Open access • Journal Article • DOI:10.1007/S10479-010-0812-0

A framework for optimization under ambiguity — Source link

David Wozabal

Institutions: University of Vienna

Published on: 01 Mar 2012 - Annals of Operations Research (Springer US)

Topics: Stochastic optimization, Stochastic programming, Optimization problem, Robust optimization and Semidefinite programming

A Framework for Optimization under Ambiguity

D. Wozabal
Department of Statistics and Decision Support Systems
University of Vienna
Universitätsstraße 5/9, A-1010 Vienna, Austria
david.wozabal@univie.ac.at

November 3, 2008

Abstract

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.

Keywords: robust optimization; portfolio management; difference of convex algorithm; semi definite programming; expected shortfall; non-convex optimization; extreme points
Classcode: 46A55; 62C20; 62G35; 65K05

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 [28] to a more general setting.

We consider the following stochastic programming problem

$$\begin{array}{rll} \sup_{x\in\mathbb{R}^m} & F(x,P) & \\ s.t. & G(x,P) & \leq 0 \\ & H(x) & \leq 0, \end{array} \tag{1}$$

where P is an element of $\mathcal{P}(\mathbb{R}^d)$ the set of probability measures on $\mathbb{R}^d$, $F : \mathbb{R}^m \times \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}$ is concave in the first variable and linear in the probability measure P, the function $G : \mathbb{R}^m \times \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}$ 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 [25] or [11]). Therefore it is reasonable to consider robust versions of problem (1), i.e. to introduce a so called ambiguity set $\mathcal{B} \subseteq \mathcal{P}(\mathbb{R}^d)$ 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 $\mathcal{B}$.

Given $\mathcal{B}$, the robust counterpart of (1) that is considered looks like

$$\begin{array}{rlll} \sup_{x\in\mathbb{R}^m} \inf_{Q\in\mathcal{B}} & F(x,Q) & & \\ s.t. & G(x,Q) & \leq 0, & \forall Q \in \mathcal{B} \\ & H(x) & \leq 0. & \end{array} \tag{2}$$

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 $\mathcal{B}$ which usually is relatively small in comparison to $\mathcal{P}(\mathbb{R}^d)$, and the worst case is relative to this set.

Like in [28] we construct the ambiguity sets as Kantorovich neighborhoods of the empirical measure $\hat{P}_n$ (constructed from a sample of size n), where

$$\hat{P}_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{\xi_i}. \tag{3}$$

In specific, we set

$$\mathcal{B} = \mathcal{B}_\epsilon(\hat{P}) = \{Q \in \mathcal{P}(\mathbb{R}^d) : d_K^r(\hat{P}, Q) \leq \epsilon\} \tag{4}$$

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

$$d_K^r(P_1, P_2) \quad = \quad \inf_{\gamma(A\times\Omega)=P_1(A),\ \gamma(\Omega\times A)=P_2(A)} \left(\int ||x-y||_1^r d\gamma(x,y) \right)^{\frac{1}{r}}, \tag{5}$$

where $\|u-v\|_1 = \sum_i |u_i - v_i|$ (see [21] for a short introduction to the subject of probability metrics and [29] for an extensive review). Here, we will work with $r = 1$. The space $(\mathcal{P}(C), d_K^1)$ is a compact, separable, complete metric space for a compact set $C \subseteq \mathbb{R}^d$. The topology on $(\mathcal{P}(C), d_K^1)$ is the weak (star) topology induced by the continuous functions on C.

In the terminology of [6] 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 (2) with $\mathcal{B} = \mathcal{B}_\epsilon(\hat{P})$ without further assumptions. The formulation (2) suffers from the drawback, that it is numerically intractable for the following three reasons:

i. The problem (2) has infinitely many constraints.

ii. The solutions of the problems (2) are elements of an infinite dimensional space (together with (i) this makes the problem infinite).

iii. The problem (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 [28]. 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 2 we give an overview of papers that deal with problems in similar settings, Section 3 is concerned with the reformulation of problem (2) in such a way that it becomes numerically tractable, in Section 4 the framework is applied to a concrete robust portfolio composition problem and in Section 5 numerical results for this particular problem are discussed. Section 6 concludes the paper.

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 [15, 16, 17, 39]) and more recent works include [9, 10, 19, 18, 22, 26, 28, 33, 34, 37]. 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 $\mathcal{B}$ to keep problem (2) computationally tractable. In [22, 37] for example a classical mean-risk portfolio composition problem is studied and it is assumed that the set $\mathcal{B}$ 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, 20, 31]).

In [33] the authors assume the ambiguity set to be of the form

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

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 (2) can be solved efficiently.

In [6, 7] 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 [9] the ambiguity sets $\mathcal{B}$ 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 $\mathcal{B}$ 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 $P(x_1) = \epsilon$, $P(x_2) = 1 - \epsilon$ and $Q(x_1) = 1 - \epsilon$, $Q(x_2) = \epsilon$, then

$$d_{KL}(P, Q) = 2\epsilon \log \frac{\epsilon}{1-\epsilon} + \log \frac{1-\epsilon}{\epsilon} \overset{\epsilon \to 0}{\longrightarrow} \infty. \tag{6}$$

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 [28] the ideas on the construction of $\mathcal{B}$ are similar to the approach presented in this paper. However, when solving problem (2) only a small fraction of the measures in $\mathcal{B}_\epsilon(\hat{P}_n)$ 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 (2) becomes much simpler, since $\mathcal{B}$ can be described by a subset of a finite dimensional space. However, we will show in section 5, considering all distributions in $\mathcal{B}_\epsilon(\hat{P}_n)$ 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 $\mathcal{B}_\epsilon(\hat{P}_n)$ 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 $\mathcal{B}_\epsilon(\hat{P}_n)$ and this reduction should be achieved without sacrificing on our goal to solve problem (2) for the whole of $P \in \mathcal{B}_\epsilon(\hat{P}_n)$.

Summarizing, our approach is different from the aforementioned in the sense, that we do not restrict the set $\mathcal{B}$ 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 $\mathcal{B}$ as *confidence balls* around the empirical measure $\hat{P}_n$.

3 Reducing the problem

As a first step necessary to achieve numerical tractability of problem (2) we reduce the set of possible distributions from the full Kantorovich ball to a subset $\mathcal{B}'$, whose elements can be described by finite dimensional vectors. In particular $\mathcal{B}'$ should fulfill

$$\inf_{Q \in \mathcal{B}} F(x, Q) = \inf_{Q \in \mathcal{B}'} F(x, Q). \tag{7}$$

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

$$\sup_{Q\in\mathcal{B}} G(x,Q) = \sup_{Q\in\mathcal{B}'} G(x,Q) \tag{8}$$

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 $(\mathcal{P}, d_K^1)$), i.e.

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

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

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

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

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 (7). Hence, instead of trying to approximate every element of $\mathcal{B}_\epsilon(\hat{P}_n)$, we only consider the extremal elements of $\mathcal{B}_\epsilon(\hat{P}_n)$.

Definition 3.1 (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 3.1 (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 [12], Theorem 25.9. □

Remark: The extremals of the set $\mathcal{P}(\mathbb{R}^d)$ are the Dirac measures δ_x with $x \in \mathbb{R}^d$. To see this fix a measure $P \in \mathcal{P}(\mathbb{R}^d)$ 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 $\mathcal{P}(\mathbb{R}^d)$.

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

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

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

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

$$F(\delta_x) \leq F(\delta_{x_1}) < F(\delta_{x_1} p + \delta_{x_2}(1-p)), \quad \forall 0 < p < 1 \tag{9}$$

Since the set $argmax_{P \in \mathcal{B}_\epsilon(\delta_{x_0})} F(P)$ has to contain an extremal point, (9) shows that there are extremal points of $\mathcal{B}_\epsilon(\delta_{x_0})$, 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 3.2 (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 P of a convex set C in the space $(\mathcal{P}(\mathbb{R}^d), d_K^1)$ it is therefore possible to find a continuous bounded function $g : \mathbb{R}^d \to \mathbb{R}$ such that

$$\{P\} = argmax_{Q \in C} \int g dQ. \tag{10}$$

The next example demonstrates that Theorem 3.1 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 $B = \left\{(x, y) \in \mathbb{R}^2 : y \leq 1\right\}$ 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

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

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 [32] to characterize the distributions that are exposed points of $\mathcal{B}$.

Theorem 3.2. *Let $f_1, \ldots, f_v$ 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 $|\mu'| \leq v+1$ and*

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

Theorem 3.3. *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 $\mathcal{B}_\epsilon(\hat{P}_n)$ with radius $\epsilon > 0$ around $\hat{P}_n$. Then all the exposed points of $\mathcal{B}_\epsilon(\hat{P}_n)$ are in $\mathcal{P}_{(n+3)}$ – the discrete distributions with at most $(n+3)$ atoms.*

Proof. In view of the above remarks, to show that all the exposed points of $\mathcal{B}_\epsilon(\hat{P}_n)$ are discrete, it is enough to show that for any measure $P \in \partial\mathcal{B}_\epsilon(\hat{P}_n)$ 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 $d_K(\hat{P}_n, \tilde{P}) \leq \epsilon$. This would establish that there is no argmax-set like (10) that consists only of a single non-discrete measure, i.e. all the exposed points have to be discrete measures.

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

$$\begin{aligned}
\int ||x-y||_1 d\gamma(x,y) &= \epsilon && (11)\\
\int g(y) d\gamma(x,y) &= c && (12)\\
\gamma(\mathbb{R}^d \times A) &= P(A), \qquad \forall A \subseteq \mathbb{R}^d \text{ measurable} &&\\
\gamma(A \times \mathbb{R}^d) &= \hat{P}_n(A), \qquad \forall A \subseteq \mathbb{R}^d \text{ measurable.} &&
\end{aligned}$$

Now invoke Theorem 3.2 for γ with conditions (11) and (12). The functions f_i are: $f_1(x,y) = ||x-y||_1$, $f_2(x,y) = g(y)$ and additionally

$$f_{i+2}(x,y) = p_i \, \mathbb{1}_{\{x_i\} \times \mathbb{R}^d}(x,y), \qquad \text{for } i = 1, \ldots, n. \tag{13}$$

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 (13)). Call the second marginal $\tilde{P}$. This together with the moment conditions (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 (11) has to be replaced by

$$\int ||x - y||_1^r d\gamma(x, y).$$

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 [13], Section 17) to extend our result to the extreme points.

Theorem 3.4 (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 3.1. *The extremal points of $\mathcal{B}$ are discrete measures with at most $(n + 3)$ points.*

Proof. Since the $exp(\mathcal{B}) \subseteq \mathcal{P}_{n+3}$ and $\overline{exp(\mathcal{B})} = ext(\mathcal{B})$, we know that for every $P \in ext(\mathcal{B})$ $\exists (P_r)_{r\in\mathbb{N}} \in exp(\mathcal{B})$, 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 $P_r \in \mathcal{P}_{(n+3)}$. □

Having characterized the extreme points of $\mathcal{B}_\epsilon(\hat{P})$ as discrete distribution with a fixed number of atoms, the robustified problem (2) can be reformulated as a semi definite problem using (7) and (8). To establish the existence of a saddle-points in (2) we use the following classical minimax theorem (see for example [35] or [38]).

Theorem 3.5 (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 quasiconcave in x and quasiconvex in y. If F is upper (or lower) semi continuous in x in every line segment and lower semi continuous in y, while D is compact then the function $F(x, y)$ possesses a saddle-value on $C \times D$ and*

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

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

Theorem 3.6. *Let*

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

then a solution to problem (2) *exists and coincides with the solution of the following reduced problem*

$$\begin{array}{lll} \sup_{x\in\mathbb{R}^m} \inf_{P\in\mathcal{B}\cap\mathcal{P}_{(n+3)}} & F(x,P) & \\ s.t. & G(x,P) & \leq 0, \quad \forall P \in \mathcal{B}_\epsilon(\hat{P}_n)\cap\mathcal{P}_{(n+3)} \\ & H(x) \leq 0 & \end{array} \tag{14}$$

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

Proof. It only remains to show that a saddle point of the problem (2) exists – from which the existence of a saddle point for (14) follows. Define $D = \mathcal{B}_\epsilon(\hat{P}_n)$ and note that since the Kantorovich ball $\mathcal{B}_\epsilon(\hat{P}_n)$ is a closed subset of the probability measures, it is a compact set in $(\mathcal{P}, d_K^1)$. Further define

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

C is closed since $G(\cdot, P)$ and $H(\cdot)$ are convex and therefore continuous. Therefore the conditions of Theorem 3.5 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 3.1 and Theorem 3.1. □

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 (2) one has to describe the set of distributions in $\mathcal{B}_\epsilon(\hat{P}_n)\cap\mathcal{P}_{(n+3)}$. The measures $Q \in \mathcal{B}_\epsilon(\hat{P}_n)\cap\mathcal{P}_{(n+3)}$ 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

$$\begin{array}{rll} \sum_{j=1}^{n+3} q_j & = 1 & \\ \sum_{j=1}^{n+3} t_{i,j} & = p_i, & \forall i = 1,\ldots,n \\ \sum_{i=1}^{n} t_{i,j} & = q_j, & \forall j = 1,\ldots,(n+3) \\ \sum_{i=1}^{n}\sum_{j=1}^{n+3} ||x_i - y_j||_1 t_{i,j} & \leq \epsilon & \\ t_{i,j} \geq 0, q_j & \geq 0 & \end{array} \tag{15}$$

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 (5) is – in the discrete setting – equivalent to solving the optimal mass transportation problem (see [30]). This fact makes it easy to handle the Kantorovich distance for discrete distributions.

However, the conditions in (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 (15). In the following numerical solution techniques for specific instances of problem (2) will be discussed.

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 (15) numerically are discussed.

The case we want to treat here is a variation of the mean-risk problem described in [28]. The 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

$$\begin{array}{lll} \max_{w \in \mathbb{R}^d} & \mathbb{E}(w^\top X^{\hat{P}_n}) & \\ s.t. & ES_a(w^\top X^{\hat{P}_n}) & \leq R \\ & \sum w_i & = 1 \\ & w & \geq 0 \end{array} \tag{16}$$

where $X^Q = (\xi_1^Q, \ldots, \xi_d^Q)$ 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 $ES_a : (w, Q) \mapsto ES_a(w^\top X^Q)$ is concave in the first and linear in the second component (see [27]).

The robust counterpart looks like

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

If we replace the operators $\mathbb{E}$ and ES_a by their explicit versions for discrete distributions and apply Theorem 3.3 (note that the required continuity properties are fulfilled for $\mathbb{E}$ and ES_a) the above problem is equivalent to

$$\begin{array}{rlll} \max_{w\in\mathbb{R}^d} \min & \sum_{j=1}^{n+3}(w^\top y_j)q_j & & \\ s.t. & \sum_{j=1}^{n+3} q_j & = 1 & \\ & \sum_{j=1}^{n+3} t_{i,j} & = p_i, & \forall i = 1,\dots,n \\ & \sum_{i=1}^{n} t_{i,j} & = q_j, & \forall j = 1,\dots,(n+3) \\ & \sum_{i=1}^{n}\sum_{j=1}^{n+3} ||x_i - y_j||_1 t_{i,j} & \leq \epsilon & \\ & \sum_{j=1}^{n+3} max(a - w^\top y_j, 0)q_j & \leq R, & \end{array} \tag{17}$$

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, $\epsilon > 0$ 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 (17) by the following iterative algorithm originally presented in [28].

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

2. Solve the outer problem

$$\begin{array}{rlll} \max_{(w,t)} & t & & \\ s.t. & \mathbb{E}(w^\top X^Q) & \leq t, & \forall Q \in \mathcal{Q}_i \\ & ES_a(w^\top X^Q) & \leq R, & \forall Q \in \mathcal{Q}_i \\ & \sum_{i=1}^{d} w_i & = 1, \quad w \geq 0 & \end{array} \tag{18}$$

   and call the solution (w_i, t_i).

3. Solve the problem

$$\min_{Q\in\mathcal{B}_\epsilon(\hat{P}_n)\cap\mathcal{P}_{n+3}} \mathbb{E}(w_i^\top X^Q) \tag{19}$$

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

4. Solve the problem

$$\max_{Q\in\mathcal{B}_\epsilon(\hat{P}_n)\cap\mathcal{P}_{n+3}} ES_a(w_i^\top X^Q) \tag{20}$$

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

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

6. If

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

(b) the optimal value of (19) equals t_i and the solution of (20) is equal to $\min_{P\in\mathcal{Q}_i} ES_a(w_i^\top X^P)$

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 (18). The measures Q_n are chosen to be the worst case with respect to the current portfolio decision w_i and therefore represent the extremal elements of $\mathcal{B}_\epsilon(\hat{P}_n)$ which are required to approximate the relevant parts of $\mathcal{B}_\epsilon(\hat{P}_n)$. 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 [28]).

The non-convexity of problem (17) is reflected in the non-convexity of problems (19) and (20). In the remainder of this section we will discuss solution techniques for these problems.

4.1 Minimizing Expectation

Problem (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 k^* when altered leads to the maximum change in expectation, that is $k^* = \arg\max_k w_k$ and therefore

$$\frac{\partial \mathbb{E}(w^\top X)}{\partial w_{k^*}} = \max_{1\le k\le d} \frac{\partial \mathbb{E}(w^\top X)}{\partial w_k}. \tag{21}$$

The aim is – starting from the empirical distribution – to obtain a distribution which minimizes the expectation over the set of discrete distributions $\mathcal{B}_\epsilon(\hat{P}_n)\cap\mathcal{P}_{n+3}$. Clearly, from (21) the optimal way to achieve this goal is to start altering scenarios in the k^*-th component. Therefore the following algorithm to find the optimum in (19) is proposed. A constant c is used as a lower bound on the return of asset k^*. 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 $k^* = \arg\max_{1\le k\le d} w_k$.
3. Set $j = 1$.
4. Pick any atom $y_i^{(j-1)}$ with $1 \le i \le n$ and change it's k^*-th component to c and call the resulting measure Q.
5. Solve the linear programming problem

$$\begin{array}{ll} \max & \lambda \\ s.t. & \lambda Q + (1-\lambda)P^{(j-1)} \in \mathcal{B}_\epsilon(\hat{P}_n)\cap\mathcal{P}_{n+3} \end{array} \tag{22}$$

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 4.1). 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 i 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 4.1. *The algorithm presented above leads to an optimal solution of problem (19).*

Proof. Let P be in $\mathcal{B}_\epsilon(\hat{P}_n) \cap \mathcal{P}_{n+3}$ and let $(t_{i,j})$ be the optimal transportation plan that transports mass from $\hat{P}_n$ to P (i.e. $t_{i,j}$ fulfill the set of constraints (15)). The difference of the expectation under $\hat{P}_n$ and P can be bounded as follows

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

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

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

Therefore the measure P' is optimal. □

Remark: If the algorithm does not terminate, i.e. if $\epsilon > 0$ 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.

4.2 Maximizing Expected Shortfall

Define the index sets $I = \{1, \ldots, n\}$, $J = \{1, \ldots, n+3\}$ and $M = \{1, \ldots, d\}$. The inner problem (20) can be written as

$$\begin{array}{lllll} \max & \sum_{j \in J} max(a - w^\top y_j, 0) q_j & & & \\ s.t. & \sum_{i \in I} t_{i,j} & = q_j, & \forall j \in J & \\ & \sum_{j \in J} t_{i,j} & = p_i, & \forall i \in I & \\ & x_{i,k} - y_{j,k} & \leq d_{i,j}^k, & \forall (i,j,k) \in I \times J \times M & \\ & y_{j,k} - x_{i,k} & \leq d_{i,j}^k, & \forall (i,j) \in I \times J & \\ & \sum_{k \in M} d_{i,j}^k & = d_{i,j}, & \forall (i,j) \in I \times J & \\ & \sum_{(i,j) \in I \times J} d_{i,j} t_{i,j} & \leq \epsilon & & \\ & t_{i,j}, q_j & \geq 0. & & \end{array} \tag{23}$$

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 (23), we rewrite it as a D.C. program (see for example [23] or [24]). Towards this we model the objective in (23) as follows: introduce the splitting variables

$$a - w^\top y_j = z_j^+ + z_j^-, \quad z_j^+ z_j^- \geq 0, \quad z_j^+ \geq 0. \tag{24}$$

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 (23) in (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 (23) as the following non-convex quadratic problem.

$$\begin{array}{lllll} \min & t & & & \\ s.t. & \sum_{j\in J}(\beta_j^2 - \alpha_j^2) & \leq t & & \\ & \sum_{i\in I} t_{i,j} & = q_j, & \forall j \in J & \\ & \sum_{j\in J} t_{i,j} & = p_i, & \forall i \in I & \\ & x_{i,k} - y_j^k & \leq d_{i,j}^k, & \forall (i,j,k) \in I \times J \times M & \\ & y_j^k - x_{i,k} & \leq d_{i,j}^k, & \forall (i,j) \in I \times J & \\ & \sum_{k\in M} d_{i,j}^k & = d_{i,j}, & \forall (i,j) \in I \times J & \\ & \sum_{(i,j)\in I\times J}(\gamma_{i,j}^2 - \delta_{i,j}^2) & \leq \epsilon & & \\ & \sum_{j\in J}(\phi_j^2 - \chi_j^2) & \geq 0 & & \\ & a - w^\top y_j & = z_j^+ + z_j^- & & \\ & t_{i,j}, z_j^+, q_j & \geq 0. & & \end{array} \tag{25}$$

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 (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 (25) then becomes

$$\begin{array}{lll} \min_{y,q,d,t,z^+,z^-} & t & \\ s.t. & \sum_{j\in J}(\beta_j^2 - \alpha_j^2) - t & \leq 0 \\ & \sum_{(i,j)\in I\times J}(\gamma_{i,j}^2 - \delta_{i,j}^2) & \leq \epsilon \\ & \sum_j(\phi_j^2 - \chi_j^2) & \geq 0, \\ & l(\bar{x}) & \leq 0. \end{array} \tag{26}$$

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 various situations (see for example [36, 2, 3, 5, 14]) and its low computational complexity makes it possible to work with large scale problems.

To apply the DCA we need to write (26) as an unconstrained D.C. problem of the form

$$\inf\{f(\bar{x}) = g(\bar{x}) - h(\bar{x}) : \bar{x} \in \mathbb{R}^N\} \tag{27}$$

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

The DCA algorithm works for problems of the form (27) by repeatedly solving the two convex optimization problems

$$\inf_{\bar{x}\in\mathbb{R}^N}\{g(\bar{x}) - (h(\bar{x}^k) + \langle \bar{x} - \bar{x}^k, \bar{y}^k\rangle)\} \tag{28}$$

and

$$\inf_{\bar{y}\in\mathbb{R}^N}\{h^*(\bar{y}) - (g^*(\bar{y}^{k-1}) + \langle \bar{x}^k, \bar{y} - \bar{y}^{k-1}\rangle)\} \tag{29}$$

and iteratively produces candidates $\bar{x}^{k+1}$, $\bar{y}^k$ 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 (28) and the dual problem (29) respectively. For a more detailed discussion of DCA algorithm and its properties see [4].

We use Corollary 2.4.1. from [8] to reformulate (26) by an exact penalty.

Theorem 4.2. *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*

$$\begin{array}{lll} \inf & f_0(x) & \\ s.t. & G(x) & \in C \\ & x & \in S, \end{array} \tag{30}$$

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 (30) and

$$\inf_{x\in X} f_0(x) + \psi(x|S) + \tau dist(G(x)|C) \tag{31}$$

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

$$0 \in int\left\{G(\bar{x}^*) + G'(\bar{x}^*)(S - \bar{x}^*) - C\right\} \tag{32}$$

is fulfilled at optimal points $\bar{x}^*$ of the system (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 $G(\bar{x}) = (0, \epsilon, 0)$. The three constraints are analyzed separately to find a common point $s \in S$ which fulfils (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

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

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

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

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

$$\sum d_{i,j}\bar{t}_{i,j}^* < \epsilon. \tag{33}$$

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 (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

$$\min_{y \in \mathbb{R}^d} \sum_{i \in N(j)} ||x_i - y||_1 \bar{t}_{i,j}^*. \tag{34}$$

Note that $\bar{y}_j^*$ can not be a solution to (34), since by optimality $\bar{y}_j^{k*} \leq \min_{i \in N(j)} x_i^k$ for all $1 \leq k \leq d$. This is the case, since if $\bar{y}_j^{k*} > x_i^k$ 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

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

instead of $\bar{y}_j^{k*}$. 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

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

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

$$\sum \bar{z}_j^{-*} z_j^+ + \sum \bar{z}_j^{+*} z_j^- < 0. \tag{35}$$

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 (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 t is unrestricted and can be chosen such that the corresponding condition is fulfilled. □

Theorem 4.3. *For some $\tau > 0$ the problem* (26) *can be equivalently reformulated to the following problem*

$$\begin{aligned} min_{\bar{x} \in \mathbb{R}^N} \quad & t + \tau \max \left\{ \sum_{j \in J} \beta_j^2 + \sum_{(i,j) \in I \times J} \delta_{i,j}^2 + \sum_{j \in J} \phi_j^2 - t, \sum_{(i,j) \in I \times J} \gamma_{i,j}^2 - \epsilon + \sum_{j \in J} \alpha_j^2 + \sum_{j \in J} \phi_j^2, \right. \\ & \left. \sum_{j \in J} \chi_j^2 + \sum_{(i,j) \in I \times J} \delta_{i,j}^2 + \sum_{j \in J} \alpha_j^2, \sum_{(i,j) \in I \times J} \delta_{i,j}^2 + \sum_{j \in J} \alpha_j^2 + \sum_{j \in J} \phi_j^2 \right\} \\ - \quad & \tau \left(\sum_{i,j} \delta_{i,j}^2 + \sum_j \alpha_j^2 + \sum_j \phi_j^2 \right) + \psi(x|S). \end{aligned} \tag{36}$$

where $\psi(\cdot|S)$ is the convex indicator function of the set S. In other words (26) *can be reformulated to a D.C. problem of the form* (27) *with*

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

Proof. From Lemma 4.1 we get that Theorem 4.2 is applicable and therefore problem (26) can be reformulated to a unconstrained problem of the form (31) with S, C and G as defined in Lemma 4.1. To reduce the problem to an unconstrained D.C. problem as in (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

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

which finally yields (36). □

Solving problem (28) From Theorem 4.3 it follows that for (26) the corresponding DCA sub-problem (28) can be written as the quadratically constrained problem

$$\begin{array}{lll} \min & \tau L - \langle \bar{x}, \bar{y}^k \rangle & \\ s.t. & \sum_j \beta_j^2 - t + \sum_{i,j} \delta_{i,j}^2 + \sum_j \phi_j^2 & \le L \\ & \sum_{i,j} \gamma_{i,j}^2 - \epsilon + \sum_j \alpha_j^2 + \sum_j \phi_j^2 & \le L \\ & \sum_j \chi_j^2 + \sum_j \alpha_j^2 + \sum_{i,j} \delta_{i,j}^2 & \le L \\ & \sum_{i,j} \delta_{i,j}^2 + \sum_j \alpha_j^2 + \sum_j \phi_j^2 & \le L \\ & l(\bar{x}) & \le 0. \end{array} \tag{37}$$

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 (29) To solve the DCA sub-problem (29), we first have to compute the conjugate h^* for

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

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 $\bar{y} = (t, (q_j), (y_j), (z_j^+), (z_j^-), (t_{i,j}), (d_{i,j}), (d_{i,j}^k))$. (29) therefore becomes

$$\begin{array}{lrll} \min & \sum_{i,j} \frac{q_{i,j}^2}{\tau} + \sum_{i,j} \frac{d_{i,j}^2}{\tau} + \sum_j \frac{(z_j^-)^2}{\tau} - \langle \bar{x}^k, \bar{y} \rangle & & \\ s.t. & t & = -1 & \\ & d_{i,j} & = -t_{i,j}, & \forall (i,j) \in I \times J \\ & z_j^+ & = z_j^- + q_j, & \forall j \in J \\ & y_j & = 0, & \forall j \in J \\ & d_{i,j}^k & = 0, & \forall (i,j,k) \in I \times J \times M. \end{array} \tag{38}$$

The solution to problem (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.

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 3.6. Such risk measures have to be convex in the portfolio weights w as well as linear in the probability measure P. One can consider for example risk functionals corresponding to some expected utility functionals (see [27]).

5 Numerical Results

In this section numerical results for problem (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 1).

All numerical experiments are carried out for the robustifications described in this paper as well as for the techniques in [28], where only the probability mass is shifted between the points of the empirical distribution. We henceforth call the problem described in Section 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.

Table 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

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 $\epsilon = 0.005$ to $\epsilon = 0.03$ in steps of 0.005. Furthermore we choose the risk parameter $R = 0.007$.

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 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 $\mathcal{B}_\epsilon(\hat{P}_n)$, i.e. the measures obtained from the inner problems of the full robustification

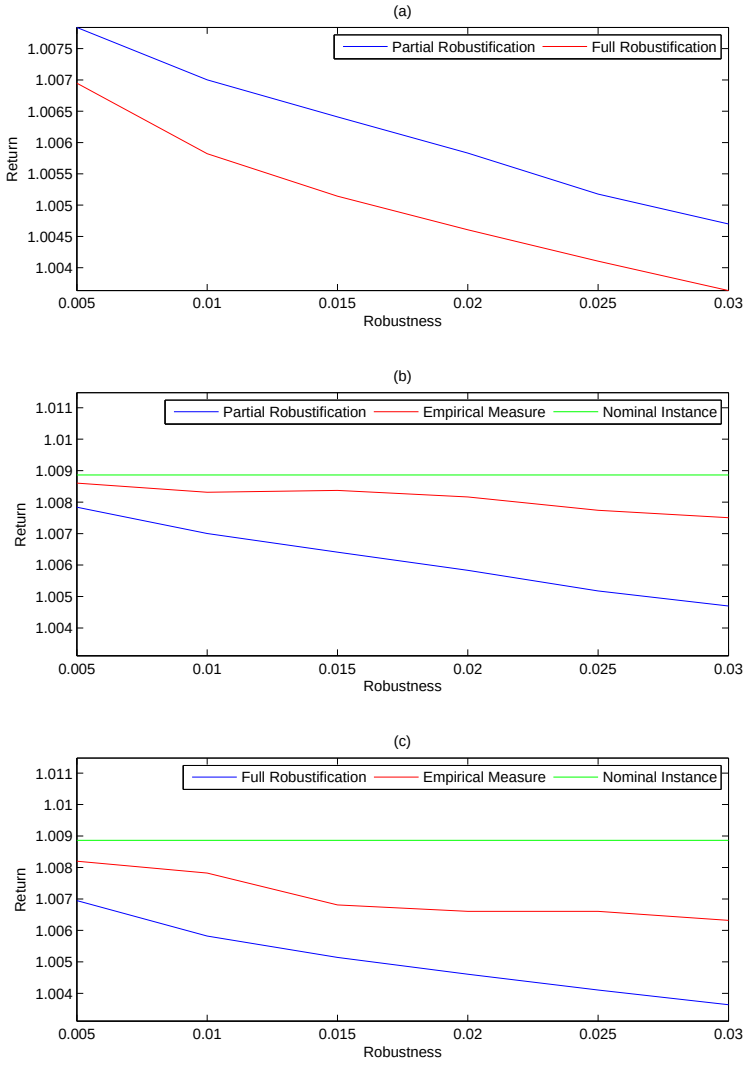
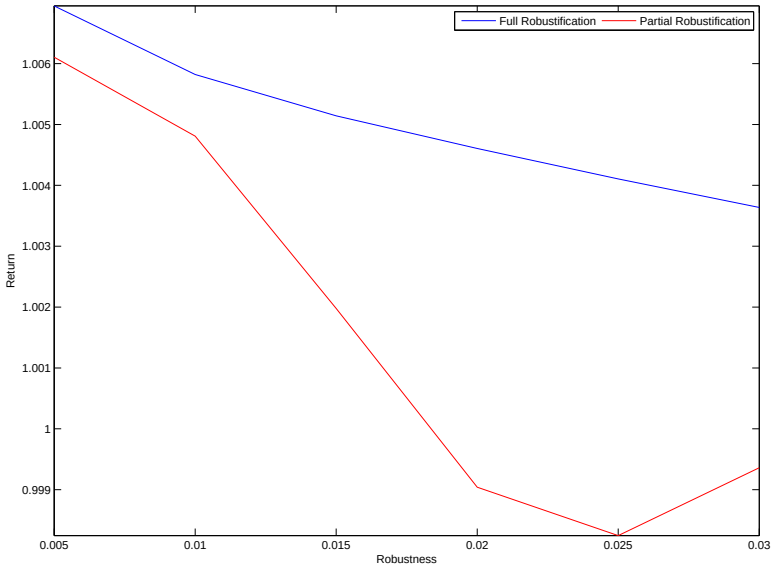
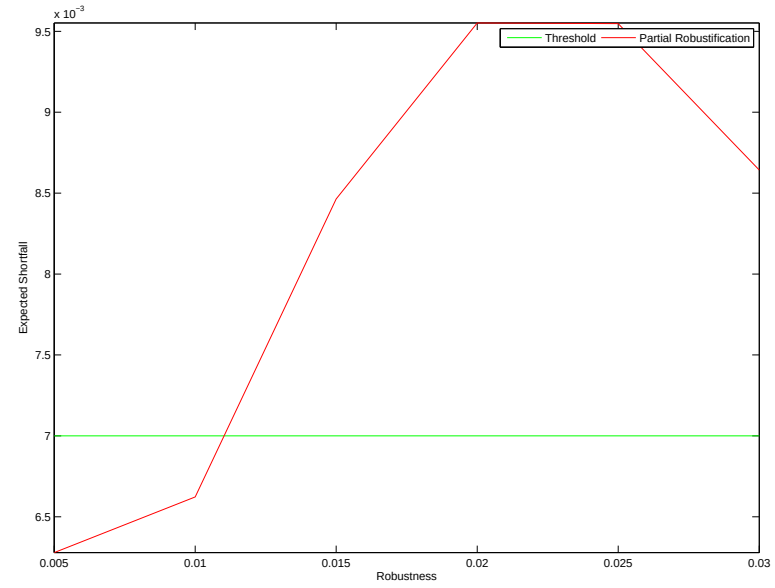


Figure 1: Figure (a) shows the expectations of robust portfolio from the two robustification techniques as the robustness parameter increases. Figure (b) and (c) show the comparisons of the respective robust portfolios with the expectation of the nominal program instance and the expectation of the robust portfolios under the empirical measure.



(a)



(b)

Figure 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.

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 $P_\epsilon^{\mathbb{E}}$ and $P_\epsilon^{ES_a}$, which give the smallest expected return and the highest Expected Shortfall for the portfolios w_ϵ. In Figure 2 (a) the expected returns of $w_\epsilon^\top \xi$ under $P_\epsilon^{\mathbb{E}}$ are compared with the respective solutions of the nominal instance given in (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 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.

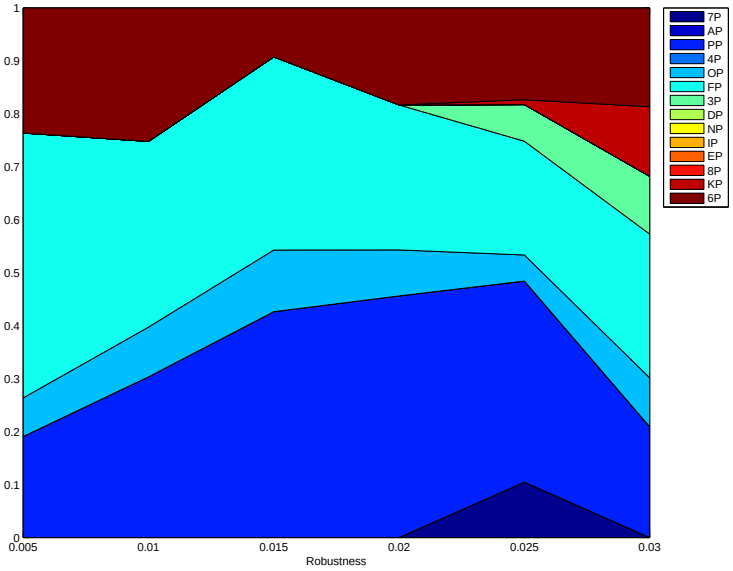
Summarizing, the Figures 1 and 2 show that the freedom to deviate from the atoms of the empirical measure actually results in portfolios, which yield lower returns in the worst case as well as under the empirical measure. However, these portfolios are clearly more robust than the measures obtained by shifting only the probabilities as Figure 2 demonstrates.

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

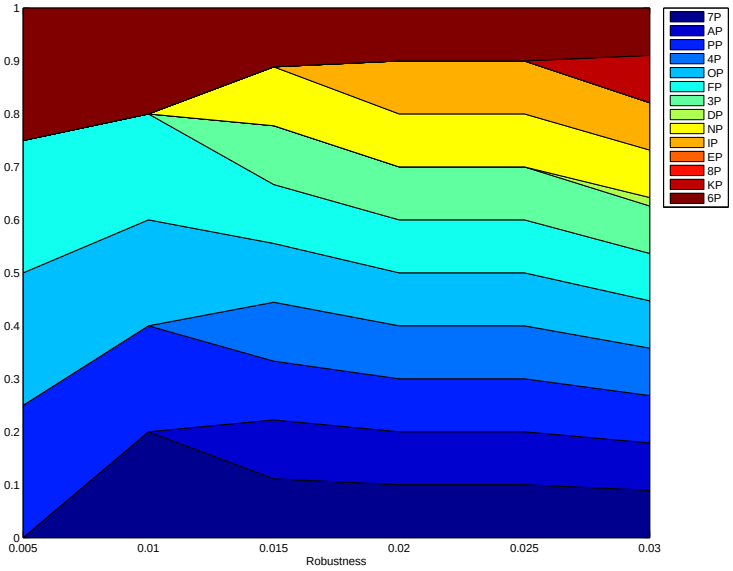
To conclude the comparison between the two methods it is investigated what effect the number of data points has on relation between partial and full robustification. It seems plausible that with a bigger number of data points the additional robustness gained by shifting the points is smaller than with 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 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 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.

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



(a)



(b)

Figure 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. .

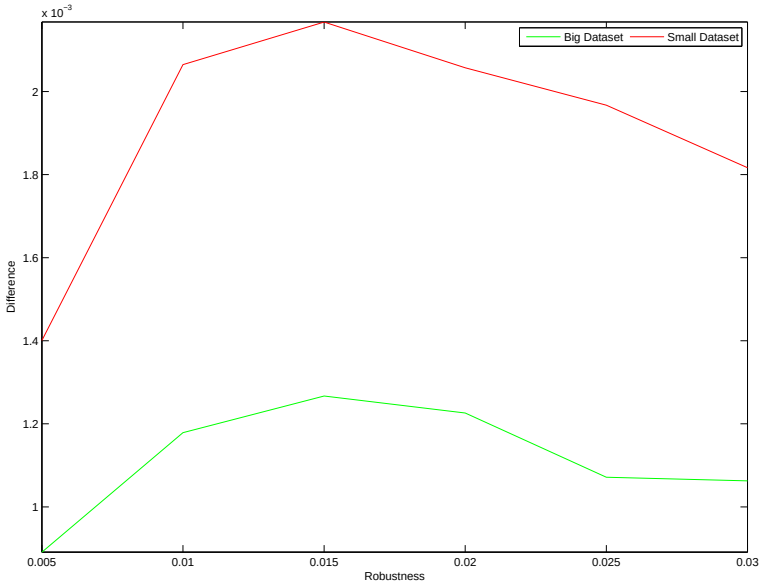


Figure 4: Differences in expected return between partial and full robustification for the whole data set from 1 and a subset of 26 points.

are reformulated to semi-definite non-convex optimization problems.

In Section 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 [28] 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.

Acknowledgements

The author wants to express his thanks to Prof. Georg Ch. Pflug for his valuable input throughout the work on this paper.

This research was partly funded by the Vienna Science and Technology Fund (Grant: MA06) and the Austrian National Bank (Grant: 12306).

References

[1] J. Affleck-Graves and B. McDonald. Nonnormalities and tests of asset pricing theories. *The Journal of Finance*, 44(4):889–908, 1989.

[2] Le Thi Hoai An and Pham Dinh Tao. A continuous approach for globally solving linearly constrained quadratic zero-one programming problems. *Optimization*, 50(1-2):93–120, 2001.

[3] Le Thi Hoai An and Pham Dinh Tao. A new algorithm for solving large scale molecular distance geometry problems. In *High performance algorithms and software for nonlinear optimization (Erice, 2001)*, volume 82 of *Appl. Optim.*, pages 285–302. Kluwer Acad. Publ., Norwell, MA, 2003.

[4] Le Thi Hoai An and Pham Dinh Tao. 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.

[5] Le Thi Hoai An, Pham Dinh Tao, and Le Dung Muu. A combined D. C. optimization – ellipsoidal branch-and-bound algorithm for solving nonconvex quadratic programming problems. *J. Comb. Optim.*, 2(1):9–28, 1998.

[6] A. Ben-Tal and A. Nemirovski. Robust convex optimization. *Math. Oper. Res.*, 23(4):769–805, 1998.

[7] A. Ben-Tal and A. Nemirovski. Robust solutions of uncertain linear programs. *Oper. Res. Lett.*, 25(1):1–13, 1999.

[8] J.V. Burke. An exact penalization viewpoint of constrained optimization. *SIAM J. Control Optim.*, 29(4):968–998, 1991.

[9] G.C. Calafiore. Ambiguous risk measures and optimal robust portfolios. *SIAM J. Optim.*, 18(3):853–877, 2007.

[10] Z. Chen and L. Epstein. Ambiguity, risk, and asset returns in continuous time. *Econometrica*, 70(4):1403–1443, 2002.

[11] V. K. Chopra and W. T. Ziemba. The effect of errors in means, variances and covariances on optimal portfolio choice. *J. Portfolio Management*, 19:6–11, 1993.

[12] G. Choquet. *Lectures on analysis. Vol. II: Representation Theory.* Edited by J. Marsden, T. Lance and S. Gelbart. W. A. Benjamin, Inc., New York-Amsterdam, 1969.

[13] G. Choquet. *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.

[14] A.R. Conn, N.I.M. Gould, and P. L. Toint. *Trust-region methods.* MPS/SIAM Series on Optimization. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2000.

[15] J. Dupačová. The minimax problem of stochastic linear programming and the moment problem. *Ekonom.-Mat. Obzor*, 13(3):279–307, 1977.

[16] J. Dupačová. On minimax decision rule in stochastic linear programming. In *Studies on mathematical programming (Papers, Third Conf. Math. Programming, Mátrafüred, 1975)*, volume 1 of *Math. Methods Oper. Res.*, pages 47–60. Akad. Kiadó, Budapest, 1980.

[17] J. Dupačová. The minimax approach to stochastic programming and an illustrative application. *Stochastics*, 20(1):73–88, 1987.

[18] Laurent El Ghaoui, Maksim Oks, and Francois Oustry. Worst-case value-at-risk and robust portfolio optimization: a conic programming approach. *Oper. Res.*, 51(4):543–556, 2003.

[19] E. Erdoğan and G. Iyengar. Ambiguous chance constrained problems and robust optimization. *Math. Program.*, 107(1-2, Ser. B):37–61, 2006.

[20] E Fama. The behavior of stock market prices. *Journal of Business*, 38(1):34–105, 1965.

[21] A.L. Gibbs and F.E. Su. On choosing and bounding probability metrics. *Int. Stat. Rev.*, 70(3):419–435, 2002.

[22] D. Goldfarb and G. Iyengar. Robust portfolio selection problems. *Math. Oper. Res.*, 28(1):1–38, 2003.

[23] R. Horst and N. V. Thoai. DC programming: overview. *Journal of Optimization Theory and Applications*, 103(1):1–43, 1999.

[24] R. Horst and H. Tuy. *Global optimization.* Springer-Verlag, Berlin, second edition, 1993.

[25] R. W. Klein and V. S. Bawa. The effect of estimation risk on optimal portfolio choice. *Journal of Financial Economics*, 3(3):215–231, June 1976.

[26] P.J. Maenhout. Robust portfolio rules and asset pricing. *Review of Financial Studies*, 17(4):951–983, 2004.

[27] G. Ch. Pflug and W. Römisch. *Modeling, Measuring and Managing Risk*. World Scientific, Singapore, 2007.

[28] G.Ch. Pflug and D. Wozabal. Ambiguity in portfolio selection. *Quantitative Finance*, 7(4):435–442, 2007.

[29] S.T. Rachev. *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.

[30] S.T. Rachev and L. Rüschendorf. *Mass transportation problems. Vol. I*. Probability and its Applications (New York). Springer-Verlag, New York, 1998.

[31] M. Richardson and T. Smith. A test for multivariate normality in stock returns. *The Journal of Business*, 66(2):295–321, 1993.

[32] W. W. Rogosinski. Moments of non-negative mass. *Proc. Roy. Soc. London Ser. A*, 245:1–27, 1958.

[33] A. Shapiro and S. Ahmed. On a class of minimax stochastic programs. *SIAM J. Optim.*, 14(4):1237–1249 (electronic), 2004.

[34] A. Shapiro and A. Kleywegt. Minimax analysis of stochastic problems. *Optim. Methods Softw.*, 17(3):523–542, 2002.

[35] M. Sion. On general minimax theorems. *Pac. J. Math.*, 8:171–176, 1958.

[36] Pham Dinh Tao and Le Thi Hoai An. A d. c. optimization algorithm for solving the trust-region subproblem. *SIAM J. on Optimization*, 8(2):476–505, 1998.

[37] R. H. Tütüncü and M. Koenig. Robust asset allocation. *Ann. Oper. Res.*, 132:157–187, 2004.

[38] Hoang Tuy. Minimax theorems revisited. *Acta Math. Vietnam.*, 29(3):217–229, 2004.

[39] J. Zackova. On minimax solutions of stochastic linear programming problems. *Casopis pro Pěstovânî Mathematiky*, 91:423–430, 1966.