On the Global Minimization of the Value-at-Risk*

Jong-Shi Pang† Sven Leyffer‡

January 7, 2004

Abstract

In this paper, we consider the nonconvex minimization problem of the value-at-risk (VaR) that arises from financial risk analysis. By considering this problem as a special linear program with linear complementarity constraints (a bilevel linear program to be more precise), we develop upper and lower bounds for the minimum VaR and show how the combined bounding procedures can be used to compute the latter value to global optimality. A numerical example is provided to illustrate the methodology.

Dedication. With great pleasure we dedicate this paper to a respected pioneer of our field, Professor Olvi L. Mangasarian, on the occasion of his 70th birthday. The two topics of this paper, LPECs and smoothing methods, are examples of the vast contributions that Olvi has made in optimization, which have benefited us in many ways and which will continue to benefit us in the future. Happy 70th birthday, Olvi!

1 Introduction

The value-at-risk (VaR) and conditional value-at-risk (CVaR) are two important risk measures that have been used extensively in recent years in portfolio selection and in risk analysis. Whereas the VaR is closely related to a particular quantile of a random variable, the CVaR is formally defined and analyzed by Rockafellar and Uryasev in two papers [32, 33] as a way to alleviate some of the computational difficulties associated with the optimization of the VaR. There is now a substantial literature on the applications and further developments of these two risk measures; a partial list of this literature relevant to optimization includes the papers [1, 18, 19, 27, 29, 34, 35, 40]. In particular, the paper [19] presents some CVaR-based algorithms for computing the VaR in a portfolio selection problem; in spite of their practical efficiency, however, these algorithms offer no guarantee of global optimality of the computed VaR.

Setting aside some criticisms of the VaR mentioned in the literature, part of which stems from the difficulty associated with the portfolio selection problem using the VaR criterion, we study the global optimization problem using a scenario formulation, which is the principal approach employed in the cited references for solving the (C)VaR minimization problem. Specifically, we consider the VaR minimization problem as an LPEC [24], a linear program with equilibrium constraints, which is a subject pioneered by Mangasarian, to whom this paper is dedicated. By exploiting the special structure of this program, we derive linear programs whose optimum objective values yield upper

*Preprint ANL/MCS-P1112-1203

†Department of Mathematical Sciences, Rensselaer Polytechnic Institute, Troy, New York 12180-3590, U.S.A. Email: pang@mts.jhu.edu. The work of this author's research was partially supported by the National Science Foundation under grant CCR-0098013.

‡Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, Illinois 60439. Supported by the Mathematical, Information, and Computational Sciences Division subprogram of the Office of Advanced Scientific Computing Research, Office of Science, U.S. Department of Energy, under Contract W-31-109-ENG-38.

and lower bounds for the optimal VaR. The bounding procedures are then used in a branch-and-cut algorithm for computing the latter value to global optimality. A numerical example is provided to illustrate the algorithm.

An LPEC is a special case of a mathematical program with equilibrium constraints (MPEC). Since the publication of the two monographs [22, 28], there has been significant computational advance in numerical methods for solving MPECs; a partial list of recent references includes [7, 11, 12, 13, 9, 10, 14, 15, 16, 20, 37, 36, 38]. In spite of such extensive efforts, the computation of globally optimal solutions to MPECs remains elusive. While some MPEC solvers are fairly robust in practice, there is no guarantee that their computed solutions are globally optimal solutions. An important reason for this lack of guarantee for global optimality is the fact that these solvers are all based on local improvement techniques and no global optimization is incorporated in their implementation. As a special MPEC, the VaR minimization problem is amenable to solution by any one of the (local) methods. In this paper, we do not stop with this routine adaptation of the existing MPEC solvers; instead, our goal is to develop a branch-and-cut algorithm for solving the minimum VaR problem to global optimality.

2 The VaR Minimization Problem

Let y denote an n-dimensional random vector whose components represent the random losses of some financial instruments. Let $X \subseteq \Re^n$ be a closed convex set (polyhedral in many practical applications) representing the set of feasible investments. For a given $x \in \Re^n$, $z \equiv x^T y$ is therefore the random loss associated with the investment vector x. For a given scalar $\beta \in (0,1)$, which denotes a confidence threshold of sustainable loss, the CVaR and VaR associated with the random variable z is given, as proved in [32, 33], by the following two deterministic quantities, respectively:

$$\begin{aligned} \mathrm{CVaR}_\beta(x) &\equiv \min_{m \in \Re} \left[m + \frac{1}{1-\beta} \mathbb{E}_y (\, x^T y - m \,)_+ \right], \\ \mathrm{VaR}_\beta(x) &\equiv \min\{\, m : m \in \mathcal{M}_\beta(x) \,\}, \end{aligned}$$

where $\mathbb{E}_y$ denotes the expectation with respect to the random vector y, the subscript plus sign denotes the nonnegative part of a scalar (i.e., the *plus function* $t_+ \equiv \max(0,t)$), and $\mathcal{M}_\beta(x)$ denotes the set of minimizers in the definition of $\mathrm{CVaR}_\beta(x)$. By the results in the cited references, $\mathrm{CVaR}_\beta(x)$ and $\mathrm{VaR}_\beta(x)$ are well-defined finite scalars for very general loss distributions. Clearly, we have

$$\mathrm{CVaR}_\beta(x) = \mathrm{VaR}_\beta(x) + \frac{1}{1-\beta} \mathbb{E}_y (x^T y - \mathrm{VaR}_\beta(x))_+ \geq \mathrm{VaR}_\beta(x), \quad \forall\, x.$$

The CVaR and VaR minimization problems are, respectively,

$$\left\{ \begin{array}{ll} \text{minimize} & \mathrm{CVaR}_\beta(x) \\ \text{subject to} & x \in X \end{array} \right\} \quad \text{and} \quad \left\{ \begin{array}{ll} \text{minimize} & \mathrm{VaR}_\beta(x) \\ \text{subject to} & x \in X \end{array} \right\}.$$

Clearly, the CVaR minimization problem can be cast equivalently as the following convex program in the joint variable (m, x):

$$\begin{array}{ll} \text{minimize} & m + \dfrac{1}{1-\beta} \mathbb{E}_y (\, x^T y - m \,)_+ \\ \text{subject to} & (\, m, x \,) \in \Re \times X. \end{array}$$

Nevertheless, the VaR problem is *not* a convex program; this fact is an acknowledged drawback of using the VaR as a criterion in portfolio selection. Our main goal in this paper is to develop remedies to this drawback.

2.1 An LPEC Formulation

In the rest of the paper, we take X to be a compact polyhedron. We adopt a scenario approach to discretize the random vector y. With this approach, the CVaR minimization problem becomes a linear program (LP) and the VaR becomes a bilevel linear program, which we reformulate as an LPEC using the optimality conditions of the lower-level LP. Specifically, let $\{y^1, \cdots, y^k\}$ be the finite set of scenario values of y, and let $\{p_1, \cdots, p_k\}$ be the associated probabilities of the respective scenarios, which, summing to one, are assumed to be all positive. The discretized CVaR minimization problem is

$$\begin{array}{ll} \text{minimize} & m + \dfrac{1}{1-\beta} \displaystyle\sum_{i=1}^{k} p_i \, (x^T y^i - m)_+ \\ \text{subject to} & (\, m, x \,) \in \Re \times X, \end{array}$$

which is equivalent to the linear program in the variables (m, x, τ):

$$\begin{array}{ll} \text{minimize} & m + \dfrac{1}{1-\beta} \displaystyle\sum_{i=1}^{k} p_i \, \tau_i \\ \text{subject to} & x \in X \\ \text{and} & \left\{ \begin{array}{lcl} \tau_i & \geq & 0 \\ \tau_i & \geq & x^T y^i - m \end{array} \right\} \quad \forall \, i = 1, \ldots, k. \end{array} \tag{1}$$

For a given $x \in X$, $\mathrm{CVaR}_\beta(x)$ is the minimum objective value of the following simple LP in the variable $(m, \tau) \in \Re^{1+k}$:

$$\begin{array}{ll} \text{minimize} & m + \dfrac{1}{1-\beta} \displaystyle\sum_{i=1}^{k} p_i \, \tau_i \\ \text{subject to} & \left\{ \begin{array}{lcl} \tau_i & \geq & 0 \\ \tau_i & \geq & x^T y^i - m \end{array} \right\} \quad \forall \, i = 1, \ldots, k. \end{array} \tag{2}$$

By letting λ_i denote the dual variable of the ith functional constraint in (2), the above LP can be solved trivially via its dual:

$$\begin{array}{ll} \text{maximize} & \displaystyle\sum_{i=1}^{k} \lambda_i \, x^T y^i \\ \text{subject to} & 0 \leq \lambda_i \leq p_i/(1-\beta), \quad \forall \, i = 1, \ldots, k \\ \text{and} & \displaystyle\sum_{i=1}^{k} \lambda_i = 1, \end{array} \tag{3}$$

which is a bounded knapsack problem that can in turn be solved by a simple sorting procedure. The optimal objective value of either (2) or (3) yields $\mathrm{CVaR}_\beta(x)$; this shows in particular that $\mathrm{CVaR}_\beta(x)$ is a convex combination of the portfolio losses $\{x^T y^1, \cdots, x^T y^k\}$.

In general, by solving either of the LPs (2) or (3), we are not guaranteed to obtain $\mathrm{VaR}_\beta(x)$ right away; to obtain the latter value, we can solve another simple LP in the variable (m, τ), with

x remaining fixed:

$$\begin{array}{ll} \text{minimize} & m \\ \text{subject to} & m + \dfrac{1}{1-\beta}\displaystyle\sum_{i=1}^{k} p_i\,\tau_i \;\le\; \mathrm{CVaR}(x) \\ \text{and} & \left\{\begin{array}{lcl} \tau_i & \ge & 0 \\ \tau_i & \ge & x^T y^i - m \end{array}\right\} \quad \forall\, i = 1,\ldots,k, \end{array}$$

which is simply the problem of finding the least element of the argmin $\mathcal{M}_\beta(x)$.

The optimality conditions of (2) are

$$\left\{\begin{array}{lllll} 0 & \le & \tau_i & \perp & \dfrac{p_i}{1-\beta} - \lambda_i \ge 0 \\ 0 & \le & \lambda_i & \perp & s_i \equiv m + \tau_i - x^T y^i \ge 0 \end{array}\right\} \quad \forall\, i = 1,\ldots,k$$

$$\text{and} \quad \sum_{i=1}^{k} \lambda_i = 1,$$

where the $\perp$ denotes the well-known complementary slackness condition. Employing these optimality conditions, we can reformulate the VaR minimization problem as the following linear program with linear complementarity constraints in the variables (m, x, τ, λ), that is, an LPEC, which in turn is a special subclass of the class of mathematical programs with equilibrium constraints [22]:

$$\begin{array}{ll} \text{minimize} & m \\ \text{subject to} & x \in X \\ & \left\{\begin{array}{lllll} 0 & \le & \tau_i & \perp & \dfrac{p_i}{1-\beta} - \lambda_i \ge 0 \\ 0 & \le & \lambda_i & \perp & s_i \equiv m + \tau_i - x^T y^i \ge 0 \end{array}\right\} \quad \forall\, i = 1,\ldots,k \\ \text{and} & \displaystyle\sum_{i=1}^{k} \lambda_i = 1. \end{array} \tag{4}$$

As an LPEC, the feasible region of (4) is the union of finitely many polyhedra. Exploiting its special structure, we state and prove in the result below that (4) attains a finite minimum objective value.

Proposition 2.1 Let X be a compact polyhedron in $\Re^n$. The LPEC (4) attains a finite minimum objective value.

Proof. Since X is compact by assumption, one can easily show that m must be bounded below on the feasible region of (4). In fact, if (m_ν, τ^ν, x^ν) is a sequence of feasible solutions with $m_\nu \to -\infty$, then $\tau_i^\nu \to \infty$ for every i. Consequently, $\lambda_i^\nu = p_i/(1-\beta)$; but this contradicts the last constraint, which requires that the sum of the λ's be equal to unity. $\square$

Let $(m_{\mathrm{VaR}}, x^{\mathrm{VaR}}, \tau^{\mathrm{VaR}}, \lambda^{\mathrm{VaR}})$ denote an optimal solution of (4). Note that whereas m_{VaR} must be unique, the triple $(x^{\mathrm{VaR}}, \tau^{\mathrm{VaR}}, \lambda^{\mathrm{VaR}})$ is not necessarily so. Our goal is to compute m_{VaR} as best as possible. Although a theoretical guarantee of global optimality is not easy to obtain, we derive valid upper and lower bounds for m_{VaR} and develop ways to tighten these bounds; obviously, when the upper and bounds coincide, then m_{VaR} is obtained.

3 Upper and Lower Bounds

In this section, we develop valid upper and lower bounds for m_{VaR}. While upper bounds are not difficult to compute, sharp lower bounds are less obvious to derive. We formally describe these bounds in the next two subsections. Here, we note that if $m_{\text{VaR}} \in [m_{\text{LB}}, m_{\text{UB}}]$, then

$$0 \leq \max\left(\frac{m_{\text{UB}} - m_{\text{VaR}}}{m_{\text{VaR}}}, \frac{m_{\text{VaR}} - m_{\text{LB}}}{m_{\text{VaR}}} \right) \leq \frac{m_{\text{UB}} - m_{\text{LB}}}{m_{\text{LB}}},$$

which gives relative accuracies of the upper and lower bound values, m_{UB} and m_{LB}, respectively, with respect to the exact minimum VaR m_{VaR}.

3.1 LP Upper Bounds

In essence, upper bounds for m_{VaR} are obtained by "breaking" the complementary slackness in (4) (i.e., restricting the feasible region) according to a given feasible solution. Let $x^0 \in X$ be given. The scalar $m_0 \equiv \text{VaR}_\beta(x^0)$ provides an upper bound for m_{VaR}. (For instance, we may take $x^0 = x^{\text{CVaR}}$ to be an optimal solution of the CVaR linear program (1).) We wish to improve on the bound m_0 by considering a restriction of the constraints in (4). Specifically, associated with the pair (m_0, x^0), let (τ^0, λ^0) satisfy

$$\left\{ \begin{array}{lllll} 0 & \leq & \tau_i^0 & \perp & \dfrac{p_i}{1-\beta} - \lambda_i^0 \geq 0 \\ 0 & \leq & \lambda_i^0 & \perp & s_i^0 \equiv m_0 + \tau_i^0 - (\,x^0\,)^T y^i \;\geq\; 0 \end{array} \right\} \quad \forall\, i = 1, \ldots, k$$

$$\text{and} \quad \sum_{i=1}^{k} \lambda_i^0 = 1.$$

Define the index sets

$$\begin{aligned} \alpha_\tau^0 &\equiv \left\{ i : \tau_i^0 > 0 = \frac{p_i}{1-\beta} - \lambda_i^0 \right\} \\ \beta_\tau^0 &\equiv \left\{ i : \tau_i^0 = 0 = \frac{p_i}{1-\beta} - \lambda_i^0 \right\} \\ \gamma_\tau^0 &\equiv \left\{ i : \tau_i^0 = 0 < \frac{p_i}{1-\beta} - \lambda_i^0 \right\} \end{aligned}$$

and

$$\begin{aligned} \alpha_\lambda^0 &\equiv \{\, i : \lambda_i^0 > 0 = m_0 + \tau_i^0 - (\,x^0\,)^T y^i \,\} \\ \beta_\lambda^0 &\equiv \{\, i : \lambda_i^0 = 0 = m_0 + \tau_i^0 - (\,x^0\,)^T y^i \,\} \\ \gamma_\lambda^0 &\equiv \{\, i : \lambda_i^0 = 0 < m_0 + \tau_i^0 - (\,x^0\,)^T y^i \,\}. \end{aligned}$$

Let δ_τ^0 and δ_λ^0 be arbitrary subsets of β_τ^0 and β_λ^0, respectively. Consider the following linear program in the variables (m, x, τ, λ):

$$\begin{array}{ll} \text{minimize} & m \\ \text{subject to} & x \in X \\ & \tau_i \geq 0 = \dfrac{p_i}{1-\beta} - \lambda_i, \quad \forall\, i \in \alpha_\tau^0 \cup \delta_\tau^0 \\ & \tau_i = 0 \leq \dfrac{p_i}{1-\beta} - \lambda_i, \quad \forall\, i \in \gamma_\tau^0 \cup (\beta_\tau^0 \setminus \delta_\tau^0) \\ & \lambda_i \geq 0 = m + \tau_i - x^T y^i, \quad \forall\, i \in \alpha_\lambda^0 \cup \delta_\lambda^0 \\ & \lambda_i = 0 \leq m + \tau_i - x^T y^i, \quad \forall\, i \in \gamma_\lambda^0 \cup (\beta_\lambda^0 \setminus \delta_\lambda^0) \\ \text{and} & \displaystyle\sum_{i=1}^k \lambda_i = 1, \end{array} \tag{5}$$

which is obtained by restricting the complementarity constraints in (4) based on the above index sets. Obviously, (5) is equivalent to a simplified LP in the variables (m, τ) only, with the λ variable being removed, that is,

$$\begin{array}{ll} \text{minimize} & m \\ \text{subject to} & x \in X \\ & \tau_i \geq 0, \quad \forall\, i \in \alpha_\tau^0 \cup \delta_\tau^0 \\ & \tau_i = 0, \quad \forall\, i \in \gamma_\tau^0 \cup (\beta_\tau^0 \setminus \delta_\tau^0) \\ & m + \tau_i - x^T y^i = 0, \quad \forall\, i \in \alpha_\lambda^0 \cup \delta_\lambda^0 \\ & m + \tau_i - x^T y^i \geq 0, \quad \forall\, i \in \gamma_\lambda^0 \cup (\beta_\lambda^0 \setminus \delta_\lambda^0). \end{array} \tag{6}$$

It is clear that (m_0, x^0, τ^0) is feasible to (6). In general, if (m, x, τ) is feasible to (6), then (m, x, τ, λ^0) is feasible to (5), and hence to (4). Consequently, (6) attains a finite global minimum. Moreover, if $\left(m_{1+\frac{1}{2}}, x^1, \tau^1\right)$ denotes an optimal solution of (6), we must have $m_{1+\frac{1}{2}} \in \mathcal{M}_\beta(x^1)$. Hence, with $m_1 \equiv \mathrm{VaR}_\beta(x^1)$, we have

$$m_0 \geq m_{1+\frac{1}{2}} \geq m_1 \geq m_{\mathrm{VaR}}.$$

One of two cases must occur: (a) $m_0 = m_1$ (no improvement), or (b) $m_0 > m_1$ (strict improvement). In case (a), no improvement is obtained with the particular choice of the pair of index sets $(\delta_\tau^0, \delta_\lambda^0)$. One can then try a new pair and solve a new LP (6), hoping to obtain a strictly improved bound for m_{VaR}. In case (b), we can replace (m_0, x^0) by the pair (m_1, x^1) and repeat the above procedure. The following result shows that if strict improvement is obtained at each iteration, then in a finite number of steps the exact minimum VaR is found.

Theorem 3.1 Let $\{x^\nu\} \subset X$ be a sequence of feasible vectors such that for each ν, $x^{\nu+1}$ is obtained from x^ν by solving a certain restricted LP as described above. If $\mathrm{VaR}_\beta(x^\nu) > \mathrm{VaR}_\beta(x^{\nu+1})$ for every ν, then a finite ν_0 exists such that $\mathrm{VaR}_\beta(x^{\nu_0}) = m_{\mathrm{VaR}}$.

Proof. The feasible region of (4) is the union of finitely many polyhedra, each being the feasible set of (5) corresponding to a particular tuple of index sets $(\alpha_\tau^0, \delta_\tau^0, \gamma_\tau^0, \alpha_\lambda^0, \delta_\lambda^0, \gamma_\lambda^0)$. Since $\mathrm{VaR}_\beta(x^\nu) >$

$\mathrm{VaR}_\beta(x^{\nu+1})$ for every ν, the tuples of index sets used to produce the sequence $\{x^\nu\}$ cannot repeat. Since there are only finitely many such tuples of index sets, in generating the sequence $\{x^\nu\}$ we must have encountered all of them; in other words, we must have searched over the entire feasible region of (4). Consequently, we must have $\mathrm{VaR}_\beta(x^{\nu_0}) = m_{\mathrm{VaR}}$ for some ν_0. $\square$

The above result is mainly of theoretical interest because rarely is one so lucky that strict improvement can be obtained with each trial choice of $(\delta_\tau^0, \delta_\lambda^0)$. Notice that the procedure described herein is based on the premise that the set $\beta_\tau^0 \cup \beta_\lambda^0$ is nonempty, which means that the pair (m_0, x^0) is a *degenerate* feasible solution of (4), degenerate with reference to the complementarity conditions. When (m_0, x^0) is nondegenerate, we will not able to continue the procedure. Consequently, this is one of the rare instances in mathematical programming where *degeneracy actually helps*: it enables one to continue the search for an improvement in a global optimization procedure.

3.2 NLP Upper Bounds

An alternative approach to obtain an upper bound is to form the equivalent nonlinear program (NLP) of the LPEC (4)

$$
\begin{array}{ll}
\text{minimize} & m \\
\text{subject to} & x \in X \\
& \left\{ \begin{array}{lll} 0 & \leq & \tau_i, \quad \dfrac{p_i}{1-\beta} - \lambda_i \geq 0 \\ 0 & \leq & \lambda_i, \quad s_i \equiv m + \tau_i - x^T y^i \geq 0 \end{array} \right\} \quad \forall\, i = 1, \ldots, k \\
& \displaystyle\sum_{i=1}^{k} \lambda_i = 1, \\
\text{and} & \displaystyle\sum_{i=1}^{k} \left[\tau_i \left(\frac{p_i}{1-\beta} - \lambda_i \right) + \lambda_i s_i \right] \leq 0,
\end{array}
\tag{7}
$$

and solve this NLP using standard solvers. The last constraint in this problem is the complementarity constraint. Note that we do not require a lower bound on the complementarity constraint in (7) because all terms in this expression are nonnegative.

It is well known that the NLP (7) fails the Mangasarian-Fromovitz constraint qualification (MFCQ) [23, 25] at *any* feasible point. This fact implies that the multiplier set of (7) is unbounded, the central path fails to exist, and active constraint normals are linearly dependent. As a consequence, solving MPECs as NLPs has been commonly regarded as numerically unsafe. Recently, however, it has been demonstrated that standard NLP solvers can be employed to solve the equivalent NLPs of MPECs reliably and efficiently. The convergence of sequential quadratic programming methods to a "stationary point" of an MPEC is analyzed in [2, 10], and the extension of interior point methods to MPECs is described in [21, 31]. For other related methods, see [7, 9, 11, 13, 15, 16], and the monographs [22, 28].

Unfortunately, solving the equivalent NLP (7) does not in itself guarantee global optimality, despite the practical success of NLP solvers. The reason is that the nonconvex nature of the complementarity constraint implies that NLP solvers may fail to find the global minimum, or even a feasible point. Nevertheless, NLP solvers have been shown to provide good solutions for many practical MPECs [9, 30], and this is the feature we wish to exploit here. In fact, for the numerical example reported in Section 5, an NLP solver finds a solution, which we show through additional techniques is a global minimum. We note that the latter proof is demonstrated not by NLP but rather by exhibiting an upper bound for m_{VaR} that coincides with a lower bound.

3.3 LP Lower Bounds

Upper bounding alone is not enough to verify global optimality of a nonconvex problem. In this subsection, we develop some valid lower bounds for m_{VaR}. As a first remark, we note that the simple LP relaxation of (4) is

$$\begin{array}{ll} \text{minimize} & m \\ \text{subject to} & x \in X \\ & \left\{ \begin{array}{lll} 0 & \leq & \tau_i, \quad \dfrac{p_i}{1-\beta} - \lambda_i \geq 0 \\ 0 & \leq & \lambda_i, \quad m + \tau_i - x^T y^i \geq 0 \end{array} \right\} \quad \forall\, i = 1, \ldots, k \\ \text{and} & \displaystyle\sum_{i=1}^{k} \lambda_i = 1, \end{array} \tag{8}$$

which does not have a finite optimal solution because we can make m tend to $-\infty$ with each $\tau_i \to \infty$. Therefore, we need to tighten this relaxation. The following lemma gives a preliminary lower bound for m_{VaR}.

Lemma 3.2 For any feasible tuple (m, x, τ, λ) to (4), an index i exists such that $m \geq x^T y^i$ for at least one index i. Consequently,

$$m \geq \min_{1 \leq j \leq k} \min_{x \in X} x^T y^j \equiv \underline{m}.$$

Proof. Let (m, x, τ, λ) be an arbitrary feasible tuple to (4). We must have

$$\tau_i = \max(0, x^T y^i - m), \quad \forall\, i = 1, \ldots, k.$$

From the first complementarity constraint in (4), we obtain

$$\tau_i \, \lambda_i = \frac{p_i}{1-\beta} \tau_i, \tag{9}$$

which, when used in the second complementarity constraint, yields

$$0 = m \, \lambda_i + \frac{p_i}{1-\beta} \tau_i - \lambda_i \, x^T y^i. \tag{10}$$

Since the sum of the λ_i is equal to unity, we deduce

$$m = \sum_{j=1}^{k} \left[\lambda_j \, x^T y^j - \frac{p_j}{1-\beta} \tau_j \right]. \tag{11}$$

which yields

$$m + \sum_{j=1}^{k} \frac{p_j}{1-\beta} \max(0, x^T y^j - m) = \sum_{j=1}^{k} \lambda_j \, x^T y^j.$$

If no index i exists such that $x^T y^i \leq m$, then $m < x^T y^j$ for all j, and the above identity yields

$$m = \frac{1-\beta}{\beta} \sum_{j=1}^{k} \left(\frac{p_j}{1-\beta} - \lambda_j \right) x^T y^j,$$

which shows that m is a convex combination of the family $\{x^T y^1, \cdots, x^T y^k\}$. This is a contradiction. The last assertion of the lemma is obvious. $\square$.

In essence, the lower bounding procedure described below aims at removing the three nonlinear terms $m\lambda_i$, $\tau\lambda_i$, and $\lambda_i x$ in (9) and (10), which are the result of the complementarity constraints, while maintaining some form of these two equations. It turns out that the first two nonlinear terms can be completely removed through some suitable substitution, whereas the third one cannot. The relaxation of (4) then employs a single variable z^i to substitute for $\lambda_i x$ and to remove the identity $z^i = \lambda_i x$ when λ_i is strictly between its lower and upper bounds. Note that the change of variables implies

$$x = \sum_{i=1}^{k} z^i.$$

Furthermore, if $x \geq 0$ in the set X, it follows that $0 \leq z^i \leq (p_i/(1-\beta))x$ for all i. More generally, if $|x| \leq a$ for all $x \in X$, where a is a given nonnegative vector and the absolute sign is meant componentwise, then $|z^i| \leq (p_i/(1-\beta))a$ for all i.

From the identity (10), we deduce that for any feasible tuple (m, x, τ, λ) to (4),

$$m \geq 0 \;\Rightarrow\; \forall\, i, \left[0 \leq (z^i)^T y^i - \frac{p_i}{1-\beta}\tau_i \leq \frac{p_i}{1-\beta} m\right]$$

$$m \leq 0 \;\Rightarrow\; \forall\, i, \left[0 \geq (z^i)^T y^i - \frac{p_i}{1-\beta}\tau_i \geq \frac{p_i}{1-\beta} m\right].$$

Assume for the moment that $m \geq 0$. This gives rise to the following LP relaxation of (4) (we assume that $|x| \leq a$ for all $x \in X$):

$$\begin{aligned}
&\text{minimize} \quad m \\
&\text{subject to} \quad x \equiv \sum_{j=1}^{k} z^j \in X, \\
&\qquad m = \sum_{j=1}^{k} \left[(z^j)^T y^j - \frac{p_j}{1-\beta}\tau_j \right] \\
&\qquad \left\{ \begin{array}{r} s_i \equiv m + \tau_i - x^T y^i \geq 0 \\ 0 \leq (z^i)^T y^i - \dfrac{p_i}{1-\beta}\tau_i \leq \dfrac{p_i}{1-\beta} m \\ \tau_i \geq 0 \\ |z^i| \leq \dfrac{p_i}{1-\beta} a \end{array} \right\}, \quad \forall\, i = 1, \ldots, k.
\end{aligned} \tag{12}$$

We are also interested in investigating the behavior of this lower bound if we branch on a disjunction. It follows from (4) that there are three possible branches for λ_i. Each branch in turn gives rise to a particular implication:

$$\begin{aligned}
0 = \lambda_i &\Rightarrow [z^i = 0 \quad \text{and} \quad \tau_i = 0] \\
\lambda_i = \frac{p_i}{1-\beta} &\Rightarrow \left[z^i = \frac{p_i}{1-\beta} x \quad \text{and} \quad s_i = 0\right] \\
0 < \lambda_i < \frac{p_i}{1-\beta} &\Rightarrow [\tau_i = 0 \quad \text{and} \quad s_i = 0].
\end{aligned}$$

Based on the above implications, we define three LPs by adding the implications to the lower bounding LP (12), respectively. In what follows, i_0 is a fixed but arbitrary index in $\{1, \ldots, k\}$.

($\mathbf{LP}^{+}_{\mathrm{I},\mathbf{i}_0}$) This corresponds to the case where $\lambda_{i_0} = 0$ and $m \geq 0$ and consists of the LP relaxation (12) with the following additional constraints:

$$z^{i_0} = 0 \quad \text{and} \quad \tau_{i_0} = 0. \tag{13}$$

($\mathbf{LP}^{+}_{\mathrm{II},\mathbf{i}_0}$) This corresponds to the case where $\lambda_{i_0} = p_{i_0}/(1-\beta)$ and $m \geq 0$ and consists of the LP relaxation (12) with the following additional constraints:

$$z^{i_0} = \frac{p_{i_0}}{1-\beta} x \quad \text{and} \quad s_{i_0} = 0. \tag{14}$$

($\mathbf{LP}^{+}_{\mathrm{III},\mathbf{i}_0}$) This corresponds to the case where $\lambda_{i_0} \in (0, p_{i_0}/(1-\beta))$ and $m \geq 0$ and consists of the LP relaxation (12) with the following additional constraints:

$$\tau_{i_0} = s_{i_0} = 0. \tag{15}$$

Let $\mathrm{LP}^{+,\mathrm{opt}}_{\mathrm{I},\mathbf{i}_0}$, $\mathrm{LP}^{+,\mathrm{opt}}_{\mathrm{II},\mathbf{i}_0}$, and $\mathrm{LP}^{+,\mathrm{opt}}_{\mathrm{III},\mathbf{i}_0}$ denote the optimal objective values of the above three LPs, respectively. Consistent with a standard convention in optimization, we define the minimum objective value of an infeasible LP to be ∞. The next result summarizes the fundamental role of the above LPs for solving the VaR minimization problem (4).

Proposition 3.3 For any index i_0, the following five statements (a)–(e) are valid.

(a) If (m, x, τ, λ) is feasible to (4) and $m \geq 0$, then with $z^i \equiv \lambda_i x$ for all i, the tuple (m, x, z, τ) is feasible to (12). It also satisfies the additional constraints (13) if $\lambda_{i_0} = 0$, (14) if $\lambda_{i_0} = p_{i_0}/(1-\beta)$, and (15) if $0 < \lambda_{i_0} < p_{i_0}/(1-\beta)$.

(b) If $m_{\mathrm{VaR}} \geq 0$, then at least one of the three LPs obtained by adding to (12) the cuts (13), or (14), or (15), must be feasible and, hence, solvable; in this case,

$$m_{\mathrm{VaR}} \geq \min\left(\mathrm{LP}^{+,\mathrm{opt}}_{\mathrm{I},\mathbf{i}_0},\ \mathrm{LP}^{+,\mathrm{opt}}_{\mathrm{II},\mathbf{i}_0},\ \mathrm{LP}^{+,\mathrm{opt}}_{\mathrm{III},\mathbf{i}_0} \right). \tag{16}$$

(c) If $m_{\mathrm{UB}} \geq m_{\mathrm{VaR}} \geq 0$ and $\infty > \mathrm{LP}^{+,\mathrm{opt}}_{\mathrm{I},\mathbf{i}_0} > m_{\mathrm{UB}}$, then for any optimal solution $(x^{\mathrm{VaR}}, \tau^{\mathrm{VaR}}, \lambda^{\mathrm{VaR}})$ of (4), we must have $\lambda^{\mathrm{opt}}_{i_0} > 0$, and thus, $s^{\mathrm{VaR}}_{i_0} \equiv \tau^{\mathrm{VaR}}_{i_0} + m_{\mathrm{VaR}} - (x^{\mathrm{VaR}})^T y^{i_0} = 0$.

(d) If $m_{\mathrm{UB}} \geq m_{\mathrm{VaR}} \geq 0$ and $\infty > \mathrm{LP}^{+,\mathrm{opt}}_{\mathrm{II},\mathbf{i}_0} > m_{\mathrm{UB}}$, then for any optimal solution $(x^{\mathrm{VaR}}, \tau^{\mathrm{VaR}}, \lambda^{\mathrm{VaR}})$ of (4), we must have $\lambda^{\mathrm{VaR}}_{i_0} < p_{i_0}/(1-\beta)$, and thus, $\tau^{\mathrm{VaR}}_{i_0} = 0$.

(e) If $m_{\mathrm{UB}} \geq m_{\mathrm{VaR}} \geq 0$ and $\infty > \mathrm{LP}^{+,\mathrm{opt}}_{\mathrm{III},\mathbf{i}_0} > m_{\mathrm{UB}}$, then for any optimal solution $(x^{\mathrm{VaR}}, \tau^{\mathrm{VaR}}, \lambda^{\mathrm{VaR}})$ of (4), we must have $\lambda^{\mathrm{VaR}}_{i_0} = 0$ or $\lambda^{\mathrm{VaR}}_{i_0} = p_{i_0}/(1-\beta)$.

Proof. Part (a) does not require a proof. For part (b), we need only to prove the bound (16). Let $(x^{\mathrm{VaR}}, \tau^{\mathrm{VaR}}, \lambda^{\mathrm{VaR}})$ be an arbitrary optimal solution of (4) corresponding to m_{VaR}. the tuple $(m_{\mathrm{VaR}}, x^{\mathrm{VaR}}, z^{\mathrm{VaR}}, \tau^{\mathrm{VaR}})$, where $z^{\mathrm{VaR},\mathbf{i}} \equiv \lambda^{\mathrm{VaR}}_i x^{\mathrm{VaR}}$, is feasible to the one of the three LPs formed from (12) plus (13), or (14), or (15). Hence (16) follows readily. To prove (c), one need only note that if $\lambda^{\mathrm{VaR}}_{i_0} = 0$, then $(m_{\mathrm{VaR}}, x^{\mathrm{VaR}}, z^{\mathrm{VaR}}, \tau^{\mathrm{VaR}})$, where $z^{\mathrm{VaR},\mathbf{i}} \equiv \lambda^{\mathrm{VaR}}_i x^{\mathrm{VaR}}$, is feasible to (13); hence $m_{\mathrm{VaR}} \geq \mathrm{LP}^{+,\mathrm{opt}}_{\mathrm{I},\mathbf{i}_0}$, which easily yields a contradiction. The proof of (d) and (e) is similar and not repeated. $\square$

We can similarly set up three other LPs to handle the case $m_{\text{VaR}} \leq 0$. It suffices to reverse the inequality signs in

$$0 \leq (z^i)^T y^i - \frac{p_i}{1-\beta}\tau_i \leq \frac{p_i}{1-\beta} m$$

and use instead

$$0 \geq (z^i)^T y^i - \frac{p_i}{1-\beta}\tau_i \geq \frac{p_i}{1-\beta} m. \tag{17}$$

Letting $\text{LP}^{-,\text{opt}}_{\text{I},\text{i}_0}$, $\text{LP}^{-,\text{opt}}_{\text{II},\text{i}_0}$, and $\text{LP}^{-,\text{opt}}_{\text{III},\text{i}_0}$ denote the optimal objective values of the resulting LPs, respectively, we can obtain a result similar to Proposition 3.3. Combining these two results, we arrive at a desired lower bound for m_{VaR}.

Corollary 3.4 It holds that

$$m_{\text{VaR}} \geq \min\left\{ \max_{1\leq j\leq k} \min\left(\text{LP}^{+,\text{opt}}_{\text{I},\text{j}},\ \text{LP}^{+,\text{opt}}_{\text{II},\text{j}},\ \text{LP}^{+,\text{opt}}_{\text{III},\text{j}} \right),\ \max_{1\leq j\leq k} \min\left(\text{LP}^{-,\text{opt}}_{\text{I},\text{j}},\ \text{LP}^{+,\text{opt}}_{\text{II},\text{j}},\ \text{LP}^{+,\text{opt}}_{\text{III},\text{j}} \right) \right\}.$$

□

The practical value of the cuts (13), (14), and (15), and their analogs with the reverse inequalities (17) built in, lies in their ability to improve the lower bound obtained from (12) making it easier to fathom nodes in the branch-and-cut framework that is described in Section 4.

3.4 Convex Hull Relaxations

Alternative lower bounds can be derived by observing that the only difference between the simple LP (8) and the LPEC (4) is the absence of the complementarity constraint. Thus, to tighten the former LP relaxation, we form a linear relaxation of the complementarity constraint,

$$\begin{aligned} 0 &= \sum_{i=1}^{k} \left\{ \tau_i \frac{p_i}{1-\beta} - \tau_i\lambda_i + \lambda_i\tau_i + \lambda_i m - \lambda_i x^T y^i \right\} \\ &= m + \sum_{i=1}^{k} \left\{ \tau_i \frac{p_i}{1-\beta} - \lambda_i x^T y^i \right\}, \end{aligned} \tag{18}$$

where we have used the fact that $\sum \lambda_i = 1$. Observe that the only nonlinear term in this expression is given by $\lambda_i x^T y^i$. Next we show how to construct the convex hull relaxation of this constraint.

We introduce new linear variables $\gamma_i = x^T y^i$ and then replace the nonlinear terms $\lambda_i\gamma_i$ by w_i. Since $w_i = \lambda_i\gamma_i$ is a simple bilinear expression, we can strengthen this LP relaxation by adding the convex hull of $w_i = \lambda_i\gamma_i$. Let L_i and U_i be valid lower and upper bounds on γ_i, respectively (these can be obtained by solving $2k$ LPs for instance); the convex hull of $w_i = \gamma_i\lambda_i$ is then given by

$$\left.\begin{aligned} w_i &\geq L_i\,\lambda_i \\ w_i &\geq \frac{p_i}{1-\beta}\gamma_i + U_i\,\lambda_i - \frac{p_i}{1-\beta}U_i \\ w_i &\leq U_i\,\lambda_i \\ w_i &\leq \frac{p_i}{1-\beta}\gamma_i + L_i\,\lambda_i - \frac{p_i}{1-\beta}L_i, \end{aligned}\right\}; \tag{19}$$

see, for example [39]. This gives rise to the LP relaxation

$$
\begin{array}{ll}
\text{minimize} & m \\
\text{subject to} & x = \sum_{i=1}^{k} z^i \in X \\
& \left\{ \begin{array}{lll} 0 & \leq & \tau_i, \quad \dfrac{p_i}{1-\beta} - \lambda_i \geq 0 \\ 0 & \leq & \lambda_i, \quad m + \tau_i - x^T y^i \geq 0 \end{array} \right\} \quad \forall\, i = 1, \ldots, k \\
& \sum_{i=1}^{k} \lambda_i = 1 \\
& \gamma_i = x^T y^i \in [L_i, U_i], \quad \forall\, i = 1, \ldots, k \\
& \left. \begin{array}{lll} w_i & \geq & L_i \lambda_i \\ w_i & \geq & \dfrac{p_i}{1-\beta} \gamma_i + U_i \lambda_i - \dfrac{p_i}{1-\beta} U_i \\ w_i & \leq & U_i \lambda_i \\ w_i & \leq & \dfrac{p_i}{1-\beta} \gamma_i + L_i \lambda_i - \dfrac{p_i}{1-\beta} L_i, \end{array} \right\} \quad \forall\, i = 1, \ldots, k \\
\text{and} & 0 = m + \sum_{i=1}^{k} \left\{ \tau_i \dfrac{p_i}{1-\beta} - w_i \right\}.
\end{array}
\tag{20}
$$

Since this LP includes the convex hull relaxation of $w_i = \gamma_i \lambda_i$, it follows that the LP is bounded whenever the original LPEC (4) is bounded.

The two LP bounds (12) and (20) are nondominating; the quality of the bounds differs from problem instance to problem instance. This is confirmed by the numerical example in Section 5, where, in one case, (12) yields a sharper lower bound, and in the other case, it is (20) that yields a better bound. Next, we show how the bounds (12) and (20) can be employed within a branch-and-cut framework to prove the optimality of a candidate solution of the LPEC (4).

4 Verifying Optimality by Branch-and-Cut

We briefly outline our approach to prove the optimality of a given candidate solution to the MPEC (4). There are two main ideas: the first is to construct as small as possible a branch-and-bound tree corresponding to a given candidate solution, and the second is to exploit the logical implications from the complementarity constraint to strengthen the LP relaxation as in Proposition 3.3. Section 5 shows how the approach works for a numerical example.

The use of branch-and-bound to solve MPECs is not new. It has been used in [3] to solve some bilevel convex programs. However, the scheme proposed here employs the special bounds derived in the preceding section that are tailored to the minimum VaR problem and that are used to define cuts that restrict the feasible region of the problem.

In general, the LPEC is initially solved with the complementarity constraint relaxed. If this problem yields a solution that is complementary, then it is also optimal. Otherwise there exists a complementarity that is violated and we can branch on this complementarity. Branching introduces two (or three in our case) child problems where the complementarity is broken. The procedure continues to solve relaxations and branch until an LPEC feasible solution is found, a problem is

infeasible, or its solution is dominated by an upper bound. This process is best envisioned as a tree search where nodes correspond to LP relaxations and edges correspond to branches.

Unfortunately, searching the entire branch-and-bound tree is likely to be inefficient. Instead, we will exploit the branch-and-cut methodology to construct the smallest tree that can be used to establish optimality of a given feasible solution. Specifically, let $(m_*, x^*, \lambda^*, \tau^*)$ be a feasible point of the LPEC (4) whose optimality we wish to prove. Note that m_* is an upper bound on the optimal value of the LPEC. We then perform several rounds of bound tightening to fix complementary expressions by solving LP relaxations. For instance, if we postulate that $\lambda_i^* = p_i/(1-\beta)$, then we solve the two remaining LP relaxations in the disjunction. If they produce bounds that are larger than the given upper bound m_*, then we can fix $\lambda_i^* = p_i/(1-\beta)$. Similar conclusions are possible for the other bounds and variables. This generates one hopes a short branch-and-bound tree. The numerical example presented next illustrates the idea.

5 Numerical Example

The numerical example has the following data: $n = 3$, $k = 27$, $\beta = 0.9$, $p_i = 1/27$ for all i,

$$X \equiv \left\{ x \in \Re^n_+ : \sum_{i=1}^n x_i = 1, \sum_{i=1}^n r_i x_i \geq f \right\},$$

where $r \equiv (-1/3, 2/3, -1)$ and $f = 1/10$. The vectors y^i are generated as follows. We generate three vectors

$$\begin{bmatrix} d^1 \\ d^2 \\ d^3 \end{bmatrix} \equiv \begin{bmatrix} 5 & 0 & -6 \\ 7 & 0 & -5 \\ 2 & 0 & -5 \end{bmatrix};$$

and then we set

$$y_1^j = d^1(i1), \quad y_2^j = d^2(i2), \quad y_3^j = d^3(i3),$$

where $j = 6(i1-1) + 3(i2-1) + i3$ for $i1, i2, i3 = 1, 2, 3$.

5.1 Finding an Upper Bound

We first solve the CVaR LP (2) and obtain the solution: CVaR $= 5.0644$, $m_{\text{CVaR}} = 4.8613$, $x^{\text{CVaR}} = (0.1097, 0.6161, 0.2742)$, $\tau_1 > 0$, $\tau_i = 0$ for all $i \geq 2$, $s_i = 0$ for $i = 1, 2, 10$, and $s_i > 0$ for all other i. The value m_{CVaR} is then verified to be the least element of $\mathcal{M}_\beta(x^{\text{CVaR}})$. Based on the pair $(m_{\text{CVaR}}, x^{\text{CVaR}})$, we solve the LP (5) by setting

$$\alpha_\tau^0 \cup \delta_\tau^0 = \{1, 2\}, \qquad \gamma_\tau^0 \cup (\beta_\tau^0 \setminus \delta_\tau^0) = \{1, \ldots, 27\} \setminus \alpha_\tau^0,$$

$$\alpha_\lambda^0 \cup \delta_\lambda^0 = \{1, 2, 10\}, \quad \gamma_\lambda^0 \cup (\beta_\lambda^0 \setminus \delta_\lambda^0) = \{1, \ldots, 27\} \setminus \alpha_\lambda^0.$$

The optimal solution for this LP is as follows: $m_1 = 4.2652$; τ_1 and τ_2 are both positive, and the remaining τ_i are zero; $s_i = 0$ for $i = 1, 2, 3, 10$, and the other s_i are all positive. This solution yields $\lambda_1 = \lambda_2$ at their common upper bound, which is 0.37037, implying that $\lambda_3 \geq 0$ and $\lambda_{10} \geq 0$ satisfy $\lambda_3 + \lambda_{10} = 1 - 2 * 0.37037 = .26926$.

The pair $(m_{\text{CVaR}}, x^{\text{CVaR}})$ also belongs to several other pieces of the feasible region of (4). For example, one such piece corresponds to the index set partitions:

$$\alpha_\tau^0 \cup \delta_\tau^0 = \{1, 10\}, \qquad \gamma_\tau^0 \cup (\beta_\tau^0 \setminus \delta_\tau^0) = \{1, \ldots, 27\} \setminus \alpha_\tau^0,$$

$$\alpha_\lambda^0 \cup \delta_\lambda^0 = \{1, 2, 10\}, \quad \gamma_\lambda^0 \cup (\beta_\lambda^0 \setminus \delta_\lambda^0) = \{1, \ldots, 27\} \setminus \alpha_\lambda^0;$$

nevertheless, solving the associated LPs (5) on these other pieces does not yield a lower objective value than 4.2652. The same upper bound $m_{\mathrm{UB}} = 4.2652$ is also obtained by solving the single equivalent NLP (7).

5.2 Branching to Verify Global Optimality

Our next task is to determine whether the value $m_{\mathrm{UB}} = 4.2652$ is globally optimal. First, we verify that $m_{\mathrm{VaR}} \geq 0$ by solving the LP

$$\begin{aligned}
\text{minimize} \quad & m \\
\text{subject to} \quad & x \equiv \sum_{j=1}^{k} z^j \in X, \\
& m = \sum_{j=1}^{k} \left[(z^j)^T y^j - \frac{p_j}{1-\beta} \tau_j \right] \\
& \left\{ \begin{array}{c} s_i \equiv m + \tau_i - x^T y^i \geq 0 \\ 0 \geq (z^i)^T y^i - \dfrac{p_i}{1-\beta} \tau_i \geq \dfrac{p_i}{1-\beta} m \\ \tau_i \geq 0, \quad 0 \leq z^i \leq \dfrac{p_i}{1-\beta} x \end{array} \right\}, \quad \forall\, i = 1, \ldots, k.
\end{aligned}$$

This LP is infeasible, and we can therefore conclude that $m_{\mathrm{VaR}} \geq 0$.

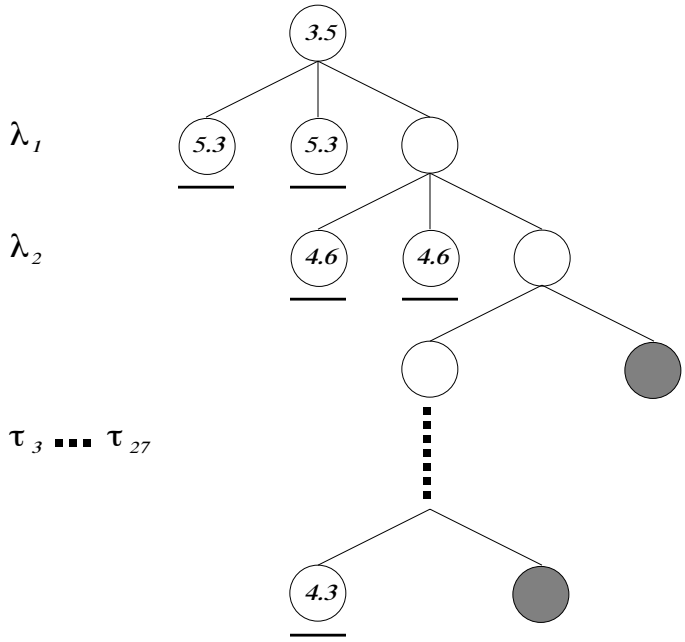

Figure 1: Branch-and-bound tree for the numerical example

Figure 1 shows the branch-and-bound tree that we construct for this example. Each node corresponds to an LP relaxation, with additional constraints included according to (13), (14), or (15). The root node shows the value of the LP relaxation (12) (all values are rounded to two digits). The variable names on the left indicate the branching variable. For each λ_i there are three branches

(which are ordered from left to write as $\lambda_i = 0$, $\lambda_i \in (0, p_i/(1-\beta))$, $\lambda_i =, p_i/(1-\beta))$. For each τ_i there are two branches (from left to right $\tau_i = 0$, and $\tau_i > 0$).

The lower bounds alone do not allow us to conclude optimality of the candidate solution. Hence, we start the construction of the branch-and-bound tree by proving that λ_1 and λ_2 must be at their upper bounds at an optimal solution. For this purpose, we solve two LPs by adding the cuts (13) and (15) to (12). These LPs have an optimal value of 5.3, which is larger than m_{UB}; we can therefore consider those nodes as fathomed. This is illustrated in the tree in Figure 1 by the bold horizontal lines under the node. Hence, we can fix λ_1 at its upper bound. Next, this process is repeated for λ_2, and we also find that λ_2 can be fixed at its upper bound.

Next, we consider proving that $\tau_3 = \ldots = \tau_{27} = 0$. First note that τ_i can be either zero or positive. If $\tau_i > 0$, then $\lambda_i = p_i/(1-\beta)$ is at its upper bound and $\tau_i = x^T y^i - m$ in (4). However, since $1 - \lambda_1 - \lambda_2 < p_i/(1-\beta)$, it follows that the LP corresponding to $\tau_i > 0$ must be inconsistent for all $i = 3, \ldots, 27$. This is represented in Figure 1 by the grey nodes. In practice, the preprocessor in AMPL detects that these LPs are inconsistent, and no solves are necessary. Hence, we conclude that $\tau_3 = \ldots = \tau_{27} = 0$. Solving the LP relaxation (12) with these τ_i fixed at zero and also $s_1 = s_2 = 0$ (since λ_1 and λ_2 are at their upper bounds), we obtain a lower bound of 4.2652, which means that our candidate solution is globally optimal.

All in all, we solved only six LPs to prove the global optimality of the candidate solution. The empty nodes are never solved, while the grey nodes can be eliminated by the preprocessor.

For this example, we have compared the two lower bounds (12) and (20) with $\beta = 0.8$ and $\beta = 0.9$. With $\beta = 0.8$, the LP relaxation gives a lower bound of $m = 0.61$, which is poorer than the convex hull relaxation, which gives $m = 1.24$. However, with $\beta = 0.9$, the LP relaxation (12) gives the tighter bound with $m = 3.48$, while (20) gives only $m = 2.45$. Hence, we use the LP relaxation (12) in the above report. (For this example, both lower bounds would actually generate identical trees except for the value at the root.) We have also used the same branch-and-cut procedure to verify global optimality in the problem with $\beta = 0.8$. Apart from the fact that the details are similar, with $\beta = 0.8$, the vector x^{CVaR} obtained by solving the CVaR minimization problem produced a VaR that is already globally optimal, as verified by the branch-and-cut procedure. In other words, the upper bounding refinement is not needed in the case where $\beta = 0.8$; for this reason, we omit the details. □

6 Approximation by Smoothing

In two pioneering papers [4, 5], Mangasarian and his then-Ph.D. student Chen developed a class of smoothing methods for solving complementarity problems. The basis of their methods is a family of smooth functions that approximate the plus function t_+. A summary of these functions can be found in [8, Subsection 11.8.2]. In what follows, we show how a smoothing approach can be applied to the VaR minimization problem.

Let ρ_ε be any nonnegative-valued, twice continuously differentiable, strictly convex function defined on the real line such that $|\rho_\varepsilon'(t)| \leq 1$ and $\rho_\varepsilon''(t) > 0$ for all t, and for some constant $c > 0$,

$$|\, t_+ - \rho_\varepsilon(t) \,| \;\leq\; c\,\varepsilon, \quad \forall\, t \;\in\; \Re, \tag{21}$$

for all $\varepsilon > 0$ sufficiently small. The latter approximating property has several consequences; among these, we have

$$\lim_{t \to \infty} \rho_\varepsilon(t) \;=\; \infty \quad \text{and} \quad \lim_{t \to -\infty} \rho_\varepsilon(t) \;=\; 0. \tag{22}$$

Two examples of such a smoothing function are

$$\rho_{\varepsilon,1}(t) \;\equiv\; \varepsilon \,\log(1 + e^{t/\varepsilon}) \quad \text{and} \quad \rho_{\varepsilon,2}(t) \;\equiv\; \frac{\sqrt{t^2 + 4\varepsilon^2} + t}{2}.$$

We leave it to the reader to verify that these functions satisfy the assumed properties.

In general, given a smoothing function ρ_ε with the cited properties, consider the ε-approximate CVaR and VaR minimization problems:

$$\left\{ \begin{array}{ll} \text{minimize} & \mathrm{CVaR}_{\beta,\varepsilon}(x) \\ \text{subject to} & x \in X \end{array} \right\} \quad \text{and} \quad \left\{ \begin{array}{ll} \text{minimize} & \mathrm{VaR}_{\beta,\varepsilon}(x) \\ \text{subject to} & x \in X \end{array} \right\},$$

where

$$\mathrm{CVaR}_{\beta,\varepsilon}(x) \equiv \min_{m \in \Re} \left[m + \frac{1}{1-\beta} \sum_{i=1}^{k} p_i \, \rho_\varepsilon(x^T y^i - m) \right],$$

and $\mathrm{VaR}_{\beta,\varepsilon}(x)$ is the unique minimizer in the definition of $\mathrm{CVaR}_{\beta,\varepsilon}(x)$. It follows from the assumed properties of ρ_ε that the minimand in $\mathrm{CVaR}_{\beta,\varepsilon}(x)$ is a coercive function of m, for fixed x, that is,

$$\lim_{m \to \pm\infty} \left[m + \frac{1}{1-\beta} \sum_{i=1}^{k} p_i \, \rho_\varepsilon(x^T y^i - m) \right] = \infty.$$

(Use the second limit in (22) to show the case $m \to \infty$ and the first limit and the nonexpansiveness of ρ_ε, which in turn is implied by the condition $|\rho_\varepsilon'(t)| \leq 1$ for all t, to show the other case $m \to -\infty$.) Therefore, $\mathrm{CVaR}_{\beta,\varepsilon}(x)$ is a well-defined, finite scalar. Moreover, by the strict convexity of ρ_ε, it follows that $\mathrm{VaR}_{\beta,\varepsilon}(x)$ is uniquely defined. The latter is a significant difference from the original $\mathrm{CVaR}_\beta(x)$ where the minimizing set $\mathcal{M}_\beta(x)$ is often not a singleton. Another important consequence with using the smooth function ρ_ε is that $\mathrm{VaR}_{\beta,\varepsilon}(x)$ can be characterized as the unique scalar m that satisfies the smooth equation:

$$1 - \beta = \sum_{i=1}^{k} p_i \, \rho_\varepsilon'(x^T y^i - m).$$

This implies, by the implicit-function theorem, that the $\mathrm{VaR}_{\beta,\varepsilon}(x)$ is a continuously differentiable function of x with gradient given by

$$\nabla \mathrm{VaR}_{\beta,\varepsilon}(x) = \frac{\displaystyle\sum_{i=1}^{k} p_i \, \rho_\varepsilon''(x^T y^i - \mathrm{VaR}_{\beta,\varepsilon}(x)) \, y^i}{\displaystyle\sum_{i=1}^{k} p_i \, \rho_\varepsilon''(x^T y^i - \mathrm{VaR}_{\beta,\varepsilon}(x))}.$$

The upshot of these properties is that $\mathrm{VaR}_{\beta,\varepsilon}(x)$ has much nicer analytic properties than $\mathrm{VaR}_\beta(x)$; furthermore, the ε-approximate VaR minimization problem is a smooth, albeit still nonconvex, linearly constrained nonlinear program in the sole variable x. As such, there are a host of efficient algorithms that one can use for computing the minimum value (to be precise, stationary values) of the ε-approximate value-at-risk.

An important question that arises is what happens to the convergence of the ε-approximation problems as $\varepsilon \downarrow 0$. Although such a question has been partially studied in a general context (see, e.g., [17]), we give a self-contained treatment to such a convergence issue for our special problem. For this purpose, we establish a preliminary boundedness lemma.

Lemma 6.1 Let $\{\varepsilon_\nu\}$ be a sequence of sufficiently small positive scalars, and let $\{x^\nu\}$ be an arbitrary sequence of vectors in X, both of which are necessarily bounded. The sequence $\{m_\nu\}$, where $m_\nu \equiv \mathrm{VaR}_{\beta,\varepsilon_\nu}(x^\nu)$ for every ν, is bounded. Moreover, if the pair (m_∞, x^∞) is the limit of a convergent subsequence $\{(m_\nu, x^\nu) : \nu \in \kappa\}$ corresponding to a sequence $\{\varepsilon_\nu\}$ of positive scalars tending to zero, then m_∞ is an element of $\mathcal{M}_\beta(x^\infty)$.

Proof. We have, for any $m \in \Re$ and any ν,

$$m_\nu + \frac{1}{1-\beta}\sum_{j=1}^{k} p_j\, \rho_{\varepsilon_\nu}((x^\nu)^T y^j - m_\nu) \;\le\; m + \frac{1}{1-\beta}\sum_{j=1}^{k} p_j\, \rho_{\varepsilon_\nu}((x^\nu)^T y^j - m),$$

which implies, by the uniform approximation property (21),

$$m_\nu + \frac{1}{1-\beta}\sum_{j=1}^{k} p_j\, ((x^\nu)^T y^j - m_\nu)_+ \;\le\; m + \frac{1}{1-\beta}\sum_{j=1}^{k} p_j\, ((x^\nu)^T y^j - m)_+ + \frac{2\,c\,\varepsilon_\nu}{1-\beta}.$$

Since the right-hand side is bounded, the boundedness of the sequence $\{m_\nu\}$ follows readily. Moreover, the second assertion of the lemma also follows easily from the last inequality. $\square$

We introduce the notation for our next result, which addresses the main convergence issue of the ε-smoothing procedure. Specifically, let $\{\varepsilon_\nu\}$ be an arbitrary sequence of positive scalars converging to zero. For each $\varepsilon > 0$, let x^ε be a (globally) optimal solution of the smooth optimization problem:

$$\begin{array}{ll} \text{minimize} & \mathrm{VaR}_{\beta,\varepsilon}(x) \\ \text{subject to} & x \in X, \end{array} \tag{23}$$

which is clearly equivalent to

$$\begin{array}{ll} \text{minimize} & m \\ \text{subject to} & x \in X \\ \text{and} & 1-\beta = \displaystyle\sum_{i=1}^{k} p_i\, \rho'_\varepsilon(x^T y^i - m). \end{array}$$

Let $m_\varepsilon \equiv \mathrm{VaR}_{\beta,\varepsilon}(x^\varepsilon)$.

Proposition 6.2 Suppose that the VaR minimization problem has a minimizer x^{VaR} such that $\mathcal{M}_\beta(x^{\mathrm{VaR}})$ is the singleton $\{m_{\mathrm{VaR}}\}$. It holds that

$$\lim_{\varepsilon \downarrow 0} m_\varepsilon = \min_{x \in X} \mathrm{VaR}_\beta(x). \tag{24}$$

Consequently, for any sequence of positive scalars $\{\varepsilon_\nu\}$ converging to zero, every accumulation point of $\{x^{\varepsilon_\nu}\}$ is a minimizer of $\mathrm{VaR}_\beta(x)$ on X.

Proof. Let $\{\varepsilon_\nu\}$ be an arbitrary sequence of positive scalars tending to zero. We have, for every ν,

$$m_{\varepsilon_\nu} \;\le\; \mathrm{VaR}_{\beta,\varepsilon_\nu}(x^{\mathrm{VaR}}).$$

By Lemma 6.1, every accumulation point of the right-hand side is an element of $\mathcal{M}_\beta(x^{\mathrm{VaR}})$, which by assumption is the singleton $\{m_{\mathrm{VaR}}\}$. Consequently, the right-hand side converges to m_{VaR} as $\varepsilon_\nu \downarrow 0$. If (m_∞, x^∞) is the limit of a convergent subsequence $\{(m_{\varepsilon_\nu}, x^{\varepsilon_\nu}) : \nu \in \kappa\}$, then $m_\infty \in \mathcal{M}_\beta(x^\infty)$, and, from the above inequality,

$$m_\infty \;\le\; m_{\mathrm{VaR}} = \mathrm{VaR}_\beta(x^{\mathrm{VaR}});$$

this shows that x^∞ is a minimizer of $\mathrm{VaR}_\beta(x)$ on X and also establishes the desired limit (24). $\square$

Proposition 6.2 is theoretically very desirable; its practical drawback is that there is no guarantee that a globally optimal solution to (23) can be computed. We have implemented the two smoothing

functions $\rho_{\varepsilon,1}$ and $\rho_{\varepsilon,2}$ using $\varepsilon = 10^{-3}$ for the example of the previous section. To solve the smoothed problem, we used five state-of-the-art NLP solvers: filter, knitro 3.0, loqo 6.06, minos 5.5, and snopt 6.6-1, which are all available on the NEOS server [6, 26]. The MPEC solver required 5 iterations to produce an upper bound for the example on hand, which turns out to be globally optimal. In contrast, the NLP solvers fail for $\rho_{\varepsilon,1}$ because the exponentials cannot be evaluated or blow up during the computation. The situation is slightly better for the square root formulation. Filter produces an optimal solution $m = 4.265827$ (which is slightly higher than the minimum VaR of 4.2652); knitro produces a local optimum $m = 4.74277$. All other solvers fail to produce a feasible point (minos and snopt), while loqo fails because it reached its iteration limit.

7 Conclusion

In this paper, we have investigated the minimization problem of the VaR as a nonconvex LPEC and developed bounding schemes that can be used to verify the global optimality of a candidate feasible solution. We have also established the convergence of a smoothing approach to compute an approximate VaR. Whereas the VaR minimization problem is special (and yet important in its own right), we maintain that the bounding schemes can be extended to more general LPECs, and possibly even to other "convex" MPECs, namely, MPECs whose only nonconvexity is the complementarity constraint. Indeed, the extension of the upper bounding scheme is fairly straightforward; it is the lower bounding scheme that is very much problem dependent. Nevertheless, we believe that for special classes of MPECs, tight lower bounds can be obtained, which can then be used in a branch-and-cut scheme either for verifying the global optimality of a candidate solution obtained from a local MPEC solver or for computing a globally optimal solution to the problem directly.

A lesson we have learned from the computational experiments in this paper is that while the NLP solvers are generally very robust, one still requires a proof such as the one given in Subsection 5.2 to ascertain the quality of the solutions they produce. For MPECs, we believe that the time is now ripe for combining existing local methods with some global branch-and-cut schemes in order to obtain solutions with proven global optimality.

Acknowledgment. Professor Stanislav Uryasev has kindly alerted us to his paper with Larsen and Mausser [19] in which CVaR-based algorithms are proposed for solving the VaR-minimization problem.

References

[1] F. Andersson, H. Mausser, D. Rosen, and S. Uryasev. Credit risk optimization with conditional value-at-risk criterion. *Mathematical Programming, Series B* 89 (2001) 273–291.

[2] M. Anitescu. On solving mathematical programs with complementarity constraints as nonlinear programs. Mathematics and Computer Science Division, Argonne National Laboratory Preprint ANL/MCS-P864-1200 (2000).

[3] J. F. Bard. Convex two-level optimization. *Mathematical Programming* 40 (1988) 15–27.

[4] C. H. Chen and O. L. Mangasarian. Smoothing methods for convex inequalities and linear complementarity problems. *Mathematical Programming* 71 (1995) 51–69.

[5] C. H. Chen and O. L. Mangasarian. A class of smoothing functions for nonlinear and mixed complementarity problems. *Computational Optimization and Applications* 5 (1996) 97–138.

[6] J. Czyzyk, M. Mesnier, and J. Moré. The NEOS Server. *IEEE Journal on Computational Science and Engineering* 5 (1998) 68–75.

[7] F. Facchinei, H. Jiang, and L. Qi. A smoothing method for mathematical programs with equilibrium constraints. *Mathematial Programming* 85 (1999) 107–134.

[8] F. Facchinei and J. S. Pang. *Finite-Dimensional Variational Inequalities and Complementarity Problems*. Springer-Verlag (New York 2003).

[9] R. Fletcher and S. Leyffer. Numerical experience with solving MPECs as NLPs. University of Dundee Report NA210 (August 2002), to appear in *Optimization Methods and Software*.

[10] R. Fletcher, S. Leyffer, D. Ralph, and S. Scholtes. Local convergence of SQP methods for mathematical programs with equilibrium constraints. University of Dundee Report NA209 (May 2002).

[11] M. Fukushima, Z. Q. Luo, and J. S. Pang. A globally convergent sequential quadratic programming algorithm for mathematical programs with linear complementarity constraints. *Computational Optimization and Applications* 10 (1998) 5–34.

[12] M. Fukushima and J. S. Pang. Some feasibility issues in mathematical programs with equilibrium constraints. *SIAM Journal on Optimization* 8 (1998) 673–681.

[13] M. Fukushima and J. S. Pang. Convergence of a smoothing continuation method for mathematical programs with complementarity constraints. In Michel Théra and Rainer Tichatschke, editors. *Ill-posed variational problems and regularization techniques (Trier 1998)* [Lecture Notes in Economics and Mathematical Systems 477], Springer (Berlin 1999) pp. 99–110.

[14] M. Fukushima and P. Tseng. An implementable active-set algorithm for computing a B-stationary point of the mathematical program with linear complementarity constraints. *SIAM Journal on Optimization* 12 (2002) 724–739.

[15] H. Jiang and D. Ralph. Smooth SQP methods for mathematical programs with nonlinear complementarity constraints. *SIAM Journal on Optimization* 10 (2000) 779–808.

[16] H. Jiang and D. Ralph. Extension of quasi-Newton methods to mathematical programs with complementarity constraints. *Computational Optimization and Applications* 25 (2003) 123–150.

[17] P. Kall. Approximation to optimization problems: an elementary review. *Mathematics of Operations Research* 11 (1986) 9–18.

[18] P. Krokhmal, J. Palmquist, and S. Uryasev. Portfolio optimization with conditional value-at-risk objective and constraints. *Journal of Risk* 4 (2002)

[19] N. Larsen, H. Mausser, and S. Uryasev. Algorithms for optimization of value-at-risk. In P. Pardalos and V.K. Tsitsiringos, editors, *Financial Engineering, E-Commerce and Supply Chain*. Kluwer Academic Publishers (Norwell 2002) pp. 129–157.

[20] S. Leyffer. The penalty interior point method fails to converge for mathematical programs with equilibrium constraints. University of Dundee Report NA208 (February 2002).

[21] X. Liu and J. Sun. Generalized stationary points and an interior point method for mathematical programs with equilibrium constraints. School of Business Preprint, Department of Decision Sciences, University of Singapore (2002).

[22] Z. Q. Luo, J. S. Pang, and D. Ralph. *Mathematical Program with Equilibrium Constraints*. Cambrige University Press (Cambridge 1996).

[23] O. L. Mangasarian. *Nonlinear Programming*, McGraw-Hill (New York 1969). [Reprinted as SIAM Classics in Applied Mathematics 10 (Philadelphia 1994).]

[24] O. L. Mangasarian. Regularized linear programs with equilibrium constraints. In M. Fukushima and L. Qi, editors. *Reformulation–Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods*. Kluwer Academic Publishers (Dordrecht 1998) pp. 259–268.

[25] O. L. Mangasarian and S. Fromovitz. The Fritz John necessary optimality conditions in the presence of equality constraints. *Journal of Mathematical Analysis and Applications* 17 (1967) 34–47.

[26] The NEOS server for optimization. `http://www-neos.mcs.anl.gov/neos/` Argonne National Laboratory Webpage, 2003

[27] W. Ogryczak and A. Ruszczynski. Dual stochastic dominance and related mean-risk models. *SIAM Journal on Optimization* 13 (2002) 60–78.

[28] J. Outrata, M. Kocvara, and J. Zowe. *Nonsmooth Approach to Optimization Problems with Equilibrium Constraints: Theory, Applications and Numerical Results.* Kluwer Academic Publishers (Dordrecht 1998).

[29] G. C. Pflug. Some remarks on the value-at-risk and the conditional value-at-risk. In S. Uryasev, editor, *Probabilistic Constrained Optimization: Methodology and Optimization*, Kluwer Academic Publishers (Norwell 2000) pp. 278–287.

[30] A. Raghunathan and L. T. Biegler. MPEC formulations and algorithms in process engineering. Technical Report, Department of Chemical Engineering, Carnegie Mellon University, Pittsburgh (2002).

[31] A. Raghunathan and L. T. Biegler. Barrier methods for mathematical programs with complementarity constraints (MPCCs). Technical Report, Department of Chemical Engineering, Carnegie Mellon University, Pittsburgh (2002).

[32] R. T. Rockafellar and S. Uryasev. Optimization of conditional value-at-risk. *Journal of Risk* 2 (2000) 21–41.

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

[34] R. T. Rockafellar, S. Uryasev, and M. Zabarankin. Deviation measures in risk analysis and optimization. Research Report #2002-7, Department of Industrial and Systems Engineering, University of Florida, Gainesville (2002).

[35] R. T. Rockafellar, S. Uryasev, and M. Zabarankin. Portfolio analysis with general deviation measures. Research Report #2003-8, Department of Industrial and Systems Engineering, University of Florida, Gainesville (2003).

[36] H. Scheel and S. Scholtes. Mathematical programs with complementarity constraints: Stationarity, optimality, and sensitivity. *Mathematics of Operations Research* 25 (2000) 1–22.

[37] S. Scholtes and M. Stöhr. Exact penalization of mathematical programs with equilibrium constraints. *SIAM Journal on Control and Optimization* 37 (1999) 617–652.

[38] S. Scholtes. Convergence properties of a regularization scheme for mathematical programs with complementarity constraints. *SIAM Journal on Optimization* 11 (2001) 918–936.

[39] M. Tawarmalani and N. V. Sahinidis. *Convexification and Global Optimization in Continuous and Mixed-Integer Nonlinear Programming: Theory, Algorithms, Software, and Applications.* Kluwer Academic Publishers (Boston, 2002).

[40] C. E. Testsuri and S. Uryasev. On relation between expected regret and conditional at-risk. Manuscript, Department of Industrial and Systems Engineering, University of Florida, Gainesville (2002).

The submitted manuscript has been created in part by the University of Chicago as Operator of Argonne National Laboratory ("Argonne") under Contract No. W-31-109-ENG-38 with the U.S. Department of Energy. The U.S. Government retains for itself, and others acting on its behalf, a paid-up, nonexclusive, irrevocable worldwide license in said article to reproduce, prepare derivative works, distribute copies to the public, and perform publicly and display publicly, by or on behalf of the Government.