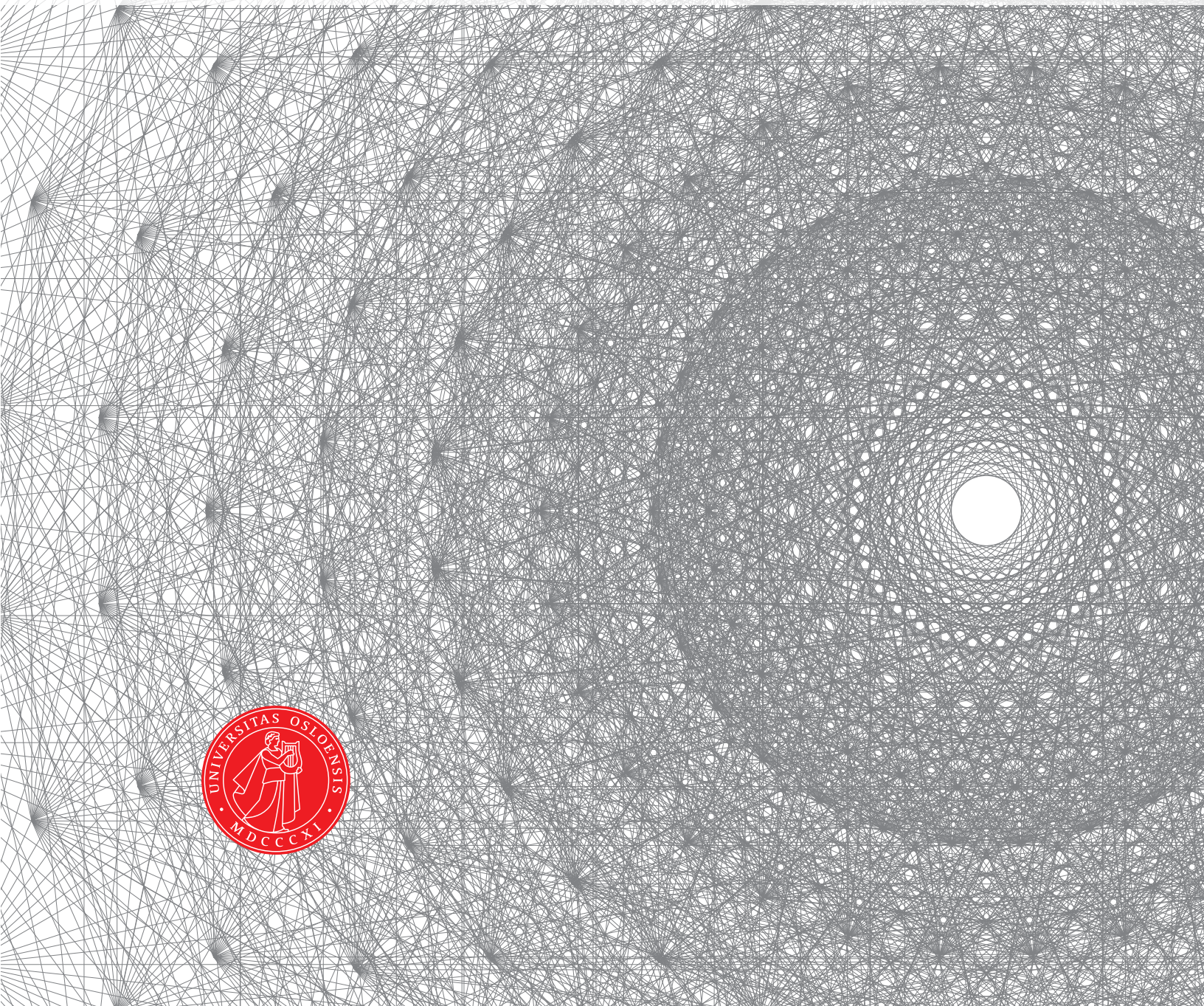
UiO : Department of Mathematics
University of Oslo

Portfolio optimization with Conditional Value-at-Risk constraints

Hilde Marianne Skjelbred Sannes

Master's Thesis, Spring 2016



Cover design by Martin Helsø

The front page depicts a section of the root system of the exceptional Lie group E_8, projected into the plane. Lie groups were invented by the Norwegian mathematician Sophus Lie (1842–1899) to express symmetries in differential equations and today they play a central role in various parts of mathematics.

Preface

Abstract

This thesis aims to study the risk measure Conditional Value-at-Risk and analyse an optimization problem of maximizing expected return subject to this risk measure. The analysis include performing Fourier-Motzkin eliminations on the system of linear constraints of the problem, so that the portfolio is the only remaining decision variable.

Acknowledgements

First of all I would like to thank my supervisor, Geir Dahl, for always taking time to answer my questions, and for pointing me in the right direction.

Also, I very much appreciate the work Katinka put in proofreading my thesis, even though she had her own exams to worry about.

I would like to thank my boyfriend, Gaute, and my family for their love and support. Gaute deserves an extra 'thanks' for checking my thesis for wrong spellings.

And lastly, but not lest, I would like to thank my fellow students for interesting conversations and for making every day enjoyable. Among these, I would like to thank Johannes for interesting discussions on the topics of this thesis.

Oslo, May 2016 Hilde.

Contents

1 Introduction

In portfolio optimization, since the future market prices are not known, one is dealing with uncertainties. This implies that making decisions involve some risk. Finding 'good' measures of risk has since the Markowitz mean-variance optimization problem in the 1950s been an essential problem. There has been several approaches, some which has been flawed. In 1999 Philippe Artzner, in 'Coherent risk measures', gave some reasonable axioms for risk measures.

The risk measure Value-at-Risk was introduced to measure an upper estimate of losses of a portfolio, but this risk measure lacked properties like for instance convexity for discrete loss distributions. This made it undesirable in optimization problems. To remedy Value-at-Risk's shortcomings, the related risk measure Conditional Value-at-Risk was introduced.

This thesis aims to analyse an optimization model with this relatively new risk measure. In doing so, the results and concepts from the main article for this thesis, by Krokhmal, Palmquist and Uryasev [KPU01], is presented. The results and concepts from this article is used to describe an optimization model of maximizing expected return subject to Conditional Value-at-Risk constraints. In this optimization model, the decision variables are the threshold, the variables representing the losses that are exceeding the threshold, and the portfolio. At the end of this thesis I perform Fourier-Motzkin eliminations on the system of constraints in the Conditional Value-at-Risk optimization model. This elimination results in new, equivalent optimization models where the portfolio is the only decision variable. In some cases, as we will see, these new optimization models not only have fewer decision variables, they also have fewer constraints.

In Chapter 2 I will give a short introduction to convex sets and functions. In

this chapter I also introduce the basics of linear programming and quadratic programming, and aim to clarify the connection between convexity and optimization. The final section in this chapter gives an introduction to Fourier-Motzkin elimination. This technique of eliminating variables in a linear system of inequalities can be used in optimization problems, where the constraints are linear. The Fourier-Motzkin elimination is of importance in this thesis, as my main contributions are performing Fourier-Motzkin eliminations on the constraints of the Conditional Value-at-Risk optimization problem.

In Chapter 3 I lie down the financial framework for this thesis, from a stochastic calculus point of view. This chapter also gives an introduction to convex risk measures. In the final section I present the risk measures Value-at-Risk and Conditional Value-at-Risk, and some of their properties. In particular, I show that Conditional Value-at-Risk is a coherent risk measure, and hence also a convex risk measure.

Chapter 4 gives an introduction to portfolio optimization problems. I give a short introduction to the classical Markowitz mean-variance optimization problem, and Markowitz's concept of efficient portfolios. In this chapter I introduce the main optimization problem of this thesis; the problem of maximizing expected return subject to the risk measure Conditional Value-at-Risk. The presentation of this model is based on the article by Krokhmal et. al. [KPU01]. In addition, some additional constraints are introduced.

In Chapter 5 I aim to analyse an linearized, one period optimization model with Conditional Value-at-Risk constraints. At the end of this chapter I perform Fourier-Motzkin elimination on this one period model, under some conditions on the scenario probabilities.

Chapter 6 contains some efficient frontiers for the Conditional Value-at-Risk problem.

In chapter 7 I give some of the Matlab code I have produced working with this thesis. Some code is left out, due to their simplicity.

1.1 My contributions

My main contribution in this thesis is the work I have done in Section 5.2, Section 5.3 and Section 5.4; the Fourier-Motzkin elimination on the system of linear constraints in the one period optimization model with Conditional Value-at-Risk constraints. In these sections I show that we can eliminate all decision variables except those representing the portfolio. This results in some new, simplified systems of linear constraints. I have also proved some results and solved some examples to clarify some of the concepts presented. Below is a list of my contributions in this thesis:

- CHAPTER 2:
  - The proof of Proposition 2.1.7.

- CHAPTER 3:
  - In the proof of Theorem 3.3.2, I proved some details, in particular that the the function $F_\alpha(\xi)$ is convex.
  - I produced some details in proving the convexity property in Theorem 3.3.3
  - The proof of Proposition 3.3.4, except the proof of the convexity property.

- CHAPTER 5: Most of this chapter consists of my ideas.
  - Section 5.2. In particular, Theorem 5.2.1 and its proof.
  - Section 5.3. In particular, Lemma 5.3.1 and its proof.
  - Section 5.4.

- CHAPTER 6: This chapter consists of the efficient frontiers I have produced.

- APPENDIX A: This appendix consists of some of the Matlab code I have produced.

2 | Convexity and optimization

Optimization is the mathematical theory of minimization or maximization problems, and is an important theory in for instance finance. The goal in optimization is to find the point x^* such that $f(x^*) \leq f(x)$ for all x. Such a point x^* is called a *global minimum*. The problem is that most numerical methods for finding such a minimum, often only find a local minimum, i.e a point $\hat{x}$ that is minimum for points 'sufficiently near' $\hat{x}$. Even though such a minimum is good locally, it may be very poor compared to the global minimum. If, however, the function f we are looking at is convex, then local minima are also global. This is a very important fact, and is why convexity is fundamental in optimization.

2.1 Basics of convexity

In this thesis, the goal is to investigate some concepts and optimization models in finance. In doing so, it is important to look at convexity, as it also is of great importance in financial optimization. In this section we will restrict the attention to convexity in $\mathbb{R}^n$, since the sets we are interested in will be the set of some real vectors.

This section is based on the report by Dahl [Dah10].

Convex sets

In optimization, the set of feasible points is often a convex set. This is the case in for instance linear programming. Here, the convexity of the feasible set plays a role in the existence of optimal solutions and how to solve optimization problems numerically. Also, for the applications in this thesis, we will later see that some risk sets are convex sets.

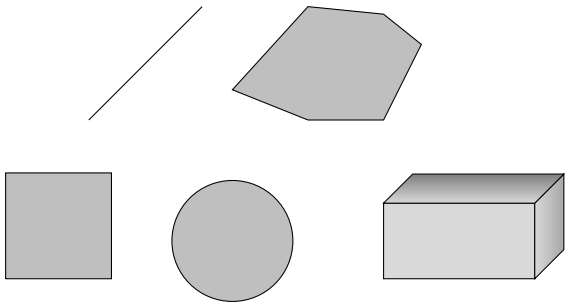


Figure 2.1: Some convex sets.

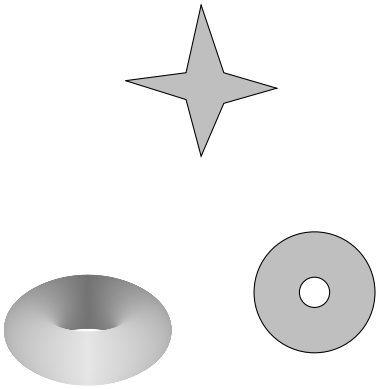


Figure 2.2: Some non-convex sets.

Definition 2.1.1. *A set $C \subseteq \mathbb{R}^n$ is called* convex *if*

$$(1-\lambda)\mathbf{x} + \lambda\mathbf{y} \in C, \textit{for all } \mathbf{x}, \mathbf{y} \in C \textit{ and } 0 \leq \lambda \leq 1$$

Geometrically, a set is convex if it contains the line segments between each point in the set. For an alternative definition of convex sets, let $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n \in \mathbb{R}^n$ be vectors and $\lambda_j \geq 0$ for $j = 1, 2, \ldots, n$ such that $\sum_{j=1}^{n} \lambda_j = 1$. Then the vector $\mathbf{x} = \sum_{j=i}^{n} \mathbf{x}_j \lambda_j$ is called a *convex combination* of the vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n \in \mathbb{R}^n$. A convex set is a set that is closed under convex combinations.

The expectation of a random variable, in a discrete setting, relates to the latter definition of convexity. Let X be the random variable taking values in $\{x_1, x_2, \ldots, x_n\}$, and let $0 \leq p_j \leq 1$ be the probability that $X = x_j$. Then the expectation $E[X] = \sum_{j=1}^{n} p_j x_j$ is a convex combination of $\{x_1, x_2, \ldots, x_n\}$.

Proposition 2.1.2. *A set is convex if and only if it contains all convex combinations of its points.*

Let $A \in \mathbb{R}^{m\times n}$, $\mathbf{x} = (x_1, x_2, \ldots, x_n)^T \in \mathbb{R}^n$ and $\mathbf{b} = (b_1, b_2, \ldots, b_m)^T \in \mathbb{R}^m$. The notation $A\mathbf{x} \leq \mathbf{b}$ should throughout this thesis be interpreted component wise. That is, for each row $\mathbf{a}_i$ in A, the inequality $\mathbf{a}_i\mathbf{x} \leq b_i$ holds for each $i = 1, 2, \ldots, m$.

Definition 2.1.3. *A* polyhedron *is a set of the form* $\{\mathbf{x} \in \mathbb{R}^n : A\mathbf{x} \leq \mathbf{b}\}$, *where* $A \in \mathbb{R}^{m\times n}$ *and* $\mathbf{b} \in \mathbb{R}^m$.

A polyhedron is a special type of convex sets, it is the solution set of a linear system. Let $S \subseteq \mathbb{R}^n$ be any set, then the *convex hull* of S is the set of all convex combinations of points in S. We denote this set by *conv(S)*. The convex hull of a set S is the smallest convex set containing S.

Definition 2.1.4. *A set P is called a* polytope *if it is the convex hull of a finite number of points.*

Often, in optimization problems, the feasible set is closed and the optimal solution lies on the boundary of this feasible set. It is therefore of interest to introduce some basic knowledge of topology. In the definition below, $||\mathbf{x}|| \in \mathbb{R}$ is the norm of vector $\mathbf{x}$, representing the length of $\mathbf{x}$.

Definition 2.1.5. *Some useful definitions:*

i) *An* open ball *is a set* $B^o(\mathbf{a}, r) = \{\mathbf{x} \in \mathbb{R}^n : ||\mathbf{x} - \mathbf{a}|| < r\}$, *where* $r \geq 0$.

ii) *Every open ball is an* open set. *Also, a set* $S \subseteq \mathbb{R}^n$ *is open if every point in S is contained in an open ball, i.e for each* $\mathbf{x} \in S$ *there is an* ϵ *such that* $\mathbf{x} \in B^o(\mathbf{a}, \epsilon) \subseteq S$.

iii) *A set S is* closed *if its set compliment* S^c *is open.*

iv) *A set S i called* bounded *if there exists a number M such that* $||\mathbf{x}|| \leq M$ *for all* $\mathbf{x} \in S$.

v) *A set is* compact *if it is closed and bounded.*

vi) *The* interior *of S, denoted* $\text{int}(S)$, *is the union of all open sets contained in S.*

vii) *The* closure *of S, denoted* $\text{cl}(S)$, *is the intersection of all closed sets containing S.*

viii) *The* boundary *of S, denoted* $\text{bd}(S)$, *is defined by* $\text{bd}(S)=\text{cl}(S)\setminus \text{int}(S)$.

Polyhedrons are closed sets. The feasible set of a linear programming problem is a polyhedron, and the optimal solution lies on this polyhedron's boundary. A set is a polytope if and only if it is a bounded polyhedron. This is an important theorem in convexity.

Convex functions

From calculus we know that a function $f\colon \mathbb{R}^n \to \mathbb{R}$ is called convex if $f'' \geq 0$, where f'' denotes the second derivative of f. Geometrically, the graph of such a function 'bends upward'. We will in this thesis use the following definition of a convex function.

Definition 2.1.6. *Let $C \subseteq \mathbb{R}^n$ be a convex set, then the real function $f\colon C \to \mathbb{R}^n$ is* convex *if*

$$f((1-\lambda)\mathbf{x} + \lambda\mathbf{y}) \leq (1-\lambda)f(\mathbf{x}) + \lambda f(\mathbf{y})$$

holds for every $\mathbf{x}, \mathbf{y} \in C$ and every $0 \leq \lambda \leq 1$.

Geometrically, for $f\colon \mathbb{R} \to \mathbb{R}$, this new definition says that the line segment between each pair $(x_1, f(x_1))$ and $(x_2, f(x_2))$ lies above the graph of f in the interval $[x_1, x_2]$.

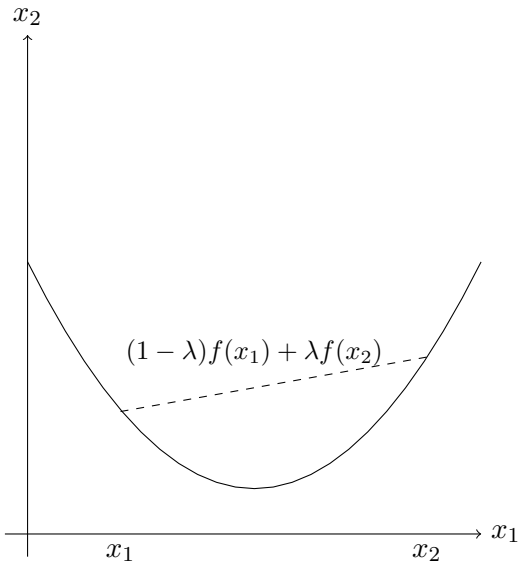


Figure 2.3: A convex function.

Proposition 2.1.7. *The sum of convex functions is a convex function.*

Proof. Let the functions $g_i\colon \mathbb{R}^n \to \mathbb{R}$ be convex for $i = 1, 2, \ldots, m$, and let $f\colon \mathbb{R}^n \to \mathbb{R}$ be such that $f(\mathbf{x}) = \sum_{i=1}^{m} g_i(\mathbf{x})$. Then for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, and $0 \leq \lambda \leq 1$

$$\begin{aligned} f((1-\lambda)\mathbf{x}+\lambda\mathbf{y}) &= \sum_{i=1}^{m} g_i((1-\lambda)\mathbf{x}+\lambda\mathbf{y}) \\ &\leq \sum_{i=1}^{m} \Big((1-\lambda)g_i(\mathbf{x})+\lambda g_i(\mathbf{y})\Big) \\ &= (1-\lambda)\sum_{i=1}^{m} g_i(\mathbf{x})+\lambda\sum_{i=1}^{m} g_i(\mathbf{y}) \\ &= (1-\lambda)f(\mathbf{x})+\lambda f(\mathbf{y}) \end{aligned}$$

This proves that f is convex.

■

Definition 2.1.8. *A function f is called* concave *if $-f$ is convex.*

The next result will be useful in showing that our main risk measure in this thesis, Conditional Value-at-Risk, is a convex function. The proof of this theorem will not be stated here.

Theorem 2.1.9. *Let g, h be convex functions on $C \subseteq \mathbb{R}^n$. Then*

$$f(\mathbf{x}) = \max\{g(\mathbf{x}), h(\mathbf{x})\}$$

is convex for $\mathbf{x} \in C$.

The following corollary is very important in convex optimization, and states that for convex functions, local minima are also global. Let $\nabla f(\mathbf{x})$ denote the gradient of f at $\mathbf{x}$.

Corollary 2.1.10. *Let $f\colon C \to \mathbb{R}$ be a differentiable convex function, defined on an open convex set $C \subseteq \mathbb{R}^n$. Let $\mathbf{x}^* \in C$. Then the following statements are equivalent*

i) $\mathbf{x}^*$ *is local minimum*
ii) $\mathbf{x}^*$ *is global minimum*
iii) $\nabla f(\mathbf{x}^*) = \mathbf{0}$*, i.e all partial derivatives at* $\mathbf{x}^*$ *are zero.*

In Corollary 2.1.10, since C is an open set, the optimal solutions, if they exist, can not lie on the boundary of C. This is simply because these points lie outside C. For the purpose of extending Corollary 2.1.10 to the case where C is a closed set, which is more typical, consider a convex optimization problem where $C \subseteq \mathbb{R}^n$ is closed and convex.

$$\max_{\mathbf{x}}\{f(\mathbf{x}) : \mathbf{x} \in C,\ C \text{ closed and convex}\} \tag{2.1}$$

Lemma 2.1.11. *Let $f\colon C \to \mathbb{R}$ be a convex function defined on a closed convex set $C \subseteq \mathbb{R}^n$. Then in problem (2.1), each local minimum is also global. Moreover, the set of minima in problem (2.1) is a closed convex subset of C.*

In the case where C is closed, optimal solutions may lie on the boundary of C.

2.2 Linear programming

The purpose of this section is to give a basic introduction to linear programming, which will be applied when analysing an one period optimization model later. This section is based on theory from Vanderbei [Van14].

A linear programming (LP) problem is an optimization problem of maximizing or minimizing a linear function subject to linear constraints. The variables whose values are to be decided in some optimal fashion, are called *decision variables*. They are usually denoted x_j, $j = 1, 2, \ldots, n$. The function to be maximized or minimized is a linear function of the decision variables, and is called the *objective function*. We can write the objective function as $\zeta = \mathbf{c}^T\mathbf{x}$, where $\mathbf{x} = (x_1, x_2, \ldots, x_n)^T$ and $\mathbf{c} = (c_1.c_2, \ldots, c_n)^T$. The *constraints* of the problem can either be linear equalities or linear inequalities:

$$\mathbf{a}^T\mathbf{x} \left\{ \begin{array}{c} \geq \\ = \\ \leq \end{array} \right\} \mathbf{b}$$

where $\mathbf{a}, \mathbf{x} \in \mathbb{R}^n$ and $\mathbf{b} \in \mathbb{R}^m$ are column vectors. It is easy to convert constraints from one form to another. We will prefer 'less-than' inequalities[1]. The *standard form* of a linear program can be formulated as follows on matrix form:

$$\begin{array}{lrcl} \text{maximize} & \mathbf{c}^T\mathbf{x} & & \\ \text{subject to} & A\mathbf{x} & \leq & \mathbf{b} \\ & \mathbf{x} & \geq & \mathbf{0} \end{array} \tag{2.2}$$

where $A \in \mathbb{R}^{m\times n}, \mathbf{b} \in \mathbb{R}^m$ and $\mathbf{c}, \mathbf{x} \in \mathbb{R}^n$. m is the number of constraints, n is the number of decision variables.

A solution $\mathbf{x}^*$ to problem (2.2) is called *feasible* if it satisfies all the constraints. Recall that a polyhedron is a convex set of form $\{\mathbf{x} \in \mathbb{R}^n \colon A\mathbf{x} \leq \mathbf{b}\}$, where $A \in \mathbb{R}^{m\times n}$ and $\mathbf{b} \in \mathbb{R}^m$. So, the set of all feasible solutions in LP

[1] Less-than inequalities are preferred from a mathematical point of view, see Vanderbei [Van14]

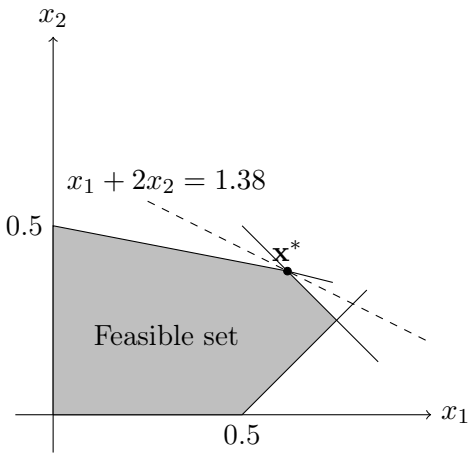


Figure 2.4: Geometrically solution to the LP problem in Example 2.2.1.

problems is a polyhedron. If the problem (2.2) has no feasible solutions, then the problem itself is called *infeasible*. A feasible solution, that in addition attains the desired maximum, is called an *optimal solution*. A problem is *unbounded* if there exists a feasible solution with arbitrary large objective value.

Example 2.2.1. Assume we have following LP problem in $\mathbb{R}^2$.

$$\begin{array}{lrcrl} \text{maximize} & x_1 & + & 2x_2 & \\ \text{subject to} & 2x_1 & + & 10x_2 & \leq 5 \\ & x_1 & + & x_2 & \leq 1 \\ & 2x_1 & - & 2x_2 & \leq 1 \\ & x_1 & , & x_2 & \geq 0 \end{array}$$

This problem can be solved geometrically in the plane. From Figure 2.4 we can see that $\mathbf{x} = (x_1, x_2)$ has feasible solutions. The feasible set is a polyhedron which is closed, so the optimal solution lies on this polyhedron's boundary. In fact, Figure 2.4 also tells us that the optimal solution is $\mathbf{x}^* = (0.62, 0.38)$, and that the optimal value is 1.38.

□

Example 2.2.1 shows geometrically the concept of feasible sets. In addition, it shows that we can find the optimal solution geometrically by finding the feasible point (or points) where the objective function leaves the feasible set. Although this is a simple method for finding optimal solutions in

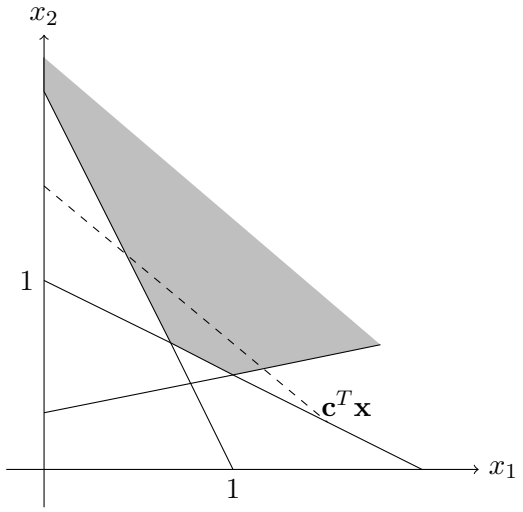


Figure 2.5: Example of a unbounded LP problem.

$\mathbb{R}^2$, when dealing with higher dimensions, this method is not good at all. The most commonly used method for solving LP problems is the *Simplex method*. It can be shown that the Simplex method always will terminate, i.e find the optimal solution or prove that no such solution exists, if the leaving variable is selected by the lexicographic rule or Bland's rule. For details on the Simplex method see Vanderbei [Van14].

Theorem 2.2.1 (Fundamental theorem of linear programming)**.** *For an arbitrary linear program in standard form, the following statements are true:*
i) If there is no optimal solution, then the problem is either infeasible or unbounded.
ii) If a feasible solution exists, then a basic feasible solution exists.
iii) If an optimal solution exists, then a basic optimal solution exists.

The first property of the fundamental theorem of LP 2.2.1 states that if a problem is neither infeasible nor unbounded, then there exists an optimal solution. This means that a LP problem is either unbounded, infeasible or it has an optimal solution. The second and third property of the fundamental theorem on LP contains the concepts of *basic feasible solutions* and *basic optimal solutions*, for definitions see Vanderbei [Van14, p. 13].

The following result, Farkas' Lemma, is important in LP, as it gives a necessary and sufficient condition for a system of linear inequalities to have

solutions.

Lemma 2.2.2 (Farkas' Lemma)**.** *The system* $A\mathbf{x} \leq \mathbf{b}$ *has no solutions if and only if there exists a* $\mathbf{y}$ *such that*

$$\begin{aligned} A^T\mathbf{y} &= \mathbf{0} \\ \mathbf{y} &\geq \mathbf{0} \\ \mathbf{b}^T\mathbf{y} &< 0 \end{aligned}$$

2.3 Quadratic programming

We will later in this thesis look at the Markowitz mean-variance portfolio optimization problem. This optimization problem has a quadratic objective function and linear constraints. For this purpose, we will consider quadratic optimization. This section is based on the book by Best [Bes10].

A general quadratic function $Q\colon \mathbb{R}^n \to \mathbb{R}$ can be written as

$$Q(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T C\mathbf{x} + \mathbf{c}^T\mathbf{x} \tag{2.3}$$

where $\mathbf{c} = (c_1, c_2, \ldots, c_n)^T$, $\mathbf{x} = (x_1, x_2, \ldots, x_n)^T$ and C is a $n \times n$ symmetric matrix. We will also assume that C is positive semidefinite, i.e that $\mathbf{s}^T C\mathbf{s} \geq 0$ for all $\mathbf{s}$.

For minimizing the quadratic function Q, iterative methods like the steepest descent and conjugate gradient method may be used. These methods are related to each other. For more on these, see Lyche [Lyc15].

A typical quadratic optimization problem will consist of a quadratic objective function and in addition some linear constraints. I will start by looking at a geometric example in the plane to illustrate the idea.

Example 2.3.1. Assume we have following optimization problem

$$\min\Big\{\frac{1}{2}\mathbf{x}^T C\mathbf{x} + \mathbf{c}^T\mathbf{x} : \mathbf{a}^T\mathbf{x} = b\Big\}$$

where $\mathbf{c} = (c_1, c_2)^T$, C is a 2×2 symmetric, semidefinite matrix, $\mathbf{a} = (a_1, a_2)^T$ and b is a scalar. The constraint function is represented as a line in the plane, and the objective function as a paraboloid, with ellipses as level curves. Then the optimal solution can be found geometrically where the objective

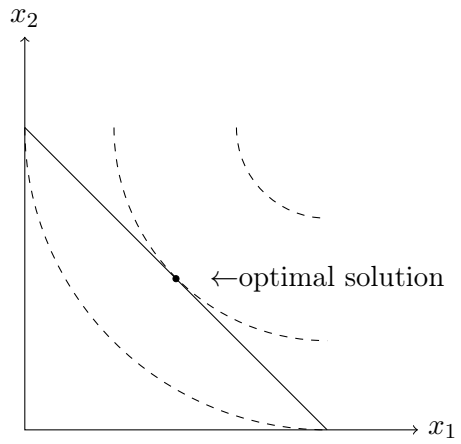


Figure 2.6: Geometrically solution to Example 2.3.1.

function and the constraint function intersect in one point. This means that the gradient of the objective function, at this optimal point $\mathbf{x}^*$, is a multiple of the linear constraint. Let $\mathbf{d}$ be the gradient of the constraint function, then

$$\nabla Q(\mathbf{x}^*) = u\mathbf{d}$$

for some scalar u.

□

For the purpose of generalizing this optimality condition to problems having n variables and m constraints, consider the problem

$$\min\Big\{\frac{1}{2}\mathbf{x}^T C\mathbf{x} + \mathbf{c}^T\mathbf{x} : A\mathbf{x} = \mathbf{b}\Big\} \tag{2.4}$$

where $\mathbf{c} \in \mathbb{R}^n$, $C \in \mathbb{R}^{n\times n}$ is symmetric and positive semidefinite, $A \in \mathbb{R}^{m\times n}$, $\mathbf{b} \in \mathbb{R}^m$. Let $A^T = (\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m)$, where $\mathbf{a}_i$ is the i'th row vector of A, and let $\mathbf{b} = (b_1, b_2, \ldots, b_m)^T$. We can write the constraints on the form

$$\mathbf{a}_i^T\mathbf{x} = b_i, \text{ for } i = 1, 2, \ldots, m$$

where $\mathbf{a}_i$ is the gradient of the i'th constraint. Then for a vector $\mathbf{u} = (u_1, u_2, \ldots, u_m)^T$ we should have following optimality condition

$$-\nabla Q(\mathbf{x}) = A^T\mathbf{u}$$

Definition 2.3.1. *The* optimality conditions *for problem (2.4) are* $A\mathbf{x}^* = \mathbf{b}$ *and that there exists a vector* $\mathbf{u}$ *such that* $-\nabla Q(\mathbf{x}^*) = A^T\mathbf{u}$

The vector $\mathbf{u}$ is called the *multiplier vector* for the problem. There is one component u_i for each constraint, and this u_i is called the *multiplier* associated with constraint i. The optimality condition $A\mathbf{x}^* = \mathbf{b}$ is often called *primal feasibility*, and the second condition $-\nabla Q(\mathbf{x}^*) = A^T\mathbf{u}$ is often called *dual feasibility*.

Theorem 2.3.2. $\mathbf{x}^*$ *is optimal for problem (2.4) if and only if* $\mathbf{x}^*$ *satisfies the optimality conditions in Definition 2.3.1.*

Proof. Assume that $\mathbf{x}^*$ satisfies the optimality conditions in Definition 2.3.1. To show that $\mathbf{x}^*$ is optimal, we must show that $Q(\mathbf{x}^*) \leq Q(\mathbf{x})$ for all $\mathbf{x}$ satisfying $A\mathbf{x} = \mathbf{b}$. Let $\mathbf{x}$ be such that $A\mathbf{x} = \mathbf{b}$. Then by Taylor' theorem we have

$$Q(\mathbf{x}) = Q(\mathbf{x}^*) + \nabla Q(\mathbf{x}^*)^T(\mathbf{x} - \mathbf{x}^*) + \frac{1}{2}(\mathbf{x} - \mathbf{x}^*)^T C(\mathbf{x} - \mathbf{x}^*)$$

From our assumption, $\nabla Q(\mathbf{x}^*) = -A^T\mathbf{u}$, so

$$\nabla Q(\mathbf{x}^*)^T(\mathbf{x} - \mathbf{x}^*) = -\mathbf{u}^T A(\mathbf{x} - \mathbf{x}^*) = -\mathbf{u}^T(\mathbf{b} - \mathbf{b}) = 0$$

Since C is positive semidefinite, we have that

$$(\mathbf{x} - \mathbf{x}^*)^T C(\mathbf{x} - \mathbf{x}^*) \geq 0$$

This gives us that $Q(\mathbf{x}^*) \leq Q(\mathbf{x})$ for all $\mathbf{x}$.

To show that $\mathbf{x}^*$ satisfies the optimal conditions, given that it is optimal, is a bit more complicated. This part of the proof will be omitted. ■

Theorem 2.3.3. $\mathbf{x}^*$ *is optimal for problem (2.4) if and only if there exists a column vector* $\mathbf{u} \in \mathbb{R}^m$ *such that* $(\mathbf{x}^*, \mathbf{u})^T$ *satisfies the linear equations*

$$\begin{bmatrix} C & A^T \\ A & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{x}^* \\ \mathbf{u} \end{bmatrix} = \begin{bmatrix} -\mathbf{c} \\ \mathbf{b} \end{bmatrix} \tag{2.5}$$

Proof. Performing the multiplication results in the equations

$$C\mathbf{x}^* + A^T\mathbf{u} = -\mathbf{c} \text{ and}$$

$$A\mathbf{x}^* = \mathbf{b}$$

In the first equation we recognize $C\mathbf{x}^* + \mathbf{c}$ to be the the gradient at $\mathbf{x}^*$, so $\nabla Q(\mathbf{x}^*) = -A^T\mathbf{u}$. These two equations are just the optimality conditions for problem (2.4), and the result follows from Theorem 2.3.2. ■

2.4 Fourier-Motzkin elimination

This section is based on the reports by Dahl [Dah14] and by Dantzig [Dan72]. The purpose of this section is to give an introduction to the *Fourier-Motzkin elimination* (FM elimination). This algorithm is used to eliminate variables from a linear system of inequalities, and at the same time find the projection of a polyhedron into a subspace. FM elimination may be used to solve LP problems, as the set of constraints can be written as a system of linear inequalities. We will in this thesis use it to analyse the portfolio optimization problem given Conditional Value-at-Risk constraints.

Assume that we have system $A\mathbf{x} \leq \mathbf{b}$, where $A = [a_{ij}] \in \mathbb{R}^{m\times n}$ and $\mathbf{b} \in \mathbb{R}^m$. Assume $\mathbf{x} = (x_1, x_2, \ldots, x_n)^T$ are variables that we want to eliminate in the given order, although any order will do. We start by eliminating variable x_1. First we divide the system into three subsystems:

$$\begin{array}{rcrcccrcll} a_{i1}x_1 & + & a_{i2}x_2 & + & \cdots & + & a_{in}x_n & \leq & b_i & \text{for } i \in I^+ \\ 0 \cdot x_1 & + & a_{i2}x_2 & + & \cdots & + & a_{in}x_n & \leq & b_i & \text{for } i \in I^0 \\ a_{i1}x_1 & + & a_{i2}x_2 & + & \cdots & + & a_{in}x_n & \leq & b_i & \text{for } i \in I^- \end{array}$$

where $I^+ = \{i : a_{i1} > 0\}$, $I^- = \{i : a_{i1} < 0\}$ and $I^0 = \{i : a_{i1} = 0\}$. The sets I^+, I^- and I^0 form a partition of the row index set $I = \{1, 2, \ldots, m\}$. We leave all inequalities with $i \in I^0$ as they are, since these don't give upper or lower bounds on x_1. For all $i \in I^+ \cup I^-$, we divide $a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n \leq b_i$ by $|a_{i1}|$ and get following system:

$$\begin{array}{rcrcccrcll} x_1 & + & a'_{i2}x_2 & + & \cdots & + & a'_{in}x_n & \leq & b'_i & \text{for } i \in I^+ \\ & + & a_{i2}x_2 & + & \cdots & + & a_{in}x_n & \leq & b_i & \text{for } i \in I^0 \\ - \; x_1 & + & a'_{i2}x_2 & + & \cdots & + & a'_{in}x_n & \leq & b'_i & \text{for } i \in I^- \end{array}$$

where $a'_{ij} = a_{ij}/|a_{i1}|$ and $b'_i = b_i/|a_{i1}|$. Combining inequalities with indexes from I^+ with inequalities with indexes from I^-, results in an equivalent system of inequalities.

$$\sum_{j=2}^{m} a'_{ij}x_j - b'_i \leq x_1 \leq b'_k - \sum_{j=2}^{m} a'_{kj}x_j \quad \text{for } i \in I^-, k \in I^+ \tag{2.6}$$

$$\sum_{j=2}^{m} a_{ij}x_j \leq b_i \quad \text{for } i \in I^0 \tag{2.7}$$

$\mathbf{x}$ is a solution to the original problem $A\mathbf{x} \leq \mathbf{b}$ if and only if $x_2, x_3, \ldots, x_n$ satisfy

$$\sum_{j=2}^{m} a'_{ij}x_j - b'_i \leq b'_k - \sum_{j=2}^{m} a'_{kj}x_j \quad \text{for } i \in I^-, k \in I^+ \tag{2.8}$$

$$\sum_{j=2}^{m} a_{ij}x_j \leq b_i \quad \text{for } i \in I^0 \tag{2.9}$$

and x_1 satisfies

$$\max_{i \in I^-}\left\{ \sum_{j=2}^{m} a'_{ij}x_j - b'_i \right\} \leq x_1 \leq \min_{k \in I^+}\left\{ b'_k - \sum_{j=2}^{m} a'_{kj}x_j \right\} \tag{2.10}$$

If either I^+ or I^- is empty, then the set of inequalities in (2.8) vanishes, and upper bounds or lower bounds in (2.10) should be set to ∞ or $-\infty$, respectively. If I^0 is empty and either I^+ or I^- is empty, we terminate. In this case, the general problem is obtained by choosing $x_2, x_3, \ldots, x_n$ arbitrarily, and then choosing x_1 according to (2.10).

Assume we now have eliminated all variables $x_1, x_2, \ldots, x_k$ (in that order) which may be all or just some of the variables. Then we have a system $l_i(x_{i+1}, \ldots, x_n) \leq x_i \leq u_i(x_{i+1}, \ldots, x_n)$, $i = 1, 2, \ldots, k$. Here $u_i(x_{i+1}, \ldots, x_n)$ and $l_i(x_{i+1}, \ldots, x_n)$ are upper and lower bounds, respectively, dependent on variables that have been eliminated after x_i and variables that haven't been eliminated (if not all have been eliminated). We may choose $x_k \in [l_k, u_k]$. Once x_k is chosen, we choose $x_{k-1} \in [l_{k-1}(x_k, \ldots, x_n), u_{k-1}(x_k, \ldots, x_n)]$. Continuing with this back-substitution produces a solution $\mathbf{x}$. In fact, all solutions of $A\mathbf{x} \leq \mathbf{x}$ may be produced this way.

Performing FM elimination on a variable can make the number of inequalities grow. Assume that I^+ has cardinality p and I^- has cardinality q. Also let r be the cardinality of I^0. Then, after eliminating the variable, the number of inequalities equals $pq + r$. If the number of inequalities in the original problem is m, the worst case occurs when $p = q, r = 0$. In this case, the number of new constraints will be $\frac{m}{2} \cdot \frac{m}{2} = \frac{m^2}{4}$.

I will now give an example to illustrate the FM elimination.

Example 2.4.1. Consider the problem in Example 2.2.1, and assume I want to eliminate x_2. Now $I^+ = \{1, 2\}, I^- = \{3, 5\}$ and $I^0 = \{4\}$ (inequality numbers 4 and 5 correspond to $x_1 \geq 0$ and $x_2 \geq 0$, respectively).

$$\begin{array}{rcrcr} 10x_2 & \leq & 5 & - & 2x_1 \\ x_2 & \leq & 1 & - & x_1 \\ 2x_2 & \geq & 2x_1 & - & 1 \\ x_2 & \geq & 0 & & \end{array}$$

Dividing by $|a_{ij}|$ gives following system.

$$\begin{array}{rcrcr} x_2 & \leq & \frac{1}{2} & - & \frac{1}{5}x_1 \\ x_2 & \leq & 1 & - & x_1 \\ x_2 & \geq & x_1 & - & \frac{1}{2} \\ x_2 & \geq & 0 & & \end{array}$$

Combining upper bounds with lower bounds will result in $pq + r = 4 + 1 = 5$ inequalities. After eliminating x_2, and some rewriting, we am left with this new system

$$\begin{array}{rcl} x_1 & \leq & \frac{5}{2} \\ x_1 & \leq & 1 \\ x_1 & \leq & \frac{5}{6} \\ x_1 & \leq & \frac{3}{4} \\ x_1 & \geq & 0 \end{array}$$

$$\max\left\{0, x_1 - \frac{1}{2}\right\} \leq x_2 \leq \min\left\{1 - x_1, \frac{1}{2} - \frac{1}{5}x_1\right\}$$

This means that $0 \leq x_1 \leq \frac{3}{4}$. Looking at Figure 2.7, we see that this interval corresponds to the feasible x_1-values. Let P be the polyhedron equal to the feasible set in Figure 2.7, then the interval $\left[0, \frac{3}{4}\right]$ is the projection of P along x_2-axis into the x_1-axis. This projection is illustrated in Figure 2.7 by a thick line along the x_1-axis.

□

Since the constraints of a LP problem can be written as a system of linear inequalities, it is sometimes useful to eliminate decision variables using

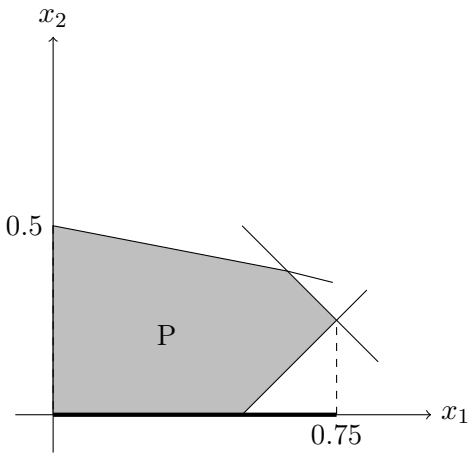


Figure 2.7: Projection of the feasible set P in Example 2.4.1, into the x_1-axis.

FM elimination. Since performing a FM elimination produces an equivalent system of linear inequalities to the original system, the new system will have a solution if and only if the original system has one. A LP problem is infeasible if we, by performing FM elimination, get something like $1 \leq 0$. One can also use Farkas' Lemma 2.2.2 to decide whether the linear system has solutions, i.e whether the LP problem is feasible. Below is an example of an infeasible LP problem. The example will illustrate that when we perform FM elimination on the constraints of this problem, we get an illogical inequality, in this case $4 \leq 2$.

Example 2.4.2. Assume we have following LP problem:

$$\begin{array}{lrcrcr} \max & 5x_1 & + & 4x_2 & & \\ \text{subject to} & x_1 & + & x_2 & \leq & 2 \\ & -2x_1 & - & 2x_2 & \leq & -8 \\ & x_1 & , & x_2 & \geq & 0 \end{array}$$

we will show that this problem is infeasible, using FM elimination. First I eliminate x_1.

$$\begin{array}{lcl} x_1 & \leq & 2 - x_2 \\ x_1 & \geq & 4 - x_2 \\ x_1 & \geq & 0 \end{array}$$

kAfter eliminating x_1 we have

$$\begin{aligned} 0 &\leq 2 - x_2 \iff x_2 \leq 2 \\ 4 - x_2 &\leq 2 - x_2 \iff 4 \leq 2 \end{aligned}$$

Since we get that $4 \leq 2$, the problem is infeasible.

□

3 | Mathematical Finance and Risk measures

3.1 Mathematical Finance

In this section I will discuss the financial framework for this thesis, using mathematical finance to describe the concepts. This chapter is based on the books by Øksendal [Øk13], McDonald and Weiss [MW13], and Çınlar [Çın11].

Definition 3.1.1. *A* σ-algebra $\mathcal{F}$ *on a set* Ω *is a collection of subsets of* Ω *such that*

i) $\emptyset \in \mathcal{F}$

ii) *If* $F \in \mathcal{F}$*, then also* $F^c \in \mathcal{F}$*. Here* $F^c = \Omega \backslash F$*.*

iii) *If* $F_1, F_2, \ldots \in \mathcal{F}$*, then* $\bigcup_{i=1}^{\infty} F_i \in \mathcal{F}$

The set Ω is in a statistical setting thought of as the sample space, the space of possible outcomes of a random experiment. In fact, Ω will be permitted to be any set containing all of the possible outcomes. This is because it may be difficult to know precisely the possible outcomes of an experiment. In finance, Ω is often the set of possible prices in a market. $\mathcal{F}$ is often called the set of events. The subsets F of Ω which belongs to $\mathcal{F}$ are called $\mathcal{F}$-*measurable* sets. A pair $(\Omega, \mathcal{F})$ is called a *measurable space*.

Definition 3.1.2. *A* measure μ *on a measurable space* $(\Omega, \mathcal{F})$ *is a function* $\mu \colon \mathcal{F} \to \mathbb{R}$ *such that*

i) $\mu(F) \geq 0$ *for all* $F \in \mathcal{F}$

ii) $\mu(\emptyset) = 0$

iii) *If* $F_1, F_2, \ldots \in \mathcal{F}$ *and* $\{F_i\}$ *is pairwise disjoint (i.e* $F_i \cap F_j = \emptyset$ *if* $i \neq j$*), then*

$$\mu\Big(\bigcup_{i=1}^{\infty} F_i\Big) = \sum_{i=1}^{\infty} \mu(F_i)$$

A measure P, which in addition satisfies the condition $P(\Omega) = 1$, is called a *probability measure*. In this case, for every $F \in \mathcal{F}$, $P(F)$ is the probability that event F will occur. Since Ω contains all possible outcomes, the probability that event Ω will occur must be equal to 1. Also, the probability of the impossible event $\emptyset$ is $P(\emptyset) = 0$. We will from now on let P be a probability measure. A triple $(\Omega, \mathcal{F}, P)$ is called a *probability space*, and is called a *complete* probability space if it is such that $\mathcal{F}$ contains all subsets G of Ω with P-outer measure zero, i.e with

$$P^*(G) = \inf\{P(F) : F \in \mathcal{F}, G \subset F\} = 0$$

Fore more on P-outer measures, see McDonald and Weiss [MW13]. We will from now on assume that all our probability spaces are complete.

The following example illustrates some of the concepts presented in this section.

Example 3.1.1. Let H denote head, and T denote tail when flipping a fair coin. When flipping a coin twice, the sample space of the experiment is $\Omega = \{HH, HT, TH, TT\}$. Let $\mathcal{F} = \mathcal{P}(\Omega)$, so that every subset of Ω is an event. Let $F \in \mathcal{F}$ be the event that the first toss is tail, i.e. that $F = \{TH, TT\}$. Then the probability that event F will occur is $P(F) = \frac{2}{4} = \frac{1}{2}$.

□

Definition 3.1.3. *Let $(\Omega, \mathcal{F}, P)$ be a probability space. A* random variable *is a real-valued function X on Ω such that $\{\omega : X(\omega) \in B\} \in \mathcal{F}$ for each $B \in \mathcal{B}$ (where $\mathcal{B}$ is the Borel σ-algebra, see [Øk13]).*

A *stochastic process* is a parametrized collection of random variables $\{X_t\}_{t \in T}$, defined on a probability space $(\Omega, \mathcal{F}, P)$. Here, and throughout this section, T is some index set, typically $[0, \infty), [0, S]$ or $\mathbb{N}$. t usually denotes the time. For a fixed $t \in T$, $\omega \to X_t(\omega)$ is a random variable. For a fixed $\omega \in \Omega$, $t \to X_t(\omega)$ is called a *path*.

Example 3.1.2 (Example 3.1.1 continued)**.** Let $(\Omega, \mathcal{F}, P)$ be the probability space described in Example 3.1.1. Let $X \colon \Omega \to \mathbb{R}$ be a random variable on Ω, such that $X(\omega)$ is the number of heads after two flips. Then

$$\begin{array}{lcll} P(X = 0) & = & P(\{TT\}) & = & 1/4 \\ P(X = 1) & = & P(\{HT, TH\}) & = & 1/2 \\ P(X = 2) & = & P(\{HH\}) & = & 1/4 \end{array}$$

□

Definition 3.1.4. *A* filtration *(on $(\Omega, \mathcal{F})$) is a family $\{\mathcal{F}_t\}_{t\in T}$ of σ-algebras such that $\mathcal{F}_t \subseteq \mathcal{F}$ for all $t \in T$, and for $s \leq t$, $\mathcal{F}_s \subseteq \mathcal{F}_t$.*

In a financial setting, one should interpret the σ- algebra $\mathcal{F}_t$ as the information available to investors at time t, in the sense that if ω is the true state of the world, and if $F \in \mathcal{F}_t$, then at time t the investors know whether $\omega \in F$. Then, at time t the filtration $\{\mathcal{F}_t\}_{t\in T}$ holds all the information about the market, up to time t.

Definition 3.1.5. *Let $\{\mathcal{F}_t\}_{t\in T}$ be an increasing family of σ-algebras of subsets of Ω. A stochastic process $\{X_t\}_{t\in T}$, such that for all $t \in T$, X_t is $\mathcal{F}_t$-measurable, is called* $\mathcal{F}_t$-adapted.

Every random variable induces a probability measure μ_X on $\mathbb{R}^n$, called the *distribution* of X. This is defined by

$$\mu_X(B) = P(X^{-1}(B))$$

Definition 3.1.6. *A random variable X is said to be an* absolutely continuous random variable *if there is a nonnegative Borel measurable function p such that $\mu_X(B) = \int_B p\, d\lambda$ for all Borel sets B. For such function, we usually write $p = p_X$ and call p_X the* density function *of X.*

In Definition 3.1.6, the integral is with respect to the Lebesgue measure λ, see McDonald and Weiss [MW13]. For functions that are Riemann integrable on $[a, b]$, the Lebesgue integral and the Riemann integral coincide on this interval. The probability that a random variable X takes on a value in the interval $[a, b]$ is the area under the density function and above this interval. For $p_X(x)$ to be a density function, it must satisfy the following conditions:

i) $p_X(x) \geq 0$ for all x

ii) $\int_{\mathbb{R}} p_X(x) d\lambda(x) = 1$

If $\int_\Omega |X(\omega)| dP(\omega) < \infty$, then the number

$$E[X] = \int_\Omega X(\omega) dP(\omega) = \int_{\mathbb{R}^n} x d\mu_X(x)$$

is called the *expectation*, and sometimes the *mean*, of X with respect to P. The expectation is the Lebesgue integral on space the $(\Omega, \mathcal{F}, P)$, see McDonald and Weiss [MW13]. So, the following properties are just properties of this integral.

Proposition 3.1.7 (Properties of the expectation)**.** *Let $(\Omega, \mathcal{F}, P)$ be a probability space, let X and Y be random variables and let $a, b \in \mathbb{R}$ be constants. Then the following properties hold for the expectation:*

i) *(*Positivity*)* $X \geq 0 \implies E[X] \geq 0$ *with equality if and only if* $X = 0$
ii) *(*Monotonicity*)* $X \geq Y \geq 0 \implies E[X] \geq E[Y]$
iii) *(*Linearity*)* $X, Y \geq 0 \implies E[aX + bY] = aE[X] + bE[Y]$
iv) *(*Insensitivity*)* $X = Y$ *almost surely* $\implies E[X] = E[Y]$

The monotonicity property in Proposition 3.1.7 can be extended, see [Çın11]. If $X \geq Y$, and if both $E[X]$ and $E[Y]$ exists, then $E[X] \geq E[Y]$. By $X = Y$ *almost surely* (*a.s.*) we mean that $X = Y$ except on a set of P- measure zero, that is, except on a set N with $P(N) = 0$.

Definition 3.1.8. *Let $(\Omega, \mathcal{F}, P)$ be a probability space, let X be a random variable such that $E[|X|] < \infty$, and let $\mathcal{H}$ be a sub-σ-algebra of $\mathcal{F}$. The* conditional expectation of X given $\mathcal{H}$*, denoted $E[X|\mathcal{H}]$, is the almost surely unique function from Ω to $\mathbb{R}^n$ satisfying:*

i) $E[X|\mathcal{H}]$ *is $\mathcal{H}$-measurable*
ii) $\int_H E[X|\mathcal{H}]dP = \int_H X dP$ *for all* $H \in \mathcal{H}$

The following proposition states some properties of the conditional expectation. Some properties are taken from Øksendal [Øk13] and some are taken from Çınlar [Çın11].

Proposition 3.1.9 (Properties of the conditional expectation)**.** *Let $(\Omega, \mathcal{F}, P)$ be a probability space, let $\mathcal{H}$ be a sub-σ-algebra of $\mathcal{F}$ and let $a, b, c \in \mathbb{R}$ be constants. The following properties hold for the conditional expectation function:*

i) *(*Monotonicity*)* $X \geq Y \implies E[X|\mathcal{H}] \geq E[Y|\mathcal{H}]$
if X, Y are $\mathcal{F}$-measurable
ii) *(*Linearity*)* $E[aX + bY + c|\mathcal{H}] = aE[X|\mathcal{H}] + bE[Y|\mathcal{H}] + c$
if X, Y are $\mathcal{F}$-measurable
iii) $E[E[X|\mathcal{H}]] = E[X]$ *if X is $\mathcal{F}$-measurable*
iv) $E[X|\mathcal{H}] = X$ *if X is $\mathcal{H}$-measurable*
v) $E[X|\mathcal{H}] = E[X]$ *if X is independent of $\mathcal{H}$*
vi) $E[Y \cdot X|\mathcal{H}] = Y \cdot E[X|\mathcal{H}]$ *if Y is $\mathcal{H}$-measurable, where $\cdot$ denotes the usual inner product.*

provided these conditional expectations exist.

Let H be an event. Then

$$E[X|H] = \frac{1}{P(H)} E[X \mathbf{1}_H]$$

where $\mathbf{1}_H$ denotes the indicator function on H. If $P(H) = 0$, then so is the integral over H, and $E[X|H] = 0$.

Definition 3.1.10. *Let X be a random variable of finite expectation. Then the* variance *of X, denoted* $\mathrm{Var}(X)$*, is defined by*

$$\mathrm{VaR}(X) = E[(X - E[X])^2]$$

The square root of the variance of X is called the standard deviation of X.

The variance of a random variable describes the variability in the distribution of X.

Definition 3.1.11. *Suppose X, Y are random variables having finite variances and are defined on the same probability space. Then the* covariance *of X and Y, denoted by* $\mathrm{Cov}(X, Y)$*, is defined by*

$$\mathrm{Cov}(X, Y) = E\Big[(X - E[X])(Y - E[Y])\Big]$$

The covariance of two random variables describes how strongly they are related. If the covariance is zero, then they are independent of each other.

The market model

The Brownian motion is a very important example of a stochastic process, and has a number of real-world applications like for instance to mathematical finance. In particular, the Brownian motion is widely applied in modelling the financial market.

Definition 3.1.12. *Let $(\Omega, \mathcal{F}, P)$ be a probability space with filtration $\{\mathcal{F}_t\}_{t\in T}$, and let $s \leq t$. A $\mathcal{F}_t$-adapted stochastic process $\{B_t\}_{t\in T}$ is called a* Brownian motion *starting in $x \in \mathbb{R}^d$ if*

i) $P(B_0 = x) = 1$

ii) $B_t - B_s$ *is independent of* $\mathcal{F}_s$

iii) $B_t - B_s$ *is normally distributed with mean x and variance $\sigma = (t - s)I$, where I is the $d \times d$ identity matrix*

We often let $x = 0$ in Definition 3.1.12, so that the Brownian motion starts at 0. The second condition in Definition 3.1.12 states that the Brownian motion has independent increments. The third condition states that the increments are normally distributed. This means that $E[B_t(\omega) - B_s(\omega)] = x$. Also, the variance of the increments $B_t(\omega) - B_s(\omega)$ equals $(t - s)$, where $t, s \in T$. The variance is proportional to the time. We will throughout this section assume

that the family of σ-algebras $\{\mathcal{F}_t\}_{t\in T}$, is generated by the Brownian motion. This means that $\mathcal{F}_t$ is the smallest σ-algebra containing all sets of the form $\{\omega\colon B_{t_1} \in F_1, \ldots, B_{t_k} \in F_k\}$, where $t_j \leq t$, and F_j, $j \leq k$, are Borel sets.

It can be shown that the Brownian motion has a continuous modification, i.e for all $T > 0$ there exists positive constants α, β, D such that $E[|B_t - B_s|^\alpha] \leq D|t-s|^{1+\beta}$ for $0 \leq s, t \leq T$. If $\{X_t\}_{t\in T}$ is a modification of $\{Y_t\}_{t\in T}$, then they have the same finite-dimensional distributions. From now on we will assume that B_t is such a modification.

We will later define a market to be an Itô process. For this purpose we want to define the Itô integral:

$$\int_0^T f(s,\omega)dB_s(\omega)$$

for a given function $f(s,\omega)$. First we need to define the $\mathcal{L}^p$-spaces.

Definition 3.1.13. *Let X be a random variable on probability space $(\Omega, \mathcal{F}, P)$, and let $p \in [1,\infty]$. Then the* $\mathcal{L}^p$-spaces *are defined by*

$$\mathcal{L}^p(P) = \mathcal{L}^p(\Omega) = \{X : \Omega \to \mathbb{R}^n\colon ||X||_p < \infty\}$$

where the $\mathcal{L}^p$-norms $||X||_p$ *are defined by*

$$||X||_p = \Big(\int_\Omega |X(\omega)|^p dP(\omega)\Big)^{\frac{1}{p}}$$

for $p \in [1,\infty)$, and for $p = \infty$

$$||X||_\infty = \inf\{N \in \mathbb{R}\colon |X(\omega)| \leq N \text{ almos surely}\}$$

The construction of the Itô integral consists of first defining it for simple functions, i.e functions on the form $\psi(t,\omega) = \sum_i \xi_i(\omega)\mathbf{1}_{[t_i,t_{i+1})}(t)$ where ξ is $\mathcal{F}_{t_i}$-measurable and $E[\xi^2] < \infty$. $\mathbf{1}_{[t_i,t_{i+1})}$ is the indicator function on $[t_i, t_{i+1})$. Then one extends it to functions in $\mathcal{L}_a^2([0,T]\times\Omega)$, which is the subspace of $\mathcal{L}^2([0,T]\times\Omega)$ that consists of adapted processes. For simple functions $\psi(t,\omega)$, the Itô integral is defined to be

$$\int_0^T \psi(s,\omega)dB_t(\omega) = \sum_{i\geq 0} \xi_i(\omega)[B_{t_i+1} - B_{t_i}]$$

where

$$t_k = \begin{cases} k2^{-n} & \text{if} \quad S \leq k2^{-n} \leq T \\ S & \text{if} \quad k2^{-n} < S \\ T & \text{if} \quad k2^{-n} > T \end{cases}$$

Definition 3.1.14. *Let $f \in \mathcal{L}_a^2([0,T] \times \Omega)$. Then the* Itô integral of f *from 0 to T is defined by*

$$\int_0^T f(s,\omega)dB_s(\omega) = \lim_{n\to\infty} \int_0^T \psi_n(s,\omega)dB_s(\omega)$$

where the limit is in $\mathcal{L}^2(P)$, and $\{\psi_n\}$ is a sequence of simple functions such that

$$E\Big[\int_0^T (f(s,\omega) - \psi_n(s,\omega))^2 ds\Big] \to 0 \text{ as } n \to \infty$$

Assume we are playing a game with payoff X_s. The game is considered to be fair if the expected payoff at time $t \geq s$, X_t, given that we have all the information up to time s, is equal to X_s. This is the idea behind the definition of a martingale.

Definition 3.1.15. *A n-dimensional stochastic process $\{M_t\}_{t\geq 0}$ on $(\Omega, \mathcal{F}, P)$ is called a* martingale *with respect to filtration $\{\mathcal{M}_t\}_{t\geq 0}$ if*

i) M_t *is $\mathcal{M}_t$-measurable for all t*
ii) $E[|M_t|] < \infty$ *for all t*
iii) $E[M_t|\mathcal{M}_s] = M_s$ *for all $s \leq t$*

The Brownian motion $\{B_t\}_{t\geq 0}$ is a martingale with respect to the filtration $\{\mathcal{F}_t\}_{t\in T}$, see Øksendal [Øk13].

The Itô integral has a t-continuous modification, i.e there exists a t-continuous stochastic process J_t on $(\Omega, \mathcal{F}, P)$ such that $P[J_t = \int_0^t f(s,\omega)dBs] = 1$. We will from now on assume that the Itô integral is such a modification. Also, it can be shown that the Itô integral $M_t(\omega) = \int_0^t f(s,\omega)dBs$ is a martingale with respect to $\mathcal{F}_t$.

Definition 3.1.16. *Let B_t be the 1-dimensional Brownian motion on probability space $(\Omega, \mathcal{F}, P)$. An 1-dimensional* Itô process *is a stochastic process $\{X_t\}_{t\geq 0}$ on $(\Omega, \mathcal{F}, P)$ of form*

$$X_t = X_0 + \int_0^t u(s,\omega)ds + \int_0^t v(s,\omega)dBs$$

where X_0 is $\mathcal{F}_0$-measurable, u and v are adapted processes, and we assume $E[\int_0^t |u(s)|ds] < \infty$ and $E[\int_0^t v^2(s)ds] < \infty$ for all t.

Sometimes we use the following notation for an Itô process:

$$dX_t = u(t)dt + v(t)dB_t$$

We are now ready to define a financial market. We will here define it to be the prices of the available assets. Then at each time t, each asset have a price $Y(t)$. Also, the future price of an asset is not known, so we are dealing with uncertainties. This implies that for each asset, its price process should be a stochastic process. The following definitions are taken from Øksendal [Øk13], but other mathematical models are also possible and actively investigated.

Definition 3.1.17. *A* market *is a $\mathcal{F}^{(m)}$-adapted $(n+1)$ dimensional Itô process $Y(t) = (Y_0(t), Y_1(t), \ldots, Y_n(t)), 0 \leq t \leq T$ which we assume has the form*

$$dY_0(t) = \rho(t,\omega)Y_0(t)dt \ ; \ Y_0(0) = 1$$

and

$$dY_i(t) = \mu_i(t,\omega)dt + \sum_{j=1}^{m} \sigma_{ij}(t,\omega)dB_j(t)$$

$$= \mu_i(t,\omega)dt + \sigma_i(t,\omega)dB(t) \ ; \ Y_i(0) = x_i$$

where σ_i is row number i of the $n \times m$ matrix $[\sigma_{ij}]$, $1 \leq i \leq n \in \mathbb{N}$

We think of the random variables $Y_i(t) = Y_i(t,\omega)$ as the price of asset number i at time t. In Definition 3.1.17 there are n *risky assets* and one *risk free asset*. The assets are called risky because of the presence of their diffusion term (their expressions include the Itô integral). This represent the randomness in the prices; we can't know what the prices will be in the future. For many applications, the risky assets are stocks. Asset number 0 is called risk free because of the absence of diffusion term. This term often represents the bank investment.

The market $\{Y(t)\}_{t\in[0,T]}$ is called *normalized* if $Y_0(t) = 1$. We can always make the market normalized by defining $\overline{Y}_0(t) = Y_0(t)^{-1}Y_i(t)$ for $1 \leq i \leq n$. This new market is called the *normalization* of $Y(t)$. The normalization corresponds to regarding the price $Y_0(t)$ of the safe investment as the unit of price, and computing the other prices in terms of this unit.

Definition 3.1.18. *A* portfolio *in the market $\{Y(t)\}_{t\in[0,T]}$ is a $(n+1)$-dimensional (t,ω)-measurable and $\mathcal{F}_t^{(m)}$-adapted stochastic process*

$$X(t,\omega) = (X_0(t,\omega), X_1(t,\omega), \ldots, X_n(t,\omega))$$

$0 \leq t \leq T$.

A portfolio, as it is defined in Definition 3.1.18, holds n different securities. The components $X_0(t,\omega), \ldots, X_n(t,\omega)$ represent the number of units of the securities number $0, \ldots, n$, respectively, which the investor holds at time t. A portfolio is an adapted stochastic process, so at time t, the investor knows how many units he or she has of each security. When an investor is choosing a portfolio for the future, the portfolio is often called a *trading strategy*. It says how the investor is investing his money, and the future value of this portfolio is clearly not known.

Definition 3.1.19. *The* value *at time t of a portfolio $X(t)$ is defined by*

$$V(t,\omega) = V^X(t,\omega) = X(t) \cdot Y(t) = \sum_{i=0}^{n} X_i(t) Y_i(t)$$

where $\cdot$ denotes the inner product in $\mathbb{R}^{n+1}$.

To calculate the value of a portfolio at time t, one simply multiplies the number of units in each security with its corresponding price at time t, and sum over these products. The value is the total value of the portfolio at time t.

Definition 3.1.20. *The portfolio $X(t)$ is called* self-financing *if*

$$\int_0^T \Big\{ |X_0(s)\rho(s)Y_0(s) + \sum_{i=1}^{n} X_i(s)\mu_i(s)| + \sum_{j=1}^{m} \Big[\sum_{i=1}^{n} X_i(s)\sigma_{ij}(s) \Big]^2 \Big\} ds < \infty \text{ a.s.} \tag{3.1}$$

and

$$V(t) = V(0) + \int_0^t X(s) dY(s) \text{ for } t \in [0,T] \tag{3.2}$$

Condition (3.1) in Definition 3.1.20 is required to make (3.2) well-defined, for more details see Øksendal [Øk13]. A portfolio is self-financing if no money is brought in or taken out from the system.

Definition 3.1.21. *A portfolio $X(t)$, which satisfies condition (3.1) in Definition 3.1.20 and which is self-financing, is called* admissible *if there exists $K = K(X) < \infty$ such that*

$$V^X(t,\omega) \geq -K \text{ for almost all } (t,\omega) \in [0,T] \times \Omega.$$

The concept of an admissible portfolio is that there must be a limit to how much debt an investor can tolerate.

Definition 3.1.22. *An admissible portfolio $X(t)$ is called an* arbitrage *in the market $\{Y(t)\}_{t\in[0,T]}$ if the corresponding value process $V^X(t)$ satisfies*

$$V^X(0) = 0 \text{ and}$$

$$V^X(T) \geq 0 \text{ a.s., and } P[V^X(T > 0)] > 0$$

In other words, a portfolio $X(t)$ is an arbitrage if it gives an increase in the value from time $t = 0$ to time $t = T$ a.s., and a strictly positive increase with positive probability. It means that we are guaranteed value increase. The existence of an arbitrage is a sign of lack of equilibrium in the market. No such market can exist in the long run, so it is therefore important to be able to determine whether a given market allows arbitrage or not.

3.2 Convex risk measures

The purpose of this section is to define and discuss what is good measures of risk. This section is based on the papers by Rockafellar [Roc07], Artzner, Delbaen, Eber and Heath [ADEH99], Krokhmal, Zabarankin and Uryasev [KZU11], Kaina and Rüschendorf [KR09]

When making decisions, and having to deal with future uncertainties, one commonly talk about risk. In financial optimization, measuring risk is very important. One is often interested in assigning a single value to a random variable, since a random variable itself is not a single quantity. Over the years, researchers have tried to find good measures of risk, some which have been flawed. One approach has been to compute the variance. The problem with this approach is that it does not distinguish between positive and negative deviations. Assume we want to compute the risk of a loss function, then negative deviations (corresponding to gain) should be welcomed, while positive deviations (corresponding to loss) should be disliked. Artzner et. al [ADEH99] gives some reasonable axioms that should be satisfied, and introduces the notion of coherent risk measure.

Let $V(X, \omega)$ be the loss of portfolio X if ω turns out to be the state of the world. A risk measure is defined on a class of random variables, in particular we can let this class be the class of losses $V(X, \omega)$ of attainable portfolios. Denote this set by $\mathcal{X}$, so that $V^X(t, \omega) \in \mathcal{X}$. For simplicity, we

will use notation $V(X, \omega)$, and sometimes just V. One should remember that we are considering the loss of a portfolio. Any risk measure has domain contained in $\mathcal{L}^0(\Omega, \mathcal{F}, P)$, the class of all measurable random variables on the probability space $(\Omega, \mathcal{F}, P)$. Sometimes we want more restrictions on the random variables, for instance that they should be bounded. Then we can let $\mathcal{X} = \mathcal{L}^\infty(\Omega, \mathcal{F}, P)$. In applications, risks are usually modelled by unbounded random variables, and for such applications, one may choose to use the space $\mathcal{L}^p(\Omega, \mathcal{F}, P),\ 1 \leq p < \infty$, as domain, see [KR09]. We will in this thesis let $p = 2$. This means that we require that $||V||_2 = (E[V^2])^{\frac{1}{2}} < \infty$.

The axioms in the definitions of convex risk measures and coherent risk measures are taken from Krokhmal et. al [KZU11] and Rockafellar [Roc07]. Since the risk measure Conditional Value-at-Risk measures the risk of loss, I have chosen to let the risk measures in this section also measure the risk of loss. This means that negative outcomes are welcomed, while positive outcomes are disliked.

Definition 3.2.1. *A* convex risk measure *is a mapping* $\rho\colon \mathcal{L}^2 \to \mathbb{R}$*, which for all* $V, U \in \mathcal{L}^2$ *satisfies the following properties:*

(R1) (Translation invariance) *For every constant function* $c \in \mathbb{R}$ *we have that* $\rho(V + c) = \rho(V) + c$

(R2) (Monotonicity) *If* $V \leq U$*, then* $\rho(V) \leq \rho(U)$

(R3) (Convexity) $\rho(\lambda V + (1 - \lambda)U) \leq \lambda\rho(V) + (1 - \lambda)\rho(U)$ *for all* $V, U \in \mathcal{L}^2$ *and* $\lambda \in [0, 1]$.

Property (R1) states that if the amount c of guaranteed loss is added to a position, then the risk increases with this amount c. Property (R2) states that if the loss V is less than loss U under every scenario, then the risk of V should be less than the risk of U. Property (R3) says that diversification[1] reduces risk.

A convex risk measure is called *coherent* if in addition it satisfies the positive homogeneity property, i.e that

$$\text{(R4) } \rho(\lambda V) = \lambda\rho(V) \text{ for all } \lambda \geq 0 \text{ and } V \in \mathcal{L}^2$$

In Artzner et. al. [ADEH99], the definition of coherent risk measures is based on sets called *acceptance sets*. Acceptance sets are sets of random variables, in our case losses. The definition of such sets will not be stated

[1] Diversification is a technique of combining a variety of instruments. It corresponds to "not putting all eggs in one basket".

here, but the concept is that if the risk of some random variable $V(\omega)$ is a positive value, then the investor should invest less in risky assets to meet capital requirements. If the risk of $V(\omega)$ is a negative value, then the investor can invest more in risky assets. We say that the risk associated with a random variable V is *acceptable* with respect to a coherent risk measure ρ when $\rho(V) \leq 0$, i.e when the risk of loss is negative. The relationship between acceptance sets and coherent risk measures can be stated as follows: The acceptance set is the set of acceptable random variables.

In the paper by Rockafellar [Roc07], there is introduced an additional axiom for coherent risk measures:

$$(Closedness)\ \ \rho(V) \leq 0 \text{ when } ||V^k - V||_2 \to 0 \text{ with } \rho(V^k) \leq 0 \quad (3.3)$$

This additional axiom (3.3) says that if the random variable V can be approximated by acceptable random variables V^k, then V is also acceptable.

3.3 Value-at-Risk and Conditional Value-at-Risk

It is the risk measure called *Conditional Value-at-Risk* (CVaR) that will be the most important risk measure in this thesis, due to its suitable properties, like for instance convexity. The related risk measure *Value-at-Risk* (VaR) will also be considered. This section is based on the articles by Krokhmal et. al. [KPU01], and Rockafellar and Uryasev [RU99]

VaR is defined to be an upper percentile of the loss distribution. α-VaR is an upper estimate of loss witch is exceeded with $(1-\alpha)100\%$ probability. For instance, if we have that 0.95-VaR equals \$1000, for some loss distribution, then there is a 5% chance of losing \$1000 or more.

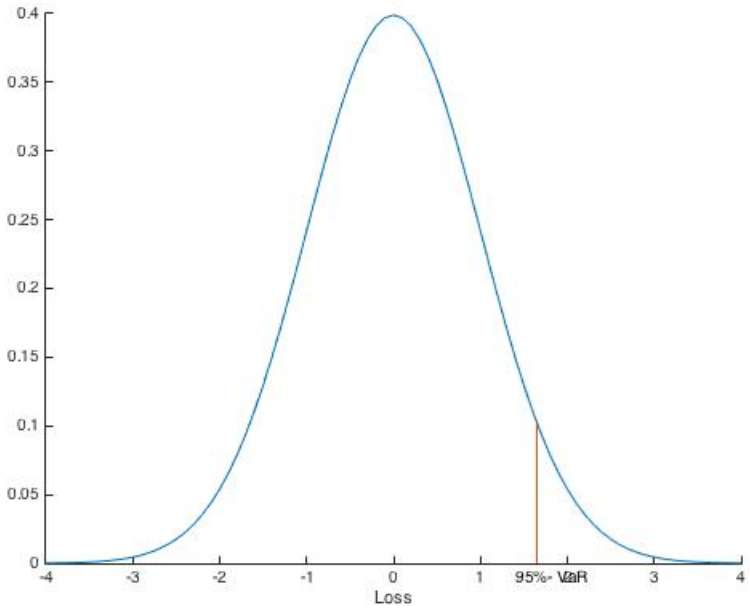


Figure 3.1: 0.95-VaR, when the loss function is normally distributed.

As a benefit of VaR being a upper α-quantile, VaR has simple representation of high losses, see Krokhmal et. al. [KPU01]. When the the underlying risk factors are normally distributed, VaR can be quite effectively estimated. For non-normal distributions, VaR may have undesirable properties such as lack of sub-additivity[2]. This implies that it is not a coherent risk measure. VaR is also non-convex for discrete distributions when it is calculated using scenarios. This makes it undesirable in optimization problems, as it is not guaranteed that local extrema are global.

CVaR is another percentile risk measure, which is closely related to VaR. As opposed to VaR, CVaR is convex and sub-additive. We will see that it is in fact a coherent risk measure. CVaR is for continuous distributions defined to be the conditional expected loss under the condition that it exceeds VaR, while it for general distributions is defined to be the weighted average of VaR and losses strictly exceeding VaR. I will only consider the continuous case in this section. If 0.95-VaR equals \$1000, then the 0.95-CVaR is the expected

[2]In some definitions of coherent risk measures, the convexity axiom is replaced by the sub-additivity axiom.

loss, given that the loss is \$1000 or more. This means that α-CVaR$\geq$ α-VaR for all α.

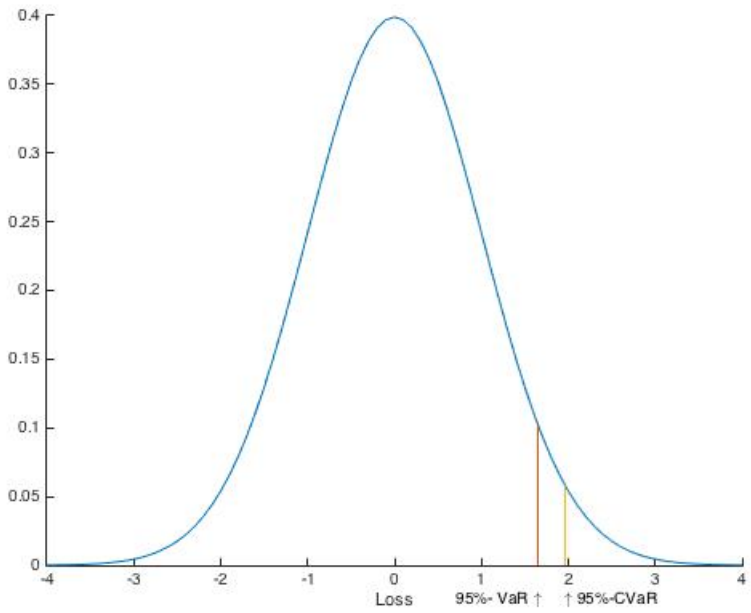


Figure 3.2: 0.95-CVaR, when the loss function is normally distributed.

When the loss distribution is normal, CVaR and VaR are equivalent in the sense that they produce the same optimal portfolio. For skewed distributions, CVaR and VaR risk optimal portfolios may be quite different, since CVaR is more sensitive right skewness in the loss distribution.

VaR and CVaR with continuous loss distributions

Assume we are given probability space $(\Omega, \mathcal{F}, P)$. Let $f(\mathbf{x}, Y(t,\omega))$ be the loss associated with the decision vector $\mathbf{x}$ to be chosen from some set $\mathcal{X} \subseteq \mathbb{R}^n$. We interpret $\mathbf{x}$ as a portfolio, with $\mathcal{X}$ as the set of available portfolios, and $Y(t,\omega)$ as the market prices. For simplicity we will use notation Y for the market prices. For each $\mathbf{x}$, the loss $f(\mathbf{x}, Y)$ is again a random variable, having a distribution in $\mathbb{R}$ induced by the distribution of Y. The assumption that Y has continuous distribution implies that it has density, which will be denoted p_Y. We will first give the probability of $f(\mathbf{x}, Y)$ not exceeding a threshold

ξ, also known as the cumulative distribution function of the loss associated with $\mathbf{x}$ (when $\mathbf{x}$ is fixed).

$$\begin{aligned}\psi(\mathbf{x},\xi) &= P(f(\mathbf{x},Y)\leq\xi)\\ &= E[\mathbf{1}_{\{\omega:f(\mathbf{x},Y)\leq\xi\}}]\\ &= \int_{\Omega}\mathbf{1}_{\{\omega:f(\mathbf{x},Y)\leq\xi\}}dP(\omega)\\ &= \int_{\{\omega:f(\mathbf{x},Y)\leq\xi\}}dP(\omega)\\ &= \int_{\{\mathbf{y}\in\mathbb{R}^n:f(\mathbf{x},\mathbf{y})\leq\xi\}}p_Y(\mathbf{y})d\mathbf{y}\end{aligned}$$

The function $\psi(\mathbf{x},\xi)$ is non-decreasing and will be assumed to be everywhere continuous with respect to ξ. This assumption is made for simplicity. Let $\alpha\in(0,1)$ be probability levels. We will denote by $\xi_\alpha(\mathbf{x})$ and $\phi_\alpha(\mathbf{x})$ the α-VaR and α-CVaR values for the loss random variable, respectively. $\xi_\alpha(\mathbf{x})$ is defined as follows:

$$\alpha\text{-VaR} = \xi_\alpha(\mathbf{x}) = \min\{\xi\in\mathbb{R}:\psi(\mathbf{x},\xi)\geq\alpha\} \tag{3.4}$$

Since $\psi(\mathbf{x},\xi)$ is continuous and non-decreasing, α-VaR comes out as the left end point of the interval consisting of the values of ξ satisfying $\psi(\mathbf{x},\xi)=\alpha$. Since the distribution of f is continuous, this interval will only consist of one point. $\phi_\alpha(\mathbf{x})$ is defined as follows:

$$\alpha\text{-CVaR} = E[f(\mathbf{x},Y)|f(\mathbf{x},Y)\geq\xi_\alpha] \tag{3.5}$$

There is an equivalent representation of ϕ_α that will be used is this thesis:

$$\begin{aligned}\phi_\alpha(\mathbf{x}) &= E[f(\mathbf{x},Y)|f(\mathbf{x},Y)\geq\xi_\alpha] \\ &= \frac{1}{P(f(\mathbf{x},Y)\geq\xi_\alpha)}E[f(\mathbf{x},Y)\mathbf{1}_{f(\mathbf{x},Y)\geq\xi_\alpha}] \\ &= \frac{1}{1-\alpha}\int_\Omega f(\mathbf{x},Y)\mathbf{1}_{f(\mathbf{x},Y)\geq\xi_\alpha}dP \\ &= \frac{1}{1-\alpha}\int_{f(\mathbf{x},\mathbf{y})\geq\xi_\alpha(\mathbf{x})} f(\mathbf{x},\mathbf{y})p_Y(\mathbf{y})d\mathbf{y} \qquad (3.6)\end{aligned}$$

The set $f(\mathbf{x},Y)\geq\xi$ in the integrals above is just notation for the set $\{\omega : f(\mathbf{x},Y)\geq\xi\}$, and $f(\mathbf{x},\mathbf{y})\geq\xi$ is notation for the set $\{\mathbf{y}\in\mathbb{R}^n : f(\mathbf{x},\mathbf{y})\geq\xi\}$. It should be clear from the context which set we are integrating over.

Example 3.3.1. Assume that the loss is normally distributed, then the 0.95 -VaR is the the minimum threshold such that $\psi(\mathbf{x},\xi)\geq 0.95$. From tables on the standard normal distribution we find that this minimum threshold is $\xi = 1.645$. So $\xi_{0.95} = 1.645$, see Figure 3.1. There is a 5% chance that the loss will exceed the mean by 1.645 standard deviations, or more.

□

Let us define a function $F_\alpha(\mathbf{x},\xi)$ on $\mathcal{X}\times\mathbb{R}$. This will be useful for our optimization problems later. We will see that F_α is convex and continuously differentiable, which eliminates the possibilities of a local minimum being different from a global minimum.

$$\begin{aligned}F_\alpha(\mathbf{x},\xi) =& \xi + (1-\alpha)^{-1}\int_{\mathbf{y}\in\mathbb{R}^n}[f(\mathbf{x},\mathbf{y})-\xi]^+ p_Y(\mathbf{y})d\mathbf{y} \qquad (3.7)\\ =& \xi + (1-\alpha)^{-1}E[[f(\mathbf{x},Y)-\xi]^+] \qquad (3.8)\end{aligned}$$

where $[t]^+ = \max\{t,0\}$. Both representations (3.7) and (3.8) will be useful later.

Before we state an important theorem, we will give a lemma that will be helpful in the proof of this theorem.

Lemma 3.3.1. *Let* $G(\xi) = \int_{\mathbf{y}\in\mathbb{R}^n}[f(\mathbf{y}-\xi)]^+ p_Y(\mathbf{y})d\mathbf{y}$. *Then* G *is a continuously differentiable function with derivative*

$$G'(\xi) = \psi(\xi) - 1$$

The following theorem show the importance of F_α; minimizing it with respect to ξ gives ϕ_α, and as a bi-product we get ξ_α.

Theorem 3.3.2. *As a function of* ξ, $F_\alpha(\mathbf{x},\xi)$ *is convex and continuously differentiable. The* α*-CVaR of the loss associated with any* $\mathbf{x} \in \mathcal{X}$ *can be determined by*

$$\phi_\alpha(\mathbf{x}) = \min_{\xi\in\mathbb{R}} F_\alpha(\mathbf{x},\xi)$$

Let

$$A_\alpha(\mathbf{x}) = \operatorname*{argmin}_{\xi\in\mathbb{R}} F_\alpha(\mathbf{x},\xi)$$

namely the set consisting of the values of ξ *for which the minimum is attained. Then* $A_\alpha(\mathbf{x})$ *is a non-empty, closed, bounded interval, and the* α*-VaR of the loss is given by*

$$\xi_\alpha(\mathbf{x}) = \textit{left endpoint of } A_\alpha(\mathbf{x})$$

In particular, we always have

$$\xi_\alpha(\mathbf{x}) \in \operatorname*{argmin}_{\xi\in\mathbb{R}} F_\alpha(\mathbf{x},\xi) \textit{ and } \phi_\alpha(\mathbf{x}) = F_\alpha(\mathbf{x},\xi_\alpha(\mathbf{x}))$$

Proof. Let us start with showing that $F_\alpha(\mathbf{x},\xi)$ is convex and continuously differentiable, as a function of ξ. To simplify notation we omit $\mathbf{x}$, for instance we write $F_\alpha(\xi)$ in stead of $F_\alpha(\mathbf{x},\xi)$.

The fact that $F_\alpha(\xi)$ is continuously differentiable follows from Lemma 3.3.1. The convexity of $F_\alpha(\xi)$ follows from the linearity and monotonicity of the expectation and the fact that the function $[f(Y)-\xi]^+$ is convex as a function of ξ.

$$[f(Y)-\xi]^+ = \max\{f(Y)-\xi, 0\} = \begin{cases} f(Y)-\xi & \text{if } f(Y) > \xi \\ 0 & \text{if } f(Y) \le \xi \end{cases}$$

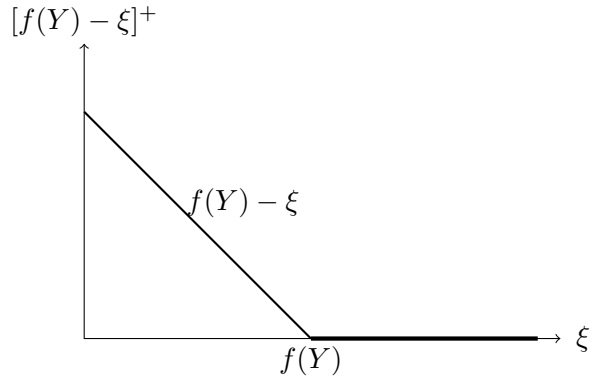


Figure 3.3: The function $[f(Y)-\xi]^+$.

We can view $f(Y)$ as a constant, and ξ as the variable. Since both $f(Y)-\xi$ and the zero function are convex functions, we know from Theorem 2.1.9 that the function $\max\{f(Y)-\xi, 0\}$ is convex. From Figure 3.3 we can also see that it 'bends upward'. Let $0 \leq \lambda \leq 1$.

$$
\begin{aligned}
&F_\alpha((1-\lambda)\xi_1 + \lambda\xi_2) \\
&= (1-\lambda)\xi_1 + \lambda\xi_2 + \frac{1}{1-\alpha} E\big[[f(Y) - (1-\lambda)\xi_1 - \lambda\xi_2]^+\big] \\
&\leq (1-\lambda)\xi_1 + \lambda\xi_2 + \frac{1}{1-\alpha} E\big[(1-\lambda)[f(Y)-\xi_1]^+ + \lambda[f(Y)-\xi_2]^+\big] \\
&= (1-\lambda)\xi_1 + \frac{1-\lambda}{1-\alpha} E\big[[f(Y)-\xi_1]^+\big] + \lambda\xi_2 + \frac{\lambda}{1-\alpha} E\big[[f(Y)-\xi_2]^+\big] \\
&= (1-\lambda)F_\alpha(\xi_1) + \lambda F_\alpha(\xi_2)
\end{aligned}
$$

The inequality above follows from the fact that $[f(Y)-\xi]^+$ is convex with respect to ξ, and from the monotonicity of the expectation. The second equality follows from the linearity of the expectation. This proves that $F_\alpha(\mathbf{x}, \xi)$ is convex as a function of ξ.

Next we want to show the property $\phi_\alpha(\mathbf{x}) = \min_{\xi\in\mathbb{R}} F_\alpha(\mathbf{x}, \xi)$, and start by rewriting $F_\alpha(\xi)$:

$$
\begin{aligned}
F_\alpha(\xi) =& \xi + \frac{1}{1-\alpha}\int_{\mathbb{R}^n} [f(\mathbf{y}) - \xi]^+ p_Y(\mathbf{y}) d\mathbf{y} \\
=& \xi + \frac{1}{1-\alpha}\bigg(\int_{\mathbb{R}^n} \mathbf{1}_{f(\mathbf{y})\geq\xi} [f(\mathbf{y}) - \xi]^+ p_Y(\mathbf{y}) d\mathbf{y} \\
& + \int_{\mathbb{R}^n} \mathbf{1}_{f(\mathbf{y})<\xi} [f(\mathbf{y}) - \xi]^+ p_Y(\mathbf{y}) d\mathbf{y} \bigg) \\
=& \xi + \frac{1}{1-\alpha}\int_{\mathbb{R}^n} \mathbf{1}_{f(\mathbf{y})\geq\xi} \big(f(\mathbf{y}) - \xi\big)\, p_Y(\mathbf{y}) d\mathbf{y} \\
=& \xi + \frac{1}{1-\alpha}\int_{f(\mathbf{y})\geq\xi} f(\mathbf{y}) - \xi\, p_Y(\mathbf{y}) d\mathbf{y} \\
=& \xi + \frac{1}{1-\alpha}\bigg(\int_{f(\mathbf{y})\geq\xi} f(\mathbf{y})\, p_Y(\mathbf{y}) d\mathbf{y} \\
& + \int_{f(\mathbf{y})\leq\xi} \xi\, p_Y(\mathbf{y}) d\mathbf{y} - \int_{\mathbb{R}^n} \xi\, p_Y(\mathbf{y}) d\mathbf{y} \bigg) \\
=& \xi + \frac{1}{1-\alpha}\bigg(\int_{f(\mathbf{y})\geq\xi} f(\mathbf{y})\, p_Y(\mathbf{y}) d\mathbf{y} - \xi(1-\psi(\xi)) \bigg) \\
=& \xi + \frac{\xi}{1-\alpha}(\psi(\xi) - 1) + \frac{1}{1-\alpha}\int_{f(\mathbf{y})\geq\xi} f(\mathbf{y})\, p_Y(\mathbf{y}) d\mathbf{y}
\end{aligned}
$$

Here I have used that $\mathbf{1}_{f(\mathbf{y})\geq\xi} = \mathbf{1}_{\mathbb{R}^n} - \mathbf{1}_{f(\mathbf{y})<\xi}$. To differentiate F_α with respect to ξ is rather difficult because of the dependence of ξ in the integrals. So, as in Rockafellar and Uryasev [RU99], we will rely on Lemma 3.3.1, and get that

$$
F'_\alpha(\xi) = 1 + \frac{1}{1-\alpha}(\psi(\xi) - 1)
$$

To minimize this, set $F'_\alpha = 0$

$$\begin{aligned} F'_\alpha(\xi) = &\quad 0 \\ 1 + (1-\alpha)^{-1}(\psi(\xi) - 1) = &\quad 0 \\ 1 - \alpha - 1 + \psi(\xi) = &\quad 0 \\ \psi(\xi) = &\quad \alpha \end{aligned}$$

The values of ξ that furnish the minimum of F_α are precisely those ξ satisfying $\psi(\xi) = \alpha$. The set of these ξ's equals the set A_α. From the definition of ξ_α we see that this is an interval where ξ_α is its left end point. ξ_α is indeed contained in A_α, which proves that it is non-empty. Since $\psi(\xi, \alpha)$ is continuous and nondecreasing in α with

$$\lim_{\alpha \to \infty} \psi(\xi, \alpha) = 1 \text{ and } \lim_{\alpha \to -\infty} \psi(\xi, \alpha) = 0$$

A_α is closed and bounded.

Now, we will get back to showing that $\min_{\xi \in \mathbb{R}} \ F(\mathbf{x}, \xi) = \phi_\alpha(\mathbf{x})$. If $A_\alpha = \{\xi_\alpha\}$ then

$$F_\alpha(\xi_\alpha) = \xi_\alpha - \frac{\xi_\alpha}{1-\alpha}(1-\alpha) + \frac{1}{1-\alpha} \int_{f(\mathbf{y}) \geq \xi_\alpha} f(\mathbf{y}) \ p_Y(\mathbf{y}) d\mathbf{y} = \phi_\alpha$$

Now, assume A_α consists of more than one point, and let $\xi^* \in A_\alpha$. Then

$$\begin{aligned} \min_{\xi \in \mathbb{R}} F_\alpha(\xi) = &\xi^* - \frac{\xi^*}{1-\alpha}(1-\alpha) + \frac{1}{1-\alpha} \int_{f(\mathbf{y}) \geq \xi^*} f(\mathbf{y}) \ p_Y(\mathbf{y}) d\mathbf{y} \\ = &\frac{1}{1-\alpha} \int_{f(\mathbf{y}) \geq \xi^*} f(\mathbf{y}) \ p_Y(\mathbf{y}) d\mathbf{y} \\ = &\frac{1}{1-\alpha} \int_{f(\mathbf{y}) \geq \xi_\alpha} f(\mathbf{y}) \ p_Y(\mathbf{y}) d\mathbf{y} \\ = &\phi_\alpha \end{aligned}$$

This finishes the proof.

■

This next theorem will be very useful in the optimization model later.

Theorem 3.3.3. *Minimizing $\phi_\alpha(\mathbf{x})$ over all $\mathbf{x} \in \mathcal{X}$ is equivalent to minimizing $F_\alpha(\mathbf{x}, \xi)$ over all $(\mathbf{x}, \xi) \in \mathcal{X} \times \mathbb{R}$, in the sense that*

$$\min_{\mathbf{x} \in \mathcal{X}} \phi_\alpha(\mathbf{x}) = \min_{(\mathbf{x}, \xi) \in \mathcal{X} \times \mathbb{R}} F_\alpha(\mathbf{x}, \xi) \tag{3.9}$$

where moreover a pair $(\mathbf{x}^*, \xi^*)$ achieves the right hand side minimum if and only if $\mathbf{x}^*$ achieves the left hand minimum and $\xi^* \in A_\alpha(\mathbf{x}^*)$. In particular, in circumstances where $A_\alpha(\mathbf{x})$ reduces to a single point, the minimization of $F_\alpha(\mathbf{x}, \xi)$ over $(\mathbf{x}, \xi) \in \mathcal{X} \times \mathbb{R}$ produce a pair $(\mathbf{x}^*, \xi^*)$, not necessarily unique, such that $\mathbf{x}^*$ minimizes $\phi_\alpha(\mathbf{x})$ and ξ^* gives the corresponding α-VaR.
Furthermore, $F_\alpha(\mathbf{x}, \xi)$ is convex with respect to $(\mathbf{x}, \xi)$, and $\phi_\alpha(\mathbf{x})$ is convex with respect to $\mathbf{x}$, when $f(\mathbf{x}, Y)$ is convex with respect to $\mathbf{x}$. In this case, if the constraints are such that $\mathcal{X}$ is a convex set, the joint minimization is an instance of convex programming.

Proof. The initial claims in Theorem 3.3.3 is a consequence of the statement $\phi_\alpha(\mathbf{x}) = \min_{\xi \in \mathbb{R}} F_\alpha(\mathbf{x}, \xi)$ in Theorem 3.3.2 and the fact that minimizing $F_\alpha(\mathbf{x}, \xi)$ over $(\mathbf{x}, \xi) \in \mathcal{X} \times \mathbb{R}$ can be carried out by first minimizing over $\xi \in \mathbb{R}$ for fixed $\mathbf{x}$ and then minimizing the result over $\mathbf{x} \in \mathcal{X}$.

To show that $F_\alpha(\mathbf{x}, \xi)$ is convex with respect to $(\mathbf{x}, \xi)$, it is enough to show that the function $[f(\mathbf{x}, Y) - \xi]^+$ is convex with respect to $(\mathbf{x}, \xi)$. In fact, all we need to show is that $f(\mathbf{x}, Y) - \xi$ is a convex function, see Theorem 2.1.9. To see that this function is convex, let $g(\mathbf{x}, \xi) = f(\mathbf{x}, Y) - \xi$ and let $0 \leq \lambda \leq 1$, then

$$\begin{aligned}
&g((1-\lambda)(\mathbf{x}_1, \xi_1) + \lambda(\mathbf{x}_2, \xi_2)) \\
&= f((1-\lambda)\mathbf{x}_1 + \lambda\mathbf{x}_2, Y) - (1-\lambda)\xi_1 - \lambda\xi_2 \\
&\leq (1-\lambda) f(\mathbf{x}_1, Y) - (1-\lambda)\xi_1 + \lambda f(\mathbf{x}_2, Y) - \lambda\xi_2 \\
&= (1-\lambda)\big(f(\mathbf{x}_1, Y) - \xi_1\big) + \lambda\big(f(\mathbf{x}_2, Y) - \xi_2\big) \\
&= (1-\lambda) g(\mathbf{x}_1, \xi_1) + \lambda g(\mathbf{x}_2, \xi_2)
\end{aligned}$$

The inequality above follows from the assumption that $f(\mathbf{x}, Y)$ is convex.

The convexity of $\phi_\alpha(\mathbf{x})$ follows from the fact that minimization of an extended real valued convex function of two vector variables, with respect to one of these variables, results in a convex function of the remaining variable, see Rockafellar and Uryasev [RU99].

■

As mentioned earlier, CVaR is a coherent risk measure, and hence also a convex risk measure.

Proposition 3.3.4. *CVaR is a coherent risk measure.*

Proof. To prove this proposition, we must show that CVaR satisfies axioms (R1)-(R4) of the definition of a coherent risk measure in Section 3.2. Let $V(\mathbf{x}, Y), V_1(\mathbf{x}, Y), V_2(\mathbf{x}, Y)$ be random variables representing loss functions. For simplicity we write V, V_1, V_2, respectively. To show this proposition we will use Theorem 3.3.2 repeatedly.

(R1) Translation invariance: From Theorem 3.3.2 we know that $\phi_\alpha(\mathrm{V}) = \min_{\xi\in\mathbb{R}} F_\alpha(V, \xi)$. So let ξ_{V+c} be such that $\phi_\alpha(V+c) = \xi_{V+c} + (1-\alpha)^{-1}E[[V + c - \xi_{V+c}]^+]$ and let ξ_V be such that $\phi_\alpha(V) = \xi_V + (1-\alpha)^{-1}E[[V - \xi_V]^+]$. Then, since $\xi_{V+c} = \xi_V + c$:

$$\begin{aligned}
\phi_\alpha(V + c) &= \xi_{V+c} + (1-\alpha)^{-1}E\big[[V + c - \xi_{V+c}]^+\big] \\
&= \xi_V + c + (1-\alpha)^{-1}E\big[[V + c - \xi_V - c]^+\big] \\
&= \xi_V + (1-\alpha)^{-1}E\big[[V - \xi_V]^+\big] + c \\
&= \phi_\alpha(V) + c
\end{aligned}$$

Before we show that the remaining axioms are satisfied, we will show convexity, monotonicity and positive homogeneity of $[V - \xi]^+$, with respect to V. Here, ξ is a constant number.

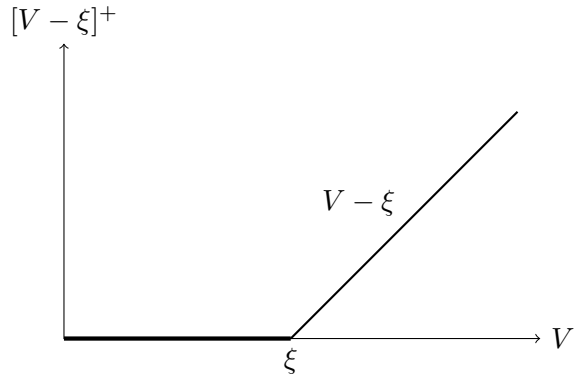


Figure 3.4: The function $[V-\xi]^+$.

$$[V-\xi]^+ = \max\{V-\xi, 0\} = \begin{cases} V-\xi & \text{if } V > \xi \\ 0 & \text{if } V \leq \xi \end{cases} \tag{3.10}$$

As a function of V, $[V-\xi]^+$ is clearly convex as $V-\xi$ and the zero function are convex, see Theorem 2.1.9. From Figure 3.4, we see that it is also a piecewise linear function that 'bends upward'. It is also clear that it is monotone, i.e if $V_1 \leq V_2$ then $[V_1-\xi]^+ \leq [V_2-\xi^+]$. From equation (3.10) we see that it is positive homogeneous, i.e. $[\lambda(V-\xi)]^+ = \lambda[V-\xi]^+$.

(R2) Monotonicity: The monotonicity of CVaR follows from the monotonicity of $[V-\xi]^+$. Assume that $V_1 \leq V_2$. Assume that ξ_1 is such that $\phi_\alpha(V_1) = \xi_1 + (1-\alpha)^{-1}E[[V_1-\xi_1]^+]$ and that ξ_2 is such that $\phi_\alpha(V_2) = \xi_2 + (1-\alpha)^{-1}E[[V_2-\xi_2]^+]$. Then

$$\begin{aligned} \phi_\alpha(V_1) &= \xi_1 + (1-\alpha)^{-1}E\big[[V_1-\xi_1]^+\big] \\ &\leq \xi_2 + (1-\alpha)^{-1}E\big[[V_2-\xi_2]^+\big] \\ &= \phi_\alpha(V_2) \end{aligned}$$

The inequality above follows from the monotonicity of $[V-\xi]^+$ and the expectation, and the fact that $\xi_1 \leq \xi_2$.

(R3) Convexity: Let $\lambda \in [0,1]$. As before, let ξ_1 be such that $\phi_\alpha(V_1) = \xi_1 +$

$(1-\alpha)^{-1}E[[V_1-\xi_1]^+]$ and ξ_2 be such that $\phi_\alpha(V_2) = \xi_2+(1-\alpha)^{-1}E[[V_2-\xi_2]^+]$. Also, let ξ be such that $\phi_\alpha((1-\lambda)V_1+\lambda V_2) = \xi + (1-\alpha)^{-1}E[[(1-\lambda)V_1 + \lambda V_2 - \xi]]$. Then

$$
\begin{aligned}
&\phi_\alpha((1-\lambda)V_1+\lambda V_2) \\
&= \xi + (1-\alpha)^{-1}E\big[[(1-\lambda)V_1+\lambda V_2-\xi]^+\big] \\
&\le (1-\lambda)\xi_1 + \lambda\xi_2 + (1-\alpha)^{-1}E\big[[(1-\lambda)(V_1-\xi_1)+\lambda(V_2-\xi_2)]^+\big] \\
&\le (1-\lambda)\xi_1 + \lambda\xi_2 + (1-\alpha)^{-1}E\big[(1-\lambda)[V_1-\xi_1]^+ + \lambda[V_2-\xi_2]^+\big] \\
&= (1-\lambda)\Big(\xi_1 + (1-\alpha)^{-1}E\big[[V_1-\xi_1]^+\big]\Big) + \lambda\Big(\xi_2 + (1-\alpha)^{-1}E\big[[V_2-\xi_2]^+\big]\Big) \\
&= (1-\lambda)\phi_\alpha(V_1) + \lambda\phi_\alpha(V_2)
\end{aligned}
$$

The first inequality follows from Theorem 3.3.2. The second inequality follows from the convexity of $[V-\xi]^+$ and the monotonicity of the expectation. The second equality follows from the linearity of the expectation.

(R4) Positive homogeneity: Let $\lambda \ge 0$, let $\xi_{\lambda V}$ be such that $\xi_{\lambda V} + (1-\alpha)^{-1}E[[\lambda V - \xi_{\lambda V}]^+]$, and let ξ_V be such that $\xi_V + (1-\alpha)^{-1}E[[V-\xi_V]^+]$. Then since $\xi_{\lambda V} = \lambda\xi_V$:

$$
\begin{aligned}
\phi_\alpha(\lambda V) = \quad & \xi_{\lambda V} + (1-\alpha)^{-1}E\big[[\lambda V - \xi_{\lambda V}]^+\big] \\
&= \lambda\xi_V + (1-\alpha)^{-1}E\big[[\lambda V - \lambda\xi_V]^+\big] \\
&= \lambda\xi_V + \lambda(1-\alpha)^{-1}E\big[[V-\xi_V]^+\big] \\
&= \lambda\phi_\alpha(V)
\end{aligned}
$$

■

4 | Portfolio optimization problems

In this chapter we will give an introduction to portfolio optimization problems. First, the classical Markowitz mean-variance problem will be given some attention, then we will study an optimization model with CVaR constrains. Both of these optimization problems can be written as problems of maximizing the expected return of a portfolio, subject to some risk measure. For such optimization problems, the relationship between the mean and the risk for optimal portfolios is often represented by a graph called an *efficient frontier*.

Both of the optimization models presented in this chapter will be represented as one-step models, that is, decisions are made for one day, one month or one year in the future. Multi-step models will not be considered, as they are too complex.

This chapter is based on the articles by Krokhmal et. al. [KZU11], Krokhmal et. al. [KPU01] and the book by Best [Bes10].

Optimization and decision making under uncertainties is a popular area, and so are the applications to finance and portfolio optimization. The basic portfolio optimization problem is to decide how much of an investor's wealth that should be optimally invested in each asset. Recall, in a deterministic setting, an optimization problem typically looks like

$$\begin{array}{ll} \max\limits_{\mathbf{x}\in S} & f(\mathbf{x}) \\ \text{subject to} & g_i(\mathbf{x}) \leq 0 \quad , i = 1, \ldots, m \end{array}$$

where $\mathbf{x} \in S$ is a decision vector in some deterministic set $S \subseteq \mathbb{R}^n$, and $f, g_i \colon S \to \mathbb{R}$ for all i. Let ω be a random element, describing uncertainties. When optimizing under uncertainties, we have to deal with $f(\mathbf{x}, \omega)$ and $g_i(\mathbf{x}, \omega)$, in stead of $f(\mathbf{x})$ and $g_i(\mathbf{x})$. Optimizing $f(\mathbf{x}, \omega)$ subject to $g_i(\mathbf{x}, \omega)$

leads to some difficulties. The decision vector $\mathbf{x}$ must be determined before one knows what the state of the world turns out to be, so the optimizer has incomplete information at the time $\mathbf{x}$ must be chosen. This means that making decisions involve some risk. Also, $f(\mathbf{x}, \omega)$ is a random variable for a fixed $\mathbf{x}$. This means that we can't replace $f(\mathbf{x})$ by $f(\mathbf{x}, \omega)$. The same goes for the constraints. The treatment of the objective function and the constraints should not be conceptual distinct. For the purpose of optimizing functions with uncertainties, risk measures are brought into the picture. As we have discussed in Section 3.2, there are several ways to measure risk.

4.1 Markowitz Mean-Variance portfolio optimization

This section is based on the book by Best [Bes10]. The goal of classical Markowitz mean-variance portfolio optimization is to choose a portfolio such that the expected return of the portfolio is large, but the variance is small. In this optimization model, the variance is the risk measure.

Let $\mathbf{x} = (x_1, x_2, \ldots, x_n)^T$ denote the portfolio, where x_j denotes the proportion of wealth to be invested in security j, for $j = 1, \ldots, n$. Note that this representation of the portfolio differ from the definition in Section 3.1, where x_j denoted the number of units to be invested in security j, rather than the percentage. Let μ_j denote the expected return of asset j, and $\sigma_{i,j}$ denote the covariance between the returns of asset i and j, $1 \leq i, j \leq n$. Let

$$\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_n)^T \text{ and } \Sigma = [\sigma_{i,j}]$$

where $\Sigma \in \mathbb{R}^{n \times n}$ is a symmetric, positive semidefinite matrix. Then the expected return and the variance of the portfolio is given by, respectively

$$\mu_p = \boldsymbol{\mu}^T \mathbf{x} \text{ and } \sigma_p^2 = \mathbf{x}^T \Sigma \mathbf{x}$$

Suppose we have two portfolios with the same expected return, but the first has smaller variance than the second one. Then the first portfolio is clearly more attractive because it bears less risk for the same expected return. This is the main idea behind Markowitz's definition of efficient portfolios.

Definition 4.1.1. *A portfolio is* variance-efficient *if for a fixed μ_p there is no other portfolio which has a smaller variance σ_p^2.*

This definition implies that a portfolio is efficient if for some fixed expected return μ_p, the variance is minimized. Variance efficient portfolios are solutions to the following optimization problem:

$$\min\{\mathbf{x}^T\Sigma\mathbf{x} : \boldsymbol{\mu}^T\mathbf{x} = \mu_p, \mathbf{e}^T\mathbf{x} = 1, \mathbf{x} \geq \mathbf{0}\} \tag{4.1}$$

where $\mathbf{e}$ is the column vector consisting entirely of ones, of length n. This optimization problem has quadratic objective function and linear constraints, and is a quadratic optimization problem as discussed in Section 2.3.

There is an equivalent definition of an efficient portfolio. Suppose we have two portfolios with the same variance, but the first portfolio has larger expected return than the second portfolio. Then the first is more attractive because it gives a higher return for the same risk.

Definition 4.1.2. *A portfolio is* expected return-efficient *if for fixed σ_p^2, there is no other portfolio with a larger μ_p.*

Definition 4.1.2 implies that a portfolio is efficient if it is the solution to the following optimization problem:

$$\max\{\boldsymbol{\mu}^T\mathbf{x} : \ \mathbf{x}^T\Sigma\mathbf{x} = \sigma_p^2, \mathbf{e}^T\mathbf{x} = 1, \mathbf{x} \geq \mathbf{0}\} \tag{4.2}$$

This optimization problem has linear objective function, quadratic and linear constraints.

There is a third, equivalent optimization problem, which also produce efficient portfolios. Let t be a scalar, and consider the following problem:

$$\min\{-t\boldsymbol{\mu}^T\mathbf{x} + \frac{1}{2}\mathbf{x}^T\Sigma\mathbf{x} : \ \mathbf{e}^T\mathbf{x} = 1, \mathbf{x} \geq \mathbf{0}\} \tag{4.3}$$

This problem has quadratic objective function and linear constraints. It is hence a quadratic optimization problem, see Section 2.3.

Definition 4.1.3. *A portfolio is* parametric-efficient *if it is an optimal solution for some non-negative parameter t.*

The three optimization problems presented in this section generates a family of optimal solutions. They are in fact identical, provided $\mathbf{x}$ is not a multiple of $\mathbf{e}$.

For solving the optimization problems where the objective is quadratic and the constraints are linear, we can use what we know about quadratic optimization from Section 2.3.

4.2 Efficient Frontiers

This section is based on the article by Krokhmal et. el [KPU01], and aims to describe the concept of *efficient frontiers*. The concept of efficient frontiers is based on Markowitz's concept of efficient portfolios. Efficient frontiers are used to describe the relationship between the expected return and the risk for optimal portfolios. Recall that a portfolio is said to be expected return-efficient if for a fixed level of risk, there is no portfolio with higher return. Similarly, a portfolio is said to be risk-efficient if for a fixed level of return, there is no portfolio with lower risk. For Markowitz mean-variance optimization, the three families of problems generate the same efficient frontiers. So, in this section we will only consider the problems generating expected return-efficient portfolios, to study the concept.

As we will see in the next section, the portfolio optimization problem with CVaR constrains is a problem of maximizing expected return subject to the risk measure CVaR. So, studying efficient frontiers for such problems is also of interest.

Let $\mathbf{x} = (x_1, x_2, \ldots, x_n)^T$ denote the portfolio, where x_j denotes the proportion of wealth to be invested in security j, for $j = 1, \ldots, n$, as in Section 4.1. Let $\rho(\mathbf{x})$ be some risk measure. In the Markowitz mean-variance optimization problem, the risk measure is the variance and in the CVaR optimization problem, the risk measure is the CVaR. Let $R(\mathbf{x})$ denote the expected return of portfolio $\mathbf{x}$

$$R(\mathbf{x}) = -E[f(\mathbf{x}, Y)]$$

where $-f(\mathbf{x}, Y)$ is the payoff function and Y is a random variable representing prices. Then solutions to the following problem are expected return-efficient portfolios:

$$\begin{array}{ll} \max\limits_{\mathbf{x}} & R(\mathbf{x}) \\ \text{subject to} & \rho(\mathbf{x}) \leq \nu \\ & \mathbf{x} \in \mathcal{X} \end{array} \tag{4.4}$$

Here, the parameter ν represents the level of risk, and is often some percentage of the initial portfolio value that one are putting up for risk exposure. For every feasible portfolio $\mathbf{x}$, one gets a pair $(\rho(\mathbf{x}), R(\mathbf{x}))$ that can be represented as a point in the plane. These points are combinations of the points

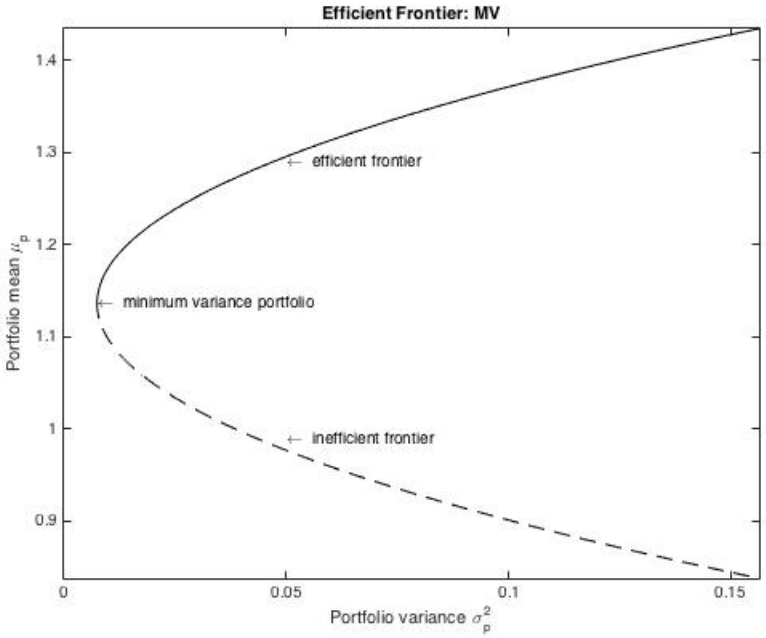


Figure 4.1: The efficient frontier for the Markowitz mean-variance optimization problem.

corresponding to invest everything in one of the assets.

Let $\mathbf{x}^*$ be the optimal solution to problem (4.4) for a fixed ν. Varying ν will generate a set of points $\{(\rho(\mathbf{x}_k^*), R(\mathbf{x}_k^*))\}$ for k in some index set I. It is the graph generated by this set of points that is called the efficient frontier. As we can see from Figure 4.1, the relationship between risk and return is not linear. This means that increasing the risk level ν by a factor a does not imply that the expected return is increased by the same factor a.

When choosing an optimal portfolio one should choose one that lies on this efficient frontier. The plot of an efficient frontier should never have any 'horizontal lines'. Points lying on a horizontal line have the same returns, but have different risks. Points on the right side of such a line correspond to portfolios that are not optimal. The same is true for 'vertical lines', as points on these have same risks, but have different returns.

Assume the risk free asset, namely the bank investment is included in the portfolio. The the efficient portfolios may be combinations of this risk free asset and risky asset. The efficient frontier for such problems may therefore

look a bit different from the case where the portfolio only consisted of risky assets. The two different cases are illustrated in Section 6.1.

4.3 Portfolio optimization with CVaR constraints

In this section we will use the main article of this thesis, Krokhmal et. al. [KPU01], to give an optimization model of maximizing expected returns subject to CVaR constraints.

The key to the model is the function $F_\alpha(\mathbf{x}, \xi)$ given in Section 3.3, due to its properties. The portfolio optimization technique we now shall look at, will optimize CVaR and at the same time calculate VaR.

Let $\phi_\alpha(\mathbf{x})$ be as in Section 3.3, $\mathbf{x} \in \mathcal{X}$, where $\mathcal{X}$ is the set of attainable portfolios, and let $R(\mathbf{x})$ be the expected return of portfolio $\mathbf{x}$. The following theorem is a special case of Theorem 3 in Krokhmal et. al. [KPU01].

Theorem 4.3.1. *Let* $\mathbf{x}$ *be the portfolio,* $R(\mathbf{x})$ *be the return of portfolio* $\mathbf{x}$ *and* $\phi_\alpha(\mathbf{x})$ *be the CVaR of* $\mathbf{x}$*. Consider the following three problems:*

$$\text{(P1)} \quad \min_{\mathbf{x}} \{\phi_\alpha(\mathbf{x}) \mid R(\mathbf{x}) \geq \gamma, \mathbf{x} \in \mathcal{X}\}$$

$$\text{(P2)} \quad \max_{\mathbf{x}} \{R(\mathbf{x}) \mid \phi_\alpha(\mathbf{x}) \leq \nu, \mathbf{x} \in \mathcal{X}\}$$

$$\text{(P3)} \quad \min_{\mathbf{x}} \{\phi_\alpha(\mathbf{x}) - \mu R(\mathbf{x}) \mid \mathbf{x} \in \mathcal{X}, \kappa \geq 0\}$$

Suppose the constraints $R(\mathbf{x}) \geq \gamma, \phi_\alpha(\mathbf{x}) \leq \nu$ *have initial points. Varying the parameters* γ, ν *and* κ *traces the efficient frontier for the problems* $(P1) - (P3)$*, accordingly. If* $\phi_\alpha(\mathbf{x})$ *is convex,* $R(\mathbf{x})$ *is concave and and the set* $\mathcal{X}$ *is convex, then the three problems* $(P1) - (P3)$ *generate the same efficient frontier.*

The proof of Theorem 4.3.1 can be found in the article by Krokhmal et. al. [KPU01].

Solutions to problems (P1)-(P3) in Theorem 4.3.1 correspond to what was defined as variance-efficient, expected return-efficient and parametric-efficient portfolios, respectively, in the Markowitz mean-variance optimization model.

In the rest of this thesis we will concentrate on problem (P2), the problem of maximizing the expected return with CVaR constraints. Theorem 3.3.3 shows that the function $F_\alpha(\mathbf{x}, \xi)$ can be used instead of $\phi_\alpha(\mathbf{x})$ in problem (P1). The next theorem shows that $F_\alpha(\mathbf{x}, \xi)$ can be used instead of $\phi_\alpha(\mathbf{x})$ in problem (P2). The same holds also for problem (P3), for more details on this see Krokhmal et. al. [KPU01].

Theorem 4.3.2. *The two optimization problems below*

$$\text{(P2)} \quad \max_{\mathbf{x}\in\mathcal{X}} \{R(\mathbf{x}) \mid \phi_\alpha(\mathbf{x}) \leq \nu, \mathbf{x} \in \mathcal{X}\}$$

$$\text{(P2')} \quad \max_{(\mathbf{x},\xi)\in\mathcal{X}\times\mathbb{R}} \{R(\mathbf{x}) \mid F_\alpha(\mathbf{x}, \xi) \leq \nu, \mathbf{x} \in \mathcal{X}\}$$

are equivalent in the sense that their objectives achieve the same maximum values. Moreover, if the CVaR constraint in (P2) is active, a pair $(\mathbf{x}^*, \xi^*)$ achieves the maximum of (P2') if and only if $\mathbf{x}^*$ achieves the maximum of (P2) and $\xi^* \in A_\alpha(\mathbf{x}^*)$. In particular, when the interval $A_\alpha(\mathbf{x}^*)$ reduces to a single point, the maximization of $R(\mathbf{x})$ over $(\mathbf{x}, \xi) \in \mathcal{X} \times \mathbb{R}$ produces a pair $(\mathbf{x}^*, \xi^*)$ such that $\mathbf{x}^*$ maximizes the return, and ξ^* gives the corresponding α-VaR.

The proof of Theorem 4.3.2 can be found in the article by Krokhmal et. al. [KPU01].

If $R(\mathbf{x})$ is linear, it is also convex. If the loss function $f(\mathbf{x}, Y)$ also is linear, then $F_\alpha(\mathbf{x}, \xi)$ is convex. In this case, problem (P2') is a convex optimization problem. This ensures that a local maximum is also global.

Discretization and linearization

The integral in $F_\alpha(\mathbf{x}, \xi)$ can be approximated in various ways. The approximation we will use is described in Krokhmal et. al [KPU01]. One can sample the probability distribution of Y according to its density p_Y. This sampling generates a collection of vectors $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_m$, where $\mathbf{y}_i \in \mathbb{R}^n$, n is the number of securities and m is the number of scenarios. There are several ways to generate scenarios. One approach is to generate scenarios using the Brownian Motion. The approach that we will use in this thesis is the one of using historical returns.

Let $\mathbf{p} \in \mathbb{R}^m$ be the column vector containing the probability of the scenarios, i.e p_i is the probability that scenario i will occur. Then we can approximate $F_\alpha(\mathbf{x}, \xi)$ by

$$\hat{F}_\alpha(\mathbf{x}, \xi) = \xi + (1-\alpha)^{-1} \sum_{i=1}^{m} p_i [f(\mathbf{x}, \mathbf{y}_i) - \xi]^+$$

If the loss function $f(\mathbf{x}, \mathbf{y}_i)$ is linear with respect to $\mathbf{x}$, then $\hat{F}_\alpha(\mathbf{x}, \xi)$ is convex and piecewise linear. Moreover, we can replace $\hat{F}_\alpha(\mathbf{x}, \xi)$ by a linear function. Let $\mathbf{z} \in \mathbb{R}^m$ be a column vector. Then we can replace $\hat{F}_\alpha(\mathbf{x}, \xi)$ by the following linear function and set of linear constraints

$$\begin{aligned} &\xi + (1-\alpha)^{-1} \mathbf{p}^T \mathbf{z} \\ &z_i \geq f(\mathbf{x}, \mathbf{y}_i) - \xi \quad i = 1, 2, \ldots, m \\ &\mathbf{z} \geq 0 \end{aligned} \tag{4.5}$$

4.4 Constraints

Because of real-world requirements and concepts, we introduce some additional constraints that should be considered. For instance, buying or selling an instrument often incurs a transaction cost, or short sales may not be allowed. Therefore, in this section, we will state some constraints that are reasonable to use. Throughout this section, n is the number of securities and m is the number of scenarios. Budget constraint: $\mathbf{e}^T \mathbf{x} = 1, \mathbf{e} = (1, 1, \ldots, 1)^T$

This section is based on the paper by Krokhmal et. al. [KPU01].

Transaction costs

we will assume, when transaction costs are considered, that they are linear and proportional to the total value of the bought/sold asset. Let $\mathbf{c} \in \mathbb{R}^n$ be the column vector of transaction costs, so that c_j is the cost of selling or buying security j . Let $c_0 = 0$ be the cost of moving money in or out of the bank. Let $\mathbf{x}_0 \in \mathbb{R}^n$ be the initial portfolio, $\mathbf{q} = (q_1, q_2, \ldots, q_n)^T$ be the initial prices, and $V_0 = \mathbf{q}^T \mathbf{x}^0$ be the initial value.

$$V_0 = \sum_{j=1}^{n} c_j q_j |x_j^0 - x_j| + \mathbf{q}^T \mathbf{x}$$

This equality can be reformulated using the following set of linear constraints:

$$V_0 = \sum_{j=1}^{n} c_j q_j (u_j^+ + u_j^-) + \mathbf{q}^T \mathbf{x}$$

$$x_j - x_j^0 = u_j^+ - u_j^- \text{ for } j = 1, \ldots, n$$

$$u_j^+, u_j^- \geq 0 \text{ for } j = 1, \ldots, n$$

CVaR constraints

Current regulations impose capital requirements on investment companies, proportional to VaR. These requirements can be met by constraining portfolio CVaR at different confidence levels, since one always have that α-CVaR$\geq$ α-VaR, for all α. Such constraints may look like this

$$\phi_\alpha(\mathbf{x}) \leq \nu V_0$$

where $\phi_\alpha(\mathbf{x})$ is the α-CVaR for the loss function, $V_0 = \mathbf{q}^T\mathbf{x}$ is the initial portfolio value, and ν is a percentage of the initial portfolio value allowed for risk exposure.

In the linearized case, the CVaR constraints will be as follows:

$$\begin{aligned} \xi + (1-\alpha)^{-1}\mathbf{p}^T\mathbf{z} &\leq \nu V_0 \\ z_i &\geq f(\mathbf{x}, \mathbf{y}_i) - \xi \qquad i = 1, 2, \ldots, m \\ \mathbf{z} &\geq 0 \end{aligned} \tag{4.6}$$

This constraint makes sure that an investor doesn't lose more than he or she can handle.

Change in individual positions and bounds on positions

Sometimes position changes are bounded. This reflects as limited liquidity of securities in the portfolio. Such a bound can be a fixed number or be proportional to the initial position in the security

$$0 \leq u_j^+ \leq \overline{u}_j^+,\ 0 \leq u_j^- \leq \underline{u_j^-}, j = 1, \ldots, n$$

One might also want to use bounds on the positions:

$$\underline{x}_j \leq x_j \leq \overline{x}_j$$

for $j = 1, 2, \ldots, n$. A typical example of a constraint of this type is the 'no short sales' constraint $\mathbf{x} \geq \mathbf{0}$.

Value constraints

Sometimes we do not allow a security, j, to constitute more than a given percentage, v_j, of the total portfolio value. Let $\mathbf{q} \in \mathbb{R}^n$ be the initial portfolio prices. Then the value constraint can be formulated as follows:

$$q_j x_j \leq v_j(\mathbf{q}^T \mathbf{x})$$

This constraint is only applicable when short positions are not allowed.

Admissibility constraint

In real life finance there must be a limit to how much debt an investor can tolerate. So there exists a $K = K(\mathbf{x}) < \infty$, such that

$$V(t, \omega) \geq -K$$

Assume we have generated scenarios $Y = (\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_m)^T$. Then the admissibility constraint will be as follows:

$$\mathbf{y}_i^T \mathbf{x} \geq -K$$

for $i = 1, \ldots, m$. This constraint is only applicable when short sales are not allowed. If no money is borrowed, the investor is ensured no debt.

5 | The model and Fourier – Motzkin elimination

This chapter contains a description of an one period portfolio optimization model with linearized CVaR constraints. Also, we will perform Fourier-Motzkin elimination on the system of linear constraints for this problem. The models presented in this chapter will be the main models for examples and production of efficient frontiers.

5.1 The one period optimization model

This section is based on the article by Krokhmal et. al. [KPU01]. In this section we will study an one period portfolio optimization model. In such a model we consider only two dates; t_0 and t_1, the start and the end of the period, respectively. At time t_0, we know the prices in the market, but we don't know the state of the world at time t_1. The t_1- prices are scenario-dependent. Throughout this chapter, m will be the number of scenarios and n will be the number of securities.

Consider the linear system (4.5) in Section 4.3. Substituting these into problem (P2) makes it a LP problem, and easy to solve using the Simplex algorithm.

Let $\mathbf{x} \in \mathbb{R}^n$ be the portfolio and play the role of the decision variable. As oppose to earlier discussions, we will now let the portfolio $\mathbf{x}$ be the proportion of the initial wealth to be invested. For this purpose we will add the constraints $\mathbf{e}^T\mathbf{x} = 1$, initial portfolio value $V_0 = 1$, and no short sales, i.e $\mathbf{x} \geq \mathbf{0}$. This approach corresponds to Rockafellar and Uryasev [RU99], and to the Markowitz Mean-Variance model in Section 4.1. Because of this interpretation of $\mathbf{x}$, the CVaR constraint should be interpreted as the expected

loss percentage, given this percentage is exceeding ξ. If one is interested in finding the numbers of units that should be invested in each security, one can calculated for each security j: $V_0 x_j/q_j^0$, where q_j^0 is the t_0-price of security j and V_0 is the value to be invested (in this case V_0 is not equal to 1). This number is usually not an integer, and in this case we should round down to nearest integer.

Let the matrix of returns be denoted by $A \in \mathbb{R}^{m\times n}$. The elements of this matrix is given by

$$a_{ij} = (q_{ij}^1 - q_i^0)/q_j^0 = q_{ij}^1/q_j^0 - 1$$

where $\mathbf{q}^0 = (q_1^0, q_2^0, \ldots, q_n^0)$ is the vector of initial prices, and $\mathbf{q}^1 \in \mathbb{R}^{m\times n}$ is the matrix of t_1-prices. The number a_{ij} should be interpreted as the percentage of return from security j under scenario i. Defining A in this way makes sense. If the t_1-price q_{ij}^1 is higher than the t_0-price q_j^0, then a_{ij} is positive, corresponding to gain. If the t_1-price q_{ij}^1 is lower than the t_0-price q_j^0, then a_{ij} is negative and corresponds to loss. $A\mathbf{x}$ is the percentage of return of portfolio $\mathbf{x}$.

Let $\mathbf{p} \in \mathbb{R}^m$ be the vector of scenario probability. Also, for simplicity, let $(1-\alpha)^{-1}\mathbf{p} = \boldsymbol{\beta}$. The reward function I will use is $R(\mathbf{x}) = A\mathbf{x}$. Then the expected return is $\mathbf{p}^T A\mathbf{x}$, and the loss is given by $-A\mathbf{x}$. As discussed in Section 4.3, when linearizing F_α, we introduce a vector $\mathbf{z} = (z_1, \ldots, z_m)^T$, where z_i denotes the loss that is exceeding the threshold ξ. $\nu \in [0,1]$ will as before be the level of risk.

The one period portfolio optimization model with CVaR constraints is as follows:

$$\begin{array}{rrcrcrcl} \max\limits_{\xi,\mathbf{z},\mathbf{x}} & & & & & \mathbf{p}^T A\mathbf{x} & & \\ \text{subject to} & & \xi & + & \boldsymbol{\beta}^T\mathbf{z} & & \leq & \nu \\ & - & \xi\mathbf{e} & - & \mathbf{z} & - & A\mathbf{x} & \leq & \mathbf{0} \\ & & & & & & \mathbf{e}^T\mathbf{x} & = & 1 \\ & & & & \mathbf{z}, & & \mathbf{x} & \geq & \mathbf{0} \end{array} \tag{5.1}$$

Problem (5.1) has feasible solutions, when the risk level ν is not very small, see Krokhmal et. al. [KPU01]. Let for instance $\mathbf{z} = \mathbf{0}$. Let also $x_1 = 1$ and $x_j = 0$ for $j = 2, \ldots, n$. Then we can choose ξ to be less than the level of

risk and larger than the loss of security 1 under each scenario, i.e

$$\max_{1\leq i\leq m}\{-a_{i1}\}\leq \xi\leq \nu$$

5.2 Fourier-Motzkin elimination

The goal in this section is to perform Fourier-Motzkin eliminations (FM eliminations) on the system of linear constraints in the one period model (5.1) from Section 5.1, to simplify this optimization model. We have seen that the number of decision variables in this model equals $m+n+1$, where n is the number of securities and m is the number of possible scenarios. I will in this section, under some conditions, find an equivalent linear system with the portfolio as the only decision variable.

Let $A\in\mathbb{R}^{m\times n}$ be the matrix of possible t_1- market returns, as defined in Section 5.1. Denote by $\mathbf{a}_i$ the i'th row of A. Let $\mathbf{p}=(p_1,p_2,\ldots,p_m)^T$, where p_i is the probability of scenario i, and let $\boldsymbol{\beta}=(1-\alpha)^{-1}\mathbf{p}$. Let $\mathbf{x}=(x_1,x_2,\ldots,x_n)^T$ represent the portfolio interpreted as in Section 5.1, and let $\mathbf{z}=(z_1,z_2,\ldots,z_m)^T$ represent the function $[f(\mathbf{x},\mathbf{a}_j)-\xi]^+$, the loss exceeding threshold ξ. $\mathbf{e}$ will denote the column vector of length m consisting entirely of ones, and $\nu\in[0,1]$ will denote the risk level.

Recall problem (5.1):

$$\begin{array}{lrcrcrcl}
\max\limits_{\xi,\mathbf{z},\mathbf{x}} & & & \mathbf{p}^TA\mathbf{x} & & & & \\
\text{subject to} & \xi & + & \boldsymbol{\beta}^T\mathbf{z} & & & \leq & \nu \\
& -\xi\mathbf{e} & - & \mathbf{z} & - & A\mathbf{x} & \leq & \mathbf{0} \\
& & & & & \mathbf{e}^T\mathbf{x} & = & 1 \\
& & & \mathbf{z}, & & \mathbf{x} & \geq & \mathbf{0}
\end{array}$$

I will now, for the purpose of performing FM eliminations, only consider the system of constraints. I start by eliminating ξ:

$$\begin{array}{ll}
\xi & \leq \nu-\boldsymbol{\beta}^T\mathbf{z} \\
\xi\mathbf{e} & \geq -\mathbf{z}-A\mathbf{x}
\end{array}$$

I now have one upper bound for ξ and m lower bounds. Combining these will result in m new inequalities containing ξ. After some simple rewriting, these new inequalities are as follows:

$$-\mathbf{z} + (\boldsymbol{\beta}^T\mathbf{z})\mathbf{e} - A\mathbf{x} \leq \nu\mathbf{e}$$

Eliminating ξ has resulted in reducing the number of inequalities by one, and the new system is as follows:

$$\begin{array}{rcccccl} - & \mathbf{z} & + & (\boldsymbol{\beta}^T\mathbf{z})\mathbf{e} & - & A\mathbf{x} & \leq \nu\mathbf{e} \\ & & & & & \mathbf{e}^T\mathbf{x} & = 1 \\ & \mathbf{z}, & & & & \mathbf{x} & \geq \mathbf{0} \end{array}$$

$$\max_{1\leq i\leq m}\{-z_i - \mathbf{a}_i\mathbf{x}\} \leq \xi \leq \nu - \boldsymbol{\beta}^T\mathbf{z}$$

Note that $\mathbf{a}_i$ is a row vector, so the product $\mathbf{a}_i\mathbf{x}$ makes sense. I will now continue and eliminate $\mathbf{z}$. For this purpose, I will write the system of those inequalities containing $\mathbf{z}$ on component form.

$$\begin{array}{rrcrcrcl} (\beta_1-1)z_1 + & \beta_2 z_2 + & \cdots + & \beta_m z_m - & \mathbf{a}_1\mathbf{x} \leq \nu \\ \beta_1 z_1 + & (\beta_2-1)z_2 + & \cdots + & \beta_m z_m - & \mathbf{a}_2\mathbf{x} \leq \nu \\ & & \ddots & & \\ \beta_1 z_1 + & \beta_2 z_2 + & \cdots + & (\beta_m-1)z_m - & \mathbf{a}_m\mathbf{x} \leq \nu \\ z_1, & z_2, & \ldots, & z_m & \geq 0 \end{array} \tag{5.2}$$

I start by eliminating z_1. From system (5.2) one can see that in the first inequality, z_1 has coefficient $\beta_1 - 1$, and in the remaining inequalities z_1 has coefficient β_1. Recall that $\beta_1 = p_1(1-\alpha)^{-1} \geq 0$, but $\beta_1 - 1$ may not be positive. If $\beta_1 - 1$ were to be negative, then dividing by it on both sides would change the inequality sign. So one wants to know when this is the case. $\beta_1 - 1 \geq 0 \iff p_1 \geq 1 - \alpha$.

The following theorem will be interpreted after the proof.

Theorem 5.2.1. *Assume $p_i \geq 1-\alpha$ for all $i = 1, 2\ldots, m$. Then eliminating $\mathbf{z}$ from system (5.2), by performing FM eliminations, will result in the following, equivalent system:*

$$\begin{array}{rr} A\mathbf{x} \geq & -\nu\mathbf{e} \\ \mathbf{e}^T\mathbf{x} = & 1 \\ \mathbf{x} \geq & \mathbf{0} \end{array} \tag{5.3}$$

with

$$\begin{array}{ccccc} \max\limits_{1\le i\le m}\{-z_i-\mathbf{a}_i\mathbf{x}\} & \le & \xi & \le & \nu-\boldsymbol{\beta}^T\mathbf{z} \\ 0 & \le & z_1 & \le & u_1(z_2,\ldots,z_m,x_1,\ldots,x_n) \\ 0 & \le & z_2 & \le & u_2(z_3,\ldots,z_m,x_1,\ldots,x_n) \\ & & \vdots & & \\ 0 & \le & z_m & \le & u_m(x_1,\ldots,x_n) \end{array}$$

where u_i is the smallest upper bound for z_i, $i=1,2,\ldots,m$, and is given by

$$\begin{aligned} u_i = \min\Big\{ & \min_{1\le j\le i-1}(\nu-\sum_{k=i+1}^{m}\beta_k z_k+\mathbf{a}_j\mathbf{x})\beta_i^{-1}, \\ & (\nu-\sum_{k=i+1}^{m}\beta_k z_k+\mathbf{a}_i\mathbf{x})(\beta_i-1)^{-1}, \\ & \min_{i+1\le j\le m}(\nu-\sum_{k=i+1}^{m}\beta_k z_k+\mathbf{a}_j\mathbf{x}+z_j)\beta_i^{-1}\Big\} \end{aligned} \tag{5.4}$$

Proof. I will prove this theorem by using induction on k. The induction hypothesis will be that using FM elimination to eliminate z_k, $k\in\{1,2,\ldots,m\}$, will result in just 'eliminating' the column containing z_k from system (5.2). If this hypothesis turns out to be true, then eliminating $z_1,z_2\ldots,z_m$ will give us the result of the theorem. I start by eliminating z_1. In doing so, I will use the following inequalities:

$$\begin{array}{ccccccccc} z_1 & \le & (\nu & - & \sum_{i=2}^m\beta_i z_i & & & + & \mathbf{a}_1\mathbf{x})(\beta_1-1)^{-1} \\ z_1 & \le & (\nu & - & \sum_{i=2}^m\beta_i z_i & + & z_2 & + & \mathbf{a}_2\mathbf{x})\beta_1^{-1} \\ & \vdots & & & & & & & \\ z_1 & \le & (\nu & - & \sum_{i=2}^m\beta_i z_i & + & z_m & + & \mathbf{a}_m\mathbf{x})\beta_1^{-1} \\ z_1 & \ge & 0 & & & & & & \end{array}$$

Combining the upper bounds with the lower bound reduces the number of inequalities with one.

$$\begin{array}{ccccccccc} 0 & \le & (\nu & - & \sum_{i=2}^m\beta_i z_i & + & & & \mathbf{a}_1\mathbf{x})(\beta_1-1)^{-1} \\ 0 & \le & (\nu & - & \sum_{i=2}^m\beta_i z_i & + & z_2 & + & \mathbf{a}_2\mathbf{x})\beta_1^{-1} \\ & \vdots & & & & & & & \\ 0 & \le & (\nu & - & \sum_{i=2}^m\beta_i z_i & + & z_m & + & \mathbf{a}_m\mathbf{x})\beta_1^{-1} \end{array}$$

$$0\le z_1\le\min\Big\{(\nu-\sum_{i=2}^{m}\beta_i z_i+\mathbf{a}_1\mathbf{x})(\beta_1-1)^{-1},\min_{2\le j\le m}(\nu-\sum_{i=2}^{m}\beta_i z_i+z_j+\mathbf{a}_j\mathbf{x})\beta_1^{-1}\Big\}$$

Since both β_1 and $\beta_1 - 1$ are nonnegative[1], one can multiply the first inequality by $\beta_1 - 1$ and the rest by β_1, without changing the inequality sign.

$$\begin{array}{rcl} 0 & \leq & \nu - \sum_{i=2}^{m} \beta_i z_i + \mathbf{a}_1\mathbf{x} \\ 0 & \leq & \nu - \sum_{i=2}^{m} \beta_i z_i + z_2 + \mathbf{a}_2\mathbf{x} \\ & \vdots & \\ 0 & \leq & \nu - \sum_{i=2}^{m} \beta_i z_i + z_m + \mathbf{a}_m\mathbf{x} \end{array}$$

The new system is now as follows:

$$\begin{array}{rcccccl} \beta_2 z_2 & + & \cdots & + & \beta_m z_m & - \mathbf{a}_1\mathbf{x} & \leq \nu \\ (\beta_2 - 1) z_2 & + & \cdots & + & \beta_m z_m & - \mathbf{a}_2\mathbf{x} & \leq \nu \\ & & \ddots & & & & \\ \beta_2 z_2 & + & \cdots & + & (\beta_m - 1) z_m & - \mathbf{a}_m\mathbf{x} & \leq \nu \\ z_2, & & \ldots, & & z_m & & \geq 0 \end{array}$$

with

$$0 \leq z_1 \leq \min\left\{ (\nu - \sum_{i=2}^{m} \beta_i z_i + \mathbf{a}_1\mathbf{x})(\beta_1 - 1^{-1}), \min_{2 \leq j \leq m} (\nu - \sum_{i=2}^{m} \beta_i z_i + z_j + \mathbf{a}_j\mathbf{x})\beta_1^{-1} \right\}$$

Comparing this new system to system (5.2), it seems like we have just eliminated the z_1-column. Now, assume we have eliminated $z_1, \ldots, z_{k-1}$, and that the corresponding columns have been deleted. I will now delete z_k from our current system, which is as follows:

$$\begin{array}{rcccccccl} \beta_k z_k & + & \beta_{k+1} z_{k+1} & + & \cdots & + & \beta_m z_m & - \mathbf{a}_1\mathbf{x} & \leq \nu \\ & & & & \vdots & & & & \\ \beta_k z_k & + & \beta_{k+1} z_{k+1} & + & \cdots & + & \beta_m z_m & - \mathbf{a}_{k-1}\mathbf{x} & \leq \nu \\ (\beta_k - 1) z_k & + & \beta_{k+1} z_{k+1} & + & \cdots & + & \beta_m z_m & - \mathbf{a}_k\mathbf{x} & \leq \nu \\ \beta_k z_k & + & (\beta_{k+1} - 1) z_{k+1} & + & \cdots & + & \beta_m z_m & - \mathbf{a}_{k+1}\mathbf{x} & \leq \nu \\ & & & & \ddots & & & & \\ \beta_k z_k & + & \beta_{k+1} z_{k+1} & + & \cdots & + & (\beta_m - 1) z_m & - \mathbf{a}_m\mathbf{x} & \leq \nu \\ z_k, & & z_{k+1}, & & \ldots, & & z_m & & \geq 0 \end{array}$$

[1]In fact, $\beta_1 > 0$. If β_1 was equal to zero, this would mean that $p_1 = 0$. Scenarios with probability equal to zero should not be considered, as they represent impossible scenarios. If $\beta_1 - 1 = 0$, then the first inequality should be deleted.

with $0 \le z_i \le u_i(z_{i+1}, \ldots, z_m, x_1, \ldots, x_n)$ for $i = 1, \ldots, k-1$. Here u_i is the smallest lower bound on z_i. Since β_k and $\beta_k - 1$ both are nonnegative, dividing by them doesn't change the inequality sign.

$$\begin{array}{rcl}
z_k & \le & (\nu - \sum_{i=k+1}^{m} \beta_i z_i + \mathbf{a}_1\mathbf{x})\beta_k^{-1} \\
 & \vdots & \\
z_k & \le & (\nu - \sum_{i=k+1}^{m} \beta_i z_i + \mathbf{a}_{k-1}\mathbf{x})\beta_k^{-1} \\
z_k & \le & (\nu - \sum_{i=k+1}^{m} \beta_i z_i + \mathbf{a}_k\mathbf{x})(\beta_k - 1)^{-1} \\
z_k & \le & (\nu - \sum_{i=k+1}^{m} \beta_i z_i + z_{k+1} + \mathbf{a}_{k+1}\mathbf{x})\beta_k^{-1} \\
 & \vdots & \\
z_k & \le & (\nu - \sum_{i=k+1}^{m} \beta_i z_i + z_m + \mathbf{a}_m\mathbf{x})\beta_k^{-1} \\
z_k & \ge & 0
\end{array}$$

Combining these upper bounds with the lower bound reduces the number of inequalities by one.

$$\begin{array}{rcl}
0 & \le & (\nu - \sum_{i=k+1}^{m} \beta_i z_i + \mathbf{a}_1\mathbf{x})\beta_k^{-1} \\
 & \vdots & \\
0 & \le & (\nu - \sum_{i=k+1}^{m} \beta_i z_i + \mathbf{a}_{k-1}\mathbf{x})\beta_k^{-1} \\
0 & \le & (\nu - \sum_{i=k+1}^{m} \beta_i z_i + \mathbf{a}_k\mathbf{x})(\beta_k - 1)^{-1} \\
0 & \le & (\nu - \sum_{i=k+1}^{m} \beta_i z_i + z_{k+1} + \mathbf{a}_{k+1}\mathbf{x})\beta_k^{-1} \\
 & \vdots & \\
0 & \le & (\nu - \sum_{i=k+1}^{m} \beta_i z_i + z_m + \mathbf{a}_m\mathbf{x}\beta_k^{-1}
\end{array}$$

$$0 \le z_k \le u_k(z_{k+1}, \ldots, x_1, \ldots, x_n)$$

where

$$\begin{aligned}
u_k = \min\Big\{ & \min_{1 \le j \le k-1} (\nu - \sum_{i=k+1}^{m} \beta_i z_i + \mathbf{a}_j\mathbf{x})\beta_k^{-1}, \\
& (\nu - \sum_{i=k+1}^{m} \beta_i z_i + \mathbf{a}_k\mathbf{x})(\beta_k - 1)^{-1}, \\
& \min_{k+1 \le j \le m} (\nu - \sum_{i=k+1}^{m} \beta_i z_i + \mathbf{a}_j\mathbf{x} + z_j)\beta_k^{-1} \Big\}
\end{aligned}$$

The expression for u_k can actually be used to calculate u_i for all $i = 1, \ldots, k$. Continue by multiplying row k by $\beta_k - 1$ and the rest by β_k.

$$\begin{array}{ccccccccc}
0 & \leq & \nu & - & \sum_{i=k+1}^{m} \beta_i z_i & + & & & \mathbf{a}_1\mathbf{x} \\
 & \vdots & & & & & & & \\
0 & \leq & \nu & - & \sum_{i=k+1}^{m} \beta_i z_i & + & & & \mathbf{a}_k\mathbf{x} \\
0 & \leq & \nu & - & \sum_{i=k+1}^{m} \beta_i z_i & + & z_{k+1} & + & \mathbf{a}_{k+1}\mathbf{x} \\
 & \vdots & & & & & & & \\
0 & \leq & \nu & - & \sum_{i=k+1}^{m} \beta_i z_i & + & z_m & + & \mathbf{a}_m\mathbf{x}
\end{array}$$

The new system is as follows:

$$\begin{array}{cccccccccc}
\beta_{k+1} z_{k+1} & + & \beta_{k+2} z_{k+2} & + & \cdots & + & \beta_m z_m & - & \mathbf{a}_1\mathbf{x} & \leq \nu \\
 & & & & \vdots & & & & & \\
\beta_{k+1} z_{k+1} & + & \beta_{k+2} z_{k+2} & + & \cdots & + & \beta_m z_m & - & \mathbf{a}_k\mathbf{x} & \leq \nu \\
(\beta_{k+1} - 1) z_{k+1} & + & \beta_{k+2} z_{k+2} & + & \cdots & + & \beta_m z_m & - & \mathbf{a}_{k+1}\mathbf{x} & \leq \nu \\
\beta_{k+1} z_{k+1} & + & (\beta_{k+2} - 1) z_{k+2} & + & \cdots & + & \beta_m z_m & - & \mathbf{a}_{k+2}\mathbf{x} & \leq \nu \\
 & & & & \ddots & & & & & \\
\beta_{k+1} z_{k+1} & + & \beta_{k+2} z_{k+2} & + & \cdots & + & (\beta_m - 1) z_m & - & \mathbf{a}_m\mathbf{x} & \leq \nu \\
z_{k+1}, & & z_{k+2}, & & \ldots, & & z_m & & & \geq 0
\end{array}$$

with

$$\begin{array}{lcccl}
\max_i\{-z_i - \mathbf{a_i x}\} & \leq & \xi & \leq & \nu - \boldsymbol{\beta}^T \mathbf{z} \\
0 & \leq & z_1 & \leq & u_1(x_1 \ldots, x_n, z_2, \ldots, z_m) \\
0 & \leq & z_2 & \leq & u_2(x_1, \ldots, x_n, z_3 \ldots, z_m) \\
 & & \vdots & & \\
0 & \leq & z_k & \leq & u_k(x_1, \ldots, x_n, z_{k+1}, \ldots, z_m)
\end{array}$$

Here the upper bounds u_i are given by (5.4). The elimination of z_k has resulted in deleting the column containing z_k, and the induction hypothesis is proved. Eliminating $z_1, z_2, \ldots, z_m$ gives the system stated in the theorem. ■

Theorem 5.2.1 says that if $p_i \geq 1-\alpha$ for $i = 1, \ldots, m$, then we can easily eliminate ξ and $\mathbf{z}$ such that the decision variable is only the portfolio. Also, the number of constraints will be reduced from $2(m+1)+n$ to $m+n+1$. The constraint $A\mathbf{x} \geq -\nu\mathbf{e}$ says that the loss percentage under each scenario must be less than or equal to the percentage of the initial portfolio value that is put up for risk exposure. The fact that one must have $p_i \geq 1-\alpha$ implies that the number of scenarios m is bounded, in fact $m \leq (1-\alpha)^{-1}$. If for instance $\alpha = 0.9$, then we must have $p_i \geq 0.1$ for all i, and the number

of scenarios must be less than or equal to 10. If there are $(1-\alpha)^{-1}$ scenarios with $p_i = 1-\alpha$ for all i, then $\beta_i - 1 = 0$ for all i. This means that m of the coefficients in system (5.2) will be zero. If we have $\alpha = 0.99$ then we must have that $p_i \geq 0.01$ in order to apply Theorem 5.2.1. In this case, the number of scenarios is bounded by 100, which also is a low number of scenarios. In the next sections we will allow some of the probabilities p_i to be less than $1-\alpha$, which implies that we can have a larger number of scenarios. This number will not increase by much, though.

Problem (5.1) is equivalent to the problem

$$\begin{array}{lrcl} \max\limits_{\mathbf{x}} & \mathbf{p}^T A\mathbf{x} & & \\ \text{subject to} & A\mathbf{x} & \geq & -\nu\mathbf{e} \\ & \mathbf{e}^T\mathbf{x} & = & 1 \\ & \mathbf{x} & \geq & \mathbf{0} \end{array} \tag{5.5}$$

with ξ and $\mathbf{z}$ chosen according to

$$\begin{array}{lll} l_\xi \leq & \xi & \leq u_\xi \\ \mathbf{0} \leq & \mathbf{z} & \leq \mathbf{u} \end{array}$$

where lower and upper bounds on ξ and $\mathbf{z}$ are given as in Theorem 5.2.1. As discussed earlier, one often wants to add additional constraints to a optimization problem. Since these constraints often are constraints on the portfolio, these can easily be added to problem (5.5). Once the solution to problem (5.5) has been found, one can find the optimal objective value since this does not depend on ξ and $\mathbf{z}$. For computing the corresponding α-CVaR value, one can for instance calculate ξ and $\mathbf{z}$ using the upper and lower bounds in Theorem 5.2.1. Using this approach, one gets a lower and a upper bound on the α-CVaR value. Another alternative for computing the corresponding α-CVaR value it to compute it using its definition. In order to use the definition, one must first compute α-VaR, but this is not difficult.

I will now give an example to illustrate Theorem 5.2.1.

Example 5.2.1. Assume we want to solve optimization problem (5.1) when we have a market consisting of two securities, and there are two possible scenarios, i.e. $m = n = 2$. The bank position is not included in the portfolio. Let $\mathbf{p} = [0.5 \;\; 0.5]^T$, $\alpha = 0.95$ and $\nu = 0.1$. Let the return matrix be

$$A = \begin{bmatrix} 0.13 & -0.11 \\ -0.12 & 0.09 \end{bmatrix}$$

Conditions for applying Theorem 5.2.1 are satisfied, since $p_1, p_2 \geq 0.05$. The linear system can be written as follows:

$$\begin{bmatrix} -0.13 & 0.11 \\ 0.12 & -0.09 \\ 1 & 1 \\ -1 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \leq \begin{bmatrix} 0.1 \\ 0.1 \\ 1 \\ -1 \end{bmatrix}$$

with $x_1, x_2 \geq 0$. Since $\mathbf{x} \in \mathbb{R}^2$, we can illustrate this problem in the plane.

From Figure 5.1 one can see that the feasible set lies on the line $x_1+x_2 = 1$. It is in fact the line segment between the points $(0.905, 0.095)$ and $(0.042, 0.958)$ (thick line). This means that the optimal solution has to be either the point $(0.905, 0.095)$, the point $(0.042, 0.958)$, or the whole line segment. In our case $\mathbf{p} = [0.5\ 0.5]^T$, which implies that $\mathbf{p}^T A\mathbf{x} = 0.005x_1 - 0.01x_2$ (dashed line). So the optimal solution is the point $(0.905, 0.095)$. This means that an investor should invest $90,5\%$ of his or her wealth in security 1 and $9,5\%$ in security 2. The expected percentage return of this portfolio is 0.36%.

To find the corresponding 0.95-CVaR value, I compute ξ and $\mathbf{z}$. Using back substitution and the expression for upper bounds (5.4) in Theorem 5.2.1 gives:

$$\begin{aligned} u_2 &= \min\{(\nu + \mathbf{a}_1\mathbf{x})\beta_2^{-1}, (\nu + \mathbf{a}_2\mathbf{x})(\beta_2 - 1)^{-1}\} \\ &= \min\{0.0207, 0\} \\ &= 0 \end{aligned}$$

Since $0 \leq z_2 \leq u_2 = 0$, we have that $z_2 = 0$.

$$\begin{aligned} u_1 &= \min\{(\nu - \beta_2 z_2 + \mathbf{a}_1\mathbf{x})(\beta_1 - 1)^{-1}, (\nu - \beta_2 z_2 + \mathbf{a}_2\mathbf{x} + z_2)\beta_1^{-1}\} \\ &= \min\{0.023, 0\} \\ &= 0 \end{aligned}$$

So, also $z_1 = 0$.

$$\begin{aligned} l_\xi &= \max\{-z_1 - \mathbf{a}_1\mathbf{x}, -z_2 - \mathbf{a}_2\mathbf{x}\} \\ &= \max\{-0.1071, 0.1\} \\ &= 0.1 \end{aligned}$$

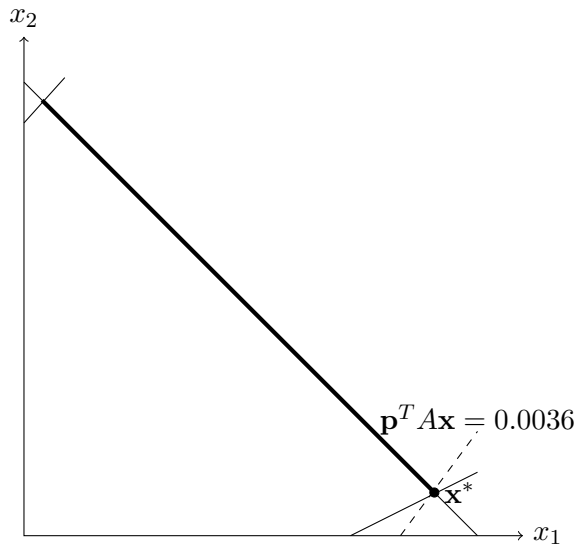


Figure 5.1: The feasible set of the problem in Example 5.2.1

.

Since $\xi \leq \nu - \boldsymbol{\beta}^T \mathbf{z} = \nu$, we have that $\xi = 0.1$. Now we can calculate 0.95-CVaR

$$\phi_{0.95} = \xi + \boldsymbol{\beta}^T \mathbf{z} = \xi$$

This means that $\phi_{0.95} = 0.1$. (To see the efficient frontier for this example, see Figure 6.3).

□

As Example 5.2.1 illustrates, the optimization model in Theorem 5.2.1 is good for finding optimal expected returns, but computing the CVaR takes more effort, when computing it using back substitution. Once the optimal solution $\mathbf{x}^*$ has been found, one continues by finding an interval for z_m. Then, one must choose a value for z_m from this interval. In the example, we were lucky since this interval happened to be a single point. Once z_m is determined, one continues with this back substitution. Since CVaR is linear and a sum of positive numbers, one can when choosing z_i's and ξ from intervals, choose in two separate cases the lowest and highest value, respectively. Then, for a given level of risk, one would get an interval containing CVaR. Computing α-CVaR using its definition might be easier, as one only needs to compute the upper percentile α-VaR, and find the expectation of the losses exceeding this value.

5.3 Fourier-Motzkin elimination with one negative coefficient

I will in this section study what happens when we eliminate $\mathbf{z}$ from system (5.2), when for some $k \in \{1, 2, \ldots, m\}$, $\beta_k - 1 < 0$ and $\beta_i - 1 \geq 0$ for all $i \neq k$. Since we can eliminate $z_1, z_2 \ldots, z_m$ in any order we like, I will assume that $k = m$.[2] When re-indexing so that $k = m$, since all coefficients $\beta_i - 1 \geq 0$ for $i = 1, \ldots, m-1$, we can easily eliminate all variables $z_1, \ldots, z_{m-1}$, see proof of Theorem 5.2.1. After eliminating $z_1, \ldots, z_{m-1}$ we are left with the following system:

$$\begin{array}{rcrcl} \beta_m z_m & - & \mathbf{a}_1\mathbf{x} & \leq & \nu \\ & & \vdots & & \\ \beta_m z_m & - & \mathbf{a}_{m-1}\mathbf{x} & \leq & \nu \\ (\beta_m - 1) z_m & - & \mathbf{a}_m\mathbf{x} & \leq & \nu \\ z_m & & & \geq & 0 \end{array} \tag{5.6}$$

with

$$\begin{array}{lcccl} \max_i\{-z_i - \mathbf{a}_i\mathbf{x}\} & \leq & \xi & \leq & \nu - \boldsymbol{\beta}^T\mathbf{z} \\ 0 & \leq & z_1 & \leq & u_1(z_2, \ldots, z_m x_1, \ldots, x_n) \\ & & \vdots & & \\ 0 & \leq & z_{m-1} & \leq & u_{m-1}(z_m, x_1, \ldots, x_n) \end{array}$$

where u_i are upper bounds given by (5.4) in Theorem 5.2.1. In system (5.6) the number of inequalities is $m + 1$. We are now ready to eliminate z_m

$$\begin{array}{rcrcl} z_m & \leq & (\nu & + & \mathbf{a}_1\mathbf{x})\beta_m^{-1} \\ & \vdots & & & \\ z_m & \leq & (\nu & + & \mathbf{a}_{m-1}\mathbf{x})\beta_m^{-1} \\ z_m & \geq & (\nu & + & \mathbf{a}_m\mathbf{x})(\beta_m - 1)^{-1} \\ z_m & \geq & 0 & & \end{array} \tag{5.7}$$

Note that in the m'th inequality, the inequality sign is changed since $\beta_m - 1$ is negative. We now have two lower bounds and $m - 1$ upper bounds, which

[2] One could also re-index so that $k = 1$. This will result in a bit more complicated systems, but it simplifies the work when one have several small p_i's

will result in $2(m-1)$ new inequalities (containing z_m). This means that the number of inequalities has increased by $m-3$. Combining the lower and the upper bounds gives the following inequalities:

$$\begin{array}{rcl} 0 & \leq & (\nu + \mathbf{a}_1\mathbf{x})\beta_m^{-1} \\ & \vdots & \\ 0 & \leq & (\nu + \mathbf{a}_{m-1}\mathbf{x})\beta_m^{-1} \\ (\nu + \mathbf{a}_m\mathbf{x})(\beta_m - 1)^{-1} & \leq & (\nu + \mathbf{a}_1\mathbf{x})\beta_m^{-1} \\ & \vdots & \\ (\nu + \mathbf{a}_m\mathbf{x})(\beta_m - 1)^{-1} & \leq & (\nu + \mathbf{a}_{m-1}\mathbf{x})\beta_m^{-1} \end{array} \tag{5.8}$$

In the case where $(\nu + \mathbf{a}_m\mathbf{x})(\beta_m - 1)^{-1} \leq 0$, we can delete the $m-1$ last inequalities. This is the case when

$$\mathbf{a}_m\mathbf{x} \geq -\nu$$

i.e. when the loss percentage under scenario m is less than the risk level. I assume this is not the case, and rewrite the $m-1$ last inequalities in system (5.8). For $1 \leq j \leq m-1$:

$$(\nu + \mathbf{a}_m\mathbf{x})(\beta_m - 1)^{-1} \leq (\nu + \mathbf{a}_j\mathbf{x})\beta_m^{-1}$$

$$\nu + \mathbf{a}_m\mathbf{x} \geq (\nu + \mathbf{a}_j\mathbf{x})\beta_m^{-1}(\beta_m - 1)$$

$$\nu + \mathbf{a}_m\mathbf{x} \geq \nu + \mathbf{a}_j\mathbf{x} - (\nu + \mathbf{a}_j\mathbf{x})\beta_m^{-1}$$

$$\beta_m\mathbf{a}_m\mathbf{x} - \beta_m\mathbf{a}_j\mathbf{x} \geq -(\nu + \mathbf{a}_j\mathbf{x})$$

$$-\beta_m\mathbf{a}_m\mathbf{x} + \beta_m\mathbf{a}_j\mathbf{x} \leq \nu + \mathbf{a}_j\mathbf{x}$$

$$-(\beta_m\mathbf{a}_m + (1-\beta_m)\mathbf{a}_j)\mathbf{x} \leq \nu$$

The new system is as follows:

$$\begin{array}{rcl} -\mathbf{a}_1\mathbf{x} & \leq & \nu \\ & \vdots & \\ -\mathbf{a}_{m-1}\mathbf{x} & \leq & \nu \\ -(\beta_m\mathbf{a}_m + (1-\beta_m)\mathbf{a}_1)\mathbf{x} & \leq & \nu \\ & \vdots & \\ -(\beta_m\mathbf{a}_m + (1-\beta_m)\mathbf{a}_{m-1})\mathbf{x} & \leq & \nu \end{array} \tag{5.9}$$

$$\max\{0, (\nu + \mathbf{a}_m\mathbf{x})(\beta_m - 1)^{-1}\} \leq z_m \leq \min_{1\leq j\leq m-1}\{(\nu + \mathbf{a}_j\mathbf{x})\beta_m^{-1}\}$$

In the first $m-1$ inequalities in system (5.9), I have used that $\beta_m \geq 0$, and multiplied by β_m on both sides of the inequality sign. These $m-1$ inequalities are equal to the $m-1$ inequalities we had in Section 5.2. The following lemma will be interpreted after its proof.

Lemma 5.3.1. *Assume $p_k < 1-\alpha$ for some $k \in \{1, 2, \ldots, m\}$ and $p_i \geq 1-\alpha$ for $i = \{1, 2, \ldots, m\}\backslash k$. Then eliminating $\mathbf{z}$ from system (5.2), by using FM elimination, will result in the following system*

$$\begin{array}{rcl} A'\mathbf{x} & \geq & -\nu\mathbf{e} \\ \beta_k A_k\mathbf{x} + (1-\beta_k)A'\mathbf{x} & \geq & -\nu\mathbf{e} \\ \mathbf{e}^T\mathbf{x} & = & 1 \\ \mathbf{x} & \geq & 0 \end{array} \tag{5.10}$$

with

$$\begin{array}{c} \max_{1\leq i\leq m}\{-z_i - \mathbf{a}_i\mathbf{x}\} \leq \xi \leq \nu - \boldsymbol{\beta}^T\mathbf{z} \\ 0 \leq z_i \leq u_i(z_{i+1}, \ldots, z_m, x_1 \ldots, x_n), \text{ for } i \neq k \\ l'_k(z_{k+1}, \ldots, z_m, x_1, \ldots, x_n) \leq z_k \leq u'_k(z_{k+1}, \ldots, z_m, x_1, \ldots, x_n) \end{array} \tag{5.11}$$

where $A' \in \mathbb{R}^{(m-1)\times n}$ is the matrix A with row k deleted, and $A_k \in \mathbb{R}^{(m-1)\times n}$ is the matrix consisting of rows equal to $\mathbf{a}_k$. The upper bounds u_i for $i \neq k$

are given by (5.4) in Theorem 5.2.1. The lower bound l'_k *and the upper bound* u'_k *are given by*

$$l'_k = \max\left\{0, (\nu - \sum_{k+1}^{m} \beta_i z_i + \mathbf{a}_k \mathbf{x})(\beta_k - 1)^{-1}\right\}$$

$$u'_k = \min\left\{\min_{1\leq j\leq k-1}(\nu - \sum_{i=k+1}^{m} \beta_i z_i + \mathbf{a}_j \mathbf{x})\beta_k^{-1}, \min_{k+1\leq j\leq m}(\nu - \sum_{i=k+1}^{m} \beta_i z_i + \mathbf{a}_j \mathbf{x} + z_j)\beta_k^{-1}\right\}$$

Proof. The system (5.10) was proved above, using Theorem 5.2.1. The lower bound l'_k follows from eliminating $z_1, \ldots, z_k$ using the proof of Theorem 5.2.1. Then 0 and $(\nu - \sum_{k+1}^{m} \beta_i z_i + \mathbf{a}_k \mathbf{x})(\beta_k - 1)^{-1}$ are the lower bounds for z_k. The upper bound u'_k follows from (5.4) and deleting $(\nu - \sum_{i=k+1}^{m} \beta_i z_i + \mathbf{a}_k \mathbf{x})(\beta_k - 1)^{-1}$, since this is now a lower bound. ■

Lemma 5.3.1 says that under scenarios $i \neq k$, the loss percentage must be less than the percentage of the initial portfolio that is up for risk exposure. In addition, an affine combination of the percentage loss under scenario k and the percentage loss under scenario i, for each $i \in \{2, 3 \ldots, m\} \backslash k$, must be less than the risk level. An *affine combination* of vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ is a vector $\sum_{i=1}^{n} a_i \mathbf{x}_i$, where $\sum_{i=1}^{n} a_i = 1$.

As in Section 5.2, constraints can easily be added to the problem of maximizing expected return subject to the constraints in Lemma 5.3.1. Assume we have found the optimal solution to this problem, then the corresponding α-CVaR value can be computed using back substitution or by using the definition.

Although we in this section have allowed one scenario probability to be less than $1 - \alpha$, the upper bound on the number of scenarios is still $(1 - \alpha)^{-1}$. Since α often is chosen to be either $0.9, 0.95$ or 0.99, the number of scenarios is still less than what is usually desire.

5.4 Fourier-Motzkin elimination with more than one negative coefficients

In this section I will study what happens when $\beta_i - 1$ in system (5.2) is negative for two indexes $k, l \in \{1, 2 \ldots, m\}$. We will see that this will result in a growth in the number of inequalities, which is because we will typically have

more lower bounds on z_k and z_l. When we get more lower bounds, the system of linear inequalities gets bigger, and the inequalities more complicated to rewrite to a simplified form. For this reason I will not try to eliminate $\mathbf{z}$ from system (5.2) when $\beta_i - 1 < 0$ for more than two i's. Although a further study is not done, one gets a sense of what will happen when more of the coefficients β_i's are less than 1. Also, one gets a sense of how these systems, after performing FM elimination, will look like.

Now, let there be two scenarios with probabilities such that $p_k < 1 - \alpha$ and $p_l < 1 - \alpha$. Arrange such that these are indexed p_{m-1} and p_m. Assume we have eliminated $z_1, \ldots, z_{m-2}$ from system (5.2), such that we are left with the following system:

$$\begin{array}{rcrcrcl}
\beta_{m-1}z_{m-1} & + & \beta_m z_m & - & \mathbf{a}_1\mathbf{x} & \leq & \nu \\
 & & & & \vdots & & \\
\beta_{m-1}z_{m-1} & + & \beta_m z_m & - & \mathbf{a}_{m-2}\mathbf{x} & \leq & \nu \\
(\beta_{m-1}-1)z_{m-1} & + & \beta_m z_m & - & \mathbf{a}_{m-1}\mathbf{x} & \leq & \nu \\
\beta_{m-1}z_{m-1} & + & (\beta_m - 1)z_m & - & \mathbf{a}_m\mathbf{x} & \leq & \nu \\
z_{m-1}, & & z_m & & & \geq & 0
\end{array} \tag{5.12}$$

with lower and upper bounds on $\xi, z_1, \ldots, z_{m-1}$ as in Theorem 5.2.1. We start by eliminating z_{m-1}.

$$\begin{array}{rcrcrcrcl}
z_{m-1} & \leq & (\nu & - & \beta_m z_m & & + & \mathbf{a}_1\mathbf{x})\beta_{m-1}^{-1} \\
 & & \vdots & & & & & & \\
z_{m-1} & \leq & (\nu & - & \beta_m z_m & & + & \mathbf{a}_{m-2}\mathbf{x})\beta_{m-1}^{-1} \\
z_{m-1} & \geq & (\nu & - & \beta_m z_m & & + & \mathbf{a}_{m-1}\mathbf{x})(\beta_{m-1}-1)^{-1} \\
z_{m-1} & \leq & (\nu & - & \beta_m z_m & + \; z_m & + & \mathbf{a}_m\mathbf{x})\beta_{m-1}^{-1} \\
z_{m-1} & \geq & 0 & & & & & &
\end{array} \tag{5.13}$$

Since the coefficient $\beta_{m-1} - 1$ is negative, we get two lower bounds, and $m - 1$ upper bounds on z_{m-1}. Combining upper and lower bounds result in the following system:

$$
\begin{array}{rl}
0 & \leq (\nu - \quad \beta_m z_m + \quad \mathbf{a}_1\mathbf{x})\beta_{m-1}^{-1} \\
 & \vdots \\
0 & \leq (\nu - \quad \beta_m z_m + \mathbf{a}_{m-2}\mathbf{x})\beta_{m-1}^{-1} \\
0 & \leq (\nu - (\beta_m - 1)z_m + \quad \mathbf{a}_m\mathbf{x})\beta_{m-1}^{-1} \\
(\nu - \beta_m z_m + \mathbf{a}_{m-1}\mathbf{x})(\beta_{m-1} - 1)^{-1} & \leq (\nu - \quad \beta_m z_m + \quad \mathbf{a}_1\mathbf{x})\beta_{m-1}^{-1} \\
 & \vdots \\
(\nu - \beta_m z_m + \mathbf{a}_{m-1}\mathbf{x})(\beta_{m-1} - 1)^{-1} & \leq (\nu - \quad \beta_m z_m + \mathbf{a}_{m-2}\mathbf{x})\beta_{m-1}^{-1} \\
(\nu - \beta_m z_m + \mathbf{a}_{m-1}\mathbf{x})(\beta_{m-1} - 1)^{-1} & \leq (\nu - (\beta_m - 1)z_m + \quad \mathbf{a}_m\mathbf{x})\beta_{m-1}^{-1}
\end{array}
\tag{5.14}
$$

$$l'_{m-1} \leq z_{m-1} \leq u'_{m-1}$$

where l'_{m-1} and u'_{m-1} are given in Lemma 5.3.1. Rewriting inequalities number m to $2(m-1)$ in system (5.14) can be done in the same fashion as in Section 5.3. For $j = 1, \ldots, m-2$:

$$(\nu - \beta_m z_m + \mathbf{a}_{m-1}\mathbf{x})(\beta_{m-1} - 1)^{-1} \leq (\nu - \beta_m z_m + \mathbf{a}_j\mathbf{x})\beta_{m-1}$$

$$\nu - \beta_m z_m + \mathbf{a}_{m-1}\mathbf{x} \geq \nu - \beta_m z_m + \mathbf{a}_j\mathbf{x} - (\nu - \beta_m z_m + \mathbf{a}_j\mathbf{x})\beta_{m-1}^{-1}$$

$$\beta_{m-1}\mathbf{a}_{m-1}\mathbf{x} - \beta_{m-1}\mathbf{a}_j\mathbf{x} \geq -\nu + \beta_m z_m - \mathbf{a}_j\mathbf{x}$$

$$\beta_{m-1}\mathbf{a}_{m-1}\mathbf{x} - \beta_{m-1}\mathbf{a}_j\mathbf{x} - \beta_m z_m + \mathbf{a}_j\mathbf{x} \leq -\nu$$

$$\beta_m z_m - (\beta_{m-1}\mathbf{a}_{m-1} + (1 - \beta_{m-1})\mathbf{a}_j)\mathbf{x} \leq \nu$$

The last inequality in system (5.14) is similar to the ones above. The difference is that, because we on the right hand side we have $(\nu - (\beta_m - 1)z_m + \mathbf{a}_m\mathbf{x})\beta_{m-1}^{-1}$, we get in addition $\beta_{m-1}z_m - z_m$ on the left hand side after the rewriting. That is:

$$(\beta_m + \beta_{m-1} - 1)z_m - (\beta_{m-1}\mathbf{a}_{m-1} + (1 - \beta_{m-1})\mathbf{a}_m)\mathbf{x} \leq \nu$$

Rewriting the system (5.14) has resulted in the following system:

$$
\begin{array}{rcrcl}
\beta_m z_m & - & \mathbf{a}_1\mathbf{x} & \leq & \nu \\
 & & \vdots & & \\
\beta_m z_m & - & \mathbf{a}_{m-2}\mathbf{x} & \leq & \nu \\
(\beta_m - 1) z_m & - & \mathbf{a}_m\mathbf{x} & \leq & \nu \\
\beta_m z_m & - & (\beta_{m-1}\mathbf{a}_{m-1} + (1-\beta_{m-1})\mathbf{a}_1)\mathbf{x} & \leq & \nu \\
 & & \vdots & & \\
\beta_m z_m & - & (\beta_{m-1}\mathbf{a}_{m-1} + (1-\beta_{m-1})\mathbf{a}_{m-2})\mathbf{x} & \leq & \nu \\
(\beta_m + \beta_{m-1} - 1) z_m & - & (\beta_{m-1}\mathbf{a}_{m-1} + (1-\beta_{m-1})\mathbf{a}_m)\mathbf{x} & \leq & \nu
\end{array}
\tag{5.15}
$$

with $0 \leq z_i \leq u_i$, $i = 1, \ldots, m-2$, and $l'_{m-1} \leq z_{m-1} \leq u'_{m-1}$ as in Lemma 5.3.1. Now we continue by eliminating z_m. The critical coefficients here are $(\beta_m - 1)$ and $(\beta_m + \beta_{m-1} - 1)$. We know that $(\beta_2 - 1) < 0$ from our assumption, but $(\beta_m + \beta_{m-1} - 1)$ might be negative as well. This happens if $p_m + p_{m-1} < 1 - \alpha$. We will assume this is not the case, i.e that $p_m + p_{m-1} \geq 1 - \alpha$. If $(\beta_m + \beta_{m-1} - 1)$ were to be negative, we would get one more lower bound on z_m, and as a result the number of inequalities would become even bigger.

$$
\begin{array}{rcrl}
z_m & \leq & (\nu + & \mathbf{a}_1\mathbf{x})\ \beta_m^{-1} \\
 & \vdots & & \\
z_m & \leq & (\nu + & \mathbf{a}_{m-2}\mathbf{x})\ \beta_m^{-1} \\
z_m & \geq & (\nu + & \mathbf{a}_m\mathbf{x})\ (\beta_m - 1)^{-1} \\
z_m & \leq & (\nu + & (\beta_{m-1}\mathbf{a}_{m-1} + (1-\beta_{m-1})\mathbf{a}_1)\mathbf{x})\ \beta_m^{-1} \\
 & \vdots & & \\
z_m & \leq & (\nu + & (\beta_{m-1}\mathbf{a}_{m-1} + (1-\beta_{m-1})\mathbf{a}_{m-2})\mathbf{x})\ \beta_m^{-1} \\
z_m & \leq & (\nu + & (\beta_{m-1}\mathbf{a}_{m-1} + (1-\beta_{m-1})\mathbf{a}_m)\mathbf{x})\ (\beta_m + \beta_{m-1} - 1)^{-1}
\end{array}
\tag{5.16}
$$

As when eliminating z_{m-1}, we now have two lower bounds. Combining these with the $2m-3$ upper bounds result in $2(2m-3)$ inequalities. Rewriting the system of inequalities after combining lower bounds with upper bounds, will require some work. For this reason we will divide the system into four subsystems. The first subsystem consists of the inequalities with 0 as lower bound. These inequalities are quite familiar, so we omit the rewriting of these. This subsystem will after eliminating z_m be as follows:

$$\begin{array}{rcl} -\mathbf{a}_1\mathbf{x} & \leq & \nu \\ & \vdots & \\ -\mathbf{a}_{m-2}\mathbf{x} & \leq & \nu \\ -(\beta_{m-1}\mathbf{a}_{m-1}+(1-\beta_{m-1})\mathbf{a}_1)\mathbf{x} & \leq & \nu \\ & \vdots & \\ -(\beta_{m-1}\mathbf{a}_{m-1}+(1-\beta_{m-1})\mathbf{a}_{m-2})\mathbf{x} & \leq & \nu \\ -(\beta_{m-1}\mathbf{a}_{m-1}+(1-\beta_{m-1})\mathbf{a}_m)\mathbf{x} & \leq & \nu \end{array} \tag{5.17}$$

The second subsystem consist of the following inequalities, for $1 \leq j \leq m-2$:

$$(\nu+\mathbf{a}_m\mathbf{x})(\beta_m-1)^{-1} \leq (\nu+\mathbf{a}_j\mathbf{x})\beta_m^{-1}$$

These inequalities are equivalent to the ones we had in system (5.8) in Section 5.2 (with $j=1,\ldots,m-1$). So for $j=1,\ldots m-2$, this subsystem of inequalities can be written as:

$$-(\beta_m\mathbf{a}_m+(1-\beta_m)\mathbf{a}_j)\mathbf{x} \leq \nu$$

The third subsystem is as follows, for $j=1,\ldots,m-2$:

$$(\nu+\mathbf{a}_m\mathbf{x})(\beta_m-1)^{-1} \leq (\nu+(\beta_{m-1}\mathbf{a}_{m-1}+(1-\beta_{m-1})\mathbf{a}_j)\mathbf{x})\beta_m^{-1}$$

$$\nu+\mathbf{a}_m\mathbf{x} \geq \nu+(\beta_{m-1}\mathbf{a}_{m-1}+(1-\beta_{m-1})\mathbf{a}_j)\mathbf{x}-(\nu+(\beta_{m-1}\mathbf{a}_{m-1}+(1-\beta_{m-1})\mathbf{a}_j)\mathbf{x})\beta_m^{-1}$$

$$\beta_m\mathbf{a}_m\mathbf{x}-\beta_m(\beta_{m-1}\mathbf{a}_{m-1}+(1-\beta_{m-1})\mathbf{a}_j)\mathbf{x} \geq -(\nu+(\beta_{m-1}\mathbf{a}_{m-1}+(1-\beta_{m-1})\mathbf{a}_j)\mathbf{x})$$

$$\beta_m\mathbf{a}_m\mathbf{x}-\beta_m(\beta_{m-1}\mathbf{a}_{m-1}+(1-\beta_{m-1})\mathbf{a}_j)\mathbf{x}+(\beta_{m-1}\mathbf{a}_{m-1}+(1-\beta_{m-1})\mathbf{a}_j)\mathbf{x}) \geq -\nu$$

$$-(\beta_m\mathbf{a}_m+(1-\beta_m)(\beta_{m-1}\mathbf{a}_{m-1}+(1-\beta_{m-1})\mathbf{a}_j)\mathbf{x} \leq \nu$$

The fourth subsystem consists of only one inequality. Rewriting this inequality require some work.

$$(\nu+\mathbf{a}_m\mathbf{x})(\beta_m-1)^{-1} \leq (\nu+(\beta_{m-1}\mathbf{a}_{m-1}+(1-\beta_{m-1})\mathbf{a}_m)\mathbf{x})(\beta_m+\beta_{m-1}-1)^{-1}$$

$$(\nu+\mathbf{a}_m\mathbf{x})(\beta_m+\beta_{m-1}-1) \geq (\nu+(\beta_{m-1}\mathbf{a}_{m-1}+(1-\beta_{m-1})\mathbf{a}_m)\mathbf{x})(\beta_m-1)$$

$$\beta_{m-1}\nu + (\beta_m + \beta_{m-1} - 1)\mathbf{a}_m\mathbf{x} \geq (\beta_{m-1}\mathbf{a}_{m-1} + (1-\beta_{m-1})\mathbf{a}_m)\mathbf{x})(\beta_m - 1)$$

$$\beta_{m-1}\nu + \beta_{m-1}\mathbf{a}_m\mathbf{x} + (\beta_m - 1)\mathbf{a}_m\mathbf{x} \geq (\beta_{m-1}\mathbf{a}_{m-1} + (1-\beta_{m-1})\mathbf{a}_m)\mathbf{x})(\beta_m - 1)$$

$$\beta_{m-1}\nu \geq \beta_m\beta_{m-1}\mathbf{a}_{m-1}\mathbf{x} - \beta_{m-1}\mathbf{a}_{m-1}\mathbf{x} - \beta_m\beta_{m-1}\mathbf{a}_m\mathbf{x}$$

$$-(\beta_m\mathbf{a}_m\mathbf{x} + (1-\beta_m)\mathbf{a}_{m-1})\mathbf{x} \leq \nu$$

We see that this inequality is similar to the ones in the first subsystem. The new system of inequalities, after eliminating $\mathbf{z}$ is as follows:

$$\begin{array}{rcll}
-\mathbf{a}_1\mathbf{x} & \leq & \nu & \\
 & \vdots & & \\
-\mathbf{a}_{m-2}\mathbf{x} & \leq & \nu & \\
-(\beta_{m-1}\mathbf{a}_{m-1} + (1-\beta_{m-1})\mathbf{a}_1)\mathbf{x} & \leq & \nu & \\
 & \vdots & & \\
-(\beta_{m-1}\mathbf{a}_{m-1} + (1-\beta_{m-1})\mathbf{a}_{m-2})\mathbf{x} & \leq & \nu & \\
-(\beta_{m-1}\mathbf{a}_{m-1} + (1-\beta_{m-1})\mathbf{a}_m)\mathbf{x} & \leq & \nu & (5.18)\\
-(\beta_m\mathbf{a}_m + (1-\beta_m)\mathbf{a}_1)\mathbf{x} & \leq & \nu & \\
 & \vdots & & \\
-(\beta_m\mathbf{a}_m + (1-\beta_m)\mathbf{a}_{m-1})\mathbf{x} & \leq & \nu & \\
-(\beta_m\mathbf{a}_m + (1-\beta_m)(\beta_{m-1}\mathbf{a}_{m-1} + (1-\beta_{m-1})\mathbf{a}_1))\mathbf{x} & \leq & \nu & \\
 & \vdots & & \\
-(\beta_m\mathbf{a}_m + (1-\beta_m)(\beta_{m-1}\mathbf{a}_{m-1} + (1-\beta_{m-1})\mathbf{a}_{m-2}))\mathbf{x} & \leq & \nu &
\end{array}$$

with lower and upper bounds on z_i, $i = 1, \ldots, m-1$ as in Lemma 5.3.1, and $l''_m \leq z_m \leq u''_m$, where

$$l''_m(x_1, \ldots, x_n) = \{0, (\nu + \mathbf{a}_m\mathbf{x})\beta_m^{-1}\}$$

and

$$\begin{aligned}
u''_m(x_1, \ldots, x_n) = \min\Big\{ & \min_{1\leq j\leq m-2}(\nu + \mathbf{a}_j\mathbf{x})\beta_m^{-1}, \\
& \min_{1\leq j\leq m-2}(\nu + (\beta_{m-1}\mathbf{a}_{m-1} + (1-\beta_{m-1})\mathbf{a}_j)\mathbf{x})\beta_m^{-1}, \\
& (\nu + (\beta_{m-1}\mathbf{a}_{m-1} + (1-\beta_{m-1})\mathbf{a}_m)\mathbf{x})(\beta_m + \beta_{m-1} - 1)^{-1}\Big\}
\end{aligned}$$

Conclusions

The analysis in the three last sections has shown that we can eliminate the decision variables ξ and $\mathbf{z}$, and get a new, equivalent optimization model where the portfolio is the only decision variable. Also, adding constraints to the new problems can easily be done. Once one has found the optimal solution for one of these new problems, one can easily find the corresponding expected return. For finding the corresponding α-CVaR, one can either calculate VaR and then use the definition of CVaR, or one may use back substitution using the lower and upper bounds found during the FM elimination.

In Section 5.2 we assumed that all the scenario probabilities were greater or equal to $1 - \alpha$. This resulted in a simplified optimization model, with both less constraints and less decision variables. The original optimization model had $2(m + 1) + n$ constraints, while after eliminating ξ and $\mathbf{z}$ there were only $m + n + 1$ constraints. The result of this section was illustrated in an example, where we could see from its efficient frontier (Figure 6.3) that the optimal solution for this example was equivalent to the solution for the original problem. The downside with this new model was that the number of scenarios was bounded by $(1 - \alpha)^{-1}$.

In Section 5.3 we allowed one of the scenario probabilities to be smaller than $1 - \alpha$. This resulted in a similar model to the one in Section 5.2, but with some additional constraint. The number of constraints after eliminating ξ and $\mathbf{z}$ was $2m - 1 + n$, so it was still less than the number of constraints in the original problem. Allowing one of the probabilities $p_k < 1 - \alpha$ did not increase the upper bound on the number of scenarios. If for instance $\alpha = 0.99$, then $p_i = 0.01$ is the smallest probability we can have for $i \neq k$. If $m = 101$, then $\mathbf{e}^T\mathbf{p} = 0.01 \cdot 100 + p_k > 1$. So, m is bounded by $(1 - \alpha)^{-1}$ still.

In Section 5.4 we allowed two scenario probabilities p_k and p_l to be smaller than $1 - \alpha$, and the result was a system of a larger number of constraints. In the case where we assumed that the coefficient $\beta_k + \beta_l - 1$ was positive, the number of constraints in the new problem became $4m - 5 + n$. In the case where this coefficient was negative, the number of constraints would be even bigger. The upper bound on the number of scenarios for this case grew with one compared to the two previous section. It was not much, but a welcomed result.

Letting several scenario probabilities to be small will make the system of constraints grow in size, as we will get at least two lower bounds for each elimination of those z_i with negative coefficients $\beta_i - 1$. It seems like we

would be able to eliminate ξ and $\mathbf{z}$ for these cases as well, but doing so would require a lot of work. Studying the previous three section, one can get a sense of how these systems will look like. A benefit of allowing more scenario probabilities to be small is that the upper bound on the number of scenarios will increase.

6 | Numerical results

In this chapter I will present some efficient frontiers. The portfolio consists of a random selection of stocks in the S&P500. It is the 13 following stocks I have chosen; ACN, AA, BA, DIS, MMM, AGG, CNX, AGN, CAT, CAM, CA, CCE and CCI. If not otherwise specified, the portfolio consists of these. The scenarios are generated from monthly historical prices, from May 1 2006 to May 1 2016.

6.1 Efficient frontiers for problems with CVaR constraints

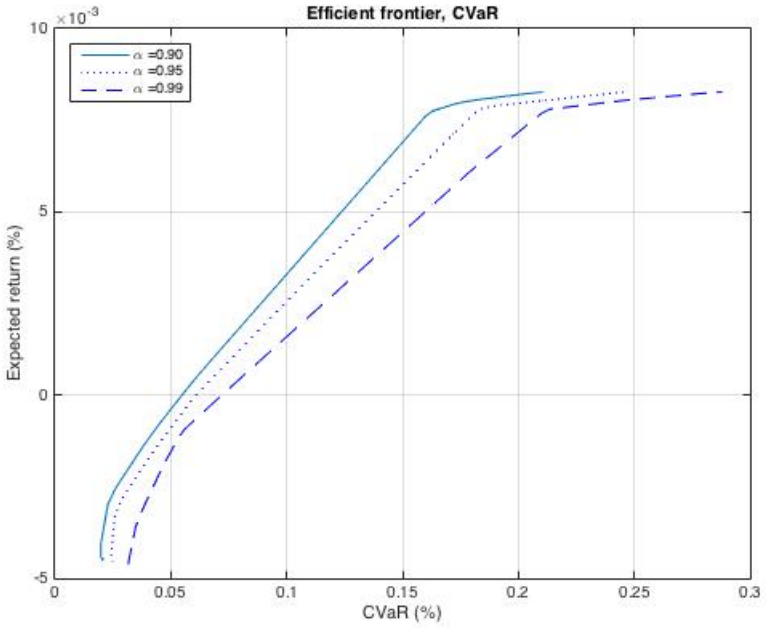


Figure 6.1: The efficient frontier of an one period model with CVaR constraints, of 13 assets from the S&P500. The graphs are plots of the efficient fronter for α-values $0.9, 0.95$ and 0.99, respectively.

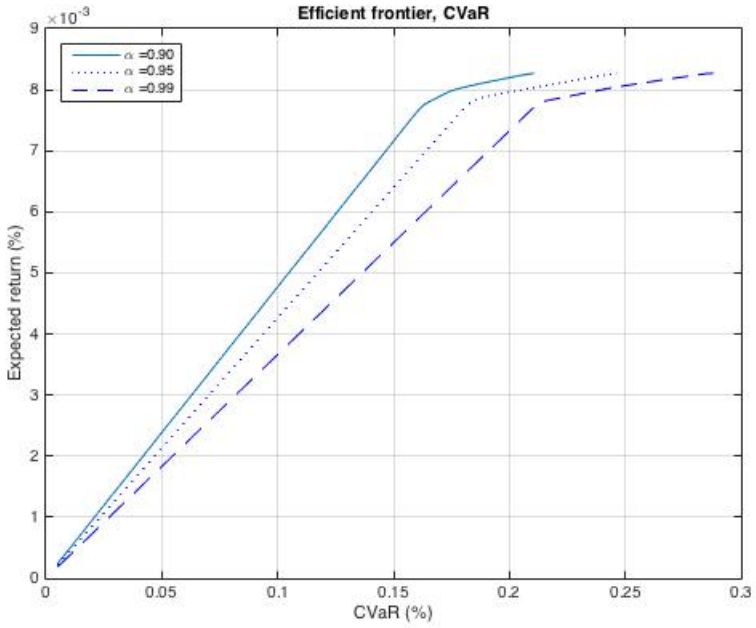


Figure 6.2: The efficient frontier of an one period model with CVaR constraints, of 13 assets from the S&P500, and the bank. The graphs are plots of the efficient fronter for α-values $0.9, 0.95$ and 0.99, respectively.

In Figure 6.1 the portfolio consists only of risky assets, while the portfolio in Figure 6.2 also includes the risk free asset. The interest rate is set to be 1%. For low expected returns (left side of the graph), we see that a combination of investing in the bank and in risky assets is preferred.

6.2 Efficient frontier for a Fourier-Motzkin example

Recall Example 5.2.1 where we found the optimal expected return and 0.95-CVaR for a risk level of 10%. Below a plot of the efficient frontier for this example. Since we only have two securities, this efficient frontier is a line.

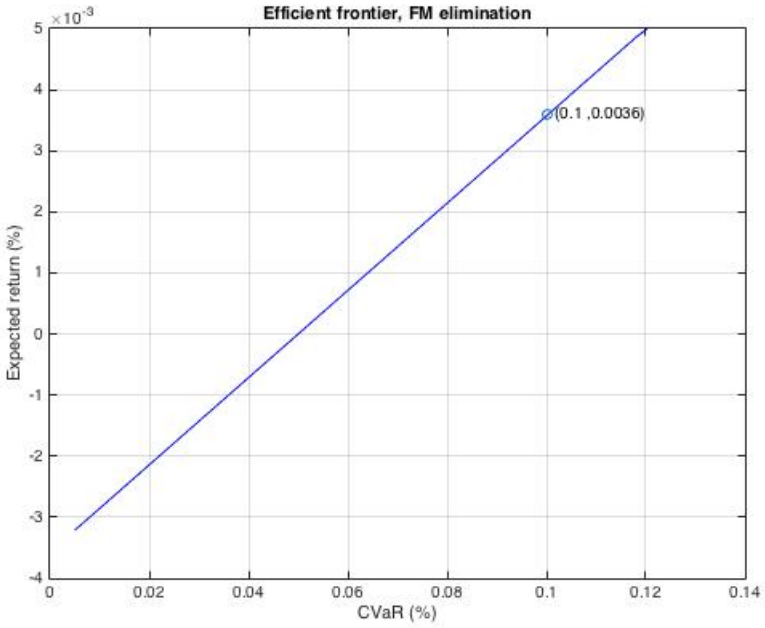


Figure 6.3: The efficient frontier for the problem in Example 5.2.1.

Figure 6.3 shows the efficient frontier for the optimization problem in Example 5.2.1. I used the original optimization problem with CVaR constraints from Section 5.1 to generate this efficient frontier. In Example 5.2.1 we found that the expected return was 0.0036 for the optimal solution, and that the corresponding 0.95-CVaR value was 0.1. This point lies, as we can see, on the efficient frontier.

Appendix A

In this chapter I will present my MATLAB code.

The following code computes the upper bound for z_k, using the equation (5.4) in Theorem 5.2.1. The vector $\mathbf{y}$ that is input for the function *uBound*, is of form $(x_1, \ldots, x_n, z_m, \ldots, z_{k-1})^T$. This program can be used when calculating CVaR using back substitution, for the problem in Section 5.2.

```
function uk= uBound(A,nu,bet,y,n)
% INPUT:
% A: matrix of returns
% nu: the risk level. it takes values in (0,1)
% bet: the beta vector. bet=p'/(1-alpha)
% y: the decision variable containg x, and
% contain some z_m,..z_{k-1}. [x,z]'
% n: the number of assets
% OUTPUT: the upper bound for z_k

m=length(bet);
z=y(n+1:length(y));
k=m-length(z);
x=y(1:n);

% bet*z=0 if all z has been eliminated
if isempty(z)
    s=0;
else
    s=bet(k+1:m)'*z;
end

% initial upper bound
```

```
uk=(nu-s+A(k,:)*x)/(bet(k)-1);

% seach for lower upper bounds
if k>1
    for i=1:k-1
        temp=(nu-s+A(i,:)*x)/bet(k);
        if temp<uk
            uk=temp;
        end
    end
end

if k<m
    for i=1:length(z)
        temp=(nu-s+A(m-i+1,:)*x-z(i))/bet(k);
        if temp<uk
            uk=temp;
        end
    end
end
if uk<10^(-9)
    uk=0;
end
% since lb=0 for all z_i:
if uk< 0
   error('Problem is infeasible');
end
end
```

The following code computes the maximum expected returns and the corresponding α-CVaR values for different risk levels. This program can be used to plot efficient frontiers, together with a program fetching prices from Yahoo. The code of these programs will be omitted here.

```
function [cvar,Fval] =CVaREF(ret,alp)
% INPUT: ret: the returns
% alp: alpha value, takes values in (0,1)
% Output:
% cvar: vector of alpha-CVaR, for different risk levels
% Fval: optimal expected rate of return, for different
% risk levels
```

```
8
9  [m,n]=size(ret);
10 p=1/m*ones(m,1);
11 bet=p/(1-alp);
12
13 % constraints
14 A=[1, bet', zeros(1,n);ones(m,1).*(-1),-eye(m),-ret];
15 b=zeros(1,m+1);
16 Aeq=[zeros(1,m+1) ones(1,n)];
17 beq=1;
18
19 % bounds on variabels
20 lb=zeros(1,m+n+1);
21 ub=[];
22
23 % objective function
24 c=p'*ret;
25 c=[zeros(1,m+1) -c]';
26
27 % levels of risk
28 nu=linspace(0.005,0.3);
29 k=length(nu);
30
31
32 % optimization
33 cvar=zeros(1,k);
34 Fval=zeros(1,k);
35 maxval=0;
36 for i=1:k
37       b(1)=nu(i);
38       [x,fval]=linprog(c,A,b,Aeq,beq,lb,ub);
39       cvar(i)=A(1,:)*x;
40       Fval(i)=-fval;
41       % Keep track of highess expected return
42       if maxval<(-fval)
43           maxval=-fval;
44       end
45 end
46
47 % ensures no horizontal lines
```

```
48 s=1;
49 for i=1:k
50     if maxval-Fval(i)>10^(-9)
51         s=s+1;
52     end
53 end
54
55 Fval=Fval(1:s);
56 cvar=cvar(1:s);
57 end
```

Bibliography

[ADEH99] Philippe Artzner, Freddy Delbaen, Jean-Marc Eber, and David Heath. Coherent measures of risk. *Mathematical finance*, 9(3):203–228, 1999.

[Bes10] Michael J. Best. *Portfolio optimization*. CRC Press, 2010.

[Çın11] Erhan Çınlar. *Probability and stochastics*, volume 261. Springer Science & Business Media, 2011.

[Dah10] Geir Dahl. An introduction to convexity. *Retrieved from University of Oslo: http://heim. ifi. uio. no/˜ geird/conv. pdf*, 2010.

[Dah14] Geir Dahl. A mini-introduction to convexity. 2014.

[Dan72] George B. Dantzig. Fourier-motzkin elimination and its dual. Technical report, DTIC Document, 1972.

[KPU01] Pavlo Krokhmal, Jonas Palmquist, and Stanislav Uryasev. Portfolio optimization with conditional value-at-risk objective and constraints. *Journal of risk*, 4:43–68, 2001.

[KR09] Mareike Kaina and Ludger Rüschendorf. On convex risk measures on L^p-spaces. *Mathematical methods of operations research*, 69(3):475–495, 2009.

[KZU11] Pavlo Krokhmal, Michael Zabarankin, and Stan Uryasev. Modeling and optimization of risk. *Surveys in Operations Research and Management Science*, 16(2):49–66, 2011.

[Lyc15] Tom Lyche. *Lecture Notes for Mat-inf 4130, 2015*. 2015.

[MW13] John N. McDonald and Neil A. Weiss. *A course in real analysis*. Academic press, 2013.

[Øk13] Bernt Øksendal. *Stochastic differential equations: an introduction with applications*. Springer Science & Business Media, 2013.

[Roc07] Ralph Tyrrell Rockafellar. Coherent approaches to risk in optimization under uncertainty. *Tutorials in operations research*, 3:38–61, 2007.

[RU99] Ralph Tyrrell Rockafellar and Stanislav Uryasev. Optimization of conditional value-at-risk. *Journal of risk*, 2:21–42, 1999.

[Van14] Robert J. Vanderbei. *Linear programming*. Springer, 2014.