On solving multistage stochastic programs with coherent risk measures

Andy Philpott [*]   Vitor de Matos[†]   Erlon Finardi[‡]

August 13, 2012

Abstract

We consider a class of multistage stochastic linear programs in which at each stage a coherent risk measure of future costs is to be minimized. A general computational approach based on dynamic programming is derived that can be shown to converge to an optimal policy. By computing an inner approximation to future cost functions, we can evaluate an upper bound on the cost of an optimal policy, and an outer approximation delivers a lower bound. The approach we describe is particularly useful in sampling-based algorithms, and a numerical example is provided to show the efficacy of the methodology when used in conjunction with stochastic dual dynamic programming.

1   Introduction

Multistage stochastic linear programming models have been studied for many years, and although there are a number of reports of practical applications (see e.g. [2],[21]) there are still very few implementations of these models in commercial settings. The classical version of this model treats uncertainty using a scenario tree that branches at each stage. Even with a small number of outcomes per stage, the size of the scenario tree grows exponentially with the number of stages. In general this makes it impossible to find the optimal solution of such problems, even using sampling approaches (see [17]).

In many circumstances multistage stochastic linear programming problems can be modelled as stochastic optimal control problems. These distinguish between control and state variables that together satisfy a state-constrained linear dynamical system with some random disturbances. When the random disturbances are stagewise independent, these models can be attacked by dynamic programming methods. This has proved to be a very powerful approach in problems involving long-term planning of hydroelectric power systems [12].

---

[*]Electric Power Optimization Centre, University of Auckland, New Zealand.

[†]Laboratório de Planejamento em Sistemas de Energia Elétrica, Universidade Federal de Santa Catarina, Brazil.

[‡]Laboratório de Planejamento em Sistemas de Energia Elétrica, Universidade Federal de Santa Catarina, Brazil.

Stochastic linear optimal control problems typically minimize the expectation of some measure of cost. In hydrothermal scheduling this is the expected fuel and shortage cost. This makes sense when decision makers wish to minimize the total average fuel and shortage cost in the long run. In practice such an approach can result in a higher probability of shortage than planners wish to accept. Then they might seek to control this probability as well as keeping costs down. One way of doing this is to add chance constraints to the model (see [16]). In general such constraints are not convex, and so this limits the scale of problems to which this approach is useful.

An alternative approach that retains convexity is to use the axiomatic approach to risk pioneered by Artzner *et al* [1]. They define a *risk measure* to be a function from the space of random variables to $\mathbb{R}$, and this is said to be *coherent* if it satisfies four key axioms. These guarantee that convex stochastic optimization problems that minimize expected cost remain convex when this is replaced by the coherent risk measure.

The use of coherent risk measures in dynamic programming models was first introduced by [14] and explored in detail by [15]. As demonstrated in [14], any time-consistent dynamic risk measure can be constructed in terms of single-period risk measures by a recursive formula. If the risk measure is coherent, then the resulting model can be solved using a dynamic programming recursion that retains convexity in the Bellman functions. This makes the model particularly attractive for optimization.

In this paper we explore the use of coherent risk measures in multistage stochastic linear programming problems. We provide a general framework for computing solutions to such models. This relies on an outer-approximation algorithm (formed by cutting planes) and an inner approximation algorithm (formed by convex combinations of feasible policies). The outer approximation provides a lower bound on the value of an optimal policy, and the inner approximation gives a candidate policy and an upper bound on its value. An important ingredient in our approach is the dual representation theorem for coherent risk measures established in [1]. This enables us to replace the risk measure computation at each stage by an expectation with an adjusted probability measure. This approach been applied in [20] to the case where the risk measure is a convex combination of expectation and conditional value at risk. Using such a "change of measure" removes the need to record an extra "value-at-risk" state variable as described in [13]. We show here how this approach can be extended to any coherent risk measure.

The theoretical basis of the approach we describe is quite general. In practical application, however, it is limited to small problem instances unless we use sampling. We show how the Stochastic Dual Dynamic Programming (SDDP) algorithm [12] for multistage stochastic linear programming can be modified to accommodate general coherent risk measures. To illustrate the approach we apply it to an instance of a large-scale hydro-thermal scheduling problem arising in Brazil.

The main contributions of our work are as follows. We formalize a general approximation procedure for computing solutions to multistage stochastic programming problems that minimize dynamic coherent risk measures. This procedure gives lower and upper bounds on the optimal value of the problem and a feasible policy with a value that lies between these. This provides a stopping criterion for the approximation procedure. In risk-averse SDDP models such as those discussed in [13] and [20], our approximation procedure can be shown to provide a reasonably tight upper bound on the policy value.

Estimating the value of a candidate policy (or an upper bound on this) by sampling appears to be very difficult for risk-averse SDDP models [20] and our approach provides a straightforward alternative. We demonstrate the effectiveness of these bounds by applying the procedure to a large-scale hydro-thermal scheduling problem.

The paper is laid out as follows. In the next section we formulate the class of multistage stochastic programs that we study, and outline the theory of coherent dynamic risk measures as applied to this class of problems. In Section 3 we describe the outer approximation and inner approximation algorithms as they apply to the risk-neutral case (where the risk measure is expectation). In Section 4 we show how these algorithms can be extended to general coherent risk measures, and derive upper and lower bounds on the optimal value of the problem. Section 5 examines how the SDDP algorithm might be adapted to include general coherent risk measures, and then Section 6 presents some results of applying the methodology to an instance of a large-scale hydro-thermal scheduling problem arising in Brazil. We conclude the paper with some discussion in Section 7.

2   Preliminaries

The type of problem we consider has T stages, denoted $t = 1, 2, \ldots, T$, in each of which a random right-hand-side vector $b_t(\omega_t) \in \mathbb{R}^m$ has a finite number of realizations defined by $\omega_t \in \Omega_t$, each with a strictly positive probability. The assumption of finite probability spaces greatly simplifies the analysis, whereby we can dispense with most measurability assumptions, such as, for example, specifying constraints that hold almost surely. Since we have in mind the solution of a large-scale finite problem that is obtained by sampling, the assumption is not too restrictive.

We assume that the outcomes ω_t are stagewise independent, and that Ω_1 is a singleton, so the first-stage problem is

$$
\begin{aligned}
z = \min \quad & c_1^\top x_1 + \rho_2(Q_2(x_1, \omega_2)) \\
\text{s.t.} \quad & A_1 x_1 = b_1, \\
& x_1 \geq 0,
\end{aligned}
\tag{1}
$$

where $x_1 \in \mathbb{R}^n$ is the first stage decision and $c_1 \in \mathbb{R}^n$ a cost vector, A_1 is a $m \times n$ matrix, and $b_1 \in \mathbb{R}^m$.

We denote by $Q_2(x_1, \omega_2)$ the optimal value of the second stage problem associated with decision x_1 and realization $\omega_2 \in \Omega_2$. In this problem ρ_2 is defined to be a one-step coherent risk measure defined on the random variables $Q_2(x_1, \omega_2)$. According to [1] a function ρ from a space $\mathbb{Z}$ of random variables to $\mathbb{R}$ is a *coherent risk measure* if ρ satisfies the following axioms for Z_1 and $Z_2 \in \mathbb{Z}$.

Subadditivity: $\rho(Z_1 + Z_2) \leq \rho(Z_1) + \rho(Z_2)$;
Monotonicity: If $Z_1 \leq Z_2$, then $\rho(Z_1) \leq \rho(Z_2)$;
Positive homogeneity: If $\lambda \in \mathbb{R}$ and $\lambda > 0$, then $\rho(\lambda Z_1) = \lambda \rho(Z_1)$;
Translation equivariance: If $\lambda \in \mathbb{R}$, then $\rho(\mathbb{I}\lambda + Z_1) = \lambda + \rho(Z_1)$.

Coherent risk measures have some attractive properties for optimization. Subadditivity and positive homogeneity implies:

Convexity: $\rho(\alpha Z_1 + (1 - \alpha)Z_2) \leq \alpha \rho(Z_1) + (1 - \alpha)\rho(Z_2)$, for $\alpha \in [0, 1]$,

which with monotonicity guarantees that convex optimization problems remain convex when the (convex) objective functions are composed with a coherent risk measure as in (1) above.

The use of coherent risk measures in dynamic programming models was introduced by [14] and is described for general Markov decision problems by [15]. Given a probability space $(\Omega, \mathcal{F}, P)$, a dynamic risk measure applies to a situation in which we have a random sequence of costs $(Z_1, Z_2, \ldots, Z_T)$ which is adapted to some filtration $\{0, \Omega\} = \mathcal{F}_1 \subset \mathcal{F}_2 \ldots \subset \mathcal{F}_T \subset \mathcal{F}$ of σ-fields, where Z_1 is assumed to be deterministic. A dynamic risk measure is then defined to be a sequence of *conditional risk measures* $\{\rho_{t,T}\}$, $t = 1, 2, \ldots, T$. Given a dynamic risk measure, we can derive a corresponding single-period risk measure using

$$\rho_t(Z_{t+1}) = \rho_{t,T}(0, Z_{t+1}, 0, \ldots, 0).$$

As demonstrated in [15, Theorem 1], any time-consistent dynamic risk measure can then be constructed in terms of single-period risk measures ρ_t by the formula

$$\rho_{t,T}(Z_t, Z_{t+1}, \ldots, Z_T) = Z_t + \rho_t(Z_{t+1} + \rho_{t+1}(Z_{t+2} + \ldots + \rho_{T-2}(Z_{T-1} + \rho_{T-1}(Z_T))\ldots))).$$

The measure $\rho_{t,T}(Z_t, Z_{t+1}, \ldots, Z_T)$ can be interpreted as a *certainty equivalent cost* or *risk-adjusted expected cost*, namely what deterministic sum would a decision maker pay at stage t to avoid all the future costs $(Z_t, Z_{t+1}, \ldots, Z_T)$ incurred by a candidate policy. In the context of maximizing returns Z_t, $\rho_{t,T}(Z_t, Z_{t+1}, \ldots, Z_T)$ can be interpreted (see [14]) as the minimum amount of money that one has to add to the position at stage t to make it acceptable.

In our setting, this construction leads us to a recursive form for the problem to be solved in the second and later stages t. Given a coherent one-step risk measure ρ_{t+1}, decisions x_{t-1} and realization ω_t, this problem can be written as

$$\begin{aligned} Q_t(x_{t-1}, \omega_t) = \min \quad & c_t^\top x_t + \rho_{t+1}(Q_{t+1}(x_t, \omega_{t+1})) \\ \text{s.t.} \quad & A_t x_t = b_t(\omega_t) - E_t x_{t-1}, \qquad [\pi_t(\omega_t)] \\ & x_t \geq 0, \end{aligned} \qquad (2)$$

where $x_t \in \mathbb{R}^n$ is the decision in stage t, c_t its cost, and A_t and E_t denote $m \times n$ matrices. Here $\pi_t(\omega_t)$ denotes the Lagrange multipliers of the constraints. We denote

$$\mathcal{X}_t(\omega) = \{x_t \geq 0 : A_t x_t = b_t(\omega) - E_t x_{t-1}\}.$$

In the last stage we assume either that $\rho_{T+1}(Q_{T+1}(x_T, \omega_{T+1})) = 0$, or there is some known (convex) polyhedral function $\mathcal{Q}_{T+1}(x_T)$ that defines $\rho_{T+1}(Q_{T+1}(x_T, \omega_{T+1}))$. (We adopt the notational convention that upper case Q_t depends on ω_t whereas its caligraphic counterpart $\mathcal{Q}_t$ does not. That is, $\mathcal{Q}_t$ is $\mathcal{F}_{t-1}$-measurable but Q_t is not.)

Observe that the coherence of ρ_{t+1} implies that it is monotonic and convex, and so $\rho_{t+1}(Q_{t+1}(x_t, \omega_{t+1}))$ is a convex function of x_t whenever $Q_{t+1}(x_t, \omega_{t+1})$ is convex in x_t for every ω_{t+1}. This means that $Q_t(x_{t-1}, \omega_t)$ is convex in x_{t-1} for every ω_t whenever $Q_{t+1}(x_t, \omega_{t+1})$ is convex in x_t for every ω_{t+1}, and so it follows by induction that for every $t = 1, 2, \ldots, T$, $Q_t(x_{t-1}, \omega_t)$ is convex in x_{t-1} for every ω_t.

4

A model of the form we descibe above has been implemented and applied to hydro-thermal scheduling problems in Brazil by [13] and [20] in the special case where $\rho_t(Z_{t+1})$ is a convex combination of expectation and conditional value at risk, i.e.

$$\rho_t(Z_{t+1}) = (1 - \lambda_t)\mathbb{E}[Z_{t+1} \mid \mathcal{F}_t] + \lambda_t \inf_u \{u + \alpha^{-1}\mathbb{E}[(Z_{t+1} - u)_+ \mid \mathcal{F}_t]\}. \qquad (3)$$

These papers build a near optimal policy for instances of this problem using SDDP. The approach of [13] represents $\rho_t(Z_{t+1})$ using a variable u that augments the state variables representing reservoir storage. The approach of [20] is to represent $\rho_t(Z_{t+1})$ by a change of probability measure that depends on the state variables representing reservoir storage. We show below how this approach extends to general coherent risk measures.

Our goal is to construct an optimal solution for the multistage problem defined by (1) and (2). Under the assumption that the random disturbances ω_t are stagewise indepen-dent, the solution has the form of a *policy* $\boldsymbol{\pi}$ defined for each stage t by a mapping from $\mathcal{X}_{t-1}(\omega_{t-1}) \times \Omega_t$ to $\mathcal{X}_t(\omega_t)$, specifying the decision $x_t(x_{t-1}, \omega_t)$ made by the policy $\boldsymbol{\pi}$ at time t

When ρ_t is the expectation operator, this problem becomes a classical multistage stochastic linear program with stagewise independent random variables. Such a model seeks to minimize the total cost on average. An important advantage of this special case is that the objective function is the sum of expectations and so it can be expressed as the expectation of a sum of random costs that is easy to estimate by sampling. This makes the application of a methodology like SDDP particularly attractive as this yields a lower bound on the cost of any feasible policy that can be compared with an estimate of its actual cost by simulation.

Of course we would like to model situations in which ρ_t is not expectation, but this presents some challenges. As we show below, it is possible to derive a lower bound on the optimal value of any policy when using outer approximation. However, estimating the objective function value of a candiadte policy using Monte Carlo sampling seems to be very difficult when ρ_t is not expectation. The absence of such an estimate is a serious disadvantage for decision makers, as well as posing a problem for SDDP, since estimating or computing the value of a candidate policy is needed to determine when to stop the algorithm, or at least to tell whether the algorithm has delivered a close to optimal policy when it is stopped. We demonstrate that it is possible to define a policy using inner approximation and provide an upper bound on its value. This together with a lower bound on the value of an optimal policy provides a suitable optimality check.

3 Policy specification under expectation

We now proceed to describe methods to build policies for the problem to be solved. One of these methods is the standard outer approximation algorithm based on Kelley's cutting plane method [8]. The outer approximation gives a lower bound on optimal solution for the problem. The other method is an inner approximation based on the algorithm developed by [5]. This can be shown to give an upper bound on optimal solution for the problem. These algorithms are simplest to describe when the risk measure is expectation,

so this section is confined to that case. As we shall see, the extension to general coherent risk measures is straightforward.

3.1 Outer approximation

One may build a policy defined at each stage t by a polyhedral outer approximation of $\mathbb{E}[Q_{t+1}(x_t, \omega_{t+1})]$. This approximation is constructed using cutting planes. In other words in each tth-stage problem, $\mathbb{E}[Q_{t+1}(x_t, \omega_{t+1})]$ is replaced by the variable θ_{t+1} which is constrained by the set of linear inequalities

$$\theta_{t+1} + \bar{\pi}_{t+1,k}^\top E_{t+1} x_t \geq \bar{g}_{t+1,k} \qquad \text{for } k = 1, 2, ...K, \tag{4}$$

where K is the number of cuts. Here $\bar{\pi}_{t+1,k} = \mathbb{E}[\pi_{t+1}(\omega_{t+1})]$, which defines the gradient $-\bar{\pi}_{t+1,k}^\top E_{t+1}$ and the intercept $\bar{g}_{t+1,k}$ for cut k in stage t, where

$$\bar{g}_{t+1,k} = \mathbb{E}[\tilde{Q}_{t+1}(x_t^k, \omega_{t+1})] + \bar{\pi}_{t+1,k}^\top E_{t+1} x_t^k,$$

and we define $\tilde{Q}_t$ by the approximate stage problem

$$\begin{aligned}
\tilde{Q}_t(x_{t-1}, \omega_t) = \min \quad & c_t^\top x_t + \theta_{t+1} \\
\text{s.t.} \quad & A_t x_t = b_t(\omega_t) - E_t x_{t-1}, \qquad [\pi_t(\omega_t)] \\
& \theta_{t+1} + \bar{\pi}_{t+1,k}^\top E_{t+1} x_t \geq \bar{g}_{t+1,k}, \quad k = 1, 2, ...K, \\
& x_t \geq 0.
\end{aligned} \tag{5}$$

Recall

$$\mathcal{X}_t(\omega) = \{x_t \geq 0 : A_t x_t = b_t(\omega) - E_t x_{t-1}\}.$$

Proposition 1 *If for any $x_t \in \mathcal{X}_t(\omega_t)$, $\bar{g}_{t+1,k} - \bar{\pi}_{t+1,k}^\top E_{t+1} x_t \leq \mathbb{E}[Q_{t+1}(x_t, \omega_{t+1})]$ for every k then*

$$\tilde{Q}_t(x_{t-1}, \omega_t) \leq Q_t(x_{t-1}, \omega_t)$$

Proof. For any $x_t \in \mathcal{X}_t(\omega_t)$ the optimal choice of θ_{t+1} satisfies

$$\begin{aligned}
c_t^\top x_t + \theta_{t+1} &= c_t^\top x_t + \max_k \{\bar{g}_{t+1,k} - \bar{\pi}_{t+1,k}^\top E_{t+1} x_t\} \\
&\leq c_t^\top x_t + \mathbb{E}[Q_{t+1}(x_t, \omega_{t+1})]
\end{aligned}$$

by hypothesis. It follows that

$$\begin{aligned}
\tilde{Q}_t(x_{t-1}, \omega_t) &= \min_{x_t \in \mathcal{X}_t(\omega_t)} \{c_t^\top x_t + \max_k \{\bar{g}_{t+1,k} - \bar{\pi}_{t+1,k}^\top E_{t+1} x_t\}\} \\
&\leq \min_{x_t \in \mathcal{X}_t(\omega_t)} \{c_t^\top x_t + \mathbb{E}[Q_{t+1}(x_t, \omega_{t+1})]\} \\
&= Q_t(x_{t-1}, \omega_t)
\end{aligned}$$

giving the desired result. ∎

Proposition 1 shows that the outer approximation property is inherited from stage to stage, so it can be used to compute a lower bound on an optimal policy for problem (1) using the following algorithm.

Outer Approximation Algorithm

1.   Let $\tilde{\mathcal{Q}}_{T+1}(x) = 0$ for every $x \in \mathcal{X}_T(\omega_T)$.

2.   For $t = 1, ..., T-1$ define J_t points $x_t^1, x_t^2, \ldots, x_t^{J_t}$.

3.   For $t = T, ..., 2$ and for each $k = 1, 2, \ldots, J_{t-1}$,

   For each $\omega_t \in \Omega_t$, compute

$$\tilde{Q}_t(x_{t-1}^k, \omega_t) = \min \quad c_t^\top x + \tilde{\mathcal{Q}}_{t+1}(x) \qquad (6)$$
$$\text{s.t.} \quad A_t x = b_t(\omega_t) - E_t x_{t-1}^k,$$
$$x \geq 0.$$

   Compute $\bar{\pi}_{t,k} = \mathbb{E}[\pi_t(\omega_t)]$, $\bar{g}_{t,k} = \mathbb{E}[\tilde{Q}_t(x_{t-1}^k, \omega_t)] + \bar{\pi}_{t,k}^\top E_t x_{t-1}^k$,
   and define $\tilde{\mathcal{Q}}_t(x) = \max_k \{\bar{g}_{t,k} - \bar{\pi}_{t,k}^\top E_t x\}$, $x \in \mathcal{X}_{t-1}(\omega_{t-1})$.

4. Solve

$$z = \min \quad c_1^\top x_1 + \tilde{\mathcal{Q}}_2(x_1)$$
$$\text{s.t.} \quad A_1 x_1 = b_1,$$
$$x_1 \geq 0.$$

This outer approximation algorithm defines a candidate policy in terms of the cutting planes that define $\tilde{\mathcal{Q}}_t(x)$. The action to be taken at stage t at any state x_{t-1} is defined by the solution to (5) with x_{t-1} chosen to be x_{t-1}^k. It is easy to see that the conditions of Proposition 1 guarantee that the value z yields a lower bound on the value of an optimal policy for the problem. In the context of minimizing expected cost, a statistical upper bound can be obtained by simulating the candidate policy and computing a sample average. This is not possible for more general risk measures as we shall see.

The outer approximation method has many variations that depend on how the points $x_t^1, x_t^2, \ldots, x_t^{J_t}$ are chosen. For example in the classical version of SDDP [12] these points are constructed in increments (forward passes) by constructing sample paths using the candidate policy obtained so far, and augmenting the set with the states visited in each iteration. In other words, SDDP performs a sequence of outer approximations (backward passes) increasing J_t at each pass. Our description above also admits algorithms that select $x_t^1, x_t^2, \ldots, x_t^{J_t}$ by other means (e.g. the quasi-Monte Carlo method described in [7]).

We now turn our attention to the dual approach to outer approximation, that we call *inner* approximation. This can be shown to give an upper bound on the optimal policy. This will become important when this is difficult to estimate using Monte Carlo simulation.

3.2   Inner approximation

Suppose at stage t that we have upper bounds $q_{t+1}^1, q_{t+1}^2, \ldots, q_{t+1}^{J_t}$ on the values of $\mathbb{E}[Q_{t+1}(x_t, \omega_{t+1})]$ at J_t points $x_t^1, x_t^2, \ldots, x_t^{J_t}$. It is easy to see that the convex hull $\mathcal{H}$ of the points

$\{(x_t^1, q_{t+1}^1), (x_t^2, q_{t+1}^2), \ldots, (x_t^{J_t}, q_{t+1}^{J_t})\}$ defines a polyhedral set that is a subset of the convex epigraph of $\mathbb{E}[Q_{t+1}(x_t, \omega_{t+1})]$. It is convenient in what follows to define the simplex

$$\Lambda_t = \{\lambda \mid \sum_{j=1}^{J_t} \lambda^j = 1, \ \lambda^j \geq 0\}.$$

Then the lower boundary of $\mathcal{H}$ is

$$\hat{\mathcal{Q}}_{t+1}(x_t) = \min \quad \sum_{j=1}^{J_t} \lambda^j q_{t+1}^j$$
$$\text{s.t.} \quad \sum_{j=1}^{J_t} \lambda^j x_t^j = x_t,$$
$$\lambda \in \Lambda_t,$$

which defines an inner (upper bound) approximation to $\mathbb{E}[Q_{t+1}(x_t, \omega_{t+1})]$. We use this approximation to find an upper bound on $c_t^\top x_t + \mathbb{E}[Q_{t+1}(x_t, \omega_{t+1})]$ under each realization of ω_t as follows.

Formally we define a set $\{x_{t-1}^1, x_{t-1}^2, \ldots, x_{t-1}^{J_{t-1}}\}$ of J_{t-1} possible values of x_{t-1}, and for each x_{t-1}^i we compute

$$q_t^i(\omega_t) = \min \quad c_t^\top x_t + \sum_{j=1}^{J_t} \lambda^j q_{t+1}^j$$
$$\text{s.t.} \quad A_t x_t = b_t(\omega_t) - E_t x_{t-1}^i,$$
$$\sum_{j=1}^{J_t} \lambda^j x_t^j = x_t,$$
$$\lambda \in \Lambda_t.$$

We denote by $\hat{x}_t$ and $\hat{\lambda}$ the minimizers of this problem.

Proposition 2 If $q_{t+1}^j \geq \mathbb{E}[Q_{t+1}(x_t^j, \omega_{t+1})]$ for all $j = 1, 2, \ldots, J_t$, then $q_t^i(\omega_t) \geq Q_t(x_{t-1}^i, \omega_t)$ for all $i = 1, 2, \ldots, J_{t-1}$.

Proof. Since $q_{t+1}^j \geq \mathbb{E}[Q_{t+1}(x_t^j, \omega_{t+1})]$ we have that

$$q_t^i(\omega_t) \ = \ c_t^\top \hat{x}_t + \sum_{j=1}^{J_t} \hat{\lambda}^j q_{t+1}^j$$

$$\geq \ c_t^\top \hat{x}_t + \sum_{j=1}^{J_t} \hat{\lambda}^j \mathbb{E}[Q_{t+1}(x_t^j, \omega_{t+1})]$$

$$\geq \ c_t^\top \hat{x}_t + \mathbb{E}[Q_{t+1}(\sum_{j=1}^{J_t} \hat{\lambda}^j x_t^j, \omega_{t+1})]$$

where the first inequality holds by hypothesis, and the second inequality follows from the convexity of $\mathbb{E}[Q_{t+1}(x_t, \omega_{t+1})]$. But the right-hand-side is $c_t^\top \hat{x}_t + \mathbb{E}[Q_{t+1}(\hat{x}_t, \omega_{t+1})]$, which is the value of a feasible solution $\hat{x}_t$ of (2), which must be at least the optimal value $Q_t(x_{t-1}^i, \omega_t)$. ∎

It is easy to see that Proposition 2 also guarantees that $\mathbb{E}[q_t^i(\omega_t)] \geq \mathbb{E}[Q_t(x_{t-1}^i, \omega_t)]$ for every i. This means that for each i, $q_t^i = \mathbb{E}[q_t^i(\omega_t)]$ will give an upper bound on

$\mathbb{E}[Q_t(x_{t-1}^i, \omega_t)]$. Thus q_t^i defines an inner approximation $\hat{\mathcal{Q}}_t(x_{t-1})$ of $\mathbb{E}[Q_t(x_{t-1}, \omega_t)]$ formed by the convex hull of the points $\{(x_{t-1}^1, q_t^1), (x_{t-1}^2, q_t^2), \ldots, (x_{t-1}^{J_{t-1}}, q_t^{J_{t-1}})\}$.

For computational purposes, the inner approximation model can be expressed purely in terms of multipliers λ^j. This gives

$$\begin{aligned} q_{t-1}^j(\omega_t) = \min \quad & \sum_{j=1}^{J_t} \left(c_t^\top x_t^j \right) \lambda^j + \sum_{j=1}^{J_t} \lambda^j q_t^j \\ \text{s.t.} \quad & \sum_{j=1}^{J_t} \left(A_t x_t^j \right) \lambda^j = b_t(\omega_t) - E_t x_{t-1}, \\ & \lambda \in \Lambda_t. \end{aligned}$$

Since the inner approximation property is inherited from stage to stage, it can be used to compute an upper bound on an optimal policy for problem (1).

Inner Approximation Algorithm

1. Let $\hat{\mathcal{Q}}_{T+1}(x) = 0$ for every $x_T \in \mathcal{X}_T(\omega_T)$.

2. For $t = 1, ..., T-1$ define J_t points $x_t^1, x_t^2, \ldots, x_t^{J_t}$.

3. Compute $q_T^1, q_T^2, \ldots, q_T^{J_{T-1}}$ by solving for each $i = 1, 2, \ldots, J_{T-1}$, and each $\omega_T \in \Omega_T$,

$$\begin{aligned} q_T^i(\omega_T) = \min \quad & c_T^\top x_T + \hat{\mathcal{Q}}_{T+1}(x) \\ \text{s.t.} \quad & A_T x_T = b_T(\omega_T) - E_T x_{T-1}^i, \\ & x_T \geq 0, \end{aligned}$$

and then computing $q_T^i = \mathbb{E}[q_T^i(\omega_T)]$.

4. For $t = T-1, ..., 2$ compute $q_t^1, q_t^2, \ldots, q_t^{J_{t-1}}$ by solving for each $i = 1, 2, \ldots, J_{t-1}$, and each $\omega_t \in \Omega_t$,

$$\begin{aligned} q_t^i(\omega_t) = \min \quad & c_t^\top x_t + \sum_{j=1}^{J_t} \lambda^j q_t^j \\ \text{s.t.} \quad & A_t x_t = b_t(\omega_t) - E_t x_{t-1}^i, \\ & \sum_{j=1}^{J_t} \lambda^j x_t^j = x_t, \\ & \lambda \in \Lambda_t. \end{aligned} \qquad (7)$$

and then computing $q_t^i = \mathbb{E}[q_t^i(\omega_t)]$.

5. Solve

$$\begin{aligned} y = \min \quad & c_1^\top x_1 + \sum_{j=1}^{J_1} \lambda^j q_2^j \\ \text{s.t.} \quad & A_1 x_1 = b_1, \\ & \sum_{j=1}^{J_1} \lambda^j x_1^j = x_1, \\ & \lambda \in \Lambda_1. \end{aligned} \qquad (8)$$

By Proposition 2, the optimal solution to (8) defines an upper bound y on the optimal value of (1). Observe that this upper bound is not statistical, but it is a valid upper bound on the optimal value of (1). Of course, if this finite problem is a sample average approximation of some "true" model, then we cannot assert that this bound is a valid upper bound on the policy as applied to the true problem.

It is tempting to suppose that the cost $\bar{z}$ incurred by a policy defined by outer approximation always satisfies $\bar{z} \leq y$, but this is not true in general. Suppose a candidate policy that is computed using some methodology (like outer approximation) is evaluated at stage t with a Bellman function that is an inner approximation of the expected future cost. If the expected future cost from the start of stage t using the candidate policy is not a convex function of the state then this makes it impossible to bound this cost using inner approximation.

To illustrate this, consider the simple example of a single reservoir supplying free water over two stages, where the demand in each period is 1 unit. One can buy water in stage 1 with cost 1, and can buy in stage 2 with cost 4. Inflows are zero. Let x_t denote the storage at the end of stage t. The optimal solution to this problem is to release $\max\{x_0 - 1, 0\}$ in stage 1, and x_1 in stage 2. The Bellman functions are therefore

$$Q_2(x_1) = \begin{cases} 4 - 4x_1, & x_1 < 1 \\ 0 & x_1 \geq 1 \end{cases}$$

$$Q_1(x_0) = \begin{cases} 5 - 4x_0, & x_0 < 1, \\ 2 - x_0, & 1 \leq x_0 < 2, \\ 0 & x_0 \geq 2. \end{cases}$$

Suppose we wish to estimate the actual value of a *myopic* policy that at any stage with reservoir level x releases $\min\{x, 1\}$. Such a policy could arise for example from a (poor) outer approximation with Bellman functions $\tilde{Q}_2(x) = \tilde{Q}_3(x) = 0$. Denote the actual cost of this policy from the end of stage t to the end of stage 2 by $v_t(x)$ where x is the storage. We can see that

$$v_1(x) = \begin{cases} 4 - 4x, & x < 1, \\ 0, & 1 \leq x, \end{cases}$$

$$v_0(x) = \begin{cases} 5 - x, & x < 1, \\ 8 - 4x, & 1 \leq x < 2, \\ 0 & x \geq 2. \end{cases}$$

Now consider an inner approximation of the Bellman function at the end of stage 1 using upper bounds W_1^1 and W_1^2 at two extreme storage levels $x_1^1 = 0$ and $x_1^2 = 2$. Suppose we choose $W_1^1 = v_1(0) = 4$, and $W_1^2 = v_1(2) = 0$. An inner approximation is then obtained by interpolation giving

$$\hat{Q}_2(x_1) = 4 - 2x_1,$$

which is a pointwise upper bound on $Q_2(x_1)$. The myopic policy evaluated with $\hat{Q}_2(x_1)$ gives expected future costs at storage levels $x_0^1 = 0$ and $x_0^2 = 2$. These are $W_0^1 = 1 + \hat{Q}_2(0) = 5$, and $W_0^2 = 0 + \hat{Q}_2(1) = 2$. The inner approximation at the start of stage 1 is then

$$\hat{Q}_1(x_0) = 5 - \frac{3}{2}x_0,$$

which is a pointwise upper bound on the cost $Q_1(x_0)$ of the optimal policy.

However we can observe that the actual cost $v_0(x_0)$ of the myopic policy starting with $x_0 = 1$ is $v_0(1) = 4$, which is larger than the cost $\hat{Q}_1(1) = \frac{7}{2}$ defined by the inner approximation. So the inner approximation cannot be used here to provide a pointwise

upper bound on the cost of a suboptimal policy, because this cost is not a convex function of x. If, on the other hand, the candidate policy were to *minimize* the objective of the problem at stage t using $\hat{\mathcal{Q}}_{t+1}(x_t)$ as a Bellman function, then the convexity of the (approximate) optimal value function means that $\hat{\mathcal{Q}}_t(x_{t-1})$ provides an upper bound on the expected value of the policy for any choice of x_{t-1}. We proceed to demonstrate this formally.

Consider a policy $\hat{\boldsymbol{\pi}}$ that is defined in terms of the set $\{(x_t^1, q_{t+1}^1), (x_t^2, q_{t+1}^2), \ldots, (x_t^{J_t}, q_{t+1}^{J_t})\}$ of points that define $\hat{\mathcal{Q}}_{t+1}(x)$. The policy $\hat{\boldsymbol{\pi}}$ defines an action $\hat{x}_t(x_{t-1}, \omega_t)$ to be taken at stage t at any state x_{t-1} and random outcome ω_t, which is the solution to

$$q_t(x_{t-1}, \omega_t) = \min \quad c_t^\top x_t + \hat{\mathcal{Q}}_{t+1}(x_t)$$
$$\text{s.t.} \quad A_t x_t = b_t(\omega_t) - E_t x_{t-1},$$
$$x_t \geq 0.$$

Observe that $\hat{Q}_t(x_{t-1}, \omega_t) = \sum_{j=1}^{J_{t-1}} \lambda^j q_t^j(\omega_t)$, where $x_{t-1} = \sum_{j=1}^{J_{t-1}} \lambda^j x_{t-1}^j$, so $\hat{Q}_t(x_{t-1}^j, \omega_t) = q_t^j(\omega_t)$, and by convexity of $q_t(x_{t-1}, \omega_t)$ we have

$$q_t(x_{t-1}, \omega_t) \leq \hat{Q}_t(x_{t-1}, \omega_t) \tag{9}$$

This enables us to show the following proposition.

Proposition 3 *The expected cost $v(\hat{\boldsymbol{\pi}})$ of policy $\hat{\boldsymbol{\pi}}$ when evaluated at x_1 is less than or equal to y.*

Proof. Let the action to be taken by policy $\hat{\boldsymbol{\pi}}$ in stage t be given by $\hat{x}_t(x_{t-1}, \omega_t)$, and let $\hat{v}_t(x_{t-1}, \omega_t)$ be the actual expected future cost of this and future actions over stages $t, t+1, \ldots, T$. We show by induction that for every x we have

$$\mathbb{E}[\hat{v}_t(x_{t-1}, \omega_t)] \leq \hat{\mathcal{Q}}_t(x_{t-1}).$$

When $t = T$, $\hat{x}_T(x_{T-1}, \omega_T)$ is the solution to

$$q_T(x_{T-1}, \omega_T) = \min \quad c_T^\top x_T + \hat{\mathcal{Q}}_{T+1}(x_T)$$
$$\text{s.t.} \quad A_T x_T = b_T(\omega_T) - E_T x_{T-1},$$
$$x_T \geq 0,$$

giving $\mathbb{E}[\hat{v}_T(x_{T-1}, \omega_T)] = \mathbb{E}[q_T(x_{T-1}, \omega_T)]$. Given an inner approximation policy defined by x_{T-1}^i, $i = 1, 2, \ldots, J_{T-1}$, any x_{T-1} can be written as

$$x_{T-1} = \sum_{i=1}^{J_{T-1}} \lambda^i x_{T-1}^i, \quad \lambda \in \Lambda_{T-1},$$

so we have

$$\begin{aligned} \mathbb{E}[\hat{v}_T(x_{T-1}, \omega_T)] &= \mathbb{E}[q_T(x_{T-1}, \omega_T)] \\ &\leq \mathbb{E}[\sum_{i=1}^{J_{T-1}} \lambda^i q_T(x_{T-1}^i, \omega_T)] \\ &= \hat{\mathcal{Q}}_T(x_{T-1}). \end{aligned}$$

11

Now suppose that for every x we have

$$\mathbb{E}[\hat{v}_{t+1}(x_t, \omega_{t+1})] \leq \hat{\mathcal{Q}}_{t+1}(x_t).$$

Recall $\hat{x}_t(x_{t-1}, \omega_t)$ is the optimal solution to

$$\begin{aligned}
q_t(x_{t-1}, \omega_t) = \min \quad & c_t^\top x_t + \hat{\mathcal{Q}}_{t+1}(x_t) \\
\text{s.t.} \quad & A_t x = b_t(\omega_t) - E_t x_{t-1}, \\
& x_t \geq 0.
\end{aligned}$$

The actual expected cost of the policy over stages $t, t+1, \ldots, T$ is then

$$\begin{aligned}
\mathbb{E}[\hat{v}_t(x_{t-1}, \omega_t)] &= \mathbb{E}[c_t^\top \hat{x}_t(x_{t-1}, \omega_t) + \mathbb{E}[\hat{v}_{t+1}(\hat{x}_t(x_{t-1}, \omega_t), \omega_{t+1}) \mid \omega_t]] \\
&\leq \mathbb{E}[c_t^\top \hat{x}_t(x_{t-1}, \omega_t) + \hat{\mathcal{Q}}_{t+1}(\hat{x}_t(x_{t-1}, \omega_t))] \\
&= \mathbb{E}[q_t(x_{t-1}, \omega_t)] \\
&\leq \hat{\mathcal{Q}}_t(x_{t-1})
\end{aligned}$$

where the first inequality follows from the inductive hypothesis, and the second inequality follows from (9).

Finally the expected cost of policy $\hat{\boldsymbol{\pi}}$ when evaluated at the solution $\hat{x}_1$ to (8) is

$$\begin{aligned}
v(\hat{\boldsymbol{\pi}}) &= c_1^\top x_1 + \mathbb{E}[\hat{v}_2(\hat{x}_1, \omega_2)] \\
&\leq c_1^\top \hat{x}_1 + \hat{\mathcal{Q}}_2(\hat{x}_1) \\
&= y
\end{aligned}$$

which gives the result. ∎

The performance of the inner approximation will depend critically on the choices of the points $x_t^1, x_t^2, \ldots, x_t^{J_t}$. In problems where x has high dimension, we cannot expect that the inner approximation will be very accurate for modest choices of J_t. Moreover to ensure that we can capture all feasible values of x, we should choose some of the J_t values to be extreme points of the range of possible x values. If $x \in \mathbb{R}^D$ then this means having $J_t \geq 2^D$, which grows very fast with D. On the other hand, we might use points $x_t^1, x_t^2, \ldots, x_t^{J_t}$ selected from those generated by some other algorithm, such as SDDP for example. This appears to give reasonable results in practice.

We conclude this section by summarizing the relationship between the bounds. Suppose that $\bar{z}$ is the actual expected value of the policy defined by the outer approximation, and $\hat{y}$ is the actual expected value of the policy defined by the inner approximation. Then we have $z \leq \bar{z}$ and $z \leq \hat{y} \leq y$. Observe that without estimating the values $\bar{z}$ and $\hat{y}$ we cannot deduce which is larger. We do have a bound on the actual expected value of the policy defined by the inner approximation, which can be used to estimate a gap between this value and z. This becomes an important issue in the general case where ρ_t is a general coherent risk measure for which we cannot estimate $\bar{z}$ or $\hat{y}$. We will return to discuss this after presenting the extension of the inner and outer aproximation algorithms to the case of general risk measures.

4 General coherent risk measures

In this section we repeat the above construction when expectation is replaced by a coherent risk measure. We begin by examining a two-stage model, where our notation will suppress the dependence on t. Recall that we restrict attention to finite probability distributions, so $\Omega = \{\omega_1, \omega_2, \ldots, \omega_M\}$ is finite. Our space of random variables can then be identified with $\mathbb{R}^M$, in the sense that Z has a finite number of outcomes $\{Z(\omega_1), Z(\omega_2), \ldots, Z(\omega_M)\}$ with probabilities $\{p_1, p_2, \ldots, p_M\}$.

Any coherent risk measure $\rho(Z)$ has a dual representation (see [1],[19]) expressing it as

$$\rho(Z) = \sup_{\zeta \in \mathfrak{A}} \sum_{m=1}^{M} p_m \zeta_m Z(\omega_m) \tag{10}$$

where $\mathfrak{A}$ is a convex subset of

$$\mathfrak{B} = \{\zeta \in \mathbb{R}^M : \sum_{m=1}^{M} p_m \zeta_m = 1, \zeta \geq 0\}.$$

In the special case where the risk measure is $\mathsf{CVaR}_{1-\alpha}[Z]$, we have

$$\mathfrak{A} = \{\zeta \in \mathfrak{B} \mid \zeta_m \leq \frac{1}{\alpha}, \ m = 1, 2, \ldots, M\}.$$

Now suppose that $Z(x, \omega)$ is a convex function of x for each $\omega \in \Omega$, and that when $x = \tilde{x}$,

$$\sup_{\zeta \in \mathfrak{A}} \sum_{m=1}^{M} p_m \zeta_m Z(\tilde{x}, \omega_m)$$

is attained at $\tilde{\zeta}$, say. Then we have the following result.

Proposition 4 *Suppose for each $\omega \in \Omega$, that $g(\tilde{x}, \omega)$ is a subgradient of $Z(x, \omega)$ at $\tilde{x}$. Then $\sum_{m=1}^{M} p_m \tilde{\zeta}_m g(\tilde{x}, \omega_m)$ is a subgradient of $\rho(Z(x, \omega))$ at $\tilde{x}$.*

Proof. For any x,

$$\begin{aligned}
\rho(Z(x, \omega)) &= \sup_{\zeta \in \mathfrak{A}} \sum_{m=1}^{M} p_m \zeta_m Z(x, \omega_m) \\
&\geq \sum_{m=1}^{M} p_m \tilde{\zeta}_m Z(x, \omega_m) \\
&\geq \sum_{m=1}^{M} p_m \tilde{\zeta}_m (Z(\tilde{x}, \omega_m) + g(\tilde{x}, \omega_m)^\top (x - \tilde{x})) \\
&= \rho(Z(\tilde{x}, \omega)) + \sum_{m=1}^{M} p_m \tilde{\zeta}_m g(\tilde{x}, \omega_m)^\top (x - \tilde{x})
\end{aligned}$$

13

which demonstrates that $\sum_{m=1}^{M} p_m \tilde{\zeta}_m g(\tilde{x}, \omega_m)$ is a subgradient at $\tilde{x}$. ∎

Now, in a multistage context, given x_{t-1} we compute an optimal solution $x_t(\omega)$ to a stage problem for each $\omega \in \Omega_t$ to yield an optimal value $c_t x_t(\omega_t) + \rho_{t+1}(Q_{t+1}(x_t(\omega_t), \omega_{t+1}))$, $\omega_t \in \Omega_t$. Since ρ_t is a coherent measure, it is monotone. So by the interchangeability principle [19, Proposition 6.36] we can compute

$$\min_{x_t(\omega)} \rho_t(c_t x_t(\omega_t) + \rho_{t+1}(Q_{t+1}(x_t(\omega_t), \omega_{t+1})))$$

by evaluating

$$\rho_t(\min_{x_t}\{c_t x_t + \rho_{t+1}(Q_{t+1}(x_t, \omega_{t+1}))\}).$$

This evaluation is then equivalent to

$$\sup_{\zeta \in \mathfrak{A}} \sum_{m=1}^{M} p_m \zeta_m \min_{x_t}\{c_t x_t + \rho_{t+1}(Q_{t+1}(x_t, \omega_{t+1}))\}$$

which is effectively a wait-and-see computation, followed by an expectation with a "changed" probability measure. In contrast to the use of a polyhedral formula for the risk measure (as used for CVaR in [13]), this construction simplifies the computation considerably. This observation was made in [20].

We can demonstrate the advantages of this "change-of-measure" approach by applying it to the two-stage model from [13]:

$$\text{SP: } \min \quad c_1^\top x_1 + (1-\lambda)\mathbb{E}[c_2^\top x_2] + \lambda \text{CVaR}_{1-\alpha}[c_2^\top x_2]$$
$$\text{s.t.} \quad A_1 x_1 = b_1,$$
$$A_2 x_2(\omega) + E_2 x_1 = b_2(\omega), \qquad \text{for all } \omega \in \Omega_2,$$
$$x_1 \geq 0, \qquad x_2(\omega) \geq 0, \qquad \text{for all } \omega \in \Omega_2.$$

written as the following linear program

$$\text{SP: } \min \quad c_1^\top x_1 + (1-\lambda)\mathbb{E}[c_2^\top x_2] + \lambda u_2 + \lambda \alpha^{-1}\mathbb{E}[v_2]$$
$$\text{s.t.} \quad A_1 x_1 = b_1,$$
$$A_2 x_2(\omega) + E_2 x_1 = b_2(\omega), \qquad \text{for all } \omega \in \Omega_2,$$
$$v_2(\omega) \geq c_2^\top x_2(\omega) - u_2, \qquad \text{for all } \omega \in \Omega_2,$$
$$x_1 \geq 0, \qquad x_2(\omega) \geq 0, \qquad v_2(\omega) \geq 0, \qquad \text{for all } \omega \in \Omega_2.$$

where x_1 and u_2 are first-stage variables. In the multistage context, the approach of [13] records x_{t-1} and u_t as state variables, and computes cutting planes as functions of these.

In the "change-of-measure" approach, we maintain only x_{t-1} as a state variable. In the two-stage context this would entail that x_1 is fixed. We then solve

$$\text{SP}(\omega): \min \quad c_2^\top x_2$$
$$\text{s.t.} \quad A_2 x_2 = b_2(\omega) - E_2 x_1$$
$$x_2 \geq 0,$$

for each ω to yield $x_2(\omega)$, and $Z(\omega_m) = c_2^\top x_2(\omega_m)$. We then compute the probabilities $p_1 \zeta_1, p_2 \zeta_2, \ldots, p_M \zeta_M$ that correspond to $Z(\omega_1), Z(\omega_2), \ldots, Z(\omega_M)$. Here

$$\rho(Z) = (1-\lambda) \sum_{m=1}^{M} p_m Z(\omega_m) + \lambda \sup_{\zeta \in \mathfrak{A}} \sum_{m=1}^{M} p_m \zeta_m Z(\omega_m),$$

where

$$\mathfrak{A}=\{\zeta \in \mathfrak{B} \mid \zeta_m \leq \frac{1}{\alpha}, \ m = 1, 2, \ldots, M\}.$$

This is a straightforward calculation once $Z(\omega_m)$ is known. Without loss of generality suppose $Z(\omega_1) \leq Z(\omega_2) \leq \ldots \leq Z(\omega_M)$. Define index i so that

$$\sum_{m=i}^{M} p_m \geq \alpha, \quad \sum_{m=i+1}^{M} p_m < \alpha$$

where we choose $i = M$ if $p_M \geq \alpha$. Then

$$\rho(Z) = \sum_{m=1}^{M} p_m \zeta_m Z(\omega_m)$$

where

$$\zeta_m = \begin{cases} (1-\lambda), & m < i \\ (1-\lambda) + \frac{1}{p_i}\left(\lambda - \frac{\lambda}{\alpha}\sum_{n=i+1}^{M} p_n\right), & m = i \\ (1-\lambda) + \frac{\lambda}{\alpha}, & m > i \end{cases}.$$

This construction avoids the necessity to maintain a VaR state variable u at each stage, at the expense of having to compute ζ. Observe also that ζ will vary with the value of the state variable x_1, and so we cannot reduce the overall problem to the minimization of an expectation - it is only possible to do this at each stage t given the values of the state variables x_{t-1}.

For general coherent risk measures, the construction is almost identical to that above. The optimal value of (5), the tth stage problem in the outer approximation algorithm, can be represented at any $x_{t-1}$by

$$
\begin{aligned}
\tilde{Q}_t(x_{t-1}, \omega_t) = \min \quad & c_t^\top x_t + \theta_{t+1} \\
\text{s.t.} \quad & A_t x_t = b_t(\omega_t) - E_t x_{t-1}, & [\pi_t(\omega_t)] \\
& \theta_{t+1} + \bar{\pi}_{t+1,k}^\top E_{t+1} x_t \geq \bar{h}_{t+1,k}, & k = 1, 2, \ldots, K_{t+1}, \\
& x_t \geq 0.
\end{aligned}
\tag{11}
$$

where k counts the cuts that are added to the tth-stage problem,

$$\bar{\pi}_{t+1,k} = \sum_{m=1}^{M} p_m \tilde{\zeta}_m \pi_{t+1,k}(\omega_{t+1,m}), \tag{12}$$

$$\bar{h}_{t+1,k} = \sum_{m=1}^{M} p_m \tilde{\zeta}_m \tilde{Q}_{t+1}(x_t^k, \omega_{t+1,m}) + \bar{\pi}_{t+1,k}^\top E_{t+1} x_t^k,$$

and $\tilde{\zeta}$ is defined by the collection of solution values $\tilde{Q}_{t+1}(x_t^k, \omega_{t+1})$, $\omega_{t+1} \in \Omega_{t+1}$ and the set $\mathfrak{A}$ that characterizes the particular coherent risk measure we are using. More precisely, $\tilde{\zeta}$ is chosen to maximize $\sum_{m=1}^{M} p_m \zeta_m \tilde{Q}_{t+1}(x_t^k, \omega_{t+1,m})$ over $\mathfrak{A}$. Finally Proposition 4 shows that $\bar{\pi}_{t+1,k}^\top$ defined by (12) gives a subgradient $-\bar{\pi}_{t+1,k}^\top E_{t+1}$ of $\rho_{t+1}(\tilde{Q}_{t+1}(x_t, \omega_{t+1}))$ at x_t^k, so the inequalities in (11) define valid cutting planes for the outer approximation.

We now return to the issue of constructing bounds on the risk-adjusted cost $v(\boldsymbol{\pi})$ of any risk-averse policy $\boldsymbol{\pi}$ that is feasible for the problem defined by (1) and (2). We know that outer approximation produces a lower bound z on $v(\boldsymbol{\pi})$, and a policy $\tilde{\boldsymbol{\pi}}$ defined by a set of Bellman functions $\tilde{\mathcal{Q}}_t(x)$ for $t = 1, 2, \ldots, T$. Inner approximation produces an upper bound y on the value $v(\hat{\boldsymbol{\pi}})$ of the policy $\hat{\boldsymbol{\pi}}$ defined by the set of Bellman functions $\hat{\mathcal{Q}}_t(x)$ for $t = 1, 2, \ldots, T$.

The inductive argument in Proposition 3 extends naturally to the case where expectation is replaced by a more general risk measure. Here we use upper bounds $q_{t+1}^1, q_{t+1}^2, \ldots, q_{t+1}^{J_t}$ on the values of $\rho_{t+1}(Q_{t+1}(x_t, \omega_{t+1}))$ at J_t points $x_t^1, x_t^2, \ldots, x_t^{J_t}$. As before we denote the inner approximation using these bounds by $\hat{\mathcal{Q}}_{t+1}(x_t)$.

Proposition 5 *The risk-adjusted cost $v(\hat{\boldsymbol{\pi}})$ of policy $\hat{\boldsymbol{\pi}}$ when evaluated at x_1 is less than or equal to y.*

Proof. Let the action to be taken by policy $\hat{\boldsymbol{\pi}}$ in stage t be given by $\hat{x}_t(x_{t-1}, \omega_t)$, and let $\hat{v}_t(x_{t-1}, \omega_t)$ be the actual risk-adjusted future cost of this and future actions over stages $t, t+1, \ldots, T$. Thus

$$\hat{v}_t(x_{t-1}, \omega_t) = c_t^\top x_t + \rho_{t+1}(\hat{v}_{t+1}(x_t, \omega_{t+1}))$$

We show by induction that for every x_{t-1} we have

$$\rho_t(\hat{v}_t(x_{t-1}, \omega_t)) \leq \hat{\mathcal{Q}}_t(x_{t-1}).$$

When $t = T$, $\hat{x}_T(x_{T-1}, \omega_T)$ is the solution to

$$
\begin{aligned}
q_T(x_{T-1}, \omega_T) = \min \quad & c_T^\top x_T + 0 \\
\text{s.t.} \quad & A_T x_T = b_T(\omega_T) - E_T x_{T-1}, \\
& x_T \geq 0,
\end{aligned}
$$

giving $\rho_T(\hat{v}_T(x_{T-1}, \omega_T)) = \rho_T(q_{T-1}(x_{T-1}, \omega_T))$. Given an inner approximation policy defined by x_{T-1}^i, $i = 1, 2, \ldots, J_{T-1}$, any x_{T-1} can be written as

$$x_{T-1} = \sum_{i=1}^{J_{T-1}} \lambda^i x_{T-1}^i, \quad \sum_{i=1}^{J_{T-1}} \lambda^i = 1, \quad \lambda^i \geq 0,$$

so for each ω_T we have

$$
\begin{aligned}
\rho_T(\hat{v}_T(x_{T-1}, \omega_T)) &= \rho_T(q_T(x_{T-1}, \omega_T)) \\
&\leq \rho_T\left(\sum_{i=1}^{J_{T-1}} \lambda^i q_T(x_{T-1}^i, \omega_T)\right) \\
&= \hat{\mathcal{Q}}_T(x_{T-1}),
\end{aligned}
$$

where the inequality follows from the convexity of $q_T(x_{T-1}, \omega_T)$ and convexity and monotonicity of ρ_T.

16

Now suppose that for every x_t we have

$$\rho_{t+1}(\hat{v}_{t+1}(x_t, \omega_{t+1})) \leq \hat{\mathcal{Q}}_{t+1}(x_t).$$

Recall $\hat{x}_t(x_{t-1}, \omega_t)$ is the solution to

$$\begin{align}
\min \quad & c_t^\top x_t + \hat{\mathcal{Q}}_{t+1}(x_t) \\
\text{s.t.} \quad & A_t x_t = b_t(\omega_t) - E_t x_{t-1}, \\
& x_t \geq 0.
\end{align} \tag{13}$$

The actual risk-adjusted cost of the policy over stages $t, t+1, \ldots, T$ is then

$$\begin{align}
\rho_t(\hat{v}_t(x_{t-1}, \omega_t)) &= \rho_t(c_t^\top \hat{x}_t(x_{t-1}, \omega_t) + \rho_{t+1}(\hat{v}_{t+1}(\hat{x}_t(x_{t-1}, \omega_t), \omega_{t+1}) \mid \omega_t)) \\
&\leq \rho_t(c_t^\top \hat{x}_t(x_{t-1}, \omega_t) + \hat{\mathcal{Q}}_{t+1}(\hat{x}_t(x_{t-1}, \omega_t))) \\
&\leq \hat{\mathcal{Q}}_t(x_{t-1})
\end{align}$$

where the first inequality is the inductive hypothesis, and the second inequality follows from the fact that $\hat{\mathcal{Q}}_t(x_{t-1})$ is an inner approximation of the risk-adjusted optimal value function of (13), and ρ_t is monotone.

Finally the risk-adjusted optimal cost of policy $\hat{\boldsymbol{\pi}}$ when evaluated at the solution $\hat{x}_1$ to (8) is

$$\begin{align}
v(\hat{\boldsymbol{\pi}}) &= c_1^\top \hat{x}_1 + \rho_2(\hat{v}_2(\hat{x}_1, \omega_2)) \\
&\leq c_1^\top \hat{x}_1 + \hat{\mathcal{Q}}_2(\hat{x}_1) \\
&= y,
\end{align}$$

which gives the result. ∎

5 Sampling algorithms with stage-wise independence

We have shown that the methods described in Section 3 can be used under a general risk measure to compute policies that will lead to an upper or lower bound to problem (1). However, in several practical applications it is impossible to build a policy considering the whole scenario tree, due to its size. The Stochastic Dual Dynamic Programming (SDDP) algorithm ([12],[18],[20]) for multistage stochastic linear programming attempts to overcome this problem by sampling.

The SDDP algorithm performs a sequence of major iterations known as the *forward pass* and the *backward pass* to build an outer approximation of the Bellman function at each stage. In each forward pass, a set of N scenarios is sampled from the scenario tree and decisions are taken for each node of those N scenarios, starting in the first stage and moving forward up to the last stage. In each stage, the observed values of the decision variables x_t, and the costs of each node in all scenarios are saved.

In the backward pass SDDP amends the current policy by adding cutting planes to each stage problem, starting at the last stage and working backwards to the first. In each stage t we solve the next stage problems for all possible realizations (Ω_{t+1}). The

values of the objective functions and dual variables at optimality are averaged over all realizations to define a cut that is added to all problems at stage t. Under general risk measures, SDDP is essentially the same as the risk-neutral method, with differences in cut computation. In order to compute a cut it is necessary to add a step in the backward pass which calculates the "change-of-measure" probabilities as discussed in Section 4. The cut computation proceeds using these probabilities as in (12). The SDDP algorithm is then as follows.

1. Set $it = 0$

2. Sample N scenarios;

3. Forward Pass

   For $t = 1$, solve (11) and save $x_1(it)$ and z;

   For $t = 2, ..., T$ and $s = 1, ..., N$,

   Solve (11), where ω_t is defined by s, and save $x_t(s, it)$ and $\tilde{Q}_t(x_{t-1}(s, it), \omega_t)$.

4. Backward Pass

   For $t = T, ..., 2$, and $s = 1, ..., N$,

   For $\omega_{t,m} \in \Omega_t$, solve (11) using $x_{t-1}(s, it)$ and save $\pi_t(\omega_{t,m})$
   and $\tilde{Q}_t(x_{t-1}(s, it), \omega_{t,m})$;

   Compute $\tilde{\zeta} \in \mathfrak{A}$ that maximizes $\sum_{m=1}^{M} p_m \zeta_m \tilde{Q}_t(x_{t-1}(s, it), \omega_{t,m})$

   Calculate a cut using (12) for $k = K + s$, and add it to all nodes in stage $t - 1$.

   Set $K = K + N$.

5. Increment it. If $it < it^{\max}$, go to step 2. Otherwise, stop.

After the SDDP method has performed $it^{\max}$ iterations, we compute an upper bound on the optimal policy as follows:

Upper Bound Computation (U)

1.   For $t = T, ..., 2$, and $s = 1, ..., N$, and $it = 1, ..., it^{\max}$,

   For $\omega_{t,m} \in \Omega_t$, solve (7) using $x_{t-1}(s, it)$ and save $q_t(x_{t-1}(s, it), \omega_{t,m})$;

   Compute $\tilde{\zeta} \in \mathfrak{A}$ that maximizes $\sum_{m=1}^{M} p_m \zeta_m q_t(x_{t-1}(s, it), \omega_{t,m})$ and save the optimal value as $q_t(s, it)$;

2.   Solve (8) and save y;

Recall that the value of y computed is an upper bound on the optimal value, but not a bound on the value of the policy obtained in the outer approximation problem. It is however an upper bound on the value of the policy defined by the inner approximation.

6 Numerical experiments

In this section we present some numerical results from applying inner and outer approximation algorithms to a long-term hydrothermal scheduling problem arising in the Brazilian electricity system [3]. The hydrothermal scheduling problem seeks a policy for managing hydroelectricity reservoir storage to minimize thermal and shortage costs given uncertainty in the inflows. It can be viewed as a stochastic control problem in which the reservoir levels are the state variables and the controls are water releases, thermal generation, and load shedding. The model we consider in this paper represents the Brazilian national system comprising 158 hydroelectric power plants and 151 thermal power plants (as at January 2012). The hydroelectric power plants are aggregated in each region - South, Southeast, North and Northeast - to form four *energy equivalent reservoirs* (EERs) as discussed in [10, 11], giving a state space of dimension four.

Long-term hydrothermal scheduling problems in Brazil typically consider a 10-year horizon with 120 stages. To reduce the computational effort, we build a policy for a reduced horizon of 2 years with monthly time stages, giving a total of 24 stages. We assume that all EERs have initial storage of 15% of the maximum storage, except for the Northern EER which has 25% of its maximum storage. The only uncertainty in our model is in the inflows, which we assume to be stagewise independent.

The inflow model is a stagewise independent lognormal model based on that described in [3]. Following [9], the spatial correlation between the inflows to the four EERs is modelled by a matrix transformation of four independent time series, to give the same covariance matrix as that estimated from a historical series of monthly data going back 80 years. Each of the four factors are assumed to be stage-wise independent with a different lognormal distribution for each factor and each month, estimated from the 80 years of monthly inflow data.

In this paper we analyse two cases, one risk neutral and one risk averse. For each case, we build a policy with up to 10,000 cuts/states for all 24 stages. In the risk-averse case we used the risk measure defined by (3) with $\lambda = 0.5$ and $\alpha = 0.2$. The scenario tree for a sample average approximation (SAA) problem is created by means of randomly sampling 20 vectors from the inflow distribution at each stage.

The risk-neutral policies are evaluated using both bounds and statistical estimation. For the latter, the policies were simulated over the 24 month horizon for 10,000 scenarios. These scenarios were obtained by randomly sampling outcomes within the scenario tree. Thus our experiments investigate the solution of the SAA problem and how close it is to optimality. We do not evaluate the policy when applied to the original model.

In the risk-neutral case, Table 1 presents the lower and upper bounds obtained with the outer and inner approximation, respectively, for several numbers of cuts/states. The outer approximation algorithm is a standard variation of SDDP with 200 scenarios used in each forward pass. We run each inner approximation algorithm fifteen times, using the states visited after 200, 400, 600, 800, 1000, 1600, 2000, 3000, . . . , 10000 cuts have been added in the outer approximation. In addition to the bounds, Table 1 shows the optimality gap for the SAA problem and the computational time for both algorithms. From Table 1 it is possible to see that the gap between the bounds reduces as we increase the number of cuts in the outer approximation and states in the inner approximation. One can also

Table 1: Risk-neutral results.

Cuts States	Lower Bound (10^9 BRL)	Upper Bound (10^9 BRL)	Gap (10^9BRL)	Gap (%)	Time Outer (s)	Time Inner (s)
200	5.436	6.430	0.994	15.46	5	25
400	5.636	6.192	0.556	8.99	10	52
600	5.729	6.126	0.397	6.48	15	78
800	5.762	6.090	0.328	5.39	21	118
1000	5.787	6.074	0.288	4.74	28	167
1600	5.839	6.033	0.195	3.23	54	356
2000	5.849	6.020	0.171	2.84	77	514
3000	5.868	5.993	0.125	2.08	151	1056
4000	5.879	5.983	0.104	1.73	248	1824
5000	5.887	5.977	0.090	1.51	368	2629
6000	5.891	5.971	0.080	1.34	516	3867
7000	5.896	5.966	0.071	1.19	696	5213
8000	5.899	5.964	0.065	1.10	901	6915
9000	5.901	5.961	0.061	1.02	1140	8921
10000	5.903	5.959	0.056	0.95	1408	13401

notice that the outer approximation is much quicker than the inner approximation, owing to the use of acceleration strategies that have been shown to improve the performance of SDDP [4]. We have not developed analogous strategies for inner approximation.

In order to compare with the traditional strategies for estimating an upper bound when minimizing expected cost, Table 2 shows statistical upper bounds. The values in the second column (UB SDDP) are computed using the 200 scenarios sampled in each forward pass of SDDP. The value reported is the sample average cost plus twice its standard error (giving a one-sided test for 97.5% confidence). This is an appropriate convergence test as discussed in [18] and [7]. The upper bound values in the fourth and sixth columns are computed using a single set of 10,000 scenarios (sampled independently from the forward passes, and re-used for each estimation).

The estimates in the second column of Table 2 are relatively coarse since the standard error of the estimator is large with only 200 scenarios. With a larger sample size these values reduce to be very close to the lower bound. The results confirm that for sufficient numbers of states/cuts the expected cost of the inner-approximation policy and that of the outer-approximation policy are of comparable value. When comparing the statistical estimates in Table 2 with the values of Table 1 one can see these are slightly smaller than the deterministic upper bounds obtained from the inner approximation. This is to be expected as these are estimates of the expected cost of the solution, whereas the deterministic values are upper bounds on this value.

Table 3 shows the bounds when the methodology is applied to the risk-averse case. One can observe a similar behaviour to the risk-neutral case, but with larger gaps for the same number of cuts/states. In general, risk-averse problems need more cuts and states to specify close to optimal policies, so we might expect to see these differences

Table 2: Risk-neutral statistical upper bounds.

Cuts States	UB SDDP Algorithm (10^9 BRL)	Gap SDDP Algorithm (%)	UB Outer App. Simulation (10^9 BRL)	Gap Outer App. Simulation (%)	UB Inner App. Simulation (10^9 BRL)	Gap Inner App. Simulation (%)
200	14.339	72.60	6.510	16.72	6.061	10.40
400	8.518	40.24	6.161	8.63	5.987	5.93
600	6.628	14.58	6.119	6.46	5.967	4.03
800	6.336	9.73	6.069	5.13	5.958	3.33
1000	5.857	1.27	6.040	4.25	5.953	2.82
1600	6.255	7.20	5.979	2.38	5.938	1.69
2000	6.194	5.85	5.973	2.09	5.937	1.49
3000	6.885	16.02	5.954	1.45	5.931	1.06
4000	6.025	2.55	5.940	1.03	5.928	0.83
5000	6.647	12.33	5.938	0.88	5.925	0.66
6000	6.705	13.39	5.929	0.65	5.923	0.55
7000	6.294	6.76	5.927	0.54	5.922	0.45
8000	6.313	7.00	5.926	0.46	5.921	0.38
9000	6.244	5.85	5.920	0.33	5.919	0.32
10000	6.208	5.30	5.920	0.30	5.920	0.28

when compared with the risk-neutral policy. Adding more cuts to give a better policy will reduce the gap, but, as we can see, the time to compute an upper bound using our approach will become prohibitive. We expect the computational effort to be reduced by acceleration strategies, though we have not studied these for inner approximation. For outer approximation, these strategies make a substantial difference, even leading to faster solves in the risk-averse case as compared with corresponding times for the risk-neutral case, where "Level 1 dominance" as defined in [4] results in a smaller number of selected cuts in the risk-averse stage problems.

7   Conclusions

This paper shows how SDDP can be implemented with a general coherent risk measure. We show how upper bounds on the value of candidate policies can be computed in this setting, and used to evaluate how close these policies are to the optimal solution. The performance of these bounds has been tested in a large-scale instance of a hydro-thermal scheduling problem. These experiments are a first step towards a more efficient bounding procedure, and give researchers a target at which to aim.

The upper bounds that we obtain are not statistical - they are deterministic upper bounds on the policy value as computed for the scenario tree that we are working with. Of course, if this tree is obtained by sampling then the bounds on the policy value when tested in the "true" process become statistical. This might appear to confer a substantial advantage over statistical bounds obtained using Monte Carlo methods. Observe however

Table 3: Risk-averse results.

Cuts States	Lower Bound (10^9 BRL)	Upper Bound (10^9 BRL)	Gap (10^9 BRL)	Gap (%)	Time Outer (s)	Time Inner (s)
200	17.387	27.886	10.498	37.65	5	23
400	18.465	26.472	8.007	30.25	9	50
600	18.942	26.061	7.119	27.32	14	99
800	19.112	25.783	6.671	25.87	19	126
1000	19.247	25.687	6.440	25.07	25	158
1600	19.718	25.249	5.531	21.91	47	416
2000	19.804	25.113	5.309	21.14	63	604
3000	20.457	24.677	4.220	17.10	109	1017
4000	20.561	24.559	3.998	16.28	169	2159
5000	20.641	24.470	3.829	15.65	243	2979
6000	20.749	24.357	3.609	14.82	334	3879
7000	20.905	24.255	3.350	13.81	454	6020
8000	20.983	24.229	3.246	13.40	593	8078
9000	21.040	24.161	3.121	12.92	762	10386
10000	21.081	24.132	3.051	12.64	959	13417

that the computational effort required for an accurate inner approximation increases dramatically with the state dimension, a problem that is not faced by Monte Carlo sampling.

We have applied inner approximation to the Bellman function at every stage. Along with the outer approximation defined by the cuts, this gives two bounding functions on the optimal Bellman function. We expect that the difference between these will decrease as we approach the end of the horizon. This measure can be used to determine the stages at which the outer approximation might need improving, and provide an opportunity to avoid unnecessary cut calculations in some stages as the algorithm proceeds.

The axioms that define coherent risk measures are not universally accepted. In particular, positive homogeneity does not represent the view that risk increases nonlinearly as positions scale. The models that we study here could be extended to *convex* risk measures [6] which replace the subadditivity and positive homogeneity conditions with convexity. Convexity, monotonicity and translation equivariance are enough to give the nested structure we require. The representation theorem for convex measures gives rise to a dual problem that seeks a worst-case measure with the addition of a penalty function of the measure. If the risk measure we adopt enables the optimal dual measure to be readily identifiable from the primal solution to each stage problem, then the approach we describe in this paper can be extended to deal with convex risk measures, by shifting the cutting planes vertically by an appropriate penalty distance. In addition to that, the inner approximation and upper bound computation remain valid, as we rely only on convexity.

References

[1] P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath. Coherent measures of risk. *Mathematical Finance*, 9:203–228, 1999.

[2] D. R. Carino, T. Kent, D. H. Myers, C. Stacy, M. Sylvanus, A. L. Turner, K. Watanabe, and W. T. Ziemba. The Russell-Yasuda Kasai model: An asset/liability model for a Japanese insurance company using multistage stochastic programming. *Interfaces*, 24(1):29–49, 1994.

[3] V.L. de Matos and E. Finardi. A computational study of a stochastic optimization model for long term hydrothermal scheduling. *International Journal of Electrical Power and Energy Systems*, 2012.

[4] V.L. de Matos, A.B. Philpott, and E. Finardi. Improving the performance of stochastic dual dynamic programming. *European Journal of Operational Research*, 2012.

[5] B.H. Dias, A.L.M. Marcato, R.C. Souza, M.P. Soares, I.C. Silva Junior, E.J. de Oliveira, R.B.S. Brandi, and T.P. Ramos. Stochastic dynamic programming applied to hydrothermal power systems operation planning based on the convex hull algorithm. *Mathematical Problems in Engineering*, 2010:20, 2010.

[6] H. Follmer and A. Schied. Convex measures of risk and trading constraints. *Finance and Stochastics*, 6:429–447, 2002.

[7] T. Homem-de Mello, V.L. de Matos, and E.C. Finardi. Sampling strategies and stopping criteria for stochastic dual dynamic programming: a case study in long-term hydrothermal scheduling. *Energy Systems*, 2:1–31, 2011.

[8] J. E. Kelley Jr. The cutting-plane method for solving convex programs. *Journal of the Society for Industrial and Applied Mathematics*, 8(4):703–712, 1960.

[9] M.E.P. Maceira, V.S. Duarte, D.D.J. Penna, L.A.M. Moraes, and A.C.G. Melo. Ten years of application of Stochastic Dual Dynamic Programming in official and agent studies in Brazil: description of the NEWAVE program. In *Proceedings of the 16th PSCC, Glasgow, Scotland.*, 2008.

[10] M.E.P. Maceira, C.B. Mercio, B.G. Gorenstin, S.F.H. Cunha, C. Suanno, M.C. Sacramento, and A.S. Kligerman. Energy evaluation of the north/northeastern and south/southeastern interconnection with newave model. In *VI Symposium of Specialists in Electric Operational and Expansion Planning - SEPOPE*, 1998.

[11] M.E.P. Maceira, L.A. Terry, F.S. Costa, J.M. Damazio, and A.C.G. Melo. Chain of optimization models for setting the energy dispatch and spot price in the brazilian system. In *14th Power Systems Computation Conference 2002*, 2002.

[12] M. V. F. Pereira and L. M. V. G. Pinto. Multi-stage stochastic optimization applied to energy planning. *Mathematical Programming*, 52:359–375, 1991.

[13] A.B. Philpott and V.L. de Matos. Dynamic sampling algorithms for multi-stage stochastic programs with risk aversion. *European Journal of Operational Research*, 218(2):470 – 483, 2012.

[14] F. Riedel. Dynamic coherent risk measures. *Stochastic Processes and their Applications*, 112(2):185–200, 2004.

[15] A. Ruszczyński. Risk-averse dynamic programming for Markov decision processes. *Mathematical Programming*, 125:235–261, 2010.

[16] C. Sagastizabal and V. Guigues. Risk averse feasible policies for large-scale multistage stochastic linear programs. 2011. http://www.optimization-online.org/DBFILE/2009/02/2215.pdf.

[17] A. Shapiro. On complexity of multistage stochastic programs. *Operations Research Letters*, 34(1):1 – 8, 2006.

[18] A. Shapiro. Analysis of stochastic dual dynamic programming method. *European Journal of Operational Research*, 209(1):63 – 72, 2011.

[19] A. Shapiro, D. Dentcheva, and A. Ruszczyński. *Lectures on stochastic programming: modeling and theory*. SIAM, 2009.

[20] A. Shapiro, W. Tekaya, J.P. da Costa, and M.P. Soares. Risk neutral and risk averse stochastic dual dynamic programming method. 2012.

[21] S. W. Wallace and W. T. Ziemba. *Applications of Stochastic Programming (MPS-SIAM series on Optimization)*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2005.