SOLVING STOCHASTIC PROGRAMMING PROBLEMS WITH RISK MEASURES BY PROGRESSIVE HEDGING

R. Tyrrell Rockafellar[1]

Abstract

The progressive hedging algorithm for stochastic programming problems in single or multiple stages is a decomposition method which, in each iteration, solves a separate subproblem with modified costs for each scenario. The decomposition exploits the separability of objective functions formulated in terms of expected costs, but nowadays expected costs are not the only objectives of interest. Minimization of risk measures for cost, such as conditional value-at-risk, can be important as well, but their lack of separability presents a hurdle. Here it is shown how the progressive hedging algorithm can nonetheless be applied to solve many such problems through the introduction of additional variables which, like the given decision variables, get updated through aggregation of the independent computations for the various scenarios.

Keywords: *stochastic programming progressive hedging algorithm, problem decomposition, risk measures, conditional value-at-risk.*

Version of June 26, 2017

[1]University of Washington, Department of Mathematics, Box 354350, Seattle, WA 98195-4350;
E-mail: *rtr@uw.edu*, URL: www.math.washington.edu/∼rtr/mypage.html
This reseach was sponsored by DARPA EQUiPS grant SNL 014150709

1 Introduction

In the early days of stochastic programming one of the challenges was how to address the common practice of running multistage scenarios and solving for each of them an optimization problem with perfect hindsight. There was no clear idea of how to combine the results from such separate problems to obtain a decision policy that could properly hedge against the underlying uncertainty. The progressive hedging algorithm of Rockafellar and Wets [4] (1991) was developed to fill that gap, in the presence of convexity. It revolved around a problem formulation in which recourse decisions could respond to successive inputs of information in time and the expected value of an associated cost expression was to be minimized. Users could still use the tools they had for solving scenario problems with perfect hindsight, but the costs in those problems were modified in two ways: an additional cost term incorporating an estimated "price of information" and a proximal term aimed at discouraging the next policy proposal from straying too far from the previous one. The results for the separate scenarios were then aggregated by a projection mapping utilizing conditional expectations. This has been a helpful tool in many practical situations; see for instance Watson and Woodruff [5].

Since the introduction of the progressive hedging algorithm the applications of stochastic programming have expanded to include situations where not just a expected cost is minimized as a *risk-neutral objective*, but instead some *risk-averse* assessment of cost. Especially of interest in this respect have been the CVaR risk measures (conditional value-at-risk) in the works of Rockafellar and Uryasev [1, 2]. However, such risk measures appear incompatible with the progressive hedging algorithm because they lack the scenario-wise separability of an expected value on which that scheme relies.

Here we demonstrate in Section 3 that the progressive hedging algorithm can be applied despite that seeming obstacle through a restatement which makes use of a formula in [1, 2]. The extra parameter in that formula becomes an additional first-stage variable that, in the course of the iterations, has different values for the different scenarios, and yet through aggregation ultimately converges to the single value that is needed and corresponds to the associated level of VaR (value-at-risk) of the objective. In Section 5 we go on from that basic CVaR case to show that the same ideas lead to extensions to a much larger class of risk measures, including for instance, in approximation, all spectral risk measures.

2 Progressive hedging background

In the circumstances we are targeting, with $N \geq 1$ decision stages k, the choice of a vector $x_k \in I\!R^{n_k}$ is followed by the revelation of an information element ξ_k in some set Ξ_k, which could be the observation of a vector random variable or something else. It is assumed for simplicity that there are only finitely many scenarios $\xi = (\xi_1, \ldots, \xi_N)$ generated in this way, each having a probability $p(\xi) > 0$ that is independent of any decisions taken. The finite subset of $\Xi_1 \times \cdots \times \Xi_N$ consisting of these is denoted by Ξ. (Alternatively this structure could be articulated equivalently in terms of a "scenario tree.")

In reacting to information provided by a scenario $\xi = (\xi_1, \ldots, \xi_N)$, the decision $x_k(\xi)$ at stage k must be *nonanticipative*:

$$x_k(\xi) \text{ depends only on } (\xi_1, \ldots, \xi_{k-1}), \text{ not on } (\xi_k, \ldots, \xi_N). \tag{2.1}$$

In denoting by $x(\cdot)$ a function that assigns to each $\xi \in \Xi$ a vector

$$x(\xi) = (x_1(\xi), x_2(\xi), \ldots, x_N(\xi)) \in I\!R^{n_1} \times I\!R^{n_2} \times \cdots \times I\!R^{n_N} = I\!R^n \tag{2.2}$$

as a "decision policy" in general, the special pattern associated with $x(\cdot)$ being nonanticipative is thus that it can be expressed by

$$x(\xi) = (x_1, x_2(\xi_1), x_3(\xi_1, \xi_2), \ldots, x_N(\xi_1, \ldots, \xi_{N-1})). \tag{2.3}$$

Additional constraints on decisions are imposed by requiring

$$x(\xi) \in C(\xi) \subset I\!R^n \text{ for every } \xi \in \Xi, \tag{2.4}$$

where $C(\xi)$ is a nonempty closed set which we take to be *convex*. Of course $C(\xi)$ might be described by a system of constraints which furthermore could reflect a stage-by-stage structure, but we don't have to get into that degree of detail in what follows.

An alternating sequence of decisions and observations $x_1, \xi_1, x_2, \xi_2, \ldots, x_N, \xi_N$ results in a cost, which we denote by $f(x, \xi)$ and assume here to be finite and convex with respect to $x = (x_1, \ldots, x_N) \in I\!R^n$ for each $\xi = (\xi_1, \ldots, \xi_N) \in \Xi$. There could be additive structure to this cost over the stages, but again that needn't be our concern. Our focus is on the fact that each choice of $x(\cdot)$ yields

$$f(x(\cdot), \cdot) : \xi \to f(x(\xi), \xi), \text{ a function from } \Xi, \text{ as a probability space, to } I\!R, \tag{2.5}$$

which we can regard as a *random variable*. Classically the expected value of this random variable,

$$E[f(x(\cdot), \cdot)] = E_\xi[f(x(\xi), \xi)] = \sum\nolimits_{\xi \in \Xi} p(\xi) f(x(\xi), \xi), \tag{2.6}$$

is adopted as the cost objective $c(x(\cdot))$ to be minimized with respect to the choice of $x(\cdot)$, subject to $x(\cdot)$ being nonanticipative and satisfying the constraints (2.4). That is the framework of stochastic programming for which the progressive hedging algorithm was developed in [4] and the subject of our review of that method in this section.

The nonanticipative mappings $x(\cdot)$, having the pattern (2.3), form a linear space which we denote by $\mathcal{N}$. This space has been adequate so far, but the progressive hedging algorithm relies on reformulating the stochastic programming problem in a larger space than $\mathcal{N}$. That strategy enables nonanticipativity to be articulated as a linear constraint to which a Lagrange multiplier can be attached. The multiplier not only has a crucial role in the algorithm; it also furnishes "shadow prices" for information about future observations.

The larger space, which we denote by $\mathcal{L}$, consists of *all* mappings $x(\cdot) : \Xi \to I\!R^n$ that assign to each scenario ξ a response $x(\xi)$ as in (2.2), regardless of nonanticipativity. It has $\mathcal{N}$ as its *nonanticipative subspace*. We furnish $\mathcal{L}$ with the expectational inner product

$$\langle x(\cdot), w(\cdot) \rangle = E_\xi \Big[\sum\nolimits_{k=1}^N x_k(\xi) \cdot w_k(\xi) \Big] = \sum\nolimits_{\xi \in \Xi} p(\xi) \sum\nolimits_{k=1}^N x_k(\xi) \cdot w_k(\xi) \text{ for } x(\cdot) \in \mathcal{L},\ w(\cdot) \in \mathcal{L}, \tag{2.7}$$

where $x_k(\xi) \cdot w_k(\xi)$ is the usual inner product in $I\!R^{n_k}$, so as to make it into a Hilbert space with norm

$$||x(\cdot)|| = \sqrt{\langle x(\xi), x(\xi) \rangle} = \sqrt{E_\xi[\, ||x(\xi)||^2 \,]} \tag{2.8}$$

The orthogonal projection $P_{\mathcal{N}}$ from $\mathcal{L}$ onto $\mathcal{N}$ with respect to this norm is characterized by a conditional expectation formula,

$$x(\cdot) = P_{\mathcal{N}}[\hat{x}(\cdot)] \iff x_k(\xi) = E_{\xi | \xi_1, \ldots, \xi_{k-1}} \hat{x}_k(\xi) \, \forall k. \tag{2.9}$$

This is easy to execute and will have a key role. Also important in this picture is the subspace of $\mathcal{L}$ that is the orthogonal complement of $\mathcal{N}$ under the inner product (2.7), namely

$$\mathcal{M} = \mathcal{N}^\perp = \{\, w(\cdot) \in \mathcal{L} \,|\, \langle x(\cdot), w(\cdot) \rangle = 0, \, \forall x(\cdot) \in \mathcal{N} \}. \tag{2.10}$$

It too has a direct description in terms of conditional expectations:

$$w(\cdot) \in \mathcal{M} \iff E_{\xi | \xi_1, \ldots, \xi_{k-1}} w_k(\xi) = 0, \, \forall k, \tag{2.11}$$

and of course $\mathcal{N} = \mathcal{M}^{\perp}$ as well. The projection mapping onto $\mathcal{M}$ is $P_{\mathcal{M}} = I - P_{\mathcal{N}}$, so that

$$x(\cdot) \in \mathcal{N} \iff P_{\mathcal{M}}[x(\cdot)] = 0 \iff x(\cdot) - P_{\mathcal{N}}[x(\cdot)] = 0. \tag{2.12}$$

The stochastic programming problem under consideration can be identified in the enlarged setting with the minimization of the expectational objective (2.6), as a convex function on all of $\mathcal{L}$, subject to the linear constraint (2.12), characterizing the nonanticipativity subspace $\mathcal{N}$, and the convex constraints in (2.4), which describe a nonempty closed convex subset $\mathcal{C}$ of $\mathcal{L}$:

$$\mathcal{C} = \{\, x(\cdot) \in \mathcal{L} \,|\, x(\xi) \in C(\xi),\ \forall \xi \}. \tag{2.13}$$

The corresponding Lagrangian function has the form

$$L(x(\cdot), w(\cdot)) = E[f(x(\cdot), \cdot)] + \langle w(\cdot), P_{\mathcal{M}}[x(\cdot)] \rangle, \tag{2.14}$$

where the term $\langle w(\cdot), P_{\mathcal{M}}[x(\cdot)] \rangle = \langle x(\cdot), P_{\mathcal{M}}[w(\cdot)] \rangle$ depends only on the projection of $w(\cdot)$ in $\mathcal{M}$ and thus can be simplified to $\langle x(\cdot), w(\cdot) \rangle$ without loss of generality by supposing $w(\cdot) \in \mathcal{M}$. This explains the sense in which $w(\cdot)$ can be interpreted as a vector of "shadow prices" for nonanticipativity. It also indicates that we should be looking not only for a solution $x^*(\cdot)$ to the minimization problem but also at the same time for an element $w^*(\cdot) \in \mathcal{M}$ that furnishes along with $x^*(\cdot)$ a saddle point of the Lagrangian in (2.14):

$$x^*(\cdot) \in \operatorname*{argmin}_{x(\cdot) \in \mathcal{C}} L(x(\cdot), w^*(\cdot)), \qquad w^*(\cdot) \in \operatorname*{argmax}_{w(\cdot) \in \mathcal{M}} L(x^*(\cdot), w(\cdot)). \tag{2.15}$$

Progressive Hedging Algorithm 1 (risk-neutral case)**.** *Iterations $\nu = 1, 2, \ldots$ with a parameter $r > 0$ proceed as follows from current elements $x^\nu(\cdot) \in \mathcal{N}$ and $w^\nu(\cdot) \in \mathcal{M}$.*

(a) *Get $\hat{x}^\nu(\cdot) \in \mathcal{L}$ by solving for each scenario $\xi \in \Xi$ a strongly convex programming problem in the variable $x(\xi) \in \mathbb{R}^n$ that has a unique solution $x(\xi) = \hat{x}^\nu(\xi)$:*

$$\hat{x}^\nu(\xi) = \operatorname*{argmin}_{x(\xi) \in C(\xi)} \Big\{ f(x(\xi), \xi) + w^\nu(\xi) \cdot x(\xi) + \frac{r}{2} ||x(\xi) - x^\nu(\xi)||^2 \Big\}. \tag{2.16}$$

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

$$x^{\nu+1}(\cdot) = P_{\mathcal{N}}[\hat{x}^\nu(\cdot)], \qquad w^{\nu+1}(\cdot) = w^\nu(\cdot) + r P_{\mathcal{M}}[\hat{x}^\nu(\cdot)] = w^\nu(\cdot) + r[\hat{x}^\nu(\cdot) - x^{\nu+1}(\cdot)]. \tag{2.17}$$

The projection in (2.17) is easy to execute because it just amounts to taking $x_k^{\nu+1}(\xi)$ equal to the conditional expectation of $\hat{x}_k^\nu(\xi)$ given $\xi_1, \ldots, \xi_{k-1}$.

Convergence Theorem 1 (risk-neutral case [4])**.** *Under the assumption that the stochastic programming problem has a solution, the sequence of pairs $(x^\nu(\cdot), w^\nu(\cdot))$ generated by the progressive hedging algorithm above from any starting pair is sure to converge to one such solution $x^*(\cdot)$ and an associated $w^*(\cdot)$ furnishing a saddle point as in (2.15). In this convergence the distance expression*

$$||x^\nu(\cdot) - x^*(\cdot)||^2 + \frac{1}{r^2} ||w^\nu(\cdot) - w^*(\cdot)||^2 \tag{2.18}$$

will form a decreasing sequence.

A prime virtue of the progressive hedging algorithm is that it solves scenario subproblems in parallel. This kind of decomposition stems from the fact that in each iteration the expression being minimized subject only to having $x(\xi) \in C(\xi)$ for all ξ to obtain $\hat{x}^\nu(\cdot)$ is the expectation

$$\sum\nolimits_{\xi \in \Xi} p(\xi) \Big[f(x(\xi), \xi) + w^\nu(\xi) \cdot x(\xi) + \frac{1}{2} ||x(\xi) - x^\nu(\xi)||^2 \Big], \tag{2.19}$$

which breaks down into a separate minimization for each scenario ξ with respect $x(\xi) \in C(\xi)$. That could give the impression that other formulations of stochastic programming in which the objective term lacks such separability might not be suitable for progressive hedging, but we'll show there are ways around that.

3 Extension of progressive hedging to CVaR objectives

The classical objective for minimization in stochastic programming is formed by taking the expected value of the random variable $f(x(\cdot), \cdot)$ in (2.5), but this is only one of many possibilities. Much more broadly one can consider a cost objective of the form

$$c(x(\cdot)) = \mathcal{R}(f(x(\cdot), \cdot)) \text{ for some "risk measure" } \mathcal{R}. \tag{3.1}$$

The theory of risk measures, as articulated for scalar-valued random variables with a "cost orientation," is broad and highly developed. It is laid out broadly, for instance, in Rockafellar-Uryasev [3]. This section focuses on CVaR risk measures $\mathcal{R}$, but the ideas that are developed enable the extension of the progressive hedging to be carried out for a far larger class of risk-averse objectives, as will be explained in Section 5.

In terms of the cumulative distribution function F_X of a random variable X, a popular approach to the risk $\mathcal{R}(X)$ in X is to look at the so-called *value-at-risk* at a level $\alpha \in (0, 1)$,

$$\text{VaR}_\alpha(X) = \min\{\, z \,|\, F_X(z) \geq \alpha\}. \tag{3.2}$$

This has the attractive property that

$$\text{VaR}_\alpha(X) \leq a \iff \text{prob}\{X \leq a\} \geq \alpha. \tag{3.3}$$

In taking $\mathcal{R} = \text{VaR}_\alpha$ in (3.1), the objective in the stochastic programming problem would be to choose $x(\cdot)$ so as to achieve the lowest level a such that the probability of the cost coming out higher than a would be less than $1 - \alpha$. This could make sense as a safeguard against cost overruns. However, value-at-risk has a huge mathematical drawback. An objective function based on it will generally (aside from special circumstances) lack convexity and even suffer from discontinuities, in particular in handling random variables with finitely many outcomes.

An alternative to value-at-risk is *conditional value-at-risk*, which again depends on a choice of $\alpha \in (0, 1)$ and is described by

$$\text{CVaR}_\alpha(X) = \quad [\text{ expectation of } X \text{ in its } \alpha\text{-tail }]. \tag{3.4}$$

The "tail" in question is the interval $[\text{VaR}_\alpha(X), \infty)$ as long as no atom in the distribution of X (corresponding to a jump in F_X) resides at the point $z = \text{VaR}_\alpha(X)$, but otherwise involves "splitting that atom" in a precise manner explained in [2]. Alternatively, regardless of the presence of atoms, the formula holds that

$$\text{CVaR}_\alpha(X) = \frac{1}{1-\alpha} \int_\alpha^1 \text{VaR}_\beta(X) d\beta. \tag{3.5}$$

The rule analogous to (3.3) for conditional value-at-risk is that

$$\text{CVaR}_\alpha(X) \leq a \iff X \text{ is } \leq a \text{ on average in its } \alpha\text{-tail}. \tag{3.6}$$

This is a stronger prescription than (3.3), which merely limits violations of $X \leq a$ to the "worst $(1 - \alpha)\%$ of outcomes" of X. In (3.6), even among those outcomes, the desired upper bound must

hold as an expectation. Taking $\mathcal{R} = \mathrm{CVaR}_\alpha$ in (3.1) would mean adopting this more cautious approach to cost overruns.

The powerful advantage of $\mathrm{CVaR}_\alpha(X)$ over $\mathrm{VaR}_\alpha(X)$ as a measure of the risk in X is that it does possess continuity and even convexity. In (3.1) with $\mathcal{R} = \mathrm{CVaR}_\alpha$, we get $c(x(\cdot))$ as a *convex* function of $x(\cdot) \in \mathcal{L}$ (under our assumption that $f(\cdot, \xi)$ is a convex function on $\mathbb{R}^n$ for each $\xi \in \Xi$). This convexity is crucial to any hope of using Lagrange multipliers $w(\cdot)$ coming from $\mathcal{M}$ in a scheme of decomposition with respect to nonanticipativity. In replacing $E[f(x(\cdot), \cdot)]$ by $\mathrm{CVaR}_\alpha(f(\cdot), \cdot)$ in our stochastic programming problem and trying directly to apply the progressive hedging algorithm we find though that the separable expression in (2.19) to be minimized to determine $\hat{x}^\nu(\cdot)$ is replaced by

$$\mathrm{CVaR}_\alpha(f(x(\xi), \xi)) + \sum\nolimits_{\xi \in \Xi} p(\xi)\Big[w^\nu(\xi) \cdot x(\xi) + \frac{r}{2}||x(\xi) - x^\nu(\xi)||^2\Big]. \tag{3.7}$$

This expression lacks separability and doesn't lead to a break down into a separate problem with respect to $x(\xi) \in C(\xi)$ for each ξ.

Fortunately there is another characterization of conditional value-at-risk, originating in [1, 2], which can lead us past this obstacle. It yields $\mathrm{CVaR}_\alpha(X)$ and $\mathrm{VaR}_\alpha(X)$ at the same time:

$$\begin{array}{l} \mathrm{CVaR}_\alpha(X) = \min_{y \in \boldsymbol{R}} \Big\{ y + \frac{1}{1-\alpha} E[\max\{0, X - y\}] \Big\}, \\ \mathrm{VaR}_\alpha(X) = \mathrm{argmin}_{y \in \boldsymbol{R}} \Big\{ y + \frac{1}{1-\alpha} E[\max\{0, X - y\}] \Big\} \quad \text{“essentially,”} \end{array} \tag{3.8}$$

where the caveat in the second line refers to an adjustment that has to be made in the situation in (3.2) where there is more than one z satisfying $F_X(z) = \alpha$. Those values of z then form a closed interval having $\mathrm{VaR}_\alpha(X)$ as its left endpoint, and the “argmin” is that entire interval.

In drawing on (3.8) we can rewrite the problem with respect to (3.7) for determining $\hat{x}^\nu(\cdot)$ as the problem

$$\begin{array}{l} \text{minimize with respect to } y \in \mathbb{R} \text{ and } x(\xi) \in C(\xi) \text{ for all } \xi \text{ the expression:} \\ \sum_{\xi \in \Xi} p(\xi)\Big[y + \frac{1}{1-\alpha} \max\{0, f(x(\xi), \xi) - y\} + w^\nu(\xi) \cdot x(\xi) + \frac{r}{2}||x(\xi) - x^\nu(\xi)||^2\Big] \end{array} \quad . \tag{3.9}$$

But although this looks “more separable,” trouble remains over the requirement of having the same y for every scenario ξ.

Our strategy to get over this difficulty is to interpret y as an additional *first-stage decision variable* which must end up as a constant due to *nonanticipativity*. The initial component $x_1(\xi)$ of $x(\xi)$ is replaced then by a pair $(y(\xi), x_1(\xi))$ with the spaces $\mathcal{L}$ and $\mathcal{N}$ enlarged accordingly to spaces $\overline{\mathcal{L}}$ and $\overline{\mathcal{N}}$, where $\overline{\mathcal{L}} = \mathcal{L}_0 \times \mathcal{L}$ for

$$\mathcal{L}_0 = \{\, y(\cdot) : \Xi \to \mathbb{R}\}$$

and $\overline{\mathcal{N}} = \mathbb{R} \times \mathcal{N}$. Likewise the multiplier space $\mathcal{M}$ gets enlarged to $\overline{\mathcal{M}}$ through replacing the initial component $w_1(\xi)$ of $w(\xi)$ by a pair $(u(\xi), w_1(\xi)) \in \overline{\mathcal{L}} = \mathcal{L}_0 \times \mathcal{L}$ having not only $E_\xi[w_1(\xi)] = 0$ but also $E_\xi[u(\xi)] = 0$. The convex sets $C(\xi)$ turn into the convex sets $\overline{C}(\xi) = \mathbb{R} \times C(\xi)$ in accommodating the augmented vectors $\overline{x}(\xi) = (y(\xi), x_1(\xi), x_2(\xi), \ldots, \ldots, x_N(\xi))$. The convex cost functions $f(x, \xi) = f(x_1, \ldots, x_N, \xi)$ on $\mathbb{R}^n \times \Xi$ are replaced by the convex functions

$$\overline{f}_\alpha(y, x_1, \ldots, x_N, \xi) = y + \frac{1}{1-\alpha} \max\{0, f(x_1, \ldots, x_N, \xi) - y\} \tag{3.10}$$

We are back then in the earlier pattern of stochastic programming in which the objective is an expectation, namely $E_\xi[f_\alpha(y(\xi), x_1(\xi), \ldots, x_N(\xi), \xi)]$. This is to be minimized over the nonanticipativity subspace $\overline{\mathcal{N}}$ subject to the extended vectors $\overline{x}(\xi)$ belonging to the sets $\overline{C}(\xi)$.

The progressive hedging algorithm can be applied straightforwardly in this framework. The resulting procedure can then be recast in terms of the variables y and $u(\xi)$ along with the original $x(\xi)$ and $w(\xi)$. In place of the Lagrangian in (2.14) we have

$$\overline{L}(y(\cdot), x(\cdot), u(\cdot), w(\cdot)) = E_\xi[\overline{f}_\alpha(y(\xi), x(\xi), \xi) + u(\xi)y(\xi) + w(\xi)\cdot x(\xi)] \tag{3.11}$$

and are aiming at a solution pair $(y^*, x^*(\cdot))$ with associated multiplier pair $(u^*(\cdot), w^*(\cdot))$ characterized by the saddle point condition

$$\begin{array}{l} (y^*, x^*(\cdot)) \in \underset{y(\cdot)\in\mathcal{L}_0,\, x(\cdot)\in\mathcal{C}}{\operatorname{argmin}} \overline{L}(y(\cdot), x(\cdot), u^*(\cdot), w^*(\cdot)), \\ (u^*(\cdot), w^*(\cdot)) \in \underset{\substack{u(\cdot)\in\mathcal{L}_0,\, E[u(\cdot)]=0 \\ w(\cdot)\in\mathcal{M}}}{\operatorname{argmax}} \overline{L}(y^*, x^*(\cdot), u(\cdot), w(\cdot)). \end{array} \tag{3.12}$$

This comes out as follows.

Progressive Hedging Algorithm 2 (risk-averse case for CVaR_α)**.** *Iterations $\nu = 1, 2, \ldots$ with a parameter value $r > 0$ proceed as follows from current elements $x^\nu(\cdot) \in \mathcal{N}$, $w^\nu(\cdot) \in \mathcal{M}$, $y^\nu \in \mathbb{R}$, and $u^\nu(\cdot) : \Xi \to \mathbb{R}$ with $E_\xi[u(\xi)] = 0$.*

(a) *Get $\hat{x}^\nu(\cdot) \in \mathcal{L}$ and $\hat{y}^\nu(\cdot) : \Xi \to \mathbb{R}$ by solving for each scenario $\xi \in \Xi$ a strongly convex programming problem in the variables $x(\xi) \in \mathbb{R}^n$ and $y(\xi) \in \mathbb{R}$ that has a unique solution $(y(\xi), x(\xi)) = (\hat{y}^\nu(\xi), \hat{x}^\nu(\xi))$:*

$$\begin{array}{l} (\hat{y}^\nu(\xi), \hat{x}^\nu(\xi)) = \underset{y(\xi)\in\boldsymbol{R},\, x(\xi)\in C(\xi)}{\operatorname{argmin}} \Big\{ y(\xi) + \frac{1}{1-\alpha}\max\{0, f(x(\xi),\xi) - y(\xi)\} \\ \qquad\qquad + u^\nu(\xi)y(\xi) + \frac{r}{2}|y(\xi) - y^\nu|^2 + w^\nu(\xi)\cdot x(\xi) + \frac{r}{2}||x(\xi) - x^\nu(\xi)||^2 \Big\}. \end{array} \tag{3.13}$$

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

$$\begin{array}{ll} x^{\nu+1}(\cdot) = P_{\mathcal{N}}[\hat{x}^\nu(\cdot)], & w^{\nu+1}(\cdot) = w^\nu(\cdot) + r[\hat{x}^\nu(\cdot) - x^{\nu+1}(\cdot)], \\ y^{\nu+1} = E_\xi[\hat{y}^\nu(\xi)], & u^{\nu+1}(\cdot) = u^\nu(\cdot) + r[\hat{y}^\nu(\cdot) - y^{\nu+1}]. \end{array} \tag{3.14}$$

Convergence Theorem 2 (risk-averse case for CVaR_α)**.** *Under the assumption that the stochastic programming problem posed with $\mathcal{R} = \mathrm{CVaR}_\alpha$ in (3.1) has a solution, the sequence of elements $(y^\nu, x^\nu(\cdot), u^\nu(\cdot), w^\nu(\cdot))$ generated by the progressive hedging algorithm above from any starting element is sure to converge to one such solution $x^*(\cdot)$ paired with a parameter value y^* and an associated pair $(u^*(\cdot), w^*(\cdot))$ furnishing a saddle point as in (3.12). In this convergence the distance expression*

$$||x^\nu(\cdot) - x^*(\cdot)||^2 + ||y(\cdot) - y^*||^2 + \frac{1}{r^2}||u^\nu(\cdot) - u^*(\cdot)||^2 + \frac{1}{r^2}||w^\nu(\cdot) - w^*(\cdot)||^2 \tag{3.15}$$

will form a decreasing sequence.

Proof. This follows from the classical convergence theorem stated in Section 2 for the risk-neutral case by virtue of the above reformulation of the risk-averse minimization problem into a risk-neutral minimization problem. That is where the real contribution and innovation lies. □

Note that the value y^* produced in this computation will belong to the argmin in the formula (3.8) for the case of the random variable $X = f(x(\cdot), \cdot)$. Thus, unless that argmin is a nontrivial interval, it will equal $\mathrm{VaR}_\alpha(f(x^*(\cdot), \cdot))$.

4 Specialization to single-stage stochastic programming

Inspection of the single-stage case, where $N = 1$, may help to fix the ideas and open them to simple comparison. In that case the classical problem is to

$$\text{minimize } E_\xi[f(x,\xi)] \text{ over } x \in \cap_\xi C(\xi), \tag{4.1}$$

with the previous $x = (x_1, x_2, \ldots, x_N)$ reduced just to $x = x_1$ and meeting the nonanticipativity requirement by being a constant independent of ξ. Now each $C(\xi)$ is a nonempty closed convex subset of $\mathbb{R}^n = \mathbb{R}^{n_1}$. The progressive hedging algorithm relaxes the constancy of x by passing to functions $x(\cdot) \in \mathcal{L}$ controlled by a multiplier element $w(\cdot) \in \mathcal{L}$ having $E_\xi[w(\xi)] = 0$.

Progressive Hedging Algorithm $1'$ (single-stage risk-neutral case)**.** *Iterations $\nu = 1, 2, \ldots$ with a parameter $r > 0$ proceed as follows from current elements $x^\nu \in \mathbb{R}^n$ and $w^\nu(\cdot)$ having $E_\xi[w^\nu(\xi)] = 0$.*

(a) *Get $\hat{x}^\nu(\cdot) \in \mathcal{L}$ by solving for each scenario $\xi \in \Xi$ a strongly convex programming problem in the variable $x(\xi) \in \mathbb{R}^n$ that has a unique solution $x(\xi) = \hat{x}^\nu(\xi)$):*

$$\hat{x}^\nu(\xi) = \operatorname*{argmin}_{x(\xi)\in C(\xi)} \Big\{ f(x(\xi),\xi) + w^\nu(\xi)\cdot x(\xi) + \frac{r}{2}||x(\xi) - x^\nu||^2 \Big\}. \tag{4.2}$$

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

$$x^{\nu+1} = E_\xi[\hat{x}^\nu(\xi)], \qquad w^{\nu+1}(\xi) = w^\nu(\xi) + r[\hat{x}^\nu(\xi) - x^{\nu+1}]. \tag{4.3}$$

The accomplishment of Section 3 was to extend this procedure from the risk-neutral objective (4.1) to stochastic programming problems objectives in terms of conditional value-at-risk. In the single-stage case the altered problem is

$$\text{minimize } \mathrm{CVaR}_\alpha(f(x,\cdot)) \text{ over } x \in \cap_\xi C(\xi), \tag{4.4}$$

where

$$\mathrm{CVaR}_\alpha(f(x,\cdot)) = \min_{y\in \boldsymbol{R}} \Big\{ y + \frac{1}{1-\alpha} E_\xi[\max\{0, f(x,\xi) - y\}] \Big\}. \tag{4.5}$$

The extended algorithm for this involves not only functions $x(\cdot)$ but also $y(\cdot)$, along with multiplier elements $w(\cdot)$ and $u(\cdot)$ having expectation equal to 0.

Progressive Hedging Algorithm $2'$ (single-stage risk-averse case for CVaR_α)**.** *Iterations $\nu = 1, 2, \ldots$ with a parameter value $r > 0$ proceed as follows from current elements $x^\nu \in \mathbb{R}^n$, $y^\nu \in \mathbb{R}$, and functions $w^\nu(\cdot)$ and $u^\nu(\cdot)$ having $E_\xi[w(\xi)] = 0$ and $E_\xi[u(\xi)] = 0$.*

(a) *Get $\hat{x}^\nu(\cdot)$ and $\hat{y}^\nu(\cdot)$ by solving for each scenario $\xi \in \Xi$ a strongly convex programming problem in the variables $x(\xi) \in \mathbb{R}^n$ and $y(\xi) \in \mathbb{R}$ that has a unique solution $(y(\xi), x(\xi)) = (\hat{y}^\nu(\xi), \hat{x}^\nu(\xi))$:*

$$\begin{aligned}(\hat{y}^\nu(\xi), \hat{x}^\nu(\xi)) = \operatorname*{argmin}_{y(\xi)\in \boldsymbol{R},\, x(\xi)\in C(\xi)} \Big\{ & y(\xi) + \tfrac{1}{1-\alpha}\max\{0, f(x(\xi),\xi) - y(\xi)\} \\ & + u^\nu(\xi)y(\xi) + \tfrac{r}{2}|y(\xi) - y^\nu|^2 + w^\nu(\xi)\cdot x(\xi) + \tfrac{r}{2}||x(\xi) - x^\nu(\xi)||^2 \Big\}.\end{aligned} \tag{4.6}$$

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

$$\begin{aligned} x^{\nu+1} &= E_\xi[\hat{x}^\nu(\xi)], \qquad & w^{\nu+1}(\xi) &= w^\nu(\xi) + r[\hat{x}^\nu(\xi) - x^{\nu+1}], \\ y^{\nu+1} &= E_\xi[\hat{y}^\nu(\xi)], \qquad & u^{\nu+1}(\xi) &= u^\nu(\xi) + r[\hat{y}^\nu(\xi) - y^{\nu+1}]. \end{aligned} \tag{4.7}$$

The "max" term in (4.6) creates nonsmoothness, at least on the surface, but can be handled by familiar tricks such as introducing an "epigraphical" decision variable. It's worth seeing how this plays out, because the effect in our setting of stochastic programming is to replace the single-stage problem by a two-stage problem. The second-stage epigraphical variable will be noted by $a(\xi)$. The idea behind it is that, in the expression on the right side of (4.5), one has

$$\max\{0, f(x,\xi) - y\} \le a(\xi) \iff a(\xi) \ge 0 \text{ and } f(x,\xi) - y - a(\xi) \le 0. \tag{4.8}$$

The risk-averse stochastic programming problem (4.4) can be identified then with

$$\text{minimize } y + \frac{1}{1-\alpha} E_\xi[a(\xi)] \text{ subject to } x \in \cap_\xi C(\xi) \text{ and } a(\xi) \ge 0, \tag{4.9}$$

where the constraints can be expressed as $(y, x, a(\xi)) \in \widetilde{C}(\xi)$ for an obvious choice of $\widetilde{C}(\xi)$. The algorithm above can be recast in this framework then as well, and the objective in each iteration will be *linear*.

5 Extension of progressive hedging to other risk-averse objectives

Only the case in (3.1) of $\mathcal{R} = \text{CVaR}_\alpha$ was worked out in Section 3, but a large class of other risk measures can be handled similarly. The key is the existence of an "expectational" representation of $\mathcal{R}$ along the lines of the CVaR minimization formula in (3.8) or more elaborate formulas involving more than one auxiliary variable. This is a topic explained in [3] in terms of the concept of an associated *measure of regret* $\mathcal{V}$ such that

$$\mathcal{R}(X) = \min_{y \in \boldsymbol{R}} \Big\{ y + \mathcal{V}(X - y) \Big\}. \tag{5.1}$$

For the purposes at hand the measure of regret in this formula should be of expectational type,

$$\mathcal{V}(X) = E[v(X)] \text{ for a function } v : (-\infty, \infty) \to (-\infty, \infty], \tag{5.2}$$

where, under the axioms in [3], v is closed convex and nondecreasing with $v(0) = 0$ but $v(x) > x$ for $x \ne 0$. The case of $\mathcal{R}(X) = \text{CVaR}_\alpha(X)$ for $\alpha \in (0,1)$ fits this with

$$v(x) = \frac{1}{1-\alpha} \max\{x, 0\}. \tag{5.3}$$

The extension of progress hedging to a risk measure $\mathcal{R}$ belonging to this larger family is completely straightforward: it is just a matter of replacing the special $v(x)$ in (5.3) by a more general $v(x)$. Thus, for instance, the term $\frac{1}{1-\alpha} \max\{0, f(x(\xi),\xi) - y(\xi)\}$ in the iteration step (3.13) of the algorithm gets generalized to $v(f(x(\xi),\xi) - y(\xi))$.

But the extensions don't have to stop with the simple case of (5.1)–(5.2). They can be promulgated further to *mixed* measures of risk of the form

$$\mathcal{R}(X) = \lambda_1 \mathcal{R}_1(X) + \cdots + \lambda_m \mathcal{R}_m(X) \text{ with } \lambda_k > 0,\ \lambda_1 + \cdots + \lambda_m = 1, \tag{5.4}$$

where each $\mathcal{R}_k$ is of the type just discussed,

$$\mathcal{R}_k(X) = \min_{y_k \in \boldsymbol{R}} \Big\{ y_k + E[v_k(X - y_k)] \Big\}. \tag{5.5}$$

In this wider picture we have

$$\mathcal{R}(X) = \min_{y_1,\dots,y_m} E[h(X, y_1, \dots, y_m)] \text{ for } h(x, y_1, \dots, y_m) = \sum_{k=1}^{m} \lambda_k [y_k + v_k(x - y_k)], \tag{5.6}$$

so again the objective can be reconsituted as an expectation but now with auxiliary variables y_k instead of just one y. These can be incorporated into the progressive hedging algorithm as additional decision variables $y_k(\xi)$.

A notable example of such mixing is the class of mixed CVaR measures of risk,

$$\mathcal{R}(X) = \lambda_1 \text{CVaR}_{\alpha_1}(X) + \cdots + \lambda_m \text{CVaR}_{\alpha_m}(X) \text{ with } \lambda_k > 0,\ \lambda_1 + \cdots + \lambda_m = 1. \tag{5.7}$$

This is important because every so-called *spectral* measure of risk can be approximated arbitrarily closely by such a mixture, cf. [3].

Another prospect for extension of progressive hedging is objectives articulated with risk measures invoked not just for the final cost but nested in stages. However, this is a subject too big and demanding to get into here.

References

[1] Rockafellar, R. T., and Uryasev, S., "Optimization of conditional value-at-risk," *Journal of Risk* **2** (2000), 21–43.

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

[3] Rockafellar, R. T., and Uryasev, S., "The fundamental risk quadrangle in risk management, optimization and statistical esimation," *Surveys in Operations Research and Management Science* **18** (2013), 33–53.

[4] Rockafellar, R. T., and Wets, R. J-B, "Scenarios and policy aggregation in optimization under uncertainty." *Mathematics of Operations Research* **16** (1991), 119–147.

[5] Watson, J.-P., and Woodruff, D.L., "Progressive hedging innovations for a class of stochastic mixed-integer resource allocation problems," *Computational Management Science* **8** (2010), 355–370.