Imperial College London

Department of Computing

Portfolio Optimization as Stochastic Programming with Polynomial Decision Rules

by

Miha Troha

Supervisor: Dr. Daniel Kuhn

Second Marker: Professor Berç Rustem

Submitted in partial fulfilment of the requirements for the MSc Degree in Computing Science (Computational Management) of Imperial College London

September 2011

Abstract

Multistage stochastic programming provides a general framework for modelling real-life decision problems that involve uncertainty. However, it is computationally very demanding, even when only medium-accuracy solutions are required. Different approximation methods have been proposed in order to tackle multistage stochastic programming problems. In this work, we analyze two, namely, scenario tree approximation and decision rule approximations. The latest attracted most of the attention in the recent years. The main idea is to limit the characterization of the decisions from all measurable functions to only certain functional forms. We focus on the most recent polynomial decision rules, which were introduced in the stochastic programming framework by Bampou [7], where decisions are modeled as polynomial functions of the uncertain parameters. We propose two extensions of this work. We release the assumption that polynomial decision rules must be characterized by even degrees polynomial functions and extend it to polynomial functions of all degrees. Moreover, we release the assumption that the recourse matrix does not depend on the uncertain parameters and instead model recourse matrix by polynomial functions of the uncertain parameters.

The last extension is needed in order to tackle a special case of multistage stochastic programming problems termed portfolio optimization. Portfolio optimization is the problem of allocating capital over different assets in order to maximize the return on the investment and at the same time minimize its risk. In the stochastic programming framework, one wants to maximize the expected return of the investment, while minimizing a specific characterization of risk. The first portfolio optimization problem was introduced by Markowitz, where the variance of the return was used to characterize risk. However, variance is not a good risk measure, because it penalizes both profits as well as losses. This contradicts the reality, where investors want to minimize only the possibility of losses. In this work we present general properties, i.e. coherency proposed by Artzner and time consistency, that a good risk measure must satisfy. We prove that one of the most popular risk measures termed conditional value at risk violates time consistency and propose its time consistent alternative.

Acknowledgements

First of all, I would like to acknowledge my supervisor, Dr. Daniel Kuhn, for his great guidance and invigorating attitude through out the project. His insightful comments and creative ideas contributed a great value to this work. Moreover, I would like to thank Dimitra Bampou, a PhD student in his research group, for all the inspiring and sincere conversations, detailed explanations, and exhaustive feedback on the report. I could not have wished for a better team of supervisors.

I also acknowledge Professor Berç Rustem for his constructive comments and accepting to be the second marker for this project.

I would like to thank my parents and my sister Nina for their constant support and encouragement. Finally, I would like to thank Ana for her interest in all the challenges and problems of my thesis.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Optimization is concerned with the problem of optimizing, i.e, minimizing or maximizing, an objective function subject to some given constraints. Under one of many different classifications, optimization problems can be divided into deterministic optimization problems and optimization problems under uncertainty.

As the name suggests deterministic optimization problems deal with deterministic variables. Consider for example a company that produces two products, X and Y, at known costs. Each of these products is produced in a number of steps in which different machines are used. Some machines might be used for the production of both products. The company has some materials needed for the production of the products on stock. Imagine that the company just obtained a new customer that demanded more X and Y products than the company is able to produce. The company would like to invest some of its savings to meet at least some of the new demand. How much should the company invest into new machines, how much in the working labour and how much in the new materials? Which machines should be upgraded or replaced? What effect does the upgrade of a certain machine have on the overall profit? These kind of problems can be solved with deterministic optimization techniques. The objective function represents the profit, i.e., the difference between the income from selling the products and all the costs and investments needed to produce the products. The constraints of this problem are the capacity of the machines, the available working labour, money to invest, time, materials etc.

Now consider a similar problem, where the demand for the products is not completely known in advance. Only some estimates are available. Also machines could break down and workers might get ill. Materials are not always delivered at the preagreed time. Such problems clearly involve uncertain parameters and therefore belong to the group of the optimization problems under uncertainty. In such problems, one would like to maximize the expectation of the profit. Optimization problems under uncertainty that involve expectation in the objective function are termed stochastic programming problems. Similarly, also other, more risk averse objective functions are possible. In a special case, one could want to maximize the profit under the worst-case realisation of the uncertain parameters. This kind of optimization problems under uncertainty are termed robust optimization problems. In the above problem not all the uncertain parameters are revealed at once. They are revealed sequentially and after each revelation some decisions, as reactions to the observed uncertain parameters, are made. These kind of problems are termed multistage optimization problems.

In order to grasp optimization problems, the classification to deterministic and nondeterministic optimization problems is not enough. Clearly, in both groups there are some very difficult and some less difficult settings. Less difficult settings are considered those, where the optimal solution can be found in polynomial time. In reality, this means that problems that involve millions of variables and constraints can be solved efficiently. Deterministic linear programming problems, where the objective function and the constraints are all linear functions of the decision parameters is one of such problems. Another example is semidefinite programming, where a linear objective function is optimized over the intersection of an affine space with the cone of positive semidefinite matrices.

On the other hand, linear optimization problems under uncertainty are much more difficult.

Even linear two-stage stochastic programming problems are #P-hard [6]. Moreover, as claimed by Saphiro and Nemirovski [11], multistage stochastic problems are generally computationally intractable even when only medium-accuracy solutions are required. Different approximation methods have been proposed in order to tackle them. The most intuitive method is scenario tree approximation. It involves discretizing outcomes of the uncertain parameters in each stage. This approach approximates the computationally intractable multistage stochastic programming problem by a tractable linear programming problem. However, the complexity of the linear programming problem grows exponentially with the number of stages of the original stochastic programming problem [14]. Therefore another approach, termed decision rule approximations, has been proposed by Ben-Tal [19]. The main idea is to keep the distribution of the uncertain parameters unchanged, but only limit the characterization of the decisions to certain functional forms. By such approach, an upper bound of the optimal solution is obtained. Linear [11] and piecewise linear [29] functional forms at first attracted most of the attention.

By considering only specific functional forms of decision rules, the approximated solutions may be very suboptimal. In order to estimate suboptimality of solutions, Kuhn [3] proposed a tractable lower bound approximation by approximating the functional form of the dual decision rules as linear functions of the uncertain parameters. Georghiou [30] applied a similar approach also in the context of piecewise linear decision rules. Both the upper and the lower bound approximations are written as tractable linear programs, but the suboptimality of solutions make them unuseful for many problems.

In order to improve accuracy of solutions, polynomial decision rules have been proposed, first in the robust optimization framework by Bertsimas [31], and then applied to stochastic programming problems by Bampou [7]. However, polynomial decision rule approximations do not lead to a tractable linear program, but instead to an intractable semi-infinite programming problem, having finitely many decision variables, but infinitely many constraints. Constraints involve checking non-negativity of a polynomial on a compact bounded semi-algebraic set. Only in the recent years, it has been shown that such problems can be approximated (sometimes even solved precisely) by a tractable semidefinite programming problem [2, 23]. Bampou showed that even polynomial decision rules of small degrees outperform the piecewise linear decision rules. However, her formulation is only applicable for polynomial decision rules of even degrees and problems where the recourse matrix does not depend on the uncertain parameters. In this work we approximate the primal and the dual stochastic programming problems with polynomial decision rules, while releasing Bampou's assumptions. The formulation holds for polynomial decision rules of all degrees and the recourse matrix is modeled by polynomial functions of the uncertain parameters.

Our extension was needed in order to tackle a special case of multistage stochastic programming problems termed portfolio optimization. Portfolio optimization is the problem of allocating capital over different assets in order to maximize the return on the investment and at the same time minimize its risk [48]. Since portfolio returns are uncertain, one usually wants to maximize the expected return of the investment. Characterization of the risk is a more difficult problem. The first portfolio optimization problem was introduced by Markowitz [34], where the variance of the return was used to characterize risk. However, the main disadvantage of the variance as a risk measure is that it penalizes both profits as well as losses, since it is a measure of the dispersion of the values of the random variable around its expected value. In reality, investors want to minimize only the possibility of losses. Artzner [32] therefore proposed some general properties that a good risk measure must satisfy. Rockafellar [33] showed that a measure termed conditional value at risk satisfies those properties for one period portfolio optimization. However, when applied to multistage portfolio optimization problems, it violates an important property termed time consistency [39]. In this work we propose a time consistent version of the multistage conditional value at risk.

This work develops as follows. In Chapter 2 we present some basic concepts that were used throughout this work. We introduce some deterministic programming approaches, measure theory and optimization of polynomials. We start Chapter 3 by introducing stochastic programming problems and later show two approximation approaches used in order to tackle them. The first is scenario tree approximation and the second is decision rule approximations. In Chapter 4 we analyze portfolio optimization problems as a special case of stochastic programming problems. We examine different risk measures and construct a time consistent version of the multistage conditional value at risk. In Chapter 5 we present numerical results on two special stochastic programming problems. The first is electricity capacity expansion and the second is multistage portfolio optimization. This work is concluded in Chapter 6, where we also give some directions for the future work.

1.1 Contributions

The main contributions of this work are:

- We investigate polynomial optimization techniques proposed by Parrilo [2], Putinar [1], Lasserre [23], Schmüdgen [27] etc. and analyze the assumptions in order to determine the context, in which they can be applied.

- We investigate decision rule approximations, where most of the work is focused on the most recent polynomial decision rules proposed by Bampou [7]. Her formulation was only applicable for even degrees polynomial decision rules and problems where the recourse matrix does not depend on the uncertain parameters. We release Bampou's assumptions and formulate tractable upper and lower bound approximations, which hold for polynomial decision rules of all degrees while the recourse matrix is modeled by polynomial functions of the uncertain parameters.

- We critically assess and compare two alternatives, namely scenario tree approximation and decision rule approximations, as means of approximating multistage stochastic programming problems.

- We apply decision rule approximations on the electricity capacity expansion problem and show that they outperform the piecewise linear decision rules.

- We investigate single-stage and multistage portfolio optimization problems and the corresponding risk measures. We describe coherent [32] and time consistent [36] risk measures. Moreover, we show that multistage conditional value at risk is not time consistent and propose its time consistent alternative.

- We present that, under our extension in the second point, optimal solutions of portfolio optimization problems can be approximated by polynomial decision rule approximations. Moreover, we give some guidelines for this approximation.

- We evaluate multistage time consistent risk measure proposed in the fourth point relative to the existing single-stage alternatives.

Chapter 2

Important concepts

In this chapter general concepts that are needed to understand this work are presented. We first examine some deterministic optimization problems [24], continue with the measure theory, techniques for optimization over polynomials and finish with the function approximation theory.

2.1 Linear Programming

A Linear Programming (LP) problem is defined as a problem of maximizing or minimizing a linear function subject to linear equality and inequality constraints. If n represents the number of decision variables and m the number of constraints, then the standard LP problem is defined by

$$\begin{aligned} \underset{x}{min} \quad & c^\top x \\ s.t. \quad & Ax \geq b \end{aligned} \tag{2.1}$$

where $c \in \mathbb{R}^n$, $x \in \mathbb{R}^n$, $b \in \mathbb{R}^m$, and $A \in \mathbb{R}^{m \times n}$.

The LP is called *feasible*, if its *feasible set* $F = \{x | Ax - b \geq 0\}$ is nonempty. A point $x \in F$ is then called *a feasible solution*. The LP is *bounded below*, if it is either infeasible, or its objective function $c^\top x$ is bounded from below on F. For a feasible bounded from below LP, the quantities $c^* \equiv \underset{x \in F}{inf}\, c^\top x$ and $x^* \equiv \underset{x \in F}{argmin}\, c^\top x$ are the *optimal value* and its corresponding *optimal solution* of the problem, respectively.

The problem defined above is called *primal*. Every LP has also its *dual* formulation, which is defined by

$$\begin{aligned} \underset{y}{max} \quad & b^\top y \\ s.t. \quad & A^\top y = c \\ & y \geq 0 \end{aligned} \tag{2.2}$$

where $c \in \mathbb{R}^n$, $y \in \mathbb{R}^m$, $b \in \mathbb{R}^m$, $A \in \mathbb{R}^{m \times n}$.
Two theorems connect the optimal values of the primal and the dual problem. Weak duality theorem states that $b^\top y \leq c^\top x$ and strong duality theorem states that $b^\top y^* = c^\top x^*$.

Many efficient approaches (e.g. Simplex algorithm [43], interior point methods [44]...) have been proposed to solve LP problems. Solutions of LP problems can be found in a polynomial time.

2.2 Conic Programming

A significant part of the nice features from the LP originates from the properties of the inequality. For the LP the inequality $Ax \geq b$ is defined as a comparison of the "coordinate-wise" vector elements. It has the following properties:

- Reflexivity: $a \geq a$;
- Anti-symmetry: if $a \geq b$ and $b \geq a$, then $a = b$;
- Transitivity: if $a \geq b$ and $b \geq c$, then $a \geq c$;
- Compatibility with linear operations:
  - Homogeneity: if $a \geq b$ and λ is a non-negative real number, then $\lambda a \geq \lambda b$;
  - Additivity: if $a \geq b$ and $c \geq d$, then $a + c \geq b + d$

The coordinate-wise inequality is not the only definition of the "inequality" that fits the axioms above. It is possible to define a generic optimization problem that looks exactly the same as an LP, where the inequality is replaced with a different ordering. Specifying properly the ordering of vectors, one can obtain generic optimization problems covering many important applications which cannot be treated by the standard LP.

It is possible to show that a set K that satisfies the above axioms must be a pointed convex cone, i.e. satisfy the following conditions:

1. Pointed: If $a \in K$ and $-a \in K$, then $a = 0$.
2. Nonempty and closed under addition: If $a, a' \in K$, then $a + a' \in K$.
3. Conic set: If $a \in K$ and any $\lambda \geq 0$, then $\lambda a \in K$.

The partial ordering induced by this cone is denoted by $\geq_K$. There are many cones that satisfy the above conditions. LP can be understood as a special case where $K = \mathbb{R}^m_+$.

Let K be a pointed, closed, convex cone with nonempty interior on a set E. A general conic program (CP) is then defined as

$$\begin{aligned} &\underset{x}{min} \quad c^\top x \\ &s.t. \quad Ax \geq_K b \end{aligned} \tag{2.3}$$

where $c \in \mathbb{R}^n$, $b \in E$, and x is a linear mapping $x \mapsto Ax : \mathbb{R}^n \to E$.

Define the dual cone K^* as

$$K^* = \{\lambda \in E : \langle \lambda, a \rangle \geq 0 \; \forall a \in K\} \tag{2.4}$$

and conjugate operator A^* of the linear mapping x

$$\langle y, Ax \rangle = \langle A^* y, x \rangle \quad \forall (y \in E, x \in \mathbb{R}^n). \tag{2.5}$$

In order to derive the dual of problem 2.3 the Lagrangian duality [45] can be applied. We define the Lagrangian function as

$$L(x, \lambda) = c^\top x - \lambda^\top (Ax - b) \tag{2.6}$$

where $\lambda \in K^*$. Consider now the following function

$$g(\lambda) = \underset{x}{inf}\, L(x, \lambda) = \underset{x}{inf} \left(c^\top x - \lambda^\top (Ax - b) \right) \tag{2.7}$$

and the optimization problem

$$\begin{aligned} &\underset{\lambda}{max} \quad g(\lambda) \\ &s.t. \quad \lambda \geq_{K^*} 0. \end{aligned} \tag{2.8}$$

Due to the definition of the dual cone 2.4 $\langle \lambda, (Ax - b) \rangle \geq 0$. Thus, the optimal solution of problem 2.8 clearly provides a lower bound for the optimal solution of problem 2.3. By using the conjugate operator A^* as defined by 2.4, problem 2.8 can be rewritten as

$$\begin{aligned} \underset{\lambda}{max} \quad & b^\top \lambda \\ s.t. \quad & A^* \lambda = c \\ & \lambda \geq_{K^*} 0 \end{aligned} \tag{2.9}$$

where $\lambda \in K^*$. Problem 2.9 represents the dual of problem 2.3.

2.3 Semidefinite Programming

Semidefinite Programming (SDP) is a special case of the CP, where the ordering is defined on the semidefinite cone $K = \mathbb{S}^m_+$ in the space $E = \mathbb{S}^m$ of $m \times m$ symmetric matrices. The primal is given by

$$\begin{aligned} \underset{X}{min} \quad & \langle C, X \rangle \\ s.t. \quad & \langle A_i, X \rangle = b_i \\ & X \succeq 0 \end{aligned} \tag{2.10}$$

where $X \in \mathbb{S}^n$ is a decision variable, $b \in \mathbb{R}^m$ and $C,\ A_i \in \mathbb{S}^n$ are given symmetric matrices.

Since SDP is convex, it is also possible to define the dual, which is given by

$$\begin{aligned} \underset{y}{max} \quad & \langle b, y \rangle \\ s.t. \quad & \textstyle\sum_{i=1}^m y_i A_i \preceq C \end{aligned} \tag{2.11}$$

where $y \in \mathbb{R}^m$.

Weak and strong duality theorems exist also for the SDP. The optimal value of the primal is always bigger than or equal to the optimal value of the dual. Strong duality holds if the primal (dual) problem is bounded from below (above) and strictly feasible. Equality then holds when both solutions are optimal.

Most algorithms for solving SDP are based on the interior point methods. Solutions of SDP problems can also be found in polynomial time [19].

2.4 Sets

Sets are an important concept that we deal a lot with in this work. In this section, we review all the definitions needed.

Definition 2.1: A set $A \subseteq \mathbb{R}^n$ is *open* if, for all points $a \in A$, there exists an ϵ-neighbourhood $V_\epsilon(a) \subseteq A$.

Example 2.1: An example of an open set in $\mathbb{R}$ is $(1, 2)$. ■

Definition 2.2: The point x is a *limit point* of a set A, if every ϵ-neighbourhood $V_\epsilon(x)$ of x intersects the set A in some point other then x.

Example 2.2: For the set $(1, 2)$ in $\mathbb{R}$ the limit points are 1 and 2. ■

Definition 2.3: A set $A \subseteq \mathbb{R}^n$ is *closed* if it contains its limit points.

Example 2.3: An example of an closed set in $\mathbb{R}$ is $[1, 2]$. ■

Definition 2.4: The *complement* of a set $A \subseteq \mathbb{R}^n$ is defined as the set

$$A^C = \{x \in \mathbb{R}^n \,|x \notin A\}\,. \tag{2.12}$$

Example 2.4: The complement of $(1, 2)$ in $\mathbb{R}$ is $(-\infty, 1] \cup [2, \infty)$. ■

Definition 2.5: A set $A \subseteq \mathbb{R}^n$ is *compact* if every sequence in A has a convergent subsequence that converges to a limit in A.

Example 2.5: All in $\mathbb{R}$:

- Set $[1, 2]$ is compact. Every sequence in this set must be bounded and thus have a convergent subsequence (due to the Bolzano Weierstrass theorem[1]). The set $[1, 2]$ is closed and hence it contains all limit points of the convergent subsequences in A.
- Set $[1, 2)$ is not compact, because, for example, the sequence $a_n = 2 - \frac{1}{n}$ as $n \to \infty$ converges to 2. Also every subsequence of it must converge to 2. Since 2 is not included in the set, it can not be compact. ■

Definition 2.6: A set $A \subseteq \mathbb{R}^n$ is *bounded* if there exists a vector $M^n > 0$ such that $|a_i| \leq M_i$ for all $a_i \in A$ and $i = 1, ..., n$.

Example 2.6: Set $(1, 2)$ in $\mathbb{R}$ is bounded for $M = 2$. ■

Definition 2.7: A subset of $\mathbb{R}^n$ is a *semi-algebraic set,* if it can be written as $\{\xi \in \mathbb{R}^n : f(\xi) > 0\}$ and $\{\xi \in \mathbb{R}^n : g(\xi) = 0\}$, where f and g are real polynomials in ξ.

Example 2.7: An example of a semi-algebraic set in $\mathbb{R}^2$ is $\left\{\xi \in \mathbb{R}^2 : f(\xi_1, \xi_2) = 1 - \xi_1^2 - \xi_2^2 > 0\right\}$. ■

2.5 Measure Theory

A measure can be understood as a generalization of the interval length in $\mathbb{R}$ or the area and the volume of subspaces in $\mathbb{R}^2$ and $\mathbb{R}^3$, respectively. The generalization is needed in order to enable the integration over arbitrary sets.

Example 2.8: The volume of an n-dimensional cuboid $Q = \{x \in \mathbb{R}^n : a_k \leq x_k \leq b_k,\ k = 1, ..., n\}$ is

$$\prod_{k=1}^{n} (b_k - a_k). \tag{2.13}$$

■

Algebra: Given a set S and a collection $\mathbb{S}$ of the subsets of S, $\mathbb{S}$ is an algebra of the subsets of S if

1. $S \in \mathbb{S}$
2. $\mathbb{S}$ is closed under the complementation: if $X \in \mathbb{S}$, then its complement $X^C \in \mathbb{S}$.
3. $\mathbb{S}$ is closed under the finite union: if $X_1, X_2 \in \mathbb{S}$, then $X_1 \cup X_2 \in \mathbb{S}$.

Note that due to points 2 and 3, $\mathbb{S}$ is also closed under the finite intersection since

$$X_1 \cup X_2 \in \mathbb{S} \Longrightarrow (X_1 \cup X_2)^C \in \mathbb{S} \Longrightarrow X_1^C \cap X_2^C \in \mathbb{S} \Longrightarrow X_1 \cap X_2 \in \mathbb{S}. \tag{2.14}$$

[1] More on the Bolzano Weierstrass theorem at http://home.iitk.ac.in/~psraj/mth101/lecture_notes/lecture3.pdf

σ-algebra: The σ-algebra $\mathbb{S}$ is a collection of the subsets of a set S that is closed under countable set operations, i.e., the complement of a member (or a subset) and the union or the intersection of countably many members. Formally, an algebra $\mathbb{S}$ of the subsets of a set S is a σ-algebra, if S contains the limit of every monotone sequence of its sets

$$X_1, X_2, ... \in \mathbb{S} \Longrightarrow \cup_{i=1}^{\infty} X_i \in \mathbb{S}. \tag{2.15}$$

The pair $(S, \mathbb{S})$ is then a *measurable space* and the sets in S are said to be *measurable.*

Borel algebra and Borel Sets: The Borel algebra of a set S is the minimal σ-algebra that contains all the open sets (or closed sets) on the real line. The elements of the Borel algebra are called Borel Sets. σ-algebras, specially Borel algebras, allow us to concentrate on certain important properties of sets and thus define the concept of a measure on seemingly arbitrary sets.

Measure space, measurable sets and measure: A function $\mu : \mathbb{S} \longrightarrow \mathbb{R} \cup +\infty$ defined on a σ-algebra $\mathbb{S}$ of the subsets of S is called a measure if:

- it is non-negative, i.e., $\mu(X) \geq 0$ for all $X \in \mathbb{S}$,
- $\mu(\emptyset) = 0$,
- and μ is countably additive, i.e.,
$$\mu(X) = \sum_{i=1}^{\infty} \mu(X_i), \tag{2.16}$$
where $X \in \mathbb{S}$; $X = \cup_{i=1}^{\infty} X_i$; $X_i \cap X_j = 0$.

The triplet $(S, \mathbb{S}, \mu)$ represents a *measure space*, the sets of $\mathbb{S}$ are called *measurable sets* and the function μ is called a *measure.*

Every measure μ satisfies the following properties:

- monotonicity: $\mu(X_1) \leq \mu(X_2)$ for all $X_1, X_2 \in \mathbb{S}$ and $X_1 \subset X_2$,
- it is continuous from below: if $X_i \in \mathbb{S}$, $i \in \mathbb{N}$ and $X_1 \subset X_2 \subset ...$, then $\mu(\cup_{i=1}^{\infty} X_i) = lim_{i \longrightarrow \infty} \mu(X_i)$.

If $\mu(S) = 1$, then the measure space is called a *probability space* and μ is a *probability measure.* The sets are termed *events.* One says that a property holds *almost everywhere,* if the set for which the property does not hold is a null set or a set with measure 0. In the probability theory, analogous to *almost everywhere*, *almost certain* or *almost sure* means except for an event of probability measure 0.

Example 2.9: An example of a probability measure is the Dirac measure δ_a. Let $(\mathbb{S}, S)$ be a measurable space and $a \in S$, then the Dirac measure δ_a is defined as

$$\delta_a(X) = \begin{cases} 1 & a \in X \\ 0 & a \notin X \end{cases} \tag{2.17}$$

for any measurable set $X \subseteq S$. The function δ_a is concentrated on the point a. In terms of probability, it represents the almost sure outcome a in the sample space S. ∎

2.6 Polynomial Optimisation

In this section, we outline two important approaches for finding a global optimum of a polynomial $p(\xi) : \mathbb{R}^k \longrightarrow \mathbb{R}$ over a compact semi-algebraic set. The first approach involves a sum of squares decomposition and the second is the problem of moments.

Notation: A polynomial $p(\xi)$ of a degree at most d in variables $\xi = [\xi_1, \xi_2, ..., \xi_k]$ is a finite linear combination of monomials, where the sum is taken over a finite number of k-tuples $\alpha = (\alpha_1, \alpha_2, ..., \alpha_k)$, $\alpha_i \in \mathbb{N}_0$.

$$p(\xi) = \sum_{\alpha \in L_d} p_\alpha \xi^\alpha = \sum_{\alpha \in L_d} p_\alpha \xi_1^{\alpha_1} \xi_2^{\alpha_2} ... \xi_k^{\alpha_k}, \;\; p_\alpha \in \mathbb{R}. \tag{2.18}$$

We define $|\alpha| = \sum_{i=1}^k \alpha_i$ and the set $L_d := \left\{\alpha \in \mathbb{N}_0^k : |\alpha| \le d\right\}$.

Let the set of all polynomials in $\xi_1, \xi_2, ..., \xi_k$ with real coefficients be $\mathbb{R}[\xi_1, \xi_2, ..., \xi_k]$. We denote by $\mathbb{R}_d[\xi]$ the set of polynomials of a degree at most d and by

$$\mathcal{B}_d(\xi) := \left[1, \xi_1, \xi_2, ..., \xi_k, \xi_1^2, \xi_1\xi_2, ..., \xi_1\xi_k, \xi_2^2, ..., \xi_k^d\right], \tag{2.19}$$

whose dimension is $s(k,d) = \binom{k+d}{d}$, its canonical basis.

Given a real-valued polynomial $p(\xi) : \mathbb{R}^k \longrightarrow \mathbb{R}$ we are interested in solving the following problem

$$p_\Xi^* \;\; = \;\; \min_{\xi \in \Xi} \;\; p(\xi). \tag{2.20}$$

Ξ is a compact semi-algebraic set defined by the polynomial inequalities

$$\Xi = \left\{\xi \in \mathbb{R}^k : w_r(\xi) \ge 0, \; r = 1, ..., R\right\}, \tag{2.21}$$

where each of the polynomials $w_r(\xi)$ is of a degree d_r.

For the further argumentation we define

$$\tilde{d_r} = \left\lfloor \frac{d - d_r}{2} \right\rfloor, \tag{2.22}$$

and the polynomial

$$\Sigma'(\Xi) = \sum \mathbb{R}[\xi]^2 + \sum_{r=1}^R w_r \sum \mathbb{R}[\xi]^2. \tag{2.23}$$

If we set $w_0 = 1$ only to simplify notation, then

$$\Sigma'(\Xi) = \sum_{r=0}^R w_r \sum \mathbb{R}[\xi]^2. \tag{2.24}$$

Moreover, we define the polynomial $\Sigma(\Xi)$ by

$$\begin{aligned} \Sigma(\Xi) \;\; &= \;\; \textstyle\sum \mathbb{R}[\xi]^2 + \sum_{r=1}^R w_r \sum \mathbb{R}[\xi]^2 + \sum_{r_1=1}^R \sum_{r_2=1}^R w_{r_1} w_{r_2} \sum \mathbb{R}[\xi]^2 + ... + w_1 w_2 ... w_R \sum \mathbb{R}[\xi]^2 \\ &= \;\; \textstyle\Sigma'(\Xi) + \sum_{r_1=1}^R \sum_{r_2=1}^R w_{r_1} w_{r_2} \sum \mathbb{R}[\xi]^2 + ... + w_1 w_2 ... w_R \sum \mathbb{R}[\xi]^2. \end{aligned} \tag{2.25}$$

An important concept in the polynomial optimization are sum of squares (SOS) polynomials. An SOS polynomial is every polynomial $p(\xi)$ that has a sum of squares decomposition

$$p(\xi) = \sum \mathbb{R}[\xi]^2 = \left\{p \in \mathbb{R}[\xi]\,;\; p = \sum_{i=1}^g p_i^2, \; p_i \in \mathbb{R}[\xi] \;\; for\ some\ g \in \mathbb{N}\right\}. \tag{2.26}$$

Let $\mathcal{P}_{k,2d}(\Xi)$ ($\mathcal{P}_{k,2d}$) denote the cone of polynomials of degree $2d$ in k variables, that are non-negative on Ξ (globally non-negative). Similarly, let $\Sigma_{k,2d}(\Xi)$ ($\Sigma_{k,2d}$) denote the cone of polynomials of degree $2d$ in k variables, that are non-negative on Ξ (globally non-negative) and have an SOS decomposition.

2.6.1 Sum of squares decomposition

Globally non-negative polynomials

It is clear that SOS polynomials are non-negative. Is it possible to claim also that every non-negative polynomial has an SOS decomposition?

Proposition 2.1: Let $p \in \mathbb{R}[\xi]$, $\xi \in \mathbb{R}$. Then $p(\xi) \geq 0$, $\forall \xi \in \mathbb{R}$ if and only if $p \in \Sigma_{1,2d}$ for some $d \in \mathbb{N}$.

Proof:

1. $\Longleftarrow$: It is clear that every SOS is non-negative.

2. $\Longrightarrow$: If $p(\xi) \geq 0$, $\forall \xi \in \mathbb{R}$ then all its real roots are of even degree, because otherwise $p(\xi)$ would have a different sign when approaching the root from the left and from the right. We denote with λ_i each of the roots of the degree n_i, $i = 1, ..., l$. Complex roots of $p(\xi)$ can be arranged in the conjugate pairs $a_j + ib_j$ and $a_j - ib_j$, $j = 1, ..., h$. Since the complex roots are in pairs and the real roots are of an even degree, it is clear that $p(\xi)$ must be of an even degree.

$$\begin{aligned}
p(\xi) &= K \textstyle\prod_{i=1}^{l} (\xi - \lambda_i)^{2n_i} \prod_{j=1}^{h} ((\xi - a_j)^2 + b_j^2) \\
&= K \textstyle\prod_{i=1}^{l} \prod_{j=1}^{h} \left(\left(\underbrace{(\xi - \lambda_i)^{n_i} (\xi - a_j)}_{v_{ij}(\xi)} \right)^2 + \left(\underbrace{(\xi - \lambda_i)^{n_i} b_j}_{u_{ij}(\xi)} \right)^2 \right) \\
&= K \textstyle\prod_{i=1}^{l} \prod_{j=1}^{h} \left(v_{ij}(\xi)^2 + u_{ij}(\xi)^2 \right)
\end{aligned} \tag{2.27}$$

Note that the expression in the last line is an SOS polynomial since $K \geq 0$ and products of the SOS polynomials are also SOS polynomials. ■

If one tries to generalize the above proposition to multivariate polynomials, he sees that it does not hold. Hilbert investigated it more in details and provided the following theorem.

Theorem 2.2: $\Sigma_{k,2d} \subseteq \mathcal{P}_{k,2d}$ holds with equality only in the following cases:

- Bivariate forms: $k = 2$
- Quadratic forms: $d = 1$
- Ternary quadratic forms: $k = 3$ and $d = 2$

Until now, we have not discussed why are we interested in the SOS decomposition. The main reason is that checking the global non-negativity of a polynomial is hard (for polynomials of degree 4 is NP-hard [8]). On the other hand, in [25] it was shown that checking whether a polynomial has a SOS decomposition is equivalent to solving an SDP.

Theorem 2.3: The existence of a SOS decomposition of a polynomial in k variables of a degree $2d$ can be decided by solving a semidefinite programming feasibility problem, where the dimensions of the matrix inequality are $s(k, d) \times s(k, d)$.

Proof: [25]. ■

The reasoning behind this theorem is that if $f(\xi)$ is a polynomial of degree $2d$ and has a SOS decomposition, it can be written as

$$f(\xi) = \mathcal{B}_d(\xi)^\top Q \mathcal{B}_d(\xi),$$

where Q is a constant and positive semidefinite matrix.

We have now analyzed the relationships between a global polynomial non-negativity and a SOS decomposition. We will continue with checking non-negativity of polynomials on general semi-algebraic sets.

Non-negativity of polynomials on semi-algebraic sets

The central theorem that can be used for this purpose is the Stengles's Positivstellensatz. For our purpose we define a slightly simplified version.

Theorem 2.4: Let Ξ be the semi-algebraic set as defined by 2.21 and $p(\xi) \in \mathbb{R}[\xi]$. Then

$$p(\xi) > 0 \, \forall \xi \in \Xi \Longleftrightarrow \exists f_1, f_2 \in (\mathbb{R}[\xi]^2 \cup w_r, \; r = 0, ..., R) \; p f_1 = 1 + f_2. \tag{2.28}$$

Proof: [26]. ■

The above theorem characterizes positive polynomials on semi-algebraic sets. Note that nothing is known about the degree of the polynomials f_1 and f_2. However, we know that if the degree bound is chosen to be large enough, the solution obtained by solving the above problem will be correct.

We will continue and analyze non-negativity on compact semi-algebraic sets.

Non-negativity of polynomials on compact semi-algebraic sets

Important theorems for this problem have been proposed by Schmüdgen and Putinar.

Theorem 2.5: Let Ξ be the compact semi-algebraic set as defined by 2.21 and $p(\xi) \in \mathbb{R}[\xi]$. Then

$$p(\xi) > 0 \, \forall \xi \in \Xi \Longleftrightarrow p(\xi) \in \Sigma(\Xi). \tag{2.29}$$

Proof: [27]. ■

Under some additional assumptions it is possible to simplify the above theorem.

Assumption 2.1: Polynomial $\Sigma'(\Xi)$ is an *archimedean*, i.e. $N - \sum_{i=1}^{k} \xi_i^2 \in \Sigma'(\Xi)$ for some $N \in \mathbb{N}$.

Theorem 2.6: If $p(\xi) \in \mathbb{R}[\xi]$, such that for $\forall \xi \in \Xi$ $p(\xi) > 0$ and $\Sigma'(\Xi)$ is an archimedean, then $p \in \Sigma'(\Xi)$.

Proof: [1]. ■

It is clear that Theorem 2.6 gives a stricter characterization of $p(\xi)$ than Theorem 2.5 since $p(\xi) \in \Sigma'(\Xi) \subseteq \Sigma(\Xi)$.

Example 2.10: We will apply Theorem 2.6 to the following example:

$$\Xi = \left\{\xi \in \mathbb{R}^2 : w(\xi_1, \xi_2) = 1 - \xi_1^2 - \xi_2^2 \geq 0\right\} \tag{2.30}$$

and $p(\xi_1) = 2\xi_1 + 3 > 0$ on Ξ.

It is clear that $\Sigma'(\Xi)$ is an archimedean since

$$1 - \xi_1^2 - \xi_2^2 = 0^2 + 1^2 w(\xi_1, \xi_2). \tag{2.31}$$

Due to Theorem 2.6, $p(\xi) \in \Sigma'(\Xi)$. Note that

$$p(\xi_1) = 2\xi_1 + 3 = (\xi_1 + 1)^2 + \xi_2^2 + 1 + (1 - \xi_1^2 - \xi_2^2). \tag{2.32}$$

■

Putinar's Theorem 2.6 still does not successfully characterize all the polynomials that are non-negative on Ξ due to two reasons:

- Assumption that $\Sigma'(\Xi)$ is an archimedean is not always fulfilled.
- It only holds for the polynomials that are positive on Ξ, but not for the non-negative ones.

However, keeping that in mind, we can still use it for solving problem 2.20. Under Assumption 2.1, $p(\xi) \in \Sigma'(\Xi)$ and $p(\xi)$ can thus be written as

$$p(\xi) = \sum_{r=0}^{R} w_r s_r,\ s_r \in \Sigma_{2\tilde{d}_r}(\mathbb{R}^k),\ r = 0, ..., R\,, \tag{2.33}$$

for some $d \in \mathbb{N}$ (Equation 2.22). Problem 2.20 can then, under the above considerations, be solved by

$$\mathbb{U}_{\Xi}^{d} \begin{cases} inf & -a \\ s.t. & a \in \mathbb{R} \\ & p(\xi) - a = \sum_{r=0}^{R} w_r s_r \\ & \left. s_r \in \Sigma_{2\tilde{d}_r}(\mathbb{R}^k) \right\} r = 0, ..., R \end{cases} \tag{2.34}$$

for some $d \in \mathbb{N}$ large enough.

Note, that the above problem is formulated as an SDP and can thus be solved in polynomial time.

2.6.2 The problem of moments

Another approach to address problem 2.20 is to adopt a dual point of view and solve the equivalent problem

$$min_{\mu \in \mathcal{B}(\Xi)} \quad \int p(\xi)\mu(d\xi) \tag{2.35}$$

where $\mathcal{B}(\Xi)$ is the space of the finite probability Borel measures on Ξ.

This optimization problem is closely related to the problem of moments, where given a sequence $y = (y_\alpha)_{\alpha \in \mathbb{N}^k}$, we attempt to find a Borel measure $\mu \in \mathcal{B}(\Xi)$ supported on Ξ such that y_α is the α-th moment of μ. Before we start with the solution of the problem of moments, the concept of moment and localizing matrices must be introduced.

Moment matrix

Definition 2.8: Given a sequence $y := \{y_\alpha\}$ of length $s(k, 2d)$, which confirms with the ordering of the polynomial basis 2.19, we define the moment matrix $M_d(y)$ of a dimension $s(k,d) \times s(k,d)$ as

$$\begin{aligned} M_d(y)(1,i) = M_d(y)(1,i) \quad &= y_{i-1} \quad i = 1, ..., k+1 \\ \left.\begin{aligned} M_d(y)(1,j) = y_\alpha \\ M_d(y)(1,i) = y_\beta \end{aligned}\right\} \Rightarrow \quad & M_d(y)(i,j) = y_{\alpha+\beta} \end{aligned} \tag{2.36}$$

Example 2.11: The moment matrix $M_d(y)$ for $k = 2$ is a block matrix $\{M_{i,j}(y)\}_{0 \le i,j \le 2d}$ given by

$$M_{i,j}(y) = \begin{bmatrix} y_{i+j,0} & y_{i+j-1,1} & \cdots & y_{i,j} \\ y_{i+j-1,1} & y_{i+j-2,2} & \cdots & y_{i-1,j+1} \\ \vdots & \vdots & \ddots & \vdots \\ y_{j,i} & y_{i+j-1,1} & \cdots & y_{0,i+j} \end{bmatrix} \tag{2.37}$$

where $y_{i,j}$ represents the $(i+j)$-order moment $\int \xi^i y^i \mu(d(\xi, y))$ for some probability measure μ. For $d = 2$ and $k = 2$ we obtain

$$M_2(y) = \begin{bmatrix} 1 & | & y_{1,0} & y_{0,1} & | & y_{2,0} & y_{1,1} & y_{0,2} \\ - & - & - & - & - & - & - & - \\ y_{1,0} & | & y_{2,0} & y_{1,1} & | & y_{3,0} & y_{2,1} & y_{1,2} \\ y_{0,1} & | & y_{1,1} & y_{0,2} & | & y_{2,1} & y_{1,2} & y_{0,3} \\ - & - & - & - & - & - & - & - \\ y_{2,0} & | & y_{3,0} & y_{2,1} & | & y_{4,0} & y_{3,1} & y_{2,2} \\ y_{1,1} & | & y_{2,1} & y_{1,2} & | & y_{3,1} & y_{2,2} & y_{1,3} \\ y_{0,2} & | & y_{1,2} & y_{0,3} & | & y_{2,2} & y_{1,3} & y_{0,4} \end{bmatrix}. \tag{2.38}$$

Similarly, the moment matrix $M_d(y)$ for $k = 3$ is given through blocks $\{M_{i,j,l}(y)\}_{0 \le i,j,l \le 2d}$. ■

Proposition 2.7: Let $y := \{y_\alpha\}$ be a sequence of moments up to the order $2d$ of some probability measure μ_y. Then $M_d(y) \succeq 0$.

Proof: Let $f(\xi) \in \mathbb{R}_d[\xi]$, then

$$\langle f, M_d(y) f \rangle = \sum_\alpha f_\alpha^2 y_\alpha = \int f(\xi)^2 \mu_y(d\xi) \ge 0. \tag{2.39}$$

Note that non-negativity is one of the properties of a measure listed in the previous section. ■

Localizing matrix

Definition 2.9: Let $w_r(\xi) : \mathbb{R}^k \longrightarrow \mathbb{R}$ be any of the real-valued polynomials defined by 2.21 with a coefficient vector $w_r \in \mathbb{R}^{s(k,d_r)}$. If the entry (i,j) of the moment matrix $M_d(y)$ is y_β and $\beta(i,j)$ denotes the subscript β of y_β, then $M_d(w_r, y)$ is the localizing matrix

$$M_d(w_r, y)(i,j) = \sum_\alpha [w_r]_\alpha \, y_{\{\beta(i,j)+\alpha\}}. \tag{2.40}$$

Example 2.12: If we are given

$$M_1 = \begin{bmatrix} 1 & y_{1,0} & y_{0,1} \\ y_{1,0} & y_{2,0} & y_{1,1} \\ y_{0,1} & y_{1,1} & y_{0,2} \end{bmatrix} \tag{2.41}$$

and

$$w_r(x) = a - bx_1 - cx_1x_2 \tag{2.42}$$

then the localizing matrix $M_1(w_r, y)$ is given by

$$M_1(w_r, y) = \begin{bmatrix} a - by_{1,0} - cy_{1,1} & ay_{1,0} - by_{2,0} - cy_{2,1} & ay_{0,1} - by_{1,1} - cy_{1,2} \\ ay_{1,0} - by_{2,0} - cy_{2,1} & ay_{2,0} - by_{3,0} - cy_{3,1} & ay_{1,1} - by_{2,1} - cy_{2,2} \\ ay_{0,1} - by_{1,1} - cy_{1,2} & ay_{1,1} - by_{2,1} - cy_{2,2} & ay_{0,2} - by_{1,2} - cy_{1,2} \end{bmatrix}. \tag{2.43}$$

■

Proposition 2.8: Let $y := \{y_\alpha\}$ be a vector of moments up to the order $2d$ of some probability measure μ_y. Then $M_d(w_r, y) \succeq 0$.

Proof: Let $f(\xi) \in \mathbb{R}_d[\xi]$, then

$$\langle f(\xi), M_d(w_r, y) f(\xi) \rangle = \int f(\xi)^2 w(\xi) \mu_y(d\xi) \geq 0. \tag{2.44}$$

Note that $w_r \geq 0$, $r = 0, ..., R$ by definition in the beginning of this section. ■

Theorem 2.9: Let the closed semi-algebraic set Ξ defined by $w_1, ..., w_R \in \mathbb{R}[\xi]$ be compact. Then a sequence $y = (y_\alpha)_{\alpha \in \mathbb{N}^k}$ is a Ξ-moment sequence if and only if $M_d((y_\alpha)_{\alpha \leq 2d}) \succeq 0$ for all $r \in \mathbb{N}$ and $M_d(((w_{i_1} ... w_{i_n} y)_\alpha)_{\alpha \leq 2d}) \succeq 0$ for all possible choices $i_1, ..., i_n$ of pairwise different numbers from $\{1, ..., R\}$ and for all $d \in \mathbb{N}$.

Proof: [23]. ■

This theorem characterizes infinite moment sequences. However, in real life problems infinitely many moments of an unknown distribution are not known. Instead only moments up to an order d are given. Moreover, even if infinitely many moments are known, dealing with them is not computationally tractable. The theorem above is in this case only necessary, but not sufficient to characterize moment sequences.

Under Assumption 2.1 $p(\xi) \in \Sigma'(\Xi)$ and $p(\xi)$ can thus be written as

$$p(\xi) = \sum_{r=0}^{R} w_r s_r, \; s_r \in \Sigma_{2\tilde{d}_r}(\mathbb{R}^k), \; r = 0, ..., R \tag{2.45}$$

for some $d \in \mathbb{N}$ (Equation 2.22). The theory of moments states that under this assumption problem 2.35 can be approximated by a sequence of problems

$$\mathbb{Q}_\Xi^d \left\{ \begin{array}{ll} inf_y & \sum_\alpha p_\alpha y_\alpha \\ s.t. & \left. M_{d-\tilde{d}_r}(w_r, y) \succeq 0 \right\} r = 0, ..., R \end{array} \right. . \tag{2.46}$$

Lasserre [23] showed that if d is chosen to be large enough then $min\, \mathbb{Q}_\Xi^d = p_\Xi^*$.

However, if Assumption 2.1 does not hold then we can only claim that as $d \longrightarrow \infty$ then $inf\, \mathbb{Q}_\Xi^d \uparrow p_\Xi^*$.

2.7 Function approximation

Weierstrass theorem tells us that any continuous function on a closed and bounded interval can be uniformly approximated on that interval by polynomials to any degree of accuracy.

Theorem 2.11 (Weierstrass): Let f be a continuous real-valued function on Ξ. Then for any $\epsilon > 0$ there exists a polynomial p on Ξ such that

$$|f(\xi) - p(\xi)| < \epsilon \tag{2.47}$$

for all $\xi \in \Xi$.

2.8 Notation

2.8.1 Multidimensional matrix

Multidimensional matrix operations are defined in [9]. In this work only a multiplication of a multidimensional matrix with a vector is needed. Multiplication of a matrix $A^{d_1 \times d_2 \times \ldots d_k \times \ldots \times d_n}$ and a vector B^{d_k} over k-th dimension $k \in \{1, ..., n\}$ is denoted by $A \cdot^{(k)} B$. It is calculated by the following expression

$$a_{i_1 i_2 \ldots i_{k-1} i_{k+1} \ldots i_n} = \sum_{j=1}^{d_k} a_{i_1 i_2 \ldots i_{k-1} i_j i_{k+1} \ldots i_n} \cdot b_j \tag{2.48}$$

Multiplication is only possible if the length of the k-th dimension of A matches the length of B.

2.8.2 Polynomial multiplication

We have two polynomials $p(\xi) = p^\top \mathcal{B}_{d_1}(\xi)$ and $q(\xi) = q^\top \mathcal{B}_{d_2}(\xi)$ with degrees d_1 and d_2, respectively. Vectors p and q contain coefficients of the polynomials. Multiplication of polynomials is defined by

$$p(\xi)q(\xi) = (p *_{(d_1,d_2)} q)^\top \mathcal{B}_{d_1+d_2}(\xi). \tag{2.49}$$

The result is a polynomial of degree $d_1 + d_2$. If $k = 1$, i.e. $p(\xi)$ and $q(\xi)$ are univariate polynomials, then coefficients of the product are calculated as

$$\left[p *_{(d_1,d_2)} q\right]_r = \sum_{i=0}^{r} p_{r-i} \cdot q_{r+i} \quad \forall r = \{0, ..., d_1 + d_2\}. \tag{2.50}$$

Operator $*_{(d_1,d_2)}$ in this work denotes a multiplication of polynomials $p(\xi)$ and $q(\xi)$ with degrees d_1 and d_2, respectively.

Chapter 3

Stochastic Programming

In the previous chapter we considered deterministic optimization problems. However, real life problems are almost never deterministic. They may include uncertain parameters that arise due to future events, measurement errors, lack of reliable data etc. Two approaches to handle the uncertainty have been proposed:

- When the uncertain parameters are known only within certain bounds, one approach to tackle such problems is called *robust optimization*. Here the goal is to find a solution which is feasible for all data and optimal in the view of the worst-case realization of the uncertainty.

- When the probability distributions governing the data are known or can be estimated, one tackles such problems with *stochastic programming*. The goal in this case is to find some policy that is feasible for all (or almost all) possible data instances and maximizes the expectation of some function of the decisions and the uncertain parameters.

Stochastic programming has been successfully applied to many different areas [5] (e.g. Capacity planning, Energy, Finance, Production Control, Scheduling, Telecommunications, Sports etc.).

In this chapter we discuss general settings of stochastic programming [4], state of the art solutions and our extensions to one of the existing approaches.

3.1 General formulation and assumptions

Consider the stochastic program 3.1, where ξ is a random vector varying over a set $\Xi \subset \mathbb{R}^k$, $x \subset \mathbb{R}^n$ is a decision variable and $b_i(\xi) : \mathbb{R}^k \longrightarrow \mathbb{R}$, $i = 1, ..., m$. Uncertainty is modeled by a probability space $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k), \mathbb{P})$. Borel σ-algebra $\mathcal{B}(\mathbb{R}^k)$ represents the set of events that are assigned probabilities by the probability measure $\mathbb{P}$. We assume that the functions $g_i(x, \cdot) : \Xi \longrightarrow \mathbb{R}$ for all x, are random variables themselves, and that the probability distribution is independent of x.

$$\begin{aligned} min \quad & g_0(x, \xi) \\ s.t. \quad & x \in X \subset \mathbb{R}^n \\ & g_i(x, \xi) \leq b_i(\xi) \quad i = 1, ..., m \end{aligned} \tag{3.1}$$

If we think of taking decision on x before knowing the realisation of ξ, then problem 3.1 is not well defined, because the meaning of min as well of the constraints is not clear at all. Problem 3.1 thus needs two additional clarifications:

1. Objective function: Minimization of a function of the uncertain parameters can be understood in different ways. One might want to minimize the function in respect to the worst-case realisation of the uncertain parameters ξ. This is the case in the robust optimization. However, one might also want to minimize the expectation of a function of the uncertain parameters and decisions, i.e. $\mathbb{E}(g_0(x, \xi))$. This is the case in the stochastic programming.

2. Constraints: Since constraints are functions of the uncertain parameters, it is not clear when each constraint has to hold. One might want it to hold always, no matter what the outcome of the uncertain parameters is. On the other hand, one might want it to hold only with a certain probability α.

Based on the above observations we can reformulate problem 3.1 into the following well defined problem

$$\begin{aligned} min \quad & \mathbb{E}(g_0(x,\xi)) \\ s.t. \quad & x \in X \subset \mathbb{R}^n \\ & P(\{\xi\,|g_i(x,\xi) \leq b_i(\xi)\,\}) \geq \alpha_i \quad i=1,...,m \end{aligned} \tag{3.2}$$

Problems at the form 3.2 are complex and difficult to tackle even theoretically and most of the work in the stochastic programming introduces some simplifications. In this work we investigate stochastic programming problems under the following assumptions:

Assumption 3.1: The objective function and the constraints are linear functions of the decision variables, i.e.

$$g_0(x,\xi) = c(\xi)^\top x \tag{3.3}$$

for some $c(\xi) : \mathbb{R}^{n\times k} \longrightarrow \mathbb{R}^n$, and

$$P(\{\xi\,|g_i(x,\xi) \leq b_i(\xi)\,\}) \geq \alpha_i = P(\{\xi\,|a_i(\xi)x \leq b_i(\xi)\,\}) \geq \alpha_i \tag{3.4}$$

for some $a_i(\xi) : \mathbb{R}^{n\times k} \longrightarrow \mathbb{R}^n$, $i = 1,...,m$.

Assumption 3.2: Constraints are almost always satisfied. In other words, for all $\xi \in \Xi$ the constraints hold with probability 1 (i.e. $\alpha_i = 1$ for $i = 1,...,m$). As a consequence, we also require Ξ to be a bounded compact set.

For the further investigation it is convenient to simplify the notation of constraints

$$P(\{\xi\,|a_i(\xi)x \leq b_i(\xi)\,\}) = 1 \Longleftrightarrow a_i(\xi)x \leq b_i(\xi)\,\mathbb{P}-a.s.. \tag{3.5}$$

Under the Assumptions 3.1 and 3.2, we can rewrite problem 3.2 as

$$\begin{aligned} inf \quad & \mathbb{E}(c(\xi)^\top x) \\ s.t. \quad & x \in X \subset \mathbb{R}^n \\ & A(\xi)x \leq b(\xi)\,\mathbb{P}-a.s. \end{aligned} \tag{3.6}$$

In the discussion above, it is assumed that the decisions x are made before the realisation of the uncertain parameters ξ. However, it is also possible that the decisions x are made after the uncertain parameters ξ are revealed. In this case, one would like to know how will the future decisions depend on the realisation of the uncertain parameters ξ. Decisions are thus modeled as functions of the uncertain parameters. A model that describes this situation is the following

$$\begin{aligned} inf \quad & E(c(\xi)^\top x(\xi)) \\ s.t. \quad & x \in \mathcal{L}_{k,n} \\ & A(\xi)x(\xi) \leq b(\xi)\,\mathbb{P}-a.s. \end{aligned} \tag{3.7}$$

where $\mathcal{L}_{k,n}$ denotes the space of all Borel measurable functions from $\mathbb{R}^k$ to $\mathbb{R}$ that are bounded on compact sets (Assumption 3.2). When the decisions x are modeled as functions of the uncertain parameters ξ, they are termed *decision rules, strategy* or *policy.*

For the further argumentation, it is useful to introduce a functional slack variable $s \in \mathcal{L}_{k,m}$ and transform inequality constraints in problem 3.7 to equality constraints as

$$\begin{aligned} inf \quad & \mathbb{E}(c(\xi)^\top x(\xi)) \\ s.t. \quad & x \in \mathcal{L}_{k,n}, \quad s \in \mathcal{L}_{k,m} \\ & \left.\begin{aligned} & A(\xi)x(\xi) + s(\xi) = b(\xi) \\ & s(\xi) \geq 0 \end{aligned}\right\} \mathbb{P} - a.s. \end{aligned} \tag{3.8}$$

Above we have analyzed two special cases, where the uncertain parameters ξ are revealed before or after the decisions are made. A more general case is when the uncertain parameters are revealed sequentially and after each revelation some decisions are made. Consider for example one month portfolio optimization, where we have an opportunity to rebalance the portfolio daily. In this case, we would like to start with an initial investment and then alter it daily, based on the stock returns observed in the previous days.

Models that allow us to investigate such problems are called *recursive models.* They are examined more in details in the next section.

3.2 Recourse problems

Recourse problems are a broad and widely applied class of stochastic programming problems. Recourse is the ability to take corrective actions after some realisation of the uncertain parameters. The simplest case of recourse problems have two stages:

1. in the first stage we make a decision
2. in the second stage we observe a realisation of the uncertain parameters of the problem, but are allowed to make further decisions to avoid the constraints of the problem becoming infeasible.

Note that in the second stage the decisions that we make will depend on a particular realisation of the uncertain parameters we observed.

Example 3.1: To illustrate a simple two-stage recourse model consider the following simplified production planning problem. We have a company that produces product X with the production costs of £2/*unit*. We have to produce enough X to meet the customer demand in the next time period. However, demand D is stochastic with only two possible outcomes S_1 and S_2. Demand D is given by the following discrete probability distribution

$$D = \begin{pmatrix} S_1 & S_2 \\ 0.6 & 0.4 \\ 40 & 20 \end{pmatrix} \tag{3.9}$$

where the first line denotes the outcome, the second line its probability and the third line the realisation of the demand.

We also have the flexibility to buy X at any time from an external supplier, but this costs us £3/*unit*. How much should we choose to make now, before we know what customer demand is?

One way to think of this two-stage model is:

1. Decide on the amount to produce
2. Observe the real demand (realisation of the scenario S_1 or S_2)

3. If demand is not met, buy the remainder of the products from the external supplier.

Let $x \geq 0$ be the number of units of X to produce now (at the first stage). Uncertain parameters are modelled with two scenarios $s \in \{S_1, S_2\}$, where each occurs with probability p_s. The number of products bought from the external supplier at the second-stage is denoted by $y_s \geq 0$ when the uncertain realisation of the demand is D_s. Constraints ensure that the demand is always satisfied, i.e.

$$x + y_s \geq D_s \quad s \in \{S_1, S_2\}\,. \tag{3.10}$$

Note that we must have $\geq$ here, since the amount of x we produce may exceed customer demand. We wish to minimize cost given by

$$\mathbb{E}(2x + 3y) = 2x + \sum_{s \in \{S_1, S_2\}} 3 p_s y_s. \tag{3.11}$$

The optimal strategy is thus the solution of the following optimization problem

$$\begin{aligned} min \quad & 2x + \textstyle\sum_{s \in \{S_1,S_2\}} 3 p_s y_s \\ s.t. \quad & x \in \mathbb{R},\ y_s \in \mathbb{R}\ s \in \{S_1, S_2\} \\ & \left.\begin{aligned} & x + y_s \geq D_s \\ & x, y_s \geq 0 \end{aligned}\right\} s \in \{S_1, S_2\} \end{aligned} \tag{3.12}$$

Note that this is actually a deterministic program. However, in many cases the uncertain parameters are not discrete and thus an exact deterministic program with finitely many constraints does not exit. ■

The idea of the two-stage recourse can be extended to more stages. The uncertain parameters are now represented as $\xi = (\xi_1, ..., \xi_T)$, where subvectors $\xi_t \in \mathbb{R}^{k_t}$ are observed at time points $t \in \mathbb{T} := \{1, ..., T\}$. The history of the observations up to time t is denoted by $\xi^t := (\xi_1, ..., \xi_t) \in \mathbb{R}^{k^t}$, where $k^t := \sum_{s=1}^{t} k_s$. $\mathbb{E}_t(\cdot)$ denotes the conditional expectation with respect to $\mathbb{P}$ given the uncertain parameters ξ^t. Note that a one-stage model can be understood as a special case of the multistage model where $\xi^T = \xi$ and $k^T = k$.

The decision $x_t(\xi^t)$ is made at time t after uncertain parameters ξ^t have been revealed, but before any future outcomes $\{\xi_s\}_{s>t}$ have been observed. The objective is to find a sequence of the decision rules $x_t \in \mathbb{R}_{k^t, n_t}$, $t \in \mathbb{T}$ that is feasible for problem 3.13 and minimizes its objective function. The requirement that x_t depends only on ξ^t reflects the idea that decisions can not depend on the unknown future parameters.

$$\begin{aligned} inf \quad & \mathbb{E}(\textstyle\sum_{t=1}^{T} c_t(\xi^t)^\top x_t(\xi^t)) \\ s.t. \quad & x_t \in \mathcal{L}_{k^t, n_t} \forall t \in \mathbb{T} \\ & \textstyle\sum_{s=1}^{t} A_{ts}(\xi^t) x_s(\xi^s) \leq b_t(\xi^t)\, \mathbb{P} - a.s.,\ \forall t \in \mathbb{T} \end{aligned} \tag{3.13}$$

For the further argumentation we introduce a sequence of non-anticipative (i.e. the decision at a given stage does not depend on the future realization of the uncertain parameters) slack variables $s_t \in \mathcal{L}_{k^t, m_t}$, $t \in \mathbb{T}$ and rewrite problem 3.13 as

$$
\begin{aligned}
inf \quad & \mathbb{E}(\textstyle\sum_{t=1}^{T} c_t(\xi^t)^\top x_t(\xi^t)) \\
s.t. \quad & x_t \in \mathcal{L}_{k^t,n_t}, \quad s_t \in \mathcal{L}_{k^t,m_t} \quad \forall t \in \mathbb{T} \\
& \left.\begin{array}{l} \sum_{s=1}^{t} A_{ts}(\xi^t)x_s(\xi^s) + s_t(\xi^t) = b_t(\xi^t) \\ s_t(\xi^t) \geq 0 \end{array}\right\} \mathbb{P} - a.s., \ \forall t \in \mathbb{T}
\end{aligned}
\tag{3.14}
$$

Linear multistage stochastic programming problems are very hard to solve. Even linear two-stage stochastic programming problems are #P-hard [6] and thus approximation methods are needed in order to tackle them. Analytical solutions are computable only for very small problems. In the next sections, we will examine two different approximation methods of stochastic programs scenario tree approximation and decision rule approximations.

3.3 Scenario Tree Approximation

Scenario tree approximation is the most widely used approach when dealing with multistage stochastic problems. It involves discretizing outcomes of the uncertain parameters ξ in each stage $t \in \mathbb{T}$. Generation of a scenario tree is described in [4, 13]. In this work, we follow formulation proposed in [12]. It was developed for portfolio optimization problems, but can be applied to general one and multistage recourse problems.

A *scenario* is defined as a possible realisation of the uncertain parameters ξ^T. A set of scenarios $\mathcal{N}_T$ in the last stage ($t = T$) corresponds to the set of leaves of a scenario tree. Nodes in the tree $\mathcal{N}_t$ at a level t ($t = 1, ..., T-1$) correspond to a possible realisation of ξ^t. Each node is denoted by $e = (s, t)$, where s is a scenario and t is the level of the node in the tree. For example, the root node is $0 = (s, 1)$, where s can be any scenario, since all the scenarios have the same root node. The ancestors of the node $e = (s, t)$ are denoted by $a_i(e) = (s, i)$, $i = 1, ..., t-1$, where $a_{t-1}(e)$ denotes the parent of node $e = (s, t)$. The branching probability p_e is the conditional probability of the event $e = (s, t)$, given its parent event $a_{t-1}(e)$. The path to the event e is a partial scenario with probability $P_e = \prod p_e$ along that path. Note that probabilities P_e sum up to 1 across each level of the tree nodes $\mathcal{N}_t$ for all $t \in \mathbb{T}$. Each node at the level t corresponds to a decision $x_t(\xi^t)$, which must be determined at time t. Finally, by $\tilde{\xi}_e^t$ we denote the realisation of the uncertain parameters ξ^t that must occur in order to get to the node e.

In order to obtain a tractable approximation of the recourse problem by the scenario tree approximation, one must discretize the uncertain parameters ξ. All possible realizations of ξ are thus approximated by a discrete set of scenarios. The entire set of scenarios can be represented by a scenario tree. An example of a scenario tree with three time periods and a two-three branching structure is depicted in Figure 5.1.

3.3.1 One-stage Scenario Tree Approximation

Scenario trees can be used to obtain a tractable approximation of problem 3.8.

Proposition 3.1: Problem 3.15 represents a tractable approximation of problem 3.8.

$$
\begin{aligned}
min \quad & \textstyle\sum_{\forall e \in \mathcal{N}_2} P_e \, c(\tilde{\xi}_e)^\top x(\tilde{\xi}_e) \\
s.t. \quad & x(\tilde{\xi}_e) \in \mathbb{R}^n, \quad s(\tilde{\xi}_e) \in \mathbb{R}^m \quad \forall e \in \mathcal{N}_2 \\
& \left.\begin{array}{l} A(\tilde{\xi}_e)x(\tilde{\xi}_e) + s(\tilde{\xi}_e) = b(\tilde{\xi}_e) \\ \\ s(\tilde{\xi}_e) \geq 0 \end{array}\right\} \forall e \in \mathcal{N}_2
\end{aligned}
\tag{3.15}
$$

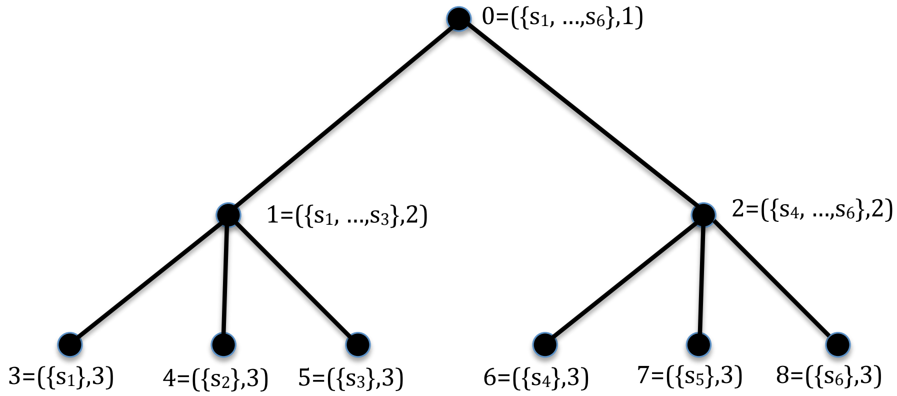

Figure 3.1: An example of a scenario tree with three time periods and two-three branching structure.

Proof: In order to obtain a tractable approximation of problem 3.8 scenario trees can be applied. One must first approximate the objective function.

$$\begin{aligned} \mathbb{E}(c(\xi)^\top x(\xi)) &\approx \textstyle\sum_{\forall e \in \mathcal{N}_2} P_e\, c(\xi \,\Big|\, \tilde{\xi}_e)^\top x(\xi \,\Big|\, \tilde{\xi}_e) \\ &= \textstyle\sum_{\forall e \in \mathcal{N}_2} P_e\, c(\tilde{\xi}_e)^\top x(\tilde{\xi}_e) \end{aligned} \tag{3.16}$$

In a similar manner it is also possible to approximate the constraints by

$$\begin{aligned} &\left.\begin{array}{l} A(\xi)x(\xi) + s(\xi) = b(\xi) \\ s(\xi) \geq 0 \end{array}\right\} \mathbb{P} - a.s. \\ \Longrightarrow\quad &\left.\begin{array}{l} A(\xi \,\Big|\, \tilde{\xi}_e) x(\xi \,\Big|\, \tilde{\xi}_e) + s(\xi \,\Big|\, \tilde{\xi}_e) = b(\xi \,\Big|\, \tilde{\xi}_e) \\ s(\xi \,\Big|\, \tilde{\xi}_e) \geq 0 \end{array}\right\} \forall e \in \mathcal{N}_2 \\ \Longleftrightarrow\quad &\left.\begin{array}{l} A(\tilde{\xi}_e)x(\tilde{\xi}_e) + s(\tilde{\xi}_e) = b(\tilde{\xi}_e) \\ s(\tilde{\xi}_e) \geq 0 \end{array}\right\} \forall e \in \mathcal{N}_2 \end{aligned} \tag{3.17}$$

By combining 3.16 and 3.17 we obtain problem 3.15. The approximation is clearly tractable, since problem 3.15 is formulated as a LP. ■

As the number of the scenarios increases, the solution of problem 3.15 converges to the optimal solution of 3.8[28].

3.3.2 Multistage Scenario Tree Approximation

A similar reasoning can also be used to obtain a tractable approximation of the multistage stochastic programs.

Proposition 3.2: Problem 3.18 represents a tractable approximation of problem 3.14.

$$\begin{aligned} min \quad & \textstyle\sum_{t=1}^{T} \sum_{\forall e \in \mathcal{N}_t} P_e c_t(\tilde{\xi}_e^t)^\top x_t(\tilde{\xi}_e^t) \\ s.t. \quad & x_t(\tilde{\xi}_e^t) \in \mathbb{R}^n, \quad s_t(\tilde{\xi}_e^t) \in \mathbb{R}^m \quad \forall t \in \mathbb{T} \quad \forall e \in \mathcal{N}_t \\ & \left.\begin{array}{l} \sum_{s=1}^{t} A_{ts}(\tilde{\xi}_e^t) x_s(\tilde{\xi}_{a_s(e)}^s) + s_t(\tilde{\xi}_e^t) = b_t(\tilde{\xi}_e^t) \\ \\ s_t(\tilde{\xi}_e^t) \geq 0 \end{array}\right\} \forall t \in \mathbb{T},\, \forall e \in \mathcal{N}_t \end{aligned} \tag{3.18}$$

Proof: In order to obtain a tractable approximation of problem 3.14 scenario trees can be applied. One must first approximate the objective function.

$$\begin{aligned} \mathbb{E}(\textstyle\sum_{t=1}^{T} c_t(\xi^t)^\top x_t(\xi^t)) &\approx \textstyle\sum_{t=1}^{T} \sum_{\forall e \in \mathcal{N}_t} P_e\, c(\xi^t \left| \tilde{\xi}_e^t \right.)^\top x(\xi^t \left| \tilde{\xi}_e^t \right.) \\ &= \textstyle\sum_{t=1}^{T} \sum_{\forall e \in \mathcal{N}_t} P_e\, c(\tilde{\xi}_e^t)^\top x(\tilde{\xi}_e^t) \end{aligned} \tag{3.19}$$

In a similar manner it is also possible to approximate the constraints by

$$\begin{aligned} & \left.\begin{array}{l} \sum_{s=1}^{t} A_{ts}(\xi^t) x_s(\xi^s) + s_t(\xi^t) = b_t(\xi^t) \\ s_t(\xi^t) \geq 0 \end{array}\right\} \mathbb{P} - a.s.,\ \forall t \in \mathbb{T} \\ \Longrightarrow \quad & \left.\begin{array}{l} \sum_{s=1}^{t} A_{ts}(\xi^t \left| \tilde{\xi}_e^t \right.) x_s(\xi^s \left| \tilde{\xi}_{a_s(e)}^s \right.) + s_t(\xi^t \left| \tilde{\xi}_e^t \right.) = b_t(\xi^t \left| \tilde{\xi}_e^t \right.) \\ s_t(\xi^t \left| \tilde{\xi}_e^t \right.) \geq 0 \end{array}\right\} \forall t \in \mathbb{T},\ \forall e \in \mathcal{N}_t \\ \Longleftrightarrow \quad & \left.\begin{array}{l} \sum_{s=1}^{t} A_{ts}(\tilde{\xi}_e^t) x_s(\tilde{\xi}_{a_s(e)}^s) + s_t(\tilde{\xi}_e^t) = b_t(\tilde{\xi}_e^t) \\ s_t(\tilde{\xi}_e^t) \geq 0 \end{array}\right\} \forall t \in \mathbb{T},\ \forall e \in \mathcal{N}_t \end{aligned} \tag{3.20}$$

By combining 3.19 and 3.20 we obtain problem 3.18. The approximation is clearly tractable, since problem 3.18 is formulated as a LP. ■

As the number of scenarios increases, the solution of problem 3.18 converges to the optimal solution of 3.14[28].

3.3.3 Advantages and disadvantages

Scenario trees have some advantages and disadvantages. One of the main advantages is that SP problem is approximated as a LP, which has been an important area of research for a long time. Many open source and commercially available solvers exist and thus LP problems can be solved efficiently. Another important advantage, which we have already discussed, is that the optimal value of problem 3.18 converges to the optimal value of problem 3.14 as discretizations are made finer.

The main disadvantage is the so called *curse of dimensionality.* Computational complexity of problem 3.18 grows exponentially with the number of stages. Even though it is often possible to reduce the number of scenarios [16], the exponential growth is unavoidable. In [14], it was shown that the number of branches starting from each node must be larger than the number of the uncertain parameters observed at that node because otherwise, arbitrage opportunities could arise. This clearly dictates the exponential growth.

The curse of dimensionality can be avoided by a different approach, referred as *decision rule approximation*s.

3.4 Decision Rule Approximations

Decision rule approximations approach is an alternative to the scenario tree approximation when solving multistage stochastic problems. Instead of discretizing the distribution of the uncertain parameters, one can restrict the functional form of the decisions.

3.4.1 Decision Rule Overview

Decisions in problems 3.8 and 3.14 can be characterized by any Borel measurable function from $\mathbb{R}^k$ to $\mathbb{R}$ that is bounded on the compact sets. This broad characterization make the multistage SP problems intractable. In the previous section, we obtained a tractable approximation with a discretization of the uncertain parameters. We found optimal decisions for each possible discrete outcome. Another approach is to keep the distribution of the uncertain parameters unchanged, but only limit the characterization of the decisions to certain functional forms. Linear, piecewise linear and polynomial functional forms have attracted most of the attention. As we have already mentioned, the decisions are in this context termed decision rules. In this section we give an overview of a such approach.

Since decision rule approximations can be applied in both robust and stochastic optimization problems, both fields evolved in parallel. Linear decision rules were first introduced by Ben-Tal [19] in the robust optimization setting. Shapiro [11] then applied a similar idea to the stochastic programming. His formulation gave a tractable upper bound approximation on the optimal value of problem 3.14. However, by considering only linear decision rules, the approximated solutions may be very suboptimal. Therefore, a computationally tractable approach to estimate the degree of suboptimality was needed. Kuhn [3] proposed a lower bound approximation by approximating the functional form of the dual decision rules as linear functions of the uncertain parameters. Primal approximation, via limiting decision rules to linear functions, underestimates the decision maker's flexibility and thus only an upper bound of the optimal solution is obtained. However, the primal problem can be transformed to the Lagrangian dual problem (remember description in Section 2.2), which represents a lower bound for the optimal value of the primal. Remember that the dual problem is defined as maximization and thus it approaches the optimal solution from below. If decision rules of the dual problem are approximated by the linear functions, this again underestimates the decision maker's flexibility. However, since flexibility is underestimated on the dual, this leads to a lower bound approximation. The main advantage of using linear decision rule approximations is that both the primal and the dual SP can be approximated by tractable linear programs. However, the resulting problems most of the times lead to very suboptimal solutions.

In order to improve the approximation quality, Chen [29] proposed the use of the piecewise linear decision rules. Georghiou [30] formulated also the dual problem to estimate the approximation error. Piecewise linear decision rules are proved to be superior relative to the linear decision rules. However, they involve multiple design parameters and are thus cumbersome to use.

A way to avoid multiple design parameters from the piecewise linear decision rules is to limit characterization of decision rules to polynomial functional forms, where the only design parameter is the degree of the polynomial functions. However, such approximation does not lead to a tractable linear program, but instead to an intractable semi-infinite programming problem, having finitely many decision variables, but infinitely many constraints. Constraints involve checking a non-negativity of a polynomial on a compact bounded semi-algebraic set. We have already seen in Section 2.6 that this can be approximated (sometimes even solved exactly) by an SDP through SOS decomposition or the problem of moments. SOS decomposition gives an upper bound on the optimal solution, while its dual, the problem of moments, gives a lower bound on the optimal solution. By such approach, we can formulate tractable primal and dual approximation. Polynomial decision rule approximations have been proposed in the robust optimization framework by Bertsimas [31]. Bampou [7] applied them to the stochastic programming problems by formulating the primal and the dual problem. The approximation error even for the polynomials of small degrees was lower than for the piecewise linear decision rules. However, the formulation was only applicable for even degrees polynomial decision rules and problems where the recourse matrix A does not depend on the uncertain parameters.

In this work we approximate the primal and the dual stochastic programming problems with polynomial decision rules, while releasing Bampou's assumptions. The formulation holds for polynomial decision rules of all degrees while the recourse matrix is modeled by polynomial functions of the uncertain parameters.

In the rest of this section we will first examine primal and dual one-stage approximations and then extend them to the multistage problems.

3.4.2 One-stage Primal Polynomial Decision Rules

Before we start with the description of our method, some assumptions under which the reasoning below holds must be defined. We require Assumptions 3.1 and 3.2 to be true. Moreover, some additional constraints need to be added.

Assumption 3.3: The objective function coefficients and the constraint functions depend polynomially on the uncertain parameters ξ. Namely, $c(\xi) = C\mathcal{B}_\theta(\xi)$ for some $C \in \mathbb{R}^{n \times s(k,\theta)}$, $b(\xi) = B\mathcal{B}_\theta(\xi)$ for some $B \in \mathbb{R}^{m \times s(k,\theta)}$ and $A(\xi) = A \cdot^{(3)} \mathcal{B}_\eta(\xi)$ (remember the definition of the multidimensional matrix multiplication in Section 2.8.1) for some $A \in \mathbb{R}^{m \times n \times s(k,\eta)}$ for all $\xi \in \Xi$. Note that the requirement that both polynomials in $c(\xi)$ and $b(\xi)$ share the same degree is nonrestrictive, but simplifies the notation.

Assumption 3.2*: The support Ξ of the probability measure $\mathbb{P}$ is a compact semi-algebraic set with nonempty interior defined by polynomial inequalities 3.21, where $w_r \in \mathbb{R}_{d_j}[\xi]$, $r = 0, ..., R$ and $w_0 = 1$.

$$\Xi = \left\{ \xi \in \mathbb{R}^k : w_r(\xi) \geq 0, \ r = 0, ..., R \right\} \tag{3.21}$$

Note that this assumption extends Assumption 3.2 where only a bounded compact set was assumed.

Example 3.2: This example is used to show that Assumption 3.3 is nonrestrictive. Let $p_1(\xi) = 3\xi + 5$ and $p_2(\xi) = 2\xi^2 + \xi + 2$. We would like both polynomials to be of the degree 2. We write

$$\begin{array}{rcccl} p_1(\xi) & = & 5 + 3\xi & = & \begin{bmatrix} 5 & 3 & 0 \end{bmatrix} \mathcal{B}_2(\xi) \\ p_2(\xi) & = & 2 + \xi + 2\xi^2 + \xi & = & \begin{bmatrix} 2 & 1 & 2 \end{bmatrix} \mathcal{B}_2(\xi) \end{array} \tag{3.22}$$

It is clear that this kind of construction is always possible. ■

In the rest of this chapter, we first discuss how to obtain computationally tractable approximation of one-stage problem 3.8, whose solution constitutes an upper bound on the value of problem 3.8.

Approximation P1: In order to derive a conservative approximation for problem 3.8, we reduce the set of admissible decision rules from the space of all continuous measurable functions to the space of polynomial functions of a degree $d - \eta$. We set

$$\begin{array}{l} x(\xi) = X\mathcal{B}_{d-\eta}(\xi) \quad for\ some\ X \in \mathbb{R}^{n \times s(k,d-\eta)} \\ s(\xi) = S\mathcal{B}_d(\xi) \quad for\ some\ S \in \mathbb{R}^{n \times s(k,d)} \end{array} \tag{3.23}$$

and require that $d \geq max\,\{\eta, \theta, d_0, d_1, ..., d_R\}$.

Note, that Approximation P1 is supported by the Weierstrass theorem and as $d \longrightarrow \infty$ the above formulation characterizes all continuous measurable functions.

Based on Approximation P1 and Assumption 3.3, the objective function in 3.8 can be rewritten as

$$\begin{aligned}
\mathbb{E}(c(\xi)^\top x(\xi)) &= \mathbb{E}((C\mathcal{B}_\theta(\xi))^\top X\mathcal{B}_{d-\eta}(\xi)) \\
&= \mathbb{E}((CT_{\theta,d-\eta}\mathcal{B}_{d-\eta}(\xi))^\top X\mathcal{B}_{d-\eta}(\xi)) \\
&= \mathbb{E}(tr(CT_{\theta,d-\eta}\mathcal{B}_{d-\eta}(\xi)\mathcal{B}_{d-\eta}(\xi)^\top X^\top)) \\
&= tr(CT_{\theta,d-\eta}\mathbb{E}(\mathcal{B}_{d-\eta}(\xi)\mathcal{B}_{d-\eta}(\xi)^\top)X^\top) \\
&= tr(CT_{\theta,d-\eta}M_{d-\eta}X^\top) \\
&= tr((CT_{\theta,d-\eta}M_{d-\eta}X^\top)^\top) \\
&= tr((M_{d-\eta}X^\top)^\top(CT_{\theta,d-\eta})^\top) \\
&= tr((CT_{\theta,d-\eta})^\top(M_{d-\eta}X^\top)^\top) \\
&= tr(T_{\theta,d-\eta}^\top C^\top X M_{d-\eta}^\top) \\
&= tr(T_{\theta,d-\eta}^\top C^\top X M_{d-\eta})
\end{aligned} \tag{3.24}$$

where $M_{d-\eta}$ denotes a moment matrix. A truncation operator $T_{d_1,d_2} : \mathbb{R}^{s(k,d_2)} \rightarrow \mathbb{R}^{s(k,d_1)}$ which maps a monomial basis $\mathcal{B}_{d_2}(\xi)$ to the reduced basis $\mathcal{B}_{d_1}(\xi)$ was introduced.

By substituting 3.23 and 3.24 into problem 3.8 we obtain the following problem

$$\begin{aligned}
&inf \quad tr(T_{\theta,d-\eta}^\top C^\top X M_{d-\eta}) \\
&s.t. \quad X \in \mathbb{R}^{n\times s(k,d-\eta)}, S \in \mathbb{R}^{n\times s(k,d)} \\
&\qquad \left.\begin{array}{l} (A \cdot^{(3)} \mathcal{B}_\eta(\xi)) \cdot (X\mathcal{B}_{d-\eta}(\xi)) + S\mathcal{B}_d(\xi) = BT_{\theta,d}\mathcal{B}_d(\xi) \\ S\mathcal{B}_d(\xi) \geq 0 \end{array}\right\} \mathbb{P} - a.s.
\end{aligned} \tag{3.25}$$

It is possible to formulate problem 3.25 in a more convenient form, by rewriting the constraints.

$$\begin{aligned}
(A \cdot^{(3)} \mathcal{B}_\eta(\xi)) \cdot (X\mathcal{B}_{d-\eta}(\xi)) &= \begin{bmatrix} a_{11}(\xi) & a_{12}(\xi) & \cdots & a_{1n}(\xi) \\ a_{21}(\xi) & a_{22}(\xi) & \cdots & a_{2n}(\xi) \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}(\xi) & a_{m2}(\xi) & \ldots & a_{mn}(\xi) \end{bmatrix} \begin{bmatrix} x_1(\xi) \\ x_2(\xi) \\ \vdots \\ x_n(\xi) \end{bmatrix} \\
&= \begin{bmatrix} \sum_{i=1}^n a_{1i}(\xi)x_i(\xi) \\ \sum_{i=1}^n a_{2i}(\xi)x_i(\xi) \\ \vdots \\ \sum_{i=1}^n a_{mi}(\xi)x_i(\xi) \end{bmatrix} \\
&= \begin{bmatrix} \sum_{i=1}^n a_{1i} *_{(\eta,d-\eta)} x_i \\ \sum_{i=1}^n a_{2i} *_{(\eta,d-\eta)} x_i \\ \vdots \\ \sum_{i=1}^n a_{mi} *_{(\eta,d-\eta)} x_i \end{bmatrix} \mathcal{B}_d(\xi) \\
&= (A *_{(\eta,d-\eta)}^{(3)} X)\mathcal{B}_d(\xi)
\end{aligned} \tag{3.26}$$

In the last step a new operator $*^{(d_3)}_{(d_1,d_2)}$ as a combination of the multidimensional matrix multiplication and the polynomial multiplication (remember the definition in Section 2.8.2) was introduced in order to keep the formulation in a matrix form. Note that d_1 and d_2 denote the degree of polynomials and d_3 the dimension for the multidimensional matrix multiplication.

By inserting 3.26 into problem 3.25, the following problem is obtained.

$$
\begin{array}{ll}
inf & tr(T^\top_{\theta,d-\eta} C^\top X M_{d-\eta}) \\
subject\ to & X \in \mathbb{R}^{n\times s(k,d-\eta)}, S \in \mathbb{R}^{n\times s(k,d)} \\
& \left.\begin{array}{l} (A *^{(3)}_{(\eta,d-\eta)} X)\mathcal{B}_d(\xi) + S\mathcal{B}_d(\xi) = BT_{\theta,d}\mathcal{B}_d(\xi) \\ S\mathcal{B}_d(\xi) \geq 0 \end{array}\right\} \mathbb{P}-a.s.
\end{array}
\tag{3.27}
$$

Proposition 3.3: For problems 3.8 and 3.27 following holds: inf 3.27 $\geq$ inf 3.8.

Proof: As problem 3.27 is obtained from 3.8 by Approximation P1, the feasible set of 3.27 is smaller than the feasible set of 3.8. It is thus clear that the solution of problem 3.27 provides an upper bound on the optimal value of problem 3.8. The feasible sets, and thus the solutions of problems 3.8 and 3.27, are in general equal as $d \longrightarrow \infty$ due to the Weierstrass theorem. ■

Problem 3.27 involves finitely many decision variables if d is finite. However, there are still infinitely many constraints, since they must hold for each possible realisation of the uncertain parameters $\xi \in \Xi$. Thus, constraints in 3.27 require a vector-valued polynomial to vanish identically on a set with nonempty interior Ξ. This is possible if and only if all the coefficients of the polynomial vanish. The equality constraint in problem 3.27 is thus equivalent to

$$
A *^{(3)}_{(\eta,d-\eta)} X + S = BT_{\theta,d} \tag{3.28}
$$

The inequality constraint in 3.27 requires that each component of a vector-valued polynomial $s(\xi)$ is non-negative on Ξ (i.e. belongs to $\mathcal{P}_d(\Xi)$). We have already discussed characterization of such polynomials in Section 2.6.1. A reasonable approach would be to apply Theorem 2.5.

Approximation P2: For every polynomial $s_i(\xi)$, $i = 1, ..., m$ positive on Ξ, 3.29 holds due to Theorem 2.5, if d is chosen to be large enough.

$$
\forall \xi \in \Xi \; s_i(\xi) > 0 \Longleftrightarrow s_i(\xi) \in \Sigma(\Xi). \tag{3.29}
$$

Note that at least $d \geq max \left\{\eta, \theta, \sum_{r=0}^{R} d_r\right\}$ must hold. ■

Real life problems usually involve many constraints $w_r(\xi)$, $r = 0, ..., R$ that define the semi-algebraic set Ξ and thus the minimum required degree d as defined by Approximation P2 can be relatively large. Our experience presented in Chapter 5.1 will show that current computers can not handle most of such problems. One possible way to get around this problem is to require Assumption 2.1 to hold. In this case, Theorem 2.6 can be applied and we can thus replace Approximation P2 with Approximation P2*.

Approximation P2*: For every polynomial $s_i(\xi)$, $i = 1, ..., m$ positive on Ξ, 3.30 holds under Assumption 2.1 due to Theorem 2.6, if d is chosen to be large enough.

$$
\forall \xi \in \Xi \; s_i(\xi) > 0 \Longleftrightarrow s_i(\xi) \in \Sigma'(\Xi). \tag{3.30}
$$

Note that at least $d \geq max \{\eta, \theta, d_0, d_1, ..., d_R\}$ must hold. ■

This approximation is more reasonable for real life problems. The minimal degree d also coincides with the requirement in Approximation P1. Note, however, that Assumption 2.1 is sometimes violated and in this case Approximation P2* does not characterize sufficiently all the polynomials positive on Ξ even as $d \longrightarrow \infty$.

For the further argumentation, we define $\Sigma_d^{'}$ in terms of degrees of the SOS polynomials by 3.31. Remember that $\tilde{d}_r = \lfloor \frac{d-d_r}{2} \rfloor$.

$$\Sigma_d'(\Xi) = \left\{ s \in \mathbb{R}_d\left[\xi\right] : \begin{array}{l} s(\xi) = \sum_{r=0}^{R} s_r(\xi) w_r(\xi) \\ s_r \in \Sigma_{2\tilde{d}_r}(\mathbb{R}^k), r = 0, ..., R \end{array} \right\} \tag{3.31}$$

Theorem 2.3 states that checking membership of a polynomial in $\Sigma_{2\tilde{d}_j}(\mathbb{R}^k)$ is equivalent to solving an SDP problem. Thus, it is clear that also checking membership in $\Sigma_d'(\Xi)$ can be formulated as an SDP problem.

Proposition 3.4: Assume that Ξ is defined as in 3.21. Then, for any $s \in \mathbb{R}_d\left[\xi\right]$ the following statements are equivalent.

1. $s \in \Sigma_d^{'}(\Xi)$
2. There exist positive semidefinite matrices $Y^r \in \mathbb{S}^{s(k,\tilde{d}_r)}$, $r = 0, ..., R$, such that $s = \sum_{r=0}^{J} \Lambda_j^*(Y^r)$, where $\Lambda_r^* : \mathbb{S}^{s(k,\tilde{d}_r)} \to \mathbb{R}^{s(k,d)}$ is a linear operator defined through

$$[\Lambda_r^*(Y^r)]_\alpha = \langle Q_\alpha^r, Y^r \rangle, \quad \alpha \in L_d \tag{3.32}$$

and $Q_\alpha^r \in \mathbb{S}^{s(k,\tilde{d}_r)}$ is a real symmetric matrix defined through

$$[Q_\alpha^r]_{\beta\gamma} = \begin{cases} [w_j]_\delta & if\ \alpha - \beta + \gamma = \delta \\ 0 & otherwise \end{cases} \tag{3.33}$$

Proof: A similar proof has been proposed in [1, 7]. We introduce the linear operators Λ_r : $\mathbb{R}^{s(k,d)} \longrightarrow \mathbb{S}^{s(k,\tilde{d}_r)}$, $r = 0, ..., R$ by

$$\Lambda_r(\mathcal{B}_d(\xi)) = \sum_{\alpha \in L_d} Q_\alpha^r \xi^\alpha = \mathcal{B}_{\tilde{d}_r}(\xi)\mathcal{B}_{\tilde{d}_r}(\xi)^\top w_r^\top \mathcal{B}_{d_r}(\xi). \tag{3.34}$$

We also define the operators Λ_r^* such that they are adjoint to Λ_r according to equation 2.5.

$$\langle Y^r, \Lambda_r(\mathcal{B}_d(\xi))\rangle \quad = \quad \langle \Lambda_r^*(Y^r), \mathcal{B}_d(\xi)\rangle \tag{3.35}$$

1. $\Longrightarrow$: Assume that $s \in \Sigma_d^{'}(\Xi)$. Then:

$$\begin{array}{rcll} s(\xi) & = & \sum_{r=0}^{R} s_r(\xi) w_r(\xi) & s_r(\xi) \in \Sigma_{2\tilde{d}_r}(\mathbb{R}^k) \\ & = & \sum_{r=0}^{R} \mathcal{B}_{\tilde{d}_r}(\xi)^\top Y^r \mathcal{B}_{\tilde{d}_r}(\xi) w_r^\top \mathcal{B}_{d_j}(\xi) & Y^r \succeq 0 \\ & = & \sum_{r=0}^{R} \left\langle Y^r \mathcal{B}_{\tilde{d}_r}(\xi)^\top, \mathcal{B}_{\tilde{d}_r}(\xi) w_r^\top \mathcal{B}_{d_j}(\xi) \right\rangle & Y^r \succeq 0 \\ & = & \sum_{r=0}^{R} \langle Y^r, \Lambda_r(\mathcal{B}_d(\xi))\rangle & Y^r \succeq 0 \\ & = & \sum_{r=0}^{R} \langle \Lambda_r^*(Y^r), \mathcal{B}_d(\xi)\rangle & Y^r \succeq 0 \end{array} \tag{3.36}$$

Since $s_r(\xi) \in \Sigma_{2\tilde{d}_r}(\mathbb{R}^k)$, then $Y^r \succeq 0$, $r = 0, ..., R$ according to Theorem 2.3.

2. $\Longleftarrow$: Assume that 2. is true and define the polynomial $s(\xi) = \sum_{r=0}^{R} \langle \Lambda_r^*(Y^r), \mathcal{B}_d(\xi) \rangle$. A reverse reasoning to the one for $\Longrightarrow$ is applied. Since $Y^r \succeq 0$, then $s_r(\xi) \in \Sigma_{2\tilde{d}_r}(\mathbb{R}^k)$, $r = 0, ..., R$ according to Theorem 2.3.

■

Above proposition shows that $\Sigma'_d(\Xi)$ has a tractable characterization.

$$\Sigma'_d(\Xi) = \left\{ s \in \mathbb{R}_d[\xi] : \begin{array}{l} s(\xi) = \sum_{r=0}^{R} \Lambda_r^*(Y^r)^\top \mathcal{B}_d(\xi) \\ Y^j \succeq 0,\ j = 1, ..., J \end{array} \right. \tag{3.37}$$

It is convenient to define cones $\mathcal{P}_d^m(\Xi)$ and $\Sigma_d^m(\Xi)$ as the sets of all $m \times s(k,d)$-matrices whose rows are all elements of $\mathcal{P}_d(\Xi)$ and $\Sigma_d(\Xi)$, respectively.

Problem 3.25 can then be approximated by the following computationally tractable problem

$$\begin{array}{ll} inf & tr(T_{\theta,d-\eta}^\top C^\top X M_{d-\eta}) \\ s.t. & X \in \mathbb{R}^{n \times s(k,d-\eta)}, S \in \mathbb{R}^{n \times s(k,d)} \\ & \left. \begin{array}{l} A *_{(\eta,d-\eta)}^{(3)} X + S = BT_{\theta,d} \\ S \in \Sigma_d^m(\Xi) \end{array} \right\} \mathbb{P} - a.s. \end{array} \tag{3.38}$$

Proposition 3.5: For problems 3.27 and 3.38 the following holds: inf 3.38 $\geq$ inf 3.27.

Proof: As problem 3.38 is obtained from 3.27 by Approximation P2*, the feasible set of 3.38 is smaller than the feasible set of 3.27. It is thus clear that solution of problem 3.38 provides an upper bound on the optimal value of problem 3.27. The feasible sets, and thus the solutions of problems 3.27 and 3.38, are in general equal under Assumption 2.1 if d is chosen to be large enough due to Theorem 2.6. ■

Problem 3.38 represents a tractable upper bound approximation of problem 3.8. In the next section we will focus on the tractable lower bound approximation.

3.4.3 One-stage Dual Polynomial Decision Rules

A similar approach as used for the upper bound approximation, can also be applied to the lower bound approximation. A related approach has been proposed in [3, 7].

We denote by $inf_{x,s}$ the infimum operator over all $x \in \mathcal{L}_{k,n}$ and over all $s \in \mathcal{L}_{k,m}$ that are almost surely non-negative, by sup_y the supremum operator over all $y \in \mathcal{L}_{k,m}$ and by sup_Y the supremum operator over all $Y \in \mathbb{R}^{m \times s(k,d)}$.

By applying the Lagrangian duality (remember definition in Section 2.2) to problem 3.8 the following equivalent problem is obtained

$$\underset{x,s}{inf}\ \underset{y}{sup}\ \ \mathbb{E}(c(\xi)^\top x(\xi) + y(\xi)^\top (A(\xi)x(\xi) + s(\xi) - b(\xi))) \tag{3.39}$$

Approximation D1: Consider only dual decision rules that are representable as polynomial functions of the uncertain parameters of a degree at most d, i.e. $y(\xi) = Y\mathcal{B}_d(\xi)$ for some matrix $Y \in \mathbb{R}^{m \times s(k,d)}$. We also require that $d \geq max\{\mu, \theta, d_0, d_1, ..., d_R\}$.

Note, that Approximation D1 is supported by the Weierstrass theorem and as $d \longrightarrow \infty$ the above formulation characterizes all continuous measurable functions.

Using Approximation D1, we write

$$\begin{aligned} & \underset{x,s}{inf}\ \underset{y}{sup}\, \mathbb{E}(c(\xi)^\top x(\xi) + y(\xi)^\top (A(\xi)x(\xi) + s(\xi) - b(\xi))) \quad \geq \\ & \underset{x,s}{inf}\ \underset{Y}{sup}\, \mathbb{E}(c(\xi)^\top x(\xi) + (Y\mathcal{B}_d(\xi))^\top \cdot (A(\xi)x(\xi) + s(\xi) - b(\xi))) \end{aligned} \tag{3.40}$$

Carrying out the inner maximization in the objective function, the following semi-infinite problem is obtained.

$$\begin{aligned} inf \quad & \mathbb{E}(c(\xi)^\top x(\xi)) \\ s.t. \quad & x \in \mathcal{L}_{k,n},\ s \in \mathcal{L}_{k,m} \\ & \mathbb{E}((A(\xi)x(\xi) + s(\xi) - b(\xi))\mathcal{B}_d(\xi)^\top) = 0 \\ & s(\xi) \geq 0\, \mathbb{P} - a.s. \end{aligned} \tag{3.41}$$

Proposition 3.6: For problems 3.8 and 3.41 following holds: inf 3.41 $\leq$ inf 3.8.

Proof: By comparison of problems 3.8 and 3.41 it is obvious that any (x, s) that is feasible in problem 3.8 will also satisfy the less restrictive constraints of problem 3.41. Thus, problem 3.41 is a relaxation of problem 3.8 and its optimal value provides a lower bound on the optimal value of problem 3.8. Remember also that the Lagrangian dual problem is defined as a maximization problem and thus it approaches the optimal solution from below. If decision rules of the dual problem are approximated by polynomial functions, this limits the feasible set and clearly leads to a lower bound approximation. Since only Approximation D1 was applied, equality in general holds as $d \longrightarrow \infty$ due to the Weierstrass theorem. ■

Problem 3.41 involves finitely many equality constraints, but involves a continuum of decision variables and inequality constraints. In order to obtain a tractable representation for problem 3.41, we introduce new decision variables $X \in \mathbb{R}^{n \times s(k,d-\eta)}$ and $S \in \mathbb{R}^{m \times s(k,d)}$, which are defined through the following constraints

$$\begin{aligned} & XM_{d-\eta} = \mathbb{E}(x(\xi)\mathcal{B}_{d-\eta}(\xi)^\top) \\ & (A *^{(3)}_{(\eta,d-\eta)} X)M_d = \mathbb{E}(A(\xi)x(\xi)\mathcal{B}_d(\xi)^\top) \\ & SM_d = \mathbb{E}(s(\xi)\mathcal{B}_d(\xi)^\top) \end{aligned} \tag{3.42}$$

Note, that decision variables $X \in \mathbb{R}^{n \times s(k,d-\eta)}$ and $S \in \mathbb{R}^{m \times s(k,d)}$ are uniquely determined by decision rules $x \in \mathcal{L}_{k,n}$ and $s \in \mathcal{L}_{k,m}$, since moment matrix is invertible (Proposition 3.7). Moreover, note that the first constraint does not imply the second constraint or vice versa.

Proposition 3.7: $M_d \succ 0$ and invertible.

Proof: $M_d \succeq 0$ due to Proposition 2.7. Let $f(\xi) \in \mathbb{R}_d[\xi]$ with non-zero coefficients, then

$$\langle f, M_d(y)f \rangle = \int f(\xi)^2 \mu_y(d\xi) \geq 0. \tag{3.43}$$

Since Ξ has non empty interior[1] (Assumption 3.2*), equality holds only when all coefficients of f are 0. Thus $M_d \succ 0$ and invertible. ■

[1]Remember that also interval $[0, 0]$ has a non empty interior in $\mathbb{R}$.

Using 3.42 we rewrite the objective function.

$$
\begin{aligned}
\mathbb{E}(c(\xi)^\top x(\xi)) &= \mathbb{E}((C\mathcal{B}_\theta(\xi))^\top x(\xi)) \\
&= \mathbb{E}((CT_{\theta,d-\eta}\mathcal{B}_{d-\eta}(\xi))^\top x(\xi)) \\
&= \mathbb{E}(tr(CT_{\theta,d-\eta}\mathcal{B}_{d-\eta}(\xi)x(\xi)^\top)) \\
&= tr(CT_{\theta,d-\eta}\mathbb{E}(x(\xi)\mathcal{B}_{d-\eta}(\xi)^\top)^\top) \\
&= tr(CT_{\theta,d-\eta}(XM_{d-\eta})^\top) \\
&= tr((CT_{\theta,d-\eta}M_{d-\eta}^\top X^\top)^\top) \\
&= tr((M_{d-\eta}^\top X^\top)^\top (CT_{\theta,d-\eta})^\top) \\
&= tr((CT_{\theta,d-\eta})^\top (M_{d-\eta}^\top X^\top)^\top) \\
&= tr(T_{\theta,d-\eta}^\top C^\top XM_{d-\eta})
\end{aligned}
\tag{3.44}
$$

Using 3.42 and the fact that M_d is invertible, the equality constraints in problem 3.41 can be written as

$$
\begin{aligned}
& \mathbb{E}((A(\xi)x(\xi) + s(\xi) - b(\xi))\mathcal{B}_d(\xi)^\top) \\
= \; & \mathbb{E}(A(\xi)x(\xi)\mathcal{B}_d(\xi)^\top + s(\xi)\mathcal{B}_d(\xi)^\top - B\mathcal{B}_\theta(\xi)\mathcal{B}_d(\xi)^\top) \\
= \; & (A *_{(\eta,d-\eta)}^{(3)} X)M_d + SM_d - BT_{\theta,d}M_d \\
= \; & A *_{(\eta,d-\eta)}^{(3)} X + S - BT_{\theta,d}
\end{aligned}
\tag{3.45}
$$

Due to 3.44 and 3.45 we can reformulate problem 3.41 to problem 3.46.

$$
\begin{aligned}
inf \quad & tr(T_{\theta,d-\eta}^\top C^\top XM_{d-\eta}) \\
s.t. \quad & X \in \mathbb{R}^{n\times s(k,d-\eta)}, S \in \mathbb{R}^{n\times s(k,d)} \\
& A *_{(\eta,d-\eta)}^{(3)} X + S = BT_{\theta,d} \\
& \exists x \in \mathcal{L}_{k,n} : \left\{ \begin{array}{l} XM_{d-\eta} = \mathbb{E}(x(\xi)B_{d-\eta}(\xi)^\top) \\ (A *_{(\eta,d-\eta)}^{(3)} X)M_d = \mathbb{E}(A(\xi)x(\xi)\mathcal{B}_d(\xi)^\top) \end{array} \right\} \\
& \exists s \in \mathcal{L}_{k,m} : \left\{ \begin{array}{l} XM_d = \mathbb{E}(s(\xi)B_d(\xi)^\top) \\ s(\xi) \geq 0 \; \mathbb{P} - a.s. \end{array} \right\}
\end{aligned}
\tag{3.46}
$$

Lemma 3.8: The penultimate constraint in 3.46 is redundant and can be omitted without affecting the problem's feasibility set.

Proof: It is clear that for any $A \in \mathbb{R}^{m\times n\times s(k,\eta)}$ and $X \in \mathbb{R}^{n\times s(k,d-\eta)}$, decision rule $x(\xi) = XB_{d-\eta}(\xi)$ satisfies the penultimate constraint.

- 1. equation:

$$\begin{aligned}\mathbb{E}(x(\xi)B_{d-\eta}(\xi)^\top) &= \mathbb{E}(XB_{d-\mu}(\xi)B_{d-\eta}(\xi)^\top)\\ &= X\mathbb{E}(B_{d-\mu}(\xi)B_{d-\eta}(\xi)^\top)\\ &= XM_{d-\eta}\end{aligned} \tag{3.47}$$

- 2. equation:

$$\begin{aligned}\mathbb{E}(A(\xi)XB_{d-\eta}(\xi)B_d(\xi)^\top) &= \mathbb{E}((A \cdot^{(3)} B_\eta(\xi))(XB_{d-\eta}(\xi))B_d(\xi)^\top)\\ &= \mathbb{E}((A *^{(3)}_{(\eta,d-\eta)} X)B_d(\xi)B_d(\xi)^\top)\\ &= (A *^{(3)}_{(\eta,d-\eta)} X)\mathbb{E}(B_d(\xi)B_d(\xi)^\top)\\ &= (A *^{(3)}_{(\eta,d-\eta)} X)M_d\end{aligned} \tag{3.48}$$

In the second line relationship 3.26 was used. ■

The last constraint in problem 3.46 involves the solution of m multidimensional moment problems.

Let $\mathcal{M}_d(\Xi)$ denote the cone of moment sequences with a representing measure supported on Ξ, where $\mathcal{N}$ denotes the set of non-negative Borel measures supported on Ξ.

$$\mathcal{M}_d(\Xi) := \left\{ \quad y \in \mathbb{R}^{n\times s(k,d)} : \quad \begin{array}{l} \exists\mu \in \mathcal{N} \\ y = \int_\Xi \mathcal{B}_d(\xi)\mu(d\xi) \end{array} \quad \right\} \tag{3.49}$$

Similarly, we define the cone

$$\mathcal{M}_d^+(\Xi) := \{ \quad y \in \mathbb{R}^{n\times s(k,d)} : \quad \Lambda_r(y) \succeq 0 \quad r = 0, ..., R \quad \} \tag{3.50}$$

where the mappings Λ_r are defined as in Proposition 3.4.

Proposition 3.9:

1. $\mathcal{P}_d(\Xi)$ and $\mathcal{M}_d(\Xi)$ are dual to each other.
2. $\Sigma'_d(\Xi)$ and $\mathcal{M}_d^+(\Xi)$ are dual to each other.
3. $\mathcal{M}_d(\Xi) \subseteq \mathcal{M}_d^+(\Xi)$.

Proof: Proof is in [7, 10]. Points 1 and 2 follow from the duality theory:

1. If $p \in \mathcal{P}_d(\Xi)$ and $y \in \mathcal{M}_d(\Xi)$ then

$$p^\top y = \int_\Xi p^\top \mathcal{B}_d(\xi)\mu(d\xi) = \int_\Xi p(\xi)\mu(d\xi) \geq 0. \tag{3.51}$$

   Note that by definition $p(\xi) \geq 0$ on Ξ and that $\mu(d\xi) \geq 0$ since it is a Borel measure. It is clear that the above reasoning holds for all $p \in \mathcal{P}_d(\Xi)$ and $y \in \mathcal{M}_d(\Xi)$.

2. If $p \in \Sigma'_d(\Xi)$ and $y \in \mathcal{M}_d^+(\Xi)$ then

$$p^\top y = \sum_{r=0}^{R} \langle \Lambda_r^*(Y^r), y\rangle = \sum_{r=0}^{R} \langle Y^r, \Lambda_r(y)\rangle \geq 0. \tag{3.52}$$

   Note that $Y^r \succeq 0$, $r = 0, ..., R$ since $p \in \Sigma'_d(\Xi)$ according to Proposition 3.4. $\Lambda_r(y) \succeq 0$, $r = 0, ..., R$ is true according to definition 3.50 of $\mathcal{M}_d^+(\Xi)$.

3. It is clear that $\Sigma'_d(\Xi) \subseteq \mathcal{P}_d(\Xi)$. Due to duality we can see that $\mathcal{M}_d(\Xi) = (\mathcal{P}_d(\Xi))^* \subseteq (\Sigma'_d(\Xi))^* = \mathcal{M}_d^+(\Xi)$.

■

By observing problem 3.46, we note that the last constraint requires each component $s_i(\xi)$, $i = 1, ..., m$, of the vector-valued function $s(\xi)$ to be the density function of a measure $\mu_i \in \mathcal{N}$ whose moments coincide with the i-th row of SM_d. Thus i-th row of SM_d must be contained in $\mathcal{M}_d(\Xi)$. Verifying the membership of $s_i(\xi)$, $i = 1, ..., m$ in the cone $\mathcal{M}_d(\Xi)$ is NP-hard [8]. However, we have shown in Section 2.6 that verifying the membership of $s_i(\xi)$, $i = 1, ..., m$ in the cone $\mathcal{M}_d^+(\Xi)$ can be solved through a tractable SDP.

Approximation D2: Verifying membership of $\mathcal{M}_d(\Xi)$ can be approximated by verifying membership of $\mathcal{M}_d^+(\Xi)$.

As we discussed in Section 2.6, the above approximation was justified by Lasserre.

It is convenient to define $\mathcal{M}_d^m(\Xi)$ and $\mathcal{M}_d^{m+}(\Xi)$ as the cones of all $m \times s(k, d)$-matrices whose rows are all contained in $\mathcal{M}_d(\Xi)$ and $\mathcal{M}_d^+(\Xi)$, respectively. Based on this reasoning, we approximate problem 3.46 by the following tractable problem

$$\begin{aligned} inf \quad & tr(T_{\theta,d-\eta}^\top C^\top X M_{d-\eta}) \\ s.t. \quad & X \in \mathbb{R}^{n \times s(k,d-\eta)}, S \in \mathbb{R}^{n \times s(k,d)} \\ & A *_{(\eta,d-\eta)}^{(3)} X + S = BT_{\theta,d} \\ & SM_d \in \mathcal{M}_d^{m+}(\xi) \end{aligned} \tag{3.53}$$

Proposition 3.10: For problems 3.46 and 3.53 the following holds: inf 3.53 $\leq$ inf 3.46.

Proof: Problem 3.53 is obtained from 3.46 by Approximation D2. We know that the feasible set of problem 3.53 is bigger than the feasible set of problem 3.46 since $\mathcal{M}_d(\Xi) \subseteq \mathcal{M}_d^+(\Xi)$ due to Preposition 2.10. It is thus clear that solution of problem 3.53 provides a lower bound on the optimal value of problem 3.46. Lasserre showed that under Assumption 2.1 equality holds for a d chosen to be large enough. However, if Assumption 2.1 is not satisfied then equality holds as $d \longrightarrow \infty$. ■

Problem 3.53 represents a computationally tractable approximation of problem 3.8. In the next section, we will focus on multistage stochastic problems.

3.4.4 Multistage Primal Polynomial Decision Rules

In this section, polynomial decision rules will be extended to obtain computationally tractable approximations of multistage stochastic programs. The goal is to formulate two tractable semidefinite programs, obtained by restricting the primal and the dual decision rules to the polynomial functions, which, when solved, provide an upper and a lower bound on the optimal value of the stochastic program.

Assumption 3.4: We require that the objective function coefficients and constraint functions depend polynomially on the uncertain parameters ξ^t. Namely, $c_t(\xi^t) = C_t P_{\theta,t} \mathcal{B}_\theta(\xi)$ for some $C_t \in \mathbb{R}^{n_t \times s(k^t,\theta)}$, $b_t(\xi^t) = B_t P_{\theta,t} \mathcal{B}_\theta(\xi)$ for some $B \in \mathbb{R}^{m_t \times s(k^t,\theta)}$ and $A_{ts}(\xi^t) = A_{ts} \cdot^{(3)} P_{\eta^s,t} \mathcal{B}_{\eta^s}(\xi)$ for some $A_{ts} \in \mathbb{R}^{m_t \times n_s \times s(k,\eta^s)}$. In order to keep the notation in a matrix form we introduced truncation operators $P_{d,t} : \mathbb{R}^{s(k,d)} \longrightarrow \mathbb{R}^{s(k^t,d)}$ for any $t \in \mathbb{T}$ and $d \in \mathbb{N}_0$ that map the monomial basis $\mathcal{B}_d(\xi)$ to the reduced basis $\mathcal{B}_d(\xi^t)$. We defined $\eta^s := max\left\{\eta^{ts}\right\}$, $t = s, ..., T$, for all $s = 1, ..., T$, where η^{ts}

corresponds to the actual degree of the polynomial $A_{ts}(\xi^t)$. A visual representation is given below

$$\begin{array}{cccc} \eta^{11} & & & \\ \eta^{21} & \eta^{22} & & \\ \vdots & \vdots & \ddots & \\ \eta^{T1} & \eta^{T2} & \cdots & \eta^{TT} \\ - & - & - & - \\ \eta^{1} & \eta^{2} & \cdots & \eta^{T} \end{array} \tag{3.54}$$

Note that this assumption is equivalent to Assumption 3.2 for one-stage stochastic problems.

Approximation MP1: We approximate the decision rules of problem 3.14 by polynomial functions of a degree at most d, where $d \geq max\,\{\eta^s, \theta, d_0, ..., d_R\}$, $s = 1, ..., T$. The decision and slack variables can thus be written as $x_t(\xi^t) = X_t P_{d-\eta^t,t} B_{d-\eta^t}(\xi)$ for some $X_t \in \mathbb{R}^{n_t \times s(k^t, d-\eta^t)}$ and $s_t(\xi^t) = S_t P_{d,t} B_d(\xi)$ for some $S_t \in \mathbb{R}^{m_t \times s(k^t,d)}$. To ensure that this approximation leads to a tractable program, we require that $\mathbb{E}_t(\mathcal{B}_d(\xi))$ is essentially polynomial in ξ^t, that is, $\mathbb{E}_t(\mathcal{B}_d(\xi)) = M_t P_{d,t} \mathcal{B}_d(\xi)$ $\mathbb{P}$-a.s for some matrix $M_t \in \mathbb{R}^{s(k,d)\times s(k^t,d)}$ for $t \in \mathbb{T}$.

Note, that Approximation MP1 is supported by the Weierstrass theorem and as $d \longrightarrow \infty$ the above formulation characterizes all continuous measurable functions.

By applying Approximation MP1 to the objective function of problem 3.14 we obtain

$$\begin{aligned} \mathbb{E}(\textstyle\sum_{t=1}^T c_t(\xi^t)^\top x_t(\xi^t)) &= \mathbb{E}(\textstyle\sum_{t=1}^T (C_t P_{\theta,t} \mathcal{B}_\theta(\xi))^\top X_t P_{d-\eta^t,t} \mathcal{B}_{d-\eta^t}(\xi)) \\ &= \mathbb{E}(\textstyle\sum_{t=1}^T (C_t P_{\theta,t} T_{\theta,d-\eta^t} \mathcal{B}_{d-\eta^t}(\xi))^\top X_t P_{d-\eta^t,t} \mathcal{B}_{d-\eta^t}(\xi)) \\ &= \mathbb{E}(\textstyle\sum_{t=1}^T tr(C_t P_{\theta,t} T_{\theta,d-\eta^t} \mathcal{B}_{d-\eta^t}(\xi) \mathcal{B}_{d-\eta^t}(\xi)^\top P_{d-\eta^t,t}^\top X_t^\top)) \\ &= \textstyle\sum_{t=1}^T tr(C_t P_{\theta,t} T_{\theta,d-\eta^t} \mathbb{E}(\mathcal{B}_{d-\eta^t}(\xi) \mathcal{B}_{d-\eta^t}(\xi)^\top) P_{d-\eta^t,t}^\top X_t^\top) \\ &= \textstyle\sum_{t=1}^T tr(C_t P_{\theta,t} T_{\theta,d-\eta^t} M_{d-\eta^t} P_{d-\eta^t,t}^\top X_t^\top). \end{aligned} \tag{3.55}$$

Similarly, by applying Approximation MP1 to the constraints we obtain

$$\begin{aligned} & \textstyle\sum_{s=1}^t A_{ts}(\xi^t) x_s(\xi^s) + s_t(\xi^t) &&= b_t(\xi^t) \\ \Longleftrightarrow\quad & \textstyle\sum_{s=1}^t (A_{ts} \cdot^{(3)} P_{\eta^s,t} \mathcal{B}_{\eta^s}(\xi)) X_s P_{d-\eta^s,s} \mathcal{B}_{d-\eta^s}(\xi) + S_t P_{d,t} \mathcal{B}_d(\xi) &&= B_t P_{\theta,t} \mathcal{B}_\theta(\xi) \\ \Longleftrightarrow\quad & \textstyle\sum_{s=1}^t (A_{ts} *^{(3)}_{(\eta^s,d-\eta^s)} X_s P_{d-\eta^s,s} P_{d-\eta^s,t}^\top) P_{d,t} \mathcal{B}_d(\xi) + S_t P_{d,t} \mathcal{B}_d(\xi) &&= B_t P_{\theta,t} T_{\theta,d} \mathcal{B}_d(\xi) \\ \Longleftrightarrow\quad & \textstyle\sum_{s=1}^t (A_{ts} *^{(3)}_{(\eta^s,d-\eta^s)} X_s P_{d-\eta^s,s} P_{d-\eta^s,t}^\top) P_{d,t} + S_t P_{d,t} &&= B_t P_{\theta,t} T_{\theta,d} \end{aligned} \tag{3.56}$$

Note that $X_s P_{d-\eta^s,s} P_{d-\eta^s,t}^\top$ maps the basis $\mathcal{B}_{d-\eta^s}(\xi^t)$ to the basis $\mathcal{B}_{d-\eta^s}(\xi^s)$. This is needed because the polynomial multiplication is defined for polynomials in the same basis. In the penultimate step there are infinitely many constraints, since it must hold for each possible realisation of the uncertain parameters $\xi \in \Xi$. Thus, a vector-valued polynomial must vanish identically on a set with nonempty interior Ξ. This is possible if and only if all the coefficients of the polynomial vanish.

By combining 3.55 and 3.56, the following problem is obtained.

$$\begin{aligned}
inf \quad & \textstyle\sum_{t=1}^{T} tr(C_t P_{\theta,t} T_{\theta,d-\eta^t} M_{d-\eta^t} P_{d-\eta^t,t}^{\top} X_t^{\top}) \\
s.t. \quad & X_t \in \mathbb{R}^{n_t \times s(k^t,d-\eta^t)},\ S_t \in \mathbb{R}^{n_t \times s(k^t,d)} \\
& \left.\begin{aligned}
&\textstyle\sum_{s=1}^{t} (A_{ts} *^{(3)}_{(\eta^s,d-\eta^s)} X_s P_{d-\eta^s,s} P_{d-\eta^s,t}^{\top}) P_{d,t} + S_t P_{d,t} = B_t P_{\theta,t} T_{\theta,d} \\
&S_t P_{d,t} \mathcal{B}_d(\xi) \geq 0
\end{aligned}\right\} \mathbb{P} - a.s.,\ \forall t \in \mathbb{T}
\end{aligned} \tag{3.57}$$

Proposition 3.11: For problems 3.14 and 3.57 the following holds: inf 3.57 $\geq$ inf 3.14.

Proof: As problem 3.57 is obtained from problem 3.14 by Approximation MP1, the feasible set of problem 3.57 is smaller than the feasible set of problem 3.14. It is thus clear that the solution of problem 3.57 provides an upper bound on the optimal value of problem 3.14. The feasible sets, and thus the solutions, are in general equal as $d \longrightarrow \infty$ due to the Weierstrass theorem. ■

Formulation 3.57 is not yet tractable because the inequality constraint requires that each component of a vector-valued polynomial $S_t P_{d,t} B_d(\xi)$, $t = 1, ..., T$ is non-negative on Ξ. We have already showed in the one-stage model how to handle such constraints.

We could apply Theorem 2.5, but the resulting degree could be too high. Thus, we again set that Assumption 2.1 holds and apply Theorem 2.6.

Approximation MP2*: For every polynomial $[S_t P_{d,t} B_d(\xi)]_i$, $t = 1, ..., T$ and $i = 1, ..., m^t$ positive on Ξ, 3.58 holds under Assumption 2.1 due to Theorem 2.6, if d is chosen to be large enough.

$$\forall \xi \in \Xi\ [S_t P_{d,t} B_d(\xi)]_i > 0 \Longleftrightarrow [S_t P_{d,t} B_d(\xi)]_i \in \Sigma'(\Xi). \tag{3.58}$$

Note that at least $d \geq max\{\eta^s, \theta, d_0, d_1, ..., d_R\}$, $s = 1, ..., T$ must hold.

The minimal degree d also coincides with the requirement in Approximation MP1. Note, however, that Assumption 2.1 can sometimes be violated and in this case Approximation MP2* does not characterize sufficiently all the polynomials positive on Ξ even as $d \longrightarrow \infty$.

By applying Approximation MP2*, we can rewrite problem 3.57 as

$$\begin{aligned}
inf \quad & \textstyle\sum_{t=1}^{T} tr(C_t P_{\theta,t} T_{\theta,d-\eta^t} M_{d-\eta^t} P_{d-\eta^t,t}^{\top} X_t^{\top}) \\
s.t. \quad & X_t \in \mathbb{R}^{n_t \times s(k^t,d-\eta^t)}, S_t \in \mathbb{R}^{n_t \times s(k^t,d)} \\
& \left.\begin{aligned}
&\textstyle\sum_{s=1}^{t} (A_{ts} *^{(3)}_{(\eta^s,d-\eta^s)} X_s P_{d-\eta^s,s} P_{d-\eta^s,t}^{\top}) P_{d,t} + S_t P_{d,t} = B_t P_{\theta,t} T_{\theta,d} \\
&S_t P_{d,t} \in \Sigma_d^m(\Xi)
\end{aligned}\right\} \forall t \in \mathbb{T}
\end{aligned} \tag{3.59}$$

Proposition 3.12: For problems 3.14 and 3.59 the following holds: inf 3.59 $\geq$ inf 3.57.

Proof: As problem 3.59 is obtained from problem 3.57 by Approximation MP2*, the feasible set of problem 3.57 is smaller than the feasible set of 3.59. It is thus clear that the solution of problem 3.59 provides an upper bound on the optimal value of problem 3.57. The feasible sets, and thus the solutions, are in general equal under Assumption 2.1 if d is chosen to be large enough due to Theorem 2.6. ■

Problem 3.59 represents a tractable upper bound approximation of problem 3.14. In the next section we discuss how a lower bound approximation for multistage stochastic problems is formulated.

3.4.5 Multistage Dual Polynomial Decision Rules

In order to obtain a tractable lower bound approximation of problem 3.14, we first rewrite it as a min-max problem similar to the Lagrangian dual presented in Section 2.2. An equivalent approach for linear decision rules has been proposed in [46]. For the t-th equality constraints $t \in \mathbb{T}$, we introduce a non-anticipative decision rule $y_t \in \mathcal{L}_{k^t,m_t}$. Problem 3.14 can thus be replaced by an equivalent problem

$$\underset{x_t,s_t\ \forall t\in\mathbb{T}}{inf} \quad \underset{y_t\ \forall t\in\mathbb{T}}{sup} \quad \mathbb{E}\left(\sum_{t=1}^{T} c_t(\xi^t)^\top x_t(\xi^t) + +y_t(\xi^t)^\top (\sum_{s=1}^{t} A_{ts}(\xi^t)x_s(\xi^s) + s_t(\xi^t) - b_t(\xi^t))\right). \tag{3.60}$$

Approximation MD1: Consider only dual decision rules that are represented as polynomial functions of the uncertain parameters of a degree at most d, i.e. $y_t(\xi^t) = Y_t P_{d,t}\mathcal{B}_d(\xi)$ for some $Y_t \in \mathbb{R}^{m_t\times s(k^t,d)}$ for all $t \in \mathbb{T}$.

Note, that Approximation MD1 is supported by the Weierstrass theorem and as $d \longrightarrow \infty$ the above formulation characterizes all continuous measurable functions.

After Approximation MD1, we write

$$\begin{aligned}
&\underset{x_t,s_t\ \forall t\in\mathbb{T}}{inf} \quad \underset{y_t\ \forall t\in\mathbb{T}}{sup}\ \mathbb{E}\left(\sum_{t=1}^{T} c_t(\xi^t)^\top x_t(\xi^t) + y_t(\xi^t)^\top \left(\sum_{s=1}^{t} A_{ts}(\xi^t)x_s(\xi^s) + s_t(\xi^t) - b_t(\xi^t)\right)\right) \geq \\
&\underset{x_t,s_t\ \forall t\in\mathbb{T}}{inf} \quad \underset{y_t\ \forall t\in\mathbb{T}}{sup}\ \mathbb{E}\left(\sum_{t=1}^{T} c_t(\xi^t)^\top x_t(\xi^t) + (Y_t P_{d,t}\mathcal{B}_d(\xi))^\top \left(\sum_{s=1}^{t} A_{ts}(\xi^t)x_s(\xi^s) + s_t(\xi^t) - b_t(\xi^t)\right)\right)
\end{aligned} \tag{3.61}$$

Carrying out the inner maximization in the objective function, we obtain the following semi-infinite problem

$$\begin{aligned}
inf \quad & \mathbb{E}(\sum_{t=1}^{T} c_t(\xi^t)^\top x_t(\xi^t)) \\
s.t. \quad & x_t \in \mathcal{L}_{k^t,n_t},\ s_t \in \mathcal{L}_{k^t,m_t} \quad \forall t \in \mathbb{T} \\
& \left.\begin{aligned} &\mathbb{E}((\sum_{s=1}^{t} A_{ts}(\xi^t)x_s(\xi^s) + s_t(\xi^t) - b_t(\xi^t))\mathcal{B}_d(\xi)^\top P_{d,t}^\top) = 0 \\ &s_t(\xi^t) \geq 0 \end{aligned}\right\} \mathbb{P} - a.s.,\ \forall t \in \mathbb{T}
\end{aligned} \tag{3.62}$$

Proposition 3.13: For problems 3.14 and 3.62 the following holds: inf 3.62 $\leq$ inf 3.14.

Proof: By comparison of problems 3.14 and 3.62 it is obvious that any (x_t, s_t), $t = 1, ..., T$ that is feasible in problem 3.14 will also satisfy the less restrictive constraints of problem 3.62. Thus, problem 3.62 is a relaxation of problem 3.14 and its optimal value provides a lower bound on the optimal value of 3.14. Remember also that the Lagrangian dual problem is defined as a maximization problem and thus it approaches the optimal solution from below. If decision rules of the dual problem are approximated by polynomial functions, this limits the feasible set and clearly leads to a lower bound approximation. Since only Approximation MD1 was applied, equality in general holds as $d \longrightarrow \infty$ due to the Weierstrass theorem. ■

Problem 3.62 involves finitely many equality constraints, but involves a continuum of decision variables and inequality constraints. In order to obtain a tractable representation for problem 3.62 we introduce new decision variables $X_t \in \mathbb{R}^{n_t\times s(k^t,d-\eta^t)}$ and $S_t \in \mathbb{R}^{m_t\times s(k^t,d)}$, $t \in \mathbb{T}$ which are defined through the following constraints

$$X_t P_{d-\eta^t,t} M_{d-\eta^t} = \mathbb{E}(x_t(\xi^t)\mathcal{B}_{d-\eta^t}(\xi)^\top)$$

$$(A_{ts} *^{(3)}_{(\eta^s,d-\eta^s)} X_s P_{d-\eta^s,s} P^\top_{d-\eta^s,t}) P_{d,t} M_d = \mathbb{E}(A_{st}(\xi^t) x_s(\xi^s)\mathcal{B}_d(\xi)^\top) \tag{3.63}$$

$$S_t P_{d,t} M_d = \mathbb{E}(s_t(\xi^t)\mathcal{B}_d(\xi)^\top)$$

In the following lemma we proof that the new constraints 3.63 do not restrict the choices of $x_t \in \mathcal{L}_{k^t,n_t}$ and $s_t \in \mathcal{L}_{k^t,m_t}$.

Lemma 3.14: For any given $x_t \in \mathcal{L}_{k^t,n_t}$ and $s_t \in \mathcal{L}_{k^t,m_t}$ there exist unique matrices X_t and S_t satisfying 3.63.

Proof:

1. Define $S_t \in \mathbb{R}^{m_t \times s(k^t,d)}$ through

$$\mathbb{E}(s_t(\xi^t)\mathcal{B}_d(\xi)^\top) P^\top_{d,t} = S_t P_{d,t} M_d P^\top_{d,t} \tag{3.64}$$

Since $P_{d,t} M_d P^\top_{d,t}$ is a principal submatrix of M_d, it is invertible. Thus, S_t is uniquely defined by

$$S_t = \mathbb{E}(s_t(\xi^t)\mathcal{B}_d(\xi)^\top) P^\top_{d,t} (P_{d,t} M_d P^\top_{d,t})^{-1} \tag{3.65}$$

Recall that $\mathbb{E}_t(\mathcal{B}_d(\xi)) = M_t P_{d,t} \mathcal{B}_d(\xi)$ $\mathbb{P}$-a.s. We write

$$\begin{aligned}
\mathbb{E}(s_t(\xi^t)\mathcal{B}_d(\xi)^\top) &= \mathbb{E}(s_t(\xi^t)\mathbb{E}_t(\mathcal{B}_d(\xi))^\top) \\
&= \mathbb{E}(s_t(\xi^t)\mathcal{B}_d(\xi)^\top) P^\top_{d,t} M_t^\top \\
&= S_t P_{d,t} M_d P^\top_{d,t} M_t^\top \\
&= S_t \mathbb{E}(P_{d,t}\mathcal{B}_d(\xi)\mathcal{B}_d(\xi)^\top P^\top_{d,t}) M_t^\top \\
&= S_t \mathbb{E}(P_{d,t}\mathcal{B}_d(\xi)\mathbb{E}_t(\mathcal{B}_d(\xi)^\top)) \\
&= S_t P_{d,t} \mathbb{E}(\mathcal{B}_d(\xi)\mathcal{B}_d(\xi)^\top) \\
&= S_t P_{d,t} M_d
\end{aligned} \tag{3.66}$$

2. A similar argumentation holds also for $x_t \in \mathcal{L}_{k^t,n_t}$. ■

Due to 3.63 the following existence constraints appear.

$$\exists x_t \in \mathcal{L}_{k^t,n_t} : \left\{ \begin{array}{l} X_t P_{d-\eta^t,t} M_{d-\eta^t} = \mathbb{E}(x_t(\xi^t)\mathcal{B}_{d-\eta^t}(\xi)^\top) \\ (A_{ts} *^{(3)}_{(\eta^s,d-\eta^s)} X_s P_{d-\eta^s,s} P^\top_{d-\eta^s,t}) P_{d,t} M_d = \mathbb{E}(A_{st}(\xi^t) x_s(\xi^s)\mathcal{B}_d(\xi)^\top) \end{array} \right\} \tag{3.67}$$

$$\exists s_t \in \mathcal{L}_{k^t,m_t} : \left\{ \begin{array}{l} S_t P_{d,t} M_d = \mathbb{E}(s_t(\xi^t)\mathcal{B}_d(\xi)^\top) \\ s_t(\xi^t) \geq 0 \, \mathbb{P} - a.s. \end{array} \right\} \tag{3.68}$$

Lemma 3.15: Constraint 3.67 is redundant and can be omitted without affecting the problem's feasibility set.

Proof: It is clear that for any $A_{ts} \in \mathbb{R}^{m_t \times n_s \times s(k,\eta^s)}$ and $X_t \in \mathbb{R}^{n_t \times s(k^t, d-\eta^t)}$, the decision rule $x_t(\xi^t) = X_t P_{d-\eta^t,t} \mathcal{B}_{d-\eta^t}(\xi) \in \mathcal{L}_{k^t,n_t}$ satisfies it.

- 1. equation:

$$\begin{aligned} \mathbb{E}(x_t(\xi^t)\mathcal{B}_{d-\eta^t}(\xi)^\top) &= \mathbb{E}\left(X_t P_{d-\eta^t,t}\mathcal{B}_{d-\eta^t}(\xi)\mathcal{B}_{d-\eta^t}(\xi)^\top\right) \\ &= X_t P_{d-\eta^t,t}\mathbb{E}(\mathcal{B}_{d-\eta^t}(\xi)\mathcal{B}_{d-\eta^t}(\xi)^\top) \\ &= X_t P_{d-\eta^t,t} M_{d-\eta^t} \end{aligned} \tag{3.69}$$

- 2. constraint

$$\begin{aligned} &\mathbb{E}\left(A_{st}(\xi^t)x_s(\xi^s)\mathcal{B}_d(\xi)^\top\right) \\ =\ &\mathbb{E}\left((A_{st} \cdot^{(3)} P_{\eta^s,t}\mathcal{B}_{\eta^s}(\xi))(X_s P_{d-\eta^s,s}\mathcal{B}_{d-\eta^s}(\xi))\mathcal{B}_d(\xi)^\top\right) \\ =\ &(A_{ts} *^{(3)}_{(\eta^s,d-\eta^s)} X_s P_{d-\eta^s,s} P^\top_{d-\eta^s,t}) P_{d,t}\mathbb{E}(\mathcal{B}_d(\xi)\mathcal{B}_d(\xi)^\top) \\ =\ &(A_{ts} *^{(3)}_{(\eta^s,d-\eta^s)} X_s P_{d-\eta^s,s} P^\top_{d-\eta^s,t}) P_{d,t} M_d \end{aligned} \tag{3.70}$$

  Note that $X_s P_{d-\eta^s,s} P^\top_{d-\eta^s,t}$ maps the basis $\mathcal{B}_{d-\eta^s}(\xi^t)$ to the basis $\mathcal{B}_{d-\eta^s}(\xi^s)$. This is needed because the polynomial multiplication is defined for polynomials in the same basis. ■

Using equations 3.63, we can rewrite the objective function of problem 3.62 as

$$\begin{aligned} \mathbb{E}(\textstyle\sum_{t=1}^T c_t(\xi^t)^\top x_t(\xi^t)) &= \mathbb{E}(\textstyle\sum_{t=1}^T (C_t P_{\theta,t} B_\theta(\xi))^\top x_t(\xi^t)) \\ &= \mathbb{E}(\textstyle\sum_{t=1}^T tr(C_t P_{\theta,t} B_\theta(\xi) x_t(\xi^t))) \\ &= \mathbb{E}(\textstyle\sum_{t=1}^T tr(C_t P_{\theta,t} T_{\theta,d-\eta^t} B_{d-\eta^t}(\xi) x_t(\xi^t))) \\ &= \textstyle\sum_{t=1}^T tr(C_t P_{\theta,t} T_{\theta,d-\eta^t}\mathbb{E}(B_{d-\eta^t}(\xi) x_t(\xi^t))) \\ &= \textstyle\sum_{t=1}^T tr(C_t P_{\theta,t} T_{\theta,d-\eta^t}(X_t P_{d-\eta^t,t} M_{d-\eta^t})^\top) \\ &= \textstyle\sum_{t=1}^T tr(C_t P_{\theta,t} T_{\theta,d-\eta^t} M_{d-\eta^t} P^\top_{d-\eta^t,t} X_t^\top) \end{aligned} \tag{3.71}$$

and the equality constraints as

$$\begin{aligned} &\mathbb{E}((\textstyle\sum_{s=1}^t A_{ts}(\xi^t)x_s(\xi^s) + s_t(\xi^t) - b_t(\xi^t))\mathcal{B}_d(\xi)^\top P^\top_{d,t}) \\ =\ &\mathbb{E}(\textstyle\sum_{s=1}^t A_{ts}(\xi^t)x_s(\xi^s)\mathcal{B}_d(\xi)^\top P^\top_{d,t} + s_t(\xi^t)\mathcal{B}_d(\xi)^\top P^\top_{d,t} - B_t P_{\theta,t}\mathcal{B}_\theta(\xi)\mathcal{B}_d(\xi)^\top P^\top_{d,t}) \\ =\ &\textstyle\sum_{s=1}^t (A_{ts} *^{(3)}_{(\eta^s,d-\eta^s)} X_s P_{d-\eta^s,s} P^\top_{d-\eta^s,t}) P_{d,t} M_d P^\top_{d,t} + S_t P_{d,t} M_d P^\top_{d,t} - B_t P_{\theta,t} T_{\theta,d}\mathbb{E}(\mathcal{B}_d(\xi)\mathcal{B}_d(\xi)^\top) P^\top_{d,t} \\ =\ &\textstyle\sum_{s=1}^t (A_{ts} *^{(3)}_{(\eta^s,d-\eta^s)} X_s P_{d-\eta^s,s} P^\top_{d-\eta^s,t}) P_{d,t} M_d P^\top_{d,t} + S_t P_{d,t} M_d P^\top_{d,t} - B_t P_{\theta,t} T_{\theta,d} M_d P^\top_{d,t} \end{aligned} \tag{3.72}$$

It is possible to additionally simplify 4.7 by taking into account also some properties of the operators $P_{d,t}$, $t \in \mathbb{T}$ and $d \in \mathbb{N}$.

1. $P_{\theta,t}T_{\theta,d} = P_{\theta,t}T_{\theta,d}P_{d,t}^{\top}P_{d,t}$. Note that $P_{\theta,t}T_{\theta,d} : \mathbb{R}^{s(k,d)} \longrightarrow \mathbb{R}^{s(k,\theta)} \longrightarrow \mathbb{R}^{s(k^t,\theta)}$ and $P_{\theta,t}T_{\theta,d}P_{d,t}^{\top}P_{d,t} : \mathbb{R}^{s(k,d)} \longrightarrow \mathbb{R}^{s(k^t,d)} \longrightarrow \mathbb{R}^{s(k,d)} \longrightarrow \mathbb{R}^{s(k,\theta)} \longrightarrow \mathbb{R}^{s(k^t,\theta)}$. It is clear that the right hand side operator is always lossless with respect to the left hand side operator.
2. $P_{d,t}M_dP_{d,t}^{\top}$ is a principal submatrix of M_d and thus invertible.
3. $(P_{d,t}M_dP_{d,t}^{\top})^{-1} = (P_{d,t}M_dP_{d,t}^{\top})^{-1}P_{d,t}P_{d,t}^{\top}$. This clearly holds because $I = P_{d,t}P_{d,t}^{\top}$.

Having the above properties in mind, we can multiply the last line in 4.7 from the right by $(P_{d,t}M_dP_{d,t}^{\top})^{-1}P_{d,t}P_{d,t}^{\top}$. We will analyze each part separately

$$\begin{aligned}
& \textstyle\sum_{s=1}^{t}(A_{ts} *^{(3)}_{(\eta^s,d-\eta^s)} X_sP_{d-\eta^s,s}P_{d-\eta^s,t}^{\top})P_{d,t}M_dP_{d,t}^{\top} \\
= & \textstyle\sum_{s=1}^{t}(A_{ts} *^{(3)}_{(\eta^s,d-\eta^s)} X_sP_{d-\eta^s,s}P_{d-\eta^s,t}^{\top})P_{d,t}M_dP_{d,t}^{\top}(P_{d,t}M_dP_{d,t}^{\top})^{-1}P_{d,t}P_{d,t}^{\top} \\
= & \textstyle\sum_{s=1}^{t}(A_{ts} *^{(3)}_{(\eta^s,d-\eta^s)} X_sP_{d-\eta^s,s}P_{d-\eta^s,t}^{\top})P_{d,t}P_{d,t}^{\top}
\end{aligned} \tag{3.73}$$

and

$$\begin{aligned}
S_tP_{d,t}M_dP_{d,t}^{\top} &= S_tP_{d,t}M_dP_{d,t}^{\top}(P_{d,t}M_dP_{d,t}^{\top})^{-1}P_{d,t}P_{d,t}^{\top} \\
&= S_tP_{d,t}P_{d,t}^{\top}
\end{aligned} \tag{3.74}$$

and

$$\begin{aligned}
-B_tP_{\theta,t}T_{\theta,d}M_dP_{d,t}^{\top} &= -B_tP_{\theta,t}T_{\theta,d}M_dP_{d,t}^{\top}(P_{d,t}M_dP_{d,t}^{\top})^{-1}P_{d,t}P_{d,t}^{\top} \\
&= -B_tP_{\theta,t}T_{\theta,d}P_{d,t}^{\top}P_{d,t}M_dP_{d,t}^{\top}(P_{d,t}M_dP_{d,t}^{\top})^{-1}P_{d,t}P_{d,t}^{\top} \\
&= -B_tP_{\theta,t}T_{\theta,d}P_{d,t}^{\top}P_{d,t}P_{d,t}^{\top} \\
&= -B_tP_{\theta,t}T_{\theta,d}P_{d,t}^{\top}
\end{aligned} \tag{3.75}$$

By combining 3.73, 3.74 and 3.75 together, we obtain the following problem

$$\begin{aligned}
inf \quad & \textstyle\sum_{t=1}^{T} tr(C_tP_{\theta,t}T_{\theta,d-\eta^t}M_{d-\eta^t}P_{d-\eta^t,t}^{\top}X_t^{\top}) \\
s.t. \quad & x_t \in \mathcal{L}_{k^t,n_t},\ s_t \in \mathcal{L}_{k^t,m_t} \quad \forall t \in \mathbb{T} \\
& \textstyle\sum_{s=1}^{t}(A_{ts} *^{(3)}_{(\eta^s,d-\eta^s)} X_sP_{d-\eta^s,s}P_{d-\eta^s,t}^{\top})P_{d,t}P_{d,t}^{\top} + S_tP_{d,t}P_{d,t}^{\top} - B_tP_{\theta,t}T_{\theta,d}P_{d,t}^{\top} = 0 \\
& \exists s_t \in \mathcal{L}_{k^t,m_t} : \left\{ \begin{array}{l} S_tP_{d,t}M_d = \mathbb{E}(s_t(\xi^t)\mathcal{B}_d(\xi)^{\top}) \\ \\ s_t(\xi^t) \geq 0\ \mathbb{P}-a.s. \end{array} \right\}
\end{aligned} \tag{3.76}$$

We can see that the last constraint in problem 3.76 requires each component $\left[s_t(\xi^t)\right]_i, t = 1, ..., T$ and $i = 1, ..., m$, of the vector-valued function $s_t(\xi^t)$ to be the density function of a measure $[\mu_t]_i \in \mathcal{N}$ whose moments coincide with the i-th row of $S_tP_{d,t}M_d$. Thus i-th row of $S_tP_{d,t}M_d$ must be contained in $\mathcal{M}_d(\Xi)$. Verifying the membership of $\left[s_t(\xi^t)\right]_i$, $t = 1, ..., T$ and $i = 1, ..., m$ in the cone $\mathcal{M}_d(\Xi)$ is NP-hard [8]. However, we have shown in Section 2.6 that verifying the membership of $\left[s_t(\xi^t)\right]_i$, $t = 1, ..., T$ and $i = 1, ..., m$ in the cone $\mathcal{M}_d^{+}(\Xi)$ can be solved through a tractable SDP.

Approximation MD2: Verifying membership of $\mathcal{M}_d(\Xi)$ can be approximated by verifying membership of $\mathcal{M}_d^{+}(\Xi)$.

As we discussed in Section 2.6, the above approximation was justified by Lasserre. This approximation is described in Proposition 3.9, which is applicable due to the following lemma.

Lemma 3.16: For any given $S_t \in \mathbb{R}^{m_t \times s(k^t,d)}$ the constraint 3.68 is equivalent to 3.77.

$$\exists \tilde{s}_t \in \mathcal{L}_{k,m_t} : \left\{ \begin{array}{l} S_t P_{d,t} M_d = \mathbb{E}(\tilde{s}_t(\xi)\mathcal{B}_d(\xi)^\top) \\ \\ \tilde{s}_t(\xi) \geq 0 \, \mathbb{P} - a.s. \end{array} \right\} \tag{3.77}$$

Proof:

1. $\Longrightarrow$: Since $\xi^t \subseteq \xi$, $t = 1, ..., T$ constraint 3.76 implies the more general constraint 3.77.
2. $\Longleftarrow$: Assume that 3.77 holds and define $s_t(\xi^t) = \mathbb{E}_t(\tilde{s}_t(\xi))$. Then

$$\begin{aligned} \mathbb{E}(s_t(\xi^t)\mathcal{B}_d(\xi)^\top) &= \mathbb{E}(\mathbb{E}_t(\tilde{s}_t(\xi))\mathcal{B}_d(\xi)^\top) \\ &= \mathbb{E}(\tilde{s}_t(\xi)\mathcal{B}_d(\xi)^\top)P_{d,t}^\top M_t^\top \\ &= S_t P_{d,t} M_d \end{aligned} \tag{3.78}$$

■

By using the above lemma, problem 3.76 can be approximated by the following semidefinite program

$$\begin{array}{ll} inf & \sum_{t=1}^T tr(C_t P_{\theta,t} T_{\theta,d-\eta^t} M_{d-\eta^t} P_{d-\eta^t,t}^\top X_t^\top) \\ s.t. & X_t \in \mathbb{R}^{n \times s(k,d-\eta^t)}, S_t \in \mathbb{R}^{n \times s(k,d)} \\ & \left. \begin{array}{l} \sum_{s=1}^t (A_{ts} *_{(\eta^s,d-\eta^s)}^{(3)} X_s P_{d-\eta^s,s} P_{d-\eta^s,t}^\top) P_{d,t} P_{d,t}^\top + S_t P_{d,t} P_{d,t}^\top = B_t P_{\theta,t} T_{\theta,d} P_{d,t}^\top \\ \\ S_t P_{d,t} M_d \in \mathcal{M}_d^{m+}(\Xi) \end{array} \right\} \begin{array}{c} \mathbb{P} - a.s., \\ \forall t \in \mathbb{T} \end{array} \end{array} \tag{3.79}$$

Proposition 3.17: For problems 3.76 and 3.79 the following holds: inf 3.79 $\leq$ inf 3.76.

Proof: Problem 3.79 is obtained from problem 3.76 by Approximation MD2, which increases the feasible set since since $\mathcal{M}_d(\Xi) \subseteq \mathcal{M}_d^+(\Xi)$ due to Preposition 2.10. It is thus clear that solution of problem 3.79 provides a lower bound an the optimal value of problem 3.76. Lasserre showed that under Assumption 2.1 equality holds for a d chosen to be large enough. However, if Assumption 2.1 is not satisfied then equality holds as $d \longrightarrow \infty$. ■

Problem 3.79 represents a computationally tractable lower bound approximation of problem 3.14.

3.4.6 Advantages and disadvantages

The main advantage of decision rule approximations is that the problem grows only quadratically with the number of decision stages. SP problem can be approximated by a tractable SDP which can be solved in polynomial time. Another important advantage is that under Assumption 2.1 the exact solution is obtained as $d \longrightarrow \infty$. If Assumption 2.1 is not satisfied than this holds only for the lower bound approximation. In contrast to the scenario tree approximation, it is possible to define the upper and the lower bound approximation, what gives us a tractable procedure to estimate the suboptimality of solutions. This is very important when solving real life problems.

An important disadvantage is that complexity grows exponentially with the degree of the polynomial decision rules d. Polynomial decision rules of the degree $d > 5$ are thus for real life problems at the current state of the art technology not feasible. Another disadvantage, that we encountered when solving real problems, is that not many fast and robust SDP solvers exist. This is expected, because SDP problems are an important area of the research only in the recent years. When the SDP solvers evolve, we will be able to approximate SP problems much more precisely.

Chapter 4

Portfolio Optimization

Portfolio optimization is the problem of allocating capital over different assets in order to maximize the return on the investment and at the same time minimize its risk [48]. Since portfolio returns are uncertain, one usually wants to maximize the expected return of the investment. Characterization of the risk is a more difficult problem. The first portfolio optimization problem was introduced by Markowitz [34], where the variance of the return was used to characterize risk. Many other characterizations were proposed later on.

In the first part of this chapter, we consider one-stage portfolio optimization problems, which are in the second part extended to the multistage portfolio optimization problems. We present state of the art characterizations of the risk and construct a risk measure with desirable properties. Finally, we explain how portfolio optimization problems can be approximated with scenario tree approximation and decision rule approximations presented in the previous chapter.

4.1 One-stage portfolio optimization

4.1.1 Problem description

Suppose there are I assets available and denote by $\xi := [\xi_1, \xi_2, ..., \xi_I]$ their random returns. Returns are modeled in a probability space $(\mathbb{R}^I, \mathcal{B}(\mathbb{R}^I), \mathbb{P})$, where the Borel σ-algebra $\mathcal{B}(\mathbb{R}^I)$ represents the set of total returns that are assigned probabilities by the probability measure $\mathbb{P}$. Suppose we have the amount W_0 to invest. In one-stage portfolio optimization, we have to decide on the amount w_i that should be invested in each of the assets i, $i = 1, ..., I$ such that $\sum_{i=1}^{I} w_i = W_0$. We denote $w := [w_1, ..., w_I]$ and the total portfolio return r_p can be calculated as $r_p = w^\top \xi$.

The two main characteristics that describe each investor are greediness and risk-aversion. Every investor tries to maximize the expected portfolio return while minimizing its risk.The expected rates of returns are $\mathbb{E}(\xi) = [\mathbb{E}(\xi_1), \mathbb{E}(\xi_2), ..., \mathbb{E}(\xi_I)]$ and thus the expected portfolio return $\overline{r}_p$ is $\overline{r}_p = w^\top \mathbb{E}(\xi)$. Description of the risk is more extensive and thus described in the next section.

4.1.2 Risk measures

Risk could be understood as the variability of the portfolio return due to market changes and uncertain events. There are many ways to describe risk of an investment. A good risk measure must reflect our preferences of what risk of an investment is and how it should behave.

Coherent risk measures

Some general properties of a good risk measure have been defined in [32].

Consider a general probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a linear space of measurable functions $\mathcal{Z} := \{Z : \Omega \longrightarrow \mathbb{R}\}$. Let the functional $\rho : \mathcal{Z} \to \mathbb{R}$ be a risk measure. The number $\rho(Z)$, when positive, is interpreted as the minimum extra cash the investor has to add to the risky position Z and invest it in a risk free asset to be allowed to proceed with his plans, i.e., have risk with value 0. Similarly, if the number $\rho(Z)$ is negative, then cash amount -$\rho(Z)$ can be withdrawn from the position.

The functional ρ is a coherent risk measure for $\mathcal{Z}$, if it satisfies the following axioms:

- Translation invariance:

  If $\alpha \in \mathbb{R}$ and $Z \in \mathcal{Z}$, then $\rho(Z+\alpha) = \rho(Z) - \alpha$. In words, adding a risk free amount α to the initial position Z, decreases the risk by α. Note, that if $\alpha = \rho(Z)$ then $\rho(Z + \rho(Z)) = 0$, which follows our interpretation of $\rho(Z)$ given above.

- Sub-additivity:

  If Z_1, $Z_2 \in \mathcal{Z}$, then $\rho(Z_1 + Z_2) \le \rho(Z_1) + \rho(Z_2)$. In words, the risk of two separate portfolios can not be smaller than the risk of having both portfolios together (diversification principle).

- Positive homogeneity:

  If $\alpha \ge 0$ and $Z \in \mathcal{Z}$, then $\rho(\alpha Z) = \alpha\rho(Z)$. In words, if an investment in a portfolio is increased (decreased), the risk increases (decreases) by the same factor.

- Monotonicity:

  If Z_1, $Z_2 \in \mathcal{Z}$ and $Z_1 \le Z_2$, then $\rho(Z_1) \ge \rho(Z_2)$. In words, if a portfolio Z_1 has worse values than portfolio Z_2 under all realisations, then the risk of Z_1 should be greater than the risk of Z_2.

Proposition 4.1: Coherent risk measures are convex functions.

Proof: Let Z_1, $Z_2 \in \mathcal{Z}$. We have to proof that, for any $\lambda = [0, 1]$,

$$\rho(\lambda Z_1 + (1-\lambda)Z_2) \le \lambda\rho(Z_1) + (1-\lambda)\rho(Z_2). \tag{4.1}$$

We write

$$\rho(\lambda Z_1 + (1-\lambda)Z_2) \le \rho(\lambda Z_1) + \rho((1-\lambda)Z_2) = \lambda\rho(Z_1) + (1-\lambda)\rho(Z_2), \tag{4.2}$$

where the inequality holds due to sub-additivity and the equality due to positive homogeneity property. ■

Variance

One of the first measures used to describe a portfolio risk is the variance σ_p^2 of the portfolio. Given the variance of each individual asset i (σ_i^2) and the covariance between assets i and j (σ_{ij}), $i, j \in \{1, ..., I\}$, the variance of the portfolio is defined as

$$\begin{aligned}
\sigma_p^2 &= \mathbb{E}\left((r_p - \overline{r}_p)^2\right) \\
&= \mathbb{E}\left(\left(\textstyle\sum_{i=1}^I w_i\xi_i - \sum_{i=1}^I w_i\mathbb{E}(\xi_i)\right)^2\right) \\
&= \mathbb{E}\left(\left(\textstyle\sum_{i=1}^I w_i\left(\xi_i - \mathbb{E}(\xi_i)\right)\right)\left(\sum_{i=1}^I w_i\left(\xi_i - \mathbb{E}(\xi_i)\right)\right)\right) \\
&= \mathbb{E}\left(\textstyle\sum_{i,j=1}^I w_i w_j\left(\xi_i - \mathbb{E}(\xi_i)\right)\left(\xi_j - \mathbb{E}(\xi_j)\right)\right) \\
&= \textstyle\sum_{i,j=1}^I w_i w_j \sigma_{ij} \\
&= w^\top \Sigma w.
\end{aligned} \tag{4.3}$$

In last line we used matrix notation, where Σ represents a covariance matrix.

The main disadvantage of the variance as a risk measure is that it penalizes both profits as well as losses, since it is a measure of the dispersion of the values of the random variable around its expected value. In reality, investors want to minimize only the possibility of losses.

Value at Risk

Value at Risk (VaR) is a risk measure that eliminates the major drawback of the variance. It is defined as a threshold loss value $\alpha \in \mathbb{R}$ that is, by the end of the investing period, only exceeded with a probability level $1-\beta \in \mathbb{R}$ (Figure 4.1).

Let $f(w,\xi)$ be the loss associated with the decision vector $w \in \mathbb{R}^I$ and the random vector $\xi \in \mathbb{R}^I$. The underlying probability distribution of $\xi \in \mathbb{R}^I$ has a density denoted by $p(\xi)$. The probability of $f(w,\xi)$ not exceeding a threshold α is given by

$$\Psi(w,\alpha) = \int_{f(w,\xi)\leq\alpha} p(\xi)d\xi. \tag{4.4}$$

The β-VaR values for the loss random variable associated with w and any specified probability level $\beta \in (0,1)$ is denoted by $\alpha_\beta(w)$ and defined by the following equation

$$\alpha_\beta(w) = min\left\{\alpha \in \mathbb{R}:\ \Psi(w,\alpha) \geq \beta\right\}. \tag{4.5}$$

Value at Risk is in general not a coherent risk measure, because it does not satisfy the sub-additivity axiom [32]. This, as a consequence, has two major drawbacks:

- It is not necessary convex (Proposition 4.1) and thus, due to the existence of local minimums, difficult to optimize.
- It does not encourage diversification and sometimes even penalizes it.

Moreover, VaR does not take into account the distribution of the loss, if it exceeds the VaR value (threshold α). To eliminate these drawbacks, conditional value at risk was proposed.

Conditional Value at Risk

Conditional Value at Risk (CVaR) is a coherent risk measure [32, 33], which takes into account the conditional expected value of loss, under the condition that it exceeds the VaR value (Figure 4.1). Mathematically, it is defined by

$$\phi_\beta(w) = (1-\beta)^{-1}\int_{f(w,\xi)\geq\alpha_\beta(w)} f(w,\xi)p(\xi)d\xi. \tag{4.6}$$

By comparing the definitions of CVaR and VaR, we can see that CVaR is a more conservative risk measure than VaR since $\alpha_\beta(w) \leq \phi_\beta(w)$.

For the further argumentation we define the function $F_\beta(w,\alpha) : \mathbb{R}^I \times \mathbb{R} \longrightarrow \mathbb{R}$ by

$$F_\beta(w,\alpha) = \alpha + (1-\beta)^{-1}\int_{\xi\in R^I} [f(w,\xi)-\alpha]^+ p(\xi)d\xi = \alpha + (1-\beta)^{-1}\mathbb{E}([f(w,\xi)-\alpha]^+), \tag{4.7}$$

where $[x]^+ = x$ if $x \geq 0$ and 0 if $x < 0$. Rockafellar [33] proved the following theorem.

Theorem 4.2: As a function of (w,α), $F_\beta(w,\alpha)$ is convex and continuously differentiable. The β-CVaR of the loss associated with any $w \in \mathbb{R}^I$ can be determined by

$$\phi_\beta(w) = \min_{\alpha\in\mathbb{R}} F_\beta(w,\alpha). \tag{4.8}$$

Moreover, minimizing the β-CVaR of the loss associated with w over all feasible $w \in \mathbb{R}^I$ is equivalent to minimizing $F_\beta(w,\alpha)$ over $(w,\alpha) \in \mathbb{R}^I \times \mathbb{R}$, in a sense that

$$\min_{w\in\mathbb{R}^I} \phi_\beta(w) = \min_{(w,\alpha)\in\mathbb{R}^I\times\mathbb{R}} F_\beta(w,\alpha). \tag{4.9}$$

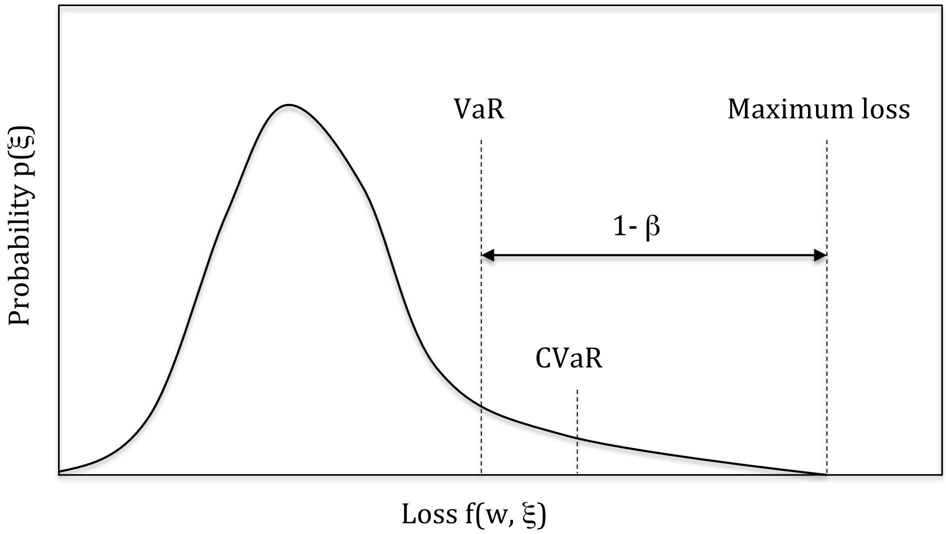

Figure 4.1: VaR and CVaR.

Proof: [33]. ■

The most important finding in the above theorem is that $F_\beta(w,\alpha)$ is convex and thus much easier to optimize with respect to the non-convex risk measures (e.g. VaR). Moreover, CVaR can be calculated directly, without first calculating VaR, even though the definition of CVaR depends on VaR. Rockafellar even showed that VaR can be calculated as a byproduct of the CVaR calculation.

Representation of $F_\beta(w,\alpha)$ needs to be altered when applied to the optimization context. We introduce a function $z \in \mathcal{L}^\infty_{I,1}$ such that $f(w,\xi)-\alpha \le z(\xi)$ and $0 \le z(\xi)$. The problem of minimizing $F_\beta(w,\alpha)$ can thus be written as

$$\begin{aligned} min \quad & \alpha + (1-\beta)^{-1}\mathbb{E}\left(z(\xi)\right) \\ s.t. \quad & \alpha \in \mathbb{R},\ w \in \mathbb{R}^I,\ z \in \mathcal{L}^\infty_{I,1} \\ & z(\xi) \ge 0 \\ & z(\xi) \ge f(w,\xi) - \alpha, \end{aligned} \tag{4.10}$$

where $\mathcal{L}^\infty_{k,n}$ denotes the space of all Borel measurable functions from $\mathbb{R}^k$ to $\mathbb{R}$ in n variables.

This representation can be used directly in portfolio optimization problems, where the loss $f(w,\xi)$ can be defined as the negative total portfolio return

$$f(w,\xi) = -\bar{r}_p = -w^\top \xi. \tag{4.11}$$

4.1.3 Mean-variance Efficient Portfolio

As described in the beginning of this chapter, portfolio selection is the optimization problem, in which we minimize our risk exposure subject to the requested expected portfolio return. The first portfolio optimization problem was formulated by Markowitz [34]. He used variance as a risk measure. The corresponding optimal portfolio for a given expected return is thus called the mean-variance efficient portfolio.

The mean-variance efficient portfolio characterizes portfolios with minimal variance given the required expected return $\bar{r}_p$. It is formulated as the following optimization problem

$$\begin{aligned} min \quad & \sigma_p^2 = w^\top \Sigma w \\ s.t. \quad & w \in \mathbb{R}^I \\ & w^\top e = W_0 \\ & w^\top E(\xi) \geq \bar{r}_p. \end{aligned} \tag{4.12}$$

If one plots the standard deviation σ_p of the solutions of problem 4.12 for all possible values of $\bar{r}_p$, an *efficient frontier* is obtained (Figure 4.2). Portfolios that lie on the efficient frontier have the minimal variance for a given expected return $\bar{r}_p$.

Bounds $\left[r_p^{min}, r_p^{max}\right]$ for all possible values of $\bar{r}_p$ can be obtained by solving two special portfolio optimization problems. In order to obtain the upper bound r_p^{max}, the maximal value of the total expected return must be found without considering risk. If w^* is the optimal solution of such problem,

$$\begin{aligned} max \quad & w^\top \mathbb{E}(\xi) \\ s.t. \quad & w \in \mathbb{R}^I \\ & w^\top e = W_0, \end{aligned} \tag{4.13}$$

then $r_p^{max} = w^{*\top}\mathbb{E}(\xi)$.

Similarly, in order to obtain the lower bound r_p^{min}, the minimal value of the variance must be found without imposing a target expected return in the constraints of the problem. If w^* is optimal solution of such problem,

$$\begin{aligned} min \quad & w^\top \Sigma w \\ s.t. \quad & w \in \mathbb{R}^I \\ & w^\top e = W_0, \end{aligned} \tag{4.14}$$

then $r_p^{min} = w^{*\top}\mathbb{E}(\xi)$.

If one then solves problem 4.12 for all $\bar{r}_p \in \left[r_p^{min}, r_p^{max}\right]$, an efficient frontier similar to the one on Figure 4.2 is obtained.

4.1.4 Mean-CVaR Efficient Portfolio

In a similar manner, the efficient frontier of other risk measures can be constructed. The mean-CVaR Efficient portfolio [33, 35] is the solution of the following optimization problem

$$\begin{aligned} min \quad & \alpha + (1-\beta)^{-1}\mathbb{E}\left(z(\xi)\right) \\ s.t. \quad & \alpha \in \mathbb{R},\ w \in \mathbb{R}^I,\ z \in \mathcal{L}_{I,1}^\infty \\ & z(\xi) \geq 0 \\ & z(\xi) \geq -w^\top \xi - \alpha \\ & w^\top e = W_0 \\ & w^\top E(\xi) \geq \bar{r}_p, \end{aligned} \tag{4.15}$$

where $\mathcal{L}_{k,n}^\infty$ denotes the space of all Borel measurable functions from $\mathbb{R}^k$ to $\mathbb{R}$ in n variables.

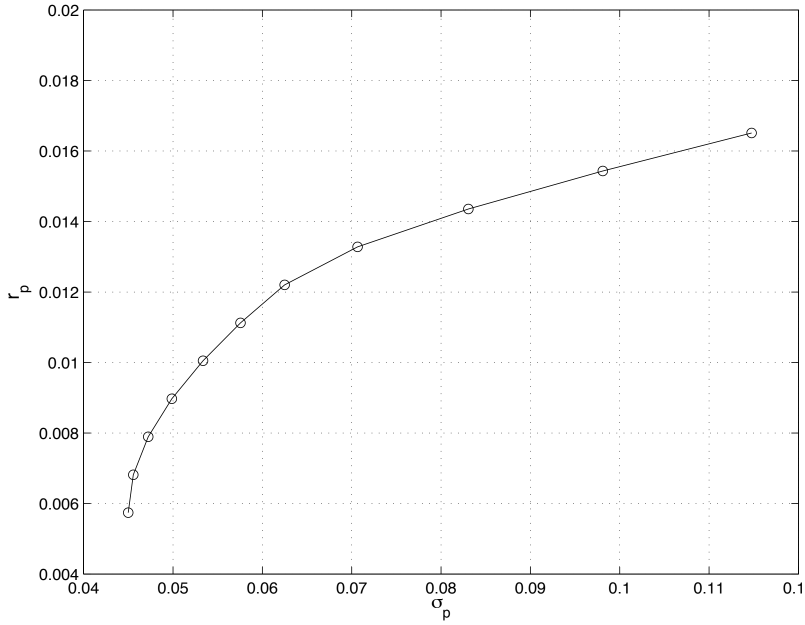

Figure 4.2: Example of the efficient frontier with $r_p^{min} = 0.0058$ and $r_p^{max} = 0.0163$

4.2 Multistage portfolio optimization

In the previous section we have discussed one-stage portfolio optimization. However, this simplified model does not reflects the reality, where investors dynamically rebalance their portfolio in time. This is the motivation to define multistage portfolio optimization problems.

4.2.1 Problem description

The classical one-stage portfolio optimization problem can be extended to multistage problems. The goal of the multistage portfolio optimization problem is to determine the optimal portfolio for a given finite investment horizon T defined by a set of stages $\mathbb{T} := \{1, ..., T\}$. After making an initial investment at time $t = 1$, the portfolio can be rebalanced at times $t = 2, ..., T - 1$, and redeemed at the end of the last period $t = T$.

Consider a filtered probability space $(\mathbb{R}^k, \mathcal{F}, (\mathcal{F})_{t\in\mathbb{T}}, \mathbb{P})$. Total returns are represented as $\xi := (\xi_1, ..., \xi_T)$, where sub vectors $\xi_t \in \mathbb{R}^{k_t}$ are the total returns at time $t \in \mathbb{T}$. The history of all the total returns up to time t is denoted by $\xi^t := (\xi_1, ..., \xi_t) \in \mathbb{R}^{k^t}$, where $k^t := \sum_{s=1}^t k_s$. The decision about the investment $w_t(\xi^t)$ is made at time t after total returns ξ^t have been revealed, but before any future outcomes $\{\xi_s\}_{s>t}$ have been observed. We associate with the process of revealing ξ_t the corresponding filtration $\mathcal{F}_1 \subset ... \subset \mathcal{F}_T$ of σ-algebras on $\mathbb{R}^k$. Finally, we denote by $\mathbb{P}_t$ the probability distribution of ξ_t. Note that, the one-stage model can be seen as a special case of the multistage model where $\xi^T = \xi$ and $k^T = k$.

We have already mentioned the two main characteristics that describe each investor, greediness and risk-aversion. Every investor tries to maximize the expected portfolio return while minimizing its risk. The expected return of a multistage portfolio $\bar{r}_p$ could be calculated as the average wealth at the stage $t = T - 1$, multiplied by the average returns in the last stage ξ_T, i.e.,

$$\bar{r}_p = \mathbb{E}\left(w_{T-1}^\top(\xi^{T-1})\xi_T\right). \tag{4.16}$$

Under the assumption that the total returns are stage-wise independent, the expected return can be written as

$$\bar{r}_p = \mathbb{E}\left(w_{T-1}^\top(\xi^{T-1})\right)\mathbb{E}(\xi_T). \tag{4.17}$$

A much harder problem is the characterization of risk, which we discuss in the following section.

4.2.2 Risk measures

We have already discussed various risk measures for one-stage portfolio optimization problem. We have seen that good risk measures must satisfy the axioms of coherency. In this section, we will first extend one-stage coherent risk measures to the multistage coherent risk measures and then present an additional requirement for characterization of good risk measures termed *time consistency.*

In the description below, we consider a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F})_{t\in\mathbb{T}}, \mathbb{P})$ with $\mathcal{F}_1 = \{\emptyset, \Omega\}$. We associate with the process of revealing ξ_t the corresponding filtration $\mathcal{F}_1 \subset ... \subset \mathcal{F}_T$ of σ-algebras on Ω. We denote by $\mathcal{L}_t^\infty := \mathcal{L}_T^\infty(\Omega, \mathcal{F}_t, \mathbb{P})$, $t \in \mathbb{T}$, the vector space of all bounded $\mathcal{F}_t$-measurable random variables. Moreover, we define $\mathcal{Z}_t$ as a space of all $\mathcal{F}_t$-measurable functions $\mathcal{Z}_t := \{Z : \mathcal{L}_t^\infty \longrightarrow \mathbb{R}\}$.

Risk in the multistage setting is given by a sequences of mappings $\rho_t : \mathcal{L}_T^\infty \longrightarrow \mathcal{L}_t^\infty$, $t = 1, ..., T-1$, where $\rho_t(X)$, $X \in \mathcal{L}_T^\infty$, can be understood as an assessment of the downside risk of position X conditional on the information ξ^t available at time t.

We will present multistage coherent risk measures through the concept of conditional convex risk measures.

Conditional convex risk measure: A mapping $\rho_t : \mathcal{L}_T^\infty \longrightarrow \mathcal{L}_t^\infty$ is called a conditional convex risk measure if it satisfies the following properties. For each $X_1, X_2 \in \mathcal{L}_T^\infty$:

- Conditional cash invariance: for all $m_t \in \mathcal{L}_t^\infty$

$$\rho_t(X_1 + m_t) = \rho_t(X_1) - m_t. \tag{4.18}$$

- Monotonicity: $X_1 \le X_2 \Longrightarrow \rho_t(X_1) \ge \rho_t(X_2)$.
- Conditional convexity: for all $\lambda \in \mathcal{L}_t^\infty$, $0 \le \lambda \le 1$,

$$\rho_t(\lambda X_1 + (1-\lambda)X_2) \le \lambda \rho_t(X_1) + (1-\lambda)\rho_t(X_2). \tag{4.19}$$

- Normalization: $\rho_t(0) = 0$.

Coherent risk measure: A conditional convex risk measure is coherent if it has in addition the following property:

- Conditional positive homogeneity: for all $\lambda \in \mathcal{L}_t^\infty$, $\lambda \ge 0$,

$$\rho_t(\lambda X_1) = \lambda \rho_t(X_1). \tag{4.20}$$

Dynamic convex risk measure: A sequence $(\rho_t)_{t\in\mathbb{T}}$ is called a dynamic convex risk measure if ρ_t is a conditional convex risk measure for each $t \in \mathbb{T}$.

Time consistent risk measure: A dynamic convex risk measure $(\rho_t)_{t\in\mathbb{T}}$ is time consistent if any of the following equivalent conditions hold:

- for all $t = 1, ..., T-1$ and all $X_1, X_2 \in \mathcal{L}_T^\infty$

$$\rho_{t+1}(X_1) \ge \rho_{t+1}(X_2) \Longrightarrow \rho_t(X_1) \ge \rho_t(X_2). \tag{4.21}$$

  In words, if a portfolio X_1 is riskier than portfolio X_2 at time $t+1$, then the portfolio X_1 must be riskier than the portfolio X_2 also at time t.

- for all $t = 1, ..., T-1$ and for all $X_1, X_2 \in \mathcal{L}_T^\infty$

$$\rho_{t+1}(X_1) = \rho_{t+1}(X_2) \Longrightarrow \rho_t(X_1) = \rho_t(X_2). \tag{4.22}$$

  In words, if a portfolio X_1 is equally risky as portfolio X_2 at time $t+1$, then the portfolio X_1 must be equally risky than the portfolio X_2 also at time t.

- $(\rho_t)_{t\in\mathbb{T}}$ is recursive: $\rho_t = \rho_t(-\rho_{t+s})$ for all $t, s \geq 0$ such that $t, t+s \in \mathbb{T}$.

  To illustrate this definition, let us consider a multistage optimization problem and for simplicity we assume that the uncertain parameters are discrete and can be described with the scenario tree of Figure 3.1. Imagine that we are at the stage $t = 2$ and at the node 1. It is clear that all the decisions we make at that node, will only depend on nodes 3, 4 and 5. At node 1 we already know that it is not possible to reach nodes 6, 7 and 8, and thus we should not consider them. The same idea holds for all the nodes in the scenario tree. Decisions in each node must only depend on scenarios that are reachable from the current node.

Even though, the idea of time consistency is quite intuitive, many measures do not satisfy it. A simple example and comparison of the decisions made with a time consistent and a time inconsistent risk measure is given in [36]. Note that neither multistage VaR nor multistage CVaR satisfy time consistency.

The last definition of time consistency is especially useful for construction of time consistent multi-period risk measure ρ'_t from one-period risk measure ρ. Construction is summarized in the following recursive procedure.

1. $\rho'_{T-1} := \rho$
2. For all $t = 1, .., T-2$ $\rho'_t := \rho(-\rho'_{t+1})$

We will use this procedure to construct a time consistent version of CVaR in the next section.

Another approach to introduce time consistent risk measures is through the Bellman's principle [40]. It is known that the Bellman's principle imply time consistency [39].

Bellman's principle: Let us assume that in every stage $t = 1, ..., T$ we can calculate real valued loss function $f_t(x_t(\xi^t), \xi_t)$ and objective function $\mathbb{F}_t(Z_t, ..., Z_T \left|\xi^t\right.) : \mathcal{Z}_t \times ... \times \mathcal{Z}_T \times \mathbb{R}^{k_t} \longrightarrow \mathbb{R}$. To slightly simplify the notation we denote $x_t := x_t(\xi^t)$, $t \in \mathbb{T}$. In every stage $t \in \mathbb{T}$ of a multistage model, we would like to minimize all the losses that might occur from the current stage t to the last stage T given the information on all the realisations of the uncertain parameters ξ^t at time t. In other words, in every stage $t = \mathbb{T}$, we would like to solve the following optimization problem

$$\underset{x_t,...,x_T}{min} \quad \mathbb{F}_t\left(f_t(x_t, \xi_t), ..., f_T(x_T, \xi_T) \left|\xi^t\right.\right). \tag{4.23}$$

It is important to note that each decision rule $x_\tau(\xi^\tau)$, $\tau = t, ..., T$ is a function of the parameters ξ^τ conditional on all the revealed parameters ξ^t up to time t.

The first stage model $t = 1$ is defined by

$$\underset{x_1,...,x_T}{min} \quad \mathbb{F}_1\left(f_1(x_1, \xi_1), ..., f_T(x_T, \xi_T) \left|\xi^1\right.\right), \tag{4.24}$$

where we included $\xi_1 = 1$ for consistency. Similarly, the last stage model $t = T$ is defined by

$$\underset{x_T}{min} \quad \mathbb{F}_T\left(f_T(x_T, \xi_T) \left|\xi^T\right.\right) . \tag{4.25}$$

The optimal value $V_t\left(x_{t-1}, \xi^t\right)$ of problem 4.23 is a function of ξ^t and last decision x_{t-1}. Let us reformulate problem 4.23 to the following equivalent problem

$$\underset{x_t}{min} \quad \left[\underset{x_{t+1},...,x_T}{inf} \mathbb{F}_t\left(f_t(x_t, \xi_t), f_{t+1}(x_{t+1}, \xi_{t+1}), ..., f_T(x_T, \xi_T) \left|\xi^t\right.\right)\right] . \tag{4.26}$$

Proposition 4.2: Optimization problem 4.26 satisfies time consistency if for $t = 1, ..., T$, the optimal value inside the parenthesis can be formulated in the form

$$\phi_t \left(f_t(x_t, \xi_t), V_{t+1}(x_t, \xi^{t+1}) \,\middle|\, \xi^t \right), \tag{4.27}$$

where $\phi_t(\cdot, \cdot \,|\cdot)$ is a real valued function.

Under Proposition 4.2, problem 4.26 can be formulated as

$$\underset{x_t}{min} \quad \phi_t \left(f_t(x_t, \xi_t), V_{t+1}(x_t, \xi^{t+1}) \,\middle|\, \xi^t \right) \ . \tag{4.28}$$

The corresponding dynamic programming equation for last stage $t = T$ is

$$V_T(x_{T-1}, \xi^T) = \underset{x_T}{inf} \, \mathbb{F}_T \left(f_T(x_T, \xi_T) \,\middle|\, \xi^T \right). \tag{4.29}$$

Similarly, for $t = T - 1, ..., 1$ the dynamic equations are

$$V_t(x_{t-1}, \xi^t) = \underset{x_t}{inf} \, \phi_t \left(f_t(x_t, \xi_t), V_{t+1}(x_t, \xi^{t+1}) \,\middle|\, \xi^t \right). \tag{4.30}$$

Example 4.1: An example of a time consistent problem is the risk neutral multistage stochastic programming problem. Consider the following general formulation

$$\underset{x_1}{min} \, f_1(x_1, \xi_1) + \mathbb{E} \left(\underset{x_2}{inf} \, f_2(x_2, \xi_2) + ... + \mathbb{E} \left(\underset{x_T}{inf} \, f_T(x_T, \xi_T) \right) ... \right) \tag{4.31}$$

where $\xi^1 = 1$ is written for consistency in the general stochastic programming formulation. The above optimization problem must be understood in the following way: for a stochastic process $\xi_1, \xi_2, ..., \xi_T$, in each stage t, we solve the problem

$$\mathbb{F}_t(Z_t, ..., Z_T \,|\xi^t) := \mathbb{E} \left(Z_t + Z_{t+1} + ... + Z_T \,\middle|\, \xi^t \right). \tag{4.32}$$

By following Proposition 4.2, we can formulate

$$\phi_t \left(f_t(x_t, \xi_t), V_{t+1}(x_t, \xi^{t+1}) \,\middle|\, \xi^t \right) := \mathbb{E} \left(f_t(x_t, \xi_t) + V_{t+1}(x_t, \xi^{t+1}) \,\middle|\, \xi^t \right). \tag{4.33}$$

Consequently, the dynamic programming equations are for $t = T$

$$V_T(x_{T-1}, \xi^T) = \mathbb{E} \left(\underset{x_T}{inf} \, f_T(x_T(\xi^T), \xi_T) \,\middle|\, \xi^T \right) \tag{4.34}$$

and for $t = 2, ..., T - 1$

$$V_t(x_{t-1}, \xi^t) = \mathbb{E} \left(\underset{x_t}{inf} \, f_t(x_t, \xi_t) + V_{t+1}(x_t, \xi^{t+1}) \,\middle|\, \xi^t \right). \tag{4.35}$$

Since it is possible to formulate dynamic equations that satisfy Proposition 4.2, the risk neutral multistage stochastic programming problems in the form 4.31 are time consistent. ■

4.2.3 Transaction costs

In multistage portfolio optimization, investors are allowed to rebalance their portfolio in every stage. However, the rebalancing is not free. Thus, the transaction costs (e.g. [12, 42]) must be taken into account. Imagine that in each stage t, $t = 1, .., T - 1$, the asset vector $w_t(\xi^t)$ denotes the current wealth, $b_t(\xi^t)$ the assets bought in this stage and $s_t(\xi^t)$ the assets sold in this stage. Without transaction costs, the portfolio balance can be calculated as

$$w_t(\xi^t) = \xi_t w_{t-1}(\xi^{t-1}) + b_t(\xi^t) - s_t(\xi^t). \tag{4.36}$$

Transaction costs are calculated as a fixed percentage c_b (c_s) of the assets bought (sold) in each time period. By including this to equation 4.36, the portfolio balance must satisfy

$$w_t(\xi^t) = \xi_t w_{t-1}(\xi^{t-1}) + (1 - c_b) b_t(\xi^t) - (1 + c_s) s_t(\xi^t). \tag{4.37}$$

4.2.4 Mean-Variance Efficient Portfolio

If the risk of a portfolio is modeled by its variance, the multistage mean-variance optimization problem [12] is given as

$$
\begin{aligned}
min \quad & w_{T-1}^{\top}(\xi^{T-1})\Sigma w_{T-1}(\xi^{T-1}) & \\
s.t. \quad & w_t \in \mathbb{W}_t,\ s_t \in \mathbb{S}_t,\ b_t \in \mathbb{B}_t & t = 1, ..., T-1 \\
& \mathbb{E}\left(w_{T-1}^{\top}(\xi^{T-1})\xi_T\right) \geq \bar{r}_p & \\
& 1^{\top} b_1(\xi^1) - 1^{\top} s_1(\xi^1) = W_0 & \\
& (1 - c_b)\, b_1(\xi^1) - (1 + c_s)\, s_1(\xi^1) = w_1(\xi^1) & \\
& 1^{\top} b_t(\xi^t) - 1^{\top} s_t(\xi^t) = 0 & t = 2, .., T \\
& w_{t-1}(\xi^{t-1})\xi_t + (1 - c_b)\, b_t(\xi^t) - (1 + c_s)\, s_t(\xi^t) = w_t(\xi^t) & t = 2, .., T
\end{aligned}
\tag{4.38}
$$

where the input parameter W_0 denotes an initial wealth to invest. We introduced sets

$$
\begin{aligned}
\mathbb{W}_t &:= \left\{w_t \in \mathcal{L}_{k^t,n} : w_{min} \leq w_t(\xi^t) \leq w_{max}\right\}, \\
\mathbb{S}_t &:= \left\{s_t \in \mathcal{L}_{k^t,n} : 0 \leq s_t(\xi^t) \leq s_{max}\right\}, \\
\mathbb{B}_t &:= \left\{b_t \in \mathcal{L}_{k^t,n} : 0 \leq b_t(\xi^t) \leq b_{max}\right\},
\end{aligned}
$$

where $b_{max} \in \mathbb{R}^I$ and $s_{max} \in \mathbb{R}^I$ denote the maximal amount of assets bought and sold, respectively. Similarly, $w_{max} \in \mathbb{R}^I$ and $w_{min} \in \mathbb{R}^I$ denote the maximal and minimal wealth allowed in each of the assets, respectively. If short-selling is not allowed, then we must set $w_{min} \geq 0$.

4.2.5 Time Inconsistent Mean-CVaR Efficient Portfolio

The one-stage mean-CVaR portfolio optimization problem can be extended to the multistage setting. A naive approach for this extension is, instead of considering just loss in one period, to consider the cumulative loss in all investment periods. Such model can be expressed as the following optimization problem

$$
\begin{array}{lll}
\textit{minimize} & \alpha(\xi^1) + (1-\beta)^{-1}\mathbb{E}\left(z(\xi^T)\right) & \\
\textit{subject to} & w_t \in \mathbb{W}_t,\ s_t \in \mathbb{S}_t,\ b_t \in \mathbb{B}_t & t = 1, ..., T-1 \\
& \alpha \in \mathcal{L}_{k^1,1}, z \in \mathcal{L}_{k,1} & \\
& -w_{T-1}^\top(\xi^{T-1})\xi_T - \alpha(\xi^1) \leq z(\xi^T) & \\
& 0 \leq z(\xi^T) & \\
& \mathbb{E}\left(w_{T-1}^\top(\xi^{T-1})\xi_T\right) \geq \bar{r}_p & \\
& 1^\top b_1(\xi^1) - 1^\top s_1(\xi^1) = W_0 & \\
& (1-c_b)\, b_1(\xi^1) - (1+c_s)\, s_1(\xi^1) = w_1(\xi^1) & \\
& 1^\top b_t(\xi^t) - 1^\top s_t(\xi^t) = 0 & t = 2, .., T-1 \\
& w_{t-1}(\xi^{t-1})\xi_t + (1-c_b)\, b_t(\xi^t) - (1+c_s)\, s_t(\xi^t) = w_t(\xi^t) & t = 2, .., T-1
\end{array}
\tag{4.39}
$$

where we used the definitions of $\mathbb{W}_t$, $\mathbb{S}_t$ and $\mathbb{B}_t$, $t = 1, ..., T-1$ from the previous section.

Proposition 4.3: Problem 4.39 is not time consistent.

Proof: In order to proof time inconsistency, we will apply Proposition 4.2. Let us first write the dynamic programming equations of the problem. For the last investment stage $t = T-1$, the value function $V_{T-1}([w_{T-2}, \alpha], \xi^{T-1})$ is defined by the optimal value of problem

$$
\begin{array}{ll}
\textit{min} & (1-\beta)^{-1}\mathbb{E}\left(\left[-w_{T-1}^\top(\xi^{T-1})\xi_T - \alpha(\xi^1)\right]^+\right) \\
\textit{s.t.} & w_{T-1} \in \mathbb{W}_{T-1},\ s_{T-1} \in \mathbb{S}_{T-1},\ b_{T-1} \in \mathbb{B}_{T-1},\ \alpha \in \mathcal{L}_{k^1,1} \\
& \mathbb{E}\left(w_{T-1}^\top(\xi^{T-1})\xi_T\right) \geq \bar{r}_p \\
& 1^\top b_{T-1}(\xi^{T-1}) - 1^\top s_{T-1}(\xi^{T-1}) = 0 \\
& w_{T-2}(\xi^{T-2})\xi_{T-1} + (1-c_b)\, b_{T-1}(\xi^{T-1}) - (1+c_s)\, s_{T-1}(\xi^{T-1}) = w_{T-1}(\xi^{T-1}).
\end{array}
\tag{4.40}
$$

Similarly, for $t = T-2, ..., 2$, value functions $V_t([w_{t-1}, \alpha], \xi^t)$ are defined by the optimal value of problems

$$
\begin{array}{ll}
\textit{min} & \mathbb{E}\left(V_{t+1}\left(\left[w_t(\xi^t), \alpha(\xi^1)\right], \xi^{t+1}\right)\right) \\
\textit{s.t.} & w_t \in \mathbb{W}_t,\ s_t \in \mathbb{S}_t,\ b_t \in \mathbb{B}_t,\ \alpha \in \mathcal{L}_{k^1,1} \\
& 1^\top b_t(\xi^t) - 1^\top s_t(\xi^t) = 0 \\
& w_{t-1}(\xi^{t-1})\xi_t + (1-c_b)\, b_t(\xi^t) - (1+c_s)\, s_t(\xi^t) = w_t(\xi^t)
\end{array}
\tag{4.41}
$$

and in the first stage $t = 1$, the corresponding problem is

$$
\begin{aligned}
\mathit{min} \quad & \alpha(\xi^1) + \mathbb{E}\left(V_2\left(\left[w_1(\xi^1), \alpha(\xi^1)\right], \xi^2\right)\right) \\
\mathit{s.t.} \quad & w_1 \in \mathbb{W}_1,\ s_1 \in \mathbb{S}_1,\ b_1 \in \mathbb{B}_1,\ \alpha \in \mathcal{L}_{k^1,1} \\
& 1^\top b_1(\xi^1) - 1^\top s_1(\xi^1) = W_0 \\
& (1 - c_b)\, b_1(\xi^1) - (1 + c_s)\, s_1(\xi^1) = w_1(\xi^1)
\end{aligned}
\tag{4.42}
$$

Note, that $\alpha(\xi^1)$ is a first stage variable. Since the decisions in the last investment stage $T-1$ depend on $\alpha(\xi^1)$, it is impossible to formulate $\phi_t(\cdot, \cdot|\cdot)$ as required by Proposition 4.2. Note, that Proposition 4.2 requires that the optimal value in stage $T-1$ depends only on decision rules from stage $T-2$, i.e., w_{T-2}, s_{T-2} and b_{T-2}. ■

In the next subsection we construct a time consistent version of problem 4.39.

4.2.6 Time Consistent Mean-CVaR Efficient Portfolio

We have already described a procedure to formulate a multistage time consistent risk measure from one-stage risk measures. In this section, we apply this procedure in terms of dynamic equations.

For the last investment stage $t = T-1$, the value function $V_{T-1}(w_{T-2}, \xi^{T-1})$ is defined by the optimal value of problem

$$
\begin{aligned}
\mathit{min} \quad & \alpha_{T-1}(\xi^{T-1}) + (1-\beta)^{-1}\mathbb{E}_{\mathbb{P}_T}\left(\left[-w_{T-1}^\top(\xi^{T-1})\xi_T - \alpha_{T-1}(\xi^{T-1})\right]^+\right) \\
\mathit{s.t.} \quad & w_{T-1} \in \mathbb{W}_{T-1},\ s_{T-1} \in \mathbb{S}_{T-1},\ b_{T-1} \in \mathbb{B}_{T-1},\ \alpha_{T-1} \in \mathcal{L}_{k^{T-1},1} \\
& \mathbb{E}\left(w_{T-1}^\top(\xi^{T-1})\xi_T\right) \geq \bar{r}_p \\
& 1^\top b_{T-1}(\xi^{T-1}) - 1^\top s_{T-1}(\xi^{T-1}) = 0 \\
& w_{T-2}(\xi^{T-2})\xi_{T-1} + (1 - c_b)\, b_{T-1}(\xi^{T-1}) - (1 + c_s)\, s_{T-1}(\xi^{T-1}) = w_{T-1}(\xi^{T-1}).
\end{aligned}
\tag{4.43}
$$

If we follow the procedure for the construction of time consistent risk measures, then the value functions $V_t(w_{t-1}, \xi^t)$, $t = T-2, ..., 2$ are defined by the optimal value of problems

$$
\begin{aligned}
\mathit{min} \quad & \alpha_t(\xi^t) + (1-\beta)^{-1}\mathbb{E}_{\mathbb{P}_{t+1}}\left(\left[V_{t+1}\left(w_t(\xi^t), \xi^{t+1}\right) - \alpha_t(\xi^t)\right]^+\right) \\
\mathit{s.t.} \quad & w_t \in \mathbb{W}_t,\ s_t \in \mathbb{S}_t,\ b_t \in \mathbb{B}_t,\ \alpha_t \in \mathcal{L}_{k^t,1} \\
& 1^\top b_t(\xi^t) - 1^\top s_t(\xi^t) = 0 \\
& w_{t-1}(\xi^{t-1})\xi_t + (1 - c_b)\, b_t(\xi^t) - (1 + c_s)\, s_t(\xi^t) = w_t(\xi^t),
\end{aligned}
\tag{4.44}
$$

and in the first stage $t = 1$ the corresponding problem is

$$
\begin{aligned}
\mathit{min} \quad & \alpha_1(\xi^1) + (1-\beta)^{-1}\mathbb{E}_{\mathbb{P}_2}\left(\left[V_2\left(w_1(\xi^1), \xi^2\right) - \alpha_1(\xi^1)\right]^+\right) \\
\mathit{s.t.} \quad & w_1 \in \mathbb{W}_1,\ s_1 \in \mathbb{S}_1,\ b_1 \in \mathbb{B}_1,\ \alpha_1 \in \mathcal{L}_{k^1,1} \\
& 1^\top b_1(\xi^1) - 1^\top s_1(\xi^1) = W_0 \\
& (1 - c_b)\, b_1(\xi^1) - (1 + c_s)\, s_1(\xi^1) = w_1(\xi^1).
\end{aligned}
\tag{4.45}
$$

By combining dynamic equations 4.43, 4.44 and 4.45 into one model and introducing variables $z_t(\xi^t)$, $t = 2, ..., T$ to avoid expressions $[x]^+$ appearing in the objective function, we obtain the following problem

$$
\begin{aligned}
min \quad & \alpha_1(\xi^1) + (1-\beta)^{-1}\mathbb{E}_{\mathbb{P}_2}\left(z_2(\xi^2)\right) & \\
s.t. \quad & w_t \in \mathbb{W}_t,\ s_t \in \mathbb{S}_t,\ b_t \in \mathbb{B}_t,\ \alpha_t \in \mathcal{L}_{k^t,1} & t = 1, ..., T-1 \\
& z_t \in \mathcal{L}_{k^t,1} & t = 2, ..., T \\
& \mathbb{E}\left(w_{T-1}^\top(\xi^{T-1})\xi_T\right) \geq \bar{r}_p & \\
& 1^\top b_1(\xi^1) - 1^\top s_1(\xi^1) = W_0 & \\
& (1-c_b)\, b_1(\xi^1) - (1+c_s)\, s_1(\xi^1) = w_1(\xi^1) & \\
& -w_{T-1}^\top(\xi^{T-1})\xi_T - \alpha_{T-1}(\xi^{T-1}) \leq z_T(\xi^T) & \\
& z_t(\xi^t) + \alpha_{t-1}(\xi^{t-1}) \geq \alpha_t(\xi^t) + (1-\beta)^{-1}\mathbb{E}_{\mathbb{P}_{t+1}}\left(z_{t+1}(\xi^{t+1})\right) & t = 2, ..., T-1 \\
& z_t(\xi^t) \geq 0 & t = 2, ..., T \\
& 1^\top b_t(\xi^t) - 1^\top s_t(\xi^t) = 0 & t = 2, .., T-1 \\
& w_{t-1}(\xi^{t-1})\xi_t + (1-c_b)\, b_t(\xi^t) - (1+c_s)\, s_t(\xi^t) = w_t(\xi^t) & t = 2, .., T-1.
\end{aligned}
\tag{4.46}
$$

Proposition 4.4: Problem 4.46 is time consistent.

Proof: Since problem 4.46 is obtained from dynamic equations 4.43, 4.44 and 4.45, we have to show that they satisfy Proposition 4.2. For each stage $t = 1, ..., T-2$, we can formulate the objective function as

$$
\phi_t\left(f_t(w_t,\xi_t), V_{t+1}(w_t,\xi^{t+1})\,\middle|\,\xi^t\right) = \alpha_t(\xi^t\,|\,\xi^t) + (1-\beta)^{-1}\mathbb{E}_{\mathbb{P}_{t+1}}\left(\left[V_{t+1}(x_t(\xi^t),\xi^{t+1}) - \alpha_t(\xi^t)\right]^+\,\middle|\,\xi^t\right). \tag{4.47}
$$

Thus, problem 4.46 is time consistent. ■

4.3 Portfolio optimization as stochastic programming

Since portfolio optimization problems are stochastic programs, they can all be approximated with scenario tree approximation and polynomial decision rule approximations.

Scenario tree approximation

Multistage mean-variance portfolio optimization with scenario tree approximation has been addressed in [12] and time inconsistent mean-CVaR in [15]. We have not found any work addressing time consistent mean-CVaR with scenario tree approximation.

Decision rule approximations

Instead of using scenario tree approximation, it is also possible to use polynomial decision rule approximations. Note, that the recourse matrix A depends on the uncertain parameters ξ and thus our extension from Chapter 3 is needed to tackle such problems. Portfolio optimization problems discussed in this chapter do not satisfy Assumption 3.2*, which is required for polynomial decision rule approximations. In order to avoid this problem, we have to determine bounds on each uncertain parameter ξ (i.e. return). We propose two approaches:

1. Under the assumption that returns are normally distributed, we can define the bounds as the γ and $1-\gamma$ percentiles of the return distribution.
2. If historical data is available, we can define the bounds as the minimal and the maximal return observed in the history.

Denote by ξ_i^{min} and ξ_i^{max} the minimal and the maximal value of the uncertain parameter ξ_i $i = 1, ..., k$ obtained with any of the above approaches, respectively. We can then formulate a compact semi-algebraic set Ξ in two ways.

1. For each ξ_i, $i = 1, ..., k$, we define two constraints, $\xi_i - \xi_i^{min} \geq 0$ and $\xi_i^{max} - \xi_i \geq 0$. In this case, the functions that define the compact semi-algebraic set Ξ are all linear. By considering equation 2.22, we can see that this definition of constraints is most useful when the degree of the polynomial decision rules d is odd. In this case $d - d_r$, $r = 1, ..., 2k$, is always even and no degree is lost by the floor function.
2. For each ξ_i $i = 1, ..., k$ we define one constraint $(\xi_i - \xi_i^{min})\,(\xi_i^{max} - \xi_i) \geq 0$. In this case, the functions that define the compact semi-algebraic set Ξ are all quadratic functions. By considering equation 2.22, we can see that this definition of constraints is most useful when the degree of the polynomial decision rules d is even. In this case $d - d_r$, $r = 1, ..., k$, is always even and no degree is lost by the floor function.

Another deviation from the stochastic programming problems in Chapter 3 is the existence of expectation in the recourse matrix. This can be solved by replacing elements of the monomial vector $\mathcal{B}_d(\xi)$ that include expectation with the corresponding elements of the moment matrix M_d.

Chapter 5

Numerical evaluation

In this chapter we evaluate the models we developed in the previous chapters for two concrete problems. The first problem is an electricity capacity expansion problem and the second is the portfolio optimization problem.

We have discussed in the previous chapters that the approximation of the solution of every stochastic programming problem is obtained in two steps:

1. Approximation of an SP problem with a tractable SDP problem (preprocessing): Preprocessing was implemented in Matlab 2010b while using optimization toolbox Yalmip [41].

2. Solving the SDP problem: Two state of the art solvers SDPT3 [20, 21] and Sedumi [22] were used for this purpose.

For both concrete problems, Monte Carlo Simulation [47] was used to evaluate the moment matrix M_d. All numerical evaluations were conducted on a 3.20GHz, Intel Core i5 CPU 650 machine with 8GB of RAM.

Let us now focus on each of the concrete problems separately.

5.1 Electricity capacity expansion model

5.1.1 Problem description

Electricity capacity expansion model is the first concrete problem that we tested the polynomial decision rule approximations on. The model is taken from [30]. Imagine that we are given five regions $R = \{1, 2, 3, 4, 5\}$ with uncertain electricity demand δ_r, $r \in R$. In each of the regions 1, 3 and 5 there is one power plant denoted by $N = \{1, 2, 3\}$, respectively. Each of the power plants can produce up to $\overline{g}_n$ units of energy at uncertain costs ζ_n, $n \in N$. In order to distribute energy among the regions, some directed transmission lines $M = \{1, 2, 3, 4, 5\}$ with maximal capacity of $\overline{f}_m$ units of energy are used (Figure 4.2).

The problem that we would like to solve is designed as a two-stage stochastic model. In the first stage, we have to decide which of the existing power plants and transmission lines should be extended. Each power plant $N = \{1, 2, 3\}$ can be extended by a factor $1 + u_n$, $u_n \in [0, 1]$, at unit cost c_n. Similarly, each of the transmission lines $M = \{1, 2, 3, 4, 5\}$ can be extended by a factor $1 + v_m$, $v_m \in [0, 1]$, at unit cost d_m. In the second stage, after the uncertain parameters, i.e. the demand for each region δ_r, $r \in R$ and the operating costs for each power plant ζ_n, $n \in N$, are revealed, the power plants are put into operation. Thus, we must decide on the number g_n of units of energy each power plant $N \in \{1, 2, 3\}$ must produce and on the number f_m of units of energy that will be transmitted through each transmission line $M \in \{1, 2, 3, 4, 5\}$ in order to satisfy the demand for each region $R = \{1, 2, 3, 4, 5\}$ almost surely.

The objective is to minimize the expectation of sum of all the expansion and the operating costs.

Model that mathematically describes the reasoning above is given by

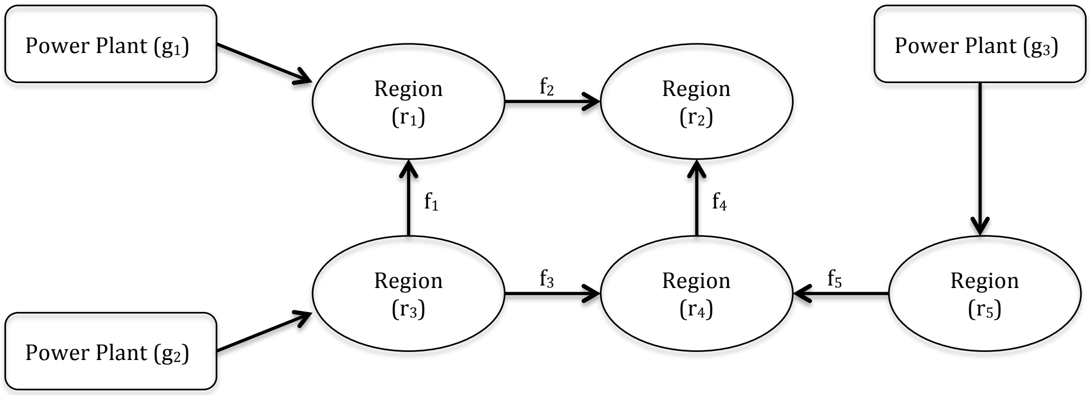

Figure 5.1: Electricity capacity expansion model.

Parameter	Value
$\overline{g_n}$	$3.5, \forall n \in N$
$\overline{f_n}$	$3.5, \forall n \in M$
c_1	1.0
c_2	0.4
c_3	1.5
d_1	5.0
d_2	0.2
d_3	0.4
d_4	0.6
d_5	0.1

Parameter	Value
δ_1	$[0.3, 1.5]$
δ_2	$[0.36, 1.8]$
δ_3	$[0.42, 2.1]$
δ_4	$[0.48, 2.4]$
δ_5	$[0.54, 2.7]$
ζ_1	$[0.2, 1]$
ζ_2	$[0.2, 0.5]$
ζ_3	$[1, 2]$

Table 5.1: Model parameters.

$$
\begin{aligned}
min \quad & \textstyle\sum_{n\in N} c_n u_n + \sum_{m\in M} d_m v_m + \mathbb{E}(\sum_{n\in N} \zeta_n g_n(\xi)) \\
s.t. \quad & u \in \mathbb{R}^3,\ v \in \mathbb{R}^5,\ g \in \mathcal{L}_{8,3},\ f \in \mathcal{L}_{8,5} \\
& \left.\begin{array}{ll}
0 \le u_n \le 1 & \forall n \in N \\
0 \le v_m \le 1 & \forall m \in M \\
0 \le g_n(\xi) \le \overline{g}_n(1+u_n) & \forall n \in N \\
0 \le f_m(\xi) \le \overline{f}_m(1+v_m) & \forall m \in M \\
g_1(\xi) + f_1(\xi) \ge f_2(\xi) + \delta_1 & \\
f_2(\xi) + f_4(\xi) \ge \delta_2 & \\
g_2(\xi) \ge f_1(\xi) + f_3(\xi) + \delta_3 & \\
f_3(\xi) + f_5(\xi) \ge f_4(\xi) + \delta_4 & \\
g_3(\xi) \ge f_5(\xi) + \delta_5 &
\end{array}\right\} \mathbb{P} - a.s.
\end{aligned}
\tag{5.1}
$$

where $\xi = (\delta, \zeta)$. Values for each parameter are shown in Table 5.1.

The objective function sums up power plant and transmission lines expansion costs and costs

Degree	LB preproc.	LB solving	UB preproc.	UB solving
1	2 s	1 s	4 s	1 s
2	2 s	16 s	5 s	8 s
3	10 min 21 s	1h 57 min 29 s	1 min 20 s	14 min 4 s
4	10 min 25 s	/	1 min 23 s	28h 34 min 50 s

Table 5.2: Preprocessing and solving time.

to produce the demanded amount of energy. The first two constraints limit the maximal possible expansion of the power plants and transmission lines. The third constraint ensures that production of each power plant does not exceed its capacity. Similarly, the fourth constraint ensures that energy transmitted through each of the transmitting lines does not exceed its capacity. Note that energy can only be transmitted in one direction. The last five constraints ensure that for each region the inputs of electricity are bigger than or equal to the sum of outputs and consumption.

5.1.2 Results

We approximated the electricity capacity expansion model for different degrees of polynomial decision rules d with both state of the art solvers, SDPT3 and Sedumi. The results obtained from each solver and the time needed to solve the problem were not significantly different. Thus, we will not make any distinction between the solvers. However, for the upper bound approximation with polynomial decision rules of the degree $d = 4$, we only obtained a result with Sedumi. SDPT3, after a few iterations, reported numerical problems and was not able to give a reasonable solution. None of the solvers was able to give a reasonable lower bound approximation for polynomial decision rules of the degree $d = 4$.

In Table 5.2 we present the time needed to approximate the SP problem with a tractable SDP problem (preprocessing) and the time needed to solve the problem. We noticed that the increase of the preprocessing time is much smaller than the increase in solving time, as the size of the input parameters increases.

We present Table 5.2 in a graph (Figure 5.2). Note, that we used an exponential scale to represent time. We can clearly see the exponential growth of the solving time for both the upper and the lower bound approximation as the degree of the polynomial decision rules d increases. Moreover, we can see that the preprocessing time changes significantly only every odd degree. The most time consuming part in the preprocessing is the construction of matrices Y (the definition in Proposition 3.4), which only change size every odd degree.

Table 5.3 shows the optimal solutions for the lower and the upper bound approximation for each degree d. The error is calculated by the following equation

$$Error_d = \frac{UB_d - LB_d}{0.5(UB_d + LB_d)}. \tag{5.2}$$

Since we did not obtain a lower bound solution for $d = 4$, we used the lower bound solution from $d = 3$, for the purpose of the error calculation, instead. The same problem has been approximated with linear and piecewise linear decision rules and an error of 41% and 16%, respectively, was reported in [30]. We can thus see that, for this example, even the quadratic polynomial decision rules outperform the piecewise linear decision rules.

Figure 5.3 depicts the optimal solutions for each degree d in a graphical form. We can clearly see how increasing the degree d improves the approximation of both, the lower and the upper bound, problems.

In all calculations we used the findings described in the end of the previous chapter. If the degree of the polynomial decision rules d was odd, then compact semi-algebraic set Ξ, was defined by linear functions. Similarly, if the degree of the polynomial decision rules d was even, then the compact semi-algebraic set Ξ, was defined by quadratic functions. Without this approach, approximations

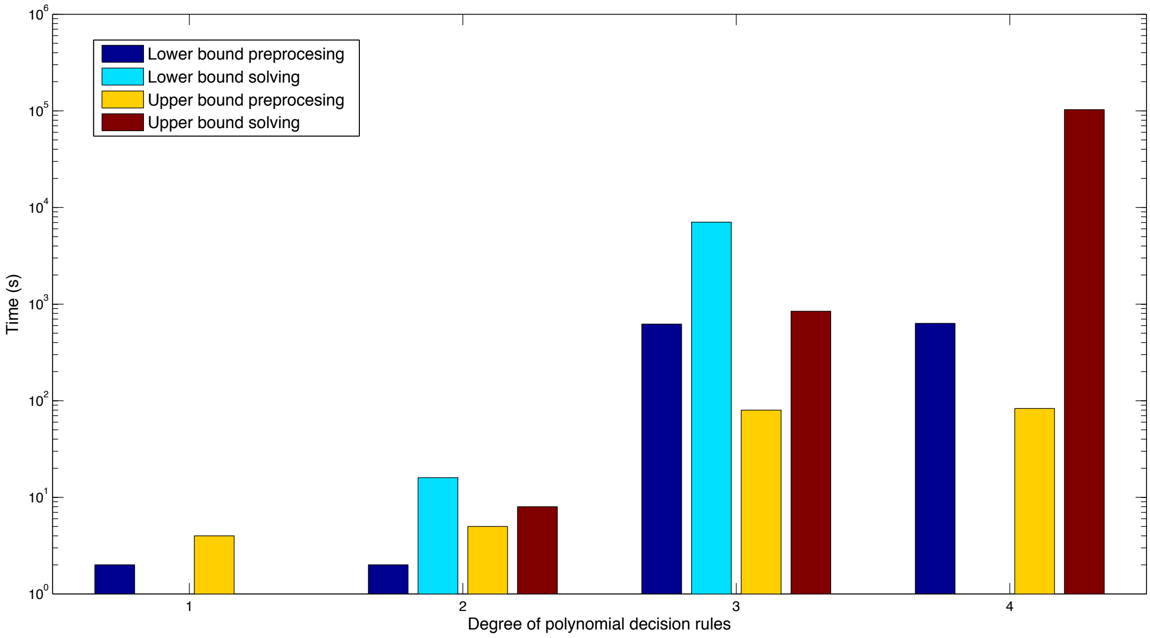

Figure 5.2: Preprocessing and solving times.

Degree	Lower bound	Upper bound	Error
1	2.024	3.053	41%
2	2.398	2.737	13%
3	2.483	2.717	9%
4	/	2.663	7%

Table 5.3: Optimal solutions.

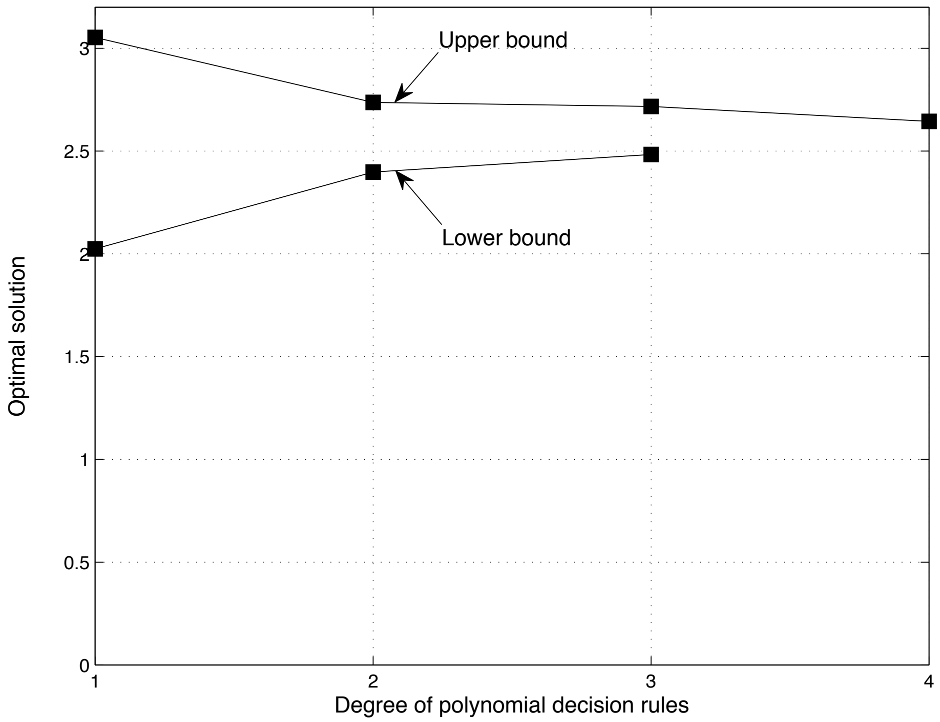

Figure 5.3: Optimal solutions.

	US stocks	Int stocks	Corp bnd	Gvnt bnd	Cash
Mean	10.80	10.37	9.49	7.90	5.61
STD	15.72	16.75	6.57	4.89	0.70

	US stocks	Int stocks	Corp bnd	Gvnt bnd	Cash
US stocks	1.00	0.601	0.247	0.062	0.094
Int stocks	0.601	1.00	0.125	0.027	0.006
Corp bnd	0.247	0.125	1.00	0.883	0.194
Gvnt bnd	0.062	0.027	0.883	1.00	0.27
Cash	0.094	0.006	0.194	0.27	1.00

Table 5.4: Expected return, standard deviations and correlation matrix.

	US stocks	Int stocks	Corp bnd	Gvnt bnd	Cash
Minimal value	-70.93	-76.72	-25.16	-17.52	1.97
Maximal value	92.53	97.46	43.16	33.32	9.25

Table 5.5: Bounds for the returns.

would only improve for every even (in case of the quadratic formulation) or every odd degree (in case of the linear formulation).

5.2 Portfolio optimization

In this section we validate single and multistage portfolio optimization problems. Based on our experience from the previous section, we expect that it would be illusionary to expect to obtain precise solutions when tens of assets are included in the portfolio. Thus, instead of using stocks and bonds directly, we decided to use indexes to represent the assets. We used real data obtained from [49]. Stock were divided into US stocks (represented by the MSCI US index) and international stocks (represented by the MSCI EAFE&C index). Bonds were similarly, grouped into US Corporate Bonds (represented by the Salomon Brothers US Corp Bnd index) and US government bonds (represented by the Salomon Brothers US Corp Bnd index). We used 3 month Treasury notes (represented by the GP Morgan US 3M index) for cash. For the further argumentation, we assume that returns are normally distributed. Expected return, standard deviation and correlation matrix for all the assets are given in Table 5.4.

Transaction costs c_b and c_s are set to 1%. In order to get bounds for each of the uncertain returns, the first approach described in Section 4.3 was applied with parameter $\gamma = 10^{-7}$. Bounds for each asset are given in Table 5.5. In all the following calculations, we used the findings described in Section 4.3. If the degree of the polynomial decision rules d was odd, then compact semi-algebraic set Ξ, was defined by linear functions. Similarly, if the degree of the polynomial decision rules d was even, then the compact semi-algebraic set Ξ, was defined by quadratic functions.

5.2.1 Single-stage Portfolio Optimization

We start our discussion with a single-stage mean-CVaR portfolio optimization problem. We first show how the approximation error changes with a degree of polynomial decision rules d, requested portfolio return $\bar{r}_p$ and parameter β (equation 3.72). In Table 5.6 we first present errors for $\beta = 0.9$, $\bar{r}_p = 0.1$ and all possible degrees d that we were able to obtain a reasonable solution for. We can see that polynomial decision rules significantly improve the solution, i.e., from an error of 24% for $d = 2$ to an error of 5%. A similar result was obtained for $\beta = 0.95$, $\bar{r}_p = 0.1$ and all possible degrees d that we were able to obtain a reasonable solution for. Note, however, that in case of $\beta = 0.95$,

Degree	Lower bound	Upper bound	Error
2	4.89	-20.93	25.82
3	1.28	-16.39	17.67
4	-6.24	-10.93	4.69

Degree	Lower bound	Upper bound	Error
2	0.02	-34.96	34.94
3	-1.24	-25.97	24.73
4	-7.86	-15.45	7.59

Table 5.6: Errors for $\bar{r}_p = 0.1$ and $\beta = 0.9$ (left) or $\beta = 0.95$ (right).

Degree	Lower bound	Upper bound	Error
2	5.51	-4.13	9.64
3	4.69	-2.73	7.42
4	2.82	0.01	2.81

Degree	Lower bound	Upper bound	Error
2	4.33	-10.06	14.39
3	3.24	-6.66	9.90
4	2.82	-2.44	5.26

Table 5.7: Errors for $\bar{r}_p = 0.08$ and $\beta = 0.9$ (left) or $\beta = 0.95$ (right).

the error is bigger than for $\beta = 0.9$. This can be explained through the VaR values which in our case correspond to the $1 - \beta$ percentiles of the normal distribution. For $\beta > 0.5$, the larger β is, the more its change affects the VaR value and consequently the CVaR value. Thus, it is more difficult to find a good approximation.

In Table 5.7, a similar test was performed for portfolio returns $\bar{r}_p = 0.08$. We see that by decreasing the expected portfolio return $\bar{r}_p$, also the error decreases. It is clear that a portfolio with smaller expected return consists of less risky assets and consequently has tighter bounds for the portfolio return. Tighter bounds and smaller standard deviation of the return lead to a more precise result.

In the rest of this section, we only focus on polynomial decision rules of degree $d = 4$, since they give the best results we were able to obtain. Figures 5.4, 5.5 and 5.6 show efficient frontiers for $\beta = 0.9$, $\beta = 0.95$ and $\beta = 0.99$, respectively.

Note that only the upper bound solution is feasible and can therefore be implemented. From the definition of CVaR in Section 4.1.2, it is expected that a higher β will have higher CVaR values. This is clearly visible in Figure 5.7. Moreover, note that CVaR increases as the required expected return increases. The main reason for this phenomena is that CVaR is a risk averse measure, which increases if standard deviation of the portfolio increases. It is clear that investing in assets with higher standard deviation leads to more risky portfolios.

In order to investigate the errors more in detail, we have to consider the nature of the function used to calculate the CVaR value. The function inside the expected value is a piecewise linear function in $\mathbb{R}^k$, where the kink appears as the function reaches 0. Parameter $\alpha(\xi^1)$ is a constant and thus only shifts the function up or down. To graphically support this reasoning, we plot $z(\xi)$ for a simplified problem where only two assets (US stocks and International stocks) are available as shown in Figure 5.9. Similarly, Figure 5.9 shows $z(\xi)$ from the top, where colours are used to represent the values of $z(\xi)$. From those figures it is clearly visible that $z(\xi)$ is a piecewise linear function with a kink approximately at $\xi_1 = 1.92 - \xi_2$. Piecewise linear functions are difficult to approximate by polynomial functions of low degrees and this is the main reason for relatively high errors even for the single-stage portfolio optimization problem. Note, however, that polynomial decision rules gave us a good approximation for the kink. Detection of the kink is very important if piecewise linear decision rules are used. Thus, a combination of polynomial decision rules and piecewise linear decision rules could lead to good approximation results. We list this idea as one of the directions for the future work.

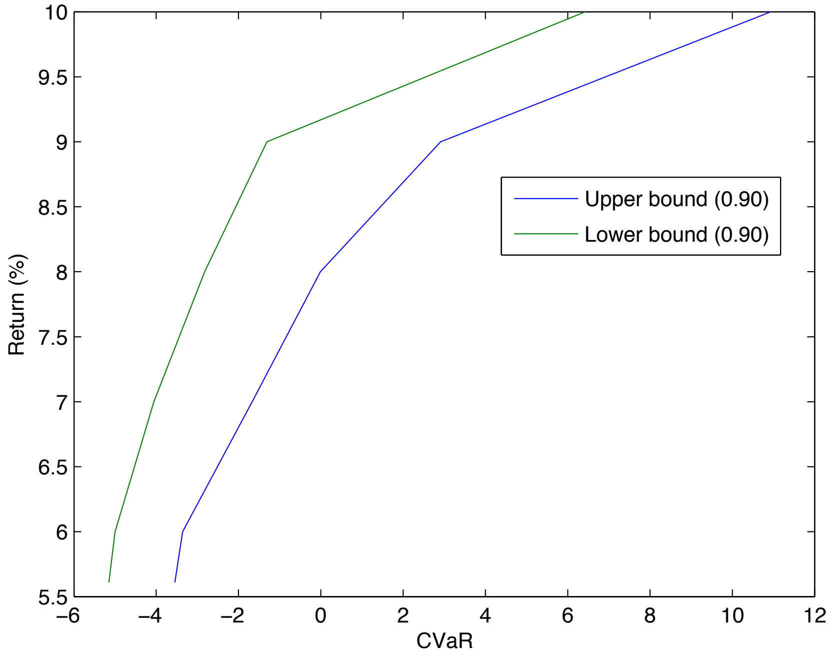

Figure 5.4: Mean-CVaR efficient frontier for $\beta = 0.9$.

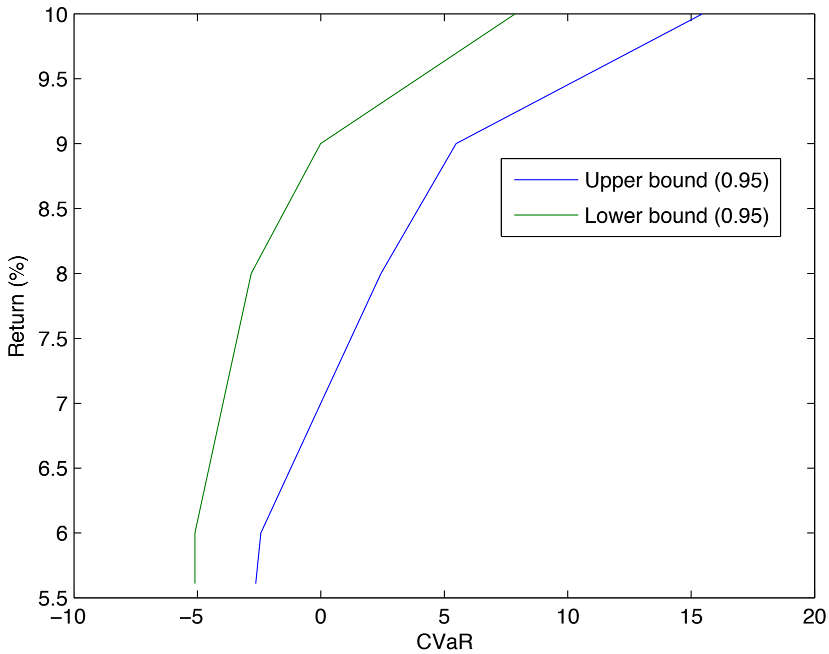

Figure 5.5: Mean-CVaR efficient frontier for $\beta = 0.95$.

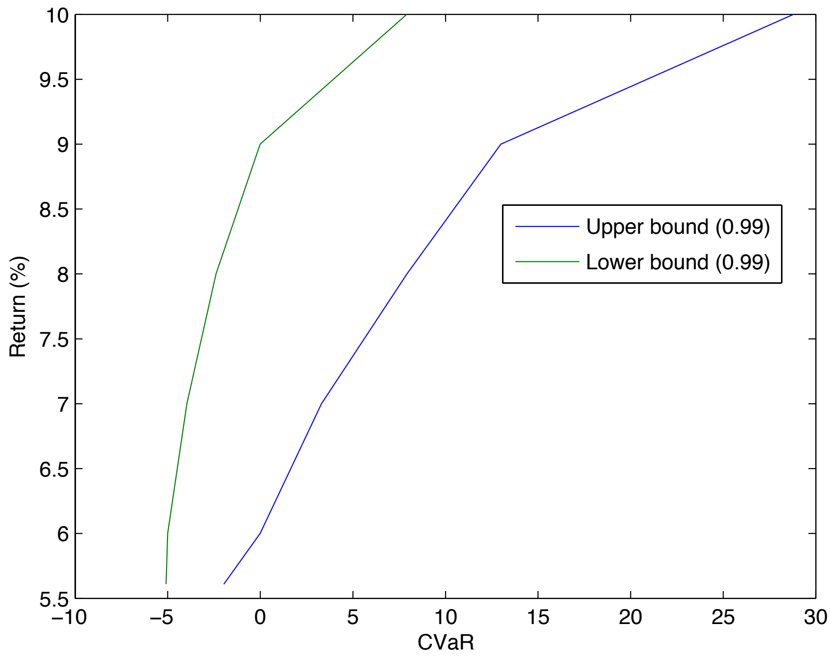

Figure 5.6: Mean-CVaR efficient frontier for $\beta = 0.99$.

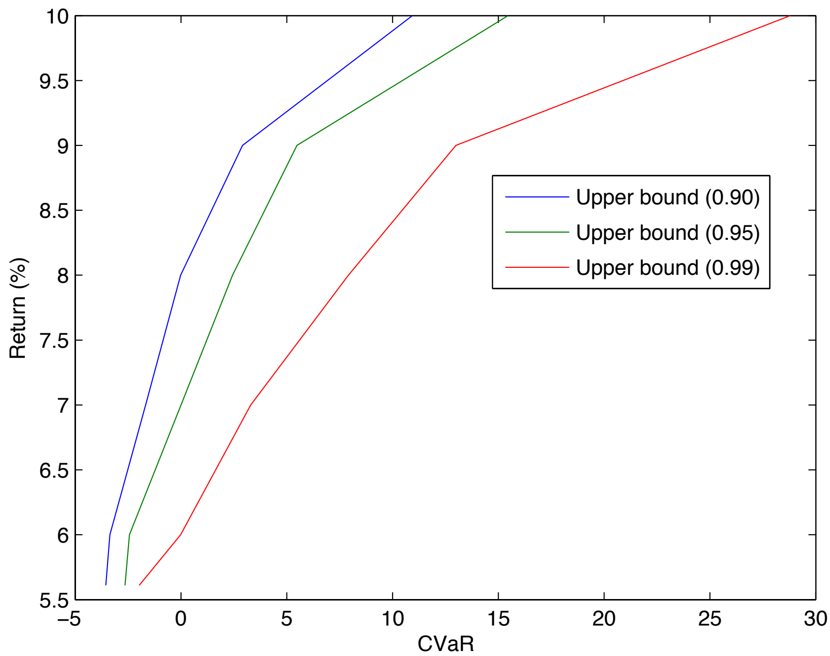

Figure 5.7: Upper bound approximations for $\beta = \{0.9, 0.95, 0.99\}$.

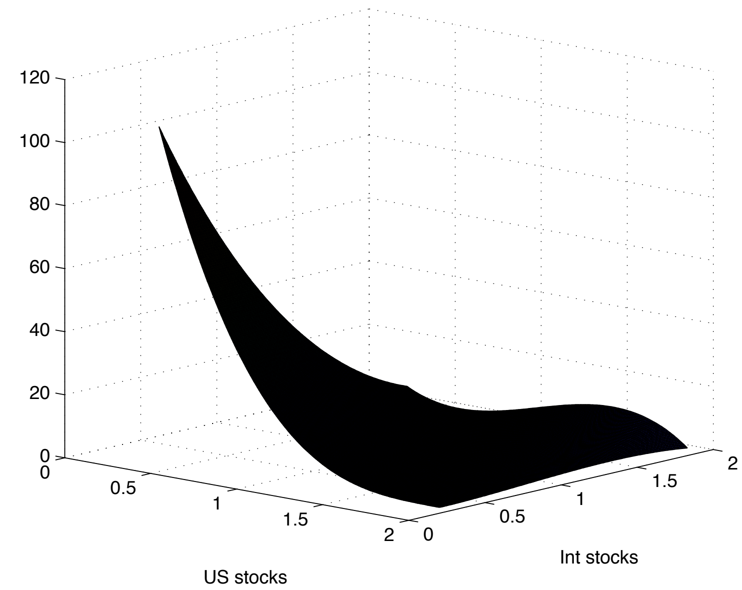

Figure 5.8: Optimal function $z(\xi)$ when only two assets, US stocks and International stocks, are used - side view.

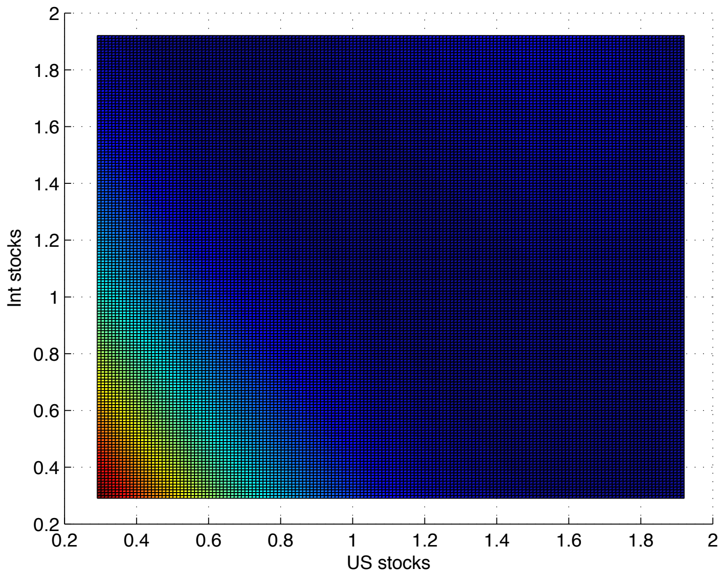

Figure 5.9: Optimal function $z(\xi)$ when only two assets, US stocks and International stocks, are used - top view.

Degree	LB	UB	Error	LB - single	UB - single	Error
2	9.36	-57.68	67.04	4.57	-34.95	39.52
3	/	-26.40	35.76	/	-14.20	18.77

Degree	LB	UB	Error	LB - single	UB - single	Error
2	8.11	-66.92	75.03	3.98	-42.48	46.46
3	/	-40.03	48.14	/	-22.56	26.54

Table 5.8: Errors for $\bar{r}_p = 0.1$ and $\beta = 0.9$ (top) or $\beta = 0.95$ (bottom).

Degree	LB	UB	Error	LB - single	UB - single	Error
2	10.89	-17.6	28.49	5.30	-9.23	14.52
3	/	-5.04	15.93	/	-2.55	7.85

Degree	LB	UB	Error	LB - single	UB - single	Error
2	10.94	-22.16	33.10	5.32	-11.77	17.09
3	/	-11.76	22.70	/	-6.06	11.38

Table 5.9: Errors for $\bar{r}_p = 0.08$ and $\beta = 0.9$ (top) or $\beta = 0.95$ (bottom).

5.2.2 Multistage portfolio optimization

Analysis of the numerical results for the multistage portfolio optimization problems is structured in a similar way. We were only able to obtain results for multistage portfolio optimization problems with two investment periods, i.e. $T = 3$, because portfolio optimization problems with more stages turned out to be infeasible for polynomial decision rules of small degrees.

We first show how the approximation error changes with a the degree of polynomial decision rules d, requested portfolio return $\bar{r}_p$ and parameter β (equation 3.72). In Table 5.8 we present errors for $\beta = 0.9$, $\bar{r}_p = 0.1$ and all possible degrees d, for which we were able to obtain a reasonable solution. Note that only solutions of degree 2 for the lower bound and up to degree 3 for the upper bound approximation were obtained. For higher degrees, SDP solvers reported numerical problems. We can again see that polynomial decision rules of higher degrees significantly improve the solution, i.e., from an error of 67% for $d = 2$ to an error of 36% for $d = 3$. A similar result was obtained also for $\beta = 0.95$, $\bar{r}_p = 0.1$ and all possible degrees d, for which we were able to obtain a reasonable solution. The values were expected and a similar reasoning than in the previous section can be applied.

Since CVaR is calculated for two investment periods, we have to calculate the corresponding one-stage CVaR, in order to compare solutions from the single and multistage portfolio optimization problems. The corresponding one-stage CVaR r_1 is calculated from the multistage CVaR r_2 as

$$r_1 = \sqrt{1 + r_2} - 1. \tag{5.3}$$

The corresponding lower and upper bound values are shown in Tables 5.8 and 5.9. We can see that the errors for the multistage portfolio optimization problem are much bigger than the corresponding single-stage errors. Note, that we obtained error smaller than 10% only for $\beta = 0.9$ and $\bar{r}_p = 0.08$.

In the rest of this section we only focus on polynomial decision rules of the degree $d = 3$ for the upper bound and polynomial decision rules of degree $d = 2$ for the lower bound approximation, since they gave us the best approximations we were able to obtain. The efficient frontiers for different β are shown on Figures 5.10, 5.11 and 5.12. The shape of the functions is similar to one-stage portfolio optimization problems. Note, however, that the lower bound approximation is worse, because the degree of the polynomial decision rules $d = 2$ is very small.

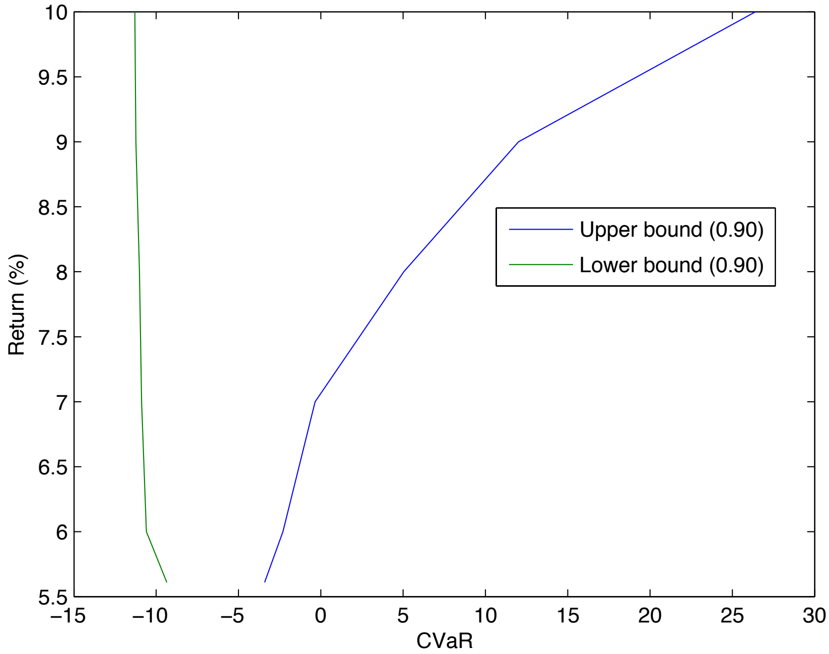

Figure 5.10: Mean-CVaR efficient frontier for $\beta = 0.90$.

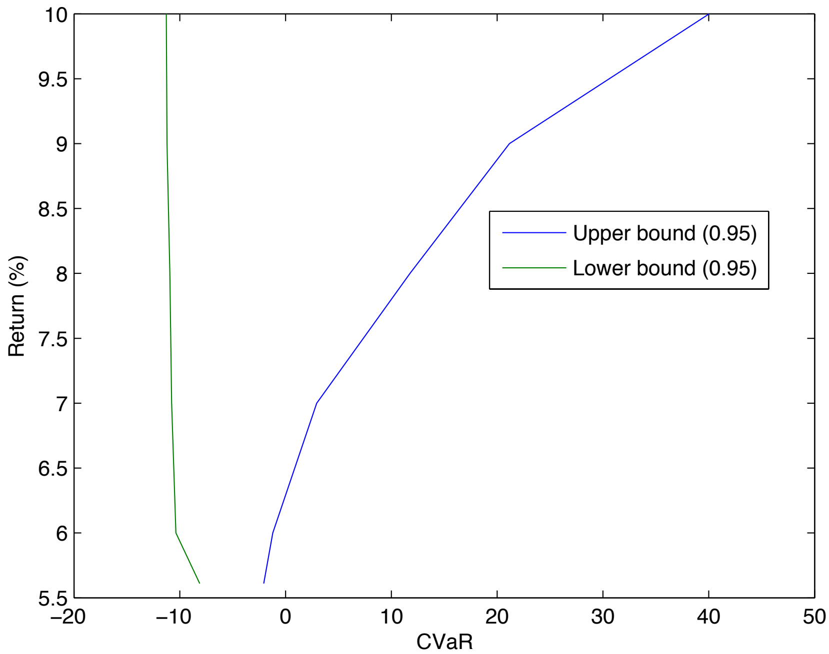

Figure 5.11: Mean-CVaR efficient frontier for $\beta = 0.95$.

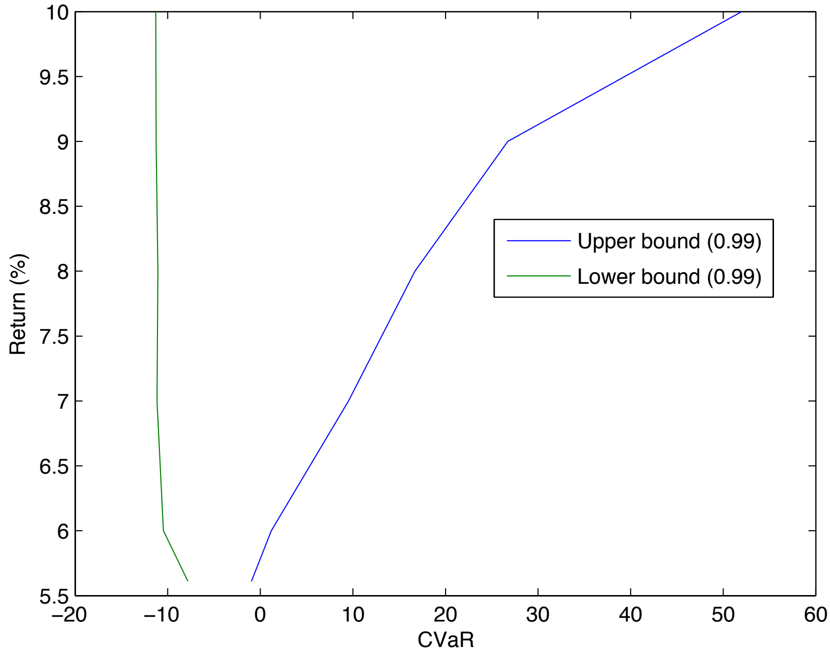

Figure 5.12: Mean-CVaR efficient frontier for $\beta = 0.99$.

On Figure 5.13 all upper bounds are shown together. We can again, as expected, see that higher βs lead to higher CVaR values.

In order to explain the main reasons for the errors obtained, we again have to consider the nature of the function used to calculate CVaR and the constraints. The objective function of the multistage stage problem is equal to the objective function of the single-stage problem. Note, however, that in the multistage case additional constraints

$$z_t(\xi^t) + \alpha_{t-1}(\xi^{t-1}) \geq \alpha_t(\xi^t) + (1-\beta)^{-1}\mathbb{E}_{\mathbb{P}_{t+1}}\left(z_{t+1}(\xi^{t+1})\right) \tag{5.4}$$

$t = 2, ..., T-1$ appear. Functions $z_t(\xi^t)$ are thus general piecewise functions and not piecewise linear functions as in the single-stage case. Side view and the top view of the function $z_2(\xi)$ for a simplified problem where only two assets, US stocks and International stocks, are used, and two investment periods, i.e. $T = 3$, are considered, is given on Figures 5.14 and 5.15. We know that $\alpha_1(\xi^1) \in \mathbb{R}$. The side and the top view of the function $\alpha_2(\xi^2)$ is shown on Figures 5.16 and 5.17, respectively. The kink in the function $z_2(\xi)$ is again clearly visible. Since we do not know anything else about the optimal functions $z_t(\xi^t)$, it is thus difficult to estimate the error from this source more precisely. Based on the figures presented, one could argue that the error must be relatively small, because functions $z_2(\xi^2)$ and $\alpha_2(\xi^2)$ clearly resemble the piecewise linear and linear function, respectively, even though the degree of polynomial decision rules is high enough to allow more flexibility. However, this is only a speculation, because there is no theoretical evidence to support this reasoning.

Consider now the constraints

$$w_{t-1}(\xi^{t-1})\xi_t + (1-c_b)\, b_t(\xi^t) - (1+c_s)\, s_t(\xi^t) = w_t(\xi^t), \tag{5.5}$$

where $t = 2, .., T-1$. Note that all decision rules in this equation are of degree $d-1$ due to Assumption 3.4. Thus, the first term is the only term of degree d and all other therms are of degree $d-1$. Since the polynomial equality must hold for all $\xi^t \in \Xi$, the vector-valued polynomial must vanish identically on a set with nonempty interior Ξ. This is possible if and only if all the coefficients of the polynomial vanish. Since only the first term is of degree d and the other terms are of degree $d-1$, all coefficients of degree $d-1$ of the decision rule $w_{t-1}(\xi^{t-1})$ must be 0. Thus,

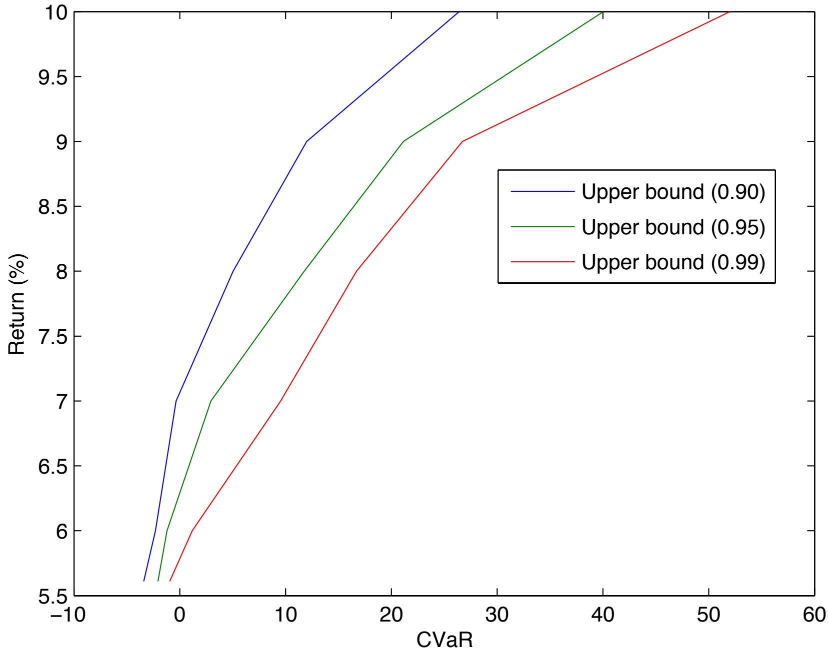

Figure 5.13: Upper bound approximations for $\beta = \{0.9, 0.95, 0.99\}$.

$w_{t-1}(\xi^{t-1})$ is actually of degree $d-2$. By recursively applying a similar reasoning, we can see that decision rule $w_{t-2}(\xi^{t-2})$ is of degree $d-3$ etc. In case of multistage portfolio optimization problems with two investment periods, for example, $w_1(\xi^1)$ is of degree $d-2$ and $w_2(\xi^2)$ is of degree $d-1$. Note, that for portfolio optimization problems with three investment periods, the minimal degree $d = 3$ of polynomial decision rules is needed in order to obtain a feasible problem. This explains the infeasibility detected for the multistage portfolio optimization problem with three investment periods, i.e. $T = 4$, which we mentioned in the beginning of this section. Since the degree of polynomial decision rules actually decreases in each stage, this has severe effects on the approximation quality. Lower bound solutions presented in this section are, for example, actually constants, i.e. polynomial functions of degree 0, which clearly does not give enough flexibility to obtain good approximations.

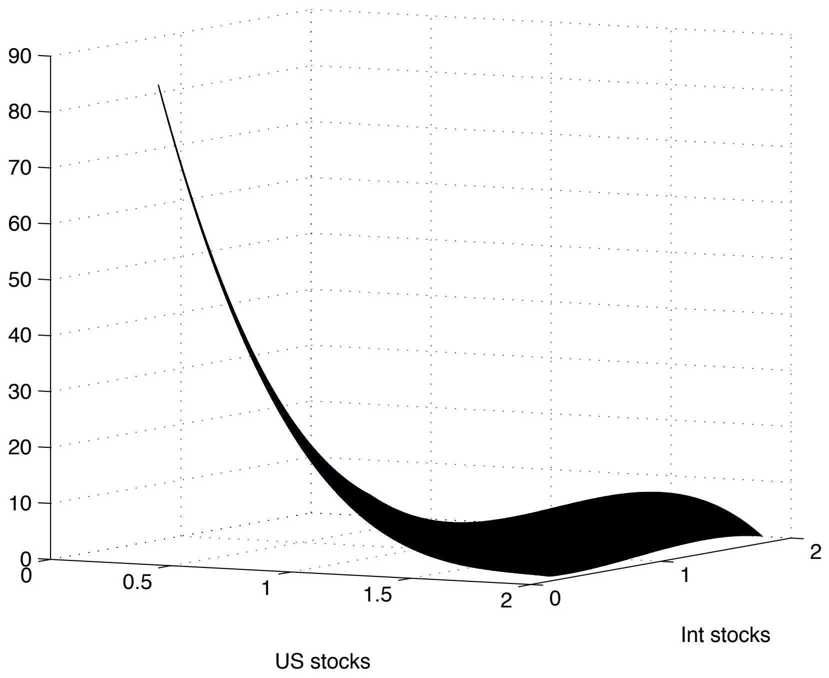

Figure 5.14: Optimal function $z_2(\xi)$ when only two assets, US stocks and International stocks, are used - side view.

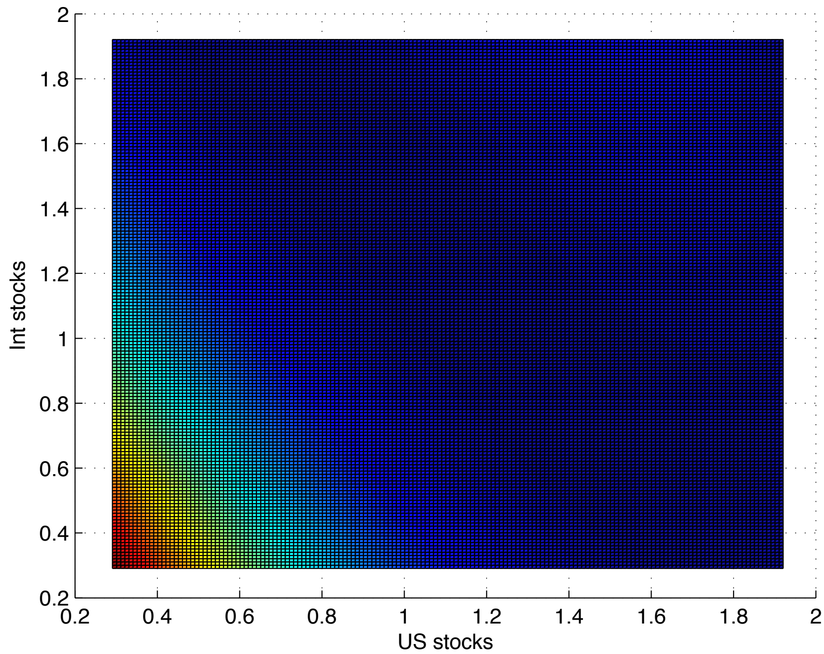

Figure 5.15: Optimal function $z_2(\xi)$ when only two assets, US stocks and International stocks, are used - top view.

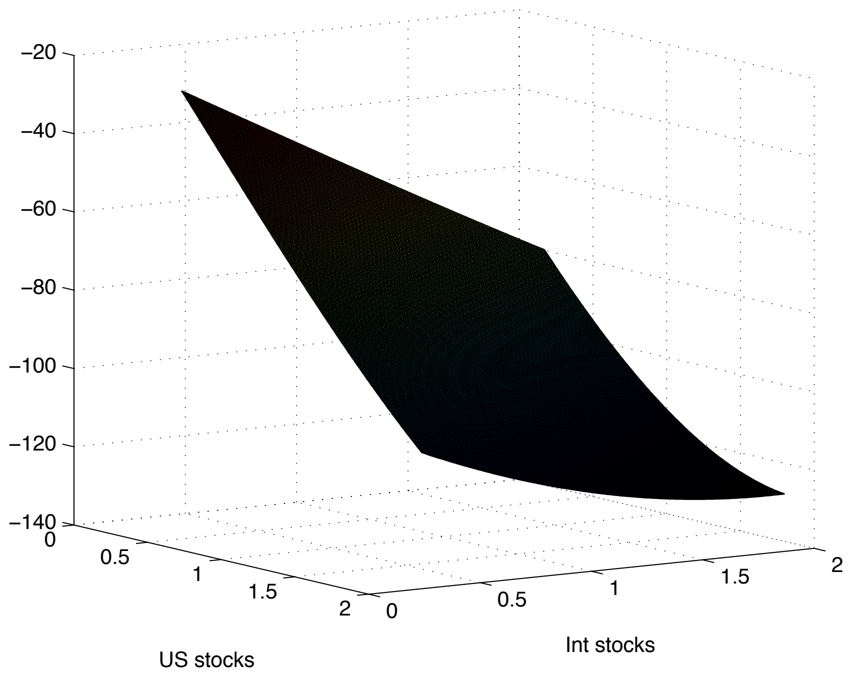

Figure 5.16: Optimal function $\alpha_2(\xi)$ when only two assets, US stocks and International stocks, are used - side view.

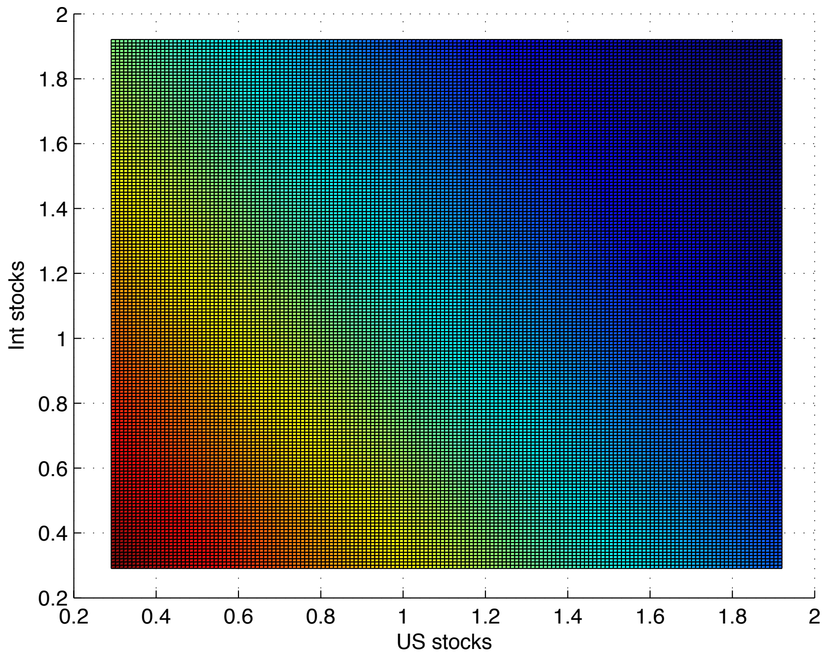

Figure 5.17: Optimal function $\alpha_2(\xi)$ when only two assets, US stocks and International stocks, are used - top view.

Chapter 6

Conclusions

We started this work by introducing tractable deterministic optimization problems, namely linear programming and semidefinite programming. We analyzed polynomial optimization techniques proposed by Parrilo, Putinar, Lasserre, Schmüdgen etc. in order to determine their assumptions and consequently their usability for stochastic programming problems. Two approximation techniques for tackling multistage stochastic problems were analyzed. We showed that scenario tree approximation leads to a tractable linear programming problem, but its complexity grows exponentially with the number of stages of the original stochastic programming problem. Moreover, there is no means for estimating the suboptimality of solutions. On the other hand, decision rule approximations grow only quadratically with the number of stages of the original stochastic programming problem, and Kuhn's lower bound approximation formulation, gives a tractable approach to estimate the approximation error. Linear decision rule approximations lead to a tractable linear programming problem and polynomial decision rules lead to a tractable semidefinite programming problem. Unfortunately, the existing, state of the art semidefinite programming solvers are still not robust and relatively slow. Thus, they are considered to be a bottleneck when stochastic programming with polynomial decision rules is applied on real-life problems. In this work we proposed an extension, which widens the spectrum of problems that can be solved by polynomial decision rule approximations. We released an assumption that the recourse matrix must be independent of the uncertain parameters and instead model it as polynomial functions of the uncertain parameters. This enables polynomial decision rule approximations to be applied on some important problems, such as portfolio optimization. Moreover, we released Bampou's assumption that polynomial decision rules must be of an even degree and formulate upper and lower bound approximations, which hold for polynomial decision rules of all degrees.

We then described single and multistage portfolio optimization problems, where the main focus was on the risk measures. We explained axioms of a coherent risk measure and listed the advantages and disadvantages of the most used risk measures, such as variance, value at risk and conditional value at risk. We then extended single-stage risk measures the the multistage setting. We showed why time consistency is an important property of the risk measures and proved that multistage conditional value at risk is not time consistent. Additionally, we also proposed its time consistent alternative.

We then showed how polynomial decision rule approximations perform on electricity capacity expansion problem. An approximation error of 6% was obtained, which is precise enough for many real life applications. We numerically evaluated also single and multistage portfolio optimization problems. An approximation error of a few percents was obtained for the single stage and an error around 20% for the multistage setting. Since multistage stochastic problems are generally computationally intractable even when only medium-accuracy solutions are required [11], this is a relatively small approximation error, even though it is still not good enough for real life applications. We explained the main sources of the error and suggest how a combination of polynomial and piecewise linear decision rules could be used together in order to obtain even better approximations.

6.1 Future Direction of Research

In this section we present some possible extensions of this work.

- Weierstrass theorem tells us that any continuous function on a closed and bounded interval can be uniformly approximated on that interval by polynomials to any degree of accuracy. Approximation theory claims that the orthogonal polynomials approximate functions better than the polynomials with an ordinary basis. If $f(\xi) \in \mathcal{L}_{k,n}$ and $p(\xi) : \mathbb{R}^k \longrightarrow \mathbb{R}$ a real valued polynomial, then our goal is to minimize the approximation error $|f(\xi) - p(\xi)|$. It is known that one can obtain polynomial $p(\xi)$ very close to the optimal one by expanding the function $f(\xi)$ in terms of Chebyshev polynomials. We thus believe that by using orthogonal Chebyshev polynomials, one could obtain better approximations. The biggest challenge of this extension is to determine multivariate orthogonal polynomials over a bounded closed semi-algebraic set Ξ. Such orthogonal polynomials would then replace the current vector of monomials $\mathcal{B}_d(\xi)$. Note, that this approach does not at all change the computational complexity of the problem and a better approximation is obtained for free.

- Another approach that could improve the approximation quality for some problems is using trigonometric decision rules instead of polynomial ones. If the optimal solution of the stochastic programming problem is a periodic function, then it can be approximated by a sum of a set of trigonometric functions as given by the Fourier series. This approximation also leads to a tractable semidefinite programming problem.

- We showed that single-stage CVaR is a piecewise linear function and thus, we expect piecewise linear decision rules to outperform polynomial decision rules, if the parameters of the piecewise linear decision rules are chosen precisely. Polynomial decision rules turned out to be useful for the kink detection, which is one of the most cumbersome parameters when piecewise linear decision rules are applied. We believe that such combination of a polynomial decision rules are piecewise linear decision rules could lead to better approximations of single-stage portfolio optimization problems.

- We showed that multistage CVaR is a piecewise function and thus, it is expected that piecewise polynomial decision rules would outperform the ordinary polynomial decision rules, if the parameters of the piecewise polynomial decision rules are chosen precisely. The main challenge in this extension is to formulate upper and lower bound approximations by using piecewise polynomial decision rules and determine good input parameters of such piecewise polynomial decision rules.

Bibliography

[1] M. Putinar, “Positive polynomials on compact semi-algebraic sets”, *Indiana University Mathematics Journal*, vol. 42, no.2, pp. 969-984, 1993.

[2] P. A. Parrilo, “Semidefinite programming relaxations for semialgebraic problems”, *Mathematical Programming*, vol. 96, no. 2, pp. 293-320, 2003.

[3] D. Kuhn, W. Wiesemann, and A. Georghiou, “Primal and dual linear decision rules in stochastic and robust optimization” *Mathematical Programming*, vol. 120, no. 2, 2009.

[4] P. Kall and S. W. Wallace, *Stochastic Programming*, John Wiley & Sons, Chichester, 1994.

[5] S. Wallace, and W. E. Ziemba, *Applications of stochastic programming*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2005.

[6] M. Dyer, and L. Stougie, “Computational complexity of stochastic programming problems”, *Math. Program.*, vol. 106, no. 3, pp. 423–432, 2006.

[7] D. Bampou, D. Kuhn, “Scenario-Free Stochastic Programming with Polynomial Decision Rules”, IEEE Conference on Decision and Control and European Control Conference, Orlando, USA, December 2011.

[8] Murty, “Some NP-complete problems in quadratic and nonlinear programming”, *Mathematical Programming*, vol. 39, pp. 117-129, 1987.

[9] A. M. G. Solo, “Multidimensional Matrix Mathematics: Multidimensional Matrix Equality, Addition, Subtraction, and Multiplication”, *Proceedings of the World Congrees on Engineering 2010*, vol 3., London, 2010.

[10] E. K. Haviland, “On the momentum problem for distribution functions in more than one dimension”, *Amarican Journal of Mathematics*, vol. 57, no. 3, pp. 562-568, 1935.

[11] A. Shapiro, A. Nemirovski, “On complexity of stochastic programming problems”, *Continous Optimization*, pp. 111-146, 2005.

[12] N. Gulpinar, B. Rustem, and R. Settergren, “Multistage stochastic mean-variance portfolio analysis with transaction costs”, *Innovations in Financial and Economic Networks*, 2003.

[13] K. Hoyland and S. W. Wallace, “Generating scenario trees for multistage decision problems”, *Management Science*, vol. 47, no. 2, pp. 295-307, February 2001.

[14] P. Klaassen, “Comment on "generating scenario trees for multistage decision problems"”, *Management Science*, vol. 48, no. 11, pp. 1512-1516, November 2002.

[15] W. N. Cho, *Robust Portfolio Optimization Using Conditional Value At Risk*, Master Thesis, Imperial College London, June 2008.

[16] H. Heitsch, W. Römisch, “Scenario reduction algorithms in stochastics programming”, *Comput. Optim. Appl.*, 2003.

[17] P. Klaassen, "Financial asset-pricing theory and stochastic programming models for asset/liability management: A synthesis", *Manage. Scie.*, pp. 31-48, 1998.

[18] T. Pennanen, "Epi-convergent siscretizations of multistage stochastic programs", *Math. Oper. Res.*, pp. 245-256, 2005.

[19] A. Ben-Tal, A. Goryashako, E. Guslitzed and A. Nemirovski, "Adjustable robust solutions of uncertain linear programs", *Math. Prog.*, pp. 351-376, 2004.

[20] K. C. Toh, M. J. Todd, and R. H. Tutuncu, "SDPT3 — a Matlab software package for semidefinite programming", *Optimization Methods and Software*, pp. 545-581, 1999.

[21] R. H. Tutuncu, K. C. Toh, and M. J. Todd, "Solving semidefinite-quadratic-linear programs using SDPT3", *Mathematical Programming Ser. B*, pp. 189–217, 2003.

[22] J. F. Sturm, "Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones", *Optimization Methods and Software*, vol. 11-12, pp. 625-653, 1999.

[23] J. B. Lasserre, "Global Optimization with Polynomials and The Problem of Moments", *SIAM Journal Optimization*, 2001.

[24] A. Ben-Tal. Conic and Robust Optimization, "Lecture Notes for the Course", University of Rome, July 2002.

[25] P. A. Parrilo, "On a decomposition for multivariable forms via LMI mothods", *In Proceedings of the American Control Conference*, 2000.

[26] J. Bochnak, M. Coste, and M. F. Roy, *Real Algebraic Geometry*, Springer-Verlag, 1998.

[27] K. Schmüdgen, "The K-moment problem for compact semi-algebraic sets", *Math. Ann.*, 1991.

[28] T. Pennanen, "Epi-convergent discretizations of multistage stochastic programs", *Math. Oper. Res.,* 2005.

[29] X. Chen, M. Sim, P. Sun, and J. Zhang, "A linear decision-based approximation approach to stochastic programming", *Operations Research*, vol. 56, no. 2, pp. 344-357, 2008.

[30] A. Georghiou, W. Wiesemann, and D. Kuhn, "Generalized decision rule approximations for stochstic programming via liftings", *Available on Optimization Online*, 2010.

[31] D. Bertsimas, D. A. Iancu, and P. A. Parrilo, "A hierarchy of near-optimal policies for multistage adaptive optimization", *Conditionally accepted in IEEE Transactions and Automatic Control*, 2009.

[32] P. Artzner, F. Delbaen, J. M. Eber, D. Heath, "Coherent Measures of Risk", *Mathematical Finance*, 2011.

[33] R.T. Rockafellar and S. Uryasev, "Optimization of conditional value at risk", *Journal of Risk*, 2000.

[34] H. Markowitz, "Portfolio Selection", *Journal of Finance*, 1952.

[35] Jonas Palmquist, S. Uryasev, and Pavlo Krokhalm, "Portfolio optimization with conditional value-at-risk objective and constraints", *Research report, University of Florida, Center for Applied Optimization*, 1999.

[36] B. Rudloff, A. Street, D. Valladao, "Time consistency and risk averse dynamic decision models: Interpretation and practical consequences", *Princeton University and PUC-Rio*, December 2010.

[37] B. Acciaio, P. I. Beatrice, “Dynamic risk measures”, *Quantitative Finance*, 2010.

[38] B. Roorda, J.M. Schumacher, “Time consistency conditions for acceptability measures, with an application to tail value at risk”, *Insurance: Mathematics and Economics*, 2007.

[39] A. Shapiro, “On a time consistency concept in risk averse multistage stochastic programming”, *Operations Research Letters - ORL*, vol. 37, no. 3, pp. 143-147, 2009.

[40] R. Bellman, *Dynamic Programming*, Princeton University Press, Princeton, NJ. Republished 2003: Dover.

[41] J. Löfberg, “YALMIP: A toolbox for modelling and optimization is MATLAB”, *Proceedings of the CACSD Conference*, Taipei, Taiwan, 2004.

[42] Andre F. Perold, “Large-scale portfolio optimization”, *Management Science*, vol. 30, no. 10, pp. 1143-1160, October 1984.

[43] Murty, G. Katta, *Linear programming*, New York: John Wiley & Sons. 1983

[44] N. Karmarkar. “A New Polynomial-Time Algorithm for Linear Programming”, *Combinatorica*, 1984.

[45] S. Boyd, L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.

[46] S. Wright, “Primal-dual aggregation and disaggregation for stochastic linear programs”, *Math. Oper. Res.,* vol. 19, no. 4, pp. 893-908, 1994.

[47] G. S. Fishman, *Monte Carlo: Concepts, Algorithms, and Applications*, New York: Springer, 1995.

[48] D. Goldfarb and G. Iyengar, “Robust Portfolio Selection Problems”, Mathematics of Operations Research, vol. 28, no. 1, pp. 1-38, 2003.

[49] G. Infanger, “Dynamic Asset Allocation Strategies using a Stochastic Dynamic Programming Approach”, *Handbook of Asset and Liability Management*, vol. 1, pp. 200-251, 2006.