ALGORITHMS BASED UPON GENERALIZED LINEAR PROGRAMMING FOR STOCHASTIC PROGRAMS WITH RECOURSE

J.L. Nazareth

October 1984
WP-84-81

Working papers are interim reports on work of the International Institute for Applied Systems Analysis and have received only limited review. Views or opinions expressed herein do not necessarily represent those of the Institute or of its National Member Organizations.

INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS
A-2361 Laxenburg, Austria

PREFACE

In this paper, the author discusses solution algorithms for a particular form of two-stage stochastic linear programs with recourse. The algorithms considered are based upon the generalized linear programming method of Wolfe.

The author first gives an alternative formulation of the original problem and uses this to examine the relation between tenders and certainty equivalents. He then considers problems with simple recourse, discussing algorithms for two cases: (a) when the distribution is discrete and probabilities are known explicitly; (b) when the probability distribution is other than discrete or when it is only known implicitly through some simulation model. The latter case is especially useful because it makes possible the transition to general recourse. Some possible solution strategies based upon generalized programming for general recourse problems are then discussed.

This paper is a product of the Adaptation and Optimization Project within the System and Decision Sciences Program.

ANDRZEJ WIERZBICKI
Chairman
System and Decision
Sciences Program

ALGORITHMS BASED UPON GENERALIZED LINEAR PROGRAMMING FOR STOCHASTIC PROGRAMS WITH RECOURSE

J.L. Nazareth

1. INTRODUCTION

We are concerned here with two-stage stochastic linear programs (SLP) with recourse, of the form

minimize $cx + \mathcal{Q}(x)$

subject to

$$Ax = b$$

$$x \geq 0 \tag{1.1a}$$

where

$$\mathcal{Q}(x) = E\{Q(x,h(w))\} \tag{1.1b}$$

and

$$Q(x,h(w)) = \inf_{y \geq 0} \{qy \mid Wy = h(w) - Tx\} \tag{1.1c}$$

In the above, only the right-hand-side h(w), is a random vector defined on a probability space whose events are denoted by w. E denotes expectation. T denotes the fixed $m_2 \times n_1$ technology matrix and W the fixed $m_2 \times n_2$ recourse matrix. A is an $m_1 \times n_1$ matrix defining the constraints, and c,b,q,x,y are vectors of appropriate dimension. We shall be concerned with problems of the form (1.1a-c) with *complete recourse* i.e. with constraints which satisfy

$$\text{pos } W \equiv \{t \mid t = Wy,\ y \geq 0\} = \mathbb{R}^{m_2} . \tag{1.1d}$$

Since T is fixed, we can define the (non-stochastic) *tender* $\chi = Tx$ and write (1.1a-c) in the equivalent form:

minimize $\quad cx + \Psi(\chi)$

subject to

$$\begin{aligned} Ax &= b \\ Tx - \chi &= 0 \\ x &\geq 0 \end{aligned} \tag{1.2a}$$

where

$$\Psi(\chi) = E\{\psi(\chi, h(w))\} \tag{1.2b}$$

and

$$\psi(\chi, h(w)) = \inf_{y \geq 0} \left\{ qy \mid Wy = h(w) - \chi \right\} \tag{1.2c}$$

We show first that an equivalent form to (1.2a) is

minimize $\quad cx + qy + \Psi(\chi)$

subject to

$$\begin{aligned} Ax &= b \\ Tx + Wy - \chi &= 0 \\ x, y &\geq 0 \quad . \end{aligned} \tag{1.3}$$

The family of algorithms that we are concerned with here were introduced in Nazareth and Wets, 1983, and are based upon the generalized linear programming (GLP) method of Wolfe (see Dantzig, 1963, Shapiro, 1979). They successively inner linearize $\Psi(\chi)$ in (1.3) and solve a sequence of master linear programming problems of the form

$$\text{minimize} \quad cx + qy + \sum_{k=1}^{K} \lambda_k \Psi(\chi^k)$$

subject to

$$\begin{aligned} Ax &= b \qquad (1.4) \\ Tx + Wy - \sum_{k=1}^{K} \lambda_k \chi^k &= 0 \\ \sum_{k=1}^{K} \lambda_k &= 1 \\ x, y, \lambda_k &\geq 0 \end{aligned}$$

The *tenders* $\chi^1, \ldots, \chi^K$ are assumed to have been previously generated and at the current cycle of the algorithm a new tender χ^{K+1} is introduced by solving the (Lagrangian) subproblem.

$$\underset{\chi \in X}{\text{minimize}} \quad \Psi(\chi) + \pi^K \chi \qquad (1.5)$$

where π^K are the dual multipliers associated with the constraints $Tx - \sum_{k=1}^{K} \lambda_k \chi^k = 0$ in the optimal solution of (1.4). χ^{K+1} the optimal solution* of (1.5), is an improving tender provided that $\Psi(\chi^{K+1}) + \pi^K \chi - \theta^K < 0$, where θ^K is the optimal dual multiplier associated with the constraint $\sum_{k=1}^{K} \lambda_k = 1$. When χ^{K+1} is introduced into the master problem (1.4), such a tender will lead to

---

*In practice (1.4) does not have to be pushed to optimality at each iteration, but this is a question of strategy, which we discuss later.

a reduction in the objective value (barring degeneracy, of course.) Since the projection of the set of vectors (x,y,χ) satisfying $Ax = b$, $Tx + Wy - \chi = 0$, $x,y, \geq 0$ onto the space of the χ vectors is $\mathbb{R}^{m_2}$ by (1.1d), χ can be assumed unrestricted in (1.5). However, it is often convenient to confine χ to some compact set X defined by simple bounds, for reasons of computational efficiency and to facilitate convergence arguments. Extensions to include lines of recession in (1.4) and relax the restriction (1.1d) will not be considered in this paper.

When the recourse is *simple* i.e., when $W = [I,-I]$, an approach based upon generalized linear programming has been suggested more than one in the literature, see, for example, Williams, 1966, Parikh, 1968. However, apart from special applications, see Ziemba, 1972, it has not been pursued in any real computational way. For problems with general recourse it has apparently not been tried at all. Moreover, it is important to recognize that the GLP approach should be combined with a suitable *problem transformation*, for example, the one involved in going from (1.1a-c) to (1.2a-c), in order to keep the degree of nonlinearity low. This was not fully appreciated, at least from an algorithmic point of view.

We turn now to the organization of our paper. In Section 2, we consider the alternative formulation of the equivalent deterministic form (1.2a), given by (1.3) and an interpretation of the solution of the above algorithm (1.4) and (1.5). In particular, we wish to see how *tenders* and *certainty equivalents* stand in relation to one another. Next we consider problems with simple recourse. We discuss algorithms for two cases: a) When the distribution is discrete and probabilities are known explicitely. Then $\Psi(\chi)$ is much more tractable. b) When the probability distribution is other than discrete or when it is only known implicitly through some simulation model involving the random elements w.

Case b) above is especially useful because it enables us to make the transition to general recourse, which is the topic of Section 4. Here $\Psi(\chi)$ is usually difficult to compute, since it involves minimization calculations and an integration. Our aim in this section is to discuss some possible solution strategies based upon generalized programming. Finally, Section 5 contains some concluding remarks.

Henceforth in this paper when, for example, the text includes equations (1.1a), (1.1b), (1.1c), (1.1d) and we refer to (1.1), we are making reference to all four equations.

2. EQUIVALENT FORMS AND AN INTERPRETATION OF THE SOLUTION

The notion of *certainty equivalent* of a SLP with recourse is well known, see Wets, 1974. Here we wish to investigate the tie between *tenders* and *certainty equivalents*, and with this in mind we first consider an alternative form for (1.2). This also turns out to be useful when formulating algorithms, as we shall see later in Section 3.

Suppose, just for the purpose of discussion, that $h(w)$ is replaced by some deterministic quantity, for example its expected value $\bar{h}$. Then to solve this simplified optimization problem, we need only solve a single stage program of the form:

$$\begin{aligned} &\text{minimize} \quad cx + qy \\ &\text{subject to} \\ &\qquad Ax = b \\ &\qquad Tx + Wy - \bar{h} = 0 \qquad\qquad (2.1) \\ &\qquad x, y \geq 0 \end{aligned}$$

Indeed, to test the feasibility and boundedness of the original SLP (1.1) we should solve problems of this form for suitably chosen $\bar{h}$, as shown by Wets, 1972.

Upon comparing (2.1) and (1.2), it is tempting to include the recourse matrix W *explicitly* in the first stage i.e., to consider the implications of having the recourse activities available to the first stage. This would often be the case in practice as pointed out by Williams, 1966. We would then have an equivalent deterministic problem of the form:

minimize $cx + qy + \Psi(\chi)$

subject to

$$\begin{aligned} &Ax = b \\ &Tx + Wy - \chi = 0 \\ &x,\ y \geq 0 \end{aligned} \tag{2.2}$$

with $\Psi(\chi)$ defined by (1.2b-c). We now want to show that (1.2) and (2.2) are equivalent forms.

Let us demonstrate this for the case when h(w) is discretely distributed. Suppose, therefore, that the distribution of h(w) is defined by vectors

$$h^1, h^2, \ldots, h^t \tag{2.3a}$$

with associated probabilities

$$f_1, f_2, \ldots, f_t, \quad \text{where} \quad \sum_{k=1}^{t} f_k = 1, f_k \geq 0 \tag{2.3b}$$

Then (1.2) can be expressed as follows:

minimize $cx + f_1qy^1 + f_2qy^2 + \ldots + f_tqy^t$

subject to

$$\begin{aligned} &Ax && = b \\ &Tx + Wy^1 && = h^1 \\ &Tx + \quad Wy^2 && = h^2 \\ &\vdots \qquad\qquad \ddots && \vdots \\ &Tx + \qquad\qquad Wy^t && = h^t \\ &x,\ y^j \geq 0 \end{aligned} \tag{2.4}$$

and (2.2) can be expressed as

minimize $cx + qy + f_1 qy^1 + f_2 qy^2 + \ldots + f_t qy^t$

subject to

$$\begin{array}{lllll} Ax & & & & = b \\ Tx + Wy + Wy^1 & & & & = h^1 \\ Tx + Wy + & Wy^2 & & & = h^2 \\ \vdots & & \ddots & & \vdots \\ Tx + Wy + & & & Wy^t & = h^t \end{array} \qquad (2.5)$$

$$x,\ y,\ y^j \geq 0$$

Any feasible solution of (2.4) gives a feasible solution of (2.5), simply by setting y = 0. Conversely, by writing $qy = \sum_{k=1}^{t} f_k(qy)$, and regrouping terms in (2.5) we obtain:

minimize $cx + f_1 q(y+y^1) + f_2 q(y+y^2) + \ldots + f_t q(y+y^t)$

subject to

$$\begin{array}{lllll} Ax & & & & = b \\ Tx + W(y+y^1) & & & & = h^1 \\ Tx + & W(y+y^2) & & & = h^2 \\ \vdots & & \ddots & & \vdots \\ Tx + & & & W(y+y^t) & = h^t \end{array} \qquad (2.6)$$

$$x,\ y,\ y^j \geq 0$$

and thus any feasible solution of (2.5) gives a feasible solution to (2.4), with the same objective value. The two problems must therefore be equivalent. We are led to the following theorem, a generalization of a result for *simple* recourse given in Parikh 1968.

THEOREM 2.1: The SLP problem with recourse given by (1.2) and (2.2) are equivalent, in the following sense:

$$(\bar{x},\bar{\chi}) \quad \text{solves} \quad (1.2) \Rightarrow (\bar{x},0,\bar{\chi}) \quad \text{solves} \ (2.2)$$

$$(\bar{x},\bar{y},\bar{\chi}) \quad \text{solves} \quad (2.2) \Rightarrow (\bar{x},\bar{\chi}-W\bar{y}) \quad \text{solves} \ (1.2)$$

We assume that (1.2) is solvable (bounded and solution attained); it will imply that (2.2) is solvable, and vice-versa.

PROOF[1]:

1. Suppose $\bar{x} \in R_+^{n_1}$, $\bar{y} \in R_+^{n_2}$, $\bar{\chi} \in R_+^{m_2}$ satisfy

$$T\bar{x} + W\bar{y} = \bar{\chi}$$

Let

$$\chi^0 = \bar{\chi} - W\bar{y} = T\bar{x}$$

Then for all $h(\cdot)$

$$\psi(\chi^0, h(\cdot)) \leq \psi(\bar{\chi}, h(\cdot)) + q\bar{y}$$

Proof of 1.:

We have to show that

$$\begin{aligned}
&\inf_{y\geq 0} (qy \mid Wy = h(\cdot) - \chi^0) \\
&\quad \leq q\bar{y} + \inf_{u\geq 0} (qu \mid Wu = h(\cdot) - \bar{\chi}) \\
&\quad = q\bar{y} + \inf_{u\geq 0} (qu \mid Wu = h(\cdot) - \chi^0 - W\bar{y}) \\
&\quad = \inf_{u\geq 0} (q(u+\bar{y}) \mid W(u+\bar{y}) = h(\cdot) - \chi^0)
\end{aligned}$$

---

[1] The formal proof of this proposition *for an arbitrary distribution*, which now follows, is due to Roger Wets.

$$= \inf_{y \geq \bar{y}}(qy \mid Wy = h(\cdot) - \chi^0)$$

But that is now evident since $\bar{y} \in R^{n_2}$ and thus the condition $y \geq \bar{y}$ is more constraining than $y \geq 0$ (except if $\bar{y} = 0$).□

2. Suppose $\bar{x},\bar{y},\bar{\chi},\chi^0$ are as in 1. Then

$$\Psi(\chi^0 = \bar{\chi} - W\bar{y}) \leq \Psi(\bar{\chi}) + q\bar{y}$$

Proof of 2.:

Use 1. + the fact: taking expectations is order preserving.□

3. Suppose $\bar{x},\bar{y},\bar{\chi}$ is any feasible solution of (2.2). Then

$$c\bar{x} + q\bar{y} + \Psi(\bar{\chi}) \geq c\bar{x} + q\cdot 0 + \Psi(\chi^0)$$

where

$$\chi^0 = \bar{\chi} - W\bar{y} = T\bar{x}$$

Proof of 3.:

Follows from 2.; add $c\bar{x}$ on each side.□

From 3. it follows that in order to find the infimum in (2.2), it suffices to restrict oneself to feasible solutions of (2.2) that have $y = 0$. But then (2.2) is exactly (1.2). Thus if $(\bar{x},\bar{\chi})$ solves (1.2), the triple $(\bar{x},0,\bar{\chi})$ solves (2.2). If $(\bar{x},\bar{y},\bar{\chi})$ solves (2.2) and $\bar{z} = c\bar{x} + q\bar{y} + \Psi(\bar{\chi})$ then 3. implies that

$$\bar{z} = c\bar{x} + q\cdot 0 + \Psi(\bar{\chi} + W\bar{y})$$

since the triple $(\bar{x},0,\bar{\chi} - W\bar{y})$ is also a feasible solution of (2.2). And the pair $(\bar{x},\bar{\chi} - W\bar{y})$ solves (1.2) since $(\bar{x},\bar{\chi} - W\bar{y})$ solves (2.2) when $y(=0)$ is deleted from the problem. This completes the proof of the theorem.□

In the light of the above proposition, we can deal henceforth with (2.2). Suppose we now apply the GLP algorithm outlined in Section 1 to (2.2). This will give Master LP problems of the form:

$$\text{minimize} \quad cx + qy + \sum_{k=1}^{K} \lambda_k \Psi(\chi^k)$$

subject to

$$\begin{aligned} Ax &= b \\ Tx + Wy - \sum_{k=1}^{K} \lambda_k \chi^k &= 0 \\ \sum_{k=1}^{K} \lambda_k &= 1 \\ x, y, \lambda_k &\geq 0 \end{aligned} \tag{2.7}$$

Let the optimal solution of (2.7) be x^*, y^*, λ^*, and note that no more than (m_2+1) compoments of λ^* are non-zero. Without loss of generality we can assume that these are the first (m_2+1) components $\lambda_1^*, \ldots, \lambda_{m_2+1}^*$, and we define

$$\chi^* = \sum_{k=1}^{m_2+1} \lambda_k^* \chi^k \tag{2.8}$$

χ^* is the *certainty equivalent*, since x^* and y^* are optimal for the LP problem

$$\text{minimize} \quad cx + qy$$

subject to

$$\begin{aligned} Ax &= b \\ Tx + Wy - \chi^* &= 0 \\ x, y &\geq 0 \end{aligned} \tag{2.9}$$

Indeed we can go further. Suppose that we approximate the distribution of h(w) by the following discrete distribution, whose values are

$$\chi^1, \chi^2, \ldots, \chi^{m_2+1} \tag{2.10a}$$

with associated probabilities

$$\lambda_1^*, \lambda_2^*, \ldots, \lambda_{m_2+1}^* \qquad (2.10b)$$

where the optimal solution λ^* to (2.7) can be interpreted as defining a probability distribution since

$$\sum_{k=1}^{K} \lambda_k^* = 1 \ , \qquad \lambda_k^* \geq 0 \ .$$

For the distribution (2.10), an equivalent form for (1.2) is

$$\text{minimize} \quad cx + \lambda_1^* qy^1 + \lambda_2^* qy^2 + \ldots + \lambda_{m_2+1}^* qy^{m_2+1}$$

subject to

$$\begin{array}{llllll} Ax & & & & = b & \\ Tx & + Wy^1 & & & = \chi^1 & \\ Tx & + & Wy^2 & & = \chi^2 & \qquad (2.11) \\ \vdots & & \ddots & & \vdots & \\ Tx & + & & Wy^{m_2+1} & = \chi^{m_2+1} & \\ & & & & & \\ x, y^j \geq 0 & & & & & \end{array}$$

For any $x \geq 0$ satisfying $Ax = b$, in particular for x^*, we know that (2.11) has a feasible solution for problems with relatively complete recourse. Let $y^{*1}, \ldots, y^{*m_2+1}$ be the corresponding components of the optimal solution of (2.11). The using Jensen's Inequality, namely $EF(x,\xi) \leq F(x,E\xi)$ we can deduce from the optimal solutions to (2.11) and (2.9) that

$$\sum_{k=1}^{m_2+1} \lambda^* qy^{*k} \leq qy^* \qquad (2.12)$$

Now in (2.11), multiply the row involving χ^i by λ_i^* and sum.

This leads to

$$Tx + W\left(\sum_{k=1}^{m_2+1} \lambda_k^* y^k\right) - \chi^* = 0 \tag{2.13}$$

When

$$y = \sum_{k=1}^{m_2+1} \lambda_k^* y^k ,$$

we have (x,y) feasible for (2.9), and thus any feasible solution of (2.11) leads to a feasible solution of (2.9). This fact combined with (2.12) implies that (2.9) and (2.11) are equivalent, and we have proved the following theorem which gives an interpretation of the optimal solution of (2.7):

THEOREM 2.2: Suppose that the nonzero components in the optimal solution of (2.7) are given by $\lambda_1^*,\ldots,\lambda_{m_2+1}^*$ with associated tenders $\chi^1,\ldots,\chi^{m_2+1}$, where, without loss of generality, we have assumed these to be the first (m_2+1) components. Then the problem (1.2) is equivalent to the associated discretized problem, obtained by replacing the distribution of h(w) by the distribution (2.10).

3. ALGORITHMS FOR SLP PROBLEMS WITH SIMPLE RECOURSE

3.1 *Discrete Distributions*

For simple recourse, the recourse problem (1.2b) takes the form

$$\psi(\chi,h(w)) = \inf_{y^+\geq 0,\, y^-\geq 0} \left\{ q^+y^+ + q^-y^- \;\middle|\; [I,-I]\begin{pmatrix} y^+ \\ y^- \end{pmatrix} = h(w) - \chi \right\} \tag{3.1}$$

Let $q = q^+ + q^- > 0$. Assume also that h(w) has a discrete distribution, say with the possible values

$$h_{i1}, h_{i2},\ldots,h_{in_i} \quad \text{where} \quad h_{il} < h_{i,l+1} \tag{3.2a}$$

with associated probabilities

$$f_{i1}, f_{i2},\ldots,f_{in_i} \tag{3.2b}$$

and let

$$\bar{h}_i = E\{h_i(\cdot)\}$$

Then $\Psi(\chi)$ is given by

$$\Psi(\chi) = \sum_{i=1}^{m_2} \Psi_i \; (\chi_i) \qquad (3.3a)$$

where

$$\Psi_i(\chi_i) = \max_{l=0,\ldots,n_i} (s_{il}\chi_i + e_{il}) \qquad (3.3b)$$

and with the convention $\sum_{t=1}^{0} = 0$

$$s_{il} = \left(\sum_{t=1}^{l} f_{it}\right) q_i - q_i^+, \; 0 \le l \le n_i \qquad (3.4a)$$

$$e_{il} = q_i^+\bar{h}_i - q_i\left(\sum_{t=1}^{l} p_{it}f_{it}\right), \; 0 \le l \le n_i \qquad (3.4b)$$

For a proof see Wets, 1983b. Note also that s_{il} form an increasing sequence with

$$- q_i^+ \le s_{il} \le q_i^- \quad , \quad 0 \le l \le n_i \qquad (3.5)$$

and e_{il} form a non-increasing sequence.

3.1.1 *Algorithmic Details.* Let us now look at the main ingrediants of an algorithm based upon generalized LP for solving the above problem.[(1)]

1. *Computing the Objective Functions:* $\Psi(\chi)$ is easily computed from (3.3) and (3.4). The objective function $cx + \Psi(\chi)$ and it is useful to explicitly introduce a scale factor $\rho > 0$, and define the objective to be $cx + \rho\Psi(\chi)$. This is simply a device for parameterizing the objective function of the recourse problem.

[(1)] The algorithm of this section 3.1.1 is quite similar to the one given in unpublished notes by Parikh, 1968.

2. *Initialization:* Motivated by the results of Section 2, in particular Theorem 2.1, we initially solve the problem

$$\text{minimize} \quad cx + \rho q^+y^+ + \rho q^-y^- + \rho\lambda_1\Psi(\chi^1)$$

subject to

$$\begin{aligned} Ax &= b \\ Tx + Iy^+ - Iy^- - \lambda_1\chi^1 &= 0 \\ \lambda_1 &= 1 \\ x, y^+, y^- &\geq 0 \end{aligned} \tag{3.6}$$

where

$$\chi^1 \equiv \bar{h} = E\{h(\cdot)\}$$

This is, of course, equivalent to (2.1), since $\lambda_1 \equiv 1$ and $\rho\lambda_1\Psi(\bar{h})$ is just a constant term, but we prefer (3.6) because it is of the same form as the master program below. From Wets, 1972, we see that successfully solving (3.6) immediately implies feasibility and boundedness of the original problem.

3. *Solving the Master Program:* This has the form

$$\text{minimize} \quad cx + \rho q^+y^+ + \rho q^-y^- + \sum_{k=1}^{K} \lambda_k \rho\Psi(\chi^k)$$

subject to

$$\begin{aligned} \sigma^K: &\quad Ax = b \\ \pi^K: &\quad Tx + Iy^+ - Iy^- - \sum_{k=1}^{K} \lambda_k\chi^k = 0 \\ \theta^K: &\quad \sum_{k=1}^{K} \lambda_k = 1 \\ &\quad x, y^+, y^-, \lambda_k \geq 0 \end{aligned} \tag{3.7}$$

Further initial tenders, other than $\chi^1 = \bar{h}$ could be introduced here. Let $\sigma^K, \pi^K, \theta^K$ denote the optimal multipliers of (3.7). Then the components of π^K satisfy

$$- q_i^+ \leq - \pi_i^K \leq q_i^- \qquad (3.8)$$

4. *Solution of the (Lagrangian) Subproblem:* This is given by

$$\underset{\chi \in X}{\text{minimize}} \quad \Psi(\chi) + \pi^K \chi \qquad (3.9)$$

Let us take $X = R^{m_2}$. Since $\Psi(\chi)$ is separable, we must solve the following for $i = 1,2,...,m_2$

$$\underset{\chi_i \in R^1}{\text{minimize}} \quad \Psi_i(\chi_i) + \pi_i^K \chi_i \qquad (3.10)$$

and since $\Psi_i(\chi_i)$ is given by (3.3b), we are dealing in (3.10) with the unconstrained minimization of a piecewise-linear function, and this is easily done.

The optimal solution χ_i^{K+1} satisfies

$$- \pi_i^K \in \partial \Psi_i(\chi_i^{K+1}) \qquad (3.11)$$

Now from (3.4a) we know that

$$- q_i^+ \leq \partial \Psi_i(\chi_i) \leq q_i^- \qquad (3.12)$$

for any χ_i in the support of the distribution of $h_i(\cdot)$. It follows from (3.8), (3.11) and (3.12) that χ_i^{K+1} can be found such that

$$h_{i1} \leq \chi_i^{K+1} \leq h_{ik_i} \qquad (3.13)$$

where h_{il} are defined by (3.2a).

5. *Adding and Deleting Tenders:* A tender χ^{K+1} is improving for (3.9) provided that

$$\Psi(\chi^{K+1}) + \pi^K \chi^{K+1} - \theta^K < 0 \ . \qquad (3.14)$$

If no such tender can be found, then the current solution is optimal. Note, in particular, that the subproblem does *not* have to be pushed to optimality. Furthermore, several improving tenders, each satisfying (3.14), could be deduced from one call to the subproblem.

We have not investigated in any detail the question of dropping columns corresponding to tenders from (3.7) when they become out-of-date. In implementations of the related Dantzig-Wolfe decomposition algorithm, see for example Ho, 1974, it is common to drop columns from (3.7), when they have not played a role in the optimal solution for some time and the same strategy could obviously be implemented here. The question is discussed further in Nazareth and Wets, 1983. Much of the theory on dropping cutting planes is also applicable, see, for example, Eaves and Zangwill, 1971.

3.1.2 *Experimental Implementation and Test Example:* We have implemented the above algorithm in an experimental code. Matrices are stored as 2-dimensional arrays and sparsity is not taken into account, so that it can only handle relatively small problems. The master program is solved using the Harwell LP code LA01BD and the subproblems (3.10) are solved by simply finding where $s_{il} + \pi_i^K$ changes sign from negative to positive. A single optimal tender is introduced at each iteration, and all tenders are retained in (3.7). The code was written in Fortran for the Vax 11/780 and validated using the test problems and solutions of Kallberg and Kusy, 1976 and Cleef, 1981.

For an illustrative example, consider the following product-mix problem due to Jim Ho. (Though only a small and highly simplified SLP problem, its full scale version comes from a real life application). The problem involves two products and three ingredients. The variables x_i, y_i, z_i are the amounts of ingredients 1 and 2. The demand for each product is a random variable with known probability distribution. The problem can be summarized as follows:

$$\text{minimize} \quad x_1 + 2y_1 + 3z_1 + x_2 + 2y_2 + 3z_2 + \Psi(\chi)$$

subject to

A matrix	Fat/Protein in Product 1:	$.3x_1 + .4y_1 + .2z_1$	≥ 3.3
	Fat/Protein in Product 2:	$.5y_2 + .6z_2$	≥ 4.0
	Amt. of Ingredient 1:	$x_1 + x_2$	$\leq 15.$
	Amt. of Ingredient 2:	$y_1 + y_2$	$\leq 12.$
T matrix	Amt. of Product 1:	$x_1 + y_1 + z_1 - \chi_1$	$= 0$
	Amt. of Product 2:	$+ x_2 + y_2 + z_2 - \chi_2$	$= 0$

$$x_i, y_i, z_i \geq 0 \qquad (3.15)$$

The penalties for under and over production are 2.0 and 1.0 units respectively and the probability distribution on demand h(w) is as follows:

product 1	levels	8	10	12
	probs	.25	.5	.25
product 2	levels	15	18	20
	probs	.2	.4	.4

The recourse function $\Psi(\chi)$ is defined by (3.1) where $q^+ = (2.0, 2.0)$ and $q^- = (1.0, 1.0)$.

The following table summarizes the progress of the algorithm

Iteration	First period cost cx	Total cost $cx + \Psi(\chi)$	
1	39.	46.06	
2	39.	44.75	
3	37.	43.575	
4	35.9	43.4727	
5	35.5	43.4625	optimal

Initial Solution: $x_1 = 6.$, $y_1 = 4.$, $z_1 = 0.1$, $x_2 = 9.$, $y_2 = 8.$, $z_2 = 0.$

Initial Tender: $\begin{pmatrix} 10 \\ 18.2 \end{pmatrix}$

Final Solution: $x_1 = 8.$, $y_1 = 2.25$, $z_1 = 0.$, $x_2 = 7.$, $y_2 = 8.$, $z_2 = 0$.

Final Tender: $0.875 \begin{pmatrix} 10 \\ 15 \end{pmatrix} + 0.125 \begin{pmatrix} 12 \\ 15 \end{pmatrix} = \begin{pmatrix} 10.25 \\ 15 \end{pmatrix}$

An implementation of the algorithm of Section 3.1.1 which is designed to solve reasonably large and sparse SLP problems with simple recourse is given in Nazareth and Wets, 1984. Such problems might typically arise when a given linear program is extended into the domain of SLP with simple recourse by allowing some of its right-hand-side elements to be random variables with known probability distribution; if the SLP arose in this way, the row of the original LP matrix corresponding to stochastic rhs elements would then define the T matrix. These considerations have influenced our design of standardized input formats for SLP problems with recourse, in which a "core" file defining elements of A,T,c,b, bounds and ranges on variables is specified in *standard MPS format*, and a "stochastics" file identifying which rows correspond to the T matrix, and defining distributions and recourse costs is specified in an *MPS-like format*. The implementation is based on the MINOS code of Murtagh and Saunders, 1978.

3.2 *When distribution of h(w) is other than discrete, or only known implicitly*

In Section 3.1, the discrete distribution of h(w) was known explicitly and this in turn led to the explicit form $\Psi(\chi)$ given by (3.3) and (3.4). When the distribution of h(w) is not discrete, then $\Psi(\chi)$ is not polyhedral and may be difficult to obtain explicitly. (In some cases it will still however, be possible to obtain $\Psi(\chi)$ quite accurately using numerical integration, in particular one dimensional integration routines when $\Psi(\chi)$ is separable). Even when h(w) has a discrete distribution, this may only be known implicitly, for example, through a simulation model involving the (explicitly) known distributions of the random variables w. When interrogated, this model would produce different observations of h(w) distributed according to its joint probability distribution, but the distribution itself is not explicitly available.

In this section we wish to consider modifications to the algorithm of Section 3.1.1 when the distribution function of h(w) is available in a form that provides samples and when estimates of $\Psi(\chi)$ are obtained from a finite set of such samples. The main modifications involve items 1 and 4, with items 2,3 and 5 remaining unchanged, and they are as follows:

1' *Computing* $\Psi(\chi)$: Suppose the distribution is sampled S times, giving observations $h^1, h^2, \dots, h^S$. Then a crude estimate of $\Psi(\chi)$ is

$$\Psi^E(\chi) = \frac{1}{S} \sum_{k=1}^{S} \psi^E(\chi, h^k) \tag{3.16a}$$

where

$$\psi^E(\chi, h^k) = \sum_{i:(h_i^k-\chi_i)\geq 0} q_i^+(h_i^k-\chi_i) - \sum_{i:(h_i^k-\chi_i)<0} q_i^-(h_i^k-\chi_i) \tag{3.16b}$$

Estimates of the subgradient $\pi(\chi)$ can also be obtained by

$$\pi_i^E(\chi, h^k) = \begin{cases} -q_i^+ & \text{if} \quad (h_i^k-\chi_i) \geq 0 \\ +q_i^- & \text{if} \quad (h_i^k-\chi_i) < 0 \end{cases} \tag{3.17a}$$

$$\pi_i^E(\chi) = \frac{1}{S} \sum_{k=1}^{S} \pi_i^E(\chi, h^k) \tag{3.17b}$$

4' *Solving the (Lagrangian) subproblem:* When minimizing (3.9) with $\Psi(\chi)$ being obtained by (3.16) above, we are dealing with a non-smooth unconstrained function with a fixed level of noise (for fixed sample size). In principle we would need to use methods suggested, for example, by Polyak, 1978 and others. In practice, however, it is possible to employ heuristic methods based upon techniques for smooth problems with good results, see Lemarechal, 1982.

3.2.2 *Results of some experimentation.* We modified the experimental code of section 3.2.2 along the above lines. Using a random number generator which produced pseudo/random numbers r, $0 \le r \le 1$, we simulate sampling from the discrete distribution (3.2), by generating a sample, say h^k as follows:

$$h_i^k = h_{it} \quad \text{if} \sum_{l=1}^{t+1} f_{il} > r \ge \sum_{l=1}^{t} f_{il}$$

$\Psi(\chi)$ was obtained by (3.16) with a fixed sample size S. Following Lemarechal, 1982, to solve the subproblem (3.9) we employed the VA13AD Harwell code based on the BFGS update, with *subgradient* estimates (3.17) used in place of the gradient.

Results are summarized in the following table: With sample size 300 for estimates of $\Psi(\chi)$ introduced into the master, and sample size 100 for estimates of $\Psi(\chi)$ and its subgradient used in the unconstrained minimization step, the progress of the algorithm during 8 iterations was as follows:

Iteration	First period cost cx	Total (estimated) cost $cx + \Psi(\chi)$	
1	39.	44.17	
2	38.14	44.86	
3	39.	44.46	
4	35.27	43.84	
5	37.14	43.53	
6	36.12	43.33	
7	35.76	42.93	
8	36.08	42.928	optimal

Initial Solution: $x_1 = 6.$, $y_1 = 4.$, $z_1 = 0.1$, $x_2 = 9.$, $y_2 = 8.$, $z_2 = 0.$

Initial Tender: $\begin{pmatrix} 10 \\ 18.2 \end{pmatrix}$

Final Solution: $x_1 = 7.62$, $y_1 = 2.54$, $z_1 = 0.$, $x_2 = 7.38$, $y_2 = 8.$, $z_2 = 0.$

Final Tender: $0.927 \begin{pmatrix}10.02\\15.25\end{pmatrix} + 0.073 \begin{pmatrix}11.91\\17.04\end{pmatrix} = \begin{pmatrix}10.14\\15.38\end{pmatrix}$

There are obviously many different strategies that could be used here e.g. progressively increase sample size, and refinement of the estimation of $\Psi^E(\chi)$.

4. GENERAL RECOURSE

In (1.2c), $\Psi(\chi,h(w))$ is now given by the solution of an LP problem defined by W. Since the computation of $\Psi(\chi)$ by (1.2b) involves a multidimensional integration over $\psi(\chi,h(w))$ it is, in general a function that is difficult to compute.

As in Section 3, we distinguish two cases a) when $\Psi(\chi)$ and possibly a subgradient of $\Psi(\chi)$ can be computed accurately, in particular, when the distribution of $h(w)$ is defined by a set of scenarios, each having a known probability. b) when $\Psi(\chi)$ and elements of $\partial\Psi(\chi)$ must be approximated in some way. Case b) is much more common, but it pays to dwell on case a), because it gives a lot of insight into methods of solution.

Our aim in this section is to give an overview of some approaches to solving (1.2) based upon generalized linear programming, and not to give specific algorithms.

4.1 *Scenarios with known probabilities*

Suppose $h^1,\ldots,h^t$ are a given set of scenarios with associated probabilities $f_1,\ldots,f_t$. Then as noted in Section 2, (1.1) can be put into the equivalent LP form.

$$\text{minimize} \quad cx + qy + f_1 qy^1 + \ldots + f_t qy^t$$

subject to

$$\begin{array}{llll} Ax & & = b & \\ Tx + Wy - \chi & & = 0 & \quad (4.1)\\ \qquad\qquad \chi + Wy^1 & & = h^1 & \\ \qquad\qquad \vdots & \ddots & & \\ \qquad\qquad \chi & +Wy^t & = h^t & \\ x, y, y^j \geq 0 & & & \end{array}$$

Note that even in the above LP formulation it is worthwhile to make the problem transformation involving χ, since otherwise Tx would repeat itself in every row involving h^i. (4.1) is a much more sparse representation than the equivalent LP in which χ is not present. If there are relatively few scenarios, it would be practical to solve (4.1) directly. What is to be gained by a method based on GLP *even in this context?*

In the GLP approach, solving (1.5) (and in the process computing the objective row coefficients of (1.4)) can be the most taxing part of the computation. Under our present assumptions, this subproblem, namely

$$\underset{\chi \in \mathbb{R}^{m_2}}{\text{minimize}} \quad \Phi(\chi) \equiv \Psi(\chi) + \pi^K \chi \tag{4.2}$$

can be expressed as:

$$\text{minimize} \quad \pi^K \chi + f_1 q y^1 + \ldots + f_t q y^t$$

subject to

$$\begin{array}{llll} \chi + Wy^1 & & & = h^1 \\ \vdots & \ddots & & \vdots \\ \chi & & + Wy^t & = h^t \\ y^j \geq 0 & & & \end{array} \tag{4.3}$$

Note that $\Phi(\chi)$ is polyhedral. Consider the following two ways of solving (4.2):

a) Use the revised simplex method to solve the equivalent LP problem (4.3) and take advantage of its very special structure. Note, in particular, that W occurs in each row but in different variables. This makes it likely that a feasible starting basis B can be found in variables $y^1,\ldots,y^t$ which is square-block diagonal with many sub-matrices on the diagonal repeating themselves. FTRAN and BTRAN operations can be done very efficiently with such a basis matrix, and subsequent iterations to find an optimal solution can be based on the *Schur Complement Update*, see Bisshop and Meeraus, 1977, and Gill et al., 1982, which retains the advantage of B.

b) Solve (4.2) using a minimization routine for non-smooth functions. Note, in particular, that the dimension of this problem is determined by the number of rows in the technology matrix T and this will often be small, even when the number t of realizations of the right-hand-side is large. An evaluation of $\Phi(\chi)$ and its subgradient, say at the point $\bar{\chi}$, which will normally be required at each iteration of the minimizer, involves the solution of the following *separable* problem:

$$\text{minimize } f_1 q y^1 + \ldots + f_1 q y^t$$

subject to

$$\begin{aligned} W^1 \qquad\qquad &= h^1 - \bar{\chi} \\ \ddots \qquad &\quad \vdots \\ W y^t &= h^t - \bar{\chi} \\ y^j \geq 0 \qquad\qquad & \end{aligned} \tag{4.4}$$

and various techniques that go under the heading of *bunching* and *sifting*, see Wets, 1983a, can now be profitably employed to substantially speedup the solution of (4.4). It is precisely these techniques, coupled with the use of the dual simplex method which give the L-shaped method for SLP, (see Birge, 1982), a substantial edge over straight LP applied to (4.1). The same would hold true for our method.

When t is large* we would not want to solve (1.4) unless a Schur Complement Update approach was attempted. Even then there might be difficulties, since n_1 could be large and

---

*Suppose T had 10 rows, and the components $h_i(w)$ were independently distributed, each with 3 possible levels. Then $t = 3^{10}$.

consequently many columns of $\binom{A}{T}$ could play a role in the optimal basis. In contrast, approaches based upon a) and b) above would still be viable. We have, for purposes of discussion, left χ unconstrained, and *minimized* $\Phi(\chi)$ in (4.2). In practice, there are three important points to note. First, not all elements of h(w) are necessarily stochastic. In this case the levels of the corresponding components of χ can be *fixed* in the solution of (4.3) as discussed in a) above, and in the solution of (4.4) as discussed in b). This reduces the dimensionality further. Recalling also the discussion after equation (1.5), we could restrict χ to the support of the distribution. This means we could often work with bound constrained problems of the form

minimize $\Psi(\chi) + \pi^K\chi$

subject to

$$\underline{l} \leq \chi \leq \underline{u} \qquad (4.5)$$

with $l_i = u_i$ for some components. As an extreme case suppose only one element of h(w) in the recourse problem was stochastic; then (4.5) is, in effect, a unidimensional problem. The second point to note is that (4.2) does not have to be pushed to optimality. All we really need is a solution χ^{K+1} which satisfies $\Psi(\chi^{K+1}) + \pi^K\chi^{K+1} - \theta^K < 0$ where θ^K is the optimal dual multiplier on the convexity row of the master (1.3). This can easily be incorporated into the methods discussed above for solving the subproblem. Thirdly, it is likely that a good set of *initial* tenders can be specified, and this will again considerably speed up the convergence of the algorithm.

4.2 $\Psi(\chi)$ *must be approximated*

One approach is to use sampling and couple this with use of the stochastic quasi-gradient method (see Ermoliev, 1983) to solve the subproblem. Another approach is to proceed by repeated apporximation of the distribution of h(w) and to compute bounds on $\Psi(\chi)$. Some preliminary suggestions are given in Birge, 1983. An important question is how to satisfactorily integrate the approximation strategy and the generalized programming algorithm, and the interpretation given in Theorem 2.2 may prove useful in this regard. We defer further discussion of this to a later date.

5. CONCLUSIONS

The methods introduced in this paper for solving SLP problems with recourse, involve the problem transformation (1.2), combined with the use of generalized linear programming. The problem transformation restricts the degree of nonlinearity to m_2, the number of rows of T and this, of course, enhances the efficiency of the GLP method. The problem transformation (1.2) is useful in other contexts. We have seen this already in (4.1) and the subsequent discussion. *We believe it could also be usefully employed within the L-shaped method*, see Van Slyke and Wets, 1969 and Birge, 1982, since each cut introduced would have at most m_2 elements rather than n_1, the dimension of x. For yet another example of such transformations, see Nazareth, 1983.

The approach discussed here could also be used to devise algorithms for solving a wider class of problems than (1.1). For example, cx, Ax - b = 0 and Tx could be replaced by nonlinear functions c(x), g(x) $\leq$ 0 and T(x) and a nonlinear programming method could then be used to solve the associated master. Also if T were stochastic we could apply GLP to (1.1), but now the degree of nonlinearity would be n_1. In practice only a few columns of T are normally stochastic. In this case, we could introduce a problem transformation $T_1x_1 - \chi_1 = 0$ where T_1 represents the nonstochastic columns of T and x_1, the corresponding x-variables. Then GLP could be applied to a transformed problem whose degree of nonlinearity is only (number of stochastic columns of T) + (number of rows of T). Both these extensions deserve further exploration.

ACKNOWLEDGEMENT

I would like to thank Roger J-B Wets who introduced me to stochastic linear programming with recourse, and gave me the benefit of his deep insight into the theory of this subject. This made possible my work on algorithms for SLP problems discussed here. Many thanks also to the referee for some helpful comments and to Elfriede Herbst for typing this paper.

REFERENCES

[1] Birge, J. (1982), "Decomposition and Partitioning methods for multistage stochastic linear programs," Tech. Report 82-6, Dept. of IE & OR, University of Michigan.

[2] Birge, J. (1983), "Using sequential approximations in the L-shaped and generalized programming algorithms for stochastic linear programs," Tech. Report 83-12, Dept. of IE & OR, University of Michigan.

[3] Bisschop, J. and Meeraus, A. (1977), "Matrix augmentation and partitioning in the updating of the basis inverse," *Mathematical Porgramming*, 18, 7-15.

[4] Cleef, H. (1981), "A solution procedure for the two-stage stochastic program with simple recourse," *Z. Operations Research*, 25, p. 1-13.

[5] Dantzig, G.B. (1963), *Linear Programming and Extensions*, Princeton University Press.

[6] Eaves, B.C. and Zangwill, W. (1971), "Generalized cutting plane algorithms," *SIAM J. Control*, 9, p. 529-542.

[7] Ermoliev, Yu. (1983), "Stochastic quasigradient methods and their application in systems optimization," *Stochastics*, 9, p. 1-36.

[8] Gill, P.E., Murray, W., Saunders, M.A. and Wright, M.H. (1982), "Sparse matrix methods in optimization," Tech. Report SOL-82-17, Systems Optimization Lab., Dept. of Operations Research, Stanford University.

[9] Ho, J. (1974), "Nested decomposition for large scale linear programs with the staircase structure," Report SOL 74-4. Systems Optimization Lab., Dept. of Operations Research, Stanford University.

[10] Kallberg, J. and Kusy, M. (1976), "Code Instruction for S.L.P.R., a stochastic linear program with simple recourse," Tech. Report, University of British Columbia.

[11] Lemarechal, C. (1982), "Numerical experiments in nonsmooth optimization," In *Progress in Nondifferentiable Optimization*, E.A. Nurminski (Ed.), IIASA Collaborative Proceedings Series CP-82-S8, p. 61-84.

[12] Murtagh, B. and Saunders, M. (1978), "Large-scale linearly constrained optimization," *Mathematical Programming* 14, p. 41-72.

[13] Nazareth, L. (1983), "Variants on Dantzig-Wolfe decomposition with applications to multistage problems," IIASA Working Paper, WP-83-61, Laxenburg, Austria.

[14] Nazareth, L. and Wets, R. J-B. (1983), "Algorithms for stochastic programs: the case of nonstochastic tenders," IIASA Working Paper, WP-83-5. (Revised version to appear in forthcoming *Mathematical Programming Study*).

[15] Nazareth, L. and R.J-B. Wets, (1984), "Stochastic programming with recourse: algorithms and implementation," IIASA Working Paper (forthcoming).

[16] Parikh, S.C. (1968), "Lecture notes on stochastic programming," unpublished, University of California, Berkeley.

[17] Polyak, B. (1978), "Nonlinear programming methods in the presence of noise," *Mathematical Programming*, 14, p. 87-97.

[18] Shapiro, J.F. (1979), "Mathematical Programming: Structures and Algorithms," John Wiley, New York.

[19] Van Slyke, R. and Wets, R. (1979), "L-shaped linear programs with applications to optimal control and stochastic linear programs," *SIAM J. on Appl. Math.*, 17, pp. 638-663.

[20] Wets, R. (1972), "Characterization theorems for stochastic programs," *Mathematical Programming*, 2. 166-175.

[21] Wets, R. (1974), "Stochastic programming", unpublished, Lecture notes, University of California, Berkeley.

[22] Wets, R. (1983a), "Stochastic programming: approximation schemes and solution techniques," In *Mathematical Programming 1982: The State-of-the-Art*, Springer-Verlag, Berlin.

[23] Wets, R. (1983b), "Solving stochastic programs with simple recourse," *Stochastics*, 10, p. 219-242.

[24] Williams, A.C. (1966), "Approximation formulas for stochastic linear programming," *SIAM J. Appl. Math.*, 14, No. 4, p. 668-677.

[25] Ziemba, W.T. (1972), "Solving nonlinear problems with stochastic objective functions," *Journal of Financial and Quantitative Analysis*, VII, p. 1809-1827.