universität wien

DISSERTATION

Titel der Dissertation

”Applications of non-convex Optimization in Portfolio Selection”

Verfasser

Mag. David Wozabal

angestrebeter akademischer Grad

Doktor der Sozial- und Wirtschaftswissenschaften (Dr.rer.soc.oec.)

Wien, im Oktober 2008

Studienkennzahl lt. Studienblatt:	084 136
Dissertationsgebiet lt. Studienblatt:	Statistik
Betreuer:	Univ.-Prof. Mag. Dr. Georg Pflug

Acknowledgements

It is a painful but undeniable truth, that the acknowledgements section is probably the most read part in 95% of the thesis written. While I hope that my thesis belongs to the fortunate 5%, I don't have much hope, especially since all it's parts except the appendices are available as separate papers. Surprisingly – as an empirical survey amongst my friends and acquaintances shows – it is also the part which is written in the morning of the day, when the thesis is going in printing.

Having said this I hope that I still manage to give credits to all people that have directly or indirectly contributed to this work.

First and foremost I want to thank my guide Prof. Georg Pflug – who mentored me throughout my work and yet gave me the freedom to work on the things that interested me. He was also the one to make this thesis possible by providing continuous funding throughout my work. I also owe him thanks for help in various administrative matters, which without his determination and boldness would have delayed my work substantially.

I feel similarly indebted to my wife Nancy – without her the completion of the thesis would have been nearly impossible. Most of the time spouses are mentioned for bearing with the author during the time the thesis was written – also in my case this was not always easy (so I was told). However, more importantly I want to thank Nancy for countless hours of discussion of all the major matters treated in this work. She gave me hints when badly needed, thought along with me and helped to solve some problems.

Next I want to mention my parents for the encouragement during my work and the faith in my ability to finish this thesis.

I also want to thank Prof. Jitka Dupačová for having agreed to be my second guide and carefully reviewing my work and my colleague Dr. Ronald Hochreiter at the Institute of Statistics for stimulating discussions and a lot of references.

Last but not least I want to mention Prof. Kriegl at the Department of Mathematics, who significantly advanced this work by giving the best maths courses I ever attended.

This work was partly supported by the *Jubiläumsfonds* of the Austrian National Bank (Grant: 12306), the *Fonds zur Förderung der wissenschaftlichen Forschung (FWF)* and the *Wiener Wissenschafts-, Forschungs- und Technologiefonds (WWTF)*.

Contents

1 Introduction **3**
- 1.1 Chapter II – A D.C. Formulation of Value-at-Risk constrained Optimization . . 4
- 1.2 Chapter III – A new method for Value-at-Risk constrained optimization using the Difference of Convex Algorithm 4
- 1.3 Chapter IV – A Framework for Optimization under Ambiguity 5
- 1.4 Appendices 5

2 A D.C. Formulation of V@R constrained Optimization **7**
- 2.1 Introduction 7
- 2.2 Reformulation of V@R$_\alpha$ as a D.C. function 10
- 2.3 A Branch-and-Bound Algorithm 14
- 2.4 Convergence & Implementation Issues 18
  - 2.4.1 The ω-rule 20
  - 2.4.2 Bisection 20
  - 2.4.3 Combination of rules 20
- 2.5 Numerical Results 22
  - 2.5.1 Run-Time Behavior 22
  - 2.5.2 Results of Portfolio optimization 23
  - 2.5.3 Local Search 26
- 2.6 Conclusion 27

3 V@R Optimization using the DCA **29**
- 3.1 Introduction 29
- 3.2 Reformulation of V@R$_\alpha$ as a D.C. function 31
- 3.3 Application of the DCA to problem (3.6) 33
- 3.4 Applications 35
  - 3.4.1 Comparison with global optima 35
  - 3.4.2 Application of DCA to large data sets 38
- 3.5 Conclusion 39

4 A Framework for Optimization under Ambiguity **41**
- 4.1 Introduction 41
- 4.2 Overview over existing literature 43
- 4.3 Reducing the problem 44
- 4.4 A concrete problem 50
  - 4.4.1 Minimizing Expectation 52
  - 4.4.2 Maximizing Expected Shortfall 54
  - 4.4.3 Other measures of risk 60
- 4.5 Numerical Results 60
- 4.6 Conclusion 64

Appendices **67**

A D.C. Functions and the DCA **69**
A.1 Difference of Convex Functions 69
A.2 The Difference of Convex Algorithm (DCA) 70

B Extreme and exposed points **79**
B.1 Extreme Points 79
B.2 Exposed points 80

CHAPTER 1

Introduction

This thesis is concerned with applications of non-convex programming to problems of portfolio selection in a single stage stochastic programming framework. It is divided into three parts each of which forms a separate technical report.

This introductory Chapter is an attempt to provide the reader with a short preview of the work presented and put the three parts in context to each other. This will hopefully provide this compilation with the required coherence in form and content and make reading easier.

First the common elements of the three papers will be described. All the three chapters are concerned with problems of portfolio optimization in a single stage stochastic optimization framework. A scenario based approach to model uncertainty is followed, i.e. the randomness in the models is always described by finitely many joint realizations of the asset returns. The advantage of this approach – apart from the relative simplicity with which complex dependencies can be described – is, that it is mostly possible to solve the optimization problems numerically, i.e. with the help of convex (usually linear or quadratic) interior point solvers.

A further common trait of all the papers is that the problems are of the mean risk type, i.e. the expected returns of the portfolios are maximized while some notion of risk is controlled for. Hence, in this work the return and the risk dimension are strictly separated – not mixed as often done in a multi-criteria portfolio optimization.

Apart from these similarities the topics of the three chapters can essentially be divided in two parts:

1. Optimization of the Value-at-Risk, when the distributions of the asset returns are known and finitely supported. Chapters 2 and 3 are devoted to this problem.

2. Solving mean risk problems in a single stage stochastic problem, where there is no precise knowledge of the involved distributions available. This topic is treated in Chapter 4.

Although the starting points for all Chapters were the respective problems and not the solution methods, it turned out that the all the problems ultimately lead to non-convex optimization problems, which can be solved using methods of difference of convex (D.C.) programming – which is connecting the papers on a methodological level.

While the problems in Chapter 2 are of moderate size and can be solved globally by a especially designed Branch-and-Bound method, in Chapter 3 and 4 the so called Difference of Convex Algorithm (DCA) is used to find local optima of bigger problems.

In the following a brief overview of the three Chapters of the thesis will be given.

1.1 Chapter II – A D.C. Formulation of Value-at-Risk constrained Optimization

The aim of In this report, which is joint work with Ronald Hochreiter and Georg Pflug, is to solve non-convex mean risk models with the Value-at-Risk as a risk measure. The paper like all other papers in this thesis assumes the random returns to have a discrete distribution (or to be approximated by a discrete distribution) with finitely many atoms. In this way dependencies can be modeled in a non-parametric way and scenarios can readily be obtained by resampling historical data or from econometric models.

It is shown that in this setting the Value-at-Risk is a D.C. function and the mentioned mean risk problem therefore corresponds to a D.C. problem. The non-convex problem of optimizing the Value-at-Risk is rather extensively treated in the literature and there are various approximative solution techniques as well as some approaches to solve the problem globally.

The reformulation as D.C. problem provides structural insight into the nature of the problem, which can be exploited to devise a Branch-and-Bound algorithm for finding global solutions for small to medium sized instances. Using this algorithm relatively small portfolio optimization problems can be solved to global optimality.

The discussion of the possibility to refine ε-optimal solutions obtained from the Branch-and-Bound framework via local search heuristics concludes the chapter.

This paper is separately available in [81] or online via *Optimization Online* (`http://www.optimization-online.org/`).

1.2 Chapter III – A new method for Value-at-Risk constrained optimization using the Difference of Convex Algorithm

In this part of the thesis the Value-at-Risk problem is once again investigated with the aim of solving problems of realistic sizes in relatively short time. Since the Value at Risk optimization can be shown to be a NP hard problem, this can only be achieved by sacrificing on the guaranteed globality of the solutions. Therefore a local solution technique for unconstrained D.C. problems called Difference of Convex Algorithm (DCA) is employed.

To use the DCA the problem has to first be transformed into a unconstrained problem by exact penalization. The DCA in its complete form requires to iteratively solve non-convex (to be precise convex maximization) problems to obtain approximations for the corresponding primal and dual solutions. To avoid having to solve these non-convex problems, there exists the so called *simplified DCA* with weaker convergence properties for which the non-convex problems are substituted by convex ones. A hybrid version of the algorithm for the considered problem, which in our case preserves the favorable convergence properties of the *complete DCA* as well as the computational tractability of the simplified DCA is proposed.

The results are tested for small instances and the solutions are shown to actually coincide with the global optima obtained in Chapter 2 in most of the cases. For realistic problem sizes the proposed method is shown to consistently outperform known heuristic approximations implemented in commercial software.

This paper is separately available in [80] or online via *Optimization Online* (`http://www.optimization-online.org/`).

1.3 Chapter IV – A Framework for Optimization under Ambiguity

The last Chapter of this Thesis is devoted to a different topic which received much attention in the recent stochastic programming literature: the topic of robust optimization. More specifically the aim is to robustify single stage stochastic optimization models with respect to *uncertainty about the distributions* of the random variables involved in the formulation of the stochastic program. The framework is designed with the application of robust portfolio selection in mind. The aim of the paper is to explore ways of explicitly taking into account ambiguity about the model, when finding a decision and imposing only very weak restrictions on possible probability models that are taken into consideration.

The main idea is the following: since in a stochastic programming framework randomness is explicitly incorporated, a prerequisite for solving such problems is the knowledge of the distribution P of the random variables used in the formulation of the problem.

If the probability measure P is unknown – as it is the case in most applications – usually an estimate $\hat{P}$ is used as a proxy for the real P. In the approach pursued here an estimate $\hat{P}$ for P serves as a reference measure, but the goal is that the solution also behaves well, if the true distribution P does not coincide with $\hat{P}$ exactly, i.e. we aim to find solutions that are to some degree stable with respect to model misspecification. Ambiguity is defined as the possible deviation from the discrete reference measure $\hat{P}$ and modeled by a so called *ambiguity set* $\mathscr{B}$. $\mathscr{B}$ is chosen such that it contains all the measures that can reasonably be assumed to be the real measure P given the available data. Following the idea to devise a general approach not restricted by assuming P to be from any specific parametric family, we define our ambiguity sets by the use of general probability metrics. Relative to these measures a worst case approach is adopted to robustify the problem.

The resulting optimization problems turn out to be infinite problems in the space of measures. To cope with these the exposed points of Kantorovich neighborhoods of discrete measures $\hat{P}$ are identified as discrete measures with at most three atoms more than $\hat{P}$. This result is subsequently used to reduce the infinite problems to non-convex semi-definite problems.

In the last part of the paper we show how to solve these problems numerically for the example of a mean risk portfolio selection problem with *Expected Shortfall under a Threshold* as the risk measure. The DCA in combination with an iterative algorithm to approximate the infinite set of constraints by finitely many ones is used to obtain numerical solutions.

This paper is separately available in [79] or online via *Optimization Online* (`http://www.optimization-online.org/`).

1.4 Appendices

The technical reports presented in Chapters 2, 3 and 4 are designed to be published in scientific journals and written accordingly. Owing to this structure at certain points the arguments are not provided in the detail this would have been possible in a classical thesis.

Therefore two technical appendices to make up for the lack of detail which might occur at certain places of the work are provided. In particular one of the appendices is devoted to D.C. functions and the Difference of Convex Algorithm which will be used in two of the three papers, but is not described in detail there. All the results (including the proofs) which where used in this work are discussed in this Appendix. This decisions was motivated by the fact that the results on the DCA are scattered among many papers and – to the best of the authors knowledge – there is no single accessible paper or textbook summarizing these results and at the same time providing self sufficient proofs. The Appendix is also used to discuss a new variant of the DCA – the *hybrid DCA* proposed in 3 – and prove that it has convergence properties comparable to the *complete DCA* while being computationally less demanding.

The second appendix is concerned with the properties of exposed and extreme points of convex sets and is crucial for the understanding of Chapter 4 of this work. The reason for the detailed discussion of this topic are similar to the ones for Appendix A. The results on extreme and exposed points are not readily available in any textbook or single scientific article – especially the rather general version of Straszewicz Theorem (i.e. exposed points are dense in the extreme points of a compact set) which is central for the arguments in Chapter 4 of this thesis, is mentioned in passing in some papers but – again to the best of my knowledge – nowhere completely proven.

CHAPTER 2

A D.C. Formulation of Value-at-Risk constrained Optimization

In this paper we present a representation of Value-at-Risk (V@R) as a difference of convex (D.C.) functions in the case where the distribution of the underlying random variable is discrete and has finitely many atoms. The D.C. representation is used to study a financial risk-return portfolio selection problem with a V@R constraint. A Branch-and-Bound algorithm that numerically solves the problem exactly is given. Numerical experiments with historical asset returns from representative market indices are performed to apply the algorithm to real-world financial market data.

2.1 Introduction

Value-at-Risk (V@R, see e.g. [45]) is an important topic for modern financial risk management, especially due to regulatory reasons in the context of Basel-II for the banking sector, as well as Solvency-II for the insurance sector. From an economic point of view, a t-day V@R at α of \$x means that the financial portfolio will incur a loss of \$x with probability $(1-\alpha)$ by the end of a t-day holding period, if the composition remains fixed over this period. Both the holding period and the confidence level is determined by either regulatory authorities (if V@R is used for the calculation of regulatory capital), or by companies themselves for the purpose of internal risk management. See [51] for a detailed discussion of V@R for quantitative risk management purposes.

The Value-at-Risk of a random variable X is defined as

$$V@R_\alpha(X) = \inf\{u : F_X(u) \geq \alpha\} = F_X^{-1}(\alpha), \quad 0 < \alpha < 1,$$

where F_X is the distribution function of X.

V@R was first proposed by the global financial services firm JPMorgan Chase & Co. as a measure of acceptability for a financial position with random return. If $\mathrm{V@R}_\alpha$ is taken to be quantile function of the return distribution of X, then $\mathrm{V@R}_\alpha$ is said to be an acceptability functional. A higher value indicates a more acceptable, i.e. better, less riskier portfolio. If on the other hand X represents the random losses, then $V@R_{1-\alpha}$ is a risk functional. High values indicate higher risk and thereby worse portfolios. See [56] for an in-depth discussion of acceptability and risk functionals. In this paper we consider X as anticipated (random) returns, and therefore consider $\mathrm{V@R}_\alpha$ as an acceptability functional.

However V@R$_\alpha$ – being the quantile of the return distribution – has the undesirable property of being non-concave. The non-concavity of the quantile function has two major drawbacks, one being of practical and the other of technical nature. The practical drawback is that Value-at-Risk may penalize diversification, i.e. given two financial positions, and their anticipated (random) future returns X and Y it might be that

$$V@R_\alpha(X+Y) < V@R_\alpha(X) + V@R_\alpha(Y), \tag{2.1}$$

which contradicts financial common sense and is a violation of the requirements specified in [11] and [12] for coherent risk measures - in particular V@R is not sup-additive.

The technical problem stemming from non-concavity is that Value-at-Risk constraint makes optimization problems computationally intractable. Maximizing Value-at-Risk or minimizing a convex function under a Value-at-Risk constraint leads to a non-convex optimization problem, which consequently is hard to solve.

To use Value-at-Risk in a decision optimization framework, we will formulate it as a Stochastic Program, and consider the following non-convex portfolio optimization problem:

$$\begin{array}{lllll} \max & \mathbb{E}\left(w^\top \xi\right) & & & \\ \text{s.t.} & \sum_{i=1}^m w_i & = & 1 & \\ & w_i & \geq & 0, & 1 \leq i \leq m \\ & V@R_\alpha\left(w^\top \xi\right) & \geq & a, & \end{array} \tag{2.2}$$

where w_i denotes the relative weight of asset i in the portfolio, ξ_i the random return of asset i and $w^\top \xi = \sum_{i=1}^m w_i \xi_i$.

Because of the shortcomings of V@R mentioned above it is often replaced by the Average Value-at-Risk (AV@R, also called Conditional Value-at-Risk) in practical applications of optimization problems of the type (2.2) (see for example [10]). The *Average Value-at-Risk* AV@R$_\alpha$ of a random variable X is defined as

$$AV@R_\alpha(X) = \frac{1}{\alpha}\int_0^\alpha F_X^{-1}(t)dt \quad \text{for} \quad 0 < \alpha \leq 1, \tag{2.3}$$

where again $X \sim F_X$, and F_X^{-1} is the inverse distribution function of X. The reason for the popularity of AV@R is threefold

1. AV@R is relatively easy to incorporate into optimization problems like (2.2). In the case of discrete random variables a linear programming formulation exists, as shown in [65] and [78].

2. AV@R is a coherent risk measure (see [55]), and does not suffer from the problems described in (2.1).

3. Additionally the Average Value-at-Risk is a concave lower minorant of the Value-at-Risk and therefore the function $w \mapsto AV@R_\alpha(w^\top \xi)$ is a lower bound for the function $w \mapsto V@R_\alpha(w^\top \xi)$.

Therefore one could replace (2.2) by the following conservative approximation

$$\begin{array}{lllll} \max & \mathbb{E}\left(w^\top \xi\right) & & & \\ \text{s.t.} & \sum_{i=1}^m w_i & = & 1 & \\ & w_i & \geq & 0, & 1 \leq i \leq m \\ & AV@R_\alpha\left(w^\top \xi\right) & \geq & a. & \end{array}$$

As mentioned above this problem can be formulated as a linear program and therefore solved efficiently with standard optimization packages.

However, due to the importance of Value-at-Risk for the financial industry, numerous approaches have been proposed to solve problems of type (2.2). The approaches can be divided into two groups: approximation methods, and global optimization methods. One of the oldest technique is to solve a different parametric optimization problem (e.g. the mean-variance problem or the $AV@R$ constrained problem), and find a point on the risk-return space on the efficient frontier, for which the V@R is exactly as required (see [60] for an overview). Obviously such an approach does not solve the V@R constrained problem (2.2). Other approximations use smoothing techniques (see [35]), concave minorants (see [47]), or quadratic approximations (see [73]).

Under some distributional assumptions, the V@R constrained problem becomes numerically tractable. In particular, if the vector ξ is multivariate normal, then the mapping $w \mapsto V@R_\alpha(w^\top \xi)$ is concave and the V@R-constraint coincides with the variance constraint (up to a constant). The same holds for the multivariate lognormal family. This property was e.g. used by [32], [13] and [58] to study V@R-constrained portfolio problems for the continuous time Geometric Brownian Motion model. In general, all elliptical distributions allow the reformulation of optimization problems involving V@R as objective or as constraint to a quadratic or quadratically constrained problem respectively (see [73]). In [52] robust optimization techniques are used to approximate V@R.

A straightforward approach to solve (2.2) exactly can be based on a mixed-integer program formulation of the problem (see for example [16]). Although this problem is hard (NP-complete in the strong sense, see [16]) some heuristic solution schemes have been designed, such as random search with threshold acceptance (see [37] and [38]), or evolutionary computation techniques (see [42]). In [20] complete enumeration on the risk-return grid is used to find near optimal portfolios for the V@R portfolio optimization problem. Pang and Leyffer [53] formulate the problem as a linear program with equilibrium constraints to derive lower and upper bounds for a Branch-and-Bound solution. Cheon et al. [22] propose a solution technique for the more general class of probabilistically constrained linear programs which is based on a Branch-Reduced-Cut algorithm.

The approach we pursue in this paper is slightly different. We consider the special case where the involved distributions are discrete with a finite number of atoms, i.e. we treat Value-at-Risk in a finite scenario-based stochastic programming framework. In the case of finitely many scenarios we give a representation of the V@R as the difference of convex (D.C.) functions, i.e. we show that the V@R is a D.C. function.

To solve problem (2.2) we use a Branch-and-Bound algorithm for optimizing D.C. functions, which is a modification of an algorithm given in [44]. The algorithm finds the global optimum

of the given D.C. problem. Numerical experiments using financial market data substantiate the applicability of the approach.

This paper is organized as follows. In Section 2.2 the D.C. reformulation of V@R is discussed. Sections 2.3 and 2.4 discuss the Branch-and-Bound algorithm, that is used to calculate global solutions, in detail. Section 2.5 presents a set of numerical results based on real-world financial market data, while Section 2.6 concludes the paper.

2.2 Reformulation of V@R$_\alpha$ as a D.C. function

We will consider the classical bi-criteria asset allocation problem based on Markowitz seminal paper [50] on portfolio selection. An investor has to choose a portfolio from a set of investment possibilities (financial assets) $\mathscr{I}$ with finite cardinality $m = |\mathscr{I}|$ to invest her available budget. The bi-criteria problem stems from the fact that the investor aims at maximizing her return while controlling the risk of the chosen portfolio at the same time. The idea to maximize return under a risk constraint became very popular in the recent years (see for example [48] or [2]) and also has been extended to a multi-stage stochastic programming setting (see [56] or [67]). We will apply this general idea to a setting, where V@R is used as the risk functional, as shown in (2.2), i.e. maximize the expected return subject to a V@R lower bound.

The assumption that $w_i \geq 0$ in (2.2) is made for the sake of convenience and could easily be relaxed in favor of an arbitrary boundedness assumption of the form $w_i \in [a_i, b_i]$ with a_i, $b_i \in \mathbb{R}$ (in particular the possibility of short sales could be incorporated). For our approach to work, we need the following assumption.

Assumption 1. *The distribution of the random asset returns $\xi = (\xi_1, \ldots, \xi_m)$ is a discrete distribution with finitely many atoms, i.e. there are $S \in \mathbb{N}$ scenarios for the joint realizations of the random variables ξ_i. The $(m \times 1)$ vector of realizations of ξ in scenario s will be denoted by ξ^s and the probability of the scenario will be denoted by p_s.*

Assuming that random variables have discrete distributions is a common approach in stochastic programming and has the advantage that the random variables can easily be described in terms of empirical data, either by using historical realizations as scenarios directly, or by applying resampling techniques. The obvious advantage over models where the random variables are assumed to follow a distribution from a specific parametric family is that certain features like heavy tails, and non-normal skewness or kurtosis can easily be captured, especially in the highly multi-variate case.

As already mentioned V@R$_\alpha(X)$ is non-concave in X, and therefore the problem (2.2) cannot be solved by standard convex programming techniques. However, it turns out that – in the discrete case – it can be reformulated to a difference of convex (D.C.) problem. To facilitate this reformulation we make Assumption 2.

Assumption 2. *Assume that all the scenarios have equal probabilities, i.e.*

$$p_s = \frac{1}{S}, \quad 1 \leq s \leq S.$$

Assumption 2 is made for the sake of simplicity of presentation. It is for example satisfied, if empirical data is used as scenarios but also if scenarios are generated from some parametric model and not explicitly re-weighted. However, Remark 2.2 shows, that the V@R functional is a D.C. function even in the case of unequal weights and the results of this paper can be extended to this case.

Given the above assumptions problem (2.2) becomes

$$\begin{array}{lll} \max & \frac{1}{S}\sum_{s=1}^{S} w^\top \xi^s & \\ \text{s.t.} & \sum_{i=1}^{m} w_i & = \quad 1 \\ & w_i & \geq \quad 0 \\ & V@R_\alpha\left(w^\top \xi\right) & \geq \quad a. \end{array} \tag{2.4}$$

Although we will restrict ourselves to the above problem in the rest of the paper, our proposed methodology can easily be adapted to other cases, e.g. problems with more involved convex portfolio constraints, or problems with V@R as the objective function.

If the above two assumptions are satisfied, we can derive a D.C. formulation of the V@R in the following way. If X follows a discrete distribution taking the values $x_1,\ldots,x_n$ with equal probability (in our case X represents the anticipated (random) returns of the portfolio, i.e. $X = w^\top \xi$), then

$$AV@R_{\frac{k}{n}}(X) = \frac{n}{k}\sum_{i=1}^{k} x_{i:n}\frac{1}{n} = \frac{1}{k}\sum_{i=1}^{k} x_{i:n}, \tag{2.5}$$

where $x_{1:n} \leq x_{2:n} \leq \cdots \leq x_{n:n}$ is the set of ordered values of X. Therefore

$$V@R_\alpha(X) = x_{k:n} = kAV@R_{\frac{k}{n}}(X) - (k-1)AV@R_{\frac{k-1}{n}}(X)$$

with $k = [\alpha n]$. Hence under the above assumptions, V@R$_\alpha(X)$ can be written as the difference of the two concave functions

$$kAV@R_{\frac{k}{n}}(X) \text{ and } (k-1)AV@R_{\frac{k-1}{n}}(X).$$

Therefore, problem (2.4) becomes

$$\begin{array}{lll} \max & \frac{1}{S}\sum_{s=1}^{S} w^\top \xi^s & \\ \text{s.t.} & \sum_{i=1}^{m} w_i & = \quad 1 \\ & w_i & \geq \quad 0 \\ & kAV@R_{\frac{k}{n}}(w^\top \xi) - (k-1)AV@R_{\frac{k-1}{n}}(w^\top \xi) & \geq \quad a, \end{array}$$

or equivalently can be written as a minimization problem with a D.C. constraint:

$$\begin{array}{lll} \min & -\frac{1}{S}\sum_{s=1}^{S} w^\top \xi^s & \\ \text{s.t.} & \sum_{i=1}^{m} w_i & = \quad 1 \\ & w_i & \geq \quad 0 \\ & (-kAV@R_{\frac{k}{n}}(w^\top \xi)) - (-(k-1)AV@R_{\frac{k-1}{n}}(w^\top \xi)) & \leq \quad -a. \end{array} \tag{2.6}$$

Value of ξ_1	Value of ξ_2	Probability
$-\frac{1}{2}-\frac{1}{2^k}$	$\frac{1}{2}-\frac{1}{2^k}$	$\frac{1}{2^{k+1}}$
$\frac{1}{2}+\frac{1}{2^k}$	$-\frac{1}{2}+\frac{1}{2^k}$	$\frac{1}{2^{k+1}}$

Table 2.1: Joint distribution of ξ_1 and ξ_2 for Remark 2.1, $k \in \mathbb{N}$

We use the latter formulation in this paper. Since we want to apply a variant of algorithm X.5 outlined in [44], we require the problem to be in so called canonical form:

$$\begin{array}{ll} \min & \Psi(x) \\ \text{s.t.} & f(x) \leq 0, \quad g(x) \geq 0, \end{array} \tag{2.7}$$

where $f : \mathbb{R}^d \to \mathbb{R}$ and $g : \mathbb{R}^d \to \mathbb{R}$ are convex functions and $\Psi : \mathbb{R}^d \to \mathbb{R}$ is a linear function. In fact every D.C. program can be reformulated to an equivalent problem in standard form (2.7) (see [44]). In our case the reformulation is done by introducing the real variable z and writing

$$\begin{array}{lllll} \min & -\frac{1}{S}\sum_{s=1}^{S} w^\top \xi^s & & & \\ \text{s.t.} & \sum_{i=1}^{m} w_i & = & 1 & \\ & w_i & \geq & 0 & \\ & (-kAV@R_{\frac{k}{n}}(w^\top \xi)) - z & \leq & -a & \\ & (-(k-1)AV@R_{\frac{k-1}{n}}(w^\top \xi)) - z & \geq & 0. & \end{array} \tag{2.8}$$

Obviously problems (2.6) and (2.8) are equivalent, and the latter is in the canonical form (2.7) with

$$\begin{aligned} \Psi(z,w) &= -\frac{1}{S}\sum_{s=1}^{S} w^\top \xi^s \\ f(z,w) &= max\left((-kAV@R_{\frac{k}{n}}(w^\top \xi)) - z + a, w_i, \sum_{i=1}^{m} w_i - 1 \right) \\ g(z,w) &= (-(k-1)AV@R_{\frac{k-1}{n}}(w^\top \xi)) - z. \end{aligned}$$

Consistent with (2.7) we will denote the vector decision variables (z,w) as x and the dimension as d. Note that we added the splitting variable z, so that $d = m+1$.

The above reformulation holds under Assumption 1 & 2. Remark 2.1 shows, that if Assumption 1 is violated, the V@R functional can not be represented as a D.C. function in general, while Remark 2.2 shows that Assumption 2 can be relaxed.

Remark 2.1. *In the following we will demonstrate, that there are random variables ξ_1 and ξ_2 such that $w \mapsto V@R_{0.5}(w\xi_1 + (1-w)\xi_2)$ is not D.C. on $[0,1]$. Let the joint distribution of ξ_1 and ξ_2 be given as in Table 2.1.*

Further Let $X_w = w\xi_1 + (1-w)\xi_2$ and $q(w) = V@R_{0.5}(X_w)$ with $w \in [0,1]$. Notice that the distribution of X_w is symmetric around 0 for all w. However, since $q(w) = \inf\{u : P\{w\xi_1 + (1-$

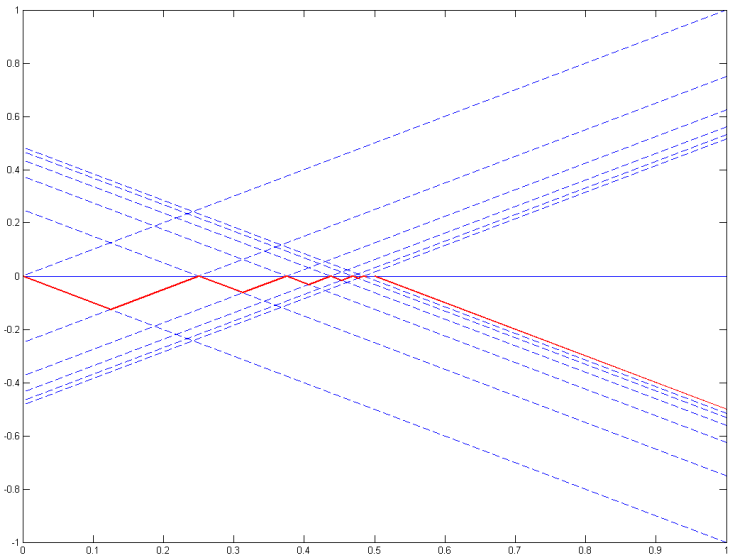


Figure 2.1: Sample trajectories $w \mapsto w\xi_1 + (1-w)\xi_2$ (dashed lines), and the V@R-function $q(w)$ (in thick lines).

$w)\xi_2 \le u\} \ge 0.5\}$ *the value of* $q(w)$ *equals the largest value of* X_w*, which is below* 0*, if there is no value exactly at* 0*. This leads to the following form for* q *in the interval* $[0,1/2]$*:*

$$\begin{array}{rcll} q\left(\frac{1}{2}-\frac{1}{2^k}\right) & = & 0, & k=1,2,\ldots \\ q\left(\frac{1}{2}-\frac{3}{2^k}\right) & = & -\frac{1}{2^k}, & k=3,4,\ldots, \end{array}$$

and linearly interpolated between these points (see Fig. 2.1). Therefore, the derivative of q *is*

$$\begin{array}{rcllll} q'(x) & = & -1 & \text{for } \frac{1}{2}-\frac{1}{2^k} < x < & \frac{1}{2}-\frac{3}{2^{k+1}} & k=1,2,\ldots \\ q'(x) & = & 1 & \text{for } \frac{1}{2}-\frac{3}{2^k} < x < & \frac{1}{2}-\frac{1}{2^k} & k=3,4,\ldots \end{array}$$

One sees that q' *is not of bounded variation. In contrast, all D.C. functions on compact intervals possess a derivative (where any value from the sub- and subdifferential respectively can be chosen), which is of bounded variation. Thus* q *cannot be a D.C. function.*

Remark 2.2. *In this remark, we show that the V@R functional is always a D.C. function, if the probability space is finite, i.e. Assumption 1 holds. Suppose that* p_s*,* $1 \le s \le S$ *are arbitrary scenario probabilities and w.l.o.g. let the scenarios be ordered according to their returns. Define*

$$\varepsilon = \alpha - \max\{\sum_{s=1}^{k} p_s : \sum_{s=1}^{k} p_s < \alpha, 1 \le k \le S\}.$$

By finiteness it follows, that $\varepsilon > 0$*. By a similar argument as above we obtain*

$$V@R_\alpha(X) = \frac{\alpha}{\varepsilon} AV@R_\alpha(X) - \frac{\alpha-\varepsilon}{\varepsilon} AV@R_{\alpha-\varepsilon}(X).$$

Notice that (2.5) *is the special case for* $\varepsilon = \frac{1}{n}$ *and* $\alpha = \frac{k}{n}$*.*

It should also be remarked, that D.C. representations are not unique. An interesting question is how the actual D.C. representation and the properties of the two convex functions influence the efficiency of solution techniques. See for example [74] or [17] for work in this direction.

To conclude this section, we give a D.C. representation of the k-minimum of finitely many concave functions. An alternative D.C. representation for V@R can be deduced as a special case of this result.

Remark 2.3. *The minimum of concave functions is concave, all other order statistics including the maximum are (in general) not. We demonstrate, that the k-minimum is a D.C. function: Let* $f = (f_1, \ldots, f_S)^\top$ *be a vector of S concave functions and let* $\bar{f} = (f_{1:S}, \ldots, f_{S:S})^\top$ *be the ordered vector, where*

$$f_{1:S} \leq f_{2:S} \leq \cdots \leq f_{S:S}.$$

The function $f_{k:S}$ *is called the k-minimum of* $f_1, \ldots, f_S$*. Let further*

$$F_k = \sum_{i_1,\ldots,i_{S-k+1}} \min(f_{i_1}, \ldots, f_{i_{S-k+1}}).$$

Here $(i_1, \ldots, i_{S-k+1})$ *iterates through all* $\begin{pmatrix} S \\ k-1 \end{pmatrix}$ *selections of* $S-k+1$*. All* F_k *obviously are concave functions. A D.C. representation of* $f_{k:S}$ *is* $f_{k:S} = g_1 - g_2$ *with*

$$\begin{aligned} g_1 &= \sum_{j+k\equiv 0(2);1\leq j\leq k} \begin{pmatrix} S-j \\ S-k \end{pmatrix} F_j \\ g_2 &= \sum_{j+k\equiv 1(2);1\leq j\leq k} \begin{pmatrix} S-j \\ S-k \end{pmatrix} F_j, \end{aligned}$$

where $m \equiv k(n)$ *means m is equal to k modulo n.*

2.3 A Branch-and-Bound Algorithm

In this section we describe a solution technique for problem (2.8). For an introduction to D.C. programming and solution algorithms we refer to [44] or [43].

The algorithm used in this paper is a variant of the Branch-and-Bound algorithm X.5 in [44]. The units that are subject to the branching and bounding operations are cones in our decision space $\mathbb{R}^d$. Table 2.2 gives an overview of the functions and sets used in the algorithm and how they are denoted.

Ψ	...	the objective function to be minimized
f, g	...	the convex functions making up the D.C. constraint $f - g \leq 0$
Ω	...	the set $\{x : f(x) \leq 0\}$
Δ	...	the set $\{x : g(x) < 0\}$
∂A	...	the boundary of the set A
d	...	the dimension of the space of decision variables
α	...	the best objective value of a feasible point found so far
x	...	the best feasible point found so far
$\mathscr{C}_n$	...	the set of all cones in iteration n
$\mathscr{D}_n$	...	the set of all cones added in iteration n by splitting up an existing cone
$C_{i,n}$	...	the i-th cone that was produced in the n-th iteration
$C'_{i,n}$	...	the predecessor of cone $C_{i,n}$, i.e. the cone that $C'_{i,n}$ is a part of
c_i	...	points in $\partial\Delta$ that define the rays of the cone C
$H_{C,\mu}$	...	the hyperplane $\{\sum_{i=1}^{d} \lambda_i c_i : \sum_{i=1}^{d} \lambda_i c_i = \mu\}$, where c_i are the defining points of C
β_C	...	lower on Ψ in cone C
β_n^*	...	the lowest lower bound in all cones in iteration n

Table 2.2: Functions and sets used in the Branch-and-Bound algorithm.

The cones C we use can be described by d linearly independent vectors $c_i \in \mathbb{R}^d$. Given the points $c_1, \ldots, c_d$ the cone C is the smallest cone containing these points, i.e.

$$C = \{\sum_{i=1}^{d} \lambda_i c_i : \lambda_i > 0,\ 1 \leq i \leq d\}. \tag{2.9}$$

Obviously positive rescaling of the points c_i does not change the cone.

Like other Branch-and-Bound algorithms the procedure involves the following steps:

1. Computing lower bounds for the objective for every cone C.
2. Deleting cones that yield higher lower bounds than the best found feasible solution so far.
3. Splitting up the cone that yields the lowest lower bound.
4. Repeating the procedure until the lowest lower bound and the value of the best feasible point coincide or there are no more cones.

We will shift our coordinate system in such a way, that the origin is a point $w \in \partial\Omega \cap \Delta$. The coordinates of all the cones, hyperplanes and points mentioned below are coordinates in this shifted coordinate system, i.e. relative to w. As indicated in Table 2.2, we represent our cones C by points $c_i \in \partial\Delta$. This is possible since the origin of our coordinate system is inside the bounded set Δ.

In the following we give a brief sketch of the algorithm. The functions *getW*, *update*, *lowerBound*, *maxShift* used in Algorithm 1 and the division rules for the cones will be discussed separately below.

Algorithm 1 Outer approximation Branch-and-Bound algorithm

w $\leftarrow$ getW(α, μ);
$\mathscr{D}_0 \leftarrow$ getPartition(w);
$\mathscr{C}_0 \leftarrow \{\}$; $\alpha_0 \leftarrow \infty$; $\beta_{0,0} \leftarrow -\infty$
for every intersection z of a ray of a cone $C \in \mathscr{D}_0$ with the set $\mathscr{C}$ **do**
  if z is feasible and $\Psi(z) \leq \alpha$ **then**
    $x \leftarrow z$;
    $\alpha_0 \leftarrow f(z)$;
  end if
end for

n $\leftarrow$ 1;
while $\alpha - \min_{C_{r,s}} \beta_{C_{r,s}} > \varepsilon$ **do**
  for every $C_{i,n-1} \in \mathscr{D}_{n-1}$ **do**
    $s_i \leftarrow maxShift(C_i)$;
    if $s_i < 1$ **then**
      $\mathscr{D}_{n-1} \leftarrow \mathscr{D}_{n-1} \setminus \{C_i\}$;
    else
      $(\beta_{C_{i,n}}, x_i) \leftarrow lowerBound(C_{i,n-1}, s_i)$;
      $\beta_{C_{i,n}} \leftarrow max(\beta_{C_{i,n}}, \beta_{C'_{i,n-1}})$;
      if $feasible(x_i)$ **then**
        $\mathscr{D}_{n-1} \leftarrow \mathscr{D}_{n-1} \setminus \{C_i\}$;
        if $f(x_i) < \alpha$ **then**
          $x \leftarrow z$;
          $\alpha_0 \leftarrow f(z)$;
        end if
      end if
    end if
  end for
  if $\mathscr{D}_{n-1} = \{\}$ and $\mathscr{C}_{n-1} = \{\}$ **then**
    α is the optimal value, x is an optimal point
    break;
  end if
  split up $C^*_{i,j}$ with $\beta_{C^*_{i,j}} = \min_{C_{r,s}} \beta_{C_{r,s}}$ into the set of cones $\mathscr{D}_n$
  $\mathscr{C}_n \leftarrow update((\mathscr{C}_{n-1} \setminus \{C^*_{i,j}\}) \cup \mathscr{D}_{n-1}, \alpha)$;
  n $\leftarrow$ (n+1);
end while

The function *getW* The function *getW* finds the point w, that serves as the origin of the coordinate system used in the algorithm. The point w should be in the set $\partial\Omega \cup \Delta$. Such a point can easily be found by solving the linear program

$$\begin{array}{lll} \min & -\frac{1}{S}\sum_{s=1}^{S} w^\top \xi^s & \\ \text{s.t.} & \sum_{i=1}^{m} w_i & = \quad 1 \\ & w_i & \geq \quad 0 \\ & -kAV@R_{\frac{k}{n}}(w^\top \xi) & \leq \quad z \end{array} \tag{2.10}$$

The function *getPartition* The function *getPartition* divides the decision space $\mathbb{R}^d$ into an initial partition of $d+1$ cones. Such a partition can be found by starting with a set of vectors $c_1, \ldots, c_{d+1}$ such that $(c_1 - c_{d+1}), \ldots, (c_d - c_{d+1})$ are linearly independent and

$$0 \in conv(c_1, \ldots, c_{d+1}),$$

where $conv(c_1, \ldots, c_{d+1})$ is the convex hull of the points c_i.

An initial conical partition can be obtained from c_1, ..., c_{d+1} by defining cones according to (2.9) for every of the $(d+1)$ subsets of $\{c_1, \ldots, c_{d+1}\}$ with d elements. One possible example for the choice of c_i is: e_1, ..., e_d, -$\mathbb{1}$, where the e_i are the standard basis vectors of $\mathbb{R}^d$ and $\mathbb{1}$ is the vector consisting of ones. More generally, we can start with any set of linearly independent vectors c_1, ..., c_d, and any $\lambda = (\lambda_1, \ldots, \lambda_{d+1}) > 0$ with $\sum_{i=1}^{d+1} \lambda_i = 1$ and set $c_{d+1} = -\frac{1}{\lambda_{d+1}} \sum_{i=1}^{d} \lambda_i c_i$.

The function *maxShift* The function *maxShift(C)* determines how far the hyperplane

$$H_{C,1} = \left\{ \sum_{i=1}^{d} \lambda_i c_i : \sum_{i=1}^{d} \lambda_i = 1 \right\}$$

can be shifted outwards (i.e. away from the point w) and still touch the set $\Omega \cap C$. This can be found out by solving the following optimization problem

$$\begin{array}{ll} \max & \sum_{i=1}^{d} \lambda_i \\ s.t. & x \in \Omega \cap C \end{array} \tag{2.11}$$

Clearly all feasible points in C lie between the hyperplanes $H_{C,1}$ and $H_{C,\mu}$, where $\mu = \sum_{i=1}^{d} \lambda_i^*$ and $\lambda^* = (\lambda_1^*, \ldots, \lambda_d^*)$ is the optimal solution to problem (2.11). If μ is smaller than 1, then there is no feasible point in the cone C. For the further processing of the cone in the algorithm it would be advantageous if μ is as small as possible. Since the idea is to capture all the feasible points in C between $H_{C,1}$ and $H_{C,\mu}$ we can improve this bound by also taking a linear overestimation of g and thereby an outer approximation of the set Δ into account, when solving problem (2.11).

The actual implementation therefore restricts the feasible set by linearly overestimating g and using this estimate to determine the highest level $H_{C,\mu}$. This is done in the following iterative procedure.

Algorithm 2 maxShift

Require: cone C given by $c_i \in \partial\Delta$, that together with w define the rays

1: find the optimal value μ_0 to the problem (2.11)
2: $n \leftarrow 0$
3: **repeat**
4: $\quad n \leftarrow (n+1)$
5: $\quad$ solve (2.11), with constraint $\sum_{i=1}^{d} \lambda_i g(\mu_{n-1} c_i) \geq z$ to get $\mu_n = \sum_{i=1}^{d} \lambda_i^*$
6: **until** $\mu_n = \mu_{n-1}$
7: **return** $(\mu_n, \sum_{i=1}^{d} \lambda_i^* c_i)$

The linear overestimation $\bar{g}$ of the function g enables us to approximate the non-convex constraint $g(x) \geq 0$. The points that do not fulfill

$$\bar{g}(x) := \sum_{i=1}^{d} \lambda_i g(\mu_n c_i) \geq 0$$

do also not fulfill $g(x) \geq 0$ and are therefore infeasible. The final value of μ_n therefore has the property, that all the feasible points in C are below the hyperplane $H_{C,\mu_n} = \left\{ \sum_{i=1}^{d} \lambda_i c_i : \sum_{i=1}^{d} \lambda_i = \mu_n \right\}$.

The function *lowerBounds* The function *lowerBounds* finds lower bounds on the optimal value in every cone C, i.e. solves the problem

$$\begin{array}{ll} \max & \Psi(x) \\ s.t. & x \in \Omega \cap C \\ & \bar{g}(x) \geq z \end{array}$$

Where the function $\bar{g}$ is the linearized version of the function g as used in the function *maxShift* for this particular cone. The point that is found here does not need to be feasible. However, if it is feasible then the cone can be deleted, since the best feasible point in this cone is found.

The function *update* The function *update* receives a set of cones and the lowest objective value α as arguments. The function deletes all cones where the lower bounds are higher than α and returns the remaining cones.

2.4 Convergence & Implementation Issues

The reason for the convergence of the algorithm is the fact that the linear approximation of $\partial\Delta \cap C$ by the hyperplane

$$H_1 = \left\{ \sum_{i=1}^{d} \lambda_i c_i : \sum_{i=1}^{d} \lambda_i = 1 \right\}$$

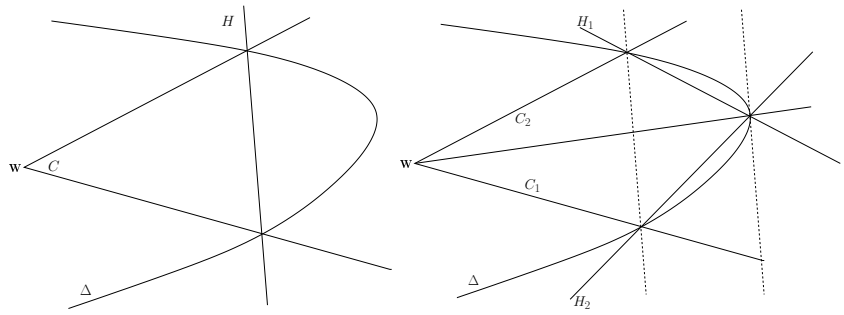


Figure 2.2: The above pictures show the splitting of the cone C in two sub-cones C_1 and C_2.

keeps improving as the size of the cone decreases. See Fig. 2.2 for a graphical depiction of this fact. It is obvious that if $\partial\Delta \cap C = H_1$ the problem reduces to a convex optimization problem, since the second (problematic) convex function is equal to a linear one in the cone C.

In the general version of the algorithm presented in [44] the approximation of $\partial\Delta \cap C$ by H_1 might never become exact as described above. Therefore the algorithm results in points that converge to the global solution, but it is not guaranteed that the algorithm will terminate after finitely many iterations. Furthermore, since the algorithm works with an outer approximation of the feasible set no statement can be made about whether a feasible solution will be found at all.

In our case the situation is a little different. The set Δ is a polyhedron and therefore the approximation of $\partial\Delta \cap C$ by a hyperplane can be exact. Since by the splitting operation the cones get smaller and smaller and finally contract to single rays, we know that after finitely many iterations the algorithm will stop, since for all C the approximations of $\partial\Delta \cap C$ are exact. A prerequisite for this is the aforementioned contraction of cones, which we will discuss in more detail below.

There are different strategies to split up cones and since this seems to be a central point determining the practical success of the implementation, we give a brief summary of the two approaches applied in this work and how to combine them.

A cone is always split up along a ray. Given a point y in the cone C, we first find the point c that lies on the line connecting w and y that intersects $\partial\Delta$. Since $w \in \Delta$ and Δ is convex this point exists and is unique. The point can be found by simple line search techniques like bisection, the secant method or newton type methods. Let $c = \sum_{i=1}^{d} \lambda_i c_i$ and $I = \{i : \lambda_i > 0\}$, then the new cones are formed by replacing single rays c_i, $i \in I$ of the old cone by the new ray (defined by c). The number of new cones $|I|$ therefore varies between 2 and d and depends on whether the new ray lies in the interior of C or on one of the faces.

When we talk about convergence of the algorithm in the following, we mean the closing of the gap between the lowest lower bound β_n^* on the objective value in iteration n and the objective value of the best found feasible solution, i.e. the convergence $\alpha - \beta_n^* \to 0$ as $n \to \infty$. There might therefore be cases when a globally optimal point x^* is already found, but since β_n^* is still smaller than $\Psi(x^*)$, we can not yet be sure about the optimality. In this case we say that the algorithm did not yet converge, although we already found the optimal solution.

2.4.1 The ω-rule

The cone that is selected for splitting is usually found by solving an optimization problem. The idea of the ω-rule is to use the result of this optimization to find a suitable new ray to split the cone. In our case the solution to the two programs *maxShift* and *lowerBounds* seem not to be suitable candidates for such rays.

In practice it turned out, that it is best to use the solution $(\lambda_1^*, \ldots, \lambda_d^*)$ of the optimization problem

$$\begin{array}{ll} \max & \sum_{i=1}^{d} \lambda_i \\ s.t. & x = \sum_{i=1}^{d} \lambda_i c_i \in \Delta. \end{array} \tag{2.12}$$

to obtain the splitting ray $\omega = \sum_{i=1}^{d} \lambda_i^* c_i$.

The point found in (2.12) is the vertex of the polyhedron Δ with the biggest distance to the hyperplane H_1. Dividing the cone along this ray potentially reduces the gap between the linear approximations of Δ in the successor cones the most.

2.4.2 Bisection

Bisection is a rather simple technique to split up a cone C into two cones. The splitting happens at the midpoint of the longest edge of the polygon $H_1 \cap C$. The splitting of cones by bisection obviously leads to shrinking cones and therefore to more and more accurate approximations of $\partial\Delta$. The advantage of bisection over the ω-rule is, that it is computationally much cheaper and it produces 2 instead of potentially d sub-cones for every cone that is split up.

2.4.3 Combination of rules

The general perception in the literature (see for example [44]) is that although bisection guarantees a steady convergence it usually is too slow if applied exclusively. The ω-rule on the other hand can lead to very fast convergence, since the splitting ray is chosen in a fashion that fits to the problem and the specific situation in the cone. However, in general it cannot be guaranteed, that the algorithm converges, when exclusively applying the ω-rule. The reason for this is that rays that are chosen might *converge to an already existing ray* and therefore lead to a smaller and smaller partition of one particular cone that does not necessarily lead to the deletion of the same. If this happens, the approximation of the set $\partial\Delta$ need not get better as it should for the convergence of the algorithm.

A solution to this problem is to combine the two strategies. As we will discuss below we apply both of the rules alternately. The repeated application of bisection ensures that the cones contract to single rays as discussed above. Since the set Δ has a finite number of vertices it is obvious that the linear approximation of Δ will become exact after a finite number of divisions. This yields the following (finite) convergence Theorem.

Theorem 2.1. *Algorithm 1 terminates after finitely many iterations. If the problem is feasible the optimal solution is found, otherwise the termination of the algorithm proves that there is no feasible point.*

Proof. The proof follows from the above discussion and Proposition X.12 in [44]. □

In the following we give an overview of possible rules that can be applied to decide on the mix of bisection and application of the ω-rule. An approach that proved to work in practice is to adopt the policy to always apply the ω-rule except in some cases where it is unlikely to yield good results. In the implementation the following heuristics helped to detect these situations and therefore to speed up the convergence of the algorithm.

1. If the point $\omega \in C$ lies on a ray of the cone C then splitting along this ray would not lead to a new cone and therefore the application of the ω-rule does not make sense.

2. A more strict version of the above rule would be to avoid the ω-rule if the point ω is not central enough. This is based on the idea, that the convergence slows down if the rays used for splitting cones are too near to already existing rays. To this end we write

$$\omega = \sum_{i=1}^{d} \lambda_i c_i = \sum_{i=1}^{d} \lambda_i' \frac{c_i}{||c_i||_2}$$

   where $\lambda_i' = \lambda_i ||c_i||_2$ and $\mu = \sum_{i=1}^{d} \lambda_i'$. A possible measure of centrality $\gamma(C,\omega)$ would be

$$\gamma(C,\omega) = \sum_{i=1}^{d} \left(\frac{1}{n} - \frac{\lambda_i'}{\mu} \right)^2$$

   $\gamma(C,w)$ varies from 0 in the case of equal weights for all the λ_i' (i.e. a centered ray) to 1 in the case of the new ray being identical with an already existing ray. In the actual implementation it proved useful to calculate $\gamma(C,\omega)$ for every cone C that gets split up and refrain from using the ω-rule, if $\gamma(C,\omega)$ is bigger than a certain threshold.

3. Define the generation $G(C)$ of cone C as the number of splits that occurred starting from a cone in the initial partition of the space to come to C. A rather simple rule to control the amount of bisection, is to use the bisection whenever

$$G(C) \equiv 0(k)$$

   i.e. $G(C)$ equals 0 modulo k.

4. Another heuristic that proved useful is to apply bisection whenever the convergence stalls. It is in general not easy to assess whether the algorithm stalls or not.

   (a) An obvious measure would be the progress in the last k iterations. Following this principle one could apply bisection if the progress is less than ε in the last k iterations, i.e.

$$\beta_{n-k}^* - \beta_n^* < \varepsilon$$

   Switch back to the ω-rule once the difference to β_{n-k}^* is greater then ε.

   (b) Another way of detecting a lack of progress in the ω-rule is to count how many β_C are in a close neighborhood of β_n^*. If there are to many cones in this neighborhood switch to bisection until the specified neighborhood is cleared.

Name	Average Return	Variance	$V@R_{.05}$	$AV@R_{.05}$
US Long Bond	0.9988	0.0002908	0.9762	0.9569
Standard & Poors 100	1.0013	0.0001658	0.9773	0.9749
Nasdaq 100	1.0008	0.0004351	0.9625	0.9547
FTSE 100	1.0026	0.0003147	0.9735	0.9614
Hang Seng	1.0033	0.0003226	0.9727	0.9637

Table 2.3: Characteristics of the weekly returns of the five indices subsequently used in portfolio optimization. Time frame: 2004-2005.

5. A sign of the malfunctioning of the ω-rule often is the splitting up of a successor of a cone that has been split up in the last iteration. This means that the splitting did not yield a substantial decrease in the lower bound in the resulting cone. If this phenomenon occurs in a certain number of consecutive iterations, it seems reasonable to switch to bisection.

It should be mentioned that certain obvious optimizations and performance enhancing tricks, that are applied in the implementation, are not discussed for the sake of brevity and to ensure the clarity of exposition.

2.5 Numerical Results

In this section the algorithm described above is applied to real-world financial market data to show the applicability of the approach. Weekly closing values of the following 5 indices have been used to calculate the discrete return scenarios: US Long Bond, Standard & Poors 100, Nasdaq 100, FTSE 100, and Hang Seng. Table 2.3 gives an overview of the characteristics of the used data.

The optimization problems were solved using the solver MOSEK, version 5. The Branch-and-Bound algorithm, as well as the general workflow are implemented using MatLab R2007a on both Windows and Linux machines. The main results were calculated on a HP ProLiant DL 585 with 32GB RAM using Red Hat Enterprise Linux 5, the results in 2.5.1 were calculated on a notebook with a Pentium Mobile Processor (1.8 GHz) 1.5GB RAM using Windows XP SP 2, i.e. the algorithm works well on standard personal computers.

2.5.1 Run-Time Behavior

The non-smooth and non-convex behavior of V@R functional increases with the α-level that is used and with the number of scenarios. The number of local optima in fact grows exponentially in those parameters. The reason for this is that the V@R problem can also be viewed as the combinatorial problem of choosing $[\alpha n]$ scenarios and then maximizing the expected return under the constraint that the maximal return in these scenarios is at least a. The best portfolio in terms of expectation among all possible choices of the $[n\alpha]$ scenarios is the global optimum. See [35] for an in-depth discussion.

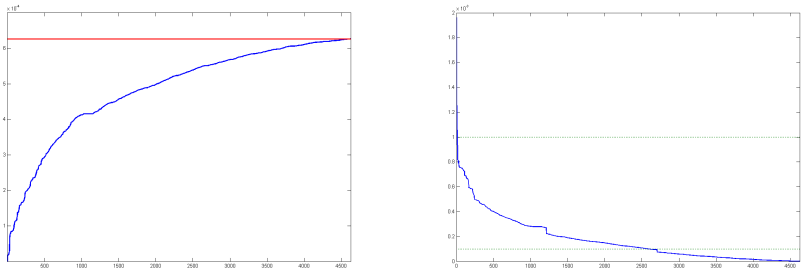


Figure 2.3: The left graph shows the increase in the lowest lower bounds (over all the remaining cones), while the right graph shows the decrease in the gap between the lowest lower bound and the best feasible point found so far.

This *decrease in convexity* of the V@R functional (i.e. the increase in the number of local minima) is also noticeable in the run time behavior of the algorithm described above. For small values of α and a small number of scenarios the algorithm terminates very fast, while for bigger α or a large number of scenarios the run-time behavior gets worse.

For all calculations in this paper we used the empirical distribution of two years of weekly data, i.e. 104 scenarios and an α of 0.05. These parameters yield a runtime between 1.4 to 7771 seconds (see Table 2.4).

Fig. 2.3 shows the convergence of the algorithm as a function of time (seconds) for $a = 0.98$. The rate of convergence is fast in the beginning and drops significantly during the last third of the iterations. It also should be noted that the optimal solution is typically found in a relatively early iteration. Closing the gap between the solution and the lowest lower bound on the objective in the Branch-and-Bound framework, i.e. the computational verification of global optimality, is however time consuming.

This behavior of the algorithm suggests the possibility of prematurely terminating the algorithm and using the (guaranteed) ε-optimal solution found so far. In the above example for an ε of 0.001 this could be done in iteration 77 (after 14.29 seconds) and for $\varepsilon = 0.0001$ after 16301 iterations (or 2592.3 seconds) as indicated by the dotted lines in the right graph. For practical purposes both of these precisions should be enough.

We will investigate the possibility of prematurely terminating the algorithm and possible refinements of ε-optimal solutions by means of local search heuristics in more detail below.

2.5.2 Results of Portfolio optimization

Next we present efficient frontiers for the bi-criteria V@R problem obtained by varying the parameter a. It turns out that the interesting range for a is between the V@R of the portfolio that consists only of the asset with the highest return (i.e. Hang Seng with $V@R_{.05}$ of 0.9727 and expected weekly return 1.0033) and the last feasible value of a, which is 0.987.

In this range we performed optimizations varying the a in 0.0005 steps. In Fig. 2.4 the dependence of the maximal returns and the portfolio compositions of the optimal portfolios on

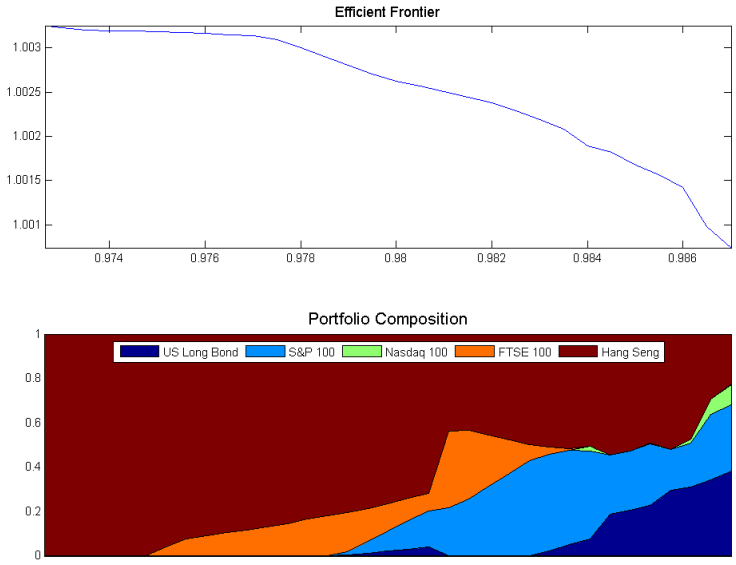


Figure 2.4: Efficient frontier and globally optimal portfolios for V@R constrained optimization.

the acceptance level a are depicted.

The non-concavity of the V@R shows in the plots in Fig. 2.4: at the points 0.984 and 0.9865 the efficient frontier has an inward kink that makes the function non-concave. Also the heavy fluctuations in the asset weights can be interpreted as a consequence of the non-convexity (convexity would ensure a continuous dependency of the solutions on the parameter a).

Table 2.4 shows the run times and number of iterations for the optimization runs in Fig. 2.4 and for the respective $\varepsilon = 0.001$ optimal solutions.

The runtime is consistent with the findings of the last section. The higher the parameter a, the more involved the problem becomes. The runtime behavior for ε-optimal solutions exhibit a similar pattern, but the runtimes as well as the number of iterations required stay at a much lower level.

Next we demonstrate the possibility of incorporating further convex constraints into problem (2.2). We used a slightly modified version of Algorithm 1 to compute optimal solutions for the following problem

$$\begin{array}{lllll} \max & \mathbb{E}\left(w^\top \xi\right) & & & \\ \text{s.t.} & \sum_{i=1}^m w_i & = & 1 & \\ & 0 \le w_i & \le & U, & 1 \le i \le m \\ & V@R_\alpha\left(w^\top \xi\right) & \ge & a. & \end{array} \tag{2.13}$$

where U is the upper bound for the investment into a single asset. Constraints like the above are frequently used in real life portfolio selection to prevent the optimization software from

a	Time	Iterations	Time ($\varepsilon = 0.001$)	Iterations ($\varepsilon = 0.001$)
0.9730	5.7438	30	1.4	10
0.9735	141.59	891	1.5	10
0.9740	121.92	643	1.5	10
0.9745	111.18	489	1.4	10
0.9750	125.16	543	1.5	10
0.9755	73.292	320	1.4	10
0.9760	68.188	303	1.5	10
0.9765	170.36	882	2.4	23
0.9770	155.43	863	2.5	23
0.9775	413.64	2993	2.4	23
0.9780	738.96	5628	2.4	23
0.9785	1414.3	12636	2.3	23
0.9790	2328.3	21849	2.5	24
0.9795	3643.3	34574	5.4	43
0.9800	3336	32246	11.8	77
0.9805	2482.7	23544	14.8	94
0.9810	2305.6	20734	17.7	104
0.9815	2313.6	20390	24.4	142
0.9820	2325.5	21441	66.7	363
0.9825	2568.4	24268	94.3	579
0.9830	3504.6	35893	61.6	376
0.9835	3749.4	39414	388.8	2976
0.9840	5500.8	59114	431.8	3330
0.9845	4830.9	52113	450.6	3502
0.9850	4880	52095	560.7	4525
0.9855	5426.6	57469	1661.5	16762
0.9860	5807.7	63225	2047.9	20820
0.9865	6555.6	71719	2566.2	26463
0.9870	7771	82968	5121.3	54972

Table 2.4: Runtime of the algorithm for the results in Fig. 2.4. Runtime and the number of iterations are reported for the exact solutions as well as for the $\varepsilon = 0.001$ optimal solutions.

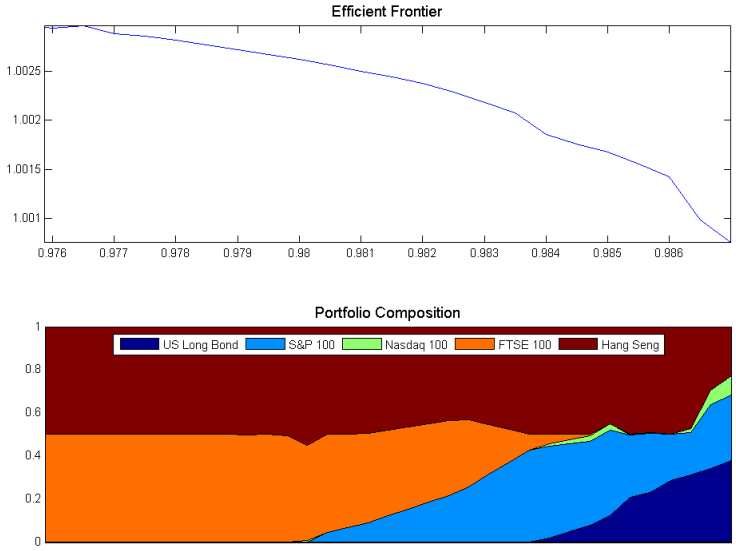


Figure 2.5: Efficient frontier and optimal portfolios for problem (2.13) with $U = 0.5$.

choosing portfolios, which consist only of very few (or in the extreme case one) assets. The reason to impose such restrictions might be of regulatory nature or stem from the discomfort an investor might feel with a portfolio that is not sufficiently diversified. In view of (2.1) it might be sensible to include such restrictions when using V@R as a measure of risk, to enforce a certain level of diversification.

Convex constraints like the above pose no difficulty to the approach outlined in this paper, they might even improve the speed of convergence, since the feasible region is reduced by further constraints.

Fig. 2.5 shows the results for problem (2.13) with $U = 0.5$.

2.5.3 Local Search

A local search algorithm was applied to evaluate, whether it might be useful to stop the Branch-and-Bound algorithm earlier, i.e. using ε-optimal solutions, and subsequently refining these solutions with a computationally inexpensive heuristic optimization technique.

The local search algorithm we used is sketched in Algorithm 3. This algorithm tries to improve an ε-optimal portfolio weight vector w by repeated shifting of asset weight by an amount δ. All shifted portfolios, given the upper weight bound U, are determined by function *shiftedPortfolios()*. The shifting from one asset to another asset, which improves the expectation of the resulting portfolio the most without violating the V@R constraint (evaluated by function *findMaxExpNVaR()*) is used for subsequent iterations. The local improvement is iteratively repeated with lower values for δ until a lower bound δ_l is reached.

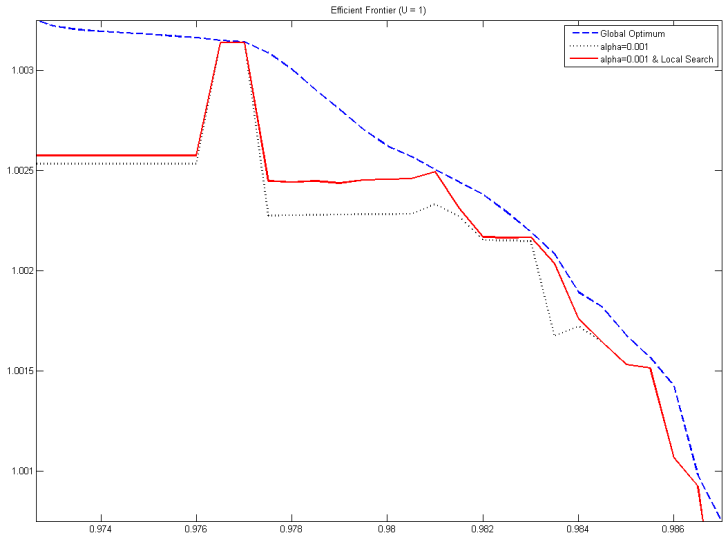


Figure 2.6: Efficient frontier for the unconstrained problem using local search

The results of the local search improvements are shown in Fig. 2.6 for the unconstrained problem, and in Fig. 2.7 for the constrained problem with $U = 0.5$ for $\delta = 0.25$, and $\delta_l = 10^{-5}$. The local search heuristic improves the quality of the solution significantly in most of the cases and sometimes even finds the global optimum. This demonstrates that a combination of the proposed Branch-and-Bound framework with local heuristics can be used to obtain good solutions in relatively short time (compared to the computation time required for global optimality).

Algorithm 3 Iterative weight-based local search

Require: δ, w, U

while $\delta > \delta_l$ **do**
  while w is locally improved **do**
    $\mathscr{W} \leftarrow shiftedPortfolios(w, \delta, U)$
    $w \leftarrow findMaxExpNVaR(\mathscr{W})$
  end while
  $\delta \leftarrow \frac{\delta}{2}$
end while

2.6 Conclusion

In this paper we presented a D.C. formulation of the Value-at-Risk functional in the case where the distribution of the random returns is discrete and finitely supported. We consequently used

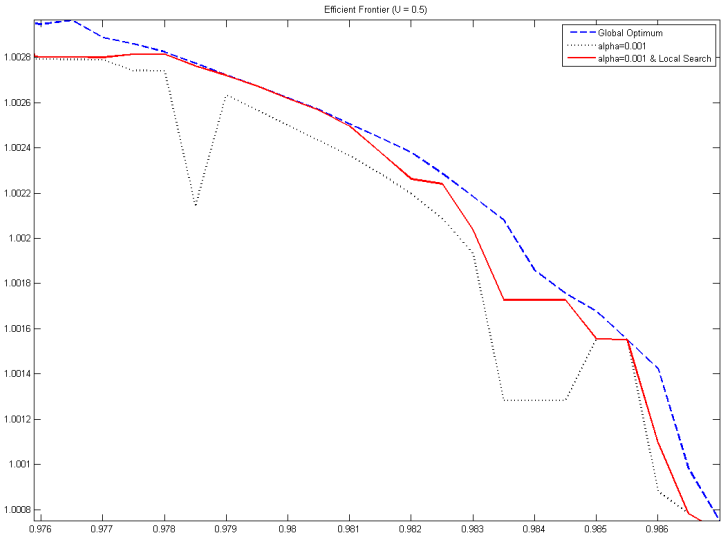


Figure 2.7: Efficient frontier for problem (2.13) with $U = 0.5$ using local search

this representation in non-convex V@R constrained portfolio selection problems and applied a Branch-and-Bound algorithm to solve these problems to global optimality. We demonstrated the applicability of the approach using real market data. We further experimented with a combination of the Branch-and-Bound solution technique with a simple local search heuristic, which improved ε-optimal solutions obtained from the Branch-and-Bound algorithm. Since both the local search and the ε-optimal solutions are computationally relatively cheap, we were able to significantly reduce computation time while still guaranteeing relatively good (at least ε-optimal) solutions.

It would be an interesting topic for future research to apply efficient approximation algorithms for D.C. optimization like the DCA (difference of convex algorithm, see [6]) in combination with Branch-and-Bound frameworks like proposed in [3] to be able to solve portfolio composition problems of the type (2.2) for larger data sets.

CHAPTER 3

A new method for Value-at-Risk constrained optimization using the Difference of Convex Algorithm

This paper treats a Value-at-Risk constrained Markowitz style portfolio selection problem. The described problem is non-convex stochastic problem and in the case of a discrete probability measure can be reformulated to a difference of convex (D.C.) program. We apply the difference of convex algorithm (DCA) to obtain solutions to the problem. Numerical results comparing the solutions found by the DCA to the respective global optima for relatively small problems as well as numerical studies for real life problems are given.

3.1 Introduction

Value-at-Risk (V@R, see e.g. [45]) is an important topic for modern financial risk management, especially due to regulatory reasons in the context of Basel-II for the banking sector, as well as Solvency-II for the insurance sector. From an economic point of view, a t-day V@R at α of \$x means that the financial portfolio will incur a loss of \$x with probability $(1-\alpha)$ by the end of a t-day holding period, if the composition remains fixed over this period.

The Value-at-Risk of a random variable X is defined as

$$V@R_\alpha(X) = \inf\{u : F_X(u) \geq \alpha\} = F_X^{-1}(\alpha), \quad 0 < \alpha < 1,$$

where F_X is the distribution function of X.

V@R was first proposed by the global financial services firm JPMorgan Chase & Co. as a measure of acceptability for a financial position with random return. If $\mathrm{V@R}_\alpha$ is taken to be the quantile function of the return distribution of X, then $\mathrm{V@R}_\alpha$ is said to be an acceptability functional. A higher value indicates a more acceptable, i.e. better, less riskier portfolio. If on the other hand X represents the random losses, then $V@R_{1-\alpha}$ is a risk functional. High values indicate higher risk and thereby worse portfolios. See [56] for an in-depth discussion of acceptability and risk functionals. In this paper X will represent anticipated (random) returns, and therefore $\mathrm{V@R}_\alpha$ is considered as an acceptability functional.

However $\mathrm{V@R}_\alpha$ – being the quantile of the return distribution – has the undesirable property of being non-concave. The non-concavity of the quantile function has two major drawbacks, one

being of practical and the other of technical nature. The practical drawback is that Value-at-Risk may penalize diversification, i.e. given two financial positions, and their anticipated (random) future returns X and Y it might be that

$$V@R_\alpha(X+Y) < V@R_\alpha(X) + V@R_\alpha(Y), \tag{3.1}$$

which contradicts financial common sense and is a violation of the requirements specified in [11] and [12] for coherent risk measures - in particular V@R is not sup-additive.

The technical problem stemming from non-concavity is that a Value-at-Risk constraint (or a V@R objective function) makes optimization problems computationally intractable (except in certain special cases where returns follow a parametric model, see for example [49] or [58]). Maximizing Value-at-Risk or minimizing a convex function under a Value-at-Risk constraint leads to a non-convex optimization problem, which consequently is hard to solve.

To use Value-at-Risk in a decision optimization framework, we will formulate the following non-convex portfolio optimization problem for $m \in \mathbb{N}$ assets:

$$\begin{array}{lrcl} \max & \mathbb{E}\left(w^\top \xi\right) & & \\ \text{s.t.} & \sum_{i=1}^m w_i & = & 1 \\ & w_i & \in & [a_i, b_i], \quad 1 \le i \le m \\ & V@R_\alpha\left(w^\top \xi\right) & \geq & a, \end{array} \tag{3.2}$$

where w_i denotes the relative weight of asset i in the portfolio, ξ_i the random return of asset i, $w^\top \xi = \sum_{i=1}^m w_i \xi_i$ and a_i, $b_i \in \mathbb{R}$ lower and upper bounds for the relative portfolio weights of the assets respectively. Note that since the returns ξ_i are random, (3.2) is a single stage stochastic optimization problem.

Because of the shortcomings of V@R mentioned above it is often replaced by the Average Value-at-Risk (AV@R, also called Conditional Value-at-Risk) in practical applications of optimization problems of the type (3.2) (see for example [10]). The *Average Value-at-Risk* AV@R$_\alpha$ of a random variable X is defined as

$$AV@R_\alpha(X) = \frac{1}{\alpha}\int_0^\alpha F_X^{-1}(t)dt = \frac{1}{\alpha}\int_0^\alpha V@R_t(X)dt, \quad \text{for} \quad 0 < \alpha \le 1, \tag{3.3}$$

where again $X \sim F_X$, and F_X^{-1} is the inverse distribution function of X. The reason for the popularity of AV@R is threefold

1. AV@R is relatively easy to incorporate into optimization problems like (3.2). In the case of discrete random variables a linear programming formulation exists, as shown in [65] and [78].

2. AV@R is a coherent risk measure (see [55]), and does not suffer from the problem described in (3.1).

3. Additionally the Average Value-at-Risk is a concave lower minorant of the Value-at-Risk and therefore the function $w \mapsto AV@R_\alpha(w^\top \xi)$ is a lower bound for the function $w \mapsto V@R_\alpha(w^\top \xi)$.

One could therefore conclude that the Value-at-Risk should be entirely replaced by the Average Value-at-Risk to circumvent the problems mentioned above. However, due to regulatory frameworks such as Basel II and Solvency II Value-at-Risk remains to be an industry standard and is widely used in portfolio planning. Therefore numerous approaches to solve problems of the form (3.2) either exactly or approximately have been proposed in the literature (for an overview see [81] and references therein).

This work builds on the difference of convex (D.C.) formulation of $V@R$ derived in [81], where a conical Branch-and-Bound algorithm for optimizing D.C. functions is used to find global optima of (3.2). Contrary to [81], where global solutions to problem (3.2) are found by a branch-and-bound algorithm, in this paper an approximate solution technique called difference of convex algorithm (DCA) to the D.C. formulation of the problem. Hence, the globality of the obtained solutions is lost, but the procedure is computationally tractable also for problems of realistic sizes. High quality solutions can be found in reasonable time also for real life portfolio selection problems.

This paper is organized as follows. In Section 3.2 D.C. reformulation of V@R is reviewed. Sections 3.3 discusses the DCA and its application to the problem at hand, while section 3.4 presents a set of numerical results based on real-world financial market data. Section 3.5 concludes the paper.

3.2 Reformulation of V@R$_\alpha$ as a D.C. function

We will consider the classical asset allocation problem based on the seminal paper [50] by Markowitz on portfolio selection. An investor has to choose a portfolio from a set of investment possibilities (financial assets) $\mathscr{I}$ with finite cardinality $m = |\mathscr{I}|$ to invest her available budget. The decision is taken in such a way that the expected return is maximized, while controlling for some kind of financial risk. Incorporation of a risk measure into portfolio optimization can either be achieved by bi-criteria optimization or by explicitly enforcing risk constraints in the optimization problem as done in (3.2).

To reformulate problem (3.2) into a D.C. problem, we need the following assumption.

Assumption 3. *The distribution of the random asset returns $\xi = (\xi_1, \ldots, \xi_m)$ is discrete with finitely many atoms, i.e. there are $S \in \mathbb{N}$ scenarios for the joint realizations of the random variables ξ_i. The $(m \times 1)$ vector of realizations of ξ in scenario s will be denoted by ξ^s and the probability of the scenario will be denoted by p_s.*

Assuming that random variables have discrete distributions is a common approach in stochastic programming and has the advantage that the random variables can easily be described in terms of empirical data, either by using historical realizations as scenarios directly, or by applying resampling techniques. The obvious advantage over models where the random variables are assumed to follow a distribution from a specific parametric family is that certain features like heavy tails, and non-normal skewness or kurtosis can easily be captured, especially in the multi-variate case.

As already mentioned V@R$_\alpha(X)$ is non-concave in X, and therefore the problem (3.2) cannot be solved by standard convex programming techniques. However, it turns out that – in the

discrete case – it can be reformulated to a difference of convex (D.C.) problem. To facilitate this reformulation we make Assumption 4.

Assumption 4. *Assume that all the scenarios have equal probabilities, i.e.*

$$p_s = \frac{1}{S}, \quad 1 \leq s \leq S.$$

Assumption 4 is made for the sake of simplicity of presentation. It is for example satisfied, if empirical data is used as scenarios but also if scenarios are generated from some parametric model and not explicitly re-weighted (in [81] it is shown, that the V@R functional is a D.C. function even in the case of unequal weights).

Given the above assumptions problem (3.2) becomes

$$\begin{array}{lrcl} \max & \frac{1}{S}\sum_{s=1}^{S} w^\top \xi^s & & \\ \text{s.t.} & \sum_{i=1}^{m} w_i & = & 1 \\ & w_i & \in & [a_i, b_i], \quad \forall 1 \leq i \leq m \\ & V@R_\alpha\left(w^\top \xi\right) & \geq & a. \end{array} \tag{3.4}$$

Although we will restrict ourselves to the above problem in the rest of the paper, our proposed methodology can easily be adapted to other cases, e.g. problems with more involved convex portfolio constraints, or problems with V@R as the objective function.

If the above two assumptions are satisfied, we can derive a D.C. formulation of the V@R in the following way. If X follows a discrete distribution taking the values $x_1, \ldots, x_n$ with equal probability (in our case X represents the anticipated (random) returns of the portfolio, i.e. $X = w^\top \xi$), then

$$AV@R_{\frac{k}{n}}(X) = \frac{n}{k}\sum_{i=1}^{k} x_{i:n}\frac{1}{n} = \frac{1}{k}\sum_{i=1}^{k} x_{i:n}, \tag{3.5}$$

where $x_{1:n} \leq x_{2:n} \leq \cdots \leq x_{n:n}$ is the set of ordered values of X. Therefore

$$V@R_\alpha(X) = x_{k:n} = kAV@R_{\frac{k}{n}}(X) - (k-1)AV@R_{\frac{k-1}{n}}(X)$$

with $k = [\alpha n]$. Hence under the above assumptions, $\text{V@R}_\alpha(X)$ can be written as the difference of the two concave functions

$$kAV@R_{\frac{k}{n}}(X) \text{ and } (k-1)AV@R_{\frac{k-1}{n}}(X).$$

Therefore, problem (3.4) becomes

$$\begin{array}{ll} \max & \frac{1}{S}\sum_{s=1}^{S} w^\top \xi^s \\ \text{s.t.} & \sum_{i=1}^{m} w_i = 1 \\ & w_i \in [a_i, b_i], \quad \forall 1 \leq i \leq m \\ & kAV@R_{\frac{k}{n}}(w^\top \xi) - (k-1)AV@R_{\frac{k-1}{n}}(w^\top \xi) \geq a, \end{array}$$

or equivalently can be written as a minimization problem with a D.C. constraint:

$$\begin{array}{ll} \min & -\frac{1}{S}\sum_{s=1}^{S} w^\top \xi^s \\ \text{s.t.} & \sum_{i=1}^{m} w_i = 1 \\ & w_i \in [a_i, b_i], \quad \forall 1 \leq i \leq m \\ & (-kAV@R_{\frac{k}{n}}(w^\top \xi)) - (-(k-1)AV@R_{\frac{k-1}{n}}(w^\top \xi)) \leq -a. \end{array} \tag{3.6}$$

3.3 Application of the DCA to problem (3.6)

The difference of convex algorithm (DCA) is an approximate solution method for unconstrained D.C. problems of the form

$$\inf\{f(x) = g(x) - h(x) : x \in \mathbb{R}^N\}, \tag{3.7}$$

where $g : \mathbb{R}^N \to \mathbb{R}$ and $h : \mathbb{R}^N \to \mathbb{R}$ are convex functions.

There are two versions of the DCA: the theoretically superior but computationally hard complete DCA and the so called simplified DCA.

The simplified DCA algorithm works by repeatedly solving the two convex optimization problems

$$\inf_{x\in\mathbb{R}^N}\{g(\bar{x}) - (h(x^k) + \langle x - x^k, y^k\rangle)\} \tag{3.8}$$

and

$$\inf_{y\in\mathbb{R}^N}\{h^*(y) - (g^*(y^{k-1}) + \langle x^k, y - y^{k-1}\rangle)\}. \tag{3.9}$$

The first problem can be viewed as a convex approximation of the original problem, while the second problem can be thought of as a convex approximation of the dual D.C. program

$$\inf_{y\in\mathbb{R}^N} h^*(y) - g^*(y).$$

Note that the solution to problem (3.8) is an element of $\partial g^*(y^k)$ and likewise the solution of (3.8) is an element of $\partial h(x^k)$.

The complete DCA differs from the simplified DCA in that it does not suffice to choose any arbitrary element of $\partial g^*(y^k)$ and $\partial h(x^k)$ but elements which satisfy

$$x^{k+1} \in \arg\min\{\langle x, y^k\rangle - h(x) : x \in \partial g^*(y^k)\} \tag{3.10}$$

$$y^k \in \arg\min\{\langle x^k, y\rangle - g^*(y) : y \in \partial h(x^k)\}. \tag{3.11}$$

However these problems are in general hard to solve, since they require the minimization of concave functions $-g^*$ and $-h$.

The DCA guarantees that $(g(x^k) - h(x^k))_{k\in\mathbb{N}}$ and $(h^*(y^k) - g^*(x^k))_{k\in\mathbb{N}}$ are decreasing sequences and the limit points of $(x^k)_{k\in\mathbb{N}}$ and $(y^k)_{k\in\mathbb{N}}$ are critical points of the primal problem (3.8) and the dual problem (3.9) respectively.

The DCA has proven to be very effective in solving many different D.C. programs (see for example [5, 8, 26, 41, 71]) and its low computational complexity makes it possible to work with large scale D.C. programs. For a more detailed discussion of DCA algorithm and its properties see [7].

To apply the DCA to problem (3.2) we first need to reformulate the problem to an unconstrained D.C. problem by an exact penalty approach: we start by defining the set C to be the set of portfolio decisions in which the convex constraints of problem (3.2) are fulfilled. Using the convex indicator function χ_C of C we can rewrite (3.6) as

$$\begin{array}{lll} \min & -\frac{1}{S}\sum_{s=1}^{S} w^\top \xi^s + \chi_C(w) & \\ \text{s.t.} & (-kAV@R_{\frac{k}{n}}(w^\top\xi)) - (-(k-1)AV@R_{\frac{k-1}{n}}(w^\top\xi)) & \leq \quad -a \end{array} \tag{3.12}$$

and thus penalize for not fulfilling the respective convex constraints represented by C.

To penalize for the last remaining constraint note that

$$\begin{aligned} max(V@R_\alpha(w^\top\xi)+a,0) &= max\left(-kAV@R_{\frac{k}{n}}(w^\top\xi)+a,-(k-1)AV@R_{\frac{k-1}{n}}(w^\top\xi)\right) \\ &+ (k-1)AV@R_{\frac{k-1}{n}}(w^\top\xi). \end{aligned} \tag{3.13}$$

using the fact that the pointwise maximum of finitely many D.C. functions $f_i = g_i - h_i$ with $1 \le i \le M \in \mathbb{N}$ can be written as

$$\max_i f_i = \max_i \left(g_i - \sum_{j\neq i} h_j\right) - \sum_j g_j.$$

Using the results on exact penalization of D.C. programs from [9], for some $\tau > 0$ we finally rewrite (3.12) as the equivalent problem

$$\begin{aligned} \min \quad - \quad & \frac{1}{S}\sum_{s=1}^{S} w^\top\xi^s + \chi_C(w) + \tau\left[max\left(-kAV@R_{\frac{k}{n}}(w^\top\xi)+a,-(k-1AV@R_{\frac{k-1}{n}}(w^\top\xi)\right)\right. \\ + \quad & \left.(k-1)AV@R_{\frac{k-1}{n}}(w^\top\xi)\right]. \end{aligned} \tag{3.14}$$

Now define

$$\begin{aligned} g(w) &= -\frac{1}{S}\sum_{s=1}^{S} w^\top\xi^s + \chi_C(w) + \tau max\left(-kAV@R_{\frac{k}{n}}(w^\top\xi)+a,-(k-1)AV@R_{\frac{k-1}{n}}(w^\top\xi)\right) \\ h(x) &= -\tau(k-1)AV@R_{\frac{k-1}{n}}(w^\top\xi), \end{aligned}$$

we have derived a problem of the form (3.7) and thus can apply the DCA. In our case (3.8) is the linear program

$$\begin{array}{lll} \min & -\frac{1}{S}\sum_{s=1}^{S} w^\top\xi^s + \tau M - \langle w, y^k\rangle & \\ \text{s.t.} & -kAV@R_{\frac{k}{n}}(w^\top\xi)+a & \le M \\ & -(k-1AV@R_{\frac{k-1}{n}}(w^\top\xi) & \le M \\ & w \in C, & \end{array} \tag{3.15}$$

since the constant terms $h(w)$ and $\langle w^k, y^k\rangle$ do not influence the solution of (3.8) but only the optimal value.

To solve problem (3.9) we observe that for a given w^k any solution w^* of

$$\inf_{w^*\in\mathbb{R}^N}\{h^*(w^*) - \langle w^k, w^*\rangle)\}.$$

is just an element in $\partial h(w^k)$. To find such elements we recall from (3.5), that

$$h(w) = -\tau(k-1)AV@R_{\frac{k-1}{n}}(w^\top\xi) = -\tau\sum_{i=1}^{k-1} r_{i:n}(w)$$

where the $r_{i:n}$ are the ordered returns depending on the portfolio w. The dependence of h on w is twofold: the return $r_i = w^\top \xi^i$ in each of the scenarios ξ^i is dependent on w and the ordering of the returns (which in turn determines which of the r_i are summed) depends on w as well. These two dependencies render the mapping $w \mapsto AV@R$ piecewise linear. For a given w^k the subgradients are therefore convex combinations of the gradient of the linear functions defining $AV@R_{\frac{k-1}{n}}$ locally. Let us observe this a little more closely and note that $\partial h(w^k) = \{\nabla h(w)\}$ if $r_{(k-1):n} < r_{k:n}$ because in this case $w \mapsto AV@R_{\frac{k-1}{n}}(w)$ is linear in a neighborhood of w^k.

If on the other hand side the following situation occurs for a given w^k

$$r_{1:n} \leq \cdots r_{s-1:n} < r_{s:n} = \cdots = r_{k-1:n} = r_{k:n} = \cdots = r_{t:n} < r_{t+1:n} \leq \cdots \leq r_{n:n}$$

then $\nabla h(w^k)$ does not exist, i.e. $\partial h(w^k)$ contains more than one element. Let us assume w.l.o.g. that the ordering of the returns above coincides with the scenario numbers, i.e. $r_{i:n} = w^{k\top}\xi^i = r_i$ then following the above remark $\partial h(w^k)$ can be written as the convex hull of the following vectors

$$V = \left\{ \sum_{i=1}^{s-1} \nabla r_i(w) + \sum_{j \in J} \nabla r_j(w) : J \subset \{s, \ldots, t\}, |J| = (k-s-2) \right\}. \tag{3.16}$$

To solve problem (3.11) in our setting we therefore just have to find

$$y^k \quad \in \quad \arg\min\{\langle x^k, y\rangle - g^*(y) : y \in V\}$$

which is easy, since $|V| < \infty$. Note that $g^*(y)$ for a $y \in V$ is just the optimal value of (3.15) for $y^k = y$. Therefore we can actually solve the non-convex problem (3.11) by enumeration. To apply the complete DCA we would have to solve (3.10) as well. Unfortunately problem (3.10) can not be solved in that fashion and we have to contend ourselves with finding an arbitrary element of $\partial g^*(y^k)$ by solving problem (3.15). We therefore apply a hybrid of simplified and complete version of the DCA.

3.4 Applications

In this section we test the algorithm by applying it to real market data. In section 3.4.1 we compare the globally optimal portfolios obtained in [81] to the optimal portfolios from the DCA. In sections 3.4.2 we compute optimal portfolios for a large data set to demonstrate the applicability of the approach to realistically sized problems, which can no longer be solved to global optimality.

All optimization problems were solved using the solver MOSEK, version 5. The calculations we performed on a notebook with a Pentium Mobile Processor (1.8 GHz) 1.5GB RAM using Windows XP SP 2.

3.4.1 Comparison with global optima

In the following we compare results obtained for problem (3.6) by applying the DCA with results we obtained in [81]. Weekly closing values of the following 5 indices have been used

Name	Average Return	Variance	$V@R_{.05}$	$AV@R_{.05}$
US Long Bond	0.9988	0.0002908	0.9762	0.9569
Standard & Poors 100	1.0013	0.0001658	0.9773	0.9749
Nasdaq 100	1.0008	0.0004351	0.9625	0.9547
FTSE 100	1.0026	0.0003147	0.9735	0.9614
Hang Seng	1.0033	0.0003226	0.9727	0.9637

Table 3.1: Characteristics of the weekly returns of the five indices subsequently used in portfolio optimization. Time frame: 2004-2005.

to calculate the discrete return scenarios: US Long Bond, Standard & Poors 100, Nasdaq 100, FTSE 100, and Hang Seng. Table 3.1 gives an overview of the characteristics of the used data.

In Figure 3.1 we present efficient frontiers for the bi-criteria V@R problem (3.6) obtained by varying the parameter a. It turns out that the interesting range for a is between the V@R of the portfolio that consists only of the asset with the highest return (i.e. Hang Seng with $V@R_{.05}$ of 0.9727 and expected weekly return 1.0033) and the last feasible value of a, which is 0.987. We choose $\alpha = 0.05$ and $a_i = 0$, $b_i = 1$ for all $i = 1, \ldots, m$.

In this range we performed optimizations varying the a in 0.0005 steps. In Fig. 3.1 the dependence of the maximal returns and the portfolio compositions of the optimal portfolios on the acceptance level a are depicted.

We see that the performance of the DCA relative to the Branch-and-Bound algorithm mainly depends on the parameter a. In particular three observations can be made looking at the graph.

1. Surprisingly the DCA finds the global optima for the respective problems for a ranging from 0.972 to 0.9785. Also the portfolio compositions obtained by the DCA and the Branch-and-Bound algorithm are very similar in this area as Figure 3.1 shows.

2. For values of a ranging from 0.9785 to 0.9835 the DCA finds solutions but these solutions are inferior to the solutions found by the global Branch-and-Bound method for some of the points in on the efficient frontier.

3. For a from 0.984 to 0.987 the DCA is not able to find feasible points for the problem (3.6).

The decreasing quality of the solutions with the parameter a are in line with the observations made in [81], where it was found that increasing a makes the problem computationally harder (expressed in terms of number of iterations and computing time needed by the Branch-and-Bound algorithm).

However, the fact that the DCA actually finds global solutions shows that the method in general yields good solutions of (3.6). Because of the comparatively low runtime the algorithm will be preferable to a global solution approach in practice. See table 3.2 for a comparison of the respective computing times of the DCA and the Branch-and-Bound algorithm.

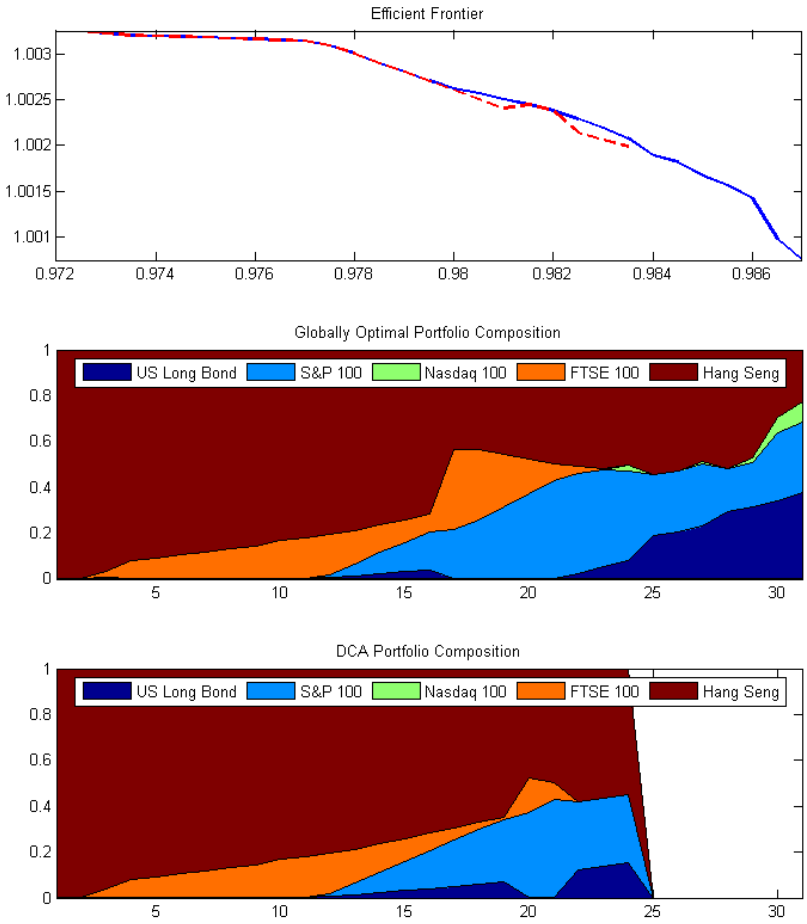


Figure 3.1: Efficient frontier and optimal portfolios for problem (3.6) computed with the Branch-and-Bound algorithm from [81] (blue) and with the DCA (dashed, red).

a	Time Branch-and-Bound	Time DCA
0.9730	5.7438	0.7463
0.9740	121.92	0.7377
0.9750	125.16	0.7713
0.9760	68.188	0.7634
0.9770	155.43	0.8295
0.9780	738.96	0.693
0.9790	2328.3	0.7054
0.9800	3336	0.7566
0.9810	2305.6	0.7560
0.9820	2325.5	0.812
0.9830	3504.6	0.7122

Table 3.2: Runtime of the DCA algorithm versus the Branch-and-Bound algorithm in seconds.

3.4.2 Application of DCA to large data sets

To demonstrate the ability of the DCA to solve bigger problems and therefore its applicability in practice, we repeat the analysis of the last section with daily return data of the assets comprising the Dow Jones Industrial index. Our scenario set consists of the 251 daily returns for 30 assets observed in the year 2007. Furthermore we choose $\alpha = 0.1$ and $a_i = 0$ and $b_i = 1$ for all $i = 1,\ldots,m$. The size of the scenario set and the choice of a relatively large α makes it impossible to solve the respective problem to global optimality.

To have a benchmark of the performance of the DCA algorithm we compare the results with the results obtained by a variant of a well known heuristic to optimize $V@R$ portfolios (see [47]). The described method is a simple yet effective technique to maximize the Value-at-Risk of a portfolio. The main idea is the approximation of the Value-at-Risk by the Average Value-at-Risk and iterative application of a truncation operation. The algorithm performs well even on vast data sets and is widely applied in the industry. Algorithm A1 from [47] adapted to our situation is described below. For a justification of the steps in the algorithm we refer to the original paper.

1. Fix C as a lower bound on expectation, a parameter for the Value-at-Risk α and a parameter for the heuristic $0 < \zeta < 1$.

2. Set $\alpha_0 = \alpha$, $i_0 = 0$ and $n = 0$.

3. Solve the problem

$$\begin{array}{lllll} max_x & AV@R_{\alpha_n}(\sum_{i=i_n}^{S} x^\top \xi^i) & & & \\ s.t. & \mathbb{E}(w^\top \xi) & \geq C & & \\ & w^\top \xi^i & \leq \gamma, & i \leq i_n & \\ & w^\top \xi^i & \geq \gamma, & i > i_n & \\ & a_i \leq x & \leq b_i, & \forall 1 \leq i \leq m & \end{array}$$

Expectations	$V@R_{0.1}$ (A1)	$V@R_{0.1}$ (DCA)	difference	active A1	active DCA
1.002537	0.984289	0.984300	0.000012	30	3
1.002468	0.984660	0.984800	0.000140	2	3
1.002409	0.985002	0.985300	0.000298	2	3
1.002337	0.985417	0.985800	0.000383	3	4
1.002264	0.985764	0.986300	0.000536	3	5
1.002163	0.985755	0.986800	0.001045	3	5
1.002058	0.986191	0.987300	0.001109	30	5
1.001945	0.986377	0.987800	0.001423	9	4
1.001811	0.987042	0.988300	0.001258	29	5
1.001657	0.987873	0.988800	0.000927	30	5
1.001504	0.988458	0.989300	0.000846	5	6
1.001409	0.989338	0.989800	0.000462	7	5

Table 3.3: Optimal expected returns, Value-at-Risk and number of active assets for daily return data for Dow Jones assets for Algorithm A1 from [47] and the DCA.

4. Call the solution of the above problem w_n and sort the scenarios ξ^i according to their returns $r_i = w_n^\top \xi^i$.

5. Set $n = n+1$, $b_n = \alpha + (1-\alpha)(1-\zeta)^n$, $i_n = [S(1-b)]$ and $\alpha_n = 1 - \frac{1-\alpha}{b_n}$.

6. If $i_n \leq [S/\alpha]$ go to step 3 otherwise exit.

The comparison between the two methods is carried out by first running the DCA algorithm for a set of given V@R constraints and subsequently using the optimal expectations as constraints in the above heuristic algorithm. In our concrete example the range of r.h.s. values a for the $V@R$ problem is chosen to be $[0.9843, 0.99]$ and is traversed by a loop in 0.0005 steps. The lower bound 0.9843 is the value of the right hand side of problem (3.2) for which the optimal portfolio consists only of the asset with the highest expected return. The upper bound is the highest value for which the DCA still finds a feasible solution for (3.2).

The results of the comparison are compiled in Table 3.3 and depicted in Figure 3.2.

The results show that the approximate algorithm from [47] yields similar performance for low values of a, while the performance decreases relative to the DCA when a gets bigger, i.e. the problem gets harder to solve. Table 3.3 also shows that the number of active assets, i.e. the number of assets whose portfolio contributions are different from 0, is in most cases smaller for solutions obtained with the DCA algorithm, implying that the DCA is closer to the real $V@R$ bound and hence chooses a more risky (and less diversified portfolio).

3.5 Conclusion

In this paper we reviewed the D.C. formulation of the Value-at-Risk functional presented in [81]. We used the representation to formulate a classical Markowitz type portfolio selection

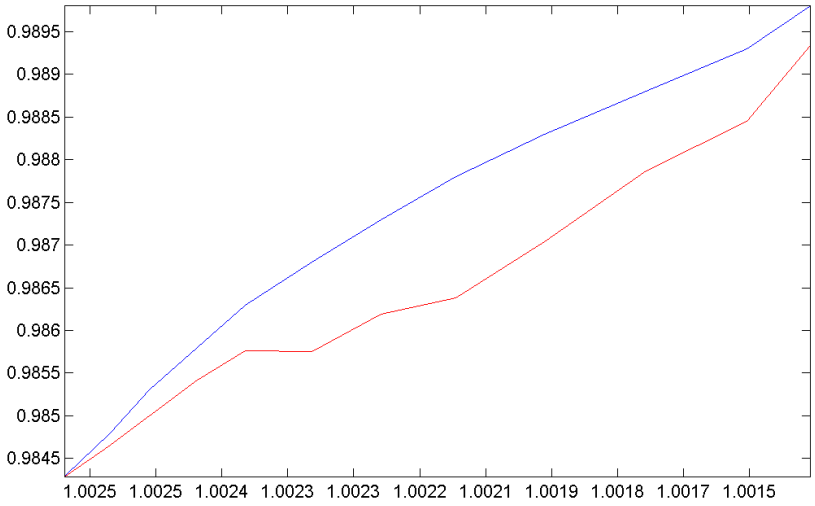


Figure 3.2: Efficient frontiers for $V@R_{0.1}$. Approximate algorithm is depicted in red, while the results from the DCA are blue.

problem with $V@R$ constraint. This problem in turn is approximately solved by the DCA, a generic approximate solution technique for D.C. problems. We demonstrated that the DCA algorithm yields good results by comparing the solutions obtained by the DCA with the global solutions from [81]. The DCA finds global optima for small values of a and generally produces slightly suboptimal portfolios for higher values of a.

To demonstrate the applicability to problems of realistic size, we tested the algorithm on a data set that consists of the daily return data of the Dow Jones assets for the year 2007 (i.e. the 30 assets and 251 data points). We demonstrate that in this setting the proposed algorithm clearly outperforms existing approximation schemes for the Value-at-Risk.

CHAPTER 4

A Framework for Optimization under Ambiguity

In this paper, single stage stochastic programs with ambiguous distributions for the involved random variables are considered. Though the true distribution is unknown, existence of a reference measure $\hat{P}$ enables the construction of non-parametric ambiguity sets as Kantorovich balls around $\hat{P}$. The resulting robustified problems are infinite optimization problems and can therefore not be solved computationally. To solve these problems numerically, equivalent formulations as finite dimensional non-convex, semi definite saddle point problems are proposed. Finally an application from portfolio selection is studied for which methods to solve the robust counterpart problems explicitly are proposed and numerical results for sample problems are computed.

4.1 Introduction

In this paper, a general framework, that can be used to deal with model uncertainty in stochastic optimization models, is developed. More specifically the aim is to robustify single stage stochastic optimization models with respect to *uncertainty about the distributions* of the random variables involved in the formulation of the stochastic program.

The paper extends work that was done in [54] to a more general setting.

We consider the following stochastic programming problem

$$\begin{array}{rll} \sup_{x\in\mathbb{R}^m} & F(x,P) & \\ s.t. & G(x,P) & \leq 0 \\ & H(x) & \leq 0, \end{array} \tag{4.1}$$

where P is an element of $\mathscr{P}(\mathbb{R}^d)$ the set of probability measures on $\mathbb{R}^d$, $F : \mathbb{R}^m \times \mathscr{P}(\mathbb{R}^d) \to \mathbb{R}$ is concave in the first variable and linear in the probability measure P, the function $G : \mathbb{R}^m \times \mathscr{P}(\mathbb{R}^d) \to \mathbb{R}$ is convex in the first variable and $H : \mathbb{R}^m \to \mathbb{R}$ is a convex function.

If the measure P is known and the involved functions are *nice* the above problem can be solved either explicitly or at least numerically, thus yielding a solution x which takes the statistical uncertainty about the random variables in the problem formulation into account. However, in many practical situations the measure P is not known exactly. It is known that many stochastic programming problems are very sensitive to changes in the underlying distributions of the

random variables (see [46] or [23]). Therefore it is reasonable to consider robust versions of problem (4.1), i.e. to introduce a so called ambiguity set $\mathscr{B} \subseteq \mathscr{P}(\mathbb{R}^d)$ which represents the ambiguity about the real measure P. More precisely it is assumed that (with high probability) the real measure P is an element of $\mathscr{B}$.

Given $\mathscr{B}$, the robust counterpart of (4.1) that is considered looks like

$$\begin{array}{rlll} \sup_{x\in\mathbb{R}^m} \inf_{Q\in\mathscr{B}} & F(x,Q) & & \\ s.t. & G(x,Q) & \leq 0, & \forall Q \in \mathscr{B} \\ & H(x) & \leq 0. & \end{array} \tag{4.2}$$

Note that the robustness is achieved by a worst case approach in the objective function as well as in the constraints. However, we work with an ambiguity set $\mathscr{B}$ which usually is relatively small in comparison to $\mathscr{P}(\mathbb{R}^d)$, and the worst case is relative to this set.

Like in [54] we construct the ambiguity sets as Kantorovich neighborhoods of the empirical measure $\hat{P}_n$ (constructed from a sample of size n), where

$$\hat{P}_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{\xi_i} \tag{4.3}$$

In specific, we set

$$\mathscr{B} = \mathscr{B}_\varepsilon(\hat{P}) = \{Q \in \mathscr{P}(\mathbb{R}^d) : d_K(\hat{P}, Q) \leq \varepsilon\} \tag{4.4}$$

where d_K is the so called Kantorovich (or Wasserstein) metric for probability measures defined as

$$d_K^r(P_1,P_2) \quad = \quad \inf_{\gamma(A\times\Omega)=P_1(A),\ \gamma(\Omega\times A)=P_2(A)} \left(\int ||x-y||^r d\gamma(x,y) \right)^{\frac{1}{r}}, \tag{4.5}$$

where $\|u-v\|_1 = \sum_i |u_i - v_i|$ (see [36] for a short introduction to the subject of probability metrics and [61] for an extensive review). Here, we will work with $r = 1$. The space $(\mathscr{P}(C), d_K^1)$ is a compact, separable, complete metric space for a compact set $C \subseteq \mathbb{R}^d$. The topology on $(\mathscr{P}(C), d_K^1)$ is the weak (star) topology induced by the continuous functions on C.

In the terminology of [14] the problem with $P = \hat{P}$ serves as the nominal or baseline instance of the problem, while the ambiguous counterpart is the robustified version, where ε quantifies the degree of ambiguity in the model.

The purpose of this paper is to solve problem (4.2) with $\mathscr{B} = \mathscr{B}_\varepsilon(\hat{P})$ without further assumptions. The formulation (4.2) suffers from the drawback, that it is numerically intractable for the following three reasons:

i. The problem (4.2) has infinitely many constraints.

ii. The solutions of the problems (4.2) are elements of an infinite dimensional space (together with (i) this makes the problem infinite).

iii. The problem (4.2) is not a standard optimization but a maximin problem. Finding a solution involves finding a saddle point of the objective function, which usually is harder than finding a maximum or a minimum.

Methods to treat problems (i.) and (iii.) were already developed and tested in [54]. The focus of the current paper is therefore problem (ii.) and the integration of the results in the already existing framework.

The paper is organized as follows: in Section 4.2 we give an overview of papers that deal with problems in similar settings, Section 4.3 is concerned with the reformulation of problem (4.2) in such a way that it becomes numerically tractable, in Section 4.4 the framework is applied to a concrete robust portfolio composition problem and in Section 4.5 numerical results for this particular problem are discussed. Section 4.6 concludes the paper.

4.2 Overview over existing literature

There exists a large and fast growing literature which deals with similar problems as discussed in this paper. The first papers dealing with subject are works by Dupačová (see [27–29, 82]) and more recent works include [19, 21, 31, 33, 39, 54, 59, 68, 69, 69, 76]. A comprehensive summary is beyond the scope of this paper, we therefore briefly discuss a few of the mentioned papers, which are in some sense similar to the approach taken here.

The approaches proposed up till now use strong conditions on $\mathscr{B}$ to keep problem (4.2) computationally tractable. In [39, 76] for example a classical mean-risk portfolio composition problem is studied and it is assumed that the set $\mathscr{B}$ consists of normal distribution that are in some sense close to the normal distribution with empirically measured mean and variance. However, the assumption of normality of asset returns – though wide spread in the literature – received criticism based on empirical evidence (see for example [1, 34, 63]).

In [68] the authors assume the ambiguity set to be of the form

$$\mathscr{B} = \{P \in \mathscr{P} : P_1 \leq P \leq P_2\}$$

where P_1 and P_2 are given measures and $P \leq Q$ iff $P(A) \leq Q(A)$ for all Borel sets A and show that under these conditions a problem similar to (4.2) can be solved efficiently.

In [14, 15] different types of finite dimensional robust problems with more general structures of the ambiguity set (referred to as ellipsoidal uncertainty) are solved. However, since the problem of an infinite dimensional decision space does not arise in these works and the core technique is to get rid of infinitely many constraints by dualization techniques and subsequent reformulation to a definite problem, it is questionable whether this approach would still yield meaningful results in general metric spaces and for general convex functions.

In [19] the ambiguity sets $\mathscr{B}$ are constructed as measures with Kullback-Leibler distance less than ε to the empirical measure $\hat{P}_n$. The Kullback-Leibler distance between two discrete measures P and Q is defined as

$$d_{KL}(P,Q) = \sum_{i=1}^{n} p_i log\left(\frac{p_i}{q_i}\right),$$

where p_i and q_i are the probabilities of the points under the measures P and Q respectively. The definition immediately implies that only those measures can be compared that have the same atoms (to be more precise the measures Q and P have to have common atoms x_i). This in turn

implies that all the measures that can be included in $\mathscr{B}$ are measures with atoms identical to the empirical distribution. We will discuss the shortcomings of this approach below.

Also note that d_{KL} does not reflect the impact a change in the distribution has on classical probability functionals. To see this consider two points x_1, x_2 and two discrete distributions P, Q on these points with $P(x_1) = \varepsilon$, $P(x_2) = 1-\varepsilon$ and $Q(x_1) = 1-\varepsilon$, $Q(x_2) = \varepsilon$, then

$$d_{KL}(P,Q) = 2\varepsilon \log \frac{\varepsilon}{1-\varepsilon} + \log \frac{1-\varepsilon}{\varepsilon} \xrightarrow{\varepsilon \to 0} \infty. \tag{4.6}$$

Note that the distance between P and Q is not dependent on the location of x_1 and x_2. If x_1 and x_2 are close, then $|F(P) - F(Q)|$ is also small for most probability functionals F (like for example the expectation) which in turn implies that the notion of distance modeled by d_{KL} is not appropriate for stochastic programming.

In [54] the ideas on the construction of $\mathscr{B}$ are similar to the approach presented in this paper. However, when solving problem (4.2) only a small fraction of the measures in $\mathscr{B}_\varepsilon(\hat{P}_n)$ are actually taken into consideration. In fact all the measures considered are discrete measures with atoms identical to those of $\hat{P}_n$. This restriction can be justified by arguing that the atoms of $\hat{P}_n$ are sufficiently dense in the set of all possible realizations of the random variables involved in the problem formulation. However, if the number of points is too small or the number of random elements is too big this assumptions becomes unrealistic. It can also be argued that the future distribution of the random variables of interest may change and therefore scenarios that were not sampled in the empirical measure actually become possible outcomes of future realizations. In the context of portfolio optimization this aspect becomes especially important if one thinks of the structural breaks in the distributions of asset returns observed frequently on the markets.

The advantage of exclusively varying the probabilities of the scenarios of the empirical measure is that the solution of problem (4.2) becomes much simpler, since $\mathscr{B}$ can be described by a subset of a finite dimensional space. However, we will show in section 4.5, considering all distributions in $\mathscr{B}_\varepsilon(\hat{P}_n)$ substantially changes the results for realistic data sets.

Thus we conclude that in order to make a decision which is robust with respect to model misspecification, we should take into account all the measures in $\mathscr{B}_\varepsilon(\hat{P}_n)$ and not only those which put mass on pre-specified points. Since we want to solve the problem numerically we have to reduce the complexity of the set $\mathscr{B}_\varepsilon(\hat{P}_n)$ and this reduction should be achieved without sacrificing on our goal to solve problem (4.2) for the whole of $P \in \mathscr{B}_\varepsilon(\hat{P}_n)$.

Summarizing, our approach is different from the aforementioned in the sense, that we do not restrict the set $\mathscr{B}$ by structural assumptions other than that all the measures considered in the robustification have to be close to the empirical measure $\hat{P}_n$ in the Kantorovich sense. This enables us to construct the sets $\mathscr{B}$ as *confidence balls* around the empirical measure $\hat{P}_n$.

4.3 Reducing the problem

As a first step necessary to achieve numerical tractability of problem (4.2) we reduce the set of possible distributions from the full Kantorovich ball to a subset $\mathscr{B}'$, whose elements can be described by finite dimensional vectors. In particular $\mathscr{B}'$ should fulfill

$$\inf_{Q \in \mathscr{B}} F(x,Q) = \inf_{Q \in \mathscr{B}'} F(x,Q). \tag{4.7}$$

This would allow us to numerically solve the problem on the right hand and substitute the solution for the problem on the left hand side. Similarly the condition

$$\sup_{Q\in\mathscr{B}} G(x,Q) = \sup_{Q\in\mathscr{B}'} G(x,Q) \tag{4.8}$$

would enable us to treat the constraints in a finite dimensional setting.

A natural approach to restrict the set of distributions would be to consider the discrete distributions on finitely many atoms. To this end we note that every distribution can be approximated in the Kantorovich distance by a discrete distribution to arbitrary precision (in other words the discrete distributions are a dense subset of $(\mathscr{P}, d_K^1)$), i.e.

$$\forall P \in \mathscr{P}(\mathbb{R}^d),\ \forall \varepsilon > 0,\ \exists Q \in \mathscr{P}^D : d_K(P,Q) < \varepsilon$$

where $\mathscr{P}^D \subseteq \mathscr{P}$ is the set of discrete distributions.

However this does not help in our setting, since we would need to approximate every $P \in \mathscr{B}_\varepsilon(\hat{P}_n)$ and therefore would need a bound $N \in \mathbb{N}$ on the number of atoms in Q that works uniformly for all $P \in \mathscr{P}(\mathbb{R}^d)$ and a given $\varepsilon > 0$, i.e.

$$\forall P \in \mathscr{P}(\mathbb{R}^d),\ \forall \varepsilon > 0\ \exists N(d) \in \mathbb{N},\ \exists Q \in \mathscr{P}^D : |Q| = N, d_K(P,Q) < \varepsilon,$$

where $|Q|$ denotes the number of atoms of the discrete distribution Q.

Such results can only be obtained by covering arguments of the domain space of the measure. This requires a restriction to a bounded support set $B \subseteq \mathbb{R}^d$ and leads to an exponential increase in N with the dimension of the space.

We therefore restrict our attention to measures, that are actually possible candidates for optimizers of the problem (4.7). Hence, instead of trying to approximate every element of $\mathscr{B}_\varepsilon(\hat{P}_n)$, we only consider the extremal elements of $\mathscr{B}_\varepsilon(\hat{P}_n)$.

Definition 4.1 (Extreme Point)**.** *Let $C \subseteq E$ be a convex set in a vector space E. A point $x \in C$ is called an extreme point of C, if $C\setminus\{x\}$ is still a convex set. We denote the set of all extreme points of C by $ext(C)$.*

This approach is motivated by the following result:

Theorem 4.1 (Bauer Minimum Principle)**.** *Let E be a Hausdorff locally convex vector space (LCS), $C \subset E$ be a non-empty compact convex set and f a concave lower semi-continuous function. Then f attains its minimum over C at an extreme point of C.*

Proof. See [24], Theorem 25.9. □

Remark: The extremals of the set $\mathscr{P}(\mathbb{R}^d)$ are the Dirac measures δ_x with $x \in \mathbb{R}^d$. To see this fix a measure $P \in \mathscr{P}(\mathbb{R}^d)$ which is not a Dirac measure. Then there exist disjoint sets A_1, A_2 with $\mathbb{R}^d = A_1 \cup A_2$ such that $P(A_i) > 0$ for $i = 1,2$. Define

$$P_i(B) = \frac{1}{P(A_i)} P(A_i \cap B), \quad \text{for } i = 1,2$$

Then it follows that $P = P(A_1)P_1 + (1 - P(A_1))P_2$ and thus P is not an extremal element of $\mathscr{P}(\mathbb{R}^d)$.

Therefore it seems plausible, that the extremals of $\mathscr{B}_\varepsilon(\hat{P}_n) \subseteq \mathscr{P}$ are also Dirac measures. However, as evident from the following example, the situation is more complicated for Kantorovich balls.

Example: Let $\varepsilon > 0, x_0, x_1, x_2 \in \mathbb{R}^d$ with $||x_0 - x_1||_1 < \varepsilon < ||x_0 - x_2||_1$. Further, let $f : \mathbb{R}^d \to \mathbb{R}$ be continuous and such that

$$\{x_1\} = \arg\max \{f(x) : ||x_0 - x||_1 \leq \varepsilon\}$$

and $f(x_2) > f(x_1)$. Define the linear functional $F : \mathscr{P} \to \mathbb{R}$ by $F(P) = \int f dP$. If $||x - x_0||_1 \leq \varepsilon$ and $x \neq x_1$, it holds that

$$F(\delta_x) \leq F(\delta_{x_1}) < F(\delta_{x_1} p + \delta_{x_2}(1-p)), \quad \forall 0 < p < 1 \tag{4.9}$$

Since the set $argmax_{P \in \mathscr{B}_\varepsilon(\delta_{x_0})} F(P)$ has to contain an extremal point, (4.9) shows that there are extremal points of $\mathscr{B}_\varepsilon(\delta_{x_0})$, which are not Dirac measures.

We therefore have to consider also discrete measures that assign mass to more than one point. Since it proves hard to characterize the points of Kantorovich balls directly, first the so called exposed points are studied. These admit a more convenient characterization in our case. In a second step the results for the exposed points are carried over to the extremals.

Definition 4.2 (Exposed Point)**.** *Let $C \subseteq E$ be a convex set in a LCS. $c \in C$ is an exposed point of C is, if it is possible to separate the set $C \setminus \{c\}$ from c via a continuous affine functional. In other words there exists a continuous affine functional l, such that $l(c) > l(y)$, $\forall y \in C$ or equivalently $argmax_{y \in C} l(y) = \{c\}$. We denote the set of exposed points of C by $exp(C)$.*

Remark: Every exposed point c of C is also an extreme point, since if there would be points $a, b \in C$ which are both not equal c and $c = \frac{a}{2} + \frac{b}{2}$ then

$$l(c) = \frac{1}{2} l(a) + \frac{1}{2} l(b) < l(c).$$

Remark: For an exposed point μ of a convex set C in the space $(\mathscr{P}(\mathbb{R}^d), d_K^1)$ it is therefore possible to find a continuous bounded function $g : \mathbb{R}^d \to \mathbb{R}$ such that

$$\{\mu\} = argmax_{\lambda \in C} \int g d\lambda. \tag{4.10}$$

The next example demonstrates that Theorem 4.1 does not hold for exposed points instead of extreme points and that not every extreme point is also exposed.

Example: Define

$$f(x) = \begin{cases} (|x| - 1)^2, & x \notin [-1, 1] \\ 0, & x \in [-1, 1] \end{cases}$$

and let

$$A = epi(f) = \{(x, y) : f(x) \leq y\}$$

Define $B = \{(x,y) \in \mathbb{R}^2 : y \leq 1\}$ and $C = A \cap B$. The points $(-1,0)$ and $(1,0)$ are extremal points of C, but not exposed points. The continuous, convex function $g(x,y) = -y$ takes its maximum value over the compact convex set C on the line segment $\{(x,0) : x \in [-1,1]\}$. All boundary points of C are extreme points and

$$exp(C) = \{(x,(|x|-1)^2) : -2 \leq x < -1\} \cup \{(x,(|x|-1)^2) : 1 < x \leq 2\}$$

Notice that the points $(1,0)$ and $(-1,0)$ are extreme points of C at which the maximum of g is attained . Although in every neighborhood of $(1,0)$ and $(-1,0)$ there are exposed points the points themselves are not exposed.

We need the following Theorem from [66] to characterize the distributions that are exposed points of $\mathscr{B}$.

Theorem 4.2. *Let $f_1, \ldots, f_n$ be given real-valued Borel measurable functions on a measurable space Ω. Let μ be a probability measure on Ω such that each f_i is integrable with respect to μ. Then there exists a probability measure μ' with finite support on Ω satisfying $|\mu'| \leq n+1$ and*

$$\mu'(f_i) = \mu(f_i), \qquad \forall i = 1, \ldots, n.$$

Theorem 4.3. *Consider a discrete measure $\hat{P}_n$ on the n points $x_1, \ldots, x_n$ in $\mathbb{R}^d$ (with respective probabilities $p_1, \ldots, p_n$) and the Kantorovich ball $\mathscr{B}_\varepsilon(\hat{P}_n)$ with radius $\varepsilon > 0$ around $\hat{P}_n$. Then all the exposed points of $\mathscr{B}_\varepsilon(\hat{P}_n)$ are in $\mathscr{P}_{(n+3)}$ – the discrete distributions with at most $(n+3)$ atoms.*

Proof. In view of the above remarks, to show that all the exposed points of $\mathscr{B}_\varepsilon(\hat{P}_n)$ are discrete, it is enough to show that for any measure $P \in \partial\mathscr{B}_\varepsilon(\hat{P}_n)$ and every continuous bounded function g, there is a discrete measure $\tilde{P}$ (on $(n+3)$ points) such that $\int g dP = \int g d\tilde{P}$ and $d_K(\hat{P}_n, \tilde{P}) \leq \varepsilon$. This would establish that there is no argmax-set like (4.10) that consists only of a single non-discrete measure, i.e. all the exposed points have to be discrete measures.

If $d_K(\hat{P}_n, P) = \varepsilon$ and $\int g dP = c$, then there exists a measure γ on $\mathbb{R}^d \times \mathbb{R}^d$ such that

$$\begin{aligned}
\int ||x-y||_1 d\gamma(x,y) &= \varepsilon && (4.11)\\
\int g(y) d\gamma(x,y) &= c && (4.12)\\
\gamma(\mathbb{R}^d \times A) &= P(A), \qquad \forall A \subseteq \mathbb{R}^d \text{ measurable} &&\\
\gamma(A \times \mathbb{R}^d) &= \hat{P}_n(A), \qquad \forall A \subseteq \mathbb{R}^d \text{ measurable} &&
\end{aligned}$$

Now invoke Theorem 4.2 for γ with conditions (4.11) and (4.12). The functions f_i are: $f_1(x,y) = ||x-y||_1$, $f_2(x,y) = g(x)$ and additionally

$$f_{i+2}(x,y) = p_i \, \mathbb{1}_{\{x_i\} \times \mathbb{R}^d}(x,y), \qquad \text{for } i = 1, \ldots, n. \tag{4.13}$$

This results in $(n+2)$ moment conditions and therefore the theorem yields a measure $\tilde{\gamma}$ on $\mathbb{R}^d \times \mathbb{R}^d$ sitting on $(n+3)$ points in $\mathbb{R}^d \times \mathbb{R}^d$. For $\tilde{\gamma}$ the functions f_i $i = 3, \ldots, (n+3)$ have the same expectations as for γ, i.e. the first marginal of $\tilde{\gamma}$ is $\hat{P}_n$ (from condition (4.13)). Call the second marginal $\tilde{P}$. This together with the moment conditions for (4.11) yields that the Kantorovich distance between $\tilde{P}$ and $\hat{P}_n$ is at most ε. The expectation of g with respect to the second marginal $\tilde{P}$ is c as required, and the support of $\tilde{P}$ consists of at most $(n+3)$ points. □

Remark: It is obvious that the same results holds for the r-th Kantorovich metric with $r > 1$. In this case the function in (4.11) has to be replaced by

$$\int ||x-y||_r d\gamma(x,y)$$

The rest of the proof remains unchanged.

Having identified the exposed points as discrete distributions with at most $n+3$ points we use the following result (see for example [25], Section 17) to extend our result to the extreme points.

Theorem 4.4 (Straszewicz's Theorem)**.** *If X be a compact, metrizable subset of a Hausdorff LCS, then the exposed points of X are dense in ext*(X).

Corollary 4.1. *The extremal points of* $\mathscr{B}$ *are discrete measures with at most* $(n+3)$ *points.*

Proof. Since the $exp(\mathscr{B}) \subseteq \mathscr{P}_{n+3}$ and $\overline{exp(\mathscr{B})} = ext(\mathscr{B})$, we know that for every $P \in ext(\mathscr{B})$ $\exists (P_r)_{r\in\mathbb{N}} \in exp(\mathscr{B})$, such that $P_r \to P$ weakly. Suppose the measure P has a support of more than $n+3$ points. In this case there exist disjoint open balls B_i $1 \leq i \leq n+4$ with $P(\partial B_i) = 0$ and $P(B_i) > 0$. Since $P_k \to P$ weakly

$$P_r(B_i) \to P(B_i), \quad \forall 1 \leq i \leq n+4.$$

This implies that there exists a $R \in \mathbb{N}$ such that $P_r(B_i) > 0$, $\forall 1 \leq i \leq n+4$, $\forall r \geq R$, which is a contradiction to $P_r \in \mathscr{P}_{(n+3)}$. □

Having characterized the extreme points of $\mathscr{B}_\varepsilon(\hat{P})$ as discrete distribution with a fixed number of atoms, the robustified problem (4.2) can be reformulated as a semi definite problem using (4.7) and (4.8). To establish the existence of a saddle-points in (4.2) one can for example use the following classical minimax theorem (see for example [70] or [77]).

Theorem 4.5 (Sion)**.** *Let C and D be two closed convex sets in two topological vector spaces X and Y respectively. Let further* $F(x,y) : C \times D \to \mathbb{R}$ *be a function which is quasiconvex in x and quasiconcave in y. If F is upper (or lower) semi continuous in y in every line segment and lower semi continuous in x, while C is compact then the function* $F(x,y)$ *possesses a saddle-value on* $C \times D$ *and*

$$\inf_{x\in C} \sup_{y\in D} F(x,y) = \sup_{y\in D} \inf_{x\in C} F(x,y).$$

We are now in a position to state the following Theorem.

Theorem 4.6. *Let*

1. $F : \mathbb{R}^m \times \mathscr{P}(\mathbb{R}^d) \to \mathbb{R}$ *be convex in the first and linear and lower semi continuous in the second component.*
2. $G : \mathbb{R}^m \times \mathscr{P}(\mathbb{R}^d) \to \mathbb{R}$ *be convex in the first and the second component.*
3. $H : \mathbb{R}^m \to \mathbb{R}$ *be convex*

then a solution to problem (4.2) *exists and coincides with the solution of the following reduced problem*

$$\begin{array}{lll} \sup_{x\in\mathbb{R}^m} \inf_{P\in\mathscr{B}\cap\mathscr{P}_{(n+3)}} & F(x,P) & \\ s.t. & G(x,P) & \leq 0, \quad \forall P \in \mathscr{B}_\varepsilon(\hat{P}_n)\cap\mathscr{P}_{(n+3)} \\ & H(x) \leq 0 & \end{array} \tag{4.14}$$

with $\mathscr{P}_{(n+3)}$ *the discrete measures with at most* $(n+3)$ *points.*

Proof. It only remains to show that a saddle point of the problem (4.2) exists – from which the existence of a saddle point for (4.14) follows. Define $C = \mathscr{B}_\varepsilon(\hat{P}_n)$ and note that since the Kantorovich ball $\mathscr{B}_\varepsilon(\hat{P}_n)$ is a closed subset of the probability measures, it is a compact set in $(\mathscr{P}, d_K^1)$. Further define

$$D = \left(\bigcap_{P\in\mathscr{B}_\varepsilon(\hat{P}_n)} \{x\in\mathbb{R}^m : G(x,P)\leq 0\} \right) \cap \{x\in\mathbb{R}^m : H(x)\leq 0\}.$$

D is closed since $G(\cdot,P)$ and $H(\cdot)$ are convex and therefore continuous. Therefore the conditions of Theorem 4.5 are fulfilled and a saddle point for the original problem exists. The saddle value of the reduced problem coincides with that of the original problem by Corollary 4.1 and Theorem 4.1. □

Notice that the numbers of points needed does not depend on the dimension of the space $\mathbb{R}^d$ or on the structure of the support of the considered measures, but only on the number of observations the discrete measure comprises of. This is a considerable reduction of complexity, since the feasible set can now be modeled as a subset in $\mathbb{R}^{(d+1)(n+3)}$.

Possible solutions to the reduced problems are elements of a finite dimensional vector space, while the number of constraints is still infinite. Thus exploiting the special structure of Kantorovich balls, we reduced the problem from an infinite problem to a semi definite problem which is generally easier to solve.

To actually solve the optimization problem (4.2) one has to describe the set of distributions in $\mathscr{B}_\varepsilon(\hat{P}_n)\cap\mathscr{P}_{(n+3)}$. The measures $Q\in\mathscr{B}_\varepsilon(\hat{P}_n)\cap\mathscr{P}_{(n+3)}$ that assign probabilities q_j to points $y_j\in\mathbb{R}^d$ where $1\leq j\leq(n+3)$ can be described in a simple way as all the measures fulfilling the following mass transportation constraints

$$\begin{array}{rll} \sum_{j=1}^{n+3} q_j & = 1 & \\ \sum_{j=1}^{n+3} t_{i,j} & = p_i, & \forall i = 1,\ldots,n \\ \sum_{i=1}^{n} t_{i,j} & = q_j, & \forall j = 1,\ldots,(n+3) \\ \sum_{i=1}^{n}\sum_{j=1}^{n+3} ||x_i - y_j||_1 t_{i,j} & \leq \varepsilon & \\ t_{i,j}\geq 0, q_j & \geq 0 & \end{array} \tag{4.15}$$

where $(t_{i,j})_{i,j}$ models the mass transportation plan between $\hat{P}_n$ and Q, i.e. $t_{i,j}$ is the amount of probability mass that is transported from atom x_i of $\hat{P}_n$ to atom y_j of Q. The second last equation restricts the *effort* of the mass transport (given by distance times transported mass) by ε.

The connection of the above constraints to the Kantorovich distance is due to the fact that finding a minimizing distribution in (4.5) is – in the discrete setting – equivalent to solving the optimal mass transportation problem (see [62]). This fact makes it easy to handle the Kantorovich distance for discrete distributions.

However, the conditions in (4.15) if incorporated into an optimization problem will render the problem non-convex. The reason for this is the (non-convex) quadratic structure of the last constraint of (4.15). In the following numerical solution techniques for specific instances of problem (4.2) will be discussed.

4.4 A concrete problem

In this section a robust Markowitz style portfolio optimization problem with *Expected Shortfall under a Threshold a* (denoted as ES_a) as a risk functional is treated by an application of the results from the last section. Techniques to solve stochastic programs with constraint set (4.15) numerically are discussed.

The case we want to treat here is a variation of the problem described in [54]. The mean problem can be described as follows: an investor faces the problem of partitioning a budget between d investment possibilities with dependent random returns. The decision is taken by maximizing the expected return of the portfolio while controlling for the risk modeled by the expected shortfall.

ES_a – the expected shortfall under a threshold a – of random variable $X : \Omega \to \mathbb{R}$ is defined as

$$ES_a(X) = \int -max(a-x,0)dP(x)$$

where P is the distribution of X. In the case that X follows a discrete distribution ES_a can be described via finitely many linear functions and therefore efficiently incorporated into optimization problems.

The original (non-robust) problem therefore looks like

$$\begin{array}{lll} \max_{w\in\mathbb{R}^d} & \mathbb{E}(w^\top X^{\hat{P}_n}) & \\ s.t. & ES_a(w^\top X^{\hat{P}_n}) & \leq R \\ & \sum w_i & = 1 \\ & w & \geq 0 \end{array} \tag{4.16}$$

where $X^Q = (\xi_1^Q, \ldots, \xi_n^Q)$ is the random vector with joint distribution Q describing the uncertain, future returns $\xi_i : \Omega \to \mathbb{R}$ of the considered assets, $w \in \mathbb{R}^d$ is the vector of portfolio weights and $w^\top X^Q = \sum_{i=1}^d w_i \xi_i^Q$ is the random return of the portfolio. Note that the $ES_a : (w,P) \mapsto ES_a(w^\top X)$ is concave in the first and linear in the second component (see [57]).

The robust counterpart looks like

$$\begin{array}{lllll} \max_{w\in\mathbb{R}^d} \min_{Q\in\mathscr{B}_\varepsilon(\hat{P}_n)} & \mathbb{E}(w^\top X^Q) & & & \\ s.t. & ES_a(w^\top X^Q) & \leq R, & \forall Q \in \mathscr{B}_\varepsilon(\hat{P}_n) \\ & \sum w_i & = 1, & w \geq 0 \end{array}$$

If we replace the operators $\mathbb{E}$ and ES_a by their explicit versions for discrete distributions and apply Theorem 4.3 (note that the required continuity properties are fulfilled for $\mathbb{E}$ and ES_a) the above problem is equivalent to

$$\begin{array}{rlll} \max_{w\in\mathbb{R}^d}\min_{y,q,t} & \sum_{j=1}^{n+3}(w^\top y_j)q_j & & \\ s.t. & \sum_{j=1}^{n+3} q_j & = 1 & \\ & \sum_{j=1}^{n+3} t_{i,j} & = p_i, & \forall i=1,\dots,n \\ & \sum_{i=1}^{n} t_{i,j} & = q_j, & \forall j=1,\dots,(n+3) \\ & \sum_{i=1}^{n}\sum_{j=1}^{n+3} ||x_i-y_j||_1 t_{i,j} & \leq \varepsilon & \\ & \sum_{j=1}^{n+3} max(a-w^\top y_j,0)q_j & \leq R, & \end{array} \tag{4.17}$$

where the points $y_j \in \mathbb{R}^d$, $1 \leq j \leq n+3$ are the atoms of the measures Q with respective probabilities q_j. The probabilities p_i, the points x_i, $\varepsilon > 0$ and the risk parameter R are the data of the problem.

The above optimization problem exhibits non-convexities. These occur in the definition of the Kantorovich distance as well as in the expression for the expected shortfall and the expectation. All these non-convexities are bi-linear in the decision variables. We thus arrived at a non-convex semi-definite problem.

We solve (4.17) by the following iterative algorithm originally presented in [54].

1. Set $i=0$ and $\mathscr{Q}_0 = \{\hat{P}_n\}$ with $\hat{P}_n \in \mathscr{P}$.

2. Solve the outer problem

$$\begin{array}{rlll} \max_{(w,t)} & t & & \\ s.t. & \mathbb{E}(w^\top X^Q) & \leq t, & \forall Q \in \mathscr{Q}_i \\ & ES_a(w^\top X^Q) & \leq R, & \forall Q \in \mathscr{Q}_i \\ & \sum_{i=1}^{d} w_i & = 1, \quad w \geq 0 & \end{array} \tag{4.18}$$

   and call the solution (w_i, t_i).

3. Solve the problem

$$\min_{Q\in\mathscr{B}_\varepsilon(\hat{P}_n)\cap\mathscr{P}_{n+3}} \mathbb{E}(w^\top X^Q) \tag{4.19}$$

   and call the solution $Q_i^{(1)}$.

4. Solve the problem

$$\max_{Q\in\mathscr{B}_\varepsilon(\hat{P}_n)\cap\mathscr{P}_{n+3}} ES_a(w^\top X^Q) \tag{4.20}$$

   and call the solution $Q_i^{(2)}$.

5. $\mathscr{Q}_{i+1} \leftarrow \mathscr{Q}_i \cup \{Q_i^{(1)}\} \cup \{Q_i^{(2)}\}$.

6. If

(a) $\mathscr{Q}_{i+1} = \mathscr{Q}_i$ or

(b) the optimal value of (4.19) equals t_n and the solution of (4.20) is equal to $\min_{P \in \mathscr{Q}_i} ES_a(w^\top X^P)$

then a saddle point is found and the algorithm stops. Otherwise $i \leftarrow i+1$ and goto 2.

The idea of the algorithm is to achieve robustness gradually by including finitely many measures in the problem (4.18). The measures Q_n are chosen to be the worst case with respect to the current portfolio decision w_n and therefore represent the extremal elements of $\mathscr{B}_\varepsilon(\hat{P}_n)$ which are required to approximate the relevant parts of $\mathscr{B}_\varepsilon(\hat{P}_n)$. This approximation gets better with every iteration and – as numerical experiments show – the respective portfolios w_i and objective values t_i stabilize. Under certain continuity conditions fulfilled for the problem at hand the algorithm either finds a saddle point (w^*, t^*) in finitely many iterations or the intermediary solutions (w_i, t_i) converge to a saddle point (see Proposition 1 in [54]).

The non-convexity of problem (4.17) is reflected in the non-convexity of problems (4.19) and (4.20). In the remainder of this section we will discuss solution techniques for these problems.

4.4.1 Minimizing Expectation

Problem (4.19) can be solved directly in an iterative manner by gradually altering the empirical distribution $\hat{P}_n$ to lower the expectation. The changes we make in the process should be optimal in the sense that the cost measured in Kantorovich distance is small compared to the impact in expectation.

To make this precise, suppose that asset i^* when altered leads to the maximum change in expectation, that is $i^* = \arg\max_i w_i$ and therefore

$$\frac{\partial \mathbb{E}(X_w)}{\partial w_{i^*}} = \max_{1 \le j \le d} \frac{\partial \mathbb{E}(X_w)}{\partial w_j}. \tag{4.21}$$

The aim is – starting from the empirical distribution – to obtain a distribution which minimizes the expectation over the set of discrete distributions $\mathscr{B}_\varepsilon(\hat{P}_n) \cap \mathscr{P}_{n+3}$. Clearly, from (4.21) the optimal way to achieve this goal is to start altering scenarios in the i^*-th component. Therefore the following algorithm to find the optimum in (4.19) is proposed. A constant c is used as a lower bound on the return of asset i^*. If no specific assumption of this type seems acceptable (i.e. also very small asset returns are plausible to the modeler) c can be set to zero.

While $\{(x_i, p_i) : 1 \le i \le n\}$ denotes the atom probability pairs of the empirical measure $\hat{P}_n$, the $\left\{(y_i^{(j)}, q_i^{(j)}) : 1 \le i \le n+1\right\}$ denotes those of the altered measures $P^{(j)}$ after the $j-th$ iteration.

1. Set $P^{(0)} = \hat{P}_n$.
2. Find $i^* = \arg\max_{1 \le j \le d} w_j$.
3. Set $j = 1$.

4. Pick any atom $y_i^{(j-1)}$ with $1 \leq i \leq n$ and change it's i^*-th component to c and call the resulting measure Q.

5. Solve the linear programming problem

$$\begin{array}{ll} \max & \lambda \\ s.t. & \lambda Q + (1-\lambda) P_m \in \mathscr{B}_\varepsilon(\hat{P}_n) \cap \mathscr{P}_{n+3} \end{array} \tag{4.22}$$

and set $P^{(j)} = \lambda^* Q + (1-\lambda^*) P^{(j-1)}$, where λ^* is the optimal value of the above problem. If $\lambda^* = 1$ set $j \leftarrow j+1$ and goto step 4 otherwise set $P' \leftarrow P^{(j)}$ and stop the algorithm.

Although the above algorithm has the flavor of a local search, the procedure leads to a global optimum (see Theorem 4.7). It is obvious that there are many different distributions that achieve the optimum. Which of these possible solutions is eventually chosen depends on the choices for k made in step 4 of the algorithm. Notice that the optimal distribution P' does not have the full $(n+3)$ points but generally only $(n+1)$ points.

Theorem 4.7. *The algorithm presented above leads to an optimal solution of problem (4.19).*

Proof. Let P be in $\mathscr{B}_\varepsilon(\hat{P}_n) \cap \mathscr{P}_{n+3}$ and let $(t_{i,j})$ be the transportation plan that transports mass from $\hat{P}_n$ to P (i.e. $t_{i,j}$ fulfill the set of constraints (4.15)). The difference of the expectation under $\hat{P}_n$ and P can be bounded as follows

$$\begin{aligned} |\mathbb{E}(w^\top X^{\hat{P}_n}) - \mathbb{E}(w^\top X^P)| &\leq \sum_{i,j} t_{i,j} \sum_{k=1}^{d} |x_{i,k} - y_{j,k}| w_k \\ &\leq \sum_{i,j} t_{i,j} ||x_i - y_j||_1 w_{i^*} = w_{i^*} \varepsilon \end{aligned}$$

The difference for the measure P', that is chosen by the above algorithm is given by

$$\mathbb{E}(w^\top X^{\hat{P}_n}) - \mathbb{E}(w^\top X^{P'}) = \sum_{i,j} t_{i,j} (x_{i,i^*} - y_{j,i^*}) w_{i^*} = w_{i^*} \varepsilon.$$

Therefore the measure P' is optimal. □

Remark: If the algorithm does not terminate, i.e. if $\varepsilon > 0$ is too big and the returns of asset i^* can be shifted to c for all the scenarios, then the algorithm can be extended in the following way: find $i_2^* = \arg\max_{j \neq i^*} w_j$ and continue with the algorithm using i_2^* instead of i^*. The optimality of the procedure still holds by essentially the same proof.

4.4.2 Maximizing Expected Shortfall

Define the index sets $I = \{1,\dots,n\}$, $J = \{1,\dots,n+3\}$ and $M = \{1,\dots,d\}$. The inner problem (4.20) can be written as

$$\begin{array}{lll}
\max_{y,q,d,t} & \sum_{j\in J} max(a - w^\top y_j, 0) q_j & \\
s.t. & \sum_{i\in I} t_{i,j} & = q_j, \quad \forall j \in J \\
& \sum_{j\in J} t_{i,j} & = p_i, \quad \forall i \in I \\
& x_{i,k} - y_{j,k} & \le d^k_{i,j}, \quad \forall (i,j,k) \in I \times J \times M \\
& y_{j,k} - x_{i,k} & \le d^k_{i,j}, \quad \forall (i,j) \in I \times J \\
& \sum_{k\in M} d^k_{i,j} & = d_{i,j}, \quad \forall (i,j) \in I \times J \\
& \sum_{(i,j)\in I\times J} d_{i,j} t_{i,j} & \le \varepsilon \\
& t_{i,j}, d^k_{i,j}, q_j & \ge 0.
\end{array} \tag{4.23}$$

with data w, p, a and ε. The above problem is non-convex due to the bilinear terms defining the Kantorovich distance and the Expected Shortfall.

Since the above problem deals with the maximization of a convex function to solve (4.23), we rewrite it as a D.C. program (see Appendix A or for example [43] or [44]). Towards this we model the objective in (4.23) as follows: introduce the splitting variables

$$a - w^\top y_j = z_j^+ + z_j^-, \quad z_j^+ z_j^- \ge 0, \quad z_j^+ \ge 0. \tag{4.24}$$

When modeled like this $z_j^+ = max(a - w^\top y_j, 0)$ and $z_j^- = min(a - w^\top y_j, 0)$ in optimal points. To simplify notation and to tackle the bilinear terms in (4.23) in (4.24) we introduce the variables

$$\begin{array}{llll}
\alpha_j = \frac{1}{2}(z_j^+ + q_j), & \beta_j = \frac{1}{2}(z_j^+ - q_j), & \forall j \in J \\
\gamma_{i,j} = \frac{1}{2}(d_{i,j} + t_{i,j}), & \delta_{i,j} = \frac{1}{2}(d_{i,j} - t_{i,j}), & \forall (i,j) \in I \times J \\
\phi_j = \frac{1}{2}(z_j^+ + z_j^-), & \chi_j = \frac{1}{2}(z_j^+ - z_j^-), & \forall j \in J.
\end{array}$$

Note that these variables help with the reformulation of the problem but need not be used in the numerical implementation.

We are now in the position to reformulate problem (4.23) as the following non-convex quadratic problem.

$$\begin{array}{lll}
\min_{y,q,d,t,z^+,z^-} & t & \\
s.t. & \sum_{j\in J} (\beta_j^2 - \alpha_j^2) & \le t \\
& \sum_{i\in I} t_{i,j} & = q_j, \quad \forall j \in J \\
& \sum_{j\in J} t_{i,j} & = p_i, \quad \forall i \in I \\
& x_{i,k} - y_j^k & \le d^k_{i,j}, \quad \forall (i,j,k) \in I \times J \times M \\
& y_j^k - x_{i,k} & \le d^k_{i,j}, \quad \forall (i,j) \in I \times J \\
& \sum_{k=1}^m d^k_{i,j} & = d_{i,j}, \quad \forall (i,j) \in I \times J \\
& \sum_{(i,j)\in I\times J} (\gamma_{i,j}^2 - \delta_{i,j}^2) & \le \varepsilon \\
& \sum_{j\in J} (\phi_j^2 - \chi_j^2) & \ge 0 \\
& a - w^\top y_j & = z_j^+ + z_j^- \\
& t_{i,j}, z_j^+, d^k_{i,j}, q_j & \ge 0.
\end{array} \tag{4.25}$$

Notice that all the non-convex constraints are defined by a difference of convex (D.C.) functions, hence the program is a D.C. program. In the remainder of this section we will use this structural property to obtain solutions of (4.25).

To simplify notation define the vector valued affine function $l(\bar{x})$ in such a way that the linear constraints in the above problem are fulfilled iff $l(\bar{x}) \leq 0$ where $\bar{x}$ is a vector containing all decision variables, i.e.

$$\bar{x} = (t, (q_j)_{j\in J}, (y_j)_{j\in J}, (z_j^+)_{j\in J}, (z_j^-)_{j\in J}, (t_{i,j})_{(i,j)\in I\times J}, (d_{i,j})_{(i,j)\in I\times J}, (d_{i,j}^k)_{(i,j,k)\in I\times J\times M})$$

Problem (4.25) then becomes

$$\begin{array}{lll} \min_{y,q,d,t,z^+,z^-} & t & \\ s.t. & \sum_{j\in J}(\beta_j^2 - \alpha_j^2) & \leq t \\ & \sum_{(i,j)\in I\times J}(\gamma_{i,j}^2 - \delta_{i,j}^2) & \leq \varepsilon \\ & \sum_j(\phi_j^2 - \chi_j^2) & \geq 0, \\ & l(\bar{x}) & \leq 0. \end{array} \tag{4.26}$$

The above problem is a D.C. problem with

$$N = 1 + 3(n+3) + (n+3)d + 2(n+3)n + (n+3)nd$$

variables. For realistic values $n \approx 100$ and $d \approx 10$ this amounts to more than 120.000 variables. These are far too many variable for the existing methods for finding global solutions of general D.C. problems. We therefore employ the so called DCA (difference of convex algorithm) which is an approximate solution technique for D.C. programs. The DCA has proven to be very effective in solving various kinds of D.C. programs (see for example [5, 8, 26, 41, 71]) and its low computational complexity makes it possible to work with large scale D.C. programs.

To apply the DCA we need to write (4.26) as an unconstrained D.C. problem of the form

$$\inf\{f(\bar{x}) = g(\bar{x}) - h(\bar{x}) : \bar{x} \in \mathbb{R}^N\} \tag{4.27}$$

where N is the dimension of $\bar{x}$ as defined above.

The DCA algorithm works for problems of the form (4.27) by repeatedly solving the two convex optimization problems

$$\inf_{\bar{x}\in\mathbb{R}^N}\{g(\bar{x}) - (h(\bar{x}^k) + \langle \bar{x} - \bar{x}^k, \bar{y}^k\rangle)\} \tag{4.28}$$

and

$$\inf_{\bar{y}\in\mathbb{R}^N}\{h^*(\bar{y}) - (g^*(\bar{y}^{k-1}) + \langle \bar{x}^k, \bar{y} - \bar{y}^{k-1}\rangle)\} \tag{4.29}$$

and iteratively produces candidates x^{k+1}, y^k for solutions of increased quality of the primal and dual D.C. problems respectively. The first problem can be viewed as a convex approximation of the original problem, while the second problem can be thought of as a convex approximation of the dual D.C. program $\inf_{y\in\mathbb{R}^N} h^*(y) - g^*(y)$.

The DCA guarantees that $(g(\bar{x}^k) - h(\bar{x}^k))_{k\in\mathbb{N}}$ and $(h^*(\bar{y}^k) - g^*(\bar{x}^k))_{k\in\mathbb{N}}$ are decreasing sequences and the limit points of $(\bar{x}^k)_{k\in\mathbb{N}}$ and $(\bar{y}^k)_{k\in\mathbb{N}}$ are critical points of the primal problem (4.28) and the dual problem (4.29) respectively. For a more detailed discussion of DCA algorithm and its properties see [7].

We use Corollary 2.4.1. from [18] to reformulate (4.26) by an exact penalty.

Theorem 4.8. *Suppose that X and Y are finite dimensional Banach spaces, the sets $S \subseteq X$ and $C \subseteq Y$ are nonempty, closed and convex and the point x^* solves the following optimization problem*

$$\begin{array}{lll} \inf & f_0(x) & \\ s.t. & G(x) & \in C \\ & x & \in S, \end{array} \tag{4.30}$$

where the function $G : X \to Y$ is strictly differentiable at $x^* \in S$, $f_0 : X \to \mathbb{R}$ is Lipschitz near x^* and

$$0 \in int\left\{G(x^*) + G'(x^*)(S - x^*) - C\right\}.$$

Then there exists an $\bar{\tau} > 0$, such that the optima of (4.30) *and*

$$\inf_{x \in X} f_0(x) + \psi(x|S) + \tau dist(G(x)|C) \tag{4.31}$$

coincide for all $\tau > \bar{\tau}$, where $\psi(\cdot|S)$ is the convex indicator function of the set S and

$$dist(G(x)|C) = \inf\{||G(x) - y|| : y \in C\}.$$

Lemma 4.1. *Set $X = \mathbb{R}^N$, $Y = (\mathbb{R}^3, ||\cdot||_\infty)$, $S = \{\bar{x} \in \mathbb{R}^N : l(\bar{x}) \leq 0\}$, $G = (G_1, G_2, G_3)$, where the $G_i : \mathbb{R}^N \to \mathbb{R}$ are functions described by the left hand sides of the first three constraints of* (4.26)*, i.e.*

$$G(\bar{x}^*) = \begin{pmatrix} \sum \left(\left(\frac{z_j^+ - q_j}{2} \right)^2 - \left(\frac{z_j^+ + q_j}{2} \right)^2 \right) - t \\ \sum \left(\frac{d_{i,j} + t_{i,j}}{2} \right)^2 - \left(\frac{d_{i,j} - t_{i,j}}{2} \right)^2 \\ \sum \left(\frac{z_j^+ + z_j^-}{2} \right)^2 - \left(\frac{z_j^+ - z_j^-}{2} \right)^2 \end{pmatrix}$$

and $C = (-\infty, 0] \times (-\infty, \varepsilon] \times [0, \infty)$. Then the constraint qualification

$$0 \in int\left\{G(\bar{x}^*) + G'(\bar{x}^*)(S - \bar{x}^*) - C\right\} \tag{4.32}$$

is fulfilled at optimal points $\bar{x}^*$ of the system (4.26)*.*

Proof. Let $\bar{x}^* = (\bar{t}^*, (\bar{q}_j^*)_j, (\bar{y}_j^*)_j, (\bar{z}_j^{+*})_j, (\bar{z}_j^{-*})_j, (\bar{t}_{i,j}^*)_{i,j}, (\bar{d}_{i,j}^*)_{i,j}, (\bar{d}_{i,j}^{k*})_{i,j,k})$. Since all the constraints G_i have to be binding at optimal points we have $G(\bar{x}) = (0, \varepsilon, 0)$. The three constraints are analyzed separately to find a common point $s \in S$ which fulfils (4.32). First note that

$$\frac{\partial G_2(x)}{\partial d_{i,j}} = t_{i,j}, \quad \frac{\partial G_2(x)}{\partial t_{i,j}} = d_{i,j}$$

and therefore

$$\nabla G_2(\bar{x}^*)(s - \bar{x}) = \sum d_{i,j} \bar{t}_{i,j}^* + \sum t_{i,j} \bar{d}_{i,j}^* - 2 \underbrace{\sum \bar{d}_{i,j}^* \bar{t}_{i,j}^*}_{\varepsilon}$$

where $\bar{d}^*_{i,j}$ and $\bar{t}^*_{i,j}$ are the corresponding components of $\bar{x}^*$ and $t_{i,j}$ and $d_{i,j}$ the corresponding components of s. We therefore need an s which fulfils

$$\sum d_{i,j}\bar{t}^*_{i,j} + \sum \bar{d}^*_{i,j} t_{i,j} < 2\varepsilon.$$

To achieve this fix $t_{i,j} = \bar{t}^*_{i,j}$ for all $(i,j) \in I \times J$, which results in $\sum \bar{d}^*_{i,j} t_{i,j} = \varepsilon$ and therefore in the modified condition

$$\sum d_{i,j}\bar{t}^*_{i,j} < \varepsilon. \tag{4.33}$$

Next define $N(j) = \left\{ i \in I : \bar{t}^*_{i,j} > 0 \right\}$. Now if there is a $j \in J$ with $\{i\} = N(j)$ and $x_i \neq \bar{y}^*_j$ then replace the point $\bar{y}^*_j$ by x_i, therefore reducing $d_{i,j}$ to zero. This yields (4.33). If there is no such point, then choose a $j' \in J$ with $|N(j')| > 1$ and define the point $y_{j'}$ as the solution of following optimization problem

$$\min_{y \in \mathbb{R}^d} \sum_{i \in N(j)} ||x_i - y||_1 \bar{t}^*_{i,j}. \tag{4.34}$$

Note that $\bar{y}^*_j$ can not be a solution to (4.34), since by optimality $\bar{y}^{k*}_j \leq \min_{i \in N(j)} x^k_i$ for all $1 \leq k \leq d$. This is the case, since if $\bar{y}^{k*}_j > x^k_i$ for some $i \in N(j)$ and some $1 \leq k \leq d$ then we could move $\bar{t}^*_{i,j}$ from x_i to a new point $\hat{y}$ with

$$\hat{y}^l = \begin{cases} \bar{y}^{k*}_j, & l \neq k \\ x^k_i, & l = k \end{cases}$$

instead of $\bar{y}^{k*}_j$. This would yield a lower expected shortfall and a lower transportation cost at the same time, contradicting the optimality of the transportation plan.

Now if we choose all other points y_j for $j \in J \setminus \{j'\}$ to be equal to $\bar{y}^*_j$, then we have

$$\varepsilon = \sum \bar{d}^*_{i,j}\bar{t}^*_{i,j} > \sum_{j \neq j'} \sum_{i \in I} \bar{d}^*_{i,j}\bar{t}^*_{i,j} + \sum_{i \in N(j')} d_{i,j}\bar{t}^*_{i,j} = \sum d_{i,j}\bar{t}^*_{i,j}$$

as required.

Next G_3 is analyzed. Arguing similarly as above it holds that

$$\nabla G_3(\bar{x}^*)(s - \bar{x}^*) = \sum \bar{z}^{-*}_j z^+_j + \sum \bar{z}^{+*}_j z^-_j - 2\underbrace{\sum \bar{z}^{+*}_j \bar{z}^{-*}_j}_{0}.$$

We therefore need z^+_j and z^-_j such that

$$\sum \bar{z}^{-*}_j z^+_j + \sum \bar{z}^{+*}_j z^-_j < 0. \tag{4.35}$$

To achieve this we choose $z^+_j = \max(a + w^\top y_j, 0) + \nu$ and $z^-_j = \min(a + w^\top y_j, 0) - \nu$ with $\nu > 0$. Since $\bar{x}^*$ is optimal either $\bar{z}^{+*}_j = 0$ or $\bar{z}^{-*}_j = 0$ holds. In the former case we get

$$\bar{z}^{+*}_j z^-_j + \bar{z}^{+*}_j z^-_j = \bar{z}^{+*}_j z^-_j \leq 0$$

and in the latter

$$\bar{z}^{+*}_j z^-_j + \bar{z}^{+*}_j z^-_j = \bar{z}^{+*}_j z^-_j \leq 0$$

and therefore (4.35) is fulfilled unless for all the $\bar{y}_j^*$ it holds that $a + w^\top \bar{y}_j^* = 0$ in which case the last two inequalities would hold with equality since then $\bar{z}_j^{+*} = \bar{z}_j^{-*} = 0$. This situation can only occur if all the points $\bar{y}_j^*$ yield the (same) return a, which again can be easily seen to be impossible for an optimal point $\bar{x}^*$.

The last remaining component of G is G_1 which is unproblematic since

$$\nabla G_1(\bar{x}^*)(s - x_0) = -\sum \bar{q}_j^* z_j^+ - \sum \bar{z}_j^{+*} q_j + t$$

and t can be chosen such that the condition

$$\sum \bar{q}_j^* z_j^+ + \sum \bar{z}_j^{+*} q_j > t$$

is fulfilled. □

Theorem 4.9. *For some $\tau > 0$ the problem* (4.26) *can be equivalently reformulated to the following problem*

$$\begin{aligned} min_{\bar{x}\in\mathbb{R}^N} \quad & t + \tau \max\left\{ \sum_{j\in J} \beta_j^2 + \sum_{(i,j)\in I\times J} \delta_{i,j}^2 + \sum_{j\in J} \phi_j^2 - t,\ \sum_{(i,j)\in I\times J} \gamma_{i,j}^2 - \varepsilon + \sum_{j\in J} \alpha_j^2 + \sum_{j\in J} \phi_j^2, \right. \\ & \left. \sum_{j\in J} \chi_j^2 + \sum_{(i,j)\in I\times J} \delta_{i,j}^2 + \sum_{j\in J} \alpha_j^2,\ \sum_{(i,j)\in I\times J} \delta_{i,j}^2 + \sum_{j\in J} \alpha_j^2 + \sum_{j\in J} \phi_j^2 \right\} \qquad (4.36) \\ - \quad & \tau \left(\sum_{i,j} \delta_{i,j}^2 + \sum_j \alpha_j^2 + \sum_j \phi_j^2 \right) + \psi(x|S). \end{aligned}$$

where $\psi(\cdot|S)$ is the convex indicator function of the set S. In other words (4.26) *can be reformulated to a D.C. problem of the form* (4.27) *with*

$$\begin{aligned} g(\bar{x}) &= \tau \max\left\{ \sum_{j\in J} \beta_j^2 + \sum_{(i,j)\in I\times J} \delta_{i,j}^2 + \sum_{j\in J} \phi_j^2 - t,\ \sum_{(i,j)\in I\times J} \gamma_{i,j}^2 - \varepsilon + \sum_{j\in J} \alpha_j^2 + \sum_{j\in J} \phi_j^2, \right. \\ &\qquad \left. \sum_{j\in J} \chi_j^2 + \sum_{(i,j)\in I\times J} \delta_{i,j}^2 + \sum_{j\in J} \alpha_j^2,\ \sum_{(i,j)\in I\times J} \delta_{i,j}^2 + \sum_{j\in J} \alpha_j^2 + \sum_{j\in J} \phi_j^2 \right\} + \psi(x|S) \\ h(\bar{x}) &= -t + \tau \left(\sum_{(i,j)\in I\times J} \delta_{i,j}^2 + \sum_{j\in J} \alpha_j^2 + \sum_{j\in J} \phi_j^2 \right). \end{aligned}$$

Proof. From Lemma 4.1 we get that Theorem 4.8 is applicable and therefore problem (4.26) can be reformulated to a unconstrained problem of the form (4.31) with S, C and G as defined in Lemma 4.1. To reduce the problem to an unconstrained D.C. problem as in (4.27), the functions g and h have to be identified. To this end we first calculate the maximum of the G_i, $1 \le i \le 3$ and 0. Note that the maximum of finitely many D.C. functions $f_i = g_i - h_i$ is again a D.C. function with the following decomposition

$$\max f_i = \max_i \{g_i + \sum_{j\neq i} h_j\} - \sum_i h_i.$$

Therefore the above maximum is of the form

$$\max \quad \left\{ \sum_j \beta_j^2 + \sum_{i,j} \delta_{i,j}^2 + \sum_j \phi_j^2 - t, \sum_{i,j} \gamma_{i,j}^2 + \sum_j \alpha_j^2 + \sum_j \phi_j^2 - \varepsilon, \sum_j \chi_j^2 + \sum_{i,j} \delta_{i,j}^2 \right.$$
$$\left. + \sum_j \alpha_j^2, \sum_{i,j} \delta_{i,j}^2 + \sum_j \alpha_j^2 + \sum_j \phi_j^2 \right\} - \left(\sum_{i,j} \delta_{i,j}^2 + \sum_j \alpha_j^2 + \sum_j \phi_j^2 \right),$$

which finally yields (4.36). □

Solving problem (4.28) From Theorem 4.9 it follows that for (4.26) the corresponding DCA sub-problem (4.28) can be written as the quadratically constrained problem

$$\begin{array}{lll} \min & \tau M - \langle \bar{x}, \bar{y}^k \rangle & \\ s.t. & \sum_j \beta_j^2 - t + \sum_{i,j} \delta_{i,j}^2 + \sum_j \phi_j^2 & \leq M \\ & \sum_{i,j} \gamma_{i,j}^2 - \varepsilon + \sum_j \alpha_j^2 + \sum_j \phi_j^2 & \leq M \\ & \sum_j \chi_j^2 + \sum_j \alpha_j^2 + \sum_{i,j} \delta_{i,j}^2 & \leq M \\ & \sum_{i,j} \delta_{i,j}^2 + \sum_j \alpha_j^2 + \sum_j \phi_j^2 & \leq M \\ & l(\bar{x}) & \leq 0. \end{array} \tag{4.37}$$

Note that $h(\bar{x}^k)$ and $-\langle \bar{x}^k, \bar{y}^k \rangle$ are omitted from the objective, since they are constant with respect to $\bar{x}$.

Solving problem (4.29) To solve the DCA sub-problem (4.29), we first have to compute the conjugate h^* for

$$h(\bar{x}) = -t + \tau \sum_j \beta_j^2 + \tau \sum_{i,j} \delta_{i,j}^2 + \tau \sum_j \phi_j^2.$$

It is easy to see that

$$h^*(\bar{y}) = \begin{cases} \sum_{i,j} \frac{q_{i,j}^2}{\tau} + \sum_{i,j} \frac{d_{i,j}^2}{\tau} + \sum_j \frac{(z_j^-)^2}{\tau}, & t = -1.\ d_{i,j} = -t_{i,j},\ z_j^+ = z_j^- + q_j,\ y_j = 0,\ d_{i,j}^k = 0 \\ \infty, & \text{otherwise,} \end{cases}$$

where $\bar{y} = (t, (q_j), (y_j), (z_j^+), (z_j^-), (t_{i,j}), (d_{i,j}), (d_{i,j}^k))$. (4.29) therefore becomes

$$\begin{array}{lll} \min & \sum_{i,j} \frac{q_{i,j}^2}{\tau} + \sum_{i,j} \frac{d_{i,j}^2}{\tau} + \sum_j \frac{(z_j^-)^2}{\tau} - \langle \bar{x}^k, \bar{y} \rangle & \\ s.t. & t = -1 & \\ & d_{i,j} = -t_{i,j}, & \forall (i,j) \in I \times J \\ & z_j^+ = z_j^- + q_j, & \forall j \in J \\ & y_j = 0, & \forall j \in J \\ & d_{i,j}^k = 0, & \forall (i,j,k) \in I \times J \times M. \end{array} \tag{4.38}$$

The solution to problem (4.38) can be found analytically and is given by $t = -1$,

$$\begin{array}{llll} z_j^- = \tau \phi_j, & q_j = \tau \alpha_j, & z_j^+ = z_j^- + q_j, & \forall j \in J \\ d_{i,j} = \tau \delta_{i,j}, & t_{i,j} = -d_{i,j}, & & \forall (i,j) \in I \times J \end{array}$$

and all the other variables equal to zero.

Table 4.1: Expectation, Expected Shortfall below 1 and standard deviation s of the weekly returns of the 14 Euro STOXX Supersector indices used in the numerical examples, time-frame: 2006

Index	$\mathbb{E}$	$\mathbf{ES_1}$	**s**
Automobiles & Parts (SXAP)	1.0063	0.0072	0.0275
Banks (SX7P)	1.0058	0.0064	0.0246
Basic Resources (SXPP)	1.0076	0.0050	0.0236
Chemicals (SX4P)	1.0052	0.0054	0.0218
Construction & Materials (SXOP)	1.0081	0.0068	0.0289
Financial Services (SXFP)	1.0096	0.0059	0.0260
Food & Beverage (SX3P)	1.0058	0.0037	0.0164
Health Care (SXDP)	1.0030	0.0069	0.0221
Industrial Goods & Services (SXNP)	1.0054	0.0069	0.0263
Insurance (SXIP)	1.0048	0.0071	0.0252
Oil & Gas (SXEP)	1.0032	0.0080	0.0263
Technology (SX8P)	1.0024	0.0092	0.0298
Telecommunications (SXKP)	1.0041	0.0063	0.0214
Utilities (SX6P)	1.0075	0.0050	0.0218

4.4.3 Other measures of risk

The methods outlined here are not specific to the expected shortfall but can be applied to any risk measure that fulfils the requirements of Theorem 4.6. Such risk measures have to be convex in the portfolio weights w as well as in the probability measure P. One can consider risk for example risk functionals corresponding to some expected utility functionals (see [57]).

4.5 Numerical Results

In this section numerical results for problem (4.17) obtained by the methods discussed in the previous sections are presented. As mentioned earlier we use the empirical measure $\hat{P}_n$ as our reference measure and the Kantorovich ball with a certain radius as our ambiguity set. The data which constitutes the empirical measure are the historical weekly returns for 14 selected Euro STOXX Supersector indices (see Table 4.1).

All numerical experiments are carried out for the robustifications described in this paper as well as for the techniques in [54], where only the probability mass is shifted between the points of the empirical distribution. We henceforth call the problem described in Section 4.4 the fully robustified problem, while we refer to the problem where only the atoms of the empirical measure are considered the partially robustified problem.

For the computations in this paper we chose $a = 1$ – we therefore control the returns smaller than 1, i.e. the losses of the portfolio. Since we are interested in the impact of robustification, we vary the radius of the Kantorovich neighborhood around the empirical measure from $\varepsilon = 0.005$ to $\varepsilon = 0.03$ in steps of 0.005. Furthermore we choose the risk parameter $q = 0.007$.

The DCA was implemented in MATLAB R2007a and the optimization problems where solved with MOSEK Version 5.0.0.87. All of the following computations can be performed on a standard PC.

Figure 4.1 shows a first comparison between the partial and the full robustification. Part (a) of the plot shows that the full robustification gives significantly lower robustified expectations than the partial robustification. As expected, mean return decreases with increasing robustness parameter for both methods.

Next we investigate the decrease in expected return under the nominal measure $\hat{P}$, when solving the fully and the partially robust version of the problem. Part (b) of the Figure shows the optimal returns of the nominal instance, the returns for the partial robustification as well as the returns of the robust portfolios under the empirical measure $\hat{P}_n$. Part (c) of the plot is an analogue of (b) with full robustification.

Obviously the drop in expected returns from the optimal return of the nominal instance is less for the empirical measure than for the respective worst case measures. It can be observed that the portfolios found by full robustification consistently yield lower returns under the empirical measure than the portfolios found by partial robustification. However, the differences between the expectations under the worst case and empirical measure seem to be comparable (except for lower values of the robustness parameters where the differences are smaller for the partial robustification).

Next it is investigated how the partially robustified portfolios perform under the real worst measures in $\mathscr{B}_\varepsilon(\hat{P}_n)$, i.e. the measures obtained from the inner problems of the full robustification routine (as described in Section 4.1 and 4.2). More specifically we solve the problems in Section 4.1 and 4.2 for the partially robustified portfolios w_ε with robustness parameter ε varying in 6 equally big steps from 0 to 0.03. We therefore obtain measures $P_\varepsilon^{\mathbb{E}}$ and $P_\varepsilon^{ES_a}$, which give the smallest expected return and the highest Expected Shortfall for the portfolios w_ε. In Figure 4.2 (a) the expected returns of $w_\varepsilon^\top \xi$ under $P_\varepsilon^{\mathbb{E}}$ are compared with the respective solutions of the nominal instance given in (4.2). It is evident that the fully robustified portfolios perform much better, especially for higher values of the robustness parameter where the worst case return of the partially robustified portfolio drops significantly.

Figure 4.2 (b) depicts a similar analysis carried out for the Expected Shortfalls. It can be observed that the worst case Expected Shortfall of the partially robust portfolio rises above the allowed 0.007 already at a robustness level of 0.01 and reaches up to 0.0095 for higher robustness levels.

Summarizing, the Figures 4.1 and 4.2 show that the freedom to deviate from the atoms of the empirical measure actually results in portfolios, which yield lower returns in the worst case as well as under the empirical measure. However, these portfolios are clearly more robust than the measures obtained by shifting only the probabilities as Figure 4.2 demonstrates.

In Figure 4.3 the differences in portfolios are shown. For both methods the number of assets increase as the robustness parameter ε grows. The portfolios are more diversified and therefore more stable. The effect of diversification however is much more pronounced with the full robustification than with the partial robustification: the partially robust portfolios encompass a maximum of seven assets, while the fully robust portfolios have up to 12 assets for $\varepsilon = 0.03$. This is another indicator for the fact that full robustification yields more stable portfolios. Another interesting observation is that for the full robustification the portfolio weights of the assets

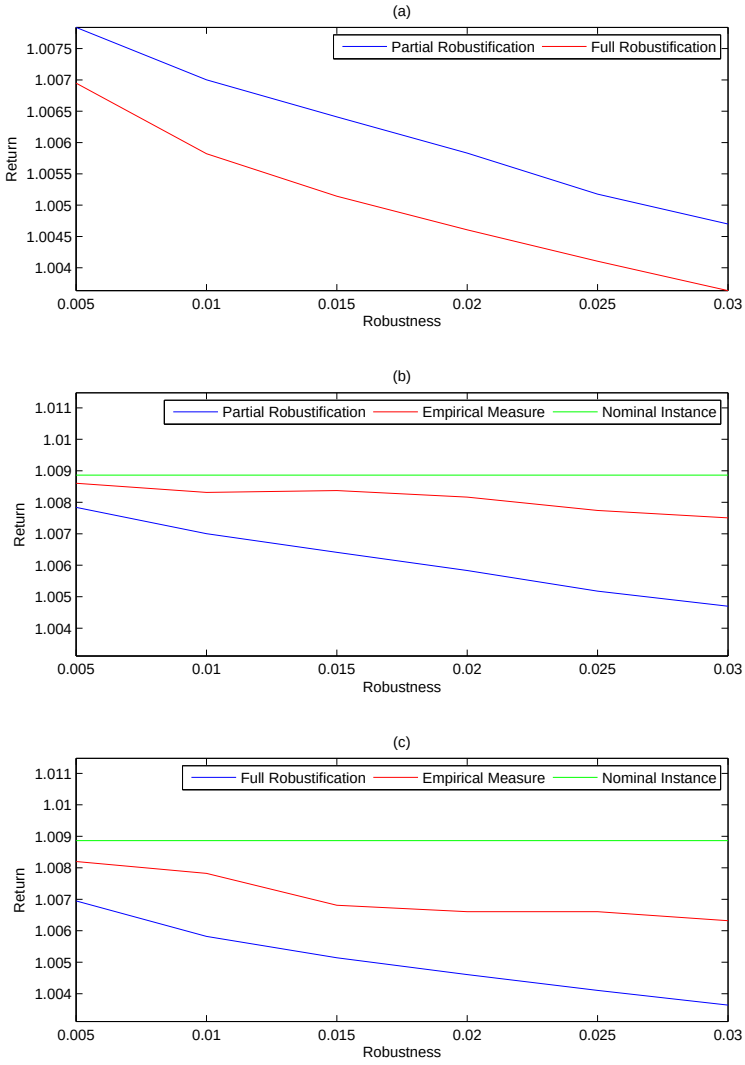
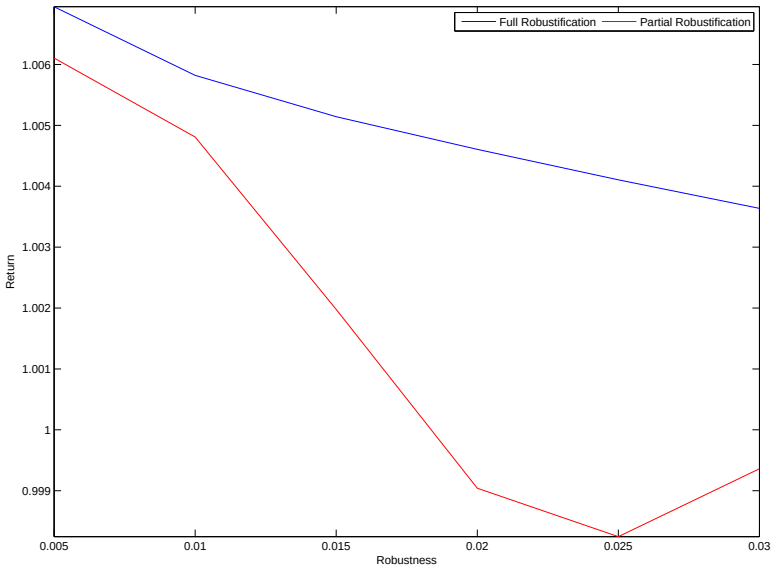
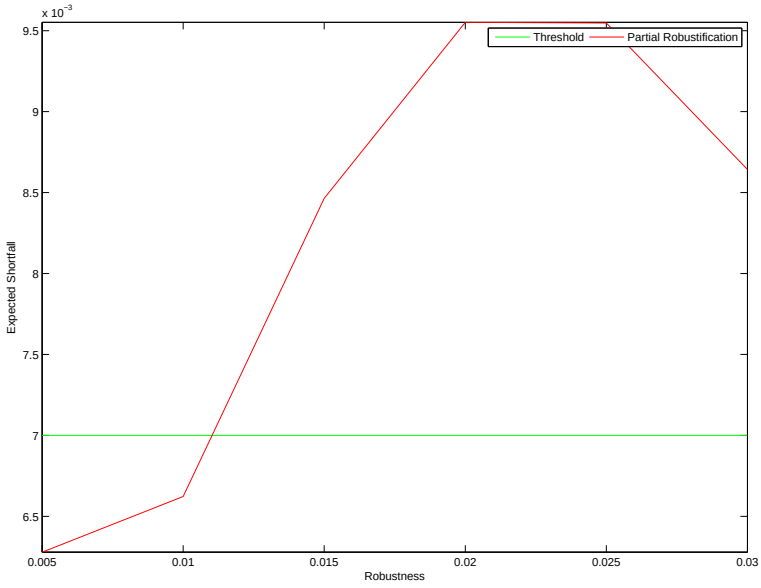


Figure 4.1: Figure (a) shows the expectations of robust portfolio from the two robustification techniques as the robustness parameter increases. Figure (b) and (c) show the comparisons of the respective robust portfolios with the expectation of the nominal program instance and the expectation of the robust portfolios under the empirical measure.



(a)



(b)

Figure 4.2: Figure (a) shows the worst case returns for the partially robust portfolios. Figure (b) shows the worst case expected shortfalls for the partially robust portfolios.

that comprise the portfolio are nearly always (except for $\varepsilon = 0.03$) uniformly weighted.

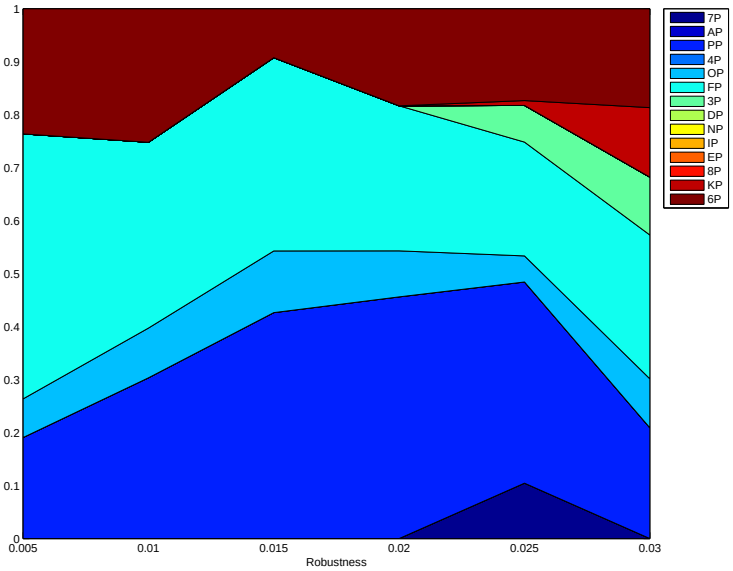
To conclude the comparison between the two methods it is investigated what effect the number of data points has on relation between partial and full robustification. It seems plausible that with a bigger number of data points the additional robustness gained by shifting the points is smaller than with a smaller data set. The intuition behind this guess is that if there are more points in the empirical measure, then only shifting probability mass is less of an restriction than if there are only very few data points with little possibility for shifting mass between them. To test this, the analysis presented in Figure 4.1 is repeated for a sub data set (in fact the first 26 data points of the above data set) and the differences between full and partial robustification are investigated. In Figure 4.4 the differences are plotted for the two data sets. As expected the bigger data set yields a smaller difference between the expected returns than the robustification of the smaller data set.

4.6 Conclusion

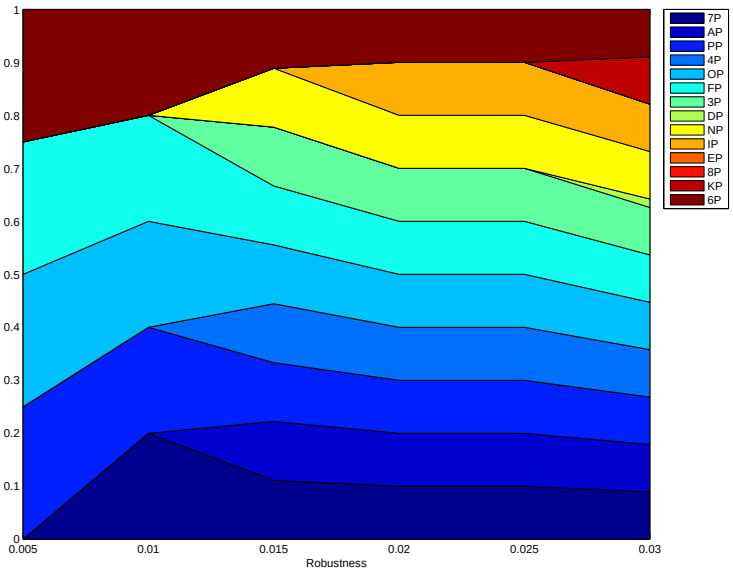
In this paper a framework for the robustification of single stage stochastic optimization problems with ambiguity about the distributions of the random variables that enter the problem formulation is presented. The robustification is achieved by considering the worst case amongst all distributions that are close to the empirical distribution, whereby the distance is measured by the Kantorovich metric. This metric is a very general distance concept that allows for non-parametric ambiguity sets and which has close ties to theoretical results from probability theory and statistics. The resulting robustified problems are computationally intractable and are reformulated to semi-definite non-convex optimization problems.

In Section 4.4 an application from the field of portfolio selection is studied in detail and techniques to cope with the inherent difficulties posed by the non-convexities in the problems are presented. The so called DCA algorithm is applied to solve the portfolio selection problem numerically.

In the last section we compare the results of the computations with earlier results obtained in [54] by partial robustification of the corresponding problem. The differences in the two methods are discussed and it is concluded that the full robustification that leads to non convex inner problems is superior in the sense that the resulting portfolio are significantly more robust with respect to variations in the chosen ambiguity sets. However, it can be expected that for an increasing number of scenarios this effect diminishes and partial robustification can be applied without too much loss in robustness.



(a)



(b)

Figure 4.3: Part (a) depicts the portfolio compositions for the partial robustifications for different values of the robustness parameter, while part (b) shows the fully robustified portfolios. .

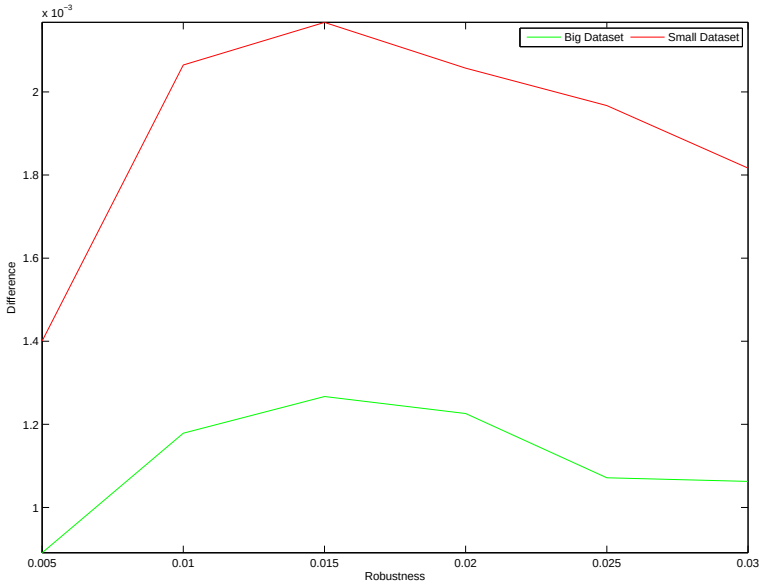


Figure 4.4: Differences in expected return between partial and full robustification for the whole data set from 4.1 and a subset of 26 points.

Appendices

APPENDIX A

D.C. Functions and the Difference of Convex Algorithm

A.1 Difference of Convex Functions

In this section we will briefly review the definition of D.C. functions and give the proofs for their most important properties in optimization. If not stated otherwise the proofs in this section are adapted from [44].

Definition A.1. *A function $f : \mathbb{R}^n \to \mathbb{R}$ is called D.C. on a convex set $C \subseteq \mathbb{R}^n$ if there are convex functions $g : C \to \mathbb{R}$ and $h : C \to \mathbb{R}$ such that*

$$f(x) = g(x) - h(x), \quad \forall x \in C.$$

We call a function a D.C. function, if it is D.C. on the whole of $\mathbb{R}^n$. A function f is called locally D.C. *if for every $x_0 \in \mathbb{R}^n$ there exists a neighborhood $\mathscr{U}$ of x_0 such that f is D.C. on this neighborhood.*

The next results demonstrate that D.C. functions are stable under all operations usually performed in optimization.

Theorem A.1. *Let $f_i = g_i - h_i$ be D.C. function for $1 \leq i \leq m \in \mathbb{N}$, then*

1. *Every linear combination of the f_i's is again a D.C. function.*
2. *$\max_i f_i$ and $\min_i f_i$ are D.C. functions.*
3. *$|f_i|$, $f_i^+(x) = \max(0, f_i(x))$ and $f_i^-(x) = -\min(0, f_i(x))$ are D.C. functions.*
4. *$f_i f_j$ is D.C. function as well as $\frac{f_i}{f_j}$ if $f_j \neq 0$.*

Proof. 1. This is trivial.

2. We prove the result for the maximum, the minimum can be argued analogously. First observe that
$$f_i = g_i - h_i = g_i + \sum_{j \neq i} h_j - \sum_{j=1}^{m} h_j$$
and the last sum is independent of i. We therefore get
$$\max_i f_i = \max_i (g_i - h_i) = \max_i \left\{ g_i + \sum_{j \neq i} h_j \right\} - \sum_{j=1}^{m} h_j$$
which is a D.C. decomposition of the maximum.

3. From part 2 it follows that $\max(f_i,0) = \max(g_i,h_i) - h_i$ and $\max(-f_i,0) = \max(g_i,h_i) - g_i$, and since $|f_i| = \max(f_i,0) + \max(-f_i,0)$, we get

$$|f_i| = 2\max(g_i,h_i) - (g_i + h_i).$$

4. See for example [40].

□

Next we give a result, that demonstrates the richness of the class of D.C. functions by establishing that the D.C. functions are dense in the set of all continuous functions.

Theorem A.2. *The set of D.C. functions on a compact set $K \subseteq \mathbb{R}^n$ is dense in the space of continuous functions $(\mathscr{C}(K), ||.||_\infty)$.*

To prove Theorem A.2 we need the following Lemma proved in [40].

Lemma A.1. *Every locally D.C. function is D.C.*

Proof of Theorem A.2. For any $f \in \mathscr{C}^2(K)$ (the space of all twice continuously differentiable functions on the compact $K \subseteq \mathbb{R}^n$) we know that for every $x_0 \in K$ all the second derivatives are bounded on a neighborhood $\mathscr{U}$ of x_0 and therefore the Hessian $\nabla^2 f$ is bounded on $\mathscr{U}$. Therefore there is a constant $c \in \mathbb{R}$ such that

$$\nabla^2(f(x) + c||x||_2^2) = \nabla^2 f(x) + c2I > 0$$

where I is the identity matrix. Therefore the function $f + c||\cdot||_2^2$ is convex and

$$f = (f + c||\cdot||_2^2) - c||\cdot||_2^2$$

is D.C. in $\mathscr{U}$ and since x_0 was arbitrary f is locally D.C.. From Lemma A.1 we get that f is a D.C. function.

Since $\mathscr{C}^2(K)$ contains the polynomials and these are dense in $\mathscr{C}(K)$ by the Theorem of Weierstraß, the D.C. functions are dense in $\mathscr{C}(K)$ as well. □

A.2 The Difference of Convex Algorithm (DCA)

In this section we present proofs for the most important results concerning the DCA and its convergence. The proofs are adapted from [71] or [72] if not stated otherwise. The results on the hybrid DCA and the finite convergence concepts for polyhedral functions are new – though kept in the spirit of the rest of the results. We start by noting that every unconstrained D.C. program of the form

$$(P) \quad \alpha = \inf\{f(x) = g(x) - h(x) : x \in X\} \tag{A.1}$$

with $g,h : X \to \mathbb{R}$ proper, convex and lower semi continuous has a "dual" of the form

$$(D) \quad \alpha = \inf\{f^*(y) = h^*(y) - g^*(y) : y \in X^*\}.$$

where f^* is the conjugate of f, i.e.

$$f^*(y) = \sup_{x \in X} \{\langle x, y\rangle - f(x)\}.$$

This can easily be seen by the following calculation

$$\begin{aligned} \inf\{f(x) = g(x) - h(x) : x \in X\} &= \inf_{x \in X} \{g(x) + \inf\{h^*(y) - \langle x, y\rangle : y \in X^*\}\} \\ &= \inf_{y \in X^*} \{h^*(y) + \inf\{g(x) - \langle x, y\rangle : x \in X\}\} \\ &= \inf\{h^*(y) - g^*(y) : y \in X^*\} \end{aligned}$$

Under the above conditions the two problems are symmetric, in the sense that $(D)^* = (P)$.

Recall that the domain of a convex function $\theta : X \to \mathbb{R} \cup \{\pm\infty\}$ denoted by $dom(\theta)$ is defined as

$$dom(\theta) = \{x \in X : \theta(x) < \infty\}.$$

We will assume throughout this appendix, that the objective values of (A.1) are finite, i.e.

$$dom(g) \subseteq dom(h)$$

and therefore also by the above duality

$$dom(h^*) \subseteq dom(g^*).$$

Definition A.2 (Critical Point)**.** *$x \in X$ is called a* critical point *of $f = g - h$, if $\partial g(x) \cap \partial h(x) \neq \emptyset$.*

The definition of critical points can be motivated by the following result.

Proposition A.1. *Every minimizer of* (A.1) *in int$(dom(h))$ is a critical point.*

Proof. Suppose $x^* \in X$ is such that

$$g(x^*) - h(x^*) = \inf\{f(x) = g(x) - h(x) : x \in X\}$$

and $y \in \partial h(x^*)$, then

$$\begin{aligned} g(x^*) - h(x^*) &\leq g(x) - h(x), \quad \forall x \in X \\ \Rightarrow g(x^*) - h(x) + \langle x - x^*, y\rangle &\leq g(x) - h(x), \quad \forall x \in X \\ \Rightarrow g(x^*) + \langle x - x^*, y\rangle &\leq g(x), \quad \forall x \in X. \end{aligned}$$

Therefore $y \in \partial h(x^*) \cap \partial g(x^*) \neq \emptyset$. □

It is trivial that the reverse argument does not hold and therefore a critical point need not be an optimum. For an in depth treatment of this necessary condition for optimality see [75].

The next Theorem discusses a necessary and a sufficient condition for local optimality in D.C. programming.

Theorem A.3. *1. If x is a local minimizer of $f = g - h$, then*

$$x \in \mathscr{P}_l := \{x \in X : \partial h(x) \subseteq \partial g(x)\}.$$

Dually if y is a local minimizer of $h^* - g^*$, then

$$y \in \mathscr{D}_l := \{y \in X^* : \partial g^*(y) \subseteq \partial h^*(y)\}.$$

2. Let x^* be a critical point of f, $y^* \in \partial g(x^*) \cap \partial h(x^*)$ and U a neighborhood of x^* such that $U \cap dom(g) \subseteq dom(\partial h)$. If for all $x \in U \cap dom(g) \exists y \in \partial h(x)$ such that

$$h^*(y) - g^*(y) \geq h^*(y^*) - g^*(y^*)$$

then x^* is a local minimizer of $g - h$.

Proof. 1. If x^* is a local minimizer, then there exists a neighborhood U of x^* such that

$$g(x) - g(x^*) \geq h(x) - h(x^*), \quad \forall x \in U \cap dom(g). \tag{A.2}$$

If $y \in \partial h(x^*)$, then from (A.2) we get

$$g(x) - g(x^*) \geq (h(x^*) + \langle x - x^*, y\rangle) - h(x^*) = \langle x - x^*, y\rangle, \quad \forall x \in U \cap dom(g).$$

Since g is a convex function this implies $y \in \partial g(x^*)$.

2. Since $y^* \in \partial g(x^*) \cap \partial h(x^*)$

$$g(x^*) + g^*(y^*) = \langle x^*, y^*\rangle = h(x^*) + h^*(y^*). \tag{A.3}$$

By assumption $\forall x \in U \cap dom(g) \exists y \in \partial h(x)$ such that

$$h^*(y) - g^*(y) \geq h^*(y^*) - g^*(y^*). \tag{A.4}$$

On the other hand from the definition of g^* we get

$$h(x) + h^*(y) = \langle x, y\rangle \leq g(x) + g^*(y) \Rightarrow g(x) - h(x) \geq h^*(y) - g^*(y). \tag{A.5}$$

Combining (A.3), (A.4) and (A.5), we get

$$g(x) - h(x) \geq g(x^*) - h(x^*), \quad \forall x \in U \cap dom(g).$$

□

The DCA is constructed such as to find points (x^*, y^*) which fulfill the necessary optimality condition given in Theorem A.3 point 1. However, before this is treated in any further detail it is shown that this condition is actually sufficient in the case of a polyhedral D.C. function, i.e. a D.C. function where one of the components is a polyhedral function, as is the case in Chapter 2 and 3.

Corollary A.1. *Suppose for x^* there exists a neighborhood U of x^* with $\partial h(x) \cap \partial g(x^*) \neq \emptyset$, $\forall x \in U \cap dom(g)$, then x^* is a local minimizer of $g-h$.*

Proof. Let $x \in U \cap dom(g)$ and $y \in \partial h(x) \cap \partial g(x^*)$. We have

$$g(x^*) + g^*(y) = \langle x^*, y \rangle \leq h(x^*) + h^*(y) \quad \Rightarrow \quad h^*(y) - g^*(y) \geq g(x^*) - h(x^*).$$

Let $y^* \in \partial h(x^*) \cap \partial g(x^*)$, then

$$g(x^*) + g^*(y^*) = h(x^*) + h^*(y^*) \Rightarrow g(x^*) - h(x^*) = h^*(y^*) - g^*(y^*)$$

and thereby

$$h^*(y) - g^*(y) \geq h^*(y^*) - g^*(y^*).$$

The conditions for Theorem A.3 part 2 are therefore fulfilled and x^* is a local optimum of $g-h$. □

Theorem A.4. *Every closed and proper locally polyhedral function f has the diff-max property, i.e. for every point $x \in dom(f)$ there exists a neighborhood U of x such that*

$$\forall y \in U : \partial f(y) \subseteq \partial f(x).$$

Proof. See [30], Theorem 2. □

Theorem A.5. *If h is a closed and proper locally polyhedral function, then the necessary local optimality condition in Theorem A.3 is also sufficient.*

Proof. Follows directly from Corollary A.1 and Theorem A.4. □

As mentioned earlier the difference of convex algorithm is motivated by the local optimality conditions in Theorem A.3 part 1. The following Theorem furnishes the connection between problems (A.6), (A.7) below and these optimality conditions.

For x^* and y^* consider the two problems

$$\inf\{h^*(y) - g^*(y) : y \in \partial h(x^*)\} = \inf\{\langle x^*, y \rangle - g^*(y) : y \in \partial h(x^*)\} \quad (A.6)$$

$$\inf\{g(x) - h(x) : x \in \partial g^*(y^*)\} = \inf\{\langle x, y^* \rangle - h(x) : x \in \partial g^*(y^*)\} \quad (A.7)$$

and call the solutions sets $\mathscr{S}(x^*)$ and $\mathscr{T}(y^*)$ respectively.

Theorem A.6. *1.* $x^* \in \mathscr{P}_l \Leftrightarrow \exists\, y^* \in \mathscr{S}(x^*) : x^* \in \partial g^*(y^*)$.

2. $y^* \in \mathscr{D}_l \Leftrightarrow \exists\, x^* \in \mathscr{T}(y^*) : y^* \in \partial h(x^*)$.

Proof. By duality its enough to show (1). ($\Rightarrow$): Suppose $x^* \in \mathscr{P}_l$. If $y \in \mathscr{S}(x^*)$, then from the fact that $y \in \partial h(x^*) \subseteq \partial g(x^*)$, it follows that $x^* \in \partial g^*(y)$. Hence, it is enough to show that the infimum in (A.6) is attained at some y^*. But this follows directly from $x^* \in \mathscr{P}_l$, since in this case $\langle x^*, y\rangle - g^*(y) = g(x^*)$ for all $y \in \partial h(x^*)$ and therefore $\mathscr{S}(x^*) = \partial h(x^*)$.
($\Leftarrow$): Let $x^* \in X$ be such that $x^* \in \partial g^*(y^*)$ for some $y^* \in \mathscr{S}(x^*)$, then

$$-\langle x^*, y\rangle \leq -g^*(y) + g^*(y^*) - \langle x^*, y^*\rangle = -g^*(y) - g(x^*), \quad \forall y \in \partial h(x^*)$$

and since $-g^*(y) \leq g(x) - \langle x, y\rangle$ for all $x \in X$, we get

$$g(x) \geq g(x^*) + \langle x - x^*, y\rangle, \quad \forall x \in X,\ \forall y \in \partial h(x^*)$$

and therefore $\partial h(x^*) \subseteq \partial g(x^*)$. □

Remark A.1. *1. By construction a solution x to* (A.7) *with parameter y is in $\partial g^*(y)$. If y itself is obtained from solving the problem* (A.6) *for some x', then $y \in \mathscr{S}(x')$. If $x = x'$ by Theorem A.6 part 1 x is a local optimum. This is the idea of the complete DCA: finding sequences of points x^k and y^k such that $x^{k+1} \in \mathscr{S}(y^k)$ and $y^{k+1} \in \mathscr{S}(x^k)$. From the above it is clear that if $x^{k+1} = x^k$ then the algorithm found a point in $\mathscr{P}_l$.*

2. If the outlined algorithm does not terminate, but the sequences $(x_k)_{k\in\mathbb{N}}$ and $(y_k)_{k\in\mathbb{N}}$ are bounded, then there exist limit points (x^*, y^*). By continuity properties of the subgradient mapping it can be argued, that $(x_k)_{k\in\mathbb{N}}$ and $(y_k)_{k\in\mathbb{N}}$ at least converge to points in $\mathscr{P}_l$ and $\mathscr{D}_l$ respectively.

Since the problems (A.6) and (A.7) might be hard to solve there exists the so called *simplified DCA*, a variant of the above algorithm which is solely based on the repeated solution of convex problems. More specifically instead of solving (A.6) and (A.7) to obtain y^k and x^{k+1}, the convex problems

$$\inf\left\{h^*(y) - \left(g^*(y^{k-1}) + \langle x^k, y - y^{k-1}\rangle\right) : y \in X^*\right\} \tag{A.8}$$

$$\inf\left\{g(x) - \left(h(x^k) + \langle x - x^k, y^k\rangle\right) : x \in X\right\} \tag{A.9}$$

are solved. These problems can be viewed as a convex approximation of the original and the dual problem respectively. Clearly when constructed as solutions to (A.8) and (A.9) y^k and x^{k+1} fulfill $y^k \in \partial h(x^k)$ and $x^{k+1} \in \partial g^*(y^k)$.

The difference between the complete and the simplified version of the DCA is therefore, that in the simplified version an arbitrary element of the respective subgradient sets is chosen, while in the complete version this element is found as a solution to a non-convex optimization problem.

In some of the applications it might be possible to find solutions to one of the problems (A.6) or (A.7) but not to the other one. We call the following algorithm *hybrid DCA*: given x^k find y^k by solving the following problem

$$\inf\left\{h^*(y) - g^*(y) : y \in \partial h(x^k)\right\} = \inf\left\{\langle x^k, y\rangle - g^*(y) : y \in \partial h(x^k)\right\} \tag{A.10}$$

and given y^k find x^{k+1} by solving the convex problem

$$\inf\left\{g(x) - \left(h(x^k) + \langle x - x^k, y^k\rangle\right) : x \in X\right\}. \tag{A.11}$$

The next Theorem summarizes some elementary properties which are valid for all the variants of the DCA.

Theorem A.7. *For the complete, the hybrid as well as the simplified DCA the following properties hold.*

1. *The DCA is a descent method for both the primal and the dual problem, i.e.*

$$\begin{aligned} g(x^{k+1}) - h(x^{k+1}) &\leq g(x^k) - h(x^k) && \text{(A.12)} \\ h^*(y^{k+1}) - g^*(y^{k+1}) &\leq h^*(y^k) - g^*(y^k). && \text{(A.13)} \end{aligned}$$

2. *Consecutive points where the objective function stays constant can be characterized as follows:*

$$g(x^k) - h(x^k) = g(x^{k+1}) - h(x^{k+1}) \Leftrightarrow x^k \in \partial g^*(y^k),\ y^k \in \partial h(x^{k+1}). \tag{A.14}$$

In particular $y^k \in \partial g(x^k) \cap \partial h(x^k)$ and $x^k \in \partial h^*(y^k) \cap \partial g^*(y^k)$ therefore x^k as well as y^k are critical points of the primal and dual problem respectively. A dual version of (A.14) *holds.*

3. *If $\alpha = \inf_{x\in X}(g(x) - h(x))$ is finite and the sequences $(x^k)_{k\in\mathbb{N}}$ and $(y^k)_{k\in\mathbb{N}}$ are bounded then every cluster point x^* of $(x^k)_{k\in\mathbb{N}}$ is a critical point and there exists a cluster point y^* of $(y^k)_{k\in\mathbb{N}}$, such that*

$$\lim_{l\to\infty}\langle x^{k_l}, y^{k_l}\rangle = \lim_{l\to\infty} g(x^{k_l}) + g^*(y^{k_l}) = \lim_{l\to\infty} h(x^{k_l}) + h^*(y^{k_l})$$

where $x^{k_l} \to x^*$ and $y^{k_l} \to y^*$.

Proof. 1. Because $y^k \in \partial h(x^k)$

$$g(x^{k+1}) - h(x^{k+1}) \leq g(x^{k+1}) - \langle x^{k+1} - x^k, y^k\rangle - h(x^k)$$

and since $x^{k+1} \in \partial g^*(y^k)$ we have $y^k \in \partial g(x^{k+1})$ and hence that

$$g(x^{k+1}) - \langle x^{k+1} - x^k, y^k\rangle - h(x^k) \leq g(x^k) - h(x^k).$$

Summarizing we get (A.12) by

$$g(x^{k+1}) - h(x^{k+1}) \leq g(x^{k+1}) - \langle x^{k+1} - x^k, y^k\rangle - h(x^k) \leq g(x^k) - h(x^k). \tag{A.15}$$

Since $y^k \in \partial h(x^k)$ and $x^{k+1} \in \partial g^*(y^k)$ it holds that

$$g(x^{k+1}) - \langle x^{k+1} - x^k, y^k\rangle - h(x^k) = h^*(y^k) - g^*(y^k)$$

which together with (A.15) yields

$$h^*(y^{k+1}) - g^*(y^{k+1}) \leq g(x^{k+1}) - h(x^{k+1}) \leq h^*(y^k) - g^*(y^k) \tag{A.16}$$

and therefore (A.13).

2. If $x^k \in \partial g^*(y^k)$, $y^k \in \partial h(x^{k+1})$ then using $x^{k+1} \in \partial g^*(y^k)$, $y^k \in \partial h(x^k)$ it follows

$$\begin{aligned} g(x^{k+1}) - h(x^{k+1}) &= (\langle x^{k+1}, y^k\rangle - g^*(y^k)) - (\langle x^{k+1}, y^k\rangle - h^*(y^k)) \\ &= h^*(y^k) - g^*(y^k) = (\langle x^k, y^k\rangle - h(x^k)) - (\langle x^k, y^k\rangle - g(x^k)) \\ &= g(x^k) - h(x^k). \end{aligned}$$

If on the other hand $g(x^{k+1}) - h(x^{k+1}) = g(x^k) - h(x^k)$, then $y^k \in \partial h(x^k)$ and $x^{k+1} \in \partial g^*(y^k)$ implies

$$0 = g(x^{k+1}) - h(x^{k+1}) - (g(x^k) - h(x^k)) \leq g(x^{k+1}) - g(x^k) - \langle x^{k+1} - x^k, y^k\rangle \leq 0$$

therefore $g(x^{k+1}) - g(x^k) = \langle x^{k+1} - x^k, y^k\rangle$ and since $x^{k+1} \in \partial g^*(y^k)$

$$g(x^k) + g^*(y^k) = \langle x^k, y^k\rangle \Rightarrow x^k \in \partial g^*(y^k).$$

In the same way we get

$$0 = g(x^k) - h(x^k) - (g(x^{k+1}) - h(x^{k+1})) \geq h(x^{k+1}) - h(x^k) + \langle x^k - x^{k+1}, y^k\rangle \geq 0$$

and therefore analogously

$$h(x^{k+1}) + h^*(y^k) = \langle x^{k+1}, y^k\rangle \Rightarrow y^k \in \partial h(x^{k+1}).$$

3. If α is finite the sequence $(g(x^k) - h(x^k))_k$ converges because of (1). Therefore

$$\begin{aligned} 0 &= \lim_{k\to\infty} \Big(g(x^{k+1}) - h(x^{k+1}) - g(x^k) + h(x^k)\Big) \\ &\leq \lim_{k\to\infty} \Big(g(x^{k+1}) - g(x^k) - \langle x^{k+1} - x^k, y^k\rangle\Big) \\ &= \lim_{k\to\infty} \Big(-g^*(y^k) + (\langle x^k, y^k\rangle - g(x^k))\Big) \\ &= \lim_{k\to\infty} \left(-\sup_{x\in\mathbb{R}^d} (\langle x, y^k\rangle - g(x)) + \langle x^k, y^k\rangle - g(x^k)\right) \leq 0 \end{aligned}$$

where the first inequality follows from $y^k \in \partial h(x^k)$, and the second from $g(x^{k+1}) + g^*(y^k) = \langle x^{k+1}, y^k\rangle$ which is true since $x^{k+1} \in \partial g^*(y^k)$. From the above we conclude that

$$\lim_{k\to\infty} g(x^k) + g^*(y^k) = \lim_{k\to\infty} \langle x^k, y^k\rangle \tag{A.17}$$

and by an analogous argument

$$\lim_{k\to\infty} h(x^k) + h^*(y^k) = \lim_{k\to\infty} \langle x^k, y^k\rangle. \tag{A.18}$$

If the sequences $(x^k)_{k\in\mathbb{N}}$ and $(y^k)_{k\in\mathbb{N}}$ are bounded there exists subsequences $(x^{k_l})_{l\in\mathbb{N}}$ and $(y^{k_l})_{l\in\mathbb{N}}$ such that $x^{k_l} \to x^*$ and by closedness of the subgradient mapping $y^{k_l} \to y^* \in \partial h(x^*)$ (see [64], Theorem 24.4).

Define the lower semi continuous function $\theta(x,y) = g(x) + g^*(y)$. then because of the semi-continuity of θ we get

$$\theta(x^*,y^*) \leq \liminf_{l\to\infty} \theta(x^{k_l},y^{k_l}) = \lim_{l\to\infty} \theta(x^{k_l},y^{k_l}) = \langle x^*,y^*\rangle$$

where the last equality follows from (A.17). By the Fenchel inequality $g(x^*) + g^*(y^*) = \langle x^*,y^*\rangle$ follows. Analogously one can show $h(x^*) + h^*(y^*) = \langle x^*,y^*\rangle$ by using (A.18).

□

The last theorem of this section is devoted to two convergence properties of the complete and the hybrid DCA.

Theorem A.8. *1. If for the hybrid DCA $g(x^{k+1}) - h(x^{k+1}) = g(x^k) - h(x^k)$, then $x^k \in \mathscr{P}_l$ and y^k is a critical point, i.e. $g^*(y^k) \cap h^*(y^k) \neq \emptyset$.*

2. If either h or g^* are polyhedral functions, then the objective values of the hybrid DCA have finite convergence.

Proof. 1. If the algorithm stops finitely, i.e. $g(x^{k+1}) - h(x^{k+1}) = g(x^k) - h(x^k)$, then $x^k \in g^*(y^k)$. By construction $y^k \in \mathscr{S}(x^k)$ for the complete DCA as well as the hybrid DCA and therefore $x^k \in \mathscr{P}_l$ by an application of Theorem A.6.

Since $y^k \in \partial h(x^k)$ and $x^k \in \partial g^*(y^k)$ (by Theorem A.7 part 2), we have $x^k \in \partial h^*(y^k) \cap \partial g^*(y^k)$ and therefore y^k is a critical point.

2. We only treat the case where h is a polyhedral function. The other case follows by switching to the dual problem.

The result relies on the fact that, since h being a polyhedral function of the form

$$h(x) = \max\left\{\langle x,a^i\rangle - \alpha_i : i \in I\right\}$$

where I is a finite index set, it holds that

$$\partial h(x) = co\left\{a^i : i \in I(x)\right\}$$

with $I(x) = \left\{i \in I : h(x) = \langle x,a^i\rangle - \alpha_i\right\}$. This means that there are only finitely many different possible subgradient sets $\partial h(x)$. Now define a mapping $T : \mathbb{R}^n \to \mathbb{R}^n$ with $T(x) \in \partial h(x)$ which selects y^k from $\partial h(x^k)$. From the above it follows that there are only finitely possible y^ks and since the sequence $(g(x^k) - h(x^k))_k$ is decreasing eventually there will be some k for which $g(x^k) - h(x^k) = g(x^{k+i}) - h(x^{k+i})$ for all $i \in \mathbb{N}$.

□

Remark A.2. *In the above Theorem the algorithm is said to converge, if $(g(x^k) - h(x^k))_k$ remains constant after a certain $K \in \mathbb{N}$. This differs slightly from the view taken in [4], where the convergence is defined as the convergence of (x^k,y^k). The latter can be achieved in case of polyhedral D.C. functions even for the simple DCA in finitely many steps via so called* natural choices *for the subgradients.*

The advantage of this approach is that the algorithm can be stopped after the first iteration in which the objective function value does not decrease. In our approach the algorithm has to be continued until it is established that the pairs (x^k, y^k) are circling through a finite selection of equivalent solutions.

Since convergence speed was not an issue in the problems treated in this work and we are primarily concerned with convergence in the objective function values there was no need to force convergence of (x^k, y^k) by special choices of T. Therefore the notion of convergence described above was used.

APPENDIX B

Extreme and exposed points

In this Appendix the concepts of extreme and exposed points, the relations between the extreme and the exposed points of a convex set and the Krein-Milman and similar theorems are discussed. In particular we give proofs for Theorems 4.1 and 4.4. The presentation is mostly from [24] and [25].

For readability we repeat the definitions of extreme and exposed points from chapter 4.

B.1 Extreme Points

Definition B.1 (Extreme Point)**.** *Let $C \subseteq E$ be a convex set in a LCS. A point $x \in C$ is called an extreme point of C, if $C \setminus \{x\}$ is still a convex set. We denote the set of extreme points of a convex set C by $ext(C)$.*

The reason extreme points are of importance for the results in chapter 4 is rooted in the following theorem which is stated in the main text as Theorem 4.1.

Theorem B.1 (Bauer Minimum Principle)**.** *Let E be a Hausdorff LCS, $C \subset E$ be a non-empty compact convex set and f a concave lower semi-continuous function. Then f attains its minimum over C on an extreme point of C.*

Proof. Define

$$\mathscr{J} = \{F \subseteq C : F \neq \emptyset, \bar{F} = F, \forall a,b \in C : (a,b) \cap F \neq \emptyset \text{ it follows } (a,b) \in F\},$$

where $(a,b) = \{\lambda a + (1-\lambda)b : \lambda \in (0,1)\}$ and $\bar{F}$ is the closure of F. Note that

$$C \in \mathscr{J} \text{ and for } (X_i) \subseteq \mathscr{J} : \bigcap \{X_i\} \neq \emptyset \Rightarrow \bigcap \{X_i\} \in \mathscr{J}. \tag{B.1}$$

Now define for each $F \in \mathscr{J}$ the set $F' = \arg\min_{x \in F} f(x) \neq \emptyset$. The set F' is contained in $\mathscr{J}$ since

$$F' = \bigcap_{n=1}^{\infty} \left\{ x \in F : f(x) \leq \inf f + \frac{1}{n} \right\}$$

and if $x_0 \in (a,b) \cap F'$ then by concavity for any $x,y \in (a,b)$ such that $x_0 = \lambda x + (1-\lambda)y$ with $0 < \lambda < 1$ it holds that $f(x_0) \geq \lambda f(x) + (1-\lambda) f(y) \geq f(x_0)$ and therefore $(a,b) \subseteq F'$.

Now we define a partial order on $\mathscr{J}$ by defining $F_1 \leq F_2 :\Leftrightarrow F_1 \supseteq F_2$. The maximal elements with respect to this order are singletons $\{x\}$. To see this take $F \in \mathscr{J}$ with $\{x,y\} \subseteq F$. Then x and y can be separated by a continuous linear functional l such that $l(x) > l(y)$ and by forming F' for l it becomes clear that $F' > F$, since $x \notin F'$. Now note that $\mathscr{J}$ is inductive, i.e. every

totally ordered subset X_i has a maximal element in $\mathscr{J}$. This is the case because X_i are nested and compact and therefore $\bigcap\{X_i\} \neq \emptyset$. It follows from (B.1) that $\bigcap\{X_i\} \neq \emptyset \in \mathscr{J}$ and therefore is a maximal element for $\{X_i\}$. An application of Zorn's lemma shows that every $F \in \mathscr{J}$ is majorized by a maximal element. In particular the set $F' = \arg\min_{x\in C} f(x) \neq \emptyset$ is majorized by the singleton $\{x\} \subseteq F$. It is easy to see that a singleton $\{x\}$ is contained in $\mathscr{J}$ iff it is an extreme point of C. □

Corollary B.1. *Let E be a Hausdorff LCS, $C \subset E$ be a non-empty compact convex set, then C has an extreme point.*

Proof. Apply B.1 to the constant function 1. □

Using Bauer Minimum Principle we can easily prove the Krein-Milman theorem which shows that a convex set can be represented by its extreme points.

Theorem B.2 (Krein-Milman)**.** *Let K be a compact, convex set in a LCS E, then $K = \overline{co(ext(K))}$.*

Proof. Clearly $\overline{co(ext(K))} \subseteq K$, since K is closed and convex. Suppose the other inclusion would not hold and fix $x \in K\backslash\overline{co(ext(K))}$. By the Hahn-Banach theorem there is a linear map $l : E \to \mathbb{R}$ such that $max\left\{l(x) : x \in \overline{co(ext(K))}\right\} < l(x)$, hence by Theorem B.1 there is an extreme point of K not contained in $\overline{co(ext(K))}$, a contradiction. □

B.2 Exposed points

As we argued in Chapter 4 it is hard to characterize the extremals of a Kantorovich neighborhood directly. We therefore resort to characterizing the exposed points which are somewhat easier to handle and then conclude that the extreme points are essentially of the same form as the exposed points. For the last step in this argument we need that the exposed points of a convex set C are dense in the extreme points. This result will be shown in the following.

Definition B.2. *An exposed point x of a convex set C is defined by the property, that it is possible to separate the convex $C\backslash\{x\}$ set from x via a linear functional. This means one can find a linear functional l such that $l(x) > l(y)$, $\forall y \in C\backslash\{x\}$ or equivalently $argmax_{y\in C} l(y) = \{x\}$.*

We will now establish the a result which sometimes is called Straszewicz Theorem and states that the exposed points are dense in the extreme points of compact, convex set C. The original result was proven for $\mathbb{R}^n$ and can be found for example in [64], Theorem 18.6. The proof is heavily based on geometrical arguments, which carry over to the setting of real Hilbert spaces. We will therefore show the result first for real Hilbert spaces and use an embedding Lemma to get the result for metrizable subsets of LCS.

Lemma B.1. *Let C be a convex set in a real Hilbert space E and $x_0 \in C$. Now define the point c to be the farthest point of C from x_0. Then c is an exposed point of C.*

Proof. Since for all $x \in C$

$$||c - x_0||^2 \geq ||x - x_0||^2 = ||x - c + x - x_0||^2 = ||x - c||^2 + 2\langle x - c, c - x_0\rangle + ||c - x_0||^2$$

we have

$$2\langle x-c, c-x_0\rangle \leq -||x-c||^2 \leq 0.$$

Now define the linear functional $f(x) = \langle x, c-x_0\rangle$. From the inequality above we get $f(x) \leq f(c)$ for all $x \in C$ with equality iff $x = c$. Therefore c is an exposed point of C. □

Lemma B.2. *Let C be a compact, convex set in a real Hilbert space E, and let f be a continuous linear functional such that $f(x) < \alpha$ for some $x \in E$. Then there exists an exposed point c of C such that $f(x) < \alpha$.*

Proof. Let $u \in E$ be the Riesz representer of f, i.e. $f(x) = \langle x, u\rangle$, and $x_0 = x + \lambda u$. Let c be a the point that is farthest away from x_0. c is an exposed point of C by Lemma B.1. We further have

$$\begin{aligned} ||x_0 - x||^2 &\leq ||x_0 - c||^2 = ||x - c + \lambda u||^2 = ||x-c|| + 2\lambda\langle x-c, u\rangle + ||\lambda u||^2 \\ &= ||x-c||^2 + 2\lambda(f(x) - f(c)) + ||x_0 - x||^2. \end{aligned}$$

Choosing λ sufficiently large we can get $f(c) < \alpha$. □

Next we proof a result, which in some sense generalizes the Krein-Milman Theorem for real Hilbert spaces.

Theorem B.3. *Let C be a compact, convex set in a real Hilbert space E, then*

$$C = \overline{co(exp(C))}.$$

Proof. Since C is closed and convex, we only have to show $C \subseteq \overline{co(exp(C))}$. If not then there would be a $x \in C \backslash \overline{co(exp(C))}$ and by the Hahn-Banach theorem a continuous linear functional f with $f(x) > \alpha \geq \max\left\{f(y) : y \in \overline{co(exp(C))}\right\}$. In this case there would be a $c \in exp(C)$ with $f(c) > \alpha$, which leads to a contradiction. □

The next two results allow us to prove the desired result for Hilbert spaces.

Lemma B.3. *If C is a relatively compact convex set in a metrizable space, then*

$$ext(\bar{C}) \subseteq \overline{ext(C)}$$

Proof. Let $x \in ext(\bar{C})$. If $x \in C$, then $x \in ext(C) \subseteq ext(\bar{C})$. If $x \notin C$, then $x = \lim_{n\to\infty} x_n$ with $x_n \in C$. To finish the proof it is enough to show, that there exists a subsequence (x_{n_k}) with $x_{n_k} \in ext(C)$. If this is not the case, then $\exists N \in \mathbb{N} : x_n \notin ext(C), \quad \forall n \geq N$. If the x_n are not extremals they can be written as $x_n = \frac{1}{2}y_n + \frac{1}{2}z_n$ with $y_n \in C$ and $z_n \in C$. Since C is relatively compact there exists a subsequence $y_{n_k} \to y \in \bar{C}$ and $z_{n_k} \to z \in \bar{C}$ and therefore $x_{n_k} \to \frac{1}{2}y + \frac{1}{2}z$ which is a contradiction to being an extreme point of $\bar{C}$. □

Remark B.1. *The reverse, i.e. $\overline{ext(C)} \subseteq ext(\bar{C})$ does not hold, since the extremals need not be a closed set. This can be seen by the following example*

$$A = \left\{(x,y,0) : x^2+y^2 \leq 1\right\},\ B = \left\{(1,0,z) : -1 \leq z \leq 1\right\},\ C = co(A \cup B)$$

The extremal points of $C = \bar{C}$ are $(1,0,-1)$ and $(1,0,1)$ and $A \backslash (1,0,0)$. Therefore

$$\overline{ext(C)} = ext(\bar{C}) \cup \{(1,0,0)\}.$$

Lemma B.4. *If C is a compact, convex set and $D \subseteq C$ has the property, that*

$$\overline{co(D)} = C$$

then $ext(C) \subseteq \bar{D}$.

Proof. By Lemma B.3 we have

$$ext(C) = ext(\overline{co(D)}) \subseteq \overline{ext(co(D))} \subseteq \bar{D}$$

□

Theorem B.4 (Straszewicz's Theorem for Hilbert Spaces)**.** *In a real Hilbert space E the exposed points of a compact, convex set C are dense in $ext(C)$.*

Proof. Since by Theorem B.3 $C = \overline{co(exp(C))}$, applying theorem B.4 to the set $exp(C)$ yields $ext(C) \subseteq \overline{exp(C)}$. □

We will now extend the previous result by to metrizable convex compact subsets of LCS.

Lemma B.5. *A subspace F of a separable metric space E is separable.*

Proof. Since E is separable and metrizable it is second countable, i.e. the topology has a countable base and therefore countably many semi-norms. The topology on a linear subspace is the initial topology with respect to the inclusion and therefore the semi-norms of F are the restrictions of the semi-norms of E. Therefore F is again second countable and metrizable and therefore also separable. □

Theorem B.5. *Let X be metrizable subset of a Hausdorff LCS vector space E, then there exists a countable set $\{l_n : n \in \mathbb{N}\}$ of affine continuous functions, that separates points in X.*

Proof. X is metrizable, therefore there exist countably many continuous semi-norms $(p_n)_{n\in\mathbb{N}}$ that generate the respective topology. Furthermore the semi norms are point separating, since the space is Hausdorff. Consider now the space $\mathscr{A}$ of rational linear combinations of functions in $\{1\} \cup \{p_n : n \in \mathbb{N}\}$. $\mathscr{A}$ is a sub-algebra of $C(X,\mathbb{R})$, is still countable and point separating and therefore dense in $(C(X,\mathbb{R},(p_K)_K)$ (where p_K is the semi-norm of uniform convergence on the compact set $K \subset X$) by the theorem of Stone-Weierstrass. It follows that $(C(X,\mathbb{R},(p_K)_K)$ is separable and therefore the affine continuous functions – being a linear subspace of $C(X,\mathbb{R})$ – are also separable by Lemma B.5. By the Theorem of Hahn-Banach the continuous affine functionals are point separating in $X \subseteq E$ and therefore there exists a countable family $(l_n)_{n\in\mathbb{N}}$, that is point separating in X. □

Theorem B.6 ([25], Problem 17(1))**.** *Let X be a compact, metrizable subset of a Hausdorff LCS vector space, then there is a an affine continuous injection of X into a Hilbert space.*

Proof. From theorem B.5 we get a countable family of point separating continuous, affine functionals on X. Since all the l_n are continuous they are bounded on X by numbers $K_n < \infty$, now define $\hat{l}_n(x) = \frac{l_n(x)}{2^n K_n}$, then the $\hat{l}_n$ are still point separating affine functionals and $\sum_{n=1}^{\infty} \hat{l}_n^2(x) < \infty$. We can now define $\iota : X \to \ell^2$ as $x \mapsto (l_n(x))_{n\in\mathbb{N}}$. ι is obviously an affine, injective functional. Since

$$\sum_{n=N}^{\infty} \hat{l}_n^2(x) \leq 2^{1-N}, \quad \forall x \in X$$

ι is the uniform limit of the functions $\iota_n(x) = (\hat{l}_i)_{i=1}^n$ in ℓ^2 and therefore continuous. □

Theorem B.7 ([25], Problem 17(2))**.** *Let X be a compact, metrizable subset of a Hausdorff LCS vector space, then the exposed points of X are dense in $ext(X)$.*

Proof. Using Lemma B.6 we get a affine continuous injection $f : X \to H$ into a real Hilbert space H. Obviously $Y = im(f)$ is again convex and compact and for every $x \in ext(X)$, $f(x) \in ext(Y)$. To see this assume $f(x) = \lambda y' + (1-\lambda)z'$ with $\lambda \in (0,1)$, $f(y) = y'$ and $f(z) = z'$, then

$$f(x) = \lambda f(y) + (1-\lambda) f(z) = f(\lambda y + (1-\lambda)z)$$

and because f is injective $x = \lambda y + (1-\lambda)z$ contradicting the fact that $x \in ext(X)$. We show that every neighborhood U of x contains exposed points of X. Since $f(x) \in ext(Y)$ by Theorem B.4 there is a $a' = f(a) \in f(U)^\circ$ exposed in Y, i.e. there exists an affine functional $l : H \to \mathbb{R}$ and an $\alpha \in \mathbb{R}$ such that $l(a') > \alpha \geq \max\{l(y) : y \in Y\}$. Since $g = l \circ f$ is affine and $g(a') = l(a) > \alpha \geq \max\{g(x) : x \in X\}$ a is an exposed point of X. □

Bibliography

[1] Affleck-Graves, J., McDonald, B.: Nonnormalities and tests of asset pricing theories. The Journal of Finance **44**(4), 889–908 (1989)

[2] Ahmed, S.: Convexity and decomposition of mean-risk stochastic programs. Mathematical Programming **106**(3, Ser. A), 433–446 (2006)

[3] An, L.T.H.: An efficient algorithm for globally minimizing a quadratic function under convex quadratic constraints. Mathematical Programming **87**(3, Ser. A), 401–426 (2000)

[4] An, L.T.H., Tao, P.D.: Solving a class of linearly constrained indefinite quadratic problems by DC algorithms. J. Glob. Optim. **11**(3), 253–285 (1997)

[5] An, L.T.H., Tao, P.D.: A new algorithm for solving large scale molecular distance geometry problems. In: High performance algorithms and software for nonlinear optimization (Erice, 2001), *Appl. Optim.*, vol. 82, pp. 285–302. Kluwer Acad. Publ., Norwell, MA (2003)

[6] An, L.T.H., Tao, P.D.: The DC (difference of convex functions) programming and DCA revisited with DC models of real world nonconvex optimization problems. Annals of Operations Research **133**, 23–46 (2005)

[7] An, L.T.H., Tao, P.D.: The DC (difference of convex functions) programming and DCA revisited with DC models of real world nonconvex optimization problems. Ann. Oper. Res. **133**, 23–46 (2005)

[8] An, L.T.H., Tao, P.D., Muu, L.D.: A combined D. C. optimization – ellipsoidal branch-and-bound algorithm for solving nonconvex quadratic programming problems. J. Comb. Optim. **2**(1), 9–28 (1998)

[9] An, L.T.H., Tao, P.D., Muu, L.D.: Exact penalty in d.c. programming. Vietnam J. Math. **27**(2), 169–178 (1999)

[10] Andersson, F., Mausser, H., Rosen, D., Uryasev, S.: Credit risk optimization with conditional value-at-risk criterion. Mathematical Programming **89**(2, Ser. B), 273–291 (2001)

[11] Artzner, P., Delbaen, F., Eber, J.M., Heath, D.: Thinking coherently. Risk **10(11)**, 68–71 (1997)

[12] Artzner, P., Delbaen, F., Eber, J.M., Heath, D.: Coherent measures of risk. Mathematical Finance **9**, 203–228 (1999)

[13] Basak, S., Shapiro, A.: Value-at-risk managemant: Optimal policies and asset prices. Review of Financial Studies **14**(2), 371–405 (2001)

[14] Ben-Tal, A., Nemirovski, A.: Robust convex optimization. Math. Oper. Res. **23**(4), 769–805 (1998)

[15] Ben-Tal, A., Nemirovski, A.: Robust solutions of uncertain linear programs. Oper. Res. Lett. **25**(1), 1–13 (1999)

[16] Benati, S., Rizzi, R.: A mixed integer linear programming formulation of the optimal mean/value-at-risk portfolio problem. European Journal of Operational Research **176**(1), 423–434 (2007)

[17] Bomze, I.M., Locatelli, M.: Undominated D.C. decompositions of quadratic functions and applications to branch-and-bound approaches. Computational Optimization and Applications **28**(2), 227–245 (2004)

[18] Burke, J.: An exact penalization viewpoint of constrained optimization. SIAM J. Control Optim. **29**(4), 968–998 (1991)

[19] Calafiore, G.: Ambiguous risk measures and optimal robust portfolios. SIAM J. Optim. **18**(3), 853–877 (electronic) (2007)

[20] Campbell, R., Huisman, R., Koedijk, K.: Optimal portfolio selection in a value-at-risk framework. Journal of Banking & Finance **25**(9), 1789–1804 (2001)

[21] Chen, Z., Epstein, L.: Ambiguity, risk, and asset returns in continuous time. Econometrica **70**(4), 1403–1443 (2002)

[22] Cheon, M.S., Ahmed, S., Al-Khayyal, F.: A branch-reduced-cut algorithm for the global optimization of probabilistically constrained linear programs. Mathematical Programming **108**(2-3, Ser. B), 617–634 (2006)

[23] Chopra, V.K., Ziemba, W.T.: The effect of errors in means, variances and covariances on optimal portfolio choice. J. Portfolio Management **Winter**, 6–11 (1993)

[24] Choquet, G.: Lectures on analysis. Vol. III: Infinite dimensional measures and problem solutions. Edited by J. Marsden, T. Lance and S. Gelbart. W. A. Benjamin, Inc., New York-Amsterdam (1969)

[25] Choquet, G.: Lectures on analysis. Vol. III: Infinite dimensional measures and problem solutions. Edited by J. Marsden, T. Lance and S. Gelbart. W. A. Benjamin, Inc., New York-Amsterdam (1969)

[26] Conn, A., Gould, N., Toint, P.L.: Trust-region methods. MPS/SIAM Series on Optimization. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA (2000)

[27] Dupačová, J.: The minimax problem of stochastic linear programming and the moment problem. Ekonom.-Mat. Obzor **13**(3), 279–307 (1977)

[28] Dupačová, J.: On minimax decision rule in stochastic linear programming. In: Studies on mathematical programming (Papers, Third Conf. Math. Programming, Mátrafüred, 1975), *Math. Methods Oper. Res.*, vol. 1, pp. 47–60. Akad. Kiadó, Budapest (1980)

[29] Dupačová, J.: The minimax approach to stochastic programming and an illustrative application. Stochastics **20**(1), 73–88 (1987)

[30] Durier, R.: On locally polyhedral convex functions. In: Trends in mathematical optimization (Irsee, 1986), *Internat. Schriftenreihe Numer. Math.*, vol. 84, pp. 55–66. Birkhäuser, Basel (1988)

[31] El Ghaoui, L., Oks, M., Oustry, F.: Worst-case value-at-risk and robust portfolio optimization: a conic programming approach. Oper. Res. **51**(4), 543–556 (2003)

[32] Emmer, S., Klüppelberg, C., Korn, R.: Optimal portfolios with bounded capital at risk. Mathematical Finance **11**(4), 365–384 (2001)

[33] Erdoğan, E., Iyengar, G.: Ambiguous chance constrained problems and robust optimization. Math. Program. **107**(1-2, Ser. B), 37–61 (2006)

[34] Fama, E.: The behavior of stock market prices. Journal of Business **38**(1), 34–105 (1965)

[35] Gaivoronski, A.A., Pflug, G.C.: Value-at-risk in portfolio optimization: properties and computational approach. Journal of Risk **7**(2), 1–31 (2005)

[36] Gibbs, A., Su, F.: On choosing and bounding probability metrics. Int. Stat. Rev. **70**(3), 419–435 (2002)

[37] Gilli, M., Këllezi, E.: A global optimization heuristic for portfolio choice with VaR and expected shortfall. In: Computational methods in decision-making, economics and finance, *Applied Optimization*, vol. 74, pp. 167–183. Kluwer (2002)

[38] Gilli, M., Këllezi, E., Hysi, H.: A data-driven optimization heuristic for downside risk minimization. The Journal of Risk **8**(3), 1–18 (2006)

[39] Goldfarb, D., Iyengar, G.: Robust portfolio selection problems. Math. Oper. Res. **28**(1), 1–38 (2003)

[40] Hartman, P.: On functions representable as a difference of convex functions. Pacific J. Math. **9**, 707–713 (1959)

[41] Hoai An, L.T., Tao, P.D.: A continuous approach for globally solving linearly constrained quadratic zero-one programming problems. Optimization **50**(1-2), 93–120 (2001)

[42] Hochreiter, R.: An evolutionary computation approach to scenario-based risk-return portfolio optimization for general risk measures. In: EvoWorkshops 2007, *Lecture Notes in Computer Science*, vol. 4448, pp. 199–207. Springer (2007)

[43] Horst, R., Thoai, N.V.: DC programming: overview. Journal of Optimization Theory and Applications **103**(1), 1–43 (1999)

[44] Horst, R., Tuy, H.: Global optimization. Second edn. Springer-Verlag, Berlin (1993). Deterministic approaches

[45] Jorion, P.: Value at Risk: The New Benchmark for Controlling Market Risk. McGraw-Hill, New York (2000)

[46] Klein, R.W., Bawa, V.S.: The effect of estimation risk on optimal portfolio choice. Journal of Financial Economics **3**(3), 215–231 (1976)

[47] Larsen, N., Mausser, H., Uryasev, S.: Algorithms for optimization of value-at-risk. In: Pardalos, P., Tsitsiringos, V. (eds.) Financial Engineering, e-Commerce and Supply Chain, pp. 129–157. Kluwer Academic Publishers, Dordrecht, Netherlands (2002)

[48] Lüthi, H.J., Doege, J.: Convex risk measures for portfolio optimization and concepts of flexibility. Mathematical Programming **104**(2-3, Ser. B), 541–559 (2005)

[49] Malevergne, Y., Sornette, D.: Value-at-risk-efficient portfolios for a class of super- and sub-exponentially decaying assets return distributions. Quant. Finance **4**(1), 17–36 (2004)

[50] Markowitz, H.M.: Portfolio selection. The Journal of Finance **7**(1), 77–91 (1952)

[51] McNeil, A.J., Frey, R., Embrechts, P.: Quantitative risk management. Princeton Series in Finance. Princeton University Press (2005)

[52] Natarajan, K., Pachamanova, D., Sim, M.: Incorporating asymmetric distributional information in robust value-at-risk optimization (2007 (to appear)). *Management Science*

[53] Pang, J., Leyffer, S.: On the global minimization of the value-at-risk. Optimization Methods and Software **19**(5), 611–631 (2004)

[54] Pflug, G., Wozabal, D.: Ambiguity in portfolio selection. Quantitative Finance **7**(4), 435–442 (2007)

[55] Pflug, G.C.: Some remarks on the Value-at-Risk and the Conditional Value-at-Risk. In: Probabilistic constrained optimization, *Nonconvex Optimization and its Applications*, vol. 49, pp. 272–281. Kluwer Academic Publishers, Dordrecht, Netherlands (2000)

[56] Pflug, G.C., Römisch, W.: Modeling, Measuring and Managing Risk. World Scientific, Singapore (2007)

[57] Pflug, G.C., Römisch, W.: Modeling, Measuring and Managing Risk. World Scientific (2007)

[58] Pirvu, T.A.: Portfolio optimization under the value-at-risk constraint. Quantitative Finance **7**(2), 125–136 (2007)

[59] P.J., M.: Robust portfolio rules and asset pricing. Review of Financial Studies **17**(4), 951–983 (2004)

[60] Puelz, A.: Value-at-risk based portfolio optimization. In: Uryasev, S., Pardalos, P. (eds.) Stochastic Optimization: Algortihms and Applications. Kluwer Academic Publishers (2001)

[61] Rachev, S.: Probability metrics and the stability of stochastic models. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. John Wiley & Sons Ltd., Chichester (1991)

[62] Rachev, S., Rüschendorf, L.: Mass transportation problems. Vol. I. Probability and its Applications (New York). Springer-Verlag, New York (1998). Theory

[63] Richardson, M., Smith, T.: A test for multivariate normality in stock returns. The Journal of Business **66**(2), 295–321 (1993)

[64] Rockafellar, R.T.: Convex analysis. Princeton Landmarks in Mathematics. Princeton University Press, Princeton, NJ (1997). Reprint of the 1970 original, Princeton Paperbacks

[65] Rockafellar, R.T., Uryasev, S.: Optimization of Conditional Value-at-Risk. The Journal of Risk **2**(3), 21–41 (2000)

[66] Rogosinski, W.W.: Moments of non-negative mass. Proc. Roy. Soc. London Ser. A **245**, 1–27 (1958)

[67] Schultz, R., Tiedemann, S.: Risk aversion via excess probabilities in stochastic programs with mixed-integer recourse. SIAM Journal on Optimization **14**(1), 115–138 (2003)

[68] Shapiro, A., Ahmed, S.: On a class of minimax stochastic programs. SIAM J. Optim. **14**(4), 1237–1249 (electronic) (2004)

[69] Shapiro, A., Kleywegt, A.: Minimax analysis of stochastic problems. Optim. Methods Softw. **17**(3), 523–542 (2002). Stochastic programming

[70] Sion, M.: On general minimax theorems. Pac. J. Math. **8**, 171–176 (1958)

[71] Tao, P.D., An, L.T.: A d. c. optimization algorithm for solving the trust-region subproblem. SIAM J. on Optimization **8**(2), 476–505 (1998). DOI http://dx.doi.org/10.1137/S1052623494274313

[72] Tao, P.D., El Bernoussi, S.: Duality in d. c. (difference of convex functions) optimization. Subgradient methods. Trends in mathematical optimization, 4th French-German Conf., Irsee/FRG 1986, ISNM 84, 277-293 (1988). (1988)

[73] Tasche, D., Tibiletti, L.: Approximations for the value-at-risk approach to risk-return analysis. The ICFAI Journal of Financial Risk Management **1**(4), 44–61 (2004)

[74] Thach, P.T., Konno, H.: On the degree and separability of nonconvexity and applications to optimization problems. Mathematical Programming **77**(1, Ser. A), 23–47 (1997)

[75] Toland, J.F.: On subdifferential calculus and duality in nonconvex optimization. Bull. Soc. Math. France Mém. (60), 177–183 (1979). Analyse non convexe (Proc. Colloq., Pau, 1977)

[76] Tütüncü, R.H., Koenig, M.: Robust asset allocation. Ann. Oper. Res. **132**, 157–187 (2004)

[77] Tuy, H.: Minimax theorems revisited. Acta Math. Vietnam. **29**(3), 217–229 (2004)

[78] Uryasev, S.: Conditional Value-at-Risk: Optimization algorithms and applications. Financial Engineering News **14**, 1–5 (2000)

[79] Wozabal, D.: A framework for optimization under ambiguity. Tech. Rep. TR2008-08, Department of Statistics and Decision Support Systems, University of Vienna, Vienna, Austria (2008)

[80] Wozabal, D.: A new method for value-at-risk constrained optimization using the difference of convex algorithm. Tech. Rep. TR2008-03, Department of Statistics and Decision Support Systems, University of Vienna, Vienna, Austria (2008)

[81] Wozabal, D., Hochreiter, R., Pflug, G.: A d.c. formulation of value-at-risk constrained optimization. Tech. Rep. TR2008-01, Department of Statistics and Decision Support Systems, University of Vienna, Vienna, Austria (2008)

[82] Zackova, J.: On minimax solutions of stochastic linear programming problems. Casopis pro Pěstovânî Mathematiky **91**, 423–430 (1966)

Abstract

APPLICATIONS OF NON-CONVEX OPTIMIZATION IN PORTFOLIO SELECTION

The thesis is concerned with application of non-convex programming to problems of portfolio optimization in a single stage stochastic optimization framework. In particular two different classes of portfolio selection problems are investigated. In both the problems a scenario based approach to modeling uncertainty is pursued, i.e. the randomness in the models is always described by finitely many joint realizations of the asset returns. The thesis is structured into three chapters briefly outlined below:

A D.C. Formulation of Value-at-Risk constrained Optimization In this Chapter the aim is to solve mean risk models with the Value-at-Risk as a risk measure. In the case of finitely supported return distributions, it is shown that the Value-at-Risk can be written as a D.C. function and the mentioned mean risk problem therefore corresponds to a D.C. problem. The non-convex problem of optimizing the Value at Risk is rather extensively treated in the literature and there are various approximative solution techniques as well as some approaches to solve the problem globally.

The reformulation as D.C. problem provides an insight into the structure of the problem, which can be exploited to devise a Branch-and-Bound algorithm for finding global solutions for small to medium sized instances.

The possibility of refining ε-optimal solutions obtained from the Branch-and-Bound framework via local search heuristics is also discussed in this Chapter.

Value-at-Risk constrained optimization using the DCA In this part of the thesis the Value-at-Risk problem is once again investigated with the aim of solving problems of realistic sizes in relatively short time. Since the Value at Risk optimization can be shown to be a NP hard problem, this can only be achieved by sacrificing on the guaranteed globality of the solutions. Therefore a local solution technique for unconstrained D.C. problems called Difference of Convex Algorithm (DCA) is employed. To solve the problem a new variant of the DCA the so called *hybrid DCA* is proposed, which preserves the favorable convergence properties of the computationally hard *complete DCA* as well as the computational tractability of the so called *simple DCA*.

The results are tested for small problems and the solutions are shown to actually coincide with the global optima obtained with the Branch-and-Bound algorithm in most of the cases. For realistic problem sizes the proposed method is shown to consistently outperform known heuristic approximations implemented in commercial software.

A Framework for Optimization under Ambiguity The last part of the thesis is devoted to a different topic which received much attention in the recent stochastic programming literature: the topic of robust optimization. More specifically the aim is to robustify single stage stochastic optimization models with respect to *uncertainty about the distributions* of the random variables

involved in the formulation of the stochastic program. The aim is to explore ways of explicitly taking into account ambiguity about the distributions when finding a decision while imposing only very weak restrictions on possible probability models that are taken into consideration.

Ambiguity is defined as possible deviation from a discrete reference measure $\hat{P}$ (in this work the empirical measure). To this end a so called *ambiguity set* $\mathscr{B}$, that contains all the measures that can reasonably be assumed to be the real measure P given the available data, is defined. Since the idea is to devise a general approach not restricted by assuming P to be an element of any specific parametric family, we define our ambiguity sets by the use of general probability metrics. Relative to these measures a worst case approach is adopted to robustify the problem with respect to $\mathscr{B}$.

The resulting optimization problems turn out to be infinite and are reduced to non-convex semi-definite problems. In the last part of the paper we show how to solve these problems numerically for the example of a mean risk portfolio selection problem with *Expected Shortfall under a Threshold* as the risk measure. The DCA in combination with an iterative algorithm to approximate the infinite set of constraints by finitely many ones is used to obtain numerical solutions to the problem.

ANWENDUNGEN NICHT KONVEXER OPTIMIERUNG IN PORTFOLIO-OPTIMIERUNGSPROBLEMEN

Die vorgelegte Arbeit befasst sich mit nicht-konvexer Optimierung in dem Gebiet der Portfolio Selection.

Thematisch lässt sich die Arbeit in zwei Teilgebiete strukturieren:

1. Das Lösen von Mean-Risk Problemen mit Value-at-Risk als Risikomaß: Es werden Methoden zum Auffinden von effizienten Portfolios für den Fall von diskret verteilten Asset Returns vorgestellt. Die behandelten Probleme sind (wegen der Nicht-Konvexität des Value-at-Risk) nicht konvex und lassen sich als Differenz von konvexen Funktionen darstellen. Es werden sowohl Branch-and-Bound als auch approximative Lösungsverfahren angewandt. Die globalen Lösungen des Branch-and-Bound werden mit den Lösungen der approximativen Verfahren verglichen.

2. Robustifizierung von Portfolio-Selection Problemen: In den letzten Jahren gibt es in der Literatur verstärkt Bemühungen Optimierungsprobleme bezüglich Unsicherheiten in den Parametern zu robustifizieren. Robustifizierte Lösungen haben die Eigenschaft, dass moderate Variationen von Parametern nicht zu dramatischen Verschlechterungen der Lösungen führen. Im Rahmen der robusten Portfolio Optimierung geht es hauptsächlich darum, Lösungen in Bezug auf Abweichungen in den Verteilungen der Gewinne der verwendeten Finanzinstrumente zu kontrollieren. In der gegenständlichen Arbeit werden mit Hilfe von Wahrscheinlichkeitsmetriken sogenannte Ambiguity Mengen definiert, welche alle Verteilungen enthalten, die aufgrund der Datenlage als mögliche Verteilungen in Frage kommen. Die verwendete Metrik, die sogenannte Kantorovich (Wasserstein) Metrik, ermöglicht es mittels Ergebnissen der nichtparametrischen Statistik, die Ambiguity Mengen als Konfidenzmengen um die empirischen Verteilungschätzer zu interpretieren. Mittels der beschriebenen Methoden werden Mean-Risk Probleme robustifiziert. Diese Probleme sind zunächst infinit und werden in einem weiteren Schritt zu nicht konvexen semi-definiten Problemen umformuliert. Die Lösung dieser Probleme basiert einerseits auf einem Algortihmus zum Lösen von semi-definiten Problemen mit unendlich vielen Nebenbedingungen und andererseits auf Methoden zum approximativen Lösen von nicht konvexen Problemen (dem sogenannten Difference of Convex Algorithm).

CURRICULUM VITAE OF DAVID WOZABAL

Personal Data

Name	David Wozabal (né Giczi)
Date of Birth	07.02.1979
Born in	Vienna, Austria

Education

- **1985-1989**: Primary School in Vienna
- **1989-1997**: High School at BRG 1060 Rahlgasse 4
- **1997-2001**: Masters of Business Informatics at the University of Vienna
- **2000-**: Masters of Mathematics at the University of Vienna
- **2004**: Civilian Service
- **2003-**: PhD in Statistics (Guide. Prof. Georg Pflug)

Academic Employment

- **2002-**: Lecturer at the Institute of Statistics and Operations Research (University of Vienna).
- **2003-2007**: AURORA Project (*Advanced Models, Applications and Software for High Performance Computing*, http://www.univie.ac.at/sor/aurora6/) funded by the Austrian *Fonds zur Förderung der wissenschaftlichen Forschung*.
- **2007**: Simulation based stochastic Optimisation Methods for Risk Management in Liberalized Energy Markets (`http://www.univie.ac.at/crm/simopt/`) funded by the Wiener Wissenschafts-, Forschungs- und Technologiefonds (Vienna Science and Technology Fund).
- **2007-**: Coupled Markov Chains for Credit Portfolios funded by the OeNB Jubiläumsfonds.

Publications and Conference Proceedings

1. G.Ch. Pflug and D.Wozabal. Ambiguity in portfolio selection. *Quantitative Finance*, 7(4):435–442, 2007.

2. R. Hochreiter, G. Ch. Pflug and D. Wozabal. Multi-stage stochastic electricity portfolio optimization in liberalized energy markets. *System Modeling and Optimization*. Volume 199 of Springer IFIP International Federation for Information Processing Series: 219-226. 2006.

3. R. Hochreiter, C. Wiesinger and D. Wozabal. Large-Scale Computational Finance Applications on the Open Grid Service Environment. *European Grid Conference 2005*. Volume 3470 of Springer Lecture Notes in Computer Science: 891-899. 2005.

4. C. Wiesinger, D. Giczi and R. Hochreiter. An open grid service environment for large-scale computational finance modeling systems. International Conference on Computational Science 2004. Volume 3036 of Springer Lecture Notes in Computer Science: 83-90. 2004.

Preprints & Technical Reports

1. David Wozabal. *A Framework for Optimization under Ambiguity*. Technical Report TR2008-08. Department of Statistics and Decision Support Systems, University of Vienna. August 2008.

2. David Wozabal, Nancy Wozabal. *Consistency of Risk Functionals*. Submitted to Journal of Non-Parametric Statistics.

3. D. Wozabal, R. Hochreiter, G. Pflug. A D.C. Formulation of Value-at-Risk constrained Optimization. Technical Report TR2008-01. Department of Statistics and Decision Support Systems, University of Vienna. January 2008. Available via `http://www.optimization-online.org`.

4. D. Wozabal. A new method for Value-at-Risk constrained optimization using the Difference of Convex Algorithm (DCA). Technical Report TR2008-03, Department of Statistics and Decision Support Systems, University of Vienna. January 2008. Available via `http://www.optimization-online.org`

Working Papers

1. Ronald Hochreiter, David Wozabal. *A multi-stage stochastic programming framework for managing risk-optimal electricity portfolios*.

2. Ronald Hochreiter, Georg Pflug, David Wozabal. *A Coupled Markov Chain Approach to CDX Pricing*.

Talks at Scientific Conferences

1. *Market Risk Control of Structured Credit Products*. International Conference on Price, Liquidity, and Credit Risks. Konstanz, Oktober 2008.

2. *A New Method for Value-at-Risk constrained Optimization using the Difference of Convex Algorithm*. CARIPLO Workshop. Edinburgh, September 2008.

3. *New Methods for Value-at-Risk Constrained Optimization using Difference of Convex (D.C.) Programming*. International Symposium on Business and Industrial Statistics. Prague, July 2008. Invited Talk.

4. *A D.C. Formulation of Value-at-Risk Constrained Optimization. Colloquium of the Department of Statistics and Operations Research*. Vienna, June 2008.

5. *A D.C. Formulation of Value-at-Risk Constrained Optimization. International Conference on Applied Mathematical Programming and Modeling*. APMOD 2008. Bratislava, May 2008.

6. *Non-Parametric Model Uncertainty in Stochastic Programming*. 11 Conference on Stochastic Programming. Vienna, August 2007.

7. *Optimization under Ambiguity*. 22 European Conference on Operations Research. Prague, July 2007.

8. *Ambiguity in Portfolio Selection*. 21 European Conference on Operations Research. Reykjavik, July 2006.

9. *A Time Series Approach to Scenario Generation*. EU-Workshop Series on Mathematical Optimization Models for Financial Institutions: Asset and Liability Management for Financial Institutions. Ayia Napa, November 2003.