The Relationship Between the L-Shaped Method and Dual Basis Factorization for Stochastic Linear Programming

John R. Birge

Technical Report 82-15

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

September, 1982

The Relationship Between the L-Shaped Method and Dual Basis Factorization for Stochastic Linear Programming

John R. Birge

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

Abstract

The basis factorization method of Strazicky for stochastic linear programs is shown to involve the same computational effort per iteration as the L-shaped method of Van Slyke and Wets. A variant of the factorization approach can then be found which is equivalent to the L-shaped method. The advantages of this decomposition approach over a standard factorization are discussed.

The Relationship Between the L-Shaped Method and Dual Basis Factorization for Stochastic Linear Programming

1. Introduction

We consider the problem

$$
(1) \qquad \begin{array}{lll} \min & cx + Q(x) & \\ \text{subject to} & Ax & = b \\ & x \geqq 0, & \end{array}
$$

where

$$Q(x) = E_\xi[\min q\, y \text{ subject to } Wy = \xi - Tx, \quad y \geqq 0].$$

and ξ is a random n_2 - vector, where A is an $m_1 \times n_1$ real matrix, W is an $m_2 \times n_2$ real matrix, T is an $m_2 \times n_1$ real matrix, and c, q, and b are correspondingly dimensioned vectors.

For $\xi \in \Xi = \{\xi^1, \xi^2, \ldots, \xi^N\}$, where $P(\xi = \xi^i) = p^i$, we have (1) is equivalent to

$$
(2) \qquad \begin{array}{llllll} \min & cx + p^1 q\, y^1 + p^2 q\, y^2 + \ldots + p^N q\, y^N & & & & \\ \text{subject to} & Ax & & & & = b \\ & Tx & + Wy^1 & & & = \xi^1 \\ & Tx & & + Wy^2 & & = \xi^2 \\ & \vdots & & \ddots & & \vdots \\ & Tx & & & + Wy^N & = \xi^N \\ & x, & y^1, y^2, \ldots, y^N \geqq 0. & & & \end{array}
$$

The dual of (2) is

$$\max \quad b^T\sigma + p^1\xi^{1^T}\pi^1 + p^2\xi^{2^T}\pi^2 + \ldots + p^N\xi^{N^T}\pi^N$$

subject to

$$(3)\qquad \begin{array}{llllll} A^T\sigma + p^1T^T\pi^1 & + p^2T^T\pi^2 & + \ldots & + p^N T^T\pi^N & \leq & c^T \\ W^T\pi^1 & & & & \leq & q^T \\ & W^T\pi^2 & & & \leq & q^T \\ & & \ddots & & & \vdots \\ & & & W^T\pi^N & \leq & q^T \end{array}$$

Kall [1979] and Strazicky [1974] observed that any feasible basis of (3) may be written as

$$(4)\qquad \left[\begin{array}{c:c} B & Y \\ \hdashline L & Z \end{array}\right]$$

where B is a block diagonal matrix. For (3),

$$B = \begin{bmatrix} \tilde{W}_1^T \;\; \tilde{I}_1 & & & \\ & \tilde{W}_2^T \;\; \tilde{I}_2 & & \\ & & \ddots & \\ & & & \tilde{W}_N^T \;\; \tilde{I}_N \end{bmatrix},$$

where $[\tilde{W}_i^T \mid \tilde{I}_i]$ is an $n_2 \times n_2$ submatrix of $[W^T \mid I]$ for all $i = 1, 2, \ldots, N$. Kall notes that we may reduce the size of $[\tilde{W}_i^T \quad \tilde{I}_i]$ by taking an $m_2 \times m_2$ nonsingular submatrix $\hat{W}_i^T$ from $\tilde{W}_i^T$. We have

$$(5) \quad \left[\begin{array}{c|c} \hat{W}_i^T & 0 \\ \hline \bar{W}_i^T & \bar{I}_i \end{array}\right] \left[\begin{array}{c} \pi_i \\ \hline \bar{\rho}^i \end{array}\right] = \left[\begin{array}{c} \hat{q} \\ \hline \bar{q} \end{array}\right] + \left[\begin{array}{c} \hat{I}_i \\ \hline 0 \end{array}\right] \hat{\rho}_i$$

or $(\hat{W}_i^T)^{-1}\,\hat{q} + (\hat{W}_i^T)^{-1}\,\hat{\rho}_i = \pi_i$, $\bar{\rho}_i = \bar{q} - (\bar{W}_i^T)(\hat{W}_i^T)^{-1}\,\hat{q} - (\bar{W}_i^T)(\hat{W}_i^T)^{-1}\,\hat{\rho}_i$. So, we can rewrite (5) as

$$(6) \quad [(\bar{W}_i^T)(W_i^T)^{-1} \mid \bar{I}_i] \begin{bmatrix} \hat{\rho}_i \\ \cdots\cdots \\ \bar{\rho}_i \end{bmatrix} = [\,\bar{q} - (\bar{W}_i^T)_{\cdot}^{\frac{1}{\cdot}} \cdot (\hat{W}_i^T)^{-1}\,\hat{q}\,].$$

(6) substantially reduces the number of rows from (5) but it has a significant draw back in terms of nonzero element storage. The sparse matrix $\bar{W}_i^T$ may be transformed into a very dense matrix $(\bar{W}_i^T)(\hat{W}_i^T)^{-1}$.

A more practical approach would be to store the inverse of $\hat{W}_i^T$ for the current basis (where $\hat{\rho}_i = 0$) and to generate elements of $(\bar{W}_i^T)(\hat{W}_i^T)^{-1}$ as needed. LU decomposition and sparse updating may be used to improve this approach. It may also be shown that this method involves the same computational effort per iteration as the L-shaped method of Van Slyke and Wets [1969] and that a modification of the dual basis factorization method will follow the same path as the L-shaped method. This modification amounts to the dual decomposition procedure proposed by Dantzig and Madansky [1961].

2. Discussion

We assume we have a feasible solution to (3), $(\sigma^0, \pi^{1,0}, \pi^{2,0}, \pi^{N,0}, \lambda^0, \rho^{1,0}, \rho^{2,0}, \ldots, \rho^{N,0})$, where $(\lambda^0, \rho^{1,0}, \rho^{2,0}, \ldots, \rho^{N,0})$ are slack variables. We also assume that $\tilde{W}_i^T = W^T$ for all i so that only columns from A^T and the identity are basic in the first set of constraints. In the pricing operation, we solve

$$A^B x_B + A^N x_N = b,$$

$$I_N x_N = 0,$$

where I_N is the set of basic identity columns. We have $x_B = (A^B)b^{-1}$ and check for $x_B \geq 0$. If some $x_B(i) < 0$, then that column in A^B is replaced and the problem is solved again. If we restrict ourselves to only checking for primal feasibility in the x variables, thenwe are solving the dual problem

$$\max \quad b^T \sigma$$

$$\text{subject to} \quad A^T \sigma \leq c^T - \sum_{i=1}^{N} p^i T^T \pi^{i,0},$$

or the primal problem

$$\min \quad (c - \sum_{i=1}^{N} p^i \pi^{i,0} T)\, x$$

$$\text{subject to} \quad Ax = b$$

$$\text{(7)} \qquad x \geq 0.$$

This is essentially the first step of the L-shaped method. The dual method involves the same steps of computing $A^B x_B = b$, $\pi A^B = c_B$ and $\rho = c_N - \pi A^N$ as in the primal method, so the computational effort is the same at each step. We note that this does not include pricing for y variables as would occur in the general dual method.

After all $x_B(i) \geq 0$ have been found, we let x^1 be the prices and we proceed to solve $\tilde{W}_i\, y^i = \xi^i - T\, x^1$ for all y^i. For every subproblem i, if $y^i(j) < 0$ then we choose a leaving column only from the identity columns in subproblem i. We relax dual feasibility in the first set of constraints. This process is equivalent to solving the subproblems

$$\begin{aligned} &\min \quad q\, y^i \\ (8)\quad &\text{subject to} \quad W\, y^i = \xi^i - T\, x^1, \\ &\qquad\qquad y^i \geqq 0, \end{aligned}$$

for all i as in the L-shaped method. We note again that the computations involved in a single iteration are the same in both methods except that we do not update the prices in the first set of constraints for the factorization method. We note also that solutions of (8) may be found quickly by finding all ξ^i for which a given basis is optimal.

After solving these problems, we obtain either an unbounded condition or all $y^i \geqq 0$. In the former case, some subproblem (8) is infeasible. We then look at the column in (3) which gave the unboundedness condition. For $y^i_j < 0$, $y^i_j = (\hat{W}_i)^{-1}(j,\cdot) \cdot (\xi^i - Tx^1)$ and the column $-[(\hat{W}_i)(j,\cdot)\cdot\bar{W}_i]^T \leqq 0$. We let $\pi = -(\hat{W}_i)(j,\cdot)$ and obtain

(9) $\quad \pi\,(\xi^i - Tx^1) > 0$, and

(10) $\quad \pi\, \bar{W}_i \leqq 0.$

(9) and (10) are the infeasibility conditions for (8) that we would find in the L-shaped method. In the dual method, we would choose a pivot from the first set of constraints so that we would force $\pi\,(\xi^i - Tx) \leqq 0$.

We introduce a new column in the main problem,

$$\begin{bmatrix} p^i(\xi^i \pi^i)^T \rho \\ \hdashline (p^i \quad T^T \pi^i)\rho \end{bmatrix}$$

where $\rho \geqq 0$. The main problem is then

$$\max \qquad b^T\sigma + p^i(\xi^i \ \pi^i)^T\rho$$

$$\text{subject to} \quad A^T\sigma + (p^i T^T \pi^i)\rho \leq \bar{c}^T,$$

$$\rho \geqq 0,$$

where $\bar{c}^T$ includes c^T and other fixed columns of π. This is equivalent to adding a constraint

$$(\pi^i \ T)\ x \leqq \pi^i \ \xi,$$

as in the L-shaped algorithm. We next solve the main problem again and repeat.

If after solving the subproblems all $y^i \geqq 0$, then either the problem is optimal or one of the first set of constraints in (3) has been violated. In this case, if we let $\theta = \sum_{i=1}^{N} p^i\ ((\xi^i)^T - (Tx_1)^T)\pi^{i,1}$, where $\pi^{i,1}$ is the initial solution of (3), then we have

$$\theta < \sum_{i=1}^{N} p_i((\xi^i)^T - (Tx_1)^T)\pi^{i,2}$$

where $\pi^{i,2}$ is the optimal subproblem i solution. We observe that either $\pi^{i,1}$ or $\pi^{i,2}$ or linear combinations of these solutions may be used to for solutions for the subproblems. We use this to obtain a substitute first period problem:

$$\max \quad b^T\sigma + \lambda_1\left(\sum_{i=1}^{N} p^i(\xi^i)^T\pi^{i,1}\right) + \lambda_2\left(\sum_{i=1}^{N} p^i(\xi^i)^T \pi^{i,2}\right)$$

$$\text{subject to} \quad A^T\sigma + \lambda_1\left(\sum_{i=1}^{N} p^i(\xi^i)^T \pi^{i,2}\right) + \lambda_2\left(\sum_{i=1}^{N} p^i(\xi^i)^T \pi^{i,2}\right) \leq c^T$$

$$(11) \qquad \lambda_1 + \lambda_2 = 1$$

$$\lambda_1 , \lambda_2 \geq 0.$$

We solve problem (11) and repeat by adding a column for feasibility of the subproblems or by adding a column for choices of subproblem solutions as in (11). We note that these are the same steps as in the L-shaped decomposition method where $\theta < \sum_{i=1}^{N} p^i((\xi^i)^T - (T x_1)^T)\pi^i$ and a constraint on θ,

$$(12) \qquad \left(\sum_{i=1}^{N} \pi^i T\right) x + \theta \geq \sum_{i=1}^{N} p^i(\xi^i)^T,$$

is added. The two methods with these specifications follow the same procedures for each iteration. We note also that these methods follow the same steps as Dantzig-Wolfe decompostion applied to the dual problem (3), (Dantzig-Mandansky [1961], Van Slyke and Wets [1969]).

3. Conclusion

We have shown that on each iteration of the L-shaped method, the number of steps is equivalent to that of the basis factorization method and that the L-shaped method may be viewed as a variant of the basis factorization approach. In general, however, the two methods will not follow the same path to optimality. By maintaining dual feasibility, the basis factorization restricts the path to optimality and requires more effort in checking for feasibility within the first set of constraints.

The decomposition variant of basis factorization also avoids two other problems inherent in the full factorization approach. For $X = B^{-1}Y$ in (4), the factorization approach uses the inverse of $(L X - Z)$ in performing simplex operations. X is composed of columns of B^{-1} since Y is composed of identity columns. The columns of B^{-1} need not be sparse and may be very dense, causing $(LX - Z)$ to be dense as well. The storage requirement for the nonzero elements of this $n_1 \times n_1$ matrix may be large.

Another difficulty in applying this factorization without decomposition is that, whenever an identity column in $\bar{I}_i$ in (5) is replaced, then $\hat{W}_i^T$ must be changed and $(LX - Z)$ changes. This pivot alters the prices x for all other blocks $j \neq i$. Therefore, a pivot step is required for each new block into which this identity column enters. By fixing x in the decomposition, whenever a new matrix $\hat{W}_i^T$ is introduced, all values ξ^j such that $y^j = (\hat{W}_i^T)^{-1}(\xi^j - Tx) \geq 0$ can be found without performing separate pivot operations. For very large N, the standard factorization scheme may be forced through a long sequence of pivots, whereas the decomposition approach may change these bases quickly. For problems with large N, then, the decomposition variant above is probably the only tractable basis factorization method.

References

G. B. Dantzig and A. Madansky, "On the solution of two-stage linear programs under uncertainty", in Proceedings, 4th Berkely Symposium on Math. Stat. and Prob., University of California Press, Berkeley, (1961) 165-176.

P. Kall, "Computational methods for solving two-stage stochastic linear programming problems", Journal of Appl. Math. and Physics, 30, (1979) 261-271.

B. Strazicky, "Some results concerning an algorithm for the discrete recourse problem", in Stochastic Programming, M.A.H. Dempster (ed.), Academic Press, New York, (1980) 263-271.

R. Van Slyke and R. Wets, "L-shaped linear programs with applications to optimal control and stochastic linear programs", SIAM Journal on App. Math. 17, (1969) 638-663.