpreted as cutting off 90% of the least significant losses. A preliminary test experiment has been performed to select α_{LogExpCR} in such a way that approximately same portion of the distribution was cut off, which yielded $\alpha_{\text{LogExpCR}} = 0.5$. For the sake of simplicity, parameter λ has been set equal to e.

2.2.1.5 Discussion of results

Tables 2.1 and 2.2 summarize the obtained results for NFIP and simulated normal data sets, respectively. Discrepancy in the quality of the decisions based on LogExpCR and CVaR measures is estimated using the value

$$\gamma = \frac{L^{\text{LogExpCR}} - L^{\text{CVaR}}}{\min\left\{L^{\text{LogExpCR}}, L^{\text{CVaR}}\right\}},$$

which represents the relative difference in total losses L^{LogExpCR} and L^{CVaR} associated with the respective decisions. For example, $\gamma = -100\%$ corresponds to the case when losses due to CVaR-based decision were twice as large as losses due to LogExpCR-based decision.

First of all, we can observe that there is a definite variation between the results obtained with NFIP data on one hand and with simulated normal data on the other. Particularly, the absolute values of γ in Table 2.2 on average are considerably smaller compared to those in Table 2.1, which indicates that in the case of normal data the risk measures under consideration result in similar decisions, while heavy-tailed historical data leads to much more differentiated decisions.

Secondly, Table 2.1 suggests that LogExpCR measure yields considerably better solutions for certain sets of parameter values. Most notably, such instances correspond to smaller values of both K and m. Intuitively, this can be explained as follows. Recall that m is the number of scenarios in the training set, and $N - m$ is the number of scenarios in the testing set, which means that larger values of m correspond to shorter testing horizon. Clearly, the fewer scenarios there are in the testing set, the fewer catastrophic losses occur during this period, and vice versa, for smaller values of m there are more exceptionally high losses in the future. Thus, the observed behavior of γ is in accordance with our conjecture that LogExpCR measures are better suited for instances with heavy-tailed loss distributions. Parameter K, in turn, corresponds to the number of counties to be selected, thus, the larger its value is, the more opportunities for diversification are available for the decision-maker, which, in turn, allows for risk reduction.

To sum up, the results of this case study suggest that under certain conditions, such as heavy-tailed loss distribution, relatively poor diversification opportunities, and sufficiently large testing horizon, risk-averse decision strategies based on the introduced log-exponential convex measures of risk can substantially outperform strategies based on linear risk measures, such as the Conditional Value-at-Risk.

2.2.2 Case Study 2: Portfolio Optimization

As heavy-tailed loss distributions are often found in financial data, we conducted numerical experiments with historical stock market data as the second part of the case study.

2.2.2.1 Model description

As the underlying decision making model we use the traditional risk-reward portfolio optimization framework introduced by Markowitz (1952). In this setting, the cost/loss outcome X is usually defined as the portfolio negative rate of return, $X(\boldsymbol{x}, \omega) = -\mathbf{r}(\omega)^\top \boldsymbol{x}$, where $\boldsymbol{x}$ stands for the vector of portfolio weights, and $\mathbf{r} = \mathbf{r}(\omega)$ is the uncertain vector of assets' returns. Then, a portfolio allocation problem can be formulated as the problem of minimizing some measure of risk associated with the portfolio while maintaining a prescribed expected return:

$$\min_{\boldsymbol{x}\in\mathbb{R}^n_+} \left\{ \rho(-\mathbf{r}^\top\mathbf{x}) \;\middle|\; \mathsf{E}(\mathbf{r}^\top\mathbf{x}) \geq \bar{r},\ \mathbf{1}^\top\mathbf{x} \leq 1 \right\}, \tag{2.11}$$

where $\bar{r}$ is the prescribed level of expected return, $\boldsymbol{x} \in \mathbb{R}^n_+$ denotes the no-short-selling requirement, and $\mathbf{1} = (1, \dots, 1)^\top$. If the risk measure used is convex, it is easy to see that (2.11) is a convex optimization problem. In this case study, we again select ρ in (2.11) as either a LogExpCR or CVaR measure.

2.2.2.2 Dataset description

We utilized historical stock market data available through Yahoo!Finance. We picked 2178 listings traded at NYSE from March, 2000 through December, 2012 (total of 3223 trading days). As it was noted above, financial data often exhibit highly volatile behavior, especially higher-frequency data, while long-term data is usually relatively normal. In order to account for such differences, we generated three types of datasets of loss distribution, which were based on two-day, two-week and one-month historical returns. Particularly, if $p_{i,j}$ is the historical close price of asset i on day j, then we define the corresponding two-day, ten-day, and one-month returns as $r_{i,j} = \frac{p_{i,j} - p_{i,j-\Delta}}{p_{i,j-\Delta}}$, where Δ takes values $\Delta = 2$, 10, and 20, respectively.

2.2.2.3 Implementation details

We utilize a training-testing framework similar to the one used in the previous section, but additionally, we also employ "rolling horizon" approach, which aims to simulate a real-life self-financing trading strategy. For a given time moment, we generate a scenario set containing, respectively, m two-day, ten-day, and one-month returns immediately preceding this date. Then, the portfolio optimization problem (2.11) is solved for each type of scenario set in order to obtain the corresponding optimal portfolios; the "realized" portfolio return over the next two-day, ten-day, or one-month time period, respectively, is then observed. The portfolio is then rebalanced using the described procedure. This rolling-horizon procedure was ran for 800 days, or about 3 years.

Recall that parameter $\bar{r}$ in (2.11) represents the "target return", i.e., the minimal average return of the portfolio. Parameter $\bar{r}$ was selected as $\bar{r} = \tau \max_i\{\mathsf{E}_\omega r_i(\omega)\}$, i.e., as a certain percentage of the maximum expected return previously observed in the market (within the timespan of the current scenario set). Parameter τ has been set to be ”low“, ”moderate“, or ”high“, which corresponds to $\tau = 0.1, 0.5, 0.8$. For each pair of n and m we repeat the experiment 20 times, selecting n stocks randomly each time. Parameters α_{LogExpCR}, α_{CVaR} and λ have been set in the same way as in Case Study 1.

2.2.2.4 Discussion of results

Obtained results are summarized in Table 2.3, and a typical behavior of the portfolio value over time is presented in Figure 2.1. As in the previous case, we report relative difference in the return over appropriate time period (2-day, 2-week, or 1-month) averaged over the testing horizon of 800 days and over 20 random choices of n assets. Note that since in this case the quality of the decision is estimated in terms of rate of return, i.e., gain, positive values in Table 2.3 correspond to the cases when the LogExpCR-based portfolio outperforms the CVaR-based portfolio.

Similarly to the previous case, we can observe that the behavior of the tails of the distribution plays a key role in the comparison: under 1-month trading frequency the differences between CVaR and LogExpCR portfolios are relatively insignificant, compared to the 2-day case. Moreover, we can again conclude that for heavy-tailed loss distributions the introduced LogExpCR measure may compare favorably against CVaR; in particular, conditions of restricted diversification options (relatively small value of n) make utilization of LogExpCR measures more beneficial compared to a linear measure such as CVaR.

2.3 Concluding Remarks

In our view, the contribution of this chapter is threefold. First, we introduce a new general representation of the classes of convex and coherent risk measures by showing that any convex (coherent) measure can be defined as an infimal convolution of the form $\rho(X) = \min_\eta \eta + \phi(X - \eta)$, where ϕ is monotone, convex, and $\phi(\eta) > \eta$ for all $\eta \neq 0$, $\phi(0) = 0$ (and positive homogeneous for coherency), and vice versa, constructed in such a way function ρ is convex (coherent). Another way to look at this result is to observe that a monotone and convex ϕ only lacks translation invariance in order to satisfy the definition of a convex risk measure, and infimal convolution operator essentially forges this additional property, while preserving monotonicity and convexity. According to this scheme, a risk measure is represented as a solution of an optimization problem, hence it can be readily embedded in a stochastic programming model.

Secondly, we apply the developed representation to construct risk measures as infimal convolutions of certainty equivalents, which allows for a direct incorporation of risk preferences as given by the utility theory of von Neumann and Morgenstern (1944) into a convex or coherent measure of risk. This is highly desirable since, in general, the risk preferences induced by convex or coherent measures of risk are inconsistent with risk preferences of rational expected-utility maximizers. It is also shown that the certainty equivalent-based measures of risk are "naturally" consistent with stochastic dominance orderings.

Finally, we employ the proposed scheme to introduce a new family of risk measures, which we call the family of log-exponential convex risk measures. By construction, LogExpCR measures quantify risk by placing emphasis on extreme or catastrophic losses; also, the LogExpCR measures have been shown to be isotonic (consistent) with respect to stochastic dominance of arbitrary order. The results of the conducted case study show that in highly risky environments characterized by heavy-tailed loss distribution and limited diversification opportunities, utilization of the proposed LogExpCR measures can lead to improved results comparing to the standard approaches, such as those based on the well-known Conditional Value-at-Risk measure.

Table 2.1: Relative difference in total loss $\gamma = \frac{L^{\text{LogExpCR}} - L^{\text{CVaR}}}{\min\{L^{\text{LogExpCR}}, L^{\text{CVaR}}\}}$ for *NFIP data* for various values of the parameters K and m. Entities in bold correspond to the instances for which LogExpCR measure outperformed CVaR.

$K \backslash m$	20	60	100	140	180	220	260	300
1	**-45038.3%**	**-3944.4%**	**-4652.2%**	**-3663.7%**	**-3663.7%**	**-220.2%**	**-220.2%**	**-220.2%**
3	**-1983.7%**	**-971.0%**	**-211.5%**	**-146.7%**	**-146.7%**	**-68.4%**	0.0%	0.0%
5	**-1284.2%**	**-464.2%**	**-85.7%**	**-13.1%**	**-13.1%**	0.0%	**-6.6%**	0.0%
7	**-853.9%**	**-342.5%**	0.0%	0.0%	0.0%	**-0.4%**	**-4.5%**	**-13.1%**
9	**-387.1%**	**-282.9%**	0.0%	0.0%	0.0%	0.0%	**-3.1%**	10.8%
11	**-369.9%**	**-181.0%**	0.0%	**-18.2%**	14.0%	**-27.8%**	5.5%	**-2.2%**
13	**-360.4%**	**-33.5%**	0.0%	**-13.0%**	1.0%	4.3%	0.0%	41.0%
15	**-353.9%**	**-27.9%**	**-3.2%**	3.1%	0.0%	**-3.6%**	4.8%	20.6%
17	**-129.8%**	**-1.1%**	**-0.2%**	3.7%	**-26.5%**	11.5%	25.4%	25.4%
19	**-66.3%**	21.6%	0.9%	0.0%	0.0%	**-2042.1%**	35.4%	23.1%
21	**-64.0%**	5.0%	0.0%	**-279.0%**	2.5%	**-1.6%**	35.0%	8.0%
23	**-57.4%**	4.8%	0.0%	0.7%	**-65.6%**	**-0.1%**	20.3%	81.8%
25	**-49.5%**	0.0%	**-82.4%**	0.0%	**-39.2%**	4.4%	76.9%	84.7%
27	**-48.2%**	0.0%	**-52.2%**	0.0%	4.7%	4.1%	68.7%	84.1%
29	**-47.0%**	**-34.3%**	33.0%	**-254.3%**	**-463.8%**	4.0%	81.8%	83.5%
31	**-41.1%**	**-31.2%**	8.7%	**-218.4%**	**-309.8%**	8.5%	79.3%	83.7%
33	**-10.6%**	46.4%	**-10.0%**	**-162.7%**	**-161.7%**	8.9%	19.6%	84.6%
35	**-9.5%**	0.0%	**-12.2%**	**-142.9%**	**-153.1%**	37.8%	53.9%	47.6%
37	**-7.7%**	12.0%	**-81.7%**	5.3%	2.7%	57.0%	15.0%	9.9%
39	0.0%	5.3%	**-102.8%**	45.8%	45.4%	43.4%	8.6%	5.6%
41	0.0%	11.4%	**-77.3%**	30.9%	43.8%	34.8%	20.7%	4.8%
43	0.0%	**-13.4%**	**-11.0%**	53.8%	4.0%	50.0%	19.8%	**-3.4%**
45	0.0%	0.0%	**-28.1%**	54.5%	**-36.1%**	26.2%	14.2%	8.5%
47	**-9.1%**	9.0%	4.5%	19.4%	6.4%	17.5%	28.2%	**-2.2%**
49	0.0%	6.4%	27.8%	**-3.5%**	**-20.1%**	7.0%	0.5%	**-8.5%**
51	0.0%	5.1%	49.6%	**-2.6%**	1.0%	**-9.5%**	**-5.0%**	2.1%
53	39.9%	**-16.2%**	24.7%	28.6%	23.0%	2.9%	**-4.9%**	**-137.2%**
55	**-7.1%**	**-8.7%**	28.0%	21.6%	**-0.9%**	5.6%	**-83.5%**	**-19.1%**
57	0.0%	3.0%	28.5%	3.6%	4.1%	2.6%	**-9.1%**	**-9.2%**
59	0.0%	20.0%	9.0%	2.0%	**-0.6%**	0.1%	25.3%	27.4%

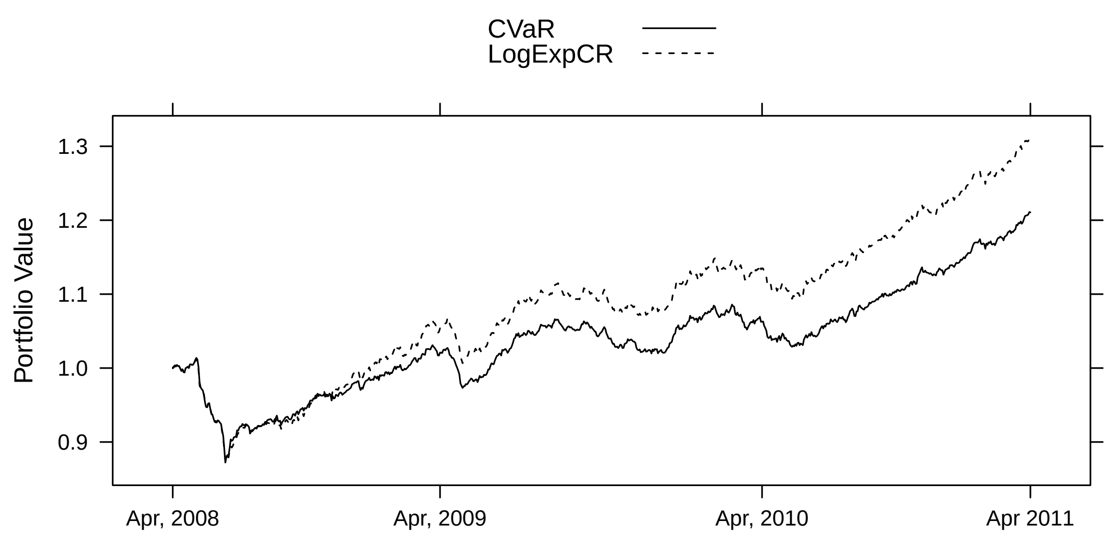
Table 2.2: Relative difference in total loss $\gamma = \frac{L^{\text{LogExpCR}} - L^{\text{CVaR}}}{\min\{L^{\text{LogExpCR}}, L^{\text{CVaR}}\}}$ for *normal data* for various values of the parameters K and m. Entities in bold LogExpCR to the instances for which LogExpCR measure outperformed CVaR.

$K\backslash m$	20	60	100	140	180	220	260	300
1	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
3	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
5	0.0%	2.2%	**-107.1%**	0.0%	0.0%	**-17.6%**	22.8%	**-58.1%**
7	0.0%	14.0%	0.0%	85.3%	28.8%	84.8%	86.4%	21.9%
9	0.0%	14.2%	27.2%	11.0%	2.9%	0.0%	0.0%	0.0%
11	17.2%	12.9%	0.0%	34.8%	33.9%	66.8%	36.0%	**-50.4%**
13	19.2%	16.6%	**-3.3%**	**-11.1%**	2.6%	18.0%	**-12.3%**	0.0%
15	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
17	0.0%	**-3.4%**	34.0%	0.0%	0.0%	312.1%	0.0%	**-355.0%**
19	43.0%	**-8.2%**	8.9%	52.3%	76.7%	**-65.9%**	**-20.3%**	0.0%
21	0.0%	4.3%	21.5%	**-45.4%**	506.1%	**-123.1%**	**-119.2%**	1.9%
23	27.3%	**-32.1%**	48.8%	75.2%	242.3%	**-63.1%**	3.0%	**-112.1%**
25	**-317.3%**	**-8.0%**	16.9%	74.7%	151.1%	**-71.3%**	**-129.0%**	**-64.7%**
27	9.7%	**-34.1%**	31.1%	**-50.8%**	96.3%	154.3%	163.8%	**-16.8%**
29	7.4%	13.7%	19.4%	78.4%	44.6%	272.6%	**-15.3%**	-31.4%
31	1.8%	10.3%	5.3%	6.4%	52.6%	234.0%	44.8%	**-5.5%**
33	10.5%	**-14.7%**	**-15.2%**	**-31.2%**	**-32.8%**	11.5%	**-15.0%**	10.1%
35	9.1%	6.0%	0.0%	0.0%	36.8%	36.9%	437.2%	0.0%
37	5.1%	**-1.0%**	0.0%	18.0%	39.1%	20.8%	119.3%	0.0%
39	0.0%	**-0.8%**	**-1.4%**	10.3%	13.2%	**-14.6%**	**-109.7%**	73.6%
41	19.8%	18.3%	0.0%	24.7%	22.1%	0.0%	44.0%	762.8%
43	7.6%	8.7%	6.4%	0.0%	**-6.1%**	0.0%	0.0%	0.0%
45	6.9%	5.9%	11.4%	7.9%	6.1%	16.9%	-20.6%	**-99.3%**
47	0.0%	1.1%	16.6%	4.0%	13.0%	0.0%	21.5%	46.7%
49	**-2.8%**	22.5%	17.7%	**-7.5%**	-11.2%	**-2.3%**	0.0%	-294.8%
51	0.0%	5.1%	17.8%	5.0%	10.4%	**-28.4%**	-0.1%	-47.4%
53	**-1.1%**	0.0%	**-6.7%**	**-0.5%**	25.4%	0.0%	0.0%	**-39.7%**
55	6.8%	0.0%	17.5%	18.3%	0.0%	-9.3%	37.8%	**-87.4%**
57	1.3%	0.0%	0.0%	**-14.5%**	**-21.8%**	0.0%	0.0%	0.0%
59	6.3%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	**-10.0%**

Table 2.3: Relative difference (in %) in average portfolio return due to LogExpCR measure and CVaR. Parameter n represents the total number of assets on the market, m is the number of time intervals in the training horizon, $\bar{r}$ is the prescribed expected rate of return. Labels “2-day”, “2-week”, and “1-month” correspond to portfolio rebalancing periods.

n	m	$\bar{r}$	2-day	2-week	1-month
20	2000	0.1	**57.3**	**29.5**	**8.3**
		0.5	**138.3**	**1.1**	–12.9
		0.8	**5.9**	–24.1	–7.4
200	2000	0.1	–17.9	–14.6	–2.2
		0.5	**11.1**	–21.1	**5.4**
		0.8	**17.6**	–13.5	–2.2

Figure 2.1: Typical behavior of portfolio value, as a multiple of the initial investment (1.0), over time.



CHAPTER 3
POLYHEDRAL APPROXIMATIONS IN P-ORDER CONE PROGRAMMING

3.1 Problem Formulation and Literature Review

In the remaining chapters we will focus our attention on mathematical programming techniques that can be employed in solving optimization problems resulting from the proposed decision making approaches discussed in Chapter 2. Our first goal will be to explore methods for solving p-order cone programming problems, as this class, in our view, presents a particularly important and interesting case. Later in the manuscript we will discuss in more details the connection between p-order cone programming problems and the modeling approach presented in Chapter 2, but for the next two chapters our main attention will be focused on designing solution techniques for such problems without regard to a particular modeling framework. With this in mind, we consider the following optimization problem

$$\min \quad \boldsymbol{c}^\top \boldsymbol{x} \tag{3.1a}$$

$$\text{s. t.} \quad \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b}, \tag{3.1b}$$

$$\left\|\boldsymbol{C}^{(k)}\boldsymbol{x} + \boldsymbol{e}^{(k)}\right\|_{p_k} \leq \boldsymbol{h}^{(k)\top}\boldsymbol{x} + f^{(k)}, \quad k = 1, \ldots, K, \tag{3.1c}$$

$$\boldsymbol{x} \in \mathbb{R}^n,$$

where $\|\cdot\|_p$ denotes the p-norm in $\mathbb{R}^N$:

$$\|\boldsymbol{a}\|_p = \begin{cases} \left(|a_1|^p + \cdots + |a_N|^p\right)^{1/p}, & p \in [1, \infty), \\ \max\left\{|a_1|, \ldots, |a_N|\right\}, & p = \infty. \end{cases}$$

We call formulation (3.1) a p-order cone programming problem (pOCP) by analogy with second-order cone programming (SOCP), which constitutes a special case of (3.1) when

$p_k = 2$ for all $k = 1, \ldots, K$.

The available literature on solving problem (3.1) with "general" values of $p_k \in (1, \infty)$, i.e., not restricted to well-studied special cases of $p_k = 1$, 2, or ∞, is relatively limited. Interior-point approaches to p-order cone programming have been considered by Xue and Ye (2000) with respect to minimization of sum of p-norms; a self-concordant barrier for p-cone has also been introduced by Nesterov (2012). Glineur and Terlaky (2004) proposed an interior-point algorithm along with the corresponding barrier functions for a related problem of l_p-norm optimization (see also Terlaky (1985)). A polyhedral approximation approach to pOCP problems was considered by Krokhmal and Soberanis (2010). In the case when p is a rational number, the existing primal-dual methods of second-order cone programming can be employed for solving p-order cone optimization problems using a reduction of p-order cone constraints to a system of linear and second-order cone constraints proposed by Nesterov and Nemirovski (1994) and Ben-Tal and Nemirovski (2001a), see also Morenko et al. (2013).

This chapter represents a continuation of the work of Krokhmal and Soberanis (2010) on polyhedral approximation approaches to solving pOCP problems. This work's contribution to the literature consists of the following: it is shown that the cutting plane method developed in Krokhmal and Soberanis (2010) for solving a special type of polyhedral approximations of pOCP problems, which allows for generation of cuts in a constant time not dependent on the accuracy of approximation, is applicable to a larger family of polyhedral approximations. Further, it is demonstrated that this constant-time cut generation procedure can be modified so as constitute an exact solution method with $O(\varepsilon^{-1})$

iteration complexity. Moreover, we demonstrate the existence of a constant-time cut generation scheme for lifted polyhedral approximations of SOCP problems that are constructed recursively, with the length of recursion controlling the accuracy of approximation, and which were introduced by Ben-Tal and Nemirovski (2001b). Finally, we illustrate that the polyhedral approximation approach and the corresponding cutting plane solution methods can be efficiently employed for obtaining exact solutions of mixed-integer extensions of pOCP problems (see below).

There is substantial literature on solution approaches for mixed integer cone programming problems. In many cases, the proposed approaches attempt to extend some of the techniques developed for mixed integer linear programming. One of such research directions concerns development of branch-and-bound schemes based on outer polyhedral approximations of cones. This potentially allows for computational savings in traversing the branch-and-bound tree due to the "warm start" capabilities of linear programming solvers. In particular, Vielma et al. (2008) proposed a branch-and-bound method for MISOCP was proposed that employed lifted polyhedral approximations of second order cones due to Ben-Tal and Nemirovski (2001b). Drewes (2009) presented subgradient-based linear outer approximations for the second order cone constraints in mixed integer programs. With respect to mixed integer nonlinear programming, a similar idea has been exploited by Bonami et al. (2008) and Tawarmalani and Sahinidis (2005).

The chapter is organized as follows: in Section 3.2 we discuss the general properties of polyhedral approximations of p-cones, Section 3.3.1 summarizes the general cutting plane method for polyhedral approximations of pOCP problems. In Sections 3.3.2 and

3.3.3 we explore fast constant-time cut generating techniques for gradient-based and lifted polyhedral approximations of pOCP and SOCP problems, respectively. The developed solution techniques are then illustrated on pOCP and SOCP problems of type (3.1), and are also employed for solving mixed-integer p-order cone programming (MIpOCP) problems

$$
\begin{aligned}
\min \quad & \boldsymbol{c}^\top \boldsymbol{x} + \boldsymbol{d}^\top \boldsymbol{z} && \text{(3.2a)}\\
\text{s. t.} \quad & \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{z} \le \boldsymbol{b}, && \text{(3.2b)}\\
& \left\| \boldsymbol{C}^{(k)}\boldsymbol{x} + \boldsymbol{D}^{(k)}\boldsymbol{z} + \boldsymbol{e}^{(k)} \right\|_{p_k} \le \boldsymbol{h}^{(k)\top}\boldsymbol{x} + \boldsymbol{g}^{(k)\top}\boldsymbol{z} + f^{(k)}, \quad k = 1, \ldots, K, && \text{(3.2c)}\\
& \boldsymbol{x} \in \mathbb{R}^n, \quad \boldsymbol{z} \in \mathbb{Z}^m, && \text{(3.2d)}
\end{aligned}
$$

which arise in the context of portfolio optimization with certainty equivalent risk measures. The corresponding discussion is presented in Section 3.4.

3.2 Polyhedral Approximations of p-Order Cones

In what follows, we consider, without loss of generality, p-cones in the positive orthant of $\mathbb{R}^{N+1}$:

$$\mathcal{K}_p^{(N+1)} = \left\{ \boldsymbol{\xi} \in \mathbb{R}_+^{N+1} \;\middle|\; \xi_0 \ge \|(\xi_1, \ldots, \xi_N)\|_p \right\}. \tag{3.3}$$

Then, by a polyhedral approximation of $\mathcal{K}_p^{(N+1)}$ we understand a polyhedral cone in $\mathbb{R}_+^{N+1+\kappa_m}$, where $\kappa_m \ge 0$ may be generally non-zero,

$$\mathcal{H}_{p,m}^{(N+1)} = \left\{ \begin{pmatrix} \boldsymbol{\xi} \\ \mathbf{u} \end{pmatrix} \in \mathbb{R}_+^{N+1+\kappa_m} \;\middle|\; \mathbf{H}_{p,m}^{(N+1)} \begin{pmatrix} \boldsymbol{\xi} \\ \mathbf{u} \end{pmatrix} \ge \mathbf{0} \right\}, \tag{3.4}$$

having the properties that:

(H1) any $(\xi_0, \ldots, \xi_N)^\top \in \mathcal{K}_p^{(N+1)}$ can be extended to some

$(\xi_0, \ldots, \xi_N, u_1, \ldots, u_{\kappa_m})^\top \in \mathcal{H}_{p,m}^{(N+1)}$;

(H2) for some prescribed $\varepsilon = \varepsilon(m) > 0$, any $(\xi_0, \dots, u_{\kappa_m})^\top \in \mathcal{H}^{(N+1)}_{p,m}$ satisfies

$$\|(\xi_1, \dots, \xi_N)\|_p \leq (1 + \varepsilon)\xi_0.$$

Here m is the parameter of the construction that controls the approximation accuracy ε. Replacing each of the p-order cone constraints in problem (3.1) by their polyhedral approximations of the form (3.4), we obtain a linear programming approximation of the pOCP problem (3.1):

$$\min \left\{ \boldsymbol{c}^\top \boldsymbol{x} \;\middle|\; \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b},\;\; \mathbf{H}^{(N_k+1)}_{p_k,m_k} \begin{pmatrix} \boldsymbol{h}^{(k)\top}\boldsymbol{x} + f^{(k)} \\ \boldsymbol{C}^{(k)}\boldsymbol{x} + \boldsymbol{e}^{(k)} \\ \boldsymbol{u}^{(k)} \end{pmatrix} \geq \mathbf{0}, \quad k = 1, \dots, K \right\}. \tag{3.5}$$

Observe that the projection of the feasible region of (3.5) on the space of variables $\boldsymbol{x}$ lies in between the feasible set of pOCP (3.1) and that of its "ε-relaxation",

$$\min \left\{ \boldsymbol{c}^\top \boldsymbol{x} \;\middle|\; \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b},\;\; \left\|\boldsymbol{C}^{(k)}\boldsymbol{x} + \boldsymbol{e}^{(k)}\right\|_{p_k} \leq (1+\varepsilon)\big(\boldsymbol{h}^{(k)\top}\boldsymbol{x} + f^{(k)}\big),\;\; k = 1, \dots, K \right\}. \tag{3.6}$$

Thus, problem (3.5) represents an ε-approximation of pOCP (3.1), given that the feasible regions of problems (3.1) and (3.6) are "close". Conditions under which the feasible sets of (3.1) and (3.6) are indeed $O(\varepsilon)$-close have been given by Ben-Tal and Nemirovski (2001b), Proposition 4.1 for the case of $p = 2$, and their argumentation carries over to the case of $p \neq 2$ practically without modifications. Specifically, if we denote by (pOCP) and (pOCP$_\varepsilon$) the initial problem (3.1) and its polyhedral ε-approximation (3.6), respectively, the following holds.

Proposition 3.1 (Ben-Tal and Nemirovski (2001b)). *Assume that* (pOCP) *is: (i) strictly* *feasible, i.e., there exist* $\bar{\boldsymbol{x}}$ *and* $r > 0$ *such that*

$$\boldsymbol{A}\bar{\boldsymbol{x}} \le \boldsymbol{b}, \quad \left\|\boldsymbol{C}^{(k)}\bar{\boldsymbol{x}} + \boldsymbol{e}^{(k)}\right\|_{p_k} \le \boldsymbol{h}^{(k)\top}\bar{\boldsymbol{x}} + f^{(k)} - r, \quad k = 1, \ldots, K, \tag{3.7a}$$

and (ii) "semibounded", i.e., there exists $R > 0$ *such that*

$$\boldsymbol{A}\boldsymbol{x} \le \boldsymbol{b}, \ \left\|\boldsymbol{C}^{(k)}\boldsymbol{x} + \boldsymbol{e}^{(k)}\right\|_{p_k} \le \boldsymbol{h}^{(k)\top}\boldsymbol{x} + f^{(k)}, \ k = 1, \ldots, K \ \Rightarrow \tag{3.7b}$$

$$\boldsymbol{h}^{(k)\top}\boldsymbol{x} + f^{(k)} \le R, \ k = 1, \ldots, K. \tag{3.7c}$$

Then for every $\varepsilon > 0$ *such that* $\gamma(\varepsilon) = R\varepsilon/r < 1$*, one has*

$$\gamma(\varepsilon)\bar{\boldsymbol{x}} + (1 - \gamma(\varepsilon))\,\mathrm{Feas}\,(\mathrm{pOCP}_\varepsilon) \subset \mathrm{Feas}\,(\mathrm{pOCP}) \subset \mathrm{Feas}\,(\mathrm{pOCP}_\varepsilon),$$

where $\mathrm{Feas}\,(\mathrm{P})$ *denotes the feasible set of a problem* (P).

Remark 10. The established inclusions essentially state, that under these conditions ε-relaxation of the pOCP problem provides both an outer and an inner approximation of the feasible region of the pOCP. For example, if we can take $\bar{\boldsymbol{x}} = 0$, then the claim reduces to $(1 - \gamma(\varepsilon))\,\mathrm{Feas}\,(\mathrm{pOCP}_\varepsilon) \subset \mathrm{Feas}\,(\mathrm{pOCP}) \subset \mathrm{Feas}\,(\mathrm{pOCP}_\varepsilon)$, where $\gamma(\varepsilon) \to 1$ for $\varepsilon \to 0$. In other words, $\mathrm{Feas}\,(\mathrm{pOCP}_\varepsilon)$ and $(1 - \gamma(\varepsilon))\,\mathrm{Feas}\,(\mathrm{pOCP}_\varepsilon)$ are "close" to $\mathrm{Feas}\,(\mathrm{pOCP})$ from the outside and from the inside respectively, and thus, for any feasible solution of the relaxation there is a close feasible solution of the initial problem.

In constructing polyhedral approximations (3.4) of p-order cones we follow the approach of Ben-Tal and Nemirovski (2001b), who developed efficient, in terms of dimensionality, polyhedral approximations for quadratic cones. The first step in the construction procedure consists in a lifted representation, dubbed by the authors "tower of variables", of a p-cone in $\mathbb{R}_+^{N+1}$, as a nested sequence of $N-1$ three-dimensional p-cones. The original construction relied on the assumption that $N = 2^d$ for some integer $d \geq 1$, which was by no means restrictive, but allowed for a simple structure of the lifted set, which could be visualized as a symmetric binary tree of three-dimensional cone inequalities that are partitioned into $d = \log_2 N$ "levels", with 2^{d-l} inequalities at a level l. Below we present a slightly different notation/representation of the "tower-of-variables" lifting technique that does not explicitly use the binary tree structure, and which simplifies its practical implementation in the case of general $N \neq 2^d$. Namely, given the $(N+1)$-dimensional p-cone, consider the set defined by intersection of $N-1$ three-dimensional p-cones in $\mathbb{R}_+^{N+1} \times \mathbb{R}_+^{N-1}$:

$$\xi_0 = \xi_{2N-1}, \quad \xi_{N+j} \geq \|(\xi_{2j-1}, \xi_{2j})\|_p, \quad j = 1, \ldots, N-1. \tag{3.8}$$

Proposition 3.2. *Projection of set* (3.8) *onto the space of variables* $(\xi_0, \ldots, \xi_N)$ *coincides with the set* (3.3)*. In other words, any* $\boldsymbol{\xi} \in \mathbb{R}_+^{N+1}$ *that satisfies* (3.3) *can be extended to* $\widetilde{\boldsymbol{\xi}} \in \mathbb{R}_+^{N+1} \times \mathbb{R}_+^{N-1}$ *that satisfies* (3.8)*, and any* $\widetilde{\boldsymbol{\xi}} \in \mathbb{R}_+^{2N}$ *satisfying* (3.8) *is such that its first* $N+1$ *components satisfy* (3.3)*.*

Proof: Follows immediately by expanding the recursion in (3.8). □

The second step of the procedure is to construct a polyhedral approximation

$$\mathcal{H}_{p,m}^{(3)} = \left\{ \begin{pmatrix} \boldsymbol{\xi} \\ \mathbf{u} \end{pmatrix} \in \mathbb{R}_+^{3+\kappa_m} \;\middle|\; \mathbf{H}_{p,m}^{(3)} \begin{pmatrix} \boldsymbol{\xi} \\ \mathbf{u} \end{pmatrix} \geq \mathbf{0} \right\} \tag{3.9}$$

for each of the three-dimensional p-cones in (3.8). Observe that if approximation (3.9) of each of the three-dimensional p-cones (3.8) contains $O(\nu)$ facets, $\nu = \nu(m)$, the total number of facets in the approximation of the original $(N+1)$-dimensional p-cone is $O(\nu N)$, i.e., it is linear in the dimensionality N of the original p-cone.

Proposition 3.3. *Consider cone* (3.3) *and its lifted representation* (3.8)*. If each of the three-dimensional cones in* (3.8) *is approximated by* (3.9) *with an accuracy* $\epsilon > 0$*, the resulting approximation accuracy* ε *of the original cone* (3.3) *satisfies* $\varepsilon \leq (1+\epsilon)^{\lceil \log_2 N \rceil} - 1 = \lceil \log_2 N \rceil \epsilon + O(\epsilon^2)$.

Proof: The vector $\widetilde{\boldsymbol{\xi}} \in \mathbb{R}_+^{2N}$ must satisfy $\xi_0 = \xi_{2N-1}$, $(1+\epsilon)\xi_{N+j} \geq \|(\xi_{2j-1}, \xi_{2j})\|_p$, $j = 1, \ldots, N-1$. Expanding the recursion, we obtain

$$\xi_0^p = \xi_{2N-1}^p \geq \frac{\xi_{2N-3}^p}{(1+\epsilon)^p} + \frac{\xi_{2N-2}^p}{(1+\epsilon)^p} \geq \frac{\xi_{2N-7}^p}{(1+\epsilon)^{2p}} + \frac{\xi_{2N-6}^p}{(1+\epsilon)^{2p}} + \frac{\xi_{2N-5}^p}{(1+\epsilon)^{2p}} + \frac{\xi_{2N-4}^p}{(1+\epsilon)^{2p}} \geq \frac{\xi_1^p}{(1+\epsilon)^{pk_1}} + \ldots + \frac{\xi_N^p}{(1+\epsilon)^{pk_N}},$$

where k_i is the number of "levels" in the "tower of variables" on the way from ξ_{2N-1} to ξ_i. It is straightforward to check that $k_i \in \{\lceil \log_2 N \rceil - 1, \lceil \log_2 N \rceil\}$ and thus, $(1+\epsilon)^{\lceil \log_2 N \rceil} \xi_0 \geq \|(\xi_1, \ldots, \xi_N)\|_p$. □

When $p = 1$ or $p = \infty$, the cone $\mathcal{K}_p^{(3)}$ is already polyhedral; in the case of $p = 2$, the problem of constructing a polyhedral approximation of the second-order cone $\mathcal{K}_2^{(3)}$ was also addressed by Ben-Tal and Nemirovski (2001b), who proposed the following *lifted* polyhedral approximation of $\mathcal{K}_2^{(3)}$,

$$u_0 \geq \xi_1, \tag{3.10a}$$

$$v_0 \geq \xi_2, \tag{3.10b}$$

$$u_i = \cos\left(\tfrac{\pi}{2^{i+1}}\right) u_{i-1} + \sin\left(\tfrac{\pi}{2^{i+1}}\right) v_{i-1}, \qquad i = 1, \ldots, m, \tag{3.10c}$$

$$v_i \geq \left| -\sin\left(\tfrac{\pi}{2^{i+1}}\right) u_{i-1} + \cos\left(\tfrac{\pi}{2^{i+1}}\right) v_{i-1} \right|, \qquad i = 1, \ldots, m, \tag{3.10d}$$

$$u_m \leq \xi_0,\ v_m \leq \tan\left(\tfrac{\pi}{2^{m+1}}\right) u_m, \tag{3.10e}$$

$$0 \leq u_i,\ v_i, \qquad i = 0, \ldots, m. \tag{3.10f}$$

Remarkably, the accuracy of the polyhedral approximation (3.10) is exponentially small in m: $\epsilon(m) = O\big(4^{-m}\big)$. The construction is based on an elegant geometric argument that utilizes a well-known elementary fact that rotation of a vector in $\mathbb{R}^2$ is an affine transformation that preserves the Euclidean norm (2-norm) and that the parameters of this affine transform depend only on the angle of rotation. An approach to constructing a framework of polyhedral relations that generalizes inductive constructions of extended formulations via projections, such as the polyhedral approximation (3.10) has been introduced by Kaibel and Pashkovich (2011).

Unfortunately, the lifted polyhedral approximation (3.10) of the second-order cone $\mathcal{K}_2^{(3)}$ does not seem to be extandable to general p-order cones $\mathcal{K}_p^{(3)}$ with $p \in (1,2) \cup (2, \infty)$. Therefore, we employ a "gradient" approximation of $\mathcal{K}_p^{(3)}$ using circumscribed planes. Given the parameter of construction $m \in \mathbb{N}$, let us call function $\varphi_m : [0, m] \mapsto [0, \pi/2]$ an *approximation function* if it is continuous and strictly increasing on $[0, m]$, and, moreover, satisfies

$$\Delta\varphi_m = \max_{i=0,\ldots,m-1}\{\varphi_m(i+1)-\varphi_m(i)\} \to 0, \quad m\to\infty.$$

Then, for the following parametrization of the p-cone surface in $\mathbb{R}^3_+$

$$\xi_1 = \xi_0\frac{\cos\theta}{(\cos^p\theta+\sin^p\theta)^{1/p}}, \quad \xi_2 = \xi_0\frac{\sin\theta}{(\cos^p\theta+\sin^p\theta)^{1/p}}, \quad \xi_0\geq 0, \quad \theta\in[0,\tfrac{\pi}{2}], \tag{3.11}$$

where θ is the polar angle, any given approximation function φ_m generates a *gradient approximation* of $\mathcal{K}_p^{(3)}$

$$\mathcal{H}^{(3)}_{p,m}(\varphi_m) = \big\{\,\boldsymbol{\xi}\in\mathbb{R}^3_+ \;\big|\; \xi_0 \geq \alpha_{p,i}[\varphi_m]\,\xi_1 + \beta_{p,i}[\varphi_m]\,\xi_2, \;\; i=0,\ldots,m\,\big\}, \tag{3.12a}$$

where

$$\begin{pmatrix}\alpha_{p,i}[\varphi_m]\\ \beta_{p,i}[\varphi_m]\end{pmatrix} = \big(\cos^p\varphi_m(i)+\sin^p\varphi_m(i)\big)^{1/p-1}\begin{pmatrix}\cos^{p-1}\varphi_m(i)\\ \sin^{p-1}\varphi_m(i)\end{pmatrix}, \quad i=0,\ldots,m. \tag{3.12b}$$

The values $\varphi_m(i)$ in (3.12) represent the polar angles at which the planes $\xi_0 = \alpha_{p,i}\xi_1+\beta_{p,i}\xi_2$ are tangent to the p-cone $\mathcal{K}_p^{(3)}$. In such a way, the properties of the polyhedral approximation (3.12) of the p-cone $\mathcal{K}_p^{(3)}$ are determined by the values of φ_m at integer values $\{0,\ldots,m\}$ of its argument; nevertheless, the computability properties of $\varphi_m(t)$ for arbitrary values $t\in[0,m]$ are also of major importance, as will be shown in the next section. The following proposition establishes the quality of the gradient polyhedral approximation (3.12), and is a generalization of a similar result established for a special choice of φ_m in Krokhmal and Soberanis (2010).

Proposition 3.4. *For large enough values of $m\in\mathbb{N}$, the polyhedral set $\mathcal{H}^{(3)}_{p,m}(\varphi_m)$ defined by the gradient approximation* (3.12) *with approximation function φ_m satisfies properties*

(H1)–(H2). Specifically, if the approximation function is such that for some $r > 0$

$$\Delta\varphi_m = O(m^{-r}), \quad m \gg 1,$$

then for any $\boldsymbol{\xi} \in \mathcal{K}_p^{(3)}$ *one has* $\boldsymbol{\xi} \in \mathcal{H}_{p,m}^{(3)}$*, and any* $\boldsymbol{\xi} \in \mathcal{H}_{p,m}^{(3)}$ *satisfies* $\|(\xi_1, \xi_2)\|_p \leq (1 + \epsilon(m))\xi_0$*, where the approximation accuracy* $\epsilon(m)$ *is polynomially small in* m*:*

$$\epsilon(m) = O\big(m^{-r\min\{p,2\}}\big), \quad m \gg 1.$$

Remark 11. One possible choice of φ_m is $\varphi_m(t) = \frac{\pi}{2m}t$, which yields a "uniform" gradient approximation of the p-cone, i.e., a gradient approximation (3.12) where the circumscribed planes are spaced "uniformly" with respect to the polar angle θ, and are tangent to the p-cone at the polar angles $\theta_i = \frac{\pi i}{2m}$. In this case, the approximation accuracy satisfies $\epsilon(m) = O\big(m^{-\min\{p,2\}}\big)$ (see Krokhmal and Soberanis (2010)) and is constant among all sectors $\big[\frac{\pi i}{2m}, \frac{\pi(i+1)}{2m}\big]$. In the case of $p \neq 2$, however, the accuracy of the uniform gradient approximation varies depending on the sector. Thus, it may be of interest to construct an approximation function φ_m that results in a constant accuracy at each sector $[\varphi_m(i), \varphi_m(i+1)]$, thereby minimizing the number of facets needed to achieve the desired accuracy. On the other hand, if the structure of the problem is such that an optimal solution is known to be located in a certain part of the cone, it might be beneficial to construct an approximation that is more accurate within this particular region and less accurate outside of it. These considerations provide an intuition on how a careful choice of φ_m may reduce the size of the problem in question. We, however, are not pursuing the problem of constructing an "optimal" approximation in this work.

For $p = 2$ and a given approximation accuracy, the lifted polyhedral approximation (3.10) due to Ben-Tal and Nemirovski (2001b) is superior to the gradient polyhedral approximation (3.12) in terms of dimensionality. However, computational studies of Glineur (2000); Krokhmal and Soberanis (2010) indicated that solving LP problems of the form (3.5) that were constructed as polyhedral approximations, either lifted or gradient, of SOCP problems, was computationally inefficient comparing to the "native" SOCP solution techniques, such as self-dual interior-point methods.

At the same time, the computational efficiency of the polyhedral approximation approach can be substantially improved by employing decomposition methods that exploit the specific structure of polyhedral approximations in (3.12), whereby the polyhedral approximation approach becomes competitive with SOCP-based solution methods for pOCP problems with $p \neq 2$. This was demonstrated for a special case of the uniform gradient polyhedral approximation in Krokhmal and Soberanis (2010). In the next section we show that analogous computational efficiencies can be achieved for more general gradient polyhedral approximations of pOCP problems, as well as for the lifted polyhedral approximation of SOCP problems.

3.3 Cutting Plane Methods for Polyhedral Approximations of SOCP and pOCP Problems

Computationally efficient methods for solving polyhedral approximations (3.4) of SOCP and pOCP problems can be constructed by taking advantage of (i) the special structure of the problem induced by the "tower-of-variables" representation of high-dimensional cones as an intersection of three-dimensional ones in a lifted space, and (ii) the special structures of polyhedral approximations of three-dimensional quadratic or p-order cones.

With respect to (i), a cutting plane method that, given a polyhedral approximation for 3D cones, utilizes the structure of the "tower-of-variables" reformulation in the approximating problems (3.4), was proposed in Krokhmal and Soberanis (2010). This method is briefly described in Section 3.3.1 below, since it is necessary in the context of (ii), namely, for exploiting the special properties of gradient and lifted polyhedral approximations of 3D cones for fast cut generation. In particular, the discussion that follows in Sections 3.3.2 and 3.3.3 demonstrates that, despite the differences in construction and properties, the lifted Ben-Tal-Nemirovski's approximation (3.10) of quadratic cones and the gradient approximation (3.12) of p-cones offer the same computational efficiency for cut generation.

3.3.1 A Cutting Plane Procedure for Polyhedral Approximations of pOCP Problems

The cutting plane algorithm described here is applicable to reformulations of pOCP problems obtained using the "tower-of-variables" lifting technique (3.8). Assuming for simplicity that problem (3.1) contains only one p-cone constraint $(K = 1)$ of dimension $N + 1$, the corresponding reformulation of (3.1) is obtained by lifting the p-cone constraint using the “tower-of-variables” method as

$$\begin{aligned}
\min \quad & \boldsymbol{c}^\top \boldsymbol{x} && && (3.13a)\\
\text{s. t.} \quad & \boldsymbol{A}\boldsymbol{x} && \leq \boldsymbol{b} && (3.13b)\\
& w_{N+j} && \geq \|(w_{2j-1}, w_{2j})\|_p, \quad j = 1, \ldots, N-1, && (3.13c)\\
& w_j && \geq |(\boldsymbol{C}\boldsymbol{x} + \boldsymbol{e})_j|, \quad j = 1, \ldots, N, && (3.13d)\\
& w_{2N-1} && = \mathbf{h}^\top \boldsymbol{x} + f, && (3.13e)
\end{aligned}$$

where $\boldsymbol{w} \in \mathbb{R}^{2N-1}$. Each of the three-dimensional p-order cones (3.13c) is subsequently replaced by its polyhedral approximation (3.9), which yields the following polyhedral approximation of pOCP (3.1):

$$\begin{aligned}
\min \quad & \boldsymbol{c}^\top \boldsymbol{x} && (3.14a)\\
\text{s. t.} \quad & \mathbf{H}^{(3)}_{p,m} \begin{pmatrix} \boldsymbol{w}_j \\ \boldsymbol{u}_j \end{pmatrix} \geq \mathbf{0}, \quad j = 1, \ldots, N-1, && (3.14b)\\
& \boldsymbol{u}_j \in \mathbb{R}^{\kappa_m}_{+}, && (3.14c)\\
& \text{(3.13b), (3.13d), (3.13e),} && (3.14d)
\end{aligned}$$

where the vectors $\boldsymbol{w}_j$ stand for the triplets $\boldsymbol{w}_j = (w_{N+j}, w_{2j-1}, w_{2j})^\top$. Constructed in such a way polyhedral approximation of the pOCP problem (3.1) possesses a special structure that can be exploited for solving the LP problem (3.14) efficiently. In particular, the following cutting plane representation for (3.14) was presented in Krokhmal and Soberanis

(2010):

$$\min \quad \boldsymbol{c}^\top \boldsymbol{x} \tag{3.15a}$$

$$\text{s. t.} \quad w_{N+j} \geq (0,\dots,0,w_{2j-1},w_{2j})\,\hat{\boldsymbol{\pi}}_i, \quad i \in \mathcal{P}_{p,m},\ j = 1,\dots,N-1, \tag{3.15b}$$

$$\text{(3.13b), (3.13d), (3.13e),} \tag{3.15c}$$

where $\mathcal{P}_{p,m}$ is the set of vertices $\hat{\boldsymbol{\pi}}_i$ of the polyhedron

$$\left\{ \boldsymbol{\pi} \geq \mathbf{0} \;\middle|\; \mathbf{H}_{p,m}^\top \boldsymbol{\pi} \leq \begin{pmatrix} 1 \\ \mathbf{0} \end{pmatrix} \right\}, \tag{3.16}$$

and the matrix $\mathbf{H}_{p,m}$ is obtained by augmenting the approximation matrix $\mathbf{H}_{p,m}^{(3)}$ with two extra rows $(0,1,0\cdots0)$, $(0,0,1,0\cdots0)$, where 1's correspond to the variables w_{2j-1} and w_{2j}:

$$\mathbf{H}_{p,m} = \begin{pmatrix} \mathbf{H}_{p,m}^{(3)} \\ 0\,1\,0\cdots0 \\ 0\,0\,1\cdots0 \end{pmatrix}.$$

Constraints (3.15b) are then generated via an iterative procedure. Assuming that problem (3.15) is bounded, consider the master problem in the form

$$\min \quad \boldsymbol{c}^\top \boldsymbol{x} \tag{3.17a}$$

$$\text{s. t.} \quad w_{N+j} \geq \varsigma_{j,i}\, w_{2j-1} + \tau_{j,i}\, w_{2j}, \quad i = 1,\dots,C_j,\ j = 1,\dots,N-1, \tag{3.17b}$$

$$\text{(3.13b), (3.13d), (3.13e),} \tag{3.17c}$$

where $\varsigma_{j,i}$ and $\tau_{j,i}$ stand for the components $\hat{\pi}_{\nu-1}$ and $\hat{\pi}_\nu$ of the vector $\hat{\boldsymbol{\pi}} \in \mathbb{R}^\nu$, and C_j is the number of constraints generated during preceding iterations. Let $(\boldsymbol{x}^*,\, \boldsymbol{w}^*) \in \mathbb{R}^{n+2N-1}$ be an optimal solution of the master (note that if (3.17) is infeasible, then (3.15) is infeasible too, and the procedure stops). For each $j = 1,\dots,N-1$, the following LP problem is solved:

$$\zeta_j^* := \max\left\{ \left(0,\ldots,0,w_{2j-1}^*,w_{2j}^*\right)\boldsymbol{\pi} \;\middle|\; \mathbf{H}_{p,m}^\top \boldsymbol{\pi} \le \begin{pmatrix}1\\ \mathbf{0}\end{pmatrix},\ \boldsymbol{\pi}\ge \mathbf{0} \right\}, \tag{3.18}$$

and it is checked whether the condition

$$w_{N+j}^* \ge \zeta_j^* = w_{2j-1}^*\pi_{\nu-1}^{*(j)} + w_{2j}^*\pi_{\nu}^{*(j)} \tag{3.19}$$

holds, where $\boldsymbol{\pi}^{*(j)}$ is an optimal solution of (3.18). If it does not, a new constraint (3.17b) is added for the variable w_{N+j} by incrementing the corresponding counter of constraints in (3.17b): $C_j := C_j + 1$, and setting $\varsigma_{j,i'} = \pi_{\nu-1}^{*(j)}$, $\tau_{j,i'} = \pi_{\nu}^{*(j)}$ for $i' = C_j$. Upon checking condition (3.19) for all variables w_{N+j}, $j = 1,\ldots,N-1$, in (3.17), the master problem (3.17) is augmented with new constraints and is solved again. If (3.19) holds for all variables w_{N+j}, and thus no new cuts are generated during an iteration, the current solution $\boldsymbol{x}^*, \boldsymbol{w}^*$ of the master problem is optimal for the original LP approximation problem (3.15). In such a way, the described cutting plane procedure obtains an optimal solution, if it exists, of the original LP approximation problem (3.15) after a finite number of iterations, with, perhaps, some anticycling scheme employed.

3.3.2 Fast Cut Generation for Gradient Approximations of p-Order Cones

The cutting-plane scheme of Section 3.3.1 exploits the properties of the "tower-of-variables" representation (3.8) of high-dimensional p-cones as a nested sequence of 3D p-cones to facilitate solving (large-scale) polyhedral approximations (3.4). In this section we show that if the gradient polyhedral approximation (3.12) is used for approximating three-dimensional p-cones in (3.14), the structure of this approximation can be utilized to achieve significant computational savings, provided that the approximation function φ_m of the gradient polyhedral approximation satisfies a certain computability condition.

Proposition 3.5. *Consider a polyhedral approximation* (3.5) *of pOCP problem* (3.1)*, obtained by reformulating each of the K p-cones in* (3.1) *using the "tower-of-variables" representation* (3.8) *and then applying the gradient polyhedral approximation* (3.12) *with parameter of construction m and approximation function φ_m. Then, if φ_m^{-1} is computable in $O(1)$ time, during an iteration of the cutting plane scheme of Section 3.3.1 new cuts can be generated in $O\big(\sum_k N_k\big)$ time that is independent of m, where N_k+1 is the dimension of k^{th} p-cone in* (3.1).

Similarly to Proposition 3.4, this result strengthens the statement in Krokhmal and Soberanis (2010). We still provide its proof here, since it is necessary for formalizing a subsequent observation in Proposition 3.6.

Proof: Proof of Proposition 3.5: When the gradient polyhedral approximation (3.12) is used, the cut-generating problem (3.18) can be formulated as

$$\max\left\{ \sum_{i=0}^{m}(\alpha_{p,i}\xi_1^* + \beta_{p,i}\xi_2^*)\,\pi_i - \sum_{i=1}^{2}\xi_i^* s_i \;\middle|\; \sum_{i=0}^{m}\pi_i \le 1,\;\; \pi_0,\ldots,\pi_m \ge 0,\;\; s_1, s_2 \ge 0 \right\}, \tag{3.20}$$

where the constants ξ_1^* and ξ_2^* stand for the corresponding elements of the current optimal solution $\boldsymbol{w}^*$ of the master problem: $\xi_1^* = w_{2j-1}^*$, $\xi_2^* = w_{2j}^*$. Disregarding the trivial case of $\xi_1^* = \xi_2^* = 0$, we assume that at least one of these parameters is positive: $\xi_1^* + \xi_2^* > 0$. It is clear that solving (3.20) amounts to finding a maximum element of the set $\{\alpha_{p,i}\xi_1^* +$

$\beta_{p,i}\xi_2^*\}_{i=0,\dots,m}$. Namely, if one has

$$i^* \in \arg\max_{i=0,\dots,m} \left\{\alpha_{p,i}\, \xi_1^* + \beta_{p,i}\, \xi_2^*\right\}, \tag{3.21a}$$

then an optimal solution $\boldsymbol{\pi}^*$ of (3.20) is given by

$$\pi_i^* = 0,\ i \in \{0,\dots,m\} \setminus i^*; \quad \pi_{i^*}^* = 1; \quad s_1 = \alpha_{p,i^*}; \quad s_2 = \beta_{p,i^*}. \tag{3.21b}$$

For fixed $\xi_1^*, \xi_2^* \geq 0$ and $p > 1$, consider the function

$$g(t) = \xi_1^* \frac{\cos^{p-1} t}{(\cos^p t + \sin^p t)^{1-1/p}} + \xi_2^* \frac{\sin^{p-1} t}{(\cos^p t + \sin^p t)^{1-1/p}}, \quad t \in [0, \tfrac{\pi}{2}],$$

with the derivative

$$g'(t) = (p-1)\frac{\sin^{p-1} t\, \cos^{p-1} t}{(\cos^p t + \sin^p t)^{2-1/p}} \left(\frac{-\xi_1^*}{\cos t} + \frac{\xi_2^*}{\sin t} \right).$$

Obviously, for $t \in [0, \frac{\pi}{2}]$ function $g(t)$ is either strictly monotone (when one of ξ_1^*, ξ_2^* is zero) or has a unique global maximum at $t^* = \arctan(\xi_2^*/\xi_1^*)$. Then, for a continuous and strictly increasing approximating function $\varphi_m : [0, m] \mapsto [0, \frac{\pi}{2}]$, the function $g(\varphi_m(\cdot))$ is also either monotone on $[0, m]$ or has a unique maximum at $\varphi_m^{-1}(\arctan(\xi_2^*/\xi_1^*))$. Consequently, if the inverse φ_m^{-1} of the approximating function is computable in $O(1)$ time, the index i^* of a maximum element of the sequence

$$g(\varphi_m(i)) = \xi_1^*\, \alpha_{p,i} + \xi_2^*\, \beta_{p,i}, \quad i = 0, \dots, m,$$

which defines an optimal solution (3.21) of cut-generating problem (3.20), can be determined in $O(1)$ time as

$$i^* \in \arg\max \left\{\varphi_m^{-1}(0), \lfloor \varphi_m^{-1}(t^*) \rfloor, \lfloor \varphi_m^{-1}(t^*) \rfloor + 1, \varphi_m^{-1}(\tfrac{\pi}{2})\right\}, \tag{3.22}$$

where $t^* = \arctan(\xi_2^*/\xi_1^*)$.

Given that each p-cone constraint of order p_k and dimensionality $N_k + 1$ requires $N_k - 1$ such operations, generation of new cuts in problem (3.17) that employs a gradient polyhedral approximation requires $O\big(\sum_k N_k\big)$ time. □

In the case when $\xi_1^*, \xi_2^* > 0$, the index i^* of the cut that may have to be added to the master is given by $\lfloor \varphi_m^{-1}(t^*) \rfloor$ or $\lfloor \varphi_m^{-1}(t^*) \rfloor + 1$. Note that as m increases (and the quality of approximation becomes finer), for any fixed $\xi_1^*, \xi_2^* > 0$ the facets corresponding to $\lfloor \varphi_m^{-1}(t^*) \rfloor$, $\lfloor \varphi_m^{-1}(t^*) \rfloor + 1$ converge to a plane tangent to the cone at the point determined by the polar angle $\theta^* = \arctan(\xi_2^*/\xi_1^*)$, so that the corresponding cut takes the form

$$\begin{aligned} w_{N+j} \geq & w_{2j-1} \frac{\cos^{p-1}\theta^*}{(\cos^p\theta^* + \sin^p\theta^*)^{1-1/p}} + w_{2j} \frac{\sin^{p-1}\theta^*}{(\cos^p\theta^* + \sin^p\theta^*)^{1-1/p}}, \\ \theta^* = & \arctan \frac{w_{2j}^*}{w_{2j-1}^*}. \end{aligned} \tag{3.23}$$

In this case, one does not need to solve the cut-generating LP (3.18) and check condition (3.19) in order to add the corresponding cut. Namely, for a current solution $\boldsymbol{w}^*$ of the master, cut (3.23) is added to the master if the condition

$$\big\|\big(w_{2j-1}^*, w_{2j}^*\big)\big\|_p \leq (1+\epsilon) w_{N+j}^* \tag{3.24}$$

is not satisfied for the respective $j = 1, \ldots, N-1$. The following proposition formalizes this procedure.

Proposition 3.6. *Given an instance of pOCP problem* (3.1) *that satisfies the conditions of Proposition 3.1, consider a cutting plane scheme for constructing an approximate solution of its lifted reformulation* (3.13)*, where the master problem has the form* (3.17)*, and for a* *given solution* $\boldsymbol{x}^*, \boldsymbol{w}^*$ *of the master, cuts of the form* (3.23) *are added if condition* (3.24) *is not satisfied for a specific* j. *Assuming that* (3.13) *is bounded, this cutting plane procedure terminates after a finite number of iterations for any given* $\epsilon > 0$, *with, perhaps, some anti-cycling scheme applied. In particular, the algorithm is guaranteed to generate at most* $O(\varepsilon^{-1})$ *cutting planes, and in the special case of* $p = 2$ *the described cutting plane algorithm is guaranteed to stop after at most* $O(\varepsilon^{-0.5})$ *iterations.*

Proof: Let $\epsilon > 0$ be the approximation accuracy for the 3D p-cones in (3.13), and w^*_{N+j}, w^*_{2j-1}, and w^*_{2j} be the elements of the current solution of the master. We will show that there exists some δ_ϵ such that if θ^*_j is located at an angular distance closer than δ_ϵ from an existing cut, then (3.23) implies (3.24), i.e., no new cut can be added withing δ_ϵ from an existing one. By (3.23), for any existing cut at polar angle θ_k the solution of the master should satisfy

$$w^*_{N+j} \geq w^*_{2j-1}\frac{\cos^{p-1}\theta_k}{(\cos^p\theta_k + \sin^p\theta_k)^{1-\frac{1}{p}}} + w^*_{2j}\frac{\sin^{p-1}\theta_k}{(\cos^p\theta_k + \sin^p\theta_k)^{1-\frac{1}{p}}} = \left\|(w^*_{2j-1}, w^*_{2j})\right\|_p \times \left(\frac{\cos\theta^*_j}{(\cos^p\theta^*_j + \sin^p\theta^*_j)^{\frac{1}{p}}}\frac{\cos^{p-1}\theta_k}{(\cos^p\theta_k + \sin^p\theta_k)^{1-\frac{1}{p}}} + \frac{\sin\theta^*_j}{(\cos^p\theta^*_j + \sin^p\theta^*_j)^{\frac{1}{p}}}\frac{\sin^{p-1}\theta_k}{(\cos^p\theta_k + \sin^p\theta_k)^{1-\frac{1}{p}}}\right),$$

where $\theta^*_j = \arctan\dfrac{w^*_{2j}}{w^*_{2j-1}}$. Let $\theta^*_j = \theta_k + \delta$, where $\delta > 0$, in which case

$$\begin{aligned} w^*_{N+j} &\geq \left\|(w^*_{2j-1}, w^*_{2j})\right\|_p \times \\ &\frac{\cos\delta\,(\cos^p\theta_k + \sin^p\theta_k) + \sin\delta\,(\sin^{p-1}\theta_k\cos\theta_k - \cos^{p-1}\theta_k\sin\theta_k)}{(\cos^p(\theta_k+\delta) + \sin^p(\theta_k+\delta))^{\frac{1}{p}}(\cos^p\theta_k + \sin^p\theta_k)^{1-\frac{1}{p}}} \qquad (3.25) \\ &= \left\|(w^*_{2j-1}, w^*_{2j})\right\|_p\big(A(\theta_k,\delta)\cos\delta + B(\theta_k,\delta)\sin\delta\big), \end{aligned}$$

where we denote

$$A(\theta_k,\delta) = \frac{(\cos^p\theta_k + \sin^p\theta_k)^{\frac{1}{p}}}{(\cos^p(\theta_k+\delta)+\sin^p(\theta_k+\delta))^{\frac{1}{p}}},$$
$$B(\theta_k,\delta) = \frac{\sin^{p-1}\theta_k\cos\theta_k - \cos^{p-1}\theta_k\sin\theta_k}{(\cos^p(\theta_k+\delta)+\sin^p(\theta_k+\delta))^{\frac{1}{p}}(\cos^p\theta_k+\sin^p\theta_k)^{1-\frac{1}{p}}}.$$

As δ approaches zero, the right-hand side in (3.25) converges uniformly to $\left\|(w^*_{2j-1}, w^*_{2j})\right\|_p$. Namely, let $K_0 = \min_\theta \|(\cos\theta, \sin\theta)\|_p = \text{const} > 0$, then

$$\begin{aligned}
&\left|A(\theta_k,\delta)\cos\delta + B(\theta_k,\delta)\sin\delta - 1\right| \leq \\
&\sin\delta|B(\theta_k,\delta)| + A(\theta_k,\delta)(1-\cos\delta) + |A(\theta_k,\delta)-1| \\
&\leq \frac{2}{K_0^p}\sin\delta + \frac{1}{K_0}(1-\cos\delta) + \\
&\frac{1}{K_0}\left|\left(\cos^p\theta_k+\sin^p\theta_k\right)^{\frac{1}{p}} - \left(\cos^p(\theta_k+\delta)+\sin^p(\theta_k+\delta)\right)^{\frac{1}{p}}\right| \\
&\leq \frac{2}{K_0^p}\sin\delta + \frac{1}{K_0}(1-\cos\delta) + \frac{2}{K_0^p}\delta \\
&\leq \frac{2\pi}{K_0^p}\left|\frac{\delta}{\pi/2}\right| + \frac{\pi^2}{4K_0}\left|\frac{\delta}{\pi/2}\right|^2 \leq \left(\frac{2\pi}{K_0^p} + \frac{\pi^2}{4K_0}\right)\left|\frac{\delta}{\pi/2}\right| =: K_1\delta,
\end{aligned}$$

where Lagrange's mean value theorem for the function $f(t) = \|(\sin t, \cos t)\|_p$ was utilized, along with the well known facts that $\sin\delta \leq \delta$ and $1-\cos\delta \leq \delta^2$.

Then, for any $\epsilon > 0$ there exists $\delta_\epsilon = \frac{1}{K_1}\frac{\epsilon}{1+\epsilon}$ such that for any θ_k and any $\delta \leq \delta_\epsilon$ condition (3.23) implies (3.24) by $w^*_{N+j} \geq (1-K_1\delta)\|(w^*_{2j-1}, w^*_{2j})\|_p \geq \frac{1}{1+\epsilon}\|(w^*_{2j-1}, w^*_{2j})\|_p$. Hence, no two cuts can be located closer than at an angular distance of δ_ϵ, whereby no more than $\left\lceil \frac{\pi}{2\delta_\epsilon} \right\rceil + 1 = O(\epsilon^{-1})$ cuts can be generated. A stronger result holds for $p = 2$, indeed, observe that in this case (3.25) can be rewritten as

$$
\begin{aligned}
w^*_{N+j} &\geq w^*_{2j-1}\cos\theta_k + w^*_{2j}\sin\theta_k \\
&= \big\|\big(w^*_{2j-1}, w^*_{2j-1}\big)\big\|_2(\cos\theta^*_j\cos\theta_k + \sin\theta^*_j\sin\theta_k) = \big\|\big(w^*_{2j-1}, w^*_{2j-1}\big)\big\|_2\cos\delta.
\end{aligned}
\tag{3.26}
$$

Again, in order for (3.26) to imply (3.24), one has to require that $\cos\delta \geq \frac{1}{1+\epsilon}$, or $\cos\delta_\epsilon = \frac{1}{1+\epsilon}$, which implies $\delta_\epsilon = O(\epsilon^{0.5})$. The statement of the proposition then follows immediately from Proposition 3.3, according to which the approximation accuracy ε of the $(N+1)$-dimensional p-cone satisfies $\varepsilon \leq O(\epsilon)$. □

Remark 12. The cutting plane procedure outlined in Proposition 3.6 can be regarded as an *exact* solution algorithm for pOCP problem (3.13), and, correspondingly, the original pOCP problem (3.1), in the sense that once an approximate solution $\boldsymbol{x}_{\varepsilon_1}$ is obtained with the given accuracy $\varepsilon = \varepsilon_1$, an (improved) solution $\boldsymbol{x}_{\varepsilon_2}$ can subsequently be constructed by setting new accuracy $\varepsilon = \varepsilon_2 < \varepsilon_1$ and resuming the cutting plane algorithm (i.e., the algorithm does not have to be restarted). In contrast, the cutting plane method of Section 3.3.1 in this case would require updating the algorithm itself, namely changing the LP problem (3.18) that is used to generate new cuts.

3.3.3 Fast Cut Generation for Lifted Polyhedral Approximation of Second-Order Cones

In this section we demonstrate that a result analogous to Proposition 3.5 can be formulated in the case of the lifted approximation (3.10) due to Ben-Tal and Nemirovski (2001b), i.e., such an approximation also allows for efficient cut-generation technique.

In accordance to the cutting plane method of Section 3.3.1, consider the master problem (3.17) that corresponds to a polyhedral approximation of the SOCP ($p = 2$) version of problem (3.13), where Ben-Tal and Nemirovski's lifted polyhedral approximation (3.10) of three-dimensional quadratic cones in the "tower-of-variables" is used. In this case, the coefficients $\varsigma_{j,i}$, $\tau_{j,i}$ in (3.17b) are found as the simplex multipliers of the first two constraints of the LP problem

$$z_j^* = \min \quad z \tag{3.27a}$$

$$\text{s. t.} \quad u_0 \geq w_{2j-1}^*, \tag{3.27b}$$

$$v_0 \geq w_{2j}^*, \tag{3.27c}$$

$$u_i = \cos\left(\frac{\pi}{2^{i+1}}\right) u_{i-1} + \sin\left(\frac{\pi}{2^{i+1}}\right) v_{i-1}, \qquad i = 1, \ldots, m, \tag{3.27d}$$

$$v_i \geq \left| -\sin\left(\frac{\pi}{2^{i+1}}\right) u_{i-1} + \cos\left(\frac{\pi}{2^{i+1}}\right) v_{i-1} \right|, \qquad i = 1, \ldots, m, \tag{3.27e}$$

$$u_m \leq z, \tag{3.27f}$$

$$v_m \leq \tan\left(\frac{\pi}{2^{m+1}}\right) u_m, \tag{3.27g}$$

$$\mathbf{u}, \mathbf{v}, z \geq 0,$$

where w_{2j-1}^*, w_{2j}^* are the components of the optimal solution of the master problem obtained during the most recent iteration. If the optimal value of (3.27) satisfies $w_{N+j}^* < z_j^*$, then a new cut of the form (3.17b) is added to the master.

It is important to note that, unlike the gradient polyhedral approximation (3.12) of p-cones, the lifted approximation (3.10) of quadratic cones due to Ben-Tal and Nemirovski is constructed *recursively*, where the parameter m represents the recursion counter and controls approximation accuracy. Intuitively, the process of constructing this lifted approximation of a 3D quadratic cone can be visualized as a sequence of “rotations” and “reflections” in $\mathbb{R}^2$. Given a vector (u_0, v_0) in the positive quadrant of the plane, during the first iteration of the recursion it is rotated clockwise by $\pi/4$ around the origin and, if the rotation puts it into the lower half-plane, it is reflected symmetrically about the horizontal axis, resulting in vector (u_1, v_1) that is again in the positive quadrant. During the second iteration, vector (u_1, v_1) is rotated clockwise by $\pi/8$ and reflected symmetrically about the horizontal axis if it falls into the lower half-plane due to the rotation. The resulting vector is designated (u_2, v_2), and so on.

In view of this, as the first step of constructing a $O(1)$ solution algorithm for the dual of (3.27), we formally show that an optimal solution of (3.27) can be obtained in $O(m)$ time by applying the above recursion procedure to vector (w^*_{2j-1}, w^*_{2j}).

To this end, let us denote by (r_i, α_i) the polar coordinates of the pair (u_i, v_i) in (3.27):

$$r_i = r_i(u_i, v_i) = \|(u_i, v_i)\|_2, \quad \alpha_i = \alpha_i(u_i, v_i) = \arg(u_i, v_i) = \arctan(v_i/u_i).$$

In what follows, we will use notations (u_i, v_i) and (r_i, α_i) interchangeably. Since one can always put $z = u_m$ in (3.27), the discussion of feasibility and optimality in (3.27) reduces to that for the pair of vectors $(\mathbf{u}, \mathbf{v}) = (u_0, \ldots, u_m; v_0, \ldots, v_m)$. First, let us make two observations.

Observation 3.7. *If* $(\mathbf{u}, \mathbf{v})$ *is feasible for* (3.27)*, then* $\alpha_i \le \dfrac{\pi}{2^{i+1}}$ *for* $i = 0, \ldots, m$*.*

Proof: Indeed, if for some i_0 one has $\alpha_{i_0} > \dfrac{\pi}{2^{i_0+1}}$, then by (3.27d)–(3.27e) $\alpha_{i_0+1} > \dfrac{\pi}{2^{i_0+2}}$, which, by continuation, yields a contradiction with (3.27g) that requires $\alpha_m \le \dfrac{\pi}{2^{m+1}}$. □

Observation 3.8. *Given a feasible* $(\mathbf{u}, \mathbf{v})$ *and* $i_0 \in \{1, \ldots, m\}$*, a feasible* $(\tilde{\mathbf{u}}, \tilde{\mathbf{v}})$ *can be constructed that satisfies* $(u_i, v_i) = (\tilde{u}_i, \tilde{v}_i)$ *for* $i \le i_0 - 1$ *and* $(\tilde{r}_i, \tilde{\alpha}_i) = \left(\tilde{r}_{i-1}, \left|\tilde{\alpha}_{i-1} - \frac{\pi}{2^{i+1}}\right|\right)$ *for* $i \ge i_0$.

Proof: For this, we only need to verify that (3.27g) is satisfied for $(\tilde{\mathbf{u}}, \tilde{\mathbf{v}})$. Due to Observation 3.7, one has $\alpha_{i_0-1} \le \dfrac{\pi}{2^{i_0}}$. Thus, by construction $\tilde{\alpha}_{i_0} \le \dfrac{\pi}{2^{i_0+1}}$, $\tilde{\alpha}_{i_0+1} \le \dfrac{\pi}{2^{i_0+2}}, \ldots,$ $\tilde{\alpha}_m \le \dfrac{\pi}{2^{m+1}}$, which is equivalent to (3.27g). □

With this in mind we can construct an optimal solution to the problem under consideration.

Lemma 3.9. *An optimal solution for the problem* (3.27) *can be obtained by setting constraints* (3.27b)–(3.27f) *to equalities, or in other words* $r_0^* = \|(w_{2j-1}^*, w_{2j}^*)\|$, $\alpha_0^* = \arg(u_0, v_0)$*, and* $r_i^* = r_{i-1}^*$, $\alpha_i^* = \left|\alpha_{i-1}^* - \frac{\pi}{2^{i+1}}\right|$ *for* $i = 1, \ldots, m$.

Proof: For a feasible $(\mathbf{u}, \mathbf{v})$, let k be the largest of those $i \in \{1, \ldots, m\}$ for which (3.27e) is a strict inequality i.e., k is such that constraint (3.27e) is non-binding for $i = k$ and binding for $i = k+1, \ldots, m$. Following Observation 3.8 with $i_0 = k$, define a feasible $(\tilde{\mathbf{u}}, \tilde{\mathbf{v}})$ which satisfies

$$\begin{aligned} (\tilde{u}_i, \tilde{v}_i) &= (u_i, v_i), \quad i = 0, \ldots, k-1, \\ (\tilde{r}_k, \tilde{\alpha}_k) &= \left(r_{k-1}, \left|\alpha_{k-1} - \frac{\pi}{2^{k+1}}\right|\right), \\ (\tilde{r}_i, \tilde{\alpha}_i) &= \left(\tilde{r}_{i-1}, \left|\alpha_{i-1} - \frac{\pi}{2^{i+1}}\right|\right), \quad i = k+1, \ldots, m. \end{aligned} \tag{3.28}$$

From the definition of k and (3.28) it follows that $\alpha_k = \tilde{\alpha}_k + \Delta$, where $\Delta > 0$ due to (3.27e). By construction one has

$$r_k = r_{k-1} \frac{\cos \tilde{\alpha}_k}{\cos(\tilde{\alpha}_k + \Delta)} > \tilde{r}_k. \tag{3.29}$$

Now let us demonstrate that $(\tilde{\mathbf{u}}, \tilde{\mathbf{v}})$ yields at least as good objective value as $(\mathbf{u}, \mathbf{v})$, or in other words, $\tilde{u}_m \le u_m$. Note that the definition of k and (3.28) immediately imply that

$$u_m = r_m \cos\alpha_m = r_k \cos\alpha_m, \quad \tilde{u}_m = \tilde{r}_m \cos\tilde{\alpha}_m = \tilde{r}_k \cos\tilde{\alpha}_m, \tag{3.30}$$

and

$$\begin{aligned} \tilde{\alpha}_m &= \left| \frac{\pi}{2^{m+1}} - \left| \frac{\pi}{2^m} - \ldots - \left| \frac{\pi}{2^{k+2}} - \tilde{\alpha}_k \right| \ldots \right| \right|, \\ \alpha_m &= \left| \frac{\pi}{2^{m+1}} - \left| \frac{\pi}{2^m} - \ldots - \left| \frac{\pi}{2^{k+2}} - \alpha_k \right| \ldots \right| \right|. \end{aligned} \tag{3.31}$$

Let us consider three cases:

(a) Assume that $\alpha_k = \tilde{\alpha}_k + \Delta \le \dfrac{\pi}{2^{m+1}}$. In this case equalities (3.31) yield the following expressions for α_m and $\tilde{\alpha}_m$:

$$\tilde{\alpha}_m = \frac{\pi}{2^{m+1}} - \tilde{\alpha}_k, \quad \alpha_m = \frac{\pi}{2^{m+1}} - (\tilde{\alpha}_k + \Delta) < \tilde{\alpha}_m,$$

which upon substitution in (3.30) provide that $u_m \ge \tilde{u}_m$.

(b) Now consider the case of $\tilde{\alpha}_k > \tilde{\alpha}_m$. Successive application of the inequality $||a| - |b|| \le |a - b|$ to the expressions in (3.31) yields that $|\alpha_m - \tilde{\alpha}_m| \le \Delta$, and consequently $\alpha_m \le \tilde{\alpha}_m + \Delta$. Thus, from (3.30) one has $u_m = r_k \cos\alpha_m \ge r_k \cos(\tilde{\alpha}_m + \Delta)$. Upon substituting expression (3.29) for r_k into the last inequality, we obtain

$$u_m \ge r_{k-1} \frac{\cos\tilde{\alpha}_k}{\cos(\tilde{\alpha}_k + \Delta)} \cos(\tilde{\alpha}_m + \Delta) =: f(\Delta).$$

Noting that $f'(\Delta) = r_{k-1} \cos(\tilde{\alpha}_k) \dfrac{\sin(\tilde{\alpha}_k - \tilde{\alpha}_m)}{\cos^2(\tilde{\alpha}_m + \Delta)} > 0$ for $\tilde{\alpha}_k > \tilde{\alpha}_m$ and $f(0) = \tilde{u}_m$, we can conclude that $u_m \ge f(\Delta) \ge f(0) = \tilde{u}_m$.

(c) Finally, suppose that both conditions of (a) and (b) are not satisfied i.e., $\tilde{\alpha}_k \le \tilde{\alpha}_m$ and $\tilde{\alpha}_k + \Delta > \dfrac{\pi}{2^{m+1}}$. Consider, the ratio of u_m and $\tilde{u}_m$ as given by (3.30), where expressions (3.29) and (3.28) are used for r_k and $\tilde{r}_k$, respectively:

$$\frac{u_m}{\tilde{u}_m} = \frac{\cos\tilde{\alpha}_k\ \cos\alpha_m}{\cos\tilde{\alpha}_m\ \cos(\tilde{\alpha}_k + \Delta)}.$$

The above assumption and Observation 3.7 imply that $\tilde{\alpha}_k \leq \tilde{\alpha}_m$ and $\alpha_m \leq \frac{\pi}{2^{m+1}} < \tilde{\alpha}_k + \Delta$, whence the last equality readily yields $u_m/\tilde{u}_m \geq 1$.

In (a)–(c) we have shown that for feasible $(\mathbf{u}, \mathbf{v})$ such that constraint (3.27e) is binding for $i = k+1, \ldots, m$, we can construct a feasible solution with at least as good objective and constraint (3.27e) binding for $i = k, \ldots, m$. Using this claim inductively, we can conclude that for any feasible $(\mathbf{u}, \mathbf{v})$ one can construct a feasible solution for which all constraints in (3.27e) are satisfied as equalities and which has objective at least as good as $(\mathbf{u}, \mathbf{v})$.

Finally, note that a similar argument can be constructed if (3.27b) or (3.27c) are not active. Indeed, the case when $v_0 > w^*_{2j}$ is completely analogous to the case when (3.27e) is not active. Similarly, if $u_0 > w^*_{2j-1}$, which essentially increases the value of r_0 and reduces the value of α_0 by some δ, let us denote as r'_0, α'_m and u'_m the new values of r_0, α_m and u_m corresponding to this case. Then we can observe that $u'_m = r'_0 \cos\alpha'_m > r'_0 \frac{\sin\alpha_0}{\sin(\alpha_0 - \delta)} \cos(\alpha_m - \delta)$. Hence, $\frac{u_m}{u'_m} = \frac{\cos\alpha_m \sin(\alpha_0 - \delta)}{\cos(\alpha_m - \delta)\sin\alpha_0} = \frac{\cos\delta - \cot\alpha_0 \sin\delta}{\cos\delta + \tan\alpha_m \sin\delta} < 1$.

Thus, we can observe that the solution, constructed by setting constraints (3.27b)–(3.27f) to equalities yields at least as good objective value as any other feasible solution.
□

By virtue of Lemma 3.9, the problem of finding optimal of (3.27) is reduced to the following: given $\alpha_0 \in [0, \frac{\pi}{2}]$ and $m \geq 1$, determine α_m from the recurrent relations

$$\alpha_i = \left|\alpha_{i-1} - \frac{\pi}{2^{i+1}}\right|, \quad i = 1, \ldots, m. \tag{3.32}$$

Clearly, this can be done in $O(m)$ time. Below we show that determining α_m from the recursion (3.32) requires $O(1)$ time.

For now, let us assume that $\alpha_0 \neq \dfrac{i\pi}{2^{m+1}}$. For $k = 1, \ldots, 2^m$, define set $A_k^{(m)} = \left(\dfrac{(k-1)\pi}{2^{m+1}}, \dfrac{k\pi}{2^{m+1}}\right)$. Note that by Observation 3.7, $\alpha_m \in A_1^{(m)}$ for any α_0.

Lemma 3.10. *If* $\alpha_0 \in A_k^{(m)}$ *and* α_m *is given by* (3.32)*, then*

$$\alpha_m = \begin{cases} \alpha_0 - \dfrac{(k-1)\pi}{2^{m+1}}, & \text{if } k \text{ is even} \\ \dfrac{k\pi}{2^{m+1}} - \alpha_0, & \text{if } k \text{ is odd.} \end{cases} \tag{3.33}$$

Proof: First, note that, by construction, the recursive formula (3.32) corresponds to the process of rotations and reflections i.e., if we treat α_i as a polar angle, then α_{i+1} is obtained by rotating α_i clockwise by $\dfrac{\pi}{2^{i+1}}$ and then, if the result is in the lower half-plane, reflecting with respect to the horizontal axis. In accordance to (3.32), a reflection is performed whenever $\alpha_{i-1} - \dfrac{\pi}{2^{i+1}} < 0$, therefore for a given α_0 we can define the number of reflections $\xi^{(m)}(\alpha_0)$ as

$$\xi^{(m)}(\alpha_0) = \left|\left\{i \,:\, \alpha_{i-1} - \frac{\pi}{2^{i+1}} < 0\right\}\right|.$$

Next, note that if $\alpha_0, \beta_0 \in A_k^{(m)}$, then $\xi^{(m)}(\alpha_0) = \xi^{(m)}(\beta_0)$ and, moreover, for any i there exists k_i such that $\alpha_i, \beta_i \in A_{k_i}^{(m)}$. Indeed, by the definition of set $A_k^{(m)}$ we have that $\text{sign}\left(\alpha_0 - \frac{\pi}{4}\right) = \text{sign}\left(\beta_0 - \frac{\pi}{4}\right)$ and thus $\alpha_1, \beta_1 \in A_{k_1}^{(m)}$, where $k_1 = k - 2^{m-1}$ if $k \geq 2^{m-1} + 1$ (no reflection) or $k_1 = 2^{m-1} - k + 1$ if $k \leq 2^{m-1}$ (one reflection). Successively repeating this argument we observe that it holds for any i.

Hence, we can define $\xi_k^{(m)}$ as the number of reflections due to (3.32) for $\alpha_0 \in A_k^{(m)}$, or $\xi_k^{(m)} = \xi^{(m)}(\alpha_0)$ for any $\alpha_0 \in A_k^{(m)}$. Let us show that if $\alpha_0 \in A_k^{(m)}$, then

$$\alpha_m = \begin{cases} \alpha_0 - \dfrac{(k-1)\pi}{2^{m+1}}, & \text{if } \xi_k^{(m)} \text{ is even,} \\ \dfrac{k\pi}{2^{m+1}} - \alpha_0, & \text{if } \xi_k^{(m)} \text{ is odd.} \end{cases}$$

Using the identity $|a| = a \operatorname{sign} a$, the recursive representation (3.32) can be written as

$$\alpha_m = \delta_m\Big(\cdots\Big(\delta_2\Big(\delta_1\Big(\alpha_0 - \frac{\pi}{4}\Big) - \frac{\pi}{8}\Big)\cdots - \frac{\pi}{2^{m+1}}\Big) = \alpha_0 \prod_{i=1}^{m} \delta_i - \delta, \tag{3.34}$$

where

$$\delta_i = \operatorname{sign}\Big(\alpha_{i-1} - \frac{\pi}{2^{i+1}}\Big) \quad \text{and} \quad \delta = \sum_{j=1}^{m} \frac{\pi}{2^{j+1}} \prod_{i=j}^{m} \delta_i.$$

According to the arguments given above, $\prod_{i=1}^{m} \delta_i$ and δ should be the same for all $\alpha_0 \in A_k^{(m)}$. Also note that $\prod_{i=1}^{m} \delta_i = \pm 1$, and for all α_0 we should have $\alpha_m \in \Big[0, \dfrac{\pi}{2^{m+1}}\Big]$. Suppose that $\prod_{i=1}^{m} \delta_i = 1$, i.e., $\alpha_m = \alpha_0 - \delta$, which is a linear translation of the interval $\Big[\dfrac{(k-1)\pi}{2^{m+1}}, \dfrac{k\pi}{2^{m+1}}\Big]$. Since the result of the translation should be contained in $\Big[0, \dfrac{\pi}{2^{m+1}}\Big]$, we have that $\delta = \dfrac{(k-1)\pi}{2^{m+1}}$. Similarly, one can conclude that

$$\delta = \begin{cases} \dfrac{(k-1)\pi}{2^{m+1}}, & \text{if } \prod\limits_{i=1}^{m} \delta_i = 1, \\[2ex] -\dfrac{k\pi}{2^{m+1}}, & \text{if } \prod\limits_{i=1}^{m} \delta_i = -1. \end{cases} \tag{3.35}$$

Now, let us show that

$$\big|\xi_j^{(m)} - \xi_{j-1}^{(m)}\big| = 1, \tag{3.36}$$

or, in other words, parity of $\xi_j^{(m)}$ alternates with j. In order to see this, consider the following inductive argument.

(i) Observe that $\xi_1^{(1)} = 1, \xi_2^{(1)} = 0$, i.e., the claim holds for $m = 1$. Indeed, the claim immediately follows from the fact that $\alpha_1 = \left|\alpha_0 - \frac{\pi}{4}\right|$ for $m = 1$.

(ii) Let $m \geq 2$ and $k \leq 2^{m-1}$, then

$$\xi_k^{(m)} = \xi_{2^m-k+1}^{(m)} + 1. \tag{3.37}$$

Indeed, $\alpha_0 \in A_k^{(m)}$ with $k \leq 2^{m-1}$ implies that $\alpha_0 < \frac{k\pi}{2^{m+1}} \leq \frac{\pi}{4}$, and hence $\alpha_1 = \frac{\pi}{4} - \alpha_0$, or, equivalently, $\alpha_1 \in A_{2^{m-1}-k+1}^{(m)}$ with one reflection performed. Similarly, for $\alpha_0 \in A_{2^m-k+1}^{(m)}$ with $k \leq 2^{m-1}$ we have that $\alpha_0 > \frac{(2^m-k)\pi}{2^{m+1}} \geq \frac{\pi}{4}$, whence $\alpha_1 = \alpha_0 - \frac{\pi}{4}$ i.e., $\alpha_1 \in A_{2^m-k+1-2^{m-1}}^{(m)} = A_{2^{m-1}-k+1}^{(m)}$, requiring no reflections. Note that both cases $\alpha_0 \in A_{2^m-k+1}^{(m)}$ and $\alpha_0 \in A_k^{(m)}$ result in $\alpha_1 \in A_{2^{m-1}-k+1}^{(m)}$ with the latter requiring one reflection, which means that $\xi_k^{(m)} = \xi_{2^m-k+1}^{(m)} + 1$.

(iii) Let $m \geq 2$ and $k \geq 2^{m-1} + 1$, then

$$\xi_k^{(m)} = \xi_{k-2^{m-1}}^{(m-1)}. \tag{3.38}$$

Similarly to the above, for $k \geq 2^{m-1} + 1$ and $\alpha_0 \in A_k^{(m)}$ it holds that $\alpha_0 > \frac{(k-1)\pi}{2^{m+1}} \geq \frac{\pi}{4}$, meaning that $\alpha_1 = \alpha_0 - \frac{\pi}{4} \in A_{k-2^{m-1}}^{(m)}$ with no reflections. Rewriting (3.32) as $2\alpha_{i+1} = \left|2\alpha_i - \frac{\pi}{2^{i+1}}\right|$, let $\beta_i = 2\alpha_{i+1}$, whence $\beta_0 = 2\alpha_1$ and $\beta_i = \left|\beta_{i-1} - \frac{\pi}{2^{i+1}}\right|$, $i = 1, \ldots, m-1$. Then, observing that $\beta_0 \in A_{k-2^{m-1}}^{(m-1)}$, it is easy to see that for $k \geq 2^{m-1}+1$, the problem of finding β_{m-1} given $\beta_0 \in A_{k-2^{m-1}}^{(m-1)}$ is equivalent to the problem of determining α_m from $\alpha_0 \in A_k^{(m)}$ and, therefore, $\xi_k^{(m)} = \xi_{k-2^{m-1}}^{(m-1)}$.

(iv) Now, assume that (3.36) holds for some $m \geq 1$ and let us show that it also holds for $m + 1$. To this end, consider the value of $\left|\xi_j^{(m+1)} - \xi_{j-1}^{(m+1)}\right|$: if $j > 2^m + 1$ (i.e., (iii) can be used for both j and $j - 1$), then by (3.38) we have that $\left|\xi_j^{(m+1)} -\right.$

$\xi_{j-1}^{(m+1)}\big| = \big|\xi_{j-2^m}^{(m)} - \xi_{j-1-2^m}^{(m)}\big| = 1$. If $j \leq 2^m$ (i.e., (ii) can be used for both j and $j-1$), then by (3.37) $\big|\xi_j^{(m+1)} - \xi_{j-1}^{(m+1)}\big| = \big|\xi_{2^{m+1}-j+1}^{(m+1)} - \xi_{2^{m+1}-j+2}^{(m+1)}\big|$. By substituting $j' = 2^{m+1} - j + 2$ we have that $\big|\xi_j^{(m+1)} - \xi_{j-1}^{(m+1)}\big| = \big|\xi_{j'}^{(m+1)} - \xi_{j'-1}^{(m+1)}\big|$, where $j' > 2^m + 1$, which reduces to the previous case. Otherwise, if $j = 2^m + 1$, then by (3.37) $\big|\xi_j^{(m+1)} - \xi_{j-1}^{(m+1)}\big| = \big|\xi_j^{(m+1)} - (\xi_j^{(m+1)} + 1)\big| = 1$. Thus, inductively we observe that (3.36) holds for any m.

Finally, by (i) and (3.38) we observe that $\xi_{2^m}^{(m)} = 0$ for all m, and thus by (3.36) $\xi_k^{(m)}$ is even iff k is even. □

Lemma 3.11. *If* $\alpha_0 = \dfrac{k\pi}{2^{m+1}}$, *then the recursive relations* (3.32) *yield*

$$\alpha_m = \begin{cases} 0, & \text{if } k \text{ is odd} \\ \dfrac{\pi}{2^{m+1}}, & \text{if } k \text{ is even.} \end{cases} \tag{3.39}$$

Proof: It is straightforward to see that for $\alpha_0 = \dfrac{\pi}{2}$ recursion (3.32) yields $\alpha_m = \dfrac{\pi}{2^{m+1}}$. Also observe that α_m defined by the recursion (3.32) is continuous with respect to α_0. Let $\alpha_0 = \dfrac{k\pi}{2^{m+1}}$, $k < 2^m$ and consider a strictly monotone sequence $\alpha_0^+(n) \downarrow \alpha_0$ with the corresponding sequence $\alpha_m^+(n)$ obtained by the recursion (3.32). For sufficiently large n we have that $\alpha_0^+(n) \in A_{k+1}^{(m)}$. If k is odd, then by Lemma 3.10 we have that $\alpha_m^+(n) = \alpha_0^+(n) - \dfrac{k\pi}{2^{m+1}} \to 0$, i.e., by continuity of α_m with respect to α_0, such α_0 yields $\alpha_m = 0$. And if k is even, then $\alpha_m^+(n) = \dfrac{(k+1)\pi}{2^{m+1}} - \alpha_0^+(n) \to \dfrac{\pi}{2^{m+1}}$, i.e., $\alpha_m = \dfrac{\pi}{2^{m+1}}$. □

Based on Lemmas 3.9 – 3.11 the following corollary can be formulated.

Corollary 3.12. *An optimal solution of problem* (3.27) *can be constructed in a constant* $O(1)$ *time that does not depend on accuracy of approximation induced by* m. *Particularly,* *if* $\alpha_0 = \arg(w_1, w_2)$ *and* $r_0 = \|(w_1, w_2)\|_2$*, then optimal value of* u_m *can be found as* $u_m = r_0 \cos \alpha_m$*, where*

$$\alpha_m = \begin{cases} \alpha_0 - \dfrac{(k-1)\pi}{2^{m+1}}, & \alpha_0 \in \left(\dfrac{(k-1)\pi}{2^{m+1}}, \dfrac{k\pi}{2^{m+1}}\right] \text{ and } k \text{ is even}, \\ \dfrac{k\pi}{2^{m+1}} - \alpha_0, & \alpha_0 \in \left(\dfrac{(k-1)\pi}{2^{m+1}}, \dfrac{k\pi}{2^{m+1}}\right] \text{ and } k \text{ is odd}, \\ \dfrac{\pi}{2^{m+1}}, & \alpha_0 = 0. \end{cases} \tag{3.40}$$

Now, let us consider the simplex multipliers of (3.27) that yield new cuts. By Lemma 3.9 we can equivalently rewrite the problem as

$$\min \quad u_m, \tag{3.41a}$$

$$\text{s. t.} \quad u_0 = w^*_{2j-1}, \tag{3.41b}$$

$$v_0 = w^*_{2j}, \tag{3.41c}$$

$$u_i = \cos\left(\frac{\pi}{2^{i+1}}\right) u_{i-1} + \sin\left(\frac{\pi}{2^{i+1}}\right) v_{i-1}, \qquad i = 1, \ldots, m, \tag{3.41d}$$

$$v_i = \delta_i\Big(- \sin\left(\frac{\pi}{2^{i+1}}\right) u_{i-1} + \cos\left(\frac{\pi}{2^{i+1}}\right) v_{i-1}\Big), \qquad i = 1, \ldots, m, \tag{3.41e}$$

$$\mathbf{u}, \mathbf{v} \geq \mathbf{0},$$

where

$$\delta_i = \operatorname{sign}\left(- \sin\left(\frac{\pi}{2^{i+1}}\right) u_{i-1} + \cos\left(\frac{\pi}{2^{i+1}}\right) v_{i-1}\right).$$

Note that for given w_1, w_2 these δ_i are constants and coincide with δ_i defined in (3.34). It is easy to see that, by construction, (3.41) has only one feasible point, which is an optimal solution for the initial problem (3.27). Again, we assume that $\delta_i \neq 0$.

Denote by y_i the simplex multipliers for constraints (3.41b) and (3.41d), and by t_i the simplex multipliers for constraints (3.41c) and (3.41e), the dual problem can be formulated as

$$
\begin{aligned}
\max \quad & w_{2j-1}^* y_0 + w_{2j}^* t_0 && && \text{(3.42a)}\\
\text{s. t.} \quad & y_{i-1} - \cos\left(\frac{\pi}{2^{i+1}}\right) y_i + \delta_i \sin\left(\frac{\pi}{2^{i+1}}\right) t_i \le 0, && i = 1, \ldots, m, && \text{(3.42b)}\\
& t_{i-1} - \sin\left(\frac{\pi}{2^{i+1}}\right) y_i - \delta_i \cos\left(\frac{\pi}{2^{i+1}}\right) t_i \le 0, && i = 1, \ldots, m, && \text{(3.42c)}\\
& y_m \le 1, && && \text{(3.42d)}\\
& t_m \le 0. && && \text{(3.42e)}
\end{aligned}
$$

Lemma 3.13. *An optimal solution of* (3.42) *can be found by setting all the constraints to equalities, in which case*

$$
\begin{aligned}
& y_m = 1, \quad t_m = 0, \\
& y_{i-1} = \cos\left(\frac{\pi}{2^{i+1}} + \delta_i \left(\frac{\pi}{2^{i+2}} + \ldots + \delta_{m-1} \frac{\pi}{2^{m+1}}\right) \ldots\right), \qquad i = 1, \ldots, m, \qquad \text{(3.43)}\\
& t_{i-1} = \sin\left(\frac{\pi}{2^{i+1}} + \delta_i \left(\frac{\pi}{2^{i+2}} + \ldots + \delta_{m-1} \frac{\pi}{2^{m+1}}\right) \ldots\right), \qquad i = 1, \ldots, m.
\end{aligned}
$$

Proof: Indeed, let $y_m = 1$, $t_m = 0$ and let us set all the constraints to equalities. Then $y_{m-1} = \cos \frac{\pi}{2^{m+1}}$, $t_{m-1} = \sin \frac{\pi}{2^{m+1}}$. Further, from the elementary trigonometry we obtain that

$$
\begin{aligned}
y_{m-2} = y_{m-1} \cos \frac{\pi}{2^m} - \delta_{m-1} t_{m-1} \sin \frac{\pi}{2^m} = \cos\left(\frac{\pi}{2^m} + \delta_{m-1} \frac{\pi}{2^{m+1}}\right),\\
t_{m-2} = y_{m-1} \sin \frac{\pi}{2^m} + \delta_{m-1} t_{m-1} \cos \frac{\pi}{2^m} = \sin\left(\frac{\pi}{2^m} + \delta_{m-1} \frac{\pi}{2^{m+1}}\right).
\end{aligned}
$$

Inductively we can see that in this case (3.43) holds. Finally, by comparing primal (3.41) and dual (3.42) we observe that by complementary slackness, (3.43) gives an optimal solution for the dual. □

Recall that in order to construct a new cut we need the values of simplex multipliers for constraints (3.27b) and (3.27c) i.e., y_0 and t_0. By Lemma 3.13, one has $y_0 = \cos \gamma$ and

$t_0 = \sin\gamma$, where

$$\gamma = \frac{\pi}{4} + \delta_1\left(\frac{\pi}{8} + \delta_2\left(\frac{\pi}{16} + \ldots + \delta_{m-1}\frac{\pi}{2^{m+1}}\right)\ldots\right).$$

Also note that by duality, $w^*_{2j-1}y_0 + w^*_{2j}t_0 = z^*$, hence $|\gamma - \alpha_0| = \arccos\frac{z^*}{\left\|\left(w^*_{2j-1}, w^*_{2j}\right)\right\|_2}$. Now, by comparing this with Lemma 3.13 and Corollary 3.12 it follows that

$$\gamma = \begin{cases} \alpha_0 - \arccos\frac{z^*}{\left\|\left(w^*_{2j-1}, w^*_{2j}\right)\right\|_2}, & \alpha_0 \in \left(\frac{(k-1)\pi}{2^{m+1}}, \frac{k\pi}{2^{m+1}}\right) \text{ and } k \text{ is even,} \\ \alpha_0 + \arccos\frac{z^*}{\left\|\left(w^*_{2j-1}, w^*_{2j}\right)\right\|_2}, & \alpha_0 \in \left(\frac{(k-1)\pi}{2^{m+1}}, \frac{k\pi}{2^{m+1}}\right) \text{ and } k \text{ is odd.} \end{cases} \tag{3.44}$$

Finally, observe that if $\delta_i = 0$ for some i, then both expressions in (3.44) can be converted into a part of a feasible solution of the dual (3.42) and since they yield the same optimal objective value, any can be taken for cut construction. In such a way, we have shown that the following proposition holds.

Proposition 3.14. *Consider the SOCP version of problem* (3.1) *with K second-order ($p_k =$ 2) cone constraints of dimension $N_k + 1$, and its polyhedral approximation* (3.5) *obtained by reformulating each second-order cone constraint using the "tower-of-variables" representation* (3.8) *and applying Ben-Tal-Nemirovski's lifted polyhedral approximation* (3.10) *with parameter of approximation m to the resulting $N_k - 1$ three-dimensional second-order cones. Then, during an iteration of the cutting plane scheme of Section 3.3.1, new cuts can be generated in a* constant $O\left(\sum_k N_k\right)$ *time that does not depend on m.*

Remark 13. While the statement of Proposition 3.14 parallels that of Proposition 3.5 for gradient polyhedral approximations of p-cones, its significance with respect to Ben-Tal-

Nemirovski's lifted polyhedral approximation of quadratic cones is substantially different, due to the fact that Ben-Tal-Nemirovski's approximation is essentially *recursive* in construction. In this sense, Proposition 3.14 and Lemma 3.10 provide a "shortcut" method for computing this recursion in a constant time that does not depend on the recursion's depth.

Remark 14. While Proposition 3.14 establishes a new approximate solution method for SOCP problems, we do not expect that it will be superior to the existing first-order and interior-point solution approaches, such as proposed in Lan et al. (2011); Lan and Monteiro (2013); Aybat and Iyengar (2012, 2013). As it has already been noticed, methods based on polyhedral approximations do not generally outperform self-dual IP SOCP methods, see Glineur (2000); Krokhmal and Soberanis (2010). However, in our view, computationally, the main advantage of the proposed cutting-plane procedure for lifted polyhedral approximation of SOCP problem is the fact that the resulting problem is a linear program of a moderate size, and thus, extensive body of literature on solving such problems can be utilized. As an example of the case, when such an approach might be advantageous, we study mixed-integer pOCP (MIpOCP) problems (3.2), see Section 3.4.3 for details. The branch-and-bound framework used there asks for repetitive solution of the polyhedral approximation of MIpOCP problem, and thus we can benefit form warm start capabilities of the solvers, since the construction of the cutting-plane procedure ensures that in each node the resulting relaxations are similar to each other.

3.4 Numerical Experiments

3.4.1 Portfolio Optimization with Higher Moment Coherent Risk Measures

3.4.1.1 Higher Moment Coherent Risk Measures

As it has been discussed above, our interest in solving optimization problems with p-order cone constraints stems from the new developments in risk averse decision making under uncertainty and stochastic optimization. Following the framework presented in Chapter 2 we consider a probability space $(\Omega, \mathcal{F}, \mathsf{P})$, and a random outcome X, which is an element of the linear space $\mathcal{L}_p(\Omega, \mathcal{F}, \mathsf{P})$ of $\mathcal{F}$-measurable functions $X : \Omega \mapsto \mathbb{R}$, where $p \geq 1$. Then, a risk measure $\rho(X)$ is defined as a mapping $\rho : \mathcal{L}_p \mapsto \mathbb{R}$. In particular, the higher moment coherent risk (HMCR) measures (Krokhmal (2007)), which we focus on in this case study, have been defined as

$$\mathsf{HMCR}_{p,\alpha}(X) = \min_{\eta \in \mathbb{R}} \quad \eta + (1-\alpha)^{-1} \big\| [X - \eta]_+ \big\|_p, \quad \alpha \in (0,1), \;\; p \geq 1, \tag{3.45}$$

where $[X]_+ = \max\{0, X\}$ and $\|X\|_p = (\mathsf{E}|X|^p)^{1/p}$. By definition, HMCR measures quantify risk in terms of higher tail moments of loss distribution, which are commonly associated with "risk". If, traditionally to stochastic programming, it is assumed that the set Ω is discrete and consists of N scenarios, $\Omega = \{\omega_1, \ldots, \omega_N\}$, with the corresponding probabilities $\varpi_1, \ldots, \varpi_N$, then expressions involving HMCR measures, e.g., $\mathsf{HMCR}_{p,\alpha}(X(\boldsymbol{x}, \omega)) \leq u$, can be implemented via $(N+1)$-dimensional p-order cone constraints. For a detailed discussion of the properties of HMCR measures, see Krokhmal (2007).

3.4.1.2 pOCP Portfolio Optimization Model

In our study, we again consider risk-reward portfolio optimization model. Namely, if just as in Chapter 2 the cost/loss outcome X is defined as the negative rate of return of the portfolio, $X(\boldsymbol{x},\omega) = -\boldsymbol{r}(\omega)^\top \boldsymbol{x}$, where $\boldsymbol{x}$ stands for the vector of portfolio weights, and $\boldsymbol{r} = \boldsymbol{r}(\omega)$ is the uncertain vector of assets' returns. Then, one may formulate the problem of minimizing the portfolio risk as given by HMCR measure, subject to the expected return constraint and the budget constraint as follows:

$$\min_{\boldsymbol{x}\in\mathbb{R}^n_+} \Big\{ \mathsf{HMCR}_{\alpha,p}(-\mathbf{r}^\top\mathbf{x}) \;\Big|\; \mathsf{E}(\mathbf{r}^\top\mathbf{x}) \geq \bar{r},\ \mathbf{1}^\top\mathbf{x} \leq 1 \Big\}, \tag{3.46}$$

where $\bar{r}$ is the prescribed level of expected return, $\boldsymbol{x} \in \mathbb{R}^n_+$ denotes the no-short-selling requirement, and $\mathbf{1} = (1,\ldots,1)^\top$. If $\boldsymbol{r}(\omega)$ is discretely distributed, $\mathsf{P}\{\boldsymbol{r}(\omega) = \boldsymbol{r}_j\} = \varpi_j$, $j = 1,\ldots,N$, then (3.46) reduces to pOCP problem with a single p-order cone constraint:

$$\begin{aligned}
\min \quad & \eta + (1-\alpha)^{-1} t \\
\text{s. t.} \quad & t \geq \|\boldsymbol{w}\|_p, \\
& \mathrm{Diag}\big(\varpi_1^{-1/p},\ldots,\varpi_N^{-1/p}\big)\,\boldsymbol{w} + (\boldsymbol{r}_1,\ldots,\boldsymbol{r}_N)^\top \boldsymbol{x} + \mathbf{1}\eta \;\geq\; \mathbf{0}, \\
& \boldsymbol{x}^\top(\varpi_1\boldsymbol{r}_1 + \ldots + \varpi_N\boldsymbol{r}_N) \geq \bar{r}, \\
& \mathbf{1}^\top\boldsymbol{x} \leq 1, \\
& \boldsymbol{x} \geq \mathbf{0},\ \boldsymbol{w} \geq \mathbf{0},
\end{aligned} \tag{3.47}$$

where $\mathrm{Diag}(a_1,\ldots,a_k)$ denotes the square $k \times k$ matrix whose diagonal elements are equal to $a_1,\ldots,a_k$ and off-diagonal elements are zero.

3.4.1.3 MIpOCP Portfolio Optimization Models

In addition to the convex portfolio optimization model (3.46), we consider two mixed-integer extensions of (3.46). One of them is a cardinality-constrained portfolio optimization problem, which allows for no more than M assets in the portfolio, where M is a given constant:

$$\min_{\boldsymbol{x}\in\mathbb{R}^n_+,\ \boldsymbol{z}\in\{0,1\}^n} \left\{ \mathsf{HMCR}_{\alpha,p}(-\mathbf{r}^\top\mathbf{x}) \,\middle|\, \mathsf{E}(\mathbf{r}^\top\mathbf{x}) \geq \bar{r},\ \ \mathbf{1}^\top\mathbf{x} \leq 1,\ \ \mathbf{x} \leq \mathbf{z},\ \ \mathbf{1}^\top\mathbf{z} \leq M \right\}, \tag{3.48}$$

Similarly to (3.46), formulation (3.48) represents a 0–1 MIpOCP problem with a single conic constraint. In addition, we consider portfolio optimization with lot-buying constraints, which reflect a common real-life trading policy that assets can only be bought in *lots* of shares,for instance, in multiples of 1,000 shares (see, e.g., Perold (1984); Bonami and Lejeune (2009); Scherer and Martin (2005) and references therein). In this case, the portfolio allocation problem can be formulated as MIpOCP with a p-order cone constraint,

$$\min_{\boldsymbol{x}\in\mathbb{R}^n_+,\ \boldsymbol{z}\in\mathbb{Z}^n_+} \left\{ \mathsf{HMCR}_{\alpha,p}(-\mathbf{r}^\top\mathbf{x}) \,\middle|\, \mathsf{E}(\mathbf{r}^\top\mathbf{x}) \geq \bar{r},\quad \mathbf{1}^\top\mathbf{x} \leq 1,\quad \boldsymbol{x} = \frac{L}{C}\,\mathrm{Diag}(\mathbf{p})\,\boldsymbol{z} \right\}, \tag{3.49}$$

where L is the size of the lot, C is the investment capital (in dollars), and vector $\mathbf{p} \in \mathbb{R}^n$ represents the prices of assets.

The following proposition ensures that the introduced portfolio optimization problems with HMCR measures (3.46)–(3.49) are amenable to the polyhedral approximation solution approach discussed in the previous sections.

Proposition 3.15. *If pOCP problem* (3.47) *is feasible, then it satisfies the approximation conditions* (3.7) *of Proposition 3.1. Moreover, the same applies to continuous relaxations of MIpOCP problems* (3.48) *and* (3.49).

Proof: Evidently, the strict feasibility condition (3.7a) can always be satisfied by selecting sufficiently large t and η in (3.47). To see that (3.47) is "semibounded" in the sense (3.7b), note that the only unrestricted variable in the problem is η, but due to the properties of the optimal solution of (3.45) (see Krokhmal (2007)) it can be bounded as $|\eta| \leq \max_{j,\boldsymbol{x}}\{|\boldsymbol{r}_j^\top \boldsymbol{x}|\} \leq \max_j \|\boldsymbol{r}_j\|_\infty$. The same arguments apply to relaxations of (3.48) and (3.49). □

3.4.1.4 Implementation and Scenario Data

We used the LP and Barrier MIP solvers of IBM ILOG CPLEX 12.2 to obtain solutions to the formulated portfolio optimization problems. All problems were coded in C++ and computations ran on a 3GHz PC with 4GB RAM in Windows XP 32bit environment. The additional details of numerical experiments are discussed in the corresponding subsections below.

In both continuous and discrete portfolio optimization problems, we used historical data for n stocks chosen at random from the S&P500 index. Namely, returns over N consequent 10-day periods starting at a (common) randomized date were used to construct the set of N equiprobable scenarios ($\varpi_j = N^{-1}$, $j = 1, \ldots, N$) for the stochastic vector $\mathbf{r}$. The values of parameters L, C, K, α, and $\bar{r}$ were set as follows: $L = 100$, $C = 100{,}000$, $M = 5$, $\alpha = 0.9$, $\bar{r} = 0.005$.

3.4.2 Cutting Plane Techniques for the Lifted and Gradient Approximations of SOCP Problems

The pOCP formulation (3.47) of portfolio selection model (3.46) was used to evaluate the performance of polyhedral approximation-based solution methods discussed in Section 3.3. Particularly, we were interested in comparing the cutting plane methods for solving gradient ($p = 2$) and lifted polyhedral approximations of SOCP problems that were presented in Sections 3.3.2 and 3.3.3, respectively. Recall that the gradient polyhedral approximation, while being applicable to cones of arbitrary order $p \in (1, \infty)$, in the case of $p = 2$ is inferior to Ben-Tal and Nemirovski's lifted polyhedral approximation of second-order cones. At the same time, the results of Sections 3.3.2 and 3.3.3 demonstrate that, in the context of the cutting plane scheme of Section 3.3.1, both types of polyhedral approximations are amenable to generation of cutting planes in a *constant time* that does not depend on the accuracy of approximation. Thus, it was of interest to compare the cutting plane techniques for gradient and lifted approximations of the SOCP ($p = 2$) version of portfolio optimization problem (3.47).

In particular, four types of solution methods were studied. First, the complete LP formulation of Ben-Tal-Nemirovski's lifted polyhedral approximation of problem (3.47) with $p = 2$ was solved using CPLEX 12.2 LP solver (referred to as "LP-lifted" below). Second, this polyhedral approximation LP was solved using the cutting plane method of Section 3.3.1 combined with the fast cut generation technique of Section 3.3.3 (referred to as "CG-lifted").

Third, the SOCP version of (3.47) was solved using the "exact" cutting plane method of Proposition 3.6 (recall that this cutting plane method derives from the corresponding scheme for gradient polyhedral approximation, but does not require a polyhedral approximation problem to be formulated). This method is referred to as "CG-exact".

Lastly, we solved a gradient polyhedral approximation of the SOCP version of (3.47) using the cutting plane method of Section 3.3.1 with the fast cut-generation scheme of Section 3.3.2. The gradient polyhedral approximation was, however, "optimized" in this case to reduce the number of approximating facets as described below, and is referred to as "CG-grad-opt".

Recall that Proposition 3.3 furnishes an expression for the approximation accuracy ε of $(N+1)$-dimensional p-cone provided that each of the three-dimensional p-cones is approximated with the same accuracy ϵ. it can be shown (see, Glineur (2000)) that in the case of the lifted approximation technique due to Ben-Tal and Nemirovski (2001b) applied to second-order cones, the size of polyhedral approximation can be reduced without sacrificing its accuracy ε by properly selecting the accuracies ϵ_i of 3D cone approximations at each level i of the "tower-of-variables". This approach can also be utilized in the case of lifting procedure (3.8) for p-cones,

$$\xi_0 = \xi_{2N-1}, \quad \xi_{N+j} \geq \|(\xi_{2j-1}, \xi_{2j})\|_p, \quad j = 1, \ldots, N-1.$$

Particularly, by introducing approximation accuracies for 3D p-cones at each "level" as

$\epsilon_1, \epsilon_2, ..., \epsilon_\ell$, where $\ell = \lceil \log_2 N \rceil$, one can observe that

$$\xi_0^p = \xi_{2N-1}^p \geq \frac{\xi_{2N-3}^p}{(1+\epsilon_1)^p} + \frac{\xi_{2N-2}^p}{(1+\epsilon_1)^p} \geq \frac{\xi_{2N-7}^p}{(1+\epsilon_1)^p(1+\epsilon_2)^p} + \frac{\xi_{2N-6}^p}{(1+\epsilon_1)^p(1+\epsilon_2)^p}$$
$$+ \frac{\xi_{2N-5}^p}{(1+\epsilon_1)^p(1+\epsilon_2)^p} + \frac{\xi_{2N-4}^p}{(1+\epsilon_1)^p(1+\epsilon_2)^p} \geq \ldots \geq$$
$$\frac{\xi_1^p}{\prod_{i=1}^{k_1}(1+\epsilon_i)^p} + \ldots + \frac{\xi_N^p}{\prod_{i=1}^{k_N}(1+\epsilon_i)^p},$$

where once again $k_i \in \{\lceil \log_2 N \rceil - 1, \lceil \log_2 N \rceil\}$ is the number of "levels" in the "tower of variables" on the way from ξ_{2N-1} to ξ_i. Then, the total number of approximation facets can be reduced by solving the following problem:

$$\min_{m_i \in \mathbb{N}_+} \left\{ \sum_{i=1}^{\ell} q_i m_i \,\middle|\, 1 + \varepsilon \geq \prod_{i=1}^{\ell} \Big(1 + \epsilon_i(m_i)\Big) \right\}, \tag{3.50}$$

where, for a given i, m_i is the number of facets in polyhedral approximation of a 3D p-cone at "level" i, $\epsilon_i = \epsilon_i(m_i)$ is the main term of the corresponding approximation accuracy, and q_i is the number of 3D p-cones thusly approximated. The objective of (3.50) represents the total number of approximation facets, while the constraint ensures that the desired approximation accuracy ε of the multidimensional p-cone is achieved. A feasible solution to (3.50) can be obtained analytically by solving its continuous relaxation with relaxed constraint $\sum_{i=1}^{\ell} \epsilon_i(m_i) \leq \ln(1+\varepsilon)$, and then taking $m_i = \lceil m_i^* \rceil$, where m_i^* is the solution of the relaxed problem. This procedure resulted in, on average, a 30% reduction in the number of approximating facets for the uniform gradient polyhedral approximation.

The results are summarized in Table 3.1, where for each combination of the number of assets n, number of scenarios N, and approximation accuracy ε, the running times are averaged over 20 instances. It has been noted that for the linear programming problems resulting from the lifted approximation, CPLEX Dual Simplex solver performed better for smaller problem instances, while CPLEX Barrier solver – for larger sizes. Thus, we used the Barrier solver for all instances except for the two smaller problem sizes (the first six rows in Table 3.1). At the same time, for the cut-generation approaches we used CPLEX Dual Simplex solver (selected by default).

It follows from Table 3.1 that the cutting plane technique of Sections 3.3.1 and 3.3.3 for solving Ben-Tal-Nemirovski's lifted approximations of SOCP problems ("CG-lifted") offers significant computational improvement over solving the "complete" LP formulation of such approximations ("LP-lifted"). This is consistent with the corresponding findings reported in Krokhmal and Soberanis (2010) for uniform gradient polyhedral approximations of pOCP problems. It is also worth noting that the performance of the cutting plane method of Section 3.3.1 in combination with fast cut generation of Section 3.3.3 ("CG-lifted") is on par with that of the "exact" cutting plane method of Proposition 3.6 ("CG-exact"). However, the cutting plane method of Section 3.3.1 and Section 3.3.3 for gradient polyhedral approximations with reduced number of facets ("CG-grad-opt") generally works slightly faster than the other two cutting plane methods, though the observed improvement is insignificant.

3.4.3 Polyhedral Approximations and Cutting Plane Techniques for Rational-Order Mixed-Integer pOCP Problems

The approaches to constructing and solving polyhedral approximations of pOCP problems (3.1) described above, can also be efficiently applied to mixed-integer extensions of pOCP (MIpOCP) (3.2); in particular, we are considering rational-order MIpOCP prob-

Table 3.1: Average running time (in seconds) for solving portfolio optimization problem (3.46)–(3.47) with $p = 2$. Symbol "$--$" indicates cases when computations exceeded 1 hour time limit, while "$***$" indicates cases for which the solver returned "Out of memory" error.

n, N	ε	LP-lifted	CG-lifted	CG-exact	CG-grad-opt
50, 500	10^{-2}	0.43	0.12	0.11	0.10
	10^{-4}	0.63	0.18	0.17	0.14
	10^{-8}	2.77	0.31	0.32	0.32
150, 1500	10^{-2}	1.83	0.96	0.98	0.89
	10^{-4}	3.85	1.24	1.18	1.09
	10^{-8}	16.29	1.67	1.65	1.64
150, 3000	10^{-2}	37.24	1.66	1.29	1.98
	10^{-4}	96.39	5.80	5.03	5.52
	10^{-8}	296.20	15.11	15.63	15.55
200, 5000	10^{-2}	151.91	9.31	10.20	7.46
	10^{-4}	230.21	23.49	22.76	22.87
	10^{-8}	791.41	48.30	47.48	47.08
200, 10000	10^{-2}	320.80	17.93	18.52	17.26
	10^{-4}	624.63	45.96	46.56	45.09
	10^{-8}	$--$	97.13	96.23	96.97
200, 20000	10^{-2}	677.14	31.56	31.15	30.21
	10^{-4}	898.74	85.95	86.43	84.12
	10^{-8}	$***$	195.99	196.20	195.36

lems, i.e., instances (3.2) where all p_k are rational: $p_k = r_k/s_k$.

In this study of MIpOCP problems (3.2), we follow the approach of Vielma et al. (2008), i.e., instead of solving a nonlinear pOCP relaxation of (3.2) at each node i of the branch-and-bound tree,

$$\begin{aligned} \min \quad & \boldsymbol{c}^\top \boldsymbol{x} + \boldsymbol{d}^\top \boldsymbol{z} \\ \text{s. t.} \quad & \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{z} \le \boldsymbol{b}, \\ & \left\| \boldsymbol{C}^{(k)}\boldsymbol{x} + \boldsymbol{D}^{(k)}\boldsymbol{z} + \boldsymbol{e}^{(k)} \right\|_{p_k} \le \boldsymbol{h}^{(k)\top}\boldsymbol{x} + \boldsymbol{g}^{(k)\top}\boldsymbol{z} + f^{(k)}, \quad k = 1, \ldots, K, \\ & \boldsymbol{x} \in \mathbb{R}^n, \quad \underline{\mathbf{z}}^{(i)} \le \boldsymbol{z} \le \overline{\boldsymbol{z}}^{(i)}, \end{aligned} \tag{3.51}$$

we solve its polyhedral approximation

$$\begin{aligned} \min \quad & \boldsymbol{c}^\top \boldsymbol{x} + \boldsymbol{d}^\top \boldsymbol{z} \\ \text{s. t.} \quad & \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{z} \le \boldsymbol{b}, \\ & \mathbf{H}^{(N_k+1)}_{p_k, m_k} \begin{pmatrix} \boldsymbol{C}^{(k)}\boldsymbol{x} + \mathbf{D}^{(k)}\mathbf{z} + \boldsymbol{e}^{(k)} \\ \boldsymbol{h}^{(k)\top}\boldsymbol{x} + \mathbf{g}^{(k)\top}\boldsymbol{z} + f^{(k)} \\ \mathbf{w}^{(k)} \end{pmatrix} \ge \mathbf{0}, \quad k = 1, \ldots, K, \\ & \boldsymbol{x} \in \mathbb{R}^n, \ \ \underline{\mathbf{z}}^{(i)} \le \boldsymbol{z} \le \overline{\boldsymbol{z}}^{(i)}, \end{aligned} \tag{3.52}$$

where $\underline{\mathbf{z}}^{(i)}$, $\overline{\mathbf{z}}^{(i)}$ are the lower and upper bounds on the relaxed values of variables $\boldsymbol{z}$, and the approximation matrix $\mathbf{H}^{(N_k+1)}_{p_k, m_k}$ is constructed using lifting procedure (3.8) and applying gradient polyhedral approximation (3.12) to the resulting 3D p-cones. In particular, we employ the fast cutting plane scheme for polyhedral gradient approximation presented in Section 3.3.2 to solve the LP problem (3.52) at each node of the tree.

Only when an integer-valued solution of (3.52) is found, in order to check its feasibility with respect to the exact nonlinear formulation (3.2) and declare incumbent or branch further, the exact pOCP relaxation (3.51) of MIpOCP must be solved with bounds on the relaxed values of variables $\boldsymbol{z}$ determined by the integer-valued solution in question (see Vielma et al. (2008) for details). To solve the pOCP relaxation (3.51) exactly, we reformulate (3.51) in the SOCP form by representing p-order cone constraints via a set of second-order cones. Such a representation is available for rational-order cones (see, e.g., Nesterov and Nemirovski (1994); Ben-Tal and Nemirovski (2001a); Alizadeh and Goldfarb (2003)), but it is generally non-unique and requires $O(N \log r)$ three-dimensional rotated quadratic cones to represent $(N+1)$-dimensional p-cone with $p = r/s$ (Krokhmal and Soberanis (2010)). We use the "economical" SOCP representation of rational-order cones due to Morenko et al. (2013), which allows for replacing an (r/s)-cone in $\mathbb{R}^{N+1}$ with exactly $\lceil \log_2 r \rceil N$ quadratic cones; in application to (3.51) with $p_k = r_k/s_k$ it yields a SOCP problem of the form

$$
\begin{aligned}
\min \quad & \boldsymbol{c}^\top \boldsymbol{x} + \boldsymbol{d}^\top \boldsymbol{z} \\
\text{s. t.} \quad & \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{z} \leq \boldsymbol{b}, \\
& \begin{pmatrix} \boldsymbol{C}^{(k)}\boldsymbol{x} + \mathbf{D}^{(k)}\mathbf{z} + \boldsymbol{e}^{(k)} \\ \boldsymbol{h}^{(k)\top}\boldsymbol{x} + \mathbf{g}^{(k)\top}\boldsymbol{z} + f^{(k)} \\ \mathbf{w}^{(k)} \end{pmatrix} \in \mathcal{S}^{N_k}_{r_k/s_k}, \quad k = 1, \ldots, K, \\
& \boldsymbol{x} \in \mathbb{R}^n, \;\; \underline{\mathbf{z}} \leq \boldsymbol{z} \leq \overline{\boldsymbol{z}},
\end{aligned}
\tag{3.53}
$$

where $\mathcal{S}^{N_k}_{r_k/s_k}$ is a set of $N_k \lceil \log_2 r_k \rceil$ "rotated" quadratic three-dimensional cones of the form $\xi_0^2 \leq \xi_1 \xi_2$ that is equivalent to the original (N_k+1)-dimensional p_k-cone. To sum up, the designed branch-and-bound method for MIpOCP primarily relies on the polyhedral approximation of the continuous relaxations with fast cutting plane generation technique, and, additionally, non-linear solver is called for reformulated exact relaxation, when a new incumbent solution is found. Note, that alternatively, the exact approximation algorithm, as described in Proposition 3.6, can be used to solve the exact relaxation for each new incumbent solution, at the same time, this choice is not crucial for the overall performance, since the bulk of the computation time is spent in the nodes of the solution tree, and exact solver is only called sporadically.

The described polyhedral approximation-based approach to solving MIpOCP problems was coded in C++ using CPLEX Concert Technology. In particular, the cutting plane scheme for solving the polyhedral approximation (3.52) of the relaxation (3.51) of the MIpOCP problem was implemented using CPLEX's callback functionality, and the SOCP reformulation (3.53) of (3.51) was solved using CPLEX Barrier solver.

The computational performance of this algorithm (referred to as BnB/CP below) was compared to that of the standard CPLEX 12.2 MIP Barrier solver, which was employed to solve MIpOCP problems in the SOCP reformulation:

$$\begin{aligned} \min \quad & \boldsymbol{c}^\top \boldsymbol{x} + \boldsymbol{d}^\top \boldsymbol{z} \\ \text{s. t.} \quad & \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{z} \le \boldsymbol{b}, \\ & \begin{pmatrix} \boldsymbol{C}^{(k)}\boldsymbol{x} + \mathbf{D}^{(k)}\mathbf{z} + \boldsymbol{e}^{(k)} \\ \boldsymbol{h}^{(k)\top}\boldsymbol{x} + \mathbf{g}^{(k)\top}\boldsymbol{z} + f^{(k)} \\ \mathbf{w}^{(k)} \end{pmatrix} \in \mathcal{S}^{N_k}_{r_k/s_k}, \quad k = 1, \ldots, K, \\ & \boldsymbol{x} \in \mathbb{R}^n, \quad \boldsymbol{z} \in \mathbb{R}^m, \end{aligned}$$

where, as before, $\mathcal{S}^{N_k}_{r_k/s_k}$ denotes the set of second-order cones equivalent to a $(N_k + 1)$-dimensional (r_k/s_k)-cone constructed in accordance to Morenko et al. (2013).

Namely, the BnB/CP algorithm and CPLEX MIP Barrier solver were applied to MIpOCP problems with $p = 3.0$ in the form of portfolio optimization with cardinality constraints (3.48) and lot-buying constraints (3.49) of various sizes (number of integer

Table 3.2: Average running times (in seconds) for BnB/CP implementation of portfolio optimization problem with cardinality constraint (3.48) and $p = 3.0$, benchmarked against IBM ILOG CPLEX 12.2 MIP Barrier solver applied to SOCP reformulation of (3.48). Better running times are highlighted in bold.

	$n = 50$		$n = 100$		$n = 200$	
N	Barrier MIP	BnB/CP	Barrier MIP	BnB/CP	Barrier MIP	BnB/CP
250	**8.43**	11.96	**13.12**	14.56	**21.45**	32.90
500	**11.67**	15.43	37.68	**36.79**	**60.11**	65.87
1000	**12.77**	19.58	38.18	**35.40**	89.36	**75.81**
1500	**33.80**	47.01	107.27	**92.63**	284.44	**190.46**

variables $n = 50, 100, 200$, dimensionality of p-cone $N = 250, \ldots, 1500$). The results are summarized in Tables 3.2 and 3.3, respectively, where the running times are averaged over 20 instances. Observe that in the case of cardinality-constrained portfolio optimization problems, the proposed BnB/CP method is inferior to the standard CPLEX MIP Barrier solver on smaller instances, and outperforms it on larger instances. This trend is confirmed by the numerical experiments on portfolio optimization problems with lot-buying constraints, which are generally harder to solve than the cardinality-constrained problems. In this latter case, the BnB/CP method dominates the standard CPLEX MIP Barrier solver on all problem instances. Moreover, it is important to point out that CPLEX 12.2 employs its own polyhedral approximations of second-order cones for solving MISOCP problems, and the results presented in Tables 3.2 and 3.3 demonstrate the contribution of the proposed fast cutting plane techniques for solving the polyhedral approximations of conic programming problems.

Table 3.3: Average running times (in seconds) for BnB/CP implementation of portfolio optimization problem with lot-buying constraints (3.49) and $p = 3.0$, benchmarked against IBM ILOG CPLEX 12.2 MIP Barrier solver applied to SOCP reformulation of (3.49). Better running times are highlighted in bold, and XX% denotes the integrality gap after 1 hour.

	$n = 50$		$n = 100$		$n = 200$	
N	Barrier MIP	BnB/CP	Barrier MIP	BnB/CP	Barrier MIP	BnB/CP
250	38.46	**27.91**	114.77	**82.92**	1020.84	**743.22**
500	99.41	**55.17**	339.63	**254.41**	2163.89	**1196.76**
1000	586.51	**506.10**	2666.62	**2395.59**	1.99%	**1.18%**

3.5 Concluding Remarks

In this chapter we discussed the use of polyhedral approximations as a solution approach to linear and mixed-integer programming problems with p-order cone constraints. In particular, we showed that the fast cutting-plane method for solving pOCP problems originally proposed by Krokhmal and Soberanis (2010) for a special case of gradient approximation of p-cones, and which allows for cut generation in a constant time independent of the approximation accuracy, can be extended to a broader class of polyhedral approximations. Moreover, a variation of this approach is proposed that constitutes essentially an exact $O(\varepsilon^{-1})$ solution method for nonlinear pOCP problems. In addition, we show that generation of cutting planes in a time that is independent of the approximation accuracy is available for the lifted polyhedral approximation of second-order cones due to Ben-Tal and Nemirovski (2001b), which is itself recursively constructed, with the number of recursion steps being dependent on the desired accuracy. Finally, it is demonstrated that the developed cutting plane techniques can be effectively applied to obtain exact solutions of mixed-integer p-order cone programming problems.

CHAPTER 4
VALID INEQUALITIES FOR P-ORDER CONE PROGRAMMING

4.1 Problem Formulation and Literature Review

In the previous chapter we have shown that involving polyhedral approximation techniques can provide promising computational results when solving mixed-integer p-order cone programming (MIpOCP) problems. In this chapter we will continue this study by exploring possible approaches for generating valid inequalities for such problems.

Two approaches for generating valid inequalities for mixed-integer second-order cone programming (MISOCP) problems have been proposed by Atamtürk and Narayanan (2010, 2011). In the first paper the authors introduce a reformulation of a second order cone constraint using a set of two-dimensional second order cones and then derive valid inequalities for the resulting mixed integer sets. The obtained cuts are termed by the authors conic mixed integer rounding cuts. In Atamtürk and Narayanan (2011), a general lifting procedure for deriving nonlinear conic valid inequalities is proposed and applied to 0-1 MISOCP problems. In a recent work of Belotti et al. (submitted), disjunctive conic cuts for MISOCP problems are introduced. For the case of general convex sets the authors are able to describe the convex hull of the intersection of a convex set and a linear disjunction. And in the particular case of the feasible set of the continuous relaxation of a MISOCP problem they derive a closed-form expression for such a convex hull, thus obtaining a new nonlinear conic cut. Among other approaches to solving mixed integer cone programming problems one can mention the split closure of a strictly convex body by Dadush et al. (2011), lift-and-project algorithm due to Stubbs and Mehrotra (1999), Chvátal-Gomory and disjunctive cuts for 0-1 conic programming by Çezik and Iyengar (2005).

It is worth noting that the vast majority of the existing literature on mixed integer cone programming problems addresses the case of self-dual cones, and particularly second-order cones, with relatively little attention paid to problems involving cones that are not self-dual, as is the case in MIpOCP with $p \in (1, 2) \cup (2, \infty)$. In this chapter, we consider derivation of valid inequalities for mixed integer problems with p-order cone constraints following the techniques of Atamtürk and Narayanan (2010, 2011) proposed for MISOCP. We derive closed form expressions for two families of valid inequalities for MIpOCP problems: mixed integer rounding conic cuts and lifted conic cuts. We also propose to use outer polyhedral approximations as a practical way of employing nonlinear lifted cuts within branch-and-cut framework.

The chapter is organized as follows. In Section 4.2 we present mixed integer rounding cuts for p-cone constrained mixed integer sets. Section 4.3 discusses (nonlinear) lifted cuts for 0-1 and mixed integer p-order cone programming problems. Computational studies of the developed techniques on randomly generated MIpOCP problems as well as portfolio optimization problems with real-life data are discussed in Section 4.4, followed by concluding remarks in Section 4.5.

4.2 Conic Mixed Integer Rounding Cuts for *p*-Order Cones

In this section we present a class of mixed integer rounding cuts for MIpOCP problems arising in the context of risk-averse stochastic optimization. We again consider mixed integer p-order cone programming problem of the form

$$
\begin{aligned}
\min \quad & \mathbf{c}_x^\top \mathbf{x} + \mathbf{c}_y^\top \mathbf{y} \\
\text{s. t.} \quad & \mathbf{D}_x \boldsymbol{x} + \mathbf{D}_y \boldsymbol{y} \le \boldsymbol{d} \\
& \|[\mathbf{A}_j \mathbf{x} + \mathbf{G}_j \mathbf{y} - \mathbf{b}_j]_+\|_{p_j} \le \mathbf{e}_j^\top \mathbf{x} + \mathbf{f}_j^\top \mathbf{y} - h_j, \quad j = 1, \ldots, k \\
& \mathbf{x} \in \mathbb{Z}_+^n, \ \mathbf{y} \in \mathbb{R}_+^q,
\end{aligned}
\tag{4.1}
$$

Operator $[\,\cdot\,]_+$ explicitly accounts for the problem structure induced by downside risk measures such as certainty equivalent measures of risk. For simplicity, we consider the case of a single p-cone constraint in (4.1), $k = 1$. Following the approach of Atamtürk and Narayanan (2010) of constructing mixed integer rounding cuts for problems of type (4.1) with $p = 2$, we rewrite the p-cone constraint in (4.1) as

$$
\begin{aligned}
& t_0 \le \mathbf{e}^\top \mathbf{x} + \mathbf{f}^\top \mathbf{y} - h \\
& t_i \ge [\mathbf{a}_i^\top \mathbf{x} + \mathbf{g}_i^\top \mathbf{y} - b_i]_+, \quad i = 1, \ldots, m \\
& t_0 \ge \|(t_1, \ldots, t_m)\|_p,
\end{aligned}
$$

where $\mathbf{a}_i$ and $\mathbf{g}_i$ denote the i-th rows of matrices $\boldsymbol{A}$ and $\mathbf{G}$, respectively. Then, the task of deriving valid inequalities for the original p-cone mixed integer set in (4.1) can be reduced to obtaining valid inequalities for the polyhedral mixed integer set

$$
T = \big\{\mathbf{x} \in \mathbb{Z}_+^n, \ \mathbf{y} \in \mathbb{R}_+^p, \ t \in \mathbb{R} \ : \ [\mathbf{a}^\top \mathbf{x} + \mathbf{g}^\top \mathbf{y} - b]_+ \le t\big\},
$$

or, without loss of generality, the set

$$
\widetilde{T} = \big\{(y^+, y^-, t, \boldsymbol{x}) \in \mathbb{R}_+^3 \times \mathbb{Z}_+^n \ : \ [\boldsymbol{a}^\top \boldsymbol{x} + y^+ - y^- - b]_+ \le t\big\}. \tag{4.2}
$$

The following two propositions provide an expression for a family of such inequalities.

Proposition 4.1. *For $\alpha \neq 0$, the inequality*

$$\sum_{j=1}^{n} \phi_{f_{|\alpha|}}\left(\frac{a_j}{|\alpha|}\right) x_j - \phi_{f_{|\alpha|}}\left(\frac{b}{|\alpha|}\right) \leq \frac{t + y^-}{|\alpha|}, \tag{4.3}$$

where $f_\alpha = \frac{b}{|\alpha|} - \left\lfloor \frac{b}{|\alpha|} \right\rfloor$ and

$$\phi_f(a) = \begin{cases} (1-f)n, & n \leq a < n+f \\ (1-f)n + (a-n) - f, & n+f \leq a < n+1 \end{cases}$$

is valid for $\widetilde{T}$.

Proposition 4.2. *Inequalities* (4.3) *with $\alpha = a_j$, $j = 1, \ldots, n$, are sufficient to cut off all fractional extreme points of the relaxation of $\widetilde{T}$.*

Proofs of Propositions 4.1 and 4.2 are furnished in the reminder of this section. It is worth noting, however, that since (4.2) is a polyhedral mixed integer set, the derived valid inequalities can also be obtained using the general theory of mixed integer rounding (MIR) inequalities; see, for example, Nemhauser and Wolsey (1988). An advantage of the direct derivation is that it provides a natural way of dealing with continuous variables y^+, y^-, t. Propositions 4.1 and 4.2 justify the usage of inequalities of type (4.3) as cuts in a branch-and-cut procedure and, following Atamtürk and Narayanan (2010), we refer to these inequalities as conic MIR cuts. The results of numerical experiments on utilization of conic MIR cuts (4.3) in MIpOCP problems are presented in Section 4.4.

Following Atamtürk and Narayanan (2010), let us first consider a simple case of the following set

$$T = \big\{(y, w, t, x) \in \mathbb{R}^3_+ \times \mathbb{Z} \,:\, [x + y - w - b]_+ \leq t\big\}.$$

Let us denote by $\text{relax}(T)$ the continuous relaxation of T and by $\text{conv}(T)$ its convex hull. It can be seen that extreme rays of $\text{relax}(T)$ are as follows: $(1,0,0,1)$, $(-1,0,0,0)$, $(1,0,1,0)$, $(-1,1,0,0)$, and its only extreme point is $(b,0,0,0)$. Let us also denote $f = b - \lfloor b \rfloor$. Clearly, the case of $f = 0$ is not interesting, hence it can be assumed that $f > 0$, whereby $\text{conv}(T)$ has four extreme points: $(\lfloor b \rfloor, 0, 0, 0)$, $(\lfloor b \rfloor, f, 0, 0)$, $(\lceil b \rceil, 0, 1-f, 0)$, $(\lceil b \rceil, 0, 0, 1-f)$. With these observations in mind we can formulate the following proposition.

Proposition 4.3. *Inequality*

$$(1-f)(x - \lfloor b \rfloor) \leq t + w \tag{4.4}$$

is valid for T and cuts off all points in $\text{relax}(T) \setminus \text{conv}(T)$.

Proof: First, let us show the validity of (4.4). The base inequality for T is

$$[x + y - w - b]_+ \leq t. \tag{4.5}$$

Now, let $x = \lfloor b \rfloor - \alpha$ and $\alpha \geq 0$. In this case, (4.5) turns into $t \geq [y - w - f - \alpha]_+$ and (4.4) becomes $t \geq -(1-f)\alpha - w$. Observing that $[y - w - f - \alpha]_+ - (-(1-f)\alpha - w) = \max\{y - f - \alpha f, (1-f)\alpha + w\} \geq 0$, one obtains that (4.5) implies (4.4) for $x \leq \lfloor b \rfloor$.

On the other hand, if $x = \lceil b \rceil + \alpha$ with $\alpha \geq 0$, we have (4.5) becomes $t \geq [y - w + (1-f) + \alpha]_+$ and (4.4) turns into $t \geq (1-f)(1+\alpha) - w$. Similarly to above, $[y - w + (1-f) + \alpha]_+ - ((1-f)(1+\alpha) - w) = \max\{y - w + (1-f) + \alpha - (1-f) - \alpha(1-f) + w, w - (1-f)(1+\alpha)\} = \max\{y + \alpha f, w - (1-f)(1+\alpha)\} \geq 0$, which means that (4.5) implies (4.4) for $x \geq \lceil b \rceil$. Hence, (4.4) is valid for T.

To prove the remaining part of the proposition, consider the polyhedron $\acute{T}$ defined by the inequalities

$$x + y - w - b \leq t, \tag{4.6}$$

$$0 \leq t, \tag{4.7}$$

$$0 \leq y, \tag{4.8}$$

$$0 \leq w, \tag{4.9}$$

$$(1 - f)(x - \lfloor b \rfloor) \leq t + w. \tag{4.10}$$

Since $\acute{T}$ has four variables, the basic solutions of $\acute{T}$ are defined by four of these inegualities at equality. They are:

- Inequalities (4.6), (4.7), (4.8), (4.9): $(x, y, w, t) = (b, 0, 0, 0)$ is infeasible if $f \neq 0$.
- Inequalities (4.6), (4.7), (4.8), (4.10): $(x, y, w, t) = (\lceil b \rceil, 0, 1 - f, 0)$.
- Inequalities (4.6), (4.7), (4.9), (4.10): $(x, y, w, t) = (\lfloor b \rfloor, f, 0, 0)$.
- Inequalities (4.6), (4.8), (4.8), (4.10): $(x, y, w, t) = (\lceil b \rceil, 0, 0, 1 - f)$.
- Inequalities (4.7), (4.9), (4.8), (4.10): $(x, y, w, t) = (\lfloor b \rfloor, 0, 0, 0)$.

Hence, $\operatorname{conv}(T)$ has exactly the same extreme points as $\acute{T}$, which completes the proof. □

In the general case, let

$$\widehat{T} = \big\{(y^+, y^-, t, \boldsymbol{x}) \in \mathbb{R}^3_+ \times \mathbb{Z}^n_+ \; : \; [\boldsymbol{a}^\top \boldsymbol{x} + y^+ - y^- - b]_+ \leq t\big\}, \tag{4.11}$$

and consider the following function

$$\phi_f(a) = \begin{cases} (1 - f)n, & n \leq a < n + f \\ (1 - f)n + (a - n) - f, & n + f \leq a < n + 1. \end{cases}$$

Proof: [of Proposition 4.1] First consider the case $\alpha = 1$. We can rewrite the base inequality for (4.11) as

$$\Big[\Big(\sum_{f_j\le f}\lfloor a_j\rfloor x_j + \sum_{f_j>f}\lceil a_j\rceil x_j\Big) + \Big(\sum_{f_j\le f} f_j x_j + y^+\Big) - \Big(\sum_{f_j>f}(1-f_j)x_j + y^-\Big) - b\Big]_+ \le t,$$

where $f_j = a_j - \lfloor a_j\rfloor$. Observe that

$$\acute{x} = \sum_{f_j\le f}\lfloor a_j\rfloor x_j + \sum_{f_j>f}\lceil a_j\rceil x_j \in \mathbb{Z},$$

$$\acute{y} = \sum_{f_j\le f} f_j x_j + y^+ \ge 0,$$

$$\acute{w} = \sum_{f_j>f}(1-f_j)x_j + y^- \ge 0.$$

Hence, we can apply simple conic MIR inequality (4.4) with variables $(\acute{x}, \acute{y}, \acute{w}, t)$:

$$(1-f)\Big(\sum_{f_j\le f}\lfloor a_j\rfloor x_j + \sum_{f_j>f}\lceil a_j\rceil x_j - \lfloor b\rfloor\Big) \le t + \sum_{f_j>f}(1-f_j)x_j + y^-.$$

Rewriting it with the help of function $\phi_f(a)$, we obtain $\sum_{j=1}^n \phi_f(a_j)x_j - \phi_f(b) \le t + y^-$. So, by Proposition 4.3 inequality (4.3) is valid for $\alpha = 1$. In order to see that the result holds for all $\alpha \ne 0$ we only need to scale the base inequality:

$$\Big[\frac{1}{|\alpha|}(\boldsymbol{a}^\top \boldsymbol{x} + y^+ - y^- - b)\Big]_+ \le \frac{t}{|\alpha|}.$$

□

Proof: [of Proposition 4.2] The set $\mathrm{relax}(\widehat{T})$ is defined by $n+3$ variables and $n+4$ constraints. Therefore, if $x_j > 0$ in an extreme point, then the remaining $n+3$ constraints must be active. Thus, the continuous relaxation has at most n fractional extreme points $(x^j, 0, 0, 0)$ of the form $x^j_j = \frac{b}{a_j} > 0$, and $x^j_i = 0$, for $i \ne j$. Such points are infeasible if $\frac{b}{a_j} \notin \mathbb{Z}$. Now, let $a_j > 0$. For such a fractional extreme point inequality (4.3) reduces to

$\phi_{f_{a_j}}(1)x_j - \phi_{f_{a_j}}\left(\frac{b}{a_j}\right) \leq \frac{t+y^-}{a_j}$, or $(1-f_{a_j})x_j - (1-f_{a_j})\left\lfloor \frac{b}{a_j} \right\rfloor \leq \frac{t+y^-}{a_j}$, which by Proposition 4.3 cuts off fractional extreme point with $x_j^j = \frac{b}{a_j}$.

Now, let us consider $a_j < 0$. In this case we observe that the inequality (4.3) reduces to $\phi_{f_{|a_j|}}(-1)x_j - \phi_{f_{|a_j|}}\left(\frac{b}{|a_j|}\right) \leq \frac{t+y^-}{|a_j|}$, or $-(1-f_{|a_j|})x_j - (1-f_{|a_j|})\left\lfloor \frac{b}{|a_j|} \right\rfloor \leq \frac{t+y^-}{|a_j|}$, which again, cuts off fractional extreme point with $x_j^j = \frac{b}{a_j}$.

□

4.3 Lifted Conic Cuts for *p*-Order Cones

4.3.1 General Framework

Atamtürk and Narayanan (2011) have studied lifting for conic mixed integer programming, where a general approach for constructing valid nonlinear conic inequalities for mixed inter conic programming problems was proposed. Namely, consider a general mixed integer conic set

$$S^n(\mathbf{b}) = \left\{ (\mathbf{x}^0, \ldots, \mathbf{x}^n) \in X^0 \times \cdots \times X^n \ : \ \mathbf{b} - \sum_{i=0}^{n} \mathbf{A}^i \mathbf{x}^i \in \mathcal{C} \right\}, \tag{4.12}$$

where $\mathbf{A}^i \in \mathbb{R}^{m \times n_i}$, $\mathbf{b} \in \mathbb{R}^m$, $\mathcal{C}$ is a proper cone (a closed, convex, pointed cone with nonempty interior), and each $X^i \subset \mathbb{R}^{n_i}$ is a mixed integer set. Sets $S^0(\mathbf{b}), \ldots, S^{n-1}(\mathbf{b})$ are restrictions of the set $S^n(\mathbf{b})$. Further, it is assumed that the following conic inequality

$$\mathbf{h} - \mathbf{F}^0 \mathbf{x}^0 \in \mathcal{K},$$

where $\mathcal{K}$ is a proper cone, is known to be valid for the restriction $S^0(\mathbf{b})$. The approach proposed in Atamtürk and Narayanan (2011) is to iteratively find a sequence $\mathbf{F}^1, \ldots, \mathbf{F}^n$, such that

$$\mathbf{h} - \sum_{j=0}^{i} \mathbf{F}^j \mathbf{x}^j \in \mathcal{K} \tag{4.13}$$

is valid for the respective restriction $S^i(\mathbf{b})$ for all i. Such a procedure is called *lifting* and the resulting inequality that is valid for the initial mixed integer set $S^n(\mathbf{b})$ is called *lifted inequality*. In order to determine the values of $\mathbf{F}^1, \ldots, \mathbf{F}^n$, the *lifting set* is introduced for $\mathbf{v} \in \mathbb{R}^m$ as

$$\Phi_i(\mathbf{v}) = \left\{ \mathbf{d} \in \mathbb{R}^s \ : \ \mathbf{h} - \sum_{j=0}^{i} \mathbf{F}^j \mathbf{x}^j - \mathbf{d} \in \mathcal{K} \ \text{ for all } \ (\mathbf{x}^0, \ldots, \mathbf{x}^i)^\top \in S^i(\mathbf{b} - \mathbf{v}) \right\}.$$

Then, a necessary and sufficient condition for (4.13) to be valid can be formulated, which essentially provides a description of the set of valid inequalities.

Proposition 4.4 (Atamtürk and Narayanan (2011)). *Inequality* (4.13) *is valid for* $S^i(\mathbf{b})$ *if and only if* $\mathbf{F}^i\mathbf{t} \in \Phi_i(\mathbf{A}^i\mathbf{t})$ *for all* $\mathbf{t} \in X^i$ *and* $i = 0, \ldots, n$.

The condition established by Proposition 4.4 is still too general to be used to derive expressions for conic cuts. For example, it can be seen that in this way the resulting inequalities are *sequence-dependent*, i.e., a change in the order in which variables $\boldsymbol{x}^i$ are introduced will change the sets $\Phi_i(\mathbf{v})$. The following theorem provides a "sequence-independent" approach to construction of lifting procedure.

Theorem 4.5 (Atamtürk and Narayanan (2011)). *If* $\Upsilon(\mathbf{v}) \subset \Phi_0(\mathbf{v})$ *for all* $\mathbf{v} \in \mathbb{R}^m$ *and* Υ *is superadditive, then* (4.13) *is a lifted valid inequality for* $S^n(\mathbf{b})$ *whenever* $\mathbf{F}^i\mathbf{t} \in \Upsilon(\mathbf{A}^i\mathbf{t})$ *for all* $\mathbf{t} \in X^i$ *and* $i = 0, \ldots, n$.

Then, the following procedure can be formulated for derivation of lifted conic inequalities:

Step 1. Compute $\Phi_0(\mathbf{v})$.

Step 2. If $\Phi_0(\mathbf{v})$ is not superadditive, find a superadditive $\Upsilon(\mathbf{v}) \subset \Phi_0(\mathbf{v})$.

Step 3. For each i find $\mathbf{F}^i$ such that for all $\mathbf{t} \in X^i$: $\mathbf{F}^i\mathbf{t} \in \Upsilon(\mathbf{A}^i\mathbf{t})$ is satisfied.

In Atamtürk and Narayanan (2011) this process was employed to obtain nonlinear lifted conic cuts for 0-1 MISOCP problems; however, no computational results were reported. Below we apply this procedure to derive nonlinear lifted conic cuts for 0-1 and mixed integer p-order cone programming problems with risk-based constraints, and also discuss polyhedral approximations of these cuts that are used in numerical implementation.

4.3.2 Lifting Procedure for 0-1 p-Order Cone Programming Problems

In the case of 0-1 p-order cone programming problem we consider the following conic set

$$S_p^n(b) = \left\{(\boldsymbol{x}, \eta_+, \eta_-, y, t) \in \{0,1\}^n \times \mathbb{R}_+^4 \ : \ \Big[\sum_{i=1}^n a_i x_i + \eta_+ - \eta_- - b\Big]_+^p + y^p \le t^p\right\},$$

where $p \in]1, \infty[$. The set $S_p^n(b)$ represents a relaxation of a high dimensional 0-1 mixed integer p-order conic set: all but one dimensions of the p-cone are aggregated into the term y^p. If necessary, by complementing the binary variables, we can assume that all $a_i \ge 0$. The restriction S_p^0 of this set can be taken as

$$S_p^0(b) = \big\{(x, y, t) \in \{0,1\} \times \mathbb{R}_+^2 \ : \ [x - b]_+^p + y^p \le t^p\big\}.$$

Notice that $S_p^0(b)$ has one extreme point $(b, 0, 0)$, which is fractional when $b \in]0,1[$. Thus, in the only interesting case we have $\lfloor b \rfloor = 0$. Using the results of the previous section, the initial valid inequality can be selected as $\big|(1-f)(x - \lfloor b \rfloor)\big|^p + y^p \le t^p$, where $f = b - \lfloor b \rfloor$ (the fact that this inequality is valid can be verified directly by simply examining the possible values of x, y, t). Now, by definition, in order to compute $\Phi_0(v)$ we need to find such d that inequality

$$\left|(1-f)(x-\lfloor b\rfloor)+d\right|^p + y^p \leq t^p \tag{4.14}$$

is satisfied for all x, y, t such that $[x-b+v]_+^p + y^p \leq t^p$.

Recalling that $\lfloor b\rfloor = 0$ and, therefore, $f = b$, we obtain that (4.14) can be rewritten as $|(1-b)x+d|^p + y^p \leq t^p$ for all x, y, t such that $[x-b+v]_+^p + y^p \leq t^p$. Given that $x \in \{0,1\}$, for $x = 0$ we have $|d| \leq [v-b]_+$, and for $x = 1$ we have $|1-b+d| \leq [1-b+v]_+$. Thus, if $v \geq b$ then $|d| \leq v-b$, and if $v < b$ then $d = 0$, i.e., $|d| \leq [v-b]_+$, whereby $\Phi_0(v) = \{d \ : \ |d| \leq [v-b]_+\}$, which is superaddive. Finally, the following proposition holds.

Proposition 4.6. *Conic inequality*

$$\left|(1-f)(x-\lfloor b\rfloor)+\sum_{i=1}^{n}\alpha_i x_i\right|^p + y^p \leq t^p \tag{4.15}$$

with $\alpha_i = [a_i - b]_+$ *is valid for the set* $S_p^n(b)$.

Proof: Since $\Phi_0(v)$ is superadditive, by Theorem 4.5 we only need to verify that the chosen values of α_i satisfy $\alpha_i x \in \Phi_0(a_i x)$ for $x \in \{0,1\}$, which follows readily from the expression for $\Phi_0(v)$. □

4.3.3 Lifting Procedure for MIpOCP Problems

Similarly, in the case of MIpOCP problem we consider the set

$$S_p^n(b) = \left\{(\boldsymbol{x}, \eta_+, \eta_-, y, t) \in \mathbb{Z}_+^n \times \mathbb{R}_+^4 \ : \ \left[\sum_{i=1}^{n} a_i x_i + \eta_+ - \eta_- - b\right]_+^p + y^p \leq t^p\right\},$$

where $p \in]1,\infty[$. Once again, the set $S_p^n(b)$ represents a relaxation of a high dimensional mixed integer p-order cone constraint. Let us also assume that values x_i are bounded, e.g., $x_i \in \{0,\ldots,M\}$ for all i. Again, without loss of generality let us suppose that $a_i > 0$. The restriction of $S_p^n(b)$ can be selected as

$$S_p^0(b) = \big\{(x,y,t) \in \mathbb{Z}_+ \times \mathbb{R}_+^2 \ : \ [x-b]_+^p + y^p \le t^p\big\}, \tag{4.16}$$

but in this case let us choose a weaker initial valid inequality, $\big[(1-f)(x-\lfloor b\rfloor)\big]_+^p + y^p \le t^p$. The problem of computing $\Phi_0(v)$ is then reduced to the problem of finding values of d such that

$$\big[(1-f)x - \lfloor b\rfloor(1-f) + d\big]_+ \le \big[x-b+v\big]_+. \tag{4.17}$$

Recall that we are only interested in a superadditive subset $\Upsilon(v)$ of such set. One of the possible choices is $\Upsilon(v) = \big\{d \ge 0 \ \big| \ d \le [v - b + \lfloor b\rfloor(1-f)]_+\big\}$. Indeed, $0 \in \Upsilon(v)$ by definition, and (4.17) is a consequence of inequality $(1-f)x - \lfloor b\rfloor(1-f) + d \le x - b + v$, which yields the above expression for $\Upsilon(v)$. Lastly, the following proposition holds.

Proposition 4.7. *Conic inequality*

$$\Big[(1-f)(x-\lfloor b\rfloor) + \sum_{i=1}^{n} \alpha_i x_i\Big]_+^p + y^p \le t^p \tag{4.18}$$

with $\alpha_i = \left[\dfrac{a_i - b + \lfloor b\rfloor(1-f)}{M}\right]_+$ *is valid for* $S_p^n(b)$.

Proof: Indeed, in accordance to Section 4.3.1 it suffices to show that for such a choice of α_i we have $\alpha_i x \in \Upsilon(a_i x)$ for all x. But for $x \ne 0$ we have

$$\Upsilon(a_i x) = \big\{d \ge 0 \ : \ d \le [a_i x - b + \lfloor b\rfloor(1-f)]_+\big\},$$

and

$$\alpha_i x = \left[\frac{a_i - b + \lfloor b \rfloor (1-f)}{M}\right]_+ x \leq [a_i - b + \lfloor b \rfloor (1-f)]_+ \leq [a_i x - b + \lfloor b \rfloor (1-f)]_+.$$

While for $x = 0$, clearly, $0 \in \Upsilon(0)$. □

4.3.4 Polyhedral Approximations of p-Order Cones

Observe that lifted cuts (4.15) and (4.18) for 0-1 and mixed integer p-order cone programming problems, respectively, have the form of p-order cones themselves. Thus, one may expect that while addition of such cuts can reduce the number of nodes explored in the branch-and-bound tree, the computational cost of solving the relaxed problem with extra p-cone constraints at the nodes may increase. In view of this, we propose to replace the nonlinear p-order cone cuts (4.15) and (4.18) with their polyhedral approximations during the branch-and-cut procedure. A detailed discussion of polyhedral approximations of p-order cones has been presented in Chapter 3.

Since in our case the lifted cuts have the form of 3-dimensional p-cones, we use a simple gradient polyhedral approximation. Particularly, a gradient polyhedral approximation for the conic set $\mathcal{K}_p^{(3)} = \{\boldsymbol{\xi} \in \mathbb{R}_+^3 \ : \ \xi_3 \geq \|(\xi_1, \xi_2)\|_p\}$, $p \in]1, \infty[$, can be constructed as

$$\mathcal{H}_{p,\ell}^{(3)} = \big\{\boldsymbol{\xi} \in \mathbb{R}_+^3 \ : \ \xi_3 \geq \alpha_i^{(p)} \xi_1 + \beta_i^{(p)} \xi_2, \ \ i = 0, \ldots, \ell\big\}, \tag{4.19}$$

where

$$\begin{bmatrix} \alpha_i^{(p)} \\ \beta_i^{(p)} \end{bmatrix} = (\cos^p \theta_i + \sin^p \theta_i)^{\frac{1-p}{p}} \begin{bmatrix} \cos^{p-1} \theta_i \\ \sin^{p-1} \theta_i \end{bmatrix}, \quad \theta_i = \frac{\pi i}{2\ell}, \quad i = 0, \ldots, \ell.$$

Here $\mathcal{H}_{p,\ell}^{(3)}$ is an approximation of $\mathcal{K}_p^{(3)}$ in the sense that $\boldsymbol{\xi} \in \mathcal{K}_p^{(3)}$ implies $\boldsymbol{\xi} \in \mathcal{H}_{p,\ell}^{(3)}$, and $\boldsymbol{\xi} \in \mathcal{H}_{p,\ell}^{(3)}$ implies $(1+\varepsilon)\xi_3 \geq \|(\xi_1, \xi_2)\|_p$, where $\varepsilon = \varepsilon(\ell)$ is the accuracy of approximation.

In the case of polyhedral approximation (4.19) it can be estimated as (see Krokhmal and Soberanis (2010) for details):

$$\varepsilon(\ell) \approx \begin{cases} \frac{1}{p}\left(1-\frac{1}{p}\right)^p\left(\frac{\pi}{2\ell}\right)^p, & p \in (1,2), \\ \frac{1}{8}(p-1)\left(\frac{\pi}{2\ell}\right)^2, & p \in (2,\infty). \end{cases}$$

For example, for $p = 4.0$ it suffices to have $\ell = 25$ facets in the approximation to ensure an accuracy of 10^{-3}.

4.4 Numerical Experiments

In this section we report the results of numerical experiments on applying the derived MIR and lifted conic cuts to MIpOCP problem instances. In this case study, three types of problem instances were considered: the first type represents the “generic” MIpOCP instances with randomly generated data, and the second and third types of instances represent two portfolio optimization problems with cardinality constraints and lot-buying constraints, respectively. Historical financial data was used for both types of portfolio optimization problems. A detailed description of each problem type is given below.

Computations were ran on a 3GHz PC with 4GB RAM, and CPLEX 12.2 solver was used. Since CPLEX cannot natively handle p-cone constraints with $p \neq 2$, a second-order cone reformulation (Nesterov and Nemirovski (1994); Alizadeh and Goldfarb (2003); Morenko et al. (2013)) was applied to p-order cone constraints with rational $p > 2$. The derived cuts were added at the root node of the branch-and-bound tree using CPLEX callback routines. In addition, each instance was solved using the default mixed integer CPLEX solver with built-in cuts. In both cases, default solver configuration was used, except the number of threads was limited to one, and QCP relaxations of the model were used at each node.

4.4.1 Problem Formulations

4.4.1.1 Randomly Generated MIpOCP Problems

The first set of problem instances consisted of randomly generated mixed integer p-order cone programming problems of general form. Specifically, the following formulation was used:

$$\begin{aligned} \min \quad & \mathbf{c}^\top \mathbf{x} + y^+ + y^- \\ \text{s. t.} \quad & \left\| [\mathbf{A}\mathbf{x} + y^+\mathbf{1} - y^-\mathbf{1} - \mathbf{b}]_+ \right\|_p \le \mathbf{e}^\top \mathbf{x} + f y^+ - g y^- - h \\ & \boldsymbol{x} \in \mathbb{Z}^n_+, \ y^+, y^- \in \mathbb{R}_+, \end{aligned} \tag{4.20}$$

where $\mathbf{A} \in \mathbb{R}^{n\times m}$, $\mathbf{c}, \mathbf{b}, \mathbf{e} \in \mathbb{R}^n$, $f, g, h \in \mathbb{R}$, and $\mathbf{1} = (1, \dots, 1)^\top$. Each of the parameters $\mathbf{A}, \mathbf{b}, \mathbf{c}, \mathbf{e}, f, g, h$ in (4.20) was selected from the uniform $U(1, 1000)$ distribution.

4.4.1.2 Portfolio Optimization with Cardinality and Lot-Buying Constraints

Similarly to the experiments presented in Chapter 3 two types of portfolio optimization models were considered. Portfolio optimization with cardinality constraints:

$$\min_{\boldsymbol{y}\in\mathbb{R}^n_+,\ \boldsymbol{x}\in\{0,1\}^n} \left\{ \mathsf{HMCR}_{\alpha,p}(-\mathbf{r}^\top\mathbf{y}) \ : \ \mathsf{E}(\mathbf{r}^\top\mathbf{y}) \ge r_0, \ \ \mathbf{1}^\top\mathbf{y} \le 1, \ \ \mathbf{y} \le \mathbf{x}, \ \ \mathbf{1}^\top\mathbf{x} \le K \right\}, \tag{4.21}$$

and lot-buying constraints:

$$\min_{\boldsymbol{y}\in\mathbb{R}^n_+,\ \boldsymbol{x}\in\mathbb{Z}^n_+} \left\{ \mathsf{HMCR}_{\alpha,p}(-\mathbf{r}^\top\mathbf{y}) \ : \ \mathsf{E}(\mathbf{r}^\top\mathbf{y}) \ge r_0, \ \ \mathbf{1}^\top\mathbf{y} \le 1, \ \ \boldsymbol{y} = \frac{L}{C}\,\mathrm{Diag}(\mathbf{p})\,\boldsymbol{x} \right\}. \tag{4.22}$$

In our computations we set $K = 5$, $\alpha = 0.9$, $L = 1{,}000$ and $C = \$100{,}000$.

For portfolio optimization problems, we used historical data for n stocks chosen at random from the S&P500 index, and returns over m consequent 10-day periods starting at a (common) randomized date were used to construct the set of m scenarios for the stochastic vector $\mathbf{r}$ in (4.21), (4.22).

4.4.2 Discussion of Results: Conic MIR Cuts

4.4.2.1 Randomly Generated MIpOCP Problems

For each pair of parameters (n, m) that determine the number of integer variables and the dimensionality of p-cone, 50 randomly generated instances of problem (4.20) were solved. The results are summarized in Table 4.1, where the average computational time (in seconds), the average number of nodes explored in the search tree, and the average number of cuts added during the solution procedure are reported. In addition, we report the percentage of cases in which addition of conic MIR cuts improves the computational time and the number of nodes explored, respectively, as compared to the default CPLEX routines. It has also been noted that randomly generated problems are relatively easy to solve; in fact, many instances were solved at the root node. Therefore, in addition to the results averaged over all instances of a given problem size (n, m), Table 4.1 presents the results averaged over "difficult" instances, i.e., instances that could not be solved at the root node by CPLEX solver with default parameter settings. As one can see, in most cases utilization of conic MIR cuts reduces the average solution time and the number of nodes explored in the solution tree, with the improvement being more noticeable for difficult instances and larger sizes of the problem. It is also worth noting that while solution times vary for different values of the parameter p, the observed improvement due to implementation of conic MIR cuts stays approximately the same.

4.4.2.2 Portfolio Optimization with Cardinality Constraints

For each problem size we generated 30 problem instances. The obtained results are summarized in Table 4.2. We can again conclude that for the majority of the instances, introduction of conic MIR cuts leads to an improved performance in comparison to the default CPLEX solution procedures, although the improvement is considerably smaller comparing to that observed on randomly generated problems. Note also that a significantly smaller number of cuts were generated in problem instances of this type; moreover, in many cases the default CPLEX optimizer did not add any cuts to the problem.

4.4.2.3 Portfolio Optimization with Lot-Buying Constraints

The results averaged over 30 instances for each problem size are summarized in Table 4.3. Note that in many instances of problems of this type, no user cuts of the proposed structure have been found. It can also be noted that regardless of the number of cuts found, solution times are rather comparable to those of the default CPLEX optimizer, which may indicate that conic MIR cuts do not make a significant difference in problems of this type.

4.4.3 Discussion of Results: Lifted Conic Cuts

4.4.3.1 Portfolio Optimization

For evaluation of the performance of lifted cuts derived in Section 4.3, we used both types of portfolio optimization problems, with parameters set up as described above. As it has been already noted, each lifted nonlinear cut was replaced by its outer gradient polyhedral approximation. Specifically, the approximation accuracy was set at 10^{-3}. Since in this case each cut results in multiple additional linear constraints, we restricted the number of lifted cuts to be added at the root node to two. The results obtained for portfolio optimization problems with cardinality constraints (4.21) and lot-buying constraints (4.22), each averaged over 30 problem instances, are summarized in Tables 4.2 and 4.3, respectively. We observed similar improvements in computational time for both types of problems. Also, it has been observed that utilization of lifted cuts in portfolio optimization with lot-buying constraints does not generally lead to a reduction in the number of nodes explored in the solution tree. Thus, based on this observation and results of the experiments of the previous section, we can suggest that the observed improvement is probably partially due to considerably less time spent while looking for cuts. In contrast, in portfolio problems with cardinality constraints we observe reductions in both the number of nodes and solution times due to utilization of lifted cuts.

4.5 Concluding Remarks

The recent progress in solving mixed integer programming problems can partially be attributed to the advances in utilization of valid inequalities for integer and mixed integer sets. Mixed integer cuts allow for tightening of the bounds given by the continuous relaxation of the problem during the branch-and-cut procedure and, as a result, can lead to reductions in the number of nodes explored in the branch-and-bound tree and in the overall computational time. Typically, valid inequalities exploit specific structure of the feasible set of the problem. This paper presents two families of valid inequalities for mixed integer p-order programming problems that arise in risk-averse stochastic optimization with downside risk measures. Particularly, we developed mixed integer rounding cuts and nonlinear lifted cuts for mixed integer p-order conic sets, extending the corresponding results for mixed integer second order programming problems by Atamtürk and Narayanan (2010, 2011). Computational studies on randomly generated problems as well as discrete portfolio optimization problems with historical data demonstrate that both conic MIR cuts and lifted conic cuts lead to improved solution times. In general, nonlinear cuts are not yet as prevalent as linear ones, partly due to the fact that additional nonlinear inequalities in the bounding (relaxed) problem tend to have deteriorating effect on the computational time of branch-and-bound procedure. In order to improve the computational tractability of the derived nonlinear lifted cuts within the branch-and-cut framework, we proposed replacing them with their polyhedral approximations; since the nonlinear lifted cuts constitute low-dimensional p-cones, the corresponding polyhedral approximations are relatively inexpensive. In this respect, our computational results are among the first successful applications of nonlinear cuts in nonlinear mixed integer programming problems.

Table 4.1: Performance of conic MIR cuts for randomly generated MIpOCP problems. The colomn "% better" represents the percentage of problem instances for which conic MIR cuts approach outperformed CPLEX with default parameters in terms of solution time and number of nodes respectively. "Difficult" instances are problem instances which cannot be solved in the root node.

					$p = 2.0$			
			all instances			"difficult" instances		
n	m		default CPLEX	conic MIR	% better	default CPLEX	conic MIR	% better
500	200	time	26.88	22.88	29.41%	58.22	43.77	61.11%
		nodes	2.0	0.75	100.00%	5.67	2.11	100.0%
		cuts	16.74	48.65	–	16.06	50.94	–
	600	time	218.0	224.72	52.83%	356.27	369.85	67.86%
		nodes	3.34	3.17	92.45%	6.32	6.0	85.71%
		cuts	73.45	53.90	–	19.08	55.82	–
	1000	time	1117.45	856.59	45.61%	2045.46	1418.66	65.22%
		nodes	1.68	0.60	96.49%	4.17	1.48	91.30%
		cuts	102.54	63.40	–	76.00	50.87	–
					$p = 3.0$			
			all instances			"difficult" instances		
n	m		default CPLEX	conic MIR	% better	default CPLEX	conic MIR	% better
500	200	time	12.60	11.10	37.25%	24.11	20.68	76.92%
		nodes	0.88	0.31	100.00%	1.23	3.46	100.0%
		cuts	11.71	49.65	–	11.38	50.94	–
	600	time	189.76	71.90	51.92%	421.64	133.0	87.50%
		nodes	6.92	2.13	100.00%	22.94	7.06	100.00%
		cuts	18.92	54.58	–	15.37	48.26	–
	1000	time	910.04	560.12	66.67%	1741.93	974.53	61.90%
		nodes	1.53	0.35	98.25%	4.14	0.95	95.24%
		cuts	32.81	63.40	–	22.0	50.87	–
					$p = 4.0$			
			all instances			"difficult" instances		
n	m		default CPLEX	conic MIR	% better	default CPLEX	conic MIR	% better
500	200	time	31.92	26.54	35.29%	62.04	48.06	52.17%
		nodes	2.29	0.98	98.04%	5.09	2.17	95.65%
		cuts	26.16	48.65	–	29.17	63.83	–
	600	time	582.88	324.86	43.40%	875.88	471.92	55.88%
		nodes	9.25	8.0	88.84%	14.41	12.47	82.36%
		cuts	76.75	53.91	–	37.87	60.01	–

Table 4.2: Performance of conic cuts for cardinality constrained portfolio optimization problems. Entries in bold correspond to the minimum solution time for each row.

		$p = 2.0$								
		default CPLEX			conic MIR cuts			lifted conic cuts		
n	m	time	nodes	cuts	time	nodes	cuts	time	nodes	cuts
100	600	360.97	31.31	0.10	315.98	31.90	3.00	**281.34**	30.59	2.00
	1000	787.16	31.15	0.00	772.44	77.90	3.00	**595.66**	30.77	2.00
	1400	916.18	37.58	0.00	766.14	55.50	3.00	**664.73**	25.8	2.00
150	600	446.11	41.8	0.00	400.02	41.20	3.00	**377.87**	40.2	2.00
	1000	1566.79	53.44	0.00	1436.57	53.20	3.00	**1326.74**	52.33	2.00
	1400	2601.84	40.69	0.00	2343.03	38.83	3.00	**2196.61**	39.92	2.00

		$p = 3.0$								
		default CPLEX			conic MIR cuts			lifted conic cuts		
n	m	time	nodes	cuts	time	nodes	cuts	time	nodes	cuts
100	600	813.62	47.93	0	**537.14**	45.63	3.00	610.98	45.35	2.00
	1000	1449.75	49.78	0	1216.24	49.90	3.00	**1213.02**	49.67	2.00
	1400	1671.64	36.38	0	1518.44	59.87	3.00	**1428.81**	40.2	2.00
150	600	488.07	41.4	0.2	415.92	40.67	3.00	**354.40**	39.8	2.00
	1000	2877.30	80.81	0.05	2661.90	83.87	3.00	**2514.82**	86.71	2.00
	1400	4307.80	70.72	0.11	4006.54	70.43	3.00	**3739.91**	69.89	2.00

		$p = 4.0$								
		default CPLEX			conic MIR cuts			lifted conic cuts		
n	m	time	nodes	cuts	time	nodes	cuts	time	nodes	cuts
100	600	1234.58	47.08	0.10	1186.99	45.83	3.00	**1062.46**	45.58	2.00
	1000	2368.82	45.05	0.00	2204.83	48.20	3.00	**2062.06**	47.87	2.00
	1400	3243.04	33.49	0.00	2630.18	34.40	3.00	**2552.70**	31.48	2.00
150	600	435.52	34.50	0.17	371.95	58.65	3.00	**340.62**	33.33	2.00
	1000	5913.61	94.71	0.00	5451.90	47.95	3.00	**5168.28**	97.57	2.00
	1400	6442.82	62.50	0.05	6087.91	31.30	3.00	**5286.47**	62.85	2.00

Table 4.3: Performance of conic cuts for lot-buying constrained portfolio optimization problems. Entries in bold correspond to the minimum solution time for each row.

$p = 2.0$										
		default CPLEX			conic MIR cuts			lifted conic cuts		
n	m	time	nodes	cuts	time	nodes	cuts	time	nodes	cuts
10	200	9.09	4.13	1.50	9.59	5.10	0.00	**8.03**	5.31	2.00
	600	45.53	4.67	2.61	40.08	5.57	0.13	**32.98**	6.17	2.00
	1000	117.78	11.47	2.37	111.44	13.97	0.33	**102.81**	14.74	2.00
20	200	42.49	20.79	3.64	37.17	23.13	0.40	**32.00**	25.36	2.00
	600	103.28	12.80	5.00	101.67	16.93	0.13	**94.96**	20.16	2.00
	1000	188.04	13.63	3.19	177.53	13.83	1.10	**168.88**	13.63	2.00
50	200	54.50	42.94	4.38	51.21	45.40	0.50	**46.55**	47.44	2.00
	600	307.66	33.19	6.19	286.28	41.27	1.50	**268.13**	46.75	2.00
	1000	640.82	49.71	3.71	**635.54**	62.03	0.00	664.29	69.35	2.00

$p = 3.0$										
		default CPLEX			conic MIR cuts			lifted conic cuts		
n	m	time	nodes	cuts	time	nodes	cuts	time	nodes	cuts
10	200	18.56	4.79	3.57	17.33	7.73	0.03	**15.50**	9.50	2.00
	600	49.60	8.33	2.22	42.32	9.73	0.03	**34.46**	10.39	2.00
	1000	96.15	10.19	2.38	94.93	12.97	0.03	**90.25**	15.38	2.00
20	200	34.05	9.06	3.11	27.11	10.97	1.10	**21.23**	12.00	2.00
	600	96.98	9.51	4.22	79.78	12.00	1.10	**66.74**	13.84	2.00
	1000	**130.59**	4.53	4.35	134.93	4.67	1.23	141.49	4.53	2.00
50	200	78.29	30.55	5.10	70.07	35.93	0.03	**57.25**	39.95	2.00
	600	316.89	37.39	5.33	275.04	38.17	0.03	**210.81**	37.67	2.00
	1000	540.25	22.58	5.37	500.46	36.87	1.00	**459.55**	47.74	2.00

$p = 4.0$										
		default CPLEX			conic MIR cuts			lifted conic cuts		
n	m	time	nodes	cuts	time	nodes	cuts	time	nodes	cuts
10	200	23.29	6.29	2.29	17.93	6.13	2.00	**13.58**	5.71	2.00
	600	44.50	3.57	2.21	41.56	3.93	7.03	**37.73**	4.21	2.00
	1000	**122.08**	8.00	2.29	123.10	10.13	25.03	125.04	12.71	2.00
20	200	49.11	7.93	4.07	43.88	16.07	0.13	**40.19**	20.40	2.00
	600	110.42	16.47	3.31	101.32	18.00	12.50	**89.95**	18.24	2.00
	1000	315.87	10.89	4.94	279.44	11.10	34.23	**256.45**	10.89	2.00
50	200	127.20	43.78	5.17	118.54	46.67	0.46	**112.06**	48.06	2.00
	600	416.48	36.76	4.68	344.87	33.93	21.40	**294.47**	29.32	2.00
	1000	993.53	44.50	5.71	825.43	46.20	33.17	**682.21**	56.59	2.00

CHAPTER 5
A SCENARIO DECOMPOSITION ALGORITHM FOR STOCHASTIC PROGRAMING PROBLEMS WITH CERTAINTY EQUIVALENT MEASURES

5.1 Introduction and Motivation

In this section, we propose an efficient algorithm for solving large-scale stochastic optimization problems with a class of "downside", or "tail" risk measures that have been proposed in Chapter 2. The presented scenario decomposition algorithm exploits the special structure of the feasible set induced by the respective risk measures as well as the properties common to the considered class of risk functionals. As an illustrative example of the general approach, we consider stochastic optimization problems with both higher-moment coherent risk measures (HMCR) and log-exponential convex risk (LogExpCR) measures.

Perhaps, the most frequently implemented risk measure in stochastic programming problems is the well known Conditional Value-at-Risk (CVaR) (Rockafellar and Uryasev, 2000, 2002). When X is piecewise linear in $\boldsymbol{x}$ and set C is polyhedral, stochastic program with CVaR objective or constraints reduces to a linear programming (LP) problem. Several recent studies addressed the solution efficiency of LPs with CVaR objectives or constraints for cases when the number of scenarios is large. Lim, Sherali, and Uryasev (2010) noted that stochastic program in this case may be viewed as a nondifferentiable optimization problem and implemented a two-phase solution approach to solve large-scale instances. In the first phase, they exploit descent-based optimization techniques to circumvent nondifferentiable points by perturbing the solution to differentiable solutions within their “relative neighborhood”. The second phase employs a deflecting subgradient search direction with a step size established by an adequate target value. They further extended this approach with a third phase that resorts to the simplex algorithm after achieving convergence by employing an advanced crash-basis dependent on solutions obtained from the first two phases.

Künzi-Bay and Mayer (2006) developed a solution technique for the problem , with measure ρ chosen as the CVaR, that utilized a specialized L-shaped method after reformulating it as a two-stage stochastic programming problem. However, Subramanian and Huang (2008) noted that the problem structure does not naturally conform to the characteristics of a two-stage stochastic program and introduced a polyhedral reformulation of the CVaR constraint with a statistics based CVaR estimator to solve a closely related version of the problem. In a followup study (Subramanian and Huang, 2009), they retained Value-at-Risk (VaR) and CVaR as unknown variables in the CVaR constraints, enabling a more efficient decomposition algorithm, as opposed to Klein Haneveld and van der Vlerk (2006), where the problem was solved as a canonical integrated chance constraint problem with preceding estimates of VaR. Espinoza and Moreno (2012) proposed a solution method that entailed generation of aggregated scenario constraints to form smaller relaxation problems whose optimal outcomes were then used to directly evaluate the respective upper bound on the objective of the original problem.

In what follows, we develop a general scenario decomposition solution framework for solving stochastic optimization problems with certainty equivalent-based risk measures by utilizing principles related to those in Espinoza and Moreno (2012). The rest of the chapter is organized as follows. In Section 5.2 we propose the scenario decomposition algorithm for stochastic programming problems with structure that is induced by the risk measures described in Chapter 2. Then, experimental studies on portfolio optimization problems with large-scale data sets that demonstrate the effectiveness of the developed technique are presented in Section 5.3 . In the reminder of this section we discuss the implementation of the risk measures discussed above in mathematical programming problems.

Given a discrete set of scenarios $\{\omega_1, \dots, \omega_N\} = \Omega$ that induce cost or loss outcomes $X(\mathbf{x}, \omega_1), \dots, X(\mathbf{x}, \omega_N)$ for any given decision vector **x**, it is easy to see that the risk constraint can be represented by the following set of inequalities:

$$\eta + (1-\alpha)^{-1} w_0 \leq h(\boldsymbol{x}), \tag{5.1a}$$

$$w_0 \geq v^{-1}\Bigg(\sum_{j\in\mathcal{N}} \pi_j v(w_j)\Bigg), \tag{5.1b}$$

$$w_j \geq X(\boldsymbol{x}, \omega_j) - \eta, \quad j \in \mathcal{N}, \tag{5.1c}$$

$$w_j \geq 0, \quad j \in \mathcal{N}, \tag{5.1d}$$

where $\mathbb{N}$ denotes the set of scenario indices, $\mathbb{N} = \{1, \dots, N\}$, and $\pi_j = \mathsf{P}(\omega_j) > 0$ represent the corresponding scenario probabilities that satisfy $\pi_1 + \cdots + \pi_N = 1$. Throughout the paper we will also assume that function v satisfies the following assumption:

(U1) *Function $v(t)$ is continuously differentiable, increasing, convex, and, moreover, such that $v(0) = 0$ and the certainty equivalent $v^{-1}\mathsf{E}v(X)$ is convex in X.*

In the above discussion it was shown that several types of risk measures emerge from different choices of the deutility function v. Here we note that the corresponding representations of constraint (5.1b) in the context of HMCR and LogExpCR measures lead to sufficiently "nice", i.e., convex, mathematical programming models. For HMCR measures inequality (5.1b) becomes

$$w_0 \geq \Big(\sum\nolimits_{j\in\mathcal{N}} \pi_j w_j^p\Big)^{1/p}, \tag{5.2}$$

which is equivalent to a standard p-order cone under affine scaling. Noteworthy instances of (5.2) for which readily available mathematical programming solution methods exist include $p = 1, 2$. In the particular case of $p = 1$, which corresponds to CVaR, the problem (5.1) reduces to a linear programming (LP) model. For instances when $p = 2$, a second-order cone programming (SOCP) model that is efficiently solvable using long-step self-dual interior point methods transpires. However, no similarly efficient solution methods exist for solving p-order conic constrained problems when $p \in (1, 2)\cup(2, \infty)$ due to the fact that the p-cone is not self-dual in this case. Additional discussion and computational considerations for such instances are given in Section 5.3.1. Lastly, the following exponential inequality corresponds to constraint (5.1b) when ρ is a LogExpCR measure:

$$w_0 \geq \ln \sum\nolimits_{j\in\mathcal{N}} \pi_j e^{w_j}, \tag{5.3}$$

which is also convex and allows for the resulting optimization problem to be solved using appropriate (e.g., interior point) methods.

5.2 Scenario Decomposition Algorithm

Large scale stochastic optimization models with CVaR measure and the corresponding solution algorithms have received considerable attention in the literature. In this section we propose an efficient scenario decomposition algorithm for solving large-scale mathematical programming problems that use certainty equivalent-based risk measures, which contain CVaR as a special case.

The algorithm relies on solving a series of relaxation problems containing linear combinations of scenario-based constraints that are systematically decomposed until an optimal solution of the original problem is found or the problem is proven to be infeasible. Naturally, the core assumption behind such a scheme is that sequential solutions of smaller relaxation problems can be achieved within shorter computation times. When the distribution of loss function $X(\boldsymbol{x}, \omega)$ has a finite support (scenario set) $\Omega = \{\omega_1, \ldots, \omega_N\}$ with probabilities $\mathsf{P}(\omega_j) = \pi_j > 0$, the stochastic programming problem with risk constraint admits the form

$$\min \quad g(\boldsymbol{x}) \tag{5.4a}$$

$$\text{s. t.} \quad \boldsymbol{x} \in C, \tag{5.4b}$$

$$\eta + (1-\alpha)^{-1} w_0 \le h(\boldsymbol{x}), \tag{5.4c}$$

$$w_0 \ge v^{-1}\Bigg(\sum_{j \in \mathcal{N}} \pi_j v(w_j)\Bigg), \tag{5.4d}$$

$$w_j \ge \; X(\mathbf{x}, \omega_j) - \eta, \quad j \in \mathcal{N}, \tag{5.4e}$$

$$w_j \ge 0, \quad j \in \mathcal{N}, \tag{5.4f}$$

where $\mathbb{N} = \{1, \ldots, N\}$. If we assume that function $g(\boldsymbol{x})$ and feasible set C are "nice" in the sense that problem $\min\{g(\boldsymbol{x}) : \boldsymbol{x} \in C\}$ admits efficient solution methods, then formulation (5.4) may present challenges that are two-fold. First, constraint (5.4d) may need a specialized solution approach, especially in the case of large N. Similarly, when N is large, computational difficulties may be associated with handling the large number of constraints (5.4e)–(5.4f). In this work we present an iterative procedure for dealing with a large number of scenario-based inequalities (5.4e)–(5.4f).

Since the original problem (5.4) with many constraints of the form (5.4e)–(5.4f) may be hard solve, a relaxation of (5.4) can be constructed by aggregating some of the scenario constraints. Let $\{\mathcal{S}_k : k \in \mathcal{K}\}$ denote a *partition* of the set $\mathbb{N}$ of scenario indices (which we will simply call scenario set), i.e.,

$$\bigcup_{k\in\mathcal{K}} \mathcal{S}_k = \mathbb{N}, \quad \mathcal{S}_i \cap \mathcal{S}_j = \emptyset \quad \text{for all} \quad i, j \in \mathcal{K},\ i \neq j.$$

The aggregation of scenario constraints by adding inequalities (5.4e) within sets $\mathcal{S}_k$ produces the following *master problem*:

$$\min \quad g(\boldsymbol{x}) \tag{5.5a}$$

$$\text{s. t.} \quad \boldsymbol{x} \in C, \tag{5.5b}$$

$$\eta + (1-\alpha)^{-1} w_0 \leq h(\boldsymbol{x}), \tag{5.5c}$$

$$w_0 \geq v^{-1}\Bigg(\sum_{j\in\mathcal{N}} \pi_j v(w_j)\Bigg), \tag{5.5d}$$

$$\sum_{j\in\mathcal{S}_k} w_j \geq \sum_{j\in\mathcal{S}_k} X(\boldsymbol{x}, \omega_j) - |\mathcal{S}_k|\eta, \quad k \in \mathcal{K}, \tag{5.5e}$$

$$w_j \geq 0, \quad j \in \mathcal{N}. \tag{5.5f}$$

Clearly, any feasible solution of (5.4) is also feasible for (5.5), and the optimal value of (5.5) represents a lower bound for that of (5.4). Since the relaxed problem contains fewer scenario-based constraints (5.5e), it is potentially easier to solve. It would then be of interest to determine the conditions under which an optimal solution of (5.5) is also optimal for the original problem (5.4). Assuming that $\boldsymbol{x}^*$ is an optimal solution of (5.5), consider the problem

$$\min \quad \eta + (1-\alpha)^{-1} w_0 \tag{5.6a}$$

$$\text{s. t.} \quad w_0 \geq v^{-1}\bigg(\sum_{j \in \mathbb{N}} \pi_j v(w_j) \bigg), \tag{5.6b}$$

$$w_j \geq X(\boldsymbol{x}^*, \omega_j) - \eta, \quad j \in \mathbb{N}, \tag{5.6c}$$

$$w_j \geq 0, \quad j \in \mathbb{N}. \tag{5.6d}$$

Proposition 5.1. *Consider problem* (5.4) *and its relaxation* (5.5) *obtained by aggregating scenario constraints* (5.4e) *over sets $\mathcal{S}_k$, $k \in \mathcal{K}$, that form a partition of* $\mathbb{N} = \{1, \dots, N\}$. *Assuming that* (5.4) *is feasible, consider problem* (5.6) *where $\boldsymbol{x}^*$ is an optimal solution of relaxation* (5.5). *Let $(\eta^{**}, \boldsymbol{w}^{**})$ be an optimal solution of* (5.6). *If the optimal value of* (5.6) *satisfies condition*

$$\eta^{**} + (1-\alpha)^{-1} w_0^{**} \leq h(\boldsymbol{x}^*), \tag{5.7}$$

then $(\boldsymbol{x}^*, \eta^{**}, \boldsymbol{w}^{**})$ is an optimal solution of the original problem (5.4).

Proof: Let $\boldsymbol{x}^\circ$ be an optimal solution of (5.4). Obviously, one has $g(\boldsymbol{x}^*) \leq g(\boldsymbol{x}^\circ)$. The statement of the proposition then follows immediately by observing that inequality (5.7) guarantees the triple $(\boldsymbol{x}^*, \eta^{**}, \boldsymbol{w}^{**})$ to be feasible for problem (5.4). □

The statement of Proposition 5.1 allows one to solve the original problem (5.4) by constructing an appropriate partition of $\mathbb{N}$ and solving the corresponding master problem (5.5). Below we outline an iterative procedure that accomplishes this goal.

Step 0: The algorithm is initialized by including all scenarios in a single partition, $\mathcal{K} = \{0\}$, $\mathcal{S}_0 = \{1, \dots, N\}$.

Step 1: For a current partition $\{\mathcal{S}_k : k \in \mathcal{K}\}$, solve the master problem (5.5). If (5.5) is infeasible, then the original problem (5.4) is infeasible as well, and the algorithm terminates. Otherwise, let $\boldsymbol{x}^*$ be an optimal solution of the master (5.5).

Step 2: Given a solution $\boldsymbol{x}^*$ of the master, solve problem (5.6), and let $(\eta^{**}, \boldsymbol{w}^{**})$ denote the corresponding optimal solution. If condition (5.7) is satisfied, the algorithm terminates with $(\boldsymbol{x}^*, \eta^{**}, \boldsymbol{w}^{**})$ being an optimal solution of (5.4) due to Proposition 5.1. If, however, condition (5.7) is violated,

$$\eta^{**} + (1-\alpha)^{-1} w_0^{**} > h(\boldsymbol{x}^*),$$

then the algorithm proceeds to Step 3 to update the partition.

Step 3: Determine the set of scenario-based constraints in (5.6) that, for a given solution of the master $\boldsymbol{x}^*$, are binding at optimality:

$$\mathcal{J} = \{j \in \mathbb{N} : w_j^{**} = X(\boldsymbol{x}^*, \omega_j) - \eta^{**} > 0\} \tag{5.8}$$

Then, the elements of $\mathcal{J}$ are removed from the existing sets $\mathcal{S}_k$:

$$\mathcal{S}_k = \mathcal{S}_k \setminus \mathcal{J}, \quad k \in \mathcal{K},$$

and added to the partition as single-element sets:

$$\{\mathcal{S}_0, \ldots, \mathcal{S}_K\} \cup \{S_{K+1}, \ldots, \mathcal{S}_{K+|\mathcal{J}|}\},$$

where, $\mathcal{S}_{K+i} = \{j_i\}$ for each $j_i \in \mathcal{J}, i = 1, \ldots, |\mathcal{J}|$, and the algorithm proceeds to Step 1.

Theorem 5.2. *Assume that in problem* (5.4) *functions* $g(\boldsymbol{x})$ *and* $X(\boldsymbol{x}, \omega)$ *are convex in* $\boldsymbol{x}$*,* $h(\boldsymbol{x})$ *is concave in* $\boldsymbol{x}$*,* v *satisfies assumption (U1), and the set* C *is convex and compact.*

Then, the described scenario decomposition algorithm either finds an optimal solution of problem (5.4) *or declares its infeasibility after at most N iterations.*

Proof: Let us show that during an iteration of the algorithm the size of the partition of the set $\mathcal{N}$ of scenarios increases by at least one.

Let $\{\mathcal{S}_k : k \in \mathcal{K}\}$ be the current partition of $\mathbb{N}$, $(\boldsymbol{x}^*, \eta^*, \boldsymbol{w}^*)$ be the corresponding optimal solution of (5.5), and $(\eta^{**}, \boldsymbol{w}^{**})$ be an optimal solution of (5.6) for the given $\boldsymbol{x}^*$, such that the stopping condition (5.7) is not satisfied,

$$\eta^{**} + (1-\alpha)^{-1} w_0^{**} > h(\boldsymbol{x}^*). \tag{5.9}$$

Let $\bar{\mathcal{S}}^*$ denote the set of constraints (5.6c) that are binding at optimality,

$$\bar{\mathcal{S}}^* = \left\{ j : w_j^{**} = X(\boldsymbol{x}^*, \omega_j) - \eta^{**} > 0,\ j \in \mathbb{N} \right\}.$$

Next, consider a problem obtained from (5.6) with a given $\boldsymbol{x}^*$ by aggregating the constraints (5.6c) that are non-binding at optimality:

$$\min \quad \eta + (1-\alpha)^{-1} w_0 \tag{5.10a}$$

$$\text{s. t.} \quad w_0 \geq v^{-1}\left(\sum_{j \in S_0} \pi_j v(w_j) \right), \tag{5.10b}$$

$$w_j \geq X(\boldsymbol{x}^*, \omega_j) - \eta, \quad j \in \bar{\mathcal{S}}^*, \tag{5.10c}$$

$$\sum_{j \in \mathcal{S}^*} w_j \geq \sum_{j \in \mathcal{S}^*} X(\boldsymbol{x}^*, \omega_j) - |\mathcal{S}^*| \eta, \tag{5.10d}$$

$$w_j \geq 0, \quad j \in \mathbb{N}, \tag{5.10e}$$

where $\mathcal{S}^* = \mathbb{N} \setminus \bar{\mathcal{S}}^*$. Obviously, an optimal solution $(\eta^{**}, \boldsymbol{w}^{**})$ of (5.6) will also be optimal for (5.10).

Next, observe that at any stage of the algorithm, the partition $\{\mathcal{S}_k : k \in \mathcal{K}\}$ is such that there exists at most one set with $|\mathcal{S}_k| > 1$, namely set $\mathcal{S}_0$, and the rest of the sets in the partition satisfy $|\mathcal{S}_k| = 1$, $k \neq 0$. Let us denote

$$\bar{\mathcal{S}}_0 = \mathbb{N} \setminus \mathcal{S}_0 = \bigcup_{k \in \mathcal{K} \setminus \{0\}} \mathcal{S}_k.$$

Assume that $\bar{\mathcal{S}}^* \subseteq \bar{\mathcal{S}}_0$. By rewriting the master problem (5.5) as

$$\min \quad g(\boldsymbol{x}) \tag{5.11a}$$

$$\text{s. t.} \quad \boldsymbol{x} \in C, \tag{5.11b}$$

$$\eta + (1-\alpha)^{-1} w_0 \leq h(\boldsymbol{x}), \tag{5.11c}$$

$$w_0 \geq v^{-1}\bigg(\sum_{j \in \mathcal{N}} \pi_j v(w_j)\bigg), \tag{5.11d}$$

$$w_j \geq X(\boldsymbol{x}, \omega_j) - \eta, \;\; j \in \bar{\mathcal{S}}_0, \tag{5.11e}$$

$$\sum_{j \in \mathcal{S}_0} w_j \geq \sum_{j \in \mathcal{S}_0} X(\boldsymbol{x}, \omega_j) - |\mathcal{S}_0|\eta, \tag{5.11f}$$

$$w_j \geq 0, \quad j \in \mathbb{N}, \tag{5.11g}$$

we observe that the components $\eta^*, \boldsymbol{w}^*$ of its optimal solution are feasible for (5.10). Indeed, from (5.11e) one has that

$$w_j^* \geq X(\boldsymbol{x}^*, \omega_j) - \eta^*, \quad j \in \bar{\mathcal{S}}^*,$$

which satisfies (5.10c), and also

$$w_j^* \geq X(\boldsymbol{x}^*, \omega_j) - \eta^*, \quad j \in \bar{\mathcal{S}}_0 \setminus \bar{\mathcal{S}}^* = \mathcal{S}^* \setminus \mathcal{S}_0.$$

Adding the last inequalities yields

$$\sum_{j \in \mathcal{S}^* \setminus \mathcal{S}_0} w_j^* \geq \sum_{j \in \mathcal{S}^* \setminus \mathcal{S}_0} X(\boldsymbol{x}^*, \omega_j) - |\mathcal{S}^* \setminus \mathcal{S}_0|\, \eta^*,$$

which can then be aggregated with (5.11f) to produce

$$\sum_{j\in\mathcal{S}^*} w_j^* \geq \sum_{j\in\mathcal{S}^*} X(\boldsymbol{x}^*, \omega_j) - |\mathcal{S}^*|\eta^*,$$

verifying the feasibility of $(\eta^*, \boldsymbol{w}^*)$ for (5.10). Since (5.11c) has to hold for $(\boldsymbol{x}^*, \eta^*, \boldsymbol{w}^*)$, we obtain that

$$\eta^{**} + (1-\alpha)^{-1}w^{**} \leq \eta^* + (1-\alpha)^{-1}w^* \leq h(\boldsymbol{x}^*),$$

which furnishes a contradiction with (5.9). Therefore, one has to have $\bar{\mathcal{S}}_0 \subset \bar{\mathcal{S}}^*$ for (5.9) to hold, meaning that at least one additional scenario from $\bar{\mathcal{S}}^*$ will be added to the partition during Step 3 of the algorithm. It is easy to see that the number of iterations cannot exceed the number N of scenarios. □

Remark 15. The fact that the proposed scenario decomposition method terminates within at most N iterations represents an important advantage over several existing cutting-plane methods that were developed in the literature for problems involving Conditional Value-at-Risk measure (Künzi-Bay and Mayer, 2006), integrated chance constraints (Klein Haneveld and van der Vlerk, 2006), and SSD constraints (Roman et al., 2006). In the mentioned works, the cutting-plane algorithms utilized supporting hyperplane representations for scenario constraints, which were themselves exponential in the size N of scenario sets. Although finite convergence of the cutting plane techniques was guaranteed by the polyhedral structure of the scenario constraints (in the case when $X(\boldsymbol{x}, \omega)$ is linear in $\boldsymbol{x}$), no estimate for the sufficient number of iterations was provided. A level-type regularization of cutting plane method for problems with SSD constraints, which allows for an estimate of the number of cuts due to Lemaréchal et al. (1995), is discussed in Fábián et al. (2011).

5.2.1 An Efficient Solution Method for Sub-Problem (5.6)

Although formulation (5.6) may be solved using appropriate mathematical programming techniques, an efficient alternative solution method can be employed by noting that (5.6) is equivalent to

$$\min\ \eta + \frac{1}{1-\alpha} v^{-1}\left(\sum_{j\in\mathbb{N}} \pi_j v\big(X(\boldsymbol{x}^*, \omega_j) - \eta\big)_+ \right), \tag{5.12}$$

which is a mathematical programming implementation of certainty equivalent representation under a finite scenario model where realizations $X(\boldsymbol{x}^*, \omega_j)$ represent scenario losses corresponding to an optimal decision $\boldsymbol{x}^*$ in the master problem (5.5). An optimal value of η in (5.6) and (5.12) can be computed directly using its properties dictated by representation.

Namely, let $X_j = X(\boldsymbol{x}^*, \omega_j)$ represent the optimal loss in scenario j for problem (5.5), and let $X_{(m)}$ be the m-th smallest outcome among $X_1, \ldots, X_N$, such that

$$X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(N)}.$$

The following proposition enables evaluation of η^{**} as a "cutoff" point within the tail of the loss distribution.

Proposition 5.3. *Given a function $v(\cdot)$ that satisfies (U1) and an $\alpha \in (0,1)$, a sufficient condition for η^{**} to be an optimal solution in problems* (5.12) *and* (5.6) *has the form*

$$\frac{\sum_{j:X_j>\eta^{**}} \pi_j v'(X_j - \eta^{**})}{v'(v^{-1}(\sum_{j\in\mathbb{N}} \pi_j v(X - \eta^{**})_+))} + \alpha - 1 = 0, \tag{5.13}$$

where v' denotes the derivative of v.

Remark 16. Note, that this claim essentially rephrases the result that we have already established in Chapter 2 (see, Proposition 2.12 and Corollary 2.13). We will still present the detailed proof of this proposition, since it will be useful later in our analysis.

Proof: The underlying assumption on v entails that $\phi(X) = (1-\alpha)^{-1}v^{-1}\mathsf{E}v(X)$ is convex, whence the objective function of (5.12)

$$\Phi_X(\eta) = \eta + \phi(X-\eta) = \eta + \frac{1}{1-\alpha}v^{-1}\Big(\sum\nolimits_{j\in\mathbb{N}} \pi_j v(X_j-\eta)_+\Big) \tag{5.14}$$

is convex on $\mathbb{R}$. Moreover, the condition $\phi(\eta) > \eta$ for $\eta \neq 0$ guarantees that the set of minimizers of $\Phi_X(\eta)$ is compact and convex in $\mathbb{R}$. Indeed, it is easy to see that $\Phi_X(\eta) = \eta$ for $\eta \geq X_{(N)}$ and $\Phi_X(\eta) \sim -\frac{\alpha\eta}{1-\alpha}$ for $\eta \ll -1$.

Now, consider the left derivative of $\Phi_X(\eta)$ at a given point $\eta = \eta^{**}$:

$$\begin{aligned}
&-(1-\alpha) + (1-\alpha)\frac{d^-}{d\eta}\Phi_X(\eta)\Big|_{\eta=\eta^{**}} = \frac{d^-}{d\eta}\Big\{v^{-1}\Big(\sum\nolimits_{j\in\mathbb{N}} \pi_j v(X_j-\eta)_+\Big)\Big\}\Big|_{\eta=\eta^{**}} \\
&= \lim_{\epsilon\to 0+}\frac{1}{-\epsilon}\Big\{v^{-1}\Big(\sum_{j:X_j\geq\eta^{**}} \pi_j v(X_j-\eta^{**}+\epsilon)\Big) - v^{-1}\Big(\sum_{j:X_j\geq\eta^{**}} \pi_j v(X_j-\eta^{**})\Big)\Big\} \\
&= \frac{d^-}{d\eta}\Big\{v^{-1}\Big(\sum_{j:X_j\geq\eta^{**}} \pi_j v(X_j-\eta)\Big)\Big\}\Big|_{\eta=\eta^{**}} = \frac{d}{d\eta}\Big\{v^{-1}\Big(\sum_{j:X_j\geq\eta^{**}} \pi_j v(X_j-\eta)\Big)\Big\}\Big|_{\eta=\eta^{**}},
\end{aligned}$$

where the last equality follows from the continuous differentiability of function $v^{-1}\big(\sum_{j:X_j\geq\eta^{**}} \pi_j v(X_j-\eta^{**})\big)$ at the point η^{**} due to the assumed properties of v. Analogously, the right derivative of $\Phi_X(\eta)$ at $\eta = \eta^{**}$ equals to

$$\frac{d^+}{d\eta}\Phi_X(\eta)\Big|_{\eta=\eta^{**}} = 1 + \frac{1}{1-\alpha}\frac{d}{d\eta}\Big\{v^{-1}\Big(\sum_{j:X_j>\eta^{**}} \pi_j v(X_j-\eta)\Big)\Big\}\Big|_{\eta=\eta^{**}},$$

where the strict inequality in summation is due to fact that $v(X_j-\eta^{**}-\epsilon)_+ = 0$ for all $\epsilon > 0$ if $\eta^{**} \leq X_j$.

Observe that $\Phi_X(\eta)$ may only be non-differentiable at points $\eta = X_j$. Indeed, for any $\eta^{**} \neq X_j$, $j \in \mathbb{N}$, the obtained expressions for left and right derivatives become equivalent, and equation (5.13) is obtained from the first order optimality conditions by computing the derivatives of the functions in braces and noting that $\sum_{j:X_j \geq \eta^{**}} \pi_j v(X_j - \eta^{**}) = \sum_{j:X_j > \eta^{**}} \pi_j v(X_j - \eta^{**}) = \sum_{j \in \mathbb{N}} \pi_j v(X_j - \eta^{**})_+$. □

Recall that the presented above scenario decomposition algorithm uses the subproblem (5.6) for determining an optimal value of η^{**}, as well as for identifying (during Step 3) the set $\mathcal{J}$ of scenarios that are binding at optimality, i.e., for which $X(\boldsymbol{x}^*, \omega_j) - \eta^{**} > 0$. This can be accomplished with the help of the derived optimality condition (5.13) as follows.

Step (i) Compute values $X_j = X(\boldsymbol{x}^*, \omega_j)$, where $\boldsymbol{x}^*$ is an optimal solution of (5.5), and sort them in ascending order: $X_{(1)} \leq \ldots \leq X_{(N)}$.

Step (ii) For $m = N, N-1, \ldots, 1$, compute values T_m as

$$
\begin{aligned}
T_N &= 1 - \alpha, \\
T_m &= 1 - \alpha - \frac{\sum_{j=m+1}^{N} \pi_j v'\big(X_{(j)} - X_{(m)}\big)}{v'\left(v^{-1}\left(\sum_{j=m+1}^{N} \pi_j v\big(X_{(j)} - X_{(m)}\big)\right)\right)}, \quad m = N-1, \ldots, 1,
\end{aligned}
\tag{5.15}
$$

until m^* is found such that

$$T_{m^*} \leq 0, \; T_{m^*+1} > 0. \tag{5.16}$$

Step (iii) If $T_{m^*} = 0$, then the solution η^{**} of (5.6), (5.12) is equal to $X_{(m^*)}$. Otherwise, η^{**} satisfies

$$\eta^{**} \in \big(X_{(m^*)}, X_{(m^*+1)}\big],$$

and its value can be found by using an appropriate numerical procedure, such as Newton's method. The set $\mathcal{J}$ in (5.8) is then obtained as

$$\mathcal{J} = \{j : X_j = X_{(k)},\ k = m^* + 1, \ldots, N\}.$$

Proposition 5.4. *Given an optimal solution $\boldsymbol{x}^*$ of the master problem* (5.5)*, the algorithm described in steps (i)–(iii) yields an optimal value η^{**} in* (5.6)*,* (5.12) *and the set $\mathcal{J}$ to be used during steps 2 and 3 of the scenario decomposition algorithm.*

Proof: First, observe that an optimal solution η^{**} of (5.6) and (5.12) satisfies $\eta^{**} \leq X_{(N)}$. Indeed, assume to the contrary that $\eta^{**} = X_{(N)} + \epsilon$ for some $\epsilon > 0$. The optimal value of (5.6) is then equal to $X_{(N)} + \epsilon$, and can be improved by selecting, e.g., $\epsilon = \epsilon/2$.

Next, observe that quantities T_m are equal, up to a factor $1 - \alpha$, to the right derivatives of function $\Phi_X(\eta)$ (5.14) at $\eta = X_{(m)}$, i.e., $T_m = (1-\alpha)\frac{d^+}{d\eta}\Phi_X(\eta)\big|_{\eta = X_{(m)}}$. The value of $T_N = 1 - \alpha$ follows directly from the fact that $\Phi_X(\eta) = \eta$ for $\eta \geq X_{(N)}$. Then, if strict inequalities in (5.16) hold, two cases are possible. Namely, an optimal η^{**} is located inside the interval $\big(X_{(m^*)}, X_{(m^*+1)}\big)$ if $\frac{d^-}{d\eta}\Phi_X(X_{(m^*+1)}) > 0$. Alternatively, $\eta^{**} = X_{(m^*+1)}$ if $\frac{d^-}{d\eta}\Phi_X(X_{(m^*+1)}) \leq 0$. Thus, we have the second statement of step (iii).

If $T_{m^*} = 0$ in (5.16), observe that necessarily $\frac{d^-}{d\eta}\Phi_X(X_{m^*}) \leq 0$ since the left derivative of Φ_X at $X_{(m)}$ differs from the expression (5.15) by an extra summand $\pi_m v'(0)$ in the numerator. If $v'(0) = 0$ then $\frac{d^-}{d\eta}\Phi_X(X_{m^*}) = \frac{d^+}{d\eta}\Phi_X(X_{m^*}) = 0$ and $\eta^{**} = X_{(m^*)}$ is a minimum due to Proposition 5.3. If $v'(0) > 0$ then $\frac{d^-}{d\eta}\Phi_X(X_{m^*}) < 0$ and $\eta^{**} = X_{(m^*)}$ is again either a unique minimizer, or represents the left endpoint of the set of minimizers. This validates the first claim of step (iii).

Once the value of η^{**} is obtained during step (iii), the set $\mathcal{J}$ in (5.8) is constructed as the set of scenario indices corresponding to $X_{(m^*+1)}, X_{(m^*+2)}, \ldots, X_{(N)}$.

Note that it is not necessary to prove that there always exists $m^* \in \{1, \ldots, N-1\}$ such that $T_{m^*} \leq 0$ and $T_{m^*+1} > 0$. If indeed it were to happen that $T_m > 0$ for all $m = 1, \ldots, N$, this would imply that set $\mathcal{J}$ must contain all scenarios, i.e., $\mathcal{J} = \mathbb{N}$, making the exact value of η^{**} irrelevant in this case, since the original problem (5.4) would have to be solved at the next iteration of the scenario decomposition algorithm. □

Remark 17. We conclude this section by noting that the presented scenario decomposition approach is applicable, with appropriate modifications, to more general forms of downside risk measures $\rho(X) = \min_\eta\{\eta + \phi((X-\eta)_+)\}$. The focus of our discussion on the case when function ϕ has the form of a certainty equivalent, $\phi(X) = v^{-1}\mathsf{E}v(X_+)$, is dictated mainly by the fact that the resulting constraint (5.4d) encompasses a number of interesting and practically relevant special cases, such as second-order cone, p-order cone, and log-exponential constraints.

5.3 Numerical Experiments

Portfolio optimization problems are commonly used as an experimental platform in risk management and stochastic optimization. In this section we illustrate the computational performance of the proposed scenario decomposition algorithm on a portfolio optimization problem, where the investment risk is quantified using HMCR or LogExpCR measures.

A standard formulation of portfolio optimization problem entails determining the vector of portfolio weights $\boldsymbol{x} = (x_1, \ldots, x_n)^\top$ of n assets so as to minimize the risk while maintaining a prescribed level of expected return. We adopt the traditional definition of portfolio losses X as negative portfolio returns, $X(\boldsymbol{x},\omega) = -\boldsymbol{r}(\omega)^\top\boldsymbol{x}$, where $\boldsymbol{r}(\omega) = (r_1(\omega),\ldots,r_n(\omega))^\top$ are random returns of the assets. Then, the portfolio selection model takes the general form

$$\min \quad \mathbb{R}\big(-\boldsymbol{r}(\omega)^\top\boldsymbol{x}\big) \tag{5.17a}$$

$$\text{s. t.} \quad \mathbf{1}^\top\boldsymbol{x} = 1, \tag{5.17b}$$

$$\mathsf{E}\big[\boldsymbol{r}(\omega)^\top\boldsymbol{x}\big] \geq \bar{r}, \tag{5.17c}$$

$$\boldsymbol{x} \geq \mathbf{0}, \tag{5.17d}$$

where $\mathbf{1} = (1,\ldots,1)^\top$, equality (5.17b) represents the budget constraint, (5.17b) ensures a minimum expected portfolio return level, $\bar{r}$, and (5.17d) corresponds to no-short-selling constraints.

The distribution of the random vector $\boldsymbol{r}(\omega)$ of assets' returns is given by a finite set of N equiprobable scenarios $\boldsymbol{r}_j = \boldsymbol{r}(\omega_j) = (r_{1j},\ldots,r_{nj})^\top$,

$$\pi_j = \mathsf{P}\big\{\boldsymbol{r} = (r_{1j},\ldots,r_{nj})^\top\big\} = 1/N, \quad j \in \mathbb{N} \equiv \{1,\ldots,N\}. \tag{5.18}$$

5.3.1 Portfolio Optimization with Higher Moment Coherent Risk Measures

In the case when risk measure ρ in (5.17) is selected as a higher moment coherent risk measure, $\rho(X) = \mathsf{HMCR}_{p,\alpha}(X)$, the portfolio optimization problem (5.17) can be written in a stochastic programming form that is consistent with the general formulation

(5.4) as

$$\min \quad \eta + (1-\alpha)^{-1} w_0 \tag{5.19a}$$

$$\text{s. t.} \quad w_0 \geq \|(w_1, \ldots, w_N)\|_p, \tag{5.19b}$$

$$\pi_j^{-1/p} w_j \geq \; -\boldsymbol{r}_j^\top \boldsymbol{x} - \eta, \quad j \in \mathbb{N}, \tag{5.19c}$$

$$\boldsymbol{x} \in C, \;\; \mathbf{w} \geq \mathbf{0}, \tag{5.19d}$$

where C represents a polyhedral set comprising the expected return, budget, and no-short-selling constraints on the vector of portfolio weights $\boldsymbol{x}$:

$$C = \left\{ \boldsymbol{x} \in \mathbb{R}^n \;:\; \sum\nolimits_{j \in \mathbb{N}} \pi_j \boldsymbol{r}_j^\top \boldsymbol{x} \geq \bar{r}, \;\; \mathbf{1}^\top \boldsymbol{x} = 1, \;\; \mathbf{x} \geq \mathbf{0} \right\}. \tag{5.20}$$

Due to the presence of p-order cone constraint (5.19b), formulation (5.19) constitutes a p-order cone programming problem (pOCP).

Solution methods for problem (5.19) are dictated by the specific value of parameter p in (5.19b). As has been mentioned, in the case of $p = 1$ formulation (5.19) reduces to a LP problem that corresponds to a choice of risk measure as the CVaR, a case that has received a considerable attention in the literature. In view of this, of particular interest are nonlinear instances of problem (5.19), which correspond to values of the parameter $p \in (1, +\infty)$.

Below we consider instances of (5.19) with $p = 2$ and $p = 3$. In the case of $p = 2$, problem (5.19) can be solved using SOCP self-dual interior point methods. In the case of $p = 3$ and, generally, $p \in (1, 2) \cup (2, \infty)$, the p-cone (5.19b) is not self-dual, and we employ two techniques for solving (5.19) and the corresponding master problem (5.5): (i) a SOCP-based approach that relies on the fact that for a rational p, a p-order cone can be equivalently represented via a sequence of second order cones, and (ii) an LP-based approach that allows for obtaining exact solutions of pOCP problems via cutting-plane methods.

Detailed discussions of the respective formulations of problems (5.19) are provided below. Throughout this section, we use abbreviations in brackets to denote the different formulations of the "complete" versions of (5.19) (i.e., with complete set of scenario constraints (5.19c)). For each "complete" formulation, we also consider the corresponding scenario decomposition approach, indicated by suffix "SD". Within the scenario decomposition approach, we present formulations of the master problem (denoted by subscript "MP"); the respective subproblems are then constructed accordingly. For example, the SOCP version of the complete problem (5.19) with $p = 2$ is denoted [SOCP], while the same problem solved by scenario decomposition is referred to as [SOCP-SD], with the master problem being denoted as [SOCP-SD]$_{\mathrm{MP}}$ (see below).

5.3.1.1 SOCP Formulation in $p = 2$ Case

In case when $p = 2$, formulation (5.19) constitutes a standard SOCP problem that can be solved using a number of available SOCP solvers, such as CPLEX, MOSEK, GUROBI, etc. In order to solve it using the scenario decomposition algorithm presented in Section 5.2, the master problem (5.5) is formulated with respect to the original problem (5.19) with $p = 2$ as follows:

$$\begin{aligned}
\min \quad & \eta + (1-\alpha)^{-1} w_0 \\
\text{s. t.} \quad & w_0 \geq \|(w_1, \ldots, w_N)\|_2, \\
& \sum_{j \in \mathcal{S}_k} \frac{\pi_j^{1/2}}{\pi^{(k)}} w_j \geq \left(\sum_{j \in \mathcal{S}_k} \frac{\pi_j}{\pi^{(k)}} \boldsymbol{r}_j^\top \right) \boldsymbol{x} - \eta, \quad k \in \mathcal{K}, \\
& \mathbf{w} \geq \mathbf{0}, \ \ \boldsymbol{x} \in C.
\end{aligned} \qquad \text{[SOCP-SD]}_{\mathrm{MP}}$$

Note that in the case of $\mathsf{HMCR}_{2,\alpha}$ measure, the function $v(t) = t^2$ is positive homogeneous of degree two, which allows for eliminating the scenario probabilities π_j from constraint (5.5d) and representing the latter in the form of a second order cone in the full formulation (5.19) and in the master problem $[\text{SOCP-SD}]_{\text{MP}}$. This affects constraints (5.5d), which then can be written in the form of the second constraint in $[\text{SOCP-SD}]_{\text{MP}}$. The subproblem (5.6) is reformulated accordingly.

5.3.1.2 SOCP Reformulation of p-Order Cone Program

One of the possible approaches for solving the pOCP problem (5.19) with $p = 3$ involves reformulating the p-cone constraint (5.19b) via a set of quadratic cone constraints. Such an exact reformulation is possible when the parameter p has a rational value, $p = q/s$. Then, a (q/s)-order cone constraint in the positive orthant $\mathbb{R}_+^{N+1}$

$$\big\{\mathbf{w} \geq \mathbf{0} \;:\; w_0 \geq (w_1^{q/s} + \ldots + w_N^{q/s})^{s/q}\big\} \tag{5.21}$$

may equivalently be represented as the following set in $\mathbb{R}_+^{N+1} \times \mathbb{R}_+^N$:

$$\big\{\mathbf{w}, \mathbf{u} \geq \mathbf{0} \;:\; w_0 \geq \|\mathbf{u}\|_1, \;\; w_j^q \leq u_j^s w_0^{q-s}, \;\; j \in \mathbb{N}\big\}. \tag{5.22}$$

Each of the N nonlinear inequalities in (5.22) can in turn be represented as a sequence of three-dimensional rotated second-order cones of the form $\xi_0^2 \leq \xi_1\xi_2$, resulting in a SOCP reformulation of the rational-order cone (5.21) (Nesterov and Nemirovski, 1994; Alizadeh and Goldfarb, 2003; Krokhmal and Soberanis, 2010). Such a representation, however, is not unique and in general may comprise a varying number of rotated second order cones for a given $p = q/s$. In this case study we use the technique of Morenko et al. (2013), which allows for representing rational order p-cones with $p = q/s$ in $\mathbb{R}^{N+1}$ via $N\lceil \log_2 q \rceil$ second order cones. Namely, in the case of $p = 3$, when $q = 3$, $s = 1$, the 3-order cone (5.21) can equivalently be replaced with $\lceil \log_2 3 \rceil N = 2N$ quadratic cones

$$\big\{\mathbf{w}, \mathbf{u}, \mathbf{v} \geq \mathbf{0} \ : \ w_0 \geq \|\mathbf{u}\|_1, \ \ w_j^2 \leq w_0 v_j, \ \ v_j^2 \leq w_j u_j, \ \ j \in \mathbb{N}\big\}. \tag{5.23}$$

In accordance with the above, a p-order cone inequality in $\mathbb{R}^{N+1}$ can be represented by a set of 3D second order cone constraints and a linear inequality when p is a positive rational number. Thus, the [SpOCP] problem (5.19) takes the following form:

$$\begin{aligned}
\min \quad & \eta + (1-\alpha)^{-1} w_0 \\
\text{s. t.} \quad & w_0 \geq \|\mathbf{u}\|_1, \\
& w_j^2 \leq w_0 v_j, \ v_j^2 \leq w_j u_j, \quad j \in \mathbb{N}, \\
& \pi_j^{-1/p} w_j \geq \ -\boldsymbol{r}_j^\top \boldsymbol{x} - \eta, \quad j \in \mathbb{N}, \\
& \boldsymbol{x} \in C, \ \mathbf{w}, \mathbf{v}, \mathbf{u} \geq \mathbf{0}.
\end{aligned} \tag{[SpOCP]}$$

The corresponding master problem sub-problem $[\text{SpOCP-SD}]_{\text{MP}}$ in the scenario decomposition method is constructed by replacing constraints of the form (5.19c) in the last problem as follows:

$$\begin{aligned}
\min \quad & \eta + (1-\alpha)^{-1} w_0 \\
\text{s. t.} \quad & w_0 \geq \|\mathbf{u}\|_1, \\
& w_j^2 \leq w_0 v_j, \ v_j^2 \leq w_j u_j, \quad j \in \mathbb{N}, \\
& \sum_{j \in \mathcal{S}_k} \frac{\pi_j^{1-1/p}}{\pi^{(k)}} w_j \geq \left(\sum_{j \in \mathcal{S}_k} \frac{\pi_j}{\pi^{(k)}} \boldsymbol{r}_j^\top \right) \boldsymbol{x} - \eta, \quad k \in \mathcal{K}, \\
& \boldsymbol{x} \in C, \ \mathbf{w}, \mathbf{v}, \mathbf{u} \geq \mathbf{0}.
\end{aligned} \tag{$[\text{SpOCP-SD}]_{\text{MP}}$}$$

5.3.1.3 An Exact Solution Method for pOCP Programs Based on Polyhedral Approximations

Computational methods for solving p-order cone programming problems that are based on polyhedral approximations represent an alternative to interior-point approaches, and can be beneficial in situations when a pOCP problem needs to be solved repeatedly, with small variations in problem data or problem structure.

Thus, in addition to the SOCP-based approaches for solving the pOCP problem (5.19) discussed above, we also employ an exact polyhedral-based approach with $O(\varepsilon^{-1})$ iteration complexity that was proposed in Chapter 3. It consists in reformulating the p-order cone $w_0 \geq \|(w_1, \ldots, w_N)\|_p$ via a set of three-dimensional p-cones

$$w_0 = w_{2N-1}, \quad w_{N+j} \geq \|(w_{2j-1}, w_{2j})\|_p, \quad j = 1, \ldots, N-1, \tag{5.24}$$

and then iteratively building outer polyhedral approximations of the 3D p-cones until the solution of desired accuracy $\varepsilon > 0$ is obtained,

$$\|(w_1, \ldots, w_N)\|_p \leq (1 + \varepsilon) w_0.$$

In the context of the lifted representation (5.24), the above ε-relaxation of p-cone inequality translates into $N - 1$ corresponding approximation inequalities for 3D p-cones:

$$\|(w^*_{2j-1}, w^*_{2j})\|_p \leq (1 + \epsilon) w^*_{N+j}, \quad j = 1, \ldots, N-1, \tag{5.25}$$

where $\epsilon = (1 + \varepsilon)^{1/\lceil \log_2 N \rceil} - 1$. Then, for a given $\varepsilon > 0$, an ε-approximate solution of pOCP portfolio optimization problem (5.19) is obtained by iteratively solving the linear programming problem

$$
\begin{aligned}
\min \quad & \eta + (1-\alpha)^{-1} w_0 \\
\text{s. t.} \quad & w_0 = w_{2N-1}, \\
& w_{N+j} \geq \alpha_p(\theta_{k_j}) w_{2j-1} + \beta_p(\theta_{k_j}) w_{2j}, \quad \theta_{k_j} \in \Theta_j, \quad j = 1, \ldots, N-1, \\
& \pi_j^{-1/p} w_j \geq \ -\boldsymbol{r}_j^\top \boldsymbol{x} - \eta, \quad j \in \mathbb{N}, \\
& \boldsymbol{x} \in C, \ \mathbf{w} \geq \mathbf{0},
\end{aligned}
\qquad \text{[LpOCP]}
$$

where coefficients α_p and β_p are defined as

$$
\alpha_p(\theta) = \frac{\cos^{p-1}\theta}{(\cos^p\theta + \sin^p\theta)^{1-\frac{1}{p}}}, \quad \beta_p(\theta) = \frac{\sin^{p-1}\theta}{(\cos^p\theta + \sin^p\theta)^{1-\frac{1}{p}}}.
$$

If, for a given solution $\boldsymbol{w}^* = (w_0^*, \ldots, w_{2N-1}^*)$ of [LpOCP], the approximation condition (5.25) is not satisfied for some $j = 1, \ldots, N-1$,

$$
\|(w_{2j-1}^*, w_{2j}^*)\|_p > (1+\epsilon) w_{N+j}^*, \tag{5.26}
$$

then a cut of the form

$$
w_{N+j} \geq \alpha_p(\theta_j^*) w_{2j-1} + \beta_p(\theta_j^*) w_{2j}, \quad \theta_j^* = \arctan \frac{w_{2j}^*}{w_{2j-1}^*}, \tag{5.27}
$$

is added to [LpOCP]. The process is initialized with $\Theta_j = \{\theta_1\}$, $\theta_1 = \pi/4$, $j = 1, \ldots, N-1$, and continues until no violations of condition (5.26) are found. In Chapter 3 it was shown that this cutting-plane procedure generates an ε-approximate solution to pOCP problem (5.19) within $O(\varepsilon^{-1})$ iterations.

The described cutting plane scheme can be employed to solve the master problem corresponding to the pOCP problem (5.19). Namely, the cutting-plane formulation of this master problem is obtained by replacing the p-cone constraint (5.19b) with cutting planes similarly to [LpOCP], and the set of N scenario constraints (5.19c) with the aggregated constraints (compare to $[\text{SpOCP-SD}]_{\text{MP}}$):

$$
\begin{aligned}
\min \quad & \eta + (1-\alpha)^{-1} t \\
\text{s. t.} \quad & w_0 = w_{2N-1}, \\
& w_{N+j} \geq \alpha_p(\theta_{k_j}) w_{2j-1} + \beta_p(\theta_{k_j}) w_{2j}, \quad \theta_{k_j} \in \Theta_j, \quad j = 1, \ldots, N-1, \\
& \sum_{j \in \mathcal{S}_k} \frac{\pi_j^{1-1/p}}{\pi^{(k)}} w_j \geq \left(\sum_{j \in \mathcal{S}_k} \frac{\pi_j}{\pi^{(k)}} \boldsymbol{r}_j^\top \right) \boldsymbol{x} - \eta, \quad k \in \mathcal{K}, \\
& \boldsymbol{x} \in C, \ \mathbf{w} \geq \mathbf{0}.
\end{aligned}
\tag{$[\text{LpOCP-SD}]_{\text{LB}}$}
$$

5.3.2 Portfolio Optimization with Log Exponential Convex Risk Measures

In order to demonstrate the applicability of the proposed method when solving problems with measures of risk other than the HMCR class, we examine an analogous experimental framework for instances when $\mathbb{R}(X) = \mathsf{LogExpCR}_{e,\alpha}(X)$. The portfolio optimization problem (5.17) may then be written as

$$\begin{aligned} \min \quad & \eta + (1-\alpha)^{-1} w_0 \\ \text{s. t.} \quad & w_0 \geq \ln \sum_{j \in \mathbb{N}} \pi_j e^{w_j}, \\ & w_j \geq \ -\boldsymbol{r}_j^\top \boldsymbol{x} - \eta, \quad j \in \mathbb{N}, \\ & \boldsymbol{x} \in C, \ \boldsymbol{w} \geq \mathbf{0}. \end{aligned} \qquad \text{[LogExpCP]}$$

Note that in contrast to pOCP and SOCP problems discussed in the preceding subsections, the above formulation is not a conic program. Since it involves a convex log-exponential constraint, we call this problem a log-exponential convex programming problem (LogExpCP) that can be solved with interior point methods.

The corresponding master problem for the scenario decomposition algorithm is obtained from [LogExpCP] by aggregating the scenario constraints in accordance to (5.5):

$$\begin{aligned} \min \quad & \eta + (1-\alpha)^{-1} w_0 \\ \text{s. t.} \quad & w_0 \geq \ln \sum_{j \in \mathbb{N}} \pi_j e^{w_j}, \\ & \sum_{j \in \mathcal{S}_k} w_j \geq -\sum_{j \in \mathcal{S}_k} \boldsymbol{r}_j^\top \boldsymbol{x} - |\mathcal{S}_k| \eta, \quad k \in \mathcal{K}, \\ & \boldsymbol{x} \in C, \ \boldsymbol{w} \geq \mathbf{0}. \end{aligned} \qquad \text{[LogExpCP-SD]}_{\text{MP}}$$

In the next section we examine the computational performances within each implementation class of problem (5.19).

5.3.3 Computational Results

The portfolio optimization problems described in Section 5.3.1 and 5.3.2 were implemented in C++ using callable libraries of three solvers, CPLEX 12.5, GUROBI 5.02, and MOSEK 6. Computations ran on a six-core 2.30GHz PC with 128GB RAM in 64-bit Windows environment. In the context of benchmarking, each adopted formulation was tested against its scenario decomposition-based implementation. Moreover, it was of particular interest to examine the performance of the scenario decomposition algorithm using various risk measure configurations, thus, the following problem settings were solved: problems [SOCP]-[SOCP-SD] with risk measure as defined by HMCR for $p = 2$; problems [SpOCP]-[SpOCP-SD] and [LpOCP]-[LpOCP-SD] with HMCR measure for $p = 3$; and problems [LogExpCP]-[LogExpCP-SD] with LogExpCR measure. The value of parameter α in the employed risk measures was fixed at $\alpha = 0.9$ throughout.

The scenario data in our numerical experiments was generated as follows. First, a set of n stocks ($n = 50$, 100, 200) was selected at random from the S&P500 index. Then, a covariance matrix of daily returns as well as the expected returns were estimated for the specific set of n stocks using historical prices from January 1, 2006 to January 1, 2012. Finally, the desired number N of scenarios, ranging from 1,000 to 100,000, have been generated as N independent and identically distributed samples from a multivariate normal distribution with the obtained mean and covariance matrix.

On account of precision arithmetic errors associated with the numerical solvers, we introduced a tolerance level $\epsilon > 0$ to specify the permissible gap in the stopping criterion (5.7):

$$\eta^{**} + (1-\alpha)^{-1} w_0^{**} \leq h(\boldsymbol{x}^*) + \epsilon. \tag{5.28}$$

Specifically, the value $\epsilon = 10^{-5}$ was was chosen to match the reduced cost of the simplex method in CPLEX and GUROBI. In a similar manner, we adjust (5.15) around m^* for precision errors as

$$T_{m^*+1}(p) - \epsilon < 0 \quad \text{and} \quad T_{m^*}(p) + \epsilon > 0.$$

Empirical observations suggest the accumulation of numerical errors is exacerbated by the use of fractional values of scenarios in assets returns, r_{ij}. To alleviate the numerical accuracy issues, the data in respective problem instances of the scenario decomposition algorithm were appropriately scaled.

The results of our numerical experiments are summarized in Tables 5.1 – 5.5. Unless stated otherwise, the reported running time values are averaged over 20 instances. Table 5.1 presents the computational times observed during solving the full formulation, [SOCP], of problem (5.19) with HMCR measure and $p = 2$, and solving the same problem using the scenario decomposition algorithm, [SOCP-SD], with the three solvers, CPLEX, GUROBI, and MOSEK. Observe that the scenario decomposition method performs better for all instances and solvers, with the exception of the largest three scenario instances when using GUROBI with $n = 50$ assets. However, this trend is tampered as the number of assets increases.

Table 5.2 reports the running times observed during solving of the second-order cone reformulation of the pOCP version of problem (5.19) with $p = 3$, in the full formulation ([SpOCP]) and via the scenario decomposition algorithm ([SpOCP-SD]). The obtained results indicate that, although the scenario decomposition algorithm is slower on smaller problem instances, it outperforms direct solution methods as the numbers of scenarios N and assets n in the problem increase. Due to observed numerical instabilities, the CPLEX solver was not considered for this particular experiment.

Table 5.1: Average computation times (in seconds) obtained by solving problems [SOCP] and [SOCP-SD] for $p = 2$ using CPLEX, GUROBI and MOSEK. All running times are averaged over 20 instances.

		CPLEX		GUROBI		MOSEK	
n	N	[SOCP]	[SOCP-SD]	[SOCP]	[SOCP-SD]	[SOCP]	[SOCP-SD]
50	1000	1.00	0.46	0.62	0.45	0.26	0.15
	2500	3.03	0.51	1.88	1.07	0.60	0.36
	5000	6.58	0.55	3.81	2.78	1.24	0.72
	10000	13.72	1.35	9.56	7.89	2.56	1.61
	25000	31.03	3.53	32.40	34.04	7.33	5.18
	50000	60.62	9.05	101.09	117.24	17.64	12.43
	100000	137.14	25.25	327.95	449.78	36.78	33.02
100	1000	2.46	0.86	1.73	0.42	0.61	0.18
	2500	6.14	0.99	4.87	1.17	1.50	0.47
	5000	13.69	1.10	11.13	3.55	3.25	1.15
	10000	27.06	2.21	21.94	9.63	6.69	3.03
	25000	72.95	8.85	71.34	37.48	20.41	6.88
	50000	157.25	20.88	185.56	129.37	44.01	16.61
	100000	319.90	58.29	464.12	467.35	79.75	41.58
200	1000	6.87	2.19	5.60	0.58	6.68	0.29
	2500	17.48	2.10	15.36	1.37	4.49	0.73
	5000	34.93	2.98	33.96	4.15	9.36	1.92
	10000	76.13	5.03	63.67	16.50	19.54	5.51
	25000	206.29	24.16	196.45	54.00	53.89	29.15
	50000	447.85	55.93	438.40	152.76	112.47	28.85
	100000	950.17	112.60	998.86	539.46	234.68	61.98

Table 5.2: Average computation times (in seconds) obtained by solving problems [SpCOP] and [SpCOP-SD] for $p = 3$ using GUROBI and MOSEK. All running times are averaged over 20 instances and symbol "—" indicates that the time limit of 3600 seconds was exceeded.

		GUROBI		MOSEK	
n	N	[SpOCP]	[SpCOP-SD]	[SpOCP]	[SpCOP-SD]
50	1000	2.58	2.73	0.18	0.63
	2500	10.63	6.61	0.49	0.96
	5000	32.01	19.27	1.06	1.70
	10000	87.27	41.34	2.31	3.49
	25000	198.56	92.39	7.14	6.70
	50000	455.63	540.09	16.36	13.70
	100000	1217.96	2080.34	35.33	30.29
100	1000	7.16	3.14	0.30	0.75
	2500	29.47	8.44	0.85	1.37
	5000	90.25	19.74	1.88	2.32
	10000	277.72	44.31	4.52	3.91
	25000	642.63	92.11	12.66	8.66
	50000	1365.37	1716.37	28.64	15.10
	100000	—	—	65.48	28.29
200	1000	17.86	3.87	0.69	1.01
	2500	78.28	8.65	1.90	1.56
	5000	276.89	22.40	4.41	2.47
	10000	799.65	49.02	9.88	4.84
	25000	2118.11	107.14	29.99	9.60
	50000	—	—	64.52	17.41
	100000	—	—	139.87	34.99

Table 5.3: Average computation times (in seconds) obtained by solving problems [LpOCP] and [LpOCP-SD] for $p = 3$ using CPLEX, GUROBI and MOSEK. All running times are averaged over 20 instances and symbol "—" indicates that the time limit of 3600 seconds was exceeded.

		CPLEX		GUROBI		MOSEK	
n	N	[LpOCP]	[LpOCP-SD]	[LpOCP]	[LpOCP-SD]	[LpOCP]	[LpOCP-SD]
50	1000	0.27	0.12	0.22	0.59	0.82	0.46
	2500	1.65	0.24	0.74	0.83	4.26	0.66
	5000	6.81	0.46	2.31	1.54	15.08	1.46
	10000	19.20	1.42	7.73	3.86	60.66	3.75
	25000	31.93	3.93	56.52	13.74	381.67	11.34
	50000	179.49	16.07	117.72	36.51	1412.81	25.47
	100000	903.36	62.79	474.68	112.72	—	54.45
100	1000	0.37	0.13	0.23	0.61	2.94	0.65
	2500	2.22	0.28	0.86	0.98	7.11	1.06
	5000	8.58	0.79	2.82	1.76	32.20	1.95
	10000	28.71	2.18	9.28	4.13	122.75	4.99
	25000	45.37	4.99	35.11	13.13	1138.99	15.34
	50000	200.12	18.80	122.21	39.78	2753.54	34.17
	100000	3336.26	82.79	1316.29	138.74	—	80.15
200	1000	0.61	0.20	0.33	0.89	15.68	1.06
	2500	3.13	0.44	1.30	1.17	20.64	1.37
	5000	13.25	1.01	3.72	2.11	70.49	2.97
	10000	47.97	3.31	13.20	4.72	322.36	8.12
	25000	195.28	6.98	94.45	14.77	2418.52	26.91
	50000	936.60	27.20	665.61	45.43	—	53.62
	100000	—	114.08	3301.44	160.92	—	123.89

Table 5.4: Average computation times (in seconds) obtained by solving a specified number of instances for problems [LogExpCP] and [LogExpCP-SD] using MOSEK solver.

n	N	MOSEK [LogExpCP]	MOSEK [LogExpCP-SD]	Instances Solved
50	1000	0.61	0.27	12
	2500	0.97	0.58	14
	5000	1.89	1.18	12
	10000	4.88	2.57	9
	25000	14.99	7.94	12
	50000	26.65	18.76	15
	100000	65.45	61.48	17
100	1000	0.57	0.25	17
	2500	1.65	0.53	16
	5000	3.69	1.14	10
	10000	9.18	2.53	15
	25000	24.61	13.83	13
	50000	50.66	39.72	19
	100000	148.54	59.02	16
200	1000	5.25	0.37	19
	2500	4.22	0.75	17
	5000	9.53	1.39	18
	10000	21.17	2.63	17
	25000	62.03	7.59	17
	50000	145.89	16.47	18
	100000	333.73	43.56	19

Table 5.5: Average number of partitioned scenarios.

n	N	MOSEK [SOCP-SD]	[SpOCP-SD]	[LpOCP-SD]	[LogExpCP-SD]
50	1000	80.3	24.8	21.3	61.8
	2500	180.8	47.8	47.0	77.8
	5000	349.3	80.3	79.0	104.6
	10000	711.6	133.4	128.3	154.3
	25000	1834.9	232.0	318.3	178.2
	50000	3582.1	445.4	675.0	841.7
	100000	6945.1	774.1	1346.4	1447.5
100	1000	87.2	32.0	27.0	81.4
	2500	191.2	73.6	74.1	107.8
	5000	367.6	107.4	102.4	192.2
	10000	711.1	148.9	156.9	229.7
	25000	1808.6	278.1	348.6	1869.1
	50000	3802.9	457.8	729.7	2418.6
	100000	7323.3	831.3	1395.8	923.4
200	1000	108.2	39.5	36.4	100.7
	2500	201.7	72.7	73.0	154.5
	5000	395.6	116.3	119.6	198.1
	10000	744.0	184.9	171.2	304.6
	25000	1805.5	308.3	347.0	464.2
	50000	3607.8	512.2	697.6	788.1
	100000	7198.9	865.0	1384.3	1153.5

Next, the same problem is solved using using the polyhedral approximation cutting-plane method described in Section 5.3.1. Table 5.3 shows the running times achieved by all three solvers for problems [LpOCP] and [LpOCP-SD] with $p = 3$. In this case, the scenario decomposition method resulted in order-of-magnitude improvements, which can be attributed to the "warm-start" capabilities of CPLEX and GUROBI's simplex solvers. Consistent with these conclusions is also the fact that the simplex-based solvers of CPLEX and GUROBI yield improved solution times on the full problem formulation comparing to the SOCP-based reformulation [SpOCP], where barrier solvers were invoked. The discrepancy between [LpOCP] and [LpOCP-SD] solution times is especially prominent for MOSEK, but in this case it appears that MOSEK's interior-point LP solver was much less effective at solving the [LpOCP] formulation using the cutting plane method.

Finally, Table 5.4 displays the running times for the discussed implementation of problems [LogExpCR] and [LogExpCP-SD]. Of the three solvers considered in this case study, only MOSEK was capable of handling problems with constraints that involve sums of univariate exponential functions. Again, the scenario decomposition-based solution method appears to be preferable in comparison to solving the full formulation. Note, however, that computational times were not averaged over 20 instances in this case due to numerical difficulties associated with the solver for many instances of [LogExpCP].

It is also of interest to comment on the number of scenarios that had to be generated during the scenario decomposition procedure in order to yield an optimal solution. Table 5.5 lists the corresponding average number of scenarios partitioned for each problem type over all instances. Although these numbers may slightly differ among the three solvers, we only present results for MOSEK as it was the only solver used to solve all the problem in Sections 5.3.1 and 5.3.2. Observe that far fewer scenarios are required relative to the total set size N. In fact, as a percentage of the total number of scenarios, the number of scenarios that were generated during the algorithm in order to achieve optimality was between 0.7% and 11% of the total scenario set size.

5.4 Concluding Remarks

In this chapter, we proposed an efficient algorithm for solving large-scale convex stochastic programming problems that involve a class of risk functionals in the form of infimal convolutions of certainty equivalents. We exploit the property induced by such risk functionals that a significant portion of scenarios is not required to obtain an optimal solution. The developed scenario decomposition technique is contingent on the identification and separation of “non-redundant” scenarios by solving a series of smaller relaxation problems. It is shown that the number of iterations of the algorithm is bounded by the number of scenarios in the problem. Numerical experiments with portfolio optimization problems based on simulated return data following the covariance structure of randomly chosen S&P500 stocks demonstrate that significant reductions in solution times may be achieved by employing the proposed algorithm. Particularly, performance improvements were observed for the large-scale instances when using HMCR measures with $p = 2, 3$, and LogExpCR measures.

CHAPTER 6
MIXED-INTEGER NONLINEAR PROGRAMMING WITH CERTAINTY EQUIVALENT MEASURES OF RISK

6.1 Introduction: Problem Formulation and Solution Approach

As presented in the previous chapter, one of the computational challenges associated with the introduction of the modeling approaches proposed in Chapter 2 comes from the large number of linear constraints linked to the scenario variables. In the current chapter we will explore the other two computational issues: the presence of the nonlinear constraint and integer variables. While our main interest is tied to the stochastic programming applications presented earlier, the solutions methods described in this work are not application-specific, i.e., can be used in other areas, as long as the problem of interest can be formulated as given below.

The general problem formulation considered in the current work can be expressed as follows.

$$\min \quad \mathbf{c}^\top \boldsymbol{x} \tag{6.1a}$$

$$\text{s. t.} \quad v_k^{-1}\left(\sum_{j=1}^{m_k} p_j^k v_k\left(\sum_{i=1}^{n} a_{ij}^k x_i + b_j^k\right)\right) \le \sum_{i=1}^{n} a_{i0}^k x_i + b_0^k, \quad k = 1, \ldots, K \tag{6.1b}$$

$$\mathbf{H}\boldsymbol{x} \le \mathbf{h} \tag{6.1c}$$

$$\boldsymbol{x} \in \mathbb{Z}_+^{n_1} \times \mathbb{R}_+^{n_2}. \tag{6.1d}$$

Here, $n = n_1 + n_2$ is the dimensionality of the mixed-integer decision vector $\boldsymbol{x}$, and $\mathbf{c}$, $\mathbf{h}$, $\mathbf{H}$ are vectors and a matrix of appropriate dimensions.

The main object of interest in problem (6.1) is the set of constraints (6.1b), where it is assumed that coefficients p_j^k are positive, $p_j^k > 0$, for all values of j and k, and functions $v_k : \mathbb{R} \mapsto \mathbb{R}$, $k = 1, \ldots, K$, have the following special properties (in accordance to the results obtained in Chapter 2)

(i) $v_k(t) = 0$ for $t \leq 0$

(ii) $v_k(t)$ are increasing and convex for $t \geq 0$

(iii) v_k are such that constraints (6.1b) are convex.

To simplify the exposition and notation, in what follows we are going to suppress index k in (6.1b), essentially assuming a single nonlinear constraint in problem (6.1), $K = 1$. Then, given the above assumptions on function v, it is straightforward to see that problem (6.1) can be rewritten in the form

$$\min \quad \mathbf{c}^\top \boldsymbol{x} \tag{6.2a}$$

$$\text{s. t.} \quad w_0 \geq v^{-1}\left(\sum_{j=1}^{m} p_j v(w_j)\right) \tag{6.2b}$$

$$w_j \geq \sum_{i=1}^{n} a_{ij} x_i + b_j, \quad j = 1, \ldots, m \tag{6.2c}$$

$$w_0 \leq \sum_{i=1}^{n} a_{i0} x_i + b_0 \tag{6.2d}$$

$$\mathbf{w} \geq \mathbf{0} \tag{6.2e}$$

$$\mathbf{Hx} \leq \mathbf{h} \tag{6.2f}$$

$$\boldsymbol{x} \in \mathbb{Z}_+^{N_1} \times \mathbb{R}_+^{N_2}, \tag{6.2g}$$

We will also assume that the set $\mathfrak{X}$ is polyhedral, so that the main challenges associated with solving problem (6.2) arise from the nonlinear constraint (6.2b) and the integrality constraints (6.2f).

In the context of risk-averse stochastic programming as it was discussed in Chapter 5, the decision making problem of minimizing certainty equivalent risk measure subject to feasibility of the decision vector can be expressed as

$$\min_{\boldsymbol{x}\in\mathcal{X},\eta\in\mathbb{R}} \eta + \frac{1}{1-\alpha} v^{-1}\Big(\sum_{j=1}^{m} p_j v([X(\boldsymbol{x},\omega_j)-\eta]_+)\Big), \tag{6.3}$$

where ω takes value ω_j with the probability p_j for $j = 1, \ldots, m$. If additionally we assume that the loss outcome is a linear function of the decision vector, i.e., $X(\boldsymbol{x}, \omega_j) = \boldsymbol{a}_j^\top \boldsymbol{x} + b_j$, and $\boldsymbol{x} \in \mathbb{Z}^{n_1} \times \mathbb{R}^{n_2}$ then (6.3) can be reformulated as

$$\min \quad \eta + \frac{1}{1-\alpha} t \tag{6.4a}$$

$$\text{s. t.} \quad t \geq v^{-1}\Big(\sum_{j=1}^{m} p_j v(w_j)\Big) \tag{6.4b}$$

$$w_j \geq \boldsymbol{a}_j^\top \boldsymbol{x} + b_j - \eta, \quad j = 1, \ldots, m \tag{6.4c}$$

$$\boldsymbol{x} \in \mathbb{Z}^{n_1} \times \mathbb{R}^{n_2}, \boldsymbol{w} \geq 0. \tag{6.4d}$$

Comparing this formulation to (6.2) we can observe that (6.4) is a special case of the problem above. Since the focus of the current work is on the computational methods addressing the nonlinear and integrality constraints in (6.2), the difference between these two formulations is essentially irrelevant to our discussion in the reminder of this chapter. Additionally, in view of this discussion, we will refer to constraint (6.2b) as a *certainty equivalent constraint*.

The two most influential ideas for solution procedures in mixed-integer programming are: branch-and-bound algorithm and valid inequalities. Development of both of this approaches in relation to problem (6.2) will be addressed in this chapter. In Section 6.2

we will present a version of branch-and-bound method targeted at the specific nonlinear constraints considered in this chapter. Next, in Section 6.3 we will address two procedures for generating inequalities valid for the feasible set of (6.2): lifted nonlinear cuts and disjunctive cuts. Finally in Section 6.4 we will present some results of numerical experiments. Relevant literature review will be presented in Sections 6.2 and 6.3.

The main contributions of this chapter in our view are twofold. First, we show that two techniques (a special implementation of a branch-and-bound and lifted nonlinear valid inequalities) considered in the previous chapters in the case of MIpOCP problems can be extended to the more general case considered here. While, both of this extensions do not require novel theoretical development, heavily relying on the results already established, the novelty of the problem formulation justifies, in our view, our interest in these extensions. Particularly, we show how these techniques can be reformulated in order to address this new application area, while still allowing for the use of the already existing theoretical basis. Secondly, we propose another numerical approach, which relies on a simple geometric idea for construction of linear disjunctive cuts. To the best of our knowledge such a process have not been considered in the literature before.

6.2 Branch-and-Bound based on Outer Polyhedral Approximations

6.2.1 Existing Methods and Approach Due to Vielma et al.

Branch-and-bound (BnB) methods for solving MINLP problems are often divided into two categories depending on the way continuous relaxations are handled. The first group consists of the methods which solve exact non-linear continuous relaxation, usually using some version of an interior point method to deal with the arising NLP problems (see, for example Gupta and Ravindran, 1985; Borchers and Mitchell, 1994; Leyffer, 2001 and references therein). Alternatively, polyhedral approximations can be employed to help with finding approximate solutions of the continuous relaxations (Duran and Grossmann, 1986; Fletcher and Leyffer, 1994; Quesada and Grossmann, 1992; Bonami et al., 2008; Vielma et al., 2008). This approach has been the basis for a few MINLP solvers such as Bonmin (Bonami et al., 2008), FilMINT (Abhishek et al., 2010) or AOA (AIMMS open MINLP solver). For example, outer approximation algorithms (AOA) solve alternating sequence of MILP master problems and NLP subproblems, while in LP-NLP-based branch-and-bound (Quesada and Grossmann, 1992, FilMINT) the solution of a single master mixed-integer linear programming (MILP) problem is terminated every time an integer valued candidate is found to solve an exact NLP, solution of which is then used to generate new outer approximations.

Another framework has been proposed in Vielma et al. (2008) for the case of MISOCP problems. The authors exploit the fact that there exists an extremely efficient lifted outer polyhedral approximation of second order cones, and thus they propose to solve full-sized approximating LP at each node of the master MILP, while, as previously, an exact NLP is solved every time a new integer solution is found. Note that in this case, the algorithm is guaranteed to find a solution that is ε-feasible at each node of the BnB tree (essentially, new approximating hyperplanes are generated in every node in order to ensure that), as opposed to LP-NLP approach, where NLP solution is used to generate new approximating facets. Hence, one of the key differences between different implementations of such branch-and-bound methods can be viewed as a trade-off between the size of approximating LPs (i.e., the accuracy of the approximation) and the number of exact NLPs that need to be solved. Note that an exact NLP, of cause, provides tighter lower bounds, and thus, more pruning capabilities, while LPs bring-in superior warm-start efficiencies, consequently speeding up the processing time in each node. In this sense, the approach due to Vielma et al. (2008) can be viewed as the most conservative in terms of the use of the exact solvers: NLPs are only solved when absolutely necessary to verify incumbent integer solutions.

The fact that this approach relies on an efficient lifted approximation scheme is essential, since it may require an exponential number of facets to achieve a guaranteed ε-feasibility for general nonlinear constraints. The main source of difficulty here can be associated with high dimensionality of the constraint, i.e., it can be seen as a manifestation of the "curse of dimensionality". In Chapter 3 we have shown that this framework can be competitive even when no such efficient approximation scheme is available by designing a branch-and-bound based on polyhedral approximations for MIpOCP problems. The key idea there was the introduction of a cutting plane generation procedure for approximately solving continuous pOCP relaxations. In the next subsection we are going to demonstrate that a similar approach is feasible in the more general setting considered in the current chapter.

6.2.2 Lifted Approximation Procedure

In the case of MISOCP problems the goal of defining an efficient approximation scheme can be accomplished with a lifted approximation due to Ben-Tal and Nemirovski (2001b). The approach proposed there can be described as a two-step procedure, where in the first step a lifting technique is used in order to reduce the dimensionality of the problem, and then an efficient way of approximating low-dimensional second-order sets is proposed. In Chapter 3 the second step of this procedure has been replaced by an efficient cutting plane generation framework. In the current chapter we again utilize the first step of the lifting procedure due to Ben-Tal and Nemirovski (2001b), and then investigate low-dimensional cutting plane generation possibilities.

The technique proposed in Ben-Tal and Nemirovski (2001b) for dimensionality reduction has been dubbed by the authors as *tower-of-variables*. It lets them represent a single $m+1$ dimensional second-order cone as an intersection of m three-dimensional cones. In Chapter 3 we demonstrated that the same procedure can be employed for p-order cones, and moreover a more compact implementation is possible. Here we observe that the structure of constraint (6.1b) allows to employ a similar process summarized in the proposition below.

Proposition 6.1. *The set in the space of variables* $w_0, \ldots, w_m$ *defined by the system of constraints*

$$w_0 = w_{2m-1},\ v(w_{m+j}) \geq \beta_{2j-1} v(w_{2j-1}) + \beta_{2j} v(w_{2j}), \quad j = 1, \ldots, m-1, \tag{6.5}$$

where $\beta_j = p_j$ *for* $j = 1, \ldots, m$ *and* $\beta_j = 1$ *for* $j = m+1, \ldots, 2m-1$ *and* $w_{m+1}, \ldots$ *are* *lifting variables, is equivalent to set defined by* (6.2b).

Proof: The claim of the proposition can be verified directly. □

Following from this proposition, the second step of the approximation procedure described above is reduced to the problem of approximating a set of 3D constraints of the form

$$v(w_0) \geq \beta_1 v(w_1) + \beta_2 v(w_2), \tag{6.6}$$

or equivalently,

$$w_0 \geq v^{-1}\Big(\beta_1 v(w_1) + \beta_2 v(w_2)\Big) =: f(w_1, w_2). \tag{6.7}$$

Due to this dimensionality reduction it can be completed using a simple gradient approximation approach, i.e., through the use of tangent planes. Before continuing with this approach, it is necessary to comment on the precise definition of approximation that we use in the current context.

Recall that for a second-order cone, $\mathcal{L}_2 = \{w \in \mathbb{R}^{m+1} \mid w_0 \geq \|(w_1, \ldots, w_m)\|_2\}$, a point $w \in \mathbb{R}^{m+1}$ is usually said to be ε-feasible if $(1+\varepsilon)w_0 \geq \|(w_1, \ldots, w_m)\|_2$ (Ben-Tal and Nemirovski, 2001b). At the same time, it will be clear from the discussion below that due to the lack of the conic property in the general case such an approach is not readily applicable. Whence, we propose a slightly different definition. Namely, we will call a point (w_0, w_1, w_3) ε-feasible towards (6.6), if

$$(1+\varepsilon)w_0 + \varepsilon \geq f(w_1, w_2). \tag{6.8}$$

Note that this close with the "multiplicative" definition of approximation for second-order conic constraints for larger values of w_1, w_2, while it relies on the "additive" part when w_1, w_2 are small.

Now, since the relaxed feasible set considered in the current work is convex, a cutting plane defined as

$$w_0 \geq f(w_1^*, w_2^*) + f'_{w_1}(w_1^*, w_2^*)(w_1 - w_1^*) + f'_{w_2}(w_1^*, w_2^*)(w_2 - w_2^*), \tag{6.9}$$

which is tangent to the 3-dimensional set (6.7) at point $(f(w_1^*, w_2^*), w_1^*, w_2^*)$, is globally feasible. Observe that

$$\begin{aligned} f(w_1, w_2) =& f(w_1^*, w_2^*) + f'_{w_1}(w_1^*, w_2^*)(w_1 - w_1^*) + f'_{w_2}(w_1^*, w_2^*)(w_2 - w_2^*) \\ &+ R_{11}(w_1^*, w_2^*)(w_1 - w_1^*)^2 + \\ &2R_{12}(w_1^*, w_2^*)(w_1 - w_1^*)(w_2 - w_2^*) + R_{22}(w_1^*, w_2^*)(w_2 - w_2^*)^2 \\ \leq& w_0 + R_{11}(w_1^*, w_2^*)(w_1 - w_1^*)^2 + \\ &2R_{12}(w_1^*, w_2^*)(w_1 - w_1^*)(w_2 - w_2^*) + R_{22}(w_1^*, w_2^*)(w_2 - w_2^*)^2 \end{aligned}$$

where $R_{ij}(w_1, w_2)$ are the coefficients of the remainder of Taylor's expansion.

Next, we will make the following regularity assumptions in addition to (i)–(iii) above:

(iv) function v is two times continuously differentiable

(v) $v(t) > 0$ for all $t > 0$

(vi) feasible region of problem (6.2) is bounded.

Assumption (iv) will simplify our further analysis and let us establish iteration complexity below. At the same time it can be noted that the existence of cutting plane (6.9) follows from convexity assumption and does not require a more strict claim (iv). In other words, the cutting plane procedure described below can still be applied even if (iv) does not hold.

Given these assumptions, it is easy to see that $v'(t) > 0$ for all $t > 0$ since v is nondecreasing, convex and strictly positive on $t > 0$. This and assumption (vi) above imply that f' and f'' are bounded on the feasible region of (6.2). Hence, usual calculus analysis (see, for example Thomas et al. (2007), Section 14.10) suggests that $R_{ij}(w_1, w_2) \leq \max_{i,j} \max_{w_1,w_2} f''_{ij}(w_1, w_2) = M < +\infty$.

Now, if (w_1, w_2) and (w_1^*, w_2^*) are sufficiently close, for instance, $|w_1 - w_1^*| \leq \delta$ and $|w_2 - w_2^*| \leq \delta$, then

$$f(w_1, w_2) \leq w_0 + 4M\delta^2.$$

Suppose that a point (w_0, w_1, w_2) satisfies a set of cutting plane constraints (6.9). If it is at least $\delta = \sqrt{\frac{\varepsilon}{4M}}$ close (in terms of ℓ_1-norm) to one of the pairs (w_1^*, w_2^*) used to define cutting planes (6.9), then it satisfies $w_3 \geq f(w_1, w_2) + \varepsilon$. Of cause, this observation is a simple manifestation of the well-known fact that a sufficiently smooth and convex set in three dimensions can be approximately described by a number of its supporting planes. We still include this explicit analysis here, because, in our opinion, it provides a clear intuitive explanation for Proposition 6.2 below.

Putting together our development so far we can formulate the following procedure for approximately solving a continuous relaxation of (6.2). Consider a master problem in the form of (6.2), where nonlinear constraint is substituted with a set of cutting planes (6.9):

$$\begin{aligned} \min \quad & \mathbf{c}^\top \boldsymbol{x} \\ \text{s. t.} \quad & w_{m+j} \geq f\left(w_1^{k_j}, w_2^{k_j}\right) + f'_{w_1}\left(w_1^{k_j}, w_2^{k_j}\right)\left(w_{2j-1} - w_1^{k_j}\right) + \\ & \qquad\qquad\qquad f'_{w_2}\left(w_1^{k_j}, w_2^{k_j}\right)\left(w_{2j} - w_2^{k_j}\right), \\ & \quad j = 1, \ldots, m-1,\ k_j = 1, \ldots, K_j, \\ & \text{(6.2c)–(6.2g)}, \end{aligned}$$

where K_j is the number of cutting planes on variables $w_{m+j}, w_{2j-1}, w_{2j}$, derived around the pairs $(w_1^{k_j}, w_2^{k_j})$, $k_j = 1, \ldots K_j$. Then, given a $\boldsymbol{w}^*$ (a current solution of the master problem), we can add new constraints (6.9) around pairs (w_{2j-1}^*, w_{2j}^*), for those j for which approximation condition $(1+\varepsilon)w_{m+j}^* + \varepsilon \geq f(w_{2j-1}^*, w_{2j}^*)$ is violated. Then the master can be resolved and the iterative process continues. The following proposition establishes some of the properties of the procedure.

Proposition 6.2. *Suppose that for a given solution $\boldsymbol{w}^*$ of the master cuts in the form of* (6.9) *are added if condition* (6.8) *is not satisfied for a specific triple* $(w_{m+j}, w_{2j-1}, w_{2j})$, $j = 1, \ldots, m-1$ *as described above. Assuming that the feasible region is bounded, this cutting plane procedure terminates after a finite number of iterations for any given* $\varepsilon > 0$. *In particular, the algorithm is guaranteed to generate at most* $O(\varepsilon^{-1.5})$ *cutting planes.*

Proof: As shown in the discussion above, a new master problem solution cannot be δ-close (in terms of L^1 norm) to one of the previous solutions and still violate condition (6.8), and thus only a finite number of new cutting planes can be generated. In other words, each generated cutting plane effectively occupies a cube with volume $8\delta^3$. If V is the volume of the feasible region for (w_0, w_1, w_2), then the maximum number of cutting planes can be estimated as $\frac{VM^{1.5}}{\varepsilon^{1.5}}$, if δ is selected as $\delta = \sqrt{\frac{\varepsilon}{4M}}$. □

Observe that this result essentially provides an exact algorithm for solving problem (6.2). Indeed, once a solution with a desired accuracy ε is found, an improved solution can be constructed by adding new cutting planes. In other words, such a procedure can yield a solution that is arbitrary close to the solution of the exact problem.

Consequently, compare this result to a similar proposition we presented earlier in Chapter 3. It has been shown there that a cutting plane approximation procedure is guaranteed to terminate with ε-feasible solution in $O(\varepsilon^{-1})$ iterations for p-order cone programing and $O(\varepsilon^{-0.5})$ in the case of second-order cones. While the claim proved here is weaker, it is in accordance with our expectation, since due to conic property, the analysis of cutting plane procedure in p-order cones can be essentially carried through in two dimensions (polar coordinates). Lack of this property in the case of certainty equivalent constraints, hence, leads to a less tight bound on the number of iterations. At the same time, all our experiments with both conic and nonconic problems suggest that in practice, only a small fraction of all possible facets is generated, i.e., the fact that this bound is very restrictive, may not hurt computational performance.

6.2.3 Branch-and-Bound Method

Now that an efficient approximation procedure for solving continuous relaxations is determined, it can be incorporated in a branch-and-bound method due to Vielma et al. (2008).

Namely, we consider a master mixed-integer linear programming (MILP) problem (we will denote it as P_1) which is constructed from problem (6.2) by substituting (6.2b) with a set of initial cutting planes of the form of (6.9). The solution procedure consists of applying a regular branch-and-bound method to P_1 with two adjustments. First, lower bounds obtained from the continuous relaxations of P_1 are found by applying the approximation scheme due to Proposition 6.2 with a preselected value of $\varepsilon = \varepsilon_1$. Note that it is not required to remove any of the added cutting planes before proceeding to the next node of the solution tree, since these constraints are globally feasible. Second, when an integer-valued solution of P_1 is found, in order to check its feasibility with respect to the exact nonlinear formulation and declare incumbent or branch further, the exact continuous relaxation of P_1 must be solved with bounds on the relaxed values of variables $\boldsymbol{x}$ determined by the integer-valued solution in question (see, Vielma et al. (2008) for more details and formal analysis). In order solve the exact relaxation, we once again employ Proposition 6.2, that is to say, we construct a second problem P_2, which represents a continuous relaxation of (6.2). In this case, we solve it using the same cutting plane procedure due to Proposition 6.2 but with $\varepsilon = \varepsilon_2 \ll \varepsilon_1$ instead. A sufficiently small value of ε_2 guarantees an essentially exact solution.

Note that it has been previously observed (see, Vielma et al. (2008); Vinel and Krokhmal (2014b) and Chapter 3 of the current work) that ε_1 can be selected to be relatively large and still provide promising computational results, which explains the relation $\varepsilon_2 \ll \varepsilon_1$ above. Note also that in this case, the described procedure can be viewed as repetitive resolving of a relatively small-scale LPs due to P_1, guided by a regular branch-and-bound, with occasional calls to a large-scale P_2.

6.3 Valid Inequalities

6.3.1 Existing Approaches

It is well-known in the literature that valid inequality theory has been essential in development of efficient solvers, particularly in mixed-integer linear programming (MILP). Building on this success various approaches to generating valid inequalities have been proposed for mixed-integer nonlinear programming (MINLP) problems. To name a few: Atamtürk and Narayanan (2010) and Atamtürk and Narayanan (2011) have proposed mixed inter rounding (MIR) and conic lifted cuts for conic programming problems; Stubbs and Mehrotra (1999) studied cutting plane theory in 0-1 mixed-convex programming; Çezik and Iyengar (2005) proposed Chvatal-Gomory cuts in conic programming; Bonami (2011) have considered lift-and-project cuts. There have also been a series of publications addressing possible approaches to designing disjunctive (or split) cuts in MINLP (for example, Burer and Saxena 2012; Cadoux 2010; Kılınç et al. 2010; Modaresi et al. 2015; Saxena et al. 2008 among others).

In this section we consider two approaches for generation of valid inequalities for the mixed-integer programming problem (6.2). First we are going to discuss lifted non-linear cuts, building on the developments presented in Chapter 4. Next, we will present a simple geometric argument that allows us to construct a class of linear disjunctive cuts valid for our feasible set.

6.3.2 Lifted Non-Linear Cuts

A lifting procedure for conic mixed-integer programming has been proposed in Atamtürk and Narayanan (2011). Authors introduced a lifting scheme, which provides a way of generating new conic valid inequalities for mixed-inter conic sets. We have employed this approach for solving MIpOCP problems in Chapter 4 and obtained promising numerical results for a class of risk-averse portfolio optimization models. While this technique has been proposed as a way to generate conic cuts for conic feasible sets, as we show below, it can be extended for certainty equivalent constraints as well. As will be clear from our discussion below, our main contribution here lies in the reformulation of the procedure in nonconic terms, while most of the proofs directly follow from the previous developments in Atamtürk and Narayanan (2011) and Chapter 4.

Consider set $V = \{\boldsymbol{x} \in \mathbb{R}^{m+1} \mid \sum_{j=1}^{m} p_j v(x_j) \leq v(x_0)\}$, which is going to play the role of a conic set in Atamtürk and Narayanan (2011). We can then define

$$T^n(\boldsymbol{b}) = \Big\{x^i \in X^i \,|\, \boldsymbol{b} - \sum_{i=0}^{n} A_i x^i \in V\Big\}, \tag{6.10}$$

where each X^i is a mixed-integer set in $\mathbb{R}^{n_i}$ and A^i and b^i are of appropriate dimensions. Suppose that $u : \mathbb{R} \mapsto \mathbb{R}$ satisfies the same assumptions as function v. Let us further suppose that inequality $h - F^0 x^0 \in U$ is valid for $T^0(\boldsymbol{b})$. Atamtürk and Narayanan (2011) show how this inequality can be lifted by computing $F^1, \ldots, F^i$ resulting in valid cuts for $T^i(\boldsymbol{b})$ $h - \sum_{\ell=0}^{i} F^\ell x^\ell \in U$. when sets V and U are proper cones. By repeating arguments of Atamtürk and Narayanan (2011) the following theorem can be shown to hold for a lifting set $\Phi^i(\boldsymbol{v})$ defined as

$$\Phi^i(\boldsymbol{v}) = \Big\{\partial \in \mathbb{R}^p \mid h - \sum_{i=0}^{n} F^i x^i - \boldsymbol{d} \in U \text{ for all } (x_0, \ldots, x^i) \in T^i(\boldsymbol{b} - \boldsymbol{v})\Big\}.$$

Theorem 6.3. *1. $\Phi^i(\boldsymbol{v})$ is closed and convex.*

2. $0 \in \Phi^i(0)$

3. $\Phi^{i+1}(\boldsymbol{v}) \subset \Phi^i(\boldsymbol{v})$

4. $F^1, \ldots, F^{i+1}$ generate a valid inequality for $T^{i+1}(\boldsymbol{b})$ iff $F^{i+1}x^i \in \Phi^i(A^{i+1}x^i)$ for all x^i.

5. If $\Omega(\boldsymbol{v}) \subset \Phi_0(v)$ is superadditive, then $F^1, \ldots, F^{i+1}$ generate a valid inequality for $T^{i+1}(\boldsymbol{b})$ whenever $F^{i+1}x^i \in \Omega(A^{i+1}x^i)$ for all x^i.

Proof: Since the arguments establishing the analogous results in Atamtürk and Narayanan (2011) do not rely on the conic assumption we believe it to be unnecessary to repeat those here. □

As it was noted above, we employed an analogous result for the case of p-order conic constraints ($V = \{\boldsymbol{x} \in \mathbb{R}^{m+1} \mid x_0 \geq \|\boldsymbol{x}\|_p\}$) in Chapter 4. As it turns out, the reasoning we used there does not rely on the conic property as well, i.e., our results from Chapter 4 can be carried through without major changes. Particularly, we can consider a set $\widehat{T}^n(b)$ as

$$\widehat{T}^n(b) = \Bigg\{(\boldsymbol{x}, \eta_+, \eta_-, y, t) \in \mathbb{Z}^n_+ \times \mathbb{R}^4_+ \ : \ v\bigg(\Big[\sum_{i=1}^{n} a_i x_i + \eta_+ - \eta_- - b\Big]_+\bigg) + v(y) \leq v(t)\Bigg\},$$

and then show that the following claim holds.

Proposition 6.4. *Inequality*

$$v\left(\left[(1-f)(x-\lfloor b\rfloor)+\sum_{i=1}^{n}\alpha_i x_i\right]_+\right)+v(y)\le v(t) \tag{6.11}$$

is valid for $\widehat{T}^n(b)$*, where* $\alpha_i = \left\lceil\dfrac{a_i - b + \lfloor b\rfloor(1-f)}{M}\right\rceil_+$*,* $f = b - \lfloor b\rfloor$*, and* M *is such that* $x_i \le M$ *for all* i*.*

This result is a very restricted application of Theorem 6.3. Indeed here we are considering the case when the set U is the same as the initial set V and moreover, not only all the analysis is restricted to three-dimensional nonlinear constraints, but also the second dimension (represented by variable y) is assumed to be continuous (in other words, integral structure of the second dimension is relaxed). Despite these simplifications we have demonstrated in Chapter 4 that such an approach may yield promising computational results in MIpOCP problems. In Section 6.4 we will numerically analyze this procedure in mixed-integer programming with certainty equivalent constraints. Moreover, two of these stipulations can be in fact viewed as natural assumptions for the task of deriving valid inequalities in our case. Observe that due to the tower-of-variables techniques presented in Section 6.2 the constraints are already represented in three-dimensional form, and furthermore, it is also highly undesirable from computational perspective to deal with U different from initial set V, since this would result in additional numerical challenges associated with the new type of nonlinearity in the problem.

6.3.3 Linear Disjunctive Cuts

Throughout this section we will use the following notation: $\bar{\boldsymbol{x}} = (x_0, \boldsymbol{x}) \in \mathbb{R}^{n+1}$. We will also reformulate certainty equivalent constraints as

$$\begin{aligned} \bar{\boldsymbol{x}} \in \mathcal{K}, \quad \mathcal{K} := \left\{ \bar{\boldsymbol{x}} \in \mathbb{R}^{n+1} \,\middle|\, F(\boldsymbol{x}) \le x_0 \right\}, \\ F(\boldsymbol{x}) := v^{-1}\Big(\sum_{j=1}^{m} v(|\boldsymbol{a}_j^\top \boldsymbol{x} + b_j|) \Big), \quad \boldsymbol{x} \in \mathbb{Z}^n_+. \end{aligned} \tag{6.12}$$

Note that we consider $F(\boldsymbol{x}) := v^{-1}\Big(\sum_{j=1}^{m} v(|\boldsymbol{a}_j^\top \boldsymbol{x} + b_j|) \Big)$ instead of possible $F(\boldsymbol{x}) := v^{-1}\Big(\sum_{j=1}^{m} v([\boldsymbol{a}_j^\top \boldsymbol{x} + b_j]_+) \Big)$ which would be in accordance with stochastic programming motivations. Such a choice simplifies some of our development below, and since it results in a relaxed set $\mathcal{K}$ any valid inequality obtained for $\mathcal{K}$ will be valid for problem (6.2) as well.

Disjunctive or split cuts have been extensively studied in the literature, especially when applied to MIP problems (Balas, 1971). This approach is based on a very intuitive idea: consider disjunction $x_k \le \pi_0 \vee x_k \ge \pi_1 = \pi_0 + 1$ with $\pi_0 \in \mathbb{Z}_+$, where $k \in 1 \dots n$ is preselected. Due to integrality condition there are no feasible solutions outside of this disjunction, hence, system (6.12) implies that

$$\bar{\boldsymbol{x}} \in \operatorname{conv}\left(\left\{ \begin{matrix} \bar{\boldsymbol{x}} \in \mathcal{K} \\ x_k \le \pi_0 \end{matrix} \right\} \bigcup \left\{ \begin{matrix} \bar{\boldsymbol{x}} \in \mathcal{K} \\ x_k \ge \pi_1 \end{matrix} \right\} \right). \tag{6.13}$$

Consequently, any inequality describing this convex hull is valid for the feasible region of (6.12). Moreover, in the case of mixed-integer linear programming (MILP) all the sets involved (including the convex hull above) are polyhedral which substantially simplifies the construction procedures, and hence, increases the effectiveness of the cuts. There also exists a considerable amount of literature on generalizing this approach for MINLP problems (Burer and Saxena, 2012; Cadoux, 2010; Kılınç et al., 2010). Recently there have also been presented various efforts to design nonlinear disjunctive cuts, see Andersen and Jensen (2013); Belotti et al. (submitted); Bienstock and Michalka (2014); Modaresi et al. (2015); Burer and Kılınç (2014).

It has been shown (see, for example, Modaresi et al., 2015) that in some cases it may be possible to describe convex hull (6.13) using a single non-linear constraint, particularly, such a description is available for second-order conic set $\mathcal{K}$. Note that many of this works consider the problem in much more involved cases of general disjunctions.

With this in mind, consider certainty equivalent system (6.12). A first question that we could ask here is whether it is more desirable to find a closed-form nonlinear description of (6.13) following one of the recent developments mentioned above, or whether a linear description would be more useful in this case. Note that if such a nonlinear description is to be found and then used in a numerical procedure to solve problem (6.2), then it is highly desirable for it to be expressed in the same form as the nonlinear constraint already present. Indeed, since adding nonlinearity to a mixed-integer programming problem can substantially increase the difficulty of solving it, adding a new type of nonlinearity may further increase this complexity. Moreover, if the computational procedures used are tailored specifically to the constraints already present in the problem, then addition of a new type of nonlinear cut that is not comparable with these approaches may be impractical. The descriptions obtained in the literature for a second-order conic set $\mathcal{K}$ express the convex hull of the disjunction in terms of quadratic sets, essentially preserving the second-order conic nonlinearity in many practical cases, thus justifying the approach.

Consequently, consider (6.13) with certainty equivalent system (6.12). We can conclude that it is desirable that its description itself is represented in terms of function F defined in (6.12). At the same time, consider supporting hyperplanes for (6.13). It is easy to realize that there exist such hyperplanes that their intersection with the convex hull (6.13) is a straight line segment in between $x_k = \pi_0$ and $x_k = \pi_1$. On the other hand, a boundary of a set defined in terms of function F does not in general contain such segments, since it, in general is nonconic. Thus, it is reasonable to expect that such a closed-form description of convex hull (6.13) cannot be expressed in terms of function F alone. With this in mind, we propose to concentrate on a more modest goal of constructing supporting hyperplanes for (6.13), or in other words, linear disjunctive cuts.

Next will propose an intuitive idea for a procedure aimed at avoiding difficulties associated with the general disjunctive cut generation techniques available in the literature by exploiting specific structural properties of (6.13). Suppose that we have selected a $\bar{\boldsymbol{x}}^0 \in \mathcal{K}$ such that $x_k^0 = \pi_0$, i.e., $\bar{\boldsymbol{x}}^0$ is located on one side of the disjunction. Given such a $\bar{\boldsymbol{x}}^0$ find $\bar{\boldsymbol{x}}^1 \in \mathcal{K}$ such that $x_k^1 = \pi_1$ and $\partial_{n-1} F(\boldsymbol{x}^1) \cap \partial_{n-1} F(\boldsymbol{x}^0) \neq \emptyset$, where subdifferential ∂_{n-1} is taken with respect to variables x_i, $i \neq k$. A linear disjunctive cut is then constructed as a constraint $\sum_i \alpha_i x_i + \beta \leq x_0$, where $(\alpha_1, \ldots, \alpha_{k-1}, \alpha_{k+1}, \ldots, \alpha_n)^T \in \partial_{n-1} F(\boldsymbol{x}^0) \cap \partial_{n-1} F(\boldsymbol{x}^1)$, while α_k and β are selected in such a way that $\sum_i \alpha_i x_i^0 + \beta = x_0^0$ and $\sum_i \alpha_i x_i^1 + \beta = x_0^1$.

For illustrative purposes first assume that F is differentiable at both $\bar{\boldsymbol{x}}^0$ and $\bar{\boldsymbol{x}}^1$. In this case requirement $\partial_{n-1} F(\boldsymbol{x}^0) \cap \partial_{n-1} F(\boldsymbol{x}^1) \neq \emptyset$ reduces to $\frac{\partial F}{\partial x_i}(\boldsymbol{x}^0) = \frac{\partial F}{\partial x_i}(\boldsymbol{x}^0) = \alpha_i$ for all $i \neq k$. In other words, the constructed hyperplane is such that it passes through both $\bar{\boldsymbol{x}}^0$

and $\bar{\boldsymbol{x}}^1$ and is tangent to the sides of the disjunction at these points due to the requirement $\alpha_i = \frac{\partial F}{\partial x_i}(\boldsymbol{x}^0)$. Hence, such an inequality is necessary valid by construction since the set $\mathcal{K}$ is convex.

Formally, the described process can be formulated as follows. Given $\bar{\boldsymbol{x}}^0 \in \mathbb{R}^n \times \mathbb{R}$, $k \in \{1, \dots, n\}$, $\pi_0, \pi_1 \in \mathbb{Z}$, and function $v : \mathbb{R} \mapsto \mathbb{R}$, find $\bar{\boldsymbol{x}}^1 \in \mathbb{R}^n \times \mathbb{R}$ and $(\boldsymbol{\alpha}, \beta) \in \mathbb{R}^n \times \mathbb{R}$ such that

$$\begin{cases} \sum_{i=1}^{n} \alpha_i x_i^0 + \beta = x_0^0 & \text{(6.14a)} \\ \sum_{i=1}^{n} \alpha_i x_i^1 + \beta = x_0^1 & \text{(6.14b)} \\ x_k^0 = \pi_0 & \text{(6.14c)} \\ x_k^1 = \pi_1 & \text{(6.14d)} \\ F(\boldsymbol{x}^0) = x_0^0 & \text{(6.14e)} \\ F(\boldsymbol{x}^1) = x_0^1 & \text{(6.14f)} \\ (\alpha_1, \dots, \alpha_{k-1}, \alpha_{k+1}, \dots, \alpha_n)^\top \in \partial_{n-1} F(\boldsymbol{x}^0) \cap \partial_{n-1} F(\boldsymbol{x}^1) & \text{(6.14g)} \end{cases}$$

where $F(\boldsymbol{x}) := v^{-1}\Big(\sum_{j=1}^m v(|\boldsymbol{a}_j^\top \boldsymbol{x} + b_j|)\Big)$, and $\partial_{n-1} F(\boldsymbol{x})$ denotes subdifferential with respect to variables x_i for $i \neq k$. Let us denote

$$\mathcal{P} := \Big\{ \bar{\boldsymbol{x}} \in \mathbb{R}^{n+1} \;\Big|\; \sum_{i=1}^{n} \alpha_i x_i + \beta \leq x_0 \Big\}.$$

Then $\mathcal{K}$ represents the initial feasible set due to constraint (6.12) and $\mathcal{P}$ is the half space valid for the linear cut $\sum_{i=1}^n \alpha_i x_i + \beta \leq x_0$. By $\partial\mathcal{K}$ and $\partial\mathcal{P}$ we will understand boundaries of these sets. Next we will show that such a linear cut is valid for $\operatorname{conv}\Big(\{\mathcal{K}, x_k \leq \pi_0\} \cup \{\mathcal{K}, x_k \geq \pi_1\}\Big)$.

Observation 6.5. *The following claims hold:*

1. $\bar{\boldsymbol{x}}^i \in \partial\mathcal{K}$ *for* $i = 0, 1$
2. $\bar{\boldsymbol{x}}^i \in \partial\mathcal{P}$ *for* $i = 0, 1$
3. *if* $\bar{\boldsymbol{x}} \notin \mathcal{P}$ *and* $x_k = \pi_0$ *then* $\bar{\boldsymbol{x}} \notin \mathcal{K}$
4. *if* $\bar{\boldsymbol{x}} \notin \mathcal{P}$ *and* $x_k = \pi_1$ *then* $\bar{\boldsymbol{x}} \notin \mathcal{K}$

Proof: Claims (1) and (2) follow immediately from (6.14a)–(6.14b) and (6.14e)–(6.14f). In order to see that (3) holds, note that (6.14) implies that on the space restricted by $x_k = \pi_0$ the set $\partial\mathcal{P}$ is a supporting hyperplane for the set $\partial\mathcal{K}$, which immediately implies (3). Analogous observation holds for (4).

□

Observation 6.6. *The following claims hold:*

1. *If* $\bar{\boldsymbol{x}} \in \partial\mathcal{P}$ *and* $x_k < \pi_0$*, then* $\bar{\boldsymbol{x}} \notin \operatorname{int}\mathcal{K}$*.*
2. *If* $\bar{\boldsymbol{x}} \in \partial\mathcal{P}$ *and* $x_k > \pi_1$*, then* $\bar{\boldsymbol{x}} \notin \operatorname{int}\mathcal{K}$*.*

Proof: First consider claim (1). Suppose that the contrary holds, i.e., that $\bar{\boldsymbol{x}} \in \operatorname{int}\mathcal{K}$. Then there exists an $\varepsilon > 0$ such that $\bar{\boldsymbol{y}} = (x_0 - \varepsilon, \boldsymbol{x}) \in \mathcal{K}$ and $\bar{\boldsymbol{y}} \notin \mathcal{P}$. Now consider the segment connecting points $\bar{\boldsymbol{y}}$ and $\bar{\boldsymbol{x}}^1$, i.e., the set $\{\lambda\bar{\boldsymbol{y}} + (1-\lambda)\bar{\boldsymbol{x}}^1 \mid \lambda \in [0,1]\} =: \mathcal{T}$. Since both $\bar{\boldsymbol{y}} \in \mathcal{K}$ and $\bar{\boldsymbol{x}}^1 \in \mathcal{K}$, then $\mathcal{T} \subset \mathcal{K}$. Since $\pi_0 < \pi_1$ and $x_k < \pi_0$, then there exists $\bar{\boldsymbol{z}} = (z_0, \boldsymbol{z}) \in \mathcal{T}$ such that $z_k = \pi_0$. At the same time, $\bar{\boldsymbol{z}} \notin \mathcal{P}$ as $\bar{\boldsymbol{y}} \notin \mathcal{P}$ while $\bar{\boldsymbol{x}}^1 \in \partial\mathcal{P}$. Thus, by Observation 6.5 (3) $\bar{\boldsymbol{x}} \notin \mathcal{K}$ which contradicts the assumption above. Hence the claim (1) holds. Clearly, (2) can be proved analogously. □

Proposition 6.7. *If $\bar{\boldsymbol{x}} \in \mathcal{K}$ and $x_k \notin [\pi_0, \pi_1]$, then $\bar{\boldsymbol{x}} \in \mathcal{P}$.*

Proof: Again, suppose the contrary, i.e., that $\bar{\boldsymbol{x}} \notin \mathcal{P}$, which means that $\sum_{i=1}^{n} \alpha_i x_i + \beta > x_0$. Then, there exists $\bar{\boldsymbol{y}} = (x_0 + \varepsilon, \boldsymbol{x}) \in \partial\mathcal{P}$ (take $\varepsilon = \sum_{i=1}^{n} \alpha_i x_i + \beta - x_0$). Moreover, by definition $\bar{\boldsymbol{y}} \in \operatorname{int} \mathcal{K}$. If $x_k < \pi_0$ then this conclusion contradicts with Observation 6.6 (1), otherwise, $x_k > \pi_1$ and the conclusion above contradicts Observation 6.6 (2). □

This result guaranties that the cut generated by (6.14) is feasible for (6.13). Moreover, it is easy to see that for any $\widetilde{\beta} > \beta$ and $\widetilde{\boldsymbol{\alpha}} = \boldsymbol{\alpha}$ the corresponding cut is not feasible due to Observation 6.5 (1). Hence, system (6.14) produces a *tight cut* in the sense that it cannot be improved by an affine transformation.

Observe that $\bar{\boldsymbol{x}}^{(1)} \in \mathbb{R}^n \times \mathbb{R}$ and $(\boldsymbol{\alpha}, \beta) \in \mathbb{R}^n \times \mathbb{R}$ are the unknowns in the system (6.14). Given a specific value of $\boldsymbol{x}^1 \in \mathbb{R}^n$ such that $\partial_{n-1} F(\boldsymbol{x}^{(0)}) \cap \partial_{n-1} F(\boldsymbol{x}^{(1)}) \neq \emptyset$ it is easy to determine the rest. Indeed, one can readily note that x_0^1 is uniquely defined by (6.14f), $(\alpha_1, \ldots, \alpha_{k-1}, \alpha_{k+1}, \ldots, \alpha_n)^\top$ can be selected according to (6.14g) and α_k and β are fixed by (6.14a) and (6.14f). Thus, the most challenging step in this procedure is the selection of $\boldsymbol{x}^1$ satisfying $\partial_{n-1} F(\boldsymbol{x}^{(0)}) \cap \partial_{n-1} F(\boldsymbol{x}^{(1)}) \neq \emptyset$. Clearly, function F defined in (6.12) is piecewise continuously differentiable, yet, the choice of such an $\boldsymbol{x}^1$ can be difficult numerically, due to the nature of subdifferentials. Consequently, we propose to employ another approximation procedure in order to achieve this goal. Namely, we consider substituting $|t| \approxeq \sqrt{t^2 + \epsilon}$, and hence, defining $\widetilde{F}(\boldsymbol{x}) := v^{-1}\Big(\sum_{j=1}^{m} v\Big(\sqrt{(\boldsymbol{a}_j^\top \boldsymbol{x} + b_j)^2 + \epsilon} \Big) \Big)$. Then, $\widetilde{F}$ is continuously differentiable and in order to find $\boldsymbol{x}^1$ we propose to solve a system of nonlinear equations with a given $\boldsymbol{x}^0$

$$\frac{\partial \widetilde{F}}{\partial x_i}(\boldsymbol{x}^1) = \frac{\partial \widetilde{F}}{\partial x_i}(\boldsymbol{x}^{(0)}), \quad i = 1, \ldots, k-1, k+1, \ldots, n. \tag{6.15}$$

After system (6.15) is solved, the validity of the found $\boldsymbol{x}^1$ can be verified directly by comparing $\partial_{n-1} F(\boldsymbol{x}^{(0)})$ and $\partial_{n-1} F(\boldsymbol{x}^{(1)})$.

Remark 18. To this end we have not commented on the existence of a solution of (6.14). While we cannot formally show that $\bar{\boldsymbol{x}}_1$ defined in (6.14) necessarily exists in general, all our numerical experience suggests that this questions can be resolved positively. In other words, it seems to be the case that an appropriate $\bar{\boldsymbol{x}}_1$ can always be found in practice.

Finally, it is necessary to comment on the selection of k, π_0, π_1 and $\bar{\boldsymbol{x}}^0$. If the cut generation procedure is implemented in a branch-and-bound setting, it can be assumed that a solution of a relaxed problem $\bar{\boldsymbol{x}}^{\text{relax}}$ is known beforehand. Hence, it is natural to select $k \in \left\{ \{1, \ldots, n\} \,\middle|\, x_k^{\text{relax}} \notin \mathbb{Z} \right\}$, $\pi_0 = \lfloor x_k^{\text{relax}} \rfloor$ and $\pi_1 = \pi_0 + 1$. Since the goal of generating a valid inequality is to cutoff $\bar{\boldsymbol{x}}^{\text{relax}}$, then it is natural to pick $\bar{\boldsymbol{x}}^0$ according to $x_i^0 = x_i^{\text{relax}}$, for $i \neq k$, $x_k^0 = \pi_0$ and $x_0^0 = F(\boldsymbol{x}^0)$.

Before concluding this section, it is worth noting that the proposed procedure does not represent a general way to generate a split closure for the feasible set (6.12). Alternatively it can be seen as a quick and simple numerical procedure to find a valid inequality that can cutoff the current non-integral solution.

6.4 Numerical Experiments

In this section we will report some of the results of numerical case studies performed in order to evaluate the proposed techniques. As it has been explained in the introduction, our main interest in the problem class considered in this chapter stems from risk-averse approaches to stochastic programming, and hence we base our numerical experiments on this application area. Next, we will discuss particular formulation used in our study.

6.4.1 Model Formulation

We use the same two types of discrete portfolio optimization problems as in the previous chapters. Namely, lot-buying constrained problem:

$$\min_{\boldsymbol{x}\in\mathbb{R}^n_+,\ \boldsymbol{z}\in\mathbb{Z}^n_+} \left\{ \rho(-\mathbf{r}^\top\mathbf{x}) \;\middle|\; \mathsf{E}(\mathbf{r}^\top\mathbf{x}) \geq \bar{r}, \quad \mathbf{1}^\top\mathbf{x} \leq 1, \quad \boldsymbol{x} = \frac{L}{C}\,\mathrm{Diag}(\mathbf{p})\,\boldsymbol{z} \right\}, \tag{6.16}$$

where $\bar{r}$ is the prescribed level of expected return, $\boldsymbol{x} \in \mathbb{R}^n_+$ denotes the no-short-selling requirement, $\mathbf{1} = (1, \ldots, 1)^\top$, L is the size of the lot, C is the investment capital (in dollars), and vector $\mathbf{p} \in \mathbb{R}^n$ represents the prices of assets; and cardinality constrained problem:

$$\min_{\boldsymbol{x}\in\mathbb{R}^n_+,\ \boldsymbol{z}\in\{0,1\}^n} \left\{ \rho(-\mathbf{r}^\top\mathbf{x}) \;\middle|\; \mathsf{E}(\mathbf{r}^\top\mathbf{x}) \geq \bar{r}, \;\; \mathbf{1}^\top\mathbf{x} \leq 1, \;\; \mathbf{x} \leq \mathbf{z}, \;\; \mathbf{1}^\top\mathbf{z} \leq Q \right\}, \tag{6.17}$$

where Q is the maximum number of assets in the portfolio.

6.4.2 Implementation Remarks, Numerical Results and Conclusions

We used historical data for n stocks chosen at random out of stocks traded on NYSE, such that historical prices are available for 5100 consecutive trading periods preceding December, 2012. Returns over m consequent 10-day periods starting at a (common) randomized date were used to construct the set of m equiprobable scenarios for the stochastic vector $\mathbf{r}$. The values of parameters L, C, Q, α, and $\bar{r}$ were set as follows: $L = 1000$,

$C = 100{,}000$, $Q = 5$, $\alpha = 0.9$, $\bar{r} = 0.005$.

We used historical data for n assets chosen at random out of stocks traded on NYSE, such that historical prices are available for 5100 consecutive trading periods preceding December, 2012. Returns over m consequent 10-day periods starting at a (common) randomized date were used to construct the set of m equiprobable scenarios for the stochastic vector $\mathbf{r}$. The values of parameters L, C, K, α, and $\bar{r}$ were set as follows: $L = 1000$, $C = 100{,}000$, $M = 5$, $\alpha = 0.9$, $\bar{r} = 0.005$.

CPLEX MIP and LP solvers have been used to implement the branch-and-bound method described in Section 6.2. Namely, callback routines have been employed in order to add approximating hyperplanes at each node of the solution tree, while a goal framework was utilized to direct branching. The exact algorithm based on approximation scheme presented in Section 6.2 has been used to verify incumbent solutions.

The two families of valid inequalities have been employed through CPLEX callback routine. In our experiments we only added cuts in the root node of the solution tree. A quasi-Newton's method has been used to solve the underlying nonlinear systems of equations when finding linear split cuts presented in Section 6.3.3.

Two sets of experiments have been performed. First, the implementation of the branch-and-bound method from Section 6.2 has been compared against AIMMS AOA implementation. The results for lot-buying and cardinality constrained problems are summarized in Table 6.1 and 6.2 respectively. Observe that our custom implementation significantly outperforms AOA method for all choices of the parameters n and m. It is worth noting that, as it is stated in AIMMS manual, their implementation is much more efficient for binary variables, which is the case in our cardinality constrained problems. This observation explains the fact that in our experiments the improvement over AOA method has been less significant for this class of the problems. Overall, we can conclude that this study confirms that the branch-and-bound approach presented here can be seen as a viable strategy for solving the considered class of MINLP problems.

In the second stage of our case study, we aimed at evaluating the effect that valid inequalities defined in Section 6.3 can play in solving problems (6.16) and (6.17). Results of this case study are summarized in Table 6.3 and 6.4. Note that for each problem size 20 instances were generated and solved with a 1 hour time limit. We report the number of instances solved within the time limit, solution time and number of nodes in the branch-and-bound tree averaged over the instances that have been solved in 1 hour by all three approaches, and the average integrality gap among instances not solved to optimality.

We can observe that in both of the models the usage of the proposed valid inequalities leads to improved solution performance, especially for larger problems sizes. It is, in our view, particularly important to note that we are able to solve more problem instances within the time limit, as well as significantly reduce the integrality gap. It can also be noted that while in the case of lot-buying constrained problems the lifted cuts presented in Section 6.3.2 exhibit the best overall performance, in cardinality constrained optimization, this approach does not provide any improvement over pure branch-and-bound.

Table 6.1: Running time of AIMMS-AOA and the proposed implementation of the branch-and-bound method in lot-buying constrained portfolio optimization. Results averaged over 20 instances.

n	5			10			20		
m	10	50	100	10	50	100	10	50	100
CGBNB	**0.93**	**0.87**	**0.34**	**0.80**	**1.46**	**1.62**	**1.51**	**2.70**	**3.99**
AIMMS	49.17	67.10	73.55	104.43	151.35	221.19	195.29	618.61	7710.85

Table 6.2: Running time of AIMMS-AOA and the proposed implementation of the branch-and-bound method in cardinality constrained portfolio optimization. Results averaged over 20 instances.

n	10			20			50		
m	500	1000	2000	500	1000	2000	500	1000	2000
CG-BNB	**0.74**	**1.72**	**5.10**	**12.03**	**22.57**	**50.64**	**108.67**	**240.38**	**263.57**
AIMMS	11.65	35.90	96.88	294.74	459.21	639.43	863.50	1489.65	2071.98

Table 6.3: Performance of two valid inequality families in lot-buying constrained portfolio optimization. The rows refer to: *no cuts* – pure branch-and-bound presented in Section 6.2, *lifted* – lifted cuts from Section 6.3.2, *split* – disjunctive cuts introduced in Section 6.3.3. Results averaged over 20 instances. *Running time* and *nodes in solution tree* columns reflect only instances solved within 1 hour time limit by all three approaches. Similarly *gap after time limit* corresponds to instances for which no optimal solution was found within the time limit for each of the methods.

n	m	number solved			running time			nodes in solution tree			gap after time limit		
		lifted	split	no cuts	lifted	split	no cuts	lifted	split	no cuts	nonlin	split	no cuts
50	500	20	20	20	11.57	**9.92**	11.01	5864.50	**4309.05**	5905.65	—	—	—
	1000	20	20	20	41.07	38.45	**28.57**	9307.70	8265.75	**6453.65**	—	—	—
	2000	20	20	20	**68.12**	**68.11**	138.37	7411.30	**6559.15**	13016.30	—	—	—
	5000	19	19	19	695.14	622.18	**581.49**	18903.58	16145.32	**15368.53**	**2.41%**	5.19%	6.25%
100	500	**19**	14	14	**400.22**	436.02	467.32	**129745.46**	173480.42	190997.69	—	—	—
	1000	**15**	13	13	**456.84**	502.90	1300.26	**77967.64**	86555.38	221685.91	**2.68%**	14.02%	6.06%
	2000	19	**20**	15	**179.06**	337.18	223.93	**11908.73**	24955.93	16974.87	3.01%	**—**	5.46%
	5000	19	20	18	673.90	**670.20**	731.66	16101.59	**13831.22**	17026.82	—	—	—
200	500	**6**	1	0	—	—	—	—	—	—	87.92%	**46.49%**	191.83%
	1000	0	0	0	—	—	—	—	—	—	**16.31%**	24.34%	22.99%
	2000	**8**	6	5	**498.57**	787.33	2153.11	**25654.00**	35918.50	138485.00	8.33%	**3.84%**	6.50%
	5000	**17**	12	12	**1408.58**	1804.24	1539.48	**19271.44**	20442.11	22581.56	—	—	—
500	500	0	0	0	—	—	—	—	—	—	**128.91%**	**128.89%**	200.04%
	1000	0	0	0	—	—	—	—	—	—	**109.42%**	114.03%	116.02%
	2000	0	0	0	—	—	—	—	—	—	29.27%	29.34%	**28.95%**
	5000	**2**	1	0	—	—	—	—	—	—	124.45%	**113.67%**	213.15%
1000	500	0	0	0	—	—	—	—	—	—	**97.01%**	98.57%	106.20%
	1000	0	0	0	—	—	—	—	—	—	**227.93%**	**227.73%**	316.26%
	2000	0	0	0	—	—	—	—	—	—	**54.65%**	55.90%	65.86%
	5000	0	0	0	—	—	—	—	—	—	**111.06%**	214.31%	219.85%

Table 6.4: Performance of two valid inequality families in cardinality constrained portfolio optimization. The rows refer to: *no cuts* – pure branch-and-bound presented in Section 6.2, *lifted* – lifted cuts from Section 6.3.2, *split* – disjunctive cuts introduced in Section 6.3.3. Results averaged over 20 instances. *Running time* and *nodes in solution tree* columns reflect only instances solved within 1 hour time limit by all three approaches. Similarly *gap after time limit* corresponds to instances for which no optimal solution was found within the time limit for each of the methods.

n	m	number solved			running time			nodes in solution tree			gap after time limit		
		lifted	split	no cuts	lifted	split	no cuts	lifted	split	no cuts	nonlin	split	no cuts
50	500	20	20	20	**108.84**	122.91	**108.67**	**25574.20**	26636.55	**25574.20**	—	—	—
	1000	20	20	20	**240.62**	252.45	**240.38**	19634.00	**19239.50**	19634.00	—	—	—
	2000	20	20	20	**263.00**	288.33	**263.57**	7651.90	**7506.10**	7651.90	—	—	—
	5000	20	20	20	152.99	**76.31**	151.91	1274.30	**994.70**	1274.30	—	—	—
100	500	6	**7**	6	2001.51	**1795.24**	1998.73	293837.33	**111602.00**	293837.33	23.63%	**20.40%**	23.62%
	1000	0	**3**	0	—	—	—	—	—	—	29.48%	**28.22%**	29.36%
	2000	3	**5**	3	2770.52	**2440.59**	2796.48	54008.00	**42317.75**	54008.00	13.08%	**11.84%**	13.07%
	5000	18	**19**	18	1043.44	**991.63**	1047.82	7770.28	**6734.63**	7770.28	4.63%	4.60%	4.61%
200	500	0	**1**	0	—	—	—	—	—	—	85.93%	**74.56%**	85.78%
	1000	0	0	0	—	—	—	—	—	—	71.87%	**52.10%**	71.86%
	2000	0	0	0	—	—	—	—	—	—	37.56%	**17.68%**	37.56%
	5000	0	0	0	—	—	—	—	—	—	8.87%	**8.82%**	8.87%
500	500	0	**1**	0	—	—	—	—	—	—	178.71%	**79.19%**	178.56%
	1000	0	0	0	—	—	—	—	—	—	126.57%	**26.28%**	126.58%
	2000	0	0	0	—	—	—	—	—	—	67.29%	**37.03%**	67.30%
	5000	0	0	0	—	—	—	—	—	—	21.63%	**13.15%**	21.63%
1000	500	0	0	0	—	—	—	—	—	—	223.31%	**123.95%**	223.31%
	1000	0	0	0	—	—	—	—	—	—	163.56%	**65.14%**	163.57%
	2000	0	0	0	—	—	—	—	—	—	92.95%	**73.52%**	92.96%
	5000	0	0	0	—	—	—	—	—	—	219.35%	**124.00%**	219.36%

6.5 Concluding Remarks

In this chapter we reconsidered some of the methods that have been previously proposed in the literature, and shown that these approaches can be naturally applied in the case of mixed-integer nonlinear programming problems . In addition we also proposed a new simple procedure for generating disjunctive cuts. The performed numerical experiments show some promising results.

REFERENCES

Abhishek, K., Leyffer, S., and Linderoth, J. (2010) "FilMINT: an outer approximation-based solver for convex mixed-integer nonlinear programs," *INFORMS J. Comput.*, **22** (4), 555–567.

Acerbi, C. (2002) "Spectral measures of risk: A coherent representation of subjective risk aversion," *Journal of Banking and Finance*, **26** (7), 1487–1503.

Alizadeh, F. and Goldfarb, D. (2003) "Second-order cone programming," *Math. Program.*, **95** (1, Ser. B), 3–51.

Andersen, K. and Jensen, A. N. (2013) "Intersection cuts for mixed integer conic quadratic sets," in: "Integer programming and combinatorial optimization," volume 7801 of *Lecture Notes in Comput. Sci.*, 37–48, Springer, Heidelberg.

Artzner, P., Delbaen, F., Eber, J.-M., and Heath, D. (1999) "Coherent measures of risk," *Math. Finance*, **9** (3), 203–228.

Atamtürk, A. and Narayanan, V. (2010) "Conic mixed-integer rounding cuts," *Math. Program.*, **122** (1, Ser. A), 1–20.

Atamtürk, A. and Narayanan, V. (2011) "Lifting for conic mixed-integer programming," *Math. Program.*, **126** (2, Ser. A), 351–363.

Aybat, N. S. and Iyengar, G. (2012) "Unified approach for minimizing composite norms," *Math. Program.*, (Ser. A).

Aybat, N. S. and Iyengar, G. (2013) "An augmented lagrangian method for conic convex programming," .

Balas, E. (1971) "Intersection cuts—a new type of cutting planes for integer programming," *Operations Res.*, **19**, 19–39.

Belotti, P., Góez, J. C., Pólik, I., Ralphs, T. K., and Terlaky, T. (submitted) "A Conic Representation of the Convex Hull of Disjunctive Sets and Conic Cuts for Integer Second Order Cone Optimization," *Math. Program.*.

Ben-Tal, A. and Nemirovski, A. (2001a) *Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications*, volume 2 of *MPS/SIAM Series on Optimization*, SIAM, Philadelphia, PA.

Ben-Tal, A. and Nemirovski, A. (2001b) "On polyhedral approximations of the second-order cone," *Math. Oper. Res.*, **26** (2), 193–205.

Ben-Tal, A. and Teboulle, M. (2007) "An Old-New Concept of Convex Risk Measures: An Optimized Certainty Equivalent," *Mathematical Finance*, **17** (3), 449–476.

Bienstock, D. and Michalka, A. (2014) "Cutting-planes for optimization of convex functions over nonconvex sets," *SIAM J. Optim.*, **24** (2), 643–677.

Birge, J. R. and Louveaux, F. (1997) *Introduction to Stochastic Programming*, Springer, New York.

Bonami, P. (2011) "Lift-and-project cuts for mixed integer convex programs," in: "Integer programming and combinatorial optimization," volume 6655 of *Lecture Notes in Comput. Sci.*, 52–64, Springer, Heidelberg.

Bonami, P., Biegler, L. T., Conn, A. R., Cornuéjols, G., Grossmann, I. E., Laird, C. D., Lee, J., Lodi, A., Margot, F., Sawaya, N., and Wächter, A. (2008) "An algorithmic framework for convex mixed integer nonlinear programs," *Discrete Optim.*, **5** (2), 186–204.

Bonami, P. and Lejeune, M. A. (2009) "An exact solution approach for portfolio optimization problems under stochastic and integer constraints," *Oper. Res.*, **57** (3), 650–670.

Borchers, B. and Mitchell, J. E. (1994) "An improved branch and bound algorithm for mixed integer nonlinear programs," *Comput. Oper. Res.*, **21** (4), 359–367.

Bullen, P. S., Mitrinović, D. S., and Vasić, P. M. (1988) *Means and their inequalities*, volume 31 of *Mathematics and its Applications (East European Series)*, D. Reidel Publishing Co., Dordrecht, translated and revised from the Serbo-Croatian.

Burer, S. and Kılınç, M. (2014) "How to Convexify the Intersection of a Second Order Cone and a Nonconvex Quadratic," *Technical report*, Department of Management Sciences, University of Iowa.

Burer, S. and Saxena, A. (2012) "The MILP Road to MIQCP," in: J. Lee and S. Leyffer (Eds.) "Mixed Integer Nonlinear Programming," volume 154 of *The IMA Volumes in Mathematics and its Applications*, 373–405, Springer New York.

Cadoux, F. (2010) "Computing deep facet-defining disjunctive cuts for mixed-integer programming," *Math. Program.*, **122** (2, Ser. A), 197–223.

Çezik, M. T. and Iyengar, G. (2005) "Cuts for mixed 0-1 conic programming," *Math. Program.*, **104** (1, Ser. A), 179–202.

Cooke, R. M. and Nieboer, D. (2011) "Heavy-Tailed Distributions: Data, Diagnostics, and New Developments," *Resources for the Future Discussion Paper*, (11-19).

Dadush, D., Dey, S. S., and Vielma, J. P. (2011) "The split closure of a strictly convex body," *Oper. Res. Lett.*, **39** (2), 121–126.

Dana, R.-A. (2005) "A representation result for concave Schur concave functions," *Math. Finance*, **15** (4), 613–634.

De Giorgi, E. (2005) "Reward-Risk Portfolio Selection and Stochastic Dominance," *Journal of Banking and Finance*, **29** (4), 895–926.

Delbaen, F. (2002) "Coherent risk measures on general probability spaces," in: K. Sandmann and P. J. Schönbucher (Eds.) "Advances in Finance and Stochastics: Essays in Honour of Dieter Sondermann," 1–37, Springer.

Drewes, S. (2009) *Mixed integer second order cone programming*, Ph.D. thesis, Technische Universität Darmstadt, Germany.

Duffie, D. and Pan, J. (1997) "An Overview of Value-at-Risk," *Journal of Derivatives*, **4**, 7–49.

Duran, M. A. and Grossmann, I. E. (1986) "An outer-approximation algorithm for a class of mixed-integer nonlinear programs," *Math. Programming*, **36** (3), 307–339.

Espinoza, D. and Moreno, E. (2012) "Fast sample average approximation for minimizing Conditional-Value-at-Risk," *Preprint Paper*.

Fábián, C. I., Mitra, G., Roman, D., and Zverovich, V. (2011) "An enhanced model for portfolio choice with SSD criteria: a constructive approach," *Quantitative Finance*, **11**, 1525–1534.

Fishburn, P. C. (1977) "Mean-Risk Analysis with Risk Associated with Below-Target Returns," *The American Economic Review*, **67** (2), 116–126.

Fletcher, R. and Leyffer, S. (1994) "Solving mixed integer nonlinear programs by outer approximation," *Math. Programming*, **66** (3, Ser. A), 327–349.

Föllmer, H. and Schied, A. (2002) "Convex measures of risk and trading constraints," *Finance Stoch.*, **6** (4), 429–447.

Föllmer, H. and Schied, A. (2004) *Stochastic Finance: An Introduction in Discrete Time*, De Gruyter studies in mathematics, Walter de Gruyter.

Frittelli, M. and Rosazza Gianin, E. (2005) "Law invariant convex risk measures," in: "Advances in mathematical economics. Volume 7," volume 7 of *Adv. Math. Econ.*, 33–46, Springer, Tokyo.

Glineur, F. (2000) "Computational experiments with a linear approximation of second order cone optimization," *Technical Report 0001*, Service de Mathématique et de Recherche Opérationnelle, Faculté Polytechnique de Mons, Mons, Belgium.

Glineur, F. and Terlaky, T. (2004) "Conic formulation for l_p-norm optimization," *J. Optim. Theory Appl.*, **122** (2), 285–307.

Gupta, O. K. and Ravindran, A. (1985) "Branch and bound experiments in convex nonlinear integer programming," *Management Sci.*, **31** (12), 1533–1546.

Hardy, G. H., Littlewood, J. E., and Pólya, G. (1952) *Inequalities*, Cambridge, at the University Press, 2d ed.

Iaquinta, G., Lamantia, F., Massab, I., and Ortobelli, S. (2009) "Moment based approaches to value the risk of contingent claim portfolios," *Annals of Operations Research*, **165** (1), 97–121.

Jorion, P. (1997) *Value at Risk: The New Benchmark for Controlling Market Risk*, McGraw-Hill.

Kaibel, V. and Pashkovich, K. (2011) "Constructing extended formulations from reflection relations," in: "Integer programming and combinatorial optimization," volume 6655 of *Lecture Notes in Comput. Sci.*, 287–300, Springer, Heidelberg.

Kılınç, M., Linderoth, J., and Luedtke, J. (2010) "Effective separation of disjunctive cuts for convex mixed integer nonlinear programs," *Technical report*.

Klein Haneveld, W. K. and van der Vlerk, M. H. (2006) "Integrated chance constraints: reduced forms and an algorithm," *Computational Management Science*, **3**, 245–269.

Kousky, C. and Cooke, R. M. (2009) "The unholy trinity: fat tails, tail dependence, and micro-correlations," *Resources for the Future Discussion Paper*, 09–36.

Kreinovich, V., Chiangpradit, M., and Panichkitkosolkul, W. (2012) "Efficient algorithms for heavy-tail analysis under interval uncertainty," *Annals of Operations Research*, **195** (1), 73–96.

Krokhmal, P., Zabarankin, M., and Uryasev, S. (2011) "Modeling and optimization of risk," *Surveys in Operations Research and Management Science*, **16** (2), 49 – 66.

Krokhmal, P. A. (2007) "Higher moment coherent risk measures," *Quant. Finance*, **7** (4), 373–387.

Krokhmal, P. A. and Soberanis, P. (2010) "Risk optimization with p-order conic constraints: A linear programming approach," *European J. Oper. Res.*, **201** (3), 653–671.

Künzi-Bay, A. and Mayer, J. (2006) "Computational aspects of minimizing conditional value-at-risk," *Computational Management Science*, **3** (1), 3–27.

Kusuoka, S. (2001) "On law invariant coherent risk measures," in: "Advances in mathematical economics, Vol. 3," volume 3 of *Adv. Math. Econ.*, 83–95, Springer, Tokyo.

Kusuoka, S. (2012) "A remark on Malliavin calculus: uniform estimates and localization," *J. Math. Sci. Univ. Tokyo*, **19** (4), 533–558 (2013).

Lan, G., Lu, Z., and Monteiro, R. D. C. (2011) "Primal-dual first-order methods with $\mathcal{O}(1/\epsilon)$ iteration-complexity for cone programming," *Math. Program.*, **126** (1, Ser. A), 1–29.

Lan, G. and Monteiro, R. D. C. (2013) "Iteration-complexity of first-order penalty methods for convex programming," *Math. Program.*, **138** (1-2, Ser. A), 115–139.

Lemaréchal, C., Nemirovskii, A., and Nesterov, Y. (1995) "New variants of bundle methods," *Mathematical Programming*, **69**, 111–147.

Levy, H. (1998) *Stochastic Dominance*, Kluwer Academic Publishers, Boston-Dodrecht-London.

Leyffer, S. (2001) "Integrating SQP and branch-and-bound for mixed integer nonlinear programming," *Comput. Optim. Appl.*, **18** (3), 295–309.

Lim, C., Sherali, H. D., and Uryasev, S. (2010) "Portfolio optimization by minimizing conditional value-at-risk via nondifferentiable optimization," *Computational Optimization and Applications*, **46** (3), 391–415.

Malevergne, Y. and Sornette, D. (2005) "Higher-Moment Portfolio Theory," *The Journal of Portfolio Management*, **31** (4), 49–55.

Markowitz, H. M. (1952) "Portfolio Selection," *Journal of Finance*, **7** (1), 77–91.

McCord, M. and Neufville, R. d. (1986) ""Lottery Equivalents": Reduction of the Certainty Effect Problem in Utility Assessment," *Management Science*, **32** (1), pp. 56–60.

Modaresi, S., Kılınç, M. R., and Vielma, J. P. (2015) "Split cuts and extended formulations for Mixed Integer Conic Quadratic Programming," *Oper. Res. Lett.*, **43** (1), 10–15.

Morenko, Y., Vinel, A., Yu, Z., and Krokhmal, P. (2013) "On p-cone linear discrimination," *European J. Oper. Res.*, **231** (3), 784789.

Nemhauser, G. L. and Wolsey, L. A. (1988) *Integer and combinatorial optimization*, Wiley-Interscience Series in Discrete Mathematics and Optimization, John Wiley & Sons Inc., New York, a Wiley-Interscience Publication.

Nesterov, Y. (2012) "Towards non-symmetric conic optimization," *Optim. Methods Softw.*, **27** (4-5, SI), 893–917.

Nesterov, Y. E. and Nemirovski, A. (1994) *Interior Point Polynomial Algorithms in Convex Programming*, volume 13 of *Studies in Applied Mathematics*, SIAM, Philadelphia, PA.

Perold, A. F. (1984) "Large-scale portfolio optimization," *Management Sci.*, **30** (10), 1143–1160.

Pflug, G. C. (2006) "Subdifferential representations of risk measures," *Math. Program.*, **108** (2-3, Ser. B), 339–354.

Prékopa, A. (1995) *Stochastic Programming*, Kluwer Academic Publishers.

Quesada, I. and Grossmann, I. E. (1992) "An LP/NLP based branch and bound algorithm for convex MINLP optimization problems," *Computers & chemical engineering*, **16** (10), 937–947.

Randolph, J. (1952) *Calculus*, Macmillan.

Rockafellar, R. T. (1997) *Convex analysis*, Princeton Landmarks in Mathematics, Princeton University Press, Princeton, NJ, reprint of the 1970 original, Princeton Paperbacks.

Rockafellar, R. T. and Uryasev, S. (2000) "Optimization of Conditional Value-at-Risk," *Journal of Risk*, **2**, 21–41.

Rockafellar, R. T. and Uryasev, S. (2002) "Conditional Value-at-Risk for General Loss Distributions," *Journal of Banking and Finance*, **26** (7), 1443–1471.

Rockafellar, R. T. and Uryasev, S. (2013) "The fundamental risk quadrangle in risk management, optimization and statistical estimation," *Surveys in Operations Research and Management Science*, **18** (12), 33 – 53.

Rockafellar, R. T., Uryasev, S., and Zabarankin, M. (2006) "Generalized deviations in risk analysis," *Finance Stoch.*, **10** (1), 51–74.

Roman, D., Darby-Dowman, K., and Mitra, G. (2006) "Portfolio construction based on stochastic dominance and target return distributions," *Mathematical Programming*, **108**, 541–569.

Rothschild, M. and Stiglitz, J. (1970) "Increasing risk I: a definition," *Journal of Economic Theory*, **2** (3), 225–243.

Rysz, M., Vinel, A., Krokhmal, P., and Pasiliao, E. L. (2014) "A scenario decomposition algorithm for stochastic programming problems with a class of downside risk measures," *INFORMS Journal on Computing*, **37**, 231–242.

Saxena, A., Bonami, P., and Lee, J. A. A. T. (2008) "Disjunctive cuts for non-convex mixed integer quadratically constrained programs," in: "Integer programming and combinatorial optimization," volume 5035 of *Lecture Notes in Comput. Sci.*, 17–33, Springer, Berlin.

Scherer, B. and Martin, R. D. (2005) *Introduction to modern portfolio optimization with NUOPT and S-PLUS*, Springer, New York.

Stubbs, R. A. and Mehrotra, S. (1999) "A branch-and-cut method for 0-1 mixed convex programming," *Math. Program.*, **86** (3, Ser. A), 515–532.

Subramanian, D. and Huang, P. (2008) "A Novel Algorithm for Stochastic Linear Programs with Conditional-value-at-risk (CVaR) Constraints," *IBM Research Report, RC24752*.

Subramanian, D. and Huang, P. (2009) "An Efficient Decomposition Algorithm for Static, Stochastic, Linear and Mixed-Integer Linear Programs with Conditional-Value-at-Risk Constraints," *IBM Research Report, RC24752*.

Tawarmalani, M. and Sahinidis, N. V. (2005) "A polyhedral branch-and-cut approach to global optimization," *Math. Program.*, **103** (2, Ser. B), 225–249.

Terlaky, T. (1985) "On l_p programming," *European J. Oper. Res.*, **22** (1), 70–100.

Thomas, G., Hass, J., Weir, M., and Giordano, F. (2007) *Thomas' Calculus*, Thomas 11e Series, Pearson Addison Wesley.

Vielma, J. P., Ahmed, S., and Nemhauser, G. L. (2008) "A lifted linear programming branch-and-bound algorithm for mixed-integer conic quadratic programs," *INFORMS J. Comput.*, **20** (3), 438–450.

Vinel, A. and Krokhmal, P. (2014a) "On valid inequalities for mixed integer p-order cone programming," *J. Optim. Theory Appl.*, **160** (2), 439–456.

Vinel, A. and Krokhmal, P. A. (2014b) "Polyhedral approximations in p-order cone programming," *Optim. Methods Softw.*, **29** (6), 1210–1237.

Vinel, A. and Krokhmal, P. A. (2015) "Certainty equivalent measures of risk," *Annals of Operations Research*, (to appear).

von Neumann, J. and Morgenstern, O. (1944) *Theory of Games and Economic Behavior*, Princeton University Press, Princeton, NJ, 1953rd edition.

Wilson, R. (1979) "Auctions of Shares," *The Quarterly Journal of Economics*, **93** (4), pp. 675–689.

Xue, G. and Ye, Y. (2000) "An efficient algorithm for minimizing a sum of p-norms," *SIAM J. Optim.*, **10** (2), 551–579.