Optimization with multivariate conditional value-at-risk constraints

Nilay Noyan and Gábor Rudolf
Manufacturing Systems/Industrial Engineering Program, Sabancı University, 34956 Istanbul, Turkey,
nnoyan@sabanciuniv.edu and grudolf@sabanciuniv.edu

Abstract: For many decision making problems under uncertainty, it is crucial to develop risk-averse models and specify the decision makers' risk preferences based on multiple stochastic performance measures (or criteria). Incorporating such multivariate preference rules into optimization models is a fairly recent research area. Existing studies focus on extending univariate stochastic dominance rules to the multivariate case. However, enforcing multivariate stochastic dominance constraints can often be overly conservative in practice. As an alternative, we focus on the widely-applied risk measure conditional value-at-risk (CVaR), introduce a multivariate CVaR relation, and develop a novel optimization model with multivariate CVaR constraints based on polyhedral scalarization. To solve such problems for finite probability spaces we develop a cut generation algorithm, where each cut is obtained by solving a mixed integer problem. We show that a multivariate CVaR constraint reduces to finitely many univariate CVaR constraints, which proves the finite convergence of our algorithm. We also show that our results can be naturally extended to a wider class of coherent risk measures. The proposed approach provides a flexible, and computationally tractable way of modeling preferences in stochastic multi-criteria decision making. We conduct a computational study for a budget allocation problem to illustrate the effect of enforcing multivariate CVaR constraints and demonstrate the computational performance of the proposed solution methods.

Keywords: multivariate risk-aversion; conditional value-at-risk; multiple criteria; cut generation; coherent risk measures; stochastic dominance; Kusuoka representation

---

1. Introduction The ability to compare random outcomes based on the decision makers' risk preferences is crucial to modeling decision making problems under uncertainty. In this paper we focus on optimization problems that feature risk preference relations as constraints. Risk measures are functionals that represent the risk associated with a random variable by a scalar value, and provide a direct way to define such preferences. Popular risk measures include semi-deviations, quantiles (under the name *value-at-risk*), and *conditional value-at-risk* (CVaR). Desirable properties of risk measures, such as *law invariance* and *coherence*, are axiomatized in Artzner et al. (1999). CVaR, introduced by Rockafellar and Uryasev (2000), is a risk measure of particular importance which not only satisfies these axioms, but also serves as a fundamental building block for other law invariant coherent risk measures (as demonstrated by Kusuoka (2001)). Due to these attractive properties, univariate risk constraints based on CVaR have been widely incorporated into optimization models, primarily in a financial context (see, e.g., Uryasev, 2000; Rockafellar and Uryasev, 2002; Fabian and Veszpremi, 2008).

Relations derived from risk measures use a single scalar-valued functional to compare random outcomes. In contrast, stochastic dominance relations provide a well-established (Mann and Whitney, 1947; Lehmann, 1955) basis for more sophisticated comparisons; for a review on these and other comparison methods we refer the reader to Shaked and Shanthikumar (1994), Müller and Stoyan (2002), and the references therein. In particular, the second-order stochastic dominance (SSD) relation has been receiving significant attention due its correspondence with risk-averse preferences. Dentcheva and Ruszczyński (2003) have proposed to incorporate such relations into optimization problems as constraints, requiring the decision-based random outcome to stochastically dominate some benchmark random outcome. Recently, such optimization models with univariate stochastic dominance constraints have been studied, among others, by Luedtke (2008); Noyan et al. (2008); Noyan and Ruszczyński (2008); Rudolf and Ruszczyński (2008); Gollmer et al. (2011), and they have been applied to various areas including financial portfolio optimization (see, e.g., Dentcheva and Ruszczyński, 2006), emergency service system design (Noyan, 2010), power planning (see, e.g., Gollmer et al., 2008), and optimal path problems (Nie et al., 2011).

For many decision making problems, it may be essential to consider multiple random outcomes of interest. In contrast to the scalar-based comparisons mentioned above, such a multi-criteria (or multi-objective) approach requires specifying preference relations among random vectors, where each dimension of a vector corresponds to a decision criterion. This is usually accomplished by extending scalar-based preferences to vector-valued random variables. Incorporating multivariate preference rules as constraints into optimization models is a fairly recent research area, focusing on problems of the general form

$$\begin{aligned} \max \quad & f(\mathbf{z}) \\ \text{s.t.} \quad & G(\mathbf{z}) \succcurlyeq \mathbf{Y} \\ & \mathbf{z} \in Z. \end{aligned}$$

Here $G(\mathbf{z})$ is the random *outcome vector* associated with the decision variable $\mathbf{z}$ according to some *outcome mapping* G, the relation $\succcurlyeq$ represents multivariate preferences, and $\mathbf{Y}$ is a *benchmark* (or *reference*) random outcome vector. A key idea in this line of research, initiated by the work of Dentcheva and Ruszczyński (2009), is to consider a family of scalarization functions and require that the scalarized versions of the random variables conform to some scalar-based preference relation. In case of linear scalarization, one can interpret scalarization coefficients as the weights representing the subjective importance of each criterion. However, in many decision making situations, especially those involving multiple decision makers, it can be difficult to exactly specify a single scalarization. In such cases one can enforce the preference relation over a given set of weights representing a wider range of views.

Dentcheva and Ruszczyński (2009) consider linear scalarization with positive coefficients and apply a univariate SSD dominance constraint to all nonnegative weighted combinations of random outcomes, leading to the concept of *positive linear SSD*. They provide a solid theoretical background and develop duality results for this problem, while Homem-de-Mello and Mehrotra (2009) propose a cutting surface method to solve a related class of problems. The latter study considers only finitely supported random variables under certain linearity assumptions, but the set of scalarization coefficients is allowed to be an arbitrary polyhedron. However, their method is computationally demanding as it typically requires solving a large number of non-convex cut generation problems. Hu et al. (2010) introduce an even more general concept of dominance by allowing arbitrary convex scalarization sets, and apply a sample average approximation-based solution method. Not all notions of multivariate stochastic dominance rely on scalarization functions. Armbruster and Luedtke (2010) consider optimization problems constrained by first and second order stochastic dominance relations based on multi-dimensional utility functions (see, e.g., Müller and Stoyan, 2002).

As we have seen, the majority of existing studies on optimization models with multivariate risk-averse preference relations focus on extending univariate stochastic dominance rules to the multivariate case. However, this approach typically results in very demanding constraints that can be excessively hard to satisfy in practice, and sometimes even lead to infeasible problems. For example, Hu et al. (2011b) solve a multivariate SSD-constrained homeland security budget allocation problem, and ensure feasibility by introducing a tolerance parameter into the SSD constraints. Other attempts to weaken stochastic dominance relations in order to extend the feasible region have resulted in concepts such as *almost stochastic dominance* (Leshno and Levy, 2002; Lizyayev and Ruszczyński, 2011) and *stochastically weighted stochastic dominance* (Hu et al., 2011a).

In this paper we propose an alternative approach, where stochastic dominance relations are replaced by a collection of *conditional value-at-risk* (CVaR) constraints at various confidence levels. This is a very natural relaxation, due to the well known fact that the univariate SSD relation is equivalent to a continuum of CVaR inequalities (Dentcheva and Ruszczyński, 2006). Furthermore, compared to methods directly based on dominance concepts, the the ability to specify confidence levels allows a significantly higher flexibility to express decision makers' risk preferences. At the extreme ends of the spectrum CVaR-based constraints can express both risk-neutral and worst case-based decision rules, while SSD relations can be approximated (and even exactly modeled) by simultaneously enforcing CVaR inequalities at multiple confidence levels. Comparison between random vectors is achieved by means of a polyhedral scalarization set, along the lines of Homem-de-Mello and Mehrotra (2009), leading to multivariate *polyhedral* CVaR *constraints*. We remark that this concept is not directly related to the risk measure introduced under the name "multivariate CVaR" by Prékopa (2012), defined as the conditional expectation of a scalarized random vector. To the best of our knowledge, incorporating the risk measure CVaR is a first for optimization problems with multivariate preference relations based on a set of scalarization weights.

The contributions of this study are as follows.

- We introduce a new multivariate risk-averse preference relation based on CVaR and linear scalarization.
- We develop a modeling approach for multi-criteria decision making under uncertainty featuring multivariate CVaR-based preferences.
- We develop a finitely convergent cut generation algorithm to solve polyhedral CVaR-constrained optimization problems. Under linearity assumptions we provide explicit formulations of the master problem as a linear program, and of the cut generation problem as a mixed integer linear program.
- We provide a theoretical background to our formulations, including duality results. We also show that on a finite probability space a polyhedral CVaR constraint can be reduced to a finite number of univariate CVaR inequalities. This important result, which is used to prove finite convergence of our cut generation algorithm, is then extended to polyhedral constraints based on a wider class of coherent risk measures.
- We adapt and extend some existing results from the theory of risk measures to fit the framework of our problems, as necessary. In particular, we prove the equivalence of relaxed SSD relations to a continuum of relaxed CVaR constraints, and show that for finite probability spaces this continuum can be reduced to a finite set. We also provide a form of Kusuoka's representation theorem for coherent risk measures which does not require the underlying probability space to be atomless.
- In a small-scale numerical study we examine the feasible regions associated with various polyhedral CVaR constraints, and compare them to their SSD-based counterparts. We also conduct a comprehensive computational study of a budget allocation problem, previously explored in Hu et al. (2011b), to evaluate the effectiveness of our proposed model and solution methods.

The rest of the paper is organized as follows. In Section 2 we review fundamental concepts related to CVaR, SSD, and linear scalarization. Then we define multivariate CVaR relations, and present a general form of optimization problems involving such relations as constraints. Section 3 contains theoretical results including optimization representations of CVaR, and finite representations of polyhedral CVaR and SSD constraints. In Section 4 we provide a linear programming formulation and duality results under certain linearity assumptions. In Section 5 we generalize our finite representation results to a class of coherent risk measures, extending Kusuoka's representation theorem to non-atomless measures in the process. In Section 6 we briefly discuss a vertex enumeration-based solution approach, then proceed to present a detailed description of a cut generation algorithm, and prove its correctness and finite convergence. Section 7 is dedicated to numerical results, while Section 8 contains concluding remarks.

2. Basic concepts and fundamental results In this section we aim to introduce a stochastic optimization framework for multi-objective (multi-criteria) decision making problems where the decision leads to a vector of random outcomes which is required to be preferable to a reference random outcome vector. We begin by discussing some widely used risk measures and associated relations which can be used to establish preferences between scalar-valued random variables. We also recall and generalize some fundamental results on the connections between these relations. Next, we extend these relations to vector-valued random variables, and present a general form of optimization problems involving them as constraints.

REMARK 2.1 *Throughout our paper larger values of random variables are considered to be preferable. In the literature the opposite convention is also widespread, especially in the context of loss functions. When citing such sources, the definitions and formulas are altered to reflect this difference.*

2.1 VaR, CVaR, and second order stochastic dominance We now present some basic definitions and results related to the risk measure CVaR. Unless otherwise specified, all random variables in this paper are assumed to be in $\mathcal{L}_1$, which ensures that the following definitions and formulas are valid. For a more detailed exposition on the concepts described below we refer to Pflug and Römisch (2007) and Rockafellar (2007).

- Let V be a random variable with a cumulative distribution function (CDF) denoted by F_V. The *value-at-risk* (VaR) at confidence level $\alpha \in (0, 1]$, also known as the α-quantile, is defined as

$$\mathrm{VaR}_\alpha(V) = \inf\{\eta \ : \ F_V(\eta) \geq \alpha\}. \tag{1}$$

- The *conditional value-at-risk* at confidence level α is defined (Rockafellar and Uryasev, 2000) as

$$\mathrm{CVaR}_\alpha(V) = \sup\left\{\eta - \frac{1}{\alpha}\mathbb{E}\left([\eta - V]_+\right) \ : \ \eta \in \mathbb{R}\right\}, \tag{2}$$

  where $[x]_+ = \max(x, 0)$ denotes the *positive part* of a number $x \in \mathbb{R}$.

- It is well-known (Rockafellar and Uryasev, 2002) that if $\mathrm{VaR}_\alpha(V)$ is finite, the supremum in the above definition is attained at $\eta = \mathrm{VaR}_\alpha(V)$, i.e.,

$$\mathrm{CVaR}_\alpha(V) = \mathrm{VaR}_\alpha(V) - \frac{1}{\alpha}\mathbb{E}\left([\mathrm{VaR}_\alpha(V) - V]_+\right). \tag{3}$$

- CVaR is also known in the literature as *average value-at-risk* and *tail value-at-risk*, due to the following expression:

$$\mathrm{CVaR}_\alpha(V) = \frac{1}{\alpha}\int_0^\alpha \mathrm{VaR}_\gamma(V)\,\mathrm{d}\gamma. \tag{4}$$

We note that the *expected shortfall* term $\mathbb{E}\left([\eta - V]_+\right)$ introduced in (2) is closely related to the the *second order distribution function* $F_{2,V} : \mathbb{R} \to \mathbb{R}$ of the random variable V defined by

$$F_{2,V}(\eta) = \int_{-\infty}^{\eta} F_V(\xi)\,\mathrm{d}\xi.$$

Using integration by parts we obtain the following well-known equality:

$$\begin{aligned} F_{2,V}(\eta) &= \int_{-\infty}^{\eta} F_V(\xi)\,\mathrm{d}\xi = \eta F_V(\eta) - \int_{-\infty}^{\eta} \xi\,\mathrm{d}F_V(\xi) = \int_{-\infty}^{\eta} (\eta - \xi)\,\mathrm{d}F_V(\xi) = \int_{-\infty}^{\infty} [\eta - \xi]_+\,\mathrm{d}F_V(\xi) \\ &= \mathbb{E}\left([\eta - V]_+\right). \end{aligned} \tag{5}$$

CVaR is a widely used risk measure with significant advantages over VaR, due to a number of useful properties. For example, in contrast to VaR, the risk measure CVaR_α is *coherent* (Pflug, 2000), and serves as a fundamental building block for other coherent measures (see Section 5 for more details).

Furthermore, for a given random variable V the mapping $\alpha \mapsto \mathrm{CVaR}_\alpha$ is continuous and non-decreasing. CVaR can be used to express a wide range of risk preferences, including risk neutral (for $\alpha = 1$) and pessimistic worst-case (for sufficiently small values of α) approaches. We now introduce notation to express some risk preference relations associated with CVaR.

- Let V_1 and V_2 be two random variables with respective CDFs F_{V_1} and F_{V_2}. We say that V_1 is CVaR-*preferable* to V_2 at confidence level α, denoted as $V_1 \succcurlyeq_{\mathrm{CVaR}_\alpha} V_2$, if

$$\mathrm{CVaR}_\alpha(V_1) \geq \mathrm{CVaR}_\alpha(V_2). \tag{6}$$

- We say that V_1 is *second-order stochastically dominant* over V_2 (or that V_1 *dominates* V_2 *in the second order*), denoted as $V_1 \succcurlyeq_{(2)} V_2$, if $F_{2,V_1}(\eta) \leq F_{2,V_2}(\eta)$ holds for all $\eta \in \mathbb{R}$.

Remark 2.2 *According to* (3) *one can view* $\mathrm{CVaR}_\alpha(V)$ *as the expected value of* $U_V(V)$*, where* $U_V(t) = \mathrm{VaR}_\alpha(V) - \frac{1}{\alpha}[\mathrm{VaR}_\alpha(V) - t]_+$ *is a* probability-dependent utility function *(Street, 2009). In this context the relation* (6) *can be interpreted in terms of expected utilities as*

$$\mathbb{E}(U_{V_1}(V_1)) \geq \mathbb{E}(U_{V_2}(V_2)).$$

We proceed by examining the close connection between CVaR-preferability and second-order stochastic dominance (SSD) relations. In preparation let us recall some basic definitions and facts from the theory of conjugate duality (for a good overview see Rockafellar (1970)).

- Denoting the *extended real line* by $\overline{\mathbb{R}} = \mathbb{R} \cup \{\infty\}$, the *Fenchel conjugate* of a function $f : \mathbb{R} \to \overline{\mathbb{R}}$ is the mapping $f^* : \mathbb{R} \to \overline{\mathbb{R}}$ defined by $f^*(\alpha) = \sup\{\alpha\eta - f(\eta) \ : \ \eta \in \mathbb{R}\}$.
- For a constant ι, the conjugate of $f + \iota$ is given by $f^* - \iota$.
- Conjugation is *order-reversing*: if the relation $f_1(\eta) \leq f_2(\eta)$ holds for all $\eta \in \mathbb{R}$, then $f_2^*(\alpha) \leq f_1^*(\alpha)$ holds for all $\alpha \in \mathbb{R}$.
- **Fenchel-Moreau theorem**: If f is lower semi-continuous and convex, then it is equal to its *biconjugate*, i.e., $f^{**} = f$.

The first part of the proposition below is a well-known result (Dentcheva and Ruszczyński, 2006; Pflug and Römisch, 2007). Our proof of the more general second part uses a straightforward extension of the arguments in (Dentcheva and Ruszczyński, 2006).

Proposition 2.1 *Let* V_1 *and* V_2 *be two random variables with respective CDFs* F_{V_1} *and* F_{V_2}.

(*i*) *An SSD constraint is equivalent to the continuum of* CVaR*-constraints for all confidence levels* $\alpha \in (0, 1]$*, i.e.,*

$$V_1 \succcurlyeq_{(2)} V_2 \quad \Longleftrightarrow \quad \mathrm{CVaR}_\alpha(V_1) \geq \mathrm{CVaR}_\alpha(V_2) \ \ \text{for all } \alpha \in (0, 1].$$

(*ii*) *Let* $\iota \in \mathbb{R}_+$ *be a tolerance parameter. Then the relaxed SSD constraint*

$$F_{2,V_1}(\eta) \leq F_{2,V_2}(\eta) + \iota \qquad \text{for all } \eta \in \mathbb{R} \tag{7}$$

is equivalent to the continuum of relaxed CVaR *constraints given by*

$$\mathrm{CVaR}_\alpha(V_1) \geq \mathrm{CVaR}_\alpha(V_2) - \frac{\iota}{\alpha} \ \ \text{for all } \alpha \in (0, 1]. \tag{8}$$

Proof. Since (i) is a special case of (ii), it suffices to prove the latter. The second order distribution function $F_{2,V}$ of a random variable V is the integral of a monotone non-decreasing function, therefore it is continuous and convex. By the Fenchel-Moreau theorem it follows that both of the functions F_{2,V_1} and

$F_{2,V_2} + \iota$ are equal to their respective biconjugates. This implies, due to the order reversing property of conjugation, that the condition (7) is equivalent to

$$F^*_{2,V_1}(\alpha) \geq F^*_{2,V_2}(\alpha) - \iota \quad \text{ for all } \alpha \in \mathbb{R}. \tag{9}$$

According to (5) we have $F_{2,V}(\eta) = \mathbb{E}([\eta - V]_+)$. Taking into account (2) it is easy to verify that

$$F^*_{2,V}(\alpha) = \begin{cases} \infty & \alpha < 0 \\ 0 & \alpha = 0 \\ \alpha \, \text{CVaR}_\alpha(V) & \alpha \in (0,1] \\ \infty & \alpha > 1 \end{cases}$$

holds for any random variable V. Substituting into (9) our claim immediately follows. □

The mapping $\alpha \mapsto \alpha \, \text{CVaR}_\alpha(V)$, which appears as the Fenchel conjugate of $F_{2,V}$ in the previous proof, is a well-studied function, known in the literature under various names.

- The function $F_V^{(-1)} : (0,1] \to \mathbb{R}$ defined by $F_V^{(-1)}(\alpha) = \inf\{\eta \; : \; F_V(\eta) \geq \alpha\}$ is called the *first quantile function* (or simply *quantile function*) of the random variable V. Note that $\text{VaR}_\alpha(V) = F_V^{(-1)}(\alpha)$ holds by definition.
- The *second quantile function* $F_V^{(-2)} : (0,1] \to \mathbb{R}$ of V is defined as
$$F_V^{(-2)}(\alpha) = \int_0^\alpha F_V^{(-1)}(\gamma) \, \mathrm{d}\gamma = \int_0^\alpha \text{VaR}_\gamma(V) \, \mathrm{d}\gamma. \tag{10}$$
This function is also known in the literature as the *generalized Lorenz curve* and the *absolute Lorenz curve*. Somewhat confusingly, the latter term is sometimes used to refer to the mean centered second quantile function $F^{(-2)}_{V-\mathbb{E}(V)}$.
- According to (4) we have $F_V^{(-2)}(\alpha) = \alpha \, \text{CVaR}_\alpha(V)$ for all $\alpha \in (0,1]$. It follows that the inequality $\text{CVaR}_\alpha(V_1) \geq \text{CVaR}_\alpha(V_2)$ is equivalent to $F_{V_1}^{(-2)}(\alpha) \geq F_{V_2}^{(-2)}(\alpha)$, while the SSD relation $V_1 \succcurlyeq_{(2)} V_2$ is equivalent to the continuum of constraints $F_{V_1}^{(-2)}(\alpha) \geq F_{V_2}^{(-2)}(\alpha)$ for all $\alpha \in (0,1]$.

It is interesting to note that when the probability space is finite, the continuum of CVaR constraints in the first part of Proposition 2.1 can be reduced to a finite number of inequalities. We conclude this section by proving a more general form of this statement, using the properties of the second quantile function outlined above. Our proof relies on the following trivial observation.

Observation 2.1 *Let $f_1, f_2 : \mathbb{R} \to \mathbb{R}$ be affine functions and consider three real numbers $a \leq b \leq c$. If we have $f_1(a) \geq f_2(a)$ and $f_1(c) \geq f_2(c)$, then the inequality $f_1(b) \geq f_2(b)$ also holds.*

Proposition 2.2 *Consider two random variables V_1 and V_2 on the (not necessarily discrete) probability space $(\Omega, \mathcal{A}, \Pi)$, let $\mathcal{Q} = \{\Pi(S) \; : \; S \in \mathcal{A}, \; \Pi(S) > 0\}$ denote the set of all non-zero probabilities of events, and let $\iota \in \mathbb{R}_+$ be a tolerance parameter. If the relation $F_{V_1}^{(-2)}(\alpha) \geq F_{V_2}^{(-2)}(\alpha) + \iota$ holds for all $\alpha \in \mathcal{Q}$, then it holds for all $\alpha \in (0,1]$.*

Proof. Assume that $F_{V_1}^{(-2)}(\alpha) \geq F_{V_2}^{(-2)}(\alpha) + \iota$ holds for all $\alpha \in \mathcal{Q}$ and consider an arbitrary confidence level $\hat{\alpha} \in (0,1]$. Since the random variables V_1 and V_2 are measurable, the values

$$\alpha_- = \max_{i \in \{1,2\}} \Pi\left(V_i < \text{VaR}_{\hat{\alpha}}(V_i)\right) \qquad \text{and} \qquad \alpha_+ = \min_{i \in \{1,2\}} \Pi\left(V_i \leq \text{VaR}_{\hat{\alpha}}(V_i)\right)$$

both belong to the set $\mathcal{Q}$, therefore by our assumption we have

$$F_{V_1}^{(-2)}(\alpha_-) \geq F_{V_2}^{(-2)}(\alpha_-) + \iota \qquad \text{and} \qquad F_{V_1}^{(-2)}(\alpha_+) \geq F_{V_2}^{(-2)}(\alpha_+) + \iota. \tag{11}$$

Furthermore, by the definition of VaR the inequalities $\alpha_- \le \hat{\alpha} \le \alpha_+$ hold, and for any $\gamma \in (\alpha_-, \alpha_+]$, $i \in \{1,2\}$ we have $\mathrm{VaR}_\gamma(V_i) = \mathrm{VaR}_{\hat{\alpha}}(V_i)$. It follows that, according to the definition in (10), the functions $F_{V_1}^{(-2)}$ and $F_{V_2}^{(-2)} + \iota$ are both affine on the interval $[\alpha_-, \alpha_+]$, with respective slopes $\mathrm{VaR}_{\hat{\alpha}}(V_1)$ and $\mathrm{VaR}_{\hat{\alpha}}(V_2)$. Recalling (11), our claim immediately follows from Observation 2.1. □

Corollary 2.1 *If $(\Omega, 2^\Omega, \Pi)$ is a finite probability space, then there exists a* finite *set $\mathcal{Q} \subset (0,1]$ of confidence levels such that for any two random variables V_1, V_2 on $(\Omega, 2^\Omega, \Pi)$ the SSD relation $V_1 \succcurlyeq_{(2)} V_2$ is equivalent to the collection of* CVaR *inequalities*

$$\mathrm{CVaR}_\alpha(V_1) \ge \mathrm{CVaR}_\alpha(V_2) \quad \text{for all } \alpha \in \mathcal{Q}.$$

Furthermore, if all elementary events in $\Omega = \{\omega_1, \dots, \omega_n\}$ have equal probability, then the SSD relation $V_1 \succcurlyeq_{(2)} V_2$ is equivalent to

$$\mathrm{CVaR}_{\frac{k}{n}}(V_1) \ge \mathrm{CVaR}_{\frac{k}{n}}(V_2) \quad \text{for all } k = 1, \dots, n.$$

Proof. Let $\mathcal{Q}$ be the set defined in Proposition 2.2, and set $\iota = 0$. Note that $|\mathcal{Q}| < 2^{|\Omega|}$, and in the equal probability case $\mathcal{Q} = \left\{\frac{1}{n}, \dots, \frac{n}{n}\right\}$. Since by part (ii) of Proposition 2.1 the SSD relation $V_1 \succcurlyeq_{(2)} V_2$ is equivalent to the the continuum of CVaR-constraints for all confidence levels $\alpha \in (0,1]$, our result now immediately follows from Proposition 2.2. □

It is easy to see that by combining Proposition 2.2 with part (ii) of Proposition 2.1 one can obtain analogous finiteness results for the relaxed SSD and CVaR constraints introduced in (7)-(8).

In the next section we extend CVaR-based preferences to allow the comparison of random vectors.

2.2 Comparing random vectors via scalarization To be able to tackle multiple criteria we need to extend scalar-based preferences to vector-valued random variables. The key concept is to consider a family of *scalarization functions* and require that all scalarized versions of the random variables conform to some preference relation. In order to eventually obtain computationally tractable formulations, we restrict ourselves to linear scalarization functions.

Definition 2.1 *Let $\preccurlyeq$ be a preordering of scalar-valued random variables, and let $C \subset \mathbb{R}^d$ be a set of* scalarization vectors. *Given two d-dimensional random vectors $\mathbf{X}$ and $\mathbf{Y}$ we say that $\mathbf{X}$ is $\preccurlyeq$-preferable to $\mathbf{Y}$ with respect to C, denoted as $\mathbf{X} \succcurlyeq^C \mathbf{Y}$, if*

$$\mathbf{c}^T\mathbf{X} \succcurlyeq \mathbf{c}^T\mathbf{Y} \text{ for all } \mathbf{c} \in C.$$

Remark 2.3 *A natural way to compare two random vectors $\mathbf{X} = (X_1, \dots, X_d)$ and $\mathbf{Y} = (Y_1, \dots, Y_d)$ is by* coordinate-wise preference*: we say that $\mathbf{X}$ is preferable to $\mathbf{Y}$ if $X_l \succcurlyeq Y_l$ for all $l = 1, \dots, d$. It is easy to see that this is a special case of Definition 2.1 obtained with the choice $C = \{\mathbf{e}_1, \dots, \mathbf{e}_d\}$, where $\mathbf{e}_l = (0, \dots, 0, 1, 0, \dots, 0) \in \mathbb{R}^d$ is the unit vector with the 1 in the lth position. In addition, whenever $\{\mathbf{e}_1, \dots, \mathbf{e}_d\} \subset C$, preference with respect to C implies coordinate-wise preference. Notably, this is the case for the positive linear SSD relation mentioned below.*

An example of the type of preference rule introduced in Definition 2.1 has been suggested under the name *positive linear SSD* by Dentcheva and Ruszczyński (2009), with the choice $C = \mathbb{R}_+^d$, and $\preccurlyeq$ representing the SSD relation $\preccurlyeq_{(2)}$. Homem-de-Mello and Mehrotra (2009) generalize this approach by allowing the $C \subset \mathbb{R}^d$ to be an arbitrary polyhedron, leading to the concept of *polyhedral linear SSD*. Their idea is motivated by the observation that, by taking C to be a proper subset of the positive orthant, polyhedral dominance can be a significantly less restrictive constraint than positive linear dominance. This reflects a wider trend in recent literature suggesting that in a practical optimization context stochastic dominance relations are often excessively hard to satisfy. Attempts to weaken stochastic dominance relations in order to extend the feasible region have resulted in the study of concepts such as *almost stochastic dominance* and *stochastically weighted stochastic dominance* (Leshno and Levy, 2002; Lizyayev and Ruszczyński, 2011; Hu et al., 2011b). Recalling Proposition 2.1, another natural way to relax the stochastic dominance relation is to require CVaR-preferability only at certain confidence levels, as opposed to the full continuum of constraints. This motivates us to introduce a special case of Definition 2.1.

Definition 2.2 (Multivariate CVaR relation) *Let* $\mathbf{X}$ *and* $\mathbf{Y}$ *be two* d*-dimensional random vectors,* $C \subset \mathbb{R}^d$ *a set of scalarization vectors, and* $\alpha \in (0,1]$ *a specified confidence level. We say that* $\mathbf{X}$ *is* CVaR-preferable *to* Y *at confidence level* α *with respect to* C*, denoted as* $X \succcurlyeq^{C}_{\mathrm{CVaR}_\alpha} Y$*, if*

$$\mathrm{CVaR}_\alpha(\mathbf{c}^T X) \geq \mathrm{CVaR}_\alpha(\mathbf{c}^T Y) \quad \textit{for all } \mathbf{c} \in C. \tag{12}$$

In our following analysis we focus on CVaR-preferability with respect to polyhedral scalarization sets. We begin by proving a close analogue of Proposition 1 in Homem-de-Mello and Mehrotra (2009), which shows that in these cases we can assume without loss of generality that the polyhedron C is compact, i.e., a polytope. In preparation, we recall the representation of CVaR as a *distortion risk measure* (see, e.g., Pflug and Römisch, 2007):

$$\mathrm{CVaR}_\alpha(V) = -\int_{-\infty}^{0} g(F_V(\eta))\,\mathrm{d}\eta + \int_{0}^{\infty} \tilde{g}(1 - F_V(\eta))\,\mathrm{d}\eta \tag{13}$$

where $g : [0,1] \to [0,1]$ is the *distortion function* defined by $g(\gamma) = \min(\frac{\gamma}{\alpha}, 1)$, while $\tilde{g}(\gamma) = 1 - g(1-\gamma)$ denotes the *dual distortion function.*

Proposition 2.3 *Let* C *be a nonempty convex set, and let* $\tilde{C} = \{\mathbf{c} \in \mathrm{cl\ cone}(C) : \|\mathbf{c}\|_1 \leq 1\}$*, where* $\mathrm{cl\ cone}(C)$ *denotes the closure of the conical hull of the set* C*. Then, for any confidence level* $\alpha \in (0,1]$*, the relations* $\succcurlyeq^{C}_{\mathrm{CVaR}_\alpha}$ *and* $\succcurlyeq^{\tilde{C}}_{\mathrm{CVaR}_\alpha}$ *coincide.*

Proof. For any non-zero vector $\mathbf{c} \in C$ we have $\frac{\mathbf{c}}{\|\mathbf{c}\|_1} \in \tilde{C}$. Since CVaR is positive homogenous it immediately follows that, for any two random variables V_1 and V_2, the relation $V_1 \succcurlyeq^{\tilde{C}}_{\mathrm{CVaR}_\alpha} V_2$ implies $V_1 \succcurlyeq^{C}_{\mathrm{CVaR}_\alpha} V_2$. On the other hand, let us assume that $V_1 \succcurlyeq^{C}_{\mathrm{CVaR}_\alpha} V_2$ and consider a non-zero vector $\tilde{\mathbf{c}} = \sum_{i=1}^{k} \lambda_i \mathbf{c}_i \in \mathrm{cone}(C)$, where $\lambda_i > 0$ and $\mathbf{c}_i \in C$ for all $i = 1, \ldots, k$. Since C is convex, we have $\frac{\tilde{\mathbf{c}}}{\sum_{i=1}^{k} \lambda_i} \in \mathcal{C}$, implying

$$\mathrm{CVaR}_\alpha(\tilde{\mathbf{c}}^T V_1) \geq \mathrm{CVaR}_\alpha(\tilde{\mathbf{c}}^T V_2) \quad \text{for all } \tilde{\mathbf{c}} \in \mathrm{cone}(C). \tag{14}$$

Finally, let $\bar{\mathbf{c}}$ be a vector in $\tilde{C}$. Since $\tilde{C} \subset \mathrm{cl\ cone}(C)$, there exists a sequence $\{\mathbf{c}_k\} \subset \mathrm{cone}(C)$ such that $\mathbf{c}_k \to \bar{\mathbf{c}}$. It follows that, for $i = 1, 2$, the sequence $\{\mathbf{c}_k^T V_i\}$ converges pointwise to $\bar{\mathbf{c}}^T V_i$. As pointwise convergence implies convergence in distribution, this means that $F_{\mathbf{c}_k^T V_i}(\eta) \to F_{\bar{\mathbf{c}}^T V_i}(\eta)$ holds at all continuity points η of $F_{\bar{\mathbf{c}}^T V_i}$. Keeping in mind that the distortion functions g, $\tilde{g}$ are bounded and continuous, by applying the bounded convergence theorem to (13) we obtain $\mathrm{CVaR}_\alpha(\mathbf{c}_k^T V_i) \to \mathrm{CVaR}_\alpha(\bar{\mathbf{c}}^T V_i)$. Therefore (14) implies the inequality $\mathrm{CVaR}_\alpha(\bar{\mathbf{c}}^T V_1) \geq \mathrm{CVaR}_\alpha(\bar{\mathbf{c}}^T V_2)$, which proves our claim. □

2.3 Optimization with multivariate CVaR constraints Let $(\Omega, 2^\Omega, \Pi)$ be a finite probability space with $\Omega = \{\omega_1, \ldots, \omega_n\}$ and $\Pi(\omega_i) = p_i$. Consider a multi-criteria decision making problem where the decision variable $\mathbf{z}$ is selected from a feasible set Z, and associated random outcomes are determined by the outcome mapping $G : Z \times \Omega \to \mathbb{R}^d$. We introduce the following additional notation:

- For a given decision $\mathbf{z} \in Z$ the random outcome vector $G(\mathbf{z}) : \Omega \to \mathbb{R}^d$ is defined by $G(\mathbf{z})(\omega) = G(\mathbf{z}, \omega)$.

- For a given elementary event ω_i the mapping $g_i : Z \to \mathbb{R}^d$ is defined by $g_i(\mathbf{z}) = G(\mathbf{z}, \omega_i)$.

Let $f : Z \to \mathbb{R}$ be an objective function, $\mathbf{Y}$ a d-dimensional *benchmark random vector*, $C \subset \mathbb{R}^d$ a polytope of scalarization vectors, and $\alpha \in (0, 1]$ a confidence level. Our goal is to provide an explicit mathematical programming formulation and, in some cases, a computationally tractable solution method to problems of the following form.

$$\begin{aligned} \max \quad & f(\mathbf{z}) \\ \text{s.t.} \quad & G(\mathbf{z}) \succcurlyeq^C_{\mathrm{CVaR}_\alpha} \mathbf{Y} \\ & \mathbf{z} \in Z \end{aligned} \tag{GeneralP}$$

While the benchmark random vector can be defined on a probability space different from Ω, in practical applications it often takes the form $\mathbf{Y} = G(\bar{\mathbf{z}})$, where $\bar{\mathbf{z}} \in Z$ is a *benchmark decision*. For risk-averse decision makers typical choices for the confidence level are small values such as $\alpha = 0.05$.

In order to keep our exposition simple, in (GeneralP) we only consider a single CVaR constraint. However, all of our results and methods remain fully applicable for problems of the more general form

$$\begin{aligned} \max \quad & f(\mathbf{z}) \\ \text{s.t.} \quad & G(\mathbf{z}) \succcurlyeq^{C_{ij}}_{\mathrm{CVaR}_{\alpha_{ij}}} \mathbf{Y}_i \qquad i = 1, \dots, M, \quad j = 1, \dots, K_i \\ & \mathbf{z} \in Z, \end{aligned} \tag{15}$$

with CVaR constraints enforced for M multiple benchmarks, multiple confidence levels, and varying scalarization sets. In addition, constraints can be replaced by the relaxed versions introduced in (8). In Section 7.2.2 we present numerical results for a budget allocation problem featuring relaxed constraints on two benchmarks, enforced at up to 9 confidence levels for each. Even more generally, our approach can be naturally extended to include *mixed* CVaR *constraints* (Rockafellar, 2007) based on risk measures of the form $\varrho(V) = \lambda_1 \mathrm{CVaR}_{\alpha_1}(V) + \cdots + \lambda_r \mathrm{CVaR}_{\alpha_r}(V)$. The necessary theoretical background for the latter extension is laid out in Section 5, while formulation (56) provides a blueprint for combining CVaR at various confidence levels in a mathematical programming context.

3. Main theoretical results In this section we provide the theoretical background necessary to develop, and prove the finite convergence of, our solution methods. We begin by expressing CVaR as the optimum of various minimization and maximization problems, then proceed to prove that in finite probability spaces one can replace scalarization polyhedra by a finite set of scalarization vectors. To conclude the section, we show that this finiteness result extends to multivariate SSD constraints, providing an alternative to the representation in Homem-de-Mello and Mehrotra (2009).

3.1 Alternative expressions of CVaR By definition, CVaR can be obtained as a result of a maximization problem. On the other hand, CVaR is also a *spectral risk measure* (Acerbi, 2002) and thus can be viewed as a weighted sum of the least favorable outcomes. This allows us to express CVaR as the optimum of minimization problems.

Theorem 3.1 *Let V be a random variable with (not necessarily distinct) realizations $v_1, \dots, v_n$ and corresponding probabilities $p_1, \dots, p_n$. Then, for a given confidence level α the optimum values of the following optimization problems are all equal to* $\mathrm{CVaR}_\alpha(V)$.

(*i*)

$$\begin{aligned} \max \quad & \eta - \frac{1}{\alpha}\sum_{i=1}^{n} p_i w_i \\ \text{s.t.} \quad & w_i \geq \eta - v_i \qquad i = 1,\ldots,n \\ & w_i \geq 0 \qquad\qquad i = 1,\ldots,n \end{aligned} \tag{16}$$

(*ii*)

$$\begin{aligned} \min \quad & \frac{1}{\alpha}\sum_{i=1}^{n} \gamma_i v_i \\ \text{s.t.} \quad & \sum_{i=1}^{n} \gamma_i = \alpha \\ & 0 \leq \gamma_i \leq p_i \quad i = 1,\ldots,n \end{aligned} \tag{17}$$

(*iii*)

$$\begin{aligned} \min \quad & \Psi_\alpha(V,K,k) \\ s.t. \quad & K \subset [n] \\ & k \in [n] \setminus K \\ & \sum_{i\in K} p_i \leq \alpha \\ & \alpha - \sum_{i\in K} p_i \leq p_k, \end{aligned} \tag{18}$$

where $[n] = \{1,\ldots,n\}$ *and*

$$\Psi_\alpha(V,K,k) = \frac{1}{\alpha}\left[\sum_{i\in K} p_i v_i + \left(\alpha - \sum_{i\in K} p_i\right) v_k\right].$$

Proof. It is easy to see that at an optimal solution of (16) we have $w_i = \max(\eta - v_i, 0) = [\eta - v_i]_+$. Therefore, by the definition given in (2), the optimum value equals $\text{CVaR}_\alpha(V)$. Problem (17) is equivalent to the linear programming dual of (16), therefore its optimum also equals $\text{CVaR}_\alpha(V)$.

Without loss of generality assume $v_1 \leq v_2 \leq \cdots \leq v_n$, and let $k^* = \min\left\{k \in [n] \ : \ \sum_{i=1}^{k} p_i \geq \alpha\right\}$. Since (17) is a continuous knapsack problem, the greedy solution given by the following formula is optimal.

$$\gamma_i^* = \begin{cases} p_i & i = 1,\ldots,k^*-1 \\ \alpha - \sum_{i=1}^{k^*-1} p_i & i = k^* \\ 0 & i = k^*+1,\ldots,n \end{cases}$$

Setting $K^* = \{1,\ldots,k^*-1\}$, the pair (K^*,k^*) is a feasible solution of (18) with objective value $\Psi_\alpha(V,K^*,k^*) = \text{CVaR}_\alpha(V)$. On the other hand, for any feasible solution (K,k) of (18) we can construct a feasible solution

$$\gamma_i = \begin{cases} p_i & i \in K \\ \alpha - \sum_{i\in K} p_i & i = k \\ 0 & i \notin K \cup \{k\} \end{cases}$$

of (17) with objective value $\Psi_\alpha(V,K,k)$. This implies that the optimum values of (17) and (18) coincide, which completes our proof. □

Remark 3.1 *The minimization problem in* (17) *is equivalent to the well-known risk envelope-based dual representation of* CVaR *(see, e.g., Rockafellar, 2007).*

Corollary 3.1 *A simple consequence of claim (i) in Theorem 3.1 is the well known fact that* CVaR-*relations can be represented by linear inequalities. For a benchmark value $b \in \mathbb{R}$ the inequality* $\mathrm{CVaR}_\alpha(V) \geq b$ *holds if and only if there exist $\eta \in \mathbb{R}$ and $\mathbf{w} \in \mathbb{R}^n$ satisfying the following system.*

$$\begin{aligned} &\eta - \tfrac{1}{\alpha}\sum_{i=1}^{n} p_i w_i \geq b && \\ &w_i \geq \eta - v_i && i = 1,\ldots,n, \\ &w_i \geq 0 && i = 1,\ldots,n \end{aligned}$$

When realizations of the random variable V are equally likely, CVaR has alternative closed form representations, presented below. These results prove useful in developing tractable solution methods (see Section 6.2.3).

Proposition 3.1 *Let V be a random variable with (not necessarily distinct) realizations $v_1, \ldots, v_n$ and corresponding equal probabilities $p_1 = \cdots = p_n = \frac{1}{n}$.*

(*i*) *Let $v_{(1)} \leq v_{(2)} \leq \cdots \leq v_{(n)}$ denote an ordering of the realizations. Then*

$$\mathrm{CVaR}_{\frac{k}{n}}(V) = \frac{1}{k}\sum_{i=1}^{k} v_{(i)}$$

holds for all $k = 1, \ldots, n$.

(*ii*) *For a confidence level $\alpha \in [\frac{k}{n}, \frac{k+1}{n})$, $k \in [n-1]$, we have*

$$\mathrm{CVaR}_\alpha(V) = \lambda_\alpha \,\mathrm{CVaR}_{\frac{k}{n}}(V) + (1-\lambda_\alpha)\,\mathrm{CVaR}_{\frac{k+1}{n}}(V),$$

where $\lambda_\alpha = \frac{k(k+1-\alpha n)}{\alpha n}$. Note that $0 < \lambda_\alpha \leq \lambda_{\frac{k}{n}} = 1$.

Proof. Since $\mathrm{VaR}_{\frac{k}{n}}(V) = v_{(k)}$, by (3) we have

$$\mathrm{CVaR}_{\frac{k}{n}}(V) = v_{(k)} - \frac{n}{k}\sum_{i=1}^{n} p_i [v_{(k)} - v_{(i)}]_+ = v_{(k)} - \frac{1}{k}\sum_{i=1}^{k}(v_{(k)} - v_{(i)}) = \frac{1}{k}\sum_{i=1}^{k} v_{(i)},$$

proving (i). For $\alpha = \frac{k}{n}$ claim (ii) trivially holds. Now suppose that $\alpha \in (\frac{k}{n}, \frac{k+1}{n})$. Then $\mathrm{VaR}_\alpha(V) = v_{(k+1)}$, and using (i) we have

$$\begin{aligned} \lambda_\alpha \,\mathrm{CVaR}_{\frac{k}{n}}(V) + (1-\lambda_\alpha)\,\mathrm{CVaR}_{\frac{k+1}{n}}(V) &= \frac{k(k+1-\alpha n)}{\alpha n}\frac{1}{k}\sum_{i=1}^{k} v_{(i)} + \frac{(k+1)(\alpha n - k)}{\alpha n}\frac{1}{k+1}\sum_{i=1}^{k+1} v_{(i)} \\ &= v_{(k+1)} - \frac{1}{\alpha n}\sum_{i=1}^{k}(v_{(k+1)} - v_{(i)}) = v_{(k+1)} - \frac{1}{\alpha}\sum_{i=1}^{n} p_i [v_{(k+1)} - v_{(i)}]_+ = \mathrm{CVaR}_\alpha(V). \end{aligned}$$

□

3.2 Finite representations of scalarization polyhedra For any nontrivial polyhedron C of scalarization vectors the corresponding CVaR-preferability constraint is equivalent by definition to a collection of infinitely many scalar-based CVaR constraints, one for each scalarization vector $\mathbf{c} \in C$. The following theorem shows that for finite probability spaces it is sufficient to consider a finite subset of these vectors, obtained as projections of the vertices of a higher dimensional polyhedron.

Theorem 3.2 *Let $\mathbf{X}$ and $\mathbf{Y}$ be d-dimensional random vectors with realizations $\mathbf{x}_1, \ldots, \mathbf{x}_n$ and $\mathbf{y}_1, \ldots, \mathbf{y}_m$, respectively. Let $p_1, \ldots, p_n$ and $q_1, \ldots, q_m$ denote the corresponding probabilities, and let*

$C \subset \mathbb{R}^d$ be a polytope of scalarization vectors. $\mathbf{X}$ *is* CVaR*-preferable to* $\mathbf{Y}$ *at confidence level* α *with respect to* C *if and only if*

$$\mathrm{CVaR}_\alpha(\mathbf{c}_{(\ell)}^T\mathbf{X}) \geq \mathrm{CVaR}_\alpha(\mathbf{c}_{(\ell)}^T\mathbf{Y}) \quad \textit{for all } \ell = 1,\dots,N,$$

where $(\mathbf{c}_{(\ell)}, \eta_{(\ell)}, \mathbf{w}_{(\ell)})$, $\ell = 1,\dots,N$, *are the vertices of the (line-free) polyhedron*

$$P(\mathbf{Y}, C) = \left\{(\mathbf{c}, \eta, \mathbf{w}) \in C \times \mathbb{R} \times \mathbb{R}_+^m \ : \ w_j \geq \eta - \mathbf{c}^T y_j, \quad j = 1,\dots,m\right\}. \tag{19}$$

Proof. If $\mathbf{X}$ is preferable to $\mathbf{Y}$, the condition trivially holds, since $\mathbf{c}_{(\ell)} \in C$ for all $\ell = 1,\dots,N$. Now assume that $\mathbf{X}$ is not preferable to $\mathbf{Y}$. Then the optimal objective value Δ of the following problem is negative:

$$\min_{\mathbf{c}\in C} \mathrm{CVaR}_\alpha(\mathbf{c}^T\mathbf{X}) - \mathrm{CVaR}_\alpha(\mathbf{c}^T\mathbf{Y}). \tag{20}$$

Using Theorem 3.1 we can reformulate this problem as

$$\begin{aligned}
\min \quad & \Psi_\alpha(\mathbf{c}^T\mathbf{X}, K, k) - \eta + \frac{1}{\alpha}\sum_{j=1}^m q_j w_j \\
\text{s.t.} \quad & K \subset [n] \\
& k \in [n] \setminus K \\
& \sum_{i\in K} p_i \leq \alpha \\
& \alpha - \sum_{i\in K} p_i \leq p_k \\
& w_j \geq \eta - \mathbf{c}^T y_j \qquad j = 1,\dots,m \\
& w_j \geq 0 \qquad j = 1,\dots,m \\
& \mathbf{c} \in C.
\end{aligned} \tag{SetBased}$$

Let $(K^*, k^*, \mathbf{c}^*, \eta^*, \mathbf{w}^*)$ be an optimal solution of (SetBased). Then, by fixing $K = K^*$ and $k = k^*$ we obtain the following problem, which clearly has the same optimal objective value Δ.

$$\begin{aligned}
\min \quad & \Psi_\alpha(\mathbf{c}^T\mathbf{X}, K^*, k^*) - \eta + \frac{1}{\alpha}\sum_{j=1}^m q_j w_j \\
\text{s.t.} \quad & w_j \geq \eta - \mathbf{c}^T y_j \qquad j = 1,\dots,m \\
& w_j \geq 0 \qquad j = 1,\dots,m \\
& \mathbf{c} \in C.
\end{aligned} \tag{FixedSet}$$

Since $\Psi_\alpha(\mathbf{c}^T\mathbf{X}, K^*, k^*)$ is a linear function of $\mathbf{c}$, (FixedSet) is a linear program with feasible set $P(\mathbf{Y}, C)$. Therefore, problem (FixedSet) has an optimal solution which is a vertex of $P(\mathbf{Y}, C)$, i.e., of the form $(\mathbf{c}_{(\ell)}, \eta_{(\ell)}, \mathbf{w}_{(\ell)})$ for some $\ell \in [N]$. Let $V = \mathbf{c}_{(\ell)}^T\mathbf{X}$; then Theorem 3.1 implies that $\mathrm{CVaR}_\alpha(\mathbf{c}_{(\ell)}^T\mathbf{X}) = \mathrm{CVaR}_\alpha(V)$ is equal to the optimal objective value of the minimization problem (18). Since (K^*, k^*) is a feasible solution of (18), we have

$$\Psi_\alpha(\mathbf{c}_{(\ell)}^T\mathbf{X}, K^*, k^*) \geq \mathrm{CVaR}_\alpha(\mathbf{c}_{(\ell)}^T\mathbf{X}). \tag{21}$$

Observe that if we fix $\mathbf{c} = \mathbf{c}_{(\ell)}$ in problem (FixedSet), it becomes

$$\Psi_\alpha(\mathbf{c}_{(\ell)}^T\mathbf{X}, K^*, k^*) - \max\left\{\eta - \frac{1}{\alpha}\mathbf{q}^T\mathbf{w} \ : \ w_j \geq \eta - \mathbf{c}_{(\ell)}^T y_j, \quad j = 1,\dots,m, \ \mathbf{w} \in \mathbb{R}_+^m\right\},$$

where by (2) the maximization term equals $\mathrm{CVaR}_\alpha(\mathbf{c}_{(\ell)}^T\mathbf{Y})$. Consequently, taking into account (21) we have

$$0 > \Delta = \Psi_\alpha(\mathbf{c}_{(\ell)}^T\mathbf{X}, K^*, k^*) - \mathrm{CVaR}_\alpha(\mathbf{c}_{(\ell)}^T\mathbf{Y}) \geq \mathrm{CVaR}_\alpha(\mathbf{c}_{(\ell)}^T\mathbf{X}) - \mathrm{CVaR}_\alpha(\mathbf{c}_{(\ell)}^T\mathbf{Y}), \tag{22}$$

which completes our proof. □

COROLLARY 3.2 *Under the conditions of the previous theorem there exists an index $\ell \in \{1, \ldots, N\}$ such that $\mathbf{c}_{(\ell)}$ is an optimal solution of problem (20).*

PROOF. Let $\mathbf{c}_{(\ell)}$ be the vector obtained as part of a vertex optimal solution to (FixedSet) like in the previous proof. By (22) we have $\mathrm{CVaR}_\alpha(\mathbf{c}_{(\ell)}^T\mathbf{X}) - \mathrm{CVaR}_\alpha(\mathbf{c}_{(\ell)}^T\mathbf{Y}) \leq \Delta$, where Δ denotes the optimal objective value of the minimization problem (20). On the other hand, $\mathbf{c}_{(\ell)}$ is a feasible solution, which proves our claim. □

REMARK 3.2 *In Theorem 3.2 the confidence levels applied to the two sides coincide. However, this is not a necessary condition, as it is easy to verify that the same proof is valid for the following asymmetric relation with any $\alpha_1, \alpha_2 \in (0, 1]$:*

$$\mathrm{CVaR}_{\alpha_1}(\mathbf{c}^T X) \geq \mathrm{CVaR}_{\alpha_2}(\mathbf{c}^T Y) \quad \textit{for all } \mathbf{c} \in C.$$

An even more general form of this theorem, featuring a wider class of risk measures, will be presented in Section 5.2.

COROLLARY 3.3 *Using our previous notation, $\mathbf{X}$ dominates $\mathbf{Y}$ in polyhedral linear second order with respect to C if and only if*

$$\mathbf{c}_{(\ell)}^T\mathbf{X} \succcurlyeq_{(2)} \mathbf{c}_{(\ell)}^T\mathbf{Y} \qquad \textit{for all } \ell = 1, \ldots, N.$$

PROOF. We show that the following statements are equivalent:

(i) $\mathbf{c}^T\mathbf{X} \succcurlyeq_{(2)} \mathbf{c}^T\mathbf{Y}$ for all $\mathbf{c} \in C$.

(ii) $\mathrm{CVaR}_\alpha(\mathbf{c}^T\mathbf{X}) \geq \mathrm{CVaR}_\alpha(\mathbf{c}^T\mathbf{Y})$ for all $\alpha \in (0, 1]$, $\mathbf{c} \in C$.

(iii) $\mathrm{CVaR}_\alpha(\mathbf{c}_{(\ell)}^T\mathbf{X}) \geq \mathrm{CVaR}_\alpha(\mathbf{c}_{(\ell)}^T\mathbf{Y})$ for all $\alpha \in (0, 1]$, $\ell = 1, \ldots, N$.

(iv) $\mathbf{c}_{(\ell)}^T\mathbf{X} \succcurlyeq_{(2)} \mathbf{c}_{(\ell)}^T\mathbf{Y}$ for all $\ell = 1, \ldots, N$.

Equivalences (i) ⇔ (ii) and (iii) ⇔ (iv) follow from the fact that, by Proposition 2.1, the SSD constraint is equivalent to the continuum of CVaR constraints for all $\alpha \in (0, 1]$. On the other hand, Theorem 3.2 states the equivalence of (ii) and (iii). □

REMARK 3.3 *The previous result is closely related to Theorem 1 of Homem-de-Mello and Mehrotra (2009), where the continuous variable η in (19) is replaced by the finite set of terms $\mathbf{c}^T\mathbf{y}_j$ for $j = 1, \ldots, m$, leading to a set of m lower-dimensional polyhedra instead of our single polyhedron $P(\mathbf{Y}, C)$.*

4. Linear programming formulation and duality In this section we develop duality results for problem (GeneralP) under certain conditions. Working under the assumption that C is a polytope we begin by introducing, for any subset $\tilde{C} \subset C$, the following relaxed problem:

$$\begin{aligned} \max \quad & f(\mathbf{z}) \\ \text{s.t.} \quad & \mathrm{CVaR}_\alpha(\mathbf{c}^T G(\mathbf{z})) \geq \mathrm{CVaR}_\alpha(\mathbf{c}^T\mathbf{Y}) \quad \text{for all } \mathbf{c} \in \tilde{C} \\ & \mathbf{z} \in Z. \end{aligned} \tag{Relax($\tilde{C}$)}$$

OBSERVATION 4.1 *Let $\hat{C}$ denote the set consisting of vectors $\mathbf{c}_{(1)}, \ldots, \mathbf{c}_{(N)}$, as defined in Theorem 3.2. Then, according to the theorem, (Relax($\hat{C}$)) is equivalent to (Relax(C)), which in turn is equivalent to our original problem (GeneralP).*

From a practical perspective the case when the probability space is finite, the mappings f and G are linear, and the set Z is polyhedral, is of particular interest. Let us introduce the following notation:

- $Z = \{\mathbf{z} \in \mathbb{R}^{r_1} \ : \ A\mathbf{z} \leq \mathbf{b}\}$ for some $A \in \mathbb{R}^{r_2 \times r_1}$ and $\mathbf{b} \in \mathbb{R}^{r_2}$.
- $f(\mathbf{z}) = \mathbf{f}^T\mathbf{z}$ for some vector $\mathbf{f} \in \mathbb{R}^{r_1}$.
- $G(\mathbf{z}, \omega) = \Gamma(\omega)\mathbf{z}$ for a random matrix $\Gamma : \Omega \to \mathbb{R}^{d \times r_1}$. In addition, let $\Gamma_i = \Gamma(\omega_i)$ for $i = 1\ldots, n$.

By Corollary 3.1 scalar-based CVaR-relations can be represented by linear inequalities. For a finite set $\tilde{C} = \{\tilde{\mathbf{c}}_{(1)}, \ldots, \tilde{\mathbf{c}}_{(L)}\}$ this allows us to formulate (Relax($\tilde{C}$)) as a linear program:

$$\begin{aligned} \max \quad & \mathbf{f}^T\mathbf{z} \\ \text{s.t.} \quad & \eta_\ell - \frac{1}{\alpha}\sum_{i=1}^n p_i w_{i\ell} \geq \mathrm{CVaR}_\alpha(\tilde{\mathbf{c}}_{(\ell)}^T \mathbf{Y}) && \ell = 1\ldots, L \\ & w_{i\ell} \geq \eta_\ell - \tilde{\mathbf{c}}_{(\ell)}^T \Gamma_i \mathbf{z} && i = 1, \ldots, n, \ \ell = 1\ldots, L \\ & w_{i\ell} \geq 0 && i = 1, \ldots, n, \ \ell = 1\ldots, L \\ & A\mathbf{z} \leq \mathbf{b}. \end{aligned} \tag{RelaxP($\tilde{C}$)}$$

The dual problem of (RelaxP($\tilde{C}$)) can be written as follows.

$$\begin{aligned} \min \quad & \lambda^T\mathbf{b} - \sum_{\ell=1}^L \mu_\ell \, \mathrm{CVaR}_\alpha(\tilde{\mathbf{c}}_{(\ell)}^T \mathbf{Y}) \\ \text{s.t.} \quad & \sum_{i=1}^n p_i \nu_{i\ell} = \mu_\ell && \ell = 1\ldots, L \\ & \nu_{i\ell} \leq \frac{1}{\alpha}\mu_\ell && i = 1, \ldots, n, \ \ell = 1\ldots, L \\ & \sum_{i=1}^n p_i \sum_{\ell=1}^L \nu_{i\ell} \tilde{\mathbf{c}}_{(\ell)}^T \Gamma_i = \lambda^T A - \mathbf{f}^T \\ & \lambda \in \mathbb{R}_+^{r_2}, \ \mu \in \mathbb{R}_+^L, \ \nu \in \mathbb{R}_+^{n \times L} \end{aligned} \tag{RelaxD($\tilde{C}$)}$$

Note that the above formulation slightly differs from the usual LP dual, since a scaling factor of p_i has been applied to each dual variable $\nu_{i\ell}$.

Observation 4.2 *The dual variable μ can be viewed as a measure supported on the finite set $\tilde{C}$, while ν can be interpreted as a random measure on the same set. For instance, the sum in the dual objective and the first set of dual constraints can be written as $\int_C \mathrm{CVaR}_\alpha(\mathbf{c}^T\mathbf{Y})\,\mu(\mathrm{d}\mathbf{c})$ and $\mathbb{E}(\nu) = \mu$, respectively. In this context, the complementary slackness conditions can be expressed as*

$$\begin{aligned} \mathrm{support}(\mu) \quad & \subset \quad \big\{\mathbf{c} \ : \ \mathrm{CVaR}_\alpha(\mathbf{c}^T\Gamma\mathbf{z}) = \mathrm{CVaR}_\alpha(\mathbf{c}^T\mathbf{Y})\big\} \\ \mathrm{support}(\nu(\omega_i)) \quad & \subset \quad \big\{\mathbf{c} \ : \ \mathbf{c}^T\Gamma_i\mathbf{z} < \mathrm{VaR}_\alpha(\mathbf{c}^T\Gamma\mathbf{z})\big\} \\ \lambda^T(A\mathbf{z} - \mathbf{b}) \quad & = \quad 0. \end{aligned}$$

This interpretation motivates us to introduce a general dual scheme. Let $\mathcal{M}_+^F(S)$ denote the set of all finitely supported finite non-negative measures on a set S. For a family of measures $\mathcal{M} \subset \mathcal{M}_+^F(C)$ consider the following dual problem:

$$\begin{aligned} \min \quad & \lambda^T\mathbf{b} - \int_C \mathrm{CVaR}_\alpha(\mathbf{c}^T\mathbf{Y})\,\mu(\mathrm{d}\mathbf{c}) \\ \text{s.t.} \quad & \mathbb{E}(\nu) = \mu \\ & \nu \leq \frac{1}{\alpha}\mu \\ & \mathbb{E}\left(\int_C \mathbf{c}^T\Gamma\,\nu(\mathrm{d}\mathbf{c})\right) = \lambda^T A - \mathbf{f}^T \\ & \lambda \in \mathbb{R}_+^{r_2}, \ \mu \in \mathcal{M}, \ \nu : \Omega \to \mathcal{M} \end{aligned} \tag{GeneralD($\mathcal{M}$)}$$

Proposition 4.1 *If our original (primal) problem (GeneralP) has a finite optimum value, then so does the dual problem (GeneralD($\mathcal{M}_+^F(\hat{C})$)), and the optimum values coincide.*

Proof. According to the interpretation of the measures μ and ν given in Observation 4.2, the problem (GeneralD($\mathcal{M}_+^F(\hat{C})$)) is equivalent to (RelaxD($\hat{C}$)), which has the same optimum value as (RelaxP($\hat{C}$)) due to linear programming duality. By Observation 4.1 this optimum coincides with that of (GeneralP), proving our claim. □

Notice that, while the above proposition provides a strong duality result, the dual problem features the set $\hat{C}$, which depends on the reference variable $\mathbf{Y}$, and can potentially consist of an exponential number of scalarization vectors. Since this set can be impractical to explicitly construct in practice, we conclude this section by providing a different dual formulation which can serve as the foundation of a column generation-type solution method.

Theorem 4.1 *If our original (primal) problem (GeneralP) has a finite optimum value, then so does the dual problem (GeneralD($\mathcal{M}_+^F(C)$)), and the optimum values coincide.*

Proof. Since $\mathcal{M}_+^F(\hat{C})$ is a subset of $\mathcal{M}_+^F(C)$, the feasible region of problem (GeneralD($\mathcal{M}_+^F(\hat{C})$)) is a subset of the feasible region of (GeneralD($\mathcal{M}_+^F(C)$)). Therefore, taking into account Proposition 4.1 the following relation holds for the respective optimum values:

$$\text{OPT}(\text{GeneralD}(\mathcal{M}_+^F(C))) \leq \text{OPT}(\text{GeneralD}(\mathcal{M}_+^F(\hat{C}))) = \text{OPT}(\text{GeneralP}).$$

On the other hand, let $(\lambda^*, \mu^*, \nu^*)$ be an optimal solution of (GeneralD($\mathcal{M}_+^F(C)$)) and consider the finite set

$$C^* = \text{support}(\mu^*) \cup \left(\bigcup_{i=1}^{n} \text{support}(\nu^*(\omega_i)) \right) \subset C.$$

Then the optimum values of (GeneralD($\mathcal{M}_+^F(C^*)$)) and (GeneralD($\mathcal{M}_+^F(C)$)) coincide, therefore we have

$$\begin{aligned}\text{OPT}(\text{GeneralD}(\mathcal{M}_+^F(C))) &= \text{OPT}(\text{GeneralD}(\mathcal{M}_+^F(C^*))) = \text{OPT}(\text{RelaxD}(C^*)) = \text{OPT}(\text{RelaxP}(C^*)) \\ &= \text{OPT}(\text{Relax}(C^*)) \geq \text{OPT}(\text{Relax}(C)) = \text{OPT}(\text{GeneralP}),\end{aligned}$$

which completes our proof.

□

5. Coherent risk measures For finite probability spaces, Theorem 3.2 shows that when the set of scalarization vectors is polyhedral, the multivariate CVaR constraints given in (12) can be reduced to finitely many univariate CVaR constraints. This fact is the key to proving the finite convergence of our cut generation method outlined in Section 6.2. Our goal here is to extend this important finiteness result to constraints based on a wider class of *coherent risk measures*.

5.1 Geometric preliminaries We now provide some necessary geometrical background to the general finiteness results that follow. The notation for this section is largely independent from that used for the rest of the paper.

Definition 5.1 *Let $\mathbf{p} \in P$ be a point belonging to some polyhedron $P \subset \mathbb{R}^n$. We say that a vector $\mathbf{d} \in \mathbb{R}^n$ is a P-*direction *of $\mathbf{p}$ if there exists $\epsilon > 0$ such that both $\mathbf{p} + \epsilon\mathbf{d}$ and $\mathbf{p} - \epsilon\mathbf{d}$ belong to P.*

Proposition 5.1 *Let $P \subset \mathbb{R}^n$ be a polyhedron.*

(*i*) *If $\mathbf{p}$ belongs to the interior of P, then every vector $\mathbf{d} \in \mathbb{R}^n$ is a P-direction of $\mathbf{p}$.*

(*ii*) *The point* $\mathbf{p}$ *is a vertex of* P *if and only if it has no non-zero* P*-directions.*

(*iii*) *If* $\mathbf{p}_1$ *and* $\mathbf{p}_2$ *are two points which belong to the relative interior of the same face of* P*, then the sets of their* P*-directions coincide.*

Proof. Claims (i) and (ii) are trivial. To prove (iii), consider a point $\mathbf{p}$ belonging to the relative interior of a face F of P. Notice that a vector is a P-direction of $\mathbf{p}$ if and only if it is an F-direction of $\mathbf{p}$. Let A denote the smallest affine subspace of R^n which contains F, then A is of the form $\mathbf{p}+S$ for some linear subspace S. The polyhedron F is full-dimensional in A, therefore by (i) the set of all F-directions (and thus the set of all P-directions) of $\mathbf{p}$ is the linear subspace S. As S is uniquely defined by F, our claim immediately follows. □

Definition 5.2 *Let* $P \subset \mathbb{R}^n \times \mathbb{R}^m$ *be a polyhedron. We call a vector* $\mathbf{x} \in \mathbb{R}^n$ *an* n*-vertex of* P *if it can be extended into a vertex, i.e., if there exists some* $\mathbf{y} \in \mathbb{R}^m$ *such that* $(\mathbf{x}, \mathbf{y})$ *is a vertex of* P.

Observe that the vectors $\mathbf{c}_{(\ell)}$ in Theorem 3.2 are the d-vertices of the polyhedron $P(\mathbf{Y}, C)$. When we extend this theorem to a more general class of risk measures, it is necessary to consider some more complicated polyhedra in place of $P(\mathbf{Y}, C)$. Given a polyhedron $P = P^{(1)} \subset \mathbb{R}^n \times \mathbb{R}^m$ we next introduce a series of "liftings".

$$P^{(k)} = \left\{ (\mathbf{x}, \mathbf{y}^{(1)}, \dots, \mathbf{y}^{(k)}) \in \mathbb{R}^n \times \mathbb{R}^m \times \cdots \times \mathbb{R}^m \ : \ (\mathbf{x}, \mathbf{y}^{(i)}) \in P \text{ for all } i = 1, \dots, k \right\} \tag{23}$$

Observation 5.1 *A vector* $(\mathbf{d}^{(0)}, \mathbf{d}^{(1)}, \dots, \mathbf{d}^{(k)})$ *is a* $P^{(k)}$*-direction of a point* $(\mathbf{x}, \mathbf{y}^{(1)}, \dots, \mathbf{y}^{(k)}) \in P^{(k)}$ *if and only if* $(\mathbf{d}^{(0)}, \mathbf{d}^{(i)})$ *is a* P*-direction of* $(\mathbf{x}, \mathbf{y}^{(i)})$ *for all* $i = 1, \dots, k$.

The following example shows that lifting a polyhedron in the above manner can introduce additional n-vertices.

Proposition 5.2 *Let* $P \subset \mathbb{R}^2 \times \mathbb{R}^1$ *be the tetrahedron depicted in Figure 1 with vertices* $(-1,0,-1)$*,* $(1,0,-1)$*,* $(0,-1,1)$*,* $(0,1,1)$*. In accordance with (23), let*

$$P^{(2)} = \left\{ (x_1, x_2, y^{(1)}, y^{(2)}) \ : \ (x_1, x_2, y^{(1)}) \in P, \ (x_1, x_2, y^{(2)}) \in P \right\}.$$

The point $(0,0)$ *is not a* 2*-vertex of* P*, but it is a* 2*-vertex of* $P^{(2)}$.

Proof. The fact that $(0,0)$ is not a 2-vertex of P can be verified by simply looking at the list of the vertices of P. We now show that $(0,0,-1,1)$ is a vertex of $P^{(2)}$, which proves our claim. Assume that $(d_1^{(0)}, d_2^{(0)}, d^{(1)}, d^{(2)})$ is a $P^{(2)}$-direction of $(0,0,-1,1)$. Then, by Observation 5.1 the vector $(d_1^{(0)}, d_2^{(0)}, d^{(1)})$ is a P-direction of the point $(0,0,-1)$. Since this point lies in the relative interior of the edge $[(-1,0,-1),(1,0,-1)] = \{(\lambda,0,-1) \ : \ \lambda \in [-1,1]\}$ of P, it is easy to see that $d_2^{(0)} = d^{(1)} = 0$. Analogously, $(d_1^{(0)}, d_2^{(0)}, d^{(2)})$ is a P-direction of the point $(0,0,1)$, which lies in the relative interior of the edge $[(0,-1,1),(0,1,1)]$, implying $d_1^{(0)} = d^{(2)} = 0$. Therefore $(0,0,-1,1)$ has no non-zero $P^{(2)}$-directions, so according to part (ii) of Proposition 5.1 it is a vertex. □

We conclude this subsection by showing the crucial result that, even though the lifting procedure can introduce new n-vertices, the set of n-vertices of the series of polyhedra $P_1, P_2, \dots$ eventually stabilizes.

Theorem 5.1 *For any given polyhedron* $P \subset \mathbb{R}^n \times \mathbb{R}^m$ *there exists a positive integer* k^* *such that for all* $k = 1, 2, \dots$ *any* n*-vertex of* $P^{(k)}$ *is also an* n*-vertex of* $P^{(k^*)}$.

Proof. Let k^* denote the number of the faces of P (including the trivial faces, i.e., the vertices and the polyhedron itself). We prove our theorem by showing that the following two statements hold:

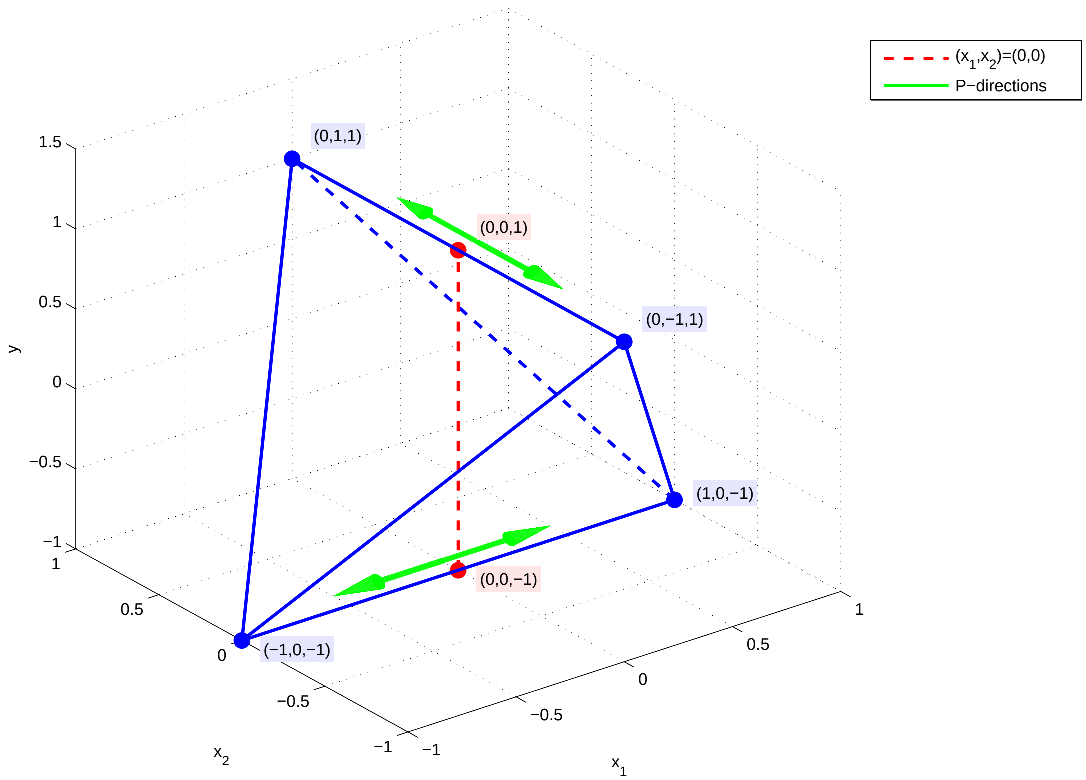

Figure 1: Tetrahedron P in Proposition 5.2

(i) For an integer $k < k^*$ any n-vertex of $P^{(k)}$ is also an n-vertex of $P^{(k+1)}$.

(ii) For an integer $k > k^*$ any n-vertex of $P^{(k)}$ is also an n-vertex of $P^{(k-1)}$.

Let us first assume $k < k^*$, and let $\mathbf{v}^{(k)} = (\mathbf{x}, \mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(k)})$ be a vertex of $P^{(k)}$. We prove (i) by showing that $\mathbf{v}^{(k+1)} = (\mathbf{x}, \mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(k)}, \mathbf{y}^{(k)})$ is a vertex of $P^{(k+1)}$. Indeed, if $\mathbf{d} = (\mathbf{d}^{(0)}, \mathbf{d}^{(1)}, \ldots, \mathbf{d}^{(k+1)})$ is a $P^{(k+1)}$-direction of $\mathbf{v}^{(k+1)}$, then by Observation 5.1 both $(\mathbf{d}^{(0)}, \mathbf{d}^{(1)}, \ldots, \mathbf{d}^{(k-1)}, \mathbf{d}^{(k)})$ and $(\mathbf{d}^{(0)}, \mathbf{d}^{(1)}, \ldots, \mathbf{d}^{(k-1)}, \mathbf{d}^{(k+1)})$ are $P^{(k)}$-directions of $\mathbf{v}^{(k)}$. According to claim (ii) of Proposition 5.1, the vertex $\mathbf{v}^{(k)}$ has no non-zero $P^{(k)}$-directions. Therefore every component of $\mathbf{d}$ is zero, which (by the same claim) implies that $\mathbf{v}^{(k+1)}$ is a vertex.

Now assume $k > k^*$, and again let $\mathbf{v}^{(k)} = (\mathbf{x}, \mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(k)})$ be a vertex of $P^{(k)}$. Then, due to our choice of k^*, by the pigeonhole principle at least two of the points $(\mathbf{x}, \mathbf{y}^{(1)}), \ldots, (\mathbf{x}, \mathbf{y}^{(k)})$ belong to the relative interior of the same face of P. Without loss of generality assume that $(\mathbf{x}, \mathbf{y}^{(k-1)})$ and $(\mathbf{x}, \mathbf{y}^{(k)})$ are two such points, and note that by claim (iii) of Proposition 5.1 their P-directions coincide. We conclude our proof by showing that $\mathbf{v}^{(k-1)} = (\mathbf{x}, \mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(k-1)})$ is a vertex of $P^{(k-1)}$. As before, let $\mathbf{d} = (\mathbf{d}^{(0)}, \mathbf{d}^{(1)}, \ldots, \mathbf{d}^{(k-1)})$ be a $P^{(k-1)}$-direction of $\mathbf{v}^{(k-1)}$. Analogously to the previous case it is easy to verify that $(\mathbf{d}^{(0)}, \mathbf{d}^{(1)}, \ldots, \mathbf{d}^{(k-1)}, \mathbf{d}^{(k-1)})$ is a $P^{(k)}$-direction of the vertex $\mathbf{v}^{(k)}$, implying that every component of $\mathbf{d}$ is zero. □

5.2 General finiteness proof The proof of the finite representation in Theorem 3.2 relied on representing CVaR both as a supremum and an infimum. Along these lines we begin this section by introducing two general classes of risk measures with similar representations.

Let $\mathcal{V}$ denote the set of all random variables $V : \Omega \to \mathbb{R}$ on the probability space $(\Omega, 2^\Omega, \Pi)$, and let

$\mathcal{L}$ denote the set of all linear functions $\Lambda : \mathcal{V} \to \mathbb{R}$ of the form $\Lambda(V) = \sum_{i=1}^n \lambda_i V(\omega_i)$. For a family of linear functions $L \subset \mathcal{L}$ we define the risk measure $\varrho_L : \mathcal{V} \to \mathbb{R}$ by

$$\varrho_L(V) = \inf_{\Lambda \in L} \Lambda(V). \tag{24}$$

Let $\mathcal{M}_+((0,1])$ denote the set of all non-negative measures on the interval $(0,1]$. For a family of measures $\mathcal{M} \subset \mathcal{M}_+((0,1])$ we define the risk measure $\rho_{\mathcal{M}} : \mathcal{V} \to \mathbb{R}$ by

$$\rho_{\mathcal{M}}(V) = \sup_{\mu \in \mathcal{M}} \int_0^1 \mathrm{CVaR}_\alpha(V)\, \mu(\mathrm{d}\alpha). \tag{25}$$

Note that for a family consisting of a single measure μ we have

$$\rho_{\{\mu\}}(V) = \int_0^1 \mathrm{CVaR}_\alpha(V)\, \mu(\mathrm{d}\alpha). \tag{26}$$

Structurally, the definitions in (24) and (25) correspond to the *risk envelope representation* and the *Kusuoka representation* of coherent risk measures, respectively (see Section 5.3). We are now going to prove a very general analogue of the finiteness result in Theorem 3.2. A key step in showing this result involves replacing a measure in representations of type (26) with a finitely supported approximating measure. In preparation, we prove that such an approximation always exists, achieving a preset level of precision. Note that for any given random variable V with finitely many realizations the mapping $\alpha \mapsto \mathrm{CVaR}_\alpha(V)$ is a bounded continuous non-decreasing function on $(0,1]$, and thus satisfies the conditions of the following lemma.

Lemma 5.1 *Let $\mu \in \mathcal{M}_+((0,1])$ be a measure and let $f_1, \dots, f_N$ be bounded continuous non-decreasing functions on $(0,1]$. Then, for any $\epsilon > 0$ there exists a finitely supported measure $\bar{\mu}$ on the interval $(0,1]$ such that*

$$\left| \int_{(0,1]} f_\ell \,\mathrm{d}\mu - \int_{(0,1]} f_\ell \,\mathrm{d}\bar{\mu} \right| < \epsilon$$

holds for all $\ell = 1, \dots, N$.

Proof. Since $f_1, \dots, f_N$ are bounded, we can assume without loss of generality that they are non-negative. Let us consider the following functions.

$$f_\ell^{(k)}(\alpha) = \begin{cases} 0 & \alpha \in (0, \frac{1}{2^k}] \\ f_\ell(\frac{i}{2^k}) & \alpha \in (\frac{i}{2^k}, \frac{i+1}{2^k}],\ i = 1, \dots, 2^k - 1 \end{cases} \qquad k = 1, 2, \dots$$

For any given ℓ the sequence $f_\ell^{(1)}, f_\ell^{(2)}, \dots$ is pointwise non-decreasing and converges pointwise to f_ℓ, therefore by Beppo Levi's monotone convergence theorem we have $\lim_{k\to\infty} \int_{(0,1]} f_\ell^{(k)} \,\mathrm{d}\mu = \int_{(0,1]} f_\ell \,\mathrm{d}\mu$. Let us now define the measures $\mu^{(k)}$ on support $\left\{\frac{1}{2^k}, \dots, \frac{2^k-1}{2^k}\right\}$ by setting $\mu^{(k)}\left(\left\{\frac{i}{2^k}\right\}\right) = \mu\left((\frac{i}{2^k}, \frac{i+1}{2^k}]\right)$. Then

$$\lim_{k\to\infty} \int_{(0,1]} f_\ell \,\mathrm{d}\mu^{(k)} = \lim_{k\to\infty} \sum_{i=1}^{2^k-1} f_\ell\left(\frac{i}{2^k}\right) \mu\left((\frac{i}{2^k}, \frac{i+1}{2^k}]\right) = \lim_{k\to\infty} \int_{(0,1]} f_\ell^{(k)} \,\mathrm{d}\mu = \int_{(0,1]} f_\ell \,\mathrm{d}\mu$$

holds for all $\ell = 1 \dots, N$, therefore for a sufficiently large choice of k the measure $\bar{\mu} = \mu^{(k)}$ will satisfy the requirements of the lemma. □

Theorem 5.2 *Let $\mathbf{X}$ and $\mathbf{Y}$ be d-dimensional random vectors with realizations $\mathbf{x}_1, \dots, \mathbf{x}_n$ and $\mathbf{y}_1, \dots, \mathbf{y}_m$, respectively. Let $p_1, \dots, p_n$ and $q_1, \dots, q_m$ denote the corresponding probabilities, and let $C \subset \mathbb{R}^d$ be a polytope of scalarization vectors. Given a family of linear functions $L \subset \mathcal{L}$ and a family of probability measures $\mathcal{M} \subset \mathcal{M}_+((0,1])$ the relation*

$$\varrho_L(\mathbf{c}^T\mathbf{X}) \geq \rho_{\mathcal{M}}(\mathbf{c}^T\mathbf{Y}) \quad \textit{for all } \mathbf{c} \in C \tag{27}$$

holds if and only if

$$\varrho_L(\mathbf{c}_{(\ell)}^T\mathbf{X}) \geq \rho_{\mathcal{M}}(\mathbf{c}_{(\ell)}^T\mathbf{Y}) \quad \forall \ell = 1,\ldots,N^*, \tag{28}$$

with $\mathbf{c}_{(1)},\ldots,\mathbf{c}_{(N^*)}$ *denoting the d-vertices of* $P^{(k^*)}(\mathbf{Y},C)$*, where* $P(\mathbf{Y},C)$ *is the polyhedron defined in* (19)*, and* k^* *is a positive integer as introduced in Theorem 5.1.*

Proof. Relation (27) trivially implies (28), since $\mathbf{c}_{(\ell)} \in C$ for all $\ell = 1,\ldots,N^*$. Now assume that (27) does not hold, implying

$$\inf_{\mathbf{c}\in\mathcal{C}} \varrho_L(\mathbf{c}^T\mathbf{X}) - \rho_{\mathcal{M}}(\mathbf{c}^T\mathbf{Y}) < 0.$$

By the definitions of ϱ_L and $\rho_{\mathcal{M}}$ this means that there exist $\Lambda^* \in L$ and $\mu^* \in \mathcal{M}$ such that

$$\inf_{\mathbf{c}\in\mathcal{C}} \Lambda^*(\mathbf{c}^T\mathbf{X}) - \rho_{\{\mu^*\}}(\mathbf{c}^T\mathbf{Y}) < 0.$$

Since C is compact, the infimum is attained at some $\mathbf{c}_{(0)} \in C$. As mentioned above, let $\mathbf{c}_{(1)},\ldots,\mathbf{c}_{(N^*)}$ denote the d-vertices of $P^{(k^*)}(\mathbf{Y},C)$. Then there exists a threshold $\epsilon > 0$ such that

$$\Lambda^*(\mathbf{c}_{(\ell)}^T\mathbf{X}) - \rho_{\{\mu^*\}}(\mathbf{c}_{(\ell)}^T\mathbf{Y}) < -\epsilon$$

holds for all indices $\ell \in \{0,\ldots,N^*\}$ that make the left-hand side negative. We now approximate the measure μ^* with a measure $\bar{\mu}$ supported on a finite set $\{\alpha_1,\ldots,\alpha_M\} \subset (0,1]$ with corresponding weights $\bar{\mu}_1,\ldots,\bar{\mu}_M$, and require

$$\left|\rho_{\{\mu^*\}}(\mathbf{c}_{(\ell)}^T\mathbf{Y}) - \rho_{\{\bar{\mu}\}}(\mathbf{c}_{(\ell)}^T\mathbf{Y})\right| < \frac{\epsilon}{2} \tag{29}$$

to hold for all $\ell = 0,\ldots,N^*$. The existence of such an approximation is guaranteed by Lemma 5.1. It follows that

$$\Lambda^*(\mathbf{c}_{(0)}^T\mathbf{X}) - \rho_{\{\bar{\mu}\}}(\mathbf{c}_{(0)}^T\mathbf{Y}) = \Lambda^*(\mathbf{c}_{(0)}^T\mathbf{X}) - \rho_{\{\mu^*\}}(\mathbf{c}_{(0)}^T\mathbf{Y}) + \Big(\rho_{\{\mu^*\}}(\mathbf{c}_{(0)}^T\mathbf{Y}) - \rho_{\{\bar{\mu}\}}(\mathbf{c}_{(0)}^T\mathbf{Y})\Big) < -\epsilon + \frac{\epsilon}{2} = -\frac{\epsilon}{2},$$

which implies

$$\inf_{\mathbf{c}\in\mathcal{C}} \Lambda^*(\mathbf{c}^T\mathbf{X}) - \rho_{\{\bar{\mu}\}}(\mathbf{c}^T\mathbf{Y}) < -\frac{\epsilon}{2}. \tag{30}$$

This infimum can be expressed as the optimum value of the following linear program:

$$\begin{aligned}
\min \quad & \sum_{i=1}^{n} \lambda_i^* \mathbf{c}^T\mathbf{x}_i - \sum_{h=1}^{M} \bar{\mu}_h \left(\eta_h - \frac{1}{\alpha_h}\sum_{j=1}^{m} q_j w_{jh}\right) & \\
\text{s.t.} \quad & w_{jh} \geq \eta_h - \mathbf{c}^T y_j & j = 1,\ldots,m,\ h = 1,\ldots,M \\
& w_{jh} \geq 0 & j = 1,\ldots,m,\ h = 1,\ldots,M \\
& \mathbf{c} \in C. &
\end{aligned}$$

Recalling the notation introduced in (23), the feasible set of this problem is $P^{(M)}(\mathbf{Y},C)$. Therefore, there exists an optimal solution $(\mathbf{c}^*,\eta^*,\mathbf{w}^*)$ which is a vertex of $P^{(M)}(\mathbf{Y},C)$. By Theorem 5.1 the vector $\mathbf{c}^*$ is a d-vertex of $P^{(k^*)}(\mathbf{Y},C)$, i.e., $\mathbf{c}^* = \mathbf{c}_{(\ell^*)}$ for some $\ell^* \in [N^*]$. Recalling (29) and (30), we have

$$\Lambda^*(\mathbf{c}_{(\ell^*)}^T\mathbf{X}) - \rho_{\{\mu^*\}}(\mathbf{c}_{(\ell^*)}^T\mathbf{Y}) = \Lambda^*(\mathbf{c}_{(\ell^*)}^T\mathbf{X}) - \rho_{\{\bar{\mu}\}}(\mathbf{c}_{(\ell^*)}^T\mathbf{Y}) - \Big(\rho_{\{\mu^*\}}(\mathbf{c}_{(\ell^*)}^T\mathbf{Y}) - \rho_{\{\bar{\mu}\}}(\mathbf{c}_{(\ell^*)}^T\mathbf{Y})\Big) < -\frac{\epsilon}{2} + \frac{\epsilon}{2} = 0.$$

Thus, relation (28) does not hold, which completes our proof. □

5.3 Functionally coherent risk measures In this section we apply the finite representation result in Theorem 5.2 to a class of coherent risk measures. In order to accomplish this, we first need to extend Kusuoka's representation of coherent risk measures to probability spaces which are not necessarily atomless. Let $\mathcal{V}(\Omega,\mathcal{A},\Pi)$ denote the set of all real valued random variables on an arbitrary probability space $(\Omega,\mathcal{A},\Pi)$, and let $\mathcal{F}(\Omega,\mathcal{A},\Pi) = \{F_V : V \in \mathcal{V}(\Omega,\mathcal{A},\Pi)\}$ denote the corresponding family of CDFs. Similarly, for a value $p \in [1,\infty]$ let $\mathcal{F}_p(\Omega,\mathcal{A},\Pi) = \{F_V : V \in \mathcal{L}^p(\Omega,\mathcal{A},\Pi)\}$. Let us recall some elementary properties of these distribution functions.

Proposition 5.3 *For any two* atomless *probability spaces* $(\Omega_1, \mathcal{A}_1, \Pi_1)$, $(\Omega_2, \mathcal{A}_2, \Pi_2)$ *and* $p \in [1, \infty]$ *the sets* $\mathcal{F}_p(\Omega_1, \mathcal{A}_1, \Pi_1)$ *and* $\mathcal{F}_p(\Omega_2, \mathcal{A}_2, \Pi_2)$ *coincide; let us denote this common family of CDFs by* F_p^*. *Furthermore, for an arbitrary space* $(\Omega, \mathcal{A}, \Pi)$ *we have* $\mathcal{F}_p(\Omega, \mathcal{A}, \Pi) \subset F_p^*$, *and equality holds if and only if* Ω *is atomless.*

Definition 5.3 *A mapping* $\rho : \mathcal{L}^p(\Omega, \mathcal{A}, \Pi) \to \mathbb{R}$ *is called* law invariant *if the value* $\rho(V)$ *depends only on the distribution of the random variable* V, *i.e., if there exists a mapping* $\varphi_\rho : \mathcal{F}_p(\Omega, \mathcal{A}, \Pi) \to \mathbb{R}$ *such that* $\rho(V) = \varphi_\rho(F_V)$ *holds for all* $V \in \mathcal{L}^p(\Omega, \mathcal{A}, \Pi)$. *Note that in this case* φ_ρ *is uniquely determined by* ρ.

Recalling formulas (1) and (4) it is easy to verify that for any confidence level α the risk measure CVaR_α is law invariant, with the corresponding mapping

$$\varphi_{\mathrm{CVaR}_\alpha} : F \mapsto \frac{1}{\alpha} \int_0^\alpha \inf \{\eta \ : \ F(\eta) \geq \gamma\} \, \mathrm{d}\gamma. \tag{31}$$

A mapping $\rho : \mathcal{L}^p(\Omega, \mathcal{A}, \Pi) \to \mathbb{R}$ is called a *coherent risk measure* (Artzner et al., 1999) if it has the following properties:

- *Monotone*: $\mathbf{X} \leq \mathbf{Y} \ \Rightarrow \ \rho(\mathbf{X}) \leq \rho(\mathbf{Y})$.
- *Superadditive*: $\rho(\mathbf{X} + \mathbf{Y}) \geq \rho(\mathbf{X}) + \rho(\mathbf{Y})$.
- *Positive homogeneous*: $\rho(\lambda \mathbf{X}) = \lambda \rho(\mathbf{X})$ for all $\lambda \geq 0$.
- *Translation invariant*: $\rho(\mathbf{X} + \lambda) = \rho(\mathbf{X}) + \lambda$.

It is well known (Pflug, 2000) that CVaR_α is a law invariant coherent risk measure for any confidence level $\alpha \in (0, 1]$. Moreover, these CVaR_α can be viewed as building blocks of coherent risk measures, as the following fundamental theorem shows. The result for $p = \infty$ is due to Kusuoka (2001) and Jouini et al. (2006), while the proof for $p \in [1, \infty)$ can be found in Shapiro et al. (2009).

Theorem 5.3 (Kusuoka representation) *Let* $(\Omega, \mathcal{A}, \Pi)$ *be an* atomless *probability space and* $p \in [1, \infty]$. *Then a mapping* $\rho : \mathcal{L}^p(\Omega, \mathcal{A}, \Pi) \to \mathbb{R}$ *is a law invariant coherent risk measure if and only if there exists a family* $\mathcal{M} \subset \mathcal{P}((0, 1])$ *of probability measures on the interval* $(0, 1]$ *such that the following representation holds for all* $V \in \mathcal{L}^p(\Omega, \mathcal{A}, \Pi)$.

$$\rho(V) = \sup_{\mu \in \mathcal{M}} \int_0^1 \mathrm{CVaR}_\alpha(V) \, \mu(\mathrm{d}\alpha) \tag{32}$$

If the underlying probability space has atoms, the above Kusuoka representation does not hold in general, as the next example shows.

Example 5.1 *Let* $\Omega = \{\omega_1, \omega_2\}$ *with* $\Pi(\omega_1) > \frac{1}{2}$. *It is easy to verify that the risk measure* ρ *defined by* $\rho(V) = V(\omega_1)$ *is coherent. In addition, the equality* $\rho(V) = \sup \left\{v \ : \ F_V(v) < \frac{1}{2}\right\}$ *shows that* ρ *is law invariant. However, it can be proven (Pflug and Römisch, 2007) that* ρ *has no Kusuoka representation.*

The risk measure ρ in the above example is "pathological" in the sense that its corresponding mapping φ_ρ cannot be coherently extended to the set $\mathcal{F}_p^*$ of all CDFs. We now formalize this intuitive notion.

Definition 5.4 *Consider a value* $p \in [1, \infty]$. *A mapping* $\phi : \mathcal{F}_p^* \to \mathbb{R}$ *is called* coherent *if* $\phi = \varphi_\rho$ *holds for some law invariant coherent risk measure* ρ *defined on an atomless probability space. Given a not necessarily atomless probability space* $(\Omega, \mathcal{A}, \Pi)$, *a mapping* $\rho : \mathcal{L}^p(\Omega, \mathcal{A}, \Pi) \to \mathbb{R}$ *is called a* functionally coherent risk measure *if there exists a coherent mapping* $\phi^* : \mathcal{F}_p^* \to \mathbb{R}$ *such that* φ_ρ *is a restriction of* ϕ^*, *i.e., we have* $\varphi_\rho = \phi^*|_{\mathcal{F}_p(\Omega, \mathcal{A}, \Pi)}$.

The class of functionally coherent risk measures preserves the desirable properties of law invariant coherent risk measures on atomless spaces without sacrificing generality. To our best knowledge the following theorem is the first published extension of Kusuoka representations to probability spaces having atoms[1].

THEOREM 5.4 *Let $(\Omega, \mathcal{A}, \Pi)$ be an arbitrary probability space and $p \in [1, \infty]$. A mapping $\rho : \mathcal{L}^p(\Omega, \mathcal{A}, \Pi) \to \mathbb{R}$ is a functionally coherent risk measure if and only if it has a Kusuoka representation of the form (32). In particular, if $(\Omega, \mathcal{A}, \Pi)$ is atomless, the class of functionally coherent risk measures coincides with the class of law invariant coherent risk measures.*

PROOF. First assume that ρ is functionally coherent. Then, according to Definition 5.4, there exists a law invariant risk measure ρ^* on an atomless space such that $\rho(V) = \varphi_{\rho^*}(F_V)$ holds for all $V \in \mathcal{L}^p(\Omega, \mathcal{A}, \Pi)$. By Theorem 5.3 this ρ^* has a Kusuoka representation. Recalling the notation introduced in (31), we then have

$$\rho(V) = \varphi_{\rho^*}(F_V) = \sup_{\mu \in \mathcal{M}} \int_0^1 \varphi_{\mathrm{CVaR}_\alpha}(F_V)\, \mu(\mathrm{d}\alpha) = \sup_{\mu \in \mathcal{M}} \int_0^1 \mathrm{CVaR}_\alpha(V)\, \mu(\mathrm{d}\alpha)$$

for some family $\mathcal{M} \subset \mathcal{P}((0,1])$, as required.

Now assume that ρ has a Kusuoka representation given by a family $\mathcal{M} \subset \mathcal{P}((0,1])$, and consider the risk measure ρ^* on some atomless probability space (for instance $([0,1], \mathcal{B}, \lambda)$ with the Lebesgue measure) given by the same Kusuoka representation. According to Theorem 5.3 the risk measure ρ^* is law invariant and coherent, therefore by definition $\varphi_{\rho^*} : \mathcal{F}_p^* \to \mathbb{R}$ is a coherent mapping. Then, since

$$\varphi_\rho(F_V) = \rho(V) = \sup_{\mu \in \mathcal{M}} \int_0^1 \mathrm{CVaR}_\alpha(V)\, \mu(\mathrm{d}\alpha) = \sup_{\mu \in \mathcal{M}} \int_0^1 \varphi_{\mathrm{CVaR}_\alpha}(F_V)\, \mu(\mathrm{d}\alpha) = \varphi_{\rho^*}(F_V)$$

holds for all $V \in \mathcal{L}_p(\Omega, \mathcal{A}, \Pi)$, we have $\varphi_\rho = \varphi_{\rho^*}|_{\mathcal{F}_p(\Omega, \mathcal{A}, \Pi)}$, which proves our claim. □

We are now ready to provide a finite representation for scalarization polyhedra associated with multivariate preference relations based on functionally coherent risk measures.

COROLLARY 5.1 *Let ρ be a functionally coherent risk measure on a finite probability space. Then the relation $X \succcurlyeq_\rho^C Y$ is equivalent to*

$$\rho(\mathbf{c}_{(\ell)}^T \mathbf{X}) \geq \rho(\mathbf{c}_{(\ell)}^T \mathbf{Y}) \qquad \textit{for all } \ell = 1, \ldots, N^*,$$

where $\mathbf{c}_{(1)}, \ldots, \mathbf{c}_{(N^*)}$ denote the d-vertices of $P^{(k^*)}(\mathbf{Y}, C)$.

PROOF. According to the well-known risk envelope representation of coherent risk measures (Artzner et al., 1999) there exists a *coherent risk envelope* $\mathcal{Q} \subset \{Q \in \mathcal{V} \;:\; Q \geq 0,\ \mathbb{E}(Q) = 1\}$ such that $\rho(V) = \inf\limits_{Q \in \mathcal{Q}} \mathbb{E}(QV)$. Introducing the set $L = \{V \mapsto \mathbb{E}(QV) \;:\; Q \in \mathcal{Q}\}$ and recalling the notation in (24) we have $\rho = \varrho_L$. On the other hand, by Theorem 5.4 the risk measure ρ has a Kusuoka representation. This provides a set of probability measures $\mathcal{M} \subset \mathcal{P}((0,1])$ for which, using the notation in (25), we have $\rho = \rho_{\mathcal{M}}$. Our assertion now follows directly from Theorem 5.2. □

[1] Bertsimas and Brown (2009) prove the related result that the integral representation of distortion risk measures (also due to Kusuoka (2001)) can be expressed via a finite sum on finite probability spaces where every outcome has equal probability.

6. Solution methods Here we develop methods to solve the multivariate CVaR-constrained optimization problem (GeneralP) in the case when the probability space is finite and the scalarization set C is polyhedral. We first briefly discuss a "brute force" approach based on vertex enumeration, which is made possible by the finite representation in Theorem 3.2. We then proceed to present a cut generation algorithm which avoids many of the pitfalls associated with an enumeration-based approach. After proving finite convergence, we provide a detailed discussion on implementing various steps of the algorithm.

6.1 Vertex enumeration In this section we consider a polytope $C = \left\{\mathbf{c} \in \mathbb{R}^d \ : \ B\mathbf{c} \leq \mathbf{h}\right\}$ of scalarization vectors, where the matrix $B \in \mathbb{R}^{r\times d}$ and the vector $\mathbf{h} \in \mathbb{R}^r$ provide a non-degenerate linear description. Let us recall that according to Observation 4.1 the formulation (Relax($\hat{C}$)), while formally a relaxation of our original problem (GeneralP), is actually equivalent to it. Here $\hat{C} = \{\mathbf{c}_{(1)}, \dots, \mathbf{c}_{(N)}\}$ denotes the set of all d-vertices of the polyhedron $P(\mathbf{Y}, C)$; assuming that we have access to this set, we can attempt a "brute force" solution by using non-linear programming techniques to tackle (Relax($\hat{C}$)). Assuming concavity of the functions f and G we obtain a convex programming problem, while under the linearity assumptions introduced in the beginning of Section 4 we arrive at the linear program (RelaxP($\hat{C}$)).

There are various methods available in the literature to construct the set $\hat{C}$ from the linear description

$$\begin{aligned} & w_j \geq \eta - \mathbf{c}^T y_j && j = 1, \dots, m \\ & w_j \geq 0 && j = 1, \dots, m \\ & B\mathbf{c} \leq \mathbf{h} \end{aligned}$$

of the polyhedron $P(\mathbf{Y}, C)$. While enumerating the vertices of a polyhedron is an NP-hard problem (Khachiyan et al., 2008), the well-known vertex enumeration algorithm in Avis and Fukuda (1992) provides a solution for the bounded case. If we construct a polytope $\hat{P}(\mathbf{Y}, C)$ by adding a non-restrictive lower bound on η and similar upper bounds on $\mathbf{w}$, the algorithm produces a complete list of its vertices in time $O\left(\hat{N}(d + m)^2\right)$, where $\hat{N}$ denotes the number of vertices of $\hat{P}(\mathbf{Y}, C)$. The fact that the polyhedron $P(\mathbf{Y}, C)$ depends only on $\mathbf{Y}$ and C (and not on the confidence level α or the feasible set Z) indicates some potential advantages of vertex enumeration: once d-vertices are generated, they can be reused in a variety of situations, including the following.

- Solving multiple problems with the same reference vector $\mathbf{Y}$ (e.g., solving a problem at various confidence levels, or for different feasible regions).
- Enforcing CVaR constraints for multiple confidence levels in an optimization problem.

However, in large-scale applications using such a vertex enumeration approach can become impractical for a variety of reasons:

- The number of vertices is potentially exponential in $(d + m)$.
- The additional bounding constraints used to construct $\hat{P}(\mathbf{Y}, C)$ introduce new vertices which are not relevant to the original problem.
- The list $\mathbf{c}_{(1)}, \dots, \mathbf{c}_{(N)}$ might contain duplicates, as a d-vertex can be part of multiple vertices of $P(\mathbf{Y}, C)$. When such duplicates occur, enumerating all vertices can create significant redundancy.

In the next section we outline a cut generation approach which addresses some of the above concerns. Some advantages of this method over vertex enumeration are listed below.

- It is typically not necessary to generate all d-vertices before arriving at the optimal solution.
- No additional vertices are introduced.
- Each d-vertex is generated at most once.

6.2 A cut-generation algorithm In this section we present an iterative algorithm which solves our original problem (GeneralP) in the case when the objective function f is continuous, the scalarization set C is a non-empty polytope and the feasible set Z is compact[2]. Each iteration consists of two steps: first we find an optimal solution $\mathbf{z}^*$ of the relaxed problem (Relax($\tilde{C}$)) for some finite set $\tilde{C} \in C$. Then given the associated outcome vector $\mathbf{X} = G(\mathbf{z}^*)$ we attempt to find a scalarization vector $\mathbf{c}^* \in C$ for which the corresponding condition

$$\mathrm{CVaR}_\alpha({\mathbf{c}^*}^T\mathbf{X}) \geq \mathrm{CVaR}_\alpha({\mathbf{c}^*}^T\mathbf{Y}) \tag{33}$$

is violated. We accomplish this by solving the *cut generation problem* (20). If the optimal objective value is non-negative, it follows that $\mathbf{z}^*$ is an optimal solution of (GeneralP). Otherwise, by Corollary 3.2 there exists an optimal solution $\mathbf{c}^*$ which is a d-vertex of the polyhedron $P(\mathbf{Y}, C)$ introduced in (19). We find such a vector and add it to the set $\tilde{C}$, which creates a tighter relaxation to be solved in the next iteration. This corresponds to introducing the constraint (33), which is a valid cut for the current solution $\mathbf{z}^*$. Note that introducing the new constraint requires calculating the parameter $\mathrm{CVaR}_\alpha({\mathbf{c}^*}^T\mathbf{Y})$. This simple calculation is automatically performed as a byproduct of solving the optimization problems presented in Sections 6.2.2-6.2.3. Algorithm 1 provides a formal description of our solution method.

Algorithm 1 Cut-Generation Algorithm

1: Initialize a set of scalarization vectors $\tilde{C} = \{c_{(1)}, \dots, c_{(L)}\} \subset C$.
2: Solve the master problem

$$\begin{aligned} \max \quad & f(\mathbf{z}) \\ \text{s.t.} \quad & \mathrm{CVaR}_\alpha(\mathbf{c}_{(\ell)}^T G(\mathbf{z})) \geq \mathrm{CVaR}_\alpha(\mathbf{c}_{(\ell)}^T \mathbf{Y}) \qquad \ell = 1, \dots, L \\ & \mathbf{z} \in Z. \end{aligned} \tag{Master}$$

3: **if** the master problem is infeasible **then**
4: Stop.
5: **else**
6: Let $\mathbf{z}^*$ be an optimal solution.
7: Given the optimal decision vector $\mathbf{z}^*$ set $\mathbf{X} = G(\mathbf{z}^*)$, and solve the cut generation problem

$$\min_{\mathbf{c} \in C} \mathrm{CVaR}_\alpha(\mathbf{c}^T\mathbf{X}) - \mathrm{CVaR}_\alpha(\mathbf{c}^T\mathbf{Y}). \tag{CutGen}$$

8: **if** the optimal objective value of the cut generation problem is nonnegative **then**
9: Stop.
10: **else**
11: Find an optimal solution $\mathbf{c}_{(L+1)}$ of the cut generation problem which is a d-vertex of $P(\mathbf{Y}, C)$. Set $\tilde{C} = \tilde{C} \cup \{\mathbf{c}_{(L+1)}\}$ and $L = L + 1$, then go to Step 2.
12: **end if**
13: **end if**

Remark 6.1 *A trivial way to perform the initialization in Step 1 is by setting $L = 0$ and $\tilde{C} = \emptyset$. However, since the cut generation problem often presents a computational bottleneck, more aggressive initialization strategies can improve the performance of the algorithm. When the master problem is comparatively easier to solve, considering a large initial scalarization set does not result in a significant burden. For instance, if the vertices $\hat{\mathbf{c}}_{(1)}, \dots, \hat{\mathbf{c}}_{(k)}$ of the scalarization polyhedron are known, setting $L = k$ and $\tilde{C} = \{\hat{\mathbf{c}}_{(1)}, \dots, \hat{\mathbf{c}}_{(k)}\}$ can provide a suitable initialization.*

[2]While the assumption of having a polyhedral scalarization set is essential to proving finite convergence, the compactness assumptions on C and Z are adopted for the ease of exposition only.

While Algorithm 1 is presented for the case of a single CVaR constraint, it can naturally be extended to problems of the more general form (15). In this case a separate cut generation problem is defined for each pair of a benchmark vector and an associated confidence level. Note that, in contrast to the method proposed in Homem-de-Mello and Mehrotra (2009) to solve SSD-constrained models, the number of cut generation problems does not depend on the number of benchmark realizations.

Theorem 6.1 *Algorithm 1 terminates after a finite number of iterations, and provides either an optimal solution of* (GeneralP)*, or a proof of infeasibility.*

Proof. Note that under our assumptions both the master problem and the cut generation problem involve the optimization of a continuous function over a compact set. It follows that the master problem either has an optimal solution or it is infeasible, while the cut generation problem always has an optimal solution since its feasible set C is non-empty. In addition, Corollary 3.2 states that at least one of the optimal solutions of the cut generation problem is a d-vertex of $P(\mathbf{Y}, C)$. Therefore, the cut generation algorithm operates as described, and can terminate in one of two ways:

- The master problem is infeasible. Since the master problem is a relaxation of (GeneralP), this constitutes a proof of infeasibility for our original problem.
- The optimum of the cut generation problem is non-negative. This implies that the current optimal solution $\mathbf{z}^*$ of the master problem is a feasible, and therefore optimal, solution of (GeneralP).

It remains to show that the algorithm always terminates in a finite number of iterations. This follows from the fact that every non-terminating iteration introduces a distinct d-vertex of the polyhedron $P(\mathbf{Y}, C)$, and the number of d-vertices is finite. □

Remark 6.2 *It is possible to introduce a dual counterpart to Algorithm 1, reminiscent of a column generation method. Under linearity assumptions the master problem will take the form of* (RelaxD($\tilde{C}$)) *and the* pricing problem *will be equivalent to* (CutGen)*. In accordance with Observation 4.2, introducing a new vector* $\mathbf{c}_{(L+1)}$ *can be interpreted as adding a new point to the finite supports of the measures* μ *and* ν*.*

6.2.1 Solving the master problem Corollary 3.1 allows us to represent CVaR constraints by linear inequalities, leading to the following formulation of (Master).

$$\begin{aligned}
\max \quad & f(\mathbf{z}) & \\
\text{s.t.} \quad & \eta_\ell - \frac{1}{\alpha}\sum_{i=1}^{n} p_i w_{i\ell} \geq \mathrm{CVaR}_\alpha(\mathbf{c}_{(\ell)}^T \mathbf{Y}) & \ell = 1, \ldots, L \\
& w_{i\ell} \geq \eta_\ell - \mathbf{c}_{(\ell)}^T g_i(\mathbf{z}) & i = 1, \ldots, n, \ \ell = 1, \ldots, L \\
& w_{i\ell} \geq 0 & i = 1, \ldots, n, \ \ell = 1, \ldots, L \\
& \mathbf{z} \in Z &
\end{aligned}$$

In the general case we can attempt to solve this problem using non-linear programming techniques, or, with appropriate assumptions on f and Z, a convex programming approach. Under the linearity assumptions of Section 4 the master problem becomes the linear program (RelaxP($\tilde{C}$)), providing a computationally tractable formulation.

6.2.2 Solving the cut generation problem In this section we consider two d-dimensional random vectors $\mathbf{X}$ and $\mathbf{Y}$ with realizations $\mathbf{x}_1, \ldots, \mathbf{x}_n$ and $\mathbf{y}_1, \ldots, \mathbf{y}_m$, respectively. Let $p_1, \ldots, p_n$ and $q_1, \ldots, q_m$ denote the corresponding probabilities, and let $C = \left\{\mathbf{c} \in \mathbb{R}^d \ : \ B\mathbf{c} \leq \mathbf{h}\right\}$ be a polytope of scalarization vectors for some matrix B and vector $\mathbf{h}$ of appropriate dimensions. The cut generation problem at confidence level $\alpha \in (0,1]$ involves either finding a vector $\mathbf{c} \in C$ such that $\mathrm{CVaR}_\alpha(\mathbf{c}^T\mathbf{X}) < \mathrm{CVaR}_\alpha(\mathbf{c}^T\mathbf{Y})$ or showing that such a vector does not exist. To accomplish this, we aim to solve the optimization problem (CutGen). Recalling Theorem 3.1, we represent $\mathrm{CVaR}_\alpha(\mathbf{c}^T\mathbf{X})$ and $\mathrm{CVaR}_\alpha(\mathbf{c}^T\mathbf{Y})$ using formulations (17) and (16), respectively. This allows us to restate (CutGen) as a quadratic program:

$$
\begin{aligned}
\min \quad & \frac{1}{\alpha}\sum_{i=1}^{n}\gamma_i\mathbf{c}^T\mathbf{x}_i - \eta + \frac{1}{\alpha}\sum_{j=1}^{m} q_j w_j \\
\text{s.t.} \quad & \sum_{i=1}^{n}\gamma_i = \alpha \\
& 0 \le \gamma_i \le p_i && i = 1,\dots,n \\
& w_j \ge \eta - \mathbf{c}^T\mathbf{y}_j && j = 1,\dots,m \\
& \mathbf{c} \in C, \;\; \mathbf{w} \in \mathbb{R}^m_+.
\end{aligned}
\tag{34}
$$

Note that this quadratic problem is not necessarily convex, and therefore can present a significant computational challenge. This motivates us to introduce an alternate *mixed integer linear programming* (MIP) formulation which is potentially more tractable.

According to (3) the supremum in the classical definition of CVaR_α is attained at VaR_α. Since the probability space is finite, $\mathrm{VaR}_\alpha(\mathbf{c}^T\mathbf{X}) = \mathbf{c}^T\mathbf{x}_k$ for at least one $k \in \{1,\dots,n\}$, implying

$$\mathrm{CVaR}_\alpha(\mathbf{c}^T\mathbf{X}) = \max_{k\in\{1,\dots,n\}} \mathbf{c}^T\mathbf{x}_k - \frac{1}{\alpha}\sum_{i=1}^{n} p_i[\mathbf{c}^T\mathbf{x}_k - \mathbf{c}^T\mathbf{x}_i]_+.$$

Representing $\mathrm{CVaR}_\alpha(\mathbf{c}^T\mathbf{Y})$ as before, we obtain the following intermediate formulation of (CutGen).

$$
\begin{aligned}
\min \quad & z - \eta + \frac{1}{\alpha}\sum_{j=1}^{m} q_j w_j \\
\text{s.t.} \quad & z \ge \mathbf{c}^T\mathbf{x}_k - \frac{1}{\alpha}\sum_{i=1}^{n} p_i[\mathbf{c}^T\mathbf{x}_k - \mathbf{c}^T\mathbf{x}_i]_+ && k = 1,\dots,n \\
& w_j \ge \eta - \mathbf{c}^T\mathbf{y}_j && j = 1,\dots,m \\
& \mathbf{c} \in C, \quad \mathbf{w} \in \mathbb{R}^m_+
\end{aligned}
\tag{35}
$$

The term $[c^T\mathbf{x}_k - c^T\mathbf{x}_i]_+$ is not linear. To obtain a MIP formulation we linearize it by introducing additional variables and constraints (a similar linearization is used in Homem-de-Mello and Mehrotra (2009)).

$$
\begin{aligned}
\min \quad & z - \eta + \frac{1}{\alpha}\sum_{j=1}^{m} q_j w_j && && (36)\\
\text{s.t.} \quad & z \ge \mathbf{c}^T\mathbf{x}_k - \frac{1}{\alpha}\sum_{i=1}^{n} p_i \upsilon_{ik} && i = 1,\dots,n, \;\; k = 1,\dots,n && (37)\\
& \upsilon_{ik} - \delta_{ik} = \mathbf{c}^T\mathbf{x}_k - \mathbf{c}^T\mathbf{x}_i && i = 1,\dots,n, \;\; k = 1,\dots,n && (38)\\
& M\beta_{ik} \ge \upsilon_{ik} && i = 1,\dots,n, \;\; k = 1,\dots,n && (39)\\
& M(1-\beta_{ik}) \ge \delta_{ik} && i = 1,\dots,n, \;\; k = 1,\dots,n && (40)\\
& \beta_{ik} \in \{0,1\} && i = 1,\dots,n, \;\; k = 1,\dots,n && (41)\\
& \upsilon \in \mathbb{R}^{n\times n}_+, \quad \delta \in \mathbb{R}^{n\times n}_+ && && (42)\\
& w_j \ge \eta - \mathbf{c}^T\mathbf{y}_j && j = 1,\dots,m && (43)\\
& B\mathbf{c} \le \mathbf{h} && && (44)
\end{aligned}
$$

$$\mathbf{w} \in \mathbb{R}^m_+ \tag{45}$$

Here M is a sufficiently large constant to make the above system feasible. Constraints (39)-(42) ensure that only one of the variables υ_{ik} and δ_{ik} is positive. Then by constraint (38) we have $\upsilon_{ik} = [\mathbf{c}^T\mathbf{x}_k - \mathbf{c}^T\mathbf{x}_i]_+$ for all pairs of i and k. The equivalence of the MIP (36)-(45) to (35) follows immediately.

Remark 6.3 *The choice of the constant M can significantly impact computational performance. In order to achieve tighter bounds, M in constraints (39) and (40) can be replaced by $M_{ki} = \max_{\mathbf{c}\in C}[\mathbf{c}^T\mathbf{x}_k - \mathbf{c}^T\mathbf{x}_i]_+$ and $\hat{M}_{ki} = \max_{\mathbf{c}\in C} \; [\mathbf{c}^T\mathbf{x}_i - \mathbf{c}^T\mathbf{x}_k]_+$, respectively.*

The above formulation (36)-(45) contains $O(n^2)$ binary variables. In the next section we show that, in the special case when scalarization vectors are non-negative and all the outcomes of $\mathbf{X}$ are equally likely, this can be reduced to $O(n)$.

6.2.3 Solving the cut generation problem in the equal probability case In this section we consider a polytope $C = \left\{\mathbf{c} \in \mathbb{R}^d_+ \; : \; B\mathbf{z} \leq \mathbf{h}\right\}$ of non-negative scalarization vectors. Since we consider larger outcomes to be preferable, the assumption of non-negativity is justified. In addition, we assume that each realization of $\mathbf{X}$ has probability $\frac{1}{n}$, and at first consider confidence levels of the form $\alpha = \frac{k}{n}$ for some $k \in \{1, \ldots, n\}$. Recalling formula (17) in Theorem 3.1 and introducing the scaled variables $\beta_i = n\gamma_i$ we have

$$\mathrm{CVaR}_{\frac{k}{n}}(\mathbf{c}^T\mathbf{X}) = \min\left\{\frac{1}{k}\sum_{i=1}^n \beta_i \mathbf{c}^T\mathbf{x}_i \quad : \quad \sum_{i=1}^n \beta_i = k, \quad \beta \in [0,1]^n\right\}.$$

The cut generation problem (34) now reads

$$\begin{aligned}
\min \quad & \frac{1}{k}\sum_{i=1}^n \beta_i \mathbf{c}^T\mathbf{x}_i - \eta + \frac{1}{\alpha}\sum_{j=1}^n q_j w_j \\
\text{s.t.} \quad & \sum_{i=1}^n \beta_i = k \\
& \beta \in [0,1]^n \\
& w_j \geq \eta - \mathbf{c}^T\mathbf{y}_j \qquad\qquad j = 1, \ldots, m \\
& \mathbf{c} \in C, \;\; \mathbf{w} \in \mathbb{R}^m_+.
\end{aligned} \tag{46}$$

We linearize the quadratic terms $\beta_i\mathbf{c}^T\mathbf{x}_i, \; i = 1, \ldots, n$, appearing in the objective function of problem (46) by introducing some additional variables and constraints. Using the notation $\delta_i = (\delta_{i1}, \ldots, \delta_{id})^T$ we obtain a MIP formulation with n binary variables.

$$\min \quad \frac{1}{k}\sum_{i=1}^n \delta_i^T\mathbf{x}_i - \eta + \frac{1}{\alpha}\sum_{j=1}^m q_j w_j \tag{47}$$

$$\text{s.t.} \quad \sum_{i=1}^n \beta_i = k \tag{48}$$

$$\beta \in \{0,1\}^n \tag{49}$$

$$0 \leq \delta_{il} \leq c_l \qquad i = 1, \ldots, n, \; l = 1, \ldots, d \tag{50}$$

$$\delta_{il} \leq M\beta_i \qquad i = 1, \ldots, n, \; l = 1, \ldots, d \tag{51}$$

$$-\delta_{il} + c_l \leq M(1 - \beta_i) \qquad i = 1, \ldots, n, \; l = 1, \ldots, d \tag{52}$$

$$w_j \geq \eta - \mathbf{c}^T\mathbf{y}_j \qquad j = 1, \ldots, m \tag{53}$$

$$B\mathbf{c} \leq \mathbf{h} \tag{54}$$

$$\mathbf{c} \in \mathbb{R}^d_+, \quad \mathbf{w} \in \mathbb{R}^m_+, \tag{55}$$

where M is again a sufficiently large constant ensuring the feasibility of the problem. It is easy to see that constraints (49)-(52) guarantee that

$$\delta_{il} = \begin{cases} c_l & \text{if } \beta_i = 1 \\ 0 & \text{if } \beta_i = 0 \end{cases}$$

holds for all $i = 1, \ldots, n$ and $l = 1, \ldots, d$. Therefore, we have $\sum_{i=1}^n \delta_i^T \mathbf{x}_i = \sum_{i=1}^n \beta_i \mathbf{c}^T \mathbf{x}_i$ which shows the equivalence of (46) and the MIP (47)-(55).

We proceed by extending the above formulation (47)-(55) to allow arbitrary confidence levels. The key observation is that for a given $\alpha \in [\frac{k}{n}, \frac{k+1}{n})$ Proposition 3.1 allows us to express $\mathrm{CVaR}_\alpha(\mathbf{c}^T\mathbf{X})$ as a convex combination of $\mathrm{CVaR}_{\frac{k}{n}}(\mathbf{c}^T\mathbf{X})$ and $\mathrm{CVaR}_{\frac{k+1}{n}}(\mathbf{c}^T\mathbf{X})$:

$$\mathrm{CVaR}_\alpha(\mathbf{c}^T\mathbf{X}) = \lambda_\alpha \,\mathrm{CVaR}_{\frac{k}{n}}(\mathbf{c}^T\mathbf{X}) + (1-\lambda_\alpha)\,\mathrm{CVaR}_{\frac{k+1}{n}}(\mathbf{c}^T\mathbf{X}),$$

where $\lambda_\alpha = \frac{k(k+1-\alpha n)}{\alpha n}$. Analogously to the previous formulation, we express $\mathrm{CVaR}_{\frac{k}{n}}$ and $\mathrm{CVaR}_{\frac{k+1}{n}}$ using the binary vectors $\beta^{(1)}$ and $\beta^{(2)}$, respectively. This leads to an alternate MIP representation of (CutGen):

$$\begin{aligned}
\min \quad & \frac{\lambda_\alpha}{k}\sum_{i=1}^n {\delta_i^{(1)}}^T \mathbf{x}_i + \frac{(1-\lambda_\alpha)}{k}\sum_{i=1}^n {\delta_i^{(2)}}^T \mathbf{x}_i - \eta + \frac{1}{\alpha}\sum_{j=1}^m q_j w_j \\
\text{s.t.} \quad & \sum_{i=1}^n \beta_i^{(1)} = k \\
& \delta_{il}^{(1)} \le M\beta_i^{(1)} && i=1,\ldots,n,\ l=1,\ldots,d \\
& 0 \le \delta_{il}^{(1)} \le c_l && i=1,\ldots,n,\ l=1,\ldots,d \\
& -\delta_{il}^{(1)} + c_l \le M(1-\beta_i^{(1)}) && i=1,\ldots,n,\ l=1,\ldots,d \\
& \beta^{(1)} \in \{0,1\}^n \\
& \sum_{i=1}^n \beta_i^{(2)} = k+1 \\
& \delta_{il}^{(2)} \le M\beta_i^{(2)} && i=1,\ldots,n,\ l=1,\ldots,d \\
& 0 \le \delta_{il}^{(2)} \le c_l && i=1,\ldots,n,\ l=1,\ldots,d \\
& -\delta_{il}^{(2)} + c_l \le M(1-\beta_i^{(2)}) && i=1,\ldots,n,\ l=1,\ldots,d \\
& \beta^{(2)} \in \{0,1\}^n \\
& w_j \ge \eta - \mathbf{c}^T\mathbf{y}_j && j=1,\ldots,m \\
& B\mathbf{c} \le \mathbf{h} \\
& \mathbf{c} \in \mathbb{R}^d_+, \quad \mathbf{w} \in \mathbb{R}^m_+.
\end{aligned} \tag{56}$$

Remark 6.4 *Similarly to the general case in Section 6.2.2, the parameter M in constraints (51) and (52), as well as their counterparts in (56), can be replaced by $M_l = \max\{c_l \;:\; \mathbf{c} \in C\}$.*

To conclude this section, we present a set of valid inequalities to strengthen the MIP formulation in (56).

Proposition 6.1 *There exists an optimal solution to the problem (56) satisfying the relations below.*

$$\sum_{i=1}^n \beta_i^{(2)} - \beta_i^{(1)} = 1$$
$$\beta^{(1)} \le \beta^{(2)}$$

Proof. Keeping in mind the knapsack structure explored in the proof of Theorem 3.1, note that in the above formulation $\mathrm{CVaR}_{\frac{k}{n}}(\mathbf{c}^T\mathbf{X})$ and $\mathrm{CVaR}_{\frac{k+1}{n}}(\mathbf{c}^T\mathbf{X})$ are expressed as the mean of k and $k+1$ smallest realizations of the random variable $\mathbf{c}^T\mathbf{X}$, respectively. The selection of realizations to be featured in these means is encoded by the binary variables $\beta^{(1)}$ and $\beta^{(2)}$. While some of the realizations $\mathbf{c}^T\mathbf{x}_1,\dots,\mathbf{c}^T\mathbf{x}_n$ might coincide, our claim immediately follows from the trivial observation that a set of k smallest realizations can always be extended to a set of $k+1$ smallest realizations by adding to it a single new realization. For example, the choice of the lexicographically smallest optimal vectors $\beta^{(1)}$ and $\beta^{(2)}$ provides a solution with the desired properties. □

6.2.4 Finding a *d*-vertex solution The provable finite convergence of Algorithm 1 depends on finding a solution to the cut generation problem which is d-vertex of the polyhedron $P(\mathbf{Y},C)$. Let $\mathbf{c}^*$ be an optimal solution obtained using one of the methods outlined in Sections 6.2.2 and 6.2.3, and let π be a permutation describing a non-decreasing ordering of the realizations of the random vector ${\mathbf{c}^*}^T\mathbf{X}$, i.e., ${\mathbf{c}^*}^T\mathbf{x}_{\pi(1)}\leq\dots\leq{\mathbf{c}^*}^T\mathbf{x}_{\pi(n)}$. Defining $k^*=\min\left\{k\in[n]\ :\ \sum_{i=1}^{k}p_{\pi(i)}\geq\alpha\right\}$ and $K^*=\{\pi(1),\dots,\pi(k^*-1)\}$, we can obtain the desired d-vertex solution $\hat{\mathbf{c}}$ by finding a vertex optimal solution $(\hat{\mathbf{c}},\hat{\eta},\hat{\mathbf{w}})$ of the linear program (FixedSet). According to Corollary 3.2 the vector $\hat{\mathbf{c}}$ is also an optimal solution of (CutGen). We remark that this step is often redundant in practice, since MIP solvers typically provide vertex solutions.

7. Computational Study In this section we demonstrate the effectiveness of our approach by presenting two numerical studies. First we examine feasible regions associated with various multivariate risk constraints on an illustrative example. Then we evaluate the effectiveness of our optimization models and solution methods by applying them to a homeland security budget allocation problem.

We used MATLAB® 7.11.0 to generate data and perform supporting calculations, AMPL (Fourer et al., 2003) to formulate models and implement solution methods, and `CPLEX 11.2` (ILOG, 2008) to solve optimization problems. All experiments were carried out on a single core of an HP Linux workstation with two Intel® Xeon®W5580 3.20 GHz CPUs and 32 GB of memory.

7.1 A small-scale study of feasibility regions In this section we present a simple two-dimensional problem to illustrate feasible regions associated with multivariate CVaR constraints, along with the effects of various parameter choices. The problem originally appeared in Hu et al. (2011a), where the authors compare the feasible regions associated with various multivariate SSD constraints: *positive linear dominance*, *weak stochastically weighted dominance*, *stochastically weighted dominance with chance*, and *relaxed strong stochastically weighted dominance*. We chose to explore the same numerical example, as this allows a direct comparison between our CVaR constraints and the dominance concepts mentioned above.

Consider the probability space $(\Omega,2^\Omega,\Pi)$ where $\Omega=\{\omega_1,\omega_2\}$ and $\Pi(\omega_1)=\Pi(\omega_2)=\frac{1}{2}$. Let $\Delta:\Omega\to\mathbb{R}$ denote the random variable with realizations $\Delta(\omega_1)=1$, $\Delta(\omega_2)=-1$, and let

$$\Gamma=\begin{bmatrix}1+0.25\Delta & 0.5\\ 0.5 & 0.5-0.25\Delta\\ 0.25 & 0.03\end{bmatrix},\quad \mathbf{Y}^{(1)}=\begin{bmatrix}0.5-0.0025\Delta\\ 0.4\\ 0.1+0.013\Delta\end{bmatrix},\quad \mathbf{Y}^{(2)}=\begin{bmatrix}0.05\\ 0.2-0.025\Delta\\ 0.01+0.013\Delta\end{bmatrix}.$$

In addition, we define the scalarization polyhedra

$$C_\vartheta=\left\{(c_1,c_2,c_3)\in\mathbb{R}^3\ :\ c_1+c_2+c_3=1,\ c_1\geq\vartheta,\ c_2\geq\vartheta,\ c_3\geq\vartheta\right\},\qquad \vartheta\in\left[0,\frac{1}{3}\right].$$

Note that C_0 is the simplex used to define positive linear dominance, while $C_{\frac{1}{3}}$ consists of the single scalarization vector $\left(\frac{1}{3},\frac{1}{3},\frac{1}{3}\right)$. We are interested in feasibility regions defined by the constraints of the

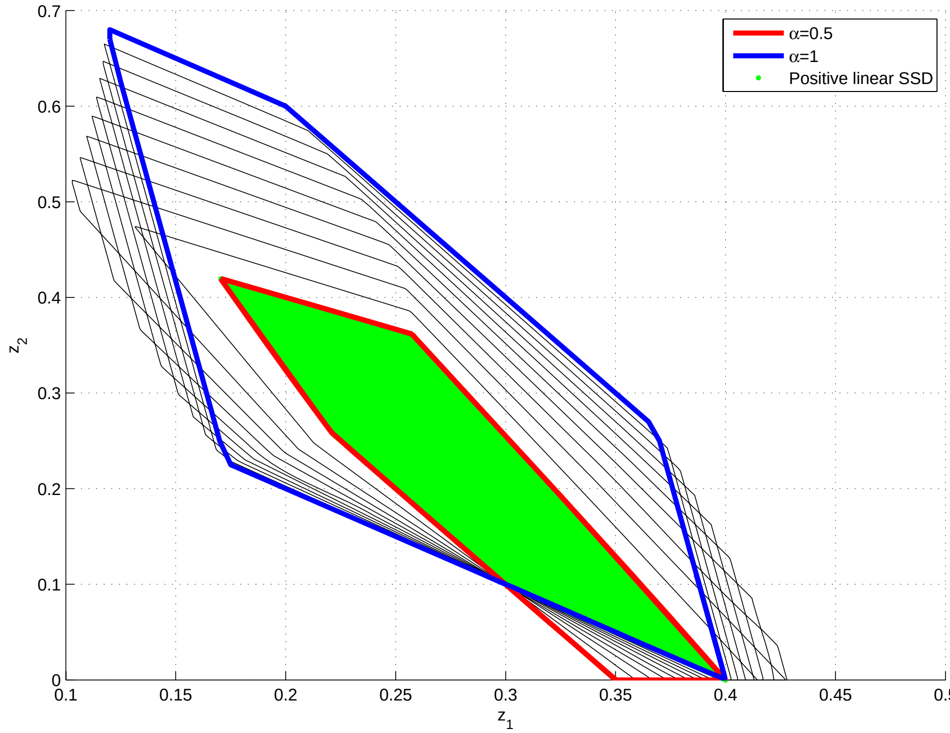

Figure 2: Feasible regions for $\alpha^{(1)} = \alpha^{(2)} \in [0.5, 1]$ and $\vartheta = 0$

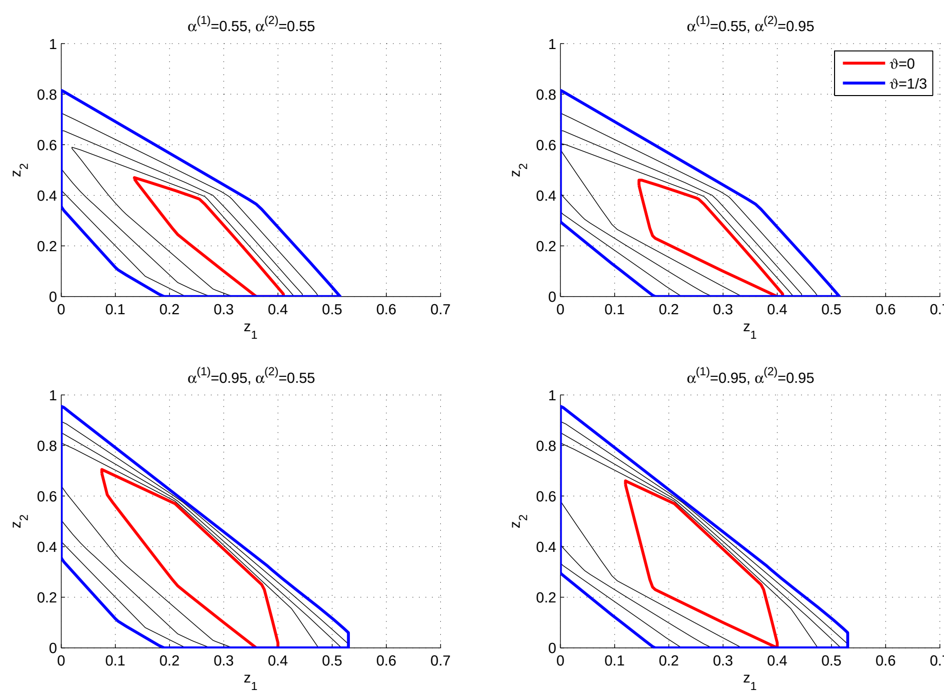

Figure 3: Feasible regions for various choices of $\alpha^{(1)}$ and $\alpha^{(2)}$ with $\vartheta \in \left[0, \frac{1}{3}\right]$

form

$$\begin{aligned} -\Gamma \mathbf{z} &\succcurlyeq^{C_\vartheta}_{\mathrm{CVaR}_{\alpha^{(1)}}} -\mathbf{Y}^{(1)} \\ \Gamma \mathbf{z} &\succcurlyeq^{C_\vartheta}_{\mathrm{CVaR}_{\alpha^{(2)}}} \mathbf{Y}^{(2)} \\ \mathbf{z} &\in \mathbb{R}^2_+, \end{aligned}$$

where $\mathbf{z} = (z_1, z_2)$ is a decision vector. Figure 2 shows the feasible regions associated with the scalarization polyhedron C_0 and confidence levels $\alpha^{(1)} = \alpha^{(2)}$ changing between 0.5 and 1. Note that these regions are not nested, i.e., CVaR-preferability at a certain confidence level does not imply preferability at other levels. In accordance with Corollary 2.1, the intersection of these regions (filled area) corresponds to the region associated with the *positive linear SSD constraint* (compare with Figure 2(a) in Hu et al. (2011b)). Figure 3 illustrates shapes of feasible regions obtained by various combinations of $\alpha^{(1)}$ and $\alpha^{(2)}$, for a range of ϑ values between 0 and $\frac{1}{3}$. Note that $\vartheta_1 \leq \vartheta_2$ implies $C_{\vartheta_1} \supset C_{\vartheta_2}$, therefore CVaR-preferability with respect to C_{ϑ_1} implies preferability with respect to C_{ϑ_2}. This results in a nested structure between the corresponding feasible regions.

Further customization of the feasible region can be achieved by requiring CVaR constraints to hold at multiple different confidence levels, and with respect to different corresponding scalarization polyhedra, for each reference variable.

7.2 Homeland security budget allocation To explore the computational performance of our methods, along with the impact of various polyhedral CVaR constraints, we examine a budget allocation problem. This problem was presented in Hu et al. (2011b) with polyhedral SSD constraints in a homeland security context, and also inspired the numerical study in Armbruster and Luedtke (2010). Our exposition below closely follows that in Hu et al. (2011b), replacing the SSD constraints with CVaR-based ones. The model concerns the allocation of a fixed budget to ten urban areas (New York, Chicago, etc.). The budget is used for prevention, response, and recovery from national catastrophes. The risk share of each area is defined based four criteria: property losses, fatalities, air departures, and average daily bridge traffic. Accordingly, we consider a random *risk share matrix* $A : \Omega \to \mathbb{R}^{4\times 10}_+$, where the entry $A_{ij} : \Omega \to \mathbb{R}$ denotes, for criterion i, the proportion of losses in urban area j relative to the total losses. The penalty for allocations under the risk share is expressed by the *budget misallocation functions* $\mathcal{M}_i : Z \to \mathcal{V}(\Omega, 2^\Omega, \Pi)$ defined as

$$M_i(\mathbf{z}) = \sum_{j=1}^{10} [A_{ij} - z_j]_+ \quad \text{for each criterion } i = 1, \ldots, 4,$$

where $Z = \left\{\mathbf{z} \in \mathbb{R}^{10}_+ \ : \ \|\mathbf{z}\|_1 = 1\right\}$ denotes the set of all feasible allocations. Let us also introduce the notation $\mathbf{M} = (M_1, M_2, M_3, M_4)^T$.

We consider two benchmark solutions: one based on average government allocations by the Department of Homeland Security Urban Areas Security Initiative, and one based on suggestions in the RAND report by Willis et al. (2005). These benchmark allocations are denoted by $\mathbf{z}^G$ and $\mathbf{z}^R$, respectively. The scalarization polyhedron is of the form $C = \left\{\mathbf{c} \in \mathbb{R}^4 \ : \ \|\mathbf{c}\|_1 = 1,\ c_i \geq c_i^* - \frac{\theta}{3}\right\}$, where $\mathbf{c}^* \in \mathbb{R}^4$ is a *center* satisfying $\|\mathbf{c}^*\|_1 = 1$, and $\theta \in [0,1]$ is a constant for which $\frac{\theta}{3} \leq \min\limits_{i\in\{1,\ldots,4\}} c_i^*$ holds. It is easy to see that if θ is positive, the polyhedron C is a 3-dimensional simplex. Denoting the vertices of C by $\hat{\mathbf{c}}_{(1)}, \ldots, \hat{\mathbf{c}}_{(4)}$ the objective function of the budget allocation problem, based on a robust approach, is defined as

$$f(\mathbf{z}) = \max_{k\in\{1,\ldots,4\}} \mathbb{E}\left(\hat{\mathbf{c}}_{(k)}^T \mathbf{M}(\mathbf{z})\right).$$

Selecting two finite sets of confidence levels $\mathcal{A}^G, \mathcal{A}^R \subset (0,1]$ we introduce an optimization problem with multivariate CVaR constraints:

$$\begin{aligned} \max \quad & -f(\mathbf{z}) \\ \text{s.t.} \quad & -\mathbf{M}(\mathbf{z}) \succcurlyeq^{C}_{\mathrm{CVaR}_\alpha} -\mathbf{M}(\mathbf{z}^G) \qquad \alpha \in \mathcal{A}^G \\ & -\mathbf{M}(\mathbf{z}) \succcurlyeq^{C}_{\mathrm{CVaR}_\alpha} -\mathbf{M}(\mathbf{z}^R) \qquad \alpha \in \mathcal{A}^R \\ & \mathbf{z} \in Z. \end{aligned} \tag{57}$$

Note that the negative signs were added in order to be consistent with our convention of preferring large values. To keep the exposition concise, we refer the reader to Hu et al. (2011b, Section 4) for a description of how the objective function f can be linearized, along with the explicit construction of the benchmarks $\mathbf{z}^G$, $\mathbf{z}^R$ and the realizations of the risk share matrix A. Unless otherwise specified, we consider the "base case" with the choices of $\theta = 0.25$, the *equality center* $\mathbf{c}^* = (\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$, and additional correlation and volatility parameters $\pi = \frac{1}{2}$, $\gamma = 3$.

7.2.1 Computational performance We use the cut generation method outlined in Section 6.2 to solve problem (57) in the case when each scenario in $\Omega = \{\omega_1, \ldots, \omega_n\}$ has probability $\frac{1}{n}$, and all confidence levels in $\mathcal{A}^G$ and $\mathcal{A}^R$ are chosen from the set $\{\frac{1}{n}, \ldots, \frac{n}{n}\}$. When necessary, the confidence levels in our tables were rounded up to a multiple of $\frac{1}{n}$ during computation. Note that we have a separate cut generation problem for each pair of a benchmark and an associated confidence level. Under our assumptions all of these cut generation problems take the form of the MIP (47)-(55). All numerical results in Sections 7.2.1-7.2.2 were obtained using batch sampling, averaging over 5 samples.

Table 1 shows the computational performance of our implementation when solving problem (57) with a single CVaR constraint based on the RAND benchmark ($\mathcal{A}^G = \emptyset$, $\mathcal{A}^R = \{\alpha\}$). We report the total number of cuts, including those introduced in the initialization step (associated with the four vertices of the scalarization polyhedron). Additional cuts are generated in each iteration except the final one, at which the algorithm terminates by proving optimality. While solving the master problem is nearly instantaneous, as the number of scenarios increases, solving the cut generation MIP becomes a computational bottleneck. It is interesting to note that CPU times are typically higher for $\alpha = 0.05$ than for $\alpha = 0.01$ when solving otherwise identical problems. The reason lies in the increased combinatorial complexity of the cut generation MIP, which involves selecting $\lceil \alpha n \rceil$ binary variables (out of a total of n) to take value 1. This point is further illustrated by Figure 4(a), which shows that CPU times are significantly lower for α values near the endpoints of the interval $[0, 1]$ despite generating a similar number of cuts. By contrast, for a fixed value of α, considering larger scalarization sets by increasing θ results in a higher number of cuts and a proportional increase of CPU time; see Figure 4(b).

7.2.2 Numerical study on the effect of risk constraints We now look at optimal solutions of problem (57) and its SSD-constrained counterpart, along with an "unconstrained" variant of the problem which features no risk constraints. To keep our presentation simple, for the purposes of discussing allocation results we have divided the set of urban areas into three groups:

- New York (highest risk);
- Chicago, Bay Area, Washington DC-MD-VA-WV, and Los Angeles-Long Beach (medium risk);
- Philadelphia PA-NJ, Boston MA-NH, Houston, Newark, and Seattle-Bellevue-Everett (lower risk).

Figure 5 shows optimal results for problem (57) with CVaR preferability required over the benchmark $\mathbf{z}^R$ at a single confidence level of 0.1, along with solutions of SSD-constrained and unconstrained versions of the problem. As the parameter θ increases, the scalarization set becomes larger, leading to more restrictive constraints. Accordingly, as illustrated in Figure 5(a), optimal objective values of the CVaR- and SSD-constrained problems diverge sharply from that of the unconstrained version. We observe that

α	n	Number of Cuts			Number of		CPU Time (sec)		
		Total	Initial	CutGen	Iterations	MIP solved	CutGen	Total	CutGen/Total
0.01	50	5.2	4	1.2	2.2	2.2	0.14	0.34	40.356%
	100	5.2	4	1.2	2.2	2.2	0.36	1.10	33.809%
	150	5	4	1	2	2	23.60	25.15	93.889%
	200	5.2	4	1.2	2.2	2.2	56.48	58.82	96.054%
	250	5	4	1	2	2	6312.35	6315.66	99.948%
	500	4.6	4	0.6	1.6	1.6	11507.10	11528.40	99.792%
0.05	50	5	4	1	2	2	12.19	12.41	98.213%
	100	4.8	4	0.8	1.8	1.8	5244.96	5245.48	99.990%
	150	4	4	0	1	1	3921.22	3922.16	99.976%
	200	4	4	0	1	1	5004.41	5006.03	99.968%
	250	4	4	0	1	1	6021.99	6024.31	99.962%
	500	5	4	1	2	2	14386.69	14413.13	99.817%

Table 1: Computational performance of the cut generation algorithm for a single CVaR constraint

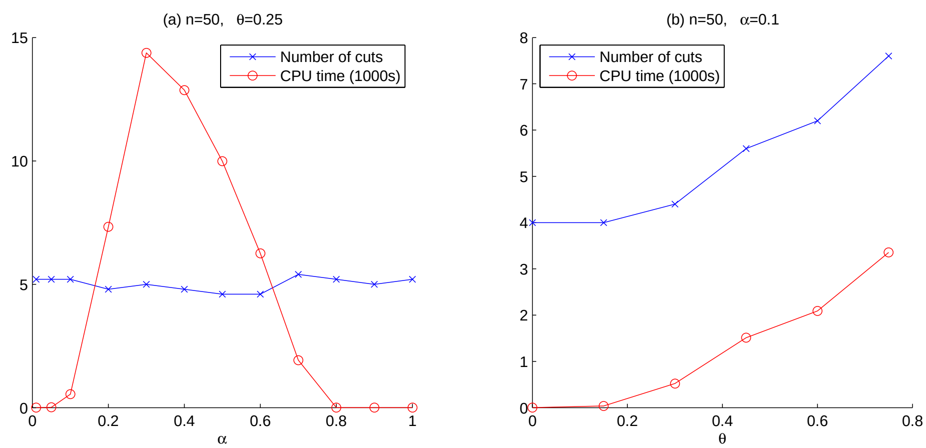

Figure 4: Computational performance of cut generation algorithm for a single benchmark

while the budget allocated to urban areas with medium risk remains relatively unchanged in all three models, under CVaR and SSD constraints there is a significant tradeoff between allocations to New York and areas with lower risk. It is interesting to note that enforcing the CVaR constraint at a single confidence level yields results very close to those obtained under SSD constraints, although the difference between the two models becomes more pronounced for larger values of θ.

We next present results for problem (57) with CVaR constraints on both benchmarks $\mathbf{z}^G$ and $\mathbf{z}^R$, enforced at multiple common confidence levels ($\mathcal{A}^G = \mathcal{A}^R = \mathcal{A}$). While problems requiring (weak) preference over a single benchmark solution are always feasible, this is not necessarily the case when considering multiple benchmarks. A natural approach to overcome this issue is to relax risk constraints by introducing a tolerance parameter ι, as described in part (ii) of Proposition 2.1. In accordance with Hu et al. (2011b), we set $\iota = 0.005$. We remark that smaller values of ι typically result in infeasible SSD-constrained problems, and at some confidence level settings we encounter infeasibility in certain problem instances even under CVaR constraints. Table 2 contains our results for the relaxed two-benchmark problem. We can see that enforcing CVaR constraints at low confidence levels yields solutions close to the unconstrained allocations, while requiring them to hold at both ends of the spectrum results

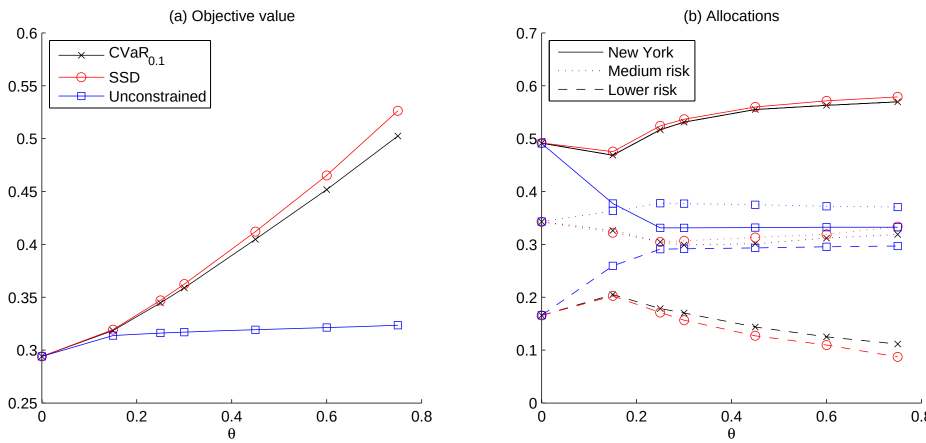

Figure 5: Optimal results for benchmark $\mathbf{z}^R$ and $n = 100$

in convergence to the SSD-constrained solution. Although the latter fact is not surprising given the equivalence established in Proposition 2.1, it is interesting to note that simply requiring CVaR to hold at the lowest and highest levels (corresponding to worst case- and expectation-based constraints) already leads to a close approximation of the SSD constraint. In line with the conclusions reached by Hu et al. (2011b) we finally observe that the budget allocated to New York, the area with the highest risk, gradually increases with the introduction of additional risk constraints (from 32.9% in the unconstrained case to a maximum of 49.3% under SSD).

Confidence levels		Objective	New York	Medium risk	Lower risk
Unconstrained model	($\mathcal{A} = \emptyset$)	0.316245	0.331189	0.377809	0.291002
0.01		0.316245	0.331189	0.377809	0.291002
0.01,0.05		0.316245	0.331189	0.377809	0.291002
0.01,0.05,0.1		0.321030	0.388341	0.355589	0.256069
0.01,0.05,0.1,0.2		0.329228	0.447436	0.332685	0.219879
0.01,0.05,0.1,0.2,0.3		0.332944	0.467037	0.326971	0.205992
0.01,	1	0.338968	0.490263	0.327498	0.182239
0.01,0.05,	1	0.338968	0.490263	0.327498	0.182239
0.01,0.05,0.1,	1	0.338968	0.490263	0.327498	0.182239
0.01,0.05,0.1,0.2,	1	0.338968	0.490263	0.327498	0.182239
0.01,0.05,0.1,0.2,0.3,	1	0.338968	0.490263	0.327498	0.182239
0.01,0.05,0.1,0.2,0.3,	0.9, 1	0.339043	0.490795	0.326833	0.182372
0.01,0.05,0.1,0.2,0.3,	0.8, 0.9, 1	0.339101	0.490891	0.327012	0.182097
0.01,0.05,0.1,0.2,0.3,	0.7, 0.8, 0.9, 1	0.339101	0.490891	0.327012	0.182097
SSD-constrained model	($\mathcal{A} = (0, 1]$)	0.339167	0.491158	0.327025	0.181817

Table 2: Optimal objective and allocations for two benchmarks, $n = 100$, $\theta = 0.25$ and $\iota = 0.005$

8. Conclusion and future research We have introduced new multivariate risk-averse preference relations based on CVaR and linear scalarization, referred to as *polyhedral* CVaR *constraints*. We have demonstrated that they provide an efficient and computationally tractable way of relaxing multivariate stochastic dominance constraints. Additionally, we have illustrated that the flexibility of our approach allows for modeling a wide range of risk preferences. In particular, unlike existing SSD-based relations, the ability to specify confidence levels allows us to focus on various aspects of the distribution (including the tails, expectation, and worst case behavior) separately or in arbitrary combinations. We have shown that our framework can be extended from CVaR to a wider class of coherent risk measures, including mixed CVaR risk measures.

We have incorporated polyhedral CVaR constraints into optimization problems, providing a novel way of modeling risk preferences in stochastic multi-criteria decision making. We have developed a finitely convergent cut generation algorithm to solve such problems on finite probability spaces. Under certain linearity assumptions we have formulated the master problem as a linear program, and the cut generation problem as a MIP, solvable by off-the-shelf software such as CPLEX. We have applied our solution methods to a budget allocation problem featuring CVaR constraints at multiple confidence levels for two benchmark solutions, and compared our results to those obtained by an SSD-based approach. While problem instances featuring up to 500 scenarios were found to be tractable, solving our MIP formulations increasingly became a computational bottleneck. Developing valid inequalities and heuristics which lead to more efficient solution of these MIPs is the topic of future research. In addition, utilizing CVaR-based Kusuoka representations, such advances could also be crucial to solving large-scale problems with polyhedral constraints featuring other coherent risk measures.

Acknowledgment. The first author has been supported by The Scientific and Technological Research Council of Turkey (TUBITAK) Career Award # 111M543 and the second author has been funded by TUBITAK-2216 Research Fellowship Programme.

References

Acerbi, C. (2002). Spectral measures of risk: a coherent representation of subjective risk aversion. *Journal of Banking and Finance*, 26(7):1505–1518.

Armbruster, B. and Luedtke, J. (2010). Models and formulations for multivariate dominance constrained stochastic programs. http://www.optimization-online.org/DB_FILE/2010/05/2613.pdf.

Artzner, P., Delbaen, F., Eber, J., and Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, 9(3):203–228.

Avis, D. and Fukuda, K. (1992). A pivoting algorithm for convex hulls and vertex enumeration of arrangements and polyhedra. *Discrete & Computational Geometry*, 8(1):295–313.

Bertsimas, D. and Brown, D. B. (2009). Constructing uncertainty sets for robust linear optimization. *Operations Research*, 57(6):1483–1495.

Dentcheva, D. and Ruszczyński, A. (2003). Optimization with stochastic dominance constraints. *SIAM Journal on Optimization*, 14(2):548–566.

Dentcheva, D. and Ruszczyński, A. (2006). Portfolio optimization with stochastic dominance constraints. *Journal of Banking and Finance*, 30(2):433–451.

Dentcheva, D. and Ruszczyński, A. (2009). Optimization with multivariate stochastic dominance constraints. *Mathematical Programming*, 117:111–127.

Fabian, C. I. and Veszpremi, A. (2008). Algorithms for handling cvar-constraints in dynamic stochastic programming models with applications to finance. *The Journal of Risk*, 10:111–131.

Fourer, R., Gay, D., and Kernighan, B. (2003). *AMPL: A Modeling Language for Mathematical Programming Second Edition.* Duxbury Press Brooks Cole Publishing Co.

Gollmer, R., Gotzes, U., and Schultz, R. (2011). A note on second-order stochastic dominance constraints induced by mixed-integer linear recourse. *Mathematical Programming*, 216(1):179–190.

Gollmer, R., Neise, F., and Schultz:, R. (2008). Stochastic programs with first-order dominance constraints induced by mixed-integer linear recourse. *SIAM Journal on Optimization*, 19(2):552–571.

Homem-de-Mello, T. and Mehrotra, S. (2009). A cutting surface method for uncertain linear programs with linear stochastic dominance constraints. *SIAM Journal on Optimization*, 20(3):1250–1273.

Hu, J., Homem-de-Mello, T., and Mehrotra, S. (2010). Sample average approximation of stochastic dominance constrained programs. *Mathematical Programming, Online First (October 2010)*. dx.doi.org/10.1007/s10107-010-0428-9.

Hu, J., Homem-de-Mello, T., and Mehrotra, S. (2011a). Concepts and applications of stochastically weighted stochastic dominance. http://www.optimization-online.org/DB_FILE/2011/04/2981.pdf.

Hu, J., Homem-de-Mello, T., and Mehrotra, S. (2011b). Risk-adjusted budget allocation models with application in homeland security. *IIE Transactions*, 43(12):819–839.

ILOG (2008). *ILOG AMPL CPLEX System Version 11 User's Guide*. ILOG CPLEX Division.

Jouini, E., Schachermayer, W., and Touzi, N. (2006). Law-invariant risk measures have the Fatou property. *Advances in Mathematical Economics*, 9(1):49–71.

Khachiyan, L., Boros, E., Borys, K., Elbassioni, K. M., and Gurvich, V. (2008). Generating all vertices of a polyhedron is hard. *Discrete & Computational Geometry*, 39(1-3):174–190.

Kusuoka, S. (2001). On law invariant coherent risk measures. *Advances in Mathematical Economics*, 3:83–95.

Lehmann, E. (1955). Ordered families of distributions. *Annals of Mathematical Statistics*, 26(3):399–419.

Leshno, M. and Levy, H. (2002). Preferred by "all" and preferred by "most" decision makers: Almost stochastic dominance. *Management Science*, 48(8):1074–1085.

Lizyayev, A. and Ruszczyński, A. (2011). Tractable almost stochastic dominance. *European Journal of Operational Research*, 218(2):448–455.

Luedtke, J. (2008). New formulations for optimization under stochastic dominance constraints. *SIAM Journal on Optimization*, 19(3):1433–1450.

Mann, H. and Whitney, D. (1947). On a test of whether one of two random variables is stochastically larger than the other. *Annals of Mathematical Statistics*, 18:50–60.

Müller, A. and Stoyan, D. (2002). *Comparison Methods for Stochastic Models and Risks*. John Wiley & Sons, Chichester.

Nie, Y., Wu, X., and Homem-de-Mello, T. (2011). Optimal path problems with second-order stochastic dominance constraints. *Networks and Spatial Economics, Online First (October 2011)*. dx.doi.org/10.1007/s11067-011-9167-6.

Noyan, N. (2010). Alternate risk measures for emergency medical service system design. *Annals of Operations Research*, 181(1):559–589.

Noyan, N., Rudolf, G., and Ruszczyński, A. (2008). Relaxations of linear programming problems with first order stochastic dominance constraints. *Operations Research Letters*, 34(6):653–659.

Noyan, N. and Ruszczyński, A. (2008). Valid inequalities and restrictions for stochastic programming problems with first order stochastic dominance constraints. *Mathematical Programming, Ser. A*, 114(2):249–275.

Pflug, G. C. (2000). Some remarks on the value-at-risk and the conditional value-at-risk. In Uryasev, S., editor, *Probabilistic Constrained Optimization: Methodology and Applications*. Kluwer Academic Publishers, Dordrecht.

Pflug, G. C. and Römisch, W. (2007). *Modelling, managing and measuring risk*. World Scientific publishing, Singapore.

Prékopa, A. (2012). Multivariate value at risk and related topics. *Annals of Operations Research*, 193:49–69.

Rockafellar, R. (1970). *Convex Analysis*. Princeton University Press, Princeton.

Rockafellar, R. (2007). Coherent approaches to risk in optimization under uncertainty. In *Tutorials in Operations Research*, pages 38–61. INFORMS.

Rockafellar, R. and Uryasev, S. (2000). Optimization of conditional value-at-risk. *The Journal of Risk*, 2(3):21–41.

Rockafellar, R. and Uryasev, S. (2002). Conditional value-at-risk for general loss distributions. *Journal of Banking and Finance*, 26(7):1443–1471.

Rudolf, G. and Ruszczyński, A. (2008). Optimization problems with second order stochastic dominance constraints: duality, compact formulations, and cut generation methods. *SIAM Journal on Optimization*, 19:1326–1343.

Shaked, M. and Shanthikumar, J. G. (1994). *Stochastic Orders and their Applications*. Associated Press, Boston.

Shapiro, A., Dentcheva, D., and Ruszczyński, A. (2009). *Lectures on stochastic programming: modeling and theory*. The society for industrial and applied mathematics and the mathematical programming society, Philadelphia, USA.

Street, A. (2009). On the conditional value-at-risk probability-dependent utility function. *Theory and Decision*, 68(1-2):49–68.

Uryasev, S. (2000). Conditional value-at-risk: Optimization algorithms and applications. *Financial Engineering News*.

Willis, H. H., Morral, A. R., Kelly, T. K., and Medby, J. J. (2005). Estimating terrorism risk. Technical report, The RAND Corporation, Santa Monica, CA.