Noname manuscript No.
(will be inserted by the editor)

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

David Wozabal

Received: date / Accepted: date

Abstract Value-at-Risk (VaR) is an integral part of contemporary financial regulations. Therefore the measurement of VaR as well as the design of VaR optimal portfolios is a highly relevant problem for financial institutions.

This paper treats a Value-at-Risk constrained Markowitz style portfolio selection problem when the distribution of returns of the considered assets are given in the form of finitely many scenarios. The problem is a non-convex stochastic optimization problem and can be reformulated as a difference of convex (D.C.) program. We apply the difference of convex algorithm (DCA) to solve 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 large real life problems are discussed.

Department of Statistics and Decision Support Systems, University of Vienna

Universitätsstraße 5/9, A-1010 Vienna, Austria

E-mail: david.wozabal@univie.ac.at

Keywords Risk Management · Stochastic Programming · Non-convex Optimization · Value-at-Risk

1 Introduction

Due to it's prominence in current regulatory frameworks for banks (Basel II, [24]) as well as for insurance companies (Solvency II) Value-at-Risk (VaR, see e.g. Jorion [25]) plays an important role in the modern financial and risk management literature. From an economic point of view, a t-day VaR at level α (with $0 < \alpha < 1$) of \$x means that the financial portfolio will incur a loss of at most \$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 can therefore be defined in terms of the quantile function F_X^{-1} of X as

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

where F_X is the distribution function of X.

The Value-at-Risk was first proposed by the global financial services firm J.P. Morgan Chase & Co. as a measure of acceptability for a financial position with random return. If VaR_α is taken to be the quantile function of the return distribution of X, then it is an acceptability functional: A higher value indicates a more acceptable, i.e. less risky portfolio. If on the other hand X represents the random losses, then $VaR_{1-\alpha}$ is a risk functional: High values indicate high risk. See Pflug and Römisch [33] for an in-depth discussion of acceptability and risk functionals. In this paper X

will represent anticipated (random) returns, and therefore VaR_α is considered to be an acceptability functional.

However, VaR_α – being the quantile of the return distribution – is non-concave. This is an undesirable property from the point of view of the axiomatic theory of risk measures as well as from a technical optimization viewpoint. The theoretical drawback is that Value-at-Risk may penalize diversification: given two financial positions and their anticipated (random) future returns X and Y it might be that

$$VaR_\alpha(X+Y) < VaR_\alpha(X) + VaR_\alpha(Y). \tag{1}$$

Note that (1) contradicts financial common sense and is a violation of the requirements specified in Artzner et al [7] and Artzner et al [8] for coherent risk measures (in particular the sup-additive property). In Artzner et al [8] the authors argue that the lack of coherence (or more specifically the lack of sup-additivity) of Value-at-Risk may indeed lead to wrong economical decisions in certain situations (see also the comprehensive discussion of this topic in McNeil et al [29]).

The technical problem stemming from non-concavity is that a Value-at-Risk constraint (or a VaR objective function) makes optimization problems computationally intractable (except in certain special cases where returns are known to be elliptically distributed, see for example Vehvilinen and Keppo [42] or McNeil et al [29]). Maximizing Value-at-Risk or minimizing a convex function under a Value-at-Risk constraint results in a non-convex optimization problem, which consequently is hard to solve.

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

$$\begin{aligned} \max\ & \mathbb{E}\left(w^\top \xi\right) & & \\ \text{s.t.}\ & \textstyle\sum_{i=1}^m w_i & = & \ 1 \\ & w_i & \in & \ [a_i, b_i],\ 1 \leq i \leq m \\ & VaR_\alpha\left(w^\top \xi\right) \geq a, & & \end{aligned} \tag{2}$$

where $w^\top = (w_1, \ldots, w_m)$, w_i denotes the relative weight and ξ_i the random return of asset i, $w^\top \xi = \sum_{i=1}^m w_i \xi_i$ and a_i, $b_i \in \mathbb{R}$ are the lower and upper bounds for the relative portfolio weights of the assets. Note that since the returns ξ_i are random, (2) is a single stage stochastic optimization problem.

Because of the shortcomings of VaR mentioned above it is often replaced by the Conditional Value-at-Risk (CVaR, also called Average Value-at-Risk or expected shortfall) in practical applications of optimization problems of the type (2) (see for example Andersson et al [6]). The *Conditional Value-at-Risk* CVaR_α of a random variable X is defined as

$$CVaR_\alpha(X) = \frac{1}{\alpha}\int_0^\alpha F_X^{-1}(t)dt = \frac{1}{\alpha}\int_0^\alpha VaR_t(X)dt \tag{3}$$

$$= \max\left\{a - \frac{1}{\alpha}\mathbb{E}([X-a]^-) : a \in \mathbb{R}\right\}, \quad \text{for} \quad 0 < \alpha \leq 1, \tag{4}$$

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

1. CVaR is relatively easy to incorporate into optimization problems like (2). In the case of discrete random variables a linear programming formulation exists, as shown in Rockafellar and Uryasev [36] and Uryasev [41].

2. CVaR is a coherent risk measure (see Pflug [32]), and does not suffer from the problem described in (1). In particular CVaR is a sup-additive acceptability functional favoring diversified portfolios over concentrated portfolios.
3. Additionally the Conditional Value-at-Risk is a concave lower minorant of the Value-at-Risk and therefore the function $w \mapsto CVaR_\alpha(w^\top \xi)$ is a lower bound for the function $w \mapsto VaR_\alpha(w^\top \xi)$.[1] Hence, portfolios constructed under a $CVaR_\alpha$ constraint will fulfill the corresponding VaR_α constraint as well, i.e. maximizing or controlling for the Conditional Value-at-Risk also influences the Value-at-Risk characteristics of the return distribution favorably.

One could therefore conclude that the Value-at-Risk should be entirely replaced by the Conditional Value-at-Risk to circumvent the problems mentioned above. However, since the Value-at-Risk is explicitly incorporated in the Basel II and Solvency II regulations, virtually every financial institution has to actively monitor and manage the Value-at-Risk of it's portfolios: most notably banks have to control credit risk as well as market and operational risk using the Value-at-Risk and have to report the resulting figures for all their major business branches. The amount of regulatory capital to be held under the current regulations directly depends on these figures. It would therefore be a competitive disadvantage to completely neglect the Value-at-Risk in decision making and exclusively use other concave acceptability functionals as guiding principle when designing portfolios.

[1] In fact it can be shown that the Conditional Value-at-Risk is the best conservative approximation of the Value-at-Risk from the class of law invariant concave acceptability measures, which are continuous from below (see [17], Theorem 4.61)

Correspondingly there is a large body of academic literature which focuses on either estimating the Value-at-Risk of complex portfolios correctly (see McNeil et al [29] for a discussion of this topic) or treats problems of actively managing the Value-at-Risk of a portfolio and thereby solving problems similar to (2) (see for example [9–12, 15, 18–21, 27, 30, 31, 34, 35, 39], or Wozabal et al [43] for an overview).

The results presented in this paper build on the difference of convex (D.C.) formulation of VaR derived in Wozabal et al [43], where a conical Branch-and-Bound algorithm for optimizing D.C. functions is used to find global optima of (2). In this paper an approximate solution technique called difference of convex algorithm (DCA) is applied 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 realistically sized problems.

This paper is organized as follows. In Section 2 D.C. reformulation of Value-at-Risk is reviewed. Sections 3 discusses the DCA and its application to the problem at hand, while Section 4 presents a set of numerical results based on real-world financial market data. Section 5 concludes the paper.

2 Reformulation of VaR_α as a D.C. function

As already outlined we will consider the classical form of an asset allocation problem based on the seminal paper Markowitz [28] 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. Generally speaking a risk measure can be included in a portfolio optimization problem in three ways: by minimizing the risk (while possibly enforcing some lower bound on expected return), by maximizing the expected return under some constraint on risk/acceptability of the portfolio or by bi-criteria optimization with both the expected return as well as the acceptability of the return distribution in the objective function. Clearly all the above approaches are essentially equivalent as they all yield portfolios on the *efficient frontier*. In this paper we consider problem (2), i.e. following the second approach. However the presented methods can be adapted in a straightforward manner to also treat the other versions of the problem.

To reformulate problem (2) into a D.C. problem, we assume that the distribution of the random asset returns $\xi = (\xi_1, \dots, \xi_m)$ 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 realization of ξ in scenario s will be denoted by $\xi^s \in \mathbb{R}^m$ while the probability of the scenario will be denoted by p_s.

Working with discrete distributions is common in stochastic programming because 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, especially in the multi-variate case, certain features like heavy tails, and non-normal skewness or kurtosis can be captured in a natural way.

Since in a non-parametric setting, analytical solutions to (2) can't be found, we resort to numerical procedures. As already mentioned $\mathrm{VaR}_\alpha(X)$ is non-concave in X, and therefore problem (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.

Using the above notation problem (2) becomes

$$\begin{aligned} \max\ & \textstyle\sum_{s=1}^{S} w^\top \xi^s p_s \\ \text{s.t.}\ & \textstyle\sum_{i=1}^{m} w_i = 1 \\ & w_i \in [a_i, b_i],\ \forall 1 \le i \le m \\ & VaR_\alpha \left(w^\top \xi\right) \ge a. \end{aligned} \tag{5}$$

The above problem was shown to be a NP hard, combinatorial problem in [10].

Although we will restrict ourselves to problem (5) in this section, the proposed methodology can easily be adapted to other cases, e.g. problems with more complicated convex portfolio constraints, problems with VaR as the objective function as demonstrated in Section 4.3.

To reformulate the above problem, we derive a D.C. formulation of the Value-at-Risk. If X follows a discrete distribution taking the values $x_1, \ldots, x_S$ with probabilities $p_1, \ldots, p_S$ (in our case X represents the anticipated (random) returns of a portfolio w, i.e. $X = w^\top \xi$), then

$$CVaR_\alpha(X) = \frac{1}{\alpha} \sum_{i=1}^{k^*} x_{i:S} p_{i:S} + \frac{1}{\alpha} x_{k^*+1:S} \varepsilon, \tag{6}$$

where $x_{1:S} \le x_{2:S} \le \cdots \le x_{S:S}$ is the set of ordered outcomes of X, $p_{i:S}$ are the corresponding scenario probabilities, $\varepsilon := \alpha - \max\left\{\sum_{i=1}^{k} p_{i:S} : \sum_{i=1}^{k} p_{i:S} < \alpha\right\}$ and $k^* := \max\left\{k : \sum_{i=1}^{k} p_{i:S} < \alpha\right\}$.[2] Note that $VaR_\alpha(X) = x_{k^*+1:S}$ and that by (6) for all $0 < \gamma <$

[2] To deal with the case $p_{1:S} > \alpha$, we define $\max \emptyset = 0$ which leads to $\varepsilon = \alpha$ and $k^* = 0$ in the aforementioned case.

ε

$$\frac{\alpha}{\gamma}CVaR_{\alpha}(X) = \frac{1}{\gamma}\sum_{i=1}^{k^*} x_{i:S}p_{i:S} + \frac{\varepsilon}{\gamma}x_{k^*+1:S}$$
$$\frac{\alpha-\gamma}{\gamma}CVaR_{\alpha-\gamma}(X) = \frac{1}{\gamma}\sum_{i=1}^{k^*} x_{i:S}p_{i:S} + \frac{\varepsilon-\gamma}{\gamma}x_{k^*+1:S}.$$

Hence, $\mathrm{VaR}_{\alpha}(X)$ can be written as the difference of the two concave functions as

$$VaR_{\alpha}(X) = x_{k^*+1:S} = \frac{\alpha}{\gamma}CVaR_{\alpha}(X) - \frac{\alpha-\gamma}{\gamma}CVaR_{\alpha-\gamma}(X).$$

Problem (5) consequently becomes

$$\begin{aligned} &\max \textstyle\sum_{s=1}^{S} w^{\top}\xi^s p_s \\ &\text{s.t.} \quad \textstyle\sum_{i=1}^{m} w_i = 1 \\ &\qquad w_i \in [a_i, b_i], \quad \forall 1 \le i \le m \\ &\qquad \tfrac{\alpha}{\gamma}CVaR_{\alpha}(w^{\top}\xi) - \tfrac{\alpha-\gamma}{\gamma}CVaR_{\alpha-\gamma}(w^{\top}\xi) \ge a, \end{aligned}$$

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

$$\begin{aligned} &\min -\textstyle\sum_{s=1}^{S} w^{\top}\xi^s p_s \\ &\text{s.t.} \quad \textstyle\sum_{i=1}^{m} w_i = 1 \\ &\qquad w_i \in [a_i, b_i], \quad \forall 1 \le i \le m \\ &\qquad -\tfrac{\alpha}{\gamma}CVaR_{\alpha}(w^{\top}\xi) - (-\tfrac{\alpha-\gamma}{\gamma}CVaR_{\alpha-\gamma}(w^{\top}\xi)) \le -a. \end{aligned} \tag{7}$$

Note that ε depends on on the order of the scenarios, i.e. on the portfolio decision w. We therefore write $\varepsilon = \varepsilon(w)$. Hence, to find $\gamma > 0$ smaller than $\varepsilon(w)$, for all possible w, we would have to solve the following bin packing problem

$$\inf_{w} \varepsilon(w) = \alpha - max\left\{ b : b = \sum_{i \in I'} p_i < \alpha,\ I' \subset I = \{1, \ldots, S\} \right\} > 0.$$

However, for practical applications it turns out that if γ is chosen small relative to the scenario probabilities p_s (for example $\gamma = \min_s p_s c$, with $c \ll 1$), then $\gamma < \varepsilon(w)$ is never violated in all the considered examples.

Remark 1 If all the scenarios have equal probabilities, i.e. $p_s = S^{-1}$, we set $k^* = \max\left\{k \in \mathbb{N} : \frac{k}{S} < \alpha\right\}$ and write[3]

$$\begin{aligned} VaR_\alpha(X) &= F_X^{-1}(\alpha) = x_{k^*+1:S} = \sum_{i=1}^{k^*+1} x_{i:S} - \sum_{i=1}^{k^*} x_{i:S} \\ &= (k^*+1)\frac{S}{k^*+1}\sum_{i=1}^{k^*+1}\frac{1}{S}x_{i:S} - k^*\frac{S}{k^*}\sum_{i=1}^{k^*}\frac{1}{S}x_{i:S} \\ &= (k^*+1)CVaR_{\frac{k^*+1}{S}}(X) - k^*CVaR_{\frac{k^*}{S}}(X). \end{aligned}$$

The assumption of equal probabilities for scenarios is often fulfilled in stochastic programming, since equally weighted scenarios naturally result from simulation driven scenario generation methods as well as in the case where historical realization of random variables are used as scenarios.

Problem (7) is a standard constrained D.C. problem. Since in the finite scenario case CVaR is a polyhedral function in w (see (4)), the problem only involves linear and polyhedral functions and therefore is a so called polyhedral D.C. problem.

There exist a variety of global solution techniques for D.C. problems (for an overview see Horst and Tuy [23]). However, global solutions to general D.C. problems are computationally difficult to obtain and existing solution techniques are restricted to problems of relatively small dimension.

[3] We additionally assume that $\alpha > \frac{1}{S}$. If $\alpha \leq \frac{1}{S}$, then $VaR_\alpha(X) = min_{1\leq s\leq S} x_s$, which is convex in X and the problem reduces to a significantly easier (convex) problem.

In the next section we therefore introduce an approximate solution method based on the so called Difference of Convex Algorithm.

3 Solution by the DCA

3.1 The Difference of Convex Algorithm

The difference of convex algorithm (DCA) was originally proposed in [38] and is an approximate solution method for unconstrained D.C. problems of the form

$$\inf\{f(x) = g(x) - h(x) : x \in \mathbb{R}^N\}, \tag{8}$$

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. Both of the versions start from an initial value x^0 and produce two sequences (x^k) and (y^k) of candidates of increasing quality for the solutions of the primal problem (8) and the dual D.C. problem

$$\inf_{y\in\mathbb{R}^N} h^*(y) - g^*(y) \tag{9}$$

respectively.

Setting $y^0 = 0$ the simplified DCA proceeds by repeatedly solving the two convex optimization problems

$$\inf_{y\in\mathbb{R}^N} \{h^*(y) - (g^*(y^{k-1}) + \langle x^k, y - y^{k-1}\rangle)\}. \tag{10}$$

and

$$\inf_{x\in\mathbb{R}^N} \{g(x) - (h(x^k) + \langle x - x^k, y^k\rangle)\}. \tag{11}$$

where the former yields y^k given $x^0,\dots,x^k$ and $y^0,\dots,y^{k-1}$ and the latter yields x^{k+1} given $x^0,\dots,x^k$ and $y^0,\dots,y^k$.

The first problem can be viewed as a convex approximation of (9), while the second problem can be thought of as a convex approximation of the original problem.

Note that the solution to problem (11) is an element of $\partial g^*(y^k)$ and likewise the solution of (10) is an element of $\partial h(x^k)$.

The complete DCA differs from the simplified DCA in that it does not suffice to choose arbitrary elements of $\partial g^*(y^k)$ and $\partial h(x^k)$ for given y^k and x^k but the elements y^k and x^{k+1} are obtained as follows

$$y^k \in \arg\min\{\langle x^k, y\rangle - g^*(y) : y \in \partial h(x^k)\} \tag{12}$$

$$x^{k+1} \in \arg\min\{\langle x, y^k\rangle - h(x) : x \in \partial g^*(y^k)\}. \tag{13}$$

Observe that the above problems are in general hard to solve, since they require the minimization of concave functions $-g^*$ and $-h$, which amounts to solving non-convex problems.

Both versions of the algorithm guarantee 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 (11) and the dual problem (10) respectively. In the case of a polyhedral D.C. problem it can be shown that the complete DCA terminates at a local solution after finitely many iterations.

The DCA has proven to be very effective in solving a wide range of different D.C. programs (see for example An and Tao [1, 2], An et al [4], Conn et al [13], Tao and An [37], Thi et al [40]) and its low computational complexity makes it possible to work with large D.C. programs. For a more detailed discussion of DCA algorithm and its properties see An and Tao [3].

3.2 Application of the DCA to problem (2)

In this section we apply the general DCA methodology to the portfolio selection problem at hand. It turns out that we can efficiently solve the non-convex problem (12) but not problem (13). We therefore apply a hybrid version of the DCA in which we solve problem (11) and problem (12) interchangeably to obtain candidate solutions for the dual and primal problem respectively.

It turns out that the proposed hybrid version of the algorithm preserves many of the desirable properties of the complete DCA, while remaining computationally simple (at least for the problems considered in this paper).

To comply with the notation introduced in Section 2, we denote candidates for the primal solution by w^j and candidates for the dual solutions by w^{*j}.

To apply the DCA to problem (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 (2) are fulfilled, i.e.

$$C = \left\{ w \in \mathbb{R}^m : \sum_{i=1}^{m} w_i = 1,\ w_i \in [a_i, b_i],\ \forall i : 1 \leq i \leq m \right\}.$$

Using the convex indicator function

$$\chi_C(w) = \begin{cases} 0, & w \in C \\ \infty, & \text{otherwise} \end{cases}$$

we can rewrite (7) as

$$\begin{aligned} &\min \; -\textstyle\sum_{s=1}^{S} w^\top \xi^s p_s + \chi_C(w) \\ &\text{s.t.} \;\; -\tfrac{\alpha}{\gamma} CVaR_\alpha(w^\top \xi) - (-\tfrac{\alpha-\gamma}{\gamma} CVaR_{\alpha-\gamma}(w^\top \xi)) \le -a. \end{aligned} \tag{14}$$

and thus penalize for not fulfilling the respective convex constraints represented by C. To penalize for the D.C. constraint note that

$$\begin{aligned} max(-VaR_\alpha(w^\top \xi) + a, 0) = max\left(-\frac{\alpha}{\gamma} CVaR_\alpha(w^\top \xi) + a, -\frac{\alpha-\gamma}{\gamma} CVaR_{\alpha-\gamma}(w^\top \xi)\right) \\ + \frac{\alpha-\gamma}{\gamma} CVaR_{\alpha-\gamma}(w^\top \xi). \end{aligned} \tag{15}$$

This follows from 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

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

Using the results on exact penalization of D.C. programs from An et al [5], for some $\tau > 0$ we finally rewrite (14) as the equivalent problem

$$\begin{aligned} \min \; -\sum_{s=1}^{S} w^\top \xi^s p_s + \chi_C(w) + \tau \Bigg[max\left(-\frac{\alpha}{\gamma} CVaR_\alpha(w^\top \xi) + a, -\frac{\alpha-\gamma}{\gamma} CVaR_{\alpha-\gamma}(w^\top \xi)\right) \\ + \frac{\alpha-\gamma}{\gamma} CVaR_{\alpha-\gamma}(w^\top \xi) \Bigg] \end{aligned} \tag{16}$$

and define

$$\begin{aligned} g(w) &= -\sum_{s=1}^{S} w^\top \xi^s p_s + \chi_C(w) + \tau max\left(-\frac{\alpha}{\gamma} CVaR_\alpha(w^\top \xi) + a, -\frac{\alpha-\gamma}{\gamma} CVaR_{\alpha-\gamma}(w^\top \xi)\right) \\ h(w) &= -\tau \frac{\alpha-\gamma}{\gamma} CVaR_{\alpha-\gamma}(w^\top \xi). \end{aligned}$$

Hence, we arrived at a problem of the form (8) to which the DCA can be applied.

We first discuss how to solve problem (12) for a given portfolio w^j, i.e.

$$\min\left\{\langle w^j, w^*\rangle - g^*(w^*) : w^* \in \partial h(w^j)\right\}.$$

To characterize the elements in $\partial h(w^j)$, we recall from (6), that

$$h(w^j) = -\tau\frac{\alpha-\gamma}{\gamma}CVaR_{\alpha-\gamma}(w^{j\top}\xi) = -\frac{\tau}{\gamma}\sum_{i=1}^{k^*} r_{i:S}(w^j)p_{i:S} - \frac{\tau(\varepsilon-\gamma)}{\gamma}r_{k^*+1:S}(w^j)$$

where the $r_{i:S} \in \mathbb{R}$ are the ordered returns depending on the portfolio w^j. The dependence of h on w^j is twofold: the return $r_i = w^{j\top}\xi^i$ in each of the scenarios ξ^i is linearly dependent on w^j and the ordering of the returns (which in turn determines which of the r_i are summed) depends on w^j as well. These two dependencies render the mapping $w \mapsto CVaR$ piecewise linear. For a given w^j the subgradients are therefore convex combinations of the gradients of the linear functions defining $CVaR_{\alpha-\gamma}$ locally. To investigate this fact, assume w.l.o.g. that the ordering of the returns above coincides with the scenario numbers, i.e. $r_{i:S} = w^{j\top}\xi^i = r_i$. Note that

$$\begin{aligned}\partial h(w^j) = \left\{\nabla h(w^j)\right\} &= \left\{-\frac{\tau}{\gamma}\sum_{i=1}^{k^*}\nabla r_i(w^j)p_i - \frac{\tau}{\gamma}\nabla r_{k^*+1}(w^j)(\varepsilon-\gamma)\right\}\\ &= \left\{-\frac{\tau}{\gamma}\sum_{i=1}^{k^*}\xi^i p_i - \frac{\tau}{\gamma}\xi^{k^*+1}(\varepsilon-\gamma)\right\}\end{aligned}$$

if $r_{k^*} < r_{k^*+1} < r_{k^*+2}$ because in this case $w \mapsto CVaR_{\alpha-\gamma}(w^\top\xi)$ is linear in a neighborhood of w^j.

If on the other hand

$$r_1 \leq \cdots \leq r_{s-1} < r_s = \cdots = r_{k^*+1} = \cdots = r_t < r_{t+1} \leq \cdots \leq r_S$$

for a given w^j and some $s,t \in \mathbb{N}$ with $s \leq k^* < t$, then $\nabla h(w^j)$ does not exist, i.e. $\partial h(w^j)$ contains more than one element.

To analyze this case let $\bar{p} = \sum_{i=1}^{s-1} p_i$ and define

$$\mathscr{J} = \left\{ I \subseteq \{s, \dots, t\} : \bar{p} + \sum_{i \in I} p_i \geq \alpha,\ \exists j \in I \text{ with } \bar{p} + \sum_{i \in I \setminus \{j\}} p_i < \alpha \right\}$$

and $J(I) = \left\{ j : \bar{p} + \sum_{i \in I \setminus \{j\}} p_i < \alpha \right\}$. Then by the above remarks $\partial h(w^j)$ can be written as the convex hull of the following vectors

$$V = \left\{ -\frac{\tau}{\gamma} \left(\sum_{i=1}^{s-1} \xi^i p_i + \sum_{i \in I \setminus \{j\}} \xi^i p_i + (\alpha - \gamma - \bar{p} - \sum_{i \in I \setminus \{j\}} p_i) \xi^j \right) : \forall I \in \mathscr{J},\ j \in J(I) \right\}.$$

To solve problem (12), we therefore just have to find

$$y^k \in \arg\min\{\langle x^k, y \rangle - g^*(y) : y \in V\}. \tag{17}$$

Note that $g^*(y)$ for a $y \in V$ is just the optimal value of (18) for $w^{*j} = y$ (see below). Therefore we can actually solve the non-convex problem (12) by enumeration. It should also be remarked that V could potentially contain a large number of elements, rendering the proposed strategy practically infeasible. However, the numerical experiments performed for this paper show that $|V| = 1$ in most of the iterations and even for the examples with many scenarios $|V|$ is of the order 10^3 at the very most.

Remark 2 If the (ordered) scenarios are equally probable, we want to find subgradients of $h(w) = -\tau k^* CVaR_{\frac{k^*}{S}}(w^\top \xi)$. Suppose

$$r_1 \leq \cdots \leq r_{s-1} < r_s = \cdots = r_{k^*+1} = \cdots = r_t < r_{t+1} \leq \cdots \leq r_S$$

as before, then for every $I \subseteq \{s, \dots, t\}$ of size $(k^* - s + 1)$ it holds that

$$\sum_{i=1}^{s-1} p_i + \sum_{i \in I} p_i = \frac{(s-1) + (k^* - s + 1)}{S} = \frac{k^*}{S}$$

while $\sum_{i=1}^{s-1} p_i + \sum_{i\in I\setminus\{k\}} p_i = \frac{k^*-1}{S}$ for every $k \in I$ (i.e. $I \in \mathscr{J}$ and $J(I) = I$). Hence, $\partial h(w^j)$ is contained in the convex hull of

$$V = \left\{ -\tau \sum_{i=1}^{s-1} \xi^i - \tau \sum_{i\in I} \xi^i : I \subset \{s,\dots,t\}, |I| = (k^* - s + 1) \right\}.$$

To apply the complete DCA we would have to solve (13) as well. Unfortunately, problem (13) can not be solved efficiently and we have to contend ourselves with finding an arbitrary element of $\partial g^*(y^j)$ by solving the following linear problem

$$\begin{aligned} \min\ & -\textstyle\sum_{s=1}^{S} w^\top \xi^s p_s + \tau M - \langle w, w^{*j} \rangle \\ \text{s.t.}\ & -\tfrac{\alpha}{\gamma} CVaR_\alpha(w^\top \xi) + a \ \le M \\ & -\tfrac{\alpha-\gamma}{\gamma} CVaR_{\alpha-\gamma}(w^\top \xi) \le M \\ & w \in C, \end{aligned} \tag{18}$$

which corresponds to (11) in our setting (since the constant terms $h(w^j)$ and $\langle w^j, w^{*j} \rangle$ do not influence the solution of (11) but only the optimal value).

As mentioned above we apply a hybrid of the simplified and complete version of the DCA. The next Theorem justifies the use of the outlined approach and establishes convergence properties similar to the complete DCA for the hybrid version.

Theorem 1 *Let $g : \mathbb{R}^N \to \mathbb{R}$, $h : \mathbb{R}^N \to \mathbb{R}$ be convex functions, $x^0 \in \mathbb{R}^N$ and $y^0 = 0$. Now define two sequences $(x^j)_{j\in\mathbb{N}_0}$, $(y^j)_{j\in\mathbb{N}_0}$ (where $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$) by*

$$\begin{aligned} y^j &\in \arg\min\{\langle x^j, y \rangle - g^*(y) : y \in \partial h(x^j)\} \\ x^{j+1} &\in \arg\min\{g(x) - (h(x^j) + \langle x - x^j, y^j \rangle) : x \in \mathbb{R}^N\}. \end{aligned}$$

Then it holds that

1. *If $g(x^{j+1}) - h(x^{j+1}) = g(x^j) - h(x^j)$, then*

$$x^j \in \mathscr{P}_l := \left\{x \in \mathbb{R}^N : \partial h(x) \subseteq \partial g(x)\right\}$$

and y^j is a critical point, i.e. $\partial g^*(y^j) \cap \partial h^*(y^j) \neq \emptyset$.

2. *If h is a polyhedral function, then the objective values of the hybrid DCA have finite convergence to a local optimum of the primal problem.*

Proof 1. If the algorithm stops finitely, i.e. $g(x^{j+1}) - h(x^{j+1}) = g(x^j) - h(x^j)$, then $x^j \in \partial g^*(y^j)$. By construction

$$y^j \in \mathscr{S}(x^j) = \arg\min\{\langle x^*, y\rangle - g^*(y) : y \in \partial h(x^*)\}$$

and therefore $x^j \in \mathscr{P}_l$ by an application of Theorem 2.3 in Tao and El Bernoussi [38]. Since $y^j \in \partial h(x^j)$ and $x^j \in \partial g^*(y^j)$ (by Theorem 3.7 in Tao and An [37] or Theorem 3 in Tao and El Bernoussi [38]), we have $x^j \in \partial h^*(y^j) \cap \partial g^*(y^j)$ and therefore y^j is a critical point.

2. The result relies on the fact that, h is a polyhedral function and therefore has a representation 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 follows 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 from $\partial h(x)$. From the above it follows that there are only finitely possible y^js and since the sequence $(g(x^j) - h(x^j))_j$ 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 $i \in \mathbb{N}$.

By part 1, $x^k \in \mathscr{P}_l$ and therefore x^k is a local optimum by Theorem 2 in Durier [14] and Corollary 3.3 in Tao and An [37].

Summarizing we can therefore obtain local solutions to problem (7) by repeatedly solving problem (17) and (18). Note that since we are dealing with a polyhedral problem all the subproblems can be solved using standard linear programming software, which makes the algorithm feasible also for relatively big problems.

4 Applications

In this section we test the algorithm by applying it to real market data. In section 4.1 we compare the globally optimal portfolios obtained in [43] to the optimal portfolios found by the DCA. In section 4.2 we compute optimal portfolios for a data set with a large number of assets while in Section 4.3 we tackle credit risk problems with a vast number of scenarios to demonstrate the applicability of the approach to realistically sized problems, which can no longer be solved to global optimality. While Sections 4.1 and 4.2 we use historical returns as scenarios and consequently are solving problems with equal scenario probabilities, in Section 4.3 the scenarios are unequally weighted.

All optimization problems were solved using the MOSEK, version 5 interfaced with MATLAB 2008a. The calculations we performed on a notebook with a Pentium Mobile Processor (1.8 GHz) 1.5GB RAM using Windows XP SP 3.

Name	Average Return	Variance	$VaR_{.05}$	$CVaR_{.05}$
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 1 Characteristics of the weekly returns of the five indices subsequently used in portfolio optimization. Time frame: 02.01.2004 - 27.12.2005 (resulting in 104 weekly return scenarios).

4.1 Comparison with global optima

In the following we compare results obtained for problem (7) by applying the DCA with results we obtained in [43]. Weekly closing values of the following 5 indices have been used to calculate the equally probable discrete return scenarios: US Long Bond, Standard & Poors 100, Nasdaq 100, FTSE 100, and Hang Seng. Table 1 gives an overview of the characteristics of the used data.

In Figure 1 we present efficient frontiers for VaR problem (7) obtained by varying the parameter a. It turns out that the interesting range for a is between the VaR of the portfolio that consists only of the asset with the highest return (i.e. Hang Seng with $VaR_{.05}$ 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. 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 0.9786 and from 0.981 to 0.985. Also the portfolio compositions obtained by the DCA and the Branch-and-Bound algorithm are very similar in this area as Figure 1 shows.
2. For some of the values of a the DCA finds solutions which are slightly inferior to the solutions found by the global Branch-and-Bound method.
3. For a close to values that make the problem infeasible (from 0.9865 onwards) the DCA is not able to find feasible points for problem (7). This problem could be rectified by providing feasible starting points to the algorithm. Since - in the current implementation - this is not the case the algorithm has to search for feasible points itself, which becomes increasingly tricky as the problem approaches infeasibility.

The decreasing quality of the solutions with the parameter a are in line with the observations made in [43], 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 (7). Because of the comparatively low runtime the algorithm will be preferable to a global solution approach in practice.

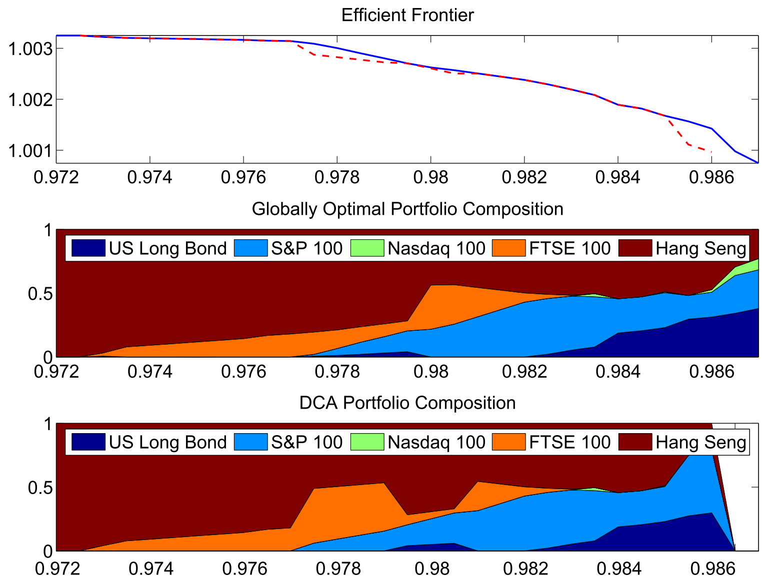


Fig. 1 The topmost plot shows the efficient frontier (x-axis: acceptability bound, y-axis: maximum return) for Branch-and-Bound algorithm from Wozabal et al [43] (blue) and the DCA (dashed, red). The corresponding optimal portfolios for the DCA and the Branch-and-Bound algorithm are depicted in the second and the third plot respectively (x-axis: acceptability bound, y-axis: portfolio composition).

See table 2 for a comparison of the respective computing times of the DCA and the Branch-and-Bound algorithm.

4.2 Application of DCA to medium sized data sets

To demonstrate the ability of the DCA to solve bigger problems and thereby its practical applicability, we repeat the analysis of the last section with daily return data of the assets comprising the S&P 500 index. Our scenario set consists of the 503 daily

a	Branch-and-Bound	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 2 Runtime of the DCA algorithm versus the Branch-and-Bound algorithm in seconds.

returns for 491 assets observed in the years 2007 and 2008.[4] Furthermore we choose $\alpha = 0.1$ and $a_i = 0$ and $b_i = 1$ for all $i = 1, \ldots, m$. The number of assets as well as 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 VaR portfolios (see Larsen et al [27]). 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 Conditional Value-at-Risk and iterative application of a truncation operation. The algorithm performs well even on vast data sets and

[4] We only use the 491 assets of the S&P 500 which where members of the index for the whole observation period.

is widely applied in the industry. Algorithm A1 from Larsen et al [27] 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

$$\begin{array}{lll} \max_{w,\gamma} CVaR_{\alpha_n}(\sum_{i=i_n}^{S} w^\top \xi^i) & & \\ s.t. \quad \mathbb{E}(w^\top \xi) \geq C & & \\ \quad\quad w^\top \xi^i & \leq \gamma, & i \leq i_n \\ \quad\quad w^\top \xi^i & \geq \gamma, & i > i_n \\ \quad\quad a_i \leq w & \leq b_i, & \forall 1 \leq i \leq m \end{array}$$

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 $n = n+1$, $b_n = \alpha + (1-\alpha)(1-\zeta)^n$, $i_n = [S(1-b_n)]$ 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 heuristic algorithm A1 for a given set of lower bounds on the expectations and subsequently using the optimal Value-at-Risk figures as constraints in the DCA algorithm. In our concrete example the range of r.h.s. values a for the heuristic algorithm is chosen to be in $[1, 1.00219]$. The upper bound is chosen as the maximal average return of the set of observed assets, i.e. the maximum expected return which can be attained by a portfolio constructed from this particular set of assets. The value 1 is the lowest sen-

Value-at-Risk ($\alpha = 0.1$)	Expectations (A1)	Expectations (DCA)	difference
0.93939	1.00219	1.00219	-0.000000
0.96098	1.00209	1.00210	0.000011
0.96441	1.00199	1.00201	0.000019
0.96961	1.00189	1.00192	0.000022
0.97230	1.00179	1.00183	0.000035
0.97740	1.00169	1.00171	0.000019
0.97964	1.00159	1.00166	0.000070
0.98069	1.00149	1.00159	0.000091
0.98170	1.00139	1.00155	0.000154
0.98342	1.00129	1.00145	0.000159
0.98366	1.00119	1.00139	0.000196
0.98450	1.00109	1.00131	0.000212
0.98474	1.00089	1.00123	0.000335
0.98521	1.00079	1.00113	0.000337
0.98616	1.00069	1.00105	0.000351
0.98738	1.00049	1.00076	0.000270
0.98799	1.00039	1.00063	0.000236
0.98802	1.00019	1.00063	0.000435
0.98859	1.00009	1.00051	0.000420

Table 3 Optimal expected returns, Value-at-Risk for daily return data for S&P 500 data for Algorithm A1 from Larsen et al [27] as well as the DCA.

sible value for a lower bound on returns. The results of the comparison are compiled in Table 3 and depicted in Figure 2.

The results show that the approximate algorithm from Larsen et al [27] 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. The differences in

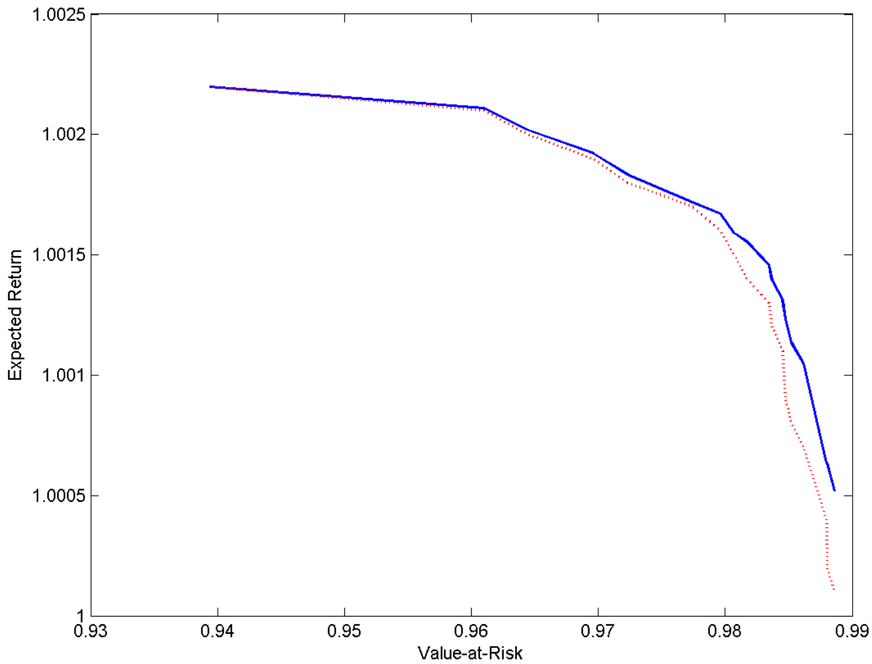


Fig. 2 Efficient frontiers for $VaR_{0.1}$. Approximate algorithm is depicted in red (dotted), while the results from the DCA are blue.

optimal expectations range from approximately 0 to approximately 0.0004, the latter being a significant difference (for expected daily returns). It can also be observed for values of a close to 0.999 that the DCA yields a returns which are up to eight times as large as the approximative algorithm A1.

We conclude that the application of the DCA is a good alternative to existing methods to find Value-at-Risk optimal portfolios also for situations with fairly many assets and large scenario sets.

4.3 A real life example: minimizing credit risk

The last set of examples that we analyze comprises of two credit risk problems. In both the examples we have a moderate number of investment possibilities but a large number of unequally weighted scenarios. To demonstrate the flexibility of the proposed approach we consider the alternative problem formulation

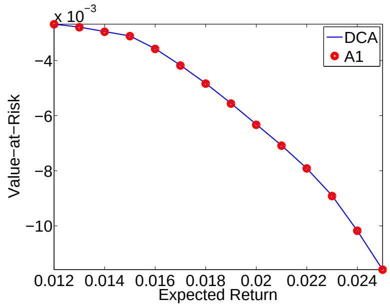
$$\begin{aligned} \max\ & VaR_\alpha\left(w^\top \xi\right) \\ \text{s.t.}\ & w_i \quad \in [0,1],\ 1 \leq i \leq m \\ & w^\top e \geq r, \end{aligned}$$

i.e. we are maximizing the Value-at-Risk under an expectation constraint, where $e = (e_1, \ldots, e_m)$ comprises of the expectations of the single assets not necessarily calculated from the scenarios. Since the above problem can be reformulated as unconstrained D.C. program using the same methods discussed in section 2, we omit the details.

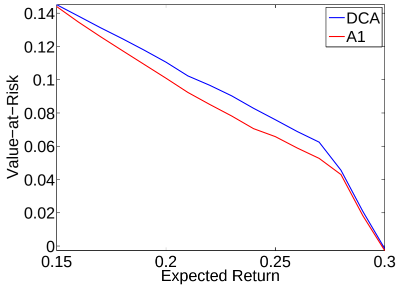
The data for the first problem is taken from the case study section of the *American Optimal Decisions* homepage and contains 100,000 return scenarios for 23 clusters of retail loans. We remove asset 17 and 23 from the asset universe, since these assets are strongly dominant in the sense that they yield the highest expectations (by far) while at the same time being the least risky (hence, inclusion of these assets would make the problem trivial). A more detailed description of the data can be found in the case study description available online.[5]

Results on this data set demonstrate the limitations of the DCA, in two ways:

[5] See the case studies section of `http://www.aorda.com/`.



(a)



(b)

Fig. 3 Efficient frontiers for the AOD case study in (a) as well as for the iTraxx example in (b).

1. the size of the scenario set makes the solution of problem (18) rather slow[6] (see the comparison of runtimes in Table 4);
2. the fact that the data set does not contain scenarios where extreme losses occur makes the tail of the return distribution rather short (the average kurtosis of the assets is 7.97). This implies that the Conditional Value-at-Risk and the Value-at-Risk are close and consequently the benchmark algorithm works very well.

The efficient frontier for a problem with $\alpha = 0.01$ and return r ranging from 0.12 to 0.25 are depicted in Figure 3, while the runtimes as well as the numerical results can be found in Table 4. It can be concluded that for this data set the DCA offers no significant advantage over algorithm A1 in terms of solution quality, while being computationally more expensive.

[6] The speed of the algorithm could be significantly improved by solving problem (18) in a computationally more efficient way (see for example [16]). Another strategy would be to explore techniques to reduce the scenario set without changing the problem. However, exploring these methods more carefully is beyond the scope of the present paper.

	DCA			A1			CVaR Optimal		
r	VaR	A	Time	VaR	A	Time	CVaR	VaR	A
0.012	-0.002681	4	1746	-0.002688	5	505	-0.006603	-0.004416	1
0.013	-0.002789	4	522	-0.002795	4	598	-0.006603	-0.004416	1
0.014	-0.002952	4	41960	-0.002960	5	704	-0.004229	-0.003200	2
0.015	-0.003110	5	54	-0.003118	6	442	-0.004049	-0.003298	2
0.016	-0.003573	4	17025	-0.003581	6	526	-0.004395	-0.003631	2
0.017	-0.004173	5	33118	-0.004185	9	527	-0.005908	-0.004757	2
0.018	-0.004830	3	1081	-0.004840	9	397	-0.007748	-0.006107	2
0.019	-0.005552	4	12180	-0.005564	8	510	-0.008822	-0.006974	2
0.02	-0.006314	3	3052	-0.006334	9	455	-0.008687	-0.007067	2
0.021	-0.007080	3	2927	-0.007091	7	392	-0.008676	-0.007271	2
0.022	-0.007913	3	987	-0.007916	4	555	-0.009033	-0.007879	2
0.023	-0.008901	2	2746	-0.008917	6	408	-0.009984	-0.008900	2
0.024	-0.010176	3	261	-0.010176	3	615	-0.011409	-0.010173	2
0.025	-0.011587	3	28	-0.011587	3	507	-0.013078	-0.011630	2

Table 4 Runtimes, objective values as well as number of assets in the optimal portfolios (A) for the AOD case study (100.000 scenarios). Additional information on the CVaR versions of the problems is given.

The second example in this section consists of 10,000 return scenarios for all the iTraxx CDX tranches for maturities 5 years, 7 years as well as 10 years (in all 5 tranches per maturity, i.e. in all 15 assets).[7] The scenarios are simulated using a model for credit migrations proposed by [26] (see [22] for details on model fitting and simulation). In contrast to the first example in this section, the return distributions of the assets have extremely fat tails (the average kurtosis of the assets is 117), characteristic for this type of assets.

7 See http://indices.markit.com/ for information on iTraxx products.

	DCA			A1			CVaR Optimal		
r	VaR	A	Time	VaR	A	Time	CVaR	VaR	A
0.15	0.145125	5	74	0.143993	3	8	0.085685	0.143933	3
0.16	0.138232	5	61	0.134642	3	20	0.079660	0.134462	3
0.17	0.131306	5	169	0.125924	4	18	0.073435	0.125626	3
0.18	0.124703	5	296	0.117537	3	9	0.067085	0.117360	3
0.19	0.117790	6	119	0.109226	3	18	0.060632	0.108758	3
0.2	0.110548	6	87	0.100896	4	16	0.054084	0.100317	3
0.21	0.102297	5	19	0.092356	4	18	0.047457	0.092094	3
0.22	0.096665	6	65	0.085141	4	10	0.040722	0.084512	3
0.23	0.090323	5	61	0.078203	4	17	0.033816	0.077385	3
0.24	0.082835	6	107	0.070598	5	15	0.026773	0.070495	3
0.25	0.075934	6	273	0.065773	5	17	0.019439	0.064447	3
0.26	0.068843	4	21	0.058947	6	10	0.011720	0.058788	3
0.27	0.062462	5	435	0.052699	5	18	0.003578	0.052066	3
0.28	0.045585	4	2	0.042977	5	9	-0.005682	0.041347	2
0.29	0.020811	5	8	0.018251	4	17	-0.025939	0.017957	3
0.3	-0.001541	5	26	-0.002739	4	7	-0.047873	-0.003724	3

Table 5 Runtimes, objective values as well as number of assets in the optimal portfolios (A) for the iTraxx examples (10.000 scenarios). Additional information on the CVaR versions of the problems is given.

The results on this data sets for $\alpha = 0.05$ and r ranging from 0.15 to 0.3 are depicted in Figure 3 and detailed in Table 5. While the DCA is clearly outperformed in terms of runtime, due to the extreme losses present in the scenarios we observe that the benchmark algorithm performs worse in terms of the objective value.

To facilitate a comparison with optimal portfolios with respect to the Conditional Value-at-Risk criterion, which is closely related to both of the solution methods as well as to the Value-at-Risk itself (see Section 1), we also provide the respective results in Table 4 and 5.

4.4 Portfolios & Diversification

When examining the results from sections 4.1 - 4.3 it can be observed that the portfolios yield a significantly higher VaR, then any of the single assets, i.e. that risk optimal portfolios can only be achieved by diversification. In the example of Section 4.1 the highest Value-at-Risk of a single asset is 0.9773 and the most conservative Value-at-Risk portfolio constructed with the DCA reaches a Value-at-Risk of 0.983 and contains three of the five available assets. The return of 1.002 is significantly higher, than the return of the least risky asset (the S&P 100).

The situation is similar for the example considered in Section 4.2. The highest Value-at-Risk (at level $\alpha = 0.1$) of a single asset in the considered data set is 0.98347 with a expected daily return of 0.99985. The typical portfolios found by the DCA in this setting consist of roughly 20 assets.

Although the number of assets with positive weights is often small relative to the number of available assets, we observe that the number of assets chosen is consistently larger than the number of assets in the CVaR optimal portfolios (see Tables 4 and 5). Hence it is fair to say that in the setting of this paper the theoretical shortcomings of Value-at-Risk do not lead to nonsensical portfolio decisions like extremely concentrated portfolios (which can be observed in certain examples, see [29] or [8]) and that diversification allows to construct portfolios which are significantly less risky than the individual assets.

Another conclusion that can be drawn from the results in Section 4.3 is that the optimal Conditional Value-at-Risk differs significantly from the optimal Value-at-Risk, hence the replacement of the Value-at-Risk by the Conditional Value-at-Risk, as sensible as it might seem from the viewpoint of the axiomatic theory of risk measures, leads to more *conservative* portfolios and possibly to a competitive disadvantage.

5 Conclusion

In this paper we reviewed the D.C. formulation of the Value-at-Risk functional presented in Wozabal et al [43]. We used the representation to formulate a classical Markowitz type portfolio selection problem with a VaR constraint. This problem in turn is approximately solved by the DCA, a generic approximate solution technique for unconstrained D.C. problems. We demonstrated that the DCA algorithm yields good results by comparing the solutions obtained by the DCA with the global solutions. In out computational studies 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 S&P 500 assets for the years 2007 and 2008 (i.e. the 491 assets and 503 data points) as well as on two data sets from credit risk applications with 100,000 and 10,000 scenarios. We show that for settings were the tails of the return distributions are sufficiently heavy the proposed algorithm outperforms a popular heuristic optimization algorithm for the Value-at-Risk. However, it also has to be mentioned that for very large data sets the computation times for the DCA become prohibitively long.

However, the DCA yields good results in acceptable time for moderately sized problems (up to approximately 10,000 scenarios) and is easy to implement and has nice convergence properties makes the algorithm attractive to financial institutions actively managing the Value-at-Risk of their portfolios.

Acknowledgements The author wants to express his thanks to Prof. Georg Ch. Pflug for his valuable input throughout the work on this paper and to two anonymous referees whose comments led to a significant improvement of the paper.

References

1. An LTH, Tao PD (2001) A continuous approach for globally solving linearly constrained quadratic zero-one programming problems. Optimization 50(1-2):93–120
2. An LTH, Tao PD (2003) 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, Kluwer Acad. Publ., Norwell, MA, pp 285–302
3. An LTH, Tao PD (2005) 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
4. An LTH, Tao PD, Muu LD (1998) A combined D. C. optimization – ellipsoidal branch-and-bound algorithm for solving nonconvex quadratic programming problems. J Comb Optim 2(1):9–28

5. An LTH, Tao PD, Muu LD (1999) Exact penalty in d.c. programming. Vietnam J Math 27(2):169–178
6. Andersson F, Mausser H, Rosen D, Uryasev S (2001) Credit risk optimization with conditional value-at-risk criterion. Mathematical Programming 89(2, Ser. B):273–291
7. Artzner P, Delbaen F, Eber JM, Heath D (1997) Thinking coherently. Risk 10(11):68–71
8. Artzner P, Delbaen F, Eber JM, Heath D (1999) Coherent measures of risk. Mathematical Finance 9:203–228
9. Basak S, Shapiro A (2001) Value-at-risk managemant: Optimal policies and asset prices. Review of Financial Studies 14(2):371–405
10. Benati S, Rizzi R (2007) A mixed integer linear programming formulation of the optimal mean/value-at-risk portfolio problem. European Journal of Operational Research 176(1):423–434
11. Campbell R, Huisman R, Koedijk K (2001) Optimal portfolio selection in a value-at-risk framework. Journal of Banking & Finance 25(9):1789–1804
12. Cheon MS, Ahmed S, Al-Khayyal F (2006) A branch-reduced-cut algorithm for the global optimization of probabilistically constrained linear programs. Mathematical Programming 108(2-3, Ser. B):617–634
13. Conn A, Gould N, Toint PL (2000) Trust-region methods. MPS/SIAM Series on Optimization, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA

14. Durier R (1988) On locally polyhedral convex functions. In: Trends in mathematical optimization (Irsee, 1986), Internat. Schriftenreihe Numer. Math., vol 84, Birkhäuser, Basel, pp 55–66
15. Emmer S, Klüppelberg C, Korn R (2001) Optimal portfolios with bounded capital at risk. Mathematical Finance 11(4):365–384
16. Fbin CI (2008) Handling cvar objectives and constraints in two-stage stochastic models. European Journal of Operational Research 191(3):888 – 911
17. Föllmer H, Schied A (2004) Stochastic finance, de Gruyter Studies in Mathematics, vol 27, extended edn. Walter de Gruyter & Co., Berlin, an introduction in discrete time
18. Gaivoronski AA, Pflug GC (2005) Value-at-risk in portfolio optimization: properties and computational approach. Journal of Risk 7(2):1–31
19. Gilli M, Këllezi E (2002) 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, Kluwer, pp 167–183
20. Gilli M, Këllezi E, Hysi H (2006) A data-driven optimization heuristic for downside risk minimization. The Journal of Risk 8(3):1–18
21. Hochreiter R (2007) An evolutionary computation approach to scenario-based risk-return portfolio optimization for general risk measures. In: EvoWorkshops 2007, Springer, Lecture Notes in Computer Science, vol 4448, pp 199–207
22. Hochreiter R, Wozabal D (2009) Evolutionary approaches for estimating a coupled markov chain model for credit portfolio risk management. In: EvoWorkshops 2007, Springer, Lecture Notes in Computer Science, vol 5488, pp 193–202

23. Horst R, Tuy H (1993) Global optimization, 2nd edn. Springer-Verlag, Berlin
24. for International Settlements B (2004) Basel II: International convergence of capital measurements and capital standards. Publication 107, Basel Committee on Banking Supervision
25. Jorion P (2000) Value at Risk: The New Benchmark for Controlling Market Risk. McGraw-Hill, New York
26. Kaniovski YM, Pflug GC (2007) Risk assessment for credit portfolios: A coupled markov chain model. Journal of Banking & Finance 31(8):2303–2323
27. Larsen N, Mausser H, Uryasev S (2002) Algorithms for optimization of value-at-risk. In: Pardalos P, Tsitsiringos V (eds) Financial Engineering, e-Commerce and Supply Chain, Kluwer Academic Publishers, Dordrecht, Netherlands, pp 129–157
28. Markowitz HM (1952) Portfolio selection. The Journal of Finance 7(1):77–91
29. McNeil AJ, Frey R, Embrechts P (2005) Quantitative risk management. Princeton Series in Finance, Princeton University Press
30. Natarajan K, Pachamanova D, Sim M (2008) Incorporating asymmetric distributional information in robust value-at-risk optimization. Manage Sci 54(3):573–585
31. Pang J, Leyffer S (2004) On the global minimization of the value-at-risk. Optimization Methods and Software 19(5):611–631
32. Pflug GC (2000) Some remarks on the Value-at-Risk and the Conditional Value-at-Risk. In: Probabilistic constrained optimization, Nonconvex Optimization and its Applications, vol 49, Kluwer Academic Publishers, Dordrecht, Netherlands,

pp 272–281

33. Pflug GC, Römisch W (2007) Modeling, Measuring and Managing Risk. World Scientific, Singapore
34. Pirvu TA (2007) Portfolio optimization under the value-at-risk constraint. Quantitative Finance 7(2):125–136
35. Puelz A (2001) Value-at-risk based portfolio optimization. In: Uryasev S, Pardalos P (eds) Stochastic Optimization: Algortihms and Applications, Kluwer Academic Publishers
36. Rockafellar RT, Uryasev S (2000) Optimization of Conditional Value-at-Risk. The Journal of Risk 2(3):21–41
37. Tao PD, An LTH (1998) A d. c. optimization algorithm for solving the trust-region subproblem. SIAM Journal on Optimization 8(2):476–505
38. Tao PD, El Bernoussi S (1988) Duality in D.C. (difference of convex functions) optimization. Subgradient methods. In: Trends in mathematical optimization (Irsee, 1986), ISNM, vol 84, Birkhäuser, Basel, pp 277–293
39. Tasche D, Tibiletti L (2004) Approximations for the value-at-risk approach to risk-return analysis. The ICFAI Journal of Financial Risk Management 1(4):44–61
40. Thi HAL, Pham DT, Thoai NV (2002) Combination between global and local methods for solving an optimization problem over the efficient set. European Journal of Operational Research 142(2):258 – 270, DOI 10.1016/S0377-2217(01)00301-0

41. Uryasev S (2000) Conditional Value-at-Risk: Optimization algorithms and applications. Financial Engineering News 14:1–5
42. Vehvilinen I, Keppo J (2003) Managing electricity market price risk. European Journal of Operational Research 145(1):136 – 147, DOI 10.1016/S0377-2217(01)00399-X
43. Wozabal D, Hochreiter R, Pflug G (2010) A d.c. formulation of value-at-risk constrained optimization. Optimization 59(3)