

Conditional Value-at-Risk in Stochastic Programs with Mixed-Integer Recourse

April 26, 2004

Abstract. In classical two-stage stochastic programming the expected value of the total costs is minimized. Recently, mean-risk models - studied in mathematical finance for several decades - have attracted attention in stochastic programming. We consider Conditional Value-at-Risk as risk measure in the framework of two-stage stochastic integer programming. The paper addresses structure, stability, and algorithms for this class of models. In particular, we study continuity properties of the objective function, both with respect to the first-stage decisions and the integrating probability measure. Further, we present an explicit mixed-integer linear programming formulation of the problem when the probability distribution is discrete and finite. Finally, a solution algorithm based on Lagrangean relaxation of nonanticipativity is proposed.

Key words. Stochastic programming – mean-risk models – mixed-integer optimization – conditional value-at-risk

1. Introduction

Most real-world problems in operations research involve uncertain data. Thus, finding optimal decisions turns into selecting a “best” random variable. Then the question comes up which criteria to use for the selection. Quickly, the matter of risk aversion becomes an issue. Further, for realistic modeling integer variables are often helpful and sometimes inevitable. This paper suggests a way of how to make such decisions by using Conditional Value-at-Risk as risk measure in the framework of recourse stochastic integer programming.

Throughout the paper, we impose a cost minimization framework. Consider a family of real random cost variables $\{Z(x, \omega)\}_{x \in X} \subseteq \mathcal{Z}$ on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. We want to decide on the variable x and so on the corresponding random variable by comparing certain scalar characteristics of the random variables, namely by so-called mean-risk models

$$\min_{x \in X} \mathbb{E}\{Z(x, \omega)\} + \rho \mathcal{R}\{Z(x, \omega)\} \quad (1)$$

Rüdiger Schultz: Institute of Mathematics, University Duisburg-Essen, Lotharstr. 65, D-47048 Duisburg, Germany, e-mail: schultz@math.uni-duisburg.de

Stephan Tiedemann: Institute of Mathematics, University Duisburg-Essen, Lotharstr. 65, D-47048 Duisburg, Germany, e-mail: tiedemann@math.uni-duisburg.de

Mathematics Subject Classification: 90C15, 90C11, 90C06

where $\mathbb{E} : \mathcal{Z} \rightarrow \mathbb{R}$ denotes the expected value, $\mathcal{R} : \mathcal{Z} \rightarrow \mathbb{R}$ a risk measure, \mathcal{Z} the space of all real random cost variables $Z : \Omega \rightarrow \mathbb{R}$ satisfying $\mathbb{E}\{|Z(\omega)|\} < \infty$, and $\rho > 0$ a suitable weight factor.

This paper deals with a risk measure called Conditional Value-at-Risk applied to random variables occurring in recourse stochastic integer programming. These random variables are essentially defined by value functions of mixed-integer linear programs being discontinuous and nonconvex, such that, in particular, convexity of the objectives in the mean-risk model is not given. Therefore, for the sake of applicability to real-world problems, it is inevitable to choose a risk measure such that, despite the poor properties of the random variables, the resulting stochastic integer programs are structurally sound and amenable to algorithmic treatment. In the following we will show that Conditional Value-at-Risk is a risk measure that satisfies these requirements.

The mean-risk model (1) aims at minimizing the weighted sum of two competing objectives. Viewed from a more general perspective, it is a scalarization of the multiobjective optimization problem

$$\min_{x \in X} (\mathbb{E}\{Z(x, \omega)\}, \mathcal{R}\{Z(x, \omega)\}). \quad (2)$$

For an introduction to multiobjective optimization we refer to [7,20]. An accepted notion of optimality in multiobjective optimization is efficiency. A point $\bar{x} \in X$ is called efficient for (2) if there is no other point $x \in X$ such that $\mathbb{E}\{Z(x, \omega)\} \leq \mathbb{E}\{Z(\bar{x}, \omega)\}$ and $\mathcal{R}\{Z(x, \omega)\} \leq \mathcal{R}\{Z(\bar{x}, \omega)\}$, with at least one strict inequality. The set of all efficient points is named efficient frontier. Every optimal solution to the mean-risk model (1) with a weight factor $\rho > 0$ is an efficient point, a so-called supported efficient point. Due to the lacking convexity of our objective functions, not all efficient points are supported, and thus cannot be computed by solving scalarizations (1). However, solving the mean-risk model (1) for various values of $\rho > 0$ has the capability to trace the supported part of the efficient frontier. In Section 5 a discrete tracing method is described and carried out for a real-life optimization problem.

The relation of stochastic dominance, one of the fundamental concepts in decision theory, introduces a partial order in the space of real random variables. This provides a basis for selecting “best” members from families of random variables. Ogryczak and Ruszczyński have studied mean-risk models and their consistency with the multiobjective criteria induced by stochastic dominance, see [23,24]. Artzner et al., cf. [4], proposed axiomatic properties of risk measures, calling risk measures fulfilling these properties coherent. Conditional Value-at-Risk is a coherent risk measure, cf. [2,4,30], and the corresponding mean-risk model is consistent with the second degree stochastic dominance relation, cf. [24].

For controlling the variability of the costs, the first risk measure that comes into mind is the variance, cf. [21]. However, the variance has several drawbacks. In general, it is neither consistent with the stochastic dominance relation nor coherent. Further, the square and the mentioned discontinuities may lead to an

objective which is not even lower semicontinuous, such that (1), with compact X , may have a finite infimum which is not attained, cf. [33].

In [33] another risk measure, the Excess Probability, see Section 2, has been analyzed. It is consistent with the first degree stochastic dominance relation. The present paper shall be seen as a continuation of that research, analyzing a risk measure with a higher degree of consistency with the stochastic dominance relation.

Section 2 gives an introduction to Conditional Value-at-Risk. In Section 3 we extend the traditional modeling in two-stage stochastic integer programming towards risk aversion and formulate a mean-risk model with Conditional Value-at-Risk. Section 4 analyzes structure and stability of the risk measure on the space of the special random variables in question, defined essentially by value-functions of mixed-integer programs with random right-hand sides. An explicit mixed-integer linear programming formulation of the model when the probability distribution is discrete and finite is presented in Section 5. Moreover, a solution algorithm is proposed, and finally, we report on some first numerical experiments.

2. Conditional Value-at-Risk (CVaR)

In this paper we want to minimize risk in the following sense: Minimize

$$\text{“the expected value of the costs in the } (1 - \alpha) \cdot 100\% \text{ worst cases”}, \quad (3)$$

where $\alpha \in (0, 1)$ is a preselected probability.

We call the value described in (3) α -Conditional Value-at-Risk (α -CVaR). Since we do not want to impose any additional assumptions on the random variables $Z(x, \omega) \in \mathcal{Z}$ or the probability space $(\Omega, \mathcal{A}, \mathbb{P})$, the precise mathematical definition requires some care. We follow the paper of Rockafellar and Uryasev ([30]):

Definition 2.1. For $Z(x, \omega) \in \mathcal{Z}$ and a preselected probability $\alpha \in (0, 1)$ we define the following:

(i) The distribution function

$$\Psi(x, \eta) := \mathbb{P}(\{\omega \in \Omega : Z(x, \omega) \leq \eta\}).$$

(ii) The α -Value-at-Risk (α -VaR)

$$\eta_\alpha(x) := \min\{\eta : \Psi(x, \eta) \geq \alpha\}. \quad (4)$$

(iii) The α -Value-at-Risk⁺ (α -VaR⁺)

$$\eta_\alpha^+(x) := \inf\{\eta : \Psi(x, \eta) > \alpha\}.$$

(iv) *The α -CVaR*

$$\phi_\alpha(x) := \text{mean of the } \alpha\text{-tail distribution of } Z(x, \omega), \quad (5)$$

where the distribution in question is the one with the distribution function $\Psi_\alpha(x, \eta)$ defined by

$$\Psi_\alpha(x, \eta) := \begin{cases} 0 & \text{for } \eta < \eta_\alpha(x), \\ [\Psi(x, \eta) - \alpha]/[1 - \alpha] & \text{for } \eta \geq \eta_\alpha(x). \end{cases} \quad (6)$$

We verify that $\Psi_\alpha(x, \cdot)$ actually is another distribution function: It is the α -tail or rather the upper $(1 - \alpha)$ -part of $\Psi(x, \cdot)$ rescaled onto $[0, 1]$, and therefore, it is nondecreasing and right-continuous, with $\lim_{\eta \rightarrow \infty} \Psi_\alpha(x, \eta) = 1$ and $\lim_{\eta \rightarrow -\infty} \Psi_\alpha(x, \eta) = 0$.

The minimum in (4) is always attained, since the distribution function $\Psi(x, \cdot)$ is nondecreasing and right-continuous. It is immediate that always $\eta_\alpha(x) \leq \eta_\alpha^+(x)$. These values are equal unless $\Psi(x, \cdot)$ is constant at α over a certain interval. When $\Psi(x, \cdot)$ is continuous and strictly increasing, $\eta_\alpha(x) = \eta_\alpha^+(x)$ is simply the unique η satisfying $\Psi(x, \eta) = \alpha$. Otherwise, it is possible that this equation has no solution or a whole range of solutions. In the former situation $\Psi(x, \cdot)$ has a probability atom at $\eta_\alpha(x)$, while in the latter, the graph of $\Psi(x, \cdot)$ has a constant segment at $\Psi(x, \eta) = \alpha$ being either the interval $[\eta_\alpha(x), \eta_\alpha^+(x))$ or $[\eta_\alpha(x), \eta_\alpha^+(x)]$, depending on whether or not $\Psi(x, \cdot)$ has a jump at $\eta_\alpha^+(x)$.

If there is no probability atom at $\eta_\alpha(x)$ and so $\Psi(x, \eta_\alpha(x)) = \alpha$, the α -CVaR (3) is equal to the conditional expectation

$$\mathbb{E}\{Z(x, \omega) \mid Z(x, \omega) \geq \eta_\alpha(x)\}, \quad (7)$$

since the α -VaR $\eta_\alpha(x)$ equals “the minimum potential costs that can occur in the $(1 - \alpha) \cdot 100\%$ worst cases”.

Note, that (7) is the usual definition of α -CVaR for continuous distribution functions (having no probability atoms at all) as it then coincides with (5) and so with (3), cf. [25, 29, 30].

But, if there is no η such that $\Psi(x, \eta) = \alpha$ and so there is a probability atom at $\eta_\alpha(x)$, which in particular may occur for discretely distributed random variables, (7) does not coincide with (3) such that then the correct definition of α -CVaR is (5). For a rigorous proof of this, including graphical examples we refer the reader to [30]. The problem is, when using (7) for the general case, one is not taking the expectation of the upper $(1 - \alpha)$ -part of the full distribution, since the probability atom at the α -VaR must be split to do so, but can not be done by taking any conditional expectation. Thus one has to make the trick as in (6): taking the correct part of the original distribution function and rescale it onto $[0, 1]$.

Another correct formalization of (3) for the general case has been worked out in [1, 2], where the α -CVaR is expressed as a difference of an expectation and a correcting exceeding part if there is a probability atom at the α -VaR. The authors also discuss the confusion that inheres the current publications

on this subject due to the latter described problems. In particular the authors mention that the name Conditional Value-at-Risk stems from the time where in general the continuity of the distribution function was assumed and thus the conditional expectation (7) was the correct definition. However, they show that in the general case there is no way to express (3) as a conditional expectation and thus decline the term Conditional Value-at-Risk. They suggest the name α -Expected Shortfall for the gain maximization framework.

In [24] the α -CVaR was defined for general distribution functions by means of the second quantile function being the convex conjugate function of the distribution function of order two. The authors call the risk measure Tail Value-at-Risk. As already mentioned, they also show that the corresponding mean-risk model is consistent with the second degree stochastic dominance relation.

As we will see in the next Proposition the risk measure η -Expected Excess

$$Q_{\mathbb{E}_\eta}(x) = \mathbb{E}\{\max\{Z(x, \omega) - \eta, 0\}\}$$

is closely related to the α -CVaR, and therefore will also be analyzed intensively in this paper. The corresponding “partner” of the α -VaR, the η -Excess Probability,

$$Q_{\mathbb{P}_\eta}(x) = \mathbb{P}(\{\omega \in \Omega : Z(x, \omega) > \eta\})$$

has been treated in [33].

Before we come to the next Section and analyze the α -CVaR for our special random variables we quote for later reference the following fundamental result, cf. [2, 24, 25, 30].

Proposition 2.2. *For $Z(x, \omega) \in \mathcal{Z}$, the α -CVaR can be expressed by the following minimization formula:*

$$\phi_\alpha(x) = \min_{\eta \in \mathbb{R}} f(\alpha, \eta, x), \quad (8)$$

where

$$f(\alpha, \eta, x) := \eta + \frac{1}{1 - \alpha} \mathbb{E}\{\max\{Z(x, \omega) - \eta, 0\}\}.$$

Further, f is convex in η and finite (hence continuous) and the optimal set of (8) is the nonempty closed interval $[\eta_\alpha(x), \eta_\alpha^+(x)]$, reducing to $\eta_\alpha(x)$ when the graph of $\Psi(x, \cdot)$ has no constant segment at $\Psi(x, \eta) = \alpha$. In particular, the α -VaR $\eta_\alpha(x)$ always is a minimizer, and thus $\phi_\alpha(x) \in \mathbb{R}$.

Note, that (8) is sometimes used as the definition of α -CVaR, cf. [2, 25].

Rockafellar and Uryasev ([29, 30]) have analyzed α -CVaR for random variables $Z(x, \omega) \in \mathcal{Z}$ that are convex in x , yielding convexity of f jointly in (η, x) and so the convexity of α -CVaR. In that case the minimization of α -CVaR is a convex optimization problem and therefore algorithmic treatment is possible by various techniques from convex analysis.

In the remaining part of the paper we will discuss α -CVaR for the discontinuous and nonconvex random variables occurring in two-stage stochastic integer programs. Then the convexity of α -CVaR is lost. However, we will show that it still has certain continuity properties both with respect to the first-stage decisions and the integrating probability measure.

3. Two-Stage Stochastic Integer Programs with CVaR

Consider the following random mixed-integer linear program

$$\begin{aligned} \min_{x,y,y'} \{ & c^\top x + q^\top y + q'^\top y' : Tx + Wy + W'y' = h(\omega), \\ & x \in X, y \in \mathbb{Z}_+^{\bar{m}}, y' \in \mathbb{R}_+^{m'} \}. \end{aligned} \quad (9)$$

Assume that all ingredients in (9) have conformable dimensions, that W, W' are rational matrices, and that $X \subseteq \mathbb{R}^m$ is a nonempty closed set, possibly with integer requirements to some components of x . The right-hand side $h(\omega) \in \mathbb{R}^s$ of the equality constraints is a random vector on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Together with (9) we have the so-called nonanticipativity information constraint, which forces the variables x to be fixed before observing the outcome of $h(\omega)$, and allows the variables (y, y') to be fixed afterwards. Therefore, x and (y, y') are called first- and second-stage variables, respectively.

The mixed-integer value function

$$\Phi(t) := \min\{q^\top y + q'^\top y' : Wy + W'y' = t, y \in \mathbb{Z}_+^{\bar{m}}, y' \in \mathbb{R}_+^{m'}\} \quad (10)$$

is a substantial object in our succeeding stochastic programming models. By integer programming theory ([22]), this function is real-valued on \mathbb{R}^s if $W(\mathbb{Z}_+^{\bar{m}}) + W'(\mathbb{R}_+^{m'}) = \mathbb{R}^s$ and $\{u \in \mathbb{R}^s : W^\top u \leq q, W'^\top u \leq q'\} \neq \emptyset$ which, therefore, will be assumed throughout the paper.

With

$$Q_{\mathbb{E}}(x) := \mathbb{E}\{c^\top x + \Phi(h(\omega) - Tx)\}$$

the traditional expectation-based stochastic program with recourse is the optimization problem

$$\min\{Q_{\mathbb{E}}(x) : x \in X\}. \quad (11)$$

Introducing the Conditional Value-at-Risk functional

$$Q_{CVaR_\alpha}(x) := \min_{\eta \in \mathbb{R}} f(\alpha, \eta, x) \quad (12)$$

with

$$f(\alpha, \eta, x) := \eta + \frac{1}{1-\alpha} \mathbb{E}\{\max\{c^\top x + \Phi(h(\omega) - Tx) - \eta, 0\}\},$$

problem (11) is extended into the mean-risk model

$$\min\{Q_{\mathbb{E}}(x) + \rho Q_{CVaR_\alpha}(x) : x \in X\}. \quad (13)$$

Here $\alpha \in (0, 1)$ denotes some preselected probability and $\rho > 0$ is a suitable weight factor.

The principal aim of the above construction is to choose the first-stage decisions x in an optimal way without anticipation of future outcomes of the random vector $h(\omega)$. More specifically, having decided on x and observed $h(\omega)$, the

remaining decisions (y, y') have to be taken optimal. This leads to the mixed-integer linear program determining the function Φ in (10). The costs of the two-stage sequential process of decision and observation are expressed by the random variable $Z(x, \omega) = c^\top x + \Phi(h(\omega) - Tx)$. The mean-risk model (13) is then used to find a “best” first-stage decision $x \in X$, and thus also a “best” random variable from the indexed family $(c^\top x + \Phi(h(\omega) - Tx))_{x \in X}$.

Note that it is necessary to have $\mathbb{E}\{|c^\top x + \Phi(h(\omega) - Tx)|\} < \infty$ for all $x \in X$ to apply Proposition 2.2 such that then the definition of Q_{CVaR_α} in (12) is justified. This is the case under reasonable assumptions as will be shown in the next Section.

If no second-stage variables are restricted to be integers, Φ is the value function of a linear program and thus convex. Then, also the objective in (13) is convex and therefore allows for application of various analytical and algorithmic techniques from convex analysis. With integers in the second stage, Φ is the value function of a mixed-integer linear program, so in general discontinuous and nonconvex, and therefore the random variables $Z(x, \omega) = c^\top x + \Phi(h(\omega) - Tx)$ are discontinuous and nonconvex in x , yielding that the objective function in (13) is not convex anymore. However, in the next Section we will show that (13) is still well-posed from formal viewpoint. We will establish structural properties of the functional Q_{CVaR_α} , and finally in Section 5 we will demonstrate that solution methodology from mixed-integer linear programming (the class our initial random optimization problem (9) belongs to) can be employed for solving (13).

The traditional expectation-based optimization problem (11) is one of the well-studied objects in stochastic programming, see [3, 12, 31, 32]. Therefore, the main focus in the following analysis will be on the functional Q_{CVaR_α} . We briefly mention that, under mild conditions, $Q_{\mathbb{E}}$ is real-valued and lower semicontinuous and that $Q_{\mathbb{E}}$ is continuous if the distribution of $h(\omega)$ has a density. Optimal values and optimal solutions to (11) behave stable under perturbations of the probability distribution of $h(\omega)$. This allows for discrete approximations of the probability distribution for which (11) can be rewritten equivalently as a block-structured mixed-integer linear program. The latter is amenable to decomposition methods splitting (11) into smaller mixed-integer linear programs that are often solvable by standard software like CPLEX, [15].

4. Structure and Stability

Before we start with the structural analysis of Q_{CVaR_α} let us fix our notions of lower and upper semicontinuity. Consider a space Ξ endowed with some notion of convergence. Recall that a function $\mathcal{G} : \Xi \rightarrow \mathbb{R}$ is called continuous at $\xi_o \in \Xi$ if for all $\varepsilon > 0$ and all sequences $\xi_n \rightarrow \xi_o$ it holds that $-\varepsilon \leq \mathcal{G}(\xi_n) - \mathcal{G}(\xi_o) \leq \varepsilon$ for almost all $n \in \mathbb{N}$, meaning for all but finitely many $n \in \mathbb{N}$. If only the lower inequality holds, the function is called lower semicontinuous, and respectively, upper semicontinuous if only the upper inequality holds.

As Q_{CVaR_α} is the value function of the non-linear parametric program

$$\min_{\eta \in \mathbb{R}} f(\alpha, \eta, x), \quad (14)$$

we need the following specialized case of standard results from non-linear parametric optimization, cf. Section 4.2 in [5].

Lemma 4.1. *Consider the problem*

$$(P_\lambda) \quad \inf\{\mathcal{F}(\theta, \lambda) : \theta \in \Theta\}, \quad \lambda \in \Lambda,$$

where Λ and Θ are nonempty spaces each endowed with some notion of convergence and $\mathcal{F} : \Theta \times \Lambda \rightarrow \mathbb{R}$. Let $\varphi(\lambda)$ and $\psi(\lambda)$ denote the infimum and the optimal set respectively of (P_λ) . Then the following holds:

- (i) φ is lower semicontinuous at $\lambda_o \in \Lambda$ if \mathcal{F} is lower semicontinuous on $\Theta \times \{\lambda_o\}$ and if for each sequence $\lambda_n \rightarrow \lambda_o$ there exists a compact subset K of Θ such that $\psi(\lambda_n) \cap K \neq \emptyset$ holds for all $n \in \mathbb{N}$.
- (ii) φ is upper semicontinuous at λ_o if $\mathcal{F}(\theta, \cdot)$ is upper semicontinuous at λ_o for all $\theta \in \Theta$.

Proof. We prove (i) by contradiction: Assume φ is not lower semicontinuous at λ_o . Then there exist a sequence $\lambda_n \rightarrow \lambda_o$ and an $\varepsilon_o > 0$ such that $\varphi(\lambda_n) < \varphi(\lambda_o) - \varepsilon_o$, w.l.o.g., for all $n \in \mathbb{N}$. By assumption there exist $\theta_n \in \psi(\lambda_n) \cap K$ for all $n \in \mathbb{N}$ such that $\mathcal{F}(\theta_n, \lambda_n) = \varphi(\lambda_n) < \varphi(\lambda_o) - \varepsilon_o$ for all $n \in \mathbb{N}$. Since K is compact, there exists $\tilde{\theta} \in K$ with, w.l.o.g., $\theta_n \rightarrow \tilde{\theta}$. Using the lower semicontinuity of \mathcal{F} on $\Theta \times \{\lambda_o\}$ we obtain, $\mathcal{F}(\tilde{\theta}, \lambda_o) \leq \mathcal{F}(\theta_n, \lambda_n) + \varepsilon_o < \varphi(\lambda_o) - \varepsilon_o + \varepsilon_o = \varphi(\lambda_o)$, yielding a contradiction to $\varphi(\lambda_o) = \inf\{\mathcal{F}(\theta, \lambda_o) : \theta \in \Theta\}$. Therefore, φ is lower semicontinuous at λ_o .

For proving (ii) we need to show that for all $\varepsilon > 0$ and all sequences $\lambda_n \rightarrow \lambda_o$ it holds that $\varphi(\lambda_n) \leq \varphi(\lambda_o) + \varepsilon$ for almost all $n \in \mathbb{N}$. Since $\varphi(\lambda_o) = \inf\{\mathcal{F}(\theta, \lambda_o) : \theta \in \Theta\}$, there exists $\tilde{\theta} \in \Theta$ with $\mathcal{F}(\tilde{\theta}, \lambda_o) \leq \varphi(\lambda_o) + \frac{\varepsilon}{2}$. By the upper semicontinuity of $\mathcal{F}(\tilde{\theta}, \cdot)$ at λ_o , and since $\varphi(\lambda_n) = \inf\{\mathcal{F}(\theta, \lambda_n) : \theta \in \Theta\}$, we obtain $\varphi(\lambda_n) \leq \mathcal{F}(\tilde{\theta}, \lambda_n) \leq \mathcal{F}(\tilde{\theta}, \lambda_o) + \frac{\varepsilon}{2} \leq \varphi(\lambda_o) + \varepsilon$ for almost all $n \in \mathbb{N}$, yielding the upper semicontinuity of φ at λ_o . \square

Verifying the assumptions of Lemma 4.1 for the non-linear program (14) will be a recurring issue in the present Section.

The mixed-integer value function Φ from (10) is crucial for the structural understanding of f . From parametric integer optimization ([6, 11]) the following is known about Φ .

Proposition 4.2. *Assume that $W(\mathbb{Z}_+^m) + W'(\mathbb{R}_+^{m'}) = \mathbb{R}^s$ and $\{u \in \mathbb{R}^s : W^\top u \leq q, W'^\top u \leq q'\} \neq \emptyset$. Then the following holds:*

- (i) Φ is real-valued and lower semicontinuous on \mathbb{R}^s .
- (ii) There exists a countable partition $\mathbb{R}^s = \cup_{i=1}^\infty \mathcal{T}_i$ such that the restrictions of Φ to \mathcal{T}_i are piecewise linear and Lipschitz continuous with a uniform constant not depending on i .

- (iii) Each of the sets \mathcal{T}_i has a representation $\mathcal{T}_i = \{t_i + \mathcal{K}\} \setminus \cup_{j=1}^N \{t_{ij} + \mathcal{K}\}$ where \mathcal{K} denotes the polyhedral cone $W'(\mathbb{R}_+^{m'})$ and t_i, t_{ij} are suitable points from \mathbb{R}^s , moreover, N does not depend on i .
- (iv) There exist positive constants β, γ such that $|\Phi(t_1) - \Phi(t_2)| \leq \beta \|t_1 - t_2\| + \gamma$ whenever $t_1, t_2 \in \mathbb{R}^s$.

For convenience, we denote by μ the image measure $\mathbb{P} \circ h^{-1}$ on \mathbb{R}^s , such that

$$f(\alpha, \eta, x) = \eta + \frac{1}{1 - \alpha} \int_{\mathbb{R}^s} \max\{c^\top x + \Phi(h - Tx) - \eta, 0\} \mu(dh).$$

To facilitate notation we introduce for all $x \in \mathbb{R}^m$,

$$M_d(x) := \{h \in \mathbb{R}^s : \Phi \text{ is discontinuous at } h - Tx\},$$

which is μ -measurable for all $x \in \mathbb{R}^m$, cf. [8], p. 225.

The following Lemma shows that in Proposition 2.2 the assumption on the finiteness of the first moment of the relevant random variables is fulfilled. This yields that $Q_{CVaR_\alpha}(x) \in \mathbb{R}$ for all $x \in X$, and therefore the definition of Q_{CVaR_α} in (12) is justified.

Lemma 4.3. *Assume that $W(\mathbb{Z}_+^m) + W'(\mathbb{R}_+^{m'}) = \mathbb{R}^s$, $\{u \in \mathbb{R}^s : W^\top u \leq q, W'^\top u \leq q'\} \neq \emptyset$ and $\int_{\mathbb{R}^s} \|h\| \mu(dh) < \infty$. Then $\mathbb{E}\{|c^\top x + \Phi(h(\omega) - Tx)|\} < \infty$ for all $x \in X$.*

Proof. The function $c^\top x + \Phi(h - Tx)$ is measurable in h , since Φ is measurable as a lower semicontinuous function, cf. Proposition 4.2(i). The second assumption implies that $\Phi(0) = 0$. Together with Proposition 4.2(iv) this provides the following estimate

$$\begin{aligned} |c^\top x + \Phi(h - Tx)| &\leq |c^\top x| + |\Phi(h - Tx)| \\ &\leq |c^\top x| + \beta \|h - Tx\| + \gamma \\ &\leq |c^\top x| + \beta \|h\| + \beta \|Tx\| + \gamma. \end{aligned}$$

Thus

$$\int_{\mathbb{R}^s} |c^\top x + \Phi(h - Tx)| \mu(dh) \leq |c^\top x| + \beta \int_{\mathbb{R}^s} \|h\| \mu(dh) + \beta \|Tx\| + \gamma < \infty$$

yielding the assertion. \square

As already mentioned the function

$$Q_E : \mathbb{R}^m \longrightarrow \mathbb{R}, \quad Q_E(x) = \int_{\mathbb{R}^s} (c^\top x + \Phi(h - Tx)) \mu(dh)$$

is well-studied. In this section we want to use the results available for Q_E to establish similar results for f . For this purpose we consider the η -Expected Excess, now also dependent on η ,

$$\begin{aligned} Q_{E_\eta} &: \mathbb{R} \times \mathbb{R}^m \longrightarrow \mathbb{R}, \\ Q_{E_\eta}(\tilde{x}) &= Q_{E_\eta}(\eta, x) := \int_{\mathbb{R}^s} \max\{c^\top x + \Phi(h - Tx) - \eta, 0\} \mu(dh). \end{aligned}$$

We will show that it is possible to express $Q_{\mathcal{E}_\eta}$ in the form of $Q_{\mathcal{E}}$ with a suitably adapted second-stage mixed-integer program fulfilling the needed assumptions. This enables us to use the results already proved for $Q_{\mathcal{E}}$, immediately conclude the corresponding results for $Q_{\mathcal{E}_\eta}$ and so for f .

Lemma 4.4. *The η -Expected Excess $Q_{\mathcal{E}_\eta}$ can be written in the following form*

$$Q_{\mathcal{E}_\eta}(\eta, x) = Q_{\mathcal{E}_\eta}(\tilde{x}) = \tilde{Q}_{\mathcal{E}}(\tilde{x}) := \int_{\mathbb{R}^{s+1}} (\tilde{c}^\top \tilde{x} + \tilde{\Phi}(\tilde{h} - \tilde{T}\tilde{x})) \tilde{\mu}(d\tilde{h}),$$

with

$$\tilde{\Phi}(\tilde{t}) = \min\{\tilde{q}^\top y + \tilde{q}'^\top \tilde{y}' : \tilde{W}y + \tilde{W}'\tilde{y}' = \tilde{t}, y \in \mathbb{Z}_+^{\tilde{m}}, \tilde{y}' \in \mathbb{R}_+^{\tilde{m}'}\},$$

$$\tilde{h}(\omega) := \begin{pmatrix} h(\omega) \\ 0 \end{pmatrix}, \text{ and } \tilde{\mu} := \mathbb{P} \circ \tilde{h}^{-1}.$$

Further, assume that q and q' are rational vectors, $W(\mathbb{Z}_+^{\tilde{m}}) + W'(\mathbb{R}_+^{\tilde{m}'}) = \mathbb{R}^s$, and $\{u \in \mathbb{R}^s : W^\top u \leq q, W'^\top u \leq q'\} \neq \emptyset$. Then it holds that $\tilde{W}(\mathbb{Z}_+^{\tilde{m}}) + \tilde{W}'(\mathbb{R}_+^{\tilde{m}'}) = \mathbb{R}^{s+1}$, $\{\tilde{u} \in \mathbb{R}^{s+1} : \tilde{W}^\top \tilde{u} \leq \tilde{q}, \tilde{W}'^\top \tilde{u} \leq \tilde{q}'\} \neq \emptyset$, and \tilde{W}, \tilde{W}' are rational matrices.

Proof. For $\tilde{t} \in \mathbb{R}^{s+1}$ let

$$\begin{aligned} \tilde{\Phi}(\tilde{t}) &:= \min \left\{ 0^\top y + \begin{pmatrix} 0 \\ 1 \end{pmatrix}^\top \begin{pmatrix} y' \\ v \\ w \end{pmatrix} : \begin{pmatrix} W \\ q^\top \end{pmatrix} y + \begin{pmatrix} W' & 0 & 0 \\ q'^\top & -1 & 1 \end{pmatrix} \begin{pmatrix} y' \\ v \\ w \end{pmatrix} = \tilde{t}, \right. \\ &\quad \left. y \in \mathbb{Z}_+^{\tilde{m}}, \begin{pmatrix} y' \\ v \\ w \end{pmatrix} \in \mathbb{R}_+^{\tilde{m}'+2} \right\} \\ &= \min \{ v : Wy + W'y' = \tilde{t}_1, v \geq q^\top y + q'^\top y' - \tilde{t}_2, \\ &\quad y \in \mathbb{Z}_+^{\tilde{m}}, y' \in \mathbb{R}_+^{\tilde{m}'}, v \in \mathbb{R}_+ \}, \end{aligned}$$

then it holds

$$\max\{c^\top x + \Phi(h - Tx) - \eta, 0\} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}^\top \begin{pmatrix} \eta \\ x \end{pmatrix} + \tilde{\Phi} \left(\begin{pmatrix} h \\ 0 \end{pmatrix} - \begin{pmatrix} 0 & T \\ -1 & c^\top \end{pmatrix} \begin{pmatrix} \eta \\ x \end{pmatrix} \right),$$

such that

$$\begin{aligned} Q_{\mathcal{E}_\eta}(\eta, x) &= \int_{\mathbb{R}^s} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}^\top \begin{pmatrix} \eta \\ x \end{pmatrix} + \tilde{\Phi} \left(\begin{pmatrix} h \\ 0 \end{pmatrix} - \begin{pmatrix} 0 & T \\ -1 & c^\top \end{pmatrix} \begin{pmatrix} \eta \\ x \end{pmatrix} \right) \right) \mu(dh) \\ &= \int_{\mathbb{R}^{s+1}} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}^\top \begin{pmatrix} \eta \\ x \end{pmatrix} + \tilde{\Phi} \left(\tilde{h} - \begin{pmatrix} 0 & T \\ -1 & c^\top \end{pmatrix} \begin{pmatrix} \eta \\ x \end{pmatrix} \right) \right) \tilde{\mu}(d\tilde{h}), \end{aligned}$$

by the definition of \tilde{h} , $\tilde{\mu}$, and by the transformation formula for integrals with respect to image measures.

Now, verify $\tilde{W}(\mathbb{Z}_+^{\tilde{m}}) + \tilde{W}'(\mathbb{R}_+^{\tilde{m}'}) = \mathbb{R}^{s+1}$ as follows: Let $(\tilde{t}_1, \tilde{t}_2)^\top \in \mathbb{R}^{s+1}$. By the first assumption of the Lemma there exist $y \in \mathbb{Z}_+^{\tilde{m}}$ and $y' \in \mathbb{R}_+^{\tilde{m}'}$ such that $Wy + W'y' = \tilde{t}_1$ and further there are $v, w \in \mathbb{R}_+$ such that $q^\top y + q'^\top y' - v + w = \tilde{t}_2$.

The next condition also holds, because by the second assumption there exists $u \in \mathbb{R}^s$ such that $W^\top u \leq q$ and $W'^\top u \leq q'$ and so with $\tilde{u} := (u, -1)^\top$ we have $\tilde{W}^\top \tilde{u} = W^\top u - q^\top \leq 0 = \tilde{q}^\top$ and $\tilde{W}'^\top \tilde{u} = (W'^\top u - q'^\top, 1, -1)^\top \leq (0, 1, 0)^\top = \tilde{q}'^\top$.

Finally, the vectors q and q' as well as the matrices W and W' are assumed to be rational such that the matrices \tilde{W} and \tilde{W}' are rational, which completes the proof. \square

In addition to the rationality assumptions on the matrices W and W' we want to assume for the remaining part of the paper that the vectors q and q' are rational, too.

Now, we are ready to derive continuity properties of $Q_{\mathcal{E}_\eta}$ and then proceed with the objective function f in (14).

Proposition 4.5. *Assume that $W(\mathbb{Z}_+^{\tilde{m}}) + W'(\mathbb{R}_+^{\tilde{m}'}) = \mathbb{R}^s$, $\{u \in \mathbb{R}^s : W^\top u \leq q, W'^\top u \leq q'\} \neq \emptyset$ and $\int_{\mathbb{R}^s} \|h\| \mu(dh) < \infty$. Then $Q_{\mathcal{E}_\eta} : \mathbb{R} \times \mathbb{R}^m \rightarrow \mathbb{R}$ is a real-valued lower semicontinuous function. If, in addition, it holds that $\mu(M_d(x)) = 0$, then $Q_{\mathcal{E}_\eta}$ is continuous at (η, x) for all $\eta \in \mathbb{R}$. The latter assumption is fulfilled for all $x \in \mathbb{R}^m$ if μ has a density. Then $Q_{\mathcal{E}_\eta}$ is continuous on $\mathbb{R} \times \mathbb{R}^m$. Further, it holds that $Q_{\mathcal{E}_\eta}(\cdot, x)$ is convex for all $x \in \mathbb{R}^m$.*

Proof. Lemma 4.4 allows us to derive these assertions from the known results for $Q_{\mathcal{E}}$. Let $\tilde{M}_d(\tilde{x}) := \{\tilde{h} \in \mathbb{R}^{s+1} : \tilde{\Phi} \text{ is discontinuous at } \tilde{h} - \tilde{T}\tilde{x}\}$. Since $\tilde{h}(\omega) = (h(\omega), 0)^\top$, we obtain $\tilde{h}^{-1}[\tilde{M}_d(\tilde{x})] = h^{-1}[\wp_{\mathbb{R}^s}(\tilde{M}_d(\tilde{x}) \cap \{\tilde{h} \in \mathbb{R}^{s+1} : \tilde{h}_{s+1} = 0\})]$, where $\wp_{\mathbb{R}^s}$ is the canonical projection on \mathbb{R}^s . Further $\wp_{\mathbb{R}^s}(\tilde{M}_d(\tilde{x}) \cap \{\tilde{h} \in \mathbb{R}^{s+1} : \tilde{h}_{s+1} = 0\}) \subseteq M_d(x)$. Indeed, by $\tilde{\Phi}(\tilde{h} - \tilde{T}\tilde{x}) = \max\{c^\top x + \Phi(h - Tx) - \eta, 0\}$ it follows that $\tilde{\Phi}$ is discontinuous at $\tilde{h} - \tilde{T}\tilde{x}$ only in the case Φ is discontinuous at $h - Tx$. Therefore $\tilde{\mu}(\tilde{M}_d(\tilde{x})) = \mathbb{P}(\tilde{h}^{-1}[\tilde{M}_d(\tilde{x})]) \leq \mathbb{P}(h^{-1}[M_d(x)]) = \mu(M_d(x))$. So, $\mu(M_d(x)) = 0$ implies $\tilde{\mu}(\tilde{M}_d(\tilde{x})) = 0$ for all $\tilde{x} \in \mathbb{R}^m$. Further, $\int_{\mathbb{R}^s} \|h\| \mu(dh) < \infty$ yields $\int_{\mathbb{R}^{s+1}} \|\tilde{h}\| \tilde{\mu}(d\tilde{h}) < \infty$ by the definition of $\tilde{\mu}$. Altogether the assumptions for the known results (cf. Proposition 3.1 and 3.2 in [31]) are verified, applying these to $Q_{\mathcal{E}_\eta}(\eta, x) = \tilde{Q}_{\mathcal{E}}(\tilde{x})$ yields the desired properties.

In view of Proposition 4.2(ii),(iii), for given $x \in \mathbb{R}^m$, $M_d(x)$ is contained in a countable union of hyperplanes, i.e., in a set of Lebesgue measure zero. If μ has a density, it is absolutely continuous with respect to the Lebesgue measure, hence $\mu(M_d(x)) = 0$.

The convexity of $Q_{\mathcal{E}_\eta}(\cdot, x)$ for all $x \in \mathbb{R}^m$ holds, because the maximum of two convex functions is convex and by the linearity of the integral. This completes the proof. \square

Proposition 4.6. *Assume that $W(\mathbb{Z}_+^{\tilde{m}}) + W'(\mathbb{R}_+^{\tilde{m}'}) = \mathbb{R}^s$, $\{u \in \mathbb{R}^s : W^\top u \leq q, W'^\top u \leq q'\} \neq \emptyset$ and $\int_{\mathbb{R}^s} \|h\| \mu(dh) < \infty$. Then $f : (0, 1) \times \mathbb{R} \times \mathbb{R}^m \rightarrow \mathbb{R}$ is a real-valued lower semicontinuous function. If, in addition, it holds that*

$\mu(M_d(x)) = 0$, then f is continuous at (α, η, x) for all $(\alpha, \eta) \in (0, 1) \times \mathbb{R}$. The latter assumption is fulfilled for all $x \in \mathbb{R}^m$ if μ has a density. Then f is continuous on $(0, 1) \times \mathbb{R} \times \mathbb{R}^m$. Further, it holds that $f(\alpha, \cdot, x)$ is convex for all $(\alpha, x) \in (0, 1) \times \mathbb{R}^m$.

Proof. It holds that $f(\alpha, \eta, x) = \eta + \frac{1}{1-\alpha} \int_{\mathbb{R}^s} \max\{c^\top x + \Phi(h - Tx) - \eta, 0\} \mu(dh) = \eta + \frac{1}{1-\alpha} Q_{E_\eta}(\eta, x)$. Further $\alpha \in (0, 1)$ and so $0 < \frac{1}{1-\alpha} < \infty$, such that the results follow immediately from Proposition 4.5. \square

We also want to extend our results to continuity results jointly with the measure μ . If indicated the functions in question are, thus, also functions dependent on the measure μ .

Let $\mathcal{P}(\mathbb{R}^s)$ denote the set of all Borel probability measures on \mathbb{R}^s . We say that a sequence $\{\mu_n\}$ in $\mathcal{P}(\mathbb{R}^s)$ converges weakly to $\mu \in \mathcal{P}(\mathbb{R}^s)$, written $\mu_n \xrightarrow{w} \mu$, if for any bounded continuous function $g : \mathbb{R}^s \rightarrow \mathbb{R}$ it holds

$$\int_{\mathbb{R}^s} g(h) \mu_n(dh) \rightarrow \int_{\mathbb{R}^s} g(h) \mu(dh) \quad \text{as } n \rightarrow \infty.$$

Fix arbitrary $p > 1$ and $C > 0$, and denote $\Delta_{p,C}(\mathbb{R}^s) := \{\nu \in \mathcal{P}(\mathbb{R}^s) : \int_{\mathbb{R}^s} \|h\|^p \nu(dh) \leq C\}$.

Proposition 4.7. *Assume that $W(\mathbb{Z}_+^{\tilde{m}}) + W'(\mathbb{R}_+^{m'}) = \mathbb{R}^s$, $\{u \in \mathbb{R}^s : W^\top u \leq q, W'^\top u \leq q'\} \neq \emptyset$ and let $\mu \in \Delta_{p,C}(\mathbb{R}^s)$ such that $\mu(M_d(x)) = 0$. Then $Q_{E_\eta} : \mathbb{R} \times \mathbb{R}^m \times \Delta_{p,C}(\mathbb{R}^s) \rightarrow \mathbb{R}$ is continuous at (η, x, μ) for all $\eta \in \mathbb{R}$.*

Proof. Using the known result for Q_E , cf. Proposition 3.8 in [31], it remains to show that $\mu \in \Delta_{p,C}(\mathbb{R}^s)$ implies $\tilde{\mu} \in \Delta_{p,C}(\mathbb{R}^{s+1})$, which is straightforward by the definition of the entities, and that $\mu(M_d(x)) = 0$ implies $\tilde{\mu}(\tilde{M}_d(\tilde{x})) = 0$, which has already been shown in the proof of Proposition 4.5. \square

Proposition 4.8. *Assume that $W(\mathbb{Z}_+^{\tilde{m}}) + W'(\mathbb{R}_+^{m'}) = \mathbb{R}^s$, $\{u \in \mathbb{R}^s : W^\top u \leq q, W'^\top u \leq q'\} \neq \emptyset$ and let $\mu \in \Delta_{p,C}(\mathbb{R}^s)$ such that $\mu(M_d(x)) = 0$. Then $f : (0, 1) \times \mathbb{R} \times \mathbb{R}^m \times \Delta_{p,C}(\mathbb{R}^s) \rightarrow \mathbb{R}$ is continuous at (α, η, x, μ) for all $(\alpha, \eta) \in (0, 1) \times \mathbb{R}$.*

Proof. Since $\alpha \in (0, 1)$ and so $0 < \frac{1}{1-\alpha} < \infty$, the results follow immediately from Proposition 4.7. \square

The next Lemma verifies the remaining assumption of Lemma 4.1.

Lemma 4.9. *Assume that $W(\mathbb{Z}_+^{\tilde{m}}) + W'(\mathbb{R}_+^{m'}) = \mathbb{R}^s$ and $\{u \in \mathbb{R}^s : W^\top u \leq q, W'^\top u \leq q'\} \neq \emptyset$. Let $(\alpha_n, x_n, \mu_n) \rightarrow (\alpha_o, x_o, \mu_o)$ with $\{\alpha_n, x_n, \mu_n\}_{n \in \mathbb{N}} \subseteq (0, 1) \times \mathbb{R}^m \times \mathcal{P}(\mathbb{R}^s)$. Then there exists a compact subset K of \mathbb{R} such that $\eta_{\alpha_n}^{\mu_n}(x_n) := \min\{\eta : \mu_n(\{h \in \mathbb{R}^s : c^\top x_n + \Phi(h - Tx_n) \leq \eta\}) \geq \alpha_n\} \in K$ for all $n \in \mathbb{N}$.*

Proof. For every such sequence there exist $\alpha', \alpha'' \in (0, 1)$, and a real number $r := \max_{n \in \mathbb{N}} \|x_n\|$ such that $(\alpha_n, x_n) \in [\alpha', \alpha''] \times B_r(0)$ for all $n \in \mathbb{N}$, where $B_r(0) := \{x \in \mathbb{R}^m : \|x\| \leq r\}$.

The second assumption implies that $\Phi(0) = 0$. Together with Proposition 4.2(iv) this provides the following estimate for $x \in B_r(0)$, $h \in \mathbb{R}^s$,

$$\begin{aligned} |c^\top x + \Phi(h - Tx)| &\leq |c^\top x| + |\Phi(h - Tx)| \leq |c^\top x| + \beta \|h - Tx\| + \gamma \\ &\leq r \|c\| + \beta \|h\| + \beta r \|T\| + \gamma \\ &\leq \beta \|h\| + \bar{r}, \end{aligned}$$

where $\beta, \gamma > 0$ and $\bar{r} := r \|c\| + \beta r \|T\| + \gamma > 0$.

Using the above estimate, we obtain for $(\alpha, x) \in [\alpha', \alpha''] \times B_r(0)$ and $\mu \in \Delta := \{\mu_n\}_{n \in \mathbb{N}}$,

$$\begin{aligned} \eta_\alpha^\mu(x) &= \min\{\eta : \mu(\{h \in \mathbb{R}^s : c^\top x + \Phi(h - Tx) \leq \eta\}) \geq \alpha\} \\ &\leq \min\{\eta : \mu(\{h \in \mathbb{R}^s : c^\top x + \Phi(h - Tx) \leq \eta\}) \geq \alpha''\} \\ &\leq \min\{\eta : \mu(\{h \in \mathbb{R}^s : \beta \|h\| + \bar{r} \leq \eta\}) \geq \alpha''\} \\ &= \min\{\eta : \mu(\{h \in \mathbb{R}^s : \|h\| \leq \beta^{-1}(\eta - \bar{r})\}) \geq \alpha''\} \\ &= \beta \cdot \min\{\eta : \mu(\{h \in \mathbb{R}^s : \|h\| \leq \eta\}) \geq \alpha''\} + \bar{r}, \end{aligned}$$

with the substitution $\eta := \beta^{-1}(\eta - \bar{r})$, and

$$\begin{aligned} \eta_\alpha^\mu(x) &= \min\{\eta : \mu(\{h \in \mathbb{R}^s : c^\top x + \Phi(h - Tx) \leq \eta\}) \geq \alpha\} \\ &\geq \min\{\eta : \mu(\{h \in \mathbb{R}^s : c^\top x + \Phi(h - Tx) \leq \eta\}) \geq \alpha'\} \\ &\geq \min\{\eta : \mu(\{h \in \mathbb{R}^s : -\beta \|h\| - \bar{r} \leq \eta\}) \geq \alpha'\} \\ &= \min\{\eta : \mu(\{h \in \mathbb{R}^s : \beta \|h\| \geq -(\eta + \bar{r})\}) \geq \alpha'\} \\ &= \min\{\eta : \mu(\{h \in \mathbb{R}^s : \|h\| \geq -\beta^{-1}(\eta + \bar{r})\}) \geq \alpha'\} \\ &= \beta \cdot \min\{\eta : \mu(\{h \in \mathbb{R}^s : \|h\| \geq -\eta\}) \geq \alpha'\} - \bar{r}, \end{aligned}$$

where $\eta := \beta^{-1}(\eta + \bar{r})$.

We complete the proof by finding an upper bound for

$$\bar{\eta}_{\alpha''}^\mu := \min\{\eta : \mu(\{h \in \mathbb{R}^s : \|h\| \leq \eta\}) \geq \alpha''\},$$

and a lower bound for

$$\begin{aligned} \tilde{\eta}_{\alpha'}^\mu &:= \min\{\eta : \mu(\{h \in \mathbb{R}^s : \|h\| \geq -\eta\}) \geq \alpha'\} \\ &= -\max\{\eta : \mu(\{h \in \mathbb{R}^s : \|h\| < \eta\}) \leq 1 - \alpha'\}, \end{aligned}$$

both independent of $\mu \in \Delta$.

Since the elements of Δ form a weakly convergent sequence, Prohorov's Theorem, cf. [8], Theorem 6.2, tells us that Δ is uniformly tight: There exists, for each $\varepsilon > 0$, a compact set $\mathcal{C} \subseteq \mathbb{R}^s$ such that $\mu(\mathcal{C}) > 1 - \varepsilon$ for all $\mu \in \Delta$. This yields the desired bounds as follows: First, let $\varepsilon := 1 - \alpha'' > 0$. Then there exists a compact set \mathcal{C} such that $\mu(\mathcal{C}) > \alpha''$ for all $\mu \in \Delta$, yielding the existence of

$b \in \mathbb{R}$ such that $\tilde{\eta}_{\alpha'}^\mu \leq b$. Secondly, let $\varepsilon := \alpha' > 0$. Then there exists a compact set \mathcal{C} such that $\mu(\mathcal{C}) > 1 - \alpha'$ for all $\mu \in \Delta$. Hence there is a constant $-a \in \mathbb{R}$ such that $\max\{\eta : \mu(\{h \in \mathbb{R}^s : \|h\| < \eta\}) \leq 1 - \alpha'\} \leq -a$, and thus $\tilde{\eta}_{\alpha'}^\mu \geq a$.

Summing up, we obtain that for all $n \in \mathbb{N}$ it holds

$$\eta_{\alpha_n}^{\mu_n}(x_n) \in [\beta a - \bar{r}, \beta b + \bar{r}] := K.$$

□

Now, we are ready to apply Lemma 4.1 to (14). First, for fixed measure μ , we study Q_{CVaR_α} as a function jointly in (α, x) , tacitly assuming proper extension of the notation in (12).

Proposition 4.10. *Assume that $W(\mathbb{Z}_+^m) + W'(\mathbb{R}_+^{m'}) = \mathbb{R}^s$, $\{u \in \mathbb{R}^s : W^\top u \leq q, W'^\top u \leq q'\} \neq \emptyset$ and $\int_{\mathbb{R}^s} \|h\| \mu(dh) < \infty$. Then $Q_{CVaR_\alpha} : (0, 1) \times \mathbb{R}^m \rightarrow \mathbb{R}$ is a real-valued lower semicontinuous function. If, in addition, it holds that $\mu(M_d(x)) = 0$, then Q_{CVaR_α} is continuous at (α, x) for all $\alpha \in (0, 1)$. The latter assumption is fulfilled for all $x \in \mathbb{R}^m$ if μ has a density. Then Q_{CVaR_α} is continuous on $(0, 1) \times \mathbb{R}^m$.*

Proof. Follows immediately from Proposition 2.2, Lemma 4.3 and Lemma 4.1 together with Proposition 4.6 and Lemma 4.9. □

Finally, we obtain the joint continuity of Q_{CVaR_α} in (α, x, μ) .

Proposition 4.11. *Assume that $W(\mathbb{Z}_+^m) + W'(\mathbb{R}_+^{m'}) = \mathbb{R}^s$, $\{u \in \mathbb{R}^s : W^\top u \leq q, W'^\top u \leq q'\} \neq \emptyset$ and let $\mu \in \Delta_{p,C}(\mathbb{R}^s)$ such that $\mu(M_d(x)) = 0$. Then $Q_{CVaR_\alpha} : (0, 1) \times \mathbb{R}^m \times \Delta_{p,C}(\mathbb{R}^s) \rightarrow \mathbb{R}$ is continuous at (α, x, μ) for all $\alpha \in (0, 1)$.*

Proof. Follows immediately from Proposition 2.2, Lemma 4.3 and Lemma 4.1 together with Proposition 4.8 and Lemma 4.9. □

Proposition 4.10 states the basic analytical properties of Q_{CVaR_α} as a function in (α, x) . As a particular conclusion of the lower semicontinuity in x established there we obtain that the mean-risk model (13) always has an optimal solution, provided X is nonempty and compact. Joint-continuity results like Proposition 4.11 are instrumental when studying the stability of stochastic programs with respect to perturbations of the underlying probability measures. For an illustration let us consider the problem

$$(P_\mu) \quad \min\{Q_{CVaR_\alpha}(x, \mu) : x \in X\}$$

where $\mu \in \mathcal{P}(\mathbb{R}^s)$ enters as a parameter, and $\alpha \in (0, 1)$ is preselected and fixed. Again we focus on the pure risk model and remark that in combination with results from, e.g., [31] the findings below readily extend into stability results for mean-risk models of the type (13).

There are two main issues motivating stability analysis of stochastic programs: numerical accessibility and incompleteness of information. The definition

of $Q_{CVaR_\alpha}(x, \mu)$ involves multivariate integration which is hard to accomplish if the dimension is substantial and the integrating measure continuous. Approximation by discrete measures then is a proven remedy. Moreover, in most practical modeling situations of stochastic programming the underlying probability measure results from subjective considerations and, therefore, is available in an approximate sense only. In both cases, stability then arises as a natural requirement: One wants to be sure that “small” perturbations of the underlying measure imply only “small” perturbations of the solutions. For putting perturbation of probability measures into mathematical terms weak convergence of probability measures is very useful. It covers modes of perturbation that are particularly important in stochastic programming. As examples we mention discretization of continuous probability measures via conditional expectations [10, 16], almost surely converging densities (Scheffé’s Theorem [8]), and estimation using empirical measures (Glivenko-Cantelli almost sure uniform convergence [26]).

$Q_{CVaR_\alpha}(x, \mu)$ being nonconvex in x in general, local solutions to (P_μ) that are not necessarily global become relevant, and it is convenient to consider the following localized optimal values and solution sets

$$\begin{aligned}\varphi_V(\mu) &:= \inf\{Q_{CVaR_\alpha}(x, \mu) : x \in X \cap \text{cl } V\}, \\ \psi_V(\mu) &:= \{x \in X \cap \text{cl } V : Q_{CVaR_\alpha}(x, \mu) = \varphi_V(\mu)\},\end{aligned}$$

where $V \subseteq \mathbb{R}^m$. Given $\mu \in \mathcal{P}(\mathbb{R}^s)$, a nonempty set $Z \subseteq \mathbb{R}^m$ is called a complete local minimizing set (CLM set, [28]) of (P_μ) with respect to V if V is open and $Z = \psi_V(\mu) \subseteq V$. A set of local minimizers thus has the CLM property if it contains all local minimizers “nearby”. Without this property pathologies may occur. Think, for example, of some real-valued function on \mathbb{R}^m that is constant on some ball B_r with radius r . Any ball $B_{\frac{r}{2}}$ around the same point but with radius $\frac{r}{2}$ then is a set of local minimizers (lacking the CLM property, of course). “Tilting” the function by adding a suitable linear function with arbitrarily small norm then creates instability since none of the points “in and near” $B_{\frac{r}{2}}$ remains locally optimal. It thus makes sense to impose the CLM property when studying stability of local solutions. Isolated local minimizers and the set of global minimizers are examples for CLM sets, while strict local minimizers not necessarily have the CLM property, cf. [28] for further details.

Proposition 4.12. *Assume that $W(\mathbb{Z}_+^m) + W'(\mathbb{R}_+^{m'}) = \mathbb{R}^s$, $\{u \in \mathbb{R}^s : W^\top u \leq q, W'^\top u \leq q'\} \neq \emptyset$ and let $\mu \in \Delta_{p,C}(\mathbb{R}^s)$ with $\mu(M_d(x)) = 0$ for all $x \in X$. Further, suppose that there exists a subset $Z \subseteq \mathbb{R}^m$ which is a CLM set for (P_μ) with respect to some bounded open set $V \subseteq \mathbb{R}^m$. Then the following holds:*

- (i) *The function $\varphi_V : \Delta_{p,C}(\mathbb{R}^s) \rightarrow \mathbb{R}$ is continuous at μ .*
- (ii) *The multifunction $\psi_V : \Delta_{p,C}(\mathbb{R}^s) \rightarrow 2^{\mathbb{R}^m}$ is Berge upper semicontinuous at μ ; i.e., for any open set $\mathcal{O} \subseteq \mathbb{R}^m$ with $\psi_V(\mu) \subseteq \mathcal{O}$ there exists a neighborhood \mathcal{N} of μ in $\Delta_{p,C}(\mathbb{R}^s)$ such that $\psi_V(\nu) \subseteq \mathcal{O}$ for all $\nu \in \mathcal{N}$.*
- (iii) *There exists a neighborhood \mathcal{N}' of μ in $\Delta_{p,C}(\mathbb{R}^s)$ such that for all $\nu \in \mathcal{N}'$ the set $\psi_V(\nu)$ is a CLM set for (P_ν) with respect to V . In particular, this implies that $\psi_V(\nu)$ is a nonempty set of local minimizers whenever $\nu \in \mathcal{N}'$.*

Proof. (i) By Proposition 4.10, $Q_{CVaR_\alpha}(\cdot, \nu)$ is lower semicontinuous on the compact set $K := X \cap \text{cl}V$ for all $\nu \in \Delta_{p,C}(\mathbb{R}^s)$ and therefore $\psi_V(\nu) \cap K \neq \emptyset$ for all $\nu \in \Delta_{p,C}(\mathbb{R}^s)$. Further, with Proposition 4.11 the function $Q_{CVaR_\alpha} : \mathbb{R}^m \times \Delta_{p,C}(\mathbb{R}^s) \rightarrow \mathbb{R}$ is continuous on $K \times \{\mu\}$. Applying Lemma 4.1, now to the setting of (P_μ) , yields the first assertion.

(ii) It holds $\psi_V(\nu) = \{x \in K : g(x, \nu) \leq 0\}$ for all $\nu \in \Delta_{p,C}(\mathbb{R}^s)$, with $g(x, \nu) := Q_{CVaR_\alpha}(x, \nu) - \varphi_V(\nu)$ being lower semicontinuous on $K \times \{\mu\}$. We complete the proof by contradiction: Assume ψ_V is not Berge upper semicontinuous at μ . Then there exists an open set $\mathcal{O} \subseteq \mathbb{R}^m$ with $\psi_V(\mu) \subseteq \mathcal{O}$ and for all neighborhoods \mathcal{N} of μ there is a $\nu \in \mathcal{N}$ such that $\psi_V(\nu) \not\subseteq \mathcal{O}$. This is equivalent to the existence of an open set $\mathcal{O} \subseteq \mathbb{R}^m$ with $\psi_V(\mu) \subseteq \mathcal{O}$ and two sequences $\nu_n \xrightarrow{w} \mu$ and $x_n \in \psi_V(\nu_n) \setminus \mathcal{O} \subseteq K \setminus \mathcal{O}$. By the compactness of $K \setminus \mathcal{O}$ there is a convergent subsequence $x_{n_k} \rightarrow x_o \in K \setminus \mathcal{O}$. However, since $x_{n_k} \in \psi_V(\nu_{n_k})$ and by the lower semicontinuity of $g(x, \nu)$ on $K \times \{\mu\}$, we obtain that for all $\varepsilon > 0$ it holds $g(x_o, \mu) \leq g(x_{n_k}, \nu_{n_k}) + \varepsilon \leq \varepsilon$ for almost all $k \in \mathbb{N}$, and so $g(x_o, \mu) \leq 0$. Therefore, $x_o \in \psi_V(\mu) \subseteq \mathcal{O}$, which yields the contradiction, and ψ_V is Berge upper semicontinuous at μ .

(iii) As shown above in the proof of part (i), it holds that $\psi_V(\nu) \neq \emptyset$ for all $\nu \in \Delta_{p,C}(\mathbb{R}^s)$. Since $Z = \psi_V(\mu)$ is a CLM set for (P_μ) with respect to V , we have $\psi_V(\mu) \subseteq V$. Recall that V is open. By the upper semicontinuity established in (ii) now there exists a neighborhood \mathcal{N}' of μ in $\Delta_{p,C}(\mathbb{R}^s)$ such that $\psi_V(\nu) \subseteq V$ whenever $\nu \in \mathcal{N}'$. Hence, $\psi_V(\nu)$ is a CLM set for all $\nu \in \mathcal{N}'$. \square

5. Algorithm and computational experiments

In this Section we consider the case where the measure μ is discrete with finitely many scenarios. As justified by the results of the last Section these models are suitable to approximate models with continuous probability distributions yielding multivariate integration problems. We start with an equivalent problem formulation that will enable us to use an already existing solution algorithm. Namely the one developed for the traditional expectation-based model, cf. [12].

Proposition 5.1. *Assume that μ is discrete with finitely many scenarios h_1, \dots, h_J and corresponding probabilities π_1, \dots, π_J . Let $\alpha \in (0, 1)$. Then the stochastic program*

$$\min\{Q_{CVaR_\alpha}(x) : x \in X\} \quad (15)$$

can be equivalently restated as

$$\begin{aligned} \min_{x, y, y', v, \eta} \left\{ \eta + \frac{1}{1-\alpha} \sum_{j=1}^J \pi_j v_j : W y_j + W' y'_j = h_j - T x, \right. \\ v_j \geq c^\top x + q^\top y_j + q'^\top y'_j - \eta, \\ x \in X, \eta \in \mathbb{R}, y_j \in \mathbb{Z}_+^m, \\ \left. y'_j \in \mathbb{R}_+^{m'}, v_j \in \mathbb{R}_+, j = 1, \dots, J \right\}. \end{aligned} \quad (16)$$

Proof. By Lemma 4.4, (15) is equivalent to

$$\min_{x, \eta} \left\{ \eta + \frac{1}{1 - \alpha} \int_{\mathbb{R}^{s+1}} (\tilde{c}^\top \tilde{x} + \tilde{\Phi}(\tilde{h} - \tilde{T}\tilde{x})) \tilde{\mu}(d\tilde{h}) : x \in X, \eta \in \mathbb{R} \right\}.$$

Since μ is discrete with finitely many scenarios, this is equivalent to (16) which is already well-known from linear two-stage models without integer requirements, see [9, 17, 27]. The same proof applies here. \square

In the following we assume that $X \subseteq \mathbb{R}^m$ is nonempty and compact, and arises as a solution set to a system of linear inequalities, possibly involving integer requirements to some components of x . By the lower semicontinuity of Q_{CVaR_α} our problem therefore always has an optimal solution.

The dimensions of problem (16) quickly become large-scale such that general-purpose mixed-integer linear programming algorithms and software fail. However, the constraint matrix of (16) has the same block-angular structure which occurs in the traditional expectation-based model: If we consider the additional variable η as a first-stage variable and v_j , $j = 1, \dots, J$, as additional second-stage variables, the second-stage variables (y_j, y'_j, v_j) for different scenarios are not linked in explicit constraints but only through the scenario-independent first-stage variables x and η . Similarly to the traditional expectation-based model, cf. [12], this allows for the following algorithmic approach to (16) via scenario decomposition, i.e., Lagrangian relaxation of nonanticipativity.

We introduce in (16) scenario-many copies x_j, η_j , $j = 1, \dots, J$, of the first-stage variables x and η , and force them to be really copies of each other, by the addition of the nonanticipativity constraints $x_1 = \dots = x_J$ and $\eta_1 = \dots = \eta_J$ (or equivalent systems). For the latter we use the notations $\sum_{j=1}^J H'_j x_j = 0$ and $\sum_{j=1}^J H''_j \eta_j = 0$ with suitable (l', m) -matrices H'_j and $(l'', 1)$ -vectors H''_j , $j = 1, \dots, J$. Problem (16) then becomes

$$\begin{aligned} \min_{x, y, y', v, \eta} \left\{ \sum_{j=1}^J \pi_j \eta_j + \frac{1}{1 - \alpha} \sum_{j=1}^J \pi_j v_j : W y_j + W' y'_j = h_j - T x_j, \right. \\ v_j \geq c^\top x_j + q^\top y_j + q'^\top y'_j - \eta_j, \\ x_j \in X, \eta_j \in \mathbb{R}, y_j \in \mathbb{Z}_+^{\tilde{m}}, \\ y'_j \in \mathbb{R}_+^{m'}, v_j \in \mathbb{R}_+, j = 1, \dots, J, \\ \left. \sum_{j=1}^J H'_j x_j = 0, \sum_{j=1}^J H''_j \eta_j = 0 \right\}. \end{aligned} \quad (17)$$

The constraint system of (17) can be decoupled by Lagrangian relaxation of $\sum_{j=1}^J H'_j x_j = 0$ and $\sum_{j=1}^J H''_j \eta_j = 0$. For this purpose, we consider for $\lambda' \in \mathbb{R}^{l'}$ and $\lambda'' \in \mathbb{R}^{l''}$ the functions

$$L_j(x_j, y_j, y'_j, v_j, \eta_j, \lambda', \lambda'') := \pi_j \eta_j + \frac{1}{1 - \alpha} \pi_j v_j + \lambda'^\top H'_j x_j + \lambda''^\top H''_j \eta_j,$$

for $j = 1, \dots, J$, and obtain the Lagrangian with $\lambda := (\lambda', \lambda'')$

$$L(x, y, y', v, \eta, \lambda) := \sum_{j=1}^J L_j(x_j, y_j, y'_j, v_j, \eta_j, \lambda).$$

Then, the Lagrangian dual of (17) is the optimization problem

$$\max\{D(\lambda) : \lambda \in \mathbb{R}^l\} \quad (18)$$

where $l := l' + l''$ and

$$D(\lambda) = \min \left\{ \sum_{j=1}^J L_j(x_j, y_j, y'_j, v_j, \eta_j, \lambda) : \begin{aligned} &W y_j + W' y'_j = h_j - T x_j, \\ &v_j \geq c^\top x_j + q^\top y_j + q'^\top y'_j - \eta_j, \\ &x_j \in X, \eta_j \in \mathbb{R}, y_j \in \mathbb{Z}_+^m, \\ &y'_j \in \mathbb{R}_+^{m'}, v_j \in \mathbb{R}_+, j = 1, \dots, J \end{aligned} \right\}.$$

Separability yields

$$D(\lambda) = \sum_{j=1}^J D_j(\lambda) \quad (19)$$

where

$$D_j(\lambda) = \min \left\{ L_j(x_j, y_j, y'_j, v_j, \eta_j, \lambda) : \begin{aligned} &W y_j + W' y'_j = h_j - T x_j, \\ &v_j \geq c^\top x_j + q^\top y_j + q'^\top y'_j - \eta_j, \\ &x_j \in X, \eta_j \in \mathbb{R}, y_j \in \mathbb{Z}_+^m, \\ &y'_j \in \mathbb{R}_+^{m'}, v_j \in \mathbb{R}_+ \end{aligned} \right\}. \quad (20)$$

Each $D_j(\lambda)$ is the pointwise minimum of affine functions in λ . Therefore $D(\lambda)$ is piecewise affine and concave. Thus, (18) is a nonsmooth concave maximization (or convex minimization) problem that can be solved by bundle methods from nondifferentiable optimization, for instance by the conic bundle method of [14] or the proximal bundle method of [18, 19].

At each iteration, these methods require the objective value of D which is given by (19) and one subgradient of D which is given by the vector

$$\left(\sum_{j=1}^J H'_j x_j^\lambda, \sum_{j=1}^J H''_j \eta_j^\lambda \right).$$

Here, x_j^λ and η_j^λ refer to the corresponding components in an optimal solution vector to (20). Note that the structure of D , cf. (19), enables substantial decomposition, since the single-scenario problems (20) can be tackled separately. Their

moderate size often allows application of general-purpose mixed-integer linear programming codes, like CPLEX, [15].

The optimal value z_{LD} of (18) is a lower bound to the optimal value z of problem (16). From integer programming theory ([22]) it is known, that, generally, one has to live with a positive duality gap. However, the lower bound obtained by the above procedure, is never worse than the bound obtained by eliminating the integer requirements. In other words, it holds that $z_{LD} \geq z_{LP}$ where z_{LP} denotes the optimal value of the LP relaxation of (16).

The results of the dual optimization provide starting points for heuristics to find promising feasible points for the unrelaxed problem (16). Our relaxed constraints being very simple ($x_1 = \dots = x_J$, $\eta_1 = \dots = \eta_J$), ideas for such heuristics come up quickly. For example, examine the x_j, η_j -components, $j = 1, \dots, J$, of solutions to (20) for optimal or nearly optimal λ , and decide for the most frequent value arising, or average and round if necessary.

If the heuristic yields a feasible solution to (16), then we obtain by the objective value of the latter an upper bound \bar{z} for z . Together with the lower bound z_{LD} this gives the quality certificate (gap) $\bar{z} - z_{LD}$. The full algorithm improves this certificate by embedding the procedure described so far into a branch-and-bound scheme for (15) seen as a nonconvex global optimization problem.

In contrast to the expectation-based model the objective function Q_{CVaR_α} is not explicitly given and therefore in some cases additional computational effort is needed: If the scenario solutions (x_j, η_j) , $j = 1, \dots, J$, differ and we determine by some heuristics a suggestion $(\bar{x}^R, \bar{\eta}^R)$, then $\bar{\eta}^R$ is not necessarily a minimizer for $Q_{CVaR_\alpha}(\bar{x}^R)$ in (12), thus yielding just an upper bound on $Q_{CVaR_\alpha}(\bar{x}^R)$. However, by Proposition 4.6 it holds that $f(\alpha, \cdot, x)$ is convex for all $(\alpha, x) \in (0, 1) \times \mathbb{R}^m$. So, we can determine a minimizer η , and thus $Q_{CVaR_\alpha}(\bar{x}^R)$, by solving a one-dimensional piecewise-linear convex, or equivalently, the following $(J+1)$ -dimensional linear optimization problem:

$$Q_{CVaR_\alpha}(\bar{x}^R) = \min \left\{ \eta + \frac{1}{1-\alpha} \sum_{j=1}^J \pi_j v_j : v_j \geq c^\top \bar{x}^R + \bar{\Phi}_j - \eta, v_j \in \mathbb{R}_+, \right. \\ \left. j = 1, \dots, J, \eta \in \mathbb{R} \right\} \quad (21)$$

where $\bar{\Phi}_j := \min\{q^\top y_j + q'^\top y'_j : W y_j + W' y'_j = h_j - T x_j, y_j \in \mathbb{Z}_+^{\bar{m}}, y'_j \in \mathbb{R}_+^{m'}, x_j \in X, x_j = \bar{x}^R\}$. Note, that if $\bar{\Phi}_j < +\infty$ for all $j = 1, \dots, J$, the suggested first-stage solution \bar{x}^R is feasible for all scenario subproblems. Otherwise, if $\bar{\Phi}_j = +\infty$ for some j , \bar{x}^R can be discarded.

If the scenario solutions (x_j, η_j) , $j = 1, \dots, J$, are identical, therefore feasible for the unrelaxed problem (16), then η_j is a minimizer corresponding to x_j and so $Q_{CVaR_\alpha}(x_j) = \eta_j + \frac{1}{1-\alpha} \sum_{j=1}^J \pi_j v_j$, where the v_j are solutions to the scenario subproblems with fixed x_j, η_j .

Let \mathbf{P} denote the list of current problems and $z_{LD} = z_{LD}(P)$ the Lagrangian lower bound for $P \in \mathbf{P}$. The algorithm then proceeds as follows.

Algorithm 5.2.

Step 1 (Initialization): Set $\bar{z} = +\infty$ and let \mathbf{P} consist of problem (17).

Step 2 (Termination): If $\mathbf{P} = \emptyset$ then the solution \hat{x} that yielded $\bar{z} = Q_{CVaR_\alpha}(\hat{x})$ is optimal.

Step 3 (Node selection): Select and delete a problem P from \mathbf{P} and solve its Lagrangian dual. If the optimal value $z_{LD}(P)$ hereof equals $+\infty$ (infeasibility of a subproblem) then go to Step 2.

Step 4 (Bounding): If $z_{LD}(P) \geq \bar{z}$ go to Step 2 (this step can be carried out as soon as the value of the Lagrangian dual rises above \bar{z}). Consider the following situations:

1. The scenario solutions (x_j, η_j) , $j = 1, \dots, J$, are identical: If $Q_{CVaR_\alpha}(x_j) < \bar{z}$ then let $\bar{z} = Q_{CVaR_\alpha}(x_j)$. Go to Step 2.
2. The scenario solutions (x_j, η_j) , $j = 1, \dots, J$ differ: Compute the average $\bar{x} = \sum_{j=1}^J \pi_j x_j$ and round it by some heuristic to obtain \bar{x}^R . Calculate $\bar{\Phi}_j$, $j = 1, \dots, J$. If, for some j , $\bar{\Phi}_j = +\infty$ then go to Step 5, otherwise solve (21) and obtain the value $Q_{CVaR_\alpha}(\bar{x}^R)$. If $Q_{CVaR_\alpha}(\bar{x}^R) < \bar{z}$ then let $\bar{z} = Q_{CVaR_\alpha}(\bar{x}^R)$. Go to Step 5.

Step 5 (Branching): Select a component $x_{(k)}$ of x and add two new problems to \mathbf{P} obtained from P by adding the constraints $x_{(k)} \leq \lfloor \bar{x}_{(k)} \rfloor$ and $x_{(k)} \geq \lfloor \bar{x}_{(k)} \rfloor + 1$, respectively (if $x_{(k)}$ is an integer component), or $x_{(k)} \leq \bar{x}_{(k)} - \varepsilon$ and $x_{(k)} \geq \bar{x}_{(k)} + \varepsilon$, respectively, where $\varepsilon > 0$ is a tolerance parameter to have disjoint subdomains. Go to Step 3.

Since X is bounded, and if all x -components are restricted to be integers, the algorithm is obviously finite. If x is mixed-integer some stopping criterion to avoid endless branching on the continuous components has to be employed.

Apart from the adaptations due to the implicit nature of the representation of Q_{CVaR_α} the algorithm follows the same lines as the algorithm for $\min\{Q_E(x) : x \in X\}$ developed in [12]. In a straightforward manner this leads to a scenario decomposition algorithm for the mean-risk model $\min\{Q_E(x) + \rho Q_{CVaR_\alpha}(x) : x \in X\}$. Using Proposition 5.1 this problem can be equivalently restated as

$$\begin{aligned} \min_{x, y, y', v, \eta} \left\{ \sum_{j=1}^J \pi_j (c^\top x + q^\top y_j + q'^\top y'_j) + \rho(\eta + \frac{1}{1-\alpha} \sum_{j=1}^J \pi_j v_j) : \right. \\ \left. W y_j + W' y'_j = h_j - T x, \quad v_j \geq c^\top x + q^\top y_j + q'^\top y'_j - \eta, \quad (22) \right. \\ \left. x \in X, \quad \eta \in \mathbb{R}, \quad y_j \in \mathbb{Z}_+^{\bar{m}}, \quad y'_j \in \mathbb{R}_+^{m'}, \quad v_j \in \mathbb{R}_+, \quad j = 1, \dots, J \right\}. \end{aligned}$$

As described in Section 1, by solving this problem for various values of $\rho > 0$, we are able to obtain supported efficient points of the corresponding multiobjective optimization problem $\min\{(Q_E(x), Q_{CVaR_\alpha}(x)) : x \in X\}$. As a systematic procedure for choosing the values of ρ we follow the tracing method described in [33] which will be outlined exemplarily at a specific problem instance below.

We have run some first numerical experiments with an optimization problem from chemical engineering. It is about a planning model of a real-life multi-product batch plant producing expandable polystyrene (EPS). A detailed description of the EPS process can be found in [13].

The process consists of three production stages: preparation, polymerization, and finishing. In the preparation stage various kinds of intermediates are produced. Depending on a finite number of recipes certain mixtures of the intermediates are fed batch-wise into the polymerization reactors. Upon completion of each polymerization its product is transferred without delay into a mixing tank of a finishing line, yielding a discontinuous inflow into these tanks. Further, each finishing line consists of a separation stage where different grain sizes of EPS are separated from each other. These grain sizes are the final products whose amounts have to comply with customer demands. Shut-down and start-up procedures for the continuously driven separation stages are time consuming, expensive, and are restricted by minimum up- and down-times of the stages.

The process is controlled by determining starting times and choices of recipes for the polymerizations and by choosing start-up and shut-down times as well as feed rates for the separation stages. A typical planning horizon is given by two weeks, with the time period discretized into five equidistant intervals. The essential source of uncertainty is customer demand. The objective function to be minimized is a weighted sum of costs coming from running the polymerizations, switching the separation stages, and compensating deficit between production and customer demand.

The described model gives rise to a multitude of two-stage stochastic integer programs, cf. [13]. For our numerical experiments we have formulated a planning model with the first-stage variables being the states of the separation stages. This focuses on the qualitative aspect that a smooth operation of the EPS process is desired, which is achieved by fixing the states of the most sensible part of the plant as early as possible.

Table 1 reports the tracing of the supported part of the efficient frontier of four EPS problem instances, using 10, 20, 50 or 100 scenarios to describe the uncertain data. The probability level α was set to 0.7 for all computations. The ρ column shows which weight factor was used for the specific instance of the mean-risk model, where $\rho = 0$ denotes the traditional expectation-based model, and $\rho = +\infty$ the model using just Conditional Value-at-Risk as the objective function. For each instance the size of the problem $\min\{Q_{\mathcal{E}}(x) + \rho Q_{CVaR_{0.7}}(x) : x \in X\}$ is displayed in the columns Cont/Int/Bin and Constraints, showing the number of continuous, integer (nonbinary), binary variables and constraints respectively. The Time column shows the time needed by our decomposition algorithm to obtain an optimal solution with a quality certificate of a relative gap less than 1%. For comparison, we employed directly CPLEX 8.1.1 (with default parameters) to tackle the large-scale mixed-integer program (22). The CPLEX column contains the relative gap achieved by CPLEX within the time displayed in the Time column. Finally, the $(Q_{\mathcal{E}}, Q_{CVaR_{0.7}})$ column contains the $Q_{\mathcal{E}}(\bar{x})$ and $Q_{CVaR_{0.7}}(\bar{x})$ values, rounded to the second decimal place, where

Table 1. Computational results for the EPS problem

Scenarios	Cont/Int/Bin	Constraints	ρ	$(Q_E, Q_{CVaR_{0.7}})$	Time (h:mm)	CPLEX
10	500/400/352	1370	0	(114.81, 135.33)	0:03	3.24%
	511/400/352	1380	0.001	(114.81, 135.33)	0:18	2.97%
	511/400/352	1380	0.987	(115.86, 134.26)	0:27	2.11%
	511/400/352	1380	2.044	(115.86, 134.26)	0:23	2.38%
	511/400/352	1380	2.414	(123.20, 131.22)	0:23	2.54%
	511/400/352	1380	1000	(123.20, 131.22)	0:22	2.72%
	511/400/352	1380	$+\infty$	(123.20, 131.22)	0:23	2.25%
20	1000/800/692	2740	0	(120.50, 151.53)	0:06	4.86%
	1021/800/692	2760	0.001	(120.50, 151.53)	0:36	4.61%
	1021/800/692	2760	0.35	(120.50, 151.53)	1:46	4.90%
	1021/800/692	2760	1000	(125.20, 138.11)	0:35	4.28%
	1021/800/692	2760	$+\infty$	(125.20, 138.11)	2:07	3.79%
	2500/2000/1712	6850	0	(124.00, 156.52)	0:19	9.07%
50	2551/2000/1712	6900	0.001	(124.00, 156.52)	1:56	4.86%
	2551/2000/1712	6900	0.093	(125.79, 137.30)	2:56	4.68%
	2551/2000/1712	6900	1000	(125.79, 137.30)	1:58	6.19%
	2551/2000/1712	6900	$+\infty$	(125.79, 137.30)	7:15	4.44%
	5000/4000/3412	13700	0	(122.10, 153.94)	0:36	8.42%
100	5101/4000/3412	13800	0.001	(122.10, 153.94)	4:43	5.05%
	5101/4000/3412	13800	0.203	(125.45, 137.41)	6:04	6.30%
	5101/4000/3412	13800	1000	(125.45, 137.41)	3:55	4.97%
	5101/4000/3412	13800	$+\infty$	(125.45, 137.41)	15:36	5.34% ¹

\bar{x} is the obtained efficient point. For the computations we have used a Linux PC with an Intel Pentium 4 CPU 3.00 GHz and 2 GB RAM.

We explain the tracing of the supported part of the efficient frontier exemplarily at the 10 scenario instance. The first computation carried out is the traditional expectation-based model ($\rho = 0$). The obtained solution is not necessarily an efficient point, since the second objective is not minimized explicitly. However, now we know that the obtained Q_E value 114.81 is the smallest possible. So, if we can find a $\rho > 0$ that yields an efficient point with this Q_E value, we can be sure that there is no efficient point with a smaller Q_E value. This is done with the computation using $\rho = 0.001$. The same procedure has been done concerning $Q_{CVaR_{0.7}}$ with the computations $\rho = +\infty$ and $\rho = 1000$. Knowing both outer efficient points, (114.81, 135.33) and (123.20, 131.22), we are able to calculate the ρ that yields the biggest chance of finding another efficient point, or proves, of course, depending on the accuracy of our computations, that there is no other supported efficient point with a different objective value vector from what we already have: The new ρ is calculated such that $(1, \rho)$ is the normal vector of the straight line passing through the points (114.81, 135.33) and (123.20, 131.22). For this $\rho = 2.044$ we obtain a new efficient point having the objective value vector (115.86, 134.26). This process is iterated for the lower and upper parts until no efficient point with a new objective value vector can be found. This is the case for the computations $\rho = 0.987$ and $\rho = 2.414$, thus terminating the tracing procedure.

We conclude that our decomposition algorithm is suitable to trace the supported part of the efficient frontier of the given nonconvex multiobjective optimization problem.

¹ CPLEX termination by “Out of memory” after 5:35 time with the displayed relative gap

Acknowledgements. We wish to thank Christoph Helmberg (Technical University of Chemnitz) and Andreas Märkert (University Duisburg-Essen) for giving us access to their implementations of the conic bundle and the scenario decomposition method, respectively.

References

1. Acerbi, C., Tasche, D.: Expected Shortfall: a natural coherent alternative to Value at Risk. *Econ. Notes* **31**, 379–388 (2002)
2. Acerbi, C., Tasche, D.: On the coherence of Expected Shortfall. *Journal of Banking and Finance* **26**, 1487–1503 (2002)
3. Artstein, Z., Wets, R.J-B: Stability results for stochastic programs and sensors, allowing for discontinuous objective functions. *SIAM J. Optim.* **4**, 537–550 (1994)
4. Artzner, P., Delbaen, F., Eber, J.-M., Heath, D.: Coherent measures of risk. *Math. Finance* **9**, 203–228 (1999)
5. Bank, B., Guddat, J., Klatte, D., Kummer, B., Tammer, K.: *Non-linear Parametric Optimization* (Akademie-Verlag, Berlin 1982)
6. Bank, B., Mandel, R.: *Parametric Integer Optimization* (Akademie-Verlag, Berlin 1988)
7. Benson, H.P.: Multi-objective optimization: Pareto optimal solutions, properties. In: C.A. Floudas, P.M. Pardalos (eds.) *Encyclopedia of Optimization*, Vol. III (Kluwer, Dordrecht 2001) 489–493
8. Billingsley, P.: *Convergence of Probability Measures* (Wiley, New York 1968)
9. Birge, J.R., Louveaux, F.: *Introduction to Stochastic Programming* (Springer, New York 1997)
10. Birge, J.R., Wets, R.J-B: Designing approximation schemes for stochastic optimization problems, in particular for stochastic programs with recourse. *Math. Program. Study* **27**, 54–102 (1986)
11. Blair, C.E., Jeroslow, R.G.: The value function of a mixed integer program: I. *Discrete Math.* **19**, 121–138 (1977)
12. Carøe, C.C., Schultz, R.: Dual decomposition in stochastic integer programming. *Operations Res. Lett.* **24**, 37–45 (1999)
13. Engell, S., Märkert, A., Sand, G., Schultz, R., Schulz, Ch.: Online scheduling of multi-product batch plants under uncertainty. In: M. Grötschel, S.O. Krumke, J. Rambau (eds.) *Online Optimization of Large Scale Systems* (Springer, Berlin 2001) 649–676
14. Helmberg, C., Kiwiel, K.C.: A spectral bundle method with bounds. *Math. Programming* **93**, 173–194 (2002)
15. ILOG CPLEX, User’s Manual (ILOG, Inc., Mountain View, CA 2002) Information available online from <http://www.cplex.com>.
16. Kall, P., Ruszczyński, A., Frauendorfer, K.: Approximation techniques in stochastic programming. In: Yu. Ermoliev, R.J-B Wets (eds.) *Numerical Techniques for Stochastic Optimization* (Springer, Berlin 1988) 33–64
17. Kall, P., Wallace, S.W.: *Stochastic Programming* (Wiley, Chichester, 1994)
18. Kiwiel, K. C.: Proximity control in bundle methods for convex nondifferentiable optimization. *Math. Programming* **46**, 105–122 (1990)
19. Kiwiel, K. C.: User’s Guide for NOA 2.0/3.0: A Fortran Package for Convex Nondifferentiable Optimization (Systems Research Institute, Polish Academy of Sciences, Warsaw 1994)
20. Korhonen, P.: Multiple objective programming support. In: C.A. Floudas, P.M. Pardalos (eds.) *Encyclopedia of Optimization*, Vol. III (Kluwer, Dordrecht 2001) 566–574
21. Markowitz, H.M.: Portfolio selection. *Journal of Finance* **7**, 77–91 (1952)
22. Nemhauser, G.L., Wolsey, L.A.: *Integer and Combinatorial Optimization* (Wiley, New York 1988)
23. Ogryczak, W., Ruszczyński, A.: From stochastic dominance to mean-risk models: Semideviations as risk measures. *European J. Oper. Res.* **116**, 33–50 (1999)
24. Ogryczak, W., Ruszczyński, A.: Dual stochastic dominance and related mean-risk models. *SIAM J. Optim.* **13**, 60–78 (2002)
25. Pflug, G.C.: Some remarks on the Value-at-Risk and the Conditional Value-at-Risk. In: S. Uryasev (ed.) *Probabilistic Constrained Optimization: Methodology and Applications* (Kluwer, Dordrecht 2000) 272–281
26. Pollard, D.: *Convergence of Stochastic Processes* (Springer, New York 1984)

27. Prékopa, A.: Stochastic Programming (Kluwer, Dordrecht 1995)
28. Robinson, S.M.: Local epi-continuity and local optimization. *Math. Programming* **37**, 208–222 (1987)
29. Rockafellar, R.T., Uryasev, S.: Optimization of Conditional Value-at-Risk. *Journal of Risk* **2**, 21–41 (2000)
30. Rockafellar, R.T., Uryasev, S.: Conditional Value-at-Risk for General Loss Distributions. *Journal of Banking & Finance* **26**, 1443–1471 (2002)
31. Schultz, R.: On structure and stability in stochastic programs with random technology matrix and complete integer recourse. *Math. Programming* **70**, 73–89 (1995)
32. Schultz, R.: Rates of convergence in stochastic programs with complete integer recourse. *SIAM J. Optim.* **6**, 1138–1152 (1996)
33. Schultz, R., Tiedemann, S.: Risk Aversion via Excess Probabilities in Stochastic Programs with Mixed-Integer Recourse. *SIAM J. Optim.* **14**, 115–138 (2003)