STOCHASTIC DECOMPOSITION: AN ALGORITHM FOR TWO-STAGE LINEAR PROGRAMS WITH RECURSE*†

JULIA L. HIGLE AND SUVRAJEET SEN

We present a cutting plane algorithm for two-stage stochastic linear programs with recourse. Motivated by Benders' decomposition, our method uses randomly generated observations of random variables to construct statistical estimates of supports of the objective function. In general, the resulting piecewise linear approximations do not agree with the objective function in finite time. However, certain subsequences of the estimated supports are shown to accumulate at supports of the objective function, with probability one. From this, we establish the convergence of the algorithm under relatively mild assumptions.

1. Introduction. One of the more common stochastic optimization problems is the two-stage linear program with recourse. This problem can be stated as follows:

\[
\begin{align*}
\text{Min} & \quad f(x) = cx + E[h(x, \omega)] \\
\text{s.t.} & \quad x \in X
\end{align*}
\]

where \( X \) is a convex polyhedral set and

\[
\begin{align*}
h(x, \omega) = \text{Min} & \quad gy \\
\text{s.t.} & \quad Wy = \omega - Tx, \\
& \quad y \geq 0.
\end{align*}
\]

The decision variable \( x \) represents a decision that must be implemented prior to the realization of the multi-dimensional random variable \( \omega \). The evaluation of the function \( h \) involves the determination of the optimal recourse variable \( y \), given the initial decision \( x \) and the random variable \( \omega \). Thus, recourse problems involve the solution of adaptive optimization problems.

In this paper, the random variable \( \omega \) is defined on a probability space \( (\Omega, \mathcal{A}, \mathbb{P}) \) with corresponding distribution function \( F_\omega \). We assume that \( h \) is finite on \( X \times \Omega \). That is, we assume that \( (P) \) has the complete recourse property (Wets [1982]). While there is some loss in generality as a result of this assumption, there are many situations in which the assumption of complete recourse is realistic. For example, in many production planning problems, if the demand exceeds the amount produced, it is generally possible to satisfy demand by subcontracting it. Naturally, this option tends to be costly.

Two-stage stochastic programs with recourse are applicable to a variety of commonly occurring decision making problems. For example, in many capacity planning problems (e.g. Murphy, Sen and Soyster [1982]), the choice of cost effective facilities

*Received September 21, 1987; revised April 28, 1990.
OR/MS Index 1978 subject classification. Primary 663: Programming/stochastic.
Key words. Stochastic programming, decomposition, cutting plane methods.

†This work was supported in part by Grant No. AFOSR-88-0076 from the Air Force Office of Scientific Research and Grant No. DDM-89-10046 from the National Science Foundation.

650

0364-765X/91/1603/0650/$01.25
Copyright © 1991, The Institute of Management Sciences/Operations Research Society of America
must be made prior to the realization of random demands. In this situation, $x$ denotes a vector of acquired capacities. The actual deployment of facilities gives rise to the recourse problem which is solved after the outcome (demand) is known. Other examples include models in production planning, facilities location etc. (Louveaux [1986]).

The solution method presented in this paper, Stochastic Decomposition, is an attempt to capture, in one algorithm, key properties of two classes of algorithms: (a) decomposition based methods, and (b) methods motivated by stochastic approximation as they apply to problem ($P$). More general stochastic optimization problems, such as those addressed by Rockafellar and Wets [1987], are beyond the scope of our development.

As an example, consider a classic representative of the first class: the L-shaped method presented by Van Slyke and Wets [1969]. This method is essentially motivated by cutting plane methods for deterministic problems. Hence, traditional mathematical programming constructs (such as dual variables) are readily available, and as a result, so are bounds and objective function estimates. However, one must acknowledge that each iterative step is computationally burdensome, especially in cases where the number of possible outcomes is large. Some computational expedients such as bunching and/or sifting have been proposed by Gartksa and Rutenberg [1973] to take advantage of piecewise linear recourse functions, when $\omega$ has finite support. Despite some savings, the number of subproblems solved can still be large. A slightly weaker criticism is that the method is well defined only for discrete random variables. In reality, if one allows a very fine discretization of a continuous random variable, this is not a technical shortcoming. However, given our earlier comment regarding the relationship between the size of the sample space and the computational effort required, discretization of a continuous sample space can result in a handicap. Similar observations can be made with regard to more recent extensions of the L-shaped method (Birge and Louveaux [1988]). Other algorithms motivated by the L-shaped method include those by Fraendorfer and Kall [1986] and Fraendorfer [1988]. With these algorithms, continuous distributions are approximated by increasingly finer discretizations. The approximations use both upper and lower bounds on the recourse function. A major step in the method involves a refinement procedure which successively improves the quality of the approximations.

Next, consider a representative of the second class: stochastic quasi-gradient (SQG) methods. Unlike the L-shaped method and its relatives, SQG methods treat both discrete and continuous random variables with equal ease. Due to their use of only one sample point at any iteration, SQG methods permit recursive on-line applications (Ermoliev [1983]). However, there are several procedural handicaps associated with SQG methods. First, there are currently no prescriptions for efficient step length rules. While convergence can be proven under lenient conditions, algorithmic choice of an appropriate step length is still an open question. Second, SQG methods usually do not provide an estimate of the objective value during the iterative scheme. This is especially unsatisfactory, because in practice, one must support a recommended plan (optimal solution) with its associated cost. This handicap is probably related to a somewhat more theoretical drawback: SQG algorithms suffer from a lack of algorithmically implemented stopping rules (Pflug [1988], and Gaivoronski [1988]). A recent recursive quadratic programming algorithm due to Ruszcynski [1987] addresses some of these issues, but questions regarding appropriate step lengths persist. Motivated by SQG methods, Gaivoronski and Nazareth [1989] offer a conceptual method for combining sampling techniques with grid linearization, for which convergence is based on successively refined statistical estimates of the objective value at each grid point.
The method of stochastic decomposition (SD) attempts to unite algorithmic concepts from both classes of algorithms mentioned above. Like decomposition based successive approximation methods, the new method iteratively generates a sequence of piecewise linear approximations, and like its deterministic counterpart, SD retains information regarding the nature of the nonlinear objective function, $f(x) = cx + E[h(x, \bar{\omega})]$, about the current iterate. On the other hand, like SQG methods, each iteration requires the solution of the subproblem at one sample point. Furthermore, no distributional assumptions are necessary for the application of SD. We believe that this marriage between stochastic and deterministic methods is a novel stochastic programming concept.

This paper is organized as follows. In §2 we outline the basic method and investigate its limiting properties. The SD algorithm is presented in §3. It is an extension of the basic method which results in stronger limiting properties. Termination criteria are discussed in §4 and computational experience is reported in §5.

2. The fundamental idea. Each iteration of the basic algorithm involves the solution of one subproblem and one linear master problem. The former is used to recursively update a piecewise linear approximation of the recourse function, while the latter is used as a means to generate successive iterates $\{x^k\}$. Given $x^k$ at iteration $k$, one specifies a subproblem using a randomly generated observation of $\bar{\omega}, \omega^k$.

Subproblem ($S^k$).

$$h(x^k, \omega^k) = \min \quad gy$$

$$\text{s.t. } Wy = \omega^k - Tx^k,$$

$$y \geq 0.$$ The dual of this subproblem is represented as follows:

$$h(x^k, \omega^k) = \max \quad \pi(\omega^k - Tx^k)$$

$$\text{s.t. } \pi W \leq g.$$ A solution to the subproblem, together with a history of past iterations, is then used to obtain an estimate of a support of the objective function $cx + E[h(x, \bar{\omega})]$. The piecewise linear approximation used after $k$ iterations is given by

$$f_k(x) = \max \{\alpha^k_t + (\beta^k_t + c) x | t = 1, \ldots, k\}.$$ It should be emphasized that, unlike other outer approximation methods, this piecewise linear approximation is not necessarily defined from supports of the objective function. Furthermore, all segments of the approximation are updated as the iterations progress, and hence the use of two descriptors $(t, k, t \leq k)$ in denoting the cuts. The structure of the master problem is as follows:

Master Problem ($M^k$).

$$\min \quad f_k(x)$$

$$\text{s.t. } x \in X.$$ The solution to the $k$th master problem, denoted $x^{k+1}$, is then used as an input to the next subproblem. Further details are given in §2.2.
2.1. Notation and assumptions. The following notation is used throughout this paper:

\[ (x^k)_{k=1}^\infty \equiv \text{the sequence of points identified,} \]
\[ (\omega^k)_{k=1}^\infty \equiv \text{the sequence of independent and identically distributed observations of } \tilde{\omega}. \]
\[ V \equiv \text{the set of all subproblem dual vertices.} \]
\[ V_k \equiv \text{the set of subproblem dual vertices identified in the first } k \text{ iterations } (V_k \subseteq V). \]
\[ \pi \equiv \text{an extreme point in the subproblem dual (} \pi \in V). \]
\[ f(x) = cx + E[h(x, \tilde{\omega})]. \]
\[ h(x, \omega) = \text{Max}\{\pi(\omega - Tx) | \pi \in V\}. \]
\[ \pi(x, \omega) \in \text{argmax}\{\pi(\omega - TX) | \pi \in V\}. \]
\[ f_k(x) = \text{Max}\{\alpha_k^k + (\beta_k^k + c)x | t = 1, \ldots, k\}. \]
\[ h_k(x, \omega) = \text{Max}\{\pi(\omega - TX) | \pi \in V_k\}. \]
\[ \pi^k_i \in \text{argmax}\{\pi(\omega' - TXk) | \pi \in V_k\}. \]
\[ \Omega \equiv \text{support of the random variable } \tilde{\omega}. \]
\[ F_{\tilde{\omega}} \equiv \text{probability distribution function associated with the random variable } \tilde{\omega}. \]
\[ X \equiv \text{the feasible region associated with the decision variable } x. \]

Unless otherwise specified, we make the following assumptions:

A1. The feasible region associated with the dual of the recourse subproblem is a nonempty compact convex polyhedral set.

A2. X and \( \Omega \) are nonempty compact sets under the Euclidean metric.

A3. For all \( x \in X \), \( h(x, \omega) \geq 0 \) wp1.

Note that the combination of A1 and A2 implies that \( \{h(x, \omega), x \in X, \omega \in \Omega\} \) is contained in a compact set.

2.2. The basic method. Consider the following algorithm.

**Step 0.** Initialize. \( k \leftarrow 0, V_0 \leftarrow \emptyset, \omega^0 \leftarrow E[\tilde{\omega}], \) and
\[
x^1 \in \text{argmin}\{cx + h(x, \omega^0) | x \in X\}.
\]

**Step 1.** \( k \leftarrow k + 1. \) Randomly generate \( \omega^k \) according to the distribution \( F_{\tilde{\omega}}. \)
(Note that \( \omega^t, t = 1, \ldots, k \) are generated independently.)

**Step 2.** Solve \( S_k \) and obtain an associated dual vertex \( \pi(x^k, \omega^k) \).
\( V_k \leftarrow V_{k-1} \cup \{\pi(x^k, \omega^k)\}. \)

**Step 3.** a. Construct the coefficients of the \( k \)th cut to be added to the master problem.
\[
\alpha_k^k + (\beta_k^k + c)x \equiv cx + \frac{1}{k} \sum_{i=1}^{k} \pi_i^k(\omega' - Tx).
\]
(Recall that \( \pi_i^k \in \text{argmax}\{\pi(\omega' - TXk) | \pi \in V_k\}. \))

b. Update the previously generated cuts:
\[
\alpha_i^k \leftarrow \frac{k - 1}{k} \alpha_i^{k-1}, \quad \beta_i^k \leftarrow \frac{k - 1}{k} \beta_i^{k-1}, \quad t = 1, \ldots, k - 1.
\]

**Step 4.** Solve the \( k \)th master problem, \( M^k \), to obtain \( x^{k+1} \). Repeat from Step 1.

Note that \( V_k \subseteq V \). As a result,
\[
\text{Max}\{\pi(\omega' - TXk) | \pi \in V_k\} \leq \text{Max}\{\pi(\omega' - TXk) | \pi \in V\}
\]
\[
\Rightarrow \pi_i^k(\omega' - TXk) \leq h(x^k, \omega').
\]
It follows that
\[ \frac{1}{k} \sum_{t=1}^{k} \pi^k_t (\omega^t - T x^k) \leq \frac{1}{k} \sum_{t=1}^{k} h(x^k, \omega^t), \]

implying that the cuts generated in Step 3a of the algorithm are derived from a statistically based estimate of a lower bound on the objective value, \( f(x^k) = cx^k + E[h(x^k, \bar{\omega})] \). Furthermore, \( \pi(\omega^t - T x) \leq h(x, \omega^t) \) whenever \( \pi \) is a feasible solution to the subproblem dual. Thus, for every \( x \in X \)
\[ \frac{1}{k} \sum_{t=1}^{k} \pi^k_t (\omega^t - T x) \leq \frac{1}{k} \sum_{t=1}^{k} h(x, \omega^t), \]

and the cut generated during the \( k \)th iteration provides a statistically valid lower bound on the objective function. As iterations progress, more observations of \( \bar{\omega} \) become available. In order to maintain consistency in the amount of information contained in the various cuts that have been generated, they must be updated. During iteration \( k + m \) (\( m > 0 \)), \( h(x, \bar{\omega}) \geq 0 \) (wp1) implies that
\[ \frac{1}{k + m} \sum_{t=1}^{k} \pi^k_t (\omega^t - T x) \leq \frac{1}{k + m} \sum_{t=1}^{k} h(x, \omega^t) \]
\[ \leq \frac{1}{k + m} \sum_{t=1}^{k+m} h(x, \omega^t), \quad \text{(wp1)}. \]

Thus, in iteration \( k + m \), the function
\[ cx + \frac{1}{k + m} \sum_{t=1}^{k} \pi^k_t (\omega^t - T x) = \alpha^{k+m} + (\beta^{k+m} + c) x \]
still provides a statistically valid lower bound on the objective function. This is precisely the result of the repeated application of the update described in Step 3b of the method. Note that any updating scheme that is consistent with the requirement of dual feasibility of the subproblem will result in a statistically valid lower bound. The update presented in Step 3b is recommended solely for its simplicity and ease of implementation.

2.3. Limiting behavior. As mentioned in the previous section, the algorithm revolves around the generation of cutting planes that provide statistically valid lower bounds on the objective function. The limiting behavior of the algorithm depends critically on the limiting behavior of these cuts. We first investigate the limiting nature of the functions, \( h_k(x, \omega) = \text{Max}\{\pi(\omega - T x) | \pi \in V_k\} \), from which the cuts are derived. The following results provide the foundation for most of what follows.

**Lemma 1.** The sequence of functions \( \{h_k\}_{k=1}^{\infty} \) converges uniformly on \( X \).

**Proof.** Note that \( V_k \subseteq V_{k+1} \subseteq V \) implies that \( h_k(x, \omega) \leq h_{k+1}(x, \omega) \leq h(x, \omega) \) for all \( k \) and for all \((x, \omega) \in X \times \Omega\). Since \( \{h_k\}_{k=1}^{\infty} \) increases monotonically and is bounded from above by the finite function \( h \), it follows that \( \{h_k\}_{k=1}^{\infty} \) converges pointwise to some function \( g \leq h \). Since \( V_k \subseteq V \), \( V_k \subseteq V_k+1 \subseteq V \) for all \( k \),
\[ \bar{V} = \lim_{k \to \infty} V_k \subseteq V. \]
By assumption A1, \( V \) is a finite set and thus so is \( \bar{V} \). Hence

\[
g(x, \omega) = \lim_{k \to \infty} h_k(x, \omega) = \lim_{k \to \infty} \{ \text{Max} \{ \pi(\omega - Tx) | \pi \in V_k \} \}
\]

\[
= \text{Max} \{ \pi(\omega - Tx) | \pi \in \bar{V} \},
\]

and it follows that \( g \) is a continuous function. Since \( X \times \Omega \) is a compact space, and \( \{ h_{k-1}^k \}_{k=1}^\infty \) is a monotonic sequence of continuous functions, it follows that \( \{ h_k \}_{k=1}^\infty \) converges uniformly to the function \( g \) (see Rudin, Theorem 7.13).

In the following theorem, the sequence of cuts generated in Step 3a of the basic method are shown to yield supports of the objective function, \( f \), in the limit.

**THEOREM 2.** Let \( \{ x_n^k \}_{n=1}^\infty \) be an infinite subsequence of \( \{ x_n^k \}_{n=1}^\infty \). If \( x_n^k \to \hat{x} \), then with probability 1,

\[
\frac{1}{k_n} \sum_{i=1}^{k_n} \pi_i^{k_n}(\omega' - Tx_n^k) \to E[h(\hat{x}, \omega)].
\]

In addition, every limit of \( (\alpha_i^{k_n}, \beta_i^{k_n} + c)_{n=1}^\infty \) defines a support of \( f(x) \) at \( \hat{x} \), with probability one.

**PROOF.** By definition,

\[
h_{k_n}(x_n^k, \omega') = \pi_i^{k_n}(\omega' - Tx_n^k) \quad \text{and}
\]

\[
\frac{1}{k_n} \sum_{i=1}^{k_n} h_{k_n}(x_n^k, \omega') = \frac{1}{k_n} \sum_{i=1}^{k_n} \pi_i^{k_n}(\omega' - Tx_n^k).
\]

By Lemma 1, there exists a function \( g \leq h \) such that \( \{ h_{k_n} \}_{n=1}^\infty \) converges uniformly to \( g \). Thus, because

\[
\frac{1}{k_n} \sum_{i=1}^{k_n} \left[ h_{k_n}(x_n^k, \omega') - g(\hat{x}, \omega') \right] \to 0 \quad \text{and} \quad \frac{1}{k_n} \sum_{i=1}^{k_n} h(x, \omega') \to E[h(x, \omega)]
\]

\[(wp1)\]

it is sufficient to show that \( g(\hat{x}, \omega') = h(\hat{x}, \omega') \) with probability one.

Let \( \omega' \) be given, and suppose that for every \( \delta > 0 \),

\[
(1) \quad P(|\hat{\omega} - \omega'| < \delta) > 0.
\]

Then, for every \( \delta > 0 \), \( |\omega^{k_n} - \omega'| < \delta \) infinitely often, with probability one. Because \( h \) is a continuous function and \( \{ h_{k_n} \}_{n=1}^\infty \) is a uniformly convergent sequence of continuous functions, for every \( \epsilon > 0 \) there exist \( \delta > 0 \) and \( N < \infty \) such that

\[
|g(\hat{x}, \omega') - h(\hat{x}, \omega')| < \epsilon/3 \quad \text{and} \quad |h_{k_n}(\hat{x}, \omega') - h_{k_n}(x, \omega)| < \epsilon/3 \quad \forall n \geq N.
\]
Thus, since \( x^{k_n} \to \hat{x} \), property (1) implies that for every \( \epsilon > 0 \) there exists a further subsequence \( \{ (x^{k_{n_i}}, \omega^{k_{n_i}}) \}_{n_i=1}^{\infty} \) and \( K < \infty \) such that

\[
|h(\hat{x}, \omega') - h(\hat{x}, \omega^{k_{n_i}})| < \epsilon / 3, \]
\[
|h(\hat{x}, \omega^{k_{n_i}}) - h(x^{k_{n_i}}, \omega^{k_{n_i}})| < \epsilon / 3 \quad \text{and} \]
\[
|h_{k_{n_i}}(x^{k_{n_i}}, \omega^{k_{n_i}}) - h_{k_{n_i}}(x^{k_{n_i}}, \omega')| < \epsilon / 3
\]

for all \( k_{n_i} \geq K \). By construction \( h_{k_{n_i}}(x^{k_{n_i}}, \omega^{k_{n_i}}) = h(x^{k_{n_i}}, \omega^{k_{n_i}}) \) (see Step 2 of the basic method). Thus, for every \( \epsilon > 0 \) there exists a subsequence \( \{ x^{k_{n_i}} \}_{n_i=1}^{\infty} \) and \( K < \infty \) such that

\[
|h(\hat{x}, \omega') - h_{k_{n_i}}(x^{k_{n_i}}, \omega')| \leq |h(\hat{x}, \omega') - h(\hat{x}, \omega^{k_{n_i}})|
\]
\[
+ |h(\hat{x}, \omega^{k_{n_i}}) - h(x^{k_{n_i}}, \omega^{k_{n_i}})|
\]
\[
+ |h(x^{k_{n_i}}, \omega^{k_{n_i}}) - h_{k_{n_i}}(x^{k_{n_i}}, \omega')|
\]
\[
< \epsilon
\]

for all \( k_{n_i} \geq K \). Hence, by the uniqueness of the sequential limit, it follows that \( g(\hat{x}, \omega') = h(\hat{x}, \omega') \).

Finally, since \( \Omega \) is a compact space,

\[
P[|\tilde{\omega} - \omega'| < \delta] = 0 \quad \text{for some } \delta > 0
\]

for only finitely many values of \( t \), with probability one. Thus, with probability one, \( g(\hat{x}, \omega') = h(\hat{x}, \omega') \) for all but a finite number of observations. Hence with probability one, we have

\[
\frac{1}{k_{n}} \sum_{t=1}^{k_{n}} \pi^{k_{n}}(\omega' - Tx^{k_{n}}) \to E[h(\hat{x}, \tilde{\omega})].
\]

Moreover, since \( h(x, \omega') = \max(\pi(\omega' - Tx) | \pi \in V) \) and \( V_k \subset V \forall k \), it follows that

\[
cx + \frac{1}{k_{n}} \sum_{t=1}^{k_{n}} h(x, \omega') \geq cx + \frac{1}{k_{n}} \sum_{t=1}^{k_{n}} \pi^{k_{n}}(\omega' - Tx) \equiv \alpha^{k_{n}} + (\beta^{k_{n}} + c)x \quad \forall x \in X.
\]

We conclude that, with probability one, the objective function \( f(x) \) is at least as large as any limiting cut that is associated with the subsequence of cuts defined by \( \{ \alpha^{k_{n}}, \beta^{k_{n}} + c \}_{n=1}^{\infty} \). Thus, any such limiting cut defines a support of \( f(x) \) at \( \hat{x} \).

Theorem 2 indicates that the sequence \( \{ f_k(x^k) \} \) accumulates at objective values. Note that in Step 4, \( x^k \) minimizes \( f_{k-1} \). As a result of the repeated updates in Step 3b, \( f_{k-1}(x^k) \) may provide a poor approximation of \( f(x^k) \). Thus, before examining the limiting behavior of the basic algorithm, we offer the following theorem which, in combination with Theorem 2, establishes the existence of a subsequence on which \( f_{k-1}(x^k) \) accumulates at objective function values.

**Theorem 3.** There exists a subsequence of \( \{ x^k \}_{k=1}^{\infty}, \{ x^{k_{n}} \}_{n=1}^{\infty} \), such that

\[
\lim_{n \to \infty} f_{k_{n}}(x^{k_{n}}) - f_{k_{n}-1}(x^{k_{n}}) = 0.
\]

**Proof.** See appendix.

With the above results, we now have the following theorem, which establishes the existence of an optimal accumulation point of \( \{ x^k \}_{k=1}^{\infty} \).
Theorem 4. There exists a subsequence of \( \{x^k\}_{k=1}^{\infty}, \{x^k\}_{n=1}^{\infty} \), such that every accumulation point of \( \{x^k\}_{n=1}^{\infty} \) is an optimal solution, with probability one.

Proof. Let \( x^* \) be an optimal solution. From Theorem 3, there exists a subsequence \( \{x^k\}_{n=1}^{\infty} \) such that \( \lim_{n \to \infty} f_k(x^k) - f_{k_{n-1}}(x^{k_{n-1}}) = 0 \). Let \( \{x^k\}_{n=1}^{\infty} \) be a further subsequence such that \( \lim_{n \to \infty} x^k = \hat{x} \). Compactness of \( X \) ensures that \( \hat{x} \in X \), and thus

\[
 f(x^*) \leq f(\hat{x}).
\]

As previously noted,

\[
f_k(x) \leq cx + \frac{1}{K} \sum_{i=1}^{k} h(x, \omega_i) \quad \forall k, x \in X,
\]

and thus,

\[
 \limsup_{k \to K} f_k(x^*) \leq cx^* + E[h(x^*, \tilde{\omega})] = f(x^*) \quad \text{(wp1)}
\]

for any index set \( K \). Note that \( x^k \) minimizes \( f_{k-1} \) on \( X \), and thus,

\[
f_{k-1}(x^k) \leq f_{k-1}(x^*) \quad \forall k.
\]

From Theorem 2, \( \lim_{n \to \infty} f_{k_n}(x^{k_n}) = f(\hat{x}) \) (wp1) and by definition, \( \lim_{n \to \infty} f_{k_n}(x^{k_n}) - f_{k_{n-1}}(x^{k_{n-1}}) = 0 \). Thus, \( \lim_{n \to \infty} f_{k_n-1}(x^{k_n}) = f(\hat{x}) \), with probability one. Combining (2), (3), and (4), we have

\[
 f(x^*) \leq f(\hat{x}) = \lim_{n \to \infty} f_{k_n-1}(x^{k_n}) \leq \limsup_{n \to \infty} f_{k_n-1}(x^*) \leq f(x^*)
\]

with probability one, and hence, the result. ■

3. Stochastic decomposition. The existence result established for the basic method relies heavily on Theorem 3 which indicates that, on some subsequence, an adequate objective function estimate is retained despite the update procedure in Step 3b. However, it is difficult to algorithmically identify such subsequences. Moreover, in the course of computational investigations, we have observed that elements outside of this subsequence can detract from the stability of both the objective value and solution estimates. In order to retain the quality of the estimates, we introduce the idea of an incumbent solution. Such a solution is defined as an iterate whose estimated objective value is "sufficiently low". Naturally, the incumbent is updated during the course of the algorithm. By re-evaluating the cuts at an incumbent solution with sufficient regularity, the cuts continue to improve in their approximation of the recourse function within a neighborhood of the incumbent.

In this section, we show that this extended version of the basic method, which we formally call the stochastic decomposition algorithm (SD), exhibits limiting behavior similar to that of the basic method. However, it possesses additional properties that allow the development of termination criteria. Unfortunately, the additional details may obscure the simplicity underlying the basic methodology. We recommend that the algorithm be read concurrently with the notation summarized below.

3.1. Further notation. In what follows, it is important to distinguish between symbols that have a bar above them, and those that do not. For example, at the start of the \( k \)th iteration, \( \bar{x}^{k-1} \) denotes the incumbent. Similarly, \( \bar{\pi}_t \) (see Step 2b of the
algorithm) will denote the dual vector obtained when the 'argmax' operation is performed using \( \tilde{x}^{k-1} \) as the input, rather than using \( x^k \) as the input. We summarize the additional notation below.

- \( \{\tilde{x}^k\}_{k=0}^0 \) is the sequence of incumbent solutions.
- \( \tilde{X} \) is the set of accumulation points associated with \( \{\tilde{x}^k\}_{k=0}^0 \).
- \( \tilde{X}^* \) is the set of optimal solutions.
- \( i_k \) is the iteration at which the incumbent, \( \tilde{x}^k \), was identified.
- \( \pi^k \in \text{argmax}(\pi(\omega^t - T\tilde{x}^{k-1})|\pi \in V_k) \).
- \( \theta^k = f_{k-1}(x^k) - f_{k-1}(\tilde{x}^{k-1}) \).

Note that, at the start of the \( k \)th iteration of SD, \( x^{k-1} \) is the incumbent solution and \( x^k \in \text{argmin}(f_{k-1}(x)|x \in X) \). Additionally, the requirement for identification of a new incumbent is that

\[
(5) \quad f_k(x^k) - f_k(\tilde{x}^{k-1}) < r(f_{k-1}(x^k) - f_{k-1}(\tilde{x}^{k-1}))
\]

where \( r \in (0,1) \) is a fixed parameter. Since \( f_k(x^k) \) and \( f_k(\tilde{x}^{k-1}) \) represent the estimates of \( f(x^k) \) and \( f(\tilde{x}^{k-1}) \) (after cuts have been added and updated), respectively, we use (5) to hypothesize that the actual value of the objective function at \( x^k \) is sufficiently lower than at \( \tilde{x}^{k-1} \). In such cases, \( x^k \) is adopted as a new incumbent. As shown in the following section, the type of descent implied by (5), which holds for the randomly generated estimates of the objective function values, plays a critical role in characterizing the limiting properties of the SD algorithm.

**Algorithm: Stochastic Decomposition (SD).**

**Step 0.** Initialize. \( V_0 \leftarrow \emptyset, \omega^0 \leftarrow E[\omega], x^1 \in \text{argmin}(cx + h(x, \omega^0)|x \in X), \tilde{x}^0 = x^1, i_0 \leftarrow 0, r \in (0,1) \) is given, \( k \leftarrow 0. \)

**Step 1.** Generate \( \omega^k \). \( k \leftarrow k + 1. \) Randomly generate an observation \( \omega^k \) according to the distribution \( F_{\omega^0}. \) (Note that \( \omega^t, t = 1, \ldots, k \) are generated independently.)

**Step 2.** Update \( V_k. \) \( V_k \leftarrow V_{k-1} \cup \{\pi(x^k, \omega^k), \pi(\tilde{x}^{k-1}, \omega^k)\}. \)
(Recall that \( \pi(x, \omega) \in \text{argmax}(\pi(\omega - Tx)|\pi \in V) \).

**Step 3.** Define \( f_k \) (approximation of \( f \)).

a. Construct the coefficients of the \( k \)th cut to be added to the master problem.
With \( \pi^k_i \in \text{argmax}(\pi(\omega^t - Tx^k)|\pi \in V_k) \),

\[
\alpha^k_i + (\beta^k_i + c)x = cx + \frac{1}{k} \sum_{t=1}^{k} \pi^k_i(\omega^t - Tx).
\]

b. Update the coefficients of the cut indexed by \( i_{k-1} \).
With \( \pi^i \in \text{argmax}(\pi(\omega^t - T\tilde{x}^{k-1})|\pi \in V_k) \),

\[
\alpha^k_{i_{k-1}} + (\beta^k_{i_{k-1}} + c)x = cx + \frac{1}{k} \sum_{t=1}^{k} \pi^i(\omega^t - Tx).
\]

c. Update the remaining cuts.
\[
\alpha^k_i \leftarrow \frac{k-1}{k} \alpha^k_{i-1}, \quad \beta^k_i \leftarrow \frac{k-1}{k} \beta^k_{i-1}, \quad t \notin \{i_{k-1}, k\}.
\]

**Step 4.** Test the incumbent.
If \( f_k(x^k) - f_k(\tilde{x}^{k-1}) < r(f_{k-1}(x^k) - f_{k-1}(\tilde{x}^{k-1})) \) then \( \tilde{x}^k \leftarrow x^k, i_k \leftarrow k. \)
Otherwise, \( \tilde{x}^k \leftarrow \tilde{x}^{k-1}, i_k \leftarrow i_{k-1}. \)

**Step 5.** Solve the master problem, \( M^k \), to obtain \( x^{k+1} \). Repeat from Step 1.
3.2. Proof of convergence. Retaining the assumptions of §2.1, we now show that there exists at least one optimal accumulation point of \( \{x^k\} \). In addition, we also show that an averaged objective value converges to the optimal objective value, with probability one. We begin by noting that, for all \( k \),

\[
\alpha^k_{ik} + \left( \beta^k_{ik} + c \right) x^k \leq f_k(x^k) \leq c x^k + \frac{1}{k} \sum_{i=1}^{k} h(x^k, \omega^i).
\]

Since \( h(x, \omega) \) is continuous in \( x \) for all \( \omega \in \Omega \), if \( \{x^k\}_{k=1}^{\infty} \) is a subsequence such that \( x^k \to x \), \( h(x^k, \omega^i) \to h(x, \omega^i) \) \( \forall i \). Thus,

\[
\lim_{n \to \infty} \alpha^k_{ik} + \left( \beta^k_{ik} + c \right) x^k \leq \lim_{n \to \infty} f_k(x^k) \leq \lim_{n \to \infty} c x^k + \frac{1}{k} \sum_{i=1}^{k} h(x^k, \omega^i).
\]

Since both the upper and lower limits are \( f(x) \) (with probability one, see Theorem 2), it follows that

\[
\lim_{n \to \infty} f_k(x^k) = f(x) \quad \text{(wp1)}.
\]

Similarly, if \( \{x^k\}_{k=1}^{\infty} \to x \), then

\[
\lim_{n \to \infty} f_{k+1}(x^k) = f(x) \quad \text{(wp1)}.
\]

As these results follow from Theorem 2, we formally state them in the following corollary.

**Corollary 5.** Let \( \{x^k\}_{k=1}^{\infty} \) be an infinite subsequence of \( \{x^k\} \). If \( x^k \to x \), then with probability one,

\[
\lim_{n \to \infty} f_k(x^k) = \lim_{n \to \infty} f_{k+1}(x^k) = f(x).
\]

As the first step in establishing the existence of an optimal accumulation point of the incumbent sequence, we explore the implication of the test for the new incumbent in Step 4 of the algorithm.

**Lemma 6.** Let \( \{k_n\}_{n=1}^{\infty} \) represent the sequence of iterations at which the incumbent is changed. If \( N \) is finite, then \( \lim_{k \to \infty} \theta^k = 0 \) (wp1). Otherwise

\[
\lim_{m \to \infty} \frac{1}{m} \sum_{n=1}^{m} \theta^k_n = 0 \quad \text{(wp1)}.
\]

**Proof.** By definition, \( \theta^k = f_{k-1}(x^k) - f_{k-1}(\bar{x}^k) \leq 0 \) for all \( k \). If \( N \) is a finite set, there exist \( \bar{x} \) and \( K < \infty \) such that \( \bar{x}^k = \bar{x} \) for all \( k \geq K \) and thus

\[
f_k(x^k) - f_k(\bar{x}) \geq r \left( f_{k-1}(x^k) - f_{k-1}(\bar{x}) \right) = r \theta^k \quad \forall k \geq K.
\]
By Theorems 2 and 3 and Corollary 5, there exists a subsequence indexed by the set \( \mathcal{H} \) such that

\[
\lim_{k \in \mathcal{H}} x^k = \hat{x},
\]

\[
\lim_{k \in \mathcal{H}} f_k(x^k) = f(\hat{x}), \quad \lim_{k \in \mathcal{H}} f_k(\bar{x}) = f(\bar{x}),
\]

\[
\lim_{k \in \mathcal{H}} f_{k-1}(x^k) = f(\hat{x}), \quad \lim_{k \in \mathcal{H}} f_{k-1}(\bar{x}) = f(\bar{x}),
\]

with probability one. Thus,

\[
\lim_{k \in \mathcal{H}} f_k(x^k) - f_k(\bar{x}) \geq r \left\{ \lim_{k \in \mathcal{H}} f_{k-1}(x^k) - f_{k-1}(\bar{x}) \right\}
\]

\[
\Rightarrow f(\hat{x}) - f(\bar{x}) \geq r \{ f(\hat{x}) - f(\bar{x}) \} = \lim_{k \in \mathcal{H}} r \theta^k.
\]

Since \( r \in (0, 1) \) and \( \theta^k \leq 0 \) for all \( k \), it follows that \( f(\hat{x}) - f(\bar{x}) = 0 \). The result now follows since

\[
\lim_{k \in \mathcal{H}} \theta^k \leq \lim_{k \to \infty} \theta^k \leq 0.
\]

Now suppose \( \mathcal{N} \) is not a finite set. By hypothesis,

\[
f_{k_n}(x^{k_n}) - f_{k_n}(\bar{x}^{k_n-1}) < r \{ f_{k_n-1}(x^{k_n}) - f_{k_n-1}(\bar{x}^{k_n-1}) \} = r \theta^{k_n} \leq 0 \quad \forall n.
\]

Noting that \( \bar{x}^{k_n-1} = \bar{x}^{k_n-1} \),

\[
f_{k_n}(\bar{x}^{k_n}) - f_{k_n}(\bar{x}^{k_n-1}) \leq r \theta^{k_n} \leq 0 \quad \forall n.
\]

Thus,

\[
\frac{1}{m} \sum_{n=1}^{m} \left\{ f_{k_n}(\bar{x}^{k_n}) - f_{k_n}(\bar{x}^{k_n-1}) \right\} \leq \frac{r}{m} \sum_{n=1}^{m} \theta^{k_n} \leq 0 \quad \forall m
\]

\[
\Rightarrow \frac{1}{m} \left\{ \sum_{n=1}^{m-1} f_{k_n}(\bar{x}^{k_n}) - f_{k_n+1}(\bar{x}^{k_n}) \right\} + \left( f_{k_m}(\bar{x}^{k_m}) - f_k(\bar{x}^{k_m}) \right) \leq \frac{r}{m} \sum_{n=1}^{m} \theta^{k_n} \leq 0 \quad \forall m.
\]

Assumptions A1 and A2 ensure that there exists \( M < \infty \) such that \( |f_{k_n}(\bar{x}^{k_n}) - f_k(\bar{x}^{k_n})| < M \) for all \( m \). Thus, since \( \bar{x}^{k_n} = \bar{x}^{k_n-1} \), Corollary 5 ensures that the left-hand side converges to zero, with probability one, and thus,

\[
\lim_{m \to \infty} \frac{1}{m} \sum_{n=1}^{m} \theta^{k_n} = 0 \quad (wp1).
\]

With these results, we can now prove the following.
THEOREM 7. Let \( \{\bar{x}^k\}_{k=1}^\infty \) represent the sequence of incumbents, and let \( X^\ast \) represent the set of optimal solutions. With probability one, there exists a subsequence \( \{\bar{x}^k\}_{k \in K} \) for which every accumulation point is contained in \( X^\ast \).

PROOF. Let \( (k_n)_{n \in N} \) represent the sequence of iterations at which the incumbent is changed. Note that if \( N \) is an infinite set,

\[
\lim_{m \to \infty} \frac{1}{m} \sum_{n=1}^{m} \theta^{k_n} \leq \lim_{n \to \infty} \theta^{k_n} \leq 0.
\]

Thus, as a result of Lemma 6, whether \( N \) is finite or infinite, there exists a subsequence indexed by the set \( K \) such that

\[
\lim_{k \in K} \theta^{k+1} = 0 \quad (wp1).
\]

Note that

\[
\theta^{k+1} = f_k(x^{k+1}) - f_k(\bar{x}^k) \leq f_k(x^\ast) - f_k(\bar{x}^k) \quad \forall k \in K.
\]

Thus, as a result of Corollary 5 and (3), it follows that if \( \bar{x} \) is an accumulation point of \( \{\bar{x}^k\}_{k \in K} \), then

\[
f(\bar{x}) \leq f(x^\ast) \quad (wp1)
\]

and thus, \( \bar{x} \in X^\ast \), with probability one. ■

Theorem 7 verifies the existence of a subsequence of \( \{\bar{x}^k\}_{k=1}^\infty \) which accumulates at optimal solutions. In the following, we offer a simple method by which such a subsequence may be identified. Of course, if the incumbent solution changes only finitely often, the subsequence is easily identified. Thus, our attention is focussed on the case in which the incumbent changes infinitely often.

LEMMA 8. Let \( \{\bar{x}^k\}_{k=1}^\infty \) be the sequence of incumbent solutions identified by the SD algorithm, and suppose that \( (k_n)_{n=1}^\infty \) represents the subsequence of iterations on which the incumbent changes. If \( N^\ast \) is the index set such that

\[
n \in N^\ast \iff \theta^{k_n} \geq \frac{1}{n} \sum_{m=1}^{n} \theta^{k_m}
\]

then \( N^\ast \) is an infinite set.

PROOF. We proceed by contradiction. Suppose \( N^\ast \) is a finite set. Then there exists \( n_0 < \infty \) such that

\[
\theta^{k_n} < \frac{1}{n} \sum_{m=1}^{n} \theta^{k_m} \quad \forall n \geq n_0.
\]

Let

\[
\bar{S} = \frac{1}{n_0} \sum_{m=1}^{n_0} \theta^{k_m},
\]

and note that since \( \theta^k \leq 0 \) for all \( k \), (7) ensures that \( \theta^{k_{n_0}} < 0 \), and thus, \( \bar{S} < 0 \). Inductively, we will establish that \( \theta^{k_{n}} < \bar{S} \) for all \( n \geq n_0 \).
By definition, $\theta^{k_{n_0}} < \bar{S}$. Thus, suppose that $\theta^{k_{n_0 + l}} < \bar{S}$, $j = 0, \ldots, l - 1$. From (7),

$$
\theta^{k_{n_0 + l}} < \frac{1}{n_0 + l} \sum_{m=1}^{n_0 + l} \theta^{k_m} = \frac{1}{n_0 + l} \left( n_0 \theta^{k_{n_0}} + \sum_{m=n_0 + 1}^{n_0 + l - 1} \theta^{k_m} + \theta^{k_{n_0 + l}} \right) 
$$

$$
\Rightarrow \theta^{k_{n_0 + l}} < \frac{1}{n_0 + l} \left( n_0 \bar{S} + (l - 1) \bar{S} + \theta^{k_{n_0 + l}} \right) = \theta^{k_{n_0 + l}} < \bar{S}.
$$

Thus, if $N^*$ is a finite set, $\theta^{k_n} < \bar{S}$ for all $n \geq n_0$ and it follows that

$$
\lim_{n \to \infty} \frac{1}{n} \sum_{m=1}^{n} \theta^{k_m} \leq \bar{S} < 0,
$$

contradicting Lemma 6. ■

With Lemma 8, we can now identify a subsequence of the incumbent solutions that accumulates at the set of optimal solutions, with probability one.

**Theorem 9.** Let $\{\bar{x}^k\}_{k=1}^{\infty}$ be the sequence of incumbent solutions identified by the SD algorithm, and suppose that $\{k_n\}_{n=1}^{\infty}$ represents the subsequence of iterations on which the incumbent changes. If $N^*$ is the index set identified in (6), then every accumulation point of $\{\bar{x}^k\}_{n \in N^*}$ is optimal, with probability one.

**Proof.** From the proof of Theorem 7, it is sufficient to show that $(\theta^{k_n})_{n \in N^*} \to 0$, with probability one. By construction,

$$
(8) \quad \frac{1}{n} \sum_{m=1}^{n} \theta^{k_m} \leq \theta^{k_n} \leq 0 \quad \forall n \in N^*.
$$

Taking the limit of (8), Lemma 6 ensures that

$$
\lim_{n \in N^*} \theta^{k_n} = 0 \quad \text{(wp1)}.
$$

Finally, we offer the following theorem in which we establish the optimality of an averaged objective value.

**Theorem 10.** Let $\{k_n\}_{n \in N}$ represent the sequence of iterations at which the incumbent is changed, and let $x^* \in X^*$ be given. If $N$ is finite, then

$$
\frac{1}{m} \sum_{k=1}^{m} f_k(\bar{x}^k) \to f(x^*) \quad \text{with probability one}.
$$

Otherwise,

$$
\frac{1}{m} \sum_{n=1}^{m} f_{k_n}(\bar{x}^{k_n}) \to f(x^*), \quad \text{with probability one}.
$$

**Proof.** If $N$ is finite, there exists $K < \infty$ such that $\bar{x}^k = \bar{x}$ for all $k \geq K$, and the result follows immediately from Corollary 5 and Theorem 7. If $N$ is infinite, then
Corollary 5 ensures that
\[ f(x^*) \leq \lim_{n \to \infty} f_{k_n}(\bar{x}^{k_n}) \quad \text{(wp1)}. \]

Furthermore, since
\[ \frac{1}{m} \sum_{n=1}^{m} \theta^{k_n} \leq \frac{1}{m} \sum_{n=1}^{m} \left\{ f_{k_n-1}(x^*) - f_{k_n-1}(\bar{x}^{k_n-1}) \right\}, \]
we have
\[ \lim_{m \to \infty} \frac{1}{m} \sum_{n=1}^{m} f_{k_n-1}(\bar{x}^{k_n-1}) \leq \lim_{m \to \infty} \left\{ \frac{1}{m} \sum_{n=1}^{m} f_{k_n-1}(x^*) - \frac{1}{m} \sum_{n=1}^{m} \theta^{k_n} \right\}. \]

From Lemma 7, we have
\[ \lim_{m \to \infty} \frac{1}{m} \sum_{n=1}^{m} \theta^{k_n} = 0. \]

Hence using (3), it follows that
\[ \lim_{m \to \infty} \frac{1}{m} \sum_{n=1}^{m} f_{k_n-1}(\bar{x}^{k_n-1}) \leq \lim_{m \to \infty} \frac{1}{m} \sum_{n=1}^{m} f_{k_n-1}(x^*) \]
\[ \leq \lim_{n \to \infty} f_{k_n-1}(x^*) = f(x^*) \quad \text{(wp1)}. \]

Since \( \bar{x}^{k_n-1} = \bar{x}^{k_n-1} \), Corollary 5 ensures that
\[ f_{k_n-1}(\bar{x}^{k_n-1}) \rightarrow f_{k_n-1}(\bar{x}^{k_n-1}) \rightarrow 0. \]

Thus
\[ f(x^*) \leq \lim_{n \to \infty} f_{k_n}(\bar{x}^{k_n}) \leq \lim_{m \to \infty} \frac{1}{m} \sum_{n=1}^{m} f_{k_n}(\bar{x}^{k_n}) \leq \lim_{m \to \infty} \frac{1}{m} \sum_{n=1}^{m} f_{k_n}(\bar{x}^{k_n}) \leq f(x^*) \]
with probability one, and the result follows.

It is interesting to compare the convergence results of §2 and §3 with those of their deterministic counterparts. For the basic algorithm of §2, comparisons with Kelley's cutting plane method (Avriel [1976]) reveal that both methods generate sequences that contain an accumulation point that is optimal. In the deterministic setting, the inclusion of a descent requirement via an incumbent solution ensures optimality of all accumulation points of the incumbent sequence. The apparent failure of the stochastic analog in §3 to exhibit this property can be traced to the manner in which the objective function approximations are updated. To obtain this property, one need only re-evaluate the cuts at every incumbent generated by the algorithm with sufficient frequency. Clearly, such an undertaking is computationally unrealistic for problems for which sampling procedures can be justified. Nonetheless, Theorem 9 offers an easily implemented procedure for identifying a subsequence for which every accumulation point is optimal. In addition, Theorem 10 indicates that the average objective value converges to the optimal value, with probability one. This suggests natural termination criteria, which we discuss in the following section.
4. Termination criteria. While the convergence results of §2 and §3 lay the theoretical foundation for SD, they provide valuable insight into appropriate choices for termination criteria as well. For example, as a result of Theorem 10, one may monitor the progress of the incumbent objective value, \( f_k(\bar{x}^k) \), as it can be compared to an average value which converges to the optimal objective value (with probability one). Similarly, as a result of Lemmas 6 and 8, one can also monitor the progress of \( \theta_k \), as it contains a subsequence which converges to zero.

Termination criteria based on the incumbent objective values are easily envisioned. For example, let

\[
\gamma^k = \frac{1}{k} \sum_{i=1}^{k} f_i(\bar{x}^i), \\
\bar{\gamma}^k = \frac{1}{m_k} \sum_{n=1}^{m_k} f_k(\bar{x}^k_n),
\]

where \( m_k \) represents the number of incumbent solutions that have been identified during the first \( k \) iterations. By Theorem 10, at least one of these statistics converges to the optimal objective value, with probability one. Thus, a termination criterion based on the convergence of these values would test whether

\[
\frac{|f_k(\bar{x}^k) - \gamma^{k-1}|}{\gamma^{k-1}} < \epsilon
\]

when the incumbent solution, \( \bar{x}^k \), has remained unchanged over an appropriately large number of iterations, where \( \epsilon > 0 \) is an acceptable tolerance level. An analogous test for the case in which \( \bar{x}^k \) has recently changed is given by

\[
\frac{|f_k(\bar{x}^k) - \bar{\gamma}^{k-1}|}{\bar{\gamma}^{k-1}} < \epsilon.
\]

Termination criteria based on the values of \( \theta_k \) may be similarly defined. Note that

\[
\theta^k = f_{k-1}(x^k) - f_{k-1}(\bar{x}^{k-1}) \leq f_{k-1}(x^*) - f_{k-1}(\bar{x}^{k-1}).
\]

From (3), \( f_{k-1}(x^*) \) accumulates at lower bounds on \( f(x^*) \), and Corollary 5 ensures that \( f_{k-1}(\bar{x}^{k-1}) \) accumulates at objective function values. Thus, when \( k \) is sufficiently large, \( \theta^k \) provides an estimate of the deviation from optimality. Consequently, the algorithm may be terminated whenever this estimated deviation is acceptably small.

Of course, since SD is an algorithm that is built upon randomized subproblems, termination criteria should include certain safeguards designed to prevent premature termination. Initially, one must determine whether a sufficient subset of the subproblem dual vertices have been obtained during the execution of the algorithm. Toward this end, it is necessary to evaluate the stability of the set \( V_k \). This may be accomplished by considering termination only during iterations for which \( V_k = V_{k-1} \), where \( K \) is an appropriately large integer. Note that since \( V_k \subseteq V_{k+1} \) for all \( k \), equality of these sets is easily established by comparing the cardinality of the sets.

Another possible variation of the previously discussed criteria involves the reinitialization of the statistics being computed whenever the cardinality of \( V_k \) increases.

5. Computational experience. The algorithm was used to solve standard test problems included in Birge [1985], which are formulated using discrete random variables. These test problems were selected because both optimal values and
TABLE I. Computational Results

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>SC205</td>
<td>0 (0)</td>
<td>n/a</td>
<td>56.7 (22.3)</td>
</tr>
<tr>
<td>SCSR58</td>
<td>0.0070 (0.0077)</td>
<td>n/a</td>
<td>146.4 (74.2)</td>
</tr>
<tr>
<td>SCAGR7</td>
<td>0.0 (0.0)</td>
<td>0.00013 (0.0)</td>
<td>161.6 (70.0)</td>
</tr>
<tr>
<td>SCTAP1_A</td>
<td>0.0014 (0.0006)</td>
<td>n/a</td>
<td>74.3 (39.7)</td>
</tr>
</tbody>
</table>

*Note: 0.0 indicates numbers less than $1 \times 10^{-4}$.*

solutions are known, and as a result, we can verify the performance of the new method. Our purpose in this section is to provide evidence of the implementability of the algorithm, and thus the computational results presented here should not be construed as representing a complete test of the algorithm's effectiveness. In implementing the algorithm, we adopted a ‘trust region’ approach through which we enclose the incumbent solution within a ‘box’. The size of the box is varied as a function of the retention/rejection of the incumbent solution. The algorithm was terminated when the trust region was so small that it appeared to contain only the terminal incumbent solution, thereby suggesting that the sequence of distinct incumbent solutions would be finite. From Theorem 7, in such cases, the terminal incumbent is necessarily optimal.

Upon termination of the algorithm, there are two types of objective value errors that are of interest. One error is statistical in nature, and represents the deviation of the estimated objective function value from the actual objective value at the terminal incumbent. A second error results from the deviation of the actual objective value at the terminal incumbent from the optimal objective value. In Table I, we list the average error for the former and latter types, as a fraction of the actual objective value at the terminal incumbent and the optimal objective value, respectively. Errors are listed only for applicable cases. Thus, when the terminal incumbent is an optimal solution, there is no error due to deviation from optimality. Because the algorithm incorporates a sampling procedure, these averages are computed over ten runs involving different streams of generated observations of $\tilde{\omega}$. The values in parentheses represent the standard deviation (as a fraction of the actual objective values) corresponding to the reported averages.

From Table I, it is clear that in solving the above problems, SD produced negligible errors. In the instances in which SD terminated at nonoptimal solutions (SCAGR7), the average deviation from the optimal value was only 0.013%, and the maximum deviation was less than 0.014%. It appears that the iteration counts exhibit a sensitivity to the sequence of observations generated, as evidenced by the relatively high standard deviations. Although this is somewhat disconcerting, the fact that the optimality of the terminal incumbents and objective value errors do not exhibit this sensitivity lends credence to the implementability of the method. Note that the minute errors presented in Table I suggest that the termination criteria used in this implementation may be more stringent than those suggested in §4.

6. Conclusions. In this paper, we have presented a novel algorithm that introduces a marriage between deterministic decomposition based methods and stochastic quasi-gradient (SQG) methods. As a consequence of this development we are able to overcome several shortcomings associated with previously known stochastic programming algorithms. To illustrate this point, recall that decomposition based methods are designed for problems governed by discrete random variables, while SQG methods allow continuous random variables. On the other hand, the latter cannot accommo-
date integer restrictions on the first stage decisions. Stochastic decomposition is sufficiently general to be applicable to all of the above situations. Additionally, like SQG methods, SD may also prove useful for real-time optimization of a multiple period problem in which the actual distributions of the random variables are not known. Finally, SD is applicable even in situations where the matrix \( T \) is replaced by a random matrix \( T(\hat{\omega}) \).

Among its other features, SD can incorporate statistically motivated termination criteria. Our computational experiments suggest that SD performs well. Even though it is a statistically motivated algorithm, it exhibits an overwhelming tendency to terminate with an optimal or a near optimal solution. This lends encouraging evidence of the usefulness of SD as an algorithmic concept.

Acknowledgements. The authors would like to acknowledge the contribution of their students Jeffrey T. Gunn, M. S. Sodhi and Mousa Mitwasi who programmed stochastic decomposition and other stochastic programming algorithms and to Wing Lowe for providing the insight to Theorem 9. We are also grateful to an anonymous referee for careful reading of an earlier version of this paper.

Appendix. In proving Theorem 3, note that since \( f_k(x) = cx + \max\{\alpha_t^k + \beta_t^k x\mid 1 \leq t \leq k\} \),

\[
    f_k(x^k) - f_{k-1}(x^k) = \max\{\alpha_t^k + \beta_t^k x^k \mid 1 \leq t \leq k\} \\
    - \max\{\alpha_t^{k-1} + \beta_t^{k-1} x^k \mid 1 \leq t \leq k - 1\}.
\]

Thus, if we define

\[
    \nu_k(x) = \max\{\alpha_t^k + \beta_t^k x \mid 1 \leq t \leq k\},
\]

then \( \lim_{k \to \infty} f_k(x^k) - f_{k-1}(x^k) = 0 \) if, and only if, \( \lim_{k \to \infty} \nu_k(x^k) - \nu_{k-1}(x^k) = 0 \). Furthermore, the definition of \((\alpha_t^k, \beta_t^k)\) ensures that

\[
    \nu_k(x) \geq \alpha_t^k + \beta_t^k x = \frac{t}{K} (\alpha_t^k + \beta_t^k x) \quad \forall x \in X, \forall t \in \{1, 2, \ldots, k\}.
\]

By construction, \( \lim_{k \to \infty} f_k(x^k) - f_{k-1}(x^k) \geq 0 \). Thus, to prove Theorem 3, we begin with the following lemmas, in which we explore the consequences which would arise if \( \lim_{k \to \infty} f_k(x^k) - f_{k-1}(x^k) > 0 \).

**Lemma A1.** Let

\[
    \gamma_k = \min_{k/2 \leq t \leq k} \left( \frac{t}{K} (\nu_t(x^t) - \nu_{t-1}(x^t)) / \|(1, -\beta_t^k)\| \right).
\]

If \( \lim_{k \to \infty} f_k(x^k) - f_{k-1}(x^k) > 0 \), then there exists \( K^* < \infty \) such that

\[
    \|(\nu_k(x^{k+1}), x^{k+1}) - \left( \frac{t}{K} \nu_{t-1}(x^t), x^t \right) \| \geq \gamma_k \quad \forall k \geq K^*, \forall k/2 \leq t \leq k.
\]
PROOF. By definition,

\[ \nu_k(x^{k+1}) = \frac{t}{k} (\alpha_i' + \beta_i' x^{k+1}) \]

\[ = \frac{t}{k} (\alpha_i' + \beta_i' x^{k+1}) + \frac{t}{k} \left( \nu_{i-1}(x^t) - \nu_{i-1}(x') \right) \]

\[ = \frac{t}{k} (\alpha_i' + \beta_i' x^t) + \frac{t}{k} \left( \nu_{i-1}(x^t) - \nu_{i-1}(x') \right) + \frac{t}{k} \beta_i' (x^{k+1} - x') \]

\[ = \frac{t}{k} (\nu_i(x^t) - \nu_{i-1}(x^t)) + \frac{t}{k} \nu_{i-1}(x^t) + \beta_i' (x^{k+1} - x'). \]

Thus,

\[ \frac{t}{k} (\nu_i(x^t) - \nu_{i-1}(x^t)) \leq \nu_k(x^{k+1}) - \frac{t}{k} \nu_{i-1}(x^t) - \beta_i' (x^{k+1} - x') \]

\[ = \left(1, -\beta_k^k\right) \left( \nu_k(x^{k+1}) - \frac{t}{k} \nu_{i-1}(x^t), x^{k+1} - x' \right) \]

\[ = \left(1, -\beta_k^k\right) \left( \nu_k(x^{k+1}), x^{k+1} - \left( \frac{t}{k} \nu_{i-1}(x^t), x^t \right) \right) \]

By hypothesis, \( \lim_{k \to \infty} f_k(x^k) - f_{k-1}(x^k) > 0 \), and thus, \( \lim_{k \to \infty} \nu_k(x^k) - \nu_{k-1}(x^k) > 0 \). It follows that there exists \( K^* < \infty \) such that \( \nu_i(x^t) - \nu_{i-1}(x^t) > 0 \) for all \( t \geq K^*/2 \), and thus for all \( k/2 < t \leq k \), \( k \geq K^* \),

\[ 0 < \frac{t}{k} (\nu_i(x^t) - \nu_{i-1}(x^t)) \leq \left(1, -\beta_k^k\right) \left( \nu_k(x^{k+1}), x^{k+1} - \left( \frac{t}{k} \nu_{i-1}(x^t), x^t \right) \right) \]

\[ = \left\| (\nu_k(x^{k+1}), x^{k+1}) - \left( \frac{t}{k} \nu_{i-1}(x^t), x^t \right) \right\| \]

\[ \geq \frac{t}{k} (\nu_i(x^t) - \nu_{i-1}(x^t))/\left\| (1, -\beta_k^k) \right\| \]

\[ \geq \gamma_k \]

for all \( k/2 < t \leq k \), for all \( k \geq K^* \). \( \blacksquare \)

Let \( \gamma = \lim_{k \to \infty} \gamma_k \). As a result of Lemma A1, we see that if \( \gamma > 0 \) then eventually,

\[ \left\| (\nu_k(x^{k+1}), x^{k+1}) - \left( \frac{t}{k} \nu_{i-1}(x^t), x^t \right) \right\| > \gamma/2, \]

and thus the open ball of radius \( \gamma/4 \) centered at \( (\nu_k(x^{k+1}), x^{k+1}) \) does not contain \( ((t/k)\nu_{i-1}(x^t), x^t) \), for any \( k/2 < t \leq k \). Since \( t/k \in [1/2, 1] \) when \( k/2 < t \leq k \), and \( \beta_k^k \) is contained within a compact set,

\[ \lim_{k \to \infty} f_k(x^k) - f_{k-1}(x^k) > 0 \iff \lim_{k \to \infty} \gamma_k = \gamma > 0. \]
In the following lemma, we expand on the results of Lemma A1 and show that if $\gamma > 0$, then for every integer $n$, there are eventually $n$ disjoint open balls of radius $\gamma/8$ that can be identified.

**Lemma A2.** If $\gamma > 0$, then, for every $n$, there exists $K_n$ such that for all $k \geq K_n$

$$\left\| \left( \frac{p}{k + n} \nu_{p-1}(x^p), x^p \right) - \left( \frac{q}{k + n} \nu_{q-1}(x^q), x^q \right) \right\| > \gamma/4$$

for all $p, q \in \{k + 1, k + 2, \ldots, k + n\}$ for all $k \geq K_n$.

**Proof.** We assume without loss of generality that $p > q$. Then

$$\frac{p}{k + n} \nu_{p-1}(x^p) - \frac{q}{k + n} \nu_{q-1}(x^q) = \frac{p}{k + n} \left( \nu_{p-1}(x^p) - \frac{q}{p} \nu_{q-1}(x^q) \right).$$

Thus,

\begin{equation}
(A.1) \quad \left\| \left( \frac{p}{k + n} \nu_{p-1}(x^p), x^p \right) - \left( \frac{q}{k + n} \nu_{q-1}(x^q), x^q \right) \right\|^2
\end{equation}

\[= \left( \frac{p}{k + n} \right)^2 \left[ \nu_{p-1}(x^p) - \frac{q}{p} \nu_{q-1}(x^q) \right]^2 + \left\| x^p - x^q \right\|^2 \]

\[\geq \left( \frac{p}{k + n} \right)^2 \left[ \nu_{p-1}(x^p) - \frac{q}{p} \nu_{q-1}(x^q) \right]^2 + \left\| x^p - x^q \right\|^2 \].

Since $\gamma = \lim_{k \to \infty} \gamma_k$, there exists $K < \infty$ such that $\gamma_k \geq \gamma/2$ for all $k \geq K$. Thus, let $K^*$ be the index identified in Lemma A1 and let $K_n = \text{Max}\{K, K^*, 2n\}$ and $k \geq K_n$.

Noting that $p, q \in \{k + 1, \ldots, k + n\}$, $p > q$, and $K_n \geq 2n$, it follows that

$$\frac{q}{p} \geq \frac{1}{2}.$$

Hence the indices $p, q$ satisfy the requirement of Lemma A1 and therefore the right-hand side of $(A.1)$ can be bounded from below by

$$\left( \frac{p}{k + n} \right)^2 \gamma_k^2.$$

But since $k + n \geq p \geq k + 1$ and $k \geq K_n \geq 2n$, it follows that

$$\frac{p}{k + n} \geq \frac{k + 1}{k + n} \geq \frac{1}{2}$$

and since $k \geq K_n \geq K$,

$$\left\| \left( \frac{p}{k + n} \nu_{p-1}(x^p), x^p \right) - \left( \frac{q}{k + n} \nu_{q-1}(x^q), x^q \right) \right\|^2 \geq (1/2 \gamma_k)^2 \geq (\gamma/4)^2.$$  

With Lemmas A1 and A2, we can now prove Theorem 3.

**Proof of Theorem 3.** Assumptions A1–A3 ensure that there exists $M < \infty$ such that $0 \leq \nu_k(x) \leq M$ for all $x \in X$, and for all $k$. Thus, there exists a compact set $S$ such that

$$\left( \frac{t}{k} \nu_k(x), x \right) \in S \quad \forall \frac{k}{2} \leq t \leq k, \forall x \in X, \forall k.$$
We proceed by contradiction. If no subsequence satisfying the hypotheses exist, then
\[
\lim_{k \to \infty} f_k(x^k) - f_{k-1}(x^k) > 0,
\]
and it follows that \( \gamma > 0 \). Thus, from Lemma A2, for every \( n \) there exists \( K_n < \infty \) such that
\[
\left\| \left( \frac{p}{k+n} \nu_{p-1}(x^n), x^n \right) - \left( \frac{q}{k+n} \nu_{q-1}(x^q), x^q \right) \right\| > \gamma/4
\]
for \( p, q \in \{k+1, \ldots, k+n\} \), for all \( k \geq K_n \). Hence, the open balls of radius \( \gamma/8 \) centered at
\[
\left\{ \left( \frac{l}{k+n} \nu_{l-1}(x^l), x^l \right) \right\}_{l=k+1}^{k+n}
\]
are mutually disjoint for all \( k \geq K_n \), for all \( n \). This contradicts the compactness of \( S \), and hence the result. ■

References


SIE DEPARTMENT, UNIVERSITY OF ARIZONA, TUCSON, ARIZONA 85721