Multistage Stochastic Planning Models Using Piecewise Linear Response Functions

John R. Birge
Department of Industrial and Operations Engineering
The University of Michigan
Ann Arbor, MI 48109-2117

Technical Report 89-11

April 1989

Multistage Stochastic Planning Models using Piecewise Linear Response Functions

John R. Birge*
The University of Michigan

1. INTRODUCTION

Random outcomes can often produce significant effects on planning decisions that consider several time periods. Multistage stochastic programs can model these decisions but implementations are generally restricted to a limited number of scenarios in each period. We present an alternative approximation scheme that can obtain lower and upper bounds on the optimal objective value in these stochastic programs.

The method is based on building response functions to future outcomes that depend separably on the variation of random parameters around the limited set of scenarios that is initially provided. For stochastic linear programs, the resulting optimization problem involves an objective with a limited number of nonlinear terms subject to linear constraints. The method can be incorporated into various alternative procedures for solving

* Department of Industrial and Operations Engineering, The University of Michigan, Ann Arbor, MI 48109, USA. His work was supported by Office of Naval Research Grant N0014-86-K-0628.

multistage stochastic linear programs with finite numbers of scenarios.

In this next section, we discuss the basic model and alternative approaches. We then describe the basic properties of piecewise linear response functions. The fourth section presents a basic model for a single scenario and randomness restricted to constraint levels. The fifth section extends this to multiple scenarios with varying scenario ranges and to possibilities for randomness among the constraint vectors.

2. THE BASIC MODEL AND APPROACHES

The multistage stochastic programming model has been applied to resource planning in a variety of contexts (see, for example, the references in Ermoliev and Wets (1988)). The general multistage stochastic programming problem is:

$$\min \mathrm{E}[\sum_{t=1}^{T} \mathrm{f}_t(\mathbf{x}_t)] \text{ s. t. } \mathbf{g}_1(\mathbf{x}_1) = 0, \mathbf{g}_t(\mathbf{x}_{t-1}, \mathbf{x}_t) = 0, \text{ a. s.}, t = 1, \ldots, T, \tag{1}$$

and $\mathbf{x}$ nonanticipative. This last constraint restricts $\mathbf{x}_t$ to depend only on the outcomes of any random variables up to time t.

In the following, we restrict the problem in (1) to the case of linear f and g and only allow randomness to enter the right-hand sides of the constraints. In this case, the problem becomes:

$$\begin{array}{lllll} \min & c_1^T x_1 & +E[c_2 x_2 + & \ldots + c_T x_T] & \\ \text{subject to} & A_1 x_1 & & & = b_0, \\ & & B_{t-1} x_{t-1} & +A_t x_t & = \xi_t, t = 2, \ldots, T, \\ & & x_t \geq 0, & t = 1, \ldots, T, & \end{array} \tag{2}$$

where the sizes of the matrices are consistent with $x_t \in \Re^{n_t}, t = 1, \ldots, T$, and $b_1 \in \Re^{m_1}$,and the random vectors, $\xi_t \in \Re^{m_t}$. Note that this implies that the decisions x_t in future periods are random and depend on the past outcomes.

Many possibilities have been suggested for solving (2). We describe these briefly below.

One approach to (2) is to assume that the data are Markovian and to fit the stochastic program into the framework of Markov decision processes. Grinold showed how this could be done in (1976). For computation, it would generally be necessary to restrict the outcomes from one period to some finite set. This may not, however, be possible in general and introduces error that may not be easily quantifiable.

A similar approach is to fit the stochastic program into the framework of a dynamic program and to approximate the value function using suitable functions. The value function is defined recursively using:

$$V_{T-1}(\mathbf{x_{T-1}}) = \mathrm{E}\ [\min\ \mathrm{f_T}(\mathbf{x_T}) \text{ subject to } \mathrm{g_T}(\mathbf{x_{T-1}}, \mathbf{x_T}) = \mathbf{0}],$$

and the recursion:

$$V_t(\mathbf{x_t}) = \mathrm{E}\ [\min\ \mathrm{f_{t+1}}(\mathbf{x_{t+1}}) + \mathbf{V_{t+1}}(\mathbf{x_{t+1}}) \text{ subject to } \mathrm{g_{t+1}}(\mathbf{x_t}, \mathbf{x_{t+1}}) = \mathbf{0}].$$

This approximation appears in Beale, Dantzig, Watson (1986), by assuming a suitable functional for V_t (for example, a quadratic function) and then fitting this function using the mean value and the mean and standard deviations of the random variables. This has an especially convenient form for production problems. The piecewise linear approximation discussed in this paper follows many of the basic ideas in this approach.

The use of scenarios is also common in implementations. In this case, a scenario corresponds to a specific realization of $\xi_2, \ldots, \xi_T$. The scenario approach uses a limited number of these realizations and combines the solutions in some suitable way to obtain a solution to the original problem. Rockafellar and Wets (1988) show how this can be implemented in a procedure that converges to the solution of the stochastic program (1) in which all of the scenarios are included.

Implicit in this approach and other procedures with finite numbers of scenarios is some discretization and aggregation of random variables. The basic idea behind this approach is to combine several periods and scenarios (realizations of the random variable) together into an aggregated problem. For example, the last $T - k + 1$ periods may be replaced by an objective, $f^k(\mathbf{x^k}) = \sum_{\mathbf{t=k}}^{\mathbf{T}} \alpha_\mathbf{t} \mathbf{f_t}(\mathbf{T_t x^k})$, and the constraints become $g^k(\mathbf{x^k}, \mathbf{T_1 x^k}) = \sum_{\mathbf{t=k}}^{\mathbf{T}} \alpha_\mathbf{t} \mathbf{g_t}(\mathbf{T_t x^k}, \mathbf{T_{t+1} x^k})$, where the T_t are appropriate shift operators to have $\mathbf{x^k}$ act on the correct underlying stochastic element.

Lower bounds are available in this approach by using expected values when the functions f are convex and the constraints are linear. Upper bounds are more difficult to achieve. For linear f, Birge (1985a) develops upper and lower bounds that assume bounded $\mathbf{x}$ and linear penalties for violations of the constraints. They are based on constructing feasible primal and dual solutions. The assumptions can be relaxed for problems with special structure, such as production problems (see Birge,1985b). It may also be possible to develop bounds on the primal and dual variables using the sublinear approximation procedure.

The piecewise linear approximation presented here is an extension of this approach in which we find functions that approximate (and bound) response to variations about the random parameter. The next section describes the basic principles behind this approach.

3. PIECEWISE LINEAR APPROXIMATIONS

The difficulty with discretization approaches is that many functions evaluations are required to obtain bounds on the optimal values (see Birge and Wets (1986a)). The piecewise linear response approach is an alternative that can reduce work considerably. The basis for this is the ray approximation procedure in Birge and Wets (1986a,b). This uses the sublinearity property of the recourse function to obtain a separable function that majorizes V_t, the value function at stage t. This approach is generalized in Birge and Wets (1987, 1989). Wallace (1987), on the other hand, formulated a procedure that applies to problems in which the recourse function involves the solution of a network problem. The piecewise linear procedure in Birge and Wallace (1988) is a combination and generalization of these two basic approaches for two-stage problems ($T = 2$).

The two-stage algorithm provides a separable piecewise linear function that can bound V_1 throughout the support of the random variables and can be easily evaluated. This procedure can be extended to multiple stages as we show in the next section. We first show the results for one period.

To simplify notation and to establish general results, we consider the following system :

$$A_1 x = b_1 + \xi, \quad A_2 x = b_2, \quad 0 \le x \le c + \phi \tag{3}$$

where $A_1 \in \Re^{m_1 \times n}$, $A_2 \in \Re^{(m-m_1)\times n}$, $(A_1|A_2)^T = A$ is the coefficient matrix, $(b_1|b_2)^T = b$ is the fixed part of the right-hand side, c is the fixed part of the bounds on the variables, ξ is the random availability of resources and ϕ is the random part of the variable capacities, where $\phi \ge 0$. We assume that there is a positive probability that $\phi = 0$. Next define $Q(\xi, \phi)$ by

$$Q(\xi, \phi) = \min\{q^T x | (3)\}. \tag{4}$$

Finally define $\chi(\xi, \phi, d^-, d^+)$ as the set of x-vectors satisfying

$$A_1 x = \xi, \ A_2 x = 0, \ d^- \le x \le \phi + d^+. \tag{5}$$

Our goal is to find an upper bound on $Q(\xi, \phi)$, or, more precisely, on $\mathrm{E}Q(\xi, \phi)$. We do this by finding a separable piecewise linear function $U(\xi, \phi)$ defined by

$$U(\xi, \phi) = Q(\bar{\xi}, 0) + H(\phi) + \sum_{i=1}^{m_1} \begin{cases} q^T x^{i+}(\xi_i - \bar{\xi}_i) & \text{if } \xi_i \ge \bar{\xi}_i, \\ q^T x^{i-}(\bar{\xi}_i - \xi_i) & \text{if } \xi < \bar{\xi}_i \end{cases}$$

where $\bar{\xi}_i = \mathrm{E}\xi_i$, and $H(\phi)$ is a piecewise linear function in ϕ. The last term can also be replaced by a general piecewise linear function, $f(\xi_i - \bar{\xi}_i)$, to obtain sharper bounds.

4. MULTISTAGE APPROXIMATIONS

The basic idea is to use the two-stage method repeatedly to approximate the objective function by separable functions. For linear problems, this leads to sublinear or piecewise linear functions as in the previous section. Functions without recession directions (e.g., quadratic functions) would require some type of nonlinear (e.g., quadratic) function that should again be easily integrable, requiring, for example, limited moment information (second moments for quadratic functions). We consider the linear case below.

The goal is to construct a problem that is separable in the components of the random vector. Let η_t be the random right-hand side that appears in period t. We apply the two-stage approach to this problem and then use the resulting solution as an input to the next stage of the problem. This random vector η depends on the first-stage decision. Its distribution is found directly (as, e.g., a piecewise normal distribution if the original random elements all have normal distributions), or it can be approximated using moment information. In this case, the distribution that solves the appropriate moment problem can be used to provide a bound.

The problem is then to find the distribution of the η_t and the form of the piecewise linear functions ψ_i^t for each component i of η. The resulting problem to solve is:

$$\min c_1 x_1 + \sum_{t=1}^{T} \sum_{i=1}^{m_t} \int \psi_i^t(\eta_t(i)) P(d\eta_t(i))$$

$$\text{subject to } \mathrm{A}_1 \mathrm{x}_1 = \mathrm{b}_1, \mathrm{x}_1 \geq 0. \qquad (SL)$$

To find η and ψ recursively, assume that we know the distribution of η_t. (Initially, this is $\eta_2 = \xi_2 - B_1 x_1$.) The piecewise linear functions are formed by solving the parametric linear programs in α:

$$\min c_t x_t \text{ subject to } \mathrm{A}_t \mathrm{x}_t = \pm \alpha \mathrm{e}(\mathrm{i}), \mathrm{x}_t \geq \beta_t, \qquad (sub(t-i))$$

to obtain $x_t^{\pm i}$ that depends on the parameter α and the lower bound β_t, which is generally zero but may have negative components if certain random variables are bounded.

If we restrict ourselves to two pieces, the function is sublinear, so that $\psi_t^i(\eta_t(i)) = (c_t x_t^{+i} 1_{\eta_t(i) \geq 0} + c_t x_t^{-i} 1_{\eta_t(i) < 0})(|\eta_t(i)|)$. We then have that $\eta_{t+1}(j)$ is given by:

$$\eta_{t+1}(j) = \xi_{t+1}(j) - B_t(j, \cdot)[\sum_{i=1}^{m_t} (x_t^{+i} 1_{\eta_t(i) \geq 0} + x_t^{-i} 1_{\eta_t(i) < 0})(|\eta_t(i)|)], \quad (6)$$

where we must find the resulting distribution. Note that the values in (6) are linear functions of η_t on the regions where η_t has constant sign. We can, therefore, construct η_{t+1} as a function of η_t by overlaying these linear transformations of random variables. For normally distributed data, this may be possible since the transformation will not affect the distribution class.

5. EXTENSIONS

The approximation given above may be difficult to compute if normal distributions are not present. Instead we may approximate the distribution of η_{t+1} . This requires bounds on the $prob.\{\eta_t(i) \geq 0\}$ and on the moments conditional on $\eta_t(i) \geq$ or < 0. Given these values, moment problems can be solved to calculate corresponding values for η_{t+1} and to bound ψ_t^i. Any other bounds on the input from period t to period $t+1$ ($B_t x_t$) can also be used to obtain bounds on the ψ values so that crude bounds can be obtained. This may be possible in special cases.

Note also that certain problems, such as networks, may have very few variables present in the $B_t(i,\cdot)$ term in (6) (because they have a close-to-simple recourse structure), so that η_{t+1} may be easily calculable for these problems. These special cases require further investigation.

Another looser but more implementable bound can be obtained by forcing a feasible and separable response in all future periods depending on a single random variable in the current period. This eliminates the problem of characterizing the distribution of inputs to all periods. It does, however, force a dependency in future periods that may increase the bound.

To develop this response function, let $X_t(\pm i)$ be an optimal solution, $x_t, \ldots, x_T$, to :

$$\begin{array}{llllll} \min & c_t^T x_t & + \ldots & & +c_T^T x_T & \\ \text{subject to} & A_t x_t & & & & = \pm e_i, \\ & B_t x_t & +A_{t+1} x_{t+1} & & & = 0, \\ & & & \cdots & & \vdots \\ & & & & A_T x_T & = 0, \\ & & x_\tau \geq 0, & \tau = t, \ldots, T. & & \end{array} \tag{7}$$

Now define $\phi^{t,i}(\hat{\xi}_t(i)) = \mathrm{E}\{C_t^T X_t(\pm i)(\xi_t(i) - \hat{\xi}_t(i))^{\pm}\}$, where $C_t = (c_t, \ldots, c_T)$. An upper bound on the objective value of (1) is obtained by solving the separable nonlinear program:

$$
\begin{array}{llllll}
\min & c_1^T x_1 & \ldots + c_T^T x_T & + \sum_t \sum_i (\phi^{t,i}(\hat{\xi}_t(i)) & & \\
\text{subject to} & A_1 x_1 & & & = b, & \\
& B_t x_t & + A_{t+1} x_{t+1} & -\hat{\xi}_{t+1} & & = 0, \\
& & & & t = 1, \ldots, T-1, & \\
& & x_t \geq 0, & t = 1, \ldots, T. & & \\
& & & \hat{\xi} \in \Xi, & &
\end{array}
\tag{8}
$$

where Ξ is the support set of the random variables. Note that if we drop the nonlinear term in the objective, and replace $\hat{\xi}$ in the constraints with a fixed valued of $\mathrm{E}[\xi]$, then we can obtain a lower bound on the optimal objective value in (1). The upper bound from (8) represents allowing some variation of the choice of the centering point $\hat{\xi}$ so that a best approximation is found. The value of (8) is an upper bound because the composition of the x_t solutions from (8) and the X_t values used in the ϕ terms yield a feasible solution for all ξ.

This procedure may also be implemented as response to several scenarios. In this case, the random vectors are partitioned and the response functions are evaluated over the partitions. The partitions may also be part of the higher level optimization problem so that in some way a "best" partition may be found.

The points used within the partitions may be chosen as expected values, in which case, the solution without penalty terms is again a lower bound on the optimal objective value. For an upper bound, this vector may be allowed to take on any value in the partition.

The use of multiple scenarios enters directly into the scenario aggregation approach of Rockafellar and Wets. This can be used to solve the top level problem and to approach a solution that is optimal for a given set of partitions and the piecewise linear penalty structure presented here.

Computations are then restricted to optimizing separable nonlinear functions subject to linear constraints. Implementations can be based on previous procedures (such as decomposition). The value of the method can, however, only be demonstrated by solving practical problems and observing the relative errors. Implementation is underway to achieve this goal.

8. REFERENCES

E.M.L. Beale, G.B. Dantzig, and R.D. Watson, "A first order approach to a class of multi–time period stochastic programming problems," *Mathematical Programming Study* 27 (1986) 103–117.

J.R. Birge, "Aggregation in stochastic linear programming," *Mathematical Programming* 31 (1985a) 25–41.

J.R. Birge, "Aggregation in stochastic production problems,"in *Proceedins of the 11th IFIP Conferenc on System Modelling and Optimization*, Springer-Verlag, Berlin (1985b), pp. 672–683.

J.R. Birge and S.W. Wallace, "A separable piecewise linear upper bound for stochastic linear programs ," *SIAM J. Control* 36 (1988) 725-739.

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 (1986a) 54–102.

J.R. Birge and R. J–B. Wets, "On–line solution of linear programs using sublinear functions," Technical Report No. 86–25, Department of Industrial and Operations Engineering, The University of Michigan (Ann Arbor, MI, 1986b).

J.R. Birge and R. J–B. Wets, "Computing bounds for stochastic programming problems by means of a generalized moment problem," *Mathematics of Operations Research* 12 (1987) 149–162.

J.R. Birge and R. J–B. Wets, "Sublinear upper bounds for stochastic linear programs," *Mathematical Programming* 43 (1989) 131–149.

Y. Ermoliev and R. Wets,*Numerical Techniques for Stochastic Optimization*, Springer, Berlin, 1988.

R.C. Grinold, "A new approach to multistage stochastic linear problems," *Mathematical Programming Study* 6 (1976) 19–29.

P. Kall, K. Frauendorfer and A. Ruszczynski, "Approximation techniques in stochastic programming," in: Y. Ermoliev and R. Wets, eds., *Numerical Techniques for Optimization*, Springer–Verlag, Berlin, 1988.

F.V. Louveaux, "Optimal investment for electricity generation: a stochastic model and a test problem," in: Y. Ermoliev and R. Wets, eds., *Numerical Techniques for Stochastic Optimization* , Springer–Verlag, Berlin, 1988.

A. Prékopa, "Boole–Bonferroni inequalities and linear programming," Rutgers Research Report #4–86, Rutgers Center for Operations Research, Rutgers University (New Brunswick, NJ, 1986).

R.T. Rockafellar and R. Wets, *Scenario aggregation in stochastic programming*, IIASA report, Laxenburg, Austria, 1988.

S.W. Wallace, "A piecewise linear upper bound on the network recourse problem," *Mathematical Programming* 38 (1987) 133-146..

R. J–B. Wets, "Large scale linear programming techniques in stochastic programming," in: Y. Ermoliev and R. Wets, eds., *Numerical Techniques for Stochastic Optimization* , Springer–Verlag, Berlin, 1988.