Scenario Reduction for Risk-Averse Stochastic Programs

Sebastián Arpón*[1], Tito Homem-de-Mello †[1], Bernardo Pagnoncelli‡[1]

[1]School of Business, Universidad Adolfo Ibañez, Santiago, Chile

Abstract

In this paper we discuss scenario reduction methods for risk-averse stochastic optimization problems. Scenario reduction techniques have received some attention in the literature and are used by practitioners, as such methods allow for an approximation of the random variables in the problem with a moderate number of scenarios, which in turn make the optimization problem easier to solve. The majority of works for scenario reduction are designed for classical risk-neutral stochastic optimization problems; however, it is intuitive that in the risk-averse case one is more concerned with scenarios that correspond to high cost. By building upon the notion of *effective scenarios* recently introduced in the literature, we formalize that intuitive idea and propose a scenario reduction technique for stochastic optimization problems where the objective function is a Conditional Value-at-Risk. The numerical results presented with problems from the literature illustrate the performance of the method and indicates the general cases where we expect it to perform well.

1 Introduction

Scenario generation and scenario reduction methods have a long history in stochastic programming. The importance of such methods stems from the need to approximate the (joint) probability distribution of the random variables in the problem with another distribution with a moderate number of outcomes — henceforth called scenarios — so that the resulting optimization problem can be efficiently solved, perhaps by means of some algorithm that decomposes the problem across scenarios. Here we make a distinction between scenario generation and scenario reduction: in the former, scenarios are chosen from the original distributions, whereas the latter starts with a finite set of scenarios and attempts to find a smaller subset. In both cases the goal is the same — to obtain an approximating problem that provides a good approximation to the original one. We will discuss shortly what is meant by a "good approximation".

Many approaches for scenario generation/reduction have been developed in the literature. These include, for example, clustering (e.g., Dupačová, Consigli, and Wallace (2000)) and moment-matching techniques, see Hoyland and Wallace (2001); Hoyland, Kaut, and Wallace (2003) and Mehrotra and Papp (2014). Another approach is based on generating scenarios via sampling from the underlying probability distributions, as in the case of the Sample Average Approximation (SAA) approach; overviews of sampling-based methods together with their properties can be found in Shapiro (2003) and Homem-de-Mello and Bayraksan (2014).

*sebastian.arpon@edu.uai.cl

†tito.hmello@uai.cl; Corresponding author

‡bernardo.pagnoncelli@uai.cl

Another class of methods, based on probability metrics, aims at finding a distribution Q with relatively few scenarios in such a way that Q minimizes a distance $d(P,Q)$ between Q and the original distribution P. Typically, these approaches rely on stability results that ensure that the difference between the optimal values of the original and approximating problems is bound by a constant times $d(P,Q)$. Several distances for distributions can be used for that purpose, such as the Wasserstein distance and the Fortet-Mourier metric. This type of approach has gained attention especially for the case of multistage problems, where the goal is to derive a scenario tree that properly approximates the original problem, though more sophisticated distances are required in that case. We refer to (Pflug, 2001; Dupačová, Gröwe-Kuska, & Römisch, 2003; Heitsch & Römisch, 2003, 2009; Pflug & Pichler, 2011) and references therein for further discussions on this type of methods.

The vast majority of scenario generation/reduction methods found in the literature assume that the underlying optimization problem has the form

$$\min_{x\in X} \mathbb{E}\left[G(x,\xi)\right], \tag{1.1}$$

where ξ is a random vector representing the uncertainty in the problem, and G is a cost function that depends on ξ as well as on the decision variables x (we shall assume, as customary in the literature, that G is convex in x and X is a convex closed set). Such a formulation assumes that the decision-maker is *risk-neutral*, i.e., large losses in one scenario can be offset by large gains in another one. In many situations, however, the decision-maker is *risk-averse* and wants to protect against the risk of (large) losses. Such problems are modeled by replacing the expectation in (1.1) with a proper risk measure. Optimization problem with risk aversion have been extensively studied in the literature in the past 10-20 years, both from theoretical and application perspectives.

The "explosion" of papers in risk-averse optimization, however, has not been accompanied by a similar effort in the scenario generation/reduction literature. This is due in part to the fact that some risk measures can be written in terms of expectations; a prominent example is the Conditional-Value-at-Risk (CVaR), also called Average-Value-at-Risk, which we review briefly here for completeness and to set up some notation that will be used throughout the paper. The Value at Risk (VaR) of a random variable Z with cdf $F(\cdot)$ with risk level $\alpha \in [0,1]$ is defined as

$$\mathrm{VaR}_\alpha[Z] := \min\{t \mid F(t) \geq \alpha\} = \min\{t \mid P(Z \leq t) \geq \alpha\}. \tag{1.2}$$

The Conditional Value-at-Risk (CVaR) of a random variable Z with cdf $F(\cdot)$ with risk level $\alpha \in [0,1]$ is defined as

$$\mathrm{CVaR}_\alpha[Z] := \frac{1}{1-\alpha}\int_\alpha^1 \mathrm{VaR}_\gamma[Z]d\gamma. \tag{1.3}$$

A key result in Rockafellar and Uryasev (2000) is the proof that CVaR can be expressed as the optimal value of the following optimization problem:

$$\mathrm{CVaR}_\alpha[Z] = \min_{\eta\in\mathbb{R}} \left\{\eta + \frac{1}{1-\alpha}\mathbb{E}\left[(Z-\eta)_+\right]\right\}, \tag{1.4}$$

where $(a)_+ := \max(a,0)$. Moreover, they show that $\mathrm{VaR}_\alpha[Z]$ is an optimal solution of the optimization problem in (1.4). [1]

Suppose now that that we want to solve a CVaR-version of problem (1.1), i.e.,

$$\min_{x\in X} \mathrm{CVaR}_\alpha\left[G(x,\xi)\right]. \tag{1.5}$$

[1] A useful property of CVaR_α that will be used in the sequel is that it coincides with the expectation when $\alpha = 0$. Moreover, $\lim_{\alpha\to 1} \mathrm{CVaR}_\alpha[Z] = \text{ess sup }(Z)$.

Such problems have received considerable attention in the literature, starting with the seminal work by Rockafellar and Uryasev (2000); see also Miller and Ruszczynski (2011) for the case where the problem is a two-stage stochastic program, and Noyan (2012) and Pineda and Conejo (2010) for discussions of applications of such models in humanitarian logistics and energy planning, respectively[2]. Then, by using representation (1.4), we can reformulate (1.5) as

$$\min_{x\in X,\ \eta\in\mathbb{R}} \left\{ \eta + \frac{1}{1-\alpha}\mathbb{E}_P\left[(G(x,\xi)-\eta)_+\right]\right\}. \tag{1.6}$$

We see that (1.6) is an optimization problem with expectations, though in an extended space. Thus, we can in principle apply any of the scenario generation/reduction techniques discussed earlier. Note also that formulation (1.6) makes it explicit the underlying distribution P used to calculate the CVaR in (1.5).

A closer look at the objective function in (1.6), however, shows why standard scenario generation/reduction methods may not work so well in that context. To see that, observe first that for any given x, the optimal $\eta = \eta(x)$ is given by $\eta(x) = \mathrm{VaR}_\alpha\left[G(x,\xi)\right]$; see Rockafellar and Uryasev (2000). Also, for any fixed x, the value of $\mathbb{E}\left[(G(x,\xi)-\eta(x))_+\right]$ is zero for all scenarios ξ such that $G(x,\xi) \le \eta(x)$, so such scenarios do not contribute to the calculation. In particular, at an optimal solution (x^*,η^*) to (1.6) we have that only the scenarios ξ such that $G(x^*,\xi) > \eta(x^*) = \mathrm{VaR}_\alpha\left[G(x^*,\xi)\right]$ contribute to the optimal objective function value. So, one could argue that generating scenarios outside the tail of $G(x^*,\cdot)$ is actually wasteful.

As presented, the above argument is of course heuristic and raises many questions. For example, while it is true that a scenario $\bar\xi$ such that $G(x^*,\bar\xi) \le \eta(x^*)$ does not contribute to the optimal objective function value, it could certainly be the case that $G(x,\bar\xi) > \eta(x))$ for values of $x \neq x^*$, in which case $\bar\xi$ does contribute to the evaluation of the objective function outside the optimal solution — so it is not clear whether such $\bar\xi$ can really be discarded. Nevertheless, the idea that only the scenarios "in the tail" must be generated has gained some traction in the literature, often in a heuristic way; examples are Pineda and Conejo (2010) and García-Bertrand and Mínguez (2014). Fairbrother, Turner, and Wallace (2015) provide a more solid argument by defining precisely the notion of *risk regions*. We review these works in more detail in Section 3.

In this paper we present a scenario reduction method for risk-averse problems of the form (1.5), in which the distribution of the underlying random vector ξ has finite support $\Xi := \{\xi_1, \ldots, \xi_n\}$ with a *moderately large* number of elements. A typical situation where this occurs is when a large random sample is drawn from the original distribution in order to represent that distribution well. We call it the *source distribution* and will represent it by P, with P_i denoting the probability of scenario ξ_i. One then aims at reducing the number of scenarios of the source distribution further so that the optimization problem can be solved in a reasonable time. While our work has some similarities with the aforementioned papers that address the same problem — in the sense that we also aim at generating scenarios in the tail of $G(x,\cdot)$ — we define precisely what is the set of scenarios that matter for the problem and how to generate a representative subset of those scenarios.

Our method relies strongly on the concept of *effective scenarios* recently introduced by Rahimian, Bayraksan, and Homem-de Mello (2018). According to that notion, a scenario is effective if its removal from the problem leads to changes in the optimal objective value (there is a precise definition for "removal"); otherwise it is called ineffective. Thus, our method tries to (i) identify the ineffective scenarios for (1.5), as those are provably unnecessary since the optimal value does not change if they are removed, and (ii) find a distribution whose support is a subset of the set of effective scenarios, and which minimizes the distance to the source distribution.

[2]Models with CVaR have also been used in the context of multistage stochastic programs, but we do not review them here as we focus on the static case in this paper.

We apply our method to the class of two-stage stochastic programs in which randomness appears only in the right-hand side. In such cases we are able to provide some heuristic techniques to further speed up the scenario reduction algorithm. We illustrate our methodology with a number of CVaR-versions of problems from the literature, where we compare the results obtained with our method to those obtained with traditional scenario generation techniques. The comparisons show that our approach typically performs well, in some cases it yields a much more accurate solution for the same number of scenarios.

2 Background material

In this section, for the sake of completeness — and to set up some notation — we review some concepts from the literature that will be used in the sequel.

2.1 Risk-averse two-stage stochastic programs

Two-stage stochastic programs constitute a fundamental class of problems in the field of stochastic programming. In those problems, the decision maker decides in a first stage, then an outcome of a random variable is revealed and finally another decision is made, which depends on the first stage decision and the outcome of the random variable. The classical formulation of two-stage stochastic linear programming with fixed recourse can be written as follows

$$\begin{aligned} \min_{x} \quad & d^{\intercal} x + \mathbb{Q}(x), \\ \text{subject to} \quad & Ax = b, \\ & x \geq 0, \end{aligned} \tag{2.1}$$

where $A \in \mathbb{R}^{r_1 \times s_1}$, $x \in \mathbb{R}^{s_1 \times 1}$, $b \in \mathbb{R}^{r_1 \times 1}$, with r_1 and s_1 representing the number of constraints and variables of the first stage problem. As stated earlier, the vector ξ is random with known source distribution P (with finite support), so the function $\mathbb{Q}(x) = \mathbb{E}_P\left[Q(x,\xi)\right]$ is written as $\mathbb{Q}(x) = \sum_{i=1}^{n} P_i\, Q(x,\xi_i)$, where $Q(x,\xi)$ is defined as follows[3]:

$$\begin{aligned} Q(x,\xi) = \min_{z_\xi} \quad & q^{\intercal} z_\xi \\ \text{subject to} \quad & W z_\xi = h^{\xi} - T^{\xi} x \\ & z_\xi \geq 0. \end{aligned} \tag{2.2}$$

In the above formulation, $T_\xi \in \mathbb{R}^{r_2 \times s_1}$, $W \in \mathbb{R}^{r_2 \times s_2}$, $z_\xi \in \mathbb{R}^{s_2 \times 1}$, $h_\xi \in \mathbb{R}^{r_2 \times 1}$ with r_2 and s_2 representing the number of constraints and variables of the second stage problem. Two-stage stochastic linear programs were first proposed by (G. B. Dantzig, 1955), and the theoretical properties of this class of problems are well known (for instance, for every fixed ξ_i the function $Q(x,\xi_i)$ is convex, and the function $\mathbb{Q}(x)$ is piecewise linear convex, see Birge and Louveaux (2011)). Applications of these models include finance, energy, natural resources, disaster relief, among others; see, for instance, Wallace and Ziemba (2005).

The development of theory of optimization with risk measures have led formulations of *risk-averse* two-stage models, where the expectation in the function $\mathbb{Q}(x) = \mathbb{E}_P\left[Q(x,\xi)\right]$ is replaced with some other coherent risk measure; see, for instance, Ahmed (2006), Römisch and Wets (2007),

[3] In the most general case, W anmd q are random as well, but for our purposes we assume that these quantities are deterministic.

Miller and Ruszczynski (2011), and Noyan (2012). In our case, as mentioned earlier, we consider problems with Conditional Value-at-Risk. That is, we replace (2.1) with

$$\begin{aligned} \min_{x} \quad & d^{\intercal}x + \mathrm{CVaR}_{\alpha}\left[Q(x,\xi)\right] \\ \text{subject to} \quad & Ax = b \\ & x \geq 0. \end{aligned} \tag{2.3}$$

Note that because of the translation-invariant property of CVaR (i.e. $\mathrm{CVaR}_{\alpha}[a + Z] = a + \mathrm{CVaR}_{\alpha}[Z]$), the above problem fits the frameworlk of (1.5), with $G(x,\xi) = d^{\intercal}x + Q(x,\xi)$. Moreover, by using formulation (1.4), we can alternatively formulate (2.3) as a standard (risk-neutral) two-stage stochastic program with extra variables, as indicated below:

$$\begin{aligned} \min_{x,\eta,\nu} \quad & d^{\intercal}x + \eta + \widehat{\mathbb{Q}}(x,\eta), \\ \text{subject to} \quad & Ax = b, \\ & x \geq 0, \end{aligned} \tag{2.4}$$

where $\widehat{\mathbb{Q}}(x,\eta) = \mathbb{E}_P\left[\widehat{Q}(x,\eta,\xi)\right]$ and $\widehat{Q}(x,\eta,\xi)$ is defined as follows:

$$\begin{aligned} \widehat{Q}(x,\eta,\xi) = \min_{z_\xi,\nu_\xi} \quad & \frac{1}{1-\alpha}\,\nu_\xi, \\ \text{subject to} \quad & W z_\xi = h^\xi - T^\xi x, \\ & \nu_\xi - q^{\intercal} z_\xi \geq -\eta \\ & z_\xi, \nu_\xi \geq 0, \end{aligned} \tag{2.5}$$

Note that, once the scenarios are fixed, the above formulation allows us to solve the optimization problem using standard decomposition techniques.

2.2 Scenario reduction via the Monge-Kantorovich problem

We review next the approach proposed in Dupačová, Gröwe-Kuska, and Römisch (2003) for scenario reduction. The idea relies upon the notion of *stability* of stochastic programs which is extensively developed in the literature. While we follow the ideas from Dupačová et al. (2003), we present the results in the context of distributions with finite support, which is the case of the models in this paper. The restriction to finite distributions allows us to simplify considerably the results and the required assumptions.

The following assumptions are made regarding problem (1.1): $X \subset \mathbb{R}^n$ is a given nonempty convex closed set, ξ is a random vector with values in Ξ and distribution P, and the function G from $\mathbb{R}^s \times \Xi$ to the extended real numbers $\hat{\mathbb{R}}$ is lower semicontinuous and convex with respect to x. Moreover, $G(\cdot,\xi)$ is bounded on X for all $\xi \in \Xi$. The optimal value, parameterized by the distribution of ξ, is denoted by

$$v(P) := \inf\left\{\mathbb{E}_P\left[G(x,\xi)\right] \,:\, x \in X\right\} \tag{2.6}$$

Given the source distribution P and an approximation $\hat{P}$, quantitative estimates of the closeness of $v(\hat{P})$ to $v(P)$ in terms of a certain probability metric can be shown. To proceed, let $\mathcal{P}(\Xi)$ denote the set of distributions whose support is contained in Ξ. Define the distance $d_{G,\rho}(P,\hat{P})$ between two probability distributions in $\mathcal{P}(\Xi)$ as follows:

$$d_{G,\rho}(P,\hat{P}) := \sup_{x\in X\cap\rho\mathbb{B}} \left|\mathbb{E}_P\left[G(x,\xi)\right] - \mathbb{E}_{\hat{P}}\left[G(x,\xi)\right]\right| \;=\; \sup_{x\in X\cap\rho\mathbb{B}} \left|\sum_{i=1}^{n} G(x,\xi_i)P(\xi_i) - \sum_{i=1}^{n} G(x,\xi_i)\hat{P}(\xi_i)\right|.$$

In the above expressions, The set $\mathbb{B}\subset\mathbb{R}^n$ is the ball centered in the origin with radius equals to 1.

By relying on results in (Rockafellar & Wets, 1998), Dupačová et al. (2003) show that there exist constants $\rho > 0$ and $\bar{\epsilon} > 0$ such that

$$\left|v(P) - v(\hat{P})\right| \;\le\; d_{G,\rho}(P,\hat{P}) \tag{2.7}$$

whenever $\hat{P}\in\mathcal{P}(\Xi)$ with $d_{G,\rho}(P,\hat{P}) < \bar{\epsilon}$. Notice that a major consequence of (2.15) is that, if we have a new distribution $\hat{P}$ that is close enough to the source distribution P, then the differences between both objective function values will be bounded by the distance between both distributions. Thus, if we can find a new distribution $\hat{P}\in\mathcal{P}(\Xi)$ which has smaller support than P but is sufficiently close to P, then we obtain a problem which is easier to solve than (1.1) and provides a good approximation to that problem in terms of optimal values.

As we can see from the definition, it is very hard to compute the value of $d_{G,\rho}(P,\hat{P})$ which is required to verify (2.7). It is possible however to derive upper bounds for that quantity in terms of P and $\hat{P}$. To do so, suppose we have a function $c:\Xi\times\Xi\mapsto\mathbb{R}_+$ that provides a "distance" between two scenarios ξ and $\hat{\xi}\in\Xi$. The function c does not necessarily represent a distance; it is only required that (i) c be symmetric, and (ii) $c(\xi,\hat{\xi}) = 0$ if and only if $\xi = \hat{\xi}$.

Given the function c described above, suppose that the function G satisfies

$$|G(x,\xi) - G(x,\hat{\xi})| \;\le\; h(\|x\|)\, c(\xi,\hat{\xi}) \tag{2.8}$$

for some nondecreasing function $h:\mathbb{R}_+\mapsto\mathbb{R}_{++}$. Since $\|x\|\le\rho$ for any $x\in X\cap\rho\mathbb{B}$, it follows from (2.8) that the function $\tilde{G}(\cdot,\xi) := G(\cdot,\xi)/h(\rho)$ (when restricted to $X\cap\rho\mathbb{B}$) belongs to the class of functions

$$\mathcal{F}_c := \left\{f:\Xi\mapsto\mathbb{R} \,:\, |f(\xi) - f(\hat{\xi})| \le c(\xi,\hat{\xi})\right\} \tag{2.9}$$

and thus by definition of $d_{G,\rho}(P,\hat{P})$ we must have

$$d_{G,\rho}(P,\hat{P}) \;=\; h(\rho)\sup_{x\in X\cap\rho\mathbb{B}} \left|\mathbb{E}_P\left[\tilde{G}(x,\xi)\right] - \mathbb{E}_{\hat{P}}\left[\tilde{G}(x,\xi)\right]\right| \;\le\; h(\rho)\sup_{f\in\mathcal{F}_c}\left(\mathbb{E}_P\left[f(\xi)\right] - \mathbb{E}_{\hat{P}}\left[f(\xi)\right]\right) \tag{2.10}$$

(note that we remove the absolute value in the right-most expression since $f\in\mathcal{F}_c$ implies that $-f\in\mathcal{F}_c$).

The key idea now is to show that the right-hand side of (2.10) can be easily computed. Indeed, notice that since $\Xi = \{\xi_1,\ldots,\xi_n\}$, it follows that we can view each function $f:\Xi\mapsto\mathbb{R}$ as a vector in $\mathbb{R}^n$; we write f_i for $f(\xi_i)$ for short, and similarly for the probabilities $P(\xi_i)$ and $\hat{P}(\xi_i)$. Moreover, we can view the function c as a matrix in $\mathbb{R}^{n\times n}$, and write c_{ij} for $c(\xi_i,\xi_j)$. With this notation, we can see that the "sup" problem on the right-hand side of (2.10) can be formulated as the linear program

$$\begin{aligned} \max_{f} \quad & \sum_{k=1}^{n}(P_k - \hat{P}_k)\, f_k \\ \text{s. to} \quad & |f_i - f_j| \le c_{ij}, \quad i,j = 1,\ldots,n. \end{aligned} \tag{2.11}$$

The dual of this linear program is given by

$$\begin{aligned} \min_{\pi} \quad & \sum_{i,j=1}^{n} c_{ij}\,\pi_{ij} \\ \text{s. to} \quad & \sum_{j=1}^{n} \pi_{kj} - \sum_{i=1}^{n} \pi_{ik} = P_k - \hat{P}_k, \quad k = 1, \ldots, n \\ & \pi \geq 0. \end{aligned} \tag{2.12}$$

These linear programs can be easily solved. Moreover, there is no duality gap, as both primal and dual are feasible; this can be seen by taking $f_i = a$ for all i in the primal (where a is any constant), and $\pi_{ij} = P_i \hat{P}_j$ in the dual. However, we wish to take a step further to obtain a quantity that represents a distance between P and $\hat{P}$. Indeed, consider the linear program below, which is known as the *Monge-Kantorovich* mass transportation problem in finite dimensions:

$$\min_{\pi} \quad \sum_{i,j=1}^{n} c_{ij}\,\pi_{ij}$$

$$\text{s. to} \quad \sum_{j=1}^{n} \pi_{kj} = P_k, \quad k = 1, \ldots, n \tag{2.13}$$

$$\sum_{i=1}^{n} \pi_{ik} = \hat{P}_k, \quad k = 1, \ldots, n \tag{2.14}$$

$$\pi \geq 0.$$

The optimal value function $\hat{\mu}_c(P, \hat{P})$ of problem (2.13)-(2.14) is known as the *Kantorovich functional*. It is useful to observe that in the particular case where $c(\xi, \hat{\xi}) = \|\xi - \hat{\xi}\|$, the functional $\hat{\mu}_c(P, \hat{P})$ becomes the well-known *Wasserstein distance* between P and $\hat{P}$; see, for instance, Rachev (1991) for a thorough discussion of these concepts.

By noticing that equation (2.12) can be written by subtracting (2.14) from (2.13), it follows that the feasibility set of the former problem contains that of the latter one, which implies that the optimal value of problem (2.12) is less than or equal to $\hat{\mu}_c(P, \hat{P})$. Together with (2.7) and (2.10), this implies that

$$\left| v(P) - v(\hat{P}) \right| \;\leq\; d_{G,\rho}(P, \hat{P}) \;\leq\; h(\rho)\, \hat{\mu}_c(P, \hat{P}), \tag{2.15}$$

The focus of the work Dupačová et al. (2003) can be described as follows: suppose that we have a fixed probability distribution P from a discrete random variable. Then, find a new distribution $\hat{P}$ with $\text{supp}(\hat{P}) \subset \Xi$ and $|\text{supp}(\hat{P})| = M < n = |\Xi|$ (where $\text{supp}(\hat{P})$ indicates the support of the distribution $\hat{P}$), such that $\hat{P}$ minimizes the optimal value of the *Monge-Kantorovich* problem

(2.13)-(2.14) in its discrete form. This amounts to solving the following problem

$$
\begin{aligned}
\min_{\pi,r} \quad & \sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij}\pi_{ij} \\
\text{s. to} \quad & \sum_{j=1}^{n} \pi_{ij} = P_i, \qquad i = 1,\dots,n \\
& \pi_{ij} \le r_j, \qquad i,j = 1,\dots,n \\
& \sum_{j=1}^{n} r_j = M \\
& \pi_{ij} \in [0,1]\,, \quad i,j = 1,\dots,n \\
& r_j \in \{0,1\}, \quad j = 1,\dots,n,
\end{aligned}
\tag{2.16}
$$

where M is the cardinality of supp$(\hat{P})$. Let (π^*, r^*) be an optimal solution to (2.16). Then, the selected scenarios are those for which $r_j^* = 1$ (call that set R). The probability of each selected scenario j is given by

$$
\hat{P}_j \;=\; \sum_{i \in C_j} \pi_{ij}, \quad j \in R, \tag{2.17}
$$

where C_j is the set of scenarios ξ_i such that scenario ξ_j is the "closest" scenario to ξ_i (in terms of c_{ij}) among those that belong to R.

Since (2.16) is an NP-Hard problem, in Dupačová et al. (2003) develop a so-called *backward recursion* heuristic to solve (2.16). Briefly speaking, the idea of the backward reduction is to reduce the scenarios one by one, as the problem of reducing the cardinality from n to $n-1$ can be solved efficiently. We refer to that paper for details.

The procedure described above is valid for any problem of the form (1.1) that satisfies the assumptions discussed in the text. While most such assumptions are standard — for example, convexity of X and of G with respect to x — it is important to recall condition (2.8) which is key to develop the stability result (2.15). Such condition can proved for particular classes of problems. For our purposes, we rely on a result in (Römisch & Wets, 2007) that ensures that the CVaR two-stage problem (2.3) does indeed satisfy (2.8); briefly speaking, this is accomplished by showing that the set of dual solutions $D(\xi)$ of the second-stage problem (2.5) is Lipschitz on Ξ when viewed as a set-valued function of ξ (with respect to the Hausdorff distance), which then implies that (2.8) holds by virtue of Proposition 3.3 in that paper.

2.3 Effective scenarios

We discuss now the concept of *effective scenarios* recently introduced by Rahimian et al. (2018). The authors present that notion in the context of distributionally robust problems of the form

$$
\min_{x \in X} \max_{P \in \mathbb{P}} \mathbb{E}_P \left[G(x, \xi) \right], \tag{2.18}
$$

where $\mathbb{P}$ is a convex set of distributions, called *ambiguity set* in the literature. Suppose that all distributions in $\mathbb{P}$ have support which is a subset of a common finite set, say, Ξ. The question posed by Rahimian et al. (2018) is, what happens to problem (2.18) if a scenario $\bar{\xi}$ is removed from Ξ? If the optimal value of (2.18) changes as a result of the removal, then $\bar{\xi}$ is said to be effective; otherwise, it is ineffective. For such notion to be well-defined, however, it is essential to specify what is meant by “removing" a scenario $\bar{\xi}$. Removing $\bar{\xi}$ amounts to solving a problem similar to (2.18) but with $\mathbb{P}$ replaced with

$$\mathbb{P}^{\bar{\xi}} := \left\{ P \in \mathbb{P} \,:\, P(\bar{\xi}) = 0 \right\}. \tag{2.19}$$

That is, we restrict our search for the worst-case distribution in the ambiguity set to those ones whose support does not include $\bar{\xi}$. The problem with the modified $\mathbb{P}^{\bar{\xi}}$ is called the *assessment problem* corresponding to $\bar{\xi}$. Note that while in the above exposition we chose to remove a single scenario $\bar{\xi}$ to simplify the presentation, we can actually remove a collection of scenarios simultaneously.

The notion of effective scenarios can be readily applied to risk-averse stochastic optimization problems of the form

$$\min_{x \in X} \mathcal{R}[G(x, \xi)], \tag{2.20}$$

where $\mathcal{R}$ is a *coherent* risk measure and as before ξ is a discrete random variable with values in Ξ. This is a consequence of the well-known duality between coherent risk measures and ambiguity sets, i.e., given a coherent risk-measure $\mathcal{R}$ there exists a set of probability distributions $\mathbb{P}$ such that $\mathcal{R}(Z) = \max_{P \in \mathbb{P}} \mathbb{E}_P[Z]$; see, for instance, Artzner, Delbaen, Eber, and Heath (1999). It is clear then that the risk-averse problem (2.20) can be written as (2.18).

A particular case of (2.20) is obtained when $\mathcal{R} = \mathrm{CVaR}_\alpha$. In that context, the notion of effective scenarios allows us to formalize the idea that we only the scenarios “in the tail" of G matter for the problem, and as a result we develop a scenario reduction algorithm based on those concepts. Such development will be described in detail in Section 4.

3 Literature Review

In this section we review the literature on scenario reduction methods in more detail, focusing particularly on works that aim at solving risk-averse problems. The goal of scenario reduction techniques is to reduce the number of scenarios of a stochastic optimization problem, up to a prescribed cardinality M, where we know that our computational capacity allows to solve it efficiently. To achieve such reduction, several techniques have been developed in the literature. An example is the Sample Average Approximation (SAA) approach, which replaces the expected value by a sample average of M scenarios drawn from the underlying distribution. SAA enjoys a number of theoretical properties (see, e.g., Shapiro (2003) and Homem-de-Mello and Bayraksan (2014)) and many successful applications of such an approach have been reported in the literature; however, the majority of such applications correspond to classical risk-neutral models. A detailed study of SAA for risk-averse problems is provided in Guigues, Krätschmer, and Shapiro (2016). Note however that, in the context of risk-averse problems, some authors have reported that standard Monte Carlo methods may require a large number of samples to provide a good approximation; see, for instance Espinoza and Moreno (2014).

As mentioned in Section 1, problem (1.6) can be written as a risk-neutral model with extra variables, so in principle the probability metrics approach discussed in Section 2.2 can be applied for that problem. Indeed, Römisch and Wets (2007) make precisely that argument, whereas Eichhorn and Römisch (2008) extend it for the case of polyhedral risk measures (of whcih CVaR is a particular case). For that reason, in Section 5 we will compare our proposed method with a standard application of the approach in Section 2.2.

Pineda and Conejo (2010) also develop scenario reduction techniques for two-stage problems of the form (2.3) based on Dupačová et al. (2003), but their approach is different from ours. The authors redefine the way in which the distance between the two probability distributions is measured.

More specifically, they define a particular function $c(\xi, \hat{\xi})$ to measure the "distance" between two scenarios ξ and $\hat{\xi}$ (recall the discussion about the function c in Section 2.2) as follows. First, a first-stage solution $\bar{x}$ is fixed, obtained by solving the two-stage problem with a single scenario equal to the average scenario. Then they set $c(\xi, \hat{\xi})$ as (adapting their notation to ours)

$$c(\xi, \hat{\xi}) \;=\; \left| (Q(\bar{x}, \xi) - \mathrm{VaR}_\alpha\left[Q(\bar{x}, \cdot)\right])_+ - \left(Q(\bar{x}, \hat{\xi}) - \mathrm{VaR}_\alpha\left[Q(\bar{x}, \cdot)\right]\right)_+ \right|$$

In words, $c(\xi, \hat{\xi})$ corresponds to the difference between the second-stage costs of scenarios ξ and $\hat{\xi}$ if both scenarios are in the tail of $Q(\bar{x}, \cdot)$. If neither scenario is in the tail, c is set to zero. Note that the approach constitutes a heuristic, as such c does not satisfy the properties discussed in Section 2.2 — in particular, it violates the property that $c(\xi, \hat{\xi}) = 0$ if and only if $\xi = \hat{\xi}$.

García-Bertrand and Mínguez (2014) also aim at solve problems of the form (1.6). However, the method proposed in that paper is not itself a scenario reduction technique, but rather an iterative algorithm to solve the optimization problem. Their procedure uses, at every iteration k, only the scenarios that are in the tail of $G(x^{k-1}, \cdot)$, i.e. those scenarios ξ such that $G(x^{k-1}, \xi) \geq \mathrm{VaR}_\alpha\left[G(x^{k-1}, \cdot)\right]$. The procedure does not fix the number of scenarios in each iteration.

Another related technique is the one proposed in Fairbrother et al. (2015). Similarly to us, their goal is to generate a set scenarios of desired cardinality M for risk-averse stochastic optimization problems. Their approach, however, is not based on probability metrics methods; instead, the authors develop the concept of a *risk region*, which is the union, over all the feasible decisions x, of the scenarios whose cost $G(x, \xi_i)$ is great than or equal to $\mathrm{VaR}_\alpha\left[G(x, \cdot)\right]$. The authors then propose to sample scenarios from that region, and show consistency of the obtained statistical estimates.

4 The proposed approach

In this section we discuss the scenario reduction method we propose for problem (1.5). We first establish the theoretical foundations for our approach, and then we describe the algorithm.

4.1 Theoretical development

As mentioned in Section 2.3, we rely on the notion of effective scenarios proposed by Rahimian et al. (2018). The main idea is to identify the ineffective scenarios of problem (1.5) — which by definition can be removed without affecting the optimal value — and then select, from the complementary set, the scenarios that will be kept to approximate the problem.

The first task is to write (1.5) in the form of a distributionally robust problem so we can apply the formulation of Section 2.3. Such dual representation of CVaR is well known, see, e.g., Shapiro, Dentcheva, and Ruszczyński (2014). Then, (1.5) can be written as

$$\min_{x \in X} \max_{\hat{P} \in \mathbb{P}} \mathbb{E}_{\hat{P}}\left[G(x, \xi)\right], \tag{4.1}$$

where

$$\mathbb{P} \;=\; \left\{ \hat{P} \in \mathcal{P}(\Xi) \;:\; \hat{P}_i \leq \frac{P_i}{1-\alpha} \right\}. \tag{4.2}$$

Let $D \subsetneq \Xi$ be an arbitrary set of scenarios such that $\sum_{i \in D} P_i < \alpha$. To simplify the notation, we shall use D interchangeably as a set of scenarios and also as the set of indices of those scenarios. To test for the effectiveness of D, we first write the corresponding assessment problem, which is the same as (4.1) but with $\mathbb{P}$ replaced with $\mathbb{P} \cap \{\hat{P} \in \mathcal{P}(\Xi) \,:\, \hat{P}_i = 0 \text{ for all } i \in D\}$. That is, the assessment problem is written as

$$\min_{x\in X} \max \left\{ \sum_{i\in I\setminus D} \hat{P}_i \, G(x,\xi) \,:\, \hat{P} \in \mathbb{P}, \sum_{i\in I\setminus D} \hat{P}_i = 1 \right\} \tag{4.3}$$

where $I := \{1, \dots, n\}$.

Proposition 1 below shows that the assessment problem has a natural interpretation in terms of conditional distributions. A related result is shown in (Rahimian et al., 2018, Proposition 5), but we adapt it here for our purposes.

Proposition 1. *The assessment problem (4.3) can be written as*

$$\min_{x\in X} \mathrm{CVaR}_{\alpha'} \left[G(x,\xi) \,|\, \Xi\setminus D \right], \tag{4.4}$$

where

$$\alpha' = \frac{\alpha - \sum_{i\in D} P_i}{1 - \sum_{i\in D} P_i}. \tag{4.5}$$

Proof. For any fixed $x \in X$, the dual formulation of the inner maximization in (4.3) can be written as

$$\begin{aligned} \min_{\mu,\lambda} \;\; & \mu + \frac{1}{1-\alpha} \sum_{i\in I\setminus D} P_i \lambda_i \\ & \mu + \lambda_i \;\; \geq G(x,\xi) \quad \forall i \in I\setminus D \\ & \lambda_i \;\; \geq 0 \quad \forall i \in I\setminus D, \end{aligned} \tag{4.6}$$

where λ_i is the multiplier of the constraint $\hat{P}_i \leq \frac{P_i}{1-\alpha}$ and μ is the multiplier of the constraint $\sum_{i\in I\setminus D} \hat{P}_i = 1$.

Let us focus on the objective function problem (4.6). We have

$$\min_{\mu,\lambda} \; \mu + \frac{1}{1-\alpha} \sum_{i\in I\setminus D} P_i \lambda_i \;\; = \;\; \min_{\mu,\lambda} \; \mu + \frac{1 - \sum_{i\in D} P_i}{1-\alpha} \sum_{i\in I\setminus D} \frac{P_i}{1 - \sum_{i\in D} P_i} \lambda_i.$$

Now define the quantities

$$P'_i := \frac{P_i}{1 - \sum_{i\in D} P_i}. \tag{4.7}$$

By using the definition of P' and of α' we can rewrite the problem (4.6) as

$$\begin{aligned} \min_{\mu,\lambda} \;\; & \mu + \frac{1}{1-\alpha'} \sum_{i\in I\setminus D} P'_i \lambda_i \\ & \mu + \lambda_i \;\; \geq G(x,\xi) \quad \forall i \in I\setminus D \\ & \lambda_i \;\; \geq 0 \quad \forall i \in I\setminus D. \end{aligned} \tag{4.8}$$

It is easy to see that problem (4.8) is the dual to

$$\max \left\{ \sum_{i\in I\setminus D} \hat{P}_i \, G(x,\xi) \,:\, \hat{P} \in \mathbb{P}', \sum_{i\in I\setminus D} \hat{P}_i = 1, \right\} \tag{4.9}$$

where

$$\mathbb{P}' \;:=\; \left\{ \hat{P} \in \mathcal{P}(\Xi) \,:\, \hat{P}_i \leq \frac{P'_i}{1-\alpha'}, \quad i \in I \backslash D \right\},$$

and thus we see that problem (4.9) is simply $\mathrm{CVaR}_{\alpha'}\left[G(x,\xi) \,|\, (\Xi\backslash D)\right]$. □

The importance of Proposition 1 lies in the fact the assessment problem (4.4) has precisely the same structure as the original problem (1.5) — we just replace P by its conditional version $P(\cdot \,|\, \Xi\backslash D)$ and adjust the level α of CVaR. The similarity between the two problems make it easier to determine conditions on the set D under which the optimal values of the problems coincide, in which case D is ineffective. Theorem 2 below, which is our main result, explores that idea. Again, a related result is stated in Theorems 1 and 4 in (Rahimian et al., 2018) in a slightly different context, so we adapt the result and its proof to our case.

Theorem 2. *Let $x^* \in X$ be an optimal solution to (1.5) and define the set*

$$D \;:=\; \{\xi \in \Xi \,:\, G(x^*,\xi) < \mathrm{VaR}_\alpha\left[G(x^*,\cdot)\right]\} \tag{4.10}$$

Then, the set D is ineffective for problem (1.5).

Proof. In order to show that D is ineffective, we must show that the optimal values of problems(1.5) and (4.4) are the same. The key tool to accomplish that goal is Lemma 1 in Rahimian et al. (2018), which uses convex analysis techniques to show that, if both the objective function of (1.5) and its subdifferential set calculated at x^* (i.e., $\mathrm{CVaR}_\alpha\left[G(x^*,,\xi)\right]$ and $\partial_x \mathrm{CVaR}_\alpha\left[G(x^*,,\xi)\right]$) coincide with their counterparts in (4.4) (also calculated at x^*), then the optimal value of problem (4.4) is the same as the optimal value of (1.5).

The expression for $\partial \mathrm{CVaR}_\alpha\left[G(x^*,,\xi)\right]$ can be derived from the dual formulation (4.1) as follows; see, e.g., (Shapiro et al., 2014, Theorem 6.14). Given $x \in X$, let $S_\alpha(x)$ denote the set of optimal solutions to the inner problem in (4.1). Then, have that

$$\partial \mathrm{CVaR}_\alpha\left[G(x^*,,\xi)\right] \;=\; \bigcup_{\hat{P} \in S_\alpha(x)} \sum_{i=1}^n \hat{P}_i \,\partial_x G(x^*,\xi_i). \tag{4.11}$$

Next, noticing that for each $x \in X$ the inner problem in (4.1) is a linear program in $\hat{P}$, it is not difficult to see that this inner problem can be solved analytically by sorting the values of $G(x,\cdot)$ and assigning the largest value allowed by (4.2) for $\hat{P}$ to the scenarios with higher cost. Thus, the set of optimal solutions to that inner problem is given by

$$S_\alpha(x) = \left\{ \hat{P} \;:\; \begin{array}{ll} \hat{P}_i = 0, & \text{if } G(x,\xi_i) < \mathrm{VaR}_\alpha\left[G(x,\cdot)\right] \\ \hat{P}_i = \frac{P_i}{1-\alpha}, & \text{if } G(x,\xi_i) > \mathrm{VaR}_\alpha\left[G(x,\cdot)\right] \\ \hat{P}_i \in \left[0, \frac{P_i}{1-\alpha}\right], & \text{if } G(x,\xi_i) = \mathrm{VaR}_\alpha\left[G(x,\cdot)\right] \\ \displaystyle\sum_{i\,:\, G(x,\xi_i) = \mathrm{VaR}_\alpha[G(x,\cdot)]} \hat{P}_i = \frac{1}{1-\alpha}\left[\displaystyle\sum_{i\,:\, G(x,\xi_i) \leq \mathrm{VaR}_\alpha[G(x,\cdot)]} P_i - \alpha \right]. & \end{array} \right. \tag{4.12}$$

Note that the problem has multiple optimal solutions when the set $\{i \,:\, G(x^*,\xi_i) = \mathrm{VaR}_\alpha\left[G(x,\cdot)\right]\}$ contains more than one element.

Similarly, by using the conditional formulation obtained in Proposition 1, we see that the set of optimal solutions to the inner problem in (4.3) is given by

$$S'_\alpha(x) := \left\{ \hat{P} \,:\, \begin{array}{ll} \hat{P}_i = 0, & \text{if } G(x,\xi_i) < \mathrm{VaR}_{\alpha'}\left[G(x,\cdot)\,|\,\Xi\backslash D\right] \text{ or } i \in D \\ \hat{P}_i = \frac{P'_i}{1-\alpha'}, & \text{if } G(x,\xi_i) > \mathrm{VaR}_{\alpha'}\left[G(x,\cdot)\,|\,\Xi\backslash D\right] \\ \hat{P}_i \in \left[0, \frac{P'_i}{1-\alpha'}\right], & \text{if } G(x,\xi_i) = \mathrm{VaR}_{\alpha'}\left[G(x,\cdot)\,|\,\Xi\backslash D\right] \\ \multicolumn{2}{c}{\displaystyle\sum_{i\notin D\,:\,G(x,\xi_i)=\mathrm{VaR}_{\alpha'}[G(x,\cdot)\,|\,\Xi\backslash D]} \hat{P}_i = \frac{1}{1-\alpha'}\left[\sum_{i\notin D\,:\,G(x,\xi_i)\le\mathrm{VaR}_{\alpha'}[G(x,\cdot)\,|\,\Xi\backslash D]} P'_i - \alpha'\right]} \end{array} \right. . \tag{4.13}$$

From (4.7) and (4.5) we see that $\frac{P'_i}{1-\alpha'} = \frac{P_i}{1-\alpha}$. Moreover, the proof of Proposition 7 in (Rahimian et al., 2018) shows that

$$\mathrm{VaR}_\alpha\left[G(x,\cdot)\right] = \mathrm{VaR}_{\alpha'}\left[G(x,\cdot)\,|\,\Xi\backslash D'\right] \quad \text{whenever} \quad D' \subseteq \{\xi \in \Xi \,:\, G(x,\xi) < \mathrm{VaR}_\alpha\left[G(x,\cdot)\right]\}.$$

It follows that the first three conditions in (4.12) and (4.13) are identical. Finally, since

$$\frac{\alpha'}{1-\alpha'} = \frac{\alpha - \sum_{i\in D} P_i}{1-\alpha}$$

we have that the last equation in (4.13) can be written as

$$\sum_{i\notin D\,:\,G(x,\xi_i)=\mathrm{VaR}_\alpha[G(x,\cdot)]} \hat{P}_i = \frac{1}{1-\alpha}\left[\sum_{i\notin D\,:\,G(x,\xi_i)\le\mathrm{VaR}_\alpha[G(x,\cdot)]} P_i + \sum_{i\in D} P_i - \alpha\right]$$

and since $G(x,\xi_i) < \mathrm{VaR}_\alpha\left[G(x,\cdot)\right]$ for any $i \in D$ we conclude that (4.13) is equivalent to

$$\sum_{i\,:\,G(x,\xi_i)=\mathrm{VaR}_\alpha[G(x,\cdot)]} \hat{P}_i = \frac{1}{1-\alpha}\left[\sum_{i\,:\,G(x,\xi_i)\le\mathrm{VaR}_\alpha[G(x,\cdot)]} P_i - \alpha\right],$$

which in turn is the last equation in (4.12).

The above argument show that the optimal solution sets of the inner problems in (4.1) and (4.3) *coincide*. It is easy to see then that the objective function values of the two problems (calculated at x^*) are equal, and from (4.11) we see that the respective subdifferential sets at x^* also coincide. The conclusion follows now from Lemma 1 in Rahimian et al. (2018). □

Together, Proposition 1 and Theorem 2 formalize the aforementioned intuitive notion that "only the scenarios in the tail matter" and provide the foundation for our scenario reduction approach: we know that the scenarios in the set D can be removed from the problem without causing any changes to the optimal value; moreover, once D is removed, we can solve a problem that has the same structure as the original one, i.e., it is still a CVaR minimization problem — we just replace the source distribution P with its conditional version $P(\cdot\,|\,\Xi\backslash D)$ and adjust the level α of CVaR. Keeping the same structure has two advantages: (i) it allows us to apply further scenario reduction techniques on the resulting problem, and (ii) it allows us to solve the optimization problem using the same decomposition techniques discussed earlier, but with potentially far fewer scenarios.

It is worthwhile pointing out that the expression for D in (4.10) involves only the scenarios that are *strictly below* the VaR term; we cannot include those that are equal to the VaR, as we cannot guarantee that the scenarios in the latter group are ineffective. This is not just a technicality —

as we shall see in Section 5, in some practical problems the set of scenarios ξ such that $G(x^*, \xi) = \mathrm{VaR}_\alpha[G(x^*, \cdot)]$ can be very large. Also, while one expects that D has large size for typical values of α such 95%, it is even possible that the set D in (4.10) be empty; as an extreme example, consider a problem in which all scenarios have the same cost at x^*, which happens for instance in a newsvendor problem with no backlog costs and a high degree of risk aversion, as the optimal solution is simply to order an amount equal to the minimum possible demand. Finding structural properties of the problem that ensure that D has a large size is an open question.

4.2 Approximating the optimal set of ineffective scenarios

The discussion in Section 4.1 shows that significant scenario reduction can be obtained if we can identify the set D. The definition of D in (4.10), however, requires knowledge of the optimal solution x^*, and obviously such x^* is not known, otherwise there would be no need to solve the optimization problem.

Our idea is then to provide a "quick" approximation to x^*, from which we can then approximate the set D. Such an argument, of course, presupposes a certain continuity of the set $D(x) := \{\xi \in \Xi : G(x, \xi) < \mathrm{VaR}_\alpha[G(x, \cdot)]\}$ in terms of x. Proposition 3 below shows a sufficient condition for that continuity to hold:

Proposition 3. *Consider the sets $D(x)$ defined above, and suppose that the function $G(\cdot, \xi)$ is continuous on X for each $\xi \in \Xi$. As before, let x^* be an optimal solution to (1.5). Suppose also that (i) the set of $\alpha-$quantiles of $G(x^*, \cdot)$ is a singleton, and (ii) the set $\{\xi \in \Xi : G(x^*, \xi) = \mathrm{VaR}_\alpha[G(x^*, \cdot)]\}$ is a singleton. Let $\{x^k\}$ be a convergent sequence of points in X such that $x^* = \lim_{k\to\infty} x^k$. Then $\lim_{k\to\infty} D(x^k) = D(x^*)$.*

Proof. Define the function $f_x(\eta)$ as follows:

$$f_x(\eta) := \eta + \frac{1}{1-\alpha}\sum_{i=1}^{n} P_i\left[G(x, \xi_i) - \eta\right]_+. \tag{4.14}$$

Note that the function $f_x(\cdot)$ is convex and continuous, and the domain of f_x is the real line. The set of minimizers of $f_x(\eta)$ over $\eta \in \mathbb{R}$ is the set of the α-quantiles of $G(x, \cdot)$ (Rockafellar & Uryasev, 2000). Moreover, the continuity assumption on G implies that $f_{x_k}(\eta) \to f_{x^*}(\eta)$ for all $\eta \in \mathbb{R}$. Thus, by Theorem 7.17 of Rockafellar and Wets (1998), we have that the sequence of functions $\{f_{x^k}(\cdot)\}$ converges to $f_{x^*}(\cdot)$ *uniformly* on every compact set C of the real line.

Note also that the sequence $\{f_{x^k}(\cdot)\}$ is *eventually level bounded* (cf. Rockafellar and Wets (1998)) since $f_x(\eta) \geq \eta$ for all η and the identity function $I(\eta) = \eta$ is level bounded, i.e., its level sets are bounded. Thus, by Theorem 7.33 from Rockafellar and Wets (1998) and the assumption that the set $\operatorname*{argmin}_\eta f_{x^*}(\eta)$ is a singleton (and therefore equal to $\mathrm{VaR}_\alpha[G(x^*, \cdot)]$) we have that

$$\mathrm{VaR}_\alpha\left[G(x^k, \cdot)\right] \to \mathrm{VaR}_\alpha\left[G(x^*, \cdot)\right]. \tag{4.15}$$

Next, let us sort the values of $\{G(x^*, \xi_i)\}$ in increasing order, and denote by $\{G(x^*, \xi_{(i)})\}$ the resulting sequence. Assumption (ii) of the proposition implies that there exists some index $i_\alpha \in \{1, \ldots, n\}$ such that $G(x^*, \xi_{(i_\alpha)}) = \mathrm{VaR}_\alpha[G(x^*, \cdot)]$ and

$$G(x^*, \xi_{(i_\alpha - 1)}) \;<\; G(x^*, \xi_{(i_\alpha)}) \;<\; G(x^*, \xi_{(i_\alpha + 1)}).$$

Finally, the continuity of $G(\cdot, \xi)$ implies that for k sufficiently large we have that

$$G(x^k, \xi_{(i_\alpha - 1)}) \;<\; G(x^k, \xi_{(i_\alpha)}) \;<\; G(x^k, \xi_{(i_\alpha + 1)}).$$

and thus $G(x^k, \xi_{(i_\alpha)}) = \mathrm{VaR}_\alpha\left[G(x^k, \cdot)\right]$ and $D(x^k) \to D(x^*)$. $\square$

The assumptions of Proposition 3 seem somewhat restrictive, but unfortunately they cannot be removed. Without assumption (i) we cannot even ensure that $\{\mathrm{VaR}_\alpha\left[G(x^k, \cdot)\right]\}$ converges at all (it may oscillate), and without assumption (ii) we may not have $D(x^k) \to D(x^*)$ even if assumption (i) holds. To see that, suppose that $\Xi = \{\xi_1, \xi_2\}$ with $P_1 = P_2 = 1/2$ and consider the function G defined as $G(x, \xi_1) = x$, $G(x, \xi_2) = 2 - x$ with $x \in X := [0, 1]$. Then, for any $\alpha > 1/2$ we have that $\mathrm{VaR}_\alpha[G(x, \cdot)] = 2 - x$ and thus $D(x) = \{\xi_1\}$ for all $x < 1$. However, we have that $\mathrm{CVaR}_\alpha[G(x, \cdot)] = 2 - x$ and thus $x^* = 1$, but $D(1) = \emptyset$ since $G(1, \xi_1) = G(1, \xi_2) = 1$.

Proposition 3 nevertheless suggests that, if we have a feasible point $\hat{x}$ that is close to x^*, then we can approximate $D(x^*)$ with $D(\hat{x}$. This is detailed in Algorithm 1 below. Recall the notation I for the index set $\{1, \ldots, n\}$.

Algorithm 1 Naive approximation of optimal set of ineffective scenarios

1: Let $\hat{x}$ be an approximating optimal solution to (1.5).
2: Compute the values of $G(\hat{x}, \xi_i)$ for all $i \in I$, and calculate $\mathrm{VaR}_\alpha[G(\hat{x}, \cdot)]$.
3: The output of this procedure is $D(\hat{x}) = \{i \in I : G(\hat{x}, \xi_i) < \mathrm{VaR}_\alpha[G(\hat{x}, \cdot)]\}$.

The major drawback of Algorithm 1 is the need to compute $G(\hat{x}, \xi_i)$ for all scenarios ξ_i. This might be time consuming, for example, in case of the two-stage problem (2.3), since computing $G(\hat{x}, \xi) = d^\intercal \hat{x} + Q(\hat{x}, \xi)$ requires solving the second-stage problem. In that case, in order to circumvent that issue we can use the dual formulation of $Q(x, \xi)$, which has the following form

$$\begin{aligned} Q(x,\xi) = \max_{\pi} \quad & \pi_\xi^\intercal \left(h^\xi - T^\xi x\right), \\ \text{subject to} \quad & \pi_\xi^\intercal W \le q, \\ & \pi_\xi \in \mathbb{R}^{r_2}. \end{aligned} \tag{4.16}$$

From (4.16) we can see that the feasible region does not depend on random variables. This means that if we find one feasible $\hat{\pi}$ we will have that $Q(\hat{x}, \xi) \ge \hat{\pi}^\intercal \left(h_\xi - T^\xi \hat{x}\right) \forall \xi$. Then to estimate the set D we adopt the following procedure. Instead of computing $Q(\hat{x}, \xi_i)$ for all of the original scenarios, we propose to compute it just for a subset $J \subsetneq \Xi$, and then store the dual multipliers $\hat{\pi}_i$, $i \in J$. The, we approximate the value of $Q(\hat{x}, \xi_i)$ by

$$\hat{Q}_J(\hat{x}, \xi_i) \ := \ \max_{j \in J} \hat{\pi}_j^\intercal \left(h^i - T^i \hat{x}\right). \tag{4.17}$$

The resulting algorithm to approximate the set of ineffective scenarios is summarized in Algorithm 2 below.

Algorithm 2 Approximating the optimal set of ineffective scenarios for two-stage problems

1: Let $\hat{x}$ be an approximating optimal solution to (2.3).
2: Select a set of scenarios $J \subsetneq I$.
3: Compute the values of $\hat{G}(\hat{x}, \xi_i) := d^\intercal \hat{x} + \hat{Q}_J(\hat{x}, \xi)$ for all $i \in I$ using (4.17), and calculate $\mathrm{VaR}_\alpha\left[\hat{G}(\hat{x}, \cdot)\right]$.
4: The output of this procedure is $\hat{D}(\hat{x}) = \left\{i \in I : \hat{G}(\hat{x}, \xi_i) < \mathrm{VaR}_\alpha\left[\hat{G}(\hat{x}, \cdot)\right]\right\}$.

Note that we do not specify how to construct the set J; of course, the larger J is, the better the approximation $\hat{Q}_J$.

4.3 Further reduction of the set of effective scenarios

In Section 4.1 we saw that if we can compute the set D given in Theorem 2, then instead of solving the original problem we can solve the reduced problem (4.4) that uses only the scenarios in the complement of D (which we denote by D^c). It makes sense to perform further scenario reduction on D^c to obtain an approximation that can be solved even faster. We propose to use the probability metrics approach described in Section 2.2, exploiting the fact that (4.4) is also a CVaR problem and as such it can be represented as an expected-value problem with extra variables. We emphasize that such an approach is related to but fundamentally different from that proposed in (Römisch & Wets, 2007), which applies the probability metrics approach to the entire scenario set Ξ.

In practice, of course, we do not know D, so we use the methods described in Section 4.2 to approximate it with another set $\hat{D}$. Our goal is then to apply the scenario reduction problem (2.16) to the complementary set $\hat{D}^c$. Before we achieve such a goal, we show that in the case where all scenarios have the same probability, then (2.16) — which can be viewed as a *facility location problem* — can be reformulated so it can be solved more efficiently. Proposition 4 states the result.

Proposition 4. *Consider problem (2.16) If all probabilities P_i are equal, then the problem can be reformulated as follows*

$$\begin{aligned}
\min_{\hat{\pi},r} \; & \sum_{i=1}^{n}\sum_{j=1}^{n}\frac{1}{n}c_{ij}\hat{\pi}_{ij} \\
& \sum_{j=1}^{n}\hat{\pi}_{ij} = 1 \qquad \forall i \in I \\
& \hat{\pi}_{ij} \le r_j \qquad \forall i,j \in I \\
& \sum_{j=1}^{n} r_j = M \\
& \hat{\pi}_{ij} \in [0,1] \quad \forall i,j \in I \\
& r_i \in \{0,1\} \quad \forall i \in I.
\end{aligned} \tag{4.18}$$

Proof. Since the first constraint in (2.16) establish that $\sum_{j=1}^{n}\pi_{ij} = P_i \quad \forall i \in I$ and $\pi_{ij} \ge 0$, then it follows that $\pi_{ij} \le \max_{k\in I}\{P_k\}$ for all $i,j \in I$, so we can write the constraint $\pi_{ij} \le r_j$ as $\pi_{ij} \le (\max_{k\in I}\{P_k\})\, r_j$. If $P_i = 1/n$ for all $i \in I$ then $\max_k\{P_k\} = 1/n$, so by defining $\hat{\pi}_{ij} := n\pi_{ij}$ we obtain (4.18). ☐

As we shall see shortly, Proposition 4 allow us to use efficient algorithms from the literature for the facility location problem. We shall assume henceforth that the assumption of Proposition 4 (i.e. $P_i = 1/n$ for all $i \in I$) is satisfied. Such an assumption holds for example when the source distribution P corresponds a random sample (obtained with standard Monte Carlo or alternative methods such as Latin Hypercube sampling or Quasi-Monte Carlo) from the original distribution.

As mentioned at the beginning of this subsection, in our approach we apply the facility location problem only to the scenarios in $\hat{D}^c$. For ease of reference, we state the resulting problem below.

$$
\begin{aligned}
\min_{\hat{\pi},r} \sum_{i\in\hat{D}^c}\sum_{j\in\hat{D}^c} \frac{1}{|\hat{D}^c|} c_{ij}\hat{\pi}_{ij} & \\
\sum_{j\in\hat{D}^c} \hat{\pi}_{ij} = 1 & \qquad \forall i \in \hat{D}^c \\
\hat{\pi}_{ij} \le r_j & \qquad \forall i,j \in \hat{D}^c \\
\sum_{j\in\hat{D}^c} r_j = M & \\
\hat{\pi}_{ij} \in [0,1] & \qquad \forall i,j \in \hat{D}^c \\
r_i \in \{0,1\} & \qquad \forall i \in \hat{D}^c.
\end{aligned}
\tag{4.19}
$$

Note that the term $1/|\hat{D}^c|$ in the objective function corresponds to the conditional probability $P_i/(1-\sum_{j\in\hat{D}} P_j)$ (see (4.7)) in the case of equal probabilities.

Problem (4.19) is a mixed-binary program and can in principle be solved with any commercial software designed for that purpose. We want however to exploit the structure of the facility location problem in order to solve it more efficiently. In our approach, we use *ZEBRA* algorithm proposed in García, Labbé, and Marín (2011). The method described in that paper solves facility location problems formulated precisely as (4.19); we refer to that paper for further details.

4.4 The procedure

In this section we summarize our proposed scenario reduction procedure for the risk-averse two-stage problem (2.3) where all the probabilities are equal. Also, in our implementation we use the Euclidean distance $c_{ij} = \|\xi_i - \xi_j\|$. The steps are outlined in Algorithm 3 below.

Algorithm 3 Risk-averse two-stage problems with scenario reduction (RASRA)

1: Select a set of scenarios J from Ξ. In our implementation this is done using the backward reduction heuristic from Dupačová et al. (2003).
2: Solve problem (2.4)-(2.5) using only the scenarios in J in order to obtain an approximating solution $\hat{x}$.
3: Use Algorithm 2 (with $c_{ij} := \|\xi_i - \xi_j\|$) to obtain an approximation $\hat{D}^c$ of the set of effective scenarios (in step 2 of Algorithm 2 we use the same set J chosen in step 1 above).
4: Solve problem (4.19) to obtain a reduced set of scenarios R of cardinality M.
5: Solve the following problem:

$$
\begin{aligned}
\min_{x,z,\eta} \quad & d^\intercal x + \eta + \frac{1}{1-\alpha'}\sum_{i\in R}\hat{P}_i\nu_i, \\
\text{subject to} \quad & Ax = b, \\
& T^i x + W z_i = h^i, \qquad i \in R \\
& q^\intercal z_i - \eta \le \nu_i, \qquad i \in R \\
& x, z_i, \nu_i \ge 0,
\end{aligned}
\tag{4.20}
$$

where α' is defined in (4.5) and $\hat{P}_i$ is given by (2.17)

5 Experimental Results

In this section we present some numerical experiments to show the performance of Algorithm 3.

5.1 Methodology

To test the performance of our algorithm RASRA we adopted the following methodology. We chose a few two-stage stochastic programs from the literature and modified their objective function to minimize CVaR_α instead of the expectation (we used $\alpha = 90\%$ in all problems). We then generated large samples (of size 10,000) from the underlying distributions. The samples were generated with the Latin hypercube sampling technique (McKay, Beckman, & Conover, 1979) so ensure that they had reasonably good coverage of the whole space. The distribution corresponding to the samples is the *source* distribution.

We first solved these problems to optimality using the source distribution, and took the optimal value to represent the "true" optimal value of the problem. As the purpose of the experiments was to test the performance of RASRA, we chose relatively small problems that could be solved to optimality (even with a large sample) so we could compare their optimal values.

With the source distribution fixed, we then compared our algorithm with the standard probability metrics approach of Dupačová et al. (2003) applied to the expectation formulation (2.4)-(2.5) of the CVaR problem. Note that for this approach it was not possible to solve (2.16) exactly due to the size of the scenario set Ξ; thus, we applied the Backward Reduction heuristics described in that paper. We applied the scenario reduction for target cardinalities M equal to 200, 300 and 400.

To compare the performance of the Backward Reduction heuristic and our RASRA algorithm, we used the following metric:

$$\mu \ := \ \frac{V(P_o) - V(P_r)}{V(P_o)} \tag{5.1}$$

where $V(P_o)$ is the "true" optimal value of (2.3) and $V(P_r)$ is the optimal value of the problem obtained with a reduced set of scenarios. As the metric indicates the percentage deviation from $V(P_o)$, it is desirable that it takes values closer to zero. The above procedure was repeated ten times, to ensure that the results were not just product of one experiment. We then calculated box-plots of the metric μ in (5.1) for the various values of M.

As discussed in Section 4.1, we expect our approach to yield good results when there are not many scenarios with cost equal to $\text{VaR}_\alpha[G(x^*, \cdot)]$. We provide an auxiliary heuristics to help us check in advance whether we should expect good results. The idea is to draw a random sample of size $M = 400$, calculate the optimal solution x_M for that sample and sort the values of $G(x_M, \xi_i)$, $i = 1, \ldots, M$ in ascending order. If we see many scenarios with cost equal to $\text{VaR}_\alpha[G(x_M, \cdot)]$, we probably cannot expect RASRA to yield very good results. Otherwise, the set D should have a large size and RASRA should perform well. We repeat the procedure four times to avoid the effects of a single sample. We call this procedure the *exploratory heuristics* in the sequel.

All the computations were done in a Intel I7-4770 (3.40GHz), 16 GB of RAM Windows 7 computer. The programming language used to implement the backward reduction heuristic of Dupačová et al. (2003) and Algorithm 3 was Python 2.7. The ZEBRA algorithm (García et al., 2011) is originally implemented in Microsoft Visual C++ 2005 and the LP solver was CPLEX 11.0.[4] This implementation has been adapted to receive our input data in Microsoft Visual C++ 2017, and uses CPLEX 12.7 as LP solver. In order to solve problem (4.20) we use GUROBI 6.5. The value of$V(P_o)$ in each experiment was calculated by solving teh problem using the Bouncing Nested

[4] An implementation of ZEBRA can be found in https://sites.google.com/site/sergiogarciaquiles/the-p-median-facility-location-problem.

Benders Solver (BNBS) method (Altenstedt, 2003) on the NEOS-Server (Czyzyk, Mesnier, & Moré, 1998; Dolan, 2001; Gropp & Moré, 1997).

We now briefly describe the problems we study and present the results. The descriptions are mostly taken from (Linderoth, Shapiro, & Wright, 2006), where these problems (with the expectation formulation) are studied.

5.2 LandS

`LandS` is an energy planning problem (originally proposed in Louveaux and Smeers (1988)) where we must minimize the cost of investment. The first stage variables represent capacities of different new technologies and first stage constraints represent minimum capacity and budget restriction. Second stage variables represent production of each of the three different modes of electricity from each of the technologies and second stage constraints include capacity constraints for each technology, and constraints imposing that the (random)demand must be satisfied.

Figure 1 shows the results for the exploratory heuristics. The figure depicts the values of $G(x_M, \xi_i)$ sorted in ascending order, and the red line indicate the value of $\mathrm{VaR}_\alpha\left[G(x_M, \cdot)\right]$. We see that the values of $G(x_M, \xi_i)$, $i = 1, \dots, M$ are strictly increasing, so there is only one scenario with cost equal to $\mathrm{VaR}_\alpha\left[G(x_M, \cdot)\right]$. Thus, we expect our algorithm to work well, as the set $\hat{D}$ of scenarios discarded by Algorithm 2 will have approximately 90% of the total number of scenarios.

Figure 1: Exploratory heuristics for `LandS` problem with 400 scenarios

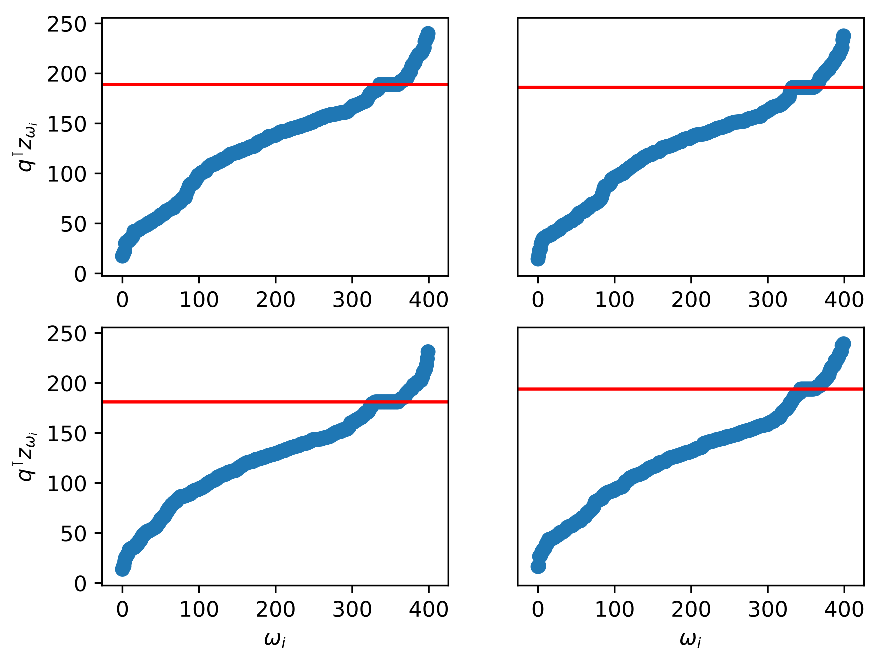
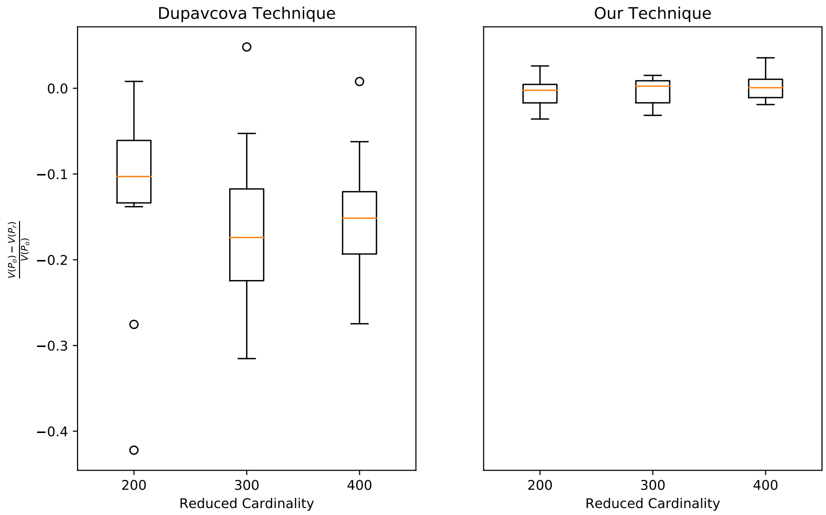


Figure 2 corroborates our expectations. We see that our RASRA algorithm yields very precise estimations of the “true" optimal value for the chosen number M of reduced scenarios. For the same values values of M, however, the Backward Reduction does not perform very well — we see from the boxplots for example that for $M = 300$ the relative gap μ was higher than 15% in at least half of the experiments.

Figure 2: Backward Reduction versus RASRA in LandS problem



5.3 gbd

The `gbd` problem (proposed originally by G. Dantzig (1963)) considers the allocation of different types of aircraft which need to be assigned to routes in a way that maximizes profit under uncertain demand . Besides the cost of operating the aircraft, there are costs associated with bumping passengers when the demand for seats outstrips the capacity. The first stage variables are the number of aircraft of each type allocated to each route and the first-stage constraints bounds number available aircraft of each type. The second-stage variables indicate the number of carried passengers and the number of bumped passengers on each route and second-stage constraints are demand balance equations for the routes.

Again, we start with the exploratory heuristics, shown in Figure 3. Here we see that there are many scenarios with cost equal to $\text{VaR}_\alpha\left[G(x_M, \cdot)\right]$, and in fact there is no scenario with cost strictly below that value. Thus, we do not expect our algorithm to perform particularly well in this case.

Figure 4 shows the box-plots for the relative gaps μ. While our method still outperforms the Backward Heuristics, we see that the box-plots for RASRA are farther form zero than in the `LandS` case.

5.4 gbd-sk3

Problem `gbd-sk3` is a modification of the `gbd` proposed in Cotton and Ntaimo (2015) which changes the distribution of the underlying random variable. This change in the distribution shows a significant difference in the graphs depicting the exploratory heuristics. We see a behavior similar to the `LandS` problem, so we expect RASRA to perform equally well.

Figure 6 again corroborates our expectations. Note however that, while RASRA outperforms the Backward Reduction method. the latter also performs very well, with largest relative gap being equal to 0.5% for $M = 200$.

Figure 3: Exploratory heuristics for gbd problem with 400 scenarios

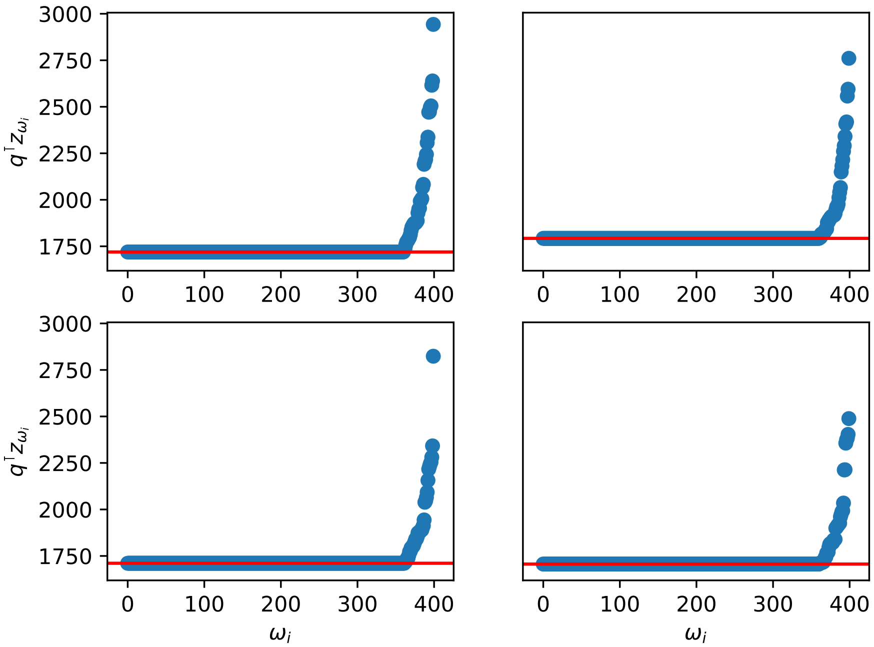
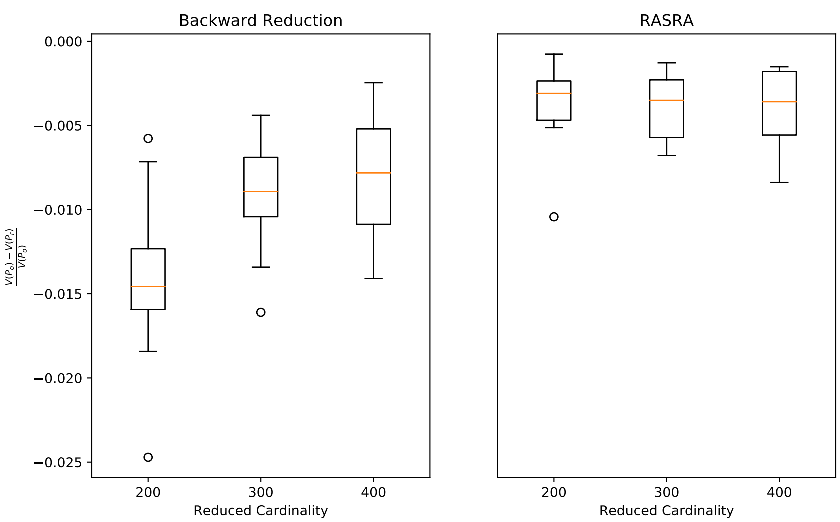


Figure 4: Backward Reduction versus RASRA in gbd problem



6 Conclusions

We have presented a new scenario reduction approach for risk-averse stochastic optimization problems where the objective function is the Conditional Value-at-Risk. Our approach, which relies on the notion of effective scenarios recently introduced in the literature, makes precise the intuitive notion that "only the scenarios in the tail matter" and alerts for the case in which many scenarios

Figure 5: Exploratory heuristics for `gbd-sk3` problem with 400 scenarios

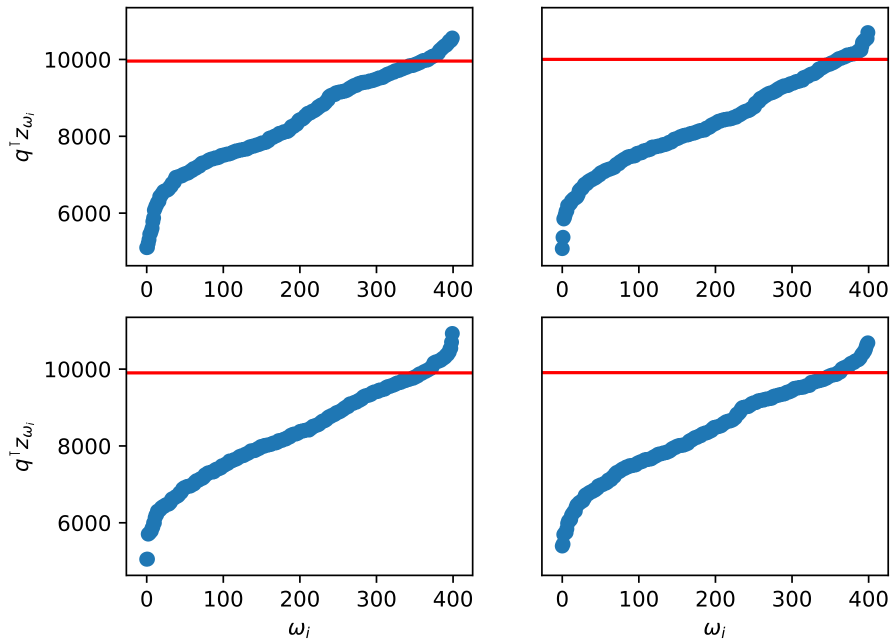
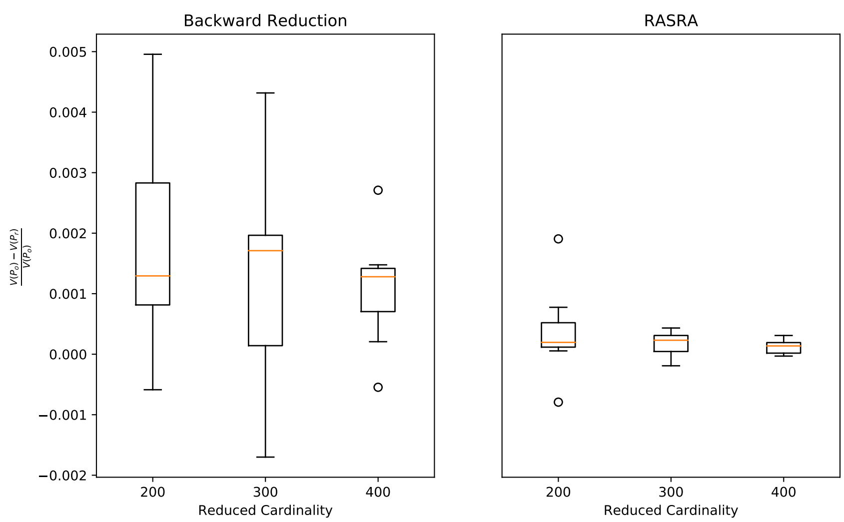


Figure 6: Backward Reduction versus RASRA in gbd-sk3 problem



have the same cost equal to the Value-at-Risk, an issue that appears to have overlooked in the literature.

Our method combines techniques from three different fronts: the concepts of effective scenarios, classical scenario reduction techniques based on probability metrics, and computational methods for facility location problems. While it has been recognized in the literature that some scenario reduction methods are related to facility location problems, it appears that the vast literature (and computational methods) on that topic has not been much exploited in the context of scenario reduction. The numerical results presented in this paper suggest that our approach can be very useful for the problems it proposes to solve.

Acknowledgments

This work has been supported by FONDECYT 1171145, Chile.

References

Ahmed, S. (2006). Convexity and decomposition of mean-risk stochastic programs. *Mathematical Programming*, *106*(3), 433–446.

Altenstedt, F. (2003). *Aspects on asset liability management via stochastic programming.* Chalmers University of Technology.

Artzner, P., Delbaen, F., Eber, J.-M., & Heath, D. (1999). Coherent measures of risk. *Math Financ*, *9*, 203–227.

Birge, J. R., & Louveaux, F. (2011). *Introduction to stochastic programming.* Springer Science & Business Media.

Cotton, T. G., & Ntaimo, L. (2015). Computational study of decomposition algorithms for mean-risk stochastic linear programs. *Mathematical Programming Computation*, *7*(4), 471–499.

Czyzyk, J., Mesnier, M. P., & Moré, J. J. (1998). The neos server. *IEEE Journal on Computational Science and Engineering*, *5*(3), 68 – 75.

Dantzig, G. (1963). *Linear programming and extensions.* Princeton, New Jersey: Princeton University Press.

Dantzig, G. B. (1955). Linear programming under uncertainty. *Management science*, *1*(3-4), 197–206.

Dolan, E. D. (2001). *The neos server 4.0 administrative guide* (Technical Memorandum No. ANL/MCS-TM-250). Mathematics and Computer Science Division, Argonne National Laboratory.

Dupačová, J., Gröwe-Kuska, N., & Römisch, W. (2003). Scenario reduction in stochastic programming. *Mathematical programming*, *95*(3), 493–511.

Dupačová, J., Consigli, G., & Wallace, S. W. (2000). Scenarios for multistage stochastic programs. *Ann Oper Res*, *100*, 25–53.

Dupačová, J., Gröwe-Kuska, N., & Römisch, W. (2003). Scenario reduction in stochastic programming: An approach using probability metrics. *Math Program*, *95*, 493–511.

Eichhorn, A., & Römisch, W. (2008). Stability of multistage stochastic programs incorporating polyhedral risk measures. *Optimization*, *57*(2), 295–318.

Espinoza, D., & Moreno, E. (2014). A primal-dual aggregation algorithm for minimizing conditional value-at-risk in linear programs. *Computational Optimization and Applications*, *59*(3), 617–638.

Fairbrother, J., Turner, A., & Wallace, S. (2015). Scenario generation for stochastic programs with tail risk measures. *arXiv preprint arXiv:1511.03074*.

García, S., Labbé, M., & Marín, A. (2011). Solving large p-median problems with a radius formulation. *INFORMS Journal on Computing*, *23*(4), 546–556.

García-Bertrand, R., & Mínguez, R. (2014). Iterative scenario based reduction technique for stochastic optimization using conditional value-at-risk. *Optimization and Engineering*, *15*(2), 355–380.

Gropp, W., & Moré, J. J. (1997). Optimization environments and the neos server. In Martin D. Buhman and Arieh Iserles (Ed.), *Approximation theory and optimization* (pp. 167 – 182). Cambridge University Press.

Guigues, V., Krätschmer, V., & Shapiro, A. (2016). Statistical inference and hypotheses testing of risk averse stochastic programs. *arXiv preprint arXiv:1603.07384*.

Heitsch, H., & Römisch, W. (2003). Scenario reduction algorithms in stochastic programming. *Comput Optim Appl*, *24*, 187–206.

Heitsch, H., & Römisch, W. (2009). Scenario tree modeling for multistage stochastic programs. *Math Program*, *118*, 371–406.

Homem-de-Mello, T., & Bayraksan, G. (2014). Monte Carlo sampling-based methods for stochastic optimization. *Surveys in Operations Research and Management Science*, *19*, 56–85.

Hoyland, K., Kaut, M., & Wallace, S. W. (2003). A heuristic for moment-matching scenario generation. *Comput Optim Appl*, *24*, 169–185.

Hoyland, K., & Wallace, S. W. (2001). Generating scenario trees for multistage decision problems. *Manage Sci*, *47*(2), 295–307.

Linderoth, J., Shapiro, A., & Wright, S. (2006). The empirical behavior of sampling methods for stochastic programming. *Annals of Operations Research*, *142*(1), 215–241.

Louveaux, F., & Smeers, Y. (1988). Optimal investments for electricity generation: A stochastic model and a test problem. In Y. Ermoliev & R. J.-B. Wets (Eds.), *Numerical techniques for stochastic optimization problems* (pp. 445–452). Berlin: Springer-Verlag.

McKay, M. D., Beckman, R. J., & Conover, W. J. (1979). A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, *21*, 239–245.

Mehrotra, S., & Papp, D. (2014). A cutting surface algorithm for semi-infinite convex programming with an application to moment robust optimization. *SIAM Journal on Optimization*, *24*(4), 1670–1697.

Miller, N., & Ruszczynski, A. (2011). Risk-averse two-stage stochastic linear programming: Modeling and decomposition. *Oper Res*, *59*, 125–132.

Noyan, N. (2012). Risk-averse two-stage stochastic programming with an application to disaster management. *Computers & Operations Research*, *39*(3), 541–559.

Pflug, G. C. (2001). Scenario tree generation for multiperiod financial optimization by optimal discretization. *Mathematical Programming, Series B*, *89*(2), 251–271.

Pflug, G. C., & Pichler, A. (2011). Approximations for probability distributions and stochastic optimization problems. In M. Bertocchi, G. Consigli, & M. A. H. Dempster (Eds.), *Stochastic optimization methods in finance and energy* (pp. 343–387). Springer.

Pineda, S., & Conejo, A. (2010). Scenario reduction for risk-averse electricity trading. *IET generation, transmission & distribution*, *4*(6), 694–705.

Rachev, S. T. (1991). *Probability metrics and the stability of stochastic models* (Vol. 269). John Wiley & Son Ltd.

Rahimian, H., Bayraksan, G., & Homem-de Mello, T. (2018). Identifying effective scenarios in distributionally robust stochastic programs with total variation distance. *Mathematical Programming*. Retrieved from https://doi.org/10.1007/s10107-017-1224-6 doi: 10.1007/s10107-017-1224-6

Rockafellar, R. T., & Uryasev, S. P. (2000). Optimization of conditional value-at-risk. *J. Risk*, *2*, 21–41.

Rockafellar, R. T., & Wets, R. J. (1998). *Variational analysis: Grundlehren der mathematischen wissenschaften.* Springer Berlin.

Römisch, W., & Wets, R.-B. (2007). Stability of ε-approximate solutions to convex stochastic programs. *SIAM Journal on Optimization*, *18*(3), 961–979.

Shapiro, A. (2003). Inference of statistical bounds for multistage stochastic programming problems. *Math. Meth. Oper. Res.*, *58*, 57–68.

Shapiro, A., Dentcheva, D., & Ruszczyński, A. (2014). *Lectures on stochastic programming : modeling and theory* (2nd ed.). SIAM.

Wallace, S. W., & Ziemba, W. T. (2005). *Applications of stochastic programming* (Vol. 5). Siam.