Mathematical Programming 75 (1996) 295–325

Implementing bounds-based approximations in convex-concave two-stage stochastic programming

N.C.P. Edirisinghe [a,*], W.T. Ziemba [b]

[a] *Management Science Program, College of Business Administration, University of Tennessee, Knoxville, TN 37996, USA*

[b] *Management Science Division, Faculty of Commerce and Business Administration, University of British Columbia, Vancouver, BC, Canada V6T 1Y8*

Received 1 August 1993; revised manuscript received 1 September 1994

Abstract

This paper is concerned with implementational issues and computational testing of bounds-based approximations for solving two-stage stochastic programs with fixed recourse. The implemented bounds are those derived by the authors previously, using first and cross moment information of the random parameters and a convex-concave saddle property of the recourse function. The paper first examines these bounds with regard to their tightness, monotonic behavior, convergence properties, and computationally exploitable decomposition structures. Subsequently, the bounds are implemented under various partitioning/refining strategies for the successive approximation. The detailed numerical experiments demonstrate the effectiveness in solving large scenario-based two-stage stochastic optimization problems through *successive scenario clusters* induced by refining the approximations.

Keywords: Stochastic programs with recourse; Bounds on expectation of saddle functions; Partitioning strategies for rectangles; L-shaped decomposition

1. Introduction

In many real-world sequential decision making problems, given a set of possible alternative actions available at the present time, one chooses a decision in the face of uncertainty of the future states of nature. Suppose the future uncertain events are modeled using random variables and the decision problem is posed in an optimization framework. This yields a stochastic optimization model, the solution of which entails

* Corresponding author. e-mail: pa144918@utkvm1.utk.edu.

PII S0025-5610(96)00017-2

choosing a feasible decision at the outset which would hedge well against all possible realizations of the random parameters with respect to an objective criterion such as minimizing total expected costs.

The focus here is on the following stochastic programming model which involves making decisions in two time periods, known as the two-stage stochastic linear programming problem with *fixed* recourse (see for instance, Kall [13] and Wets [18]):

$$\begin{aligned} Z^* := \inf_x \quad & c'x + E_\omega[\phi(x, \omega)] \\ \text{s.t.} \quad & Ax = b, \\ & x \geqslant 0, \end{aligned} \tag{1}$$

where the second-stage recourse function is given by

$$\begin{aligned} \phi(x, \omega) \equiv \phi(x, \xi, \eta) := \inf_y \quad & q(\eta)' y \\ \text{s.t.} \quad & Wy = h(\xi) - T(\xi)x, \\ & y \geqslant 0. \end{aligned} \tag{2}$$

The domain of the random $(K+L)$-vector $\omega := (\xi, \eta)$ is denoted by $\Omega (\subset \mathbb{R}^{K+L})$; marginal support of ξ is $\Xi (\subset \mathbb{R}^K)$ and that of η is $\Theta (\subset \mathbb{R}^L)$. We assume that $\Omega = \Xi \times \Theta$. A is a fixed $(m_1 \times n_1)$-matrix and c is a fixed vector in $\mathbb{R}^{n_1}$. Stochasticity is present in the technology matrix $T: \Xi \to \mathbb{R}^{m_2 \times n_1}$, right-hand side $h: \mathbb{R}^K \to \mathbb{R}^{m_2}$, and objective costs $q: \mathbb{R}^L \to \mathbb{R}^{n_2}$. $E[\cdot]$ represents mathematical expectation with respect to the true probability measure P^{tr} on Ω. The matrix W $(\in \mathbb{R}^{m_2 \times n_2})$ is deterministic, hence the term *fixed* recourse for (1)–(2).

In (1), the first period decisions x must be chosen prior to the realizations of the random events $\omega := (\xi, \eta)$. However, after the anticipative decisions x are chosen, an adaptive (or second stage recourse) decision y may be chosen based on the realized outcome of the random parameters ω to bring the system close to a specified description, described by the constraints in (2), at a minimum cost. Consequently, one needs to determine an anticipative decision x which leads to the optimal trade-off between the first period direct costs and the second period expected costs. This latter cost estimation makes the solution of (1) difficult. The random parameters may be described by a set of scenarios, which may possibly be very large in number, or by continuous probability distributions. In the former case, (1) may be written in so-called extensive form involving a very large number of variables and constraints. In the latter case, one generally resorts to nonlinear programming approaches with multidimensional integration or quasi-gradient methods with sampling from the distribution; see for example, Ermoliev and Wets [8] and the references therein.

Another promising approach for solving (1) is the use of approximations. For an illustrative survey, see Frauendorfer [10] and the references therein. These methods advocate approximating either the probability distribution or the objective function itself, or both. The aim is to obtain deterministic optimization problems which are easily-solvable lower and upper approximations on (1). In *bounds-based approximations*, lower and upper bounds on the expectation functional $E\phi(x) := E[\phi(x, \omega)]$ of the second-stage

recourse problem (2) are determined which then would be used in the first-stage minimization to obtain approximate solutions. When approximating the underlying probability distribution P^{tr} for the latter purpose, one determines a random vector ζ to approximate the original random vector $\omega := (\xi, \eta)$ such that:

(i) The resulting problem with ζ in place of ω is relatively easy to solve.

(ii) The value of the resulting problem serves as a lower (or upper) bound for the optimal value Z^*.

(iii) It is possible to revise the approximation ζ such that the resulting bound can monotonically improve with each revision, along with a sequence of approximate x solutions which converges to an exact optimal solution of (1).

Property (i) is usually accomplished by requiring ζ to have a discrete distribution, with a relatively small number of atoms, together with an assurance that the approximate problem inherits the required convexity properties (for the minimization). What is even more challenging is to obtain these approximate problems as linear programs having decomposable structures. For property (ii), one usually determines ζ such that certain moments of it coincide with those of ω. In particular, by casting the bounding problem in a generalized moment problem setting, one solves the semi-infinite dual of the moment problem to determine required approximations; see for example, Kall [14]. Finally, property (iii) is assured by refining these bounds to a desired degree of accuracy by repeatedly applying them on subsets of the domain of the random vector ω; see for example, Huang et al. [12].

In the sequel, we assume that (1)–(2) has *complete recourse*, i.e.,

$$\{z \in \mathbb{R}^{m_2} : Wy = z, \ y \geqslant 0\} = \mathbb{R}^{m_2}, \tag{3}$$

which also implies that the recourse problem (2) is feasible w.p. 1 for any $x \in X$ where

$$X := \{x \in \mathbb{R}^{n_1} : Ax = b, \ x \geqslant 0\}, \tag{4}$$

Moreover, assuming that $\pi W \leqslant q(\eta)$ is feasible w.p. 1, ϕ is obtained as a proper convex function in $h(\xi) - T(\xi)x$, for $x \in X$, as well as a proper concave function in $q(\eta)$. Furthermore, q, h, and T are considered to be affine transformations of the random vectors, i.e.,

$$T(\xi) = T_0 + \sum_{k=1}^{K} T_k \xi_k, \qquad h(\xi) = H\xi + h_0, \quad \text{and} \quad q(\eta) = Q\eta + q_0, \tag{5}$$

where $H \in \mathbb{R}^{m_2 \times K}$ and $Q \in \mathbb{R}^{n_2 \times L}$ are fixed matrices while $h_0 \in \mathbb{R}^{m_2}$ and $q_0 \in \mathbb{R}^{n_2}$ are fixed vectors. Consequently, the recourse function $\phi(x,.,.)$ is obtained as a proper convex-concave saddle function for each $x \in X$.

This paper is concerned with implementing the bounds-based approximations derived in Edirisinghe and Ziemba [5], using first and cross moment information of the underlying distribution. On one hand, these approximations correspond to discretizing the probability distribution as mentioned under property (i), thus, the bounds are sharp, and on the other hand, these bounds are obtained as solutions to the respective moment problems, thus satisfying a tightness measure according to an optimization criterion.

While the latter bounds are applicable under convex polyhedral (bounded) Ω, by restricting the domain Ω to be of rectangular shape, one achieves two important implementational advantages. First, when refining the current approximation (see property (iii)), the partitioning of rectangles is rather straightforward. Secondly, and more importantly, the resulting lower bounding approximation can be shown to involve no more than $(L+1)$-blocks of constraints/variables.

The paper is organized as follows. In Section 2, we summarize the first and cross moment bounds developed in Edirisinghe and Ziemba [5] along with their association with the underlying moment problems. Section 3 establishes that these approximations are monotonic and they converge to true values under a successive refinement strategy that utilize partitioning of the domain. In Section 4, under rectangular domains, the approximating models are shown to admit a two-stage decomposition. Such decompositions are particularly attractive when using parallel computation to solve problems with larger dimensions of uncertainty. An implementation of the first and cross moment bounding approximations for solving scenario-based two-stage stochastic programs is discussed in Section 5. Various refining strategies for successively improving the approximations are discussed and tested computationally on randomly generated stochastic programs. Section 6 concludes the paper with some remarks.

2. Bounds using first and cross moments

Assume that Ξ and Θ are compact, convex polyhedral sets, being finitely generated by extreme points given by

$$\Xi = \mathrm{co}\{u^1, u^2, \ldots, u^I\} \quad \text{and} \quad \Theta = \mathrm{co}\{v^1, v^2, \ldots, v^J\},$$

where $\mathrm{co}\{\cdot\}$ is the convex hull. Denote the first moments of the underlying probability distribution by $\bar{\omega} := (\bar{\xi}, \bar{\eta})$ and the cross moments by $m_{kl} := E[\xi_k \eta_l]$, for all $k = 1, \ldots, K$ and $l = 1, \ldots, L$. Using the above moment information and Ω being a polytope, the following lower and upper approximations on (1) hold:

Theorem 2.1 (Edirisinghe and Ziemba [5]).

$$Z^* \leqslant \inf_{x \in X} \left\{ c'x + \sup_{\rho, \beta} \sum_{i=1}^{I} \sum_{j=1}^{J} \rho_{ij} \phi(x, u^i, \beta^i) \right\} \tag{6}$$

$$\text{s.t.} \sum_{j=1}^{J} v^j \rho_{ij} - \beta^i \sum_{j=1}^{J} \rho_{ij} = 0, \quad i = 1, \ldots, I,$$

$$\rho \in \mathscr{C},$$

and

$$Z^* \geqslant \inf_{x \in X} \left\{ c'x + \inf_{\rho, \alpha} \sum_{i=1}^{I} \sum_{j=1}^{J} \rho_{ij} \phi(x, \alpha^j, v^j) \right\} \tag{7}$$

$$\text{s.t.} \sum_{i=1}^{L} u^i \rho_{ij} - \alpha^j \sum_{i=1}^{L} \rho_{ij} = 0, \quad j = 1, \ldots, J,$$

$$\rho \in \mathscr{C},$$

where the polyhedral set $\mathscr{C}$ is defined by

$$\mathscr{C} := \left\{ \rho \in \mathbb{R}^{I \times J} : \sum_{i=1}^{I} \sum_{j=1}^{J} (u^i, v^j) \rho_{ij} = (\bar{\xi}, \bar{\eta}), \right.$$

$$\sum_{i=1}^{I} \sum_{j=1}^{J} u_k^i v_l^j \rho_{ij} = m_{kl} \quad \forall k = 1, \ldots, K, \quad l = 1, \ldots, L,$$

$$\left. \sum_{i=1}^{I} \sum_{j=1}^{J} \rho_{ij} = 1, \quad \rho_{ij} \geqslant 0 \right\}. \tag{8}$$

For the evaluation of the preceding bounds, observe that a probability measure $\rho \in \mathscr{C}$ as well as the variables $\beta^i \in \mathbb{R}^L$ $(i = 1, \ldots, I)$ and $\alpha^j \in \mathbb{R}^K$ $(j = 1, \ldots, J)$ are required to be determined optimally. It can also be shown that $\beta^i \in \Theta$ and $\alpha^j \in \Xi$ for any i and j. Consequently, the bounds in Theorem 2.1 are attained according to certain discrete probability measures (approximating P^{tr}) on the boundary of Ω, implying that the bounds are sharp. Nevertheless, the yardstick of tightness of bounds is usually the underlying moment optimization problems involving the first and cross moment conditions. That is, the bounds would be examined with respect to the upper and lower approximations due to moment problems given by

$$Z^* = \inf_{x \in X} \left\{ c'x + \int_{\Omega} \phi(x, \omega) P^{\mathrm{tr}}(\mathrm{d}\omega) \right\} \leqslant \inf_{x \in X} \left\{ c'x + \sup_{P \in \mathscr{P}} \int_{\Omega} \phi(x, \omega) P(\mathrm{d}\omega) \right\} \tag{9}$$

and

$$Z^* = \inf_{x \in X} \left\{ c'x + \int_{\Omega} \phi(x, \omega) P^{\mathrm{tr}}(\mathrm{d}\omega) \right\} \geqslant \inf_{x \in X} \left\{ c'x + \inf_{P \in \mathscr{P}} \int_{\Omega} \phi(x, \omega) P(\mathrm{d}\omega) \right\}, \tag{10}$$

where $\mathscr{P}$ denotes the set of all probability measures on Ω, characterized by the (known) first and cross moment conditions. The validity of the inequalities in (9) and (10) is due to $P^{\mathrm{tr}} \in \mathscr{P}$.

Solution of the above moment problems is not generally easy. However, the polyhedral nature of Ω and the saddle property of ϕ may be exploited for solving the semi-infinite duals of these moment problems; see Edirisinghe and Ziemba [5]. In particular, it develops that these moment problems have solutions which are the same

(optimal) discrete probability measures ρ as determined in the upper and lower approximations in Theorem 2.1, which thus implies that the bounds in the latter theorem are tight.

Moment problems involving saddle functions have been solved in literature under various assumptions on the shape of Ω as well as utilizing various higher-order moment conditions. For instance, when all joint moments of the ξ vector and all joint moments of the η vector, along with certain higher degree cross moments between ξ and η are specified, and with random vectors having compact multidimensional rectangular domains, moment problem solutions have been obtained by Frauendorfer [9]. Also, extensions to the case when Ω is unbounded polyhedral can be found in Edirisinghe and Ziemba [6].

When implementing the bounds in Theorem 2.1, however, the more important aspect is whether (6) and (7) can be solved using linear programming methodology. Although it appears at the first glance that the upper bound in (6) requires optimization with respect to $\beta^i \in \mathbb{R}^L$ (for $i = 1, \ldots, l$), it develops due to the structure of the recourse problem (2) that this upper approximation may be recast as a linear program. In particular, this means that the optimization with respect to β^i in (6) no longer becomes necessary; see Edirisinghe and Ziemba [5]. Moreover, in Section 4, we show a further decomposition of this linear program for efficient computation. In contrast, the lower approximation in (7) remains a non-convex problem owing to the optimization with respect to α^j. As a possible remedy to this situation, the following LP formulation has been derived as a weaker lower bound using first and cross moments.

Theorem 2.2 (Edirisinghe and Ziemba [5]).

$$Z^* \geqslant \min_{x \in X} \{c'x + \phi_L(x)\}, \tag{11}$$

where

$$\phi_L(x) := \min_{y \geqslant 0} \sum_{j=1}^{J} q(v^j)' y^j \tag{12}$$

$$\text{s.t.} \quad W\left(\sum_{j=1}^{J} y^j\right) = h(\bar{\xi}) - T(\bar{\xi})x,$$

$$W\left(\sum_{j=1}^{J} v_l^j y^j\right) = \bar{\eta}_l(h_0 - T_0 x) + \sum_{k=1}^{K} m_{kl}(h_k - T_k x),$$

$$l = 1, \ldots, L.$$

Observe that the lower bound $\phi_L(x)$ is convex polyhedral in x. While this lower bound is generally weaker than that in (7), at least when the polyhedral domain Θ is a simplex, it has been shown that (7) and (11) are identical. Section 4 derives a decomposition of this lower approximating LP for efficient computation, and the implementation is considered in Section 5.

3. Monotonicity and convergence

One advantage of using bounds-based approximations, such as those discussed above, is the ability to generate ever-improving sequences of upper and lower bounds on the optimal value Z^* of (1), by applying the bounding strategies on subsets of the domains. The goal of this section is to show that the preceding first/cross moment bounds when applied repeatedly on partitions of the domain Ω, have the property that as these partitions become finer the bounds become tighter. The exposition here is for the upper bound in (6) and the lower bound in (11) under compact polyhedral domains. These are the bounds that will be further discussed and implemented in the remainder of this article.

Consider a partition $\mathcal{S}$ on Ω, consisting of sub-cells Ω_r $(= \Xi_r \times \Theta_r)$, for $r = 1, \ldots, R$, with positive probability mass p_r such that the closure of Ω_r is a polytope and

$$\Omega_r \cap \Omega_{r'} = \emptyset, \quad r, r' \in \{1, \ldots, R\}, \ r \neq r', \tag{13}$$

$$\bigcup_{r=1}^{R} \Omega_r = \Omega. \tag{14}$$

Since a chosen sub-cell Ω_r can be an open set, when applying the bounding strategies on this cell, the extreme points of the closure of Ω_r are considered. Also denote the required first/cross conditional moments on Ω_r for $r = 1, \ldots, R$ by

$$\bar{\xi}^r := E[\xi \mid (\xi, \eta) \in \Omega_r],$$

$$\bar{\eta}^r := E[\eta \mid (\xi, \eta) \in \Omega_r],$$

$$m_{kl}^r := E[\xi_k \eta_l \mid (\xi, \eta) \in \Omega_r] \quad \forall (k, l).$$

Let $\phi_U(x)$ denote the resulting upper bound when applying the upper bounding strategy in (6) on the expectation of the recourse function, i.e., $E\phi(x) \leqslant \phi_U(x)$ where

$$\phi_U(x) := \sup_{\rho, \beta} \left\{ \sum_{i,j} \rho_{ij} \phi(x, u^i, \beta^i) : \sum_j v^j \rho_{ij} - \beta^i \sum_j \rho_{ij} = 0, \right.$$

$$\left. i = 1, \ldots, I, \ \rho \in \mathcal{C} \right\}. \tag{15}$$

While $\phi_U(x)$ has been shown to be convex polyhedral in Edirisinghe and Ziemba [5], the following equivalent formulation of (15) will be used to show that this upper bound is monotonic.

Lemma 3.1.

$$\phi_U(x) = \inf_{\pi} \ \pi_0 + \sum_{k=1}^{K} \pi_k^1 \bar{\xi}_k + \sum_{l=1}^{L} \pi_l^2 \bar{\eta}_l + \sum_{k,l} \pi_{kl} m_{kl} \tag{16}$$

$$\text{s.t.} \quad \pi_0 + \sum_{k=1}^{K} \pi_k^1 \xi_k + \sum_{l=1}^{L} \pi_l^2 \eta_l + \sum_{k,l} \pi_{kl} \xi_k \eta_l \geqslant \phi(x, \xi, \eta)$$

$$\forall (\xi, \eta) \in \Omega.$$

Proof. Due to Theorem 2.12 in Edirisinghe and Ziemba [5], $\phi_U(x)$ is the value of the underlying moment problem with first and cross moments, i.e.,

$$\phi_U(x) = \sup_{P \in \mathscr{P}} \int_{\Omega} \phi(x, \xi, \eta) P(\mathrm{d}\xi, \mathrm{d}\eta), \tag{17}$$

where $\mathscr{P}$ is the set of all probability measures on Ω with first moments $(\bar{\xi}, \bar{\eta})$ and cross moments m_{kl}. Now due to Theorem 1.1 of the latter reference, since Ω is compact and ϕ is continuous, applying the strong duality result on (17), (16) is obtained as the dual problem of (17). □

Theorem 3.2. *Under a partition* (13)–(14) *of the domain* Ω, *the upper bound* $\phi_U(x)$ *in* (15) *behaves monotonically for all* $x \in X$.

Proof. Let $\phi_U^r(x)$ denote the resulting upper bound when applying (15) for the expectation of $\phi(x, \xi, \eta)$ conditional on $(\xi, \eta) \in \mathrm{Cl}(\Omega_r)$, the closure of Ω_r. To prove the required claim, it must be shown that

$$E\phi(x) \leqslant \sum_{r=1}^{R} p_r \phi_U^r(x) \leqslant \phi_U(x). \tag{18}$$

By Lemma 3.1,

$$\begin{aligned}
\phi_U^r(x) = \inf_{\pi(r)} \quad & \pi^0(r) + \sum_{k=1}^{K} \bar{\xi}_k^r \pi_k^1(r) + \sum_{l=1}^{L} \bar{\eta}_l^r \pi_l^2(r) + \sum_{k,l} m_{kl}^r \pi_{kl}^3(r) \\
\text{s.t.} \quad & \pi^0(r) + \sum_{k=1}^{K} \xi_k \pi_k^1(r) + \sum_{l=1}^{K} \eta_l \pi_l^2(r) + \sum_{k,l} \xi_k \eta_l \pi_{kl}^3(r) \\
& \geqslant \phi(x, \xi, \eta) \quad \forall (\xi, \eta) \in \mathrm{Cl}(\Omega_r).
\end{aligned}$$

Since, $\bar{\xi} = \sum_r p_r \bar{\xi}^r$, $\bar{\eta} = \sum_r p_r \bar{\eta}^r$, and $m_{kl} = \sum_r p_r m_{kl}^r$,

$$\begin{aligned}
\phi_U(x) = \inf_{\pi} \quad & \sum_r p_r \pi^0 + \sum_{k,r} p_r \bar{\xi}_k^r \pi_k^1 + \sum_{l,r} p_r \bar{\eta}_l^r \pi_l^2 + \sum_{k,l,r} p_r m_{kl}^r \pi_{kl}^3 \\
\text{s.t.} \quad & \pi^0 + \sum_k \xi_k \pi_k^1 + \sum_l \eta_l \pi_l^2 + \sum_{k,l} \xi_k \eta_l \pi_{kl}^3 \geqslant \phi(x, \xi, \eta) \\
& \forall (\xi, \eta) \in \mathrm{Cl}(\Omega_r), \quad r = 1, \ldots, R, \\
\geqslant \inf_{\pi(r)} \quad & \sum_r p_r \pi^0(r) + \sum_{k,r} p_r \bar{\xi}_k^r \pi_k^1(r) + \sum_{l,r} p_r \bar{\eta}_l^r \pi_l^2(r) \\
& + \sum_{k,l,r} p_r m_{kl}^r \pi_{kl}^3(r) \\
\text{s.t.} \quad & \pi^0(r) + \sum_k \xi_k \pi_k^1(r) + \sum_l \eta_l \pi_l^2(r) + \sum_{k,l} \xi_k \eta_l \pi_{kl}^3(r) \\
& \geqslant \phi(x, \xi, \eta) \quad \forall (\xi, \eta) \in \mathrm{Cl}(\Omega_r), \quad r = 1, \ldots, R \\
= \sum_{r=1}^{R} & p_r \phi_U^r(x),
\end{aligned}$$

thus proving the second inequality in (18). To prove the first inequality in (18), note that $E\phi(x) = \sum_r p_r E[\phi(x, \xi, \eta) \mid \Omega_r]$. Since $E[\phi(x, \xi, \eta) \mid \Omega_r] \leqslant \phi_U^r(x)$ holds for $r = 1, \ldots, R$, the proof is completed. □

Next, we turn to the issue of monotonicity of the convex lower bound $\phi_L(x)$ in (12). Since this lower bound may not generally represent a solution of the moment problem in (10), to show its monotonic behavior, one has to proceed differently than in the case of the upper bound.

Considering the partition $\mathscr{S}$ in (13)–(14), let $\phi_L^r(x)$ denote the resulting lower bound when applying (12) on (the closure of) the sub-cell Ω_r, $r = 1, \ldots, R$. Consequently, one has to show the inequalities

$$E(\phi) \geqslant \sum_{r=1}^{R} p_r \phi_L^r(x) \geqslant \phi_L(x) \tag{19}$$

to prove the monotonic behavior of $\phi_L(x)$.

Theorem 3.3. *Under a partition* (13)–(14) *of the domain* Ω, *the lower bound* $\phi_L(x)$ *in* (12) *behaves monotonically for all* $x \in X$.

Proof. Since $\mathrm{Cl}(\Omega_r) \subseteq \Omega$ holds for any $r = 1, \ldots, R$, one may apply the lower bounding strategy (12) with conditional moment information $\bar{\xi}^r$, $\bar{\eta}^r$, and m_{kl}^r, but with extreme points of $\mathrm{Cl}(\Omega_r)$ being replaced by those of Ω to determine a weaker lower bound as follows:

$$\begin{aligned}
& E[\phi(x, \omega) \mid \omega \in \Omega_r] \geqslant \phi_L^r(x) \qquad (20) \\
& \geqslant \min_{y^{rj} \geqslant 0} \quad \sum_{j=1}^{J} q(v^j) y^{rj} \\
& \text{s.t} \quad W\left(\sum_{j=1}^{J} y^{rj} \right) = h(\bar{\xi}^r) - T(\bar{\xi}^r) x, \\
& \qquad W\left(\sum_{j=1}^{J} v_l^j y^{rj} \right) = \bar{\eta}_l^r (h_0 - T_0 x) + \sum_{k=1}^{K} m_{kl}^r (h_k - T_k x) \quad \forall l.
\end{aligned}$$

The second inequality in (20) can be easily shown due to that each extreme point of $\mathrm{Cl}(\Omega_r)$ can be expressed as a convex combination of the extreme points v^j, $j = 1, \ldots, J$, of Ω.Therefore, it follows that

$$\begin{aligned}
E\phi(x) & \geqslant \sum_{r=1}^{R} p_r \phi_L^r(x) \qquad (21) \\
& \geqslant \min_{y^{rj} \geqslant 0} \quad \sum_{r=1}^{R} p_r \sum_{j=1}^{J} q(v^j) y^{rj}
\end{aligned}$$

$$
\begin{aligned}
\text{s.t.}\quad & W\left(\sum_{j=1}^{J} y^{rj}\right) = h(\bar{\xi}^r) - T(\bar{\xi}^r)x \quad \forall r = 1,\dots,R,\\
& W\left(\sum_{j=1}^{J} v_l^j y^{rj}\right) = \bar{\eta}_l^r(h_0 - T_0 x) + \sum_{k=1}^{K} m_{kl}^r(h_k - T_k x)\\
& \qquad \forall l, \quad \forall r = 1,\dots,R\\
\geqslant \min_{y^{rj}\geqslant 0}\ & \sum_{j=1}^{J} q(v^j)\left[\sum_{r=1}^{R} p_r y^{rj}\right]\\
\text{s.t.}\quad & W\left(\sum_{j=1}^{J}\left[\sum_{r=1}^{R} p_r y^{rj}\right]\right) = h(\bar{\xi}) - T(\bar{\xi})x,\\
& W\left(\sum_{j=1}^{J} v_l^j\left[\sum_{r=1}^{R} p_r y^{rj}\right]\right) = \bar{\eta}_l(h_0 - T_0 x)\\
& \qquad + \sum_{k=1}^{K} m_{kl}(h_k - T_k x) \quad \forall l\\
= \phi_L(x).&
\end{aligned}
$$

The last inequality above is due to taking a row aggregation of the constraints of (21) for $r = 1,\dots,R$ with aggregating multipliers being the probabilities p_r, along with using that $h(\cdot)$ and $T(\cdot)$ are linear affine. This completes the proof. □

3.1. Convergence

Pointwise-convergence of bounds-based approximations to the expected recourse function $E\phi(x)$, under partitioning the domain, is usually claimed by relying on *weak* convergence of approximating probability measures, see for instance, Frauendorfer [9]. In particular, to claim convergence of the upper bound, consider a sequence of partitions $\mathscr{S}^\nu$, $\nu = 1, 2,\dots$, where each $\mathscr{S}^\nu$ is described by a collection of (polyhedral) cells $\{\Omega_r, r = 1,\dots,R_\nu\}$ of Ω satisfying (13)–(14) such that $\mathscr{S}^\nu \subset \mathscr{S}^{\nu+1}$. Suppose,

$$\lim_{\nu\to\infty}\left(\max_{1\leqslant r\leqslant R_\nu} P^{\text{tr}}[\,\omega \mid \omega\in\Omega_r\,]\right) = 0. \tag{22}$$

Observe that the upper bound $\phi_U(x)$ in (15) is the expectation of $\phi(x,\omega)$ with respect to a probability measure ρ defined on the discrete points $(u^i, \beta^i)\in\Omega$, $i = 1,\dots,I$, these points being optimally determined in the solution of (15). With ρ^ν denoting the upper approximating probability measure at the partitioning iteration ν, it follows due to (22) that $\{\rho^\nu\}$ weakly-converges to P^{tr}. Then, due to Birge and Wets [2, Section 2.11], from the compactness of Ω it follows that $E_{\rho^\nu}[\phi]$ is both pointwise- and epi-convergent to $E_{P^{\text{tr}}}[\phi]$.

An important issue is whether a sequence of approximate first-stage solutions so generated converges to a true optimal solution of the stochastic program (1). With regard to the sequence of upper bounding functions $\phi_U^\nu(x)$ which pointwise- and epi-converge

to $E\phi(x)$, one may apply Theorem 9 of Wets [17] to conclude that the sequence of approximate first-stage solutions is guaranteed to belong to the set of optimal solutions. That is, for the upper bounding solutions x_U^ν of (6), we have $\lim_{\nu\to\infty} x_U^\nu = x^*$ where x^* is an optimal solution of the stochastic program (1).

Convergence argument for the convex lower bound in (12) is slightly complicated by the fact that this lower bound generally does not correspond to an approximating probability measure. To claim pointwise-convergence of $\phi_L^\nu(x)$ to $E\phi(x)$, we first reformulate the lower bound as follows:

Proposition 3.4. *The first/cross moment lower bound in* (12) *can be restated as*

$$\phi_L(x) = \inf_{\eta^i}\{\Phi(x,\eta^1,\dots,\eta^{n_2}) : \eta^i\in\Theta,\ i = 1,\dots,n_2\}, \tag{23}$$

where

$$
\begin{aligned}
\Phi(x,\eta^1,\dots,\eta^{n_2}) := \min_{z^0,z^l}\ & q_0' z^0 + \sum_{l=1}^{L} q_l' z^l\\
\text{s.t.}\ & Wz^0 = h(\bar{\xi}) - T(\bar{\xi})x,\\
& Wz^l = \gamma_l \quad \forall l = 1,\dots,L,\\
& z_i^l - \eta_i^l z_i^0 = 0, \qquad z_i^0 \geqslant 0 \quad \forall i = 1,\dots,n_2,
\end{aligned}
\tag{24}
$$

and $\gamma_l := \bar{\eta}_l(h_0 - T_0 x) + \sum_{k=1}^{K} m_{kl}(h_k - T_k x)$ *and* q_l *is the l-th column of* Q.

Proof. Since $\sum_j q(v^j)' y^j = q_0'\sum_j y^j + \sum_l q_l'\sum_j v_l^j y^j$, (12) may be expressed as

$$
\begin{aligned}
\phi_L(x) = \min_{z^0,z^l}\ & q_0' z^0 + \sum_{l=1}^{L} q_l' z^l\\
\text{s.t.}\ & Wz^0 = h(\bar{\xi}) - T(\bar{\xi})x,\\
& Wz^l = \gamma_l \quad \forall l,\\
& \mathscr{Y}(z) \neq \emptyset,
\end{aligned}
\tag{25}
$$

where $\mathscr{Y}(z) := \{(y^1,\dots,y^J) : \sum_j y^j = z^0,\ \sum_j v_l^j y^j = z^l,\ y^j\geqslant 0,\ \forall j\}$. Notice that $\mathscr{Y}(z)\neq\emptyset$ if and only if $\eta^i\in\Theta$ for $i = 1,\dots,n_2$ where η^i is defined by $z_i^l - \eta_i^l z_i^0 = 0$ for $l = 1,\dots,L$. Substituting this in (25) yields the desired result. □

Theorem 3.5. *Let* $\mathscr{S}^\nu$, $\nu = 1, 2,\dots$, *be a sequence of partitions where each* $\mathscr{S}^\nu$ *is described by a collection of (polyhedral) cells* $\Omega_r(=\Xi_r\times\Theta_r)$, $r = 1,\dots,R_\nu$, *of* Ω *satisfying* (13)–(14) *such that* $\mathscr{S}^\nu\subset\mathscr{S}^{\nu+1}$. *Suppose*

$$\lim_{\nu\to\infty}\left[\max_{r=1,\dots,R_\nu}\left(\max_{\omega^1,\omega^2\in\Omega_r}\|\,\omega^1 - \omega^2\,\|\right)\right] = 0, \tag{26}$$

i.e., the diameters of the cells of the partitions become arbitrarily small. Then, the sequence of lower bounding functions $\{\phi_L^\nu(x)\}$ *is pointwise-convergent, i.e., for all* x,

$$E\phi(x) = \lim_{\nu\to\infty}\phi_L^\nu(x).$$

Furthermore, the sequence $\{\phi_L^\nu(x)\}$ *is epi-convergent.*

Proof. Let $\phi_L^{r,\nu}(x)$ denote the (convex) lower bound on the cell Ω_r, $r = 1,\ldots,R_\nu$, of a partition $\mathcal{S}^\nu$. Thus, the overall lower bound on $E\phi(x)$ corresponding to the partition $\mathcal{S}^\nu$ is $\phi_L^\nu(x) = \sum_{r=1}^{R_\nu} p_r \phi_L^{r,\nu}(x)$. For notational simplicity, suppressing the superscript ν, denote the conditional moments in cell Ω_r by $\bar{\xi}^r$, $\bar{\eta}^r$, and m_{kl}^r. Due to Proposition 3.4,

$$\phi_L^{r,\nu}(x) = \inf_{\eta^i}\{\Phi_r(x, \eta^1,\ldots,\eta^{n_2}) : \eta^i \in \Theta_r,\ i = 1,\ldots,n_2\}, \tag{27}$$

where Φ_r is defined according to (24) for the cell Ω_r. Let us first fix $\eta^1,\ldots,\eta^{n_2} \in \Theta_r$. Since

$$\begin{aligned}&\Phi_r(x,\eta,\ldots,\eta) - \Phi_r(x,\eta^1,\ldots,\eta^{n_2})\\ &\quad= [\Phi_r(x,\eta,\ldots,\eta) - \Phi_r(x,\eta,\ldots,\eta,\eta^{n_2})]\\ &\qquad+\cdots+[\Phi_r(x,\eta,\eta^2,\ldots,\eta^{n_2}) - \Phi_r(x,\eta^1,\ldots,\eta^{n_2})]\end{aligned}$$

holds for any $\eta \in \Theta_r$,

$$\begin{aligned}&|\Phi_r(x,\eta,\ldots,\eta) - \Phi_r(x,\eta^1,\ldots,\eta^{n_2})|\\ &\quad\leqslant |\Phi_r(x,\eta,\ldots,\eta) - \Phi_r(x,\eta,\ldots,\eta,\eta^{n_2})|\\ &\qquad+\cdots+|\Phi_r(x,\eta,\eta^2,\ldots,\eta^{n_2}) - \Phi_r(x,\eta^1,\ldots,\eta^{n_2})|.\end{aligned}$$

For sufficiently large ν, since (26) implies that there exists small $\varepsilon > 0$ such that $\|\eta - \eta^i\| \leqslant \varepsilon$ for any $\eta \in \Theta_r$, the continuity of Φ_r in η^i implies that there exists small $\varepsilon_i > 0$ such that

$$|\Phi_r(x,\eta,\ldots,\eta,\eta^{i+1},\ldots,\eta^{n_2}) - \Phi_r(x,\eta,\ldots,\eta,\eta^i,\eta^{i+1},\ldots,\eta^{n_2})| \leqslant \varepsilon_i \quad \forall i = 1,\ldots,n_2,$$

and thus,

$$|\Phi_r(x,\eta,\ldots,\eta) - \Phi_r(x,\eta^1,\ldots,\eta^{n_2})| \leqslant \varepsilon + \cdots + \varepsilon_{n_2}.$$

Defining $\delta_1 := \varepsilon_1 + \cdots + \varepsilon_{n_2}$, this yields

$$\Phi_r(x,\eta^1,\ldots,\eta^{n_2}) \geqslant \Phi_r(x,\eta,\ldots,\eta) - \delta_1 \quad \forall \eta \in \Theta_r. \tag{28}$$

Furthermore, it follows due to (24) that

$$\begin{aligned}\Phi_r(x,\eta,\ldots,\eta) = \min_{z^0\geqslant 0}\ & q(\eta)' z^0\\ \text{s.t.}\ & Wz^0 = h(\bar{\xi}^r) - T(\bar{\xi}^r)x,\\ & \eta_l W z^0 = \gamma_l^r \quad \forall l = 1,\ldots,L\\ \geqslant \min_{z^0\geqslant 0}\ & q(\eta)' z^0 \qquad (29)\\ \text{s.t.}\ & Wz^0 = h(\bar{\xi}^r) - T(\bar{\xi}^r)x\\ = \phi(x,\bar{\xi}^r,\eta), & \qquad (30)\end{aligned}$$

where $\gamma_l^r := \bar{\eta}_l^r(h_0 - T_0 x) + \sum_{k=1}^K m_{kl}^r(h_k - T_k x)$. The inequality in (29) is due to dropping the constraints involving γ_l^r. Furthermore, since $\|(\xi,\eta) - (\bar{\xi}^r,\eta)\| \leqslant \varepsilon$ holds due to (26) for any $(\xi,\eta) \in \Omega_r$, the continuity of $\phi(x,.,.)$ implies that there exists small $\delta_2 > 0$ such that $|\phi(x,\xi,\eta) - \phi(x,\bar{\xi}^r,\eta)| \leqslant \delta_2$ and hence, $\phi(x,\bar{\xi}^r,\eta) \geqslant \phi(x,\xi,\eta) - \delta_2$ for any $(\xi,\eta) \in \Omega_r$. Combining with (28) and (30), this yields

$$\Phi_r(x,\eta^1,\ldots,\eta^{n_2}) \geqslant \phi(x,\xi,\eta) - (\delta_1 + \delta_2) \quad \forall(\xi,\eta) \in \Omega_r. \tag{31}$$

Since (31) holds for any fixed $\eta^1,\ldots,\eta^{n_2} \in \Theta_r$, by referring to (27) and taking the infimum w.r.t. $\eta^i \in \Theta_r$,

$$\phi_L^{r,\nu}(x) \geqslant \phi(x,\xi,\eta) - (\delta_1 + \delta_2) \quad \forall(\xi,\eta) \in \Omega_r. \tag{32}$$

Since (32) holds for all cells Ω_r in $\mathcal{S}^\nu$, taking expectations and since $\nu \to \infty$ ensures that $\varepsilon \to 0$, $\delta_1 \to 0$, and $\delta_2 \to 0$, it follows that $\lim_{\nu\to\infty}\phi_L^\nu(x) \geqslant E\phi(x)$ for all x. However, since $\phi_L^\nu(x)$ is a valid lower bound on $E\phi(x)$, i.e., $\phi_L^\nu(x) \leqslant E\phi(x)$, $\forall\nu$, we must also have $\lim_{\nu\to\infty}\phi_L^\nu(x) \leqslant E\phi(x)$ for all x, thus proving the pointwise-convergence of the lower bound.

Next observe due to Theorem 3.3 that $\{\phi_L^\nu(x)\}$ is monotonic, and furthermore, $E\phi(x)$ is lower-semicontinuous. These along with the pointwise-convergence of $\{\phi_L^\nu(x)\}$ satisfy the necessary conditions of Proposition 3.12 of Attouch and Wets [1], thus ensuring that ϕ_L^ν epi-converges to $E\phi$. □

Now, applying Theorem 9 of Wets [17], one may conclude due to epi-convergence that the sequence of first-stage solutions of the lower approximation (11) converges to an optimal solution of the stochastic program (1).

There remain two issues to be addressed. First, how efficiently can the resulting upper and lower approximations be solved at each partitioning iteration, and second, strategies for refinement of a partition of Ω, based on the information gathered from previous partitioning steps. The former is discussed in the next section while the latter is addressed in Section 5 under implementational details.

4. Decomposition of the approximations

A standard practice in applying bounds for sequentially approximating (1) is to first apply the lower bounding strategy to determine an approximate first stage (lower bounding) solution, x_L, and then apply the upper bounding strategy on each cell of a partition of Ω, fixing $x = x_L$. This approach is valid because (2) is assumed to have complete recourse. There are at least two advantages with this practice. First, the upper bounding problem becomes separable by each cell of the partition. Secondly, it is possible to examine how far apart the lower and upper bounds are for each cell (since the same x solution is used). Then, based on the relative gap of the bounds (or some other measure; see Frauendorfer and Kall [11]), cells which need to be further partitioned could be chosen.

Another important element in sequentially refining approximations is choosing a suitable (geometric) partitioning scheme for the domain Ω. While the first/cross moment bound $\phi_U(x)$ in (15) and $\phi_L(x)$ in (12) are applicable under general (bounded)

polyhedral Ω, the implementation to be described in this paper incorporates a rectangular decomposition of Ω. The main computational advantage is that in refining the bounds on rectangular domains, partitioning planes can be chosen perpendicular to coordinate axes to yield sub-rectangles. However, the major drawback of this approach is that the number of extreme points in the (partitioned) domain increases exponentially. In particular, the lower bounding computations are adversely affected by this due to the necessity that an (improved) approximate x solution needs to be determined, which thus results in an exponentially-growing lower approximation model with each partitioning step.

In this section, it will be shown under rectangular domains that the lower approximation in (11)–(12) can be reformulated such that it affords no more than a linear growth in model size with respect to the number of random variables. Furthermore, it follows that the resulting structure in the lower bound permits efficient solution through LP decomposition techniques.

4.1. Lower approximation

The goal here is to take advantage of Ω being rectangular to cast the first/cross moment lower bound (12) in a decomposable format that would necessitate solving linear programs of size no larger than that of deterministic recourse problems. In its present form, observe that $\phi_L(x)$ in (12) is $(L+1)$ times larger in the number of constraints and 2^L times larger in the number of variables than a deterministic recourse problem. That the computations of this lower bounding model is affordable for larger dimensions of uncertainty is evident in the following result:

Theorem 4.1. *If Θ, the domain of η, is a multidimensional rectangle of the form $\prod_{l=1}^{L}[b_{l0}, b_{l1}]$, then (12) can be simplified as*

$$\begin{aligned} \phi_L(x) = \min_{z^0} \quad & q_0' z^0 + \sum_{l=1}^{L} \varphi_l(x, z^0) \\ \text{s.t.} \quad & W z^0 = h(\bar{\xi}) - T(\bar{\xi}) x, \\ & z^0 \geqslant 0, \end{aligned} \tag{33}$$

where

$$\begin{aligned} \varphi_l(x, z^0) := \min_{z^l} \quad & q_l' z^l \\ \text{s.t.} \quad & W z^l = \gamma_l, \\ & b_{l0} z^0 \leqslant z^l \leqslant b_{l1} z^0, \end{aligned} \tag{34}$$

$\gamma_l := \bar{\eta}_l(h_0 - T_0 x) + \sum_{k=1}^{K} m_{kl}(h_k - T_k x)$, and q_l is the l-th column of Q.

Proof. Consider the equivalent formulation of $\phi_L(x)$ in Proposition 3.4. Since Θ is rectangular, $\eta^i \in \Theta$ in (23) is equivalent to $b_{l0} \leqslant \eta_l^i \leqslant b_{l1}$. Substituting this in (24) yields the desired result. □

The lower bounding model (33) is amenable to solution by the standard L-shaped method for solving two-stage linear programs; see Van Slyke and Wets [16]. Note that $\phi_L(x)$ in (33) now involves only L linear subproblems in (34), each of the size of a deterministic recourse problem with simple bounds on variables. This linearity in L allows solution of the lower approximation more efficiently for larger dimensions of uncertainty, and in particular, would need only L processors in a parallel computing environment. Finally, observe that in the absence of uncertainty in the objective coefficients, i.e., $q(\eta) \equiv q_0$, the model in (33) retains only the first stage of the minimization, which is the standard Jensen mean model.

Given a lower bounding solution x_L of the model (11) where $\phi_L(x)$ is evaluated according to (33), that the upper approximation in (6) can be cast in a decomposable format is shown next. Also the steps of an algorithm for solving it is briefly sketched.

4.2. Upper approximation

In the upper bounding strategy (15), substitute for ϕ from the recourse function definition (2) to obtain

$$\phi_U(x_L) = \max_{\substack{\rho \in \mathscr{B} \\ \sum_j v^j \rho_{ij} - \beta^i \sum_j \rho_{ij} = 0}} \left[\sum_{i=1}^{I} \min_{\substack{y^i \geqslant 0 \\ \text{s.t.}}} \sum_{j=1}^{J} \rho_{ij} q(\beta^i)' y^i \quad W y^i = h(u^i) - T(u^i) x_L \right]. \tag{35}$$

Defining

$$f_i(\rho) := \min_{\substack{y^i \geqslant 0 \\ \text{s.t.}}} \left(\sum_{j=1}^{J} \rho_{ij} q(v^j) \right)' y^i \quad W y_i = h(u^i) - T(u^i) x_L, \tag{36}$$

thus it follows that

$$\phi_U(x_L) = \max_{\rho \in \mathscr{B}} \sum_{i=1}^{I} f_i(\rho). \tag{37}$$

Since $f_i(\rho)$ is piece-wise linear concave in ρ, we propose to solve (37) by generating these linear pieces iteratively. This is the fundamental idea in the L-shaped decomposi-

tion of two-stage linear programs. The resulting decomposition algorithm to solve the upper approximation is as follows:

Outline of solution method:

- Begin with any feasible $\rho^1 \in \mathscr{C}$ (using the phase-I method) and set $k = 1$.
- *Step* 1. At some iteration k, given $\rho = \rho^k$, solve the i-th LP subproblem in (36). There are two cases:
  (a) Eq. (36) is *unbounded* for some i:
  For a direction of unboundedness $\sigma^{i,k}$, add the *boundedness cut*

$$\sum_j \left(q(v^j)\sigma^{i,k}\right)\rho_{ij} \geqslant 0 \tag{38}$$

  to the *master* LP problem (40) and go to Step 2.
  (b) Eq. (36) is *finite* for all i:
  For an optimal solution $y^{i,k}$ of (36), add the *optimality cut*

$$\theta_i - \sum_j \left(q(v^j)\, y^{i,k}\right)\rho_{ij} \leqslant 0 \quad \forall i = 1, \ldots, I \tag{39}$$

  to the *master* LP problem (40) and go to Step (2).
- *Step* 2. Solve the master problem:

$$\begin{aligned} F^k(x_L) := \max_{\rho,\theta} \quad & \sum_i \theta_i \\ \text{s.t.} \quad & \sum_j \left(q(v^j)'\sigma^{i,r}\right)\rho_{ij} \geqslant 0, \quad r = 1, \ldots, k, \\ & \theta_i - \sum_j \left(q(v^j)'\, y^{i,r}\right)\rho_{ij} \leqslant 0, \\ & \quad r = 1, \ldots, k \quad \forall i, \\ & \rho \in \mathscr{C}. \end{aligned} \tag{40}$$

  Go to Step (1).
- *Optimality Criterion*:
  At some iteration k, for an optimal solution (θ^k, ρ^k) of (40) and for this ρ^k, in Step 1 under case (b), for LP subproblem optimal solutions $y^{i,k}$, if it holds that

$$\theta_i^k \leqslant \sum_j \rho_{ij}^k q(v^j)\, y^{i,k} \quad \forall i, \tag{41}$$

  then STOP with $\phi_U(x_L) = F^k(x_L)$.

Details of the derivation of the above algorithm is skipped due to its similarity to the standard L-shaped algorithm. What is important to notice here is that each of the LP subproblems is of the size of a *deterministic recourse problem* while the master linear problem, without any added cuts, has only $(K+1)*(L+1)$ constraints and $I*(J+1)$ variables, thus permitting solution of the upper approximation for larger dimensions of uncertainty. Moreover, it becomes possible to parallelize subproblem solution on I number of processors.

5. Implementation of first/cross moment bounds

Computational results using an implementation of the preceding first/cross moment bounds for rectangular domains are reported here. These results pertain to stochastic programs having complete recourse, generated randomly with stochastic right hand sides and stochastic objective coefficients. The code by Kall and Keller [15] was used to generate these problems with a slight modification to include stochastic objective coefficients such that the resulting recourse problems are feasible and bounded for any realization of the random vector in the prescribed rectangle. These rectangles are specified by lower and upper limits a_{k0} and a_{k1} on ξ coordinates and b_{l0} and b_{l1} on η coordinates.

In the prescribed rectangle Ω, an (empirical) probability distribution is created by generating a (large) number of discrete realizations of ω. This is done by assigning realizations to the 2^{K+L} extreme points of Ω and then taking (random) convex combinations of these extreme realizations. Thus, the stochastic problems so generated may be viewed as having random vectors described by a large number of (randomly generated) equally-likely *scenarios*. Elements of H, T, and Q are generated according to (5) for user specified dimensions, with all deterministic coefficients being generated in the range $[-10, 10]$.

The experimental code RAWFC (**R**ectangular **A**pproximation **W**ith **F**irst and **C**ross moments), written in Fortran 77, generates two-stage stochastic programs according to user specifications and solves them using the first/cross moment lower approximation model in (33) and the upper approximation model in (37). It uses IBM's OSL subroutine to solve all linear programming subproblems when evaluating the approximations. In this experimental code, the number of scenarios has a maximum limit of 6000, along with a maximum Ω dimension of 8. Thus, for this maximum allowable size, each of the 8 basic random variables may be viewed as being allowed to take 3 possible outcomes on average. Also the code allows partitioning of the rectangular domain Ω along a specified coordinate direction to yield sub-rectangles, up to a maximum of 20 such partitions.

At an intermediate partitioning iteration, three types of questions arise: first, which cell or cells of the current partition should be further sub-divided?; secondly, once a cell is chosen, should it be sub-divided by partitioning a coordinate of ξ or η?; and finally, what should be the partitioning point for the chosen coordinate?

Recall from Section 4 that the standard practice is to first solve the lower bounding approximation and then evaluate the upper bound for this lower bounding first-stage solution, x_L. Consequently, the first issue may be resolved by examining the gap between the lower and upper bounds and choosing a cell Ω_r of the current partition $\mathscr{S}^\nu$ of Ω, according to the criterion

$$\max_{r=1,\ldots,R_\nu} \delta(\Omega_r) := \left\{P^\pi(\omega \in \Omega_r)\left[\phi_U^{r,\nu}(x_L) - \phi_L^{r,\nu}(x_L)\right]\right\}, \tag{42}$$

where ν is the partitioning iteration index. While there are alternative criteria in the choice of a cell (see Frauendorfer and Kall [11]), the criterion in (42) chooses the partitioning cell having the maximum *weighted difference* of lower and upper bounds, being weighted by the probability associated with the cell.

For notational convenience, we suppress both the partitioning iteration index and the index for cell number, and thus in the ensuing discussion, Ω denotes the current cell (or domain) to be partitioned. In order to determine whether to partition a ξ coordinate or an η coordinate, first consider indexing the following $(K+L+1)$ extreme points of Ω:

$$w^0 := (u^0, v^0) := (a_{10}, a_{20}, \ldots, a_{K0}, b_{10}, b_{20}, \ldots, b_{L0})$$

and

$$w^t := \begin{cases} (u^t, v^0) \equiv (a_{10}, \ldots, a_{t1}, \ldots, a_{K0}, b_{10}, b_{20}, \ldots, b_{L0}) \\ \quad \text{for } t = 1, \ldots, K, \\ (u0, v^t) \equiv (a_{10}, \ldots, a_{K0}, b_{10}, b_{20}, \ldots, b_{(t-K)1}, \ldots, b_{L0}), \\ \quad \text{for } t = K+1, \ldots, K+L. \end{cases}$$

As in Frauendorfer and Kall [11], the degree of nonlinearity of the recourse function ϕ between the two points w^0 and w^t can be evaluated, respectively, by considering dual solutions (for $t = 1, \ldots, K$, and primal solutions for $t = K+1, \ldots, K+L$) of the LPs defining $\phi(x_L, w^0)$ and $\phi(x_L, w^t)$.

We use the following approach. Solve the LP determining $\phi(x_L, w^0)$ and let y^0 and π^0 be its optimal primal and dual solutions, respectively. Then, for each $t = 1, \ldots, K$, under sensitivity analysis, determine the allowable increase and the allowable decrease for all constraint right-hand sides such that the current dual solution π^0 remains optimal. Similarly, for each $t = K+1, \ldots, K+L$, determine the allowable increase and decrease for each objective coefficient such that the current primal solution y^0 remains optimal. These are already available from the OSL subroutine on return. Then, for each w^t, applying the *100% rule* in sensitivity analysis in linear programming (for multiple changes in the right-hand side or objective coefficients) – see Bradley et al. [3] – determine the set $\overline{\mathcal{T}}$ of indices t such that ϕ is linear between w^0 and w^t. For the ease of exposition, we assume that $\overline{\mathcal{T}}$ is empty and thus for all indices in $\mathcal{T} := \{1, \ldots, K+L\}$, ϕ is nonlinear. Consequently, solve the linear programs determining $\phi(x_L, w^t)$ for $t \in \mathcal{T}$. For $t \leqslant K$, let π^t be its dual solution and for $t > K$, let y^t be its primal solution.

For indices $t \in \mathcal{T}$, the coordinate axis along which ϕ is *mostly nonlinear* is determined by approximating the function ϕ between w^0 and w^t by the two-piece linear function ψ_t resulting from the solutions of $\phi(x_L, w^0)$ and $\phi(x_L, w^t)$; see Fig. 1 for the case $t \leqslant K$. Thus, ψ_t is convex for $1 \leqslant t \leqslant K$ and concave for $K+1 \leqslant t \leqslant K+L$. The degree of nonlinearity associated with each ψ_t is assessed using the measure *terminal-nonlinearity* defined by

$$\Delta_t := \begin{cases} \min\{(\pi^0 - \pi^t)[h(u_0) - T(u_0)x_L], \\ \quad (\pi^t - \pi^0)[h(u^t) - T(u^t)x_L]\} & \text{for } t \leqslant K, \\ \min\{q(v^t)(y^0 - y^t), q(v^0)(y^t - y^0)\} & \text{for } t > K. \end{cases} \tag{43}$$

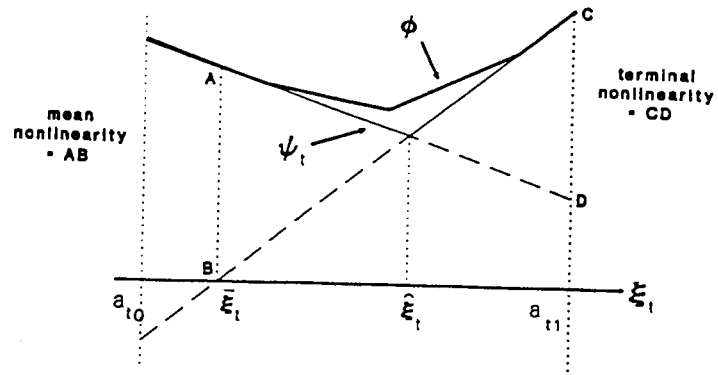

Fig. 1. Two-piece linear approximation.

Observe that (43) is the measure introduced in Frauendorfer and Kall [11] when solving stochastic programs with random right-hand sides. Furthermore, the latter reference considers terminal-nonlinearity among all pairs of *adjacent* vertices. Here, we consider only those that are adjacent to the vertex w^0. Also define the measure *mean-nonlinearity*, denoted $\overline{\Delta}_t$, as the difference of the heights of the two linear pieces defining ψ_t at the (conditional) mean point $\bar{\xi}_t$ for $t \leqslant K$ (or $\bar{\eta}_{t-K}$ for $t > K$); see Fig. 1. Defining

$$\overline{w}^t := \begin{cases} (\bar{u}^t, v^t) & \text{if } 1 \leqslant t \leqslant K, \\ (u^t, \bar{v}^t) & \text{if } K+1 \leqslant t \leqslant K+L, \end{cases}$$

where $\bar{u}^t := (a_{10}, \ldots, \bar{\xi}_t, \ldots, a_{K0})$ for $t \leqslant K$ and $\bar{v}^t := (b_{10}, \ldots, \bar{\eta}_{t-K}, \ldots, b_{L0})$ for $t > K$, let

$$\overline{\Delta}_t := \begin{cases} |(\pi^t - \pi^0)[h(\bar{u}^t) - T(\bar{u}^t)x_L]| & \text{for } t \leqslant K, \\ |q(\bar{v}^t)(y^t - y^0)| & \text{for } t > K. \end{cases} \tag{44}$$

Prior to discussing strategies for partitioning an ω coordinate, it is necessary to determine the *intersection point* of the two linear pieces defining ψ_t; see Fig. 1. One can easily show that this intersection point is

$$\hat{\xi}_t := \frac{(\pi^0 - \pi^t)\left[(h_0 - T_0 x_L) + \sum_{k \neq t}(h_k - T_k x_L)a_{k0}\right]}{(\pi^t - \pi^0)(h_t - T_t x_L)}, \quad t \leqslant K, \tag{45}$$

and

$$\hat{\eta}_t := \frac{\left(q_0 + \sum_{l \neq t-K} q_l b_{l0}\right)(y^0 - y^t)}{q_{t-K}(y^t - y^0)}, \quad K < t \leqslant K+L. \tag{46}$$

For the cell Ω_r $(r = 1, \ldots, R_\nu)$ chosen from the current partition $\mathcal{P}^\nu$ of Ω according to (42) for further sub-division, the coordinate axis to be partitioned to yield sub-rectangles is chosen by one of the following *partitioning strategies*:

Strategy 1: For t_ν which is the solution of $\arg\max_{t \in \mathcal{T}} \Delta_t$, partition the coordinate ω_{t_ν} at the conditional mean-point $\bar{\omega}_{t_\nu}$ (of cell Ω_r).

Strategy 2: For t_ν which is the solution of $\arg\max_{t\in\mathcal{T}} \Delta_t$, partition the coordinate ω_{t_ν} at the intersection-point $\hat{\omega}_{t_\nu}$.

Strategy 3: For t_ν which is the solution of $\arg\max_{t\in\mathcal{T}} \{\Delta_t/\overline{\Delta}_t\}$, partition the coordinate ω_{t_ν} at the conditional mean-point $\overline{\omega}_{t_\nu}$.

Strategy 4: For t_ν which is the solution of $\arg\max_{t\in\mathcal{T}} \{\lambda\Delta_t - (1-\lambda)\overline{\Delta}_t\}$, for some $\lambda \in [0, 1)$, partition the coordinate ω_{t_ν} at the conditional mean-point $\overline{\omega}_{t_\nu}$.

Under a sub-simplicial refinement of the domains Ξ and Θ, Frauendorfer [10] utilized Strategy 1, but with partitioning at the *mid-point* of the corresponding edge. We use the conditional mean point because it probabilistically balances the scenarios in the resulting two cells. However, with partitioning at the (conditional) mean-point, it is preferable to have smaller mean-nonlinearity $\overline{\Delta}_t$ as it implies that the partitioning point is closer to the intersection point $\hat{\omega}_t$. In the latter event, in case ϕ coincides with ψ_t, the lower and upper bounds in the subsequent iteration can be expected to become much closer. This indeed is the motivation for Strategy 2. Strategies 3 and 4 attempt to trade-off mean-nonlinearity with terminal-nonlinearity in two different ways. In the numerical experiments, we investigate Strategy 4 with $\lambda = 0.5$.

The code RAWFC also uses the following *cell redefining strategy* in obtaining the new extreme points after partitioning a cell. When partitioning a cell with a number of discrete realizations (scenarios) in it, the resulting cells may be reduced in volume by considering the smallest rectangle containing all the scenarios; see Fig. 2. Since the (unnecessary) zero probability spaces are removed, the resulting bounds can become tighter. Table 1 reports the effect of *cell-redefining* (C-R) with each coordinate axis being partitioned (at its conditional mean point) from the case of "no partitioning". As

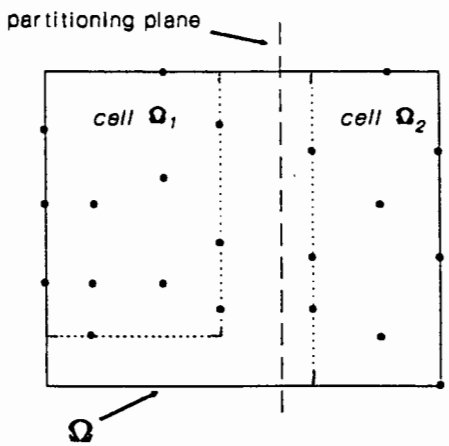

Fig. 2. Redefining cells of a partition.

Table 1
Effect of the Cell-Redefining strategy. $m_1 = m_2 = 5$, $n_1 = n_2 = 10$, $K = L = 2$, # scenarios = 625

Partition axis	No. C–R		with C–R	
	LB	UB	LB	UB
no partition	1008.451	8501.089	1008.451	8501.089
ξ_1	2742.453	7316.946	2742.453	7155.438
ξ_2	3334.557	7316.641	3334.557	7263.926
η_1	1185.088	8437.903	1185.091	8437.799
η_2	1171.392	8483.692	1172.667	8483.684

can be seen from (33) and illustrated by the numerical example in Table 1, the C-R strategy can improve the lower bound only when η coordinates are partitioned.

In view of the discussion on convergence of approximations in Section 3.1, to ensure that the volumes of the cells of partitions become arbitrarily small, one may be motivated to mix the foregoing partitioning strategies with one that chooses a partitioning coordinate having the longest edge, and then, sub-dividing the cell at the *mid-point* of this edge. This was done with simplicial decompositions in Frauendorfer [10, p. 121]. However, with discrete scenarios and C-R strategy in use, there seems to be only little advantage in doing so, since zero probability spaces are successively removed from the resulting partitions. This is also verified in our experiments (although not reported here) that with partitioning at the mid-point the improvement in the relative gap seemed very slow, specially in the initial iterations, compared to partitioning at the conditional mean or the intersection-point.

In this preliminary investigation, we generate complete recourse problems with relatively small number of constraints and variables, but allow the number of scenarios to vary depending on the number of (basic) random variables in the problem. For instance, with 6 random variables, to represent a case where each is allowed to take 4 different outcomes, 4096 dependent scenarios are generated randomly. The problem characteristics in Table 2 will be used for the illustration here. In particular, we wish to pursue the sensitivity of the (rectangular) first/cross moment bounds on partitioning

Table 2
Randomly generated example problems

Problem #	K	L	(m_1, n_1, m_2, n_2)	Uncertainty [a]	Scenarios
1	2	3	(5, 10, 5, 10)	$[-1, 1]$ [b]	1024
2	4	1	(5, 10, 5, 10)	$[-1, 1]$	1024
3	2	3	(5, 10, 5, 10)	$[-5, 5]$	1024
4	4	1	(5, 10, 5, 10)	$[-5, 5]$	1024
5	3	3	(3, 5, 3, 5)	$[-1, 1]$	4096
6	3	3	(3, 5, 3, 5)	$[-5, 5]$	4096
7	2	2	(10, 20, 50, 70)	$[-5, 5]$	625
8	1	1	(50, 100, 50, 100)	$[-5, 5]$	100
9	2	2	(10, 20, 10, 20)	$[-5, -5]$	* [c]

[a] Measured by lower and upper limits of random variables.
[b] All random variables have the same interval.
[c] Number of scenarios is increased from 128 to 4096.

Table 3
Effect of partitioning strategies – Problem #1

Partitioning strategy	Partition cell (coordinate)	LB	$\|x_L^\nu - x_L^{\nu+1}\|_2$	UB	Relative gap (%) [a]
no partition		856.331		1246.739 (1112.073) [b]	45.59
1	1 (ξ_2)	856.520	0.00	1122.138	31.01
	3 (ξ_2)	857.884	1.94026	1022.648	19.21
	2 (ξ_1)	859.324	0.00	970.396	12.93
	5 (ξ_2)	883.610	0.42754	955.705	8.16
	7 (ξ_1)	899.516	0.00	941.553	4.67
2	1 (ξ_2)	884.602	2.36779	1024.811	15.85
	2 (ξ_1)	906.708	0.00	914.620	0.87
3	1 (ξ_1)	865.632	0.00	1156.156	33.56
	3 (ξ_2)	872.572	1.06378	1035.655	18.69
	2 (ξ_2)	872.573	0.00	992.121	13.70
	7 (ξ_2)	872.599	0.00	967.059	10.83
	5 (ξ_2)	875.489	1.30401	939.587	7.32
	4 (ξ_1)	892.966	0.00	924.435	3.52
4	1 (ξ_1)	865.632	0.00	1156.156	33.56
($\lambda = 0.5$)	3 (ξ_1)	894.231	0.56233	1044.496	16.80
	2 (ξ_2)	894.232	0.00	989.721	10.68
	7 (ξ_2)	894.317	0.41588	949.160	6.13
	5 (ξ_2)	900.055	1.38958	923.790	2.53
grand-LP: 5125×10250		LB lp (max.) [c]: 145×570		UB lp (max). [d]: 52×52	

[a] Defined by [(UB-LB)/LB]×100%.
[b] When solving the upper bound (6) for its x solution.
[c] The maximum size of the lower bounding LP solved for computations in this table.
[d] The maximum size of the upper bounding LP solved for computations in this table.

strategies (with using the C-R strategy) when both the dimension of uncertainty as well as the range of uncertainty are enlarged.

In Tables 3–9, improvement in the current x solution, x_L, is also reported for each partitioning iteration. The UB column reports the upper bound when the solution x_L is used. For the "no partitioning" case, the upper bound when solving the model (6) for its solution x_U is provided within parentheses. Cells in a partition are numbered consecutively so that partitioning the original domain (i.e., cell 1) yield cells 2 and 3, with cell 2 referring to the lower side of the coordinate being partitioned. The code automatically terminates when the *relative gap* of the bounds is less than 5%; however, for the reporting purposes, no more than 8 partitioning iterations are tabulated although 20 such iterations are allowed in the code.

As it develops, Strategy 2 with partitioning at the intersection points seems to work best, in our limited computational experience. As in Problem #1 (see Table 3), Strategy 2 is able to take the best advantage of the coordinate that shows the maximum

Table 4
Effect of partitioning strategies – Problem #2

Partitioning strategy	Partition cell (coordinate)	LB	$\|x_L^\nu - x_L^{\nu+1}\|_2$	UB	Relative gap (%)
no partition		144.795		222.192 (213.492)	53.45
1	1 (ξ_2)	146.086	0.01835	205.486	40.66
	3 (ξ_2)	146.187	0.00	193.677	32.49
	2 (ξ_1)	146.685	0.14599	192.906	31.51
	5 (ξ_3)	147.281	0.00	190.134	29.10
	6 (ξ_3)	149.331	0.00	187.662	25.67
	7 (ξ_1)	149.331	0.00	185.328	24.11
	4 (ξ_3)	150.093	0.14599	182.693	21.72
	9 (ξ_1)	151.051	0.00	180.553	19.53
2	1 (ξ_2)	145.795	0.00876	217.473	49.19
	2 (ξ_1)	146.814	0.03731	211.919	44.35
	4 (ξ_3)	149.946	0.06442	186.369	24.29
	6 (ξ_4)	153.280	0.70188	184.561	20.41
	7 (ξ_3)	153.722	0.00	181.259	17.91
	10 (ξ_3)	153.735	0.00	177.972	15.77
	9 (ξ_4)	154.316	0.30877	175.247	13.56
	3 (ξ_1)	158.062	0.00	172.365	9.05
3	1 (ξ_4)	151.539	0.79543	226.668	49.58
	3 (ξ_4)	151.745	0.24114	210.367	38.63
	2 (ξ_3)	151.745	0.00	198.251	30.65
	5 (ξ_1)	152.799	0.00	197.079	28.98
	7 (ξ_3)	152.799	0.00	190.438	24.63
	4 (ξ_3)	153.242	0.31429	185.418	20.99
	8 (ξ_2)	153.242	0.00	183.029	19.44
	6 (ξ_1)	153.242	0.00	181.774	18.62
4	1 (ξ_4)	151.539	0.79542	226.668	49.58
($\lambda = 0.5$)	3 (ξ_2)	151.629	0.12761	216.413	42.73
	2 (ξ_3)	151.629	0.00	203.815	34.42
	5 (ξ_1)	152.233	0.42874	195.553	28.46
	7 (ξ_2)	152.233	0.00	192.224	26.27
	4 (ξ_1)	152.286	0.06973	190.120	24.84
	8 (ξ_3)	152.286	0.00	187.820	23.33
	6 (ξ_2)	152.285	0.00	186.090	22.20
grand-LP: 5125×10250		LB lp (max.): 95×190		UB lp (max.): 122×170	

terminal-nonlinearity, compared to partitioning at the conditional mean of the same coordinate. Also it developed that the range of uncertainty or the dimension of uncertainty has only little effect on the performance of Strategy 2, while Strategy 3 displayed the poorest performance on average. Comparison of Strategy 1 and 2 on slightly larger problems is given in Table 9.

Table 5
Effect of range of uncertainty – Problem #3

Partitioning strategy	Partition cell (coordinate)	LB	$\|x_L^\nu - x_L^{\nu+1}\|_2$	UB	Relative gap (%)
no partition		2178.092		3542.409	62.64
				(3072.993)	
1	1 (ξ_1)	2224.170	0.01891	3209.889	44.32
	3 (ξ_1)	2327.952	0.00	3028.499	30.09
	2 (ξ_2)	2422.207	1.46132	2654.958	9.61
	5 (ξ_1)	2457.682	0.01311	2610.022	6.29
	4 (ξ_2)	2472.582	0.03360	2539.100	2.69
2	1 (ξ_1)	2341.368	0.01183	3037.838	29.75
	2 (ξ_2)	2447.920	2.20355	2683.243	9.61
	5 (ξ_2)	2468.124	0.61415	2567.749	4.04
3	1 (ξ_2)	2334.306	1.37424	3039.043	30.19
	2 (ξ_1)	2339.297	0.20459	2855.773	22.08
	3 (ξ_1)	2339.333	0.00188	2688.268	14.92
	5 (ξ_1)	2395.096	0.27543	2631.221	9.86
	7 (ξ_1)	2447.931	0.09590	2561.316	4.63
4	1 (ξ_1)	2224.170	0.01891	3209.889	44.32
($\lambda = 0.5$)	3 (ξ_1)	2327.952	0.00	3028.499	30.09
	2 (ξ_2)	2422.207	1.46132	2654.958	9.61
	5 (ξ_2)	2436.839	0.00	2618.432	7.45
	4 (ξ_2)	2447.300	0.13528	2567.738	4.92
grand-LP: 5125 × 10250		LB lp (max.): 125 × 490		UB lp (max.): 52 × 52	

As for x solutions, much of the adjustment in the approximate x solution seems to occur in the initial iterations and thereafter only minor changes take place. This raises the question whether convergence of the x solutions takes place much earlier than that of the bounds. If it is so, one may use it to the computational advantage by performing multiple partitions, i.e., partitioning multiple cells for the current lower bounding solution x_L^ν.

To demonstrate the effectiveness with multiple partitioning, we use Problem #9 (see Table 1) which has a starting set of 128 scenarios. We successively increased the number of scenarios by adding more scenarios to the existing set, up to a total of 4096 scenarios. These problems were solved by RAWFC using partitioning strategy 2 and multiple partitioning of cells. We arbitrarily fixed the multiple-partitioning strategy to be the one which picks all cells whose $\delta(\Omega_r)$, the weighted-difference in (42), is within 60% of the $\max_{r=1,\ldots,R_\nu}\{\delta(\Omega_r)\}$. These problems were solved until the maximum partition-iteration limit of 20 is reached in RAWFC. The grand-LPs corresponding to these stochastic programs were also solved to optimality using OSL. All tests were run on a SUN SparcStation 2 and the CPU times are compared. It is evident from the results in Table 10 that RAWFC determines near-optimal solutions in a fraction of the time required to solve the grand-LP. Although the grand-LP cpu time corresponds to solving

Table 6
Effect of range of uncertainty – Problem #4

Partitioning strategy	Partition cell (coordinate)	LB	$\|x_L^\nu - x_L^{\nu+1}\|_2$	UB	Relative gap (%)
no partition		1023.689		3100.624	202.89
				(1861.573)	
1	1 (ξ_2)	1040.383	2.11049	2299.015	120.98
	3 (ξ_1)	1040.901	0.00	2235.983	114.81
	2 (ξ_3)	1044.882	0.17584	2114.645	102.38
	4 (ξ_3)	1057.836	2.22791	1611.486	52.34
	5 (ξ_2)	1072.128	0.00	1556.558	45.18
	7 (ξ_1)	1072.129	0.00	1532.228	42.92
	6 (ξ_1)	1072.483	0.15489	1516.282	41.38
	9 (ξ_2)	1105.535	0.30643	1533.398	38.70
2	1 (ξ_2)	1079.642	0.93756	2657.161	146.12
	2 (ξ_3)	1089.890	1.13427	1974.815	81.19
	4 (ξ_1)	1109.554	0.00	1930.428	73.98
	6 (ξ_2)	1112.405	0.00	1853.424	66.61
	8 (ξ_2)	1116.665	0.00	1793.880	60.65
	5 (ξ_1)	1132.132	0.91346	1610.099	42.22
	3 (ξ_1)	1166.390	1.55856	1390.326	19.20
	12 (η_1)	1180.348	0.49274	1381.181	17.02
3	1 (ξ_4)	1044.587	1.84099	2472.550	136.70
	2 (ξ_1)	1059.750	2.46087	1838.251	73.46
	3 (ξ_3)	1059.802	0.00	1795.004	69.37
	7 (ξ_1)	1060.290	0.00	1769.209	66.86
	4 (ξ_3)	1060.290	0.00	1736.550	63.78
	5 (ξ_2)	1060.445	0.14488	1697.038	60.03
	6 (ξ_1)	1060.445	0.00	1667.349	57.23
	8 (ξ_2)	1063.502	0.00	1636.700	53.90
4	1 (ξ_3)	1034.106	0.79917	2385.687	130.70
($\lambda = 0.5$)	3 (ξ_1)	1039.594	0.93251	2170.292	108.76
	2 (ξ_2)	1044.537	0.20920	2030.037	94.35
	4 (ξ_3)	1061.535	1.98828	1663.134	56.67
	5 (ξ_2)	1061.658	0.00	1617.429	52.35
	7 (ξ_1)	1063.568	0.17501	1588.981	49.40
	6 (ξ_1)	1063.589	0.01192	1562.897	46.95
	9 (ξ_2)	1088.189	0.62206	1481.365	36.13
grand-LP: 5125 × 10250		LB lp (max.): 95 × 190		UB lp (max.): 112 × 170	

until optimality is reached, a premature termination with an objective value as good as the best upper bound (produced by RAWFC) still required much larger cpu time. For instance, in the case of 1024 scenarios, to obtain the best RAWFC-upper bound of 2805.28, over 95% of the total OSL-iterations (which produces the grand-LP optimal value of $Z^* = 2676.32$) are required. Notice that the current lower bounding solution

Table 7
Effect of partitioning strategies – Problem #5

Partitioning strategy	Partition cell (coordinate)	LB	$\|x_L^r - x_L^{r+1}\|_2$	UB	Relative gap (%)
no partition		80.812		243.986 (120.181)	201.92
1	1 (ξ_2)	84.968	0.49872	181.327	113.41
	2 (ξ_3)	91.518	0.73575	121.086	32.31
	3 (ξ_1)	91.518	0.00	120.710	31.90
	4 (ξ_3)	92.759	0.14974	113.123	21.95
	7 (ξ_3)	92.759	0.00	111.025	19.69
	6 (ξ_1)	92.759	0.00	108.727	17.22
	5 (ξ_1)	92.759	0.00	106.904	15.25
	8 (ξ_2)	93.275	0.06250	105.664	13.28
2	1 (ξ_2)	82.221	0.16931	203.022	146.92
	2 (ξ_3)	90.252	0.92912	125.051	38.56
	4 (ξ_2)	90.960	0.07435	121.436	33.50
	6 (ξ_3)	92.755	0.20943	112.304	21.07
	5 (ξ_1)	92.755	0.00	107.650	16.06
	9 (ξ_1)	92.755	0.00	105.165	13.38
	3 (ξ_1)	93.122	0.00	103.704	11.36
	8 (ξ_2)	93.498	0.04732	102.950	10.11
3	1 (ξ_3)	89.706	1.05385	129.837	44.73
	2 (ξ_3)	92.024	0.24927	116.008	26.06
	3 (ξ_1)	92.023	0.00	112.243	21.97
	4 (ξ_3)	92.981	0.11209	108.895	17.12
	7 (ξ_1)	93.173	0.00	106.472	14.27
	5 (ξ_1)	93.172	0.00	103.814	11.42
	8 (ξ_2)	93.390	0.02172	103.555	10.89
	6 (ξ_3)	93.389	0.00	102.157	9.39
4 ($\lambda = 0.5$)	1 (ξ_3)	89.706	1.05385	129.837	44.73
	2 (ξ_3)	92.024	0.24927	116.008	26.06
	3 (ξ_1)	92.023	0.00	112.243	21.97
	4 (ξ_3)	92.981	0.11209	108.895	17.12
	7 (ξ_1)	93.173	0.00	106.472	14.27
	5 (ξ_1)	93.172	0.00	103.814	11.42
	8 (ξ_2)	93.390	0.02172	103.555	10.89
	6 (ξ_3)	93.389	0.00	102.157	9.39
grand-LP: 12291 × 20485		LB lp (max.): 111 × 365		UB lp (max.): 88 × 56	

from RAWFC is within 5% of the optimal solution of the latter problem, although the relative gap of RAWFC-bounds is 8.8%. Also it appears that as the number of scenarios is increased, the performance of RAWFC does not deteriorate, but rather, it displayed an enhanced performance.

The preceding numerical exercises ascertain that while a stochastic program may be described with a large set of scenarios, by clustering scenarios in an appropriate manner,

Table 8
Effect of range of uncertainty – Problem #6

Partitioning strategy	Partition cell (coordinate)	LB	$\|x_L^r - x_L^{r+1}\|_2$	UB	Relative gap (%)
no partition		871.748		2064.859 (1516.065)	136.864
1	1 (ξ_3)	1022.027	0.11975	1435.897	40.50
	2 (ξ_3)	1055.916	0.02838	1393.596	31.98
	3 (ξ_3)	1055.916	0.00	1312.768	24.33
	7 (ξ_1)	1059.123	0.00	1292.667	22.05
	4 (ξ_1)	1064.698	0.01235	1281.754	20.39
	5 (ξ_1)	1064.698	0.00	1240.955	16.56
	10 (ξ_3)	1070.967	0.00586	1226.731	14.54
	8 (ξ_3)	1073.605	0.00	1211.105	12.81
2	1 (ξ_3)	961.983	0.07162	1514.698	57.46
	2 (ξ_3)	1055.810	0.07643	1341.693	27.08
	5 (ξ_1)	1055.807	0.00	1268.042	20.10
	3 (ξ_1)	1070.638	0.00	1226.240	14.53
	4 (ξ_1)	1072.558	0.00190	1220.183	13.76
	11 (ξ_1)	1074.079	0.00132	1217.744	13.38
	13 (ξ_1)	1075.686	0.00137	1215.930	13.04
	15 (ξ_1)	1076.663	0.01640	1215.886	12.93
3	1 (ξ_3)	1022.027	0.11975	1435.897	40.50
	2 (ξ_3)	1055.916	0.02838	1393.596	31.98
	3 (ξ_3)	1055.916	0.00	1312.768	24.33
	7 (ξ_3)	1067.710	0.00	1274.824	19.40
	4 (ξ_2)	1069.183	0.00373	1270.639	18.84
	5 (ξ_1)	1069.184	0.00	1230.382	15.08
	10 (ξ_2)	1072.660	0.00708	1229.15	14.59
	9 (ξ_1)	1080.479	0.00	1216.974	12.63
4 ($\lambda = 0.5$)	1 (ξ_3)	1022.027	0.11975	1435.897	40.50
	2 (ξ_3)	1055.916	0.02838	1393.596	31.98
	3 (ξ_3)	1055.916	0.00	1312.768	24.33
	7 (ξ_1)	1059.123	0.00	1292.667	22.05
	4 (ξ_2)	1060.734	0.00373	1288.458	21.47
	5 (ξ_1)	1060.735	0.00	1248.201	17.67
	10 (ξ_2)	1064.471	0.00708	1246.892	17.14
	8 (ξ_3)	1067.048	0.00	1231.277	15.39
grand-LP: 12291 × 20485		LB lp (max.): 111 × 365		UB lp (max.): 88 × 56	

representative scenarios may be determined for each cluster in order to compute near-optimal solutions effectively. This is illustrated for Problem #2 in Figs. 3 and 4, respectively, for Strategies 1 and 2. The 'boxes' in these figures represent the final cell numbers (with the number of scenarios in parentheses) at $\nu = 8$. Observe that even with a cluster of as large as 558 scenarios in cell 8, Strategy 2 yields a relative gap of about 9% compared to a gap of 20% with Strategy 1, after 8 clustering steps. Thus, what is

Table 9
Strategies 1 and 2 on larger problems – Problem #7 and Problem #8

Problem	Partition cell (coord.)	LB	$\|x_L^r - x_L^{r+1}\|_2$	UB	Relative gap (%)
#7	no partition	237801.594		336920.640 (284051.469)	41.68
	$1(\xi_1)$	238006.141 [a]	0.60591	283427.226	19.08
	$3(\xi_1)$	238639.844	3.60666	261606.010	9.62
	$5(\xi_1)$	244671.875	3.37458	252938.194	3.38
	$1(\xi_1)$	238756.781 [b]	0.26646	271942.337	13.90
	$2(\xi_1)$	238963.719	0.25248	259744.366	8.70
	$3(\xi_2)$	242432.875	19.99126	253447.170	4.11
grand-LP: 31260×43770		LB lp (max.): 610×1140		UB lp (max.): 216×289	
#8	no partition	25357.725		138993.720 (54866.770)	448.13
	$1(\xi_1)$	29853.332 [c]	18.28679	76887.134	157.55
	$3(\xi_1)$	33303.734	20.46054	47013.183	41.17
	$5(\xi_1)$	36207.090	1.58927	40571.068	12.05
	$2(\xi_1)$	37587.883	4.81065	39310.941	4.58
	$1(\xi_1)$	31046.775 [d]	34.19303	50674.521	63.22
	$2(\xi_1)$	35362.020	5.14173	45535.817	28.77
	$3(\xi_1)$	37182.500	6.31890	40196.565	8.11
	$4(\xi_1)$	37551.168	5.77253	39031.668	3.94
grand-LP: 5050×10100		LB lp (max.): 550×1100		UB lp (max.): 104×204	

[a] These figures are for Strategy 1.
[b] These figures are for Strategy 2.
[c] These figures are for Strategy 1.
[d] These figures are for Strategy 2.

Table 10
Effect of increasing # of scenarios with multiple partitioning – Problem #9

# scen.	Relative gap (%)	# multiple partition iter. [a]	RAWFC CPU sec.	grand-LP size	grand-LP CPU sec. [b]
128	3.95	6	390.61	190×2580	220.88
256	9.89	6	339.71	2570×5140	1042.19
512	7.08	6	384.41	5130×10260	2396.48
1024	8.80	6	435.13	10250×20500	12447.50
2048	2.80	5	333.15	20490×40980	* [c]
4096	3.26	5	136.60	40970×81940	*

[a] Total number of single partitions reached the allowable maximum of 20.
[b] Solving grand-LP to optimality using OSL.
[c] Not solved due to memory limitations.

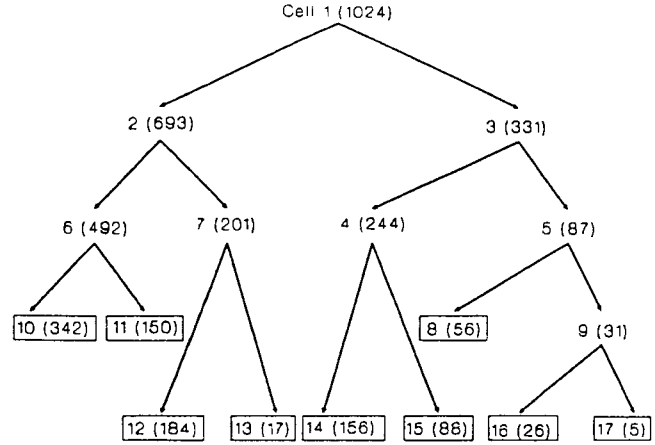

Fig. 3. Scenario clustering of Strategy 1, Problem 2.

paramount in bounds-based sequential approximation methods is the proper choice of a refinement procedure to cluster scenarios based on partitioning the underlying domain. The strategies described in this section illustrate this point.

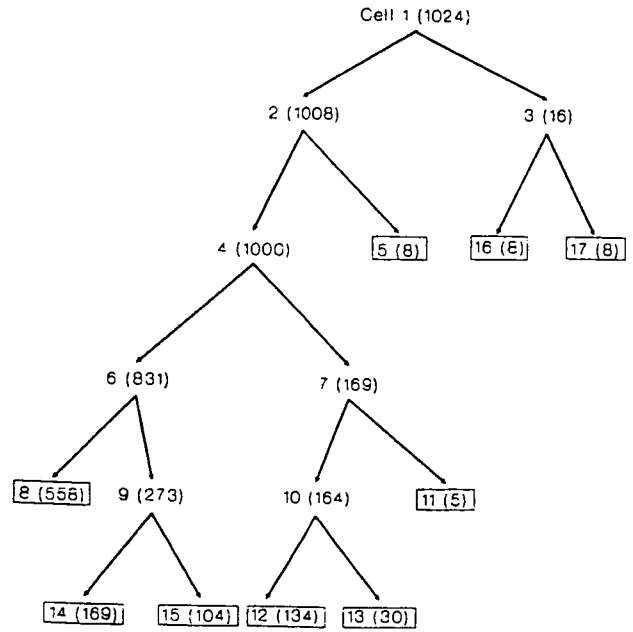

Fig. 4. Scenario clustering of Strategy 2, Problem 2.

6. Concluding remarks

It was demonstrated how stochastic programs can be solved using bounds-based approximations, in this case, utilizing rectangular domains and first/cross moment bounds. For this purpose, monotonicity and convergence of the bounds were first established, and then decomposition of the bounding computations was pursued for assuring efficiency in solution. In particular, the lower bounding computations enjoy the partitioning convenience offered by rectangular domains while affording only a linear growth in the model size with each partition.

However, evaluation of the upper approximation still involves an exponential number of sub-problems which may offset the partitioning convenience in rectangular domains, especially when solving problems with larger dimensions of uncertainty. One possible remedy is to restrict the domains to be simplices, thereby assuring a linear growth in the upper bounding model size with partitioning. Frauendorfer [10] derived such simplicial bounds which can be computed without any optimization on probability measures, and thus, the resulting approximations too possess the required decomposable structure. However, refining these approximations using a partitioning procedure is more cumbersome compared to that in rectangular domains. See Frauendorfer [10] for details of performing a simplicial decomposition of the domains for refinement of these bounds.

Alternatively, Edirisinghe [4] derives simplicial bounds on the joint domain using the complete second moment information, which are thus generally tighter than the first/cross moment approximations on simplicial domains. Edirisinghe and You [7] report an implementation of these approximations for solving two-stage stochastic problems more efficiently.

Acknowledgements

The authors would like to thank the two anonymous referees for their numerous suggestions to improve the presentation of the paper. Computer programming assistance received from G.-M. You at the University of Tennessee is thankfully acknowledged. This research was partially supported by the NSERC grant 5-87147.

References

[1] H. Attouch and R.J-B. Wets, "Approximation and convergence in nonlinear optimization," in: O. Mangasarian, R. Meyer and S. Robinson, eds., *Nonlinear Programming 4* (Academic Press, New York, 1981) pp. 367–394.

[2] J.R. Birge and R.J.-B. Wets, "Designing approximation schemes for stochastic optimization problems, in particular for stochastic programs with recourse," *Mathematical Programming Study* 27 (1986) 54–102.

[3] S.P. Bradley, A.C. Hax and T.L. Magnanti, *Applied Mathematical Programming* (Addison-Wesley, Reading, MA, 1977).

[4] N.C.P. Edirisinghe, "New second-order approximations in stochastic linear programming," Working Paper, College of Business, University of Tennessee (1993); to appear in *Operations Research.*

[5] N.C.P. Edirisinghe and W.T. Ziemba, "Bounds for two-stage stochastic programs with fixed recourse," *Mathematics of Operations Research* 19 (1994) 292–313.

[6] N.C.P. Edirisinghe and W.T. Ziemba, "Bounding the expectation of a saddle function, with application to stochastic programming," *Mathematics of Operations Research* 19 (1994) 314–340.

[7] N.C.P. Edirisinghe and G.-M. You, "Second-order scenario approximation and refinement in optimization under uncertainty," *Annals of Operations Research* 64 (1996) 143–178.

[8] Y. Ermoliev and R.J.-B. Wets, *Numerical Techniques for Stochastic Optimization* (Springer, New York, 1988).

[9] K. Frauendorfer, "SLP problems: Objective and right-hand side being stochastically dependent – Part II," Manuscript, IOR, University of Zurich (1988).

[10] K. Frauendorfer, *Stochastic Two-Stage Programming* (Springer, Berlin, 1992).

[11] K. Frauendorfer and P. Kall, "A solution method for SLP problems with arbitrary multivariate distributions – The independent case," *Problems of Control and Information Theory* 17 (1988) 177–205.

[12] C.C. Huang, W.T. Ziemba, and A. Ben-Tal, "Bounds on the expectation of a convex function of a random variable: With applications to stochastic programming," *Operations Research* 25 (1977) 315–325.

[13] P. Kall, *Stochastic Linear Programming* (Springer, Berlin, 1976).

[14] P. Kall, "Stochastic programming with recourse: Upper bounds and moment problems – A review," in: J. Guddat, P. Kall, K. Lommatzsch, M. Vlach and K. Zimmermann, eds., *Advances in Mathematical Optimization and Related Topics* (Akademie Verlag, Berlin, 1988) pp. 86–103.

[15] P. Kall and E. Keller, "GENSLP: A program for generating input for stochastic linear programs with complete fixed recourse," Manuscript, Institut für Operations Research der Universität Zürich, Zürich CH-8006, Switzerland (1985).

[16] R. Van Slyke and R.J.-B. Wets, "L-shaped linear programs with application to optimal control and stochastic optimization," *SIAM Journal on Applied Mathematics* 17 (1969) 638–663.

[17] R.J-B. Wets, "Convergence of convex functions, variational inequalities and convex optimization problems," in: R.W. Cottle, F. Giannesi and J-L. Lious, eds., *Variational Inequalities and Complimentary Problems* (Wiley, New York, 1980) pp. 375–403.

[18] R.J-B. Wets, "Stochastic programming: Solution techniques and approximation schemes," in: A. Bachem, M. Groetschel and B. Korte, eds., *Mathematical Programming: State-of-the-art* (Springer, Berlin, 1982) pp. 566–603.