Aggregation Bounds in Stochastic Linear Programming

John R. Birge

Technical Report 83-1

Department of Industrial and Operations Engineering
The University of Michigan
Ann Arbor, Michigan 48109

January 1983

Aggregation Bounds in Stochastic Linear Programming

John R. Birge

Department of Industrial and Operations Engineering
The University of Michigan
Ann Arbor, Michigan 48109

Abstract

Stochastic linear programs become extremely large and complex as additional uncertainties and possible future outcomes are included in their formulation. Row and column aggregation can significantly reduce this complexity, but the solutions of the aggregated problem only provide an approximation of the true solution. In this paper, error bounds on the value of the optimal solution of the original problem are obtained from the solution of the aggregated problem. These bounds apply for aggregation of both random variables and time periods.

1. Introduction

Stochastic linear programming models of dynamic decision making problems become extremely complex as the numbers of stochastic random variables and time periods increase. Techniques for solving these problems have been limited to simplified versions of the full problem. For example, Ashford [1] and Beale, Forrest, and Tomlin [3] have proposed an algorithm for the specific structure of production planning problems. Also, in Birge [4], a method for general problems is implemented but with only up to three stages or time periods.

In large, long range problems with general structure, it appears that some approximation technique is needed. In this paper, we use the aggregation method of Zipkin [9, 10] to reduce the multi-stage problem to one which can more easily be solved. We assume that penalties exist that provide us with the framework for bounds on the value of the full problem. We assume that the model is sufficiently general to allow for an infinite time-horizon and for continuously distributed random right-hand side vectors.

In Section 2 of the paper, the problem is outlined and our initial assumptions are presented. Section 3 presents our main results for aggregating over random variables and time periods. Section 4 presents alternative assumptions for the bounding conditions, and Section 5 summarizes our results.

2. The Multi-Stage Problem and Assumptions

Beale [2] and Dantzig [5] first proposed that uncertainties in linear programs be incorporated into stochastic linear programs. This formulation later evolved into the multi-stage stochastic program. This problem has many applications in decision making problems (see, for example, Dempter [6] and Wets [8]), and may be formulated as

$$\max \phi = c_1 x_1 + E[\sum_{t=2}^{T} \rho^{t-1} c_t x_t(\xi_2, \xi_3, \ldots, \xi_t)] \tag{1}$$

Subject to

$$A_1 x_1 \leq b;$$

$$B_{t-1} x_{t-1}(\xi_2, \xi_3, \ldots, \xi_{t-1}) + A_t x_t(\xi_2, \xi_3, \ldots, \xi_t) \leq \xi_t,$$

$$t = 2, 3, \ldots, T;$$

$$x_t \geq 0, \; t = 1, 2, \ldots, T; \; \xi_t \in \Xi_{t-1}, t = 1, 2, \ldots, T;$$

where $x_t \in R^{n_t}$ represents decisions made in time period t after the realization of random vectors $\xi_1, \xi_2, \ldots, \xi_t$ where $\xi_t \in R^{m_t}$, $c_t \in R^{n_t}$, $b \in R^{m_1}$, $A_t \in R^{m_t \times n_t}$, and $B_t \in R^{m_t \times n_{t-1}}$. ρ is a discount factor that may be used to yield a finite sum when T approaches infinity.

In our analysis, we assume that $c_t = c$ for all t, $A_t = A$ for all t, and $B_t = B$ for all t. This may be done for general problems by increasing the number of variables in each time period and bounding those which do not appear in a specific period. We also note that our model is not completely general in that uncertainties only appear in the right-hand side. Again, we may formulate general problems with constraint uncertainties in this manner (Birge [4]), but the number of variables will grow very quickly.

In computing bounds for (1), we will aggregate realizations of a single random variable within a time period and over several time periods. Our approach follows Zipkin [9, 10], and we will use consistent notation where possible. In order to use Zipkin's results, we first must make certain assumptions about the values of the optimal primal and dual variables in (1), x^* and π^*, respectively.

Assumption 1. There exists some set of upper bound vectors, u_t, $t = 1, 2,\ldots, T$, such that $x_t\ (\xi_2, \xi_3,\ldots,\xi_t) \leq u_t$, for all t and all $\xi_2, \xi_3,\ldots, \xi_t$.

Assumption 2. There exists a partition of $x_t(\xi_2, \xi_3,\ldots,\xi_t)$ and A_t: $x_t(\xi_2, \xi_3,\ldots, \xi_t) = (z_t(\xi_2, \xi_3,\ldots,\xi_t), y_t(\xi_2, \xi_3,\ldots,\xi_t))$ and $A_t = [A_t\vdots -I]$, such that

$$B_{t-1}x_{t-1}(\xi_2, \xi_3,\ldots,\xi_{t-1}) + A_t z_t(\xi_2, \xi_3,\ldots, \xi_t)$$

$$- y_t(\xi_2, \xi_3,\ldots,\xi_t) \leq \xi_t,$$

for all t, and where the objective function coefficients of y_t in (1) are penality functions, $- p_t$, where $p_t \geq 0$.

These two assumptions restrict the class of problems to which we apply our results, but they are sufficiently general to include most realistic examples. For instance, most problems can be essentially bounded by some criteria as in Assumption 1. Assumption 2 is stating that a feasible solution always exists as is true in almost all properly formulated models. In Section 4, more general conditions are given that enhance the class of applicable problems.

3. Aggregation Bounds

Aggregations will be performed both for rows and columns over random variables and time periods. We define partitions of the columns and rows in (1) such that $\sigma = \{S_\alpha\}$ where $\bigcup_\alpha S_\alpha$ is the set of all columns and

$P = \{R_\beta\}$ where $\bigcup_\beta R_\beta$ is the set of all constraints.

We then assign weighting functions to these partitions, g^α and f^β, such that for all columns $\alpha(\eta)$ in S_α,

$$\int_{\alpha(\eta)\,\varepsilon\,S_\alpha} g^\alpha\, \alpha(\eta)\, d(\alpha(\eta)) = 1,$$

where $g^\alpha(\alpha(\eta)) \geq 0$, and, for all rows $\beta(\eta)$ in R_β,

$$\int_{\beta(\eta)\,\varepsilon\,R_\beta} f^\beta(\beta(\eta))\, d(\beta(\eta)) = 1,$$

where $f^\beta(\beta(\eta)) \geq 0$. These weighting functions are generalizations of the weighting vectors used by Zipken. They are used here because the columns and rows in (1) are not necessarily discretely represented. If discrete distributions are assumed then the weighting functions take on point mass values and become weighting vectors.

We will first partition (1) so that random variables are aggregated together. For this problem, we have a discrete number of partitions. We define

$$S_i = \{\text{Columns corresponding to } x_1(i)\},$$

and

$$S_{tn+i} = \{\text{Columns corresponding to } x_t(\xi_2,\ \xi_3,\ldots,\xi_t)\,(i)\},$$

for $t = 1, 2,\ldots,T$. The rows are then taken as

$$R_i = \{\text{Rows corresponding to the right-hand sides, } b_i\},$$

$$R_{tn+i} = \{\text{Rows corresponding to the right-hand sides, } \xi_t(i)\},$$

for $t = 1, 2,\ldots, T$.

The weighting functions then correspond exactly to the distribution functions on the random variables. We then formulate the expected value aggregate problem of (1) as

$$\text{Max} \quad \Phi = cX_1 + \sum_{t=2}^{T} \rho^{t-1} cX_t \tag{2}$$

Subject to

$$\begin{aligned} AX_1 \qquad\qquad & \leq b; \\ BX_{t-1} + AX_t & \leq \bar{\xi}_t; \end{aligned}$$

$$X_t \geq 0, \quad t = 1, 2, \ldots, T;$$

where $\bar{\xi}_t(i) = \int_{\Xi_t} \xi_t(i) d\, F(\xi_t)$

$$= \int_{i(\eta) \,\varepsilon\, R^i} \xi_t(i) f^i_{t_i}(i(\eta)) d(i(\eta)).$$

Problem (2) is a vastly simplified linear program from (1). The optimal primal and dual solutions of (2) are $(X^*, \Pi^*) = ((Z^*, Y^*), \Pi^*)$. From these values, we obtain a fixed weight approximate solution of (1), $(\tilde{x}, \tilde{\pi}) = ((\tilde{z}, \tilde{y}), \tilde{\pi})$, where

$$\tilde{z}_t(\xi_2, \xi_3, \ldots, \xi_t) = Z_t^* \, dF(\xi_2, \xi_3, \ldots, \xi_t),$$

$$\tilde{\pi}_t(\xi_2, \xi_3, \ldots, \xi_t) = \Pi_t^* \, dF(\xi_2, \xi_3, \ldots, \xi_t),$$

and $\tilde{y}_t(\xi_2, \xi_3, \ldots, \xi_t)$ is defined such that $\tilde{x}$ is feasible in (1). This definition is necessary because some infeasibilities may have been caused by aggregating rows in (2). We can now state bounds on the optimal solution

ϕ^* of (1). We let $\tilde{\phi} = c_1\tilde{x}_1 + E[\sum_{t=0}^{T} \rho^t c\, \tilde{z}_t(\xi_2, \xi_3, \ldots, \xi_t)]$.

Proposition 1. Given Assumptions 1 and 2, for the optimal values ϕ^* of (1) and $\tilde{\phi}$ as defined above,

$$\tilde{\phi} - \varepsilon^- \leq \phi^* \leq \tilde{\phi} + \varepsilon^+, \tag{3}$$

where $\varepsilon^+ = \sum_{t=1}^{T} \sum_{j=1}^{n} \int_{\Xi_2 \times \Xi_3 \times \ldots \times \Xi_{t+1}} \max\{\rho^t c_j - \Pi_t^* A_{.j} - \Pi_{t+1}^* B_{.j}, 0\} u_j^t dF(\xi_2, \xi_3, \ldots, \xi_t)$

and the third term in the maximand vanishes for t = T, and

$$\varepsilon^- = \sum_{i=1}^{T} \sum_{j=1}^{m} [\int_{\Xi_2 \times \Xi_3 \times \ldots \times \Xi_{t+1}} \max\{\xi_{t+1} - A_{j.} Z_{t+1}^* + B_{j.} Z_t^*, 0\} \cdot p_t(j) dF(\xi_2, \xi_3, \ldots, \xi_t)].$$

Proof: The bounds essentially represent penalties for primal and dual infeasibilities. They are direct extensions of Proposition 2 in [10]. To obtain this, we must only show that the assumptions for that proposition are met. Those assumptions in our case generalize to

a) For some partition $\sigma' = \{S'_{\alpha'}\}$ of the columns of (1), and for a positive function, $d(x_i)$, defined for all x_i, and for nonnegative values, $\{p_{\alpha'}\}$ for all α', there exists an optimal solution x^* of (1) such that

$$\int_{x_i : i \in S'_{\alpha'}} d(x_i)\, x_i^*\, dx_i \leq p_{\alpha'}, \text{ for all } \alpha', \text{ and}$$

b) For some partition $P' = \{R'_{\beta'}\}$ of the rows of (1), for a positive function $e(\pi_j)$, defined for π_j for all rows, and for nonnegative values $\{q_{\beta'}\}$ for all β', there exists an optimal dual solution π^* of (1) such that

$$\int_{\pi_j : j \varepsilon S'_{\beta'}} e(\pi_j)\, \pi^*_j \, d\pi_j \leq q_{\beta'}, \quad \text{for all } \beta'.$$

Our Assumption 1 clearly implies (a) above since $x_t(\xi_2, \xi_3, \ldots, \xi_t) \leq u_t$ for all t.

To show (b), we note that the feasibility of π^* implies by Assumption 2 that $- p_t + \dot{\pi}^*_t \leq 0$ or $\pi^*_t \leq p_t$ for all t, implying (b). Given these assumptions, we obtain the bounds in (3) as straight forward generalizations of the results in [10]. ∎

Although Problem (2) is much simpler than (1), there may still be difficulties in solving it because of the number of periods. We propose a further simplification by aggregating variables for future time periods. We will only present results for the most extreme aggregation in which all of periods 2 through T have been collapsed into a single second period. Other cases involving some partial aggregation of these periods are also possible. For our partition, we let S_i for i = 1, 2,...,n be defined as for (2) and let

S_{n+i} = {Columns corresponding to $x_t(\xi_2, \xi_3, \ldots, \xi_t)(i)$ for all ξ and t}.

We let R_i be the same for i = 1, 2,...,m, and define

R_{n+i} = {Rows corresponding to the right-hand sides, $\xi_t(i)$ for all ξ and t}.

The weighting functions consist of both distribution functions and time period factors $\gamma(t)$ for the columns and $\delta(t)$ for the rows such that

$$\sum_{t=2}^{T} \gamma(t) = 1, \text{ and}$$

$$\sum_{t=2}^{T} \delta(t) = 1.$$

To obtain the two period aggregate problem from (1), we use (2) and find

$$\begin{aligned}
&\max && \Phi = c\,\hat{X}_1 + \hat{c}\,\hat{X}_2 \\
&\text{subject to} && A\,\hat{X}_1 + \hat{A}\,\hat{X}_2 \le b \\
& && B\,\hat{X}_1 + \hat{A}\,\hat{X}_2 \le \hat{\bar{\xi}}\,; \\
& && \hat{X}_1,\ \hat{X}_2 \ge 0,
\end{aligned} \tag{3}$$

where

$$\hat{c} = \sum_{t=2}^{T} \gamma(t)\,\rho^{t-1} c,$$

$$\hat{A} = \sum_{t=2}^{T} \delta(t)\gamma(t)\,A + \sum_{t=3}^{T} \delta(t)\sigma(t-1)B + \delta(2)B; \text{ and}$$

$$\hat{\bar{\xi}} = \sum_{t=2}^{T} \delta(t)\bar{\xi}_t.$$

Problem (3) is a simplified version of (2). The weighting factors $\gamma(t)$ and $\delta(t)$ may be used to discount the future or emphasize effects in certain periods. From the optimal primal and dual solutions, $(\hat{X}^*, \hat{\Pi}^*) = (\hat{Z}^*, \hat{Y}^*), \hat{\Pi}^*)$ of (3), we obtain a fixed weight aggregation solution of (1) for $t > 1$, $(\hat{x}, \hat{\pi}) = ((\hat{z}, \hat{y}), \hat{\pi})$, by

$$\hat{z}_t = (\xi_2, \xi_3, \ldots, \xi_t) = \hat{Z}_2^* \gamma(t) dF(\xi_2, \xi_3, \ldots, \xi_t);$$

$$\hat{\pi}_t = (\xi_2, \xi_3, \ldots, \xi_t) = \hat{\Pi}_2^* \delta(t) dF(\xi_2, \xi_3, \ldots, \xi_t);$$

and $\hat{y}_t(\xi_2, \xi_3, \ldots, \xi_t)$ again defined such that $\hat{x}$ is feasible in (1). For $\hat{\phi} = c\,\hat{x}_1 + E\,[\sum_{t=0}^{T} \rho^t\, c\, \hat{z}_t(\xi_2, \xi_3, \ldots, \xi_t)]$ we obtain the bounds of the following proposition.

<u>Proposition 2</u>. The optimal ϕ^* of (1) is bounded by

$$\hat{\phi} - \varepsilon^- \leq \phi^* \leq \hat{\phi} + \varepsilon^+,$$

where

$$\varepsilon^+ = \sum_{t=1}^{T} \sum_{j=1}^{n} \{ \int_{\Xi_2 \times \ldots \times \Xi_{t+1}} \max\{\rho^t c_j - \delta(t)\hat{\Pi}_t^* A - \delta(t+1)\hat{\Pi}_{t+1} B, 0\}$$

$$\cdot\, u_j^t dF(\xi_2, \xi_3, \ldots, \xi_{t+1})\},$$

where 0 is substituted for $\delta(T+1)\hat{\Pi}_{T+1}^* B$ and where $\delta(1) = 1$, and

$$\varepsilon^- = \sum_{t=1}^{T} \sum_{j=1}^{m} [\int_{\Xi_2 \times \ldots \times \Xi_{t+1}} \max\{ \xi_{t+1} - A\,\gamma(t+1)\hat{Z}^* - B\gamma(t)\hat{Z}^*, 0\}$$

$$\cdot\, p_t(j) dF(\xi_2, \xi_3, \ldots, \xi_{t+1})].$$

Proof: These inequalities follow directly as in the proof of Proposition 1 from the definitions of the variables. ■

Other combinations of these two forms of aggregation may also be used. One may, for instance, aggregate time periods without aggregating random variables. Conditional means may be used for aggregating random variables and other groups of variables within time periods may be aggregated.

4. Alternative Assumptions.

Assumptions 1 and 2 in the previous section may not yield adequate bounds on the optimal solution value. These penalties and variable bounds may not be readily apparent, and it may be necessary to look for alternative assumptions. In this section, we describe certain assumptions that may make the bounds we have presented more useful.

We first let $\hat{\sigma} = \{\hat{S}_k : \hat{k} = 1,\ldots,\hat{K}\}$ be a partition of $\{1, 2,\ldots,n\}$ and and $\hat{P} = \{\hat{R}_{\hat{\ell}} : \hat{\ell} = 1,\ldots\hat{L}\}$ be a partition of $\{1, 2,\ldots m\}$. The propositions above allow us to replace Assumptions 1 and 2 with new conditions.

Proposition 3. If there exists a partition $\hat{\sigma}$ such that

$$\sum_{i\in S_{\hat{k}}} \sum_{j\in R(\hat{k})} (B_{ji} + A_{ji}) > 0, \tag{4.1}$$

$$\sum_{i\notin S_{\hat{k}}} \sum_{j\in R(\hat{k})} (B_{ji} + A_{ji}) \geq 0, \tag{4.2}$$

and $$\sum_{j\in R(k)} \xi(j) \geq 0 \quad \text{for all } \xi^t, \tag{4.3}$$

where $R(\hat{k}) \subseteq \{1, 2,\ldots,m\}$ for each $\hat{k} = 1,\ldots,\hat{K}$, then Assumption (a) is satisfied.

Proof: For x^* feasible in (1), we must have

$$B x^*_{t-1} + A x^*_t \leq \xi^t,$$

so $\sum_{i=1}^{n} (B_{ji} x^*_{t-1}(i) + A_{ji} x^*_t(i)) \leq \xi^t(j)$, for all j and t. Clearly, the conditions of (4.1) , (4.2), and (4.3) satisfy the conditions in Assumption (a) for all t. ■

Proposition 4. If there exists a partition $\hat{p}$ of {1, 2,...,m} such that

$$\sum_{j \in R_{\hat{\ell}}} \sum_{i \in S(\hat{\ell})} (B_{ji} + A_{ji}) < 0, \tag{5.1}$$

$$\sum_{j \notin R_{\hat{\ell}}} \sum_{i \in S(\hat{\ell})} (B_{ji} + A_{ji}) \leq 0, \text{ and} \tag{5.2}$$

$$\sum_{i \in S(\hat{\ell})} c_i \leq 0, \tag{5.3}$$

where $S(\hat{\ell}) \subseteq \{1, 2,...,n\}$ for each $\hat{\ell} = 1, 2,...,\hat{L}$, then Assumption (b) is satisfied.

Proof: For π^* dual feasible in (1), we need

$$\rho^t c - \pi^*_t A - \pi^*_{t+1} B \leq 0,$$

or

$$\pi^*_t(-A) + \pi^*_{t+1}(-B) \leq -\rho^t c.$$

Since $\pi^*_t \geq 0$ for all t for (1), conditions (5.1), (5.2), and (5.3) lead to Assumption (b). ■

It may still be difficult to determine whether the alternative conditions (4) and (5) can be met, but they can be checked for different subsets of the rows and columns. For the partitions, $\hat{\sigma} = \{\{1,\ldots,n\}\}$ and $\hat{p} = \{\{1,\ldots,m\}\}$, the conditions may be checked individually for each row or column, respectively.

When it can be shown that these conditions are satisfied, new bounds involving ε^- and ε^+ may be found for problems (3) and (4). They may also suggest different aggregation schemes to reduce the error.

5. Conclusion

The principle of aggregation has been applied to multi-stage stochastic linear programs and bounds have been found on the value of the optimal solution. The procedure assumes that the primal and dual variables are bounded by external penalties. We also presented alternative assumptions that may hold in specific cases.

These bounds are *a posteriori* in their requiring the solution of the aggregated problem. *A priori* bounds may also be found as in [10]. These bounds would follow the same derivation as we have given here.

For both *a posteriori* and *a priori* bounds, the bounds may be made much tighter by using specifics of the problem. Queyranne and Kao [7] have for example applied aggregation to a stochastic programming example in manpower planning and used the characteristics of that problem to provide precise results. We anticipate that many applications may have a special structure that will allow for more precise bounding than in the results presented here.

References

[1] Ashford, R. W., "A stochastic programming algorithm for production planning", Scicon report, 1982.

[2] Beale, E. M. L., "On minimizing a convex function subject to linear inequalities", J. Ray, Stat. Soc. 17, (1955), 173-184.

[3] Beale, E. M. L., Forrest, J. J. H., and Taylor, C., "Multi-time period stochastic programming", in Stochastic Programming, M. A. H. Dempster, ed., Academic Press, New York, 1980.

[4] Birge, J. R., "Decomposition and partitioning methods for multi-stage stochastic linear programs", Department of Industrial and Operations Engineering, The University of Michigan, Technical Report 82-6, 1982.

[5] Dantzig, G. B., "Linear programming under uncertainty", Management Science, 1, (1955), 197-206.

[6] Dempster, M. A. H., "Introduction to stochastic programming", in Stochastic Programming, M. A. H. Dempster, ed., Academic Press, New York, 1980.

[7] Queyranne, M. and Kao, E. P. C., "Aggregation in a two-stage stochastic program for manpower planning in the service sector", University of Houston report, 1981.

[8] Wets, R. J. B., "Stochastic programming: solution techniques and approximation schemes", in Mathematical Programming: State-of-the-Art 1982, Springer-Verlag, Berlin, 1983.

[9] Zipkin, P., "Bounds on the effect of aggregating variables in linear programming", Operations Research 28, (1980), 403-418.

[10] Zipkin, P., "Bounds for row-aggregation in linear programming", Operations Research, 28, (1980), 903-916.