Handling CVaR objectives and constraints in two-stage stochastic models

Csaba I. Fábián*

Abstract

Based on the polyhedral representation of Künzi-Bay and Mayer (2006), we propose decomposition frameworks for handling CVaR objectives and constraints in two-stage stochastic models.

For the solution of the decomposed problems we propose special Level-type methods.

Keywords: Stochastic programming; Finance; Convex programming; Decomposition methods

Introduction

Value-at-Risk (VaR) is a widely accepted risk measure. Given a probability $\alpha$, VaR answers the question: what is the maximum loss with the confidence level $\alpha \times 100\%$? Using stochastic programming terminology, VaR constraints can be formulated as chance constraints or probabilistic constraints. These constraints were introduced and studied by Charnes et al. (1958) and by Prékopa (1973). Portfolio optimization problems with VaR objectives or constraints are generally hard to solve due to non-convexity. A characteristic cause of difficulty is that the technological coefficients are random since they represent returns of the different assets. For the special case when the asset returns have joint normal distribution, Kataoka (1963) proved a convexity result, and Prékopa (1974) a more general one for joint constraints.

Prékopa (1973) introduced constraints on conditional expectations, and proved tractability of such constraints under certain conditions. He also considered the case of random technological coefficients. In terms of portfolio optimization, he defined a specially normalized loss function (loss divided by standard deviation of yield). He proved that under a normally distributed return vector, a constraint on the conditional expectation of this normalized loss function is tractable.

The Conditional Value-at-Risk (CVaR) measure involves both an $\alpha$-quantile and conditional expectation. It is the conditional mean value of the worst $(1 - \alpha) \times 100\%$ losses. CVaR has been proposed by Rockafellar and Uryasev (2000). They derived a representation of CVaR as the optimum of a special minimization problem. This representation proves that CVaR is tractable under general circumstances. In case of discrete finite distributions, CVaR optimization problems can be formulated as linear programming problems. CVaR is gaining field in financial applications. Andersson et al. (2001) examined one-stage CVaR models for credit risk optimization. In one model type they minimized CVaR under a constraint on expected return, in another they constructed the risk/return efficient frontier.

An overview of VaR and CVaR optimization models and solution methods can be found in Prékopa (2003) and in Kall and Mayer (2005). Dynamic versions of the above mentioned models have also been developed:

The two-stage model with probabilistic and conditional expectation constraints for the solvability of the second-stage problem was first introduced and studied by Prékopa (1973).

*Department of Operations Research, Loránd Eötvös University of Budapest. P.O.Box 120, Budapest, H-1518 Hungary. E-mail: fabian@cs.elte.hu. Fax: +36 (1) 381 2199. This research has been supported by OTKA, Hungarian National Fund for Scientific Research, project No. 47340.
Krokhmal, Palmquist, and Uryasev (2002) and Rockafellar and Uryasev (2002) introduced the idea of using CVaR in dynamic models. Krokhmal, Palmquist, and Uryasev (2002) is also the first paper dealing with CVaR constraints. The authors observe that portfolio optimization with multiple CVaR-constraints for different time frames and at different confidence levels allows the shaping of distributions according to the decision maker’s preferences.

Topaloglou (2004) and Topaloglou, Vladimirou, Zenios (2006) developed elaborate multi-stage financial models. The objective was minimization of end-of-horizon CVaR under a constraint on expected return. The problems were formulated as linear programming problems. The authors solved one-stage and two-stage problems using general-purpose solvers. Their experimental results confirm the intuition that the two-stage model should produce superior results owing to its additional flexibility to incorporate rebalancing decisions at an intermediate stage.

Research for exploiting the special structure of CVaR-optimization problems started recently: Künzi-Bay and Mayer (2006) proposed a polyhedral representation of CVaR, and on the basis of this, developed a special method for the minimization of CVaR in one-stage stochastic problems. They implemented the method and their experimental results show the clear superiority of their approach over general-purpose methods.

Ahmed (2006) examined the complexity of mean/risk stochastic programming under different risk measures. He proved that the problem is tractable with Quantile-deviation as risk measure, and hence with CVaR as a special case. He proposed a decomposition scheme and a parametric cutting-plane algorithm to generate the efficient frontier. His approach focuses on classic two-stage models having the decision/observation/decision pattern.

In this paper we consider two-stage CVaR-minimization and CVaR-constrained problems having the decision/observation/decision/observation pattern. We propose decomposition and solution schemes based on the Künzi-Bay – Mayer polyhedral representation of CVaR. The solution methods are special Level-type methods. The decomposition schemes and the solution methods proposed in this paper are different from those suggested by Ahmed.

In Section 1 we cite the representation result of Rockafellar and Uryasev. In Section 2 we outline the polyhedral representation proposed by Künzi-Bay and Mayer.

In Section 3 we present a two-stage prototype model that prescribes the minimization of the CVaR of the end-of-horizon yield. From an algorithmic point of view, the two-stage models of Topaloglou (2004) and Topaloglou, Vladimirou, Zenios (2006) fit this prototype model. We present a decomposition scheme for the two-stage problem, and show that the master problem can be solved by an inexact version of the Level Method of Lemaréchal, Nemirovskii, and Nesterov (1995). The inexact version was proposed by Fábián (2000).

In Section 4 we propose an improvement on the prototype model: we penalize first-period risk. Experimental results of Jobst and Zenios (2001) confirm the need for such modification. In Section 5 we formulate a two-stage model with risk constraint. (Decision makers may interpret and quantify right-hand sides of constraints easier than penalties in the objective function.)

We decompose the constrained problem in such a manner that we obtain a two-stage stochastic programming problem having relatively complete recourse. (Relatively complete recourse means that no extra action is needed to enforce feasibility of the second-stage problems.) The first-stage problem can again be solved by the Inexact Level Method. For the solution of the second-stage problem we need a sharp-constrained version of the Constrained Level Method of Lemaréchal, Nemirovskii, and Nesterov (1995). This method is introduced in the Appendix.

1 Conditional value-at-risk. Formulas for discrete distributions

Let us consider a one-period financial investment.

\( w \) denotes the total wealth at the end of the examined period. This is a random variable.
$w^R$ denotes a benchmark for end-of-period wealth (i.e., the wealth that we intend to accumulate by the end of the examined period). We assume it is a parameter that has been set by the decision maker.

Then the loss relative to the benchmark can be expressed as: $w^R - w$. Given a probability $\alpha$, a heuristic definition of the risk measures is the following.

$\alpha$-Value-at-Risk (VaR) answers the question: what is the maximum loss with the confidence level $\alpha \ast 100\%$?

$\alpha$-Conditional Value-at-Risk (CVaR) is the (conditional) mean value of the worst $(1 - \alpha) \ast 100\%$ losses.

Rockafellar and Uryasev (2000) proved that $\alpha$-CVaR can be computed as the optimal objective value of the following problem:

$$\min_{z \in \mathbb{R}} z + \frac{1}{1 - \alpha} \mathbb{E} \left( [w^R - w - z]^+ \right). \hspace{1cm} (1)$$

$\alpha$-VaR is the optimal value of $z$ in case the problem has a unique solution. Generally, the set of the optimal solutions of (1) is a closed interval whose left endpoint is $\alpha$-VaR.

In this paper we assume that $w$ has a discrete distribution. Let the realizations be $w^{(1)}, \ldots, w^{(N)}$ with probabilities $p_1, \ldots, p_N$, respectively. Problem (1) takes the form

$$\min_{z \in \mathbb{R}} z + \frac{1}{1 - \alpha} \sum_{j=1}^N p_j [w^R - w^{(j)} - z]^+. \hspace{1cm} (2)$$

Rockafellar and Uryasev (2002) showed (in their Proposition 8) that the above problem can be solved by just sorting the values $w^{(j)}$. We are going to reach the same conclusion through a different approach.

Rockafellar and Uryasev (2000) proposed transforming (2) into a linear programming problem by introducing new variables $y_j$ ($j = 1, \ldots, N$):

$$\min \hspace{0.2cm} z + \frac{1}{1 - \alpha} \sum_{j=1}^N p_j y_j$$

such that

$$z + y_j \geq w^R - w^{(j)}, \quad y_j \geq 0 \hspace{0.5cm} (j = 1, \ldots, N). \hspace{1cm} (3)$$

The dual of (3) can be written as

$$\max \hspace{0.2cm} w^R - \frac{1}{1 - \alpha} \sum_{j=1}^N \pi_j w^{(j)}$$

such that

$$0 \leq \pi_j \leq p_j \hspace{0.5cm} (j = 1, \ldots, N),$$

$$\sum_{j=1}^N \pi_j = 1 - \alpha. \hspace{1cm} (4)$$

(In the objective function, we used $\frac{1}{1 - \alpha} \sum_{j=1}^N \pi_j w^R = w^R$ that is a consequence of the constraint $\sum_{j=1}^N \pi_j = 1 - \alpha$.)

Apart from lower and upper bounds for individual variables, problem (4) has but a single constraint. Hence it can be solved without using linear programming algorithms. We only need sorting the objective
coefficients $w^{(j)}$; assume that $w^{(1)} \leq \cdots \leq w^{(j)} \leq \cdots \leq w^{(N)}$ holds. Let $J \in \{1, \ldots, N-1\}$ be such that

$$P_J := \sum_{j=1}^{J-1} p_j < 1 - \alpha \quad \text{and} \quad \sum_{j=1}^{J} p_j \geq 1 - \alpha.$$ 

The optimal dual solution is

$$\hat{\pi}_j = \begin{cases} 
  p_j & \text{for } 1 \leq j < J, \\
  (1 - \alpha) - P_J & \text{for } j = J, \\
  0 & \text{for } J < j \leq N.
\end{cases} \quad (5)$$

**Remark 1** The dual problem (4) clearly expresses the (conditional) mean value of the worst $(1 - \alpha) \times 100\%$ losses. Indeed,

- the dual variable $0 \leq \pi_j \leq p_j$ can be interpreted as the weight of the $j$th scenario in an event $\mathcal{E}$,
- the constraint $\sum_{j=1}^{N} \pi_j = 1 - \alpha$ determines the probability of the event $\mathcal{E}$,
- the term $\frac{1}{1 - \alpha} \sum_{j=1}^{N} \pi_j w^{(j)}$ in the objective function can be interpreted as the conditional expectation of the end-period wealth given that $\mathcal{E}$ occurs. By maximizing the gap between the benchmark $w^B$ and the above conditional expectation, we find the worst $(1 - \alpha) \times 100\%$ cases.

(This is a straightforward proof of the validity of the Rockafellar-Uryasev representation formula (2) in case of discrete distributions.)

The benchmark wealth $w^B$ appears as a constant in the objective function of the dual problem (4). Hence its setting does not affect the respective optimal solutions of either (4) or (3). From a purely mathematical point of view, we could set $w^B = 0$.

## 2 Minimizing CVaR in a one-stage model

The end-of-period wealth $w$ will be the yield of a portfolio selected by the decision maker. Assume there are $n$ assets.

$$x = (x_1, \ldots, x_i, \ldots, x_n)^T$$ represents the amounts of money invested in the different assets at the beginning of the examined period. (I.e., $x$ is a portfolio.)

The sum of the components of $x$ should be equal to the initial capital that we denote by $w_0$. We will formalize this constraint as $1^T x = w_0$. By $1$ we will denote a vector having 1 in each component.

We may impose further constraints on the portfolios (e.g., we may prescribe proportions between different positions or impose lower/upper bounds on certain positions). Let $X \subset \mathbb{R}^n$ denote the set of the feasible portfolios. We assume that $X$ is a convex polyhedron.

$$r = (r_1, \ldots, r_i, \ldots, r_n)^T$$ denotes the returns for different assets. This is a random vector.

There are $N$ realizations,

$$r^{(1)}, \ldots, r^{(j)}, \ldots, r^{(N)},$$ where $$r^{(j)} = \left(r_1^{(j)}, \ldots, r_i^{(j)}, \ldots, r_n^{(j)}\right)^T.$$ The probabilities of the different realizations are $p_1, \ldots, p_N$. 

The total wealth at the end of the examined period is \( w = r^T x \). This is a random variable with realizations \( w(j) = r(j)^T x \).

In this paper we assume that no short positions are allowed, i.e., \( x \geq 0 \) is imposed among the constraints of \( X \). We assume moreover that \( X \) is determined by a homogeneous system of linear inequalities, i.e., it prescribes bounds on proportions between different positions. We also assume non-negative returns \( r(j) \geq 0 \) (1 \( \leq j \leq N \)). (These are technical assumptions that we need for the convergence proof of the proposed methods – see Remark 6 referenced in Remarks 8 and 12 for the convergence of Inexact Level Method; and Remark 11 for the convergence of the Sharp-Constrained Level Method. The decomposition schemes work without these assumptions, and the methods can be implemented under milder assumptions.)

Using (2) we can compute CVaR as a function of \( x \):

\[
C(x) := \min_{z \in \mathbb{R}} z + \frac{1}{1-\alpha} \sum_{j=1}^{N} p_j \left[ w^B - r(j)^T x - z \right]^+.
\]

\( C(x) \) is a polyhedral convex function.

Suppose we want to find a balance between the end-of-period expected wealth \( E(r)^T x \) and the risk as measured by \( C(x) \). A customary formulation of the problem is:

\[
\min c^T x + C(x)
\]

\[
such that
\]

\[
x \in X, \; 1^T x = w_0,
\]

where \( c := -\lambda^{-1} E(r) \) with a parameter \( \lambda > 0 \) that is interpreted as risk aversion of the decision maker. We assume a known fixed \( \lambda \).

Künzi-Bay and Mayer (2006) observe that problems of minimizing CVaR fit the prototype of the two-stage stochastic programming problem. We sketch their approach for the special problem (7). The master problem is:

\[
\min c^T x + z + \frac{1}{1-\alpha} \sum_{j=1}^{N} p_j C_j(x, z)
\]

\[
such that
\]

\[
z \in \mathbb{R}, \; x \in X, \; 1^T x = w_0.
\]

The function value \( C_j(x, z) \) is defined by the second-stage stage problem:

\[
C_j(x, z) := \min y
\]

\[
such that \; y \geq w^B - r(j)^T x - z, \; y \geq 0.
\]

The dual of this second-stage problem has a single variable and takes the form: \( \max_{0 \leq u \leq 1} \left( w^B - r(j)^T x - z \right) u \).

The optimal solution is either \( u = 1 \) or \( u = 0 \) depending on the sign of the objective coefficient \( w^B - r(j)^T x - z \).

On the basis of these observations, Künzi-Bay and Mayer propose an interesting equivalent formulation of
the CVaR-minimization problem (7):

$$\min c^T x + z + \frac{1}{1-\alpha} v$$

such that

$$z, v \in \mathbb{R}, \quad x \in X, \quad 1^T x = w_0,$$

$$\sum_{j \in J} p_j \left( w^B - r^{(j)} T x - z \right) \leq v \quad (J \subset \{1, \ldots, N\}).$$

(8)

A subset $J \subset \{1, \ldots, N\}$ represents the possibility that we have 1 as optimal dual solution for the second-stage problems belonging to $j \in J$; and 0 for those belonging to $j \notin J$. The cut belonging to the empty set $J = \emptyset$ just prescribes the non-negativity of $v$. (Künzi-Bay and Mayer observe that the formulation (8) is the CVaR-analogue of a polyhedral representation result of Klein Haneveld and Van der Vlerk (2006).)

Künzi-Bay and Mayer adapt the L-shaped method to their special two-stage problem. They use cuts of the type that appear in the formulation (8).

An alternative method for the solution of the CVaR-minimization problem (7) is obtained by computing the CVaR function using the dual formulation (4):

$$C(x) = \max w^B - \frac{1}{1-\alpha} \sum_{j=1}^{N} \pi_j r^{(j)} T x$$

such that

$$0 \leq \pi_j \leq p_j \quad (j = 1, \ldots, N),$$

$$\sum_{j=1}^{N} \pi_j = 1 - \alpha.$$  

(9)

Let $(\hat{\pi}_1, \ldots, \hat{\pi}_N)$ denote a feasible solution of problem (9). Then the linear function

$$l(x) := w^B - \frac{1}{1-\alpha} \sum_{j=1}^{N} \hat{\pi}_j r^{(j)} T x$$

obviously satisfies $l(x) \leq C(x) \quad (x \in \mathbb{R}^n)$.

Given $\hat{x} \in \mathbb{R}^n$, let $(\hat{\pi}_1, \ldots, \hat{\pi}_N)$ denote now an optimal solution of problem (9: $x = \hat{x}$). Then the linear function $l(x)$ constructed according to (10) will be a support function of $C(x)$ at $\hat{x}$. As shown in Section 1, problem (9: $x = \hat{x}$) can be solved by just sorting the potential yields $r^{(j)} T \hat{x}$ ($1 \leq j \leq N$). Easy computation of the support functions suggests a direct cutting-plane or bundle-type approach for the solution of the CVaR-minimization problem (7).

Remark 2 The above representation yields an easily computable Lipschitz constant for the polyhedral convex function $C(x)$. Namely,

$$R := \max_{1 \leq j \leq N} \| r^{(j)} \|.$$  

Indeed, let $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^n$. Suppose that we have $C(\mathbf{x}_1) \geq C(\mathbf{x}_2)$. Let $\hat{x} := \mathbf{x}_1$, and let $l(x)$ in the form of (10) be a support function of $C(x)$ at $\hat{x}$. Then we have

$$|C(\mathbf{x}_1) - C(\mathbf{x}_2)| \leq |l(\mathbf{x}_1) - l(\mathbf{x}_2)| \leq \frac{1}{1 - \alpha} \sum_{j=1}^{N} \hat{\pi}_j \left| r^{(j)} T (\mathbf{x}_1 - \mathbf{x}_2) \right| \leq \frac{1}{1 - \alpha} \sum_{j=1}^{N} \hat{\pi}_j R \| \mathbf{x}_1 - \mathbf{x}_2 \|.$$
In the right-hand-side expression, we have \( \sum_{j=1}^{N} \hat{\pi}_j = 1 - \alpha \).

In order to show similarity of the above alternative approach to the approach proposed by Künzi-Bay and Mayer, let us assume for the remaining part of this section that we have \( p_1 = \ldots = p_N = \frac{1}{N} \). Assume moreover that \((1 - \alpha)\) is an integer multiple of \( \frac{1}{N} \). Hence \( \frac{1 - \alpha}{1/N} = N(1 - \alpha) \) is an integer. The corresponding special form of (9) will be denoted by (9: \( p = \frac{1}{N} \)).

**Proposition 3** The following polyhedral representation is equivalent to problem (9: \( p = \frac{1}{N} \)).

\[
\begin{align*}
\min_{\nu} & \quad \nu \\
\text{such that} & \quad \nu \in \mathbb{R}, \\
& \quad w^B - \frac{1}{N(1-\alpha)} \sum_{j \in J} r^{(j)T} x \leq \nu \quad (J \subset \{1, \ldots, N\}, |J| = N(1 - \alpha)).
\end{align*}
\]  

(11)

**Proof.** Given a set \( \hat{J} \subset \{1, \ldots, N\} \), \(|\hat{J}| = N(1 - \alpha)\) that represents a cut in (11), let

\[
\hat{\pi}_j := \begin{cases} \\
1/N & \text{if } j \in \hat{J}, \\
0 & \text{otherwise}, \\
\end{cases} \quad (1 \leq j \leq N).
\]

This is a feasible solution of (9: \( p = \frac{1}{N} \)), hence \( l(x) \leq C(x) \quad (x \in \mathbb{R}^n) \) holds with the linear function \( l(x) \) constructed after (10).

Conversely, given \( \hat{x} \in \mathbb{R}^n \), let us construct a linear support function \( l(x) \) of \( C(x) \) at \( \hat{x} \). As we saw above, it can be done according to (10) where \((\hat{\pi}_1, \ldots, \hat{\pi}_N)\) is an optimal solution of the problem (9: \( p = \frac{1}{N}, x = \hat{x} \)).

We saw in Section 1 that an optimal solution \((\hat{\pi}_1, \ldots, \hat{\pi}_N)\) can be found in the special form:

either \( \hat{\pi}_j = 1/N \) or \( \hat{\pi}_j = 0 \) \( \quad (1 \leq j \leq N).\)

The number of the positive components is \( N(1 - \alpha) \). With this optimal solution, let

\[
\hat{J} := \{ j \mid 1 \leq j \leq N, \hat{\pi}_j = 1/N \}.
\]

The corresponding cut is valid in (11). \( \Box \)

**Corollary 4** The following polyhedral representation is equivalent to problem (7):

\[
\begin{align*}
\min_{c^T x + \nu} & \\
\text{such that} & \quad \nu \in \mathbb{R}, \quad x \in X, \quad 1^T x = w_0, \\
& \quad \frac{1}{N(1-\alpha)} \sum_{j \in J} \left( w^B - r^{(j)T} x \right) \leq \nu \quad (J \subset \{1, \ldots, N\}, |J| = N(1 - \alpha)).
\end{align*}
\]  

(12)

(In the last constraint, we used \( w^B = \frac{1}{N(1-\alpha)} \sum_{j \in J} w^B \) that is a consequence of \(|J| = N(1 - \alpha)|.\)
Let us consider now the polyhedral representation (8) proposed by Künzi-Bay and Mayer. In order to adapt it to the supposed case of approximate distribution, let us substitute \( p_j = \frac{1}{N} \) \( (j = 1, \ldots, N) \). Moreover, let us introduce a new variable \( \nu := z + \frac{1}{N} v \) instead of \( v \). We get:

\[
\min c^T x + \nu
\]

such that

\[
z, \nu \in \mathbb{R}, \quad x \in X, \quad 1^T x = w_0,
\]

\[
\frac{1}{N(1-\alpha)} \sum_{j \in J} \left( w^B - r(j)^T x \right) - \left( \frac{|J|}{N(1-\alpha)} - 1 \right) z \leq \nu \quad (J \subset \{1, \ldots, N\}).
\] (13)

There is a strong similarity between the representations (13) and (12). In (13) the cuts are generated in the space \( \mathbb{R}^{n+2} \ni (x, z, \nu) \). In (12) the cuts are generated in the space \( \mathbb{R}^{n+1} \ni (x, \nu) \). Equivalence of the two representations mean that the feasible region of (12) is a projection of the feasible region of (13).

Künzi-Bay and Mayer implemented their special decomposition method. They solved several CVaR-minimization test problems with their experimental solver called CVaRMin. They also solved the test problems with general-purpose LP solvers. These solvers were used to solve the LP-equivalents of the relevant two-stage stochastic programming problems. (The LP-equivalent form contains the constraints of the first-stage problem, and \( 2N \) additional constraints. These are individual constraints for each scenario, as opposed to the aggregate cuts that appear in the representation (8).) Additionally, they also solved the test problems in two-stage recourse forms, by employing a benchmark stochastic problem solver. Their experimental results show the clear superiority of the solver CVaRMin in case of CVaR-problems. For the largest test problems, CVaRMin was by at least one order of magnitude faster than either of the other solvers involved.

Its simplicity makes the representation (12) an attractive alternative in case of one-stage problems. For the decomposition of two-stage models, however, we need the more detailed representation of Künzi-Bay and Mayer:

### 2.1 A parametric version of the polyhedral representation (8)

Initial wealth will be a considered a parameter and denoted by \( \omega \geq 0 \). Let us moreover use a parameter \( \zeta \) instead of the decision variable \( z \).

\[
D(\omega, \zeta) := \min c^T x + \frac{1}{1-\alpha} v
\]

such that

\[
x \in X, \quad v \in \mathbb{R},
\]

\[
1^T x = \omega,
\]

\[
\sum_{j \in J} p_j \left( w^B - r(j)^T x - \zeta \right) \leq v \quad (J \subset \{1, \ldots, N\}).
\] (14)

The cutting plane method proposed by Künzi-Bay and Mayer can be easily adapted to the above problem:

0. **Initialize.**

- Set the stopping tolerance \( \epsilon \).
- The initial master problem contains the constraints \( x \in X, \ 1^T x = \omega \);
- and the cut belonging to the set \( J_0 := \emptyset \). (This cut enforces non-negativity of \( v \).)
- Set the iteration counter \( \iota := 1 \).
1. Solve master problem.

Let \((x^*, v^*)\) be an optimal solution of the current cutting-plane model problem. Let

\[ J_\iota := \{ j \mid w^B - r^{(j)T} x^* - \zeta > 0 \}. \]

2. Check for optimality

If

\[ \sum_{j \in J_\iota} p_j \left( w^B - r^{(j)T} x^* - \zeta \right) - v^* \leq (1 - \alpha) \epsilon, \]

then \(x^*\) is an \(\epsilon\)-optimal solution; stop.

3. Append cut

Append the cut belonging to the set \(J_\iota\).
Increment \(\iota\), and repeat from step 1.

Alternatively, the problem can be solved by a bundle-type method.

**Remark 5** In the original method, Künzi-Bay and Mayer used a relative stopping tolerance. For the decomposition method we are going to describe in Section 3.2, we need an absolute stopping tolerance. Moreover, in the initialization step, Künzi-Bay and Mayer set \(J_0 := \{1, \ldots, N\}\). We modified this to \(J_0 := \emptyset\) because the latter setting enforces non-negativity of the variable \(v\). We will need non-negativity in the discussion of Remark 6.

Suppose that after \(\kappa\) iterations, the cutting-plane model problem looks like this:

\[
D_\kappa(\omega, \zeta) := \min c^T x + \frac{1}{1 - \alpha} v \\
\text{such that} \\
x \in X, \quad v \in \mathbb{R}, \\
1^T x = \omega, \\
\sum_{j \in J_\iota} p_j \left( w^B - r^{(j)T} x - \zeta \right) \leq v \quad (\iota = 0, \ldots, \kappa).
\]

Given the parameter values \(\hat{\omega}, \hat{\zeta}\), let us solve the linear programming problem (15: \(\omega = \hat{\omega}, \zeta = \hat{\zeta}\)). A linear support function \(L_\kappa(\omega, \zeta)\) of \(D_\kappa(\omega, \zeta)\) at \((\hat{\omega}, \hat{\zeta})\) can be constructed using the optimal dual variables. The function \(D_\kappa(\omega, \zeta)\) is obviously a polyhedral convex lower approximation of the convex function \(D(\omega, \zeta)\). Hence we have \(L_\kappa(\omega, \zeta) \leq D_\kappa(\omega, \zeta) \leq D(\omega, \zeta)\).

Given \(\hat{\omega}\) and \(\hat{\zeta}\), suppose the solution actually terminated with the cutting-plane model problem (15: \(\omega = \hat{\omega}, \zeta = \hat{\zeta}\)). It means that \(D(\hat{\omega}, \hat{\zeta}) - D_\kappa(\hat{\omega}, \hat{\zeta}) \leq \epsilon\) holds. Then the linear support function \(L_\kappa(\omega, \zeta)\) constructed from the optimal dual variables of the final approximate problem, is also an \(\epsilon\)-support function of \(D(\omega, \zeta)\) at \((\hat{\omega}, \hat{\zeta})\).

To summarize the above discussion: Given \(\hat{\omega}, \hat{\zeta}\), and a tolerance \(\epsilon\), an \(\epsilon\)-support function of \(D(\omega, \zeta)\) at \((\hat{\omega}, \hat{\zeta})\) can be constructed through the approximate solution of the problem (14: \(\omega = \hat{\omega}, \zeta = \hat{\zeta}\)).

**Remark 6** We can easily construct upper bounds on both the \(\zeta\)-direction and the \(\omega\)-direction slope of the function \(D_\kappa(\omega, \zeta)\).

The parameter \(\zeta\) appears in the right-hand sides of the polyhedral cuts (with positive coefficients \(\leq 1\)). The dual variables corresponding to these cuts are non-positive, and their sum is \(-1/(1 - \alpha)\). It follows that in absolute value, the \(\zeta\)-direction slope of \(D_\kappa(\omega, \zeta)\) is at most \(1/(1 - \alpha)\).
It is easily seen that \( \mathcal{D}_r \) is monotone decreasing in \( \zeta \). Since we assumed non-negative return \( r \geq 0 \), it follows that \( \mathcal{D}_r(\omega, \zeta) \leq 0 \), and it is monotone decreasing in \( \omega \). (Note that we have \( c := -\lambda^{-1}E(r) \).

Due to the assumed homogeneity of \( X \), and the possible setting \( w^0 = 0 \) (see Remark 1), we have \( \mathcal{D}_r(\rho \omega, \rho \zeta) = \rho \mathcal{D}_r(\omega, \zeta) \) for \( \rho > 0 \).

Considering the \( \omega \)-direction slope, let \( \omega < \zeta \). From the above mentioned properties of \( \mathcal{D}_r \), we have

\[
0 \leq \mathcal{D}_r(\omega, \zeta) - \mathcal{D}_r(\omega, \zeta) \leq \mathcal{D}_r \left( \frac{\tilde{\omega}}{\omega}, \frac{\tilde{\zeta}}{\zeta} \right) - \mathcal{D}_r(\omega, \zeta) = \left( \frac{\tilde{\omega}}{\omega} - 1 \right) \mathcal{D}_r(\omega, \zeta).
\]

Moreover, we obviously have

\[
0 \geq \mathcal{D}_r(\omega, \zeta) \geq D_0(\omega, \zeta) \geq \min_{x \geq 0, 1^T x = \omega} -\lambda^{-1}E(r)^T x \geq -\lambda^{-1}\|E(r)\| \omega.
\]

The third inequality is the consequence of the assumption \( X \subset \mathbb{R}_+^n \), and of the non-negativity of \( v \) in problem 15 (see Remark 5).

From the above chains of inequalities it follows that in absolute value, the \( \omega \)-direction slope of \( \mathcal{D}_r(\omega, \zeta) \) is not greater than \( \lambda^{-1}\|E(r)\| \).

3 A two-stage prototype model

Suppose that we wish to plan investments for two time periods. The investment can be restructured at the beginning of each period. We are going to use the following notation:

- \( w_0 \) represents initial capital. This is a given parameter.
- \( x_1 = (x_{11}, \ldots, x_{1i}, \ldots, x_{1n})^T \) represents the portfolio selected at the beginning of the first time period.
  - This is a decision vector. Feasible portfolios are represented by \( x_1 \in X, \ 1^T x_1 = w_0 \).
- \( r_1 = (r_{11}, \ldots, r_{i1}, \ldots, r_{in})^T \) denotes return occurring in course of the first time period. This is a random vector having a known distribution. We assume that each component of \( r_1 \) is non-negative.
- \( w_1 := r_1^T x_1 \) denotes the wealth at the end of the first time period.
- \( x_2 = (x_{21}, \ldots, x_{12}, \ldots, x_{n2})^T \) represents the portfolio selected at the beginning of the second time period.
  - This is a decision vector. Feasible portfolios are represented by \( x_2 \in X, \ 1^T x_2 = w_1 \).
- \( r_2 = (r_{11}, \ldots, r_{12}, \ldots, r_{n2})^T \) denotes return occurring in course of the second time period. This is a random vector. We assume that each component of \( r_2 \) is non-negative, and that the joint distribution of the random vectors \( r_1 \) and \( r_2 \) is known.
- \( w_2 := r_2^T x_2 \) denotes the wealth at the end of the second time period.

Concerning the set \( X \subset \mathbb{R}_+^n \) of feasible portfolios, the assumptions of Section 2 will be imposed for the two-stage model also.

**Description of the random parameters.** Assume there are \( N \) realizations of the random vector \( r_1 \), namely

\[
r_1^{(1)}, \ldots, r_1^{(j)}, \ldots, r_1^{(N)} \quad \text{occurring with probabilities} \quad p_1^{(1)}, \ldots, p_1^{(j)}, \ldots, p_1^{(N)}, \quad \text{respectively.}
\]

Assume, moreover, that the random vector \( r_2 \) given that \( r_1 = r_1^{(j)} \) has a known discrete distribution for each \( 1 \leq j \leq N \). Assume there are \( M_j \) realizations of this random vector, namely

\[
r_2^{(j1)}, \ldots, r_2^{(jk)}, \ldots, r_2^{(jM_j)} \quad \text{occurring with probabilities} \quad p_2^{(j1)}, \ldots, p_2^{(jk)}, \ldots, p_2^{(jM_j)}, \quad \text{respectively.}
\]
Scenarios can be identified by the elements of the set
\[ S := \{ (j, k) \mid 1 \leq j \leq N, 1 \leq k \leq M_j \}, \]
where \((j, k)\) means that the first- and second-period returns \(r_1^{(j)}\) and \(r_2^{(jk)}\) realize. The probability of this event will be denoted by
\[ p_{(jk)} := p_1^{(j)} p_2^{(jk)} \quad ((j, k) \in S). \]
Realizations of the random wealth at the end of the first period will be denoted by
\[ w_1^{(j)} := r_1^{(j)} x_1 \quad (j = 1, \ldots, N). \]
Realizations of the random wealth at the end of the second period will be denoted by
\[ w_2^{(jk)} := r_2^{(jk)} x_2 \quad ((j, k) \in S). \]

**Problem formulation.** Suppose we wish to find a trade-off between the expectation \(E(w_2)\) of the final wealth, and the conditional value-at-risk \(CVaR(w_2)\). This aim is expressed by the objective
\[ \min -\lambda^{-1} E(w_2) + CVaR(w_2), \quad (16) \]
where the risk aversion parameter \(\lambda > 0\) has been determined by the decision maker. Suppose that the benchmark \(w^B\) for the final wealth has also been determined by the decision maker.

The objective of minimizing end-of-horizon \(CVaR\) had been used earlier by Topaloglou (2004) and by Topaloglou, Vladimirou, Zenios (2006). They developed elaborate financial models, even in multi-stage frameworks. They solved one-stage and two-stage problems with commercially available, general-purpose modeling systems and optimization solvers. From an algorithmic point of view, their two-stage models fit the above described prototype. Hence the special decomposition framework and method we are going to describe can also be applied to their financial problems.

Using the polyhedral representation (8) of Künzi-Bay and Mayer, the prototype problem can be formulated as follows:
\[ \min -\lambda^{-1} \sum_{(j, k) \in S} p_{(jk)} w_2^{(jk)} + z + \frac{1}{1-\alpha} v \]
subject to
\[ z, v \in \mathbb{R}, \quad x_1 \in X, \]
\[ w_1^{(j)} \in \mathbb{R}, \quad x_2^{(j)} \in X, \quad (j = 1, \ldots, N), \]
\[ w_2^{(jk)} \in \mathbb{R} \quad ((j, k) \in S), \]
\[ 1^T x_1 = w_0, \]
\[ 1^T x_2^{(j)} = w_1^{(j)} = r_1^{(j)} x_1 \quad (j = 1, \ldots, N), \]
\[ w_2^{(jk)} = r_2^{(jk)} x_2^{(j)} \quad ((j, k) \in S), \]
\[ \sum_{(j, k) \in \mathcal{L}} p_{(jk)} \left( w^B - w_2^{(jk)} - z \right) \leq v \quad (\mathcal{L} \subset S). \]

This is a linear programming problem that may grow to a gigantic size depending on the cardinality of the scenario set \(S\).
3.1 Decomposition of the two-stage prototype problem

The first-stage problem will be:

$$\min \quad z + \sum_{j=1}^{N} p_1^{(j)} D^{(j)} (w_1^{(j)}, z)$$

such that

$$z \in \mathbb{R}, \quad x_1 \in X,$$

$$w_1^{(j)} \in \mathbb{R} \quad (j = 1, \ldots, N),$$

$$1^T x_1 = w_0,$$

$$w_1^{(j)} = r_1^{(j)} x_1 \quad (j = 1, \ldots, N),$$

where the functions $D^{(j)} : \mathbb{R}^2 \to \mathbb{R} \quad (j = 1, \ldots, N)$ are defined by the second-stage problem:

$$D^{(j)}(\omega, \zeta) := \min \quad -\lambda^{-1} \sum_{k=1}^{M_j} p_2^{(jk)} w_2^{(k)} + \frac{1}{1-\alpha} v$$

such that

$$x_2 \in X, \quad v \in \mathbb{R},$$

$$w_2^{(k)} \in \mathbb{R} \quad (k = 1, \ldots, M_j),$$

$$1^T x_2 = \omega,$$

$$w_2^{(k)} = r_2^{(jk)} x_2 \quad (k = 1, \ldots, M_j),$$

$$\sum_{k \in \mathcal{K}} p_2^{(jk)} \left( w_2^B - w_2^{(k)} - \zeta \right) \leq v \quad (\mathcal{K} \subset \{1, \ldots, M_j\}).$$

Theorem 7 The two-stage problem (18 - 19) is equivalent to the polyhedral representation problem (17).

Proof. Let us construct the equivalent linear programming form (ELPF) of the two-stage problem (18 - 19) in the usual manner. It means defining a separate set of second-stage variables for each $j = 1, \ldots, N$. Let

$$x_2^{(j)} \in X, \quad v^{(j)} \in \mathbb{R},$$

$$w_2^{(jk)} \in \mathbb{R} \quad (k = 1, \ldots, M_j)$$

denote the set of the second-stage variables corresponding to $j$. There is a straightforward matching between the variables of the ELPF problem and those of the polyhedral representation problem (17). There is but
one non-trivial case: variable \( v \) of the polyhedral representation problem will correspond to the weighted sum \( \sum_{j=1}^{N} p_{1}^{(j)} v^{(j)} \) of the ELPF-variables \( v^{(j)} \).

The objective function of the ELPF problem is

\[
z + \sum_{j=1}^{N} p_{1}^{(j)} \left\{ -\lambda^{-1} \sum_{k=1}^{M_{j}} p_{2}^{(jk)} w_{2}^{(jk)} + \frac{1}{1 - \alpha} v^{(j)} \right\} = -\lambda^{-1} \sum_{j=1}^{N} \sum_{k=1}^{M_{j}} p_{1}^{(j)} p_{2}^{(jk)} w_{2}^{(jk)} + z + \frac{1}{1 - \alpha} \sum_{j=1}^{N} p_{1}^{(j)} v^{(j)},
\]

and this is obviously equivalent to the objective function of the polyhedral representation problem (17).

Now we show that cuts in the polyhedral representation problem (17) are just aggregates of ELPF-cuts. In preparation for this end, let us select a subset \( K^{(j)} \subset \{ 1, \ldots, M_{j} \} \) for each \( j = 1, \ldots, N \). The corresponding ELPF-cuts are

\[
\sum_{k \in K^{(j)}} p_{2}^{(jk)} \left( w_{2}^{(jk)} - w_{2}^{(jk)} - z \right) \leq v^{(j)}.
\]

(In case of an empty set \( K^{(j)} = \emptyset \), the corresponding cut is \( 0 \leq v^{(j)} \).) Aggregating these ELPF-cuts with the weights \( p_{1}^{(j)} \), we get

\[
\sum_{j=1}^{N} p_{1}^{(j)} \sum_{k \in K^{(j)}} p_{2}^{(jk)} \left( w_{2}^{(jk)} - w_{2}^{(jk)} - z \right) \leq \sum_{j=1}^{N} p_{1}^{(j)} v^{(j)}. \tag{20}
\]

Let

\[
L := \bigcup_{j=1}^{N} \left\{ (j, k) \in S \mid k \in K^{(j)} \right\}.
\]

With this \( L \), the aggregate ELPF-cut (20) takes the form

\[
\sum_{(j, k) \in L} p_{2}^{(jk)} \left( w_{2}^{(jk)} - w_{2}^{(jk)} - z \right) \leq v,
\]

that is a cut in the polyhedral representation problem (17).

In order to show that a single ELPF-cut corresponding to the set \( \hat{K}^{(j)} \) is also valid in the polyhedral representation problem; let us select \( \hat{K}^{(j)} := \emptyset \) for \( j \neq \hat{j} \) in the aggregation.

Conversely, given a cut in the polyhedral representation problem (17), i.e., a subset \( L \subset S \), we can construct the sets

\[
K^{(j)} := \{ k \mid (j, k) \in L \} \quad (j = 1, \ldots, N).
\]

In the manner of (20), let us aggregate the ELPF-cuts corresponding to the above sets. This results just the cut corresponding to \( L \) in the polyhedral representation problem (17). \( \square \)

### 3.2 A solution method for the decomposed prototype problem

Following the idea of Künzi-Bay and Mayer, the two-stage problem (18 - 19) could be solved as a special three-stage stochastic programming problem. We propose an alternative method.

The second-stage problem (19) has the form of the parametric polyhedral representation problem (14). (Indeed, let us substitute in (14)

\[
r := \left( r_{2} \mid r_{1} = r_{1}^{(j)} \right), \quad c := -\lambda^{-1} \operatorname{E} \left( r_{2} \mid r_{1} = r_{1}^{(j)} \right) = -\lambda^{-1} \sum_{k=1}^{M_{j}} p_{2}^{(jk)} r_{2}^{(jk)} \tag{21}
\]

3.2 to obtain (19). )
Having fixed the values \( \hat{z} \), \( \hat{w}^{(j)} \) (\( j = 1, \ldots, N \)) of the first-stage variables, and given a tolerance \( \epsilon \), we can construct \( \epsilon \)-support functions \( L^{(j)}(\omega, \zeta) \) of \( D^{(j)}(\omega, \zeta) \) at the point \( (\omega = \hat{w}_1^{(j)}, \zeta = \hat{z}) \); by the method described in Section 2.1. Then the linear function \( \zeta + \sum_{j=1}^{N} p^{(j)} L^{(j)}(\omega, \zeta) \) will be an \( \epsilon \)-support function of the objective function of the first-stage problem (18), at the point \( (\omega = \hat{w}_1^{(j)}, \zeta = \hat{z}) \).

Using the above described procedure as an oracle that provides information of the first-stage objective function, the first-stage problem (18) can be solved by an inexact version of the Level Method of Lemaréchal, Nemirovskii, and Nesterov (1995). This inexact version was proposed by Fábián (2000).

The Inexact Level Method is a bundle-type method that successively builds a cutting-plane model of the objective function. An oracle is used to provide objective function data, in the form of \( \epsilon \)-support functions. The accuracy tolerance \( \epsilon \) of the cuts is gradually decreased as the optimum is approached. (At each step, we have an estimate of how closely the optimum has been approached. The successive cut is generated with an accuracy tolerance derived from that estimate.) This way we can find a balance between the computational effort of optimization and of approximation.

**Remark 8** To obtain an \( \epsilon \)-optimal solution, the Inexact Level Method needs no more than \( c \left( \frac{D \Lambda}{\epsilon} \right)^2 \) iterations, where \( c \) is a constant that depends only on the parameters of the algorithm, \( D \) is the diameter of the feasible domain, and \( \Lambda \) is a common Lipschitz-constant for the approximate support functions constructed in course of the procedure.

In the present case, we assumed \( X \subset \mathbb{R}_{+}^n \), hence the constraints \( x \in X, \ 1^T x = w_0 \) determine a bounded polyhedron. Bounds for the other variables can be easily found also.

In course of the procedure, approximate support functions are constructed as support functions of the objective function of the approximating problem (15). The slope of this latter function was examined in Remark 6. Formulas given there are easily updated according to (21).

### 3.3 Approximating the efficient frontier

If the decision maker can not determine the value of the risk aversion parameter \( \lambda \), then we can help him by building an approximation of the efficient frontier. Points from this frontier can be found by solving the two-stage problem (18 - 19) with different settings of the parameter \( \lambda \).

Künzi-Bay and Mayer proposed a warm-start procedure for the approximation of the efficient frontier in case of the one-stage problem (7). This procedure can be adapted to the present case.

### 4 An improvement on the two-stage prototype model

Jobst and Zenios (2001) describe an experiment: they solved one-stage portfolio selection problems with time periods of one year. The objective was the minimization of a risk measure. As risk measure, they used MAD in one problem, and CVaR in another problem. They then simulated the returns of the optimal portfolios, at the time points 3, 6, 9, and 12 months after the beginning of the time period.

With the MAD-optimal portfolio, they found that catastrophic losses are probable after the first 3 months. But with the CVaR-optimal portfolio, worst losses were limited to around 10% of the total wealth throughout. However, an interesting phenomenon occurred: the experimental distribution of the 9-month losses had a tail much heavier than the experimental distribution of the end-of-period losses had.

From the above phenomenon we conclude that the two-stage prototype model needs a correction: Beside restricting final risk, we also need restricting the risk at the end of the first period. We may penalize end-of-first-period risk CVaR(\( w_1 \)) by including a new term in the objective function (16). The augmented objective will be:

\[
\min -E(w_2) + \lambda \text{CVaR}(w_2) + \lambda' \text{CVaR}(w_1),
\]

where the additional risk aversion parameter \( \lambda' > 0 \) also needs to be determined by the decision maker.
To include the new term into the prototype problem (17), we can use either the polyhedral representation (8) of Künzi-Bay and Mayer, or the simpler representation (12). The polyhedral cuts are then inherited by the first-stage problem (18). This problem can also be solved by the methods described in Section 3.

If the decision maker cannot determine the values of the parameters $\lambda$ and $\lambda'$, then we can help him by building an approximation of the efficient frontier. In the present case it is a 3-dimensional convex surface. Points from this surface can be found by minimizing the objective function (22) with different settings of the parameters $\lambda$ and $\lambda'$. Though a systematic search in the space of the $(\lambda, \lambda')$ values may prove rather time-consuming.

Alternatively, the decision maker may be able to set constraints on first-period and second-period risk in the form of $\text{CVaR}(w_1) \leq \gamma'$, $\text{CVaR}(w_2) \leq \gamma$. A motivation for using risk constraints is that decision makers interpret and quantify right-hand sides easier than penalties in the objective function. Another motivation is provided by Krokhmal, Palmquist, and Uryasev (2002) who observe that optimization with multiple CVaR-constraints for different time frames and at different confidence levels allows the shaping of distributions according to the decision maker’s preferences.

In the present paper we will treat in detail the handling of the constraint

$$\text{CVaR}(w_2) \leq \gamma. \tag{23}$$

The scheme we propose can be generalized to multiple constraints.

5 Handling a CVaR-constraint

We are going to maximize expected end-of-period return under the risk constraint (23). However, we will formulate the problem in minimization form which is more convenient in convex programming context:

$$\min -E(w_2) \quad \text{such that} \quad \text{CVaR}(w_2) \leq \gamma. \tag{24}$$

The parameter $\gamma$ is set by the decision maker. We assume that the parameter is set in such a manner that the problem is feasible, the Slater condition holds, and (24) is a genuine constrained problem (i.e., the constraint is satisfied with equality in an optimal solution).

The parameter $\gamma$ can be calibrated by exploring the efficient frontier. This is a concave, piecewise linear line in the (risk, return) coordinate system. (For a simple verification of this fact, let us consider (24) a parametric linear programming problem.)

We can easily construct supporting lines to the efficient frontier. Given $\lambda > 0$, a supporting line of slope $\lambda$ can be constructed by solving the unconstrained problem (16). The lower cover of such supporting lines is an upper approximation of the efficient frontier. The convex hull of the touching points gives a lower approximation. In calibrating the model, the decision maker ought to take into consideration the slope of the efficient frontier also. (Slope is interpreted as risk aversion.)

Let $\lambda^* > 0$ be such that the supporting line of slope $\lambda^*$ touches the efficient frontier at a point whose risk-coordinate is $\gamma$. (Of course there may be other touching points beside this.) If we should know $\lambda^*$, we could reduce the constrained problem to the unconstrained problem (16: $\lambda = \lambda^*$).

Although the exact value $\lambda^*$ is not known, it should be of a reasonable magnitude in a properly calibrated model. (An excessively large $\lambda^*$ would imply unreasonable risk aversion.) We can easily find an upper bound $\overline{\lambda} > \lambda^*$, because the slope of the efficient frontier is monotone decreasing in $\gamma$.

5.1 Transformation of the constrained problem into penalized form

In what follows, we assume that an upper bound $\overline{\lambda}$ of reasonable magnitude is known for the slope of the efficient frontier. We will transform the CVaR-constrained problem (24) into the following penalized form:

$$\min -E(w_2) + \overline{\lambda}[\text{CVaR}(w_2) - \gamma]. \tag{25}$$
The advantage of this penalized form is that its decomposition yields a two-stage problem having a relatively complete recourse. (Relatively complete recourse means that no extra action is needed to enforce feasibility of the second-stage problems.)

**Proposition 9** The set of the optimal solutions of the constrained problem (24) is identical to the set of the optimal solutions of the penalized problem (25).

**Proof.** Let \( \mathcal{H} \) denote the convex polyhedron of the feasible solutions of the unconstrained problem (17). Let the set \( H \subset \mathbb{R}^2 \) represent the projection of \( \mathcal{H} \) onto the (risk, return) plane, where risk means the end-of-horizon risk \( \text{CVaR}(w_2) \), and return the end-of-horizon return \( E(w_2) \). This is not a linear projection, but we have seen that the north-west contour of \( H \) is a concave, piecewise linear line, and coincides with the efficient frontier of the constrained problem (24).

Due to the assumptions made after the formulation of the constrained problem (24), all optimal solutions of this problem projects to the same point which will be denoted by \( h^* \in \mathbb{R}^2 \). Obviously \( h^* \in H \), and falls on the efficient frontier.

Let \( l^* \) denote the supporting line that touches the efficient frontier in \( h^* \). \( \lambda^* > 0 \) was defined to be the slope of \( l^* \). Let \( R^* \subset \mathbb{R}^2 \) denote a specific level set of the objective function (25):

\[
R^* := \{ (r_1, r_2) \in \mathbb{R}^2 \mid -r_1 + \bar{x} [r_2 - \gamma]_+ \leq -h^*_1 + \bar{x} [h^*_2 - \gamma]_+ \}, \quad \text{where} \ (h^*_1, h^*_2) = h^*.
\]

The above two-dimensional objects are easily depicted. Since \( \bar{x} > \lambda^* > 0 \), the supporting line \( l^* \) obviously separates \( R^* \) from \( H \). It is easily seen that \( R^* \cap H = \{ h^* \} \).

The level sets belonging to objective levels lower than \(-h^*_1 + \bar{x} [h^*_2 - \gamma]_+ \), are disjunct from \( H \).

Introducing a new variable \( y \) for the quantity \( [ \text{CVaR}(w_2) - \gamma ]_+ \), the polyhedral representation of problem (25) is the following:

\[
\begin{align*}
\text{min} & \quad -\sum_{(j,k) \in S} p_{(j,k)} w_{2}^{(jk)} + \bar{x} y \\
\text{such that} & \\
& \sum_{(j,k) \in L} p_{(j,k)} \left( w^B - w_2^{(jk)} - z \right) \leq v \quad (L \subset S), \\
& y \in \mathbb{R}, \ y \geq 0, \quad z + \frac{1-\alpha}{\alpha} v - y \leq \gamma,
\end{align*}
\]

where dots stand for the variable declarations and cash balance equations identical to those in the unconstrained prototype problem (17).

Let us substitute \( \tilde{v} := v - (1 - \alpha) y \) for the variable \( v \), and \( \tilde{y} := (1 - \alpha) y \) for the variable \( y \). Considering that \( v \geq 0 \) obviously holds for any feasible solution of the problem (26), and that \( y = 0 \) holds for any optimal
solution, we may also add the bound \( \tilde{v} \geq 0 \):

\[
\min \quad - \sum_{(j,k) \in S} p_{(j,k)} w_{2}^{(j,k)} + \bar{\lambda} \frac{1}{1-\alpha} \tilde{y}
\]

such that

\[
z, \tilde{v}, \tilde{y} \in \mathbb{R}, \quad \tilde{v}, \tilde{y} \geq 0, \quad x_1 \in X,
\]

\[
w_{1}^{(j)} \in \mathbb{R}, \quad x_{2}^{(j)} \in X, \quad (j = 1, \ldots, N),
\]

\[
w_{2}^{(j,k)} \in \mathbb{R}, \quad ((j, k) \in S),
\]

\[
1^T x_1 = w_0,
\]

\[
1^T x_{2}^{(j)} = w_{1}^{(j)} = r_{1}^{(j)} \quad (j = 1, \ldots, N),
\]

\[
w_{2}^{(j,k)} = r_{2}^{(j,k)} x_{2}^{(j)} \quad ((j, k) \in S),
\]

\[
\sum_{(j,k) \in L} w_{B}^{(j,k)} - w_{2}^{(j,k)} - z - \tilde{y} \leq \tilde{v} \quad (L \subset S),
\]

\[
z + \frac{1}{1-\alpha} \tilde{v} \leq \gamma.
\]

The aim of introducing the slack variable was to ensure that the forthcoming decomposed problem would have a relatively complete recourse.

### 5.2 Decomposition of the penalized two-stage problem

The first-stage problem will be:

\[
\min \quad \sum_{j=1}^{N} p_{1}^{(j)} F^{(j)} \left( w_{1}^{(j)}, z, v^{(j)} \right)
\]

such that

\[
z \in \mathbb{R}, \quad x_1 \in X,
\]

\[
w_{1}^{(j)}, v^{(j)} \in \mathbb{R}, \quad v^{(j)} \geq 0 \quad (j = 1, \ldots, N),
\]

\[
1^T x_1 = w_0,
\]

\[
w_{1}^{(j)} = r_{1}^{(j)} \quad (j = 1, \ldots, N),
\]

\[
z + \frac{1}{1-\alpha} \sum_{j=1}^{N} p_{1}^{(j)} v^{(j)} \leq \gamma,
\]

where the functions \( F^{(j)} : \mathbb{R}^3 \to \mathbb{R} \quad (j = 1, \ldots, N) \) are defined by the second-stage problem:
The objective function of the ELPF problem is obviously equivalent to the objective function of the polyhedral representation problem (27).

Cuts in the ELPF problem have the form
\[
\sum_{k \in K^{(j)}} p_2^{(jk)} \left( w^ B - w_2^{(k)} - \zeta \right) - y \leq \nu \quad (K \subset \{1, \ldots, M_j\}).
\]

where \(1 \leq j \leq N\) and \(K^{(j)} \subset \{1, \ldots, M_j\}\). Equivalence between aggregate ELPF-cuts and the cuts of the polyhedral representation problem can be established like in the proof of Theorem 7.

It remains to be shown only that any non-negative value of the variables \(\tilde{v}\) and \(\tilde{y}\) in problem (27) can be represented in the forms (30) with non-negative values of the ELPF-variables \(v^{(j)}\) and \(y^{(j)}\), respectively.

The objective function of the ELPF problem is obviously equivalent to the objective function of the polyhedral representation problem (27).

Cuts in the ELPF problem have the form
\[
\sum_{k \in K^{(j)}} p_2^{(jk)} \left( w^ B - w_2^{(j)} - \zeta \right) - y^{(j)} \leq \nu^{(j)},
\]

where \(1 \leq j \leq N\) and \(K^{(j)} \subset \{1, \ldots, M_j\}\). Equivalence between aggregate ELPF-cuts and the cuts of the polyhedral representation problem can be established like in the proof of Theorem 7.

Theorem 10 The two-stage problem (28 - 29) is equivalent to the polyhedral representation problem (27).

Proof. Let us construct the equivalent linear programming form (ELPF) of the two-stage problem (28 - 29) in the usual manner. It means defining a separate set of second-stage variables for each \(j = 1, \ldots, N\). Let
\[
y^{(j)} \geq 0, \quad x_2^{(j)} \in X,
\]
\[
w_2^{(j)} \in \mathbb{R} \quad (k = 1, \ldots, M_j),
\]
\[
l^T x_2 = \omega,
\]
\[
w_2^{(k)} = r_2^{(jk)} x_2 \quad (k = 1, \ldots, M_j),
\]
\[
\sum_{k \in K} p_2^{(jk)} \left( w^ B - w_2^{(k)} - \zeta \right) - y \leq \nu \quad (K \subset \{1, \ldots, M_j\}).
\]

i.e., the variables \(\tilde{v}\) and \(\tilde{y}\) of the polyhedral representation problem will correspond to weighted sums of the ELPF-variables \(v^{(j)}\) and \(y^{(j)}\), respectively.

The objective function of the ELPF problem is obviously equivalent to the objective function of the polyhedral representation problem (27).

Cuts in the ELPF problem have the form
\[
\sum_{k \in K^{(j)}} p_2^{(jk)} \left( w^ B - w_2^{(j)} - \zeta \right) - y^{(j)} \leq \nu^{(j)},
\]

where \(1 \leq j \leq N\) and \(K^{(j)} \subset \{1, \ldots, M_j\}\). Equivalence between aggregate ELPF-cuts and the cuts of the polyhedral representation problem can be established like in the proof of Theorem 7.
We have \( \chi^{(j)} \geq 0 \) since \( K^{(j)} = \emptyset \) can be selected. The aggregate of these deepest cuts is valid in the polyhedral problem (27), hence
\[
\chi - \tilde{y} \leq \tilde{v} \quad \text{holds with} \quad \chi := \sum_{j=1}^{N} p_1^{(j)} \chi^{(j)}.
\]
Assumed \( \chi > 0 \), let
\[
y^{(j)} := \frac{\chi^{(j)}}{\chi}, \quad v^{(j)} := \frac{\chi^{(j)}}{\chi} \quad (1 \leq j \leq N).
\]
These are obviously feasible values in the ELPF-problem. (In the trivial case of \( \chi = 0 \), let \( y^{(j)} := \tilde{y}/N, v^{(j)} := \tilde{v}/N \).)

The second-stage problem (29) is obviously feasible for any setting of the parameters \( \zeta, \nu \), and for each relevant setting of \( \omega \). I.e., the two-stage problem (28 - 29) has relatively complete recourse.

### 5.3 A solution method for the decomposed problem

The first-stage problem (28) can be solved by the Inexact Level Method. (The method was outlined in Section 3.2.) In the present case the objective function is
\[
\min \sum_{j=1}^{N} p_1^{(j)} F^{(j)} \left( w_1^{(j)}, z, v^{(j)} \right).
\]

The Inexact Level Method needs estimate function values and \( \epsilon \)-subgradients of the functions \( F^{(j)} \). Formally:

having fixed the values \( \hat{w}_1^{(j)}, \hat{z}, \hat{v}^{(j)} \) \( (1 \leq j \leq N) \) of the first-stage variables, and given a tolerance \( \epsilon > 0 \), we must construct a linear function \( L^{(j)}(\omega, \zeta, \nu) \) such that
\[
L^{(j)}(\omega, \zeta, \nu) \leq F^{(j)}(\omega, \zeta, \nu) \quad ((\omega, \zeta, \nu) \in \mathbb{R}^3) \quad \text{and} \quad L^{(j)}(\hat{w}_1^{(j)}, \hat{z}, \hat{v}^{(j)}) \geq F^{(j)}(\hat{w}_1^{(j)}, \hat{z}, \hat{v}^{(j)}) - \epsilon.
\]

The genuine variables of the second-stage problem (29) are only \( y \geq 0 \) and \( x_2 \in X \) because \( w_2^{(k)} \) \( (k = 1, \ldots, M_j) \) are determined by \( x_2 \). The feasible set \( x_2 \in X \), \( 1^T x_2 = \hat{w}_1^{(j)} \) is compact. Though the variable \( y \) formally takes values from an infinite interval, an upper bound \( \tilde{y} \) can be determined from the parameter values \( (\hat{w}_1^{(j)}, \hat{z}, \hat{v}^{(j)}) \). Hence the feasible domain of the second-stage problem is contained in a compact set.

The second-stage problem (29) can be written in the form
\[
\min \varphi(x) \quad \text{subject to} \quad x \in X, \quad \psi(x) \leq 0,
\]
with
\[
\begin{align*}
X := & \left\{ (x_2, y) \mid x_2 \in X, \ 1^T x_2 = \hat{w}_1^{(j)}, \ 0 \leq y \leq \tilde{y} \right\}, \\
\varphi(x) := & - \sum_{k=1}^{M_j} p_2^{(jk)} r_2^{(jk)T} x_2 + \frac{1}{1-\alpha} y, \\
\psi(x) := & \max_{K \subset \{1, \ldots, M_j\}} \left\{ \sum_{k \in K} p_2^{(jk)} \left( w^B - r_2^{(jk)T} x_2 - \hat{z} \right) - y - \hat{v}^{(j)} \right\}.
\end{align*}
\]
Remark 11 Here $\mathcal{X} \subset \mathbb{R}^n$ is a convex bounded polyhedron, and $\varphi, \psi$ are convex functions. A common Lipschitz constant for $\varphi$ and $\psi$ can easily be constructed using $\max_{k=1,\ldots,M} \| r^{(2k)}_j \|$.

A bound $D$ on the diameter of the feasible domain, and a Lipschitz constant $\Lambda$ are needed for the convergence proof of the second-stage solution method that we are going to propose.

The second-stage solution method is applied several times in course of the solution of a two-stage problem. We can easily construct global bounds $D$ and $\Lambda$ that are valid for each second-stage problem solved in course of the solution of a two-stage problem.

In case $\psi(x) \leq 0$ holds for each $x \in \mathcal{X}$, then problem (34) reduces to unconstrained minimization, and can be solved by a cutting-plane or bundle-type method. These methods construct a convex polyhedral lower approximating function to the objective function. Minimization of this approximating function is called the model problem. The parameters $\omega, \zeta, \nu$ appear in the right-hand side of the model problem. When $\epsilon$-optimal solution has been reached, the optimal dual variables of the model problem yield the $\epsilon$-support function (33).

In what follows we assume that (34) is really a constrained problem. For the solution we propose another modified version of the Constrained Level Method of Lemaréchal, Nemirovskii, and Nesterov (1995).

Let $\Phi^*$ denote the optimal objective value of the constrained convex problem (34). Given $\epsilon > 0$, the original Constrained Level Method finds an $\epsilon$-optimal solution $x^*$ in the sense

$$x^* \in \mathcal{X}, \quad \psi(x^*) \leq \epsilon, \quad \varphi(x^*) \leq \Phi^* + \epsilon. \tag{35}$$

Due to this feature, the original Constrained Level Method requires no Slater condition. In contrast, we will construct an $\epsilon$-optimal solution $x^*$ in the sense

$$x^* \in \mathcal{X}, \quad \psi(x^*) \leq 0, \quad \varphi(x^*) \leq \Phi^* + \epsilon. \tag{36}$$

To accomplish this, we will need the Slater condition

$$\min_{x \in \mathcal{X}} \psi(x) < 0. \tag{37}$$

This obviously holds if the upper bound $\overline{y}$ is selected large enough.

The modified method is described in the Appendix. It constructs a linear programming model problem (38). The optimum of this model problem is a lower approximation of the optimum of the original convex problem (34). The parameters $\omega, \zeta, \nu$ appear in the right-hand side of the model problem.

Suppose the method had stopped with an $\epsilon$-optimal solution in the sense (36). Then optimal dual variables of the linear programming model problem yield the $\epsilon$-support function (33).

Remark 12 The first-stage problem is solved by the Inexact Level Method. Convergence proof of this method needs a common Lipschitz-constant for the $\epsilon$-support functions constructed in course of the procedure.

The arguments of Remark 8 can be adapted to the present case: let us substitute problem (38) instead of (15). Problem (38) is obtained from the second-stage problem (29) by omitting some of the polyhedral cuts. (Although the parameters are not explicitly written in (38), this problem is an approximation of (29) in the same manner as (15) is an approximation of (14) and (19).)

This, together with the global bounds mentioned in Remark 11, would enable us constructing an upper bound on the total number of iterations of the whole two-stage solution procedure.

Experimental results with a similar solution method will be cited in Section 6.

6 Conclusions and prospects

The paper contains the following new results:
We develop a polyhedral representation for CVaR, which is similar to but simpler than the Künzi-Bay Mayer (2006) representation. Its simplicity makes the representation (12) an attractive alternative in case of one-stage problems. For the decomposition of two-stage models, however, the more detailed representation of Künzi-Bay and Mayer is needed.

We present a decomposition framework for the two-stage CVaR minimization prototype problem (16). From an algorithmic point of view, the two-stage financial models of Topaloglou (2004) and Topaloglou, Vladimirou, Zenios (2006) fit this prototype.

The decomposition is based on the polyhedral representation of Künzi-Bay and Mayer (2006). In this framework, the second-stage problems are parametric one-stage CVaR-minimization problems. They can be approximately solved by the cutting-plane method proposed by Künzi-Bay and Mayer, or by a bundle-type method.

Using this solution method as an oracle that provides approximate information of the master objective function, the master problem can be solved by an inexact version of the Level Method of Lemaréchal, Nemirovskii, and Nesterov (1995). The inexact version was proposed by Fábián (2000).

In contrast to usual models that minimize the CVaR of the end-of-the-horizon yield, we present a model that penalizes the first-period risk also. This brings an additional term in the objective function (22). This problem can also be decomposed and solved according to the scheme proposed for the prototype problem (16).

We present a decomposition framework for the a two-stage CVaR-constrained problem. (We impose a constraint on end-of-horizon risk, in the form of (23).) The decomposition is based on the polyhedral representation of Künzi-Bay and Mayer (2006).

We transform the two-stage constrained problem into such form that the above mentioned decomposition yields a relatively-complete-recourse problem. (If we had applied the decomposition in direct form, then infeasibility might have occurred in the second-stage problems. In such cases, enforcing feasibility requires an extra effort in course of the solution process.)

In previous work, Fábián and Szőke (2006) obtained good solution results for general two-stage stochastic programming problems by first transforming the problems into complete recourse forms. The first-stage problems were then solved as constrained convex problems. (In such a constrained problem, the constraint function value is the expectation of a second-stage infeasibility measure. The constraint function can be evaluated in the same manner as the expected recourse function.) In that Level Decomposition approach, feasibility and optimality issues are taken into consideration simultaneously, and regularization extends to both. (Feasibility cuts of the traditional solution methods may cause the scope of optimization to alternate between minimizing the objective function and finding a solution that satisfies existing feasibility cuts. No feasibility cuts are imposed in the Level Decomposition approach.) The Level Decomposition approach is practicable only if the transformation to complete recourse can be done in a computationally efficient way.

In the present portfolio management context, the slope of the efficient frontier is interpreted as a measure of the decision maker’s risk aversion. Hence in the neighborhood of well-calibrated parameter values, the efficient frontier has a reasonably moderate slope. This feature allows us to construct relatively-complete-recourse problems by penalizing violations of CVaR constraints. (The problem is put into the form (25), where $\lambda$ is a bound on the slope of the efficient frontier. This problem is then decomposed into the relatively-complete-recourse form (28 - 29).)

In the present case, the transformation into relatively-complete-recourse form is such that we need no extra constraint on the expectation of the second-stage infeasibility measure. Due to this specific feature, the present first-stage problem (28) can be solved by an unconstrained convex optimization method. We recommend the Inexact Level Method that requires information on the objective function in the form of $\epsilon$-subgradients.
We develop a sharp-constrained version of the Constrained Level Method of Lemaréchal, Nemirovskii, and Nesterov (1995). The Constrained Level Method solves the constrained convex problem (34). Given a tolerance $\epsilon > 0$, the original method constructs an $\epsilon$-optimal solution $x^*$ in the sense (35).

The sharp-constrained version constructs an $\epsilon$-optimal solution in the sense (36), hence it can be used to construct $\epsilon$-subgradients to the first-stage objective function (32).

The sharp-constrained version is simpler than the original Constrained Level Method, but it can be only used if the Slater condition (37) holds.

In the above described solution schemes, we use inexact methods for the solution of the first-stage problems. This way we can find a balance between the computational effort of optimization and of approximation. At the beginning of the solution procedure, rough approximate solutions of the second-stage problems are sufficient. As the optimum of the first-stage problem is approached, the accuracy of the current second-stage solutions is gradually increased.

This solution procedure is an adaptation of the Level Decomposition scheme that was devised by Fábián and Szőke (2006) for the solution of stochastic programming problems having the decision/observation/decision pattern. The Künzi-Bay – Mayer polyhedral representation enables the adaptation of the Level Decomposition scheme to the present problems having the decision/observation/decision/observation pattern.

Fábián and Szőke (2006) successfully used the Level Decomposition method for general two-stage stochastic programming problems. A stochastic programming problem with medium-sized first-stage and second-stage problems and almost-complete-recourse was solved with a growing number of scenarios. 6-digit optimal solution of the 600,000-scenario problem was found in 13 minutes on a regular desktop computer. (The first-stage problem had 188 variables and 62 constraints; and each of the 600,000 second-stage problems had 272 variables and 104 constraints.)

In the present case, the second-stage problems have a special structure that is effectively exploited by the method of Künzi-Bay and Mayer.

A definitive validation of the numerical efficiency of the Level Decomposition scheme adapted to the present specific problems will require numerical experiments which will be the subject of subsequent research. However, the computational evidence published by Künzi-Bay and Mayer (2006) on the one hand, and by Fábián and Szőke (2006) on the other hand, support our confidence in the efficiency of the scheme.

6.1 Further fields of potential application

Quadratic problems. Roman, Mitra, and Darby-Dowman (2006) propose and study portfolio optimization models in which variance is minimized under constraints on expected yield and CVaR. They construct an approximation of the efficient frontier. In making the final choice, the decision maker plays a key role. This approach represents a compromise between regulators’ requirements for short tails and classical fund managers’ requirements for small variance. The approach could be extended to two-stage models, and the decomposition scheme of Section 5 could be adapted to the resulting convex quadratic problems.

Steinbach (1999) proposed and studied multiperiod mean/risk portfolio optimization models. He measured risk with semivariance of end-of-horizon portfolio yield, which resulted in convex quadratic problems. Gondzio and Grothey (2003) built an interior-point solver that exploits the special structure of such problems, and achieves high efficiency in their solution. The decomposition scheme of Section 5 could be adapted to convex quadratic problems, and integrated with the interior-point approach of Gondzio and Grothey, for the solution of these multi-stage problems: decomposition could be applied in early stages, and structure-exploiting interior-point approach in further stages.

Robust optimization. CVaR is a coherent risk measure according to the definition of Artzner et al. (1999). There is a close relationship between robust optimization, and optimization with coherent risk
measures; see, e.g., Bertsimas and Brown (2005). From this point of view, the two-stage CVaR-constrained problem (24) can be considered a two-stage robust optimization problem, and the proposed solution methods robust optimization methods.

Acknowledgement

I thank the Editors and the anonymous Referees for the valuable comments and suggestions that led to substantial improvement of the paper.

References


Appendix: A Sharp-Constrained Level Method

We introduce a Sharp-Constrained version of the Constrained Level Method of Lemaréchal, Nemirovskii, and Nesterov (1995). Minor modifications are needed in the original constructions and proofs.

The methods solve the constrained convex problem (34), where $\mathcal{X}$ is a bounded convex polyhedron; and $\varphi, \psi$ are convex functions, Lipschitz-continuous relative to $\mathcal{X}$. It is also assumed that the problem is feasible, and is a genuine constrained problem (i.e., there exits $x \in \mathcal{X}$ such that $\psi(x) > 0$).

Given a tolerance $\epsilon > 0$, the original method constructs an $\epsilon$-optimal solution $x^*$ in the sense (35). The sharp-constrained version constructs an $\epsilon$-optimal solution in the sense (36). The sharp-constrained version is simpler than the original Constrained Level Method, but requires the Slater condition (37).

A sequence of feasible points is generated, and at each point, linear support functions are constructed to the functions $\varphi$ and $\psi$. These linear support functions are used to build the cutting-plane models of $\varphi$ and $\psi$.

Suppose that we have already constructed the points $x_1, \ldots, x_\kappa \in \mathcal{X}$. Suppose we have constructed the linear support function $l_\iota$ for $\varphi$ at the point $x_\iota$ ($\iota = 1, \ldots, \kappa$). Linear support functions for $\psi$ will be denoted by $l^{\#}_\iota$ ($\iota = 1, \ldots, \kappa$).

The cutting-plane models of $\varphi$ and $\psi$ will be

$$\varphi_\kappa(x) := \max_{1 \leq i \leq \kappa} l_i(x) \quad \text{and} \quad \psi_\kappa(x) := \max_{1 \leq i \leq \kappa} l^{\#}_i(x) \quad (x \in \mathbb{R}^n).$$
Obviously we have $\phi_\kappa \leq \phi$ and $\psi_\kappa \leq \psi$. A linear programming model of the constrained convex problem (34) will be constructed as

$$\begin{align*}
\min & \quad \phi_\kappa(x) \\
\text{subject to} & \quad x \in \mathcal{X}, \\
& \quad \psi_\kappa(x) \leq 0.
\end{align*}$$ (38)

Let $\Phi_\kappa^*$ denote the optimal objective value of the model problem (38). Obviously we have $\Phi_\kappa^* \leq \Phi^*$. (Lower valued function is minimized over a broader set.)

The best point associated with $x_1, \ldots, x_\kappa$ will be constructed in the form of a convex combination of the former iterates:

$$x_\kappa^* := \sum_{i=1}^{\kappa} \theta_i x_i.$$ (39)

The weights $\theta_1, \ldots, \theta_\kappa$ will be determined through the solution of the following linear programming problem:

$$\begin{align*}
\min & \quad \sum_{i=1}^{\kappa} \theta_i \phi(x_i) \\
\text{subject to} & \quad \theta_i \geq 0 \quad (i = 1, \ldots, \kappa), \\
& \quad \sum_{i=1}^{\kappa} \theta_i = 1, \\
& \quad \sum_{i=1}^{\kappa} \theta_i \psi(x_i) \leq 0.
\end{align*}$$ (40)

(The values $\phi(x_i)$ are known since they had been computed for the construction of the linear support functions $l_i(x_i)$. Similarly, the values $\psi(x_i)$ are known for each $i = 1, \ldots, \kappa$.)

Unlike the original Constrained Level Method, the present Sharp-Constrained version needs special arrangement to ensure feasibility of problem (40): Let $x^\psi$ denote an optimal solution of the problem

$$\min_{x \in \mathcal{X}} \psi(x).$$

Assume that we have selected $x^\psi$ as starting point, i.e., we have set $x_i := x^\psi$ for $i = 1$. Then $\theta_1 = 1$, $\theta_i = 0$ ($i = 2, \ldots, \kappa$) is a feasible solution, due to the Slater condition (37). Hence an optimal solution also exists (because the feasible domain is compact). Let $\Phi_\kappa^*$ denote the optimal objective value of problem (40). From the convexity of the set $\mathcal{X}$, and of the functions $\psi$ and $\phi$, it follows that

$$x_\kappa^* \in \mathcal{X}, \quad \psi(x_\kappa^*) \leq 0, \quad \phi(x_\kappa^*) \leq \Phi_\kappa^*.$$ (41)

Consequently $\Phi^* \leq \Phi_\kappa^*$ holds. (Indeed, $x_\kappa^*$ is a feasible solution of the convex programming problem (34), and $\Phi^*$ denotes the optimum of (34).)

**Remark 13** The inequality $\Phi^* \leq \Phi_\kappa^*$ does not necessarily hold with the original Constrained Level Method: In the original method, the objective in problem (40) is the simultaneous maximization of an optimality measure and a feasibility measure. Hence the best point $x_\kappa^*$ may only satisfy the non-positivity constraint with a tolerance.

$\Phi^*$ is unknown but the optimum $\Phi_\kappa^*$ of the model problem provides a lower bound. If we can achieve $\Phi_\kappa^* - \Phi_\kappa^* \leq \epsilon$ then $x_\kappa^*$ will be an $\epsilon$-optimal solution of the convex programming problem (34) in the sense (36).

The linear programming dual of problem (40) can be written in the form

$$\max_{\beta \geq 0} h_\kappa(\beta),$$ (42)
where
\[ h_\kappa(\beta) := \min_{1 \leq \iota \leq \kappa} \varphi(x_\iota) + \beta \psi(x_\iota). \]

Hence we have
\[ \max_{\beta \geq 0} h_\kappa(\beta) = \Phi_\kappa^* \] (43)
from the duality theorem. Our aim is to direct the search for new iterates \( x_\iota \) in such a way that \( \max_{\beta \geq 0} h_\kappa(\beta) \leq \Phi_\kappa^* + \epsilon \) will hold with \( \kappa \) as small as possible. In order to be able to deploy the scheme worked out by Lemaréchal, Nemirovskii, and Nesterov, we must re-formulate problem (42) so that the feasible domain would be compact.

Let us introduce the notation
\[ \Phi := \min_{x \in X} \varphi(x) \quad \text{and} \quad \beta := \frac{\varphi(x^\psi) - \Phi}{-\psi(x^\psi)}. \]

The denominator in the above fraction is positive due to the Slater condition (37). The numerator is non-negative due to the definition of \( \Phi \). It follows that \( \beta \geq 0 \).

**Proposition 14** Assume that we have selected \( x^\psi \) as starting point, i.e., we have set \( x_1 := x^\psi \) for \( \iota = 1 \). Then \( h_\kappa(\beta) < \Phi \) holds for any \( \beta > \beta_\lambda \).

**Proof.** From the selection of the starting point, we have
\[ h_1(\beta) = \varphi(x^\psi) + \beta \psi(x^\psi) = \varphi(x^\psi) + \frac{\varphi(x^\psi) - \Phi}{-\psi(x^\psi)} \psi(x^\psi) = \Phi. \]

Due to \( \psi(x^\psi) < 0 \) we have \( h_1(\beta) < h_1(\beta) \) for \( \beta > \beta_\lambda \). Moreover we have \( h_k \leq h_1 \) by definition. Summing these up, we get that
\[ h_\kappa(\beta) \leq h_1(\beta) < h_1(\beta) = \Phi \quad \text{holds for } \beta > \beta_\lambda. \]

A direct consequence of Proposition 14 is that optimal solutions of problem (42) fall into the interval \([0, \beta_\lambda] \).

Indeed, we have \( \Phi \leq \varphi(x^*_\kappa) \leq \Phi_\kappa^* = \max_{\beta \geq 0} h_\kappa(\beta) \) by (41) and (43). It follows that the dual problem (42) can be written in the form
\[ \max_{\beta \geq 0} h_\kappa(\beta). \] (44)

For the above objects we can directly apply the scheme worked out by Lemaréchal, Nemirovskii, and Nesterov. The framework of the Sharp-Constrained Level Method is the following:

**Initialize.**
- Set the stopping tolerance \( \epsilon > 0 \).
- Set the parameters \( \theta, \mu \) (\( 0 < \theta, \mu < 1 \)).
- Let \( \kappa := 1 \) (iteration counter).
- Set the starting point \( x_1 \in \arg \min_{x \in X} \psi(x) \).
- Set upper limit for the feasible interval of the dual problem as \( \beta := -\left( \varphi(x_1) - \min_{x \in X} \varphi(x) \right) / \psi(x_1) \).
Update bundle.
Given the point \( x_\kappa \), construct the linear support functions \( l_\kappa \) and \( l_\kappa^\# \) for \( \varphi \) and \( \psi \), respectively.

Define the model functions
\[
\varphi_\kappa(x) := \max_{1 \leq \iota \leq k} l_\iota(x), \quad \psi_\kappa(x) := \max_{1 \leq \iota \leq k} l_\iota^\#(x).
\]

Compute the optimum of the model problem
\[
\Phi_\kappa^\star := \min \{ \varphi_\kappa(x) \mid x \in \mathcal{X}, \psi_\kappa(x) \leq 0 \}.
\]

Define the function
\[
h_\kappa(\beta) := \min_{1 \leq \iota \leq k} \varphi_\kappa(x_\iota) + \beta \psi_\kappa(x_\iota),
\]
and compute its maximum
\[
\Phi_\kappa^* := \max_{\beta \geq 0} h_\kappa(\beta).
\]

Check for optimality.
If \( \Phi_\kappa^* - \Phi_\kappa^\star < \epsilon \), then near-optimal solution found.
Construct best point according to (39) and stop.

Find dual iterate.
Determine the interval
\[
[\underline{\beta}_\kappa, \overline{\beta}_\kappa] := \{ 0 \leq \beta \leq \beta \mid h_\kappa(\beta) \geq \Phi_\kappa^\star \}.
\]
Compute \( \beta_\kappa \):
- for \( \iota = 1 \), let \( \beta_1 := \frac{1}{2}(\overline{\beta}_1 + \underline{\beta}_1) \),
- for \( \iota > 1 \), let
\[
\beta_\kappa := \begin{cases} 
\beta_{\kappa-1}, & \text{if } \frac{\beta_\kappa + \frac{\theta}{1-\theta}(\overline{\beta}_\kappa - \beta_\kappa)}{2} \leq \beta_{\kappa-1} \leq \beta_\kappa - \frac{\theta}{1-\theta}(\overline{\beta}_\kappa - \beta_\kappa), \\
\frac{1}{2}(\overline{\beta}_\kappa + \underline{\beta}_\kappa), & \text{otherwise}.
\end{cases}
\]

Find primal iterate.
Let \( x_{\kappa+1} \) be the optimal solution of the convex quadratic programming problem
\[
\min \| x - x_\kappa \|^2 \\
\text{subject to} \quad x \in \mathcal{X}, \quad \varphi_\kappa(x) + \beta_\kappa \psi_\kappa(x) \leq (1 - \theta) \Phi_\kappa^\star + \theta h_\kappa(\beta_\kappa)
\]

Loop.
Increment \( \kappa \).
\( \to \) Update bundle.

We give explanations and sketch proofs on the basis of Lemaréchal, Nemirovskii, and Nesterov (1995).
Find dual iterate. A dual iterate $\beta_\kappa$ is selected such that $h_\kappa(\beta_\kappa)$ is ‘sufficiently close’ to $\max_{\beta \geq \beta \geq 0} \min h_\kappa(\beta)$. (The term ‘sufficiently close’ will be specified in (46).) First consider the set

$$I_\kappa := \{ 0 \leq \beta \leq \beta \mid h_\kappa(\beta) \geq \Phi^*_\kappa \}.$$  

$I_\kappa$ is an interval since $h_\kappa$ is a concave function. In the algorithmic description the notation $I_\kappa = [\beta_\kappa, \beta_\kappa]$ is used. This interval is not empty since we have $\max_{\beta \geq \beta \geq 0} \min h_\kappa(\beta) = \overline{\Phi}^*_\kappa \geq \Phi^*_\kappa$. Let the subinterval $\hat{I}_\kappa \subset I_\kappa$ be obtained by shrinking $I_\kappa$: the center of $\hat{I}_\kappa$ will be the same as the center of $I_\kappa$, and for the lengths, $|\hat{I}_\kappa| = (1 - \mu)|I_\kappa|$ will hold with some preset parameter $0 < \mu < 1$. Owing to the concavity of $h_\kappa$, it follows that

$$h_\kappa(\beta) - \Phi^*_\kappa \geq \frac{\mu}{2} \left( \overline{\Phi}^*_\kappa - \Phi^*_\kappa \right) \quad (46)$$

holds for any $\beta \in \hat{I}_\kappa$. The dual iterate $\beta_\kappa$ will be selected from $\hat{I}_\kappa$. The aim is to leave the dual iterate unchanged as long as possible.

Find primal iterate. The primal iterate $x_{\kappa+1}$ is selected by applying an unconstrained-Level-Method-type iteration to the convex function $\varphi + \beta_\kappa \psi$. (A complete description of the unconstrained Level Method can be found in Lemaréchal, Nemirovskii, and Nesterov (1995).) For $1 \leq i \leq \kappa$, the linear function $l_i + \beta_\kappa l_i^\#$ is a support function of $\varphi + \beta_\kappa \psi$ at $x_i$. Hence $\varphi_\kappa + \beta_\kappa \psi$ is an appropriate cutting-plane model of $\varphi + \beta_\kappa \psi$.

The best function value, i.e., the lowest function value taken among the known iterates is $BFV := \min_{1 \leq i \leq \kappa} \varphi(x_i) + \beta_\kappa \psi(x_i) = h_\kappa(\beta_\kappa)$.

A lower function level is selected specially as $LFL := \Phi^*_\kappa$. The lower function level in the Level Method must satisfy two requirements:

1. $LFL \leq BFV$ should hold.
2. There should exist a feasible point whose function value is lower than or equal to $LFL$.

In the present case,

1. $LFL = \Phi^*_\kappa \leq h_\kappa(\beta_\kappa) = BFV$ holds owing to the selection $\beta_\kappa \in I_\kappa$ of the dual iterate.
2. Let $x_\kappa$ denote the minimizer of the model problem (38). Obviously we have $\varphi(x_\kappa) + \beta_\kappa \psi(x_\kappa) \leq \Phi^*_\kappa$.

The gap between $BFV$ and $LFL$ is $GAP := h_\kappa(\beta_\kappa) - \Phi^*_\kappa$. Consider the level set of the model function $\varphi_\kappa(x) + \beta_\kappa \psi_\kappa(x)$ belonging to the level $LFL + \theta GAP$, where $0 < \theta < 1$ is some preset parameter. We have $LFL + \theta GAP = (1 - \theta)\Phi^*_\kappa + \theta h_\kappa(\beta_\kappa)$, hence the level set will be

$$X_\kappa := \{ x \in X \mid \varphi_\kappa(x) + \beta_\kappa \psi_\kappa(x) \leq (1 - \theta)\Phi^*_\kappa + \theta h_\kappa(\beta_\kappa) \}.$$ 

This is a convex polyhedron. The next primal iterate $x_{\kappa+1}$ will be the projection of the former iterate $x_\kappa$ onto the level set $X_\kappa$. (I.e., the point of $X_\kappa$ that is closest to $x_\kappa$ in Euclidean distance. This can be determined through the solution of a convex quadratic programming problem.)

Convergence. Since the dual iterate is left unchanged as long as possible, the method consists of runs of the (unconstrained) Level Method. First we estimate the length of a single run. Let $\varepsilon$ be a positive number (different from the stopping tolerance $c$ of the constrained method). Lemaréchal, Nemirovskii, and Nesterov proved that to obtain $GAP < \varepsilon$ in the unconstrained Level Method, it suffices to perform

$$c \left( \frac{D\Delta}{\varepsilon} \right)^2 \quad (47)$$

28
iterations, where \( D \) is the diameter of the feasible polyhedron, \( \Lambda \) is a Lipschitz constant of the objective function (presently \( \varphi + \beta \kappa \psi \)), and \( c \) is a constant that depends only on the parameter \( \theta \). Obviously we have \( \Lambda \leq (1 + \beta) \Lambda \) where \( \Lambda \) is a common Lipschitz constant of the functions \( \varphi \) and \( \psi \).

In the present case, the \( \text{GAP} \) after \( \kappa \) iterations is \( h_\kappa(\beta_\kappa) - \Phi^* \). If it has decreased below \( \varepsilon \), then we have

\[
\varepsilon > \frac{\mu}{2} (\Phi^* - \Phi^*)
\]

from (46). Here we must have \( \Phi^* - \Phi^* > \varepsilon \) with the stopping tolerance \( \epsilon \) of the constrained method, otherwise the whole procedure would have been stopped. Hence \( \varepsilon > \frac{\varepsilon}{2} \) must hold, which from (47) gives the bound

\[
c \left( \frac{2}{\mu} \right)^2 (1 + \beta)^2 \left( \frac{D \Lambda}{\epsilon} \right)^2
\]

on the length of a single run of unconstrained Level Methods.

On the other hand, a bound can also be constructed on the number of the runs: Let \( |I^{(\sigma)}| \) denote the length of the interval (45) at the beginning of the \( \sigma \)th run. Then

\[
|I^{(\sigma + 1)}| \leq \frac{1}{2 - \mu} |I^{(\sigma)}|
\]

holds due to the selection of the dual iterate. (Indeed, the shrunk interval \( \hat{I}^{(\sigma + 1)} \subset I^{(\sigma + 1)} \) must be contained in one of the halves of \( I^{(\sigma)} \), otherwise the dual iterate would not have changed.)

Hence the length of the interval decreases by a geometric progression. Since \( I_\kappa \) is the support of the function \( h_\kappa(\beta) - \Phi^* \), and \( \Phi^* - \Phi_\kappa^* \) is the maximum of this function, the latter must decrease with the length of the support. Indeed, the function \( h_\kappa(\beta) \) is Lipschitz continuous with the constant

\[
\max_{\mathbf{x} \in X} |\psi(\mathbf{x})| \leq D \Lambda.
\]

(The function \( \psi \) is assumed to take negative as well as positive values over \( X \).)

The length of the initial interval is not larger than \( \beta \). Owing to the geometric progression, the number of the runs can not be larger than

\[
c' \ln \left( \frac{D \Lambda}{\epsilon} \right),
\]

where \( c' \) is a constant that depends only on the parameter \( \mu \).

The above bound, combined with (48), gives the following efficiency estimate: To obtain an \( \varepsilon \)-optimal solution in the sense (41), it suffices to perform

\[
c'' (1 + \beta)^2 \left( \frac{D \Lambda}{\epsilon} \right)^2 \ln \left( \frac{D \Lambda}{\beta \epsilon} \right)
\]

iterations, where \( c'' \) is a constant that depends only on the parameters of the algorithm.