Mathematical Programming (2020) 181:453–472
https://doi.org/10.1007/s10107-019-01462-4

FULL LENGTH PAPER

Series B

Minimizing buffered probability of exceedance by progressive hedging

R. Tyrrell Rockafellar[1] · Stan Uryasev[2]

Received: 31 March 2018 / Accepted: 23 December 2019 / Published online: 13 January 2020

Abstract

Stochastic programming problems have for a long time been posed in terms of minimizing the expected value of a random variable influenced by decision variables, but alternative objectives can also be considered, such as minimizing a measure of risk. Here something different is introduced: minimizing the buffered probability of exceedance for a specified loss threshold. The buffered version of the traditional concept of probability of exceedance has recently been developed with many attractive properties that are conducive to successful optimization, in contrast to the usual concept, which is often posed simply as the probability of failure. The main contribution here is to demonstrate that in minimizing buffered probability of exceedance the underlying convexities in a stochastic programming problem can be maintained and the progressive hedging algorithm can be employed to compute a solution.

Keywords Convex stochastic programming problems · Probability of failure · Probability of exceedance · Buffered probability of failure · Buffered probability of exceedance · Quantiles · Superquantiles · Conditional value-at-risk · Progressive hedging algorithm

Mathematics Subject Classification 90C15

✉ R. Tyrrell Rockafellar
rtr@uw.edu
http://sites.math.washington.edu/~rtr/mypage.html

Stan Uryasev
uryasev@ufl.edu
http://www.ise.ufl.edu/uryasev

1 Department of Mathematics, University of Washington, Box 354350, Seattle, WA 98195-4350, USA

2 Department of Systems and Industrial Engineering, University of Florida, Gainesville, FL 32611, USA

1 Introduction

Since its inception in the 1970s, most of the work and applications in stochastic programming, whether one-stage, two-stage or multistage, have centered on the minimization of an expected cost of some type. In recent years the alternative of minimizing a measure of risk has aroused interest, but there is another idea to explore as well. Many practical problems in stochastic modeling center on "probability of failure" because of stipulations in contracts or regulations. Failure can be modeled in terms of some random variable X representing "hazard", "loss" or "cost" by tying it to the instances of X coming out > 0. The probability of failure is then the probability of such an instance, which is $1 - F_X(0)$ for the cumulative distribution function F_X of X. Could the minimizing of probability of failure be an attractive goal in some situations? A difficulty is its often poor mathematical behavior as a function of the decision variables on which X may depend, especially when F_X might only be a step function with vital discontinuities, as well as an inability to take advantage of underlying convexities in a problem's formulation.

In the face of such shortcomings, a better behaved, yet more conservative, substitute for probability of failure has been developed, called *buffered* probability of failure [12–17], which supports an approach to risk and reliability in engineering that can address the *scope* of a "failure" together with its likelihood. It does this by looking at the conditional-value-at-risk $\mathrm{CVaR}_\alpha(X)$ of X at various probability levels α [23,24]. Recall that $\mathrm{CVaR}_\alpha(X)$ is the expected value in the α-tail distribution of X (in the sense clarified in [24] to allow for discontinuities in F_X). In its dependence on α, it rises continuously and monotonically from EX at $\alpha = 0$ to the essential supremum $\sup X$ at $\alpha = 1$. As long as $EX < 0 < \sup X$, there is a unique probability level $\alpha \in (0, 1)$ such that $\mathrm{CVaR}_\alpha(X) = 0$, and $1 - \alpha$ is by definition then the buffered probability of failure associated with X. The buffered probability of failure is taken to be 0 if $\sup X \leq 0$ and 1 if $0 \leq EX$.

Here we imbed failure in the broader setting of "exceedance" with a parameter τ replacing 0, which facilitates thinking in many situations. For instance, X might stand for the ultimate cost of a project that has to be brought to completion in uncertain circumstances, and various dollar cost thresholds τ might be of concern with respect to potential cost overruns. The *τ-probability of exceedance* $\mathrm{POE}_\tau(X)$ is the probability that $X > \tau$; equivalently it's the probability of failure of $X - \tau$. Likewise, the *τ-buffered probability of exceedance* $\mathrm{bPOE}_\tau(X)$ is the buffered probability of failure of $X - \tau$, so that

$$\begin{aligned} &\text{if } EX < \tau < \sup X, \text{ then } \mathrm{bPOE}_\tau(X) = 1 - \alpha \text{ for the } \alpha \in (0,1) \text{ giving } \mathrm{CVaR}_\alpha(X) = \tau, \\ &\text{while if } \tau \geq \sup X, \text{ then } \mathrm{bPOE}_\tau(X) = 0, \text{ but if } \tau \leq EX, \text{ then instead } \mathrm{bPOE}_\tau(X) = 1. \end{aligned} \tag{1.1}$$

The systematic investigation of buffered probability of exceedance (bPOE) in contrast to traditional probability of exceedance (POE) was launched in [4,6,31].

From (1.1) it's evident that the function $\tau \mapsto 1 - \mathrm{bPOE}_\tau(X)$ is essentially the inverse of the function $\alpha \mapsto \mathrm{CVaR}_\alpha(X)$. This parallels the fact that the function

$\tau \mapsto 1 - \mathrm{POE}_\tau(X)$ is essentially the inverse of the quantile function $\alpha \mapsto q_\alpha(X)$, since $q_\alpha(X) = \min\left\{\tau \,\middle|\, F_X(\tau) \geq \alpha\right\}$ (with "essentially" referring to adjustments needed when F_X isn't continuous or persistently increasing). Alternative terminology enhances this parallelism. In finance, the quantile $q_\alpha(X)$ is the *value-at-risk* $\mathrm{VaR}_\alpha(X)$. On the other hand, starting in [12], the conditional-value-at-risk $\mathrm{CVaR}_\alpha(X)$ has been dubbed the α-*superquantile* of X and denoted by $\overline{q}_\alpha(X)$ in order to liberate engineering applications from language tied to finance. For a rigorous account of these "generalized inverse" relationships and their connections with convex analysis, see [20]. Additional background on superquantiles and risk can be found in [18,19,25].

With bPOE, it is possible to take into account outcomes close to a threshold, rather than merely counting outcomes exceeding it, since (generally speaking) bPOE considers tail outcomes that average to some specific value. For instance, 4% of land-falling hurricanes in the United States have cumulative damage exceeding \$50 billion, so POE = 0.04 for the threshold τ = \$50 billion. However \$50 billion is estimated to be the *average* damage from the worst 10% of hurricanes, so bPOE = 0.1 for the threshold τ = \$50 billion. Thus, bPOE can be an important supplement to POE, and it would be good to calculate it routinely alongside of POE, which typically gets all the attention but may be misleading about the seriousness of risks. POE is behind common terms like 100-year storm or 50-year flood, but the buffered version is arguably superior in assessments of hazards like in the hurricane example, as underscored in [1].

In addition to applications of bPOE in the case of $\tau = 0$, which come under the heading of buffered probability of failure [12–17], and the valuable import of bPOE for natural hazards, the concept has put to use in areas such as finance. Portfolio optimization algorithms and performance functions for cash flow matching have been explored in [30]. Paper [33] established a connection of bPOE with the monotone Sharpe ratios and related measures of investment performance. Paper [2] demonstrated how to use bPOE in optimization of PDE-constrained systems with uncertain coefficients and relevant applications. Applications to machine learning have been looked at in [6,7]. A deterministic variant of bPOE has been employed to define *cardinality of upper average* (CUA); this refers to the number of the largest components of a vector having average value beyond some threshold. The CUA characteristic was introduced in [8] and applied to network optimization problems. Further applications of bPOE are in the offing and the range of potential benefits could be enormous. Technical support has moreover been provided in analytical formulas for bPOE for various distributions in [9]. Statistical estimators of bPOE and their convergence rates have been given in [5]. A comprehensive study of the properties of bPOE in [4] is available as an introduction and a source of key facts. Higher-moment buffered probabilities were studied in [3].

With this motivation, our aim in the present paper is to determine the extent to which stochastic programming problems might successfully be solved if the usual expectational objective, taking the form of the expected value of an expression of loss under uncertainty, is replaced by a bPOE objective. More specifically, our focus is on whether the progressive hedging algorithm (PHA) of [28] can be adapted to such a different objective. A top issue there is whether convexity of the loss expression with respect to decision variables can be put to work; such convexity is behind the viability of that method. Also crucial is the "stochastic separability" that's inherent in expectations with their risk neutrality but can be lacking in risk-averse problem formulations. For that, at least, a previous effort in [11] at extending PHA to CVaR-type objectives can suggest guidelines. There it was possible, by utilizing a CVaR formula from [23,24], to translate a CVaR-type objective back to an expectational objective through the introduction of auxiliary variables. Convexity-preserving properties of that formula opened the way then for PHA.

An analogous formula for bPOE, which came to light in [6], offers a tantalizing prospect of replicating that advance. Unfortunately, in the presence of convexity in the problem's ingredients, that formula only manages to achieve quasi-convexity in the expected loss expression that it furnishes. Nevertheless we will be able here to go further and actually attain convexity by employing facts brought out in [4].

In proceeding toward that goal we must first, in Sect. 2, review some fundamentals about stochastic programming and PHA in order to have the right platform for our extension. Afterwards, in Sect. 3, we will be able to present and justify the details of our contribution. Because PHA operates with discrete probability, special attention must be devoted to the peculiarities of bPOE in such a setting.

A numerical study is offered in Sect. 4. Although the study is focused on a single-stage model, a multistage model is taken as the basis for our general explanations and developments, as was the case for the work on CVaR adaptations in [11].

Multistage models can raise issues of time consistency in their justification. Time consistency wasn't addressed directly in [11], which only looked at the CVaR associated with a terminal cost, but an accommodation could easily be made in that by passing to a "nested CVaR" objective and utilizing the same tricks. Here likewise for bPOE, the centerpiece is a terminal expression instead of some kind of nested expression. But the situation is different, as can be understood by recalling that bPOE is allied in spirit with probability of failure, and probabilities aren't additive. Whether some "nested" approach to failure probabilities or exceedance probabilities might make sense in some context is questionable, and anyway needn't be explored here. Multistage models concern initial decisions with subsequent opportunities for recourse decisions, and notions of "failure" are naturally associated with the final result of all those decisions.

2 Background in stochastic programming and progressive hedging

At the computationally oriented level of stochastic programming with convexity that we get into as the foundation for progressive hedging, the probability framework is elementary. There is a discrete probability space Ξ consisting of finitely many $scenarios/\xi$ with known probabilities $\pi(\xi) > 0$, adding to 1. The random variables X to be encountered will be real-valued functions on Ξ with staircase-type F_X. A time component enters, though, because decisions are to be made in stages $k = 1, \ldots, N$ that capture the evolution of information. For this we pose the scenarios as elements $\xi = (\xi_1, \ldots, \xi_N)$ of a product space $\Xi_1 \times \cdots \times \Xi_N$, with $\xi_k \in \Xi_k$ standing for the information revealed in stage k.

Scenario-dependent decisions in this context take the form of *policies*

$$x(\cdot) : \xi \mapsto x(\xi) = (x_1(\xi), \ldots, x_N(\xi)) \in \mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_N} = \mathbb{R}^n, \tag{2.1}$$

where $x_k(\xi) \in \mathbb{R}^{n_k}$ is the decision component in stage k. A critical constraint on policies is that they must be *nonanticipative*, which means that the kth-stage decision can only depend on the information revealed in previous stages (and in particular the first-stage decision must be scenario-independent). Another constraint will be that

$$x(\xi) \in C(\xi) \text{ for all scenarios } \xi \tag{2.2}$$

with respect to some specification of nonempty closed convex sets $C(\xi) \subset \mathbb{R}^n$. Within the finite-dimensional linear space

$$\mathcal{L} = \{\text{all policies } x(\cdot) \text{ in (2.1)}\} \tag{2.3}$$

the concern is therefore with policies $x(\cdot) \in \mathcal{C} \cap \mathcal{N}$, where

$$\begin{aligned} \mathcal{C} &= \{\text{the policies } x(\cdot) \text{ satisfying (2.2)}\}, \\ \mathcal{N} &= \{\text{the policies such that } x_k(\xi) \text{ depends only on} (\xi_1, \ldots, \xi_{k-1})\}. \end{aligned} \tag{2.4}$$

Here $\mathcal{C}$ is a nonempty closed convex subset of $\mathcal{L}$ and $\mathcal{N}$ is a subspace of $\mathcal{L}$.

The contemplated optimization will be to minimize over $\mathcal{C} \cap \mathcal{N}$ an objective functional $\mathcal{F}$ defined on $\mathcal{L}$. The objective will be tied to a "cost" structure dictated by scenario-dependent convex functions on $\mathbb{R}^n$:

$$f(\cdot, \xi) : \mathbb{R}^n \to \mathbb{R} \text{ for each scenario } \xi. \tag{2.5}$$

Such structure associates with any policy $x(\cdot)$ a "cost" random variable

$$X_f(x(\cdot)) : \xi \in \Xi \;\mapsto\; f(x(\xi), \xi) \in \mathbb{R}, \tag{2.6}$$

but a random variable itself can't be minimized. It must first be transformed in one way or another into a scalar value associated with $x(\cdot)$ to define as $\mathcal{F}(x(\cdot))$.

The long-standing approach to that is to pass to the expectation of $X_f(x(\cdot))$. The optimization problem is then to

$$\begin{aligned} &\text{minimize } \mathcal{F}(x(\cdot)) \text{ subject to } x(\cdot) \in \mathcal{C} \cap \mathcal{N}, \text{ where} \\ &\mathcal{F}(x(\cdot)) = E\big[X_f(x(\cdot))\big] = E_\xi[f(x(\xi), \xi)] = \sum_{\xi \in \Xi} \pi(\xi) f(x(\xi), \xi). \end{aligned} \tag{2.7}$$

With this choice the functional $\mathcal{F}$ on $\mathcal{L}$ is convex and finite, hence also continuous (due to the finite-dimensionality of $\mathcal{L}$). But the expectation choice isn't the only attractive possibility. One could instead choose

$$\mathcal{F}(x(\cdot)) = \mathcal{R}\big(X_f(x(\cdot))\big) \text{ for some risk measure } \mathcal{R}, \tag{2.8}$$

which was the topic in [11] in the case of $\mathcal{R}$ being CVaR_α, or a mixture $\lambda_1 \mathrm{CVaR}_{\alpha_1} + \cdots + \lambda_m \mathrm{CVaR}_{\alpha_m}$ with weights $\lambda_i > 0$ adding to 1. Coming up in next section of this paper will be a bPOE version of $\mathcal{F}$ obtained by applying $\mathrm{bPOE}_\tau(X)$ to $X = X_f(x(\cdot))$. The discussion of the moment, though, concentrates on the choice of $\mathcal{F}$ in (2.7) and how it fits with progressive hedging.

Very important for that is the subspace representation of the nonanticipativity constraint as $x(\cdot) \in \mathcal{N}$, because it leads to the introduction of "multipliers" able to dualize that constraint in computations.[1] When $\mathcal{L}$ is furnished with the expectational inner product

$$\langle x(\cdot), x'(\cdot)\rangle = E_\xi[x(\xi)\cdot x'(\xi)] = \sum_{\xi\in\Xi}\pi(\xi)\sum_{k=1}^N \big[x_k(\xi)\cdot x_k'(\xi)\big], \tag{2.9}$$

it becomes a (finite-dimensional) Hilbert space, and the "multipliers" in question lie in the orthogonal complement of $\mathcal{N}$. In denoting the conditional expectation with respect to $\xi = (\xi_1, \ldots, \xi_N)$ given its initial components $\xi_1, \ldots, \xi_{k-1}$ by $E_{\xi\,|\,\xi_1,\ldots,\xi_{k-1}}$, that complementary subspace $\mathcal{M} = \mathcal{N}^\perp$ is

$$\mathcal{M} = \Big\{\, w(\cdot) \in \mathcal{L} \,\Big|\, E_{\xi\,|\,\xi_1,\ldots,\xi_{k-1}} w_k(\xi) = 0,\ \forall k \Big\}. \tag{2.10}$$

The key observation is that finding a solution $\bar{x}(\cdot)$ to (2.7) can be recast very generally as a problem involving dual elements in $\mathcal{M}$ alongside of primal elements in $\mathcal{N}$:

$$\begin{aligned} &\text{find } \bar{x}(\cdot) \in \mathcal{N},\ \bar{w}(\cdot) \in \mathcal{M}, \text{ such that } \forall \xi \in \Xi,\\ &\bar{x}(\xi) \in \operatorname*{argmin}_{x(\xi)\in\mathbb{R}^n}\Big\{ f(x(\xi),\xi) + \bar{w}(\xi)\cdot x(\xi)\Big\}, \end{aligned} \tag{2.11}$$

where the argmin condition is equivalent to

$$\bar{x}(\cdot) \in \operatorname*{argmin}_{x(\cdot)\in\mathcal{L}} \Big\{ \mathcal{F}(x(\cdot)) + \langle \bar{w}(\cdot)\cdot x(\cdot)\rangle\Big\}. \tag{2.12}$$

Namely, if $\bar{x}(\cdot)$ and $\bar{w}(\cdot)$ satisfy (2.11), then $\bar{x}(\cdot)$ solves (2.7). Conversely, if $\bar{x}(\cdot)$ solves (2.7) and a constraint qualification is satisfied, then there will exist a companion $\bar{w}(\cdot)$ to $\bar{x}(\cdot)$ for which (2.11) holds. An example of a constraint qualification that works is the existence of some $\tilde{x}(\cdot) \in \mathcal{N}$ such that $\tilde{x}(\cdot) \in \mathrm{ri}\,\mathcal{C}$, which is the same as $\tilde{x}(\xi) \in \mathrm{ri}\,C(\xi)$ for all ξ. When the convex sets $C(\xi)$ are polyhedral, even that isn't needed.

Progressive hedging takes advantage of this recasting of the stochastic programming problem (2.7) and makes use of the mappings

$$P_\mathcal{N} = \text{projection onto } \mathcal{N}, \quad P_\mathcal{M} = \text{projection onto } \mathcal{M}, \quad \text{satisfying } P_\mathcal{N} + P_\mathcal{M} = I. \tag{2.13}$$

[1] This represention goes back to [27] and was extended from stochastic programming to stochastic variational inequalities in [29].

Progressive hedging algorithm [28]. *With a parameter value $r > 0$ and iterations indexed by $\nu = 1, 2, \ldots$, proceed as follows from current elements $x^\nu(\cdot) \in \mathcal{N}$ and $w^\nu(\cdot) \in \mathcal{M}$.*

(a) *Calculate $\widehat{x}^\nu(\xi) \in C(\xi)$ for each scenario $\xi \in \Xi$ by solving a strongly convex optimization problem in the variable $x \in \mathbb{R}^n$ that has this as its unique solution:*

$$\widehat{x}^\nu(\xi) = \operatorname*{argmin}_{x \in C(\xi)} \left\{ f(x, \xi) + w^\nu(\xi) \cdot x + \frac{r}{2} ||x - x^\nu(\xi)||^2 \right\}, \tag{2.14}$$

thereby determining a function $\widehat{x}^\nu(\cdot) \in \mathcal{L}$ that is not necessarily in $\mathcal{N}$.

(b) *Update then to iteration $\nu + 1$ by*

$$x^{\nu+1}(\cdot) = P_{\mathcal{N}}[\widehat{x}^\nu(\cdot)], \qquad w^{\nu+1}(\cdot) = w^\nu(\cdot) + r P_{\mathcal{M}}[\widehat{x}^\nu(\cdot)], \tag{2.15}$$

which means taking for each scenario $\xi \in \Xi$ and stage k

$$x_k^{\nu+1}(\xi) = E_{\xi \mid \xi_1, \ldots, \xi_{k-1}} \widehat{x}_k(\xi), \qquad w_k^{\nu+1}(\xi) = w^\nu(\xi) + r[\widehat{x}^\nu(\xi) - x^{\nu+1}(\xi)]. \tag{2.16}$$

The equivalence of (2.15) with the easy rule in (2.16) corresponds, from the perspective of (2.10), to the fact that $\mathcal{N} = \mathcal{N}^{\perp\perp} = \mathcal{M}^\perp$ and $P_{\mathcal{M}} = I - P_{\mathcal{N}}$. The algorithm combines the accessibility of that calculation with the simpicity of only needing to solve nice single-scenario problems in (2.14), which can be done by means of convex programming software that bypasses stochastic issues. Of course, solving these subproblems could bring up additional multipliers associated with constraint systems specifying the sets $C(\xi)$, but that doesn't have to come into the picture here.[2]

Convergence of the progressive hedging algorithm requires knowing that a solution $\bar{x}(\cdot)$ to the given stochastic programming problem does exist and can be combined with a multiplier vector $\bar{w}(\cdot)$.[3] Then, according to [28, Theorem 5.1], the sequence of pairs $(x^\nu(\cdot), w^\nu(\cdot))$ generated by the procedure from any starting pair is sure to converge to a particular pair $(\bar{x}(\cdot), \bar{w}(\cdot))$ satisfying (2.13), with $\bar{x}(\cdot)$ accordingly being a solution to (2.7). In this convergence the expression

$$r||x^\nu(\cdot) - \overline{x}(\cdot)||^2 + r^{-1}||w^\nu(\cdot) - \overline{w}(\cdot)||^2 \tag{2.17}$$

will be decreasing. (If a solution didn't exist, this expression would tend to ∞.)

The role of r in the expression (2.17) yields an important insight about progressive hedging. It reveals a trade-off in the algorithm's behavior with respect to the primal elements $x^\nu(\cdot)$ and the dual elements $w^\nu(\cdot)$. Clearly, a high value of r emphasizes primal convergence whereas a low value of r emphasizes dual convergence. This may need

[2] An extension of progressive hedging to iterate also on such multipliers with extra proximal terms is available in [22].

[3] The second part of this has already been addressed; the existence part can be handled by compactness of the sets $C(\xi)$ or more generality some joint aspects of these sets and growth properties of the functions $f(\cdot, \xi)$. For more on these issues, see [28, §4].

to be tuned to the circumstances in a particular application, although $r = 1$ is always available. See [32] for more on this topic. Some numerical experience in a related application of progressive hedging to solving monotone stochastic complementarity problems can be found in [21].

Remark It was natural to assume finiteness of the cost expressions in (2.6) for the sake of passing to random variables in (2.7), but as far as progressive hedging is concerned, such finiteness isn't needed. The indicated properties of the algorithm, recalled above from [28], hold even if $f(\cdot,\xi)+\delta_{C(\xi)}$ (indicator) is just a lower semicontinuous proper convex function on $\mathbb{R}^n$ for each $\xi \in \Xi$. This fact will enter the developments in the next section.

3 Adapting to buffered probability of exceedance

The task we have set for ourselves is ascertaining whether the progressive hedging algorithm can be adapted to the version of stochastic programming in which, instead of (2.7), the problem is

$$\text{minimize } \mathcal{F}(x(\cdot)) \text{ subject to } x(\cdot) \in \mathcal{C} \cap \mathcal{N}, \text{ where } \mathcal{F}(x(\cdot)) = \text{bPOE}_\tau\big(X_f(x(\cdot))\big). \tag{3.1}$$

The general meaning of the buffered probability of exceedance now in our spotlight has already been explained in Sect. 1. However, because our setting is one of discrete probability, the random variable $X_f(x(\cdot))$ will be of special type, taking on only finitely many values, namely $f(x(\xi),\xi)$ with probability $\pi(\xi)$ for the finitely many scenarios $\xi \in \Xi$. In particular, such a random variable $X = X_f(x(\cdot))$ has values $\max X$ and $\min X$ that are attained with positive probability. In accordance with (1.1),

$$\text{bPOE}_\tau(X) \begin{cases} \in (0,1] & \text{when } \tau < \max X, \\ = 0 & \text{when} \tau \geq \max X, \end{cases} \tag{3.2}$$

where moreover,

$$\text{bPOE}_\tau(X) = 0 \iff \text{POE}_\tau(X) = 0 \iff \text{prob}\{X \geq \tau\} = 0. \tag{3.3}$$

These observations induce us to avoid expending energy on "trivial" cases of problem (3.1) by making the following *nondegeneracy assumptions*:

$$\nexists x(\cdot) \in \mathcal{C} \cap \mathcal{N} \text{such that} f(x(\xi),\xi) \leq \tau \text{ for all } \xi \in \Xi, \tag{3.4}$$

in which case $\text{bPOE}_\tau\big(X_f(x(\cdot))\big) = 0$, and on the other hand

$$\exists x(\cdot) \in \mathcal{C} \cap \mathcal{N} \text{ such that } f(x(\xi),\xi) < \tau \text{ for some } \xi \in \Xi, \tag{3.5}$$

in which case $\mathrm{bPOE}_\tau\big(X_f(x(\cdot))\big) < 1$. A policy $x(\cdot)$ as in (3.4) would solve (3.1) in the extreme sense of eliminating any risk at all of $f(x(\xi), \xi)$ exceeding τ. No trade-offs would come into play, and the problem would lack interest. Without a policy as in (3.5), we would be in the opposite extreme of every policy $x(\cdot)$ giving the worst possible bPOE value 1; this would constitute a sort of infeasibiity in problem (3.1).

It may be hard to imagine how progressive hedging could be applied to a bPOE objective in (3.1), because PHA relies on the separability with respect to ξ that's enjoyed by the expectational objective in the traditional format (2.7). That difficulty was overcome, however, in the extension of PHA in [11] to the problem

$$\text{minimize } \mathcal{F}(x(\cdot))\text{subject to } x(\cdot) \in \mathcal{C} \cap \mathcal{N}, \text{ where } \mathcal{F}(x(\cdot)) = \mathrm{CVaR}_\alpha\big(X_f(x(\cdot))\big). \tag{3.6}$$

That extension invoked an optimization formula for CVaR developed in [23,24],

$$\mathrm{CVaR}_\alpha(X) = \min_{\tau\in\mathbb{R}} \left\{ \tau + \frac{1}{1-\alpha} E\big[\max\{0, X-\tau\}\big]\right\}, \tag{3.7}$$

to translate (3.6) into a problem with one more decision variable and an objective again written as an expectation. We pursue that lead by putting our hopes on an analogous formula for bPOE in [4, Proposition 2.2]:

$$\mathrm{bPOE}_\tau(X) = \begin{cases} \min\limits_{\lambda\geq 0} E_\xi\Big[\max\{0, \lambda(X(\xi)-\tau)+1\}\Big] & \text{when } \tau \neq \max X,\\ 0 & \text{when } \tau = \max X. \end{cases} \tag{3.8}$$

The exception in this formula is clarified in [4, Proposition 3.2] as being triggered because

$$\min_{\lambda\geq 0} E_\xi\Big[\max\{0, \lambda(X(\xi)-\tau)+1\}\Big] = \mathrm{prob}\{X=\tau\} > 0 \text{ when } \tau = \max X, \tag{3.9}$$

in discrepancy with the value 0 that $\mathrm{bPOE}_\tau(X)$ should have for $\tau = \max X$ by (3.3). Under our nondegeneracy assumption (3.4), though, this can't come up in problem (3.1). Our nondegeneracy assumption (3.5) instead excludes $\lambda = 0$ from coming into play at the minimum in problem (3.1), since the expression being minimized in (3.8) equals 1 when $\lambda = 0$.

Problem Reformulation 1 *Under our assumptions* (3.4) *and* (3.5), *problem* (3.1) *can be solved by solving*

$$\textit{minimize } E_\xi\Big[\max\{0, \lambda(f(x(\xi),\xi)-\tau)+1\}\Big] \textit{with respect to } \lambda \in [0,\infty),\ x(\cdot) \in \mathcal{C}\cap\mathcal{N}. \tag{3.10}$$

Namely, solution pairs $\lambda, x(\cdot)$, *to* (3.108) *necessarily have* $\lambda > 0$ *and yield solutions* $x(\cdot)$ *to* (3.1).

Anyway, even in the degenerate case, if in solving the substitute problem (3.10) for a given τ, we ended up with a policy $x(\cdot)$ corresponding to the exception in (3.9), i.e., such that $\max_\xi f(x(\xi),\xi) = \tau$, then we would know we had thereby solved the original problem with 0 as the bPOE minimum.

Our next step is to interpret (3.10) as a problem in the expectational format of stochastic programming by *interpreting λ as an additional first-stage decision variable*. For that we introduce, in parallel to (2.2)–(2.3),

$$\begin{aligned}
\widetilde{\mathcal{L}} &= \left\{ \widetilde{x}(\cdot) : \xi \mapsto (\lambda(\xi), x(\xi)) \in \mathbb{R} \times \mathbb{R}^n \right\} \\
\widetilde{\mathcal{C}} &= \left\{ \widetilde{x}(\cdot) \in \widetilde{\mathcal{L}} \,\middle|\, (\lambda(\xi), x(\xi)) \in \widetilde{C}(\xi) = [0,\infty) \times C(\xi) \right\}, \\
\widetilde{\mathcal{N}} &= \left\{ \widetilde{x}(\cdot) \in \widetilde{\mathcal{L}} \,\middle|\, \lambda(\cdot) \equiv \text{constant},\ x(\cdot) \in \mathcal{N} \right\}.
\end{aligned} \tag{3.11}$$

Problem Reformulation 2 *In the notation (3.11) problem (3.10) fits the expectational pattern of multistage stochastic programming in seeking to*

$$\text{minimize } E_\xi[\widetilde{f}(\widetilde{x}(\xi),\xi)] \text{ over policies } \widetilde{x}(\cdot) \in \widetilde{\mathcal{C}} \cap \widetilde{\mathcal{N}} \tag{3.12}$$

for the cost expressions

$$\widetilde{f}(\widetilde{x},\xi) = \widetilde{f}(\lambda, x, \xi) = \max\{0, \lambda(x-\tau)+1\}. \tag{3.13}$$

This appears to bring us very close to the goal of being able to solve the bPOE minization problem by the progressive hedging algorithm, but that proceedure requires more than just an expectated-cost objective. It relies on *convexity* of the cost expressions, but $\widetilde{f}(\lambda, x, \xi))$ in (3.13) isn't convex jointly in λ and x. The expectation in (3.12) does exhibit *quasi-convexity* with respect to $\widetilde{x}(\cdot)$, as shown in [4, Proposition 3.4], but that's not enough.

The way around this obstacle is to make a change of variables as proposed in [4, Proposition 4.5]:

$$(\lambda, x) \longleftrightarrow (\lambda, y) \quad \text{for } \lambda > 0,\ y = \lambda x,\ x = \lambda^{-1} y. \tag{3.14}$$

Problem Reformulation 3 *Problem (3.10), in not having to take $\lambda = 0$ into account under our nondegeneracy assumptions (3.4) and (3.5), can be recast as:*

$$\text{minimize } E_\xi\Big[\max\{0, \lambda(f(\lambda^{-1}y(\xi),\xi)-\tau)+1\}\Big] \text{ over } \lambda \in (0,\infty),\ \lambda^{-1}y(\cdot) \in \mathcal{C} \cap \mathcal{N}. \tag{3.15}$$

Specifically, solution pairs λ, $x(\cdot)$, for problem (3.10) correspond to solutions pairs λ, $y(\cdot)$, for problem (3.15) through the change of variables (3.14).

This change of variables will serve in bootstrapping quasi-convexity into convexity, as will soon be seen, but technicalities have to be handled carefully en route. The version of optimization in (3.15) falls short of the format needed for progressive hedging. With that format as the goal, we would like proceed from (3.15) in the way we did from the problem statement in (3.10) to the one in (3.12), but that's not possible when the expressions in (3.15) are even undefined when $\lambda \leq 0$. We must first manage to make an appropriate extension of them.

Two additional assumptions of mild character will have a part in this. We suppose the constraint sets $C(\xi)$ and cost functions $f(\cdot,\xi)$ satisfy the level-boundedness condition that

$$\big\{x \in C(\xi) \,\big|\, f(x,\xi) \leq \beta\big\} \text{ is bounded for all } \xi \in \Xi,\ \beta \in \mathbb{R}. \tag{3.16}$$

We suppose further that

$$\exists\, \xi \in \Xi \text{ such that } 0 \notin C(\xi), \tag{3.17}$$

or in other words that the do-nothing policy $x(\cdot) = 0$ isn't feasible; it doesn't lie in $\mathcal{C} \cap \mathcal{N}$. In this setting we introduce the sets

$$K(\xi) = \big\{(\lambda, y) \,\big|\, \lambda \geq 0,\ y \in \lambda C(\xi)\big\} \cup \big\{(0, y) \,\big|\, y \in 0^+C(\xi)\big\}, \tag{3.18}$$

where $0^+C(\xi)$ is the recession cone of $C(\xi)$ in convex analysis [10, §8], and the functions

$$h(\lambda, y, \xi) = \begin{cases} \max\{0, \lambda f(\lambda^{-1}y, \xi) - \lambda\tau + 1\} & \text{when } \lambda > 0, \\ \max\{0, f0^+(y,\xi) + 1\} \text{when } \lambda = 0, \text{but } \infty & \text{when } \lambda < 0, \end{cases} \tag{3.19}$$

where $f0^+(\cdot,\xi)$ is the recession function associated with the convex function $f(\cdot,\xi)$, cf. [10, §8].

Example: linear constraints and linear objective. If

$$C(\xi) = \big\{x \in \mathbb{R}^n \,\big|\, x \geq 0,\ A(\xi)x \geq a(\xi),\ B(\xi)x = b(\xi)\big\},$$

the correponding cone in (3.18) is

$$K(\xi) = \big\{(\lambda, y) \in \mathbb{R} \times \mathbb{R}^n \,\big|\, (\lambda, y) \geq (0,0),\ A(\xi)y - \lambda a(\xi) \geq 0, B(\xi)y - \lambda b(\xi) = 0\big\}.$$

If $f(x,\xi) = c(\xi)\cdot x - d(\xi)$, then in (3.19) one has $h(\lambda, y, \xi) = \max\{0, c(\xi)\cdot y - \lambda d(\xi) + 1\}$ when $\lambda \geq 0$, but $h(\lambda, y, \xi) = \infty$ when $\lambda < 0$.

Theorem 1 *Under the additional assumptions* (3.16) *and* (3.17)*, problem* (3.15) *can be stated equivalently in the notation* (3.18)–(3.19) *as the problem*

$$\textit{minimize } E_\xi[h(\lambda, y(\xi), \xi)] \textit{ over } (\lambda, y(\cdot)) \in \mathbb{R} \times \mathcal{N} \textit{ having } (\lambda, y(\xi)) \in K(\xi) \textit{ for all } \xi \in \Xi, \tag{3.20}$$

and it is sure to have at least one solution pair λ, $y(\cdot)$. *The sets* $K(\xi)$ *are closed convex cones, and with respect to their indicators* $\delta_{K(\xi)}$, *the functions*

$$\overline{h} : (\lambda, y) \mapsto h(\lambda, y, \xi) + \delta_{K(\xi)}(\lambda, y) \; \textit{for} \; \xi \in \Xi \tag{3.21}$$

are lower semicontinuous proper convex functions on $\mathbb{R} \times \mathbb{R}^n$. *With those functions,* (3.20) *can be written as*

$$\textit{minimize } E_\xi[\overline{h}(\lambda, y(\xi), \xi)] \textit{ over } (\lambda, y(\cdot)) \in \mathbb{R} \times \mathcal{N}. \tag{3.22}$$

Proof In terms of the functions

$$g(x, \xi) = f(x, \xi) - \tau + \delta_{C(\xi)}(x) \text{ for } x \in \mathbb{R}^n, \; \xi \in \Xi, \tag{3.23}$$

we have

$$\lambda g(\lambda^{-1} y, \xi) = \lambda f(\lambda^{-1} y) - \lambda \tau + \delta_{K(\xi)}(\lambda, y) \text{ when } \lambda > 0, \tag{3.24}$$

and consequently

$$\overline{h}(\lambda, y, \xi) = \max\{0, \lambda g(\lambda^{-1} y, \xi) + 1\} \text{ when } \lambda > 0. \tag{3.25}$$

Problem (3.15) can be therefore be stated equivalently as

$$\text{minimize } E_\xi[\overline{h}(\lambda, y(\xi)] \text{ over } (\lambda, y(\cdot)) \in (0, \infty) \times \mathcal{N}. \tag{3.26}$$

We have to show, among other things, that the definition of $\overline{h}$ on all of $\mathbb{R} \times \mathbb{R}^n$ in (3.21) permits the equivalence with (3.15) to persist when $(0, \infty)$ is replaced in (3.26) by $(-\infty, \infty)$.

This revolves around what happens to the objective in (3.26) as $\lambda \searrow 0$, and considerations of that get into convexity issues. It helps first to look at the sets

$$K_0(\xi) = \left\{(\lambda, y) \,\middle|\, \lambda y \in C(\xi)\right\} \cup \{0, 0\} \subset \mathbb{R} \times \mathbb{R}^n, \tag{3.27}$$

which are convex cones. According to [10, Theorem 8.2], the closure of $K_0(\xi)$ is the cone $K(\xi)$ in (3.19). Next, the closed convexity of $C(\xi)$ and finite convexity of $f(\cdot, \xi)$ make $g(\cdot, \xi)$ be convex with closed level sets $\left\{x \,\middle|\, g(y, \xi) \le \beta\right\}$. The closedness of the level sets means that $g(\cdot, \xi)$ is lower semicontinuous. Our assumption (3.16) makes these level sets also be bounded, and this will soon be needed.

It's known from convex analysis (cf. [10, page 67]) that the convexity of $g(\cdot, \xi)$ ensures the convexity of $\lambda g(\lambda^{-1} y, \xi)$ with respect to $(\lambda, y) \in (0, \infty) \times \mathbb{R}^n$. That convexity holds then also for $\overline{h}(\lambda, y, \xi)$ in (3.25), because the pointwise max of two convex functions is convex. Similarly, the lower semicontinuity of $g(\cdot, \xi)$ ensures that $\overline{h}(\lambda, y, \xi)$ is lower semicontinuous with respect to $(\lambda, y) \in (0, \infty) \times \mathbb{R}^n$. On the other hand, $\overline{h}(\lambda, y, \xi) = \infty$ by definition for $(\lambda, y) \in (-\infty, 0) \times \mathbb{R}^n$.

In convex analysis [10, §7–§8], these properties on $[(-\infty, 0) \cup (0, \infty)] \times \mathbb{R}^n$ determine a unique nature extension of $\overline{h}(\cdot, \cdot, \xi)$ to a lower semicontinuous convex function on $\mathbb{R} \times \mathbb{R}^n$. which is obtained by taking limits of $\overline{h}(\lambda, y, \xi)$ as $(\lambda, y) \in (0, \infty) \times \mathbb{R}^n$ approaches the hyperplane $\{0\} \times \mathbb{R}^n$. Our claim is that this extension agrees with the values for $\overline{h}(0, y, \xi)$ coming from its definition (3.21) through (3.18)–(3.19).

In considering what happens to a sequence of function values $\overline{h}(\lambda^\nu, y^\nu, \xi) < \infty$ indexed by $\nu = 1, 2, \ldots$ and associated with a sequence of pairs $(\lambda^\nu, y^\nu) \in (0, \infty) \times \mathbb{R}^n$ that approaches $(0, y)$ for some $y \in \mathbb{R}^n$, we can concentrate through (3.24) on the sequence of values $\lambda^\nu g((\lambda^\nu)^{-1} y^\nu, \xi)$. The liminf over all such sequences will generate through (3.25) the value that $\overline{h}(\cdot, \cdot, \xi)$ should have at $(0, y)$, and this liminf is known from [10, page 67] to be $g0^+(y, \xi)$ for the recession function $g0^+(\cdot, \xi)$ associated with the convex function $g(\cdot, \xi)$. This now is where the boundedness of the level sets of g is important. That boundedness implies that $g0^+(y, \xi)$ equals ∞ when $y \neq 0$, although it equals 0 when $y = 0$.

We have hereby confirmed the equivalence of (3.15) with (3.20), the identification of (3.20) with (3.22), and the assertions about the functions $\overline{h}$ in (3.21). Only the claim about the existence of a solution to (3.20), or (3.21), remains. The existence of a feasible solution with a bPOE value < 1 follows from our assumption (3.5), so we need only be concerned with the objective values < 1. The level sets of the objective in (3.22) for such values are bounded because those of the functions $g(\cdot, \xi)$ are bounded, and through the lower semicontinuity they are also closed. In light of finite dimensionality, the minimum in (3.22) must therefore be attained. □

By virtue of Theorem 1 we can pass to a form of our problem that parallels the one in (3.11) but rigorously incorporates the change of variables undertaken in (3.14). Again λ is interpreted as a first-stage decision variable in a multistage stochastic programming problem, and we are looking at extended policies

$$\widetilde{y}(\cdot) = (\lambda(\cdot), y(\cdot)) \tag{3.28}$$

that are nonanticipative in the sense of $\lambda(\xi)$ being independent of ξ, or in other words, $\widetilde{y}(\cdot)$ belonging to the nonanticipative subspace $\widetilde{\mathcal{N}}$ of the policy space $\widetilde{\mathcal{L}}$ as in (3.11). Where before we had the constraints $x(\xi) \in C(\xi)$, we now have the constraints $(\lambda, y(\xi)) \in K(\xi)$, where $K(\xi)$, given by (3.19) is the the closure of the convex cone $K_0(\xi)$ in (3.27), as revealed in the proof of Theorem 1. To cover this we introduce

$$\mathcal{K} = \left\{\widetilde{y}(\cdot) \in \widetilde{\mathcal{L}} \,\middle|\, \widetilde{y}(\xi) \in K(\xi)\right\} \tag{3.29}$$

for policies (3.28). Then, by also introducing

$$\mathcal{H}(\widetilde{y}(\cdot)) = E_\xi\left[\widetilde{h}(\widetilde{y}(\xi), \xi)\right] \text{for} \widetilde{h}(\widetilde{y}, \xi) = \overline{h}(\lambda, y, \xi) \text{as in (3.21)}, \tag{3.30}$$

we are ready for a problem statement in the expectational mold of stochastic programming, but in terms of extended policies.

Problem Reformulation 4 *Under the nondegeneracy assumptions* (3.4) *and* (3.5), *along with the additional assumptions* (3.16) *and* (3.17), *solving the bPOE problem* (3.1) *is equivalent to solving the problem*

$$\text{minimize } \mathcal{H}(\widetilde{y}(\cdot)) \text{ subject to } \widetilde{y}(\cdot) \in \mathcal{K} \cap \widetilde{N}. \tag{3.31}$$

This problem meets the requirements of the progressive hedging algorithm and is sure to have at least one solution $\widetilde{y}(\cdot)$. Any such solution has its λ-constant positive and translates back to a solution $\widetilde{x}(\cdot)$ to (3.12) *through the change of variables* (3.14), *and that manner it yields a solution $x(\cdot)$ to* (3.1).

The claim about the problem being suitable for progressive hedging is justified by the remark at the end of Sect. 2, in view of the properties established in Theorem 1. Note that because only positive λ actually comes up, the part of the definition of $K(\xi)$ concerned with the recession cone in (3.19), although important in getting $K(\xi)$ to be closed, never has to be encountered numerically (under our assumptions). It can be replaced in practice by $\{(0, 0)\}$.

The corresponding implementation of the progressive hedging algorithm requires knowing not only the nonanticipativity subspace $\widetilde{\mathcal{N}}$, but also its orthogonal complement $\widetilde{\mathcal{M}} = \widetilde{\mathcal{N}}^{\perp}$ in $\widetilde{\mathcal{L}}$. Since we have merely augmented the policies in $\mathcal{N}$ by a first-stage component as a constant function, the complement is given by

$$\widetilde{\mathcal{M}} = \left\{(\theta(\cdot), w(\cdot)) \in \widetilde{\mathcal{L}} \,\middle|\, w(\cdot) \in \mathcal{M},\ E_\xi[\theta(\xi)] = 0\right\}. \tag{3.32}$$

Projections onto $\widetilde{\mathcal{N}}$ and $\widetilde{\mathcal{M}}$, which likewise are essential, can readily identified as well:

$$P_{\widetilde{\mathcal{N}}}(\widetilde{y}(\cdot)) \text{ for } \widetilde{y}(\cdot) = (\lambda(\cdot), y(\cdot)) \text{ equals } \big(E_\xi[\lambda(\xi)], P_{\mathcal{N}}(y(\cdot))\big), \tag{3.33}$$

and then $P_{\widetilde{\mathcal{M}}}(\widetilde{y}(\cdot)) = \widetilde{y}(\cdot) - P_{\widetilde{\mathcal{N}}}(\widetilde{y}(\cdot))$.

Progressive hedging algorithm, bPOE version *Proceed as follows with a parameter $r > 0$ from current elements $(\lambda^\nu, y^\nu(\cdot)) \in \mathbb{R} \times \mathcal{N}$ and $(\theta^\nu(\cdot), w^\nu(\cdot))$ with $w^\nu \in \mathcal{M}$ and $E_\xi[\theta(\xi))] = 0$.*

(a) *Determine $(\widehat{\lambda}^\nu(\cdot), \widehat{y}^\nu(\cdot)) \in \widetilde{\mathcal{L}}$ by calculating $(\widehat{\lambda}^\nu(\xi), \widehat{y}^\nu(\xi))$ for each scenario $\xi \in \Xi$ as the unique solution $(\lambda(\xi), y(\xi)) \in \mathbb{R}^{n+1}$ to an optimization problem that is strongly convex:*

$$(\widehat{\lambda}^\nu(\xi), \widehat{y}^\nu(\xi)) = \underset{(\lambda, y) \in K(\xi)}{\operatorname{argmin}}\ h^\nu(\lambda, y, \xi), \tag{3.34}$$

where h^ν is obtained from the function h in (3.19) by

$$h^\nu(\lambda, y, \xi) = h(\lambda, y, \xi) + \theta^\nu(\xi)\lambda + w^\nu(\xi)\cdot y + \frac{r}{2}|\lambda - \lambda^\nu|^2 + \frac{r}{2}||y - y^\nu(\xi)||^2. \tag{3.35}$$

(b) *Update then to iteration* $\nu + 1$ *by*

$$y^{\nu+1}(\cdot) = P_{\mathcal{N}}[\widehat{y}^{\nu}(\cdot)], \qquad w^{\nu+1}(\cdot) = w^{\nu}(\cdot) + r P_{\mathcal{M}}[\widehat{y}^{\nu}(\cdot)],$$
$$\lambda^{\nu+1} = E_{\xi}[\widehat{\lambda}^{\nu}(\xi)], \qquad \theta^{\nu+1}(\xi) = \theta^{\nu}(\xi) + r[\lambda^{\nu+1} - \widehat{\lambda}^{\nu}(\xi)], \quad (3.36)$$

which for $y^{\nu+1}(\cdot)$ *and* $w^{\nu+1}(\cdot)$ *means taking for each scenario* $\xi \in \Xi$ *and stage* $k = 1, \ldots, N$

$$y_k^{\nu+1}(\xi) = E_{\xi \mid \xi_1,\ldots,\xi_{k-1}} \widehat{y}_k^{\nu}(\xi), \qquad w_k^{\nu+1}(\xi) = w_k^{\nu}(\xi) + r\big[\widehat{y}_k^{\nu}(\xi) - y_k^{\nu+1}(\xi)\big]. \quad (3.37)$$

The preceding developments in levels of reformulation, drawing on Theorem 1 and the additional capabilities of progressive hedging noted in the remark at the end of Sect. 2, furnish the following assurances about this procedure, in summary.

Theorem 2 *Under assumptions* (3.4)–(3.5) *and* (3.16)–(3.7)*, the bPOE version of the progressive hedging algorithm, regardless of where it starts, will converge to some particular triple* $\lambda \in (0, \infty)$*,* $y(\cdot) \in \mathcal{N}$*,* $w(\cdot) \in \mathcal{M}$*, and then* $x(\cdot) = \lambda^{-1} y(\cdot)$ *will solve the bPOE minimization problem* (3.1).

4 Numerical experiments in the one-stage case

When $N = 1$ in our stochastic programming framework, the problem to be solved is much simpler. The nonanticipative policies reduce to constant functions from Ξ to $\mathbb{R}^n$. Although the policy format remains central to progressive hedging, the underlying bPOE problem to be solved can be stated without resorting to the policy-dependent random variables $X_f(x(\cdot))$ in (2.6). It just concerns the vector-dependent random variables

$$f(x, \cdot): \quad \xi \in \Xi \mapsto f(x, \xi) \in \mathbb{R} \text{ for } x \in \mathbb{R}^n \quad (4.1)$$

and seeks to

$$\text{minimize bPOE}_{\tau}(f(x, \cdot)) \text{ over } x \in \bigcap\nolimits_{\xi \in \Xi} C(\xi). \quad (4.2)$$

Besides testing the computational performance of PHA for the bPOE minimization problem (3.1), we test it also for the corresponding CVaR minimization problem (3.6), since the tight “inverse” connection between bPOE and CVaR, explained in Sect. 1, suggests this might be interesting. The CVaR version of PHA that we employ is the one developed in [11]. Numerical experiments weren’t offered in [11], and that’s another reason for providing them here in a PHA comparison between bPOE and CVaR. The one-stage CVaR problem corresponding to (4.2) is

$$\text{minimize CVaR}_{\alpha}(f(x, \cdot)) \text{ over } x \in \bigcap\nolimits_{\xi \in \Xi} C(\xi). \quad (4.3)$$

In both the bPOE and CVaR problems we simplify further by specializing to cost structure in terms of linear functions with random coefficients:

$$f(x,\xi)=\sum_{i=1}^{n} c_i(\xi)x_i - d(\xi) \text{ for } x=(x_1,\dots,x_n)\in\mathbb{R}^n, \tag{4.4}$$

and scenario-independent constraint structure:

$$C(\xi)= \text{ the same } C\subset\mathbb{R}^n \text{ for all } \xi, \tag{4.5}$$

moreover with C given by linear constraints. This puts us in the pattern of the example of specialization offered just ahead of Theorem 1.

We solved several bPOE minimization problems with PHA, and for two of them we solved the corresponding CVaR minimization problem. We demonstrated numerically that the solutions for bPOE and CVaR coincide for values of τ and α coordinated with each other through the inverse relationship, which in constraint mode is reflected by

$$\text{bPOE}_\tau(X)\le 1-\alpha \quad\Longleftrightarrow\quad \text{CVaR}_\alpha(X)\le\tau. \tag{4.6}$$

The selected problems are based on the financial optimization dataset (PROBLEM 1: problem_min_cvar_dev_2p9) considered in the link.[4] Its dataset has 1000 scenarios, corresponding here to potential elements ξ of Ξ. It has $x\in\mathbb{R}^{10}$ restricted by nonnegativity and two scenario-independent linear constraints, which here serve to determine the set C. Specifically

$$(x_1,\dots,x_{10})\in C \quad\Longleftrightarrow\quad x_i\ge 0,\ \ \sum_{i=1}^{10} r_i x_i\ge r,\ \ \sum_{i=1}^{10} x_i=1. \tag{4.7}$$

But the problem's objective, the minimization of CVaR deviation (see the definition of that in [26]), has been altered to suit our present purposes.

In the bPOE case we carried out the change of variables from x to $\lambda^{-1}y$ in the manner explained in Sect. 3. The resulting optimization problem, specialized to $N=1$ from (3.20), then has 11 decision variables, all first-stage, comprised of $\lambda\in\mathbb{R}$ and the components of $y\in\mathbb{R}^{10}$. As in the example preceding Theorem 1, this produced from our constraint set C described by (4.7) the cone K described by

$$(\lambda,y_1,\dots,y_{10})\in K \quad\Longleftrightarrow\quad \lambda\ge 0,\ \ y_i\ge 0,\ \ \sum_{i=1}^{10} r_i y_i - r\lambda\ge 0,\ \ \sum_{i=1}^{10} y_i-\lambda=0. \tag{4.8}$$

The problem further has, in progressive hedging, 11 dual variables comprised of θ and the components of $w\in\mathbb{R}^{10}$.

[4] http://www.ise.ufl.edu/uryasev/research/testproblems/financial_engineering/portfolio-optimization-cvar-vs-st_dev/.

Table 1 Results for the bPOE minimization problem with PHA

# scenarios	Threshold in bPOE	# of interations	Minimum bPOE	Solution precision
10	− 0.035	4	2.6159E− 01	9.40E− 09
20	− 0.03	7	3.2343E− 01	2.80E-09
30	− 0.025	7	2.7641E− 01	3.29E− 09
100	− 0.016499	6	2.0000E− 01	4.94E− 06
1000	− 0.025	7	3.5837E− 01	2.05E− 06

To test performance of PHA for bPOE, we solved 5 instances of the problem with 10, 20, 30, 100 and 1000 scenarios (taken from the original matrix with 1000 scenarios). We used $\lambda^1 = 0$, $y^1 = \overline{0}$, $\theta^1(\xi) = 0$, $w^1(\xi) = \overline{0}$ for the initial iteration of the algorithm, where $\overline{0}$ is the 10-dimension zero vector. The initial values turn out to be quite important for fast convergence of the progressive hedging algorithm. Watson and Woodruff [32] proposed solving the subproblems (2.14) on the first iteration without the quadratic terms, for a class of discrete stochastic optimization problems with the expectational objective. But for the bPOE minimization problems here, that suggested initiation yields a large first step for variables λ, y, and then convergence becomes very slow. More exactly, in iteration 2, values of λ^2 and the components of y^2 are around 400–800 when the quadratic terms are excluded. Without excluding the quadratic terms, they instead have the order of magnitude 10^2. The parameter r was always $r = 1$. The stopping criterion was set to $E_\xi \|\hat{z}^\nu(\xi) - z^{\nu+1}(\xi)\| < 10^{-5}$, where $z = (\lambda, y) \in \mathbb{R}^{11}$ is an 11-dimensional vector.

In our experiments with PHA for CVaR minimization we found instead that it is not so sensitive to dropping the quadratic terms from the initial subproblem in the way of Watson and Woodruff [32].

For benchmarking PHA performance, we have employed the Portfolio Safeguard (PSG) package.[5] PSG has precoded CVaR and bPOE functions, which can be set in analytic format in optimization problems. See the PSG-based case study (data and codes in Text, MATLAB, and R environment) comparing CVaR and bPOE minimization problems at the link.[6]

Solution results are shown in Table 1 for several numbers of scenarios.

- Column 1: the number of scenarios of the linear random function;
- Column 2: the threshold τ in bPOE;
- Column 3: the number of iterations of the algorithm;
- Column 4: minimum value of bPOE obtained with PHA;
- Column5: the difference between the PHA minimum value and the one obtained directly with PSG.

CVaR and bPOE constraints are equivalent via (4.6). To verify correctness of the bPOE optimization algorithm we considered one case with 30 scenarios (see the corresponding row in Table 1). We minimized CVaR with PHA, as proposed in [11].

5 Portfolio Safeguard (PSG), http://www.aorda.com.

6 http://www.ise.ufl.edu/uryasev/research/testproblems/financial_engineering/case-study-cash-matching-with-bpoe-and-cvar-functions/.

Table 2 Results of solving the CVaR minimization problems

# Scenarios	Confidence level CVaR value	# of interations	Minimum in CVaR	Solution precision
30	0.72359	1899	− 2.5000E−02	3.41E−07
100	0.8	1621	− 1.6499E−02	4.39E−05

This CVaR minimization problem contains the same variables and constraints as the bPOE minimization problem. We set the confidence level for CVaR equal to $\alpha = 1 - Minimum\_bPOE = 1 - 0.27641 = 0.72359$ (see, minimum bPOE value in Table 1). According to (3.19) the CVaR minimum value should be equal to the threshold in bPOE, i.e., -0.025 (the threshold is specified in Table 1). This minimum CVaR value -0.025 was obtained with PHA for superquantile minimization, as further reported in Table 2. This fact confirms that both the bPOE and the CVaR optimization problems were solved correctly.

Further we solved the CVaR minimization problem with confidence level $\alpha = 0.8$ for the case with 100 scenarios. Afterward the minimum CVaR value, which is equal to -0.016499 (see Table 2) was used as the threshold τ in the bPOE minimization problem. According to (4.6) the minimum bPOE value should be equal to $1-\alpha = 0.2$. We obtained exactly this minimum value of bPOE by PHA. We have already mentioned Table 2 twice. Here is a description of the columns in Table 2.

- Column 1: the number of scenarios of the linear random function;
- Column 2: the confidence level α in CVaR;
- Column 3: the number of iterations of PHA;
- Column 4: the minimum value of CVaR obtained with PHA;
- Column 5: the difference between the PHA solution and solution obtained directly with PSG.

We have indicated that for benchmarking the PHA solutions by a different means of computation, made feasible by the simplicity of the problems being tested, we used the PSG package. Here is the information on how that PSG optimization of CVaR and bPOE was carried out. We utilized the Partial Moment (PM) function in the PSG package, which is defined for a random variable X by

$$\mathrm{PM}_\tau(X) = E[\max\{0, X - \tau\}], \tag{4.9}$$

applying it to the x-dependent random variables in (4.1) to have

$$\mathrm{PM}_\tau(f(x,\cdot)) = E[\max\{0, f(x,\cdot) - \tau\}]. \tag{4.10}$$

In the case of benchmarking CVaR we drew on the formula (3.7) that was developed for it in [23,24], getting its application to $X = f(x,\cdot)$ to take the form

$$\mathrm{CVaR}_\alpha(f(x,\cdot)) = \min_{\tau \in R}\left\{\tau + \frac{1}{1-\alpha}\mathrm{PM}_0\left(f(x,\cdot) - \tau\right)\right\}. \tag{4.11}$$

Therefore,

$$\min_{x\in C} \mathrm{CVaR}_\alpha(f(x,\cdot)) = \min_{(\tau,x)\in\mathbb{R}\times C} \left\{\tau + \frac{1}{1-\alpha}\mathrm{PM}_0(f(x,\cdot)-\tau)\right\}, \tag{4.12}$$

where $f(x,\xi)-\tau$ is a linear function jointly in the variables x and τ.

In the case of benchmarking bPOE, we are in the realm of the one-stage version of (3.20) with the function h given by (3.19), but because of the linearity of f in (4.5), this reduces to

$$h(\lambda, y, \xi) = \mathrm{PM}_0(f(y,\cdot) - \lambda\tau + 1) \text{ if } \lambda \geq 0, \text{ but } = \infty \text{ otherwise}. \tag{4.13}$$

Consequently,

$$\min_{x\in C} \mathrm{bPOE}_\tau(f(x,\cdot)) = \min_{(\lambda,y)\in K} \mathrm{PM}_{-1}\left((c_0(\xi)-\tau)\lambda + \sum_{i=1}^{n} c_i(\xi)y_i\right), \tag{4.14}$$

where K is a polyhedral convex cone described in (4.8).

In this way the objective in both optimization problems (4.2), (4.3), under the simplification in (4.4), is expressible through the PM function. PM minimization for a linear function with random coefficients can be reduced to convex and linear programming. PSG algorithms for minimizing CVaR and bPOE were thereby obtained for the benchmarking in our study.

Acknowledgements This research was sponsored for both authors by DARPA EQUiPS Grant SNL 014150709. The authors are grateful also to Dr. Viktor Kuzmeno for help with conducting the numerical case study.

References

1. Davis, R.A., Uryasev, S.: Analysis of tropical storm damage using buffered probability of exceedance. Nat. Hazards (2016). https://doi.org/10.1007/s11069-016-2324-y
2. Kouri, D.P., Shapiro, A.: Optimization of PDEs with uncertain inputs. In: Antil, H., Kouri, D., Lacasse, M.D., Ridzal, D. (eds.) Frontiers in PDE-Constrained Optimization. The IMA Volumes in Mathematics and Its Applications, vol. 163. Springer, New York, NY (2018)
3. Kouri, D.P.: Higher-moment buffered probability. Optm. Lett. **13**, 657–672 (2019)
4. Mafusalov, A., Uryasev, S.: Buffered probability of exceedance: mathematical properties and optimization. SIAM. J. Optim. **28**, 1077–1103 (2018)
5. Mafusalov, A., Shapiro, A., Uryasev, S.: Estimation and asymptotics for buffered probability of exceedance. Eur. J. Oper. Res. **270**, 826–836 (2018)
6. Norton, M., Uryasev, S.: Maximization of AUC and Buffered AUC in binary classification. Math. Program. (2018). https://doi.org/10.1007/s10107-018-1312-2
7. Norton, M., Mafusalov, A., Uryasev, S.: Soft margin support vector classification as buffered probability minimization. J. Mach. Learn. Res. **18**, 1–43 (2017)
8. Norton, M., Mafusalov, A., Uryasev, S.: Cardinality of upper average and application to network optimization. SIAM J. Optim. **28**(2), 1726–1750 (2018)
9. Norton, M., Khokhlov, V., Uryasev, S.: Calculating CVaR and bPOE for common probability distributions with application to portfolio optimization and density estimation. Ann. Oper. Res. arXiv:1811.11301 (submitted for publication)

10. Rockafellar, R.T.: Convex Analysis. Princeton University Press, Princeton (1970)
11. Rockafellar, R.T.: Solving stochastic programming problems with risk measures by progessive hedging. Set-valued Variational Anal. **26**, 759–768 (2017)
12. Rockafellar, R.T., Royset, J.O.: On buffered failure probability in design and optimization of structures. J. Reliab. Eng. Syst. Saf. **95**, 499–510 (2010)
13. Rockafellar, R.T., Royset, J.O.: Superquantiles and their applications to risk, random variables, and regression. In: Tutorials in Operations Research, pp. 151–167. INFORMS (2013)
14. Rockafellar, R.T., Royset, J.O., Miranda, S.J.: Superquantile regression with applications to buffered reliability, uncertainty quantification and conditional value-at-risk. Eur. J. Oper. Res. **234**, 140–154 (2014)
15. Rockafellar, R.T., Royset, J.O.: Engineering decisions under risk averseness. J. Risk Uncertain. Eng. Syst. Part A Civ. Eng. **1**(2), 04015005 (2015). https://doi.org/10.1061/AJRUA6.0000816
16. Rockafellar, R.T., Royset, J.O.: Risk measures in engineering design under uncertainty. In: Proceedings of the 12th International Conference on Applications of Statistics and Probability in Civil Engineering (ICASP), Vancouver (2015). https://doi.org/10.14288/1.0076159
17. Rockafellar, R.T., Royset, J.O., Harajli, M.M.: Importance sampling in the evaluation and optimization of buffered probability of failure. In: Proceedings of the 12th International Conference on Applications of Statistics and Probability in Civil Engineering (ICASP), Vancouver (2015). https://doi.org/10.14288/1.0076214
18. Rockafellar, R.T., Royset, J.O.: Measures of residual risk with connections to regression, risk tracking, surrogate models and ambiguity. SIAM J. Optim. **25**, 1179–1208 (2015)
19. Rockafellar, R.T., Royset, J.O.: Superquantile/CVaR risk measures: second-order theory. Ann. Oper. Res. **262**, 3–29 (2018)
20. Rockafellar, R.T., Royset, J.O.: Random variables, monotone relations and convex analysis. Math. Program. B **128**, 297–331 (2014)
21. Rockafellar, R.T., Sun, J.: Soving monotone stochastic variational inequalities and complementarity problems by progressive hedging. Math. Program. B **174**, 453–471 (2018)
22. Rockafellar, R.T., Sun, J.: Solving Lagrangian variational inequalities with applications to stochastic programming. Math. Program. (2020). https://doi.org/10.1007/s10107-019-01458-0
23. Rockafellar, R.T., Uryasev, S.: Optimization of conditional value-at-risk. J. Risk **2**, 21–43 (2000)
24. Rockafellar, R.T., Uryasev, S.: Conditional value-at-risk for general loss distributions. J. Bank. Finance **26**, 1443–1471 (2002)
25. Rockafellar, R.T., Uryasev, S.: The fundamental risk quadrangle in risk management, optimization and statistical esimation. Surv. Oper. Res. Manag. Sci. **18**, 33–53 (2013)
26. Rockafellar, R.T., Uryasev, S., Zabarankin, M.: Generalized deviations in risk analysis. Finance Stoch. **10**, 51–74 (2006)
27. Rockafellar, R.T., Wets, R.J.-B.: Nonanticipativity and $\mathcal{L}^1$-martingales in stochastic optimization problems. Stoch. Syst. Model. Identif. Optim. Math. Programm. Study **6**, 170–187 (1976)
28. Rockafellar, R.T., Wets, R.J.-B.: Scenarios and policy aggregation in optimization under uncertainty. Math. Oper. Res. **16**, 119–147 (1991)
29. Rockafellar, R.T., Wets, R.J.-B.: Stochastic variational inequalities: single-stage to multistage. Math. Program. B **165**, 291–330 (2017)
30. Shang, D., Kuzmenko, V., Uryasev, S.: Cash flow matching with risks controlled by buffered probability of exceedance and conditional value-at-risk. Ann. Oper. Res. **260**, 501–514 (2016)
31. Uryasev, S.: Buffered probability of exceedance and buffered service level: definitions and properties, Research Report 2014-3. ISE Department, University of Florida (2014)
32. Watson, J.-P., Woodruff, D.L.: Progressive hedging innovations for a class of stochastic mixed-integer resource allocation problems. Comput. Manag. Sci. **8**, 355–370 (2010)
33. Zhitlukhin, M.: Monotone Sharpe ratios and related measures of investment performance (2018). arXiv:1809.10193