Fall 2009

Risk optimization with p-order conic constraints

Policarpio Antonio Soberanis
University of Iowa

RISK OPTIMIZATION WITH P-ORDER CONIC CONSTRAINTS

by

Policarpio Antonio Soberanis

<u>An Abstract</u>

Of a thesis submitted in partial fulfillment of the
requirements for the Doctor of Philosophy
degree in Mechanical and Industrial Engineering
in the Graduate College of
The University of Iowa

December 2009

Thesis Supervisor: Assistant Professor Pavlo Krokhmal

ABSTRACT

My dissertation considers solving of linear programming problems with p-order conic constraints that are related to a class of stochastic optimization models with risk objective or constraints that involve higher moments of loss distributions. The general proposed approach is based on construction of polyhedral approximations for p-order cones, thereby approximating the non-linear convex p-order conic programming problems using linear programming models. It is shown that the resulting LP problems possess a special structure that makes them amenable to efficient decomposition techniques. The developed algorithms are tested on the example of portfolio optimization problem with higher moment coherent risk measures that reduces to a p-order conic programming problem. The conducted case studies on real financial data demonstrate that the proposed computational techniques compare favorably against a number of benchmark methods, including second-order conic programming methods.

Abstract Approved: ____________________
Thesis Supervisor

____
Title and Department

____
Date

RISK OPTIMIZATION WITH P-ORDER CONIC CONSTRAINTS

by

Policarpio Antonio Soberanis

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

December 2009

Thesis Supervisor: Assistant 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

Policarpio Antonio Soberanis

has been approved by the Examining Committee
for the thesis requirement for the Doctor of Philosophy
degree in Mechanical and Industrial Engineering
at the December 2009 graduation.

Thesis Committee: ______________________________
Pavlo Krokhmal, Thesis Supervisor

____
Peter O'Grady

____
Andrew Kusiak

____
Yong Chen

____
Samuel Burer

To my wife, Ivy Soberanis, my mother Juliet Soberanis and my father Anthony Soberanis. Thank you for all your support, encouragement and understanding.

ACKNOWLEDGMENTS

I would like to thank Dr. Pavlo Krokhmal for his dedication and efforts in assiting me. I would also like to thank the members of my dissertation committee, Dr. Andrew Kusiak, Dr. Samuel Burer, Dr. Yong Chen and Dr. Peter O'Grady. Your thoughtful comments and insight was very much apprecaited in assisting me with the completion of my dissertation.

I would also like to acknowledge my high school mathematics teacher, Mr. Fajri Bajar. Thank you for believing in me and giving me the confidence to pursue my dream.

ABSTRACT

My dissertation considers solving of linear programming problems with p-order conic constraints that are related to a class of stochastic optimization models with risk objective or constraints that involve higher moments of loss distributions. The general proposed approach is based on construction of polyhedral approximations for p-order cones, thereby approximating the non-linear convex p-order conic programming problems using linear programming models. It is shown that the resulting LP problems possess a special structure that makes them amenable to efficient decomposition techniques. The developed algorithms are tested on the example of portfolio optimization problem with higher moment coherent risk measures that reduces to a p-order conic programming problem. The conducted case studies on real financial data demonstrate that the proposed computational techniques compare favorably against a number of benchmark methods, including second-order conic programming methods.

TABLE OF CONTENT

LIST OF TABLES

Table

LIST OF FIGURES

Figure

CHAPTER 1
RISK-AVERSE OPTIMIZATION BASED ON COHERENT RISK MEASURES

1.1 Introduction

Stochastic optimization is concerned with selecting an optimal decision vector, $\mathbf{x} \in \mathbb{R}^n$ under uncertainties, when the outcome X of the decision $\mathbf{x}$ also depends on some random event $\omega \in \Omega$: $X = X(\mathbf{x}, \omega)$. Assuming that smaller values of X are preferred, or, in other words, X represents a cost or loss due to the decision $\mathbf{x}$, a traditional formulation of a stochastic programming problem involves minimization of the expected cost of $\mathbf{x}$:

$$\min_{\mathbf{x}\in\mathcal{S}} \mathsf{E}\big[X(\mathbf{x}, \omega)\big], \tag{1.1}$$

where, for simplicity it can be assumed that the feasible set $\mathcal{S}$ in (1.1) is deterministic, i.e., does not depend on the random element ω. In many situations, however, a decision that minimizes the expected loss or is otherwise based on the average outcome may not be satisfactory, and a more conservative, or risk-averse decision strategy may be preferred. A variety of approaches and techniques for implementing risk-averse preferences in stochastic programming, and in decision-making problems under uncertainty in general, have been developed in the literature. Below we present a brief overview of some of these techniques, not pretending to be exhausting.

In particular, we focus our attention on the so-called *risk measures*, which are most closely related to the stochastic optimization models with p-order conic constraints that constitute the primary subject of this work.

Analysis and modeling of decision-maker's preferences, including risk preferences, have been actively studied in financial literature, specifically in the context of portfolio optimization. From the historical and methodological prospective, risk measures relate to the *risk-reward* optimization paradigm that has developed from the seminal works of Markowitz (1952, 1959), and complements another classical approach to decision-making under uncertainty, the *utility theory* of von Neumann and Morgenstern (1944).

In this section, we will look at the major developments in risk measurements. Starting with the work of Markowitz, we will discuss the various aspects of how the field has changed as we progress to higher moment coherent risk measures. We will look at examples from Downside Risk measures, Value-at-Risk and Conditional Value-at-Risk, and finally Coherent Risk Measures. We will also discuss the strength and weaknesses of these different risk measures with regard to portfolio optimization.

1.2 Risk Measures in Decision Making Under Uncertainty and Stochastic Optimization

1.2.1 Markowitz Mean-Variance (MV) Model

Initially, maximization of portfolio instruments was based on maximization of expected return. As the field progressed, the use of risk (regret) and reward (satisfaction) based on an investor's preferences (risk-averse nature) has been the driving force behind much of the modern day research. The measure of reward that is most commonly associated with an investment portfolio is widely accepted to be the expected return. The measure of risk, however, is something that is still being debated and is an area of active research.

The modern theory of risk management was given its initial foundation back in 1952 by Markowitz (1952). In this landmark paper, Markowitz suggested that investment's or portfolio's risk can be identified with the volatility, or, more technically, variance of the cumulative return of a portfolio's assets. Defined in such a way, risk can be minimized while ensuring that the portfolio still guarantees some level of performance, measured by the expected return of the portfolio's assets.

On a more fundamental level, the Markowitz model proposed an approach to decision making under uncertainty markedly different from the prevailing framework of utility theory, which advocated that a rational decision under uncertainty is the one that maximizes the expected utility. Namely, the Markowitz paradigm postulates that *any decision under uncertainties can be viewed as a tradeoff between the risk and reward.*

The standard Markowitz mean-variance portfolio optimization model in application to portfolio optimization can be formulated as follows. Consider a set of instruments, $\{1, \ldots, n\}$ that are characterized by random rates of return $r_i = r_i(\omega)$, $i = 1, \ldots, n$. Given an investment vector $\mathbf{x} = (x_1, \ldots, x_n)$, where $x_i \in [0, 1]$ represents the proportion of one's wealth invested in asset $i = 1, \ldots, n$, the cumulative portfolio's return X has the form

$$X(\mathbf{x}, \mathbf{r}(\omega)) = \sum_{i=1}^{n} x_i r_i(\omega) = \mathbf{x}^\top \mathbf{r}(\omega)$$

According to the original Markowitz model (Markowitz, 1952), an optimal portfolio allocation decision $\mathbf{x}$ minimizes the investment risk, embodied by the variance of X, while ensuring that the expected portfolio's return exceeds some prescribed level r_0:

$$
\begin{aligned}
\min \quad & \sum_{i=1}^{n}\sum_{j=1}^{n} \sigma_{ij} x_i x_j \\
\text{s. t.} \quad & \sum_{i=1}^{n} x_i E(r_i) \geq r_0 \\
& \sum_{i=1}^{n} x_i = 1 \\
& x_i \geq 0, \quad i = 1, \ldots, n
\end{aligned}
\tag{1.2}
$$

The objective function of (1.2) is equal to the variance $\sigma^2(X)$ of portfolio return:

$$
\sigma^2(X) = E\big[(X - E[X])^2\big] = \sum_{i=1}^{n}\sum_{j=1}^{n} \sigma_{ij} x_i x_j,
$$

where σ_{ij} is the covariance of the returns r_i and r_j of assets i and j:

$$
\sigma_{ij} = \mathrm{Cov}(r_i, r_j) = E\big[(r_i - E[r_i])(r_j - E[r_j])\big]
$$

The first constraint of (1.2) stipulates that the expected return $E[X] = E[\mathbf{r}^\top \mathbf{x}]$ of the portfolio does not fall below certain level r_0, while the second and third constraints ensure that the entire available wealth is invested in the portfolio.

Observe that (1.2) is a quadratic programming problem that is moreover convex, due to the fact that the covariance matrix $\{\sigma_{ij}\}_{i,j=1,\ldots,n}$ is positive semidefinite.

In general, the Markowitz MV model can be stated using the notations adopted above as the problem of minimization of risk expressed by the variance of decision's cost $\sigma^2(X(\mathbf{x}, \omega))$ while requiring that the average cost of the decision does not exceed a predefined threshold c_0:

$$
\min_{\mathbf{x}} \{\sigma^2(X(\mathbf{x}, \omega)) \mid E(X(\mathbf{x}, \omega)) \leq c_0, \;\; \mathbf{x} \in \mathcal{C}\} \tag{1.3}
$$

where $\mathcal{C} \subset \mathbb{R}^n$ is the set of feasible decisions $\mathbf{x}$.

Although the original Markowitz's approach is still widely used today, there have been numerous criticisms of the model – specifically its inability to distinguish the risks of a "good", or "positive" deviation from a bad or negative deviation from the average. In other words, the variance $\sigma^2(X)$ as a proxy for risk penalizes equally the "desirable" instances when $X > E[X]$ and "undesirable" cases when $X < E[X]$ (assuming here that X represents, for instance, portfolio return).

This inability to reward a decision for a positive change in the return led to the development of the so-called *downside risk models* that can replace the symmetric variance functional $\sigma^2(X)$ in problem (1.3). Next we outline the most notable developments in this area, including the semivariance risk models, lower partial moments, Value-at-Risk, etc.

1.2.2 Semivariance Risk

The shortcomings of variance $\sigma^2(X)$ as a risk measure has been recognized as far back as by Markowitz himself, who proposed to use *semi-variance* $\sigma_+^2(X)$ for a more accurate estimation of risk exposure (Markowitz, 1959):

$$\sigma_+^2(X) = E\big[(X - E[X])_+^2\big] \tag{1.4}$$

where $(X)_\pm$ denotes the positive (negative) part of X:

$$(X)_\pm = \max\{0, \pm X\}$$

Note that here and in what follows, we present risk models and concepts in the form that implicitly assumes that X represents a cost or loss, and whose large positive values must be avoided. An alternative view, often found in the literature, regards X

as a *wealth* variable, whereby the relevant definition of semivariance as a risk function would take the form that only takes into account the situations when the value of $X = X(\omega)$ drops below its expected level $E[X]$:

$$\sigma_-^2(X) = E\big[(X - E[X])_-^2\big]$$

Application of semivariance risk models to decision making under uncertainty has recently been studied by Ogryczak and Ruszczyński (1999, 2001, 2002).

A drawback of the semivariance approach is that risk is quantified through the shortfalls from the expected level. In many situations, a decision maker may be interested to regard risk as shortfall from a certain benchmark level. This led to the development of a number of approaches to the measurement and optimization of risk which discussed next.

1.2.3 Downside Risk Measures

The transition from the classical Markowitz MV model (1.2)–(1.3) to semivariance risk model was a natural extension of the fact that the MV model penalized for both positive and negative variation in the portfolio's return. The next family of risk models have been motivated by practical considerations, when investors would set up a certain benchmark level of wealth, a, and associate the investment's risk with underachievement, or shortfall of this goal. Then, for instance, it is convenient to consider the risk of the decision X as the average shortfall with respect to the benchmark level a:

$$\mathrm{ER}(X) = E(X - a)^+ \tag{1.5}$$

The risk function defined in (1.5) is known in the literature as *Expected Regret* (Dembo and Rosen, 1999). A generalization of (1.5) is the so-called *Lower Partial Moment* function (see Bawa, 1975; Fishburn, 1977):

$$\mathrm{LPM}_p(X, a) = E((X - a)^+)^p, \quad p \geq 1, \;\; a \in \mathbb{R} \tag{1.6}$$

where we again note that X represent a cost or a loss, and thus its "positive" shortfall $(X - a)^+$ is of interest.

A requirement that the risk when measured by the lower partial moment function $\mathrm{LPM}_p(X, a)$ should not exceed some level $b > 0$ can be expressed as a risk constraint of the form

$$E[(X - a)_+^p] \leq b$$

In the special case of $p = 1$ the above constraint reduces to the form

$$E[(X - a)_+] \leq b \tag{1.7}$$

which is known as *Integrated Chance Constraints* (ICC) (Testuri and Uryasev, 2003; van der Vlerk, 2003). The Integrated Chance Constraints have been proposed as a computationally efficient alternative to chance, or probabilistic constraints (Prékopa, 1995; Birge and Louveaux, 1997)

$$P\{X(\mathbf{x}, \omega) \geq a\} \leq \alpha, \quad \alpha \in (0, 1) \tag{1.8}$$

Probabilistic constraints are extremely popular in a wide spectrum of disciplines, from finance to reliability theory, due to their intuitive interpretation: (1.8) defines the set of such $\mathbf{x} \in \mathbb{R}^n$ for which the probability of the decision cost $X(\mathbf{x}, \omega)$ exceeding some prescribed level a is no more than α.

Despite this transparent interpretation, chance constraints (1.8) have rather poor properties from the mathematical programming viewpoint as they are generally non-convex. In the field of financial risk management, the chance constraints are directly related to the well-known *Value-at-Risk* measure that is discussed next.

1.2.4 Value-at-Risk Measure

Perhaps the most famous risk measure in the area of financial risk management is the *Value-at-Risk (VaR)* measure (see, for instance, Morgan, 1994; Jorion, 1997; Duffie and Pan, 1997, and references therein). Methodologically, if X represents the potential financial loss, then, for instance, its 0.95% VaR ($\mathrm{VaR}_{0.95}(X)$) defines the risk of X as as the amount that can be lost with probability no more than 5%, over a given time horizon (e.g., 1 week). Mathematically, Value-at-Risk with confidence level α is defined as the α-quantile of the probability distribution of X (see, e.g., Rockafellar and Uryasev, 2002b):

$$\mathrm{VaR}_\alpha(X) = \inf\{\zeta \mid P[X \leq \zeta] \geq \alpha\} \tag{1.9}$$

From the above definition it easy to see that probabilistic constraint (1.8) can be expressed as a constraint on the Value-at-Risk of $-X(\mathbf{x}, \omega)$:

$$\mathrm{VaR}_\alpha(-X(\mathbf{x}, \omega)) \leq -a$$

Currently VaR is widely adopted by the banking and financial industry and it still remains a risk measure that is used to this day by a number of firms that do risk modeling including Cargill's Risk Analysis Department.

The VaR measure, despite the easy-to-interpret definition, turned out to have a number of modeling and implementation issues. This is mainly due to VaR's non-convexity as a function of the decision variables. This is a serious limitation, not only in the context of mathematical programming, where convexity guarantees well-behaved models, but also from the risk-management prospective, since it violates the fundamental principle of *risk reduction by diversification*. It is a well-known fact that diversification enables one to reduce the risk of investment loss. However, when risk is measured using VaR, it is possible that diversification may *increase* VaR, instead of *reducing* it!

These limitations led to the development of a better-behaved alternative to VaR, which is now known as the *Conditional Value-at-Risk (CVaR)* measure.

1.2.5 Conditional Value-at-Risk (CVaR)

The Conditional Value-at-Risk measure has been designed as a risk measure that would remedy the shortcomings of VaR while preserving its intuitive practical meaning. For random cost or loss X that has a continuous distribution, Rockafellar and Uryasev (2000) have defined CVaR with confidence level α as the conditional expectation of losses X exceeding the VaR_α level:

$$\mathrm{CVaR}_\alpha(X) = E[X \mid X \geq \mathrm{VaR}_\alpha(X)] \tag{1.10}$$

In accordance with this definition, for example, the 95% Conditional Value-at-Risk is defined as the *average* of 5% of worst-case losses, or the average loss that can occur in 5% of worst-case scenarios.

In contrast to VaR, Conditional Value-at-Risk possesses superior mathematical properties, including, but not limited to, convexity and continuity with respect to the confidence level α, etc.

It is important to note that in the case of discretely distributed X, definition (1.10) does not guarantee convexity of CVaR with respect to X. Thus, to preserve the nice properties of CVaR in the case of general loss distributions, a more intricate definition of CVaR has been introduced in Rockafellar and Uryasev (2002a), which presents $\mathrm{CVaR}_\alpha(X)$ as a convex combination of $\mathrm{VaR}_\alpha(X)$ and conditional expectation of losses strictly exceeding the $\mathrm{VaR}_\alpha(X)$ level:

$$\mathrm{CVaR}_\alpha(X) = \lambda_\alpha(X)\mathrm{VaR}_\alpha(X) + (1 - \lambda_\alpha(X))E[X \mid X > \mathrm{VaR}_\alpha(X)] \tag{1.11}$$

Despite such a seemingly complex definition, computation and minimization of CVaR can be accomplished very efficiently using the following formula due to Rockafellar and Uryasev (2000, 2002a):

$$\mathrm{CVaR}_\alpha(X) = \min_{\eta \in \mathbb{R}} \; \eta + (1 - \alpha)^{-1} E(X - \eta)^+, 0 < \alpha < 1 \tag{1.12}$$

Moreover, it turns out that the set of optimal solutions η^* that deliver minimum to (1.12) contains $\mathrm{VaR}_\alpha(X)$ as its left-hand point! In other words, one can compute both $\mathrm{VaR}_\alpha(X)$ and $\mathrm{CVaR}_\alpha(X)$ in one shot using the representation (1.12).

The last representation is a special case of a more general representation of an entire class of risk measures that have nice properties similar to CVaR, the so-called *Coherent Risk Measures*, that we discuss below. It is also worth mentioning that there are other risk measures that are similar to CVaR in construction or may coincide with it in certain cases (*Conditional Drawdown-at-Risk* (Chekhlov et al., 2005; Krokhmal et al., 2002b), *Expected Shortfall* and *Tail VaR* Acerbi and Tasche (2002), etc).

1.3 Coherent Measures of Risk

Historically, development of risk models used in the Markowitz risk-reward framework has been application-driven, or "ad-hoc" to a large degree, meaning that new risk models have been designed in an attempt to represent particular risk preferences or attitudes in decision making under uncertainty. As a result, some risk models, while possessing certain attractive properties have been lacking some seemingly fundamental features, which undermined their applicability in many problems. The most famous example of this is the Value-at-Risk measure, which has been heavily criticized by both academician and practitioners for its lack of convexity and other shortcomings.

Thus, an entirely different "axiomatic" approach to the construction of risk models has been proposed by Artzner et al. (1999), who undertook the task of determining the set of requirements, or axioms that a "good" risk function must satisfy. From a number of such potential requirements they identified four, and called the functionals that satisfied these four requirements *coherent measures of risk*. Since the pioneering work of Artzner et al. (1999), the axiomatic approach has become the dominant method in risk analysis, and a number of new classes of risk measures, tailored to specific preferences and applications, have been developed in the literature. Examples of such risk measures include *deviation measures* (Rockafellar et al., 2006), *spectral risk measures* (Acerbi, 2002), and others.

In the the formal axiomatic framework of risk analysis, a risk measure $\mathscr{R}(X)$ of a random outcome X is defined as a functional $\mathscr{R} : \mathcal{X} \mapsto \mathbb{R}$, where $\mathcal{X}$ is some functional space. For a discussion of risk measures on general spaces see, for example, Ruszczyński and Shapiro (2006); in this work we select $\mathcal{X}$ to be the well-known $\mathscr{L}_p$

space defined on probability space $(\Omega, \mathscr{F}, P)$

$$\mathcal{X} = \mathscr{L}_p(\Omega, \mathscr{F}, P), \quad p \geq 1$$

where Ω is the set of random events, $\mathscr{F}$ is the corresponding sigma algebra, and P is the probability measure. In practice this would mean that our analysis applies to all random outcomes $X = X(\mathbf{x}, \omega)$, $\omega \in \Omega$, of the decision $\mathbf{x} \in \mathbb{R}^n$ that have finite moments of order p:

$$E|X(\mathbf{x}, \omega)|^p < \infty.$$

Then, a *coherent risk measure* is defined as a functional $\mathscr{R} : \mathcal{X} \mapsto \overline{\mathbb{R}}$ that satisfies the following four axioms (Artzner et al. (1999); Delbaen (2002)):

(A1) *monotonicity*: $X \leq 0 \Rightarrow \mathscr{R}(X) \leq 0, \quad \forall X \in \mathcal{X}$,

(A2) *convexity*: $\mathscr{R}(\lambda X + (1-\lambda)Y) \leq \lambda \mathscr{R}(X) + (1-\lambda)\mathscr{R}(Y), \forall X, Y \in \mathcal{X}, 0 \leq \lambda \leq 1$

(A3) *positive homogeneity*: $\mathscr{R}(\lambda X) = \lambda \mathscr{R}(X), \quad \forall X \in \mathcal{X}, \lambda > 0$,

(A4) *translation invariance*: $\mathscr{R}(X + a) = \mathscr{R}(X) + a, \quad \forall X \in \mathcal{X}, a \in \mathbb{R}$

Let us now briefly recount the meaning of each axiom. The monotonicity axiom (A1) generally means that larger realizations of X bear more risk (see also property (P2) below).

The convexity axiom (A2) is a key property from both the methodological and computational perspectives. In the mathematical programming context, it means that $\mathscr{R}(X(\mathbf{x}, \omega))$ is a convex function of the decision vector $\mathbf{x}$ whenever the cost $X(\mathbf{x}, \omega)$ is convex in $\mathbf{x}$. This, in turn, entails that the minimization of risk over a convex set of decisions $\mathbf{x}$ constitutes a convex programming problem, which is amenable to efficient solution procedures. Moreover, convexity of coherent risk measures has important implications from the methodological risk management viewpoint: given the positive homogeneity (A3), convexity implies sub-additivity

(A2′) $\mathscr{R}(X+Y) \leq \mathscr{R}(X) + \mathscr{R}(Y) \quad \text{for all} \quad X, Y \in \mathcal{X}$ *(sub-additivity)*

which is a mathematical expression of the fundamental risk management principle of *risk reduction via diversification.*

The positive homogeneity axiom (A3) ensures that if all realizations of X increase or decrease uniformly by a positive factor, the corresponding risk $\mathscr{R}(X)$ scales accordingly. In some applications, such a behavior of $\mathscr{R}$ may not be desirable, and a number of authors dropped the positive homogeneity from the list of required properties of risk measures (see, e.g., Schied and Follmer, 2002; Ruszczyński and Shapiro, 2006).

Finally, the translation invariance axiom (A4) implies that addition of a constant term to the cost or loss profile X changes its risk by the same amount.

The following useful properties of coherent risk measures can be derived from the fundamental axioms (A1)–(A4), see Delbaen (2002) and Ruszczyński and Shapiro (2006):

(P1) $\mathscr{R}(0) = 0$, and, in general, $\mathscr{R}(a) = a, \forall a \in \mathbb{R}$

(P2) $X \leq Y \Rightarrow \mathscr{R}(X) \leq \mathscr{R}(Y)$, and, in particular, $X \leq a \Rightarrow \mathscr{R}(X) \leq a, \forall a \in \mathbb{R}$

(P3) $\mathscr{R}(X - \mathscr{R}(X)) = 0$

(P4) if $\mathcal{X}$ is a Banach lattice then $\mathscr{R}(X)$ is continuous in the interior of its effective domain.

Note that here and throughout the text the inequalities of the form

$$X \geq a, \quad X \leq Y$$

are assumed to hold *almost surely.* In other words, the above expressions are equivalent to

$$P\{X \geq a\} = 1, \quad P\{X \leq Y\} = 1$$

respectively.

Next we discuss several examples of coherent and non-coherent risk measures.

Example 1. Expected value of random variable $\mathscr{R}(X) = E[X]$ is a coherent measure of risk. Indeed, the fact that $\mathscr{R}(X) = E[X]$ satisfies properties (A1)–(A4) follows directly from the elementary statistical properties of the expectation operator.

Example 2. The so-called *Maximum Loss* measure:

$$\text{MaxLoss}(X) = \sup X$$

which associates the risk of X with the largest value that X can assume, is a coherent measure of risk

Example 3. Conditional Value-at-Risk (CVaR) measure as defined by (1.11) is a coherent measure of risk. The definition (1.10)

Example 4. Variance $\mathscr{R}(X) = \sigma^2(X)$ is not coherent as it is evident that despite being convex (A2), it fails axioms (A1), (A3), and (A4).

Example 5. Value-at-Risk measure $\mathscr{R}(X) = \text{VaR}_\alpha(X)$ is not a coherent measure of risk. Although it satisfies (A1), (A3), and (A4), $\text{VaR}_\alpha(X)$ fails the all-important convexity property (A2).

What makes the class of coherent risk measures particularly appealing for modeling of risk-averse preferences in stochastic optimization problems is the fact that the expectation operator $E[X]$ satisfies (A1)–(A4) and is therefore a coherent risk measure itself. Furthermore, properties (A1)–(A4) play a pivotal role in determining the characteristics of stochastic optimization problems of type (1.1) and the corresponding solution algorithms (see, e.g., Birge and Louveaux, 1997; Prékopa, 1995). This opens possibilities for implementing risk averse preferences in many stochastic optimization models simply by replacing $E[\cdot]$ with an appropriately selected coherent risk measure $\mathscr{R}(\cdot)$. Note, however, that success of such an approach would ultimately depend on the particular form of $\mathscr{R}$, and its amenability to efficient incorporation in mathematical programming models. Next we discuss two classes of coherent risk measures that involve higher-order moments of loss distributions, and whose implementation in mathematical programming problems constitutes the objective of the present endeavor.

1.3.1 Higher Moment Coherent Risk Measures

Constructive representations for coherent measures of risk that can be efficiently applied in stochastic optimization context have been proposed in (Krokhmal, 2007):

$$\mathscr{R}(X) = \inf_{\eta} \; \eta + \phi(X - \eta) \tag{1.13}$$

(similar constructs have been investigated by Ben-Tal and Teboulle, 2007, see also Ben-Tal and Teboulle, 1986). Formally, the following result holds:

Theorem 1.3.1 (Krokhmal (2007))**.** *Let function $\phi : \mathcal{X} \mapsto \mathbb{R}$ satisfy axioms (A1)-(A3) and be a lower semicontinuous (lsc) function such that $\phi(\eta) > \eta$, $\forall \eta \in \mathbb{R}$, $\eta \neq 0$. Then the optimal value of the stochastic programming problem* (1.13) *is a* *proper coherent risk measure, and the infimum is attained for all X, so* $\inf_{\eta}$ *in (1) may be replaced by* $\min_{\eta\in\mathbb{R}}$.

A family of coherent risk measures that quantify risk in terms of tail moments of loss distributions was then introduced as a special case of (1.13). Namely, let $\mathcal{X} = \mathscr{L}_p(\Omega, \mathscr{F}, P)$, and for some $0 < \alpha < 1$ choose the functional ϕ in (1.13) as

$$\phi(X) = (1-\alpha)^{-1}\|(X)^+\|_p,$$

where $\|X\|_p = (E|X|^p)^{1/p}$. It is clear that ϕ satisfies the conditions of Theorem 1.3.1. Then, a class of *higher moment coherent risk measures* (HMCR) that quantify risk in terms of tail moments of loss distributions is introduced as

$$\mathrm{HMCR}_{p,\alpha}(X) = \min_{\eta\in\mathbb{R}} \eta + (1-\alpha)^{-1}\|(X-\eta)^+\|_p, \quad p \geq 1, \quad \alpha \in (0,1) \tag{1.14}$$

From the definition of $\|X\|_p$ it is clear that $\|X\|_p < \|X\|_q$ for $1 \leq p < q$. Thus, the class of HMCR measures are monotonic with respect to the order of p:

$$\mathrm{HMCR}_{p,\alpha}(X) \leq \mathrm{HMCR}_{q,\alpha}(X) \text{ for } p < q \text{ and } X \in \mathscr{L}_q$$

Risk measures similar to (1.14) on more general spaces have been been discussed independently by Cheridito and Li (2007). The HMCR family contains, as a special case of $p = 1$, the aforementioned Conditional Value-at-Risk measure (Rockafellar and Uryasev, 2000, 2002b).

The importance of HMCR measures is in measuring the "mass" in the right-hand tail of loss distribution via the tail moments $\|(X-\eta)^+\|_p$. It is widely acknowledged that the "risk" is associated with higher moments of the loss distributions (e.g., "fat tails" are attributable to high kurtosis, etc). However, the HMCR measures are not the only coherent risk measures that quantify risk in terms of higher moments of loss distributions.

1.3.2 Semi-Moment Coherent Risk Measures (SMCR)

Another family of coherent measures of risk that employ higher moments of loss distributions has been considered by Fischer (2003) and Rockafellar et al. (2006):

$$\mathsf{SMCR}_{p,\beta}(X) = EX + \beta \big\|(X - EX)^+\big\|_p, \quad p \geq 1, \quad \beta \geq 0. \tag{1.15}$$

We call (1.15) the *semi-moment coherent risk measures* (SMCR) as, similarly to semivariance, they are based on central semi-moments of loss distributions.

In contrast to SMCR measures (1.15), the HMCR measures (1.14) are *tail* risk measures. By this we mean that in (1.15) the "tail cutoff" point, about which the partial moments are computed, is always fixed at $E[X]$, whereas in (1.14) the location of tail cutoff is determined by

$$\eta_{p,\alpha}(X) = \textit{left end point of } \arg\min_{\eta} \big\{\eta + (1-\alpha)^{-1}\big\|(X-\eta)^+\big\|_p\big\}$$

and is adjustable by means of the parameter α. Further, it can be verified that $0 < \alpha_1 < \alpha_2 < 1$ implies

$$\eta_{p,\alpha_1}(X) \leq \eta_{p,\alpha_2}(X)$$

and, in addition, one has

$$\eta_{p,\alpha}(X) \to \sup X \quad \text{as} \quad \alpha \to 1$$

(see Krokhmal, 2007).

As it is shown below, the HMCR measures (1.14) and the semi-moment based coherent risk measures (1.15) can be treated very similarly from the mathematical programming perspective.

1.3.3 Connection to Utility Theory

In general, coherent risk measures are inconsistent with the utility of von Neumann and Morgenstern (1944), in the sense that the minimum-risk solution as obtained by minimizing risk using a coherent risk measure may not be attractive to a rational utility maximizer (see an example in Giorgi, 2005).

Recall that the von Neumann-Morgenstern theory of utility (von Neumann and Morgenstern, 1944) states that given a person's preference relation, "$\succeq$" that satisfies the axioms of completeness, transitivity, continuity and independence, there exists a function $u : \mathbb{R} \mapsto \mathbb{R}$, such that an outcome X is preferred to outcome Y $(X \succeq Y)$ if and only if $E[u(X)] \geq E[u(Y)]$. If additionally, the function u is non-decreasing and concave, the corresponding preference is said to be risk averse.

It is possible to introduce risk measures that will be consistent with the utility theory by means of the second-order stochastic dominance (SSD) ordering. Namely, a random outcome X dominates outcome Y by the second-order stochastic dominance if

$$\int_{-\infty}^{z} P[X \leq t]dt \leq \int_{-\infty}^{z} P[Y \leq t]dt, \quad \forall z \in \mathbb{R}.$$

Using the concept of second-order stochastic dominance (SSD), Rothschild and Stiglitz (1970) showed that if X dominates Y by the second-order stochastic dominance $(X \succeq_{SSD} Y)$ then it follows that the relation $E[u(X)] \geq E[u(Y)]$ holds true for all non-decreasing concave functions u where the inequality is strict for at least one such u.

By replacing the monotonicity axiom (A1) in the definition of coherent risk measures (see §1.3) with the requirement of second-order stochastic dominance *isotonicity* (Giorgi, 2005; Pflug, 2000):

$$(-X) \succeq_{SSD} (-Y) \Rightarrow \mathscr{R}(X) \leq \mathscr{R}(Y)$$

we can obtain risk measures consistent with the SSD ordering and the utility theory of von Neumann and Morgenstern (1944). More precisely, we consider risk measures $\mathscr{R} : \mathcal{X} \mapsto \overline{\mathbb{R}}$ that satisfy the following set of axioms:

(S1) *SSD isotonicity*: $(-X) \succeq_{SSD} (-Y) \Rightarrow \mathscr{R}(X) \leq \mathscr{R}(Y), \forall X, Y \in \mathcal{X}$,

(A2) *convexity*: $\mathscr{R}(\lambda X + (1-\lambda)Y) \leq \lambda\mathscr{R}(X) + (1-\lambda)\mathscr{R}(Y), \forall X, Y \in \mathcal{X},\ 0 \leq \lambda \leq 1$,

(A3) *positive homogeneity*: $\mathscr{R}(\lambda X) = \lambda\mathscr{R}(X), \forall X \in \mathcal{X}, \lambda > 0$,

(A4) *translation invariance*: $\mathscr{R}(X + a) = \mathscr{R}(X) + a, \forall X \in \mathcal{X}, a \in \mathbb{R}$

One should note that (S1) requires X and Y to be integrable. For more on the topological properties of sets defined by stochastic dominance relations see Dentcheva and Ruszczyński (2004). With the properties of stochastic dominance defined, we can once again find an analog of formula (1.13) that would allow for the construction of risk measures that adhere to the rules of the vNM utility theory.

Theorem 1.3.2 (Krokhmal (2007))**.** *Let function $\phi : (X) \mapsto \mathbb{R}$ satisfy axioms (S1), (A2), (A3) and be a lsc function such that $\phi(\eta) > \eta$ for all real $\eta \neq 0$. Then the optimal value of the stochastic programming problem*

$$\rho(X) = \min_{\eta \in \mathbb{R}} \eta + \phi(X - \eta) \tag{1.16}$$

exists and is a proper function that satisfies (S1), (A2)–(A4).

Obviously, by solving the risk-minimization problem

$$\min_{\mathbf{x}\in\mathcal{C}} \rho\big(X(\mathbf{x},\omega)\big)$$

where $\rho(X)$ is a risk measure that is both coherent and SSD-compatible in the sense of (S1), one obtains a solution that acceptable to any risk-averse rational utility maximizer, and also bears the lowest risk in terms of coherence preference metrics. Notably, one class of coherent risk measures is represented by the HMCR measures (1.14).

Example 1. The higher-moment coherent risk measures

$$\mathrm{HMCR}_{p,\alpha}(X) = \min_{\eta\in\mathbb{R}} \eta + (1-\alpha)^{-1}\|(X-\eta)^+\|_p, \quad p \geq 1, \quad \alpha \in (0,1)$$

satisfy both the coherence properties (A1)–(A4) and the SSD isotonicity property (S1), and are therefore compatible with SSD ordering and utility theory, are the HMCR measures (1.14). For the particular case of Conditional Value-at-Risk ($p = 1$), this fact has been observed by Pflug (2000).

1.4 Deviation Measures

The introduction of the axiomatic approach to the development of risk measures by Artzner et al. (1999) has fueled the development of many new types of risk measures that can be tailored to specific applications and preferences (see Acerbi (2002), Rockafellar et al. (2006) and Ruszczyński and Shapiro (2006)). In this section we will look at the development of *deviation measures* which have been introduced by Rockafellar et al. (2006). The definition of a deviation measure is a mapping $\mathscr{D} : \mathcal{X} \mapsto [0, +\infty]$ that satisfies the following axioms:

(D1) $\mathscr{D} > 0$ for any non-constant $X \in \mathcal{X}$, whereas $\mathscr{D}(X) = 0$ for constant X,

(D2) $\mathscr{D}(\lambda X + (1-\lambda)Y) \leq \lambda \mathscr{D} X + (1-\lambda)\mathscr{D}(Y) \quad \forall X, Y \in \mathcal{X}, \quad \forall \lambda \in (0,1)$

(D3) $\mathscr{D}(\lambda X) = \lambda \mathscr{D}(X) \quad \forall X \in \mathcal{X},\ \lambda > 0$

(D4) $\mathscr{D}(X + a) = \mathscr{D}(X),\ \forall X \in \mathcal{X},\ a \in \mathbb{R}$

If, moreover, $\mathscr{D}(X)$ also satisfies

(D5) $\mathscr{D}(X) \leq \text{ess.sup}\ X - E[X],\ \forall X \in \mathcal{X}$

then $\mathscr{D}(X)$ is characterized by the one-to-one correspondence:

$$\mathscr{D}(X) = \mathscr{R}(X - E[X]) \tag{1.17}$$

with *expectation-bounded* coherent risk measures (i.e., risk measures that satisfy (A1)–(A4) and $\mathscr{R}(X) > E[X]$, for all nonconstant $X \in \mathcal{X}$ and $\mathscr{R}(X) = E[X]$ for all constant $X \in \mathcal{X}$ (Rockafellar et al., 2006)).

With this result we can formulate an analog of formula (1.13) for deviation measures.

Theorem 1.4.1 (Krokhmal (2007))**.** *Let function $\phi : \mathcal{X} \mapsto \mathbb{R}$ satisfy axioms (A1)-(A3), and be a lsc function such that $\phi(x) > E[X]\ \forall X \neq 0$. Then the optimal value of the stochastic programming problem*

$$\mathscr{D}(X) = -E[X] + \inf_{\eta}\{\eta + \phi(X - \eta)\} \tag{1.18}$$

is a deviation measure and the infimum is attained for all X, so that $\inf\limits_{\eta}$ in (1.18) may be replaced by $\min\limits_{\eta \in \mathbb{R}}$.

The above results can be applied to deviation measures in the same way that we applied the results of the previous theorems to coherent risk measures. Namely, it allows one to consider the *higher moment deviation measures*

$$\mathsf{HMD}_{p,\,\alpha}(X) = \mathsf{HMCR}_{p,\alpha}(X - EX) \tag{1.19}$$

and *semi-moment deviation measures*

$$\mathsf{SMD}_{p,\,\beta}(X) = \mathsf{SMCR}_{p,\beta}(X - EX) = \beta\|(X - EX)^{+}\|_p \tag{1.20}$$

with the latter obviously reducing to the well-known Lower Partial Moment measures of risk.

1.5 Implementation of Coherent and Deviation Measures in Mathematical Programming Problems with p-Order Conic Constraints

The traditional method of modeling of uncertainty in stochastic programming (see, e.g., Birge and Louveaux, 1997; Prékopa, 1995) is by introducing a finite set of scenarios $\{\omega_1, \ldots, \omega_J\} \subseteq \Omega$, whereby each decision $\mathbf{x}$ results in a range of outcomes $X(\mathbf{x}, \omega_1), \ldots, X(\mathbf{x}, \omega_J)$ that have respective probabilities $\varpi_1, \ldots, \varpi_J$, where $\varpi_j = \mathsf{P}\{\omega_j\} \in (0, 1)$ and $\varpi_1 + \ldots + \varpi_J = 1$. Within this framework, the HMCR measures (1.14), SMCR measures (1.15), as well as their deviation counterparts (1.19), (1.20) can be implemented in the objective and/or constraints of a stochastic optimization problem using conic constraints of order $p \geq 1$.

Implementation of the HMCR measures (1.14) in stochastic programming models is facilitated by means of the following general result regarding risk measures that admit representation in the form (1.13).

Theorem 1.5.1 (Krokhmal (2007)). *Consider the following stochastic optimization problems with risk objective and constraints*

$$\min_{\mathbf{x}\in\mathcal{C}} \ \mathscr{R}\big(X(\mathbf{x},\omega)\big) \tag{1.21a}$$

$$\min_{\mathbf{x}\in\mathcal{C}} \ \big\{\, g(\mathbf{x}) \ \big|\mathscr{R}\big(X(\mathbf{x},\omega)\big) \leq c \,\big\} \tag{1.21b}$$

where $X(\mathbf{x},\omega)$ *is convex in* $\mathbf{x}$ *over some closed convex set* $\mathcal{C} \subset \mathbb{R}^n$*,* $\mathscr{R}(X)$ *is a risk measure, and* $g(\mathbf{x})$ *is a given convex function on* $\mathcal{C}$*. Assuming that risk measure* $\mathscr{R}$ *has representation* (1.13) *where* ϕ *satisfies the conditions of Theorem 1.3.1, introduce the following counterparts of* (1.21a)–(1.21b)*:*

$$\min_{(\mathbf{x},\eta)\in\mathcal{C}\times\mathbb{R}} \ \eta + \phi\big(X(\mathbf{x},\omega) - \eta\big) \tag{1.22a}$$

$$\min_{(\mathbf{x},\eta)\in\mathcal{C}\times\mathbb{R}} \ \big\{\, g(\mathbf{x}) \ \big| \ \eta + \phi\big(X(\mathbf{x},\omega) - \eta\big) \leq c \,\big\} \tag{1.22b}$$

Then, optimization problems (1.21a) *and* (1.22a) *are equivalent in the sense that they achieve minima at the same values of the decision variable* $\mathbf{x}$ *and their optimal objective values coincide. The same holds for the pair* (1.21b), (1.22b)*. Further, if the risk constraint in* (1.21b) *is binding at optimality,* $(\mathbf{x}^*, \eta^*)$ *achieves the minimum of* (1.22b) *if and only if* $\mathbf{x}^*$ *is an optimal solution of* (1.21b) *and* $\eta^* \in \arg\min_\eta \big\{\, \eta + \phi\big(X(\mathbf{x},\omega) - \eta\big) \,\big\}$.

A similar result also exists for stochastic programming implementation of deviation measures that admit representation (1.18), see Krokhmal (2007).

As an illustration of how incorporation of coherent risk measures or deviation measures that involve higher moments of loss distributions leads to mathematical programming problems with p-order conic constraints, we consider handling an objective or constraint containing HMCR measure. According to the above Theorem it can be expressed using the constraint

$$\mathsf{HMCR}_{p,\alpha}\big(X(\mathbf{x},\omega)\big) \le u$$

with u being either a variable or a constant, correspondingly. By virtue of Theorem 1.5.1 applied to definition (1.14), the latter constraint can be written in the form

$$\eta + (1-\alpha)^{-1}\big\|\big(X(\mathbf{x},\omega)-\eta\big)^+\big\|_p \le u,$$

which, in turn, can be represented by the following set of inequalities

$$u \ge \eta + (1-\alpha)^{-1}t \tag{1.23a}$$

$$t \ge \big(w_1^p + \ldots + w_J^p\big)^{1/p} \tag{1.23b}$$

$$w_j \ge \varpi_j^{1/p}\big(X(\mathbf{x},\omega_j)-\eta\big), \qquad j = 1,\ldots,J \tag{1.23c}$$

$$w_j \ge 0, \qquad j = 1,\ldots,J \tag{1.23d}$$

Constraint (1.23b) defines a $(J+1)$-dimensional cone of order p, and is central to practical implementation of coherent risk measures (1.14), (1.15) and deviation measures (1.19), (1.20) that involve higher moments of distributions in decision making models.

For instance, it is easy to see that implementation of the SMCR measures (1.15) in stochastic programming models can be reduced to essentially the same system of inequalities, complemented by the equality constraint

$$\eta = \sum_{j=1}^{J} \varpi_j X(\mathbf{x},\omega_j) \tag{1.24}$$

Observe that linearity of the loss function $X(\mathbf{x},\omega)$ in $\mathbf{x}$, which is the case in many important applications, immediately guarantees convexity of both sets (1.23) and (1.23)–(1.24), and moreover, the addition of the linear constraint (1.24) eliminates the free variable η in the system of constraints (1.23).

Similarly, the deviation counterparts of HMCR and SMCR risk measures can be incorporated into stochastic programming models by means of p-order conic constraints.

Many models in stochastic optimization are formulated as linear programming problemss of form (1.1) where $X(\mathbf{x},\omega)$ is a linear function of $\mathbf{x}$ and $\mathcal{S}$ is a polyhedron in $\mathbb{R}^n$. Then, in the spirit of the foregoing discussion, such models can be equipped with risk-averse preferences by replacing the expectation operator in the objective function of (1.1) with an appropriately chosen coherent risk measure or deviation measure. In view of this, our interest is in the development of computational procedures that would facilitate the incorporation of coherent and deviation measures based on higher moments of loss distributions in otherwise linear decision models.

The main objective of the presented research endeavor is the development of fast and robust solution algorithms for stochastic optimization problems that can be reduced to linear programming problems with p-order conic consraints:

$$\min \quad \mathbf{c}^\top \mathbf{x} \tag{1.25a}$$

$$\text{s. t.} \quad \mathbf{A}\mathbf{x} \le \mathbf{b} \tag{1.25b}$$

$$\left\|\mathbf{D}^{(k)}\mathbf{x} - \mathbf{f}^{(k)}\right\|_{p_k} \le \mathbf{h}^{(k)\top}\mathbf{x} - g^{(k)}, \quad k = 1,\ldots,K \tag{1.25c}$$

In the discussion that follows we present the details of the proposed approaches to solving the p-order conic programming problems of type (1.25), and illustrate the developed algorithms on a portfolio optimization case study.

CHAPTER 2
A POLYHEDRAL APPROXIMATION APPROACH TO SOLVING *P*-ORDER CONIC PROGRAMMING PROBLEMS

2.1 Introduction

In the previous chapter we have surveyed a number of decision-making models under uncertainty that are based on coherent and deviation measures involving higher moments of loss distributions. The corresponding stochastic optimization problems can, in many cases, be reduced to linear programming problems with p-order conic constraints. In this section we advocate an approach to solving such problems using polyhedral approximations of p-cones, and we develop the corresponding approximations. We also discuss a special case when a p-order conic programming problem can be reduced to second order conic programming problem.

2.2 A polyhedral approximation approach to solving *p*-order conic programming problems

As it has been mentioned in Chapter 1, in this work we are concerned with solving linear programming problems with p-order conic constraints

$$\min \quad \mathbf{c}^\top \mathbf{x} \tag{2.1a}$$

$$\text{s. t.} \quad \mathbf{A}\mathbf{x} \leq \mathbf{b} \tag{2.1b}$$

$$\left\|\mathbf{D}^{(k)}\mathbf{x} - \mathbf{f}^{(k)}\right\|_{p_k} \leq \mathbf{h}^{(k)\top}\mathbf{x} - g^{(k)}, \quad k = 1, \ldots, K, \tag{2.1c}$$

where $\|\cdot\|_p$ denotes the p-norm:

$$\|\mathbf{a}\|_p = \big(|a_1|^p + \cdots + |a_m|^p\big)^{1/p}, \quad \mathbf{a} \in \mathbb{R}^m, \quad p \geq 1.$$

Formulation (2.1) is a generalization of the well-known class of second-order conic programming (SOCP) problems, and therefore we will call (2.1) a p-order conic programming (pOCP) problem.

The key feature of the pOCP problem (2.1) is the p-order conic constraints, which can be stated in a simple form as

$$t \geq (\xi_1^p + \ldots + \xi_J^p)^{1/p}, \tag{2.2}$$

where we can assume without loss of generality that all variables are nonnegative. Depending on the value of the parameter p, the following cases can be identified.

$\underline{p = 1}$: In this case the p-cone constraint (2.2) reduces to a linear inequality

$$t \geq \xi_1 + \ldots + \xi_J.$$

This particular case and the associated stochastic optimization models of type (2.1) have been studied extensively in the context of the Conditional Value-at-Risk measure (Rockafellar and Uryasev, 2000, 2002b; Krokhmal et al., 2002a). In general, the amenability of the 1-norm, also known as the "Manhattan distance," etc., to implementation via linear constraints has been exploited in a variety of approaches and applications too numerous to cite here.

$\underline{p = \infty}$: In this case $\|a\|_\infty = \sup_i |a_i|$, whereby the p-cone constraint (2.2) reduces to a polyhedral set defined by a system of J linear inequalities

$$t \geq \xi_j, \quad j = 1, \ldots, J.$$

Constraints of this type have also been heavily researched in the literature.

$\underline{p = 2}$: In this instance, constraint (2.2) represents a second-order (quadratic, "ice cream", or Lorentz) cone,

$$t \geq (\xi_1^2 + \ldots + \xi_J^2)^{1/2}. \tag{2.3}$$

The second-order conic programming (SOCP) problem, which deals with optimization problems that contain constraints of form (2.3), constitutes a well-developed subject of convex programming. A number of efficient SOCP algorithms have been developed in the literature (e.g., Nesterov and Todd, 1997, 1998, and others. See an overview in Alizadeh and Goldfarb, 2003), and some of them were implemented into software solver codes such as MOSEK and SeDuMi (Andersen et al., 2003; Sturm, 1998).

$\underline{p \in (1,2) \cup (2,+\infty)}$: This is the "general" case that constitutes the focus of our research. In addition to stochastic programming applications described in Chapter 1, the general p-order conic programming has been considered in the context of Steiner minimum tree problem on a given topology (Xue and Ye, 2000); a p-cone relaxation of integer programming problems is discussed in Burer and Chen (2008).

From the computational standpoint, the case of general p, when the cone defined by (2.2) is not self-dual, has received much less attention in the literature compared to the conic quadratic programming. Interior-point approaches to p-order conic programming have been considered by Xue and Ye (2000) with respect to minimization of the sum of p-norms; a self-concordant barrier for p-cones has also been introduced in Nesterov (2006). 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). In the case when p is a rational number, the existing primal-dual methods of SOCP can be employed for solving p-order conic optimization problems using a reduction of p-order conic constraints to a system of linear and second-order conic constraints proposed in Nesterov and Nemirovski (1994) and Ben-Tal and Nemirovski (2001a).

Our approach to solving pOCP problems (2.1) in the case of $p_k \in (1,2) \cup (2,\infty)$ consists of constructing polyhedral approximations for the p-order constraints and subsequent solving of the resulting LP problem using a Benders-type decomposition method (see Chapter 3). In many respects, the proposed approach builds upon the work of Ben-Tal and Nemirovski (2001b) where an efficient lifted polyhedral approximation for the second-order ($p = 2$) cones was developed, and whose motivation was to devise a practical method of solving SOCP problems of the form (2.1) that would utilize the powerful machinery of LP solvers. Given the possibility of reformulating a pOCP problem (2.1) in the case when $p_k \in \mathbb{Q}$ as a SOCP problem (constructive formulas for which are developed in this chapter), the following can be considered as our main motivation for pursuing the polyhedral approximation approach to solving pOCP problems:

- As it will be seen below, a SOCP reformulation in the case of a rational p results in a SOCP problem with much larger number of conic constraints than in the original pOCP problem. Currently, SOCP solvers are more effective in handling SOCP problems with few conic constraints of high dimensionality than in the case when a problem contains a large number of quadratic conic constraints. Actually, this observation also served as a motivation for the development of polyhedral approximations for second-order cones in the work of Ben-Tal and Nemirovski (2001b).

- Many stochastic programming models are formulated as linear programming problems. Due to the specific structure of SP models induced by modeling the uncertainties using discrete scenario sets, a number of iterative algorithms based on some sort of scenario decomposition have been developed for stochastic programming problems in the literature. Good performance of these methods in practice can often be attributed to the "warm start" capabilities of simplex-based LP optimization algorithms. In contrast, interior-point methods, including the SOCP algorithms, are currently lacking the "warm start" capabilities, which may jeopardize the effectiveness of implementation of pOCP-based stochastic optimization models. Thus the development of polyhedral approximations of pOCP problem (2.1) potentially allows for implementing the HMCR, SMCR, and corresponding deviation measures, etc., in large-scale, multi-stage stochastic optimization problems that can be efficiently tackled by the existing decomposition-based SP algorithms that exploit the scenario and stage-specific structure of such problems.

Next we discuss the details of the proposed method of solving the p-order conic programming problems of type (2.1) that relies on the construction of polyhedral approximations for p-order cones. Without loss of generality, we restrict our attention to a p-order cone in the positive orthant of $(J+1)$-dimensional space:

$$\mathcal{K}_p^{(J+1)} = \left\{ \boldsymbol{\xi} \in \mathbb{R}_+^{J+1} \;\middle|\; \xi_{J+1} \geq \left(\xi_1^p + \ldots + \xi_J^p\right)^{1/p} \right\}, \tag{2.4}$$

where $\mathbb{R}_+ = [0, +\infty)$. By a polyhedral approximation of $\mathcal{K}_p^{(J+1)}$ we understand a (convex) polyhedral cone in $\mathbb{R}^{J+1+\kappa_m}$, where κ_m may be generally non-zero:

$$\mathcal{H}_{p,m}^{(J+1)} = \left\{ \begin{pmatrix} \boldsymbol{\xi} \\ \mathbf{u} \end{pmatrix} \in \mathbb{R}_+^{J+1+\kappa_m} \;\middle|\; \mathbf{H}_{p,m}^{(J+1)} \begin{pmatrix} \boldsymbol{\xi} \\ \mathbf{u} \end{pmatrix} \geq \mathbf{0} \right\} \tag{2.5}$$

having the properties that

(H1) any $(\xi_1, \ldots, \xi_{J+1})^\top \in \mathcal{K}_p^{(J+1)}$ can be extended to some $(\xi_1, \ldots, \xi_{J+1}, u_1, \ldots, u_{\kappa_m})^\top \in \mathcal{H}_{p,m}^{(J+1)}$;

(H2) for some prescribed $\varepsilon > 0$, any $(\xi_1, \ldots, u_{\kappa_m})^\top \in \mathcal{H}_{p,m}^{(J+1)}$ satisfies

$$\left(\xi_1^p + \ldots + \xi_J^p\right)^{1/p} \le (1+\varepsilon)\xi_{J+1} \tag{2.6}$$

Here m is the parameter of construction that controls the approximation accuracy ε. Replacing each of the p-order conic constraints in problem (2.1) by their polyhedral approximations (2.5), we obtain an LP approximation of the pOCP problem (2.1)

$$\min \left\{ \mathbf{c}^\top \mathbf{x} \;\middle|\; \mathbf{A}\mathbf{x} \le \mathbf{b},\;\; \mathbf{H}_{p_k,m_k}^{(J_k+1)} \begin{pmatrix} \mathbf{D}^{(k)}\mathbf{x} - \mathbf{f}^{(k)} \\ \mathbf{h}^{(k)\top}\mathbf{x} - g^{(k)} \\ \mathbf{u}^{(k)} \end{pmatrix} \ge \mathbf{0}, \quad k = 1, \ldots, K \right\}. \tag{2.7}$$

Extending the arguments of Ben-Tal and Nemirovski (2001b) to the case of $p \in (1,2) \cup (2,+\infty)$, we observe that the projection of the feasible region of (2.7) on the space of variables $\mathbf{x}$ lies between the feasible set of problem (2.1) and that of its "ε-approximation",

$$\min \left\{ \mathbf{c}^\top \mathbf{x} \;\middle|\; \mathbf{A}\mathbf{x} \le \mathbf{b},\;\; \left\|\mathbf{D}^{(k)}\mathbf{x} - \mathbf{f}^{(k)}\right\|_{p_k} \le (1+\varepsilon)\left(\mathbf{h}^{(k)\top}\mathbf{x} - g^{(k)}\right), \;\; k = 1, \ldots, K \right\}. \tag{2.8}$$

Thus, problem (2.7) represents an ε-approximation of (2.1) if the feasible regions of problems (2.1) and (2.8) are "close". Conditions under which the feasible sets of (2.1) and (2.8) 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.

In constructing the polyhedral approximations, of the form defined by (2.5) for p-order cone (2.4), we follow the approach of Ben-Tal and Nemirovski (2001b) (see also Nesterov and Nemirovski, 1994; Ben-Tal and Nemirovski, 2001a), which allows for reducing the dimensionality of approximation (2.5) by replacing the $(J+1)$-dimensional conic constraint with an equivalent system of 3-dimensional conic constraints, and then constructing a polyhedral approximation for each of the 3D cones.

2.3 Dimension Reduction Techniques

In this section we discuss constructive techniques for representing the p-order cone (2.4) of an arbitrary dimension $J+1$ using conic constraints in $\mathbb{R}^3$, such that the total number of the 3-dimensional conic constraints is $O(J)$.

2.3.1 "Tower-of-variables"

The "tower-of-variables" technique was originally proposed by Ben-Tal and Nemirovski (2001b) for the construction of a polyhedral approximation of the second-order ($p = 2$) conic constraints, and it applies to p-order conic constraints as well. It essentially represents a p-order cone in $(J+1)$-dimensional space as the intersection of $J-1$ 3-dimensional p-order cones.

To demonstrate this technique when applied to p-order cones, we will assume for simplicity that $J = 2^d$ for some $d \in \mathbb{Z}_+$, in which case a (2^d+1)-dimensional p-conic set can be shown to have an equivalent "lifted" representation through 2^d-1 p-conic sets in $\mathbb{R}^3$. This assumption, however, is not restricting in any way, as will be demonstrated by Proposition 1 whose proof furnishes constructive formulas that generalized the original "tower-of-variables" representation of Ben-Tal and Nemirovski (2001b) to arbitrary $J \geq 2$ and $p \geq 1$.

Consider a p-cone in the positive orthant $\mathbb{R}^{J+1}$

$$t \geq (\xi_1^p + \ldots + \xi_J^p)^{1/p}, \tag{2.9}$$

where $J = 2^d$. By introducing new (non-negative) variables

$$\xi_j^{(\ell)}, \quad j = 0, \ldots 2^{d-\ell}, \quad \ell = 0, \ldots d,$$

where

$$\begin{aligned} \xi_j &\equiv \xi_j^{(0)}, \quad j = 1, \ldots, 2^d \\ t &\equiv \xi_1^{(d)} \end{aligned} \tag{2.10}$$

it is easy to see that the above p-cone inequality is equivalent to the following set of p-cone inequalities:

$$\xi_j^{(\ell)} \geq \left(\left(\xi_{2j-1}^{(\ell-1)} \right)^p + \left(\xi_{2j}^{(\ell-1)} \right)^p \right)^{1/p}, \quad \ell = 1, \ldots, d, \quad j = 1, \ldots, 2^{d-\ell} \tag{2.11}$$

This is an equivalent representation of the p-cone set (2.9) in $\mathbb{R}^{J+1}$ in the sense that the collection of variables in (2.10) can be extended to a feasible point of (2.11) if and only if $(\boldsymbol{\xi}, t)$ satisfies (2.10).

The set of constraints (2.11) can be visualized as a “tower” or “pyramid” consisting of $d+1$ “levels” denoted by the superscript $\ell = 0, \ldots, d$, with $2^{d-\ell}$ variables $\xi_j^{(\ell)}$ at level ℓ, such that the $2^d = J$ variables $\xi_j^{(0)} \equiv \xi_j$ represent the “foundation” of the “tower”, and the variable $\xi_1^{(d)} \equiv t$ represents its “top”, or “apex”.

Example 2. Given a p-cone, $t \geq (w_1^p + \ldots + w_8^p)^{1/p}$ in $\mathbb{R}^9$, the described above “tower-of-variables” method can be used to equivalently represent it via $8 - 1 = 7$

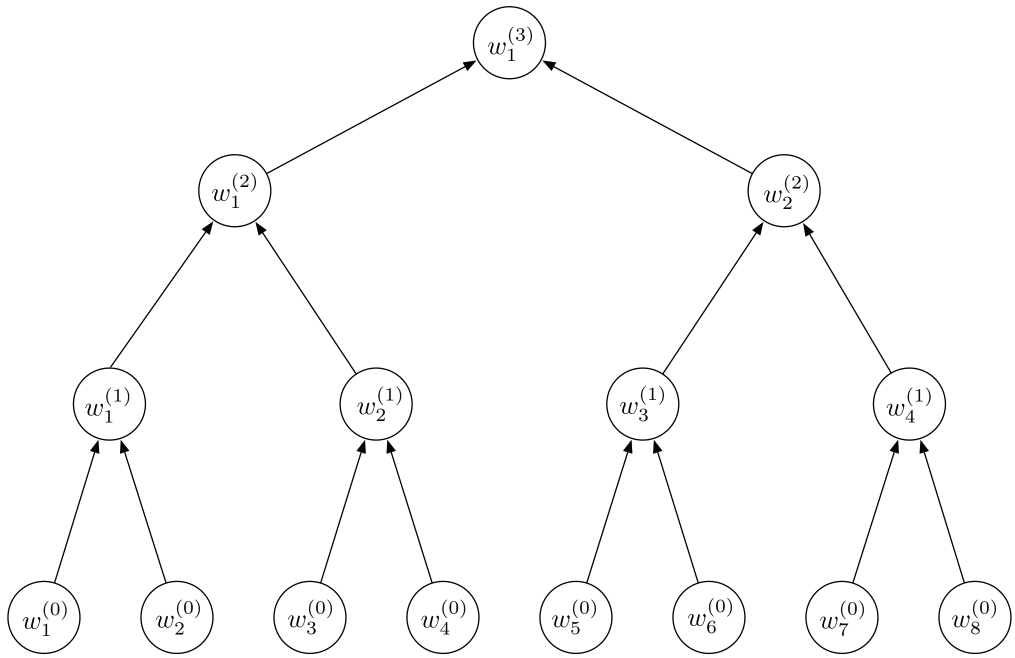

Figure 2.1: Tower of variables with 8 scenarios.

three-dimensional p-cones as follows (see also Figure 2.1):

$$\begin{aligned}
w_1^{(1)} &\geq \Big((w_1^{(0)})^p + (w_2^{(0)})^p \Big)^{1/p} \\
w_2^{(1)} &\geq \Big((w_3^{(0)})^p + (w_4^{(0)})^p \Big)^{1/p} \\
w_3^{(1)} &\geq \Big((w_5^{(0)})^p + (w_6^{(0)})^p \Big)^{1/p} \\
w_4^{(1)} &\geq \Big((w_7^{(0)})^p + (w_8^{(0)})^p \Big)^{1/p} \\
w_1^{(2)} &\geq \Big((w_1^{(1)})^p + (w_2^{(1)})^p \Big)^{1/p} \\
w_2^{(2)} &\geq \Big((w_3^{(1)})^p + (w_4^{(1)})^p \Big)^{1/p} \\
w_1^{(3)} &\geq \Big((w_1^{(2)})^p + (w_2^{(2)})^p \Big)^{1/p}
\end{aligned}$$

As it has been mentioned earlier, the construction for the tower of variables is not limited to cones in $\mathbb{R}^{2^d+1}$. Next we demonstrate how this construction can be extended to the case $J \neq 2^d$, so that the number of 3D p-cones required to equivalently represent a $(J+1)$-dimensional p-cone is still equal to $J-1$.

Proposition 1. *The construction of the tower of variables can be extended to general $J \neq 2^d$ such that a p-cone in $\mathbb{R}^{J+1}$ is represented as intersection of a set of $J-1$ p-cones in $\mathbb{R}^3$.*

Proof. Consider a p-cone in the positive orthant of $\mathbb{R}^{J+1}$:

$$t \geq \left(\xi_1^p + \ldots + \xi_J^p\right)^{1/p}, \tag{2.12}$$

For an integer $J > 2$, let

$$J = \sum_{k=0}^{\bar{d}} \delta_k 2^k, \tag{2.13}$$

where

$$\bar{d} = \lceil \log_2 J \rceil,$$

and $\delta_k \in \{0, 1\}$, i.e., δ_k is the k-th digit in the binary representation of the integer J. Let $\mathcal{D}_J$ denote the (ordered) set of those k in (2.13) for which δ_k are non-zero:

$$\mathcal{D}_J = \{\, k_0 < k_1 < \ldots < k_{s-1} \mid \delta_{k_i} = 1 \,\}, \quad s = |\mathcal{D}_J| = \sum_{k=1}^{\bar{d}} \delta_k. \tag{2.14}$$

Then, for each k such that $1 \leq k \leq \bar{d}$ and $\delta_k \neq 0$, the following constraints can be written:

$$\begin{aligned} &\xi_j^{(\ell)} \geq \left(\left(\xi_{2j-1}^{(\ell-1)}\right)^p + \left(\xi_{2j}^{(\ell-1)}\right)^p \right)^{1/p}, \\ &\qquad j = 1 + \sum_{r=k+1}^{\bar{d}} \delta_r 2^{r-\ell}, \ldots, \sum_{r=k}^{\bar{d}} \delta_r 2^{r-\ell}, \quad \ell = 1, \ldots, k, \quad k \in \mathcal{D}_J \backslash \{0\}. \end{aligned} \tag{2.15}$$

For every $k \in \mathcal{D}_J \backslash \{0\}$ constraints (2.15) define a "sub-tower of variables", each having $k+1$ "levels" including the "foundation" ($\ell = 0$) comprised of 2^k variables $\xi_j^{(0)}$ and the "top" variable

$$\xi_{j_k}^{(k)}, \quad \text{where} \quad j_k = \sum_{r=k}^{\bar{d}} \delta_r 2^{r-k}. \tag{2.16}$$

To complete our representation of $(J+1)$-dimensional p-order conic constraint, we must formulate the corresponding constraints that "connect" the "top" variables (2.16). This can be accomplished in a recursive manner as follows

$$\xi_{\nu_{r+1}}^{(\kappa_{r+1})} \geq \left(\left(\xi_{j_{k_r}}^{(k_r)}\right)^p + \left(\xi_{\nu_r}^{(\kappa_r)}\right)^p \right)^{1/p}, \quad r = 1, \ldots, \sum_{k=0}^{\bar{d}-2} \delta_k, \tag{2.17}$$

where

$$\kappa_1 = k_0, \quad \nu_1 = j_{k_0} \quad \text{and} \quad \kappa_{r+1} = k_r + 1, \quad \nu_{r+1} = \frac{j_{k_r}+1}{2} \;\; \text{for} \;\; r = 1, \ldots, \sum_{k=0}^{\bar{d}-2} \delta_k.$$

It is straightforward to verify that the projection of the set defined by (2.15), (2.17) on the space of variables $\xi_j^{(0)} \equiv \xi_j$, $(j = 1, \ldots, J)$, $\xi_1^{\bar{d}} = t$ is equal to the set (2.12). It is also evident that when $J = 2^d = 2^{\bar{d}}$, the set $\mathcal{D}_J$ will contain just one element: $\mathcal{D}_J = \{\bar{d}\}$, whereby (2.15) reduces to "tower of variables" (2.11) with constraints (2.17) being absent.

Observe that set (2.15) comprises $\sum_{k=1}^{\bar{d}} \delta_k$ "sub-towers", each containing $2^k - 1$ constraints, and set (2.17) consists of $\sum_{k=1}^{\bar{d}-2} \delta_k$ constraints, therefore the representation (2.15), (2.17) of the p-order conic constraint in $\mathbb{R}^{J+1}$ contains

$$\sum_{k=1}^{\bar{d}} \delta_k (2^k - 1) + \sum_{k=1}^{\bar{d}-2} \delta_k = \sum_{k=1}^{\bar{d}} \delta_k 2^k - \delta_{\bar{d}-1} - \delta_{\bar{d}} = J - 1 \tag{2.18}$$

three-dimensional p-order constraints. Indeed,

$$\delta_{\bar{d}-1} + \delta_{\bar{d}} = 1$$

since $\delta_{\bar{d}} = 1$, $\delta_{\bar{d}-1} = 0$ if $J = 2^d = 2^{\bar{d}}$, whereas $\delta_{\bar{d}} = 0$, $\delta_{\bar{d}-1} = 1$ for $J < 2^{\bar{d}}$. □

Remark 1. *The importance of the tower of variables technique lies in the fact that it allows for drastic reductions in the dimensionality of polyhedral approximations that will be presented later on in this chapter. Indeed, it can be shown that the number of facets needed to approximate a second-order cone is* exponential in the cone's dimensions *(Ben-Tal and Nemirovski, 2001b). By representing a p-cone in $\mathbb{R}^{J+1}$ by a sequence of $J-1$ p-cones in $\mathbb{R}^3$, it becomes possible to develop an approximation for p-cone whose dimensionality will increase linearly with the number of dimensions of the original p-cone.*

2.3.2 pOCP to SOCP Reformulation

In the case when the parameter p is a positive rational number, $p = r/s$, the $(J+1)$-dimensional p-order cone

$$t \geq (\xi_1^p + \xi_2^p + \cdots + \xi_J^p)^{1/p}$$

can be represented by a set of linear inequalities and 3D second-order conic constraints (Nesterov and Nemirovski, 1994; Ben-Tal and Nemirovski, 2001a; see also Alizadeh and Goldfarb, 2003), which opens possibilities for handling p-order conic constraints using SOCP methods. Below we demonstrate that the number of 3D second-order conic constraints needed to represent an (r/s)-order cone in $\mathbb{R}^{J+1}$ is no more than $O(J \log r)$.

Let $p = r/s$, $r > s$. Then, using the non-negativity of variables ξ_j and t, we can rewrite inequality (2.9) as follows:

$$\begin{aligned} & t^{\frac{r}{s}} \geq \xi_1^{\frac{r}{s}} + \cdots + \xi_J^{\frac{r}{s}} \\ \Leftrightarrow \quad & t^{\frac{r}{s}+1-1} \geq \xi_1^{\frac{r}{s}} + \cdots + \xi_J^{\frac{r}{s}} \\ \Leftrightarrow \quad & t \geq \sum_{j=1}^{J} \xi_j^{\frac{r}{s}} t^{1-\frac{r}{s}} \end{aligned}$$

Introducing new non-negative variables u_j such that $u_j \geq \xi_j^{\frac{r}{s}} t^{1-\frac{r}{s}}$, $j = 1, \ldots, J$, we arrive at the following system of inequalities as a representation of (2.9):

$$t \geq \sum_{j=1}^{J} u_j \tag{2.19a}$$

$$u_j^s t^{r-s} \geq \xi_j^r, \quad j = 1, \ldots, J \tag{2.19b}$$

$$u_j \geq 0, \quad j = 1, \ldots, J \tag{2.19c}$$

Each constraint (2.19b) can now be equivalently represented by a system of "rotated" second order conic inequalities using similar principles that were employed in the "tower of variables" construction. Such a representation, however, is not unique. One way of doing that is to rewrite each inequality in (2.19b) as

$$u_j^s t^{r-s} \xi^{R-r} \geq \xi_j^R, \quad j = 1, \ldots, J \tag{2.20}$$

where

$$R = 2^\rho, \quad \rho = \lceil \log_2 r \rceil$$

Then, by invoking the "tower of variables" principle, each of J constraints (2.20) can be expressed via $2^\rho - 1$ inequalities of the form $z^2 \le xy$:

$$\xi_j^2 \le v_{j,1}^{(\rho-1)} v_{j,2}^{(\rho-1)} \tag{2.21a}$$

$$(v_{j,k}^{(l)})^2 \le v_{j,2k-1}^{(l-1)} v_{j,2k}^{(l-1)}, \quad l = 2, \ldots, \rho - 1; \quad k = 1, \cdots, 2^{\rho-l} \tag{2.21b}$$

$$(v_{j,k}^{(1)})^2 \le u_j^2, \quad k = 1, \ldots, \left\lfloor \frac{s}{2} \right\rfloor \tag{2.21c}$$

$$(v_{j,k}^{(1)})^2 \le u_j t, \quad k = \left\lfloor \frac{s}{2} \right\rfloor + 1, \ldots, \left\lceil \frac{s}{2} \right\rceil \tag{2.21d}$$

$$(v_{j,k}^{(1)})^2 \le t^2, \quad k = \left\lceil \frac{s}{2} \right\rceil + 1, \ldots, \left\lfloor \frac{r}{2} \right\rfloor \tag{2.21e}$$

$$(v_{j,k}^{(1)})^2 \le t\xi_j, \quad k = \left\lfloor \frac{r}{2} \right\rfloor + 1, \ldots, \left\lceil \frac{r}{2} \right\rceil \tag{2.21f}$$

$$(v_{j,k}^{(1)})^2 \le \xi_j^2, \quad k = \left\lceil \frac{r}{2} \right\rceil + 1, \ldots, \frac{R}{2}. \tag{2.21g}$$

Similar to the "tower of variables" construction in (2.11), the set of inequalities in (2.21) can be visualized as a binary tree whose nodes represent the variables in (2.21). Each inequality in (2.21) can then be viewed as a subgraph with two arcs that connect the "parent" node (the variable at the left-hand side of a constraint) to two "child" nodes (the variables at the right-hand side of same constraint). In such a way, the variable ξ_j in (2.21a) represents the root node of the binary tree ("top of the tower"), and the variables u_j, t, and ξ_j in (2.21c)–(2.21g) represent the leaf nodes of the tree ("bottom of the tower").

Figure 2.3 presents an illustration of the "tower-of-variables" principle employed in the construction of set (2.21) on the example of $p = \frac{5}{3}$.

An inequality of the form

$$z^2 \le xy, \tag{2.22}$$

or, generally,

$$\mathbf{z}^\top \mathbf{z} \le xy,$$

is known as a "rotated" second order cone, for it is equivalent to the standard second order cone:

$$(2z)^2 + (x-y)^2 \leq (x+y)^2, \tag{2.23}$$

or, correspondingly,

$$4\mathbf{z}^\top \mathbf{z} + (x-y)^2 \leq (x+y)^2.$$

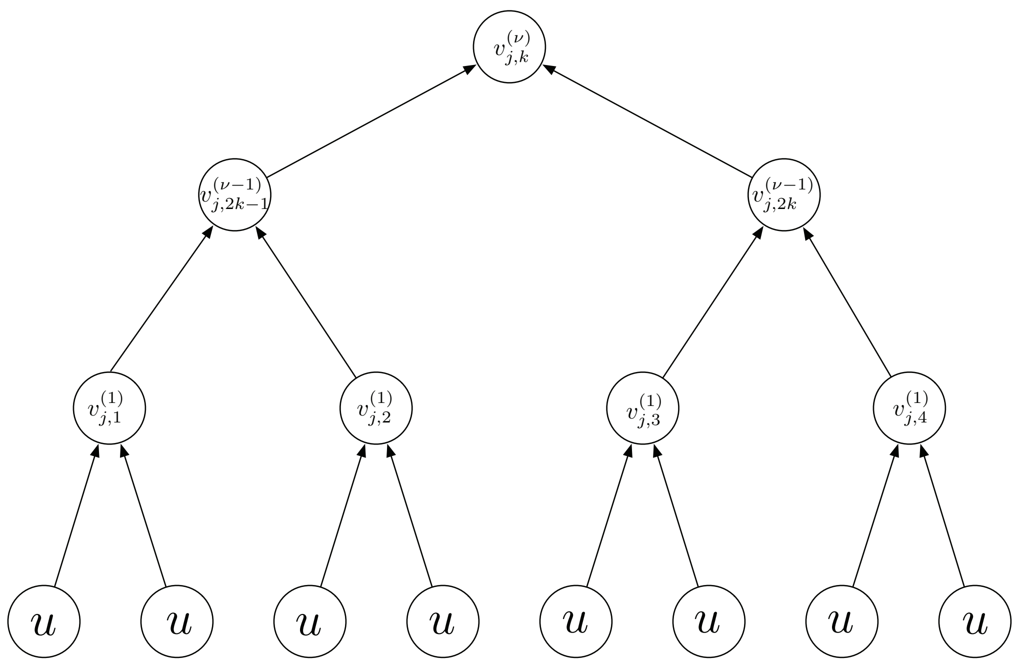

Figure 2.2: Subtree with linear constraints.

Note, however, that not all the inequalities of (2.21) are "true" second order cones. In fact, we have the following result:

Proposition 2. *When p is a positive rational number, $p = r/s$, such that $r > s$ and r and s do not have common multiples except 1, a p-order conic constraint in the* *positive orthant of* $\mathbb{R}^{J+1}$ *can be equivalently represented by a set of linear constraints and* $O(J \log r)$ *second-order conic constraints in* $\mathbb{R}^3$*.*

Proof. Consider the constraints (2.21c)-(2.21g) at the bottom of the tower of variables as constructed by the system of inequalities in (2.21). It is easy to see that only (2.21d) and (2.21f) are "true" second order conic constraints while constraints (2.21c), (2.21e), and (2.21g) consist of all positive variables and thus, taking a square root would reduce these inequalities to linear inequalities.

Next, observe that among inequalities (2.21b), only those that have inequalities (2.21d) or (2.21f) as children will evolve into true second order conic constraints (see Figure 2.3). Assume without loss of generality that for some $\nu > 1$ the subtree of (2.21) with the root node $v_{j,1}^{(\nu)}$ has all of its 2^ν leaf nodes as variables u_j of (2.21c) (see Figure 2.2), and consider the set $\mathcal{C}$ defined by the constraints of this subtree:

$$\mathcal{C} = \left\{ v_k^{(l)}, u \geq 0 \;\middle|\; \begin{array}{lll} \left(v_k^{(l)}\right)^2 \leq v_{2k-1}^{(l-1)} v_{2k}^{(l-1)}, & k = 1, \ldots, 2^{\nu-l}, & l = 2, \ldots, \nu \\ \left(v_k^{(1)}\right)^2 \leq u^2, & k = 1, \ldots, 2^{\nu-1} & \end{array} \right\} \tag{2.24}$$

where the subscript j is omitted for brevity. Then, the projection of the set $\mathcal{C}$ on the space of variables $(v_k^{(\nu)}, u)$ is equal to

$$\mathcal{C}' = \left\{ v_k^{(\nu)} \leq u \right\}$$

First, let us show that $\mathcal{C}$ reduces to $\mathcal{C}'$. The inequalities at the base of the subtree are:

$$\left(v_k^{(1)}\right)^2 \leq u^2, \quad k = 1, \ldots, 2^{\nu-1} \tag{2.25}$$

which reduce to:

$$v_k^{(1)} \leq u, \quad k = 1, \ldots, 2^{\nu-1}.$$

The next level on the tower would yield the constraints:

$$\left(v_k^{(2)}\right)^2 \leq v_{2k-1}^{(1)} v_{2k}^{(1)}, \quad k = 1, \ldots, 2^{\nu-2} \tag{2.26}$$

On substituting (2.25) into (2.26) we obtain that

$$v_k^{(2)} \leq u \quad k = 1, \ldots, 2^{\nu-2}$$

Continuing this process by chain substitution of the conic inequalities defining the set $\mathcal{C}$ into each other we get

$$v_1^{(\nu)} \leq \left(v_1^{(\nu-1)} v_2^{(\nu-1)}\right)^{1/2} \leq \ldots \leq \left(\prod_{k=1}^{2^{\nu-1}} v_k^{(1)}\right)^{1/2^{\nu-1}} \leq u$$

Next we show that $\mathcal{C}'$ can be extended to $\mathcal{C}$. for each $\left(v_1^{(\nu)}, u\right) \in \mathcal{C}'$ one can always select $\tilde{v}_1^{(\nu-1)} = \tilde{v}_2^{(\nu-1)} = \ldots = \tilde{v}_1^{(1)} = \ldots = \tilde{v}_{2^{\nu-1}}^{(1)} = u$ so that

$$\left(v_1^{(\nu)}; \tilde{v}_1^{(\nu-1)}, \tilde{v}_2^{(\nu-1)}; \ldots; \tilde{v}_1^{(1)}, \ldots, \tilde{v}_{2^{\nu-1}}^{(1)}; u\right) \in \mathcal{C}$$

Thus, we have shown that $\mathcal{C}'$ can be extended to $\mathcal{C}$ and vice versa.

Hence, among the constraints (2.21) only those are the "true" conic constraints that correspond to root nodes of subtrees whose leaf nodes are not comprised of the same variable. Then, it is easy to see that these second-order conic constraints in (2.21) correspond to nodes that belong to paths from the root note ξ_j in (2.21a) to $v_{j,k}^{(1)}$ in (2.21d) and (2.21f). Assuming that r and s do not have common multiples other than 1, so that they cannot be even simultaneously, the number of second-order conic constraints among (2.21c)–(2.21g) equals to

$$n_{rs} = \left(\lceil s/2 \rceil - \lfloor s/2 \rfloor\right) + \left(\lceil r/2 \rceil - \lfloor r/2 \rfloor\right) \in \{1, 2\}.$$

According to the above observation on the linear nature of "subtowers" that have leaf nodes comprised of the same variable, the number of "true" second-order conic constraints at level $l = 2$ will be also no more than n_{rs}, and so on.

Given that there are $\rho + 1$ "levels" in the tower-of-variables (2.21) and that in the case of $n_{rs} = 2$ the two paths share the common root ξ_j in (2.21a), there are $n_{rs}(\rho - 1) + 1$ conic constraints in (2.21). Hence the p-order cone (2.9) in $\mathbb{R}^{J+1}$ can be equivalently represented by a set of linear constraints and $J(n_{rs}(\rho - 1) + 1) = O(J \log_2 r)$ rotated second-order conic constraints in $\mathbb{R}^3$. $\square$

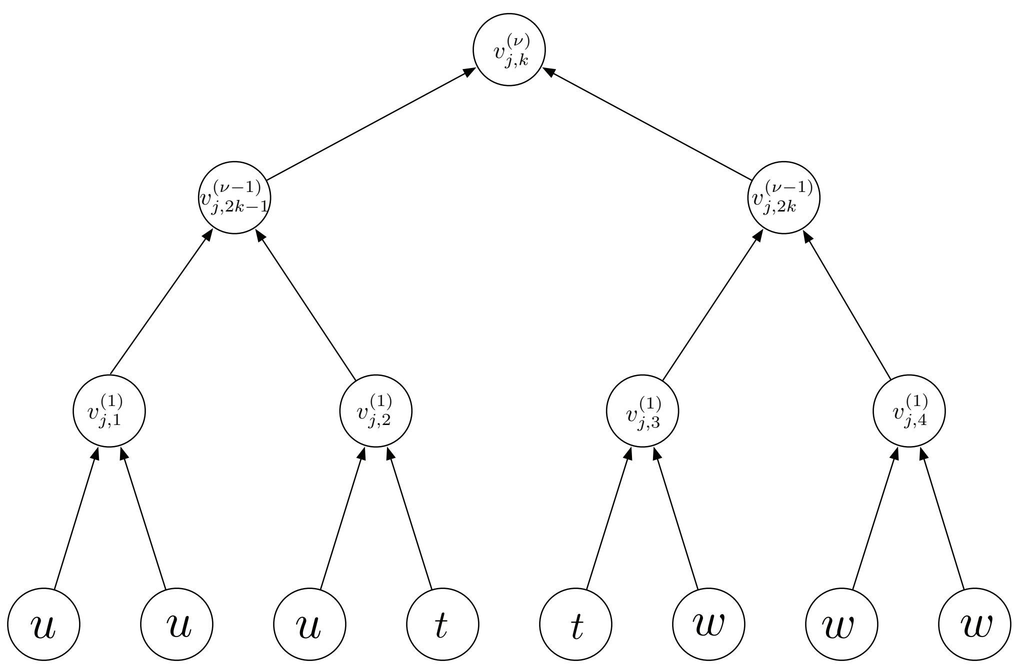

Figure 2.3: Subtree with conic constraints.

Proposition 2 provides constructive formulas for reformulating the pOCP problem (2.1) with rational parameters p_k as a SOCP problem. However, current interior point solvers for second order conic programming are less effective when the number of second order conic constraints in the problem is large. Therefore, getting a bound on the number of second order cones that result from the pOCP→SOCP reformulation is key to the efficiency of this approach. We also note that in many cases it is possible to make the pOCP→SOCP reformulation (2.19)–(2.21) in the case of a rational p even more economical in the number of true second-order rotated cones by rearranging the order of variables u_j, t, ξ_j in (2.21a)–(2.21g). Next we discuss the technical details of implementing the pOCP constraints via SOCP constraints using commercially available solvers.

2.3.3 Implementation using SOCP solvers

Many existing SOCP solvers, such as CPLEX, MOSEK, etc., require additional variables to be introduced in order for the solver to process the "standard" and "rotated" second order conic constraints properly. For example, to implement a "rotated" second order cone (2.22) to be used with CPLEX Barrier solver one has to replace (2.22) with the following system of constraints:

$$(2z)^2 + v_0^2 \leq v_1^2 \tag{2.27a}$$

$$v_0 = x - y \tag{2.27b}$$

$$v_1 = x + y \tag{2.27c}$$

While the MOSEK solver is capable of handling the rotated second-order cone constraints directly, it requires that no variable may appear in different conic constraints simultaneously. This necessitates introduction of additional auxiliary variables and linear constraints in the problem. For instance, the following two rotates second-order constraints

$$z^2 \leq xy, \quad x^2 \leq uv$$

must be implemented for use with the MOSEK solver as, for instance,

$$z^2 \leq xy, \quad t^2 \leq uv, \quad t = x.$$

A simplification that results in reduction in the number of second-order constraints in (2.21) is available for integer values of the parameter $p \in \mathbb{Z}_+$, namely $p = \frac{r}{1}, r \in \mathbb{N}$. Since the variable u enters the inequality (2.20) linearly, we can exploit the fact that either r is even and therefore, $r-s$ is odd and $R-r$ is even, or r is odd and therefore $r-s$ is even and $R-r$ is odd. With this in mind, we were able to reduce the number of "true" second order conic constraints at the "bottom" of the tower of variables to 1:

$$\xi_j^R \leq \begin{cases} u_j t^{r-s} \xi^{R-r} & \text{if } p \text{ is even} \\ u_j \xi^{R-r} t^{r-s} & \text{if } p \text{ is odd} \end{cases} \tag{2.28}$$

This led to our problem generating exactly $J(2(\varrho-1)+1)$ "rotated" conic constraints all of which were generated by either an inequality of type (2.21d) or (2.21f) at the "bottom" level of the tower of variables. In order to deal with the "rotated" conic constraint we needed to use the construction of the system of inequalities in (2.27) to transform each "rotated" second order conic constraint into a standard second order conic constraint. In either of the two cases in (2.28), we had only one set of conic constraints being generated due to the commutativity of the variables.

2.4 Polyhedral Approximations of 3D *p*-Order Cones

The "tower-of-variables" technique and the pOCP→SOCP reformulation discussed in the previous sections enables one to represent a p-order cone in multi-dimensional space R^{J+1} as an intersection of $O(J)$ 3-dimensional p-order cones or second-order cones. This, in turn, allows for reducing the problem of constructing a polyhedral approximation for multidimensional p-cone $\mathcal{K}_p^{(J+1)}$ to the problem of polyhedral approximation of 3-dimensional p-cone $\mathcal{K}_p^{(3)}$. In this section we will discuss the construction of polyhedral approximations for 3D p-cones.

2.4.1 Lifted polyhedral approximation of 3D second-order cone $\mathcal{K}_2^{(3)}$

Ben-Tal and Nemirovski (2001b) considered the problem of constructing an "efficient" polyhedral approximation for a second order cone $\mathcal{K}_2^{(3)}$. Their motivation was the fact that, although efficient interior point methods have been developed for SOCP problems that scale well with increased *dimensionality* of the second-order cones, the effectiveness of these algorithms (or their current implementations) generally deteriorates when the *number* of conic constraints is large. This is not the case for linear programming problems, where the existing solution methods are capable of handling equally well LP instances with large number of variables and large number of constraints.

Our approach to handling pOCP problems builds, to a large degree, on the approach of Ben-Tal and Nemirovski (2001b), which is outlined below. It is interesting to note, however, that a computational study by Glineur (2000) showed that despite the mathematical elegance of Ben-Tal and Nemirovski's polyhedral approximation, the resulting LP approximations generally did not outperform the interior-point SOCP solvers. Nonetheless, as we will demonstrate in Chapter 4, the polyhedral approximation approach works well for the case of general p-order conic programming problems, and is capable of outperforming the SOCP-based approach to pOCP problems that was presented above. In addition, Ben-Tal and Nemirovski's polyhedral approximation has recently been proven quite effective in solving mixed integer conic quadratic problems (Vielma, Ahmed, and Nemhauser, 2008).

The general form of the conic quadratic programming problem that was considered by Ben-Tal and Nemirovski (2001b) was as follows:

$$\min_{\mathbf{x}} \left\{ \mathbf{e}^\top \mathbf{x} \mid \mathbf{A}\mathbf{x} \geq \mathbf{b}, \|\mathbf{A}_l \mathbf{x} - \mathbf{b}_l\|_2 \leq \mathbf{c}_l^\top \mathbf{x} - d_l, \quad l = 1, \ldots, m \right\}, \tag{2.29}$$

where $\|\mathbf{y}\|_2 = \sqrt{\mathbf{y}^\top \mathbf{y}}$ is the standard euclidean norm. The significance of their contribution was the fact that the constructed LP approximation of the SOCP problem (2.29) has the same "size", up to a factor $O(\log 1/\varepsilon)$, where ε is the accuracy of the approximating LP:

$$\min_{\mathbf{x}, \{\mathbf{u}_l\}_{l=1}^m} \quad \mathbf{e}^\top \mathbf{x} \tag{2.30a}$$

$$\text{s. t.} \quad \mathbf{A}\mathbf{x} \geq \mathbf{b} \tag{2.30b}$$

$$\mathbf{H}^{k_l} \begin{pmatrix} \mathbf{A}_l \mathbf{x} - \mathbf{b}_l \\ \mathbf{c}_l^\top \mathbf{x} - d_l \\ \mathbf{u}_l \end{pmatrix} \geq \mathbf{0}, \quad l = 1, \ldots, m \tag{2.30c}$$

The definition of a polyhedral approximation links (2.29) and (2.30) in the sense that both problems have the same objective and the projection of the feasible set of (2.30) onto the $\mathbf{x}$-space is in between the feasible set of (2.29) and that of its ε-relaxation given by:

$$\min_{\mathbf{x}} \left\{ \mathbf{e}^\top \mathbf{x} \mid \mathbf{A}\mathbf{x} \geq \mathbf{b}, \|\mathbf{A}_l \mathbf{x} - \mathbf{b}_l\|_2 \leq (1+\varepsilon)\big(\mathbf{c}_l^\top \mathbf{x} - d_l\big), \quad l = 1, \ldots, m \right\} \tag{2.31}$$

Note that if ε is "small", then (2.30) is a "good" approximation of (2.29). Moreover, it has been shown that the constructed approximation is the "tightest", in the sense that the bound on the size of approximating LP problem cannot be reduced.

Two key parts in Ben-Tal and Nemirovski's method of constructing a polyhedral approximation to a second-order cone is the "tower-of-variables" method of representing a quadratic cone in 2^d+1 dimensions via intersection of 2^d-1 3-dimensional quadratic cones

$$\mathbf{L}^2 = \left\{\boldsymbol{\xi} \in \mathbb{R}^3 | \sqrt{|\xi_1|^2 + |\xi_2|^2} \leq \xi_3\right\}$$

and then constructing a "lifted" polyhedral approximation of $\mathbf{L}^2$ that employs additional variables u_j, v_j:

$$\begin{aligned}
&(a)\begin{cases} u_0 \geq |\xi_1| \\ v_0 \geq |\xi_2| \end{cases} \\
&(b)\begin{cases} u_j = \cos\left(\frac{\pi}{2^{j+1}}\right) u_{j-1} + \sin\left(\frac{\pi}{2^{j+1}}\right) v_{j-1} \\ v_j \geq \left| -\sin\left(\frac{\pi}{2^{j+1}}\right) u_{j-1} + \cos\left(\frac{\pi}{2^{j+1}}\right) v_{j-1}\right| \end{cases} \qquad j = 1, \ldots, m \\
&(c)\begin{cases} u_m \leq \xi_3 \\ v_m \leq \tan\left(\frac{\pi}{2^{m+1}}\right) u_m \end{cases}
\end{aligned} \tag{2.32}$$

where $m \in \mathbb{Z}_+$ is the parameter of the construction that controls approximation accuracy.

We should note here that the system of inequalities above can be straightforwardly written as a system of linear homogeneous inequalities $\mathbf{H}_{2,m}^{(3)} \begin{pmatrix} \boldsymbol{\xi} \\ \mathbf{u} \end{pmatrix} \geq \mathbf{0}$, where $\mathbf{w}$ is the vector of the $2(m+1)$ variables $u_j, v_j,\ j = 0, \ldots, m$. Of particular importance for this approximation of the 3 dimensional Lorentz cone is the following result concerning the error of the approximation:

Proposition 3 (Ben-Tal and Nemirovski (2001b)). *The convex polyhedral set*

$$\mathcal{H}_{2,m}^{(3)} = \left\{ \begin{pmatrix} \boldsymbol{\xi} \\ \mathbf{u} \end{pmatrix} \in \mathbb{R}^{3+\kappa_m} \;\middle|\; \mathbf{H}_{2,m}^{(3)} \begin{pmatrix} \boldsymbol{\xi} \\ \mathbf{u} \end{pmatrix} \geq \mathbf{0} \right\}$$

where the matrix $\mathbf{H}_{2,m}^{(3)}$ *is defined by inequalities* (2.32) *is a polyhedral* $\varepsilon(m)$*-approximation of* $\mathbf{L}^2$ *satisfying (H1) and (H2) with*

$$\varepsilon(m) = \frac{1}{\cos\left(\frac{\pi}{2^{m+1}}\right)} - 1 = O\left(\frac{1}{4^m}\right).$$

By expressing the Lorentz cone $\mathbf{L}^k = \{\boldsymbol{\xi} \in \mathbb{R}^{k+1} \mid (\xi_1^2 + \cdots + \xi_k^2)^{1/2} \leq \xi_{k+1}\}$ of dimension $k+1$ as a system of conic quadratic constraints of dimension 3 each, we can use the construction of (2.32) to approximate the $\mathbf{L}^k$ cone by approximating each of the cones of dimension 3. With this generalization it was shown that the following system of inequalities is as an approximation for the Lorentz cone $\mathbf{L}^k$ of dimension $k+1$:

$$\begin{array}{ll}
(a_{\ell,i}) \begin{cases} u_{\ell,i}^0 \geq |\xi_{2i-1}^{\ell-1}| \\ \nu_{\ell,i}^0 \geq |\xi_{2i}^{\ell-1}| \end{cases} & \\
(b_{\ell,i}) \begin{cases} u_{\ell,i}^j = \cos\left(\frac{\pi}{2^{j+1}}\right) u_{\ell,i}^{j-1} + \sin\left(\frac{\pi}{2^{j+1}}\right) \nu_{\ell,i}^{j-1} \\ \nu_{\ell,i}^j \geq |-\sin\left(\frac{\pi}{2^{j+1}}\right) u_{\ell,i}^{j-1} + \cos\left(\frac{\pi}{2^{j+1}}\right) \nu_{\ell,i}^{j-1}| \end{cases} & j = 1, \cdots, m_\ell \\
(c_{\ell,i}) \begin{cases} u_{\ell,i}^{m_\ell} \leq \xi_i^\ell \\ \nu_{\ell,i}^{m_\ell} \leq \tan\left(\frac{\pi}{2^{m+1}}\right) \xi^{m_\ell} \end{cases} & \\
\quad i = 1, \cdots, 2^{d-\ell}, \quad \ell = 1, \cdots, d. &
\end{array} \tag{2.33}$$

One of the wonderful properties that make this approximation very attractive for practical usage is the quality, β, of the approximation:

$$\beta = \beta(m_1, \cdots, m_d) = \prod_{\ell=1}^{d} \frac{1}{\cos\left(\frac{\pi}{2^{m_\ell+1}}\right)} - 1 \tag{2.34}$$

Given $\varepsilon \in (0, 1]$ and setting

$$m_\ell = \left\lfloor O(1)\ell \ln \frac{2}{\varepsilon} \right\rfloor, \quad \ell = 1, \cdots, d,$$

with properly chosen absolute constant $O(1)$, it can be ensured that

$$\beta(m_1, \cdots, m_d) \leq \varepsilon$$

An important aspect of the ε-approximation (2.30) based on (2.33) is whether or not its solution is "close enough" to the solution of the original SOCP problem (2.29).

As it turns out, there are examples that can be constructed where (2.29) is infeasible, while all problem of the form (2.31) are feasible. However, there is a a simple sufficient condition that will ensure that the feasible sets of (2.29) and (2.31)are "$O(\varepsilon)$-close" to each other (see Ben-Tal and Nemirovski (2001b), Proposition 4.1). Thus, we have (2.30) being a good approximation of (2.29) provided that the feasible sets of (2.29) and (2.31) are within $O(\varepsilon)$ of each other.

2.4.2 Gradient Approximation of 3D general p-order cones

The "lifted" polyhedral approximation of 3D quadratic cones, due to Ben-Tal and Nemirovski (2001b), does not seem to be extensible to the general p-order cones with $p \in [1, +\infty)$. In light of this, we develop a simple "gradient" approximation of the 3D p-cone $\mathcal{K}_p^{(3)}$ by circumscribed planes. With an external polyhedral approximation we create a convex hull that approximates the p-cone, thus allowing the use of linear programming techniques for handling the pOCP problem (2.1). Next we demonstrate the construction of a gradient polyhedral approximation to cone $\mathcal{K}_p^{(3)}$ located in the positive orthant of $\mathbb{R}^3_+$.

In the positive quadrant of $\mathbb{R}^2_+$, the projection of the p-cone

$$x^p + y^p = z_0^p \quad (= \text{const}) \tag{2.35}$$

can be parameterized using the polar coordinates as

$$\begin{aligned} x &= z_0\rho(\theta)\cos\theta \\ y &= z_0\rho(\theta)\sin\theta, \quad \theta \in \left[0, \frac{\pi}{2}\right] \end{aligned} \tag{2.36}$$

Substituting the parameterization for x and y into the equality $x^p + y^p = z_0^p$ and solving for $\rho(\theta)$ yields:

$$\rho(\theta) = \frac{z_0}{(\cos^p\theta + \sin^p\theta)^{1/p}} \tag{2.37}$$

Therefore we have:

$$\begin{aligned} x &= z_0\frac{\cos\theta}{(\cos^p\theta + \sin^p\theta)^{1/p}} \\ y &= z_0\frac{\sin\theta}{(\cos^p\theta + \sin^p\theta)^{1/p}}, \quad \theta \in \left[0, \frac{\pi}{2}\right] \end{aligned} \tag{2.38}$$

Next, observe that the plane tangent to the surface $z^p = x^p + y^p$ of the p-cone at a point $(x_0, y_0, z_0) \in \mathbb{R}^3_+$ is given by

$$(z_0)^{p-1}z = (x_0)^{p-1}x + (y_0)^{p-1}y.$$

Using the parametrization (2.38) of x_0, y_0, we arrive at the following expression for a plane that is tangent to the p-cone $x^p + y^p = z^p$ at the polar angle θ:

$$z = x\frac{\cos^{p-1}\theta}{(\sin^p\theta + \cos^p\theta)^{\frac{p-1}{p}}} + y\frac{\sin^{p-1}\theta}{(\sin^p\theta + \cos^p\theta)^{\frac{p-1}{p}}}, \quad \theta \in \left[0, \frac{\pi}{2}\right].$$

Then, a polyhedral approximation of the p-cone $\mathcal{K}_p^{(3)} \subset \mathbb{R}_+^3$ given by $m+1$ circumscribed tangent planes can be written as

$$\hat{\mathcal{H}}_{p,m}^{(3)} = \left\{ \boldsymbol{\xi} \in \mathbb{R}_+^3 \;\middle|\; \xi_3 \geq \alpha_i^{(p)} \xi_1 + \beta_i^{(p)} \xi_2, \quad i = 0, \ldots, m \right\} \tag{2.39}$$

where the coefficients $\alpha_i^{(p)}$, $\beta_i^{(p)}$ have the form

$$\begin{aligned} \alpha_i^{(p)} &= \frac{\cos^{p-1}\theta_i}{(\cos^p \theta_i + \sin^p \theta_i)^{\frac{p-1}{p}}} \\ \beta_i^{(p)} &= \frac{\sin^{p-1}\theta_i}{(\cos^p \theta_i + \sin^p \theta_i)^{\frac{p-1}{p}}} \end{aligned} \tag{2.40}$$

and $0 = \theta_0 < \theta_1 < \cdots < \theta_{m-1} < \theta_m = \frac{\pi}{2}$ is a partition of the segment $\left[0, \frac{\pi}{2}\right]$. As it is shown in Chapter 3, of particular importance is the special case of the gradient approximation (2.39) where the parameters θ_i represent a "uniform" partition of $\left[0, \frac{\pi}{2}\right]$:

$$\theta_i = \frac{\pi i}{2m}, \quad i = 0, \ldots, m. \tag{2.41}$$

The following proposition establishes approximation quality for the uniform gradient approximation (2.39)–(2.41) of the cone $\mathcal{K}_p^{(3)}$.

Proposition 4. *The set* $\hat{\mathcal{H}}_{p,m}^{(3)}$ (2.39)–(2.40) *defined by the uniform partition* (2.41) *is a convex polyhedral approximation of the p-cone* $\mathcal{K}_p^{(3)}$ *that satisfies properties (H1)–(H2) with approximation accuracy*

$$\varepsilon = \begin{cases} O(m^{-2}), & \textit{for } p \in [2, \infty) \\ O(m^{-p}), & \textit{for } p \in (1, 2) \end{cases}$$

Proof. Since the polyhedral set (2.39)–(2.40) is formed by intersection of halfspaces tangent to the p-cone $\mathcal{K}_p^{(3)}$, (H1) is obviously satisfied.

To demonstrate (H2), we need to show that a finite ε exists such that

$$\boldsymbol{\xi} \in \hat{\mathcal{H}}^{(3)}_{p,m} \implies \|(\xi_1, \xi_2)\|_p \leq (1+\varepsilon)\xi_3 \tag{2.42}$$

holds for any $\boldsymbol{\xi} \in \hat{\mathcal{H}}^{(3)}_{p,m}$. The accuracy in (2.42) can be chosen as the smallest ε that satisfies:

$$\varepsilon \geq \left\| \left(\frac{\xi_1}{\xi_3}, \frac{\xi_2}{\xi_3} \right) \right\|_p - 1.$$

for any $\boldsymbol{\xi} \in \hat{\mathcal{H}}^{(3)}_{p,m}$. Since we are concerned with an ordered pair $\left(\frac{\xi_1}{\xi_3}, \frac{\xi_2}{\xi_3}\right)$, the problem can be reduced to a two dimensional one by letting $(x, y) = \left(\frac{\xi_1}{\xi_3}, \frac{\xi_2}{\xi_3}\right)$, thus reducing the last inequality to

$$\varepsilon \geq \|(x, y)\|_p - 1 \tag{2.43}$$

where (x, y) belongs to the polygon

$$\mathcal{H}' = \left\{ (x, y) \mid 1 \geq \alpha_i^{(p)} x + \beta_i^{(p)} y, \quad i = 0, \ldots, m \right\}$$

From geometric considerations (see Figure 2.4), the approximation error will be largest for vertices of the polygon $\mathcal{H}'$. For a segment $[\theta_i, \theta_{i+1}]$, with $\theta_0 = 0 < \theta_1 < \cdots < \theta_{m-1} < \theta_m = \frac{\pi}{2}$, define the largest approximation error on the segment as ε_i. Thus, ε in (2.43) would be defined as:

$$\varepsilon = \max_{i=0,\ldots,m} \varepsilon_i \tag{2.44}$$

Consider the segment $[\theta_i, \theta_{i+1}]$. We construct the tangent lines to the p-curve $x^p + y^p = 1$ at the points (x_1, y_1) and (x_2, y_2) where:

$$x_1 = \frac{\cos\theta_i}{(\cos^p \theta_i + \sin^p \theta_i)^{1/p}} \qquad y_1 = \frac{\sin\theta_i}{(\cos^p \theta_i + \sin^p \theta_i)^{1/p}}$$
$$x_2 = \frac{\cos\theta_{i+1}}{(\cos^p \theta_{i+1} + \sin^p \theta_{i+1})^{1/p}} \qquad y_2 = \frac{\sin\theta_{i+1}}{(\cos^p \theta_{i+1} + \sin^p \theta_{i+1})^{1/p}}$$

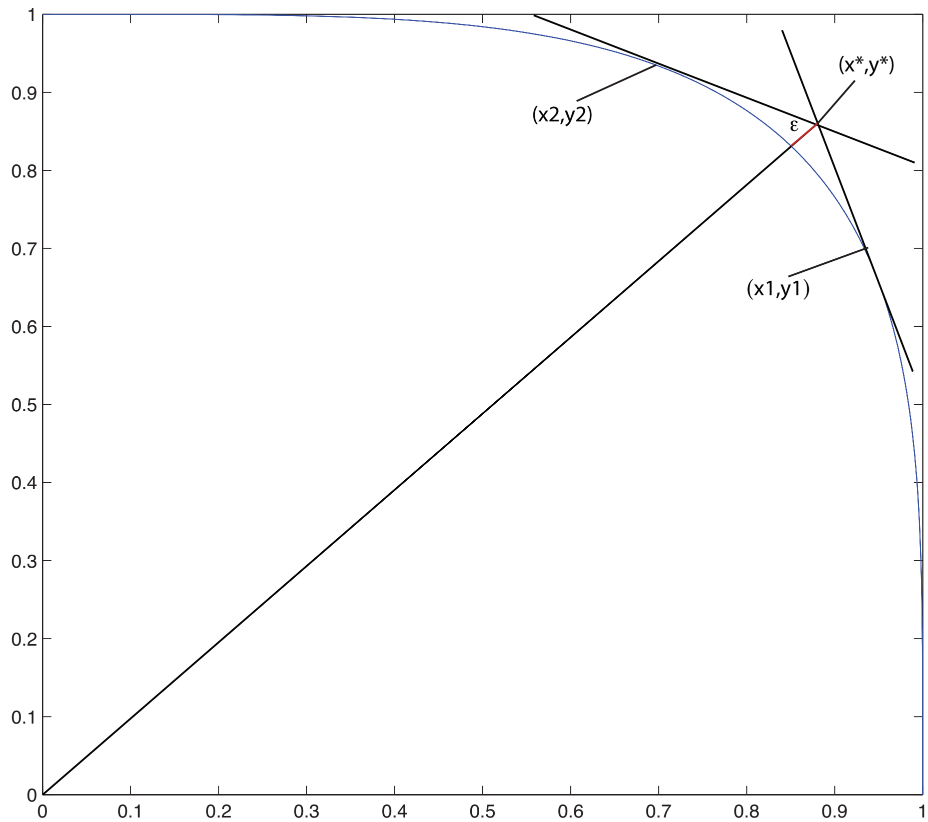

Figure 2.4: Gradient Approximation (graphical representation of ε).

correspond to the points on the p-curve at the polar angles θ_i and θ_{i+1} respectively. The equation of the tangent line to the p-curve $x^p + y^p = 1$ at the point (x_1, y_1) and (x_2, y_2) is given by $x(x_1)^{p1} + y(y_1)^{p-1} = 1$ and $x(x_2)^{p-1} + y(y_2)^{p-1} = 1$ respectively. The set of simultaneous equations:

$$\begin{aligned} x(x_1)^{p-1} + y(y_1)^{p-1} &= 1 \\ x(x_2)^{p-1} + y(y_2)^{p-1} &= 1 \end{aligned} \tag{2.45}$$

can be solved to obtain the intersection, (x^*, y^*), of these two tangent lines:

$$x^* = \frac{y_2^{p-1} - y_1^{p-1}}{(x_1y_2)^{p-1} - (x_2y_1)^{p-1}}$$
$$y^* = \frac{x_1^{p-1} - x_2^{p-1}}{(x_1y_2)^{p-1} - (x_2y_1)^{p-1}} \tag{2.46}$$

Thus, the vertex of $\mathcal{H}'$ located within the segment $[\theta_i, \theta_{i+1}]$ is determined by the solution to (2.45) and has the form (2.46).

It can be seen from geometrical considerations that the approximation error ε_i at a segment $[\theta_i, \theta_{i+1}]$ is determined by the local curvature at this segment, and the curvature of the p-curve is monotonic on $\left[0, \frac{\pi}{4}\right]$. Thus, the "local" approximation errors ε_i will be largest at either the segment $[\theta_0, \theta_1] = \left[0, \frac{\pi}{2m}\right]$ or $[\theta_{m/2-1}, \theta_{m/2}] = \left[\frac{\pi}{4} - \frac{\pi}{2m}, \frac{\pi}{4}\right]$, where it can be assumed without loss of generality that m is an even number.

Let us first consider the local approximation error ε_0 at the segment $\left[0, \frac{\pi}{2m}\right]$. Denoting $\frac{\pi}{2m} = t$, $t \ll 1$, we consider the points $(x_1, y_1) = (1, 0)$ and $(x_2, y_2) = (x_2(t), y_2(t))$. Then, the vertex (x^*, y^*) of the polygon $\mathcal{H}'$ within the segment $[\theta_0, \theta_1] = \left[0, \frac{\pi}{2m}\right]$ is given by

$$\begin{aligned} x^*(t) &= 1 \\ y^*(t) &= \cot^{p-1} t \left(\left(1 + \tan^p t\right)^{\frac{p-1}{p}}\right) - 1\right). \end{aligned}$$

The asymptotic expression for $y^*(t)$ at small $t \ll 1$ can be obtained as follows

$$\begin{aligned} y^*(t) &\approx t^{1-p}\left((1+t^p)^{\frac{p-1}{p}} - 1\right) \\ &\approx t^{1-p}\left(1 + \left(\frac{p-1}{p}\right)t^p - 1\right) \\ &= t^{1-p}\left(\frac{p-1}{p}t^p\right) \\ &= \frac{p-1}{p}t \end{aligned}$$

Then, the approximation error ε_0 at the segment $[\theta_0, \theta_1]$

$$\varepsilon_0 = \|(x^*(t), y^*(t))\|_p - 1$$

can be asymptotically estimated as

$$\begin{aligned}\varepsilon_0 &= (1 + (y^*)^p)^{1/p} - 1 \approx 1 + \frac{1}{p}(y^*)^p - 1\\ &= \frac{1}{p}(y^*)^p\\ &\approx \frac{1}{p}\left(\frac{p-1}{p}t\right)^p\end{aligned}$$

so that, finally,

$$\varepsilon_0 \approx \frac{1}{p}\left(1 - \frac{1}{p}\right)^p \left(\frac{\pi}{2m}\right)^p = O(m^{-p}) \quad \text{for } \ m \gg 1.$$

The approximation error $\varepsilon_{m/2}$ at the segment $[\theta_{m/2-1}, \theta_{m/2}] = \left[\frac{\pi}{4} - \frac{\pi}{2m}, \frac{\pi}{4}\right]$ can be obtained in a similar manner as

$$\varepsilon_{m/2} \approx \frac{p-1}{8}\left(\frac{\pi}{2m}\right)^2 = O(m^{-2}), \quad \text{for } \ m \gg 1.$$

Thus, the uniform gradient approximation (2.39)–(2.41) satisfies (H2) with the approximation accuracy given by

$$\varepsilon \approx \begin{cases} \frac{p-1}{8}\left(\frac{\pi}{2m}\right)^2, & \text{for } p \geq 2\\ \\ \frac{1}{p}\left(1 - \frac{1}{p}\right)^p \left(\frac{\pi}{2m}\right)^p, & \text{for } 1 < p < 2\end{cases} \tag{2.47}$$

$\square$

The gradient polyhedral approximation (2.39)–(2.41) of the p-cone $\mathcal{K}_p^{(3)}$ requires much larger number of facets than Ben-Tal and Nemirovski's lifted polyhedral approximation (2.32) of of the quadratic cone $\mathcal{K}_2^{(3)}$ to achieve the same level of accuracy.

However, as it is shown in Chapter 3, the special structure of the uniform gradient approximation (2.39)–(2.41) makes the corresponding LP approximations of the pOCP problem (2.1) particularly amenable to a cutting plane decomposition algorithm.

CHAPTER 3
A CUTTING PLANE ALGORITHM FOR POLYHEDRAL APPROXIMATIONS OF P-ORDER CONIC PROGRAMMING PROBLEMS

3.1 Introduction

In Chapter 2 we have presented the rationale for solving the p-order conic programming problems

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

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

$$\left\| \mathbf{D}^{(k)}\mathbf{x} - \mathbf{f}^{(k)} \right\|_{p_k} \leq \mathbf{h}^{(k)\top}\mathbf{x} - g^{(k)}, \quad k = 1, \ldots, K \tag{3.1c}$$

by constructing their polyhedral approximations, effectively reducing the pOCP problem (3.1) to a LP of the form

$$\min \left\{ \mathbf{c}^\top \mathbf{x} \;\middle|\; \mathbf{A}\mathbf{x} \leq \mathbf{b}, \;\; \mathbf{H}^{(J_k+1)}_{p_k, m_k} \begin{pmatrix} \mathbf{D}^{(k)}\mathbf{x} - \mathbf{f}^{(k)} \\ \mathbf{h}^{(k)\top}\mathbf{x} - g^{(k)} \\ \mathbf{u}^{(k)} \end{pmatrix} \geq \mathbf{0}, \quad k = 1, \ldots, K \right\} \tag{3.2}$$

where the set

$$\mathbf{H}^{(J_k+1)}_{p_k, m_k} \begin{pmatrix} \boldsymbol{\xi}^{(k)} \\ \mathbf{u}^{(k)} \end{pmatrix} \geq \mathbf{0} \tag{3.3}$$

defines a polyhedral approximation of the p-cone $\mathcal{K}^{(J_k+1)}_{p_k} \subset \mathbb{R}^{J_k+1}_+$. A key strategy that allows one to avoid polyhedral approximations of (3.1) that are exponentially large in the dimensionalities (J_k+1) of the p_k-conic constraints in (3.1) is the "tower-of-variables" technique, which represents a (J_k+1)-dimensional p-cone as intersection of $J_k - 1$ three-dimensional p-cones. (Alternatively, in the case when all p_k are rational: $p_k = r_k/s_k$, one can use the SOCP reformulation of the pOCP problem (3.2) that will require $O(J_k \log r_k)$ three-dimensional second order cones to represent each $(J_k + 1)$-dimensional p_k-cone. We will compare the performances of both these methods in Chapter 4.)

Nevertheless, this approach generally leads to a linear programming problem (3.2) whose size is much larger than the size of the original pOCP problem (3.1); on the other hand, the constructed LP approximation possesses a special structure induced by the "tower-of-variables" transformation, which can be exploited to construct an efficient solution procedure. Next we present a cutting-plane formulation of the approximating problem (3.2) for the case when the approximation (3.3) of $(J_k + 1)$-dimensional p_k-cones has been constructed using the "tower-of-variables" approach.

3.2 A Cutting Plane Formulation for Polyhedral Approximations of pOCP Problems

For simplicity, we assume from now on that the pOCP problem (3.1) contains a single p-order conic constraint ($K = 1$) of dimension $2^d + 1$, $d \in \mathbb{Z}_+$. Evidently, the presented approach and obtained results are generalizable to a pOCP with K p_k-order conic constraints of general dimensions $(J_k + 1)$ in a straightforward way. Invoking the "tower-of-variables" transformation, we reformulate the pOCP problem

(3.1) as follows:

$$\min \quad \mathbf{c}^\top \mathbf{x} \tag{3.4a}$$

$$\text{s. t.} \quad \mathbf{A}\mathbf{x} \leq \mathbf{b} \tag{3.4b}$$

$$w_j^{(\ell)} \geq \left\| \left(w_{2j-1}^{(\ell-1)}, w_{2j}^{(\ell-1)} \right) \right\|_p \qquad j = 1, \ldots, 2^{d-\ell}, \quad \ell = 1, \ldots, d \tag{3.4c}$$

$$w_1^{(d)} \leq \mathbf{h}^\top \mathbf{x} - g \tag{3.4d}$$

$$w_j^{(0)} \geq \left| (\mathbf{D}\mathbf{x} - \mathbf{f})_j \right|, \qquad j = 1, \ldots, 2^d \tag{3.4e}$$

$$w_j^{(\ell)} \geq 0, \qquad j = 1, \ldots, 2^{d-\ell}, \quad \ell = 0, \ldots, d \tag{3.4f}$$

Problem (3.4c) is an equivalent reformulation of the original pOCP problem (3.1) using $2^d - 1$ three-dimensional p-order conic constraints. In accordance with our polyhedral approximation solution approach, we replace each of the 3-dimensional p-cones

$$w_j^{(\ell)} \geq \left\| \left(w_{2j-1}^{(\ell-1)}, w_{2j}^{(\ell-1)} \right) \right\|_p$$

with its polyhedral approximation:

$$\mathbf{H}_{p,m}^{(3)} \begin{pmatrix} \mathbf{w}_j^{(\ell)} \\ \mathbf{u}_j^{(\ell)} \end{pmatrix} \geq \mathbf{0} \tag{3.5}$$

where for given ℓ and j we denote $\mathbf{w}_j^{(\ell)} = \left(w_j^{(\ell)}, w_{2j-1}^{(\ell-1)}, w_{2j}^{(\ell-1)} \right)^\top$, and m is the parameter of construction controlling the approximation accuracy. In such a way, the LP approximation (3.2) of the pOCP problem (3.1) that is based on the "tower-of-variables" transformation and the polyhedral approximations (3.5) of three-dimensional p-cones can now be stated as follows:

$$
\begin{aligned}
\min \quad & \mathbf{c}^\top \mathbf{x} && && \text{(3.6a)}\\
\text{s. t.} \quad & \mathbf{A}\mathbf{x} \le \mathbf{b} && && \text{(3.6b)}\\
& \mathbf{H}^{(3)}_{p,m} \begin{pmatrix} \mathbf{w}^{(\ell)}_j \\ \mathbf{u}^{(\ell)}_j \end{pmatrix} \ge \mathbf{0}, && \ell = 1, \dots, d \quad j = 1, \dots, 2^{d-\ell} && \text{(3.6c)}\\
& w^{(d)}_1 \le \mathbf{h}^\top \mathbf{x} - g && && \text{(3.6d)}\\
& w^{(0)}_j \ge \left| (\mathbf{D}\mathbf{x} - \mathbf{f})_j \right|, && j = 1, \cdots, 2^d && \text{(3.6e)}\\
& w^{(\ell)}_j \ge 0, && j = 1, \dots, 2^{d-\ell}, \;\; \ell = 0, \dots, d && \text{(3.6f)}
\end{aligned}
$$

The linear programming problem (3.6) has a large number of constraints induced by the polyhedral approximation (3.5). However, only a few of these constraints will be binding at optimality. This, in turn, potentially allows one to solve the linear programming problem (3.6) iteratively, by generating only those linear constraints that comprise (3.6c) that are "necessary" to achieve optimality.

To this end, we want to construct a "cutting plane" reformulation of problem (3.6) that would be amenable to iterative generation of linear constraints using a Bender's-type approach. Therefore, we replace the constraint (3.6c) with the following set:

$$
\begin{aligned}
w^{(\ell)}_j \ge \quad \min \quad & v_3 && \text{(3.7a)}\\
\text{s. t.} \quad & \mathbf{H}^{(3)}_{p,m} \begin{pmatrix} \mathbf{v} \\ \mathbf{y} \end{pmatrix} \ge \mathbf{0} && \text{(3.7b)}\\
& v_1 \ge w^{(\ell-1)}_{2j-1} && \text{(3.7c)}\\
& v_2 \ge w^{(\ell-1)}_{2j} && \text{(3.7d)}\\
& \mathbf{v}, \mathbf{y} \ge \mathbf{0} && \text{(3.7e)}
\end{aligned}
$$

where we suppress the indices ℓ and j of the auxiliary variables $\mathbf{v}$ and $\mathbf{y}$ in order to unclutter the notation. The following simple fact is important for further developments:

Proposition 5. *The linear programming problem in* (3.7) *is always feasible for any* $w_{2j-1}^{(\ell-1)}, w_{2j}^{(\ell-1)} \geq 0$.

Proof. For a given $w_{2j-1}^{(\ell-1)}, w_{2j}^{(\ell-1)}$, consider $\mathbf{v} = \left(w^*, w_{2j-1}^{(\ell-1)}, w_{2j}^{(\ell-1)}\right)$ where $w^* = \left\|\left(w_{2j-1}^{(\ell-1)}, w_{2j}^{(\ell-1)}\right)\right\|_p$. Since, obviously, $\mathbf{v} \in \mathcal{K}_p^{(3)}$, by property (H1) there exists $\mathbf{y} \geq 0$ such that $\mathbf{H}_{p,m}^{(3)} \begin{pmatrix} \mathbf{v} \\ \mathbf{y} \end{pmatrix} \geq 0$. Therefore, the linear programming problem given by (3.7) is always feasible for any $w_{2j-1}^{(\ell-1)}, w_{2j}^{(\ell-1)}$. □

Next we demonstrate that the set defined by (3.7) is a polyhedral approximation of the 3D p-cone $w_j^{(\ell)} \geq \left\|\left(w_{2j-1}^{(\ell-1)}, w_{2j}^{(\ell-1)}\right)\right\|_p$ featuring the same approximation accuracy as the set (3.6c). Namely, we have the following proposition:

Proposition 6. *The set given by* (3.7) *in the space of variables*

$$\mathbf{w}_j^{(\ell)} = \left(w_j^{(\ell)}, w_{2j-1}^{(\ell-1)}, w_{2j}^{(\ell-1)}\right)$$

is a polyhedral approximation of the p-cone $w_j^{(\ell)} \geq \left\|\left(w_{2j-1}^{(\ell-1)}, w_{2j}^{(\ell-1)}\right)\right\|_p$ *with accuracy* ε_m*, where* ε_m *is the accuracy of the the approximation* (3.6c) *(i.e. (H1) and (H2) are satisfied with accuracy* ε_m*).*

Proof. In order to show that (H1) is satisfied, consider some $\mathbf{w}_j^{*(\ell)} \in \mathcal{K}_p^{(3)}$, where

$$\mathbf{w}_j^{*(\ell)} = \left(w_j^{*(\ell)}, w_{2j-1}^{*(\ell-1)}, w_{2j}^{*(\ell-1)}\right)^\top.$$

We need to show that such a $\mathbf{w}_j^{*(\ell)}$ also satisfies (3.7). From the properties of the polyhedral approximation $\mathcal{H}_{p,m}^{(3)}$ of the p-cone $\mathcal{K}_p^{(3)}$, there exists $\mathbf{u}^*$ such that

$\mathbf{H}_{p,m}^{(3)} \begin{pmatrix} \mathbf{w}_j^{*(\ell)} \\ \mathbf{u}^* \end{pmatrix} \geq \mathbf{0}$. This implies that $\left(\mathbf{w}_j^{*(\ell)}, \mathbf{u}^*\right)$ is a feasible solution of the linear programming problem in (3.7). Thus, we have $w_j^{*(\ell)} \geq v_3^*$ where $(\mathbf{v}^*, \mathbf{y}^*)$ is an optimal solution of the linear programming problem in (3.7). This implies that $\mathbf{w}_j^{*(\ell)} = \left(w_j^{*(\ell)}, w_{2j-1}^{*(\ell-1)}, w_{2j}^{*(\ell-1)}\right)^\top$ belongs to the set defined by (3.7).

In order to show that (H2) is satisfied, consider any $\mathbf{w}_j^{*(\ell)}$ that satisfies (3.7). Let $(\mathbf{v}^*, \mathbf{y}^*)$ be the corresponding optimal solution of the linear programming problem (3.7). By property (H2) of the approximation $\mathcal{H}_{p,m}^{(3)}$ of $\mathcal{K}_p^{(3)}$ we have that $\|(v_1^*, v_2^*)\|_p \leq (1 + \varepsilon_m) v_3^*$ and

$$\left\| \left(w_{2j-1}^{*(\ell-1)}, w_{2j}^{*(\ell-1)} \right) \right\|_p \leq \|(v_1^*, v_2^*)\|_p \leq v_3^*(1 + \varepsilon_m) \leq w_j^{(\ell)}(1 + \varepsilon_m)$$

Therefore, we have that $w_j^{*(\ell)}(1 + \varepsilon_m) \geq \left\| \left(w_{2j-1}^{*(\ell-1)}, w_{2j}^{*(\ell-1)} \right) \right\|_p$ which means that we have found ε_m so that (H2) is satisfied. $\square$

To develop a cutting-plane representation of the approximating problem (3.6), let us now rewrite the LP in (3.7) as follows:

$$\min \quad v \tag{3.8a}$$

$$\text{s. t.} \quad \widetilde{\mathbf{H}}_m \begin{pmatrix} v \\ \mathbf{y} \end{pmatrix} \geq \begin{pmatrix} \mathbf{0} \\ w_{2j-1}^{(\ell-1)} \\ w_{2j}^{(\ell-1)} \end{pmatrix} \tag{3.8b}$$

$$v \geq 0, \ \mathbf{y} \geq \mathbf{0} \tag{3.8c}$$

where $\widetilde{\mathbf{H}}_m$ is reformulated so that $\mathbf{y}$ includes the variables v_1 and v_2. Then, the dual of problem (3.8) can be written as follows:

$$\begin{aligned} \max \quad & \left(\mathbf{0}, w_{2j-1}^{(\ell-1)}, w_{2j}^{(\ell-1)}\right) \boldsymbol{\pi} && \text{(3.9a)} \\ \text{s. t.} \quad & \widetilde{\mathbf{H}}_m^\top \boldsymbol{\pi} \leq \begin{pmatrix} 1 \\ \mathbf{0} \end{pmatrix} && \text{(3.9b)} \\ & \boldsymbol{\pi} \geq \mathbf{0} && \text{(3.9c)} \end{aligned}$$

The feasible set of the dual problem in (3.9) can be represented as a sum of the convex hull of its vertices $\hat{\boldsymbol{\pi}}_i$ and a (convex) cone generated by its extreme rays, or directions $\bar{\boldsymbol{\pi}}_k$ (see, for instance, Prékopa, 1995). Namely, any $\boldsymbol{\pi}$ that is feasible to (3.9) we can write as

$$\boldsymbol{\pi} = \sum_{i \in \mathscr{P}_m} \lambda_i \hat{\boldsymbol{\pi}}_i + \sum_{k \in \mathscr{Q}_m} \mu_k \bar{\boldsymbol{\pi}}_k, \quad \sum_{i \in \mathscr{P}_m} \lambda_i = 1, \quad \lambda_i \geq 0, \quad \text{and } \mu_k \geq 0 \tag{3.10}$$

where $\mathscr{P}_m$ is the set of extreme points of the feasible set

$$\left\{ \boldsymbol{\pi} \;\middle|\; \widetilde{\mathbf{H}}_m^\top \boldsymbol{\pi} \leq \begin{pmatrix} 1 \\ \mathbf{0} \end{pmatrix}, \boldsymbol{\pi} \geq \mathbf{0} \right\} \tag{3.11}$$

of the dual problem (3.9) and $\mathscr{Q}_m$ is the set of its extreme rays.

Proposition 7. *Observe that any extreme ray $\bar{\boldsymbol{\pi}}_k$, $k \in \mathscr{Q}_m$ of the set* (3.11) *must satisfy*

$$\left(\mathbf{0}, w_{2j-1}^{(\ell-1)}, w_{2j}^{(\ell-1)}\right) \bar{\boldsymbol{\pi}}_k \leq 0$$

for any $w_{2j-1}^{(\ell-1)}, w_{2j}^{(\ell-1)} \geq 0$.

Proof. To see this, assume that the contrary holds for some $k \in \mathscr{Q}_m$ and select the corresponding μ_k very large. As $\mu_k \to \infty$ the dual problem (3.9) becomes unbounded.

With this, the primal problem would become infeasible. However, by Proposition 5 we know that the primal problem is feasible for any non-negative $w_{2j-1}^{(\ell-1)}$ and $w_{2j}^{(\ell-1)}$. $\square$

Because of this observation, we can now replace (3.7) with the following:

$$w_j^{(\ell)} \geq \left(\mathbf{0}, w_{2j-1}^{(\ell-1)}, w_{2j}^{(\ell-1)}\right) \hat{\boldsymbol{\pi}}_i, \quad i \in \mathscr{P}_m \tag{3.12}$$

since the maximum of the dual problem will be achieved at some vertex $\hat{\boldsymbol{\pi}}_i$ of its feasible set (3.11). This leads to the following "cutting plane" formulation of the LP approximation to the pOCP problem (3.1):

$$\min \quad \mathbf{c}^\top \mathbf{x} \tag{3.13a}$$

$$\text{s. t.} \quad \mathbf{A}\mathbf{x} \leq \mathbf{b} \tag{3.13b}$$

$$w_j^{(\ell)} \geq \left(\mathbf{0}, w_{2j-1}^{(\ell-1)}, w_{2j}^{(\ell-1)}\right) \hat{\boldsymbol{\pi}}_i, \quad i \in \mathscr{P}_m,\ j = 1, \ldots, 2^{d-\ell},\ \ell = 1, \ldots, d \tag{3.13c}$$

$$w_1^{(d)} \leq \mathbf{h}^\top \mathbf{x} - g \tag{3.13d}$$

$$w_j^{(0)} \geq \left| (\mathbf{D}\mathbf{x} - \mathbf{f})_j \right|, \qquad j = 1, \ldots, 2^d \tag{3.13e}$$

$$w_j^{(\ell)} \geq 0, \qquad j = 1, \ldots, 2^{d-\ell},\ \ \ell = 0, \ldots, d \tag{3.13f}$$

where constraints (3.13c) are amenable to iterative generation and thus can be viewed as "cutting planes" that "cut off" the portions of the feasible region where an optimal solution cannot be achieved.

In the next section we present the corresponding cutting plane algorithm for solving problem (3.13) iteratively.

3.3 A Cutting Plane Algorithm for Polyhedral Approximations of p-Order Conic Programming Problems

Assume without loss of generality that problem (3.13) is bounded, and consider the master problem corresponding to the cutting plane formulation (3.13):

$$
\begin{aligned}
\min \quad & \mathbf{c}^\top \mathbf{x} && \text{(3.14a)}\\
\text{s. t.} \quad & \mathbf{A}\mathbf{x} \leq \mathbf{b} && \text{(3.14b)}\\
& w_j^{(\ell)} \geq \sigma_{i,\nu-1} w_{2j-1}^{(\ell-1)} + \tau_{i,\nu} w_{2j}^{(\ell-1)}, \qquad i = 1,\dots,C_j^{(\ell)}, && \\
& \qquad\qquad j = 1,\dots,2^{d-\ell},\ \ell = 1,\dots,d && \text{(3.14c)}\\
& w_1^{(d)} \leq \mathbf{h}^\top \mathbf{x} - g && \text{(3.14d)}\\
& w_j^{(0)} \geq \left| (\mathbf{D}\mathbf{x} - \mathbf{f})_j \right|, \qquad j = 1,\dots,2^d && \text{(3.14e)}\\
& w_j^{(\ell)} \geq 0, \qquad j = 1,\dots,2^{d-\ell},\ \ \ell = 0,\dots,d && \text{(3.14f)}
\end{aligned}
$$

where $\sigma_{i,\nu-1}$ and $\tau_{i,\nu}$ stand for the last two components $\hat{\pi}_{\nu-1}$ and $\hat{\pi}_\nu$ of the vector $\hat{\boldsymbol{\pi}}_i \in \mathbb{R}^\nu$. Let $(\mathbf{x}^*, \mathbf{w}^*)$ be an optimal solution to the master problem (3.14) after a given iteration (note that if (3.14) is infeasible, then (3.13) is infeasible too, and the procedure stops). Then, for any $\ell = 1,\dots,d$, $j = 1,\dots,2^{d-\ell}$ solve the subproblem:

$$
\begin{aligned}
\max \quad & \left(\mathbf{0}, w_{2j-1}^{*(\ell-1)}, w_{2j}^{*(\ell-1)}\right) \boldsymbol{\pi} && \text{(3.15a)}\\
\text{s. t.} \quad & \widetilde{\mathbf{H}}_m^\top \boldsymbol{\pi} \leq \begin{pmatrix} 1 \\ \mathbf{0} \end{pmatrix} && \text{(3.15b)}\\
& \boldsymbol{\pi} \geq \mathbf{0} && \text{(3.15c)}
\end{aligned}
$$

and, given its optimal solution $\boldsymbol{\pi}^* = \boldsymbol{\pi}_j^{*(\ell)}$ check if the condition

$$
w_j^{*(\ell)} \geq \left(\mathbf{0}, w_{2j-1}^{(\ell-1)*}, w_{2j}^{(\ell-1)*}\right) \boldsymbol{\pi}_j^{*(\ell)} \tag{3.16}
$$

is satisfied. If condition (3.16) is violated for some ℓ, j then we add a new constraint (3.14c) for the variable $w_j^{(\ell)}$ by incrementing the corresponding counter of constraints in (3.14c): $C_j^{(\ell)} = C_j^{(\ell)} + 1$, and setting

$$\sigma_{j,i'}^{(\ell)} = \pi_{j,\nu-1}^{*(\ell)}, \quad \tau_{j,i'}^{(\ell)} = \pi_{j,\nu}^{*(\ell)} \quad \text{for} \quad i' = C_j^{(\ell)} \tag{3.17}$$

After checking condition (3.16) for all variables $w_j^{(\ell)}$, the master problem (3.14) is augmented with new constraints and is solved again. If (3.16) holds for all variables $w_j^{(\ell)}$, and thus no new cuts are generated during the given iteration, the current solution $\mathbf{x}^*, \mathbf{w}^*$ of the master problem is optimal for the original LP approximation problem (3.13). In such a way, the proposed cutting-plane procedure obtains an optimal solution, if it exists, of the original LP approximation problem (3.13) after a finite number of iterations with, perhaps, some anticycling scheme employed.

A starting solution for the cut generation procedure can be constructed, for example, by solving the master problem (3.14) with constraints

$$w_j^{(\ell)} \geq w_{2j-1}^{(\ell-1)}, \quad w_j^{(\ell)} \geq w_{2j}^{(\ell-1)}, \quad j = 1, \ldots, 2^{d-\ell}, \ \ell = 1, \ldots, d,$$

in place of constraints (3.14c). Indeed, note that inequality

$$w_j^{(\ell)} \geq \left\| \left(w_{2j-1}^{(\ell-1)}, w_{2j}^{(\ell-1)} \right) \right\|_p$$

implies

$$w_j^{(\ell)} \geq \max \left\{ w_{2j-1}^{(\ell-1)}, w_{2j}^{(\ell-1)} \right\}.$$

More efficient methods of generating an initial solution can be suggested by exploiting the particular structure of the feasible region of the pOCP problem (3.1).

3.4 Cut Generation Efficiencies

Effectiveness of the described cutting-plane scheme depends, in part, on how fast the set of cuts (3.14c) can be updated. Here we demonstrate that the gradient approximation (2.39)–(2.40), introduced in Chapter 2, admits quite an efficient generation of cuts. Indeed, when the gradient polyhedral approximation (2.39)–(2.40) is used in (3.5), problem (3.9) takes the form

$$\begin{aligned}
\max \quad & w_{2j-1}^{*(\ell-1)} \pi_{m+1} + w_{2j}^{*(\ell-1)} \pi_{m+2} \\
\text{s. t.} \quad & \sum_{i=0}^{m} \alpha_i^{(p)} \pi_i \geq \pi_{m+1} \\
& \sum_{i=0}^{m} \beta_i^{(p)} \pi_i \geq \pi_{m+2} \\
& \sum_{i=0}^{m} \pi_i \leq 1 \\
& \pi_i \geq 0, \quad i = 0, \ldots, m+2
\end{aligned} \tag{3.18}$$

which is the dual of the problem obtained from (3.7) by using the gradient approximation (2.39)–(2.40)

$$\begin{aligned}
\min \quad & u_3 \\
\text{s. t.} \quad & u_3 \geq \alpha_i^{(p)} u_1 + \beta_i^{(p)} u_2, \quad i = 0, \ldots, m, \\
& u_1 \geq w_{2j-1}^{*(\ell-1)}, \\
& u_2 \geq w_{2j}^{*(\ell-1)}, \\
& u_1, u_2, u_3 \geq 0.
\end{aligned} \tag{3.19}$$

Clearly, an optimal solution of problem (3.18) is given by

$$\pi_{m+1}^* = \alpha_{i^*}^{(p)}, \;\; \pi_{m+2}^* = \beta_{i^*}^{(p)}, \quad \text{and} \quad \pi_i^* = \begin{cases} 1, & i = i^*, \\ 0, & i \in \{0, \ldots, m\} \backslash i^*, \end{cases} \tag{3.20a}$$

where the index i^* is such that

$$i^* \in \operatorname*{arg\,max}_{i=0,\dots,m} \Big\{ \alpha_i^{(p)} w_{2j-1}^{*(\ell-1)} + \beta_i^{(p)} w_{2j}^{*(\ell-1)} \Big\}. \tag{3.20b}$$

In other words, the cut-generating problem (3.15) reduces to selection of a maximum element in a set of $m+1$ numbers, and therefore can be solved in linear $O(m)$ time. However, as we show next, the special structure contained in the gradient approximation (2.39)–(2.40) and, correspondingly, in problem (3.15), allows for a more efficient solution.

Proposition 8. *Consider the pOCP problem* (3.1) *with K conic constraints of dimension J_k+1 and order $p_k \in (1,\infty)$. Assume that each conic constraint is approximated using the "tower-of-variables" approach and the gradient polyhedral approximation* (2.39)–(2.40) *with parameter of approximation m. Then, during an iteration of the decomposition scheme described above, new cuts can be generated in $O\big(\sum_k J_k \log m\big)$ time. If the "uniform" polyhedral approximation* (2.41) *is used, the cuts can be generated in a constant $O\big(\sum_k J_k\big)$ time.*

Proof. To prove the first statement of the proposition, we consider the sequence

$$\gamma_i = \xi_1^* \, \alpha_i^{(p)} + \xi_2^* \, \beta_i^{(p)}, \quad i = 0,\dots,m \tag{3.21}$$

for some non-negative $\xi_1^*, \xi_2^* \geq 0$ such that $\xi_1^* + \xi_2^* > 0$ (the case when both $\xi_1^* = \xi_2^* = 0$ is trivial). Let us call a sequence $\{c_n\}$ *strictly quasiconcave* if it is generated by a continuous strictly quasiconcave function $f(\cdot)$: $c_n = f(t_n)$, where $t_{n-1} < t_n$. An important characteristic of a strictly quasiconcave function is that every local maximum is also its global maximum (see, e.g., Bazaraa et al., 2006), hence every local maximum of a strictly quasiconcave sequence will be its global maximum as well. It is easy to see that $\{\gamma_i\}$ is a strictly quasiconcave sequence. Indeed, using the definition (2.40) of the coefficients $\alpha_i^{(p)}, \beta_i^{(p)}$ as functions of the polar angle θ, the sequence $\gamma_i\,(i = 0, \ldots, m)$ can be viewed as being generated by the function

$$\begin{aligned} f(\theta) &= \xi_1^* \alpha^{(p)}(\theta) + \xi_2^* \beta^{(p)}(\theta) \\ &= \xi_1^* \left(\cos^p\theta + \sin^p\theta\right)^{\frac{1-p}{p}} \cos^{p-1}\theta + \xi_2^* \left(\cos^p\theta + \sin^p\theta\right)^{\frac{1-p}{p}} \sin^{p-1}\theta \end{aligned} \tag{3.22}$$

evaluated at discrete points $0 \equiv \theta_0 < \theta_1 < \ldots < \theta_m \equiv \frac{\pi}{2}$. The derivative of the function $f(\theta)$ is

$$f'(\theta) = (p-1)\frac{\sin^{p-1}\theta\,\cos^{p-1}\theta}{(\cos^p\theta + \sin^p\theta)^{2-1/p}}\left(\frac{-\xi_1^*}{\cos\theta} + \frac{\xi_2^*}{\sin\theta}\right), \quad p > 1. \tag{3.23}$$

Clearly, function $f(\theta)$ is strictly quasiconcave on $[0, \pi/2]$ since it is continuous and is either monotonic on $[0, \pi/2]$ (when one of the parameters ξ_1^*, ξ_2^* is zero) or has a unique global maximum at

$$\theta^* = \arctan(\xi_2^*/\xi_1^*). \tag{3.24}$$

Thus, the function $f(\theta)$ has a unique global maximum (e.g, either θ^*, or 0, or $\frac{\pi}{2}$), which can be found by solving the equation $f'(\theta) = 0$ using binary search.

Similarly, although the maximum of the corresponding sequence $f(\theta_i) = \gamma_i\,(i = 0, \ldots, m)$ may be not unique (i.e., two adjacent elements may have the same maximal value), the largest element(s) in the sequence can be determined using a binary search that requires $O(\log_2 m)$ time. Consequently, generation of new cuts for the polyhedral approximation (2.39)–(2.40) of a p-order conic constraint in $\mathbb{R}^{J_k+1}$ requires solving of $J_k - 1$ instances of problem (3.18), which means that in the case of K p_k-order conic constraints, cut generation for the gradient polyhedral approximation (2.39)–(2.40) can be done in $O\big(\sum_k J_k \log m\big)$ time.

The computational time needed to determine the maximum element(s) of sequence (3.21) can be improved drastically if the points $\theta_i (i = 0, \ldots, m)$ are uniformly spaced on $[0, \pi/2]$: $\theta_i = \frac{\pi i}{m}$. Then, the index i^* of the maximum element(s) of $\gamma_i (i = 0, \ldots, m)$ is determined from

$$i^* \in \arg\max \Big\{ \xi_1^* \alpha_t^{(p)} + \xi_2^* \beta_t^{(p)},\ \xi_1^* \alpha_{t+1}^{(p)} + \xi_2^* \beta_{t+1}^{(p)}, \Big\}, \quad \text{where } \ t = \left\lfloor \frac{2m}{\pi} \arctan \frac{\xi_2^*}{\xi_1^*} \right\rfloor \tag{3.25}$$

Indeed, given that the constants γ_i represent the values of function $f(\theta)$ at equally spaced points $\theta_i = \frac{\pi i}{2m}$, the integer t in (3.25) identifies the segment $[\theta_t, \theta_{t+1}] = \left[\frac{\pi t}{2m}, \frac{\pi(t+1)}{2m}\right]$ that contains the extremum θ^* (3.24) of $f(\theta)$. Hence, the largest element of sequence γ_i is selected among the values of function $f(\theta)$ evaluated at the endpoints of the segment $[\theta_t, \theta_{t+1}]$. Note that (3.21) can have at most two optimal solutions, which corresponds to the case of $g(\theta_t) = g(\theta_{t+1})$.

Thus, a solution of (3.18) can be obtained in a *constant* $O(1)$ time that does not depend on the number of facets m in the uniform gradient polyhedral approximation (2.39)–(2.41). Given that each conic constraint of order p_k and dimensionality $J_k + 1$ requires $J_k - 1$ such operations, generation of new cuts in problem (3.14) that employs a uniform gradient polyhedral approximation (2.39)–(2.41) requires $O\big(\sum_k J_k\big)$ time. □

The significance of Proposition 8 lies in the fact that increasing m, and, correspondingly, the quality of the uniform gradient approximation (2.39)–(2.41) , comes at no cost with regard to the time needed to generate new cuts during an iteration of the decomposition scheme described above. Of course, this does not mean that the number of iterations needed to obtain an optimal solution of (3.14) (if it exists) remains constant with respect to m.

Observe also that when the polyhedral approximation (2.39)–(2.40) is employed, the specific form (3.20) of optimal solutions of the subproblem (3.15) allows one to write the cuts (3.14c) in the master problem (3.14) in the form

$$w_j^{(\ell)} \geq \alpha_i^{(p)} w_{2j-1}^{(\ell-1)} + \beta_i^{(p)} w_{2j}^{(\ell-1)}, \quad i \in \mathcal{C}_j^{(\ell)}, \; j = 1, \ldots, 2^{d-\ell}, \; \ell = 1, \ldots, d. \tag{3.26}$$

Here $\mathcal{C}_j^{(\ell)}$ are subsets of $\{0, \ldots, m\}$ and contain the indices i of cuts that have been generated for the variable $w_j^{(\ell)}$. Knowing the exact values of the coefficients in cuts (3.14c) without having to solve problem (3.15), one can potentially improve the numerical accuracy of the cutting-plane scheme.

CHAPTER 4
CASE STUDIES: PORTFOLIO OPTIMIZATION WITH P-ORDER CONIC CONSTRAINTS

4.1 Introduction

In this chapter we discuss the computational efficiency of the developed algorithmic approaches to solving p-order conic programming (pOCP) problems. In particular, we will be comparing the cutting plane algorithm for polyhedral approximations of pOCP problems developed in Chapter 3 with the method of reformulating pOCP problems as second-order conic programming (SOCP) in the case of rational values of the parameter p presented in Chapter 2. In addition, we will consider the computational efficiency of the "full" LP implementations of the polyhedral approximations of pOCP problems. The computational comparisons will be conducted on an example of a portfolio optimization problem with p-order conic constraints.

Later, we will consider methodological aspects of the coherent and deviation measures that involve higher moments of loss distributions, such as HMCR, SMCR, HMD, and SMD measures (see Chapter 1). We will be looking at how the risk measures based on higher moments perform when compared to each other and to more conventional risk measures such as the CVaR, and the Mean-Variance models. This comparison will be done at a later date after updating the current data set to incorporate the most recent market volatility that was experienced with the current market melt down. In addition, we will consider performance of cardinality-constrained portfolios based on the higher moment risk measures.

4.2 Computational Performance of p-Order Conic Programming Algorithms

In this section we conduct numerical comparisons of the approaches discussed in Chapters 2 and 3 to solving problems of type (2.1) on an example of a portfolio optimization problem with p-order conic constraint.

Tracing back to Markowitz (1952, 1959), portfolio optimization problems are typically stated in the form where portfolio (investment) risk is minimized while requiring a certain level of expected return on the investment, or, alternatively, the portfolio's expected return is maximized subject to a constraint on portfolio risk. Yet another formulation is employed in the literature where a "composite" objective representing a linear combination of risk and reward (e.g., expected return) is optimized (the so-called *mean-risk* models, see, e.g., Ogryczak and Ruszczyński, 1999, 2001, 2002). The particular formulation is usually chosen depending on the preferences of the decision-maker (investor) and the application at hand; Krokhmal et al. (2002a) discuss the conditions at which all three formulations are equivalent.

The portfolio selection models that will be employed in the case study have the general form:

$$\min_{\mathbf{x}} \quad \mathscr{R}(-\mathbf{r}^\top \mathbf{x}) \tag{4.1a}$$

$$\text{s. t.} \quad \mathbf{e}^\top \mathbf{x} = 1, \tag{4.1b}$$

$$\mathsf{E}\mathbf{r}^\top \mathbf{x} \geq r_0, \tag{4.1c}$$

$$\mathbf{x} \geq \mathbf{0}, \tag{4.1d}$$

where $\mathbf{x} = (x_1, \ldots, x_n)^\top$ is the vector of portfolio weights, $\mathbf{r} = (r_1, \ldots, r_n)^\top$ is the random vector of return on portfolio instruments, and $\mathbf{e} = (1, \ldots, 1)^\top$. The risk measure $\mathscr{R}(X)$ in (4.1a) can be taken to be HMCR, SMCR, etc., of the negative portfolio return, $-\mathbf{r}^\top\mathbf{x}$. Constraint (4.1b) is the budget constraint while (4.1d) together with (4.1b) ensure that all of the available funds are invested. Constraint (4.1c) imposes a minimal required level r_0 of expected return of the resulting portfolio.

For the sake of simplicity and in order to conduct a useful comparison of the effects of the risk measure selection in (4.1a) we purposely do not include any additional trading or institutional constraints such as transaction cost, liquidity constraints, etc.. In keeping with what is traditionally done in portfolio optimization problems, the distribution of random returns r_i of asset i is modeled sing a set of J discrete equiprobable scenarios $\{r_{i1}, \cdots, r_{iJ}\}$.

With the risk measure $\mathscr{R}(X)$ replaced with the $\mathrm{HMCR}_{p,\alpha}(X)$ measure, the portfolio selection problem for our case study is transformed into a linear programming problem with a single p-order conic constraint (4.2e)

$$
\begin{aligned}
\min \quad & \eta + \frac{J^{-\frac{1}{p}}}{1-\alpha} t && && (4.2\text{a}) \\
\text{s. t.} \quad & \sum_{i=1}^{n} x_i = 1, && && (4.2\text{b}) \\
& \frac{1}{J}\sum_{j=1}^{J}\sum_{i=1}^{n} r_{ij}x_i \geq r_0, && && (4.2\text{c}) \\
& w_j \geq -\sum_{i=1}^{n} r_{ij}x_i - \eta, && j = 1, \cdots, J, && (4.2\text{d}) \\
& t \geq (w_1^p + \cdots + w_J^p)^{1/p}, && && (4.2\text{e}) \\
& x_i \geq 0 && i = 1, \cdots, n, && (4.2\text{f}) \\
& w_j \geq 0 && j = 1, \cdots, J && (4.2\text{g})
\end{aligned}
$$

Problem (4.2) will be solved through its various polyhedral approximations (e.g., "lifted" Ben-Tal and Nemerovki's approximation, "gradient" approximation) as well as the "exact" SOCP reformulation in the case of rational p.

Remark 2. *In the context of benchmarking the LP approximations of the pOCP portfolio optimization problem* (4.2) *against its SOCP-based implementation in the case of a rational p, the adopted formulation of the portfolio optimization problem* (4.1), (4.2) *has several notable characteristics. Firstly, the conic constraint* (4.2e) *is feasible as long as constraints* (4.2b)–(4.2d) *are feasible; in other words, feasibility of problem* (4.2) *is determined by the budget constraint and constraint on the expected return. Secondly, the rather simple structure of linear constraints* (4.2b)–(4.2d) *that correspond to linear constraints* $\mathbf{Ax} \leq \mathbf{b}$ *of the general pOCP problem* (2.1) *allows for placing more weight on the efficiency of handling of p-order constraints by a particular computational scheme, rather than on solver's efficiency in handling of linear constraints* $\mathbf{Ax} \leq \mathbf{b}$*, in the interpretation of the computational results that follows next.*

Remark 3. *Note that with the addition of the constraint*

$$\eta = \frac{1}{J}\sum_{j=1}^{J}\sum_{i=1}^{n} r_{ij}x_i$$

we can use the construct of problem (4.2) *to implement the SMCR risk measure. This reduces the problem by one variable and as such its effects on the computation time can be considered negligible. Therefore we proceed with the computational results for the HMCR model as shown in* (4.2).

4.2.1 Set of Instrument and Scenario Data

In order to take advantage of the construct of the HMCR risk measures, which quantify risk in terms of higher tail moments of loss distribution, the portfolio optimization case studies were conducted using return data of the fifty S&P500 stocks with the so-called "heavy tails". In order to look at computation time comparisons, for scenario generations we used 10-day historical returns over $J = 2^8, \cdots, 2^{13}$ overlapping periods, calculated using daily closing prices from October 30, 1998 to January 18, 2006. The particular sizes of the scenario set has been chosen to accommodate the linear approximation techniques in problem (3.6), and the sizes of the considered scenario sets were limited only by availability of the data. From this set of S&P500 stocks, we selected $n = 50$ instruments by picking those with the highest value of kurtosis of biweekly returns, calculated over a specific period.

4.3 Computation Time Comparisons

In this section we present the computational efficiency, as measured by the average running time,in seconds, of the developed algorithms for solving pOCP problems on the example of the portfolio optimization problem with p-order conic constraint (4.2). We compare the solution time of the cutting plane algorithm (CPA), the second order conic programming (SOCP) reformulation, the full algorithm implementation (FA) of the "tower of variables" construction, the Ben-Tal–Nemirovski (BN) lifted polyhedral model for $p = 2$ and the gradient approximation (GA) with $p = 2$. For CPA, SOCP, and GA formulations, the value of the parameter p in (4.2) varied as $p = 2, 3, 4, 5$; the implementation based on Ben-Tal and Nemirovski's (BN) approximation applies only to $p = 2$. The confidence level α and minimum required expected return have been fixed at $\alpha = 0.9$ and $r_0 = 0.5\%$ for all algorithms.

A total of 76 instances of problem (4.2) corresponding to 76 bi-weekly rebalancing periods from December, 2002 to January 2006 have been solved for each implementation and each scenario size.

The computer that was used to perform the scenario runs was a Dell XPS with a Dual Core Pentium processor and 2GB of RAM. The machine was running

Windows XP with CPLEX 10.0.0. The ILOG Concert Technology implementation of the CPA algorithm utilized the CPLEX linear programming solver, and the SOCP implementation used CPLEX Barrier solver. The BN and GA implementations were done using AMPL. The accuracy of the polyhedral approximations were chosen at $\varepsilon < 10^{-5}$ that is consistent with the standard CPLEX computation accuracy.

Both the GA and CPA implementations are solving the following linear approximation of the pOCP portfolio optimization problem (4.2) based on the "uniform" gradient approximation developed in Chapter 2:

$$
\begin{aligned}
\min \quad & \eta + \frac{1}{1-\alpha}\frac{t}{\sqrt[p]{J}} && \text{(4.3a)}\\
\text{s. t.} \quad & (4.2\text{b}), (4.2\text{c}) \text{ and } (4.2\text{d}) && \text{(4.3b)}\\
& w_j^{(\ell)} \ge \alpha_i^{(p)} w_{2j-1}^{(\ell-1)} + \beta_i^{(p)} w_{2j}^{(\ell-1)} \qquad i = 0, \ldots, m, & \\
& \qquad \ell = 0, \ldots, d,\ j = 1, \ldots, 2^{d-\ell} && \text{(4.3c)}\\
& w_1^{(d)} = t && \text{(4.3d)}\\
& w_j^{(0)} = w_j, \qquad j = 1, \cdots, J && \text{(4.3e)}\\
& x_i \ge 0 \qquad i = 1, \cdots, n, && \text{(4.3f)}\\
& w_j^{(\ell)} \ge 0 \qquad \ell = 0, \cdots, d, \quad j = 1, \cdots, 2^{d-\ell}. && \text{(4.3g)}
\end{aligned}
$$

where the coefficients $\alpha_i^{(p)}$, $\beta_i^{(p)}$ have the form (2.40).

For our computational comparisons, we use this idea to solve the linear programming problem as either a "full" linear programming problem that contains all of the constraints or through the cutting plane algorithm which generates the necessary constraints for the optimization problem to reach optimality.

Here we would also like to discuss generation of the starting solution to the master problem (3.14) that is solved as a part of the cutting plane implementation of the LP (4.3): note that this problem is unbounded when constraints (3.14c) that in the case of uniform gradient approximation have the form

$$w_j^{(\ell)} \geq \alpha_i^{(p)} w_{2j-1}^{(\ell-1)} + \beta_i^{(p)} w_{2j}^{(\ell-1)}$$

are absent (i.e., when $\mathcal{C}(\ell)_j = \emptyset$ for all variables $w_j^{(\ell)}$). An initial feasible solution to the master problem of (4.3) can be constructed as follows. First, a vector $\mathbf{x}^*$ that satisfies (4.2b) and (4.2d) is selected (this can be done by distributing portfolio weights equally among all instruments whose expected return exceeds r_0), and then value η^* is chosen in such way that there is at least one $w_j^{*(0)} = \max\left\{0, \sum_j r_{ij} x_i^* - \eta^*\right\} > 0$ for both $j \leq J/2$ and $j > J/2$. This last condition ensures that, when cuts for variables $w_j^{(\ell)}$ are generated by solving (3.18), the coefficients $\alpha_i^{(p)}$ and $\beta_i^{(p)}$ in the cut for the "top" variable $w_1^{(d)}$ are both non-zero, which, in turn, guarantees that the resulting master problem will not be unbounded.

Alternatively, one can start with a master problem in which the set of cuts $\mathcal{C}_j^{(\ell)}$ for each variable $w_j^{(\ell)}$ is non-empty: e.g., $\mathcal{C}_j^{(\ell)} = \left\{\lfloor m/2 \rfloor\right\}$. Evidently, in this case the master problem will be bounded and feasible as long as constraints (4.3b) and (4.3c) are feasible. The downside of this method is that one starts with a larger problem as compared to the case described above; on the other hand, it does not require the introduction of new columns into problem (4.3) during iterations. In the context of the present case study, the last method turned out to be more efficient by requiring fewer iterations and reaching an optimal solution of (4.3) faster.

4.3.1 Computational Results for $p = 2$

In this section we compared the numerical efficiency of the approximate and exact methods of solving the portfolio optimization problem (4.2) with second-order conic constraint ($p = 2$). Precisely, we compared the full LP polyhedral approximations of (4.2) based on the Ben-Tal and Nemirovski's lifted approximation (BN), the uniform gradient approximation (GA), the cutting-plane algorithm implementation (CPA) of the GA linear programming problem, and the native SOCP implementation of (4.2) with $p = 2$.

As expected, the worst computational performance was demonstrated by the gradient-based GA polyhedral approximation implemented as a "full" LP problem. The more efficient BN implementation that relies on the lifted Ben-Tal and Nemirovski's approximation clearly outperformed the simpler GA implementation.

The results of the computational studies are represented in Figure 4.1 and in Table 4.1, which report average running times in seconds for the four algorithms (BN, GA, CPA, SOCP) applied to problem (4.2) with $p = 2$.

One of the unexpected results observed during this computational comparison is that the SOCP implementation was outperformed by the polyhedral approximation that employs the cutting plane (CPA) algorithm. This can be attributed to some inefficiencies in the current CPLEX barrier solver. We will further check this particular result by running the CPA implementation against the "native" SOCP implementation using MOSEK's more efficient primal-dual interior point SOCP solver (see Figure 4.2).

We would also like to mention that the solutions to the portfolio optimization problem (4.2) obtained by different algorithms were consistent up to the specified approximation accuracy.

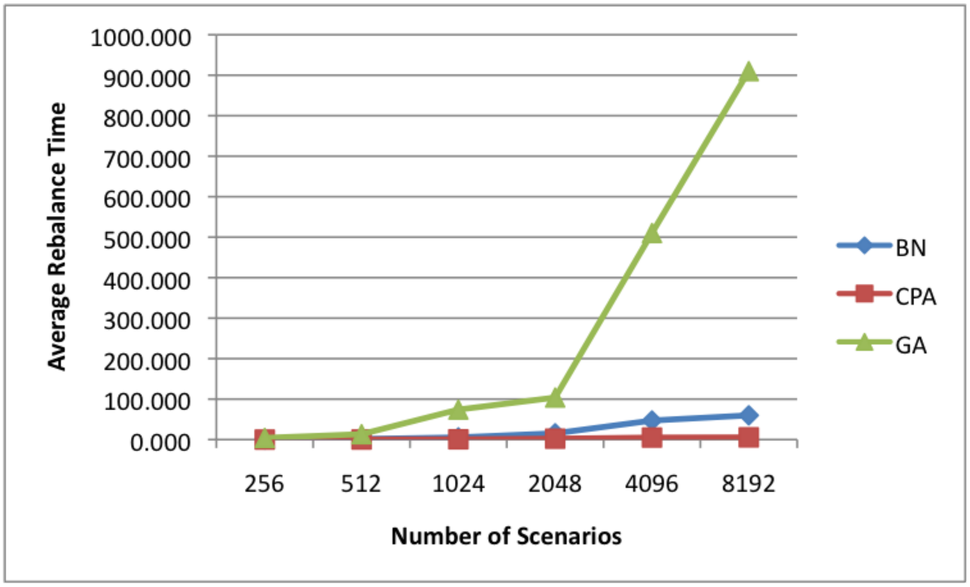

Figure 4.1: Average runing times (in seconds) for the CPA, BN and GA approximations and the SOCP reformulation for $p = 2$.

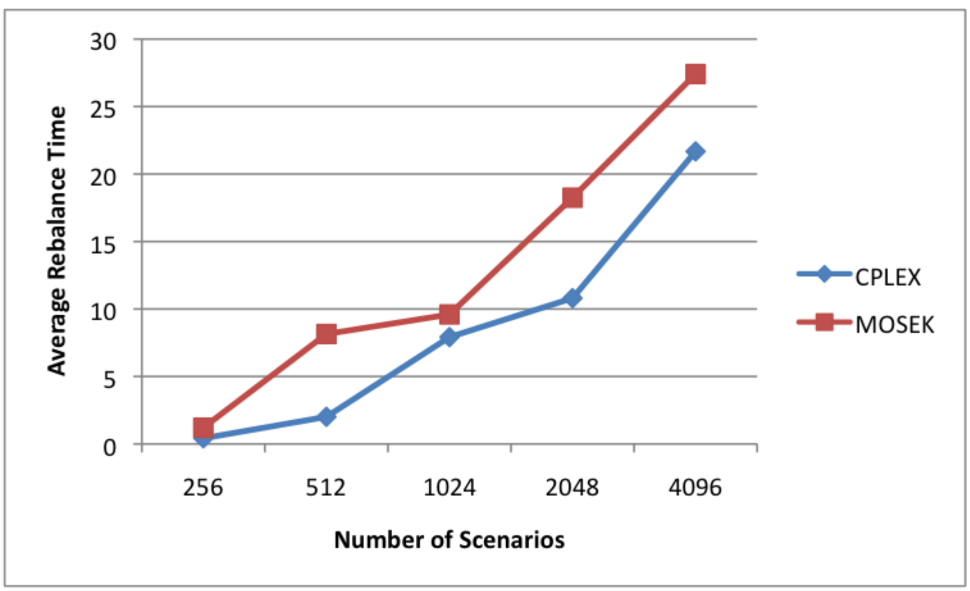

Figure 4.2: Average running times (in seconds) of the cutting plane (CPA) algorithm and the SOCP reformulation using the MOSEK solver.

J	BN	CPA	GA
256	0.438	0.040	4.133
512	1.472	0.156	13.332
1024	5.347	0.671	74.092
2048	15.487	2.231	103.860
4096	47.067	4.953	510.181
8192	59.892	5.469	910.035

Table 4.1: Average running times (in seconds) for the BN, GA, and CPA approximations of problem (4.2) with $p = 2$, $\alpha = 0.9$, and $r_0 = 0.05\%$.

4.3.2 Computational Results for $p \neq 2$

The previous subsection discussed numerical comparisons for the case $p = 2$. Our main interest, however, is in the case $p \neq 2$. Given the results of the $p = 2$ comparisons, it is evident that the most efficient methods (at least in application to the portfolio optimization problem (4.2)) are represented by the cutting plane algorithm (CPA) applied to the uniform polyhedral approximation of (4.2), and its SOCP reformulation. Thus, in this section we discuss the computational performance of these two methods in application to the pOCP problem (4.2) with the values of parameter p varied as $p = 3, 4$, and 5. The SOCP reformulation of the pOCP problem (4.2) was based on the pOCP→SOCP transformation for rational values of p that has been presented in Chapter 2. Other than this, the setup of the numerical experiments has remained the same as in the previous subsection.

The average running times of the CPA and SOCP implementations on 76 instances of the portfolio optimization problem (4.2) are reported in Figures 4.3 – 4.5

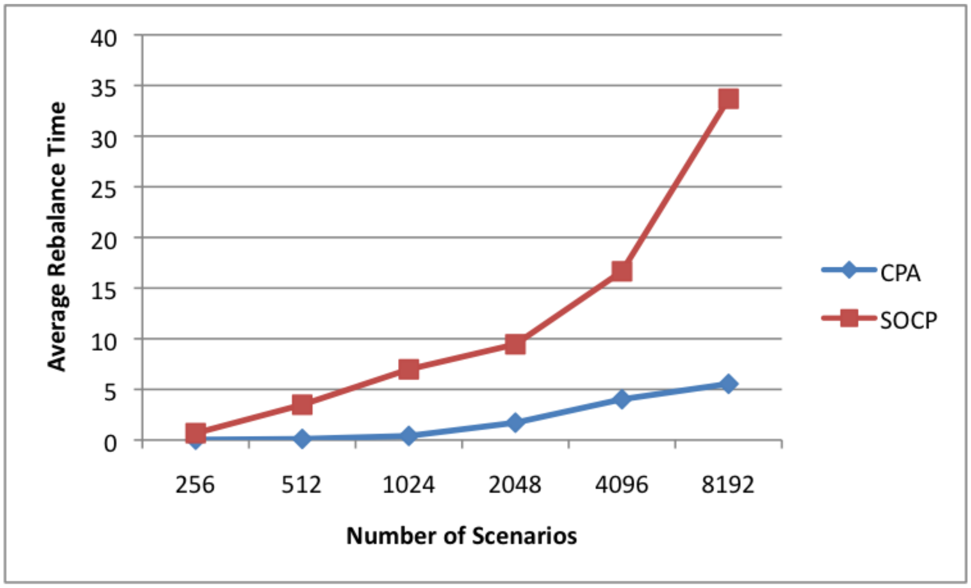

Figure 4.3: Average running times (in seconds) of the cutting plane (CPA) algorithm and the SOCP reformulation for $p = 3$.

and Table 4.2.

The main conclusion of this computational study is that the developed cutting plane algorithm (CPA) as applied to the uniform gradient approximation of pOCP problem consistently outperforms the corresponding interior-point algorithm based on "exact" SOCP reformulation of the problem.

We believe that the following factors can be contributing to the observed differences in efficiencies of the polyhedral approximation/cutting plane procedure and the SOCP-based solution approach. First, the specific structure of the uniform gradient approximations of p-cones developed in Chapter 2 allows for generating the cuts in *constant* time that does not depend on the accuracy of approximation, i.e. the number of facets that are used to approximate each 3D p-cone (see Proposition 8).

This, coupled with the "warm-start" capabilities of the simplex LP solver that

	p=3		p=4		p=5	
J	**CPA**	**SOCP**	**CPA**	**SOCP**	**CPA**	**SOCP**
256	0.034	0.67	0.034	0.636	0.033	0.726
512	0.109	3.485	0.103	3.416	0.103	3.68
1024	0.407	6.979	0.388	7.409	0.375	5.619
2048	1.712	9.45	1.645	9.153	1.582	14.85
4096	4.026	16.655	3.78	16.886	3.593	29.592
8192	5.546	33.692	5.048	40	4.837	77.637

Table 4.2: Average running times (in seconds) for the CPA approximation and SOCP reformulation of problem (4.2) with $p = 3, 4, 5$, and $\alpha = 0.9$, $r_0 = 0.05\%$.

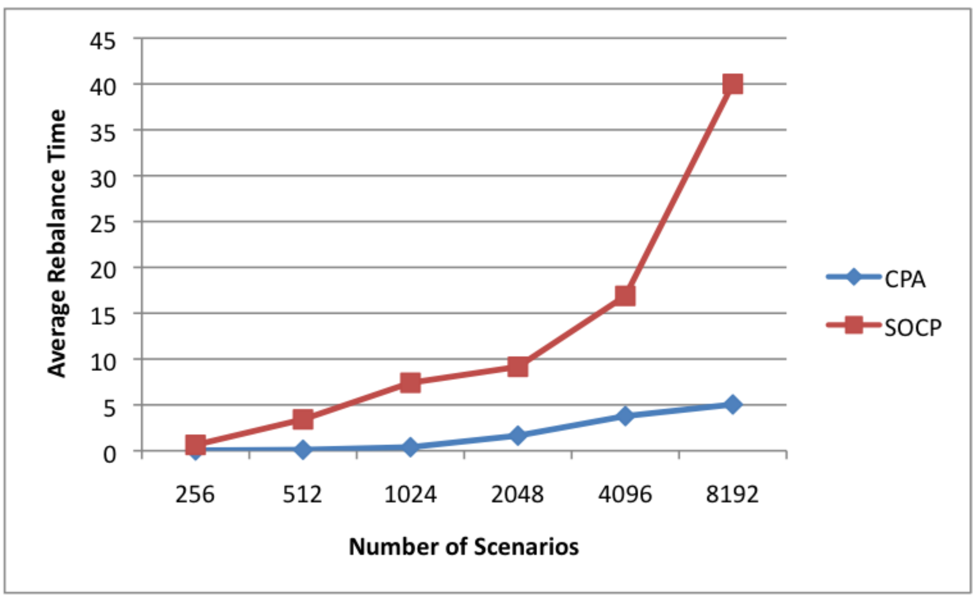

Figure 4.4: Average running times (in seconds) of the cutting plane (CPA) algorithm and the SOCP reformulation for $p = 4$.

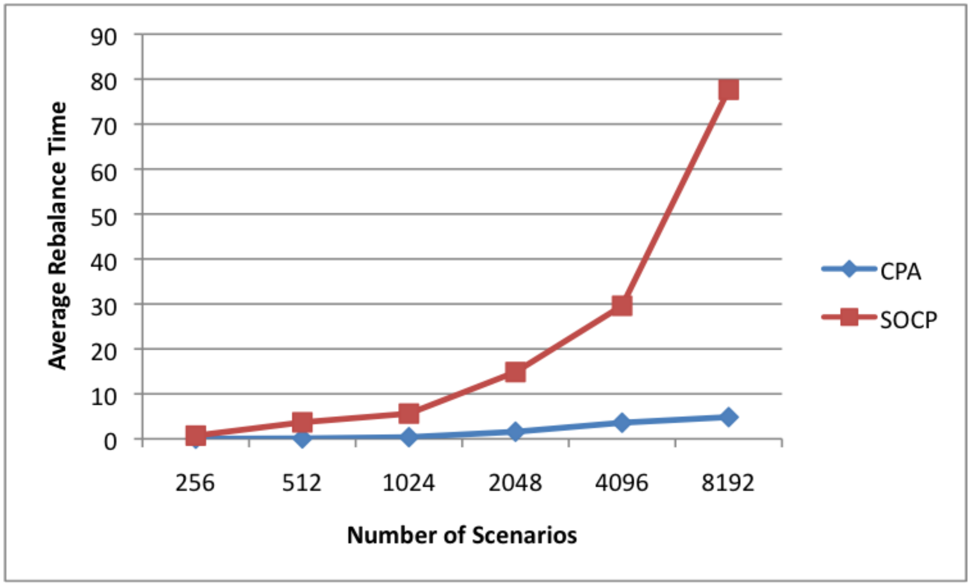

Figure 4.5: Average running times (in seconds) of the cutting plane (CPA) algorithm and the SOCP reformulation for $p = 5$.

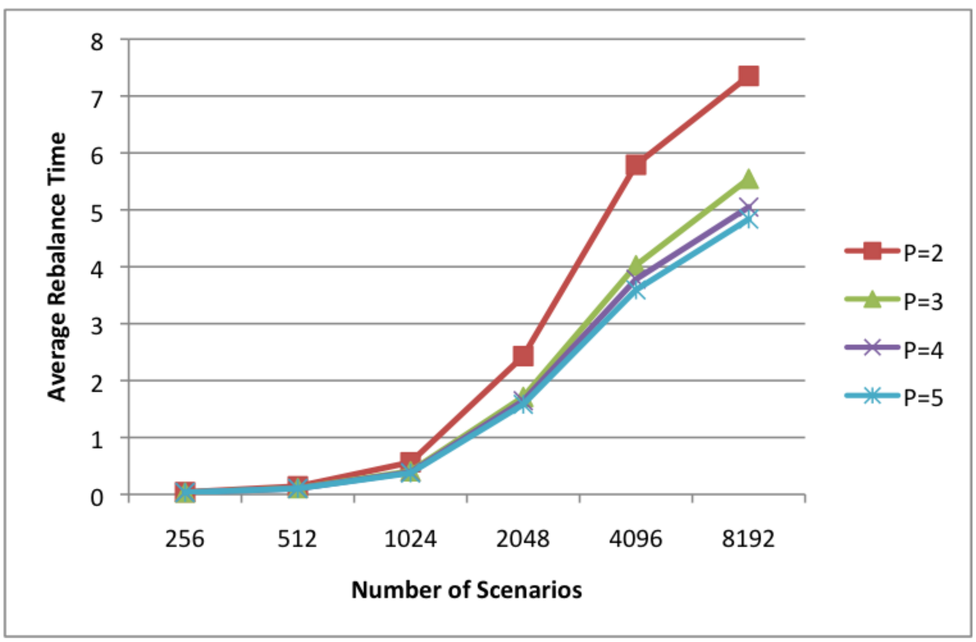

Figure 4.6: Average running times (in seconds) of the cutting plane algorithm for $p = 2, 3, 4$ and 5.

are utilized during the iterative cutting plane procedure, can be considered as major factors in the superior computational performance of the CPA algorithm.

The reasons for relatively poor performance of the SOCP-based algorithms may include the fact that, in general, most current SOCP solvers do not perform as well on instances of SOCP problems with a large number of quadratic conic constrains, as compared to problems with a few (but possibly high-dimensional) cones. Secondly, the inferior computational results may be due to possible performance limitations of the CPLEX Barrier solver.

In order to verify the last assumption regarding the computational efficiency of the CPLEX barrier solver we also used the MOSEK solver, with an interior-point algorithm, to determine the solution time of the SOCP problem. We saw no marked improvement in solution time over the cutting plane algorhtim when using the MOSEK solver in conjunction with the SOCP reformulation (see Figure 4.2).

Finally, we discuss the performance of the cutting plane algorithm at various $p = 2, 3, 4$ and 5 (see Figure 4.6). Namely, we see that for larger values of p, the computational time of the cutting plane algorithm generally decreases. This can be attributed to the fact that as p increases the $p$cone approaches the polyhedral $p = \infty$ cone, which can be natively handled by linear constraints.

4.4 Conclusions

In this chapter we conducted numerical experiments so as to determine the computational efficiency of the developed methods for solving pOCP problem on the example of a portfolio optimization problem with p-order constraints of small to medium dimensionality. The main conclusion of this case study is that the proposed approach based on polyhedral approximations of p-cones and subsequent solving of the resulting LP problem using a cutting plane algorithm turned out to be quite efficient. Namely, on the type and dimensionality of problems considered in this case study, the cutting plane algorithm that is based on gradient polyhedral approximations outperformed the "exact" SOCP implementations of the original pOCP problems. This can be attributed to two factors: first, the efficiency of cut generation procedure that was employed in the cutting plane algorithm (recall that in Chapter 3 we showed that for uniform gradient approximations the cuts can be generated in a *constant* time that does not depend on the accuracy of approximation). The second key factor that allowed the approximate cutting-plane implementation to outperform the interior-point SOCP solver seems to be the "warm-start" capability of linear programming solvers.

In addition, we note that the solution times for cutting plane algorithm seem to improve as the value of the parameter p increases. This can be explained by the fact that for large p, the corresponding p-cones become very close to the $p = \infty$ cone, which can be handled using linear programming techniques "natively".

CHAPTER 5
MIXED INTEGER P-ORDER CONIC PROGRAMMING

5.1 Introduction

Continuing our work within the general theme of p-order conic programming problems and the corresponding stochastic programming models, we now extend the results obtained in the context of polyhedral approximations for p-order conic programming problems (Chapters 2 and 3) to mixed integer pOCP problems, i.e. linear problems with p-order conic constraints where some of the variables are restricted to be integer-valued. At the final stages of our research endeavor into risk optimization with p-order conic constraints, we will conduct a case study intended to elucidate the methodological advantages (and disadvantages) of using the various risk measures that involve higher moments of loss distributions and can be incorporated in stochastic optimization models using p-cone constraints. To estimate the practical "merits" of decision models based on p-order conic programming, we will conduct a simulated "out-of-sample" portfolio optimization case study using real-life financial data. Both continuous and discrete models will be considered.

5.2 Mixed Integer p-Order Conic Programming Problems

Discrete decision making models, where decision vector(s) are required to be integer-valued are among some of the most difficult yet important problems in operations research and management science. In this context, we are considering the mixed-integer version of the general pOCP problem (2.1), where some decision variables may be restricted to integer-only values. By denoting the integer-valued part of the decision vector as $\mathbf{y} \in \mathbb{Z}^m$, and the real-valued components of the decision vector as $\mathbf{x} \in \mathbb{R}^n$, we formulate the general mixed integer p-order conic programming problem (MIpOCP) as follows:

$$\max \quad \mathbf{c}^\top \mathbf{x} + \mathbf{d}^\top \mathbf{y} \tag{5.1a}$$

$$\text{s. t.} \quad \mathbf{D}\mathbf{x} + \mathbf{E}\mathbf{y} \leq \mathbf{f} \tag{5.1b}$$

$$\|\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{y} + \mathbf{e}\|_p \leq \mathbf{a}^\top \mathbf{x} + \mathbf{b}^\top \mathbf{y} + e_0 \tag{5.1c}$$

$$\mathbf{x} \in \mathbb{R}^n \tag{5.1d}$$

$$\mathbf{y} \in \mathbb{Z}^m \tag{5.1e}$$

where, for simplicity, it can be assumed that only a single p-order conic constraint (5.1c) is present. Our goal is to develop an efficient exact (e.g., branch-and-bound) solution algorithm for the MIpOCP problem (5.1). The proposed approach to the MIpOCP problem follows the work of Vielma, Ahmed, and Nemhauser (2008) who developed a branch-and-bound algorithm for solving mixed integer SOCP problems.

To address the MIpOCP problem (5.1), we assume that the cone order p is a rational number: $p \in \mathbb{Q}$, (i.e. $p = \frac{r}{s}, r, s \in \mathbb{Z}_+$). With this assumption in place, we can reformulate the p-cone constraint(5.1c) as a set of 3 dimensional second order conic constraints and complemented linear constraints using the techniques developed in Chapter 2. A key property of this transformation that was established in Proposition 2.2 is that the number of second order conic constraints necessary to represent a p-cone of rational order with dimension $J + 1$ is equal to $O(J \log r)$. In such a way, the MIpOCP problem (5.1) can be reformulated as a mixed integer second-order conic programming problem (MISOCP) as follows:

$$
\begin{aligned}
\max \quad & \mathbf{c}^\top \mathbf{x} + \mathbf{d}^\top \mathbf{y} && \text{(5.2a)}\\
\text{s. t.} \quad & \mathbf{D}\mathbf{x} + \mathbf{E}\mathbf{y} \le \mathbf{f} && \text{(5.2b)}\\
& \mathbf{Q}\mathbf{x} + \mathbf{T}\mathbf{y} + \mathbf{S}\mathbf{u} \le \mathbf{0} && \text{(5.2c)}\\
& \left\| \mathbf{A}^{(i)}\mathbf{x} + \mathbf{B}^{(i)}\mathbf{y} + \mathbf{C}^{(i)}\mathbf{u} + \mathbf{e}^{(i)} \right\|_2 \le \mathbf{a}^{(i)\top}\mathbf{x} + \mathbf{b}^{(i)\top}\mathbf{y} + \mathbf{h}^{(i)\top}\mathbf{u} + e_0^{(i)}, \quad i \in \mathcal{I}_{r,s} && \text{(5.2d)}\\
& \mathbf{x} \in \mathbb{R}^n,\ \mathbf{y} \in \mathbb{Z}^m,\ \mathbf{u} \in \mathbb{R}^\nu && \text{(5.2e)}
\end{aligned}
$$

where $|\mathcal{I}_{r,s}| = O(J \log r)$. Next, a polyhedral approximation of (5.2) is formed by replacing each quadratic cone by its "lifted" polyhedral approximation due to Ben-Tal and Nemirovski (2001b) ($MILPP_{BN}$) or its gradient approximation ($MILPP_{GA}$), resulting in the following mixed integer linear programming problem:

$$
\begin{aligned}
\max \quad & \mathbf{c}^\top \mathbf{x} + \mathbf{d}^\top \mathbf{y} && \text{(5.3a)}\\
\text{s. t.} \quad & \mathbf{D}\mathbf{x} + \mathbf{E}\mathbf{y} \le \mathbf{f} && \text{(5.3b)}\\
& \mathbf{Q}\mathbf{x} + \mathbf{T}\mathbf{y} + \mathbf{S}\mathbf{u} \le \mathbf{0} && \text{(5.3c)}\\
& \mathbf{H}_m^{(3)} \begin{pmatrix} \mathbf{A}^{(i)}\mathbf{x} + \mathbf{B}^{(i)}\mathbf{y} + \mathbf{C}^{(i)}\mathbf{u} + \mathbf{e}^{(i)} \\ \mathbf{a}^{(i)\top}\mathbf{x} + \mathbf{b}^{(i)\top}\mathbf{y} + \mathbf{h}^{(i)\top}\mathbf{u} + e_0^{(i)} \\ \mathbf{v}^{(i)} \end{pmatrix} \ge \mathbf{0}, \quad i \in \mathcal{I}_{r,s} && \text{(5.3d)}\\
& \mathbf{x} \in \mathbb{R}^n,\ \mathbf{y} \in \mathbb{Z}^m,\ \mathbf{u} \in \mathbb{R}^\nu,\ \mathbf{v} \in \mathbb{R}^{\kappa_m} && \text{(5.3e)}
\end{aligned}
$$

Following Vielma, Ahmed, and Nemhauser (2008), the branch-and-bound algorithm for problem (5.2) solves a continuous relaxation of the mixed integer linear program (5.3) at each node of the branch-and-bound tree. If an integer-valued solution is found, its feasibility to (5.2) is tested by solving the corresponding continuous relaxation of (5.2) at this node. If the current solution is, despite integrality, infeasible to the continuous relaxation of (5.2), further branching is performed on this node. This will be looked at further in the Case Studies chapter and the results will be compared with that of the MISOCP reformulation. We will also look at the portfolio performance as compared to the traditional Mean-Variance and CVaR risk measures. We will now look at the various ways in which we can create the continuous linear relaxation of our MIpOCP problem by using one of the two polyhedral approximations that we are familiar with from Chapter 2.

5.3 Polyhedral Approximations of the MIpOCP Problem

As stated above in problem (5.2), the MIpOCP problem can be reformulated as a MISOCP problem with $O(J \log r)$ second order cones. Since the Branch and Bound algorithm that will be employed requires a continuous relaxation solution at each node of the algorithm, we can either employ the Ben-Tal and Nemerovski approximation of the p-order conic constraint or the gradient approximation in order to reduce the MIpOCP problem to a linear programming problem. As was stated before, the algorithm that will be used to solve our MIpOCP problems will be a mixture of the Branch-and-Bound Algorithm developed by Vielma, Ahmed, and Nemhauser (2008) in conjunction with the polyhedral approximation developed by Ben-Tal and Nemirovski (2001b).

5.3.1 Ben-Tal Nemerovski's Lifted Polyhedral Approximation

After the reformulation of the MIpOCP problem as a MISOCP problem, one possible approximation that we can use is its "lifted" polyhedral approximation. This is done by replacing each of the corresponding second order cones in the MISOCP with its approximation developed by Ben-Tal and Nemirovski (2001b) ($MILPP_{BN}$). This approximation is particularly effective because it allows for a compact and elegant approximation to the p-cone problem.

5.3.2 Gradient Approximation

As before, in order to implement the branch-and-bound algorithm we must be able to obtain a continuous relaxation solution at each node of the algorithm. In order to create the linear programming relaxation of the p-order conic programming problem, we employ our gradient approximation as it was defined in Chapter 3 ($MILPP_{GA}$). Since the relaxation is used to help with the pruning process for the branch-and-bound algorithm, a tight approximation is not needed.

5.4 Branch-and-Bound Algorithm for MIpOCP

As was stated before, the algorithm that will be used to solve our MIpOCP problems will be an adaptation of the branch-and-bound algorithm developed by Vielma, Ahmed, and Nemhauser (2008) in conjunction with our uniform gradient polyhedral approximation. This is the natural choice since we are interested in finding solutions for cones of ration order.

5.4.1 Pseudo Code for Branch-and-Bound Algorithm For MIpOCP problem

The branch-and-bound algorithm that was developed by Vielma, Ahmed, and Nemhauser (2008) was designed for the case $p = 2$. It is this algorithm that we have adapted to the case of rational p. In order to take advantage of the branch-and-bound algorithm for rational p, we use our uniform gradient polyhedral approximation so that we can construct the MILPP for any cone of rational order. Thus, we use the $MILPP_{GA}$ in our branch-and-bound algorithm.

1. Set global upper bound UB:$= +\infty$
2. Set nodes $l_i^0 := -\infty, u_i^0 := +\infty, \quad \forall i \in \{1, \cdots, n\}$
3. Set $\mathrm{LB}^0 = -\infty$
4. Set node list $\mathscr{N} := \{(l^0, u^0, \mathrm{LB}^0)\}$
5. **while** $\mathscr{N} \neq \varnothing$ **do**
6. Select and remove node $(l^k, u^k, \mathrm{LB}^k) \in \mathscr{N}$
7. Initialize and solve the master problem $MILPP_{GA}$ (l^k, u^k):

$$
\begin{aligned}
\max \quad & \mathbf{c}^\top \mathbf{x} + \mathbf{d}^\top \mathbf{y} \\
\text{s. t.} \quad & \mathbf{D}\mathbf{x} + \mathbf{E}\mathbf{y} \leq \mathbf{f} \\
& \mathbf{Q}\mathbf{x} + \mathbf{T}\mathbf{y} + \mathbf{S}\mathbf{u} \leq \mathbf{0} \\
& \mathbf{H}_m^{(3)} \begin{pmatrix} \mathbf{A}^{(i)}\mathbf{x} + \mathbf{B}^{(i)}\mathbf{y} + \mathbf{C}^{(i)}\mathbf{u} + \mathbf{e}^{(i)} \\ \mathbf{a}^{(i)\top}\mathbf{x} + \mathbf{b}^{(i)\top}\mathbf{y} + \mathbf{h}^{(i)\top}\mathbf{u} + e_0^{(i)} \\ \mathbf{v}^{(i)} \end{pmatrix} \geq \mathbf{0}, \quad i \in \mathcal{I}_{r,s} \\
& \mathbf{x} \in \mathbb{R}^n, \ \mathbf{y} \in \mathbb{Z}^m, \ \mathbf{u} \in \mathbb{R}^\nu, \ \mathbf{v} \in \mathbb{R}^{\kappa_m}
\end{aligned}
$$

8. **if** $MILPP_{GA}$ (l^k, u^k) is feasible **and** $\mathrm{OBJ}_{\mathrm{MISOCP}} < \mathrm{UB}$ **then**
9. Let $(\hat{x}^k, \hat{y}^k, \hat{w}^k)$ be the optimal solution of $MILPP_{GA}$ (l^k, u^k);
10. **if** $\hat{y}^k \in \mathbb{Z}$ **then**

Solve MISOCP(x^*, y^*, w^*)

if MISOCP(x^*, y^*, w^*) is feasible and $\mathrm{OBJ}_{\mathrm{MISOCP}} < \mathrm{UB}$ **then**

  UB := $\mathrm{OBJ}_{\mathrm{MISOCP}}$

end

if $l^k \neq u^k$ **and** $\mathrm{OBJ}_{\mathrm{MISOCP}} < \mathrm{UB}$ **then**

  Solve MISOCP(l^k, u^k)

  if MISOCP(l^k, u^k) is feasible and $\mathrm{OBJ}_{\mathrm{MISOCP}} < \mathrm{UB}$ **then**

    Let $(\bar{x}^k, \bar{y}^k, \bar{w}^k)$ be the optimal solution of MISOCP(l^k, u^k)

    if $\bar{y}^k \in \mathbb{Z}$ **then**

      UB := $\mathrm{OBJ}_{\mathrm{MISOCP}}$

    else

      Pick i_0 in $\{i \in \{1, \cdots, n\}$: $\bar{x}_i^k \notin \mathbb{Z}\}$

      Let $l_i = l_i^k, u_i = u_i^k \quad \forall i \in \{1, \cdots, n\} \backslash \{i_0\}$

      Let $u_{i_0} = \lfloor \bar{x}_{i_0}^k \rfloor$, $l_{i_0} = \lfloor \bar{x}_{i_0}^k \rfloor + 1$

      $\mathscr{N} := \mathscr{N} \cup \{(l^k, u, \mathrm{OBJ}_{\mathrm{MISOCP}}), (l, u^k, \mathrm{OBJ}_{\mathrm{MISOCP}})\}$

    end

  end

end

else

Pick i_0 in $\{i \in \{1, \cdots, n\}$: $\hat{x}_i^k \notin \mathbb{Z}\}$

Let $l_i = l_i^k, u_i = u_i^k \quad \forall i \in \{1, \cdots, n\} \backslash \{i_0\}$

Let $u_{i_0} = \lfloor \hat{x}_{i_0}^k \rfloor$, $l_{i_0} = \lfloor \hat{x}_{i_0}^k \rfloor + 1$

$\mathscr{N} := \mathscr{N} \cup \{(l^k, u, \mathrm{OBJ}_{MILPP_{GA}}), (l, u^k, \mathrm{OBJ}_{MILPP_{GA}})\}$

end

end

Remove every node $(l^k, u^k, \mathrm{UB}^k)$ such that $\mathrm{UB} \leq \mathrm{LB}^k$

end

In order for the branch-and-bound algorithm to be effective, it must be shown that it will terminate in a finite number of steps and that it will terminate with the optimal solution to the mixed integer nonlinear programming problem. Vielma et al. (2008), showed that the following proposition holds:

Proposition 9. *For any polyhedral relaxation (MILPP) of the nonlinear programming problem using a bounded polyhedron* $\mathbf{H}_m^{(3)}$*, the lifted linear programming branch-and-bound algorithm above terminates with lower bound equal to the optimal objective value of the mixed integer nonlinear programming problem.*

Proof. The arguments of Vielma et al. (2008) carry over to our case practically without modifications. (See Vielma, Ahmed, and Nemhauser (2008), page 441) □

5.5 Case Studies: Portfolio Optimization With Integrality And p-order Constrains

In this section we discuss the computational efficiency of the developed branch-and-bound algorithm for solving the p-order conic programming problem with integrality constraints (MIpOCP). We will be comparing the MISOCP (5.2) reformulation of the problem with the branch-and-bound algorithm that was developed in Chapter 5. The developed branch-and-bound algorithm will be tested for efficiency with respect to solution time based on the size of the scenario set. Also of interest are the optimal portfolios that the algorithm will yield.

As before, we will later consider methodological aspects of the coherent and deviation measures that involve higher moments of loss distributions, such as HMCR, SMCR, HMD and SMD measures (see Chapter 1). We will be looking at how the risk measures based on higher moments of tail loss perform when compared to each other and to more conventional risk measures such as the CVaR, and the Mean-Variance models.

5.6 Computational Performance of Mixed Integer p-Order Conic Programming Branch-and-Bound Algorithms

In this section we conduct numerical comparisons of the approaches discussed above to solving problems of type (5.1). An application of the branch-and-bound method for MIpOCP (5.1) will be demonstrated on a portfolio optimization problem with cardinality constraints. Cardinality constrained portfolio allocation problems typically arise in situations when no more than k assets are allowed to be in the portfolio, or, equivalently, each asset is not allowed to exceed a certain fraction of the portfolio value. In the case when no more than k assets are allowed in the portfolio, the cardinality-constrained portfolio optimization problem corresponding to the problem (4.1) considered in Chapter 4 can be written by introducing new binary decision variables $\mathbf{z}$:

$$\min_{\mathbf{x},\,\mathbf{z}} \quad \mathscr{R}(-\mathbf{r}^\top \mathbf{x}) \tag{5.4a}$$

$$\text{s. t.} \quad \mathbf{e}^\top \mathbf{x} = 1 \tag{5.4b}$$

$$\mathsf{E}\mathbf{r}^\top \mathbf{x} \geq r_0 \tag{5.4c}$$

$$\mathbf{e}^\top \mathbf{z} \leq k \tag{5.4d}$$

$$\mathbf{0} \leq \mathbf{x} \leq \mathbf{z} \tag{5.4e}$$

$$\mathbf{z} \in \{0, 1\}^n \tag{5.4f}$$

where $\mathbf{x} = (x_1, \ldots, x_n)^\top$ is the vector of portfolio weights, $\mathbf{r} = (r_1, \ldots, r_n)^\top$ is the random vector of return on portfolio instruments, and $\mathbf{e} = (1, \ldots, 1)^\top$. The risk measure $\mathscr{R}(X)$ in (5.4a) can be taken to be HMCR, SMCR, etc., of the negative portfolio return, $-\mathbf{r}^\top \mathbf{x}$. Constraint (5.4b) is the budget constraint while (5.4d) together with (5.4b) ensure that all of the available funds are invested. Constraint (5.4c) imposes a minimal required level r_0 of expected return of the resulting portfolio. Obviously, the meaning of variables $\mathbf{z}$ is $z_i = 1$ if asset i is present in the portfolio, and $z_i = 0$ otherwise.

With the risk measure $\mathscr{R}(X)$ replaced with the $\mathrm{HMCR}_{p,\alpha}(X)$ measure, the portfolio selection problem for our case study is transformed into a mixed integer linear programming problem with a single p-order conic constraint (5.5e)

$$
\begin{aligned}
\min \quad & \eta + \frac{J^{-\frac{1}{p}}}{1-\alpha} t && && (5.5a)\\
\text{s. t.} \quad & \sum_{i=1}^{n} x_i = 1, && && (5.5b)\\
& \frac{1}{J}\sum_{j=1}^{J}\sum_{i=1}^{n} r_{ij}x_i \geq r_0, && && (5.5c)\\
& w_j \geq -\sum_{i=1}^{n} r_{ij}x_i - \eta, && j = 1,\cdots,J, && (5.5d)\\
& t \geq (w_1^p + \cdots + w_J^p)^{1/p}, && && (5.5e)\\
& \sum_{i=1}^{n} z_i \leq k && && (5.5f)\\
& x_i \geq 0 && i = 1,\cdots,n, && (5.5g)\\
& w_j \geq 0 && j = 1,\cdots,J && (5.5h)
\end{aligned}
$$

5.6.1 Set of Instrument and Scenario Data

In order to take advantage of the construct of the HMCR risk measures, which quantify risk in terms of higher tail moments of loss distribution, the portfolio optimization case studies were conducted using return data of the fifty S&P500 stocks with the so-called "heavy tails". In order to look at computation time comparisons, for scenario generations we used 10-day historical returns over $J = 2^7, \cdots, 2^{10}$ overlapping periods, calculated using daily closing prices from October 30, 1998 to January 18, 2006. The particular sizes of the scenario set has been chosen to accommodate the linear approximation techniques in problem (5.1), and the sizes of the considered scenario sets were limited only by availability of the data. From this set of S&P500 stocks, we selected $n = 50$ instruments by picking those with the highest value of kurtosis of biweekly returns, calculated over a specific period.

5.7 Computational Time Comparisons

In this section we present the computational efficiency, as measured by the average running time, of the developed algorithm for solving MIpOCP problems on the example of the portfolio optimization problem with integrality and p-order conic constraint (5.5). We compare the solution time of the branch-and-bound algorithm with $MILPP_{GA}$ as the continuous linear relaxation that is solved at each node of the branch-and-bound tree and the second order conic programming (SOCP) reformulation. For the branch-and-bound with gradient approximation and SOCP reformulation, the value of the parameter p in (5.5) varied as $p =, 3, 4, 5$. The confidence level α, number of instruments in the portfolio, k, and minimum required expected return have been fixed at $\alpha = 0.9$, $k = 5$ and $r_0 = 0.5\%$ respectively for all algorithms.

A total of 10 instances of problem (5.5) corresponding to 10 bi-weekly rebalancing periods from December, 2002 to January 2006 have been solved for each implementation and each scenario size.

The computer that was used to perform the scenario runs was a Dell XPS with a Dual Core Pentium processor and 2GB of RAM. The machine was running Windows XP with CPLEX 10.0.0. The ILOG Concert Technology implementation of the CPA algorithm utilized the CPLEX linear programming solver, and the SOCP implementation used CPLEX Barrier solver. The BN and GA implementations were done using AMPL. The accuracy of the polyhedral approximations were chosen at $\varepsilon < 10^{-5}$ that is consistent with the standard CPLEX computation accuracy.

	p=3		p=4		p=5	
J	**MIpOCP**	**MISOCP**	**MIpOCP**	**MISOCP**	**MIpOCP**	**MISOCP**
128	2.7	86.8	2.8	83.3	2.5	113.1
256	3.8	88.0	3.8	251.9	3.8	325.6
512	32.7	2114.5	22.3	1613.5	23.8	2483.0
1024	9.2	717.0	5.0	664.6	5.5	831.8

Table 5.1: Average running times (in seconds) for the MIpOCP with Branch-and-Bound and SOCP reformulation of problem (5.5) with $p = 3, 4, 5$, and $\alpha = 0.9$,$r_0 = 0.05\%$.

5.7.1 Computational Results

The main conclusion of this computational study is that the developed branch-and-bound algorithm as applied with the uniform gradient approximation of the pOCP problem consistently outperforms the corresponding interior-point algorithm based on the “exact” SOCP reformulation of the problem.

We believe that the following factors can be contributing to the observed differences in efficiencies of the polyhedral approximation/cutting plane procedure and the SOCP-based solution approach. First, the reasons for relatively poor performance of the SOCP-base algorithms may include the fact that, in general, most current SOCP solvers do not perform as well on instances of SOCP problems with a large number of quadratic conic constrains, as compared to problems with a few (but possibly high-dimensional) cones. Secondly, the inferior computational results may be due to possible performance limitations of the CPLEX Barrier solver. Again, we plan to verify these findings by implementing the SOCP reformulations of pOCP problems using the more advanced MOSEK interior-point solver.

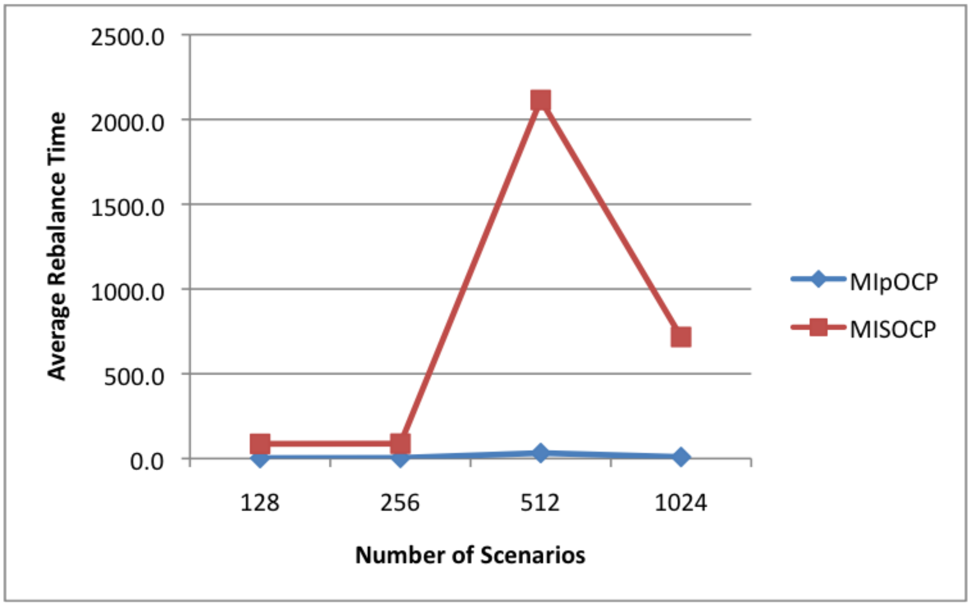

Figure 5.1: Average runtime (in seconds) of the MIpOCP with Branch-and-Bound and MISOCP algorithm for $p = 3$.

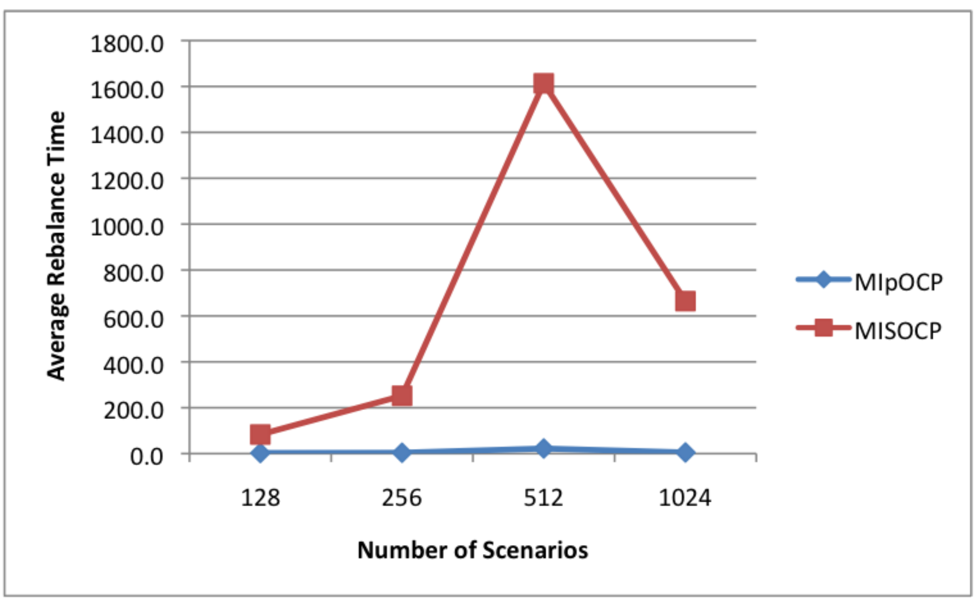

Figure 5.2: Average runtime (in seconds) of the MIpOCP with Branch-and-Bound and MISOCP algorithm for $p = 4$.

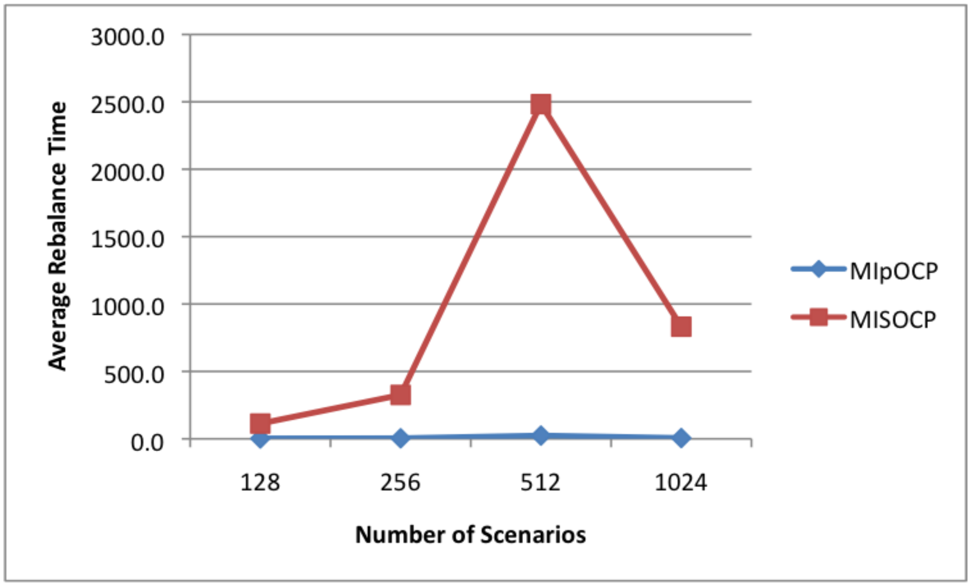

Figure 5.3: Average runtime (in seconds) of the MIpOCP with Branch-and-Bound and MISOCP algorithm for $p = 5$.

5.8 Conclusions

In this chapter we developed the branch-and-bound algorithm for solving the mixed integer p-order conic programming problem and conducted numerical experiments so as to determine the computational efficiency of the developed methods for solving MIpOCP problems on the example of a portfolio optimization problem with integrality constraints and p-order constraints of small to medium dimensionality. The main conclusion of this case study is that the proposed approach based on the branch-and-bound algorithm turned out to be quite efficient.

In addition, we note that the difference in the solution times for the branch-and-bound algorithm and the SOCP reformulation became larger as parameter J increases. This would be expected given the size in the problem, however, the major difference in time can be attributed to the handling of SOCP problems that contain many second order cones of low dimensionality as opposed to one second order cone of a high dimension.

Finally, we address the spike in the numerical data which we attribute to the use of real life data. It is not guaranteed that the solution time will always increase as the scenario size increases. Some problems are inherently more difficult to solve. This spike in the numerical data was probably due to the problem having a hard time meeting the requirement of having 5 instruments in the portfolio.

CHAPTER 6
CASE STUDY: A COMPARATIVE ANALYSIS OF PORTFOLIO REBALANCING STRATEGIES BASED ON HIGHER MOMENT COHERENT RISK MEASURES

6.1 Introduction

In this chapter we will look at the performance of the risk measures base on higher moments of loss when compared to other industry standard risk measures such as CVaR and Mean-Variance. We will see that in the case of the given portfolio optimization model, the HMCR risk measures consistently outperforms the other risk measures.

We revisit the portfolio optimization problems that were considered earlier in Chapter 4 and Chapter 5 to determine the effectiveness of the HMCR risk measure. First we will consider the general portfolio optimization problem:

$$\min_{\mathbf{x},\,\mathbf{z}} \quad \mathscr{R}(-\mathbf{r}^\top \mathbf{x}) \tag{6.1a}$$
$$\text{s. t.} \quad \mathbf{e}^\top \mathbf{x} = 1 \tag{6.1b}$$
$$\mathsf{E}\mathbf{r}^\top \mathbf{x} \geq r_0 \tag{6.1c}$$
$$\mathbf{e}^\top \mathbf{z} \leq k \tag{6.1d}$$
$$\mathbf{0} \leq \mathbf{x} \leq \mathbf{z} \tag{6.1e}$$
$$\mathbf{z} \in \{0,1\}^n \tag{6.1f}$$

where $\mathbf{x} = (x_1, \ldots, x_n)^\top$ is the vector of portfolio weights, $\mathbf{r} = (r_1, \ldots, r_n)^\top$ is the random vector of return on portfolio instruments, and $\mathbf{e} = (1, \ldots, 1)^\top$. The risk measure $\mathscr{R}(X)$ in (5.4a) will be taken to be HMCR, SMCR and CVaR, of the negative portfolio return, $-\mathbf{r}^\top \mathbf{x}$. Constraint (6.1d) will be included when we consider the portfolio optimization problem with cardinality constraint. Obviously, the meaning of variables $\mathbf{z}$ is $z_i = 1$ if asset i is present in the portfolio, and $z_i = 0$ otherwise.

With the risk measure $\mathscr{R}(X)$ replaced with the $\mathrm{HMCR}_{1,\alpha}(X)$ measure (CVaR), the portfolio selection problem for our case study is transformed into a linear programming problem:

$$
\begin{aligned}
\min \quad & \eta + \frac{J^{-1}}{1-\alpha} t && && (6.2a)\\
\text{s. t.} \quad & \sum_{i=1}^{n} x_i = 1, && && (6.2b)\\
& \frac{1}{J}\sum_{j=1}^{J}\sum_{i=1}^{n} r_{ij} x_i \geq r_0, && && (6.2c)\\
& w_j \geq -\sum_{i=1}^{n} r_{ij} x_i - \eta, && j = 1, \cdots, J, && (6.2d)\\
& t \geq w_1 + \cdots + w_J, && && (6.2e)\\
& x_i \geq 0 && i = 1, \cdots, n, && (6.2f)\\
& w_j \geq 0 && j = 1, \cdots, J && (6.2g)
\end{aligned}
$$

With the risk measure $\mathscr{R}(X)$ replaced with the $\mathrm{HMCR}_{2,\alpha}(X)$ measure, and the addition of the constraint $\eta = \mathsf{E}(X)$ (SMCR), the portfolio selection problem for our case study is transformed into a linear programming problem with a single 2-order conic constraint:

$$
\begin{aligned}
\min \quad & \eta + \frac{J^{-\frac{1}{2}}}{1-\alpha} t && && \text{(6.3a)}\\
\text{s. t.} \quad & \sum_{i=1}^{n} x_i = 1, && && \text{(6.3b)}\\
& \eta = \frac{1}{n} \sum_{i=1}^{n} x_i && && \text{(6.3c)}\\
& \frac{1}{J} \sum_{j=1}^{J} \sum_{i=1}^{n} r_{ij} x_i \geq r_0, && && \text{(6.3d)}\\
& w_j \geq -\sum_{i=1}^{n} r_{ij} x_i - \eta, && j = 1, \cdots, J, && \text{(6.3e)}\\
& t \geq (w_1^2 + \cdots + w_J^2)^{1/2}, && && \text{(6.3f)}\\
& x_i \geq 0 && i = 1, \cdots, n, && \text{(6.3g)}\\
& w_j \geq 0 && j = 1, \cdots, J && \text{(6.3h)}
\end{aligned}
$$

With the risk measure $\mathscr{R}(X)$ replaced with the $\mathrm{HMCR}_{3,\alpha}(X)$ measure (HMCR3), the portfolio selection problem for our case study is transformed into a linear programming problem with a single 3-order conic constraint:

$$
\begin{aligned}
\min \quad & \eta + \frac{J^{-\frac{1}{3}}}{1-\alpha} t && && (6.4a)\\
\text{s. t.} \quad & \sum_{i=1}^{n} x_i = 1, && && (6.4b)\\
& \frac{1}{J} \sum_{j=1}^{J} \sum_{i=1}^{n} r_{ij} x_i \geq r_0, && && (6.4c)\\
& w_j \geq -\sum_{i=1}^{n} r_{ij} x_i - \eta, && j = 1, \cdots, J, && (6.4d)\\
& t \geq (w_1^3 + \cdots + w_J^3)^{1/3}, && && (6.4e)\\
& x_i \geq 0 && i = 1, \cdots, n, && (6.4f)\\
& w_j \geq 0 && j = 1, \cdots, J && (6.4g)
\end{aligned}
$$

With the risk measure $\mathscr{R}(X)$ replaced with the $\mathrm{HMCR}_{3,\alpha}(X)$ measure and the inclusion of the cardinality constraint (HMCR3-Int), the portfolio selection problem with cardinality constraint for our case study is transformed into a mixed integer linear programming problem with a single 3-order conic constraint (6.5e)

$$\min \quad \eta + \frac{J^{-\frac{1}{3}}}{1-\alpha} t \tag{6.5a}$$

$$\text{s. t.} \quad \sum_{i=1}^{n} x_i = 1, \tag{6.5b}$$

$$\frac{1}{J} \sum_{j=1}^{J} \sum_{i=1}^{n} r_{ij} x_i \geq r_0, \tag{6.5c}$$

$$w_j \geq -\sum_{i=1}^{n} r_{ij} x_i - \eta, \qquad j = 1, \cdots, J, \tag{6.5d}$$

$$t \geq (w_1^3 + \cdots + w_J^3)^{1/3}, \tag{6.5e}$$

$$\sum_{i=1}^{n} z_i \leq k \tag{6.5f}$$

$$x_i \geq 0 \qquad i = 1, \cdots, n, \tag{6.5g}$$

$$w_j \geq 0 \qquad j = 1, \cdots, J \tag{6.5h}$$

6.2 Out-of-Sample Simulation Case Studies

One of the primary goals of this research theme is to try and reflect a "true to life" performance of the HMCR and related higher-order measures in risk management applications. To this end, we will conduct the so-called *out-of-sample* experiments. This method determines the merits of a constructed solution using the *out-of-sample* data that have not been included in the scenario model that was used to generate the solution. By using this method, the out-of-sample setup simulates a common situation when the true realization of uncertainties are unknown to the decision-maker, and the decision $\mathbf{x}$ must be made using the "known" (in-sample) data ω_0, but the outcome $X(\mathbf{x}, \omega)$ of the decision will be evaluated using the "unknown-at-the-time", or out-of-sample data $\hat{\omega}$: $X = X(\mathbf{x}, \hat{\omega})$.

We will employ the out-of-sample method to compare simulated historic performances of several self-financing portfolio rebalancing strategies that will be based on risk measures that involve higher moments of loss distributions,such as HMCR, SMCR, and the corresponding deviation measures that have been presented in Chapter 1, as well some of the industry standard risk measures such as CVaR and Mean-Variance models. In addition, we will consider the effect of cardinality constraints as described above on the efficiency of the corresponding trading strategies. The data set for this case study will be updated to incorporate the most recent market recession, which is expected to make the results of this case study more interesting from the practical risk management viewpoint.

6.2.1 Set of Instrument and Scenario Data

In order to take advantage of the construct of the HMCR risk measures, which quantify risk in terms of higher tail moments of loss distribution, the portfolio optimization case studies were conducted using return data of the fifty S&P500 stocks with the so-called "heavy tails". In order to look at portfolio performance, for scenario generations we used 10-day historical returns over $J = 2^{10}$ overlapping periods, calculated using daily closing prices from October 30, 1998 to October 30, 2009. The particular sizes of the scenario set has been chosen to accommodate the linear approximation techniques in problems (2.1) and (5.1). From this set of S&P500 stocks, we selected $n = 50$ instruments by picking those with the highest value of kurtosis of biweekly returns, calculated over a specific period. The experiments were conducted with $r_0 = 0.5\%$ and $\alpha = .9$, $r_0 = 0.5\%$ and $\alpha = .95$, $r_0 = 1\%$ and $\alpha = .9$, $r_0 = 1\%$ and $\alpha = .95$ and finally $r_0 = 1.3\%$ and $\alpha = .9$, the latter of which will be used to assess the performance of a particulary agressive strategy.

6.2.2 Portfolio Performance

In all cases the clear winner is the $HMCR_{3,\alpha}(\cdot)$ as $\mathscr{R}(\cdot)$ with cardinality constraint. In general we see that the portfolio based on the $HMCR_{3,\alpha}(\cdot)$ risk measure dominates the $SMCR_{\alpha}(\cdot)$ and the $CVaR_{\alpha}(\cdot)$.

Interestingly, the portfolio optimization with cardinality constraint outperformed all the other problem formulations (see figures 6.1 and 6.2). This can be attributed to the inherent increase in risk, and thus increase in reward, that manifests with a less diversified portfolio. We should note that as r_0 increased, the branch-and-bound algorithm became more unstable and required a more refined LP relaxation. If we look at figures 6.3 and 6.4), we see that the cardinality constrained portfolio fell below

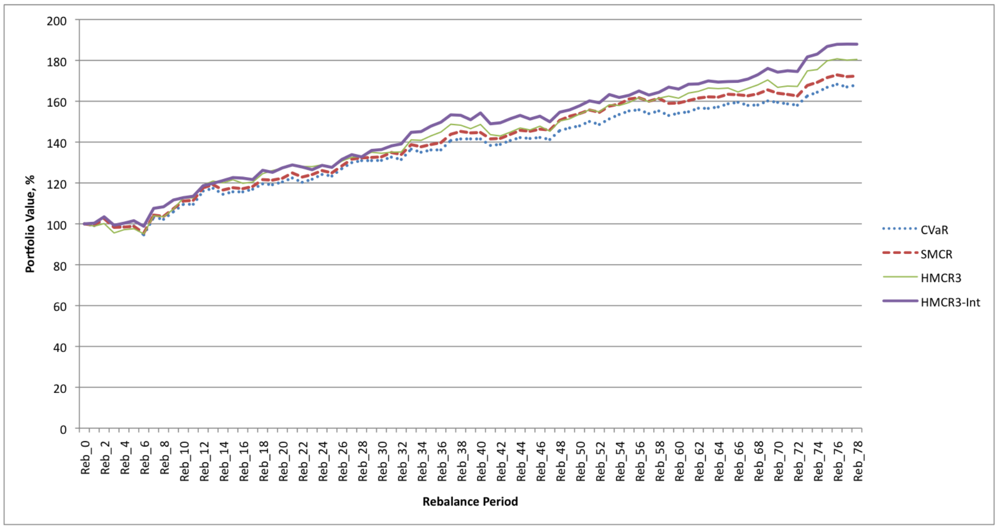

Figure 6.1: Portfolio performance comparison for CVaR, SMCR and HMCR=3, $r_0 = 0.5\%$ and $\alpha = .9$.

all the other problem formulations. This can be attributed to instances of the LP being infeasible due to the increase in r_0 from 0.5% to 1%. We see this in the areas of the graph where the portfolio flatlines, meaning that there was no change from one instance to the next due to a problem with infeasibility.

We also notice that with an increase in r_0 the value of the portfolio increases in the case of all the risk measures. However, as r_0 increases, the portfolio values, based on its particular risk measure, become harder to discern from each other. If we look at a particularly aggressive strategy where $r_0 = 1.3\%$ and $\alpha = .9$ (see figure 6.5) we see that the portfolio's are almost identical.

6.3 Conclusions

In this chapter we conducted numerical experiments so as to determine the effectiveness of the higher moment coherent risk measures as compared to other industry standard risk measures. We see that in all cases the HMCR3 risk measure was able

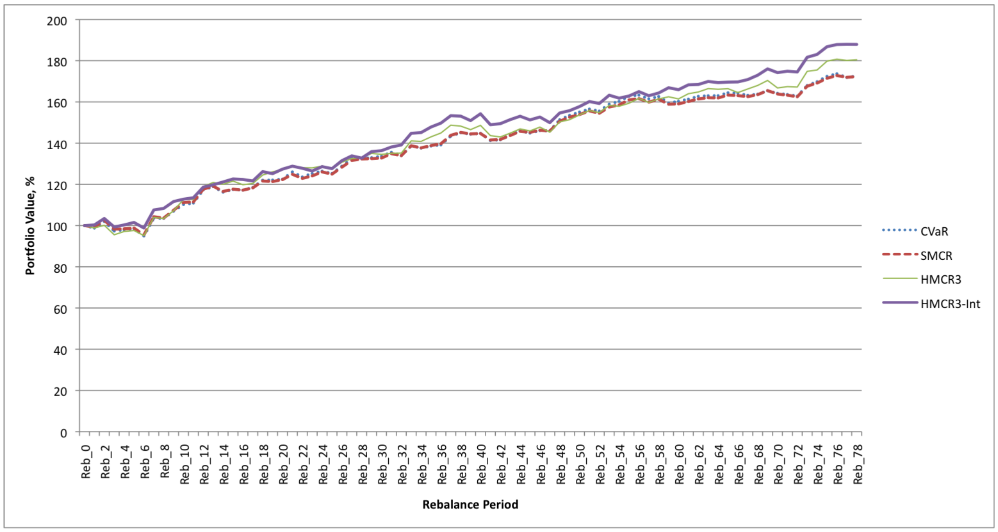

Figure 6.2: Portfolio performance comparison for CVaR, SMCR and HMCR=3, $r_0 = 0.5\%$ and $\alpha = .95$.

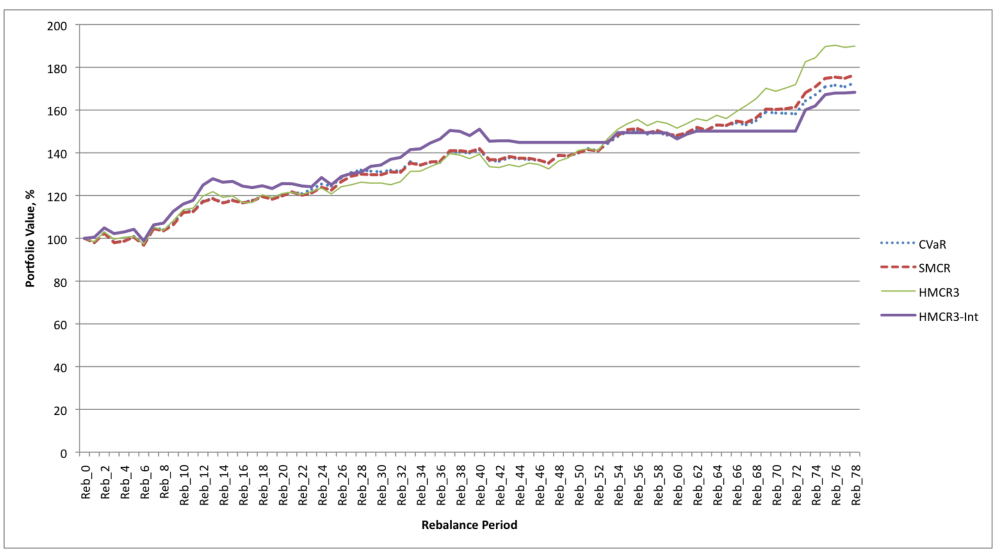

Figure 6.3: Portfolio performance comparison for CVaR, SMCR and HMCR=3, $r_0 = 1\%$ and $\alpha = .9$.

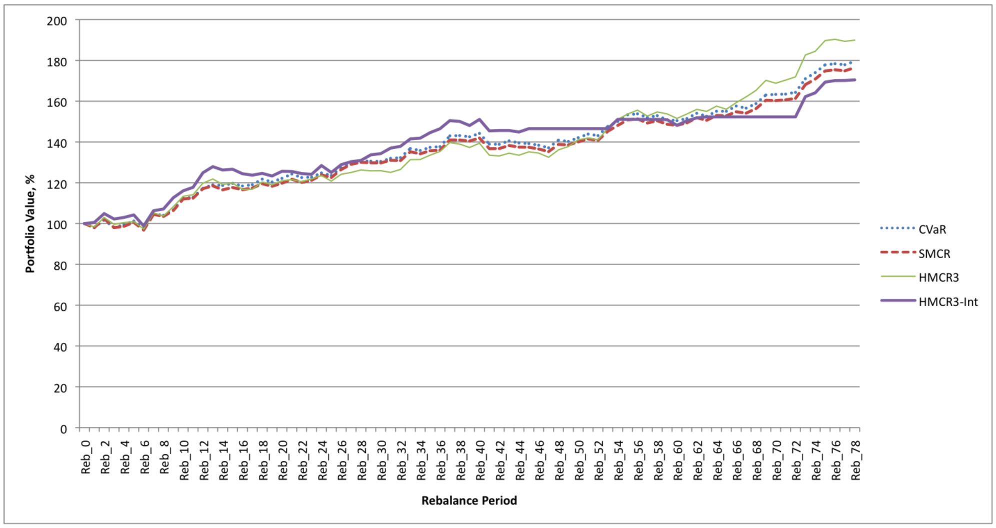

Figure 6.4: Portfolio performance comparison for CVaR, SMCR and HMCR=3, $r_0 = 1\%$ and $\alpha = .95$.

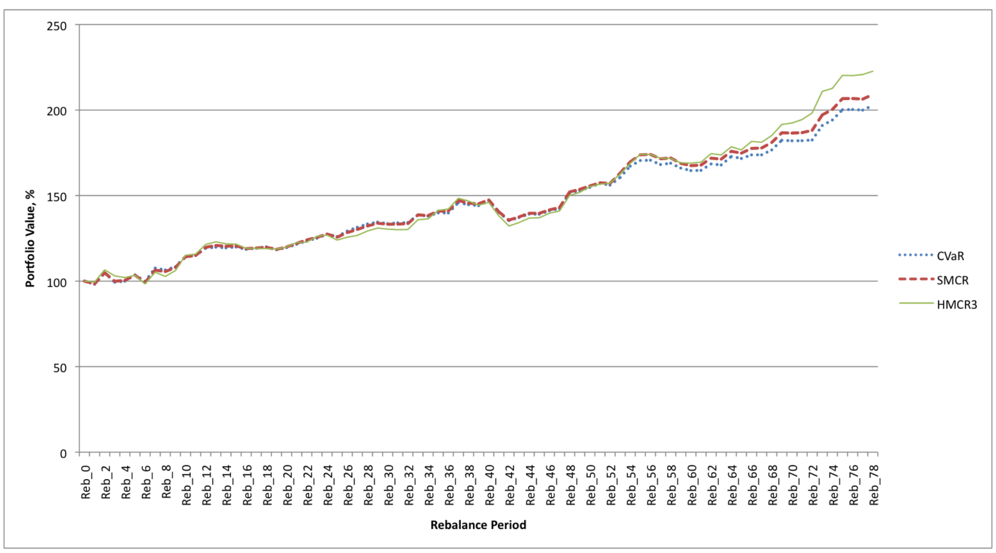

Figure 6.5: Portfolio performance comparison for CVaR, SMCR and HMCR=3, $r_0 = 1.3\%$ and $\alpha = .9$.

to outperform the CVaR and SMCR risk measures. This is accentuated by our choice of "heavy tailed" stock data which takes advantage of risk measures that are based on tail moments of loss distribution.

As it was noted earlier, the infeasibility of the MIpOCP problem lies with the budget and initial expected return constraint. We saw that as the initial expected return, r_0, increased, the branch-and-bound problem had some feasibility issues. In order to correct this, we would have to increase the accuracy of our approximation in order to ensure that the problem has a feasible solution. If, however, there are no feasibility issues, we saw that the HMCR3-Int problem formulation led to the best performing portfolio. This can be attributed to the fact that limiting the portfolio to fewer instruments effectively increases the risk and thus the reward for the portfolio.

CHAPTER 7
CONCLUSIONS AND FUTURE WORK

The goal of this research was to develop an efficient algorithmic method for solving the p-order conic programming problem. This was motivated by the implementation of higher moment coherent risk measures in stochastic programming problems. The mathematical representation of HMCR measures in stochastic programming led to a linear programming problem with a p-order conic constraint. Given the presence of a p-order conic constraint in a stochastic programming problem, it was beneficial to consider a linear approximation to the p-order cone in order to reduce our pOCP problem to a linear programming problem. One of the major justifications for seeking such a representation was to take advantage of the "warm start" capabilities of linear programming solvers, which allows for quicker solutions to multistage stochastic programming problems. Motivated by the need to solve the pOCP problem efficiently, we considered different approximations for the p-order cone to see if we could improve the solution time. Also, given the importance of integrality in practical applications, we also considered the mixed integer p-order conic programming problem (MIpOCP).

During the course of this endeavor, we showed that the pOCP problem can be reformulated as a SOCP problem with $O(J \log r)$ second order cones. We also saw that there existed an eloquent mathematical reformulation of the SOCP problem using the "lifted" polyhedral approximation developed by Ben-Tal and Nemirovski (2001b). Although the "lifted" polyhedral approximation is a very efficient linear approximation with excellent approximation error, it was shown that the practical merits of the approximation did not help with the solution time of SOCP problems. It was shown that, despite the efficiency of the approximation, the current interior point SOCP solvers performed just as well due, in part, to the self-duality of the second order cone. Another issues was the inability to extend the "lifted" polyhedral approximation to values of $p > 2$.This led to the development of our uniform gradient approximation.

The gradient approximation was shown to be a very good approximation if the number of subdivisions were large enough. As the number of subdivisions increased, so did the size of the resulting linear programming problem. This motivated the development of a cutting plane algorithm to generate the constraints for the facets of the linear approximation as needed. Given the special structure of the uniform gradient approximation, we were able to generate the cuts in $O(J)$ time. This meant that, by exploiting this special structure, we could use a subdivision of any size while the time to generate the cuts would remain constant.

Another aspect of the pOCP problem that was considered was the incorporation of integrality constraints. This led to the development of another algorithm for the mixed integer p-order conic programming problem (MIpOCP). The algorithm that was developed was an adaptation of a branch-and-bound algorithm that Vielma, Ahmed, and Nemhauser (2008) developed for solving mixed integer second order conic programming problems (MISOCP). We employed our polyhedral approximation to represent the mixed integer linear programming problem ($MILPP_{GA}$) that was used in the branch-and-bound algorithm. The $MILPP_{GA}$ is used for pruning in the branch-and-bound algorithm and as such, a tight approximation is not needed. The MISOCP reformulation was compared to the MIpOCP with branch-and-bound to determine the efficiency of the algorithm.

The numerical experiments for the pOCP problem indicated that the cutting plane algorithm offered an efficient alternative to solving the pOCP problem. This could be seen clearly in the large difference in the solution times in the portfolio optimization case study. It was also noted that the solution times decreased as $p \to \infty$. This was attributed to the fact that at $p = \infty$ the problem becomes a linear programming problem and it is no longer an approximation but a reformulation. The MIpOCP numerical experiments also showed a marked difference in solution time between the MISOCP reformulation and the MIpOCP with branch-and-bound. This could be attributed to the effectiveness of the branch-and-bound algorithm and its ability to branch on integer solutions.

When comparing the portfolio performance for risk measures $HMCR_{3,\alpha}(\cdot)$, $SMCR_{\alpha}(\cdot)$ and $CVaR_{\alpha}(\cdot)$ we saw that the $HMCR_{3,\alpha}(\cdot)$ risk measure dominated the others most of the time. We also saw that integrality, along with $HMCR_{3,\alpha}(\cdot)$, led to a less diversified portfolio, based on the number of stocks that we limited the portfolio to, and this led to greater returns over the long run when compared to all the other portfolios. The data was chosen to reflect the most up to date closing stock price information that was available. It included the current economic meltdown in order to see how the self-balancing scheme would work to correct itself. Based on the data in Chapter 6, we see that the risk measures that are based on higher moments of tail loss generally outperform the current industry standard CVaR and SMCR risk measures. This was accentuated by our choice of data in which we used stocks from the S&P 500 that had the highest kurtosis.

Overall, the algorithms that were developed showed marked improvements in their solution time when compared with trying to solve the problem through direct implementation. We also saw that the overall performance of the HMCR risk measure outperformed the SMCR and CVaR risk measures when dealing with the "heavy tailed" stock data. This can be attributed to the definition of the HMCR risk measure as a quantification of risk as tail moments of loss distribution. The underlying methodology that was employed was to solve the problem in stages and exploit the special structure that is inherent in the formulation of the pOCP problem.

It was our intention to incorporate the cutting plane algorithm, along with the branch-and-bound algorithm, to develop an efficient MIpOCP solver. There were, however, limitations in the CPLEX C++ API that did not allow us to employ the cutting plane algorithm at each node of the branch-and-bound algorithm. It is our intention to investigate this further to see if there are any other options to incorporate this idea in the future.

BIBLIOGRAPHY

Acerbi, C. (2002) "Spectral measures of risk: A coherent representation of subjective risk aversion," *Journal of Banking and Finance*, **26** (7), 1487–1503.

Acerbi, C. and Tasche, D. (2002) "On the coherence of expected shortfall," *Journal of Banking and Finance*, **26** (7), 1487–1503.

Alizadeh, F. and Goldfarb, D. (2003) "Second-order cone programming," *Mathematical Programming*, **95** (1), 3–51.

Andersen, E. D., Roos, C., and Terlaky, T. (2003) "On implementing a primal-dual interior-point method for conic quadratic optimization," *Mathematical Programming*, **95** (2), 249–277.

Artzner, P., Delbaen, F., Eber, J.-M., and Heath, D. (1999) "Coherent Measures of Risk," *Mathematical Finance*, **9** (3), 203–228.

Bawa, V. S. (1975) "Optimal Rules For Ordering Uncertain Prospects," *Review of Financial Studies*, **2** (1), 95–121.

Bazaraa, M. S., Sherali, H. D., and Shetty, C. M. (2006) *Nonlinear Programming: Theory and Algorithms*.

Ben-Tal, A. and Nemirovski, A. (2001a) *Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications*, volume 2.

Ben-Tal, A. and Nemirovski, A. (2001b) "On polyhedral approximations of the second-order cone," *Mathematics of Operations Research*, **26** (2).

Ben-Tal, A. and Teboulle, M. (1986) "Expected Utility, Penalty Functions, and Duality in Stochastic Nonlinear Programming," *Management Science*, **32** (11), 1145–1466.

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.

Birge, J. R. and Louveaux, F. (1997) *Introduction to Stochastic Programming*.

Burer, S. and Chen, J. (2008) "A p-Cone Sequential Relaxation Procedure for 0-1 Integer Programs," *Working paper*.

Chekhlov, A., Uryasev, S., and Zabarankin, M. (2005) "Drawdown Measure in Portfolio Optimization," *International Journal of Theoretical and Applied Finance*, **8** (1), 13–58.

Cheridito, P. and Li, T. (2007) "Risk Measures on Orlicz Hearts," *Working paper*.

Delbaen, F. (2002) "Coherent risk measures on general probability spaces," 1–37.

Dembo, R. and Rosen, D. (1999) "The practice of portfolio replication. A practical overview of forward and inverse problems," *Annals of Operations Research*, **85**, 267–284.

Dentcheva, D. and Ruszczyński, A. (2004) "Semi-Infinite Probabilistic Optimization: First Order Stochastic Dominance Constraints," *Optimization*, **53** (5–6), 583–601.

Duffie, D. and Pan, J. (1997) "An Overview of Value-at-Risk," *Journal of Derivatives*, **4**, 7–49.

Fischer, T. (2003) "Risk capital allocation by coherent risk measures based on one-sided moments," *Insurance: Mathematics and Economics*, **32** (1), 135–146.

Fishburn, P. C. (1977) "Mean-risk analysis with risk associated with below-target returns," *The American Economic Review*, **67** (2), 116–126.

Giorgi, E. D. (2005) "Reward-Risk Portfolio Selection and Stochastic Dominance," *Journal of Banking and Finance*, **29** (4), 895–926.

Glineur, F. (2000) "Computational experiments with a linear approximation of second order cone optimization," *Technical Report 0001*.

Glineur, F. and Terlaky, T. (2004) "Conic Formulation for l_p-Norm Optimization," *Journal of Optimization Theory and Applications*, **122** (2), 285–307.

Jorion, P. (1997) *Value at Risk: The New Benchmark for Controlling Market Risk*.

Krokhmal, P. (2007) "Higher Moment Coherent Risk Measures," 21.

Krokhmal, P., Palmquist, J., and Uryasev, S. (2002a) "Portfolio Optimization with Conditional Value-At-Risk Objective and Constraints," *Journal of Risk*, **4** (2), 43–68.

Krokhmal, P., Uryasev, S., and Zrazhevsky, G. (2002b) "Risk Management for Hedge Fund Portfolios: A Comparative Analysis of Linear Rebalancing Strategies," *Journal of Alternative Investments*, **5** (1), 10–29.

Markowitz, H. (1952) "Portfolio Selection," *Journal of Finance*, **7** (1), 77–91.

Markowitz, H. M. (1959) *Portfolio Selection*.

Morgan, J. (1994) *Riskmetrics*.

Nesterov, Y. (2006) "Towards nonsymmetric conic optimization," Working paper.

Nesterov, Y. E. and Nemirovski, A. (1994) *Interior Point Polynomial Algorithms in Convex Programming*, volume 13.

Nesterov, Y. E. and Todd, M. J. (1997) "Self-scaled barriers and interior-point methods for self-scaled cones," *Mathematics of Operations Research*, **22**, 1–42.

Nesterov, Y. E. and Todd, M. J. (1998) "Primal-dual interior-point methods for self-scaled cones," *SIAM Journal on Optimization*, **8**, 324–364.

Ogryczak, W. and Ruszczyński, A. (1999) "From stochastic dominance to mean-risk models: Semideviations as risk measures," *European Journal of Operational Research*, **116**, 33–50.

Ogryczak, W. and Ruszczyński, A. (2001) "On consistency of stochastic dominance and mean–semideviation models," *Mathematical Programming*, **89**, 217–232.

Ogryczak, W. and Ruszczyński, A. (2002) "Dual stochastic dominance and related mean-risk models," *SIAM Journal on Optimization*, **13** (1), 60–78.

Pflug, G. (2000) "Some Remarks on the Value-at-Risk and the Conditional Value-at-Risk," .

Prékopa, A. (1995) *Stochastic Programming*.

Rockafellar, R. and Uryasev, S. (2002a) "Conditional Value-at-Risk for General Loss Distribution," *Journal of Banking and Finance*, **26**, 1443–1471.

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. (2002b) "Conditional Value-at-Risk for General Loss Distributions," *Journal of Banking and Finance*, **26** (7), 1443–1471.

Rockafellar, R. T., Uryasev, S., and Zabarankin, M. (2006) "Generalized Deviations in Risk Analysis," *Finance and Stochastics*, **10** (1), 51–74.

Rothschild, M. and Stiglitz, J. (1970) "Increasing risk I: a definition," *Journal of Economic Theory*, **2** (3), 225–243.

Ruszczyński, A. and Shapiro, A. (2006) "Optimization of Convex Risk Functions," *Mathematics of Operations Research*, **31** (3), 433–452.

Schied, A. and Follmer, H. (2002) "Robust preferences and convex measures of risk," 39–56.

Sturm, J. F. (1998) "Using SeDuMi 1.0x, a MATLAB toolbox for optimization over symmetric cones," *Manuscript*.

Terlaky, T. (1985) "On l_p Programming," *European Journal of Operational Research*, **22** (1), 70–100.

Testuri, C. and Uryasev, S. (2003) "On Relation Between Expected Regret and Conditional Value-at-Risk," .

van der Vlerk, M. H. (2003) "Integrated Chance Constraints in an ALM Model for Pension Funds," *Working paper*.

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 Journal on Computing*, **20** (3), 438–450.

von Neumann, J. and Morgenstern, O. (1944) *Theory of Games and Economic Behavior*.

Xue, G. and Ye, Y. (2000) "An efficient algorithm for minimizing a sum of p-norms," *SIAM Journal on Optimization*, **10** (2), 551–579.