Working Paper

COMBINING GENERALIZED PROGRAMMING AND SAMPLING TECHNIQUES FOR STOCHASTIC PROGRAMS WITH RECOURSE

A. Gaivoronski
J.L. Nazareth

September 1986
WP-86-44

International Institute for Applied Systems Analysis
A-2361 Laxenburg, Austria

COMBINING GENERALIZED PROGRAMMING AND SAMPLING TECHNIQUES FOR STOCHASTIC PROGRAMS WITH RECOURSE

A. Gaivoronski
J.L. Nazareth

September 1986
WP-86-44

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

FOREWORD

This paper deals with an application of generalized linear programming techniques for stochastic programming problems, particularly to stochastic programming problems with recourse. The major points which needed a clarification here were the possibility to use the estimates of the objective function instead of the exact values and to use the approximate solutions of the dual subproblem instead of the exact ones.

In this paper conditions are presented which allow to use estimates and approximate solutions and still maintain convergence. The paper is a part of the effort on the development of stochastic optimization techniques at the Adaptation and Optimization Project of the System and Decision Sciences Program.

Alexander B. Kurzhanski
Chairman
System and Decision Sciences Program

CONTENTS

1 Introduction 1

2 A conceptual Algorithm 4

3 Extension 16

References 19

COMBINING GENERALIZED PROGRAMMING AND SAMPLING TECHNIQUES FOR STOCHASTIC PROGRAMS WITH RECOURSE

A. Gaivoronski and J.L. Nazareth

1. INTRODUCTION

Generalized Programming Techniques of Wolfe (see Dantzig [1]) enjoyed early use for solving stochastic programs with simple recourse (Williams [14], Parikh [8]) and there has recently been renewed interest in their relevance for solving more general classes of stochastic programs (see Nazareth & Wets [7] – stochastic programs with recourse and nonstochastic tenders, Ermoliev, Gaivoronski & Nedeva [4] – stochastic programs with incomplete information). Our interest here is in stochastic programs with recourse of the form:

$$\text{minimize } E_w[< c(w), x > + Q(x,w)]$$

s.t.

$$Ax = b$$

$$l \leq x \leq u \tag{1.1a}$$

where

$$Q(x,w) = \min_{l \leq y \leq u} \{< q(w), y > | W(w)y = h(w) - T(w)x\} \tag{1.1b}$$

where w is an element of some probability space $(\mathbf{W}, \mathbf{B}, \mathbf{P})$, $A\,(m_1 \times n_1)$ is a fixed matrix, $T(\cdot)$ $(m_2 \times n_2)$ are random matrices, $c(\cdot)$ (n_1), $q(\cdot)$ (n_2) and $h(\cdot)$ (m_2) are random vectors and $b\,(m_1)$ a fixed vector. We assume complete recourse i.e. (1.1b) always has a solution. E_w denotes expectation. Define $c = E_w[c(w)]$. Then we can express (1.1a, b) as

$$\text{minimize } <c, x> + Q(x) = F(x)$$

s.t.

$$Ax = b \tag{1.2a}$$

$$l \leq x \leq u$$

where

$$Q(x) \triangleq E_w[Q(x,w)] \tag{1.2b}$$

The set of constraints in (1.2a) we shall denote by X. Properties of (1.2a, b) have been extensively studied (see Wets [13]) and, in particular, $Q(x)$ can be shown to be convex but is, in general, nonsmooth.

The generalized programming approach applied to (1.2a) involves *inner* or *grid linearization* of this convex program and requires coordinated solution of a *master program* and a (Lagrangian) *subproblem* defined as follows:

Master:

$$\text{minimize} \sum_{j=1}^{k} \langle c, x^j \rangle \lambda_j + \sum_{j=1}^{k} Q(x^j)\lambda_j$$

s.t.

$$\pi^k: \quad \sum_{j=1}^{k} (Ax^j)\lambda_j \quad = b \tag{1.3a}$$

$$v^k: \quad \sum_{j=1}^{k} \lambda_j \quad = 1$$

$$\lambda_j \geq 0$$

where π^k, v^k are the dual multipliers associated with the optimal solution of (1.3a)

Subproblem: Find $x^{k+1} \in R^{n_1}$ such that $l \leq x^{k+1} \leq u$ and

$$\langle \sigma^k, x^{k+1} \rangle + Q(x^{k+1}) < v^k \tag{1.3b}$$

by partially optimizing the problem

$$\underset{l \leq x \leq u}{\text{minimize}} \langle \sigma^k, x \rangle + Q(x) \tag{1.3c}$$

where π^k and v^k denote the dual multipliers associated with the optimal solution of the master program (1.3a) and

$$\sigma^k \underline{\Delta} (c - A^T \pi^k) \tag{1.4}$$

We temporarily ignore all considerations related to initialization of (1.3a), unboundedness of the solution in (1.3b), recognition of optimality and so on. (1.3a–c) show only the essential features of the method, namely, that the master sends (prices) π^k to the subproblem which, in turn, uses these quantities to identify an *improving* (grid) point x^{k+1}.

In many practical applications, the probability distribution of the random events is *discrete* with relatively few points in the distribution and randomness is often restricted to certain components of (1.1a–b), for example, to $h(\cdot)$. In such cases *judicious* computation enables $Q(x^j)$ and its subgradients to be found exactly, see Nazareth [6]. These quantities are required both to define the objective function of the master (1.3a) and during the solution of (1.3c) to give an improving point satisfying (1.3b). More generally however, $Q(x^j)$ can only be approximated in (1.2b), for example, by a sampling procedure, and exact computation of its value or of its subgradients is out of the question because, it would be too expensive. We then seek to replace $Q(x^j)$ in (1.3a) by an estimate, say Q_k^j. The generalized programming approach, extended in this manner, still continues to appear viable and deserves further investigation, for the following reasons:

a) It is well known (and in the nature of a "folk theorem") that fairly crude approximations of the underlying distribution in (1.1a–b) (which then permit exact solution of the resulting approximated recourse program) often produce quite reasonable estimates of the "optimal" first stage decision. This can be interpreted to mean that fairly crude estimates Q_k^j in the master program will often be adequate to guide the algorithm to a "reasonable" neighborhood of the desired solution of the original recourse problem (1.1a–b).

The (Lagrangian) subproblem (1.3c) does not *have to be optimized* at each cycle. For example, all that is needed in the case of exact estimates $Q(x^{k+1})$, to produce an improving point is that the condition (1.3b) be satisfied. This suggests therefore that one seek to reexpress this condition in the terms of estimates Q_{k+1}^{k+1} and combine it with utilizing stochastic estimates of subgradients, stochastic quasigradient procedures (see Ermoliev and Gaivoronski [5]) which are generally effective, when they are applied to a problem that does not have to be pushed all the way to optimality.

Our paper can be viewed as a study of generalized programming in the presence of noise (whose magnitude decreases as the number of iterations increases) and with the special characteristics of recourse problems taken into consideration. In section 2 we state a conceptual algorithm and establish convergence under appropriate assumptions thereby extending the standard proofs (see, for example, Shapiro [10], for the case when $Q(x)$ is known exactly). Some considerations concerning implementation are briefly discussed. Finally extension to other stochastic programming problems is considered in section 3.

2. A CONCEPTURAL ALGORITHM

We use the term "conceptual" here in the sense of Polak [9], and study the following algorithm for solving (1.1a–b). It will be convenient to assume that all bounds l and u are finite so the $l \leq z \leq u$ is a compact set.

The algorithm generates sequence of points $x^0 \cdots x^k \cdots$ which depend on element w of some probability space $(\mathbf{W}, \mathbf{B}, \mathbf{P})$ where $w \in \mathbf{W} \subset R^p$, $\mathbf{B}$ – σ - field, $\mathbf{P}$ – probability measure. The sequence x^k converges to the solution of the problem (1.1) in a certain probabilistic sense.

Step 1: (Initialize): Choose a set of m_1 grid points $x^1, \ldots, x^{m_1}$ so that the constraints

$$\sum_{j=1}^{m_1} (Ax^j)\lambda_j = b \tag{2.1}$$

$$\sum_{j=1}^{m_1} \lambda_j = 1$$

$$\lambda_j \leq 0$$

have a feasible solution. Set $k \rightarrow m_1$.

Step 2. (Form estimates)

Define a subset N_k of integers, $N_k \subset \{1, \ldots, k\}$, this being the set of grid points indices for which estimates will be made. Define an integer $s(k)$, which controls the precision of estimates. Generally speaking $s(k)$ is the number of observations of the function $Q(x, \omega)$ used to form the estimate. Obtain the new estimates Q_k^j of $Q(x^j)$, $j \in N_k$ and for $j \notin N_k$ take $Q_k^j = Q_{k-1}^j$. It will be assumed that for $j \in N_k$

$$Q_k^j - Q(x^j) = \rightarrow 0 \text{ as } s(k) \rightarrow \infty$$

in some suitable probabilistic sense. Initially for $k = m_1$, let $N_k = \{1, \ldots, m_1\}$. For subsequent k, the set N_k, integer $s(k)$ and estimates Q_k^j can be selected in a number of different ways, some of which will be specified later.

Step 3: (Solve Master):

$$\text{minimize} \sum_{j=1}^{k} (<c, x^j> + Q_k^j)\lambda_j \tag{2.2}$$

s.t.

$$\pi^k: \quad \sum_{j=1}^{k} (Ax^j)\lambda_j \quad = b$$

$$v^k: \quad \sum_{j=1}^{k} \lambda_j \quad = 1$$

$$\lambda_j \geq 0$$

Let π^k and v^k be the associated optimal dual multipliers and λ_j^k – optimal primal variables. Define $\Lambda^k = \{j : \lambda_j^k > 0\}$. In some versions of our method it is necessary at this point to redefine the set N_k and go to step 2 (examples will be given later). Otherwise, go to step 4.

Step 4: (Define new grid point x^{k+1}).

Define

$$\sigma^k \triangleq (c - A^T \pi^k)$$

and consider the (Lagrangian) subproblem

$$\underset{l \leq x \leq u}{\text{minimize}} <\sigma^k, x> + Q(x) \tag{2.3}$$

The new point x^{k+1} is taken to be an "approximate" solution to this problem, more precisely, it is necessary that for almost all $w \in \mathbf{W}$ there exists a subsequence $k_{r(w)}$ such that

$$<\sigma^{k_r}, x^{k_r+1}> + Q(x^{k_r+1}) - \min_{l \leq x \leq u} [<\sigma^{k_r}, x> + Q(x)] \rightarrow 0$$

as $r \rightarrow \infty$.

Note, that it is *not* necessary that

$$<\sigma^k, x^{k+1}> + Q(x^{k+1}) - \min_{l \le x \le u} [<\sigma^k, x> + Q(x)] \to 0$$

for the whole sequence x^k. This makes it possible, for instance, to use random search techniques for getting x^{k+1}. Some particular methods of choosing the point x^{k+1} with this property will be specified at the end of this section.

Step 5: (Iterate): $k \to k + 1$. Go to step 2.

This algorithm has two important differences from the usual generalized linear programming algorithm. Firstly, it does not require exact values of the objective function (step 2). It is only necessary to have estimates of the objective values at the grid points whose precision gradually increases. Secondly is is not necessary to minimize the Lagrangian subproblem at step 4, precisely; it is only necessary that the current point x^{k+1} regularly comes to the vicinity of such a solution.

Both modifications are necessary in order to make use of generalized linear programming in a stochastic setting.

In order to prove convergence of this algorithm let us consider it's dual reformulation. Take

$$\varphi(x, \pi) = <c, x> + Q(x) - <\pi, Ax - b> \tag{2.4}$$

$$\varphi^k(j, \pi) = <c, x^j> + Q_k^j - <\pi, Ax^j - b> \tag{2.5}$$

$$\psi(\pi) = \min_{l \le x \le u} \varphi(x, \pi) \tag{2.6}$$

$$\psi^k(\pi) = \min_{1 \le j \le k} \varphi^k(j, \pi) \tag{2.7}$$

Then algorithm (2.1)–(2.3) can be considered as a maximization method for the concave function $\psi(\pi)$ by successive polyhedral approximation of $\psi(\pi)$ by $\psi^k(\pi)$. At step 1 the initial polyhedral approximation is constructed, in step 3 the current polyhedral approximation $\psi^k(\pi)$ is maximized, optimal dual multipliers π^k being the solution of the problem

$$\max \psi^k(\pi)$$

In steps 2 and 4 the polyhedral approximation is updated.

Theorem 1. Make the following assumptions:

1. Initial points $x^1, \ldots, x^{m_1}$ are such that

$$b \in \text{int co}\{Ax^j,\ j = 1{:}\ m_1\}$$

where int means interior and co convex hull.

2.

$$\max\left\{|Q_k^k - Q(x^k)|,\ \max_{j \in \Lambda^k} |Q_k^j - Q(x^j)|\right\} = \varepsilon_k \to 0 \quad a.s.$$

3

$$\lim_{r \to \infty} \inf_{i \le r} \left\{ <\sigma^i, x^{i+1}> + Q(x^{i+1}) \right.$$

$$\left. - \min_{l \le x \le u} [<\sigma^i, x> + Q(x)] \right\} = 0 \quad a.s.$$

Then $F(\bar{x}^k) \to \min_{x \in X} F(x)$ a.s. where $\bar{x}^k = \sum_{j \in \Lambda^k} \lambda_j^k x^j$ and all accumulation points of the sequence $\bar{x}^k$ are solutions of (1.2) a.s.

Proof. Due to the assumption 2 we have

$$\sup_{k,j \in \Lambda^k} |Q_k^j - Q(x^j)| < C < \infty \quad a.s.$$

This together with boundedness of x^k gives:

$$|<c, x^j> + Q_k^j| < C_1 < \infty$$

This together with assumption 1 implies the boundedness of the sequence π^k, which can be seen as follows: Indeed, $\psi^k(\pi^k) = \max_{\pi} \psi^k(\pi)$ and therefore

$$\psi^k(\pi^k) \ge \psi^k(0) \ge -C_1$$

which follows from (2.5). On the other hand

$$\psi^k(\pi^k) \le \min_{1 \le j \le m_1} [<c, x^j> + Q_k^j - <\pi^k, Ax^j - b>]$$

$$\le \max_{1 \le j \le m_1} [<c, x^j> + Q_{s(k)}^j] + \min_{1 \le j \le m_1} [-<\pi^k, Ax^j - b>]$$

$$\le C_1 - \max_{1 \le j \le m_1} <\pi^k, Ax^j - b>$$

$$= C_1 - \|\pi^k\| \max_{1 \le j \le m_1} \left[\frac{\pi^k}{\|\pi^k\|}, Ax^j - b\right]$$

$$\leq C_1 - \|\pi^k\| \min_{|e|=1} \max_{1 \leq j \leq m_1} \langle e, Ax^j - b \rangle \leq C_1 - \|\pi^k\| \delta$$

for some $\delta > 0$ due to assumption 1.

Thus,

$$C_1 - \|\pi^k\| \delta \geq - C_1$$

which gives

$$\|\pi^k\| \leq \frac{2C_1}{\delta} \ .$$

Therefore the sequence π^k is bounded. According to the assumption 3 of the theorem for almost all $w \in \mathbb{W}$ exist subsequence $k_r(w)$ such that

$$\langle \sigma^{k_r}, x^{k_r+1} \rangle + Q(x^{k_r+1}) - \min_{l \leq x \leq u} [\langle \sigma^{k_r}, x \rangle + Q(x)] \to 0$$

Using equality $\sigma^{k_r} = c - A^T \pi^{k_r}$ we obtain

$$\langle c, x^{k_r+1} \rangle + Q(x^{k_r+1}) - \langle \pi^{k_r}, Ax^{k_r+1} - b \rangle - \psi(\pi^{k_r}) \to 0 \qquad (2.8)$$

Due to the boundedness of the sequence π^k we may now assume without loss of generality that $\pi^{k_r} \to \pi^*$ and therefore $\|\pi^{k_r} - \pi^{k_r+1}\| \to 0$. Further more from the definitions (2.4), (2.6) of the function $\psi(\pi)$ and boundedness of the admissible set X follows that the function $\psi(\pi)$ satisfies Lipshitz condition uniformly on π and therefore

$$|\psi(\pi^{k_r}) - \psi(\pi^{k_r+1})| \leq C_2 \|\pi^{k_r} - \pi^{k_r+1}\| \to 0$$

as $r \to 0$ where $C_2 < \infty$.

Thus (2.8) implies

$$\langle c, x^{k_r+1} \rangle + Q(x^{k_r+1}) - \langle \pi^{k_r+1}, Ax^{k_r+1} - b \rangle - \psi(\pi^{k_r+1}) \leq \tau_r$$

where $\max \{0, \tau_r\} \to 0$ as $r \to \infty$.

Consequently

$$\langle c, x^{k_r+1} \rangle + Q^{k_r+1}_{k_{r+1}} - \langle \pi^{k_r+1}, Ax^{k_r+1} - b \rangle - \psi(\pi^{k_r+1})$$

$$\leq \tau_r + Q^{k_r+1}_{k_{r+1}} - Q(x^{k_r+1}) \qquad (2.9)$$

But

$$<c, x^{k_r+1}> + Q_{k_{r+1}}^{k_r+1} - <\pi^{k_{r+1}}, Ax^{k_r+1} - b>$$

$$\geq \min_{1 \leq j \leq k_{r+1}} [<c, x^j> + Q_{k_{r+1}}^j - <\pi^{k_{r+1}}, Ax^j - b>] = \psi^{k_{r+1}}(\pi^{k_{r+1}}) \tag{2.10}$$

Inequalities (2.9) and (2.10) give

$$\psi^{k_{r+1}}(\pi^{k_{r+1}}) - \psi(\pi^{k_{r+1}}) \leq \tau_r + Q_{k_{r+1}}^{k_r+1} - Q(x^{k_r+1})$$

where together with assumption 2 mean

$$\psi^{k_r}(\pi^{k_r}) - \psi(\pi^{k_r}) \leq \tau_r^1 \tag{2.11}$$

where $\max\{0, \tau_r^1\} \to 0$ as $r \to \infty$ *a.s.*.

On the other hand

$$\psi^{k_r}(\pi^{k_r}) = \max_{\pi} \psi^{k_r}(\pi)$$

$$\geq \max_{\pi} \min_{j \in \Lambda^{k_r}} [<c, x^j> + Q(x^j) - <\pi, Ax^j - b>] - \varepsilon_{k_r}$$

$$\geq \max_{\pi} \psi(\pi) - \varepsilon_{k_r} \tag{2.12}$$

Inequality (2.11) now gives

$$\psi(\pi^{k_r}) + \tau_r^1 \geq \max_{\pi} \psi(\pi) - \varepsilon_{k_r}$$

which implies

$$\psi(\pi^{k_r}) - \max_{\pi} \psi(\pi) \to 0$$

and

$$\psi^{k_r}(\pi^{k_r}) \leq \max_{\pi} \psi(\pi) + \tau_r^2$$

where $\max\{0, \tau_r^2\} \to 0$ as $r \to \infty$ *a.s.*. The last inequality together with (2.12) gives

$$\psi^{k_r}(\pi^{k_r}) - \max_{\pi} \psi(\pi) \to 0 \tag{2.13}$$

Taking now arbitrary $k > k_r$ we get:

$$\psi^k(\pi^k) - \psi^{k_r}(\pi^{k_r})$$

$$\leq \max_{\pi} \min_{j \in \Lambda^k \cup \Lambda^{k_r}} [<c, x^j> + Q(x^j) - <\pi, Ax^j - b>]$$

$$- \max_{\pi} \min_{j \in \Lambda^{k_r}} [<c, x^j> + Q(x^j) - <\pi, Ax^j - b>]$$

$$+ 2 \max_{k_r \leq i \leq k} \varepsilon_i \leq 2 \max_{k_r \leq i \leq k} \varepsilon_i$$

which together with (2.12) and (2.13) gives

$$\psi^k(\pi^k) \longrightarrow \max_{\pi} \psi(\pi) \quad a.s.$$

The problem of maximization of $\psi^k(\pi)$ is dual to (2.2) and therefore

$$\sum_{j \in \Lambda^k} (<c, x^j> + Q_k^j) \lambda_j^k \longrightarrow \min_{x \in X} F(x) \tag{2.14}$$

Finally due to convexity of $F(x)$

$$F(\bar{x}^k) \leq \sum_{j \in \Lambda^k} F(x^j) \lambda_j^k$$

$$\leq \sum_{j \in \Lambda^k} (<c, x^j> + Q_k^j) + \varepsilon_k$$

which together with (2.14) gives

$$F(\bar{x}^k) \longrightarrow \min_{x \in X} F(x) \quad a.s.$$

which completes the proof.

We now study in turn each of the assumptions upon which the preceding theorem depends. Assumption 1 of the theorem can always be satisfied if matrix A is of rank m_1.

Let us consider in more detail assumption 2, which deals with precision of function values estimates at "essential" points. It's fulfillment depends on the rule used at step 2 to determine the set N_k of current new estimates, the integer $s(k)$ which controls accuracy and the method of obtaining estimates. Consider two such rules which guarantee that condition 2 is satisfied.

1. This is the simplest *ad hoc* rule. Before starting the algorithm define a sequence $\{k_p\}_{p=1}^{\infty}$, $k_{p+1} > k_p$ and take $s(m_1) = s_0$

$$N_k = \{1, \ldots, k\},\ s(k) = s(k-1) + 1$$

if $k = k_p$

$$N_k = \{k\},\ s(k) = s(k-1) \text{ otherwise}$$

in other words for $k = k_p$ estimates at all grid points are updated with increased accuracy while for $k \neq k_p$ the estimate is made only at the latest point x^k to enter the set of grid points. The estimates themselves should possess only the property that

$$|Q(x^j) - Q_k^j| = \varepsilon_k \to 0 \quad a.s.$$

$$\text{as } s(k) \to \infty$$

An example of such estimate is

$$Q_k^j = \frac{1}{s_0} \sum_{i=1}^{s_0} Q(x^j, w^i) \text{ if } k = m_1 \tag{2.15}$$

$$Q_k^j = \left[1 - \frac{1}{s(k)}\right] Q_{k-1}^j + \frac{1}{s(k)} Q(x^j, w^{s(k)}) \quad \text{for } j < k$$

$$Q_k^k = \frac{1}{s(k)} \sum_{k=1}^{s(k)} Q(x^k, w^i) \text{ if } k = k_p$$

where w^i are independent observations of random parameters from (1.1)

2. The previous rule does not discriminate between recent points and old ones, which might become redundant. Furthermore it is better to base decisions on whether to increase precision on information which becomes available during iterations. The following *adaptive precision rule* takes account of these factors.

Let us define for each estimate Q_k^j of the function value $Q(x^j)$, the number k_j such that

$$x^j \in N_{k_j},\ x^j \bar{\in} N_i \text{ for } k_j < i \leq k$$

i.e. k_j is step number when the estimate of Q_k^j was last updated. Then the precision of the estimate Q_k^j characterized by number $s(k_j)$:

$$|Q_k^j - Q(x^j)| = \eta_{s(k_j)} \to 0$$

$$\text{as } s(k_j) \longrightarrow \infty$$

The steps 2 and 3 of the method with this adaptive precision rule are specified as follows:

Step 2 (Form estimates). There are two possibilities

(i) Preceding step was step 3. Then

$$N^k = \{j : j \in \Lambda_k, \quad \text{and} \quad s(k_j) < s(k)\}$$

$s(k)$ remains the same. For $j \in N^k$ get estimates Q_k^j such that

$$Q_k^j - Q(x^j) = \eta_{s(k)} \longrightarrow 0 \quad a.s. \tag{2.16}$$

$$\text{as } s(k) \longrightarrow \infty$$

go to step 3

(ii) Preceding step was step 5. Take $s(k) = s(k-1)$ and get estimate Q_k^k with the property (2.16). Put $Q_k^j = Q_{k-1}^j$, $j < k$. If

$$\sigma^{k-1} x^k + Q_k^k \geq v^{k-1} \tag{2.17}$$

then take $s(k) = s(k) + 1$

$$N_k = \Lambda^{k-1} \cup \{k\}$$

and update estimates for $j \in N_k$ such that (2.16) is satisfied. If (2.17) is not satisfied don't do any additional estimation and go to step 3.

Step 3 (Solve Master). Solve (2.2) and take $\Lambda_k = \{j : \lambda_j^k > 0\}$ where λ_j^k – solutions of (2.2). If $s(k_j) = s(k)$ for all $j \in \Lambda_k$ then take $\Lambda^k = \Lambda_k$ and go to step 4 otherwise go to step 2.

Thus, in this modification it is always assured that through repetition of steps 2 and 3 that we get such set Λ^k that for all $j \in \Lambda^k$ precision of estimates Q_k^j corresponds to number $s(k)$. In this case besides property (2.16) some mild "independence" conditions should be satisfied. Let us define by $\mathbf{B}_k$ σ-field generated by $\{x^1, \ldots, x^k, Q_k^1, \ldots, Q_k^k\}$ at the moment when $k_j = k$ for all $j \in \Lambda^k$. It is necessary that exists $\sigma > 0$ and for any $s(k)$ exists $\beta_{s(k)} > 0$ such that

$$\mathbf{P}(Q_k^k - Q(x^k) > \beta_{s(k)} | \mathbf{B}_k) > \sigma \tag{2.18}$$

$$\mathbf{P}(Q_k^k - Q(x^k) < -\beta_{s(k)} | \mathbf{B}_k) > \sigma$$

These conditions are satisfied, for instance for the estimates of the type (2.15):

$$Q_k^j = \frac{s(k_j)}{s(k)} Q_k^j + \frac{1}{s(k)} \sum_{t=s(k_j)+1}^{s(k)} Q(x^j, w^t) \tag{2.19}$$

This formula is also valid for the first estimate at the point x^k if we take in this case $s(k_k) = 0$. It is assumed that values w^t of the random parameters are independent. Estimates (2.19) satisfy property (2.18) except in the trivial case $Q(x^j) \equiv Q(x^j, w)$, for almost all w.

Theorem 2. Suppose that conditions 1 and 3 of theorem 1 are satisfied and, in addition, (2.16), (2.18) are fulfilled and π^k is bounded a.s. Then (2.17) is satisfied infinitely often with probability 1 and, consequently, for precision control rule, based on (2.17) assumption 2 of the theorem 1 is satisfied.

Proof. Suppose that exists set $\mathbf{W}_1 \subset \mathbf{W}$ such that for $w \in \mathbf{W}_1$ condition (2.17) is satisfied only on finite number of iterations. This means that for any $\omega \in \mathbf{W}_1$ there exists $k_1(\omega)$ such that for $k > k_1(\omega)$ we have $s(k) = s(\omega) \equiv \text{const}$. Therefore any number l can enter the set N_k only once for $k > k_1(\omega)$. Therefore for $\omega \in \mathbf{W}_1$ transition from the step 3 to step 2 can occur only finite number of times. Thus, for almost all $\omega \in \mathbf{W}_1$ exists $k_2(\omega) \geq k_1(\omega)$ such that for $k > k_2(\omega)$ there are no transitions from step 3 to step 2, i.e., only new estimates Q_k^k will be made for $k > k_2(\omega)$. Therefore for $k > k_2(\omega)$ we have

$$\psi^k(\pi) \geq \psi^{k+1}(\pi)$$

where $\psi^k(\pi)$ is defined in (2.7) and

$$\psi^k(\pi^k) \geq \psi^{k+1}(\pi^{k+1})$$

According to the assumption 3 of the theorem for almost all $w \in \mathbf{W}_1$ exists sequence $k_\tau(w)$ such that

$$\langle c, x^{k_\tau+1} \rangle + Q(x^{k_\tau+1}) - \langle \pi^{k_\tau}, Ax^{k_\tau+1} - b \rangle - \psi(\pi^{k_\tau}) = \gamma_\tau \to 0$$

Due to boundedness of the sequence π^k we can assume without loss of generality that $\pi^k \to \pi^*$. Taking into account the fact that $\psi(\pi)$ and $\psi^k(\pi)$ satisfy the Lipshitz condition uniformly over π and k we obtain for $\omega \in \mathbf{W}_1$ and $k_\tau > k_2(\omega)$:

$$\psi^{k_\tau+1}(\pi^*) \leq \psi^{k_\tau+1}(\pi^*)$$

$$\leq \langle c, x^{k_\tau+1} \rangle + Q_{k_\tau+1}^{k_\tau+1} - \langle \pi^*, Ax^{k_\tau+1} - b \rangle$$

$$\leq \langle c, x^{k_r+1}\rangle + Q(x^{k_r+1}) - \langle \pi^{k_r}, Ax^{k_r+1} - b\rangle$$

$$+ Q_{k_r+1}^{k_r+1} - Q(x^{k_r+1}) + C_2\|\pi^* - \pi^{k_r}\|$$

$$= \psi(\pi^{k_r}) + \gamma_r + Q_{k_r+1}^{k_r+1} - Q(x^{k_r+1}) + C_2\|\pi^* - \pi^{k_r}\|$$

$$\leq \psi(\pi^*) + \gamma_r + Q_{k_r+1}^{k_r+1} - Q(x^{k_r+1}) + \tilde{\gamma}_r \qquad (2.20)$$

where $\tilde{\gamma}_r = 2C_2\|\pi^* - \pi^{k_r}\| \longrightarrow 0$ as $r \longrightarrow \infty$.

Condition (2.18) gives for $k_r > k_2(\omega)$

$$\mathbf{P}(Q_{k_r+1}^{k_r+1} - Q(x^{k_r+1}) < -\bar{\beta} \mid \mathbf{B}_{k_r+1}) > \sigma$$

for some $\sigma > 0$ and $\bar{\beta} = \beta_{s(\omega)} > 0$. Therefore for almost all $\omega \in \mathbf{W}_1$ exist $k_r > k_2(\omega)$ such that

$$Q_{k_r+1}^{k_r+1} - Q(x^{k_r+1}) < -\bar{\beta}$$

and $\gamma_r + \tilde{\gamma}_r < \bar{\beta}/2$. This together with (2.20) gives for sufficiently large r:

$$\psi^{k_r}(\pi^*) \leq \psi(\pi^*) - \bar{\beta}/2$$

and therefore $\psi^{k_r}(\pi^{k_r}) \leq \psi(\pi^{k_r})$ for sufficiently large r and $w \in \mathbf{W}_1$. Hence

$$\sigma^{k_r} x^{k_r+1} + Q_{k_r+1}^{k_r+1} = \psi(\pi^{k_r}) + \langle \pi^{k_r}, b\rangle$$

$$+ Q_{k_r+1}^{k_r+1} - Q(x^{k_r+1}) + \gamma_r \geq \psi^{k_r}(\pi^{k_r}) + \langle \pi^{k_r}, b\rangle$$

$$+ Q_{k_r+1}^{k_r+1} - Q(x^{k_r+1}) + \gamma_r = v^{k_r} + Q_{k_r+1}^{k_r+1} - Q(x^{k_r+1}) + \gamma_r \qquad (2.21)$$

The condition (2.18) implies

$$\mathbf{P}(Q_{k_r+1}^{k_r+1} - Q(x^{k_r+1}) > \bar{\beta}/\mathbf{B}_{k_r+1}) > \sigma$$

with $\sigma > 0$, $\bar{\beta} > 0$, $k_r > k_2(\omega)$. Therefore for almost all $\omega \in \mathbf{W}_1$ exist $k_r > k_2(\omega)$ and $|\gamma_r| < \bar{\beta}/2$. This gives together with (2.21):

$$\sigma^{k_r} x^{k_r+1} + Q_{k_r+1}^{k_r+1} \geq v^{k_r} + \bar{\beta}/2$$

for almost all $\omega \in \mathbf{W}_1$ and some $k_r > k_2(\omega)$. We arrived in contradiction with our ini-

tial assumption. Therefore assumption 2 of the theorem 1 is satisfied. Proof is completed.

Let us now consider in more detail Assumption 3 of the Theorem 1 and the specific procedures for selection of the point x^{k+1} at step 4 of the algorithm. These procedures should satisfy assumption 3 of the theorem; namely with probability 1 exists a subsequence k_r such that

$$\varphi(x^{k_r+1}, \pi^{k_r}) - \min_{l \le x \le u} \varphi(x, \pi^{k_r}) \to 0. \tag{2.22}$$

The best choice is $\varphi(x^{k+1}, \pi^k) = \min_{l \le x \le u} \varphi(x, \pi^k)$ but this is not feasible because of inaccessibility of exact function values $\varphi(x, \pi)$. We shall consider two procedures which do not require objective function values.

1 *Random search.* Take probability measure R with nonzero density in the set $l \le x \le u$ and take successive points $x^1 \cdots x^k$ as independent observations of random variable x with distribution R. Then (2.22) is fulfilled due to continuity of $\varphi(x, \pi)$.

2 *Stochastic quasi-gradient method.* (Ermoliev [3]) This method will produce sequence of points x^k such that

$$\varphi(x^k, \pi^k) - \min_{l \le x \le u} \varphi(x, \pi^k) \to 0 \ . \tag{2.23}$$

On each iteration the following calculation, are performed at the step 4 of the algorithm:

$$x_k^{s+1} = P_X(x_k^s - \rho_s \xi_k^s) \tag{2.24}$$

$$s = 0, \ldots, m_k - 1, x_k^0 = x^k, x^{k+1} = x_k^{m_k}$$

$$E(\xi_k^s / x_k^0, \ldots, x_k^s) = c + Q_x(x_k^s)$$

$$P_X(z)_i = \begin{cases} l_i & \text{if} \quad z_i < l_i \\ u_i & \text{if} \quad z_i > u_i \\ z_i & \text{otherwise} \end{cases}$$

In particular, it is possible to take

$$\xi_k^s = c - T^T(w^s)d^s$$

where d^s are optimal dual multipliers of the following problem:

$$Q(x_k^s, u^s) = \min_{l \le y \le u} [<q(w^s), y> \mid W(\omega^s)y = h(\omega^s) - T(\omega^s)x_k^s] \qquad (2.25)$$

and ω^s are independent observations of random parameters.

If problem (2.25) has bounded solutions for all w, $\sum_{s=0}^{\infty} \rho_s = \infty$, $\sum_{s=0}^{\infty} \rho_s^2 < \infty$ and $m_k \to \infty$ as $k \to \infty$ then (2.23) is satisfied and, consequently, assumption 3 of the theorem 1 is satisfied too.

3. EXTENSION

Method, described in this section is applicable not only to the stochastic programs with recourse (1.1) but to more general problems of stochastic programming as well. Consider the following problem:

$$\text{mimimize } Ef(x, w) \qquad (2.26)$$

$$\text{subject to } p(x) \le 0,\ x \in X$$

$$p(x) = (p_1(x), \ldots, p_{m_1}(x))$$

The method and results remain essentially the same if we denote $Ef(x, w) = Q(x)$ and substitute everywhere in the above discussion $Q(x)$ for $<c, x> + Q(x)$ and $p(x)$ for $Ax - b$. The initial points should satisfy now

$$\sum_{j=1}^{m_1} \lambda_j p(x_j) \le 0$$

$$\sum_{j=1}^{m_1} \lambda_j = 1$$

$$\lambda_j \ge 0$$

Master problem (2.2) obtains the form

$$\text{minimize } \sum_{j=1}^{k} Q_k^j \lambda_j$$

s.t.

$$\pi^k: \ \sum \lambda_j p(x_j) \le 0$$

$$v^k: \sum_{j=1}^{k} \lambda_j = 1$$

$$\lambda_j \geq 0$$

where Q_k^j are estimates of $Ef(x^j, w)$. Subproblem (2.3) becomes

$$\min_{x \in X} Ef(x, w) - \langle \pi^k, p(x) \rangle$$

The theorem 1a is proved similarly to the theorem 1:

Theorem 1a. Take the following assumptions

1 Function $Ef(x, w)$, $p(x)$ are convex, the set X is compact.

2 Exists $\tilde{x} \in X$ such that $p(\tilde{x}) < 0$ and initial points $x^1, \ldots, x^{m_1}$ are such that

$$\max_{\substack{|e|=1, \\ e_i \geq 0}} \min_j \langle e, p(x^j) \rangle < 0$$

3

$$\max \{|Q_k^k - Q(x^k)|, \max_{j \in \Lambda^k} |Q_k^j - Ef(x^j, w)|\} = \varepsilon_k \to 0 \quad a.s.$$

4

$$\lim_{r \to \infty} \inf_{i \leq r} \{Ef(x^{i+1}, w) - \langle \pi^i, p(x^{i+1}) \rangle\}$$

$$- \min_{x \in X} [Ef(x, w) - \langle \pi^i, p(x) \rangle] = 0 \quad a.s.$$

Then $Ef(\bar{x}^k, w) \to \min\{Ef(x, w) | p(x) \leq 0, x \in X\}$ where $\bar{x}^k = \sum_{j \in \Lambda^k} \lambda_j^k x^j$ and all accumulation points of the sequence $\bar{x}^k$ are solutions of the problem (2.26).

Although our primary concern here is with a conceptual algorithm, let us conclude this section with a brief discussion of some considerations which apply in order to make the algorithm implementable.

a) Purging Strategy for Grid Points: The above algorithm assumes that all grid points are retained but, when storage is limited, it will be necessary to periodically remove grid points. This subject has been extensively studied, see Eaves and Zangwill [2], Topkis [11] in the context of cutting plane algorithms, and similar considerations apply here.

b) Variance of Estimates: When developing estimates Q_k^j using, for example, (2.15) or (2.19), we can also maintain and update the variance of estimates for each grid point x^j. These can than be usefully employed in refining the decision rules at Steps 2 and 3.

c) Induced Constraints: When the assumption of complete recourse (i.e. that (1.1b) always has a solution) cannot be verified *a priori*, then it may happen that for some combination of grid point x^j and random parameters w^i (in (2.15) and (2.19)), the problem (1.1b) is infeasible. Following Van-Slyke and Wets [12], an induced constraint or feasibility cut must then be deduced and introduced into the problem (1.1a) and correspondingly into the master program (2.2). This extension requires further study.

There are also a number of special cases of the general problem (1.1a, b) which permits refinements, with a view to enhancing efficiency, of the algorithm described above. One case of practical interest is stochastic programs with recourse and non-stochastic tenders (see Nazareth and Wets [7], where $T(m_1 \times n_1)$ is a *fixed* matrix. The master/subproblem pair corresponding to (2.2) and (2.3) can then be reformulated as follows:

Master:

$$\text{minimize } \langle c, x \rangle + \sum_{j=1}^{k} Q_k^j \lambda_j$$

s.t.

$$\sigma^k: Ax = b$$

$$\pi^k: Tx - \sum_{j=1}^{k} \chi^j \lambda_j = 0 \qquad (2.27)$$

$$v^k: \sum_{j=1}^{k} \lambda_j = 1$$

$$l \leq x \leq u, \ \lambda_j \geq 0$$

where

$$Q(\chi, w) = \min_{l \leq y \leq u} \{\langle q(w), y \rangle | W(w)y = h(w) - \chi\} \qquad (2.28)$$

and

$$Q(\chi) = E_w[Q(\chi, w)] \ .$$

Q_k^j is an estimate of the value $Q(\chi^j)$ at the grid point χ^j, and σ^k, π^k and v^k are the dual multipliers associated with the optimal solution of the master (2.27).

Subproblem: Consider the (Lagrangian) subproblem,

$$\underset{L \le \chi \le U}{\text{minimize}} \; \langle \varphi^k, \chi \rangle + Q(\chi) \tag{2.29}$$

where L and U are any suitable bounds implied by $\chi \equiv Tx$ and $l \le x \le u$. χ^{k+1} is again taken to be an "approximate" solution to (2.29), in the sense discussed in Step 4, after expression (2.3).

It frequently happens that $m_1 \ll n_1$ i.e. that only a few elements of the problem are stochastic. In this case, the above reformulation can considerably enhance efficiency, because the optimization in the subproblem (2.29) and the linear program in (2.28) which must be solved to obtain estimates Q_k^j are both in a space of relatively low dimension.

REFERENCES

[1] Dantzig, G.B., *Linear Programming and Extensions*, Princeton University Press (1980).

[2] Eaves, B.C. and W. Zangwill, "Generalized cutting plane algorithms," *SIAM J. Control* 9(1971).

[3] Ermoliev, Yu., *Methods of Stochastic Programming* (in Russian), Nauka, Moscow (1976).

[4] Ermoliev, Yu., A. Gaivoronski and C. Nedeva, Stochastic programming problems with incomplete information on objective functions, *SIAM J. Control and Optimization* vol. 23, No. 5.

[5] Ermoliev, Yu. and A. Gaivoronski, Stochastic quasigradient methods and their implementation. IIASA Working Paper WP-84-55, 1984.

[6] Nazareth, J.L., Design and implementation of a stochastic programming optimizer with recourse and tenders. IIASA Working Paper WP-85-063, 1985.

[7] Nazareth, J.L. and R.J-B. Wets, "Algorithms for stochastic programs: the case of non-stochastic tenders," A. Prékopa and R.J-B. Wets, eds., *Mathematical Programming Study* 28(1986).

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

[9] Polak, E., Theory of optimal control and mathematical programming. N.Y. McGraw-Hill Book Co. 1970.

[10] Shapiro, J.F., "Mathematical programming: structures and algorithms", John Wiley, New York (1979).

[11] Topkis, D., "Cutting plane methods without nested constraint sets", *Operations Research* 18(1970).

[12] Van Slyke, R. and R.J-B. Wets, "L-shaped linear program with applications to optimal control and stochastic linear programs", *SIAM Journal on Applied Mathematics* **17**(1969) 638–663.

[13] Wets, R., "Programming under uncertainty: the complete problem", *Z. Wahrsch. verw. Gebiete* **4**(1966) 316–339.

[14] Williams, A.C., "Approximation formulas for stochastic linear programming", *SIAM Journal on Applied Mathematics* **14**(1966) 668–677.