RUTCOR
RESEARCH
REPORT

ALGORITHMS FOR CVaR OPTIMIZATION IN DYNAMIC STOCHASTIC PROGRAMMING MODELS WITH APPLICATIONS TO FINANCE

Csaba I. Fábián[a] Anna Veszprémi[b]

RRR 24-2006, SEPTEMBER, 2006

RUTCOR
Rutgers Center for
Operations Research
Rutgers University
640 Bartholomew Road
Piscataway, New Jersey
08854-8003
Telephone: 732-445-3804
Telefax: 732-445-5472
Email: rrr@rutcor.rutgers.edu
http://rutcor.rutgers.edu/∼rrr

[a]Department of Operations Research, Loránd Eötvös University of Budapest. P.O.Box 120, Budapest, H-1518 Hungary. E-mail: fabian@cs.elte.hu. This research has been supported by OTKA, Hungarian National Fund for Scientific Research, project No. 47340.
[b]Department of Algorithms and their Applications, Loránd Eötvös University of Budapest.

Rutcor Research Report
RRR 24-2006, SEPTEMBER, 2006

Algorithms for CVaR optimization in dynamic stochastic programming models with applications to finance

Csaba I. Fábián Anna Veszprémi

Abstract. We propose a decomposition scheme and solution methods for multistage CVaR-minimization or CVaR-constrained problems. This scheme meets the need for handling multiple CVaR-constraints for different time frames and at different confidence levels. Hence it allows the shaping of distributions according to the decision maker's preferences.
With minor modifications, the proposed scheme can be used to decompose further types of risk constraints in dynamic portfolio management problems. We consider integrated chance constraints, second-order stochastic dominance constraints, and constraints involving a special value-of-information risk measure.
We also suggest application to further financial problems. We propose a dynamic risk-constrained optimization model for option pricing. Moreover we propose special mid-term constraints for use in asset-liability management.

1 Introduction

Value-at-Risk (VaR) is a widely accepted risk measure that is not easily managed in optimization problems. An alternative risk measure, namely Conditional Value-at-Risk (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 makes CVaR tractable in optimization problems. An overview of VaR- and CVaR-optimization models and methods can be found in Prékopa (2003) and Kall and Mayer (2005). We list the stages of the evolution of dynamic CVaR models:

Andersson et al. (2001) examined single-stage 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. They demonstrated that diverse risk measures (including VaR) can be effectively controlled through CVaR optimization.

Krokhmal, Palmquist, and Uryasev (2002) and Rockafellar and Uryasev (2002) introduced the idea of using CVaR in multistage models. They 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. Claessens and Kreuser (2004) utilize these ideas in an asset/liability management tool.

Topaloglou, Vladimirou, and Zenios (2004) developed elaborate multistage financial models. The objective was minimization of end-of-horizon CVaR under a constraint on expected return. The authors constructed the respective efficient frontiers of one- and two-stage problems. The two-stage risk/return profile is clearly dominating, and the difference grows with increasing targets of expected return. Moreover, the authors made backtesting with real market data, solving consecutive problems in a rolling horizon manner. (The single-stage model had a horizon of one month, and the two-stage model two months. In the two-stage case, the optimal first-stage portfolios were evaluated in the rolling horizon scheme.) The performance of the two-stage model proved superior in each examined point of time. The differences are more evident when higher target returns are imposed, forcing the selection of riskier portfolios. (Examining optimal solutions reveals that the two-stage model produces more diversified portfolios.)

According to our knowledge, the solution methods applied in the above mentioned projects do not exploit the special structure of CVaR-optimization problems. Research in this direction 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, however, focuses on classic two-stage models having the decision/observation/decision pattern.

In this paper we propose a decomposition scheme and solution methods for multistage CVaR-minimization or CVaR-constrained problems. The proposed scheme and methods can handle problems with multiple CVaR-constraints for different time frames and at different confidence levels.

The decomposition scheme is based on the Künzi-Bay – Mayer polyhedral representation. The solution methods are special Level-type methods. In the constrained case, the method constructs an approximation of the efficient frontier in the neighborhood of the right-hand-side vector of the constraints.

In a former project, Fábián and Szőke (2006) obtained good solution results for general two-stage problems by first transforming the problems into complete recourse forms. The complete recourse problems were then solved as constrained convex problems. (In such a constrained problem, the constraint function value is the expectation of an infeasibility measure. The constraint function can be evaluated in the same manner as the expected recourse function.) In that approach, feasibility and optimality issues are taken into consideration simultaneously, and regularization extends to both. Though that approach is practicable only if the transformation to complete recourse can be done in a computationally efficient way.

In 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 transform portfolio management problems into complete recourse forms in a computationally efficient way.

With minor modifications, the proposed scheme can be used to decompose further risk constraints in dynamic portfolio management problems. We describe adaptations to

– Integrated Chance Constraints. A generic ICC was formulated by Klein Haneveld et al. (2005).

– Second-order Stochastic Dominance constraints. A single-stage portfolio optimization model involving an SSD constraint was proposed by Dentcheva and Ruszczyński (2006). We propose multistage generalization of the SSD-constrained model.

– constraints involving a special value-of-information risk measure proposed by Pflug (2006).

We suggest application to further financial problems:

– Option pricing. Ryabchenko et al. (2004) propose a grid-based replication model for the pricing of a European option in an incomplete market. We start from this method as a basis but apply a direct approach: we propose a dynamic risk-constrained optimization model.

– Asset-liability management models with special mid-term constraints were proposed by Drijver et al. (2002) and Klein Haneveld et al. (2005). We propose another type of mid-term constraints that serves similar purpose but is easier to handle.

The paper is organized as follows: In Section 1.1 we cite the representation result of Rockafellar and Uryasev, and the polyhedral representation formula of Künzi-Bay and Mayer. In Section 1.2 we present a one-stage prototype model, and sketch the solution method proposed by Künzi-Bay and Mayer.

In Section 2 we present a decomposition framework for the minimization of CVaR in dynamic stochastic models. We propose a solution scheme based on an inexact version of the Level Method of Lemaréchal, Nemirovskii, and Nesterov (1995). The inexact version was proposed by Fábián (2000).

In Section 3 we formulate and decompose a dynamic model with a constraint on end-of-horizon CVaR. In Section 4 we formulate a dynamic model with constraints on the respective risks of the different time periods. We also present a solution scheme based on an asymmetric version of the Constrained Level Method of Lemaréchal, Nemirovskii, and Nesterov (1995). The asymmetric method is described in Appendix B.

The results of the paper are summarized and potential fields of application are suggested in Section 5.

1.1 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.

$w^{\mathcal{B}}$ 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: $w^{\mathcal{B}} - w$. Given a probability α, a heuristic definition of the risk measures is the following.

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

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

Rockafellar and Uryasev (2000) proved that α-VaR and α-CVaR can be computed through the solution of the following problem:

$$\min_{z \in \mathbb{R}} \; z + \frac{1}{1-\alpha} \mathrm{E}\left([w^{\mathcal{B}} - w - z]_+ \right). \tag{1}$$

Namely, VaR is the optimal value of z; and CVaR is the optimal objective value.

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

$$\min_{z \in \mathbb{R}} \; z + \frac{1}{1-\alpha} \sum_{i=1}^{N} p_i \left[w^{\mathcal{B}} - w^{(i)} - z \right]_+. \tag{2}$$

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

$$\begin{gathered} \min \; z + \tfrac{1}{1-\alpha} \sum_{i=1}^{N} p_i y_i \\ \text{such that} \\ z + y_i \geq w^{\mathcal{B}} - w^{(i)}, \quad y_i \geq 0 \qquad (i = 1, \dots, N). \end{gathered} \tag{3}$$

The dual of (3) can be written as

$$\begin{gathered} \max \; w^{\mathcal{B}} - \tfrac{1}{1-\alpha} \sum_{i=1}^{N} \pi_i \, w^{(i)} \\ \text{such that} \\ 0 \leq \pi_i \leq p_i \quad (i = 1, \dots, N), \\ \sum_{i=1}^{N} \pi_i = 1 - \alpha. \end{gathered} \tag{4}$$

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

the dual variable $0 \leq \pi_i \leq p_i$ *can be interpreted as the weight of the* i*th scenario in an event* $\mathcal{E}$*,*

the constraint $\sum_{i=1}^{N} \pi_i = 1 - \alpha$ *determines the probability of the event* $\mathcal{E}$*,*

the term $\frac{1}{1-\alpha} \sum_{i=1}^{N} \pi_i \, w^{(i)}$ *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* $w^{\mathcal{B}}$ *and the above conditional expectation, we find the worst* $(1-\alpha)*100\%$ *cases.*

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

The benchmark wealth $w^{\mathcal{B}}$ *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 wiev, we could set* $w^{\mathcal{B}} = 0$*.*

Rockafellar and Uryasev (2002) showed (in their Proposition 8) that problem (2) can be solved by just sorting the values $w^{(i)}$. The same conclusion is reached through observing that the dual problem (4) has but a single constraint.

Künzi-Bay and Mayer (2006) consider (3) as a two-stage problem: z is the first-stage variable, and y_i are the second-stage variables. Based on this approach, they proposed the following equivalent formulation:

$$\begin{aligned} &\min\ z + \tfrac{1}{1-\alpha} v \\ &\text{such that}\ \ z, v \in \mathbb{R}, \\ &\sum_{i \in \mathcal{I}} p_i \left(w^{\mathcal{B}} - w^{(i)} - z \right) \le v \quad (\mathcal{I} \subset \{1, \dots, N\}) . \end{aligned} \tag{5}$$

A subset $\mathcal{I} \subset \{1, \dots, N\}$ represents the possibility that $w^{\mathcal{B}} - w^{(i)} - z$ is positive for $i \in \mathcal{I}$; and non-positive for $i \notin \mathcal{I}$. The constraint belonging to $\mathcal{I} = \emptyset$ imposes non-negativity of v.

1.2 A static stochastic model

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

$\boldsymbol{x} \in \mathbb{R}^n$ represents a portfolio (i.e., the amounts of money invested in the different assets).

The sum of the components of $\boldsymbol{x}$ should be equal to the initial capital that we denote by w_0. We will formalize this constraint as $\mathbf{1} \cdot \boldsymbol{x} = w_0$. By $\mathbf{1}$ we will denote a vector having 1 in each components ; and the scalar product of vectors will be denoted by a dot.

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.

$\boldsymbol{r}$ denotes the returns for the different assets. This is an n-dimensional random vector.

Assume there are N realizations, $\boldsymbol{r}^{(1)}, \dots, \boldsymbol{r}^{(N)}$, occurring with probabilities $p_1, \dots, p_N$, respectively.

The end-of-period total wealth w is computed by the scalar product $\boldsymbol{r} \cdot \boldsymbol{x}$. This is a random variable with realizations $w^{(i)} = \boldsymbol{r}^{(i)} \cdot \boldsymbol{x}$.

In this paper we assume that no short positions are allowed, i.e., $\boldsymbol{x} \ge \mathbf{0}$ 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. (These are technical assumptions required by the convergence proof of the proposed methods – see Remark 3. The decomposition scheme works without these assumptions, and the methods can be implemented under milder assumptions.)

The decision maker needs to find a balance between expected return $\mathrm{E}(w)$ and risk. The latter will be measured by Conditional-Value-at-Risk $\mathrm{CVaR}(w)$. A customary objective is

$$\max\ \mathrm{E}(w) \ - \lambda\, \mathrm{CVaR}(w), \tag{6}$$

where the parameter $\lambda > 0$ measures *risk aversion* of the decision maker. (We assume a known fixed λ.) The Künzi-Bay – Mayer formulation (5) yields the following polyhedral representation of this problem:

$$\max \ \mathrm{E}(\boldsymbol{r}) \cdot \boldsymbol{x} \ - \lambda \left(z + \tfrac{1}{1-\alpha} v \right)$$
$$\text{such that } \ \boldsymbol{x} \in X, \ \mathbf{1} \cdot \boldsymbol{x} = w_0, \quad z, v \in \mathrm{I\!R}, \tag{7}$$
$$\sum_{i \in \mathcal{I}} p_i \left(w^{\mathcal{B}} - \boldsymbol{r}^{(i)} \cdot \boldsymbol{x} - z \right) \ \leq v \quad (\ \mathcal{I} \subset \{1, \ldots, N\} \ ) \, .$$

Künzi-Bay and Mayer propose a special cutting-plane method for such problems: The constraint belonging to a subset $\mathcal{I}$ is interpreted as a cut. A model problem is constructed that includes only certain cuts. The model problem is solved, and a new cut is added. Namely, the cut that is deepest at the optimal solution. It is easily constructed in the present case. This process is iterated while the depth of the currently added cut is significant. (Künzi-Bay and Mayer observe that from a purely mathematical point of view, their method can be considered as a version of the general method of Klein Haneveld and Van der Vlerk (2006), as specialized for CVaR minimization.)

Künzi-Bay and Mayer implemented their method, and solved several CVaR problems with their specialized solver called CVaRMin. Their experimental results clearly demonstrate the superiority of CVaRMin over general-purpose linear or stochastic problem solvers. For the largest test problems, CVaRMin was by at least one order of magnitude faster then either of the other solvers involved.

In the sequel we extend the Künzi-Bay – Mayer decomposition idea to multistage models.

2 A dynamic prototype model

Topaloglou, Vladimirou, Zenios (2004) developed elaborate multistage financial models. The objective is minimizing end-of-horizon CVaR under a constraint on expected end-of-horizon return. (These models take into account transaction costs and hedging possibilities.) From an algorithmic point of view, these financial problems fit the prototype to be described in this section. Hence the special decomposition framework and solution method we propose can also be applied to these financial problems.

In order to keep notation as simple as possible we describe a three-stage model that easily generalizes to further stages.

w_0 represents initial capital. This is a given parameter.

$\boldsymbol{x}_1$ represents the portfolio selected at the beginning of the first time period. This is a decision vector. Feasible portfolios are represented by $\boldsymbol{x}_1 \in X, \ \mathbf{1} \cdot \boldsymbol{x}_1 = w_0$.

$\boldsymbol{r}_1$ denotes the random return occurring in course of the first time period.

$w_1 := \boldsymbol{r}_1 \cdot \boldsymbol{x}_1$ denotes the wealth at the end of the first time period.

$\boldsymbol{x}_2$ represents the portfolio selected at the beginning of the second time period. This is a decision vector. Feasible portfolios are represented by $\boldsymbol{x}_2 \in X,\ \mathbf{1} \cdot \boldsymbol{x}_2 = w_1$.

$\boldsymbol{r}_2$ denotes random return occurring in course of the second time period.

$w_2 := \boldsymbol{r}_2 \cdot \boldsymbol{x}_2$ denotes the wealth at the end of the second time period.

$\boldsymbol{x}_3$ represents the portfolio selected at the beginning of the third time period. This is a decision vector. Feasible portfolios are represented by $\boldsymbol{x}_3 \in X,\ \mathbf{1} \cdot \boldsymbol{x}_3 = w_2$.

$\boldsymbol{r}_3$ denotes the random return occurring in course of the third time period.

$w_3 := \boldsymbol{r}_3 \cdot \boldsymbol{x}_3$ denotes the wealth at the end of the third time period.

Description of the random parameters. Assume that the (discrete finite) joint distribution of the random vectors $\boldsymbol{r}_1$, $\boldsymbol{r}_2$, and $\boldsymbol{r}_3$ is known:

The realizations of $\boldsymbol{r}_1$ are $\boldsymbol{r}_1^{(i)}$ $(1 \leq i \leq N)$, occurring with respective probabilities $p_1^{(i)}$.

For $1 \leq i \leq N$, the realizations of $\boldsymbol{r}_2$ given that $\boldsymbol{r}_1 = \boldsymbol{r}_1^{(i)}$

are $\boldsymbol{r}_2^{(ij)}$ $(1 \leq j \leq M^{(i)})$, occurring with respective probabilities $p_2^{(ij)}$.

For $1 \leq i \leq N,\ 1 \leq j \leq M^{(i)}$, the realizations of $\boldsymbol{r}_3$ given that $\boldsymbol{r}_1 = \boldsymbol{r}_1^{(i)}$ and $\boldsymbol{r}_2 = \boldsymbol{r}_2^{(ij)}$

are $\boldsymbol{r}_3^{(ijk)}$ $(1 \leq k \leq K^{(ij)})$, occurring with respective probabilities $p_3^{(ijk)}$.

Scenarios can be identified by the elements of the set

$$\mathcal{S} := \left\{ (i,j,k) \;\middle|\; 1 \leq i \leq N,\ 1 \leq j \leq M^{(i)},\ 1 \leq k \leq K^{(ij)} \right\},$$

where (i,j,k) means that the first-, second-, and third-period returns $\boldsymbol{r}_1^{(i)}$, $\boldsymbol{r}_2^{(ij)}$, and $\boldsymbol{r}_3^{(ijk)}$ realize. The probability of this event will be denoted by

$$p_{(ijk)} := p_1^{(i)} p_2^{(ij)} p_3^{(ijk)} \qquad ((i,j,k) \in \mathcal{S}).$$

2.1 Problem formulation

We try to find a balance between end-of-horizon expected return and Conditional-Value-at-Risk. Denoting end-of-horizon wealth by $w := w_3$, the problem can be formulated as

$$\max\ \mathrm{E}(w)\ - \lambda\, \mathrm{CVaR}(w). \tag{8}$$

We assume a known fixed λ.

The Künzi-Bay – Mayer polyhedral representation of this problem is:

$$
\begin{aligned}
\max \quad & \sum_{(i,j,k)\in\mathcal{S}} p_{(ijk)}\, \boldsymbol{r}_3^{(ijk)} \cdot \boldsymbol{x}_3^{(ij)} \quad - \lambda \left(z + \tfrac{1}{1-\alpha} v \right) \\
\text{such that} \quad & \boldsymbol{x}_1 \in X, \quad \mathbf{1} \cdot \boldsymbol{x}_1 = w_0, \\
\boldsymbol{x}_2^{(i)} \in X, \quad \mathbf{1} \cdot \boldsymbol{x}_2^{(i)} \; &= \; \boldsymbol{r}_1^{(i)} \cdot \boldsymbol{x}_1 \qquad (i = 1, \ldots, N), \\
\boldsymbol{x}_3^{(ij)} \in X, \quad \mathbf{1} \cdot \boldsymbol{x}_3^{(ij)} \; &= \; \boldsymbol{r}_2^{(ij)} \cdot \boldsymbol{x}_2^{(i)} \qquad (i = 1, \ldots, N,\ j = 1, \ldots, M^{(i)}), \\
z, v \in \mathbb{R}, \qquad & \sum_{(i,j,k)\in\mathcal{R}} p_{(ijk)} \left(w^{\mathcal{B}} - \boldsymbol{r}_3^{(ijk)} \cdot \boldsymbol{x}_3^{(ij)} - z \right) \le v \quad (\mathcal{R} \subset \mathcal{S}).
\end{aligned}
\tag{9}
$$

Decomposition. The first-stage problem will be:

$$
\begin{aligned}
\max \quad & \sum_{i=1}^{N} p_1^{(i)}\, \mathcal{D}^{(i)} \left(\boldsymbol{r}_1^{(i)} \cdot \boldsymbol{x}_1, \;\; z \right) \quad - \lambda z \\
\text{such that} \quad & \boldsymbol{x}_1 \in X, \quad \mathbf{1} \cdot \boldsymbol{x}_1 = w_0, \\
& z \in \mathbb{R},
\end{aligned}
\tag{10}
$$

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

$$
\begin{aligned}
\mathcal{D}^{(i)}(\omega, \zeta) := \quad \max \quad & \sum_{j=1}^{M^{(i)}} p_2^{(ij)}\, \mathcal{D}^{(ij)} \left(\boldsymbol{r}_2^{(ij)} \cdot \boldsymbol{x}_2, \;\; \zeta \right) \\
\text{such that} \quad & \boldsymbol{x}_2 \in X, \quad \mathbf{1} \cdot \boldsymbol{x}_2 = \omega,
\end{aligned}
\tag{11}
$$

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

$$
\begin{aligned}
\mathcal{D}^{(ij)}(\omega, \zeta) := \quad \max \quad & \sum_{k=1}^{K^{(ij)}} p_3^{(ijk)}\, \boldsymbol{r}_3^{(ijk)} \cdot \boldsymbol{x}_3 \quad - \lambda \tfrac{1}{1-\alpha} v \\
\text{such that} \quad & \boldsymbol{x}_3 \in X, \quad \mathbf{1} \cdot \boldsymbol{x}_3 = \omega, \\
v \in \mathbb{R}, \qquad & \sum_{k\in\mathcal{K}} p_3^{(ijk)} \left(w^{\mathcal{B}} - \boldsymbol{r}_3^{(ijk)} \cdot \boldsymbol{x}_3 - \zeta \right) \le v \quad \left(\mathcal{K} \subset \{1, \ldots, K^{(ij)}\} \right).
\end{aligned}
\tag{12}
$$

Proposition 2 *The three-stage problem (10 - 12) is equivalent to the polyhedral representation problem (9).*

As a heuristic proof, let us observe that in each $\mathcal{D}^{(ij)}$ function, the parameter ω represents initial wealth for the remaining period of time, and ζ represents a 'tolerable' loss level tentatively set in the first-stage problem. The optimal value of the variable v is the (conditional) expectation of losses above the 'tolerable' level. – A formal proof can be found in Appendix A.1.

2.2 Solution method

In view of the effectiveness of the solution method developed by Künzi-Bay and Mayer (2006), we propose a generalized version.

Solution of the third-stage problems. We assume a fixed (i, j), $1 \leq i \leq N$, $1 \leq j \leq M^{(i)}$. Given the parameter values $\hat{\omega}, \hat{\zeta}$, and a tolerance $\epsilon_3 > 0$, any cutting-plane or bundle-type method can be used to find an ϵ_3-optimal solution for the problem (12: $\omega = \hat{\omega}, \zeta = \hat{\zeta}$). Such a method builds successive cutting-plane models of the objective function.

In the present case, the model problems contain only certain cuts $\mathcal{K} \subset \{1, \ldots, K^{(ij)}\}$. The genuine variables are $\boldsymbol{x}_3 \in X$ only, since v just marks the upper cover of the existing cuts. The objective functions of the respective model problems are concave upper approximations of the original objective function over the set of the feasible $\boldsymbol{x}_3$ values. These model problems are solved and used to construct iterate solutions. Let $\boldsymbol{x}_3{}^\star \in X$ denote the current iterate, and $\mathcal{O}^\star$ the corresponding objective value in the original problem (12: $\omega = \hat{\omega}, \zeta = \hat{\zeta}$). Obviously we have $\mathcal{D}^{(ij)}(\hat{\omega}, \hat{\zeta}) \geq \mathcal{O}^\star$.

On the other hand, the model problems can be considered as linear programming problems that have the parameters ω, ζ in the right-hand side. (In this view the variables are $(\boldsymbol{x}_3, v)$.) Let $\overline{\mathcal{D}}^{(ij)}(\omega, \zeta)$ denote the optimum of the current model problem as a function of the parameters. $\overline{\mathcal{D}}^{(ij)}(\omega, \zeta)$ is obviously a polyhedral concave upper approximating function of $\mathcal{D}^{(ij)}(\omega, \zeta)$.

In course of the solution of the present problem (12: $\omega = \hat{\omega}, \zeta = \hat{\zeta}$), the model problems are solved with the parameter setting $\omega = \hat{\omega}, \zeta = \hat{\zeta}$ only. We have $\overline{\mathcal{D}}^{(ij)}(\hat{\omega}, \hat{\zeta}) \geq \mathcal{D}^{(ij)}(\hat{\omega}, \hat{\zeta}) \geq \mathcal{O}^\star$ at any iterate, and the method terminates when $\mathcal{O}^\star \geq \overline{\mathcal{D}}^{(ij)}(\hat{\omega}, \hat{\zeta}) - \epsilon_3$ holds. Let us consider the final model problem (i.e., let the current model problem be the terminal one). Using the optimal dual variables, we can construct a linear support function to $\overline{\mathcal{D}}^{(ij)}(\omega, \zeta)$ at $(\hat{\omega}, \hat{\zeta})$. Let $L^{(ij)}(\omega, \zeta)$ denote this support function. From the termination criterion, we have $\overline{\mathcal{D}}^{(ij)}(\hat{\omega}, \hat{\zeta}) \geq \mathcal{D}^{(ij)}(\hat{\omega}, \hat{\zeta}) \geq \overline{\mathcal{D}}^{(ij)}(\hat{\omega}, \hat{\zeta}) - \epsilon_3$. It follows that $L^{(ij)}(\omega, \zeta)$ is an ϵ_3-support function to $\mathcal{D}^{(ij)}(\omega, \zeta)$ at $(\hat{\omega}, \hat{\zeta})$.

Remark 3 *We can construct a single upper bound on the slopes of such support functions:*

In the linear programming model problem, the parameter ζ appears in the right-hand sides of the polyhedral cuts (with coefficients ≤ 1). The corresponding dual variables are non-negative, and their sum is $\lambda \frac{1}{1-\alpha}$.

Another consequence of ζ appearing in the right-hand sides, is that $\overline{\mathcal{D}}(\omega,\zeta)$ is monotone increasing in ζ. Due to the assumed homogeneity of X, and the possible setting $w^{\mathcal{B}} = 0$ (see Remark 1), we have $\overline{\mathcal{D}}(\rho\omega,\rho\zeta) = \rho\overline{\mathcal{D}}(\omega,\zeta)$ for $\rho > 0$.

Considering the ω-direction slope, let $o < \omega$. From the above mentioned properties of $\overline{\mathcal{D}}$, we have

$$\overline{\mathcal{D}}(\omega,\zeta) - \overline{\mathcal{D}}(o,\zeta) \;\le\; \overline{\mathcal{D}}(\omega,\zeta) - \overline{\mathcal{D}}\left(\frac{o}{\omega}\omega, \frac{o}{\omega}\zeta\right) \;=\; \left(1 - \frac{o}{\omega}\right)\overline{\mathcal{D}}(\omega,\zeta) \;\le\; \left(1 - \frac{o}{\omega}\right) \max_{(i,j,k)\in\mathcal{S}} \left\|\boldsymbol{r}_3^{(ijk)}\right\| \max_{\boldsymbol{1}\cdot\boldsymbol{x}_3=\omega} \|\boldsymbol{x}_3$$

Due to the assumption that no short positions are allowed, we have $\|\boldsymbol{x}_3\| \le \varrho\omega$, with a constant ϱ depending only on the dimension of the vector $\boldsymbol{x}_3$. Hence

$$\overline{\mathcal{D}}(\omega,\zeta) - \overline{\mathcal{D}}(o,\zeta) \;\le\; (\omega - o)\,\varrho \max_{(i,j,k)\in\mathcal{S}} \left\|\boldsymbol{r}_3^{(ijk)}\right\|.$$

Solution of the second-stage problems. We assume a fixed i, $1 \le i \le N$. Given the parameter values $\hat{\omega}, \hat{\zeta}$, and a tolerance $\epsilon_2 > 0$, we can use an approximate method to find an ϵ_2-optimal solution for the problem (11: $\omega = \hat{\omega}, \zeta = \hat{\zeta}$). We recommend an inexact version of the Level Method of Lemaréchal, Nemirovskii, and Nesterov (1995). The 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. It needs an *oracle* to provide objective function data, in the form of ϵ_3-support functions. The accuracy tolerance ϵ_3 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.)

In the present case, the oracle includes the solution of the third-stage problems (i,j), $1 \le j \le M^{(i)}$: As we have seen in the previous paragraph, we can construct an ϵ_3-support function to the third-stage function $\mathcal{D}^{(ij)}$ at any given point from $\mathbb{R}^2$. By aggregating appropriate support functions, we obtain an ϵ_3-support function to the second-stage objective function. The lower cover of existing cuts gives a polyhedral concave upper approximating model of the objective function. (Convergence proof of the method requires an upper bound on the slopes of the support functions. Remark 3 provides such an upper bound.)

The model problems can be considered as linear programming problems that have the parameters ω, ζ in the right-hand side. (A new variable needs to introduced to mark the lower cover of existing cuts.) Let $\overline{\mathcal{D}}^{(i)}(\omega,\zeta)$ denote the optimum of the current model problem as a function of the parameters. This is a polyhedral concave upper approximating function of $\mathcal{D}^{(i)}(\omega,\zeta)$.

Using optimal dual variables of the final model problem, we can construct an ϵ_2-support function to $\mathcal{D}^{(i)}(\omega,\zeta)$ at $(\hat{\omega},\hat{\zeta})$, like we did in the third stage.

Remark 4 *We can construct a single upper bound on the slopes of such second-stage support functions.*

The bound on the ζ-direction slope is inherited from the third-stage support functions.

The construction of the bound on the ω-direction slope is analogous to that described in Remark 3.

Solution of the first-stage problem. The Inexact Level Method can be used to find an ϵ_1-optimal solution. (The tolerance $\epsilon_1 > 0$ is a prescribed by the decision maker.)

3 Decomposing a CVaR constraint

We will maximize end-of-horizon return under a constraint on end-of-horizon Conditional-Value-at-Risk. Formally, the problem is

$$\max \ \mathrm{E}(w) \qquad \text{such that} \quad \mathrm{CVaR}(w) \leq \gamma, \tag{13}$$

where γ is set by the decision maker. We assume that the problem is feasible, and the Slater condition holds.

The parameter γ can be calibrated by exploring the efficient frontier. This is a concave curve in the (risk, return) coordinate system. Given $\lambda > 0$, a supporting line of slope λ can be constructed by solving the unconstrained problem (8). 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, decision maker ought to take into consideration the slope of the efficient frontier also. (Slope is interpreted as risk aversion.)

Let $\lambda^\star > 0$ be such that the supporting line of slope $\lambda^\star$ touches the curve at a point whose risk-coordinate is γ. (Of course there may be other touching points beside this.) If we should know $\lambda^\star$, we could reduce the constrained problem to the unconstrained problem (8: $\lambda = \lambda^\star$).

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

Assume that an upper bound $\overline{\lambda}$ of reasonable magnitude is known. Let us transform the CVaR-constrained problem (13) into the following form:

$$\max \quad \mathrm{E}(w) \ - \overline{\lambda}\left[\, \mathrm{CVaR}(w) - \gamma \,\right]_+ . \tag{14}$$

Proposition 5 *The set of the optimal solutions of the constrained problem (13) is identical to the set of the optimal solutions of the penalized problem (14).*

Proof. Let us contrast the efficient frontier with the contour lines of the objective function of (14) in the (risk, return) coordinate system. □

The polyhedral representation of problem (14) is:

$$\begin{array}{lll} & \max \quad \sum_{(i,j,k)\in\mathcal{S}} p_{(ijk)}\, \boldsymbol{r}_3^{(ijk)} \cdot \boldsymbol{x}_3^{(ij)} \quad - \overline{\lambda} y & \\ & \text{such that} & \\ & \vdots & \\ z, v \in \mathrm{I\!R}, & \sum_{(i,j,k)\in\mathcal{R}} p_{(ijk)} \left(w^{\mathcal{B}} - \boldsymbol{r}_3^{(ijk)} \cdot \boldsymbol{x}_3^{(ij)} - z \right) \leq v \quad (\, \mathcal{R} \subset \mathcal{S} \,), & \\ & y \in \mathrm{I\!R},\ \ y \geq 0, \qquad z + \frac{1}{1-\alpha} v - y \leq \gamma, & \end{array} \tag{15}$$

where dots stand for the cash balance equations identical to those in (9).

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 in any feasible solution of (15), and that $y = 0$ holds in any optimal solution, we may also add the bound $\tilde{v} \geq 0$. We get:

$$\begin{array}{c} \max \quad \sum_{(i,j,k)\in\mathcal{S}} p_{(ijk)}\, \boldsymbol{r}_3^{(ijk)} \cdot \boldsymbol{x}_3^{(ij)} \quad - \overline{\lambda}\, \frac{1}{1-\alpha} \tilde{y} \\ \text{such that} \quad \boldsymbol{x}_1 \in X, \quad \boldsymbol{1} \cdot \boldsymbol{x}_1 = w_0, \\ \boldsymbol{x}_2^{(i)} \in X, \quad \boldsymbol{1} \cdot \boldsymbol{x}_2^{(i)} = \boldsymbol{r}_1^{(i)} \cdot \boldsymbol{x}_1 \qquad (i = 1, \ldots, N), \\ \boldsymbol{x}_3^{(ij)} \in X, \quad \boldsymbol{1} \cdot \boldsymbol{x}_3^{(ij)} = \boldsymbol{r}_2^{(ij)} \cdot \boldsymbol{x}_2^{(i)} \qquad (i = 1, \ldots, N,\ j = 1, \ldots, M_i), \\ z, \tilde{v} \in \mathrm{I\!R},\ \ \tilde{v} \geq 0, \qquad z + \frac{1}{1-\alpha} \tilde{v} \leq \gamma, \\ \tilde{y} \in \mathrm{I\!R},\ \ \tilde{y} \geq 0, \qquad \sum_{(i,j,k)\in\mathcal{R}} p_{(ijk)} \left(w^{\mathcal{B}} - \boldsymbol{r}_3^{(ijk)} \cdot \boldsymbol{x}_3^{(ij)} - z \right) - \tilde{y} \leq \tilde{v} \quad (\, \mathcal{R} \subset \mathcal{S} \,). \end{array} \tag{16}$$

(We interchanged the last two rows.) The aim of introducing the slack variable was to ensure that the forthcoming decomposed problem would have a (relatively) complete recourse.

In the decomposed form, the first-stage problem will be:

$$\begin{array}{l} \max \quad \sum_{i=1}^{N} p_1^{(i)}\, \mathcal{F}^{(i)} \left(\boldsymbol{r}_1^{(i)} \cdot \boldsymbol{x}_1,\ \ z,\ v_1^{(i)} \right) \\ \text{such that} \quad \boldsymbol{x}_1 \in X, \quad \boldsymbol{1} \cdot \boldsymbol{x}_1 = w_0, \\ z \in \mathrm{I\!R}, \quad v_1^{(i)} \in \mathrm{I\!R},\ v_1^{(i)} \geq 0\ \ (i = 1, \ldots, N), \qquad z + \frac{1}{1-\alpha} \sum_{i=1}^{N} p_1^{(i)} v_1^{(i)} \leq \gamma, \end{array} \tag{17}$$

where the functions $\mathcal{F}^{(i)} : \mathrm{I\!R}^3 \to \mathrm{I\!R}$ $(i = 1, \ldots, N)$ are defined by the second-stage problem:

$$\mathcal{F}^{(i)}(\omega,\zeta,\nu) := \max \quad \sum_{j=1}^{M^{(i)}} p_2^{(ij)}\, \mathcal{F}^{(ij)}\left(\boldsymbol{r}_2^{(ij)} \cdot \boldsymbol{x}_2,\ \ \zeta,\ v_2^{(j)} \right)$$

$$\text{such that} \quad \boldsymbol{x}_2 \in X, \quad \mathbf{1} \cdot \boldsymbol{x}_2 = \omega, \tag{18}$$

$$v_2^{(j)} \in \mathrm{I\!R},\ v_2^{(j)} \geq 0 \ \ (j = 1, \ldots, M^{(i)}), \qquad \sum_{j=1}^{M^{(i)}} p_2^{(ij)} v_2^{(j)} = \nu,$$

where the functions $\mathcal{F}^{(ij)} : \mathrm{I\!R}^3 \to \mathrm{I\!R}$ $\ (i = 1, \ldots, N,\ j = 1, \ldots, M^{(i)})$ are defined by the third-stage problem:

$$\mathcal{F}^{(ij)}(\omega,\zeta,\nu) := \max \quad \sum_{k=1}^{K^{(ij)}} p_3^{(ijk)}\, \boldsymbol{r}_3^{(ijk)} \cdot \boldsymbol{x}_3 \quad - \overline{\lambda}\, \frac{1}{1-\alpha} y$$

$$\text{such that} \quad \boldsymbol{x}_3 \in X, \quad \mathbf{1} \cdot \boldsymbol{x}_3 = \omega,$$

$$y \in \mathrm{I\!R},\ \ y \geq 0, \qquad \sum_{k \in \mathcal{K}} p_3^{(ijk)} \left(w^{\mathcal{B}} - \boldsymbol{r}_3^{(ijk)} \cdot \boldsymbol{x}_3 - \zeta \right) - y \ \leq \nu \quad \left(\mathcal{K} \subset \{1, \ldots, K^{(ij)}\} \right) \tag{19}$$

Proposition 6 *The three-stage problem (17 - 19) is equivalent to the polyhedral representation problem (16).*

As a heuristic proof, let us observe that in each $\mathcal{F}^{(ij)}$ function, the parameter ω represents initial wealth for the remaining period of time, ζ represents a 'tolerable' loss level, and ν represents an upper bound on the (conditional) expectation of losses above the 'tolerable' level. (The 'tolerable' loss level ζ is tentatively set in the first-stage problem, and the bound ν is tentatively set in the parent problem.) – A formal proof can be found in Appendix A.2.

4 Multiple CVaR-constraints

Jobst and Zenios (2001) describe an experiment: they solved one-stage portfolio selection problems with the same time period of one year. The objective was the minimization of a risk measure. As risk measure, they used Mean Absolute Deviation (MAD) in one experiment, and CVaR in another experiment. 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.

With the CVaR-optimal portfolio, worst losses were limited to around 10 % of the total wealth throughout. Even so, the experimental distribution of the 9-month losses had a tail significantly heavier than the experimental distribution of the end-of-horizon losses had.

We suggest shaping the distributions of intermediate returns by imposing additional CVaR-constraints on the respective yields of the different time periods.

As an example let $w' := w_2$ denote the wealth at the end of the second time-period, and let us consider the problem

$$\max \ \mathrm{E}(w) \quad \text{such that} \quad \mathrm{CVaR}(w) \leq \gamma, \quad \mathrm{CVaR}(w') \leq \gamma', \tag{20}$$

where (γ, γ') are set by the decision maker. We assume that the problem is feasible and the Slater condition holds.

Remark 7 *Let us observe that the forthcoming discussion easily adapts to the case when $w' := w$, and we wish to impose an additional CVaR constraint (with different confidence level) on the end-of-horizon risk.*

Moreover, the simultaneous handling of several CVaR constraints is possible in the same manner.

Calibration. The parameters can be calibrated by exploring the efficient frontier. This is a concave surface in the $(\mathrm{CVaR}(w), \mathrm{CVaR}(w'), \mathrm{E}(w))$ space. Given $(\lambda, \lambda') > \mathbf{0}$, a supporting plane to the efficient frontier can be constructed by solving the unconstrained problem

$$\max \ \mathrm{E}(w) \ - \lambda \, \mathrm{CVaR}(w) \ - \lambda' \mathrm{CVaR}(w'). \tag{21}$$

(The decomposition scheme and solution methods described in Section 2 generalize to the above problem in a straightforward manner.) The lower cover of such supporting planes is an upper approximation of the efficient frontier.

Assume that unconstrained problems of the type (21) have been solved with the parameters $(\lambda_\ell, \lambda'_\ell)$ $(\ell = 1, \dots, L)$, and let $(\varsigma_\ell, \varsigma'_\ell, \eta_\ell)$ denote the $(\mathrm{CVaR}(w), \mathrm{CVaR}(w'), \mathrm{E}(w))$ coordinates of the respective optimal solutions. The convex hull

$$\mathcal{C} := \mathrm{Conv} \, \{ \, (\varsigma_\ell, \varsigma'_\ell, \eta_\ell) \ : \ 1 \leq \ell \leq L \, \}$$

gives a lower approximation of the efficient frontier.

Using these approximations of the efficient frontier, the decision maker will be able to set the parameters (γ, γ') in problem (20). 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 $(\lambda^\star, \lambda^{\star\prime}) > \mathbf{0}$ be such that the supporting plane of slope $(\lambda^\star, \lambda^{\star\prime})$ touches the frontier at a point whose risk-coordinates are (γ, γ'). (Of course there may be other touching points beside this.) If we should know $(\lambda^\star, \lambda^{\star\prime})$, then we could reduce the constrained problem to the unconstrained problem (21: $\lambda = \lambda^\star$, $\lambda' = \lambda^{\star\prime}$). Although the exact values $(\lambda^\star, \lambda^{\star\prime})$ are not known, they should be of a reasonable magnitude in a correctly calibrated model. (An excessively large $\lambda^\star$ or $\lambda^{\star\prime}$ would imply unreasonable risk aversion.)

Remark 8 *The above defined objects can be used in constructing upper bounds for the slopes: From the upper approximation of the efficient frontier, we obtain an upper bound $\overline{\eta}$ for the optimal objective value of (20). Suppose that $\mathcal{C}$ contains a point $(\varsigma, \varsigma', \eta)$ such that $\varsigma < \gamma$ and $\varsigma' = \gamma'$. Then $(\overline{\eta} - \eta)/(\gamma - \varsigma)$ is clearly an upper bound for $\lambda^\star$. (The convex hull $\mathcal{C}$ can be extended in a desired direction by solving the unconstrained problem (21) with appropriately tuned parameters.)*

Equivalent formulation. Assume that upper bounds $\overline{\lambda} > \lambda^\star$ and $\overline{\lambda}' > \lambda^{\star\prime}$ of reasonable magnitude are known, and let us consider the problem:

$$\max \quad \mathrm{E}(w) \; - \overline{\lambda}[\,\mathrm{CVaR}(w) - \gamma\,]_+ \; - \overline{\lambda}'[\,\mathrm{CVaR}(w') - \gamma'\,]_+. \tag{22}$$

Proposition 9 *The set of the optimal solutions of the constrained problem (20) is identical to the set of the optimal solutions of the penalized problem (22).*

A formal proof can be found in Appendix A.3.

The polyhedral representation of problem (22) can be constructed in a way analogous to (15) in Section 3: Let us introduce the notation

$$\mathcal{S}' := \left\{ (i,j) \;\middle|\; 1 \le i \le N,\; 1 \le j \le M^{(i)} \right\} \quad \text{and} \quad p'_{(ij)} := p_1^{(i)} p_2^{(ij)} \quad ((i,j) \in \mathcal{S}'),$$

and extend (15) by adding the variables and cuts

$$z', v' \in \mathrm{I\!R}, \qquad \sum_{(i,j)\in\mathcal{Q}} p'_{(ij)} \left(w^{\mathcal{B}} - \boldsymbol{r}_2^{(ij)} \cdot \boldsymbol{x}_2^{(i)} - z' \right) \le v' \quad (\,\mathcal{Q} \subset \mathcal{S}'\,),$$

$$y' \in \mathrm{I\!R}, \;\; y' \ge 0, \qquad z' + \tfrac{1}{1-\alpha} v' - y' \le \gamma'.$$

Moreover, the term $-\overline{\lambda}' y'$ is added to the objective function.

Carrying on the analogy to Section 3, the extended problem can be transformed into the form

$$
\begin{array}{c}
\max \quad \sum_{(i,j,k)\in\mathcal{S}} p_{(ijk)}\, \boldsymbol{r}_3^{(ijk)} \cdot \boldsymbol{x}_3^{(ij)} \quad - \overline{\lambda}\, \frac{1}{1-\alpha}\tilde{y} \quad - \overline{\lambda}'\, \frac{1}{1-\alpha}\tilde{y}' \\
\text{such that} \quad \boldsymbol{x}_1 \in X, \quad \mathbf{1}\cdot\boldsymbol{x}_1 = w_0, \\
\boldsymbol{x}_2^{(i)} \in X, \quad \mathbf{1}\cdot\boldsymbol{x}_2^{(i)} \;=\; \boldsymbol{r}_1^{(i)}\cdot\boldsymbol{x}_1 \qquad (i=1,\ldots,N), \\
\boldsymbol{x}_3^{(ij)} \in X, \quad \mathbf{1}\cdot\boldsymbol{x}_3^{(ij)} \;=\; \boldsymbol{r}_2^{(ij)}\cdot\boldsymbol{x}_2^{(i)} \qquad (i=1,\ldots,N,\ j=1,\ldots,M_i), \\
z, \tilde{v} \in \mathbb{R}, \ \ \tilde{v} \ge 0, \qquad z \;+\frac{1}{1-\alpha}\tilde{v} \;\le \gamma, \\
\tilde{y} \in \mathbb{R}, \ \ \tilde{y} \ge 0, \qquad \sum_{(i,j,k)\in\mathcal{R}} p_{(ijk)} \left(w^{\mathcal{B}} - \boldsymbol{r}_3^{(ijk)}\cdot\boldsymbol{x}_3^{(ij)} \;- z \right) - \tilde{y} \;\le \tilde{v} \quad (\,\mathcal{R}\subset\mathcal{S}\,), \\
z', \tilde{v}' \in \mathbb{R}, \ \ \tilde{v}' \ge 0, \qquad z' \;+\frac{1}{1-\alpha}\tilde{v}' \;\le \gamma', \\
\tilde{y}' \in \mathbb{R}, \ \ \tilde{y}' \ge 0, \qquad \sum_{(i,j)\in\mathcal{Q}} p'_{(ij)} \left(w^{\mathcal{B}} \;- \boldsymbol{r}_2^{(ij)}\cdot\boldsymbol{x}_2^{(i)} \;- z' \right) - \tilde{y}' \;\le \tilde{v}' \quad (\,\mathcal{Q}\subset\mathcal{S}'\,).
\end{array}
\tag{23}
$$

The above problem yields a relatively-complete-recourse decomposition.

Decomposition. The first-stage problem will be:

$$
\begin{array}{l}
\max \quad \sum_{i=1}^{N} p_1^{(i)}\, \mathcal{F}^{(i)}\left(\,\boldsymbol{r}_1^{(i)}\cdot\boldsymbol{x}_1, \ \ z, \ v_1^{(i)}, \ z', \ v_1^{(i)\prime} \right) \\
\text{such that} \quad \boldsymbol{x}_1 \in X, \quad \mathbf{1}\cdot\boldsymbol{x}_1 = w_0, \\
z \in \mathbb{R}, \quad v_1^{(i)} \in \mathbb{R}, \ v_1^{(i)} \ge 0 \ \ (i=1,\ldots,N), \qquad z + \frac{1}{1-\alpha}\sum_{i=1}^{N} p_1^{(i)} v_1^{(i)} \;\le \gamma, \\
z' \in \mathbb{R}, \quad v_1^{(i)\prime} \in \mathbb{R}, \ v_1^{(i)\prime} \ge 0 \ \ (i=1,\ldots,N), \qquad z' + \frac{1}{1-\alpha}\sum_{i=1}^{N} p_1^{(i)} v_1^{(i)\prime} \;\le \gamma',
\end{array}
\tag{24}
$$

where the functions $\mathcal{F}^{(i)} : \mathrm{I\!R}^5 \to \mathrm{I\!R} \quad (i = 1, \ldots, N)$ are defined by the second-stage problem:

$$\begin{aligned}
\mathcal{F}^{(i)}(\omega, \zeta, \nu, \zeta', \nu') := \quad & \max \quad \sum_{j=1}^{M^{(i)}} p_2^{(ij)} \, \mathcal{F}^{(ij)} \left(\boldsymbol{r}_2^{(ij)} \cdot \boldsymbol{x}_2, \; \zeta, \; v_2^{(j)} \right) \quad - \overline{\lambda}' \, \tfrac{1}{1-\alpha} y' \\
& \text{such that} \quad \boldsymbol{x}_2 \in X, \quad \boldsymbol{1} \cdot \boldsymbol{x}_2 = \omega, \\
& v_2^{(j)} \in \mathrm{I\!R}, \; v_2^{(j)} \geq 0 \;\; (j = 1, \ldots, M^{(i)}), \qquad \sum_{j=1}^{M^{(i)}} p_2^{(ij)} v_2^{(j)} = \nu, \\
& y' \in \mathrm{I\!R}, \;\; y' \geq 0, \qquad \sum_{j \in \mathcal{J}} p_2^{(ij)} \left(w^{\mathcal{B}} - \boldsymbol{r}_2^{(ij)} \cdot \boldsymbol{x}_2 - \zeta' \right) - y' \; \leq \nu' \\
& \qquad \left(\mathcal{J} \subset \{1, \ldots, M^{(i)}\} \right),
\end{aligned} \tag{25}$$

where the functions $\mathcal{F}^{(ij)} : \mathrm{I\!R}^3 \to \mathrm{I\!R} \quad (i = 1, \ldots, N, \; j = 1, \ldots, M^{(i)})$ are defined by the third-stage problem:

$$\begin{aligned}
\mathcal{F}^{(ij)}(\omega, \zeta, \nu) := \quad & \max \quad \sum_{k=1}^{K^{(ij)}} p_3^{(ijk)} \, \boldsymbol{r}_3^{(ijk)} \cdot \boldsymbol{x}_3 \quad - \overline{\lambda} \, \tfrac{1}{1-\alpha} y \\
& \text{such that} \quad \boldsymbol{x}_3 \in X, \quad \boldsymbol{1} \cdot \boldsymbol{x}_3 = \omega, \\
& y \in \mathrm{I\!R}, \;\; y \geq 0, \qquad \sum_{k \in \mathcal{K}} p_3^{(ijk)} \left(w^{\mathcal{B}} - \boldsymbol{r}_3^{(ijk)} \cdot \boldsymbol{x}_3 - \zeta \right) - y \; \leq \nu \quad \left(\mathcal{K} \subset \{1, \ldots, K^{(ij)}\} \right)
\end{aligned} \tag{26}$$

Proposition 10 *The three-stage problem (24 - 26) is equivalent to the polyhedral representation problem (23).*

The proof is analogous to the formal proof of Proposition 6. Equivalence can be established separately between

- the $(\mathcal{R} \subset \mathcal{S})$-cuts in problem (23) and the aggregates of $\mathcal{K}^{(ij)}$-cuts from the different third-stage problems, and
- the $(\mathcal{Q} \subset \mathcal{S}')$-cuts in problem (23) and the aggregates of $\mathcal{J}^{(i)}$-cuts from the different second-stage problems.

Non-negativity of the variables $v_1^{(i)}, v_2^{(ij)}, y^{(ij)}$ on the one hand, and of $v_1^{(i)\prime}, y^{(i)\prime}$ on the other hand, can also be treated separately.

4.1 Framework of the solution method

The framework will be the same as the one described in Section 2.2: Suppose we want to find an near-optimal solution of the first-stage problem. (The accuracy is prescribed by the decision maker.) The first-stage solution method will require objective function information in the form of approximate cuts of increasing accuracy.

In order to find an ϵ-supporting linear function to the first-stage objective, we must find ϵ-optimal solutions of the appropriate second-stage problems. The second-stage solution method will require objective function information in the form of approximate cuts of increasing accuracy. Etc.

Due to risk constraints, however, the present subproblems are constrained convex programming problems.

We do not describe the whole framework in detail, but discuss only the second-stage step: Assume we can construct approximate support functions to the third-stage functions $\mathcal{F}^{(ij)}$ at any point with any prescribed accuracy. We show that this will enable construction of ϵ-support functions to the second-stage functions $\mathcal{F}^{(i)}$.

Problem formulation. Let us consider the second-stage problem (25) with fixed i, and fixed parameter values $(\hat{\omega}, \hat{\zeta}, \hat{\nu}, \hat{\zeta}', \hat{\nu}')$. The problem can be written in the form

$$\begin{array}{c} \min \; \varphi(\boldsymbol{x}) \\ \text{subject to} \\ \boldsymbol{x} \in \mathcal{X}, \\ \psi(\boldsymbol{x}) \leq 0. \end{array} \tag{27}$$

The feasible polyhedron in the above constrained convex problem is

$$\mathcal{X} := \left\{ \, (\boldsymbol{x}_2, \boldsymbol{v}_2, y') \;\middle|\; \boldsymbol{x}_2 \in X, \; \mathbf{1} \cdot \boldsymbol{x}_2 = \hat{\omega}, \;\; \boldsymbol{v}_2 \geq \mathbf{0}, \; \hat{\boldsymbol{p}} \cdot \boldsymbol{v}_2 = \hat{\nu}, \;\; y' \geq 0 \, \right\},$$

where the vector $\boldsymbol{v}_2$ consists of the components $v_2^{(j)}$ $(j = 1, \ldots, M^{(i)})$, and the vector $\hat{\boldsymbol{p}}$ consists of the components $p_2^{(ij)}$ $(j = 1, \ldots, M^{(i)})$. An upper bound $\overline{y}'$ of y' can be constructed because

$$\max \; \sum_{j \in \mathcal{J}} p_2^{(ij)} \left(w^{\mathcal{B}} - \boldsymbol{r}_2^{(ij)} \cdot \boldsymbol{x}_2 - \hat{\zeta}' \right) \quad \text{such that} \quad \mathcal{J} \subset \{1, \ldots, M^{(i)}\}, \quad \boldsymbol{x}_2 \in X, \; \mathbf{1} \cdot \boldsymbol{x}_2 = \hat{\omega}$$

is finite due to the compactness of the feasible domain. Imposing the bound $\overline{y}' \geq y'$ makes $\mathcal{X}$ compact.

The objective and constraint functions are

$$\varphi(\boldsymbol{x}) := \; -\sum_{j=1}^{M^{(i)}} p_2^{(ij)} \, \mathcal{F}^{(ij)} \left(\boldsymbol{r}_2^{(ij)} \cdot \boldsymbol{x}_2, \;\; \hat{\zeta}, \; v_2^{(j)} \right) \; + \overline{\lambda}' \, \tfrac{1}{1-\alpha} y',$$

$$\psi(\boldsymbol{x}) := \; \max_{\mathcal{J} \subset \{1, \ldots, M^{(i)}\}} \; \sum_{j \in \mathcal{J}} p_2^{(ij)} \left(w^{\mathcal{B}} - \boldsymbol{r}_2^{(ij)} \cdot \boldsymbol{x}_2 - \hat{\zeta}' \right) \; - y' - \hat{\nu}'.$$

If we select a bound $\overline{y}'$ large enough, then the Slater condition

$$\min_{\boldsymbol{x}\in\mathcal{X}} \psi(\boldsymbol{x}) < 0 \tag{28}$$

will obviously hold. We assume this henceforth.

Oracle providing function information for a constrained convex programming algorithm. Given $\delta > 0$, assume that we can construct δ-support functions to the third-stage functions $\mathcal{F}^{(ij)}$ $(j = 1, \ldots, M^{(i)})$. Assume moreover that there exists a single upper bound on the slopes of such δ-support functions.

Such δ-support functions enable construction of bounded-slope δ-support functions to φ.

Remark 11 *Construction of δ-support functions to the third-stage functions $\mathcal{F}^{(ij)}$ requires a solution algorithm for the third-stage problems, analogous to the one being presented.*

A single upper bound on the slopes of the δ-support functions can be constructed in the manner described in Remark 3.

On the other hand, we can obviously construct exact support functions to ψ, and $\max_{1\le j\le N^{(i)}} \|\boldsymbol{r}_2^{(ij)}\|$ is an upper bound on the slopes of such support functions.

An algorithm for constrained convex programming. In case $\psi(\boldsymbol{x}) \le 0$ holds for each $\boldsymbol{x} \in \mathcal{X}$, then problem (27) reduces to unconstrained minimization, and can be solved by the Inexact Level Method. An ϵ-support function can be constructed to $\mathcal{F}^{(i)}$ in the manner described in Section 2.2.

In what follows we assume that (27) is really a constrained problem. For the solution we propose a modified version of the Constrained Level Method of Lemaréchal, Nemirovskii, and Nesterov (1995).

Let $\Phi^\star$ denote the optimal objective value of the constrained convex problem (27). Given $\epsilon > 0$, the original Constrained Level Method finds an ϵ-optimal solution $\boldsymbol{x}^\star$ in the sense

$$\boldsymbol{x}^\star \in \mathcal{X}, \quad \psi(\boldsymbol{x}^\star) \le \epsilon, \quad \varphi(\boldsymbol{x}^\star) \le \Phi^\star + \epsilon.$$

Due to this feature, the original Constrained Level Method requires no Slater condition, but $\varphi(\boldsymbol{x}^\star) \ll \Phi^\star$ may occur. Near-optimal solutions in the above sense can not be used to construct support functions to $\mathcal{F}^{(i)}$.

In contrast, we have provided for the Slater condition (28), and will construct an ϵ-optimal solution $\boldsymbol{x}^\star$ in the sense

$$\boldsymbol{x}^\star \in \mathcal{X}, \quad \psi(\boldsymbol{x}^\star) \le 0, \quad \varphi(\boldsymbol{x}^\star) \le \Phi^\star + \epsilon. \tag{29}$$

The modified method is described in Appendix B. We call it asymmetric Level Method, because it works with approximate objective cuts but observes the constraint without tolerance (and requires exact cuts for the constraint function).

The method builds cutting-plane models φ_κ and ψ_κ of φ and ψ, respectively. (The model functions are polyhedral convex lower approximations of the objective and constraint function, respectively.) Based on these objects, a cutting-plane model problem (38) is constructed. The optimum of this model problem is a lower approximation of the optimum of the original convex problem (27). (Indeed, lower-valued function is minimized over a broader set.)

Suppose the method had terminated with an ϵ-optimal solution in the sense (29).

Parametric model problem. Let us consider the terminal model problem (38) as a parametric problem: The parameters ω, ν appear in the right-hand side of the inequality system determining $\mathcal{X}$. In the cutting-plane model functions φ_κ and ψ_κ, the parameters ζ and ζ', ν' appear in the right-hand sides of the cuts. Let $\underline{\mathcal{F}}^{(i)}(\omega, \zeta, \nu, \zeta', \nu')$ denote the optimum of the model problem as a function of the parameters. This is a polyhedral convex lower approximation of $\mathcal{F}^{(i)}(\omega, \zeta, \nu, \zeta', \nu')$. Using the optimal values of the dual variables of the cutting-plane model problem (38), a supporting linear function $L^{(i)}$ can be constructed to $\underline{\mathcal{F}}^{(i)}$ at $(\hat{\omega}, \hat{\zeta}, \hat{\nu}, \hat{\zeta}', \hat{\nu}')$.

The termination criterion of the asymmetric Level Method implies $\mathcal{F}^{(i)}(\hat{\omega}, \hat{\zeta}, \hat{\nu}, \hat{\zeta}', \hat{\nu}') - \underline{\mathcal{F}}^{(i)}(\hat{\omega}, \hat{\zeta}, \hat{\nu}, \hat{\zeta}', \hat{\nu}') \leq \epsilon$. This in turn means that $L^{(i)}$ is an ϵ-support function to $\mathcal{F}^{(i)}$ at $(\hat{\omega}, \hat{\zeta}, \hat{\nu}, \hat{\zeta}', \hat{\nu}')$.

Remark 12 *Such ϵ-support functions enable construction of ϵ-support functions to the objective function of the first-stage problem (24).*

This problem can then be solved by an unconstrained inexact Level Method. Convergence proof of this method requires a single upper bound on the slopes of the ϵ-support functions. This can be constructed in the manner described in Remarks 3 and 4.

5 Conclusions and prospect

We proposed a decomposition scheme and solution methods for multistage portfolio management problems with CVaR constraints. In 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 *complete recourse* problems by penalizing violations of CVaR constraints.

In a former project, Fábián and Szőke (2006) obtained good solution results for general two-stage problems by first transforming the problems into complete recourse forms. The complete recourse problems were then solved as constrained convex problems. (In such a constrained problem, the constraint function value is the expectation of an infeasibility measure. The constraint function can be evaluated in the same manner as the expected recourse function.) In that 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 new approach.) The new approach is practicable only if the transformation to complete recourse can be done in a computationally efficient way, as in the case of the present portfolio models.

Fábián and Szőke solved general two-stage stochastic programming problems with 20,000 scenarios. (Even larger problems were solved by integrating a distribution-approximation procedure into the decomposition scheme.) Effectiveness of the Künzi-Bay – Mayer approach to the present special subproblems implies that the present multistage decomposition scheme has practical relevance.

5.1 Decomposition of further risk constraints

With minor modifications, the proposed scheme can be used to decompose further risk constraints in portfolio management problems:

Integrated Chance Constraints were introduced by Klein Haneveld (1986) for static models. Drijver et al. (2002) and Klein Haneveld et al. (2005) use single-stage ICCs in dynamic models. Fábián and Szőke (2006) proposed two-stage ICCs to formulate non-traditional stochastic constraints.

Klein Haneveld and van der Vlerk (2006) proposed a polyhedral representation of generic ICCs. As we mentioned in Section 1.2, the Künzi-Bay – Mayer polyhedral representation is the CVaR analogue of this. Based on the polyhedral representation, Klein Haneveld and van der Vlerk developed a cutting-plane method for simple recourse problems. They implemented the method and for simple recourse problems, their special solver proved orders of magnitude faster then existing complete recourse solvers.

A generic ICC has the form $\mathrm{E}([w^{\mathcal{B}} - w]_+) \leq \gamma$. In portfolio management context, $w^{\mathcal{B}}$ is a user-defined parameter that represents the wealth we intend to accumulate by a certain point of time, w is a random variable that represents our yield at this point of time, and γ is a user-defined tolerance.

CVaR constraints in problems (16) and (17-19) can be transformed into ICCs by fixing the variable z to 0. (Let us moreover set the parameter $\alpha = 0$ since ICCs do not involve quantiles.) The Künzi-Bay – Mayer polyhedral cuts transform into Klein Haneveld – van der Vlerk cuts. This gives a decomposition scheme for generic ICCs.

Second-order Stochastic Dominance establishes a partial ordering between random variables. Let w and $\widehat{w}$ denote the respective random yields of two different portfolios. If $\mathrm{E}(U(w)) \geq \mathrm{E}(U(\widehat{w}))$ holds for any monotonic and concave (integrable) utility function U, then w is said to dominate $\widehat{w}$. This is equivalent to $\mathrm{E}([w^{\mathcal{B}} - w]_+) \leq \mathrm{E}([w^{\mathcal{B}} - \widehat{w}]_+)$ holding for any $w^{\mathcal{B}} \in \mathbb{R}$. (Consistency between SSD and CVaR is established in Ogryczak and Ruszczyński (2002).)

Dentcheva and Ruszczyński (2006) propose a single-stage portfolio optimization model involving an SSD constraint. The motivation is the following: assume we want to emulate a certain stock index $\widehat{w}$. This is achieved by constructing a portfolio whose yield w dominates $\widehat{w}$.

In general, an SSD constraint can be expressed by a continuum of ICCs. If, however, $\widehat{w}$ has a finite discrete distribution with realizations $\widehat{w}^{(s)}$ $(s = 1, \ldots, S)$, then the SSD constraint is equivalent to the finite set of ICCs

$$\mathrm{E}([\widehat{w}^{(s)} - w]_+) \;\leq\; \mathrm{E}([\widehat{w}^{(s)} - \widehat{w}]_+) \qquad (s = 1, \ldots, S).$$

Assuming discrete finite distributions, Dentcheva and Ruszczyński formulate the SSD-constrained problem in linear programming form. The LP problems have a specific structure. For such specific problems, they develop a duality theory in which the dual objects are utility functions. Based on this duality theory, they construct a dual problem that consists of the minimization of a weighted sum of polyhedral convex functions. Domains, function values, subgradients are easily computable. The authors adapted the Regularized Decomposition method of Ruszczyński (1986) to these special dual problems.

As an alternative, we propose a direct approach. Since the SSD constraint is equivalent to a finite set of ICCs, the above mentioned cutting-plane scheme of Klein Haneveld and van der Vlerk (2006) can be easily adapted to the present problem.

Moreover, we propose a multistage generalization of the SSD-constrained model. To extend the notation introduced in Section 2, let $\widehat{w}_1, \widehat{w}_2, \widehat{w}_3$ denote the stock index at the respective ends of the first, second, and third periods. We want our first-period yield w_1 to dominate $\widehat{w}_1$, our second-period yield w_2 to dominate $\widehat{w}_2$, and our third-period yield w_3 to dominate $\widehat{w}_3$. The resulting problem can be solved by the proposed decomposition scheme.

A value-of-information risk measure for multiperiod income processes was introduced by Pflug (2006). If the random process is represented by a scenario tree, then the value-of-information risk measure is computed as a weighted sum of *conditional-value-at-risk deviations*. The conditional-value-at-risk deviation of a random income Y is defined as $\mathrm{CVaRD}(Y) := \mathrm{E}(Y) + \mathrm{CVaR}(Y)$. At each non-terminal node of the tree, the single-stage CVaRD of next period's income is computed relative to the knowledge available at that node. (Confidence level is determined by deterministic parameters.)

Pflug defines the random income process in case of a special multiperiod portfolio optimization problem. He proposes minimizing the value-of-information risk measure under a constraint on expected end-of-horizon wealth. He constructs a linear programming problem. The number of the decision variables and constraints in this problem is a multiple of the number of nodes in the scenario tree. Hence in a realistic application this LP problem will be very large.

We propose relaxing the constraint on expected end-of-horizon wealth and including it in the objective function. The decomposition scheme and solution method described in Section 2 can be easily adapted to the resulting problem.

Alternatively, we propose maximizing expected end-of-horizon wealth under a constraint on the value-of-information risk measure. The decomposition scheme and solution method described in Section 3 can be easily adapted to the resulting problem.

5.2 Further fields of potential application

Option pricing. Ryabchenko et al. (2004) propose an ingenious grid-based replication model for the pricing of a European option in an incomplete market. Replication models represent option price as the value of a portfolio consisting of the underlying stock and a risk-free asset. The replicating portfolio is dynamically rebalanced as the value of the underlying stock varies. (Background information and references can be found in the cited paper.)

Ryabchenko et al. approximate optimal rebalancing strategy through a grid in the (time $\times$ underlying stock price) plane. They use a set of sample paths to model underlying stock behavior. Variables in this model belong to grid points, hence the number of the variables is independent of the number of the sample paths. Sample paths only affect the objective function and a single constraint, and these contain only expectations. The objective is the minimization of the average of squared approximation error. The mentioned constraint prescribes that the approximation error should average out to 0. Further constraints express distribution-independent monotonicity and convexity characteristics of option prices. (A special convexity result allows formulating the problem in continuous convex quadratic form.) The paper reports remarkably good test results obtained with only 100 - 200 sample paths.

The present decomposition scheme enables a direct approach. Underlying stock behavior is modeled by a scenario tree. We construct a self-financing portfolio, i.e., no re-financing is allowed at decision points. We assume an incomplete market, hence the final net result of the option seller is random. We impose a CVaR constraint on final net result (or multiple CVaR constraints with different confidence levels). Moreover we prescribe that expected final net result should cover administration costs of the seller. (Transaction costs can also be taken into consideration.) We minimize the required initial investment under these conditions.

The models presented in sections 3 and 4 need but slight modifications: Option price will be represented by initial investment w_0, hence it will be a decision variable. Option value at maturity will be represented by benchmark wealth $w^{\mathcal{B}}$, hence it will be a random parameter. There will be two assets ($n = 2$), only one of them having random returns. Instead of maximizing end-of-horizon portfolio yield $\mathrm{E}(w)$, we will impose a constraint on expected final net result $\mathrm{E}(w - w^{\mathcal{B}})$, and minimize w_0.

Asset-Liability Management. Drijver et al. (2002) and Klein Haneveld et al. (2005) model the operation of a company-owned pension fund. In this case, liabilities consist of future pension payments. Funding comes from revenues of investments and regular contributions by active members. The funding ratio is A/L, where A is the value of the assets, and L is the value of liabilities. Seen over a number of years, the funding ratio may fall short occasionally, but if this happens too often or if the shortage is too large, then the owner company is required to make a remedial contribution to the fund.

The cited authors represent random parameters by a scenario tree and build a multistage stochastic programming model. Time periods are years, and decisions are made at the points of time $t = 0, 1, \ldots, T$. At each non-terminal decision node, a short-term risk constraint is imposed. It is an integrated chance constraint on subsequent year's funding ratio in the form of $\mathrm{E}([\alpha L_{t+1} - A_{t+1}]_+) \leq \beta$, where α, β are user-defined parameters, and the expectation is composed relative to the knowledge available at the relevant node. Moreover, in the regulation prescribing that 'funding ratio should not fall short too often', the term 'too often' is interpreted as 'in two consecutive years'. Hence at each (non-root) node an additional, mid-term constraint is imposed: if previous years's funding ratio fall short, then any shortage in current year's funding ratio has to be corrected by a remedial contribution at the present node. These mid-term constraints are expressed by binary variables.

As an alternative to these mid-term constraints, we propose a direct generalization of the above described short-term constraints: Given a decision node that belongs to the point of time $t < T - 1$, we impose a risk constraint on the funding ratio of year $t + 2$. If we use integrated chance constraints, then this alternative mid-term constraint takes the form $\mathrm{E}([\alpha L_{t+2} - A_{t+2}]_+) \leq \beta$, where the expectation is composed relative to the knowledge available at the relevant node.

These alternative mid-term constraints are clearly less direct than thsoe proposed in the cited papers. But the heuristic meaning is similar: at each decision point, special care must be taken of short- and mid-term consequences. From computational point of view, the alternative formulation requires no binary variables, and the alternative mid-term constraints fit into the proposed decomposition scheme.

References

[1] AHMED, S. (2006). Convexity and decomposition of mean-risk stochastic programs. *Mathematical Programming,* to appear, DOI: 10.1007/s10107-005-0638-8.

[2] ANDERSSON, F., H. MAUSSER, D.ROSEN, and S. URYASEV (2001). Credit risk optimization with Conditional Value-at-Risk criterion. *Mathematical Programming,* Series B **89**, 273-291.

[3] CLAESSENS, S. and J. KREUSER (2004). A framework for strategic foreign reserves risk management. In: *Risk management for Central Bank foreign reserves* (editors: C. Bernadell, P. Cardon, J. Coche, F.X. Diebold, and S. Manganelli), 47-74. European Central Bank, Frankfurt, Germany.

[4] DENTCHEVA, D. and A. RUSZCZYŃSKI (2006). Portfolio optimization with stochastic dominance constraints. *Journal of Banking & Finance* **30**, 433-451.

[5] DRIJVER, S.J., W.K. KLEIN HANEVELD, and M.H. VAN DER VLERK (2002) ALM model for pension funds: numerical results for a prototype model. *Research Report* 02A44, Research School SOM, University of Groningen, the Netherlands.

[6] FÁBIÁN, C.I. (2000). Bundle-type methods for inexact data. *Central European Journal of Operations Research* **8** (special issue, editors: T. Csendes and T. Rapcsák), 35-55.

[7] FÁBIÁN, C.I. and Z. SZŐKE (2006). Solving two-stage stochastic programming problems with level decomposition, *Computational Management Science* to appear, DOI: 10.1007/s10287-006-0026-8.

[8] JOBST, N.J. and S. ZENIOS (2001). The tail that wags the dog: Integrating credit risk in asset portfolios. *Working Paper* #01-05. HERMES, European Center of Excellence on Computational Finance and Economics, University of Cyprus, Nicosia.

[9] KALL, P. and J. MAYER. (2005). *Stochastic Linear Programming: Models, Theory, and Computation.* Springer, International Series in Operations Research and Management Science.

[10] KLEIN HANEVELD, W.K. (1986). Duality in Stochastic Linear and Dynamic Programming. *Lecture Notes in Economics and Math. Systems* **274** Springer-Verlag, New York.

[11] KLEIN HANEVELD, W.K., M.H. STREUTKER, and M.H. VAN DER VLERK (2005). An ALM model for pension funds using integrated chance constraints. *Research Report* 05A03, Research School SOM, University of Groningen, the Netherlands.

[12] KLEIN HANEVELD, W.K. and M.H. VAN DER VLERK (2006). Integrated chance constraints: reduced forms and an algorithm. *Computational Management Science* **3**, 245-269.

[13] KROKHMAL, P., J. PALMQUIST, and S. URYASEV (2002). Portfolio optimization with Conditional Value-at-Risk objective and constraints. *Journal of Risk* **4**, 11-27.

[14] KÜNZI-BAY, A. and J. MAYER (2006). Computational aspects of minimizing conditional value-at-risk. *Computational Management Science* **3**, 3-27.

[15] LEMARÉCHAL, C., A. NEMIROVSKII, and YU. NESTEROV. (1995). New variants of bundle methods. *Mathematical Programming* **69**, 111-147.

[16] OGRYCZAK, W. and A. RUSZCZYŃSKI (2002). Dual stochastic dominance and related mean-risk models. *SIAM Journal on Optimization* **13**, 60-78.

[17] PFLUG, G.CH. (2006). A value-of-information approach to measuring risk in multiperiod economic activity. *Journal of Banking & Finance* **30**, 695-715.

[18] PRÉKOPA, A. (2003). Probabilistic programming. In: *Stochastic Programming, Handbooks in Operations Research and Management Science* **10** (A. Ruszczyński and A. Shapiro, eds.), Elsevier, Amsterdam, 267-351.

[19] ROCKAFELLAR, R.T. and S. URYASEV (2000). Optimization of conditional value-at-risk. *Journal of Risk* **2**, 21-41.

[20] Rockafellar, R.T. and S. Uryasev (2002). Conditional value-at-risk for general loss distributions. *Journal of Banking & Finance* **26**, 1443-1471.

[21] Ruszczyński, A. (1986). A Regularized Decomposition Method for Minimizing the Sum of Polyhedral Functions. *Mathematical Programming* **35**, 309-333.

[22] Ryabchenko, V., S. Sarykalin, and S. Uryasev (2004). Pricing European options by numerical replication: quadratic programming with constraints. *Asia-Pacific Financial Markets* **11**, 301-333.

[23] Topaloglou, N., H. Vladimirou, and S. Zenios (2004). Dynamic stochastic programming models for international portfolio management. *Working Paper* #04-02. HERMES, European Center of Excellence on Computational Finance and Economics, University of Cyprus, Nicosia.

A Formal proofs

A.1 Proof of Proposition 2

Let us construct the equivalent linear programming form (ELPF) of the three-stage problem (10 - 12) in the usual manner. It means defining a separate set of second-stage variables for each $i = 1, \ldots, N$; and of third-stage variables for each $i = 1, \ldots, N,\ j = 1, \ldots, M^{(i)}$. Let $\boldsymbol{x}_2^{(i)}$ denote the second-stage variables corresponding to i ; and $\boldsymbol{x}_3^{(ij)},\ v^{(ij)}$, the third-stage variables corresponding to (i, j).

There is a straightforward matching between the variables of the ELPF problem and those of the polyhedral representation problem (9). There is but one non-trivial case: variable v of the polyhedral representation problem will correspond to the weighted sum $\sum_{i=1}^{N} \sum_{j=1}^{M^{(i)}} p_1^{(i)} p_2^{(ij)} v^{(ij)}$ of the ELPF-variables $v^{(ij)}$.

The objective function of the ELPF problem is

$$\sum_{i=1}^{N} p_1^{(i)} \left\{ \sum_{j=1}^{M^{(j)}} p_2^{(ij)} \left(\sum_{k=1}^{K^{(ij)}} p_3^{(ijk)} \boldsymbol{r}_3^{(jk)} \cdot \boldsymbol{x}_3^{(jk)} \ - \ \lambda \frac{1}{1-\alpha} v^{(ij)} \right) \right\} \quad - \lambda z,$$

and this is obviously equivalent to the objective function of the polyhedral representation problem (9).

Now we show that cuts in the polyhedral representation problem (9) are just aggregates of ELPF-cuts: Let us select a subset $\mathcal{K}^{(ij)} \subset \{1, \ldots, K^{(ij)}\}$ for each $i = 1, \ldots, N,\ j = 1, \ldots, M^{(i)}$. The corresponding ELPF-cuts are

$$\sum_{k \in \mathcal{K}^{(ij)}} p_3^{(ijk)} \left(w^{\mathcal{B}} - \boldsymbol{r}_3^{(ijk)} \cdot \boldsymbol{x}_3^{(ijk)} - z \right) \ \leq v^{(ij)}.$$

(In case of an empty set $\mathcal{K}^{(ij)} = \emptyset$, the corresponding cut is $0 \leq v^{(ij)}$.) Aggregating these ELPF-cuts with the weights $p_1^{(i)} p_2^{(ij)}$, we get

$$\sum_{i=1}^{N} \sum_{j=1}^{M^{(i)}} p_1^{(i)} p_2^{(ij)} \sum_{k \in \mathcal{K}^{(ij)}} p_3^{(ijk)} \left(w^{\mathcal{B}} - \boldsymbol{r}_3^{(ijk)} \cdot \boldsymbol{x}_3^{(ijk)} - z \right) \ \leq \ \sum_{i=1}^{N} \sum_{j=1}^{M^{(i)}} p_1^{(i)} p_2^{(ij)} v^{(ij)}. \tag{30}$$

Introducing

$$\mathcal{R} := \bigcup_{i=1}^{N} \bigcup_{j=1}^{M^{(j)}} \left\{ (i, j, k) \in \mathcal{S} \ \middle| \ k \in \mathcal{K}^{(ij)} \right\},$$

the aggregate ELPF-cut (30) takes the form

$$\sum_{(i,j,k) \in \mathcal{R}} p_{(ijk)} \left(w^{\mathcal{B}} - \boldsymbol{r}_3^{(ijk)} \cdot \boldsymbol{x}_3^{(ijk)} - z \right) \ \leq v,$$

that is a cut in the polyhedral representation problem (9).

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

Conversely, given a cut in the polyhedral representation problem (9), i.e., a subset $\mathcal{R} \subset \mathcal{S}$, we can construct the sets

$$\mathcal{K}^{(ij)} := \{ \; k \mid (i,j,k) \in \mathcal{R} \; \} \qquad (i = 1,\ldots,N, \; j = 1,\ldots,M^{(i)}).$$

In the manner of (30), let us aggregate the ELPF-cuts corresponding to the above sets. This results just the cut corresponding to $\mathcal{R}$ in the polyhedral representation problem (9).

A.2 Proof of Proposition 6

Let us construct the equivalent linear programming form (ELPF) of the three-stage problem (17 - 19) in the usual manner. It means defining a separate set of second-stage variables for each $i = 1,\ldots,N$; and of third-stage variables for each $i = 1,\ldots,N, \; j = 1,\ldots,M^{(i)}$. Let $\boldsymbol{x}_2^{(i)}, \;\; v_2^{(ij)}$ $(j = 1,\ldots,M^{(i)})$ denote the second-stage variables corresponding to i ; and $\boldsymbol{x}_3^{(ij)}$, $y^{(ij)}$, the third-stage variables corresponding to (i,j).

There is a straightforward matching between the variables of the ELPF problem and those of the polyhedral representation problem (16). There are but two non-trivial cases:

$$\tilde{v} \longleftrightarrow \sum_{i=1}^{N} p_1^{(i)} v_1^{(i)} \;=\; \sum_{i=1}^{N} p_1^{(i)} \left\{ \sum_{j=1}^{M^{(i)}} p_2^{(ij)} v_2^{(ij)} \right\} \qquad \text{and} \qquad \tilde{y} \longleftrightarrow \sum_{i=1}^{N} \sum_{j=1}^{M^{(i)}} p_1^{(i)} p_2^{(ij)} y^{(ij)}, \quad (31)$$

i.e., the variables $\tilde{v}$ and $\tilde{y}$ of the polyhedral representation problem will correspond to the weighted sums of the ELPF-variables $v_1^{(i)}, v_2^{(ij)}$, and $y^{(ij)}$, respectively.

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

Cuts in the ELPF problem have the form

$$\sum_{k \in \mathcal{K}^{(ij)}} p_3^{(ijk)} \left(w^{\mathcal{B}} - \boldsymbol{r}_3^{(ijk)} \cdot \boldsymbol{x}_3^{(ijk)} - z \right) \; - y^{(ij)} \; \leq v_2^{(ij)}, \qquad (32)$$

where $1 \leq i \leq N$, $1 \leq j \leq M^{(i)}$ and $\mathcal{K}^{(ij)} \subset \{1,\ldots,K^{(ij)}\}$. Equivalence between aggregate ELPF-cuts and the cuts of the polyhedral representation problem can be established like in A.1.

It remains to be shown only that any non-negative values of the variables $\tilde{v}$ and $\tilde{y}$ in the polyhedral problem (16) can be represented in the forms (31) with *non-negative* values of the ELPF-variables $v_1^{(i)}$, $v_2^{(ij)}$, and $y^{(ij)}$.

Suppose the variables in the polyhedral problem (16) had been fixed at feasible values. Thus the left-hand-side sums in the cuts (32) are determined. Let us select the deepest cut for each $1 \leq i \leq N$ and $1 \leq j \leq M^{(i)}$, i.e., let

$$\sigma^{(ij)} := \quad \max \sum_{k \in \mathcal{K}^{(ij)}} p_3^{(ijk)} \left(w^{\mathcal{B}} - \boldsymbol{r}_3^{(ijk)} \cdot \boldsymbol{x}_3^{(ijk)} - z \right) \quad \text{such that} \quad \mathcal{K}^{(ij)} \subset \{1,\ldots,K^{(ij)}\}.$$

We have $\sigma^{(ij)} \geq 0$ since $\mathcal{K}^{(ij)} = \emptyset$ can be selected.

The aggregate of these deepest cuts is valid in the polyhedral problem (16), hence

$$\sigma - \tilde{y} \leq \tilde{v} \qquad \text{holds with} \quad \sigma := \sum_{i=1}^{N} \sum_{j=1}^{M^{(i)}} p_1^{(i)} p_2^{(ij)} \, \sigma^{(ij)}. \tag{33}$$

Assumed $\sigma > 0$, let

$$y^{(ij)} := \tilde{y}\,\frac{\sigma^{(ij)}}{\sigma}, \quad v^{(ij)} := \tilde{v}\,\frac{\sigma^{(ij)}}{\sigma}, \qquad (1 \leq i \leq N,\ 1 \leq j \leq M^{(i)}).$$

These are obviously feasible values in the ELPF-problem. (The trivial case of $\sigma = 0$ can be easily dealt with.)

A.3 Proof of Proposition 9

Let $\mathcal{Y}$ denote the feasible domain of the unconstrained problem (21). Feasible solutions will be denoted by $\boldsymbol{\chi} \in \mathcal{Y}$. The end-of-horizon expected return $\eta := \mathrm{E}(w)$ is considered a component of $\boldsymbol{\chi}$. Similarly, the vector of risk measures $\boldsymbol{\varsigma} := (\mathrm{CVaR}(w), \mathrm{CVaR}(w'))$ is considered a sub-vector.

We assume that the constrained problem (20) is feasible, and the Slater condition holds. The Lagrangian function of this problem is:

$$\mathcal{V}(\boldsymbol{\chi}, \boldsymbol{\lambda}) := \eta - \boldsymbol{\lambda} \cdot (\boldsymbol{\varsigma} - \boldsymbol{\gamma}) \qquad (\boldsymbol{\chi} \in \mathcal{Y},\ \boldsymbol{\lambda} \geq \mathbf{0}),$$

where $\boldsymbol{\gamma} := (\gamma, \gamma')$ is the right-hand-side vector of the CVaR constraints.

Let $\boldsymbol{\chi}^\star \in \mathcal{Y}$ denote an optimal solution of the constrained problem. From the Karush-Kuhn-Tucker theorem, there exists $\boldsymbol{\lambda}^\star \geq \mathbf{0}$ such that

$$\mathcal{V}(\boldsymbol{\chi}^\star, \boldsymbol{\lambda}^\star) \geq \mathcal{V}(\boldsymbol{\chi}, \boldsymbol{\lambda}^\star) \quad (\boldsymbol{\chi} \in \mathcal{Y}), \qquad \text{and} \qquad \boldsymbol{\lambda}^\star \cdot (\boldsymbol{\varsigma}^\star - \boldsymbol{\gamma}) = 0, \tag{34}$$

where $\boldsymbol{\varsigma}^\star$ is the relevant sub-vector of $\boldsymbol{\chi}^\star$.

Let us define

$$\mathcal{W}(\boldsymbol{\chi}, \lambda) := \eta - \boldsymbol{\lambda} \cdot [\boldsymbol{\varsigma} - \boldsymbol{\gamma}]_+ \qquad (\boldsymbol{\chi} \in \mathcal{Y},\ \boldsymbol{\lambda} \geq \mathbf{0}),$$

where the positive part of the vector $\boldsymbol{\varsigma} - \boldsymbol{\gamma}$ is meant componentwise.

Let $\overline{\boldsymbol{\lambda}} > \boldsymbol{\lambda}^\star$, and consider the problem

$$\max\ \mathcal{W}(\boldsymbol{\chi}, \overline{\boldsymbol{\lambda}}) \qquad \text{such that} \quad \boldsymbol{\chi} \in \mathcal{Y}. \tag{35}$$

(The above problem is obviously equivalent to the penalized problem (22) where the components of $\overline{\boldsymbol{\lambda}}$ had been selected to be respective upper bounds of the components of $\boldsymbol{\lambda}$.)

We are going to show that $\boldsymbol{\chi}^\star$ is also an optimal solution of the problem (35): Indeed, for any $\boldsymbol{\chi} \in \mathcal{Y}$, we have

$$\begin{aligned}
\mathcal{W}(\boldsymbol{\chi}, \overline{\boldsymbol{\lambda}}) &\leq \mathcal{W}(\boldsymbol{\chi}, \boldsymbol{\lambda}^\star) && \text{because} \quad \overline{\boldsymbol{\lambda}} > \boldsymbol{\lambda}^\star, \\
&\leq \mathcal{V}(\boldsymbol{\chi}, \boldsymbol{\lambda}^\star) && \text{obviously}, \\
&\leq \mathcal{V}(\boldsymbol{\chi}^\star, \boldsymbol{\lambda}^\star) && \text{from the inequality in (34)}, \\
&= \mathcal{W}(\boldsymbol{\chi}^\star, \overline{\boldsymbol{\lambda}}) && \text{from the equality in (34),} \\
& && \text{and from } \boldsymbol{\chi}^\star \text{ satisfying the CVaR constraint.}
\end{aligned} \tag{36}$$

Conversely, let $\overline{\boldsymbol{\chi}} \in \mathcal{Y}$ denote an optimal solution of the problem (35). We are going to show that $\overline{\boldsymbol{\chi}}$ is also an optimal solution of the constrained problem (20). Indeed, substituting $\boldsymbol{\chi} = \overline{\boldsymbol{\chi}}$ in the chain (36), each link should hold with equality.

The first link is $\mathcal{W}(\overline{\boldsymbol{\chi}}, \overline{\boldsymbol{\lambda}}) = \mathcal{W}(\overline{\boldsymbol{\chi}}, \boldsymbol{\lambda}^\star)$. Using the definition of $\mathcal{W}$:

$$0 \;=\; \mathcal{W}(\overline{\boldsymbol{\chi}}, \overline{\boldsymbol{\lambda}}) - \mathcal{W}(\overline{\boldsymbol{\chi}}, \boldsymbol{\lambda}^\star) \;=\; (\boldsymbol{\lambda}^\star - \overline{\boldsymbol{\lambda}}) \cdot [\,\overline{\boldsymbol{\varsigma}} - \boldsymbol{\gamma}\,]_+,$$

where $\overline{\boldsymbol{\varsigma}}$ is the relevant sub-vector of $\overline{\boldsymbol{\chi}}$. Since $\overline{\boldsymbol{\lambda}} > \boldsymbol{\lambda}^\star$, the right-hand side is 0 only if $\overline{\boldsymbol{\varsigma}} - \boldsymbol{\gamma} \leq \mathbf{0}$ holds. It means that $\overline{\boldsymbol{\chi}}$ is a feasible solution of the constrained problem (20). Optimality of $\overline{\boldsymbol{\chi}}$ in (20) follows from the second and third links in the chain (36) of equalities:

From $\mathcal{W}(\overline{\boldsymbol{\chi}}, \boldsymbol{\lambda}^\star) = \mathcal{V}(\overline{\boldsymbol{\chi}}, \boldsymbol{\lambda}^\star)$ it follows that $\boldsymbol{\lambda}^\star \cdot [\,\overline{\boldsymbol{\varsigma}} - \boldsymbol{\gamma}\,]_- = 0$ holds. Since $\overline{\boldsymbol{\chi}}$ satisfies the CVaR constraint, it follows that

$$\boldsymbol{\lambda}^\star \cdot (\,\overline{\boldsymbol{\varsigma}} - \boldsymbol{\gamma}\,) \;= 0. \tag{37}$$

The third link is $\mathcal{V}(\overline{\boldsymbol{\chi}}, \boldsymbol{\lambda}^\star) = \mathcal{V}(\boldsymbol{\chi}^\star, \boldsymbol{\lambda}^\star)$. Taking (37) and the equality in (34) into account, it follows that $\overline{\boldsymbol{\chi}}$ and $\boldsymbol{\chi}^\star$ have the same objective values in the constrained problem (20).

B An asymmetric Level Method

In this section we introduce an asymmetric version of the Constrained Level Method of Lemaréchal, Nemirovskii, and Nesterov (1995). This version works with approximate objective cuts but observes the constraint without tolerance (and requires exact cuts for the constraint function). Minor modifications are needed in the constructions and proofs of Lemaréchal, Nemirovskii, and Nesterov.

A sequence of feasible points is generated. At each point, we have a support function to ψ, and an approximate support function to φ. These linear functions are used to build cutting-plane models of ψ and φ.

Suppose that we have already constructed the points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_\kappa \in \mathcal{X}$. At the point $\boldsymbol{x}_\iota$ $(\iota = 1, \ldots, \kappa)$, let $l_\iota^\sharp$ denote the support function constructed to ψ ; and let l_ι denote the δ_ι-support function constructed to φ. (Here $\delta_\iota > 0$ is a tolerance, and a δ_ι-support function means a linear function satisfying $l_\iota \leq \varphi$ and $l_\iota(\boldsymbol{x}_\iota) + \delta_\iota \geq \varphi(\boldsymbol{x}_\iota)$.)

The cutting-plane models of φ and ψ will be

$$\varphi_\kappa(\boldsymbol{x}) := \max_{1 \leq \iota \leq \kappa} l_\iota(\boldsymbol{x}) \quad \text{and} \quad \psi_\kappa(\boldsymbol{x}) := \max_{1 \leq \iota \leq \kappa} l_\iota^\sharp(\boldsymbol{x}) \qquad (\boldsymbol{x} \in \mathbb{R}^n).$$

Obviously we have $\varphi_\kappa \leq \varphi$ and $\psi_\kappa \leq \psi$. A linear programming model of the constrained convex problem (27) will be constructed as

$$\begin{array}{c} \min\; \varphi_\kappa(\boldsymbol{x}) \\ \text{subject to} \\ \boldsymbol{x} \in \mathcal{X}, \\ \psi_\kappa(\boldsymbol{x}) \leq 0. \end{array} \tag{38}$$

Let $\underline{\Phi}_\kappa^\star$ denote the optimal objective value of the model problem (38). Obviously we have $\underline{\Phi}_\kappa^\star \leq \Phi^\star$. (Lower-valued function is minimized over a broader set.)

The best point associated with $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_\kappa$ will be constructed in the form of a convex combination of the former iterates:

$$\boldsymbol{x}_\kappa^\star := \sum_{\iota=1}^{\kappa} \varrho_\iota \boldsymbol{x}_\iota. \tag{39}$$

The weights $\varrho_1, \ldots, \varrho_\kappa$ will be determined through the solution of the following linear programming problem:

$$\begin{array}{c} \min \sum_{\iota=1}^{\kappa} \varrho_\iota \left(\varphi_\kappa(\boldsymbol{x}_\iota) + \delta_\iota \right) \\ \text{subject to} \\ \varrho_\iota \geq 0 \quad (\iota = 1, \ldots, \kappa), \qquad \sum_{\iota=1}^{\kappa} \varrho_\iota = 1, \\ \sum_{\iota=1}^{k} \varrho_\iota \psi(\boldsymbol{x}_\iota) \leq 0. \end{array} \tag{40}$$

(The values $\psi(\boldsymbol{x}_\iota)$ are known since they had been computed for the construction of the linear support functions $l_\iota^\sharp$. On the other hand, $\varphi_\kappa(\boldsymbol{x}_\iota) + \delta_\iota$ is an upper approximation for $\varphi(\boldsymbol{x}_\iota)$ for each $\iota = 1, \ldots, \kappa$.)

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

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

Assume that we have selected $\boldsymbol{x}^\psi$ as starting point, i.e., we have set $\boldsymbol{x}_\iota := \boldsymbol{x}^\psi$ for $\iota = 1$. Then $\varrho_1 = 1,\ \varrho_\iota = 0\ (\iota = 2, \ldots, \kappa)$ is a feasible solution, due to the Slater condition (28). Hence an optimal solution also exists (because the feasible domain is compact). Let $\overline{\Phi}_\kappa^\star$ denote the optimal objective value of problem (40). From the convexity of the set $\mathcal{X}$, and of the functions ψ and φ, it follows that

$$\boldsymbol{x}_\kappa^\star \in \mathcal{X}, \quad \psi(\boldsymbol{x}_\kappa^\star) \leq 0, \quad \varphi(\boldsymbol{x}_\kappa^\star) \leq \overline{\Phi}_\kappa^\star. \tag{41}$$

Consequently $\Phi^\star \leq \overline{\Phi}_\kappa^\star$ holds. (Indeed, $\boldsymbol{x}_\kappa^\star$ is a feasible solution of the convex programming problem (27), and $\Phi^\star$ denotes the optimum of (27).)

Remark 13 *The inequality $\Phi^\star \leq \overline{\Phi}_\kappa^\star$ 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 $\boldsymbol{x}_\kappa^\star$ may only satisfy the non-positivity constraint with a tolerance.*

$\Phi^\star$ is unknown but the optimum $\underline{\Phi}_\kappa^\star$ of the model problem provides a lower bound. If we can achieve $\overline{\Phi}_\kappa^\star - \underline{\Phi}_\kappa^\star \leq \epsilon$ then $\boldsymbol{x}_\kappa^\star$ will be an ϵ-optimal solution of the convex programming problem (27) in the sense (29).

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

$$\max_{\beta \geq 0} h_\kappa(\beta), \tag{42}$$

where

$$h_\kappa(\beta) := \min_{1 \le \iota \le \kappa} \{\ \varphi_\kappa(\boldsymbol{x}_\iota) + \beta\psi(\boldsymbol{x}_\iota) + \delta_\iota\ \}.$$

Hence we have

$$\max_{\beta \ge 0}\ h_\kappa(\beta) = \overline{\Phi}^\star_\kappa \tag{43}$$

from the duality theorem. Our aim is to direct the search for new iterates $\boldsymbol{x}_\iota$ in such a way that $\max_{\beta \ge 0}\ h_\kappa(\beta) \le \underline{\Phi}^\star_\kappa + \epsilon$ will hold with κ 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

$$\underline{\Phi} := \min_{\boldsymbol{x} \in \mathcal{X}} \varphi(\boldsymbol{x}) \quad \text{and} \quad \overline{\beta} := \frac{\varphi(\boldsymbol{x}^\psi) - \underline{\Phi}}{-\psi(\boldsymbol{x}^\psi)}.$$

The denominator in the above fraction is positive due to the Slater condition (28). The numerator is non-negative due to the definition of $\underline{\Phi}$. It follows that $\overline{\beta} \ge 0$.

Proposition 14 *Assume that we have selected $\boldsymbol{x}^\psi$ as starting point, i.e., we have set $\boldsymbol{x}_\iota := \boldsymbol{x}^\psi$ for $\iota = 1$. Then $h_\kappa(\beta) < \underline{\Phi}$ holds for any $\beta > \overline{\beta}$.*

Proof. From the selection of the starting point, we have

$$h_1(\overline{\beta}) \ = \ \varphi(\boldsymbol{x}^\psi) + \overline{\beta}\psi(\boldsymbol{x}^\psi) = \ \varphi(\boldsymbol{x}^\psi) + \frac{\varphi(\boldsymbol{x}^\psi) - \underline{\Phi}}{-\psi(\boldsymbol{x}^\psi)}\,\psi(\boldsymbol{x}^\psi) \ = \ \underline{\Phi}.$$

Due to $\psi(\boldsymbol{x}^\psi) < 0$ we have $h_1(\beta) < h_1(\overline{\beta})$ for $\beta > \overline{\beta}$. Moreover we have $h_k \le h_1$ by definition. Summing these up, we get that

$$h_\kappa(\beta) \le h_1(\beta) < h_1(\overline{\beta}) = \underline{\Phi} \quad \text{holds for } \beta > \overline{\beta}.$$

□

A direct consequence of Proposition 14 is that optimal solutions of problem (42) fall into the interval $\left[0, \overline{\beta}\right]$. Indeed, we have $\underline{\Phi} \le \varphi(\boldsymbol{x}^\star_\kappa) \le \overline{\Phi}^\star_\kappa = \max_{\beta \ge 0}\ h_\kappa(\beta)$ by (41) and (43). It follows that the dual problem (42) can be written in the form

$$\max_{\overline{\beta} \ge \beta \ge 0}\ h_\kappa(\beta). \tag{44}$$

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

Initialize.

Set the stopping tolerance $\epsilon > 0$.

Set the parameters θ, μ, and $\tau \quad (0 < \theta, \mu < 1;\ 0 < \tau < (1-\theta)^2)$.

Let $\kappa := 1$ (iteration counter).

Set the starting point $\boldsymbol{x}_1 \in \arg\min_{\boldsymbol{x}\in\mathcal{X}} \psi(\boldsymbol{x})$.

Set starting tolerance δ_1.

Set upper limit for the feasible interval of the dual problem as $\overline{\beta} := \frac{\varphi(\boldsymbol{x}_1) - \min_{\boldsymbol{x}\in\mathcal{X}} \phi(\boldsymbol{x})}{-\psi(\boldsymbol{x}_1)}$.

Update bundle.

Given the point $\boldsymbol{x}_\kappa$, construct the support function $l^\sharp_\kappa$ to ψ, and the δ_κ-support function l_κ to φ.

Define the model functions $\varphi_\kappa(\boldsymbol{x}) := \max_{1\le\iota\le k} l_\iota(\boldsymbol{x}), \quad \psi_\kappa(\boldsymbol{x}) := \max_{1\le\iota\le k} l^\sharp_\iota(\boldsymbol{x})$.

Compute the optimum of the model problem $\underline{\Phi}^\star_\kappa := \min \{\varphi_\kappa(\boldsymbol{x}) \,|\, \boldsymbol{x} \in \mathcal{X},\ \psi_\kappa(\boldsymbol{x}) \le 0\}$.

Define the function $h_\kappa(\beta) := \min_{1\le\iota\le k} \{ \varphi_\kappa(\boldsymbol{x}_\iota) + \beta\psi(\boldsymbol{x}_\iota) + \delta_\iota\}$,

and compute its maximum $\overline{\Phi}^\star_\kappa := \max_{\overline{\beta}\ge\beta\ge 0} h_\kappa(\beta)$.

Check for optimality.

If $\overline{\Phi}^\star_\kappa - \underline{\Phi}^\star_\kappa < \epsilon$, then near-optimal solution found ; construct best point according to (39) and stop.

Find dual iterate.

Determine the interval $[\underline{\beta}_\kappa, \overline{\beta}_\kappa] := \left\{ 0 \le \beta \le \overline{\beta} \ \middle|\ h_\kappa(\beta) \ge \underline{\Phi}^\star_\kappa \right\}$.
Compute β_κ:

– for $\iota = 1$, let $\beta_1 := \frac{1}{2}(\overline{\beta}_1 + \underline{\beta}_1)$,

– for $\iota > 1$, let $\beta_\kappa := \begin{cases} \beta_{\kappa-1}, & \text{if } \underline{\beta}_\kappa + \frac{\mu}{2}(\overline{\beta}_\kappa - \underline{\beta}_\kappa) \le \beta_{k-1} \le \overline{\beta}_\kappa - \frac{\mu}{2}(\overline{\beta}_\kappa - \underline{\beta}_\kappa), \\ \frac{1}{2}(\overline{\beta}_\kappa + \underline{\beta}_\kappa), & \text{otherwise.} \end{cases}$

Find primal iterate.

Let $\boldsymbol{x}_{\kappa+1}$ be the optimal solution of the convex quadratic programming problem

$$\begin{array}{c} \min \ \|\boldsymbol{x} - \boldsymbol{x}_\kappa\|^2 \\ \text{subject to} \\ \boldsymbol{x} \in \mathcal{X}, \\ \varphi_\kappa(\boldsymbol{x}) + \beta_\kappa \psi_\kappa(\boldsymbol{x}) \ \leq \ (1-\theta)\underline{\Phi}_\kappa^\star + \theta h_\kappa(\beta_\kappa) \end{array}$$

Loop.

Set $\delta_{\kappa+1} := \tau h_\kappa(\beta_\kappa)$.

Increment κ.

$\rightarrow$ *Update bundle.*

We give explanations and sketch proofs on the basis of Lemaréchal, Nemirovskii, and Nesterov (1995) and Fábián (2000).

Find dual iterate. A dual iterate β_κ is selected such that $h_\kappa(\beta_\kappa)$ is 'sufficiently close' to $\max\limits_{\overline{\beta} \geq \beta \geq 0} h_\kappa(\beta)$. (The term 'sufficiently close' will be specified in (46).) First consider the set

$$I_\kappa := \left\{ \, 0 \leq \beta \leq \overline{\beta} \ \middle| \ h_\kappa(\beta) \geq \underline{\Phi}_\kappa^\star \, \right\}. \tag{45}$$

I_κ is an interval since h_κ is a concave function. In the algorithmic description the notation $I_\kappa = [\underline{\beta}_\kappa, \overline{\beta}_\kappa]$ is used. This interval is not empty since we have $\max\limits_{\overline{\beta} \geq \beta \geq 0} h_\kappa(\beta) = \overline{\Phi}_\kappa^\star \geq \underline{\Phi}_\kappa^\star$. Let the subinterval $\hat{I}_\kappa \subset I_\kappa$ be obtained by shrinking I_κ: the center of $\hat{I}_\kappa$ will be the same as the center of I_κ, 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_κ, it follows that

$$h_\kappa(\beta) - \underline{\Phi}_\kappa^\star \ \geq \ \frac{\mu}{2} \left(\, \overline{\Phi}_\kappa^\star - \underline{\Phi}_\kappa^\star \, \right) \tag{46}$$

holds for any $\beta \in \hat{I}_\kappa$. The dual iterate β_κ 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 $\boldsymbol{x}_{\kappa+1}$ is selected by applying an unconstrained Inexact-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), and of the inexact version, in Fábián (2000).) For $1 \leq \iota \leq \kappa$, the linear function $l_\iota + \beta_\kappa l_\iota^\sharp$ is a δ_ι-support function of $\varphi + \beta_\kappa \psi$ at $\boldsymbol{x}_\iota$. Hence $\varphi_\kappa + \beta_\kappa \psi_\kappa$ is an appropriate cutting-plane model of $\varphi + \beta_\kappa \psi$.

The *best upper approximation*, i.e., the lowest known upper approximation of the function values is $BUA := \min\limits_{1 \leq \iota \leq \kappa} \{ \, \varphi_\kappa(\boldsymbol{x}_\iota) + \beta_\kappa \psi(\boldsymbol{x}_\iota) + \delta_\iota \} \ = h_\kappa(\beta_\kappa)$.

A *lower model value* is a certain value of the cutting-plane model function. We select it specially as $LMV := \underline{\Phi}^\star_\kappa$. The lower model value in the Inexact Level Method must satisfy two requirements:

(1) $BUA \geq LMV$ should hold.

(2) There should exist a point in $\mathcal{X}$ whose model-function value is lower than or equal to LMV.

In the present case,

(1) $BUA = h_\kappa(\beta_\kappa) \;\geq\; \underline{\Phi}^\star_\kappa = LMV$ holds owing to the selection $\beta_\kappa \in I_\kappa$ of the dual iterate.

(2) Let $\underline{\boldsymbol{x}}_\kappa$ denote the minimizer of the model problem (38). Obviously we have $\varphi_\kappa(\underline{\boldsymbol{x}}_\kappa) + \beta_\kappa \psi_\kappa(\underline{\boldsymbol{x}}_\kappa) \leq \underline{\Phi}^\star_\kappa$.

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

$$\mathcal{X}_\kappa := \left\{\, \boldsymbol{x} \in \mathcal{X} \;\middle|\; \varphi_\kappa(\boldsymbol{x}) + \beta_\kappa \psi_\kappa(\boldsymbol{x}) \;\leq (1-\theta)\underline{\Phi}^\star_\kappa + \theta h_\kappa(\beta_\kappa) \,\right\}.$$

This is a convex polyhedron. The next primal iterate $\boldsymbol{x}_{\kappa+1}$ will be the projection of the former iterate $\boldsymbol{x}_\kappa$ onto the level set $\mathcal{X}_\kappa$. (I.e., the point of $\mathcal{X}_\kappa$ that is closest to $\boldsymbol{x}_\kappa$ 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) Inexact Level Method. First we estimate the length of a single run. Let ε be a positive number (different from the stopping tolerance ϵ of the constrained method). Fábián proved that to obtain $GAP < \varepsilon$ in the Inexact Level Method, it suffices to perform

$$c\left(\frac{D\overline{\Lambda}}{\varepsilon}\right)^2 \tag{47}$$

iterations, where D is the diameter of the feasible polyhedron, $\overline{\Lambda}$ is a Lipschitz constant of the objective function (presently $\phi + \beta_\kappa\psi$), and c is a constant that depends only on the parameter θ. (The proof is based on the convergence proof of the Level Method given by Lemaréchal, Nemirovskii, and Nesterov.) Obviously we have $\overline{\Lambda} \leq (1 + \overline{\beta})\Lambda$ where Λ is a common Lipschitz constant of the functions ϕ and ψ.

In the present case, the GAP after κ iterations is $h_\kappa(\beta_\kappa) - \underline{\Phi}^\star_\kappa$. If it has decreased below ε, then we have

$$\varepsilon > \frac{\mu}{2}\left(\,\overline{\Phi}^\star_\kappa - \underline{\Phi}^\star_\kappa\,\right)$$

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

$$c\left(\frac{2}{\mu}\right)^2\left(1+\overline{\beta}\right)^2\left(\frac{D\Lambda}{\epsilon}\right)^2 \tag{48}$$

on the length of a single run of (unconstrained) Inexact 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 σth run. Then

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

holds due to the selection of the dual iterate. (Indeed, the shrunken 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_κ is the support of the function $h_\kappa(\beta) - \underline{\Phi}^\star_\kappa$, and $\overline{\Phi}^\star_\kappa - \underline{\Phi}^\star_\kappa$ 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

$$\max_{\boldsymbol{x}\in\mathcal{X}}|\psi(\boldsymbol{x})| \leq D\Lambda.$$

(The function ψ is assumed to take negative as well as positive values over $\mathcal{X}$.)

The length of the initial interval is not larger than $\overline{\beta}$. Owing to the geometric progression, the number of the runs can not be larger than

$$c'\,\ln\left(\overline{\beta}\,\frac{D\Lambda}{\epsilon}\right),$$

where c' is a constant that depends only on the parameter μ.

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

$$c'' \left(1+\overline{\beta}\right)^2 \left(\frac{D\Lambda}{\epsilon}\right)^2 \ln\left(\overline{\beta}\,\frac{D\Lambda}{\epsilon}\right)$$

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