STOCHASTIC PROGRAMMING: SOLUTION TECHNIQUES AND APPROXIMATION SCHEMES

Roger J-B. Wets

September 1982
WP-82-84

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

ABSTRACT

Solutions techniques for stochastic programs are reviewed. Particular emphasis is placed on those methods that allow us to proceed by approximation. We consider both stochastic programs with recourse and stochastic programs with chance-constraints.

Supported in part by a Guggenheim Fellowship.

STOCHASTIC PROGRAMMING: SOLUTION TECHNIQUES AND APPROXIMATION SCHEMES

Roger J-B. Wets

1. INTRODUCTION

Optimization problems involving parameters only known in a statistical sense give rise to stochastic optimization models. When dealing with such problems it is important to be aware of their intrinsic dynamic nature since it plays an important role in the modeling process as well as in the design of solution techniques. Briefly the general model is as follows. First an observation of a random phenomena $\xi_1 \in R^{\nu_1}$ is made. Based on this information, a decision $x_1 \in R^{n_1}$ is chosen at some cost $f_1(x_1,\xi_1)$. Then a new observation is made that yields $\xi_2 \in R^{\nu_2}$. On the basis of the information (ξ_1,ξ_2) gained so far, one selects a decision x_2 in R^{n_2} with associated cost $f_2(\xi_1,\xi_2,x_1,x_2)$. This continues up to the time horizon N, possibly ∞. At each stage, the decisions $x_1,x_2,\ldots$ are subject to constraints that may, and usually do, depend on the actual realizations $\xi_1,\xi_2,\ldots$, as well as reliability type constraints that follow from criteria

that the modeler might find difficult to include in the cost functions. The problem is to find *recourse functions* (decision rules, policies, control laws):

$$\xi_1 \longmapsto x_1(\xi_1)$$

$$(\xi_1, \xi_2) \longmapsto x_2(\xi_1, \xi_2)$$

....

that satisfy the constraints and that minimize the expected cost. It is assumed that utility factors have been incorporated in the cost functions.

The development of mathematical programming techniques for studying and solving certain classes of stochastic optimization problems was initiated in the mid 50's by E.M. Beale [1], G. Dantzig [2], G. Tintner [3] and A. Charnes and W. Cooper [4]. The models introduced then, as well as those to be considered here, involve typically only 2(=N) stages with no (truly) random phenomena preceding the choice of x_1, but the basic features of the general model were already ubiquitous. The basic reason for such limitations is that numerous applications require only 2 or 3 stages, either per se or as a consequence of modeling choices. However, the number of decision variables and constraints is liable to be quite large as is the case in typical applications of linear or nonlinear programming. It is this class of problems that is at the core of our concerns, i.e., those problems that can be viewed as "stochastic extensions" of the linear (or

slightly nonlinear) programming model. Multistage problems, say $N > 3$, present no significant theoretical difficulties but they are for all practical purposes computationally intractable, unless they possess structural properties that can be successfully exploited, see for example [5-9]. An excellent overview of the field of Stochastic Programming and its connections to other stochastic optimization problems has been provided by M. Dempster [10, Introduction].

We consider the following class of problems

$$(1.1) \qquad \begin{aligned} &\text{Find } x \geq 0,\ \alpha \in [0,1] \text{ with } P[A(\omega)x \geq b(\omega)] \geq \alpha\ , \\ &\text{such that } Z(x) + \rho(\alpha) \text{ is minimized} \\ &\text{where } Z(x) = cx + E\{\inf_{y \geq 0} q(\omega)y \mid Wy = p(\omega) - T(\omega)x\}\ . \end{aligned}$$

The vectors b, q, p and the matrices A,T are random, whereas c and W are fixed; their sizes are consistent with: $x \in R^{n_1}$, $y \in R^{n_2}$, $b(\omega) \in R^{m_1}$ and $p(\omega) \in R^{m_2}$, and $\rho:[0,1] \to \bar{R}$ is a non-negative monotone nonincreasing lower semicontinuous convex function. A more complete model would involve a number of *chance-constraints*, i.e., several constraints of the type

$$(1.2) \qquad P[A(\omega)x \geq b(\omega)] \geq \alpha$$

but this extension is easy to work out and would add nothing to the substance of our development. Also, the *recourse cost function*

$$(1.3) \qquad Q(x) = E\{Q(x,\omega)\} = \int Q(x,\omega)P(d\omega)\ ,$$

determined by the *recourse problem*

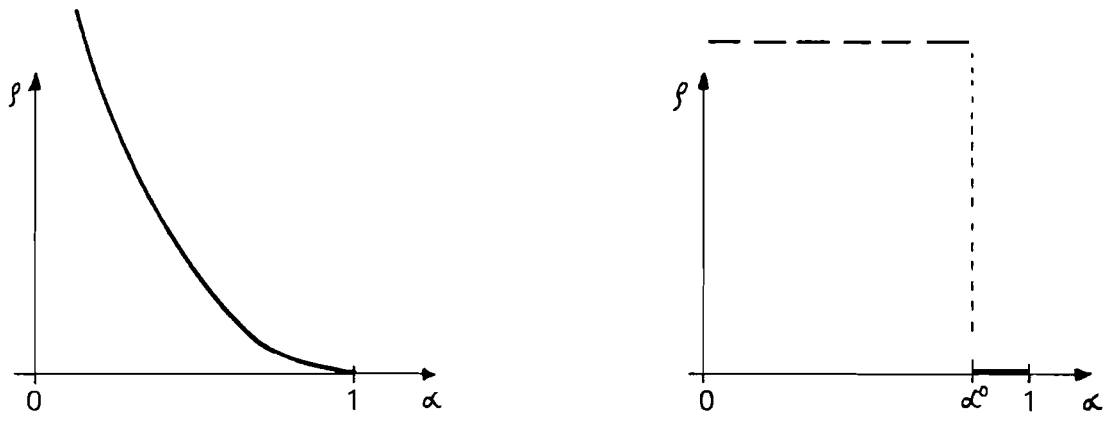
$$(1.4) \qquad Q(x,\omega) = \inf\ q(\omega)y$$
$$\text{subject to } Wy = p(\omega) - T(\omega)x$$
$$y \geq 0\ ,$$

could involve more general constraints on y, convex rather than linear objective, ..., but little would be added to the arguments except that some technical questions would need to be taken care of. When W is random rather than fixed we need a more general theory than that sketched out here; see [11.12], but since our computational capabilities do not yet include such a case, for exposition sake we limit ourselves to fixed W; we then refer to (1.4) as a problem with *fixed recourse*.

The function ρ is not a common feature of the stochastic programming models found in the literature. It represents a cost associated with the relaxation of the constraint

$$(1.5) \qquad A(\omega)x \geq b(\omega) \qquad \text{for all } \omega.$$

Typically it has the form:



1.6 Figure: Reliability Cost.

In the first case the modeler presumably has some cost information about the price he needs to pay to weaken reliability considerations. For the second function, he supposedly has been given a reliability level α° that must be attained at all cost. Problem (1.1) then becomes

$$(1.7) \qquad \begin{aligned} &\text{Find } x \geq 0 \text{ with } P[A(\omega)x \geq b(\omega)] \geq \alpha^\circ \\ &\text{such that } Z(x) = cx + Q(x) \ , \end{aligned}$$

a more common formulation of stochastic programs with (linear) chance-constraints. If moreover $\alpha^\circ = 1$, then the chance-constraints can be replaced, as we shall see, by deterministic constraints and (1.1) takes on the usual form of a stochastic program with recourse.

We take as premise that the probability distribution P of the random elements is given. We shall not consider here the case when there is insufficient statistical information about the random variables of the problem to derive their distribution with a sufficiently high level of confidence. The study of such problems is very recent and there are only limited results available at this time.

We also assume that the random variables of the problem are such that the function $\omega \mapsto Q(\omega,x)$ is bounded below by a summable (finite integral) so that

$$x \mapsto Q(x) : R^{n_1} \to R \cup \{+\infty\} \ ;$$

the function $\omega \longmapsto Q(x,\omega)$ is always measurable, details appear

in [13,14]. In particular this implies that almost surely $Q(x,\omega) > -\infty$, or equivalently the system $\pi W \leq q(\omega)$ is solvable for almost all $q(\omega)$. In fact, let us go one step further and assume that the random variables are such that $\mathcal{Q}(x) = +\infty$ if and only if $Q(x,\omega) = +\infty$ with (strictly) positive probability, i.e., if and only if the linear system

$$Wy = p(\omega) - T(\omega)x \quad , \quad y \geq 0$$

is unsolvable with positive probability. To achieve all of the above it suffices, for example, that the random elements have finite second moments, a condition always satisfied in practice. What precedes are our working assumptions and will be considered as part of the definition of the stochastic program (1.1).

Section 2 reports on computational methods and solution strategies, and Section 3 is devoted to approximation techniques and associated error bounds. In the remainder of this section, we review briefly the main properties of the stochastic program (1.1). We start with its region of feasibility. Let

$$K_1 = \{x \geq 0 \,|\, P[A(\omega)x \geq b(\omega)] \geq \alpha^{\circ} \ , \ \rho(\alpha) < +\infty\} \ , \tag{1.8}$$

with the *induced constraints* given by

$$K_2 = \{x \,|\, \mathcal{Q}(x) < +\infty\} \ . \tag{1.9}$$

The feasibility region K is simply

$$K = K_1 \cap K_2 \ . \tag{1.10}$$

One refers to (1.1) as a stochastic program with *complete recourse* if $K_2 = R^{n_2}$, i.e., there exists a feasible recourse decision whatever be the first stage decision and the random event observed. In general, it may be difficult to compute K or even to determine if a given x belongs or does not belong to K, in particular K_1 might be hard to calculate. Some characterizations of K_1 and K_2 are given here below.

We start with K_1. Let

$$(1.11) \qquad \kappa(\omega) = \{x \geq 0 \,|\, A(\omega)x \geq b(\omega)\} ,$$

and thus

$$(1.12) \qquad \kappa^{-1}(x) = \{\omega \,|\, A(\omega)x \geq b(\omega)\} .$$

By α° we denote the lower bound of α such that $\rho(\alpha) < +\infty$. We have

$$K_1 = \{x \geq 0 \,|\, P[x \in \kappa(\omega)] = P[\kappa^{-1}(x)] \geq \alpha^\circ\} .$$

For each ω, the set $\kappa(\omega)$ is convex but in general K_1 itself is not convex.

1.13 PROPOSITION. *If* $\alpha^\circ = 0$, *then* $K_1 = R_+^{n_1}$. *On other hand if* $\alpha^\circ = 1$, K_1 *is a closed convex set given by*

$$(1.14) \qquad K_1 = \bigcap_{(A,b)\in\sum} \{x \geq 0 \,|\, Ax \geq b\}$$

where $\sum \subset R^{m_1(n_1+1)}$ *is the (image) support of* $A(\cdot)$, $b(\cdot)$, *i.e., the smallest closed subset of* $R^{m_1(n_1+1)}$ *such that* $P[A(\omega),b(\omega)) \in \sum] = 1$. *Moreover if* A *is fixed, or more generally if* $A(\cdot)$ *has finite support (a finite number of possible values) then* K_1 *is a convex polyhedron.*

PROOF. The statement involving $\alpha^\circ = 0$ is trivial. When $\alpha^\circ = 1$, the fact that K_1 is closed and convex follows from (1.14) and that in turn follows from Theorem 2 of [15], with the f function of [15] defined on $R^{n_1} \times R^{m_1} \times R^{m_1 \times (n_1+1)}$ as follows

$$f(x,s,\omega) = A(\omega)x - b(\omega) + s$$

and $Y = R^{n_1}_+$. That K_1 is polyhedral if A is fixed is argued as follows: for each $b(\omega)$, the set $\kappa(\omega) = \{x \geq 0 \,|\, Ax \geq b(\omega)\}$ is a convex polyhedron with each possible face $A_i x \geq b_i(\omega)$ (and $x_j \geq 0$) parallel to the corresponding face determined by the same row A_i but another realization $b_i(\omega')$. The same argument remain valid when $A(\cdot)$ has finite support because we can argue as above for each possible value of $A(\omega)$, and then observe that the finite intersection of polyhedra is also a polyhedron. □

The next proposition completes the results of Proposition 1.13. We state it for the record, its proof would take us too far astray of our main concerns.

1.15 PROPOSITION. *Suppose* $\alpha^\circ = 1$, $b(\cdot)$ *and* $A(\cdot)$ *are independent and the convex hull of* $\sum_A \subset R^{m_1 \times n_1}$, *the support of* $A(\cdot)$, *is polyhedral. Then* K_1 *is a convex polyhedron.*

It is much more difficult to characterize the set K_1 when $0 < \alpha^\circ < 1$. Basically this comes from the fact that

$$P(\kappa^{-1}(x_1)) \geq \alpha^\circ \quad \text{and} \quad P(\kappa^{-1}(x_2)) \geq \alpha^\circ$$

does not imply that

$$P(\kappa^{-1}(x_1) \cap \kappa^{-1}(x_2)) \geq \alpha^\circ \quad ,$$

i.e., there does not exist any subset of events, or possible values of A and b, that can be singled out to yield an expression of the type (1.14). In general the set K_1 is *not* convex and examples can be constructed with K_1 disconnected, even with A fixed. For example, let

$$\kappa(\omega) = \{x \mid x + 3 \geq b(\omega),\ x \leq b(\omega)\} = [b(\omega) - 3,\ b(\omega)]$$

with

$$P[b(\omega) = 0] = P[b(\omega) = 2] = P[b(\omega) = 4] = \frac{1}{3} \quad .$$

Then for $\alpha^\circ = 2/3$, we get

$$K_1 = [-1,0] \cup [1,2] \quad .$$

However, when only $b(\cdot)$ is random, there is a general theory that originates with A. Prékopa [16], who also derived many of the major results; cf. [17] and [18] for surveys.

We say that a *probability measure* P on R^m is *quasi-convex* if for any pair U,V of convex (measurable) sets and for any $\lambda \in [0,1]$ we have

$$P((1 - \lambda)U + \lambda V) \geq \text{Min } \{P(U) , P(V)\}$$

1.17 THEOREM. *Suppose* A *is fixed and the (marginal) probability distribution of* b *is quasi-concave. Then* K_1 *is a closed convex set for any* α°.

PROOF. If K_1 is empty the assertion is immediate. Suppose $x_0, x_1 \in K_1$, then with $x_\lambda = (1 - \lambda)x_0 + \lambda x_1$

$$\kappa^{-1}(x_\lambda) \supset (1 - \lambda)\kappa^{-1}(x_0) + \lambda\kappa^{-1}(x_1) \quad ,$$

since $b(\omega_0) \leq Ax_0$ and $b(\omega_1) \leq Ax_1$ implies that

$$(1 - \lambda)b(\omega_0) + \lambda b(\omega_1) \leq Ax_\lambda \quad .$$

The monotonicity and quasi-concavity of P now yields

$$\begin{aligned} P(\kappa^{-1}(x_\lambda)) &\geq P((1 - \lambda)\kappa^{-1}(x_0) + \lambda\kappa^{-1}(x_1)) \\ &\geq \text{Min } \{P(\kappa^{-1}(x_0)), P(\kappa^{-1}(x_1))\} \end{aligned}$$

But this implies that $P(\kappa^{-1}(x_\lambda)) \geq \alpha^\circ$ since both x_0 and x_1 belong to K_1. Hence $x_\lambda \in K_1$.

To see that K_1 is closed simply observe that if $\{x_k, k = 1,\ldots\}$ is a sequence in K_1 which converges to $\bar{x}$, we have that for each k,

$$\{b \mid b \leq Ax^k = t_k\} = t_k - R_+^{m_1} \quad .$$

Since for each k, $P[t_k - R_+^{m_1}] \geq \alpha^\circ$, it follows that $P[\bar{t} - R_+^{m_1}] \geq \alpha^\circ$ where $\bar{t} = A\bar{x}$. The proof is complete since the last relation implies that $\bar{x} \in K_1$. □

A large class of quasi-concave probability measures can be identified by means of the following result of Borell [19]. *Suppose* h *is a density function of a continuous distribution function defined on* R^m *and* $h^{-1/m}$ *is convex, then the probability measure defined on the Borel subsets* S *of* R^m *by*

$$\text{meas } S = \int_S h(s)ds$$

is quasi-concave. In particular this implies that if the density is of the form

$$h(s) = e^{-Q(s)}$$

where Q is a convex function, the resulting measure is quasi-concave, since

$$[e^{-Q(s)}]^{-1/m} = e^{mQ(s)}$$

is convex as the composition of a convex function with a non-decreasing convex function $s \longmapsto e^s$. Probability measures whose densities are given by an expression of the type $e^{-Q(s)}$ are in fact logarithmic concave, a subclass of the quasi-concave measures, the first class of measures investigated by A. Prékopa [16]. Density functions giving rise to logarithmic concave measures are the (non-degenerate) multi-normal, the multivariate, Dirichlet and Wishart distributions. The multivariate t and F densities (as well as some multivariate Pareto density) engender quasi-concave measures that are not logarithmic concave.

When also A is random, the situation is much more complex. For all practical purposes we have only one result. It is an observation made by Van de Panne and Popp [20], later extended by Prékopa [21] but involving assumptions that appear difficult to verify. Before we come to the little we know, we want to point out the source of the difficulties. Let us consider the "two"-dimensional case: Suppose here that $a(\cdot)$ and $b(\cdot)$ are real-valued random variables and

$$P[a(\omega)x \geq b(\omega)] \geq \alpha^\circ$$

is the chance-constraint for some $0 < \alpha^\circ < 1$. To each $x \in R^1$ corresponds $\kappa^{-1}(x)$ a half-space in R^2 given by the expression

$$\kappa^{-1}(x) = \{(a,b) \mid ax - b \geq 0\} \quad .$$

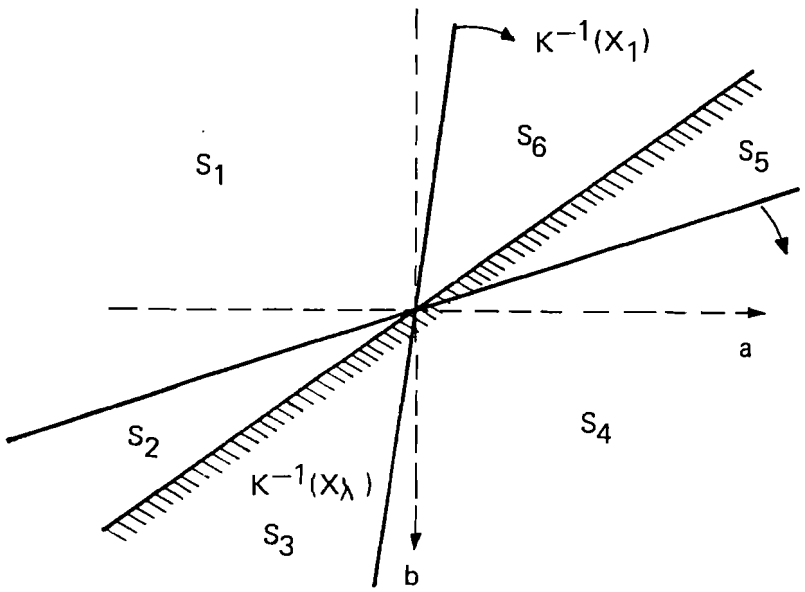
The feasibility set

$$K_1 = \{x \mid P[\kappa^{-1}(x)] \geq \alpha^\circ\}$$

is convex if for any given pair (x_0, x_1) in K_1 and any $\lambda \in [0,1]$ we have

$$P[\kappa^{-1}(x_\lambda)] \geq \alpha^\circ$$

where $x_\lambda = (1 - \lambda)x_0 + \lambda x_1$.



1.18 Figure: Half-spaces Generated by x_0, x_1, x_λ.

Figure 1.18 exemplifies a decomposition of the $\{(a,b)\}$-space through $\kappa^{-1}(x_0)$, $\kappa^{-1}(x_1)$ and $\kappa^{-1}(x_\lambda)$. Note that $\kappa^{-1}(x)$ always contains the vertical positive axis. Let

$S_4 = \kappa^{-1}(x_0) \cap \kappa^{-1}(x_1)$, $S_3 = (\kappa^{-1}(x_\lambda) \cap \kappa^{-1}(x_0)) \setminus S_4$,
$S_5 = (\kappa^{-1}(x_\lambda) \cap \kappa^{-1}(x_1)) \setminus S_4$, $S_6 = \kappa^{-1}(x_1) \setminus (S_4 \cup S_5)$,
$S_2 = \kappa^{-1}(x_0) \setminus (S_4 \cup S_3)$ and $S_1 = R^2 \setminus \cup_{i=2}^{6} S_i$. For $i = 1,\ldots,6$ let $\mu_i = P(S_i)$. Since both x_0 and x_1 belong to K_1 we have

$$\mu_2 + \mu_3 + \mu_4 \geq \alpha^\circ \quad , \tag{1.19}$$

and

$$\mu_6 + \mu_5 + \mu_4 \geq \alpha^\circ \quad . \tag{1.20}$$

The convex combination x_λ of x_0 and x_1 belongs to K_1 if

$$\mu_3 + \mu_5 + \mu_4 \geq \alpha^\circ \tag{1.21}$$

which is implied by

$$\mu_3 + \mu_5 \geq \min [\mu_2 + \mu_3 , \mu_6 + \mu_5]. \tag{1.22}$$

If α° is relatively large, i.e., much larger than .5 if not nearly 1, then (1.19) and (1.20) imply that μ_4 must be of the order of α°; recall that $\sum_{i=1}^{6} \mu_i = 1$, $\mu_i \geq 0$. Thus the inequality (1.22) will be satisfied whenever the probability mass is "sufficiently unimodal". On the other hand, if for example, the distribution is discrete with a sufficient number of points,

linearly independent, not "uniformly" distributed on the plane and with the probability mass sufficiently well-spread out, it will always be possible to find x_0, x_1 and x_λ such that (1.19) and (1.20) hold, but (1.21) and thus also (1.22) fail.

Precise and verifiable conditions that would yield the convexity of K_1 when the matrix A is random have not yet been found although the problem has now been around for the last two decades. It might appear that we exaggerate the importance of convexity for K_1. In this connection, it should be pointed out that the search for a convexity result does not stem purely from computational considerations but from model validation questions. In many ways the chance-constraint

$$(1.23) \qquad P[A(\omega)x \geq b(\omega)] \geq \alpha^\circ$$

is often accepted as the natural generalization of the standard deterministic linear constraints. Little attention is paid to the consequences of this "simple" extension. If we interpret the decision variables $x \in R^{n_1}$ as activity levels, then non-convexity implies that we can choose two programs of activity levels satisfying the constraints but any combinations of these programs is totally unacceptable. Moreover, from what precedes we know that this will occur whenever $A(\cdot)$ and $b(\cdot)$ lack "unimodal properties", in particular if they are discreetly distributed with the probability mass sufficiently will spread out. To some extent this appears to be an irredeemable condemnation of the modeling of stochastic constraints through chance-constraints, at least if more than the right-hand sides of the constraints

are random. However, there is little doubt that there are many situations when it is convenient to rely on chance-constraints to quantify certain of the criteria used by decision makers. Since well-formulated practical problems cannot lead us to unreasonable mathematical constructs, we introduce the following concept:

1.24 DEFINITION. *We say that the probability measure* P *is* α°*-consistent if for all* $\alpha \in [\alpha^\circ, 1]$, *the set* K_1 *is a closed convex set.*

1.25 PROPOSITION. [19] *Suppose that the chance-constraint is actually of the form*

$$P[\sum_{j=1}^{n_1} a_j(\omega)x_j \geq b(\omega)] \geq \alpha^\circ \quad ,$$

where the $a_j(\cdot)$ *and* $b(\cdot) =: a_0(\cdot)$ *are normal random variables, with mean* $\bar{a}_j$, *variance* σ_j *and covariance* $\rho_{jk}\sigma_j\sigma_k$. *Then the corresponding probability measure is* α°*-consistent for all* $\alpha^\circ \geq \frac{1}{2}$, or *equivalently the set* K_1 *is convex for all* $\alpha^\circ \in [1/2, 1]$.

PROOF. For any given x, define the random variable

$$\zeta(x,\omega) = \sum_{j=1}^{n_1} a_j(\omega)x_j - b(\omega) \quad .$$

This is a normal random variable. Setting

$$x_0 = 1 \text{ and } b(\omega) = a(\omega) \quad ,$$

we get that its mean μ and its variance σ^2 are given by

$$\mu(x) = \sum_{j=0}^{n_1} \bar{a}_j x_j \quad ,$$

and

$$\sigma^2(x) = \sum_{j=0}^{n_1} \sigma_j^2 x_j^2 + 2 \sum_{j=0}^{n_1} \sum_{k>j}^{n_1} \rho_{jk} \sigma_j \sigma_k x_j x_k \quad .$$

The chance constraint is then equivalent to

$$1 - \Phi(-\mu(x)/\sigma(x)) \geq \alpha^\circ$$

where Φ is the distribution function of (standard) normal with mean 0 and variance 1. Which can also be expressed as

$$\Phi^{-1}(\alpha^\circ)\sigma(x) - \mu(x) \leq 0$$

recalling that $\Phi^{-1}(1 - \alpha) = -\Phi^{-1}(\alpha)$.

This yields the convexity of K_1, since the form $\sigma^2(x)$ is positive semidefinite in x and $\Phi^{-1}(\alpha^\circ) \geq 0$ precisely when $\alpha^\circ \geq \frac{1}{2}$. □

As indicated already earlier the preceding proposition (with some extensions [21]) is basically the only known result about α°-consistent probability measure for problems involving random matrix A. On the other hand, there are clear indications that a probability measure with "appropriate unimodal" properties is always α°-consistent for $\alpha^\circ < 1$ sufficiently large. For

example, the next approximation result due to S. Sinha [22] points in that direction.

1.26 PROPOSITION. *Let*

$$K_1 = \{x \geq 0 \,|\, P[\textstyle\sum_{j=1}^{n_1} a_j(\omega)x_j \geq b(\omega)] \geq \alpha^\circ\] \quad .$$

Define

$$K_1' = \{x \geq 0 \,|\, (1 - \alpha^\circ)^{-2}(\textstyle\sum_{k=0}^{n_1} \sigma_{jk}x_jx_k)^{\frac{1}{2}} - \sum_{j=0}^{n_1} \mu_j x_j \leq 0,\ x_0 = 1\}$$

where $a_0(\cdot) = b(\cdot)$, μ_j *is the expectation of* $a_j(\cdot)$ *and* σ_{jk} *the covariance of* $a_j(\cdot)a_k(\cdot)$. *Then we always have that* K_1' *is closed and convex and* $K_1 \supset K_1'$.

PROOF. With $a_0(\cdot) = b(\cdot)$ the chance-constraint can be expressed as

$$P[\textstyle\sum_{j=0}^{n_1} a_j(\omega)x_1 = \zeta(x,\omega) \geq 0] \geq \alpha^\circ \quad , \quad x_0 = 1 \quad .$$

We now use one side of Chebyshev's inequality, viz.,

$$P[\zeta(x,\omega) \geq \bar{\zeta}(x) - k^{-2}\sigma_\zeta(x)] \geq 1 - k$$

where $\bar{\zeta}(x)$ is the expectation of $\zeta(x,\cdot)$ and $\sigma_\zeta^2(x)$ its variance, to obtain the next inequality that implies that the chance-constraint

$$\bar{\zeta}(x) - \frac{1}{(1-\alpha)^2}\,\sigma_\zeta(x) \geq 0 \quad , \quad x_0 = 1 \quad .$$

This can also be expressed as

$$(1 - \alpha^{\circ})^{-2}\left(\sum_{j=0}^{n_1} \sum_{k=0}^{n_1} \sigma_{jk}x_j x_k\right)^{\frac{1}{2}} - \sum_{j=0}^{n_1} \mu_j x_j \leq 0 \quad , \quad x_0 = 1 \quad .$$

From this it follows that $K_1' \subset K_1$. The set K_1' is clearly closed and also convex since the quadratic form $\sum_{j=0}^{n_1} \sum_{k=0}^{n_1} \sigma_{jk}x_j x_k$ is positive semidefinite. □

It should be pointed out that in general K_1' is a very crude approximation to K_1 and usually will delete from K_1 those points that are associated with the optimum. There are however many practical situations in which only the means and (co)variances of the random parameters of the problem are known, in which case K_1' is the best available approximation to K_1. The points deleted are then the result of insufficient information.

We now consider K_2, and here because we are able to associate to the stochastic constraints

$$T(\omega)x \text{ "=" } p(\omega)$$

a discrepancy cost proportional to the recourse activities needed to correct the observed differences, a more flexible modeling tool, we are led to a much less hectic situation, at least in general.

1.27 THEOREM. *The set* K_2 *is a closed convex set given by the relation*

$$(1.28) \qquad K_2 = \cap_{(p,T)\in\Xi}\{x \mid p - Tx \in W(R_+^{n_2})\}$$

where $\Xi \subset R^{m_2(n_2+1)}$ *is the support of* $p(\cdot)$, $T(\cdot)$, *i.e., the smallest closed subset of* $R^{m_2(n_2+1)}$ *such that* $P[(p(\omega), T(\omega)) \in \Xi] = 1$. *Moreover, if either* p *and* T *are independent and the convex hull of the support of* $T(\cdot)$ *is polyhedral, or if* $T(\cdot)$ *has finite support, then* K_2 *is also polyhedral.*

For the proof of this theorem, we refer to [13, Section 4]; note also that Sections 4 and 5 of [13] give constructive descriptions of K_2.

Next we turn to the recourse cost function Q as defined (1.3). Since the right-hand side of (1.4) is a linear function of x, it follows from parametric programming that for all ω,

$$x \longmapsto Q(x,\omega)$$

is a convex polyhedral function. From this and the integrability conditions introduced in connection with the definition of the original problem (1.1), it follows:

1.29 THEOREM. *The function* Q *is Lipschitz (finite) and convex on* K_2. *Moreover, for all* $x \in K_2$

$$1.30 \qquad \delta Q(x) = \int \delta Q(x,\omega) P(d\omega) + \delta\psi_{K_2}(x)$$

where ψ_{K_2} *is the indicator function of* K_2, *i.e.*, 0 *on* K_2 *and* $+\infty$ *on its complement. If* P *is absolutely continuous (with respect to the Lebesgue measure) then* Q *is differentiable at every point in the interior of* K_2.

PROOF. The first assertions are proved in [13, Theorems 7.6 and 7.7]. Formula (1.30) follows from a more general result of Rockafellar [23, Corollary 1B], consult [24]. The differentiability follows from (1.30), the fact that $\delta\psi_{K_2}(x) = \{0\}$ on int K_2 and that $\{\omega | Q(x,\omega)$ is not differentiable$\}$ is a set of zero measure because P is absolutely continuous and $Q(\cdot,\omega)$ is differentiable at every $x \in K_2$, except possibly on a set of zero Lebesgue measure. Thus δQ is a singleton for every $x \in$ int K_2 which yields the differentiability at x since Q is convex. □

Combining the properties of K_1, K_2 and Q we have the following

1.31 THEOREM. *Suppose the probability measure* P *is* α°*-consistent, then the stochastic program* (1.1) *is a convex programming problem whose objective function is Lipschitz on the convex closed set* $K = K_1 \cap K_2$. *The set* K *is polyhedral if for example* $\alpha^\circ = 1$ *and* T *is fixed or* $T(\cdot)$ *takes on a finite number of possible values.*

Many variants and extensions of the stochastic program (1.1) have been studied in connection with various applications. Theorem 1.31, except for the assertion about the solution set being polyhedral, remains valid under much more general conditions; for example, when the costs are convex-Lipschitz rather than linear and the constraints have similar properties, when there are more

than 2 stages [12], when the recourse decision must be selected subject to (conditional) chance-constraints involving stochastic variables not yet observed [25, Section V], and so on. In this context, let us just mention a model studied by Prékopa [26] which has an additional reliability constraint for the induced constraints. The set K_2 is redefined as

$$K_2^+ = \{x \mid P[T(\omega)x + Wy(\omega) = p(\omega),\ y(\omega) \geq 0] \geq \alpha'\}$$

and the objective is rendered finite by defining it as follows:

$$Q^+(x) = E\ Q^+(x,\omega)$$

where

$$Q^+(x,\omega) = \inf\ [q(\omega)\cdot y + r(s) \mid Wy + s = p(\omega) - T(\omega,x),\ y > 0\};$$

here $s \in R^{m_2}$ and r is a finite positive convex penalty function. The set K_2^+ can be reexpressed as

$$K_2^+ = \{x \mid P[p(\omega) - T(\omega)x \in W(R_+^{n_2})] \geq \alpha'\}\ .$$

The chance-constraints are thus linear and the results known about K_1 also apply to K_2^+. We are essentially in the setting of problem (1.1). Note also that this is a problem with complete recourse, and hence Q^+ is finite valued.

2. ALGORITHMIC PROCEDURES

Attention will be focused on methods to evaluate and minimize Q; we content ourselves with a few brief remarks concerning feasibility. For the chance-constraint(s) (1.8), we assume that the hypotheses of the problem are such that K_1 can be expressed as

$$(2.1) \qquad K_1 = \{x \geq 0 \,|\, g_{1l}(x,\alpha^\circ) \leq 0 \ , \ l = 1,\ldots,L_1\} \ .$$

where for all l, the functions $(x,\alpha) \longmapsto g_{1l}(x,\alpha)$ are quasi-convex. This certainly includes the case when both A and b are fixed, but also those cases for which we have convexity characterizations for K_1, e.g., with A fixed and $b(\cdot)$ random and P is quasi-concave, then with

$$(2.2) \qquad g_1(x,\alpha) = -P[\kappa^{-1}(x)] + \alpha$$

we have the above representation for K_1. These linear or non-linear constraints are handled as usual in constrained optimization. At least if explicit expressions are available for them. If this is not the case, as would usually occur when g_1 is defined through an expression of the type (2.2), solution procedures must be adapted to the "computable" quantities of that function. For example, computing $P[\kappa^{-1}(x)]$ presuppose the availability of a multidimensional integration routine. We would also need an associate calculus for the multivariate distribution of $A(\cdot)$ and

$b(\cdot)$ that allows us to obtain the gradient (or subgradient) of the function $x \longmapsto P[\kappa^{-1}(x)]$ if the algorithmic procedures requires such information. In [27] Prékopa *et al.* report on a case where all these questions were confronted.

Similarly, we assume that the induced constraints K_2 can be represented by a finite number of constraints, viz.,

$$(2.3) \qquad K_2 = \{x | g_{2l}(x) \leq 0 \quad , \quad l = 1,\ldots,L_2\} \quad ,$$

where naturally, for all $l = 1,\ldots,L_2$, the functions

$$x \longmapsto g_{2l}(x)$$

are convex, cf. Theorem 1.27. Again, explicit expressions for the functions g_{2l} are not easy to come by. However, it is usually possible, as done first in [28], to construct these constraints as needed, i.e., suppose an algorithmic procedure generates an $\hat{x}$ that does not belong to K_2, i.e.,

$$p(\omega) - T(\omega)\hat{x} \notin W(R_+^{n_2})$$

with positive probability. Then there exist a supporting hyperplane, corresponding to a facet, of the polyhedral convex cone $W(R_+^{n_2})$, say

$$\{t \in R^{m_2} | st = 0\} \quad ,$$

such that

$$P[s(p(\omega) - T(\omega)\hat{x}) < 0] > 0 \quad .$$

The constraint

$$(2.4) \qquad \inf_{(p,T)\in\Xi} s(p - Tx) \geq 0$$

where Ξ is again the support of the random $p(\cdot)$ and $T(\cdot)$, is not satisfied by $\hat{x}$ but does not eliminate any feasible points. There are only a finite number of these constraints since $W(R_+^{n_2})$ has only a finite number of facets. In general (2.4) is not a linear constraint, but in practice these constraints are very often linear [13, Section 5]. For example, if T is fixed then (2.4) becomes

$$(2.5) \qquad (sT)x \leq \inf_{p\in\Xi_p} ps$$

where Ξ_p is the support of $p(\cdot)$. The $\inf_{p\in\Xi_p}$ either exists in which (2.5) yields a valid linear constraint or this infimum is $-\infty$ in which case there are no points satisfying this constraint which means that the original stochastic program is infeasible.

Taking into account (2.1) and (2.3), we see that the problem to be solved is given by

$$(2.6) \qquad \text{Find } x \geq 0 \ , \ \alpha \in [0,1] \ \text{ such that}$$

$$g_{1l}(x,\alpha) \leq 0 \quad , \quad l = 1,\ldots,L_1$$

$$g_{2l}(x) \leq 0 \quad , \quad l = 1,\ldots,L_2$$

$$\text{and } z = cx + Q(x) + \rho(\alpha) \ \text{ is minimized}$$

where $\mathcal{Q}$ is a finite convex-Lipschitz function on K_2, defined by (1.3) and (1.4), and repeated here for easy reference,

$$\mathcal{Q}(x) = E\{Q(x,\omega)\} \quad ,$$

and

$$Q(x,\omega) = \inf_y\{q(\omega)y \mid Wy = p(\omega) - T(\omega)x \ , \ y \geq 0\} \ .$$

At least in theory, any standard convex programming package could be used to solve problem (2.6), but usually computing the value of $\mathcal{Q}$, its subgradients or even more so, second order information about $\mathcal{Q}$ requires computational resources far beyond the advantages to be gained from knowing an optimal solution to (2.6). For these reasons any solution method involving line minimization or of the Quasi-Newton type must be quickly discarded, except possibly for special classes of stochastic programs, such as stochastic programs with simple recourse whose random variables obey specific probability laws [29]. We shall not deal with those cases here; because of their special nature, the work on algorithmic procedures for stochastic programs with simple recourse, when $W = (I,-I)$, and extensions thereof, is following a course of its own that is being reviewed separately, see [30]. Here we shall be mostly concerned with the case when no advantage is taken of any special structure of the recourse matrix W, or other components of the stochastic program (1.1).

If the probability distribution of the random elements of the stochastic program is anything but discrete with finite

support, the evaluation of Q or its subgradient given by formula (1.30), involves--at least in principle--the solution of an infinite number of linear programs to describe the function $w \longmapsto Q(x,\omega)$, followed by a multidimensional integration. The material impossibility to work out these operations exactly has led to the development of approximations schemes. To date the only proposed schemes that have been exploited computationally are discretization schemes which consist in the replacement of the original random variables by approximating random variables whose support is finite; henceforth we reserve the term *discrete* to designate this type of random variables. The next section is concerned with the convergence and the error bounds that can be associated with various approximations, the rest of this section deals with solution procedures for (2.6) for discretely distributed random variables.

Let $\{(q_k, p_k, T_k), k = 1,\ldots,N\}$ be the (possible) values of the random variables $(q(\cdot), p(\cdot), T(\cdot))$ and let

$$f_k = P[(q(\omega), p(\omega), T(\omega)) = (q_k, p_k, T_k)]$$

be the associated probabilities. In this case, problem (2.6) is equivalent to

(2.7) Find $x \geq 0$, $\alpha \in [0,1]$ and $y_k \geq 0$, $k = 1,\ldots,N$ such that

$$\begin{aligned}
& g_{11}(x,\alpha) && \leq 0 \quad 1 = 1,\ldots,L_1 \\
& T_1 x + Wy_1 && = p_1 \\
& T_2 x + Wy_2 && = p_2 \\
& \vdots \qquad\qquad \ddots && \vdots \\
& T_N x \qquad\qquad Wy_N && = p_N
\end{aligned}$$

and

$$cx + f_1 q_1 y_1 + f_2 q_2 y_2 + \ldots + f_N q_N y_N + \rho(\alpha) = z \text{ is minimized.}$$

Except possibly for some nonlinearity in ρ or the constraints involving g_{11}, this is a large scale linear program with dual block angular structure. How large, clearly depends on N the number of realizations of the random variables. Note that there was no need to include the induced constraints

$$g_{21}(x) \leq 0 \quad , \qquad l = 1,\ldots,L_2 \quad ,$$

they are automatically incorporated in (2.7), which will be feasible only if for some x there exist for all $k = 1,\ldots,N$, y_k such that

$$Wy_k = p_k - T_k x \quad , \qquad y_k \geq 0 \quad .$$

Again here any large scale programming technique can be specialized--note that the matrices that appear along the diagonal are the same--to solve this type of problem. In fact various such possibilities have been worked out, consult for example [31], [32, Section 3]. Here we retain only those based on compact basis and decomposition techniques, that have been implemented and exhibit at this date the greatest promise.

To somewhat simplify the presentation and to keep our discussion in the realm of large scale linear programming, we assume that there are no terms involving α and suppose that K_1 is given by linear relations of the type

$$K_1 = \{x \geq 0 \,|\, Ax = b\} \quad ,$$

where A and b are fixed matrices. Problem (2.7) then reads

(2.8) Find $x \geq 0$ and $y_k \geq 0$, $k = 1,\ldots,N$ such that

$$Ax = b \quad ,$$

$$T_k x + W y_k = p_k \qquad k = 1,\ldots,N$$

and

$$cx + \sum_{k=1}^{N} f_k q_k y_k = z \text{ is minimized} \quad .$$

A version of the dual of this problem is then

(2.9) Find $\sigma \in R^{m_1}$ and $\pi_k \in R^{m_2}$, $k = 1,\ldots,N$ such that

$$\sigma A + \sum_{k=1}^{N} f_k \pi_k T_k \leq c$$

$$\pi_k W \leq q_k \qquad k = 1,\ldots,N$$

and

$$\sigma b + \sum_{k=1}^{N} f_k \pi_k p_k = w \text{ is minimized.}$$

Problem (2.9) is not quite the usual (formal) dual of (2.8). To obtained the standard form, set

$$\hat{\pi}_k = f_k \pi_k$$

and substitute in (2.9). The dual problem has block-angular structure, the diagonal consisting of identical matrices W.

The compact basis technique, as worked out by B. Strazicky [33] and further analyzed by P. Kall [34], who also implemented the technique as part of an approximation scheme, exploits the structure of the bases of this dual problem to obtain a working basis with

$$n_1^2 + Nn_2^2$$

elements, a number substantially smaller than

$$(n_1 + Nn_2)^2$$

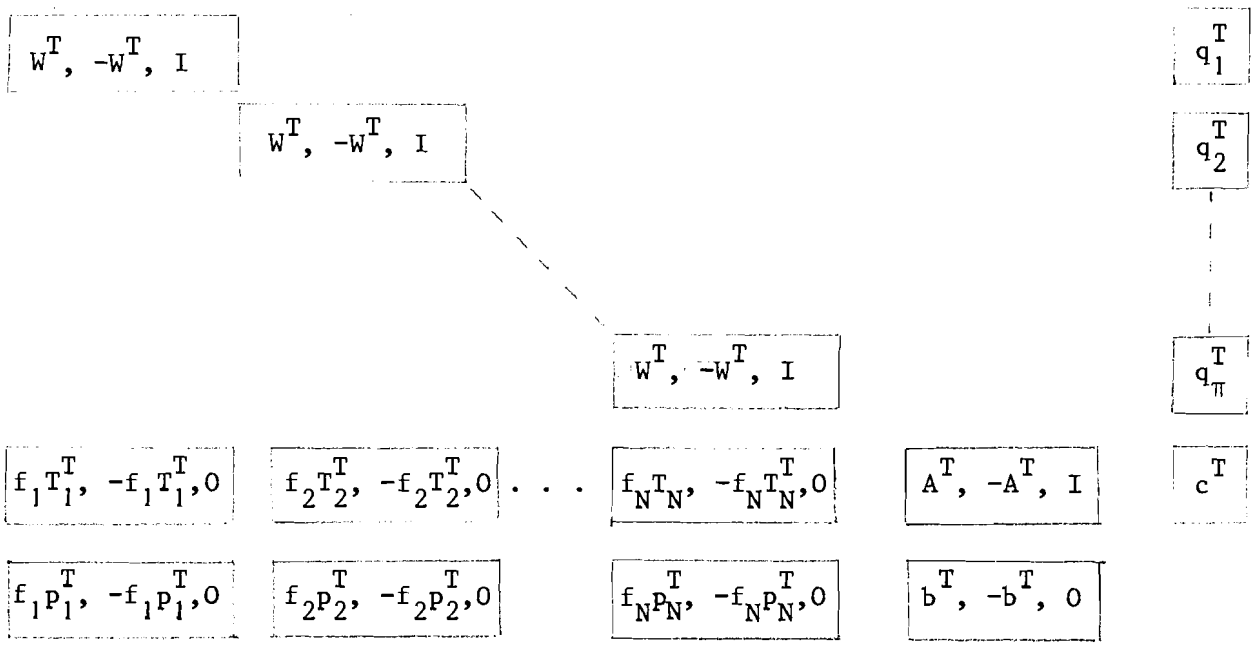
which would be the size of the basis for the standard simplex method. What makes this basis reduction possible is the following observation. Including the slack variables, the constraints of problem (2.9) involve N systems of the type

$$(2.10) \qquad \pi_k^+ W - \pi_k^- W + s_k I = q_k \quad , \quad \pi_k^+ \geq 0, \ \pi_k^- \geq 0, \ s_k \geq 0 \quad .$$

Now assuming that (2.9) is feasible (and bounded) it follows that each basic solution will have at least n_2 basic variables among those associated to the k-subsystem. (In case of degeneracy the pivoting rule can easily be adjusted to guarantee the above.) Any basis generated by the iteration of the simplex method will thus contain at least n_2 columns that "intersect" the k-subsystem.

To see this, it helps to consider the detached coefficients form (2.9):

(2.11)



Let $\hat{B}^T$ be a (feasible) basis for this problem whose restriction to the k-subsystem, we denote by

$$[B_k^T, L_k^T] \quad ,$$

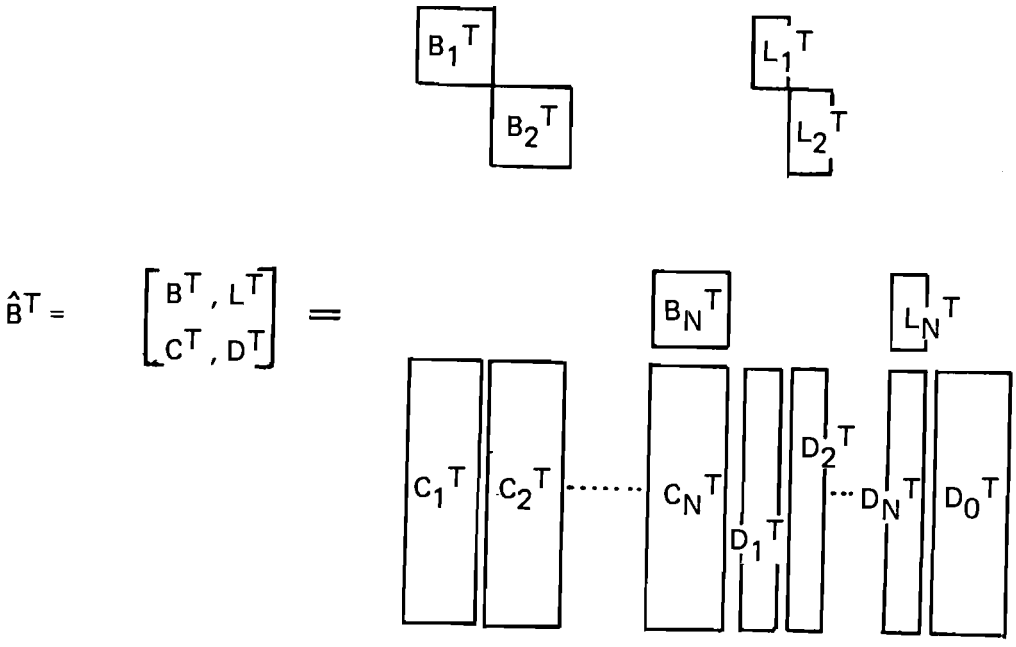
i.e., $[B_k^T, L_k^T]$ is for all k, a submatrix of

$$[W^T, -W^T, I] \quad .$$

The matrix B_k^T is supposed to be invertible (at least n_2 of the columns of the submatrix are linearly independent). The columns of L_k^T are linear combinations of the columns of B_k^T, we can thus express L_k^T as follows:

(2.12) $$L_k^T = B_k^T E_k^T \quad .$$

Recall that naturally L_k^T may be empty when exactly n_2 columns of the k-subsystem are in the basis $\hat{B}^T$. Schematically, and up to a rearrangement of the columns, the basis is of the form



where C_k^T is the submatrix of

$$[f_k T_k^T, -f_k T_k^T, 0]$$

that corresponds to B_k^T and D_k^T the one that corresponds to L_k^T. The D_0^T matrix comes from the columns of

$$[A^T, -A^T, I]$$

that are in the basis. Observe that the $n_1 x n_1$-matrix D^T is invertible. This structure of $\hat{B}$ is to be exploited to reduce the simplex operations, that usually require the inverse of $\hat{B}$, to operations requiring essentially no more than the inverses of B_k.

The simplex multipliers associated with basis $\hat{B}$, denoted by

$$\begin{bmatrix} y \\ x \end{bmatrix} = \begin{bmatrix} y_1 \\ \vdots \\ y_N \\ x \end{bmatrix}$$

are given by the relation

$$(2.14) \qquad \hat{B} \begin{bmatrix} y \\ x \end{bmatrix} = \begin{bmatrix} B & C \\ L & D \end{bmatrix} \begin{bmatrix} y \\ x \end{bmatrix} = \begin{bmatrix} \rho \\ \beta \end{bmatrix}$$

where $[\rho^T, \beta^T]$ is the appropriate rearrangement of the subvector of the coefficients of the objective of (2.11) that corresponds to the columns of $\hat{B}^T$ with β^T being the subvector whose components correspond to the columns of D^T. The (dual feasible) basis is optimal if the vectors

$$(x, y_k, k = 1,\ldots,N)$$

are primal feasible, i.e., satisfy the constraints of (2.8). To obtain x and y we see that (2.14) yields

$$y = B^{-1}(\rho - Cx)$$
$$x = D^{-1}(\beta - Ly)$$

from which we get

$$x = (D - LB^{-1}C)^{-1}(\beta - LB^{-1}\rho)$$

and for $k = 1,\ldots,N$,

$$(2.15) \qquad y^k = B_k^{-1}(\rho_k - C_k x)$$

where ρ_k^T is the subvector of the objective of (2.11) corresponding to the columns in B_k. We have used the fact that B is block diagonal with invertible matrices B_k on the diagonal. Going one step further and using the representation (2.12) for the matrices L_k, we get the equation

$$(2.16) \qquad x = (D - \sum_{k=1}^{N} E_k C_k)^{-1}(\beta - \sum_{k=1}^{N} E_k \rho_k)$$

for x. What is important to notice is that to obtain x and y through (2.16) and (2.15), we only need to know the inverse of the N $(n_2 \times n_2)$-matrices B_k and of the matrix (D - EC).

Similarly to obtain the values of the variables σ and (π_k^+, π_k^-), $k = 1,\ldots,N$; associated to this basis, exactly the same inverses is all that is really required, as can easily be verified. One now needs to work out the updating procedures in order to show that the steps of the simplex method can be performed in this compact form, i.e., that the updating procedures involve only the restricted inverses. This has been carried out in [33]. Experimental computational results are also mentioned in [33]; with only the vector p random, i.e., q and T fixed and

$$m_1 = 30 \ , \ n_1 = 40 \ , \ m_2 = 6 \ , \ n_2 = 5 \quad \text{and} \quad N = 540 \ ,$$

the run time on a CDC 3300 was 20 minutes. Further computational experience involving (generated) problems with random T is reported in [35].

A number of improvements suggest themselves. In [33] it is observed that in problem (2.9) the variables π_k and σ are not restricted in sign and that it is not really necessary to express each π_k as

$$\pi_k = \pi_k^+ - \pi_k^- \quad , \quad \pi_k^+ \geq 0 \ , \ \pi_k^- \geq 0 \quad ,$$

which doubles the number of variables. In fact the π_k should be treated as sign-unrestricted variables with the corresponding columns, i.e., all of W^T, always part of the basis. In fact if the rows of W are linearly independent, then for all k, the columns of W^T could always be left in B_k^T. This means that the only changes that would occur in the matrices

$$(B_k^T \ , \ L_k^T) \quad ,$$

from one basis to the next, would be columns of the identity $I_{(n_2)}$ shuffling in and out of the new basis. This feature was not exploited in the implementation of the algorithm and one may reasonably expect that there would be substantial savings if one did, especially if the inverses can be stored in product form. In fact one could go much further, as we show next.

Since for all k-subsystems, the columns of W^T will be contained in $(B_k^T \ , \ L_k^T)$, we can always keep them in B_k^T. We have

$$(2.17) \qquad B_k^T = [W^T \ , \ I(k1)] \quad , \quad L_k^T = I(k2)$$

where I(k1) consists of $(n_2 - m_2)$ columns of the $(n_2 \times n_2)$-identity and I(k2), possibly empty, consists of a few of the remaining columns of the same identity matrix. Schematically, and up to some rearrangement of the rows, we have that

$$B_k^T = \begin{bmatrix} S_k^T & 0 \\ V_k^T & \begin{matrix} 1 & & \\ & \ddots & \\ & & 1 \end{matrix} \end{bmatrix} = [W^T,\ I(k1)]$$

To know the inverse of B_k^T it really suffices to know S_k^{-1}. The inverse is given by

$$(B_k^T)^{-1} = \begin{bmatrix} (S_k^T)^{-1} & 0 \\ -V_k^T(S_k^T)^{-1} & I \end{bmatrix}$$

as can easily be checked. Thus rather than keeping and updating an $n_2 \times n_2$- matrix for each subsystem, it appears that all the information that is really needed can be manipulated in an $m_2 \times m_2$- matrix. As for standard linear programs we expect m_2 to be usually much smaller than n_2. This should result in substantial savings that would drastically reduced the number of essential operations by simplex iteration as calculated by Kall [34, equations (29) and (30)]. We could pursue the detailed analysis still further taking advantage of the fact that the matrices $D_1,\ldots,D_k$ are all zero, that a number of the S_k are bound to be identical if N is large, and so on. We shall however not do this here, basically because the operations would mimic very closely those of the algorithm to be described next. It is conjectured that a version of this compact basis technique that would fully exploit the structural properties of the dual problem (2.9) would then exhibit the same computational complexity as this second algorithm.

The suggestion of using the decomposition principle to solve stochastic programs goes back to G. Dantzig and A. Madansky [36], the procedure they sketched out took advantage of the structure of the dual problem (2.9). This approach via decomposition was elaborated by R. Van Slyke and myself in [37] relying on a cutting hyperplane algorithm (outer linearization, Benders' decomposition) which can be interpreted as a *partial* decomposition method [37, Section 3]. In view of the matrix layout of the problem to be solved, and the explicit use made of this structure, we refer to it as the *L-shaped* algorithm. Recent work by J. Birge [32] extends the method to multistage problems, he also reports on computational experiments with large scale problems;

see also [38] and [39]. (For an alternative use of decomposition techniques, consult [30, Section 6].)

To describe the method it is useful to think of problem (2.8) in the following form:

(2.18) Find $x \geq 0$ such that
$Ax = b$, and
$cx + Q(x) = z$ is minimized

where

$$Q(x) = \sum_{k=1}^{N} f_k(\inf q_k y \mid W_y = p_k - T_k x \ , \ y \geq 0) \quad .$$

Infeasibility and unboundedness are ignored, they can usually be handled by an appropriate coding of the initialization step, see [40]. The L-shaped algorithm given here is actually a variant of the one in [37, Section 5], in the sense that we are working with a more general class of stochastic programs than those under consideration in [37]. The method consists of 3 steps that can be interpreted as follows. In Step 1 we solve an approximation to (2.18) using an outer-linearization of Q. The two types of constraints (2.19) and (2.20) that appear in this linear program come from

(i) feasibility cuts (determining $K_2 = \{x \mid Q(x) < +\infty\}$),

and

(ii) linear approximations to Q on its domain of finiteness.

These constraints are generated systematically through Steps 2 and 3, when a proposed solution x^ν of the linear program of Step 1 fails to be in K_2 (Step 2) or if the approximating problem does not yet match the function Q at x^ν (Step 3). The row-vectors generated during Step 3 are actually subgradients of Q at x^ν. The convergence is based on the fact that there are only a finite number of constraints of type (2.19) and (2.20) that can be generated since each one corresponds to some basis of W and either some point (p_k, T_k) or to a (finite) number of weighted averages of these points.

Step 1. Set $\nu = \nu + 1$. Solve the linear program

Find $x \geq 0$, $\theta \in R$ such that

$$Ax = b$$

$$D_l x \geq d_l , \quad l = 1,\ldots,s \tag{2.19}$$

$$E_l x + \theta \geq e_l , \quad l = 1,\ldots,t \tag{2.20}$$

and

$$cx + \theta = z \text{ is minimized.}$$

Let (x^ν, θ^ν) be an optimal solution. If no constraints of the form (2.20) are present, θ is set equal to $-\infty$ and ignored in the computation. Initially set $s = t = \nu = 0$.

Step 2. For $k = 1,\ldots,N$ solve the linear program

(2.21) Find $y \geq 0$, $v^+ \geq 0$, $v^- \geq 0$ such that

$$Wy + Iv^+ - Iv^- = p_k - T_k x^\nu \quad , \text{ and}$$

$$ev^+ + ev^- = w^1 \text{ is minimized,}$$

until for some k, the optimal value $w^1 > 0$. Let σ^ν be the associated simplex multipliers and define

$$D_{s+1} = \sigma^\nu T_k$$

and

$$d_{s+1} = \sigma^\nu p_k \quad ,$$

to generate a cut of type (2.19). Return to Step 1 with a new constraint of type (2.19) and set $s = s + 1$. If for all k, $w^1 = 0$ go to Step 3.

Step 3. For all $k = 1,\ldots,N$, solve the linear program

(2.22) Find $y \geq 0$ such that
$Wy = p_k - T_k x^\nu$, and
$q_k y = w^2$ is minimized

Let π_k^ν be the multipliers associated with the optimal solution of the problem k. Define

$$E_{l+1} = \sum_{k=1}^{N} f_k \pi_k^\nu T_k$$
$$e_{l+1} = \sum_{k=1}^{N} f_k \pi_k^\nu p_k$$

and

$$w^{2\nu} = \sum_{k=1}^{N} f_k (p_k - \pi_k^\nu T_k) = e_{l+1} - E_{l+1} x^\nu$$

If $\theta^\nu \geq w^{2\nu}$ stop, x^ν is an optimal solution. Otherwise, return to Step 1 with a new constraint of type (2.20) and set $t = t + 1$.

The separation of Steps 2 and 3 is not just for expository reasons. Problem (2.21) is the counterpart of Phase I of the simplex method for (2.22). Thus, in practice these two operations would not be separated if we proceeded precisely as indicated here. However, there are many cases in which Step 2 can be modified to solving only 1 linear program. Details can be found in [13, Section 5], here let us just suggest the reasons for this simplification. Let $\prec$ be the ordering induced by the closed convex cone $W(R_+^{n_2})$ on R^{m_2}, i.e.,

$$a_1 \prec a_2 \quad \text{if} \quad a_2 - a_1 \in W(R_+^{n_2}) \quad .$$

Then for all $k = 1,\ldots,N$, the system of equations

$$(2.23) \qquad Wy = p_k - T_k x^\nu \quad , \qquad y \geq 0$$

is feasible, if there exists $\alpha \in R^{m_2}$ such that for all $k = 1,\ldots,N$

$$\alpha \prec p_k - T_k x^\nu$$

and the system of equations

$$(2.24) \qquad Wy = \alpha \ , \ y \geq 0$$

is feasible. There always exist such a lower bound. If in

addition, we can choose α such that

$$\alpha = p_{k'} - T_{k'}x^{\nu}$$

for some k', then we have that (2.23) is feasible for all k if and only if (2.24) is feasible. Although in general such a unique α will not exist, it does exist in many instances. And even when a single α will not do, it will usually be sufficient to consider a few such lower bounds, much fewer in any case than all possible (p_k, T_k). Let us remark that computating a lower bound with respect to $\prec$ may be too difficult, but it really suffices to work with lower bounds with respect to a cone ordering induced by any closed convex cone contained in $W(R_+^{n_2})$--for example, a cone generated by a subset of the columns of W--and that cone could be an orthant, cf. [13, Theorem 4.17].

The work to solve Step 3 can also be significantly reduced if we use sifting or bunching procedures as we now explain. By *bunching* we mean the following. With given x, let B be a submatrix of W that is optimal for some k, i.e., corresponding to some realization (q_k, p_k, T_k). Then from the optimality conditions for linear programming it follows that this basis will be also be optimal, when solving problems (2.22), for all k such that

$$B^{-1}(p_k - T_k x) \geq 0 \quad ,$$

and

$$q_k - \gamma_k B^{-1} W \geq 0 \quad ,$$

where γ_k is the subvector of q_k whose elements are the coefficients of the variables that are in the basis. Since B^{-1} is already available, verifying if the above inequalities are satisfied involves relatively little work, especially if only p or q varies with k. Moreover, because of the nature of the problem at hand it is reasonable to expect that only a small number of bases in W will suffice to bunch all the realizations. If problem (2.7) is the result of a discretization of the random variables of the stochastic program, a refinement of the discretization will only increase the work by that required to bunch a larger number of realizations, the basic steps of the algorithm remain unaffected.

In fact the preceding suggest the following overall procedure to solve stochastic programs with arbitrary distributions for the random variables. First, solve an approximation of the original program with only a few samples of $q(\cdot)$, $p(\cdot)$ and $T(\cdot)$, for example, such that

$$(2.25) \qquad (q_k, p_k, T_k) = E\{(q(\omega), p(\omega), T(\omega) \mid (q(\omega), p(\omega), T(\omega) \in \Omega_k\}$$

where Ω_k is part of a partition of the sample space. Using the L-shaped algorithm solve the resulting program (2.8), keeping the bases used to perform the last bunching. (If storage limitations make the storing of all these bases impossible, it is always possible to record them through a listing of the corresponding index vectors.) Next increase the number of samples, either systematically via (2.25) using a finer partitioning of the sample space or through a sampling procedure (Monte-Carlo). Then bunch

this extended sample using the bases already available, if some samples escape this bunching process proceed with the unassigned samples as usual in Step 3, using for value of x the optimal solution of the previous discretization. Continue until the optimal solution of this new (approximating) problem is attained. Then repeat--by which we mean: refine the discretization, use the stored bases of W to bunch this larger sample and so on--until the solution reached satisfy acceptable error bounds, see Section 3.

By *sifting* we mean procedures that rely on a systematic arrangement of the vectors

$$\{q_k \ , \ k = 1,\ldots,N\}$$

and

$$\{p_k - T_k x \ , \ k = 1,\ldots,N\}$$

in order to facilitate the parametric analysis of the linear program (2.22) to be solved in Step 3. This would then substantially simplify and shorten the time required to perform Step 3. In [39] S. Gartska and D. Ruthenberg describe such a sifting procedure. A variant of the arrangement they suggest was compared to a bunching procedure for the case when only p and/or q are random. Sifting appears to be better than bunching if the number of different bases needed to sift or bunch the sample is small, otherwise the need to rearrange the vectors $\{p_k - T_k x \ , \ k = 1,\ldots,N\}$ with each new x, appears to cancel out

whatever advantage one may gain from this preconditioning of the data to accelerate Step 3. However, there appears to be room here for substantial improvements.

We terminate this section with a short discussion of stochastic quasigradient methods. Solving stochastic programs is only one of their potential applications. The method whose roots lie in the theory of stochastic approximation originates with E. Kiefer and I. Wolfowitz who propose a method for unconstrained optimization, cf. also the related work of H. Robbins and S. Monro for solving systems of nonlinear equations. Applications to statistics and probability were further developed by V. Dupač, V. Fabian, J. Sachs and many others, who provided also rates of convergence as well as various characterizations of the limit distributions. H. Kushner and his students extended the methods to encompass systems that can only be measured up to some noise. The use of stochastic quasigradient methods to solve constrained optimization problems start with the work of V. Fabian and Y. Ermoliev in the mid 60's. Since then there have been numerous contributions by B. Poljak, H. Kushner and D. Clark, K. Marti, J-B. Hiriart-Urruty, L. Schmetterer, G. Pflug and A. Ruszcynski, ... the main impetus coming from Y. Ermoliev and his collaborators from the Institute of Cybernetics in Kiev: N. Shor, L. Bajenow, A. Gupal, E. Nurminiski and A. Gaivoronsky. A recent survey has been provided by Y. Ermoliev [41]. Of direct interest to the problem at hand is [42; Chapter I, §4], [43], [44], [41, Section 7] and especially the interesting monograph [45] of K. Marti who uses the method to obtain a constructive proof of duality results, for example; cf. also [46] and [47] where two applications of the

stochastic quasigradient methods to stochastic programming problems are described in detail.

In our setting, the method works basically as follows. Let $F:S \to R$ with S a closed convex subset of R^n be defined by

$$F(x) = E\{f(x,\omega)\}$$

and let us assume that for all ω, $x \longmapsto f(x,\omega)$ is convex. The algorithm generates a sequence $\{x_1,x_2,\ldots\}$ of points of S through the recursive formula

$$x_{\nu+1} = \text{prj}_S\ (x_\nu - \rho_\nu \xi_\nu)$$

where prj_S denotes the projection on S, $\{\rho_\nu,\ \nu = 1,\ldots\}$ is a sequence of scalars and ξ_ν is a stochastic quasigradient of F at x_ν, by which one means a realization of a random n-vector $\underset{\sim}{\xi}_\nu$ satisfying

$$E\{\underset{\sim}{\xi}_\nu | x_1,x_2,\ldots,x_\nu\} \in \partial F(x^\nu) \quad .$$

Typically ξ_ν is obtained as a subgradient of $f(\cdot,\omega_\nu)$ at x_ν where ω_ν is a sample of the random elements determining f or more generally

$$\xi_\nu = \frac{1}{L} \sum_{l=1}^{L} v_l$$

where each $v_l \in \partial f(x_\nu, \omega_l)$ with the ω_l, $l = 1,\ldots,L$ a collection of independent samples.

The method converges with probability 1 if the selected scalars ρ_ν satisfy conditions such as

$$\rho_\nu \geq 0 \ , \ \sum_\nu \rho_\nu > \infty \quad \text{and} \quad \sum_\nu \rho_\nu^2 < \infty \quad .$$

For example, $\rho_\nu = \nu^{-1}$ is such a sequence. The proof can be derived from a modified super-martingale convergence argument [42, Theorem 3], consult also [48].

The application of the method to stochastic programming problem (2.6) works essentially as follows. To facilitate the exposition, let us consider the case when α is not a variable and thus (2.6) can be reexpressed as

$$\begin{aligned} &\text{find } x \in K = \{x \geq 0 \,|\, g_{11}(x) \leq 0,\ g_{21}(x) \leq 0\} \ , \\ &\text{such that } cx + Q(x) = z \text{ is minimized,} \end{aligned} \tag{2.26}$$

where

$$Q(x) = E\{Q(x,\omega) = \inf_{y>0} \ (q(\omega)y \,|\, Wy = p(\omega) - T(\omega)x)\}$$

Samples of $q(\cdot)$, $p(\cdot)$ and $T(\cdot)$ corresponding to a (sample) event will be indicated by subscripting ω, i.e.,

$$Q(x,\omega_k) = \text{Min } \{q_k y \,|\, W_y = p_k - T_k x,\ y \geq 0\}$$

with the sample values (q_k, p_k, T_k) corresponding to the event ω_k. Finally, note that for all $x \in K$, $Q(x,\omega_k)$ is finite and

$$\partial Q(x,\omega_k) = \{-vT_k \mid v \text{ optimal multiplier for (2.22)}\} .$$

The sequence of solution $\{x^1,x^2,\ldots\}$ is produced by the recursion:

$$x^{\nu+1} = \mathrm{prj}_K(x^\nu - \rho_\nu g^\nu) \tag{2.27}$$

where for some $M \geq 1$

$$g^\nu = c + \frac{1}{M}\sum_{\mu=1}^{M} g_\mu \quad , \tag{2.28}$$

and for $\mu = 1,\ldots,M$,

$$g_\mu \in \partial Q(x^\nu,\omega_\mu) \quad .$$

The scalars ρ_ν are assumed to satisfy the appropriate conditions to ensure convergence (with probability 1).

There are three possible stumbling blocks in the implementation of the stochastic quasigradient method to solve (2.26):

the projection on K,

the choice of the step-size ρ_ν,

the stopping criterion.

The *projection* of a point on the closed convex set K is easy only if K is "simple" by which we mean a set such as a bounded interval or a sphere, ... If K is an arbitrary convex polyhedron, then one does usually need to solve a quadratic program of some type in order to obtain $x^{\nu+1}$ as given by (2.27). If K is a general convex set then an even more sophisticated nonlinear programming

technique must be used to get $x^{\nu+1}$.

The choice of the *step-size* ρ_ν is theoretically prescribed by convergence requirements. However, when we use the stochastic gradient method to solve (2.26) we are interested in its short run properties rather than its long run properties, and there is at present no theory that guides us in the choice of the step-size. In practice, some ρ is chosen at the outset and kept a fixed value until the user intervenes to change it or some heuristic is built in the code to adjust the size of ρ when certain phenomena are observed. That takes us very far away from the convergence requirements. How to remedy this is not clear at this time.

Finding a good *stopping criterion* is still very much an open question. Because the function Q is difficult to evaluate--and that is why we are using stochastic quasigradient methods in the first place--it is out of the question to use value comparisons between Q at x^ν and at $x^{\nu+1}$. Y. Ermoliev has suggested that the following quantity

$$\frac{1}{M+1} \sum_{h=\nu-M}^{\nu} Q(x^h, \omega_h) = \hat{Q}(x^\nu)$$

be used as an estimate for $Q(x^\nu)$ with M a relatively large number. The algorithm is to terminate when no improvement is observed in the value of $\hat{Q}$ after ρ has already been reduced to its computationally desirable lower bound.

To conclude this too brief discussion of the stochastic quasigradient method, we would like to point out the connections between this solution method and the L-shaped algorithm. To do

so we work with version (2.18) of the stochastic program, i.e., the discrete case with linear constraints determining K_1. The straightforward implementation of the stochastic quasigradient method would run into difficulties if there are induced constraints on x that cut the feasibility region, i.e., if $x \in K_1 = \{x \geq 0 | Ax = b\}$ does not automatically imply that there is a feasible recourse y_k for all (p_k, T_k). Assume this does not occur, in stochastic programming parlance this means that the *relatively complete recourse* condition is satisfied [13, Section 6]. This also means that Step 2 of the L-shaped algorithm can be skipped. Both algorithms require the calculation of the subgradient of Q. For the L-shaped algorithm this is done in Step 3, whereas in the stochastic quasigradient algorithm only on estimate of the subgradient is needed. Naturally, if all points $\{(q_k, p_k, T_k), k = 1,\ldots,N$ are used to obtain (2.28) then not an estimate but an actual subgradient of Q is utilized by the stochastic quasigradient method. But this would be contrary to the strategy of the method which consists in moving forward as soon as an estimate of a direction of descent is made available, and one hopes that after N(= number of different sample values) steps, with x^ν adjusted at each step, the decrease in the objective will be more substantial than if all samples were used to compute a (reliable) subgradient of Q. Assuming this to be true, and ignoring some of the difficulties that may arise from step-size and projections, the question would then be if the advantage gained from bunching (or sifting), which can be used to speed up Step 3 in the L-shaped algorithm, would not totally offset the fact that at each step of the stochastic quasigradient method, the recourse problem (2.22) must be solved

with a new value of x and (q, p, T). Naturally there too, one should take advantage of the fact that already a basis is available but usually a few pivot operations will be required to reach the new optimal basis.

If one was going to compare the two algorithms it would not be sufficient to measure their respective performance on (discrete) problems of type (2.18) but also when (2.18) is part of an approximation scheme for the original problem, since the stochastic quasigradient method takes no advantage of the shape of the probability distribution of the random variables and is in no way hampered by having continuous distributions. In this connection, one should also mention a recent, still unpublished, result of A. Gaivoronsky which shows that under certain conditions on f, the number of steps required by the stochastic quasigradient method to find the minimum of $E\{f(x,\omega)\}$ is smaller (in a sense which can be made precise) than the number of points required to compute for fixed x^0, the value of $E\{f(x^0,\omega)\}$ by a sampling technique.

3. APPROXIMATIONS AND ERROR BOUNDS

Section 2 dealt with algorithmic procedures for solving stochastic programs whose random variables are discretely distributed. It was suggested that in the case of arbitrary distributions we could proceed by approximation through discretization, obtaining a sequence of approximate solutions through successive refinements of the discretization. This was originally proposed by P. Kall [48] and P. Olsen [49] for recourse problems and by G. Salinetti [50] for chance-constraints. Although approximation through discretization will be the prominent theme of this section, it is by no means the only possibility, see for example [51] where it is suggested that the distribution functions be approximated by piecewise linear distribution functions, or [52] where the multivariate distribution is approximated by linear combination of lognormal univariate distributions, and also the resourceful applications of stochastic programming [53], [54], [55] and [56] where it is the structure of the problem itself that is approximated.

As in Sections 1 and 2 we start with a brief study of

$$K_1 = \{x \geq 0 \,|\, P[A(\omega)x \geq b(\omega)] \geq \alpha^\circ\} ,$$
$$= \{x \geq 0 \,|\, P[x \in \kappa(\omega)] \geq \alpha^\circ\} .$$

Let us assume that there exist matrices A^- and A^+ and vector b^- and b^+ such that for all ω

(3.1) $A^- \leq A(\omega) \leq A^+$ and $b^- \leq b(\omega) \leq b^+$,

(3.2) no row of (A,b) is identically zero for all $A \in [A^- , A^+]$ and $b \in [b^- , b^+]$,

(3.3) the interior of $\kappa^- = \{x \geq 0 \mid A^- x \geq b^+\}$ is nonempty.

Then it is easy to construct sequences of random matrices $A_\nu^-(\cdot)$ and $A_\nu^+(\cdot)$, and vectors $b_\nu^-(\cdot)$ and $b_\nu^+(\cdot)$ taking on only a finite number of values (discretely distributed), satisfying the same bounds as $A(\cdot)$ and $b(\cdot)$, and such that the sequences are monotone with

$$\{(A_\nu^-(\cdot) , b_\nu^-(\cdot)) , \nu = 1,\dots \} \text{ increasing}$$

and

$$\{A_\nu^+(\cdot) , b_\nu^+(\cdot) , \nu = 1,\dots \} \quad \text{decreasing,}$$

both sequences converging uniformly to $(A(\cdot), b(\cdot))$. Relying on the results for the almost sure convergence of measurable multifunctions and the properties of perturbed polyhedra, G. Salinetti [49] proves the following:

(3.4) THEOREM. *Suppose* $A(\cdot)$ *and* $b(\cdot)$ *satisfy conditions* (3.1)-(3.3) *and the sequences* $\{(A_\nu^+(\cdot) , b_\nu^+(\cdot)) , \nu = 1,\dots \}$ *and* $\{(A_\nu^-(\cdot) , b_\nu^-(\cdot)) , \nu = 1,\dots \}$ *are constructed to have the monotonicity and uniform convergence properties indicated here above. Then the sets defined by*

$$K_{1\nu}^{-} = \{x \geq 0 \,|\, P[A_\nu(\omega)x \geq b_\nu(\omega)] \geq \alpha^\circ\}$$

and

$$K_{1\nu}^{+} = \{x \geq 0 \,|\, P[A_\nu(\omega)x \geq b_\nu(\omega)] \geq \alpha^\circ\}$$

determine monotonic sequences of sets, with the $K_{1\nu}^{-}$ *converging from below and the* $K_{1\nu}^{+}$ *converging from above to* K_1.
Recall that a sequence of subsets $\{S_\nu, \nu = 1,\ldots\}$ of R^n is said to converge to a set S if

$$\begin{aligned} S &= \{\lim x_\nu \,|\, x_\nu \in S_\nu \quad \text{for all } \nu\} \\ &= \{\lim x_k \,|\, x_k \in S_{\nu_k} \quad \text{for all k, for some subsequence } \{\nu_k\}\} \ . \end{aligned}$$

The efficiency of this approximation scheme depends clearly on the choice of the discretizations but also, presuming that these approximations are part of an overall iterative procedure, on the possibility of using already available bases to simplify subsequent calculations. At this time there are no computational results available that allow us to verify the practicality of this approximation scheme.

Approximation of the recourse problem, more specifically the function Q, in particular through discretization has been extensively studied, in particular by K. Marti [57], [45], and W. Römisch [58] in addition to P. Kall and P. Olsen already mentioned earlier. Following another line of attack B. Van Cutsem [59] initiated the use of set-convergence to study the

convergence of the solutions of stochastic linear programming. Eventually this, as well as developments in many other areas of Nonlinear Analysis, led to the theory of epi-convergence which provides a unifying framework for the approximation of optimization problems.

Let $\{f; f^\nu; \nu = 1,\dots\}$ be a collection of lower semicontinuous functions defined on R^n and with values in $\bar{R} = [-\infty,\infty]$. The sequence $\{f^\nu, \nu = 1,\dots\}$ is said to *epi-converge* to f if for all $x \in R^n$, we have

$$(3.5)\quad \lim_{\nu\to\infty} \inf f^\nu(x_\nu) \geq f(x) \text{ for all } \{x_\nu, \nu = 1,\dots\} \text{ converging to } x \ ,$$

and

$$(3.6)\quad \text{there exists } \{x_\nu, \nu = 1,\dots\} \text{ converging to } x \text{ such that } \lim_{\nu\to\infty} \sup f^\nu(x_\nu) \leq f(x) \ .$$

It is easy to verify that (3.5) actually implies

$$\lim_{k\to\infty} \inf f^{\nu_k}(x_k) \geq f(x)$$

for any subsequence of functions $\{f^{\nu_k}, k = 1,\dots\}$ and sequence $\{x_k, k = 1,\dots\}$ converging to x. The name epi-convergence comes from the fact that the f_ν epi-converge to f if and only if the sets epi f_ν converge to epi f, where epi h is the epigraph of the function h,

$$\text{epi } h = \{(x,\alpha) \in R^{n+1} \mid \alpha \geq h(x)\} \ .$$

Our interest in epi-convergence stems from the following properties [60]:

(3.7) THEOREM. *Suppose a sequence of lower semicontinuous functions* $\{f^{\nu}, \nu = 1,\dots\}$ *epi-converges to* f. *Then if for some sequence* $\{f^{\nu_k}, k = 1,\dots\}$

$$x_k \in \operatorname{argmin} f^{\nu_k} = \{x \mid f^{\nu_k}(x) \leq \inf f^{\nu_k}\}$$

and $x = \lim_{k\to\infty} x_k$ *it follows that*

$$x \in \operatorname{argmin} f ,$$

and $\inf f = \lim_{k\to\infty} \inf f^{\nu_k}$. *Moreover, if* $\operatorname{argmin} f \neq \phi$ *and* $\inf f$ *is finite, then* $\inf f = \lim_{\nu\to\infty} \inf f^{\nu}$ *if and only if*

$$x \in \operatorname{argmin} f$$

implies that there exist sequences $\{\varepsilon_{\nu} \geq 0, \nu = 1,\dots\}$ *with* $\lim_{\nu\to\infty} \varepsilon_{\nu} = 0$ *and* $\{x_{\nu}, \nu = 1,\dots\}$ *converging to* x *such that for all* ν

$$x_{\nu} \in \varepsilon_{\nu}\text{-}\operatorname{argmin} f^{\nu} = \{x \mid f^{\nu}(x) \leq \inf f^{\nu} + \varepsilon_{\nu}\} \quad .$$

To use this in the context of stochastic programming, recall that the function $\mathcal{Q}$ is given by the following expression

$$\mathcal{Q}(x) = E\{Q(x,\omega)\} = \int Q(x,\omega) P(d\omega)$$

and approximating the probability distribution P by P_ν yields the function Q^ν defined by

$$(3.8) \qquad Q^\nu(x) = \int Q(x,\omega) P_\nu(d\omega) \quad .$$

In what follows we take the P_ν and P to be distributions defined on the sample space of $(q(\cdot), p(\cdot), T(\cdot))$ and identify ω with $(q(\omega), p(\omega), T(\omega))$. In order to bypass some technical difficulties we shall assume that the support of the P_ν and P are contained in a bounded set S, certainly not a significant practical restriction.

(3.9) THEOREM. *Suppose* $\{P_\nu, \nu = 1,\dots\}$ *is a sequence of probability measures that converge in distribution to* P *(= weak convergence). Then for all* $x \in K_2$ *the functions* $\{Q^\nu, \nu = 1,\dots\}$ *epi-converge to* Q. *Among other things, it follows that for all* ν, x_ν *is an optimal solution to the problem:*

find $x \in K$ *that minimizes* $cx + Q^\nu(x)$

and x^* *is a cluster point of the sequence* $\{x_\nu, \nu = 1,\dots\}$ *then* x^* *solves:*

find $x \in K$ *that minimizes* $cx + Q(x)$.

PROOF. For any $x \in K_2$, the function

$$\omega \qquad Q(x,\omega)$$

is continuous [13, Proposition 7.5] and thus also bounded on S, from which we have that for every $x \in K_2$,

$$\lim \mathcal{Q}^\nu(x) = \mathcal{Q}(x) \quad ,$$

as follows from the theory of weak convergence of probability measures, cf. Portemanteau theorem. Thus condition (3.6) for epi-convergence is fulfilled.

The function $x \longmapsto Q(x,\omega)$ is (finite) Lipschitz on K_2. Thus with Lipschitz constant $L(\omega)$, we have

$$|Q(x_1,\omega) - Q(x_2,\omega)| \leq L(\omega) \cdot \text{dist}(x_1,x_2)$$

for any pair x_1, x_2 in K_2. Actually $L(\omega)$ can be chosen independent of ω[13, Proof of Theorem 7.7]. Let $\{x_\nu, \nu = 1,\dots\}$ be any sequence of points in K_2 that converges to $x \in K_2$. We get

$$Q(x,\omega) - L \cdot \text{dist}(x,x_\nu) \leq Q(x_\nu,\omega) \quad .$$

Integrating both sides with respect to P_ν yields

$$\mathcal{Q}^\nu(x) - L \cdot \text{dist}(x,x_\nu) \leq \mathcal{Q}^\nu(x_\nu) \quad .$$

and hence

$$\begin{aligned} \mathcal{Q}(x) &= \lim_{\nu\to\infty} \mathcal{Q}^\nu(x) - L \cdot \lim_{\nu\to\infty} \text{dist}(x,x_\nu) \\ &= \liminf_{\nu\to\infty} (\mathcal{Q}^\nu(x) - L \cdot \text{dist}(x,x_\nu)) \\ &\leq \liminf_{\nu\to\infty} \mathcal{Q}^\nu(x_\nu) \quad , \end{aligned}$$

which gives us (3.5) and thus completes the proof of the epi-convergence of the functions Q^ν to Q. The remaining assertions directly follow from the definition--the epi-convergence of the Q^ν to Q, implies the epi-convergence of $c \cdot + Q^\nu$ to $c \cdot + Q$--and Theorem 3.7. □

From the Lipschitz bound used in the proof of Theorem 3.9, it is actually possible to obtain an estimate of the rate of convergence. For example, P. Kall [48] shows that if only p and T are random, then

$$(3.10) \qquad |Q(x) - Q^\nu(x)| \leq \gamma \int \|p(\omega) - T(\omega)x\| \, |P - P_\nu(d\omega)|$$

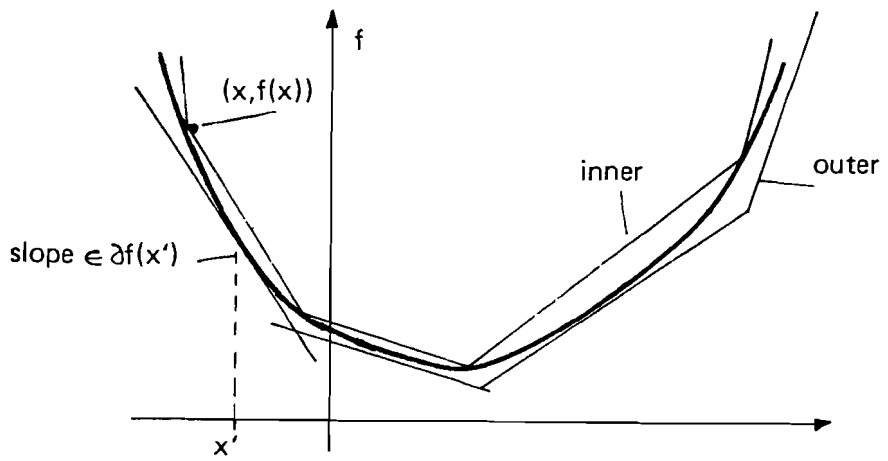
where $\|\cdot\|$ indicates the vector norm and

$$\gamma = \max [\det V | V \text{ is an invertible submatrix of } W] \quad .$$

This is somewhat better than the constant L that appears in the proof of the theorem. It allows us to compute an a priori bound, but in order to get a good approximation bound via (3.10) one needs a discretization with extremely fine mesh which would render the approximate problem (2.8) extemely large. This is why another approach is advocated.

Approximating a convex function f from R^n into R can be done in many ways, but if in addition we seek to obtain upper and lower bounds on the infimum of this function one is naturally led to proceed via outer- and inner-linearization of the function f. The infimum of outer- and inner-linearization providing

respectively the desired lower and upper bounds. For Q, the question is how to choose the sequence P_ν so that the approximations are of outer or inner type.



3.10 FIGURE: Inner/Outer Linearizations of f.

We only consider the case when p and T are random and q is fixed. (The case q random must be dealt with separately, for the properties to exploit in that case, consult [13, Section 7].) Again, identifying ω with $(p(\omega), T(\omega))$, we have that the function

$$\omega \longmapsto Q(x,\omega)$$

is a convex (polyhedral) function for all x [13, Proposition 7.5]. Let $S^\nu = \{S_h^\nu , h = 1,\ldots,H^\nu\}$ be a finite partition of Ξ, the support of the distribution of $p(\cdot)$ and $T(\cdot)$, and define P_ν as follows:

$$P_\nu(S) = \sum_{\{h \mid \omega_h^\nu \in S\}} P(S_h^\nu) \quad ,$$

where

$$\omega_h^\nu = (p_h^\nu , T_h^\nu) = E\{(p(\cdot) , T(\cdot) \mid (p(\omega) , T(\omega) \in S_h^\nu\} \quad .$$

The distribution P_ν is thus a discrete distribution whose probability mass points are the conditional expectation of $(p(\cdot), T(\cdot))$ given S_h^ν for $h = 1,\ldots,H^\nu$.

3.11 PROPOSITION. *Suppose the sequence of distribution* $\{P_\nu, \nu = 1,\ldots\}$ *are defined as here above through partitions* $\{S^\nu, \nu = 1,\ldots\}$ *such that* $S^{\nu+1}$ *is a refinement of* S^ν, *i.e.,*

$$S^{\nu+1} \supset S^\nu \quad .$$

Then $\{Q^\nu, \nu = 1,\ldots\}$ *is a sequence of monotone increasing functions such that for all* ν

$$Q^\nu \leq Q \quad .$$

PROOF. The result follows from the convexity of $Q(x,\cdot)$ through Jensen's inequality. Indeed we always have

$$\int_{S_h^\nu} Q(x,\omega) P_\nu(d\omega) = Q(x,\omega_h^\nu) \leq \int_{S_h^\nu} Q(x,\omega) P(d\omega)$$

from which we get

$$Q^{\nu}(x) = \sum_{h=1}^{H_{\nu}} Q(x,\omega_h^{\nu})P(S_h^{\nu}) \leq \sum_h \int_{S_h^{\nu}} Q(x,\omega)P(d\omega) = Q(x)$$

To see that $Q^{\nu} \leq Q^{\nu+1}$ simply repeat the argument using again the convexity of $Q(x,\cdot)$ to obtain

$$Q(x,\omega_h^{\nu}) \leq \sum_{\{k \mid S_k^{\nu+1} \subset S_h^{\nu}\}} Q(x,\omega_k^{\nu+1}) \frac{P(S_k^{\nu+1})}{P(S_h^{\nu})} \quad . \qquad \square$$

Clearly if the approximating functions are defined as in Proposition 3.11 such that the P_{ν} converge in distribution to P and there is a bounded sequence $\{x_{\nu}, \nu = 1,\dots\}$ such that

$$x_{\nu} \in \operatorname{argmin}_{x \in K} cx + Q^{\nu}(x) \quad ,$$

then any cluster point x^* of the sequence solves the problem

$$\text{find } x \in K \text{ that minimizes } cx + Q(x) \quad .$$

Moreover, the sequence $\{cx_{\nu} + Q(x_{\nu}) = z_{\nu}, \nu = 1,\dots\}$ is monotone and

$$\lim_{\nu \to \infty} z_{\nu} = \inf_{x \in K}(cx + Q(x)) \quad . \tag{3.12}$$

All of this follows directly from Theorem 3.9 and Proposition (3.11). A. Mandansky [61] was the first to use Jensen's inequality to obtain a lower bound for the infimum; see also [62] for a careful treatment of the nonlinear case. The use of conditional expectations to refine these bounds is due to P. Kall [48] and C. Huang, W. Ziemba and A. Ben-Tal [63].

Proposition (3.11) may seem to suggest that better lower bounds for the infimum of the stochastic program requires a global refinement of the partition of Ξ. But clearly it really suffices to choose the refinement so as to improve the approximation of Q in the neighborhood of the infimum. How to achieve this while relying only on a rough partition, is still an open question. The same question needs to be raised in the context of deriving upper bounds that we discuss next.

Upper bounds on Q are obtained through inner linearizations as indicated in Figure 3.10. The basic idea is the following. Again with $\omega = (p(\omega), T(\omega))$, we have that $\omega \longmapsto Q(x,\omega)$ is convex, and finite valued on Ξ when $x \in K_2$. Let us suppose that Ξ is bounded and denoted by ext Ξ, the extreme points of its convex hull, con Ξ. Since $Q(x,\cdot)$ is convex on con Ξ which is bounded, it follows that its supremum is attained at some extreme point of con Ξ, say $e_x \in$ ext Ξ. We get

$$(3.13) \qquad Q(x) = \int Q(x,\omega)P(d\omega) \leq Q(x,e_x) \quad .$$

If $x^1,\ldots,x^\nu$ are a collection of points in K_2 and $e^1,\ldots,e^\nu$ are the corresponding extreme points of Ξ that yield the preceding inequality, we get that for all $x \in$ con $(x^1,\ldots,x^\nu)$, the convex hull of $x^1,\ldots,x^\nu$

$$(3.14) \qquad Q(x) \leq \text{Min } [\textstyle\sum_{l=1}^{\nu} \lambda_l Q(x^l,e^l) \mid \sum \lambda_{l=1}^{\nu} , \lambda_l \geq 0] \quad .$$

The bounds can be substantially improved by considering partitions of Ξ. Let $S = \{S_h, h = 1,\ldots,H\}$ be a partition of Ξ with

with $e_{x,h}$ the extreme point of the convex hull of S_h at which $Q(x,\cdot)$ attains its supremum. Then, for $x \in K_2$

$$(3.15) \qquad Q(x) \leq \sum_{h=1}^{H} Q(x,e_{x,h})P(S_h)$$

Using this bound rather than (3.13) naturally yields an improved version of (3.14).

These bounds, due to P. Kall and D. Stoyan [35], can be much sharpened when the function Q has separability properties. Once again we start with the fact that $Q(x,\cdot)$ is convex to obtain

$$Q(x,\omega) \leq \int_{\text{ext}\,\Xi} Q(x,e)\,\mu_\omega(de)$$

where μ_ω is a probability measure on ext Ξ such that

$$\int e\mu_\omega(de) = \omega \quad ,$$

i.e., such that the convex combination (generated by μ_ω) yields ω. Thus, we have

$$(3.16) \qquad Q(x) \leq \int_{\Xi} \int_{\text{ext}\,\Xi} Q(x,e)\,\mu_\omega(de)P(d\omega) \quad .$$

The problem with this bound, generally much tighter than (3.14), or even the improved version resulting from (3.15), is that it is usually quite difficult to find a manageable expression for μ_ω as a function of ω. Expect if, for example,

$$Q(x,\omega) = \sum_{i=1}^{M} Q_i(x,\omega_i) \quad ,$$

where for all i = 1,...,M

$$\omega_i \longmapsto Q_i(x,\omega_i):R \to R \quad .$$

Then

$$\int Q(x,\omega)P(d\omega) = \sum_{i=1}^{M} \int Q_i(x,\omega_i)P_i(d\omega_i)$$

where P_i is the marginal distribution of ω_i. We can find an expression for the bound (3.16) by obtaining bounds for each $Q_i(x,\cdot)$ separately. Let α_i and β_i respectively be the upper and lower bounds of the support of ω_i, recall that Ξ was assumed to be bounded. By convexity of $Q_i(x,\cdot)$ we have for every $\omega_i \in [\alpha_i,\beta_i]$

$$(3.17) \qquad Q_i(x,\omega_i) \leq (1 - \lambda_{\omega_i})Q_i(x,\alpha_i) + \lambda_{\omega_i} Q_i(x,\beta_i)$$

where $\lambda_{\omega_i} = (\omega_i - \alpha_i)/(\beta_i - \alpha_i)$. Integrating the above on both sides with respect to P_i, and summing over i, we get

$$(3.18) \qquad Q(x) \leq \sum_{i=1}^{M} \left[\frac{\beta_i - E\omega_i}{\beta_i - \alpha_i} Q_i(x,\alpha_i) + \frac{E\omega_i - \alpha_i}{\beta_i - \alpha_i} Q_i(x,\beta_i) \right] \quad .$$

In stochastic programming one refers to this inequality as the *Edmundson-Madansky* inequality. A refinement of this bound can be detained by breaking up

$[\alpha_i,\beta_i]$ into subintervals, say $[\alpha_i^k,\beta_i^k]$ and for each

one rewriting (3.17) using the extreme points of the subinterval.

With m_{ik} denoting the conditional expectation of ω_i when $\omega_i \in [\alpha_i^k, \beta_i^k)$ and $P_{ik} = P_i[\alpha_i^k, \beta_i^k)$, by integrating and summing we get

$$(3.19) \qquad Q(x) \leq \sum_i \sum_k P_{ik}\left[\frac{\beta_i^k - m_{ik}}{\beta_i^k - \alpha_i^k} Q_i(x, \alpha_i^k) + \frac{m_{ik} - \alpha_i^k}{\beta_i^k - \alpha_i^k} Q_i(x, \beta_i^k)\right] .$$

A. Madansky [61] was the first to suggest the use of (3.18), the refinement is due to C. Huang, I. Vertinsky and W. Ziemba [64], see also [35] where the connection with the theory of partial ordering of distribution functions is exhibited. Another way to obtain these inequalities, which loads them with a rich interpretation, is through the minimax approach to stochastic programming investigated first by M. Isofescu and R. Theodorescu [65] and developed by J. Dupačova [66], [67] where stochastic programs are viewed as games against nature: the inf is with respect to x and the sup with respect to a given class of distribution functions. She obtains (3.18) as the result of considering for the class of distributions those satisfying given moment conditions [68].

If the function $\omega \longmapsto Q(x,\omega)$ is concave, which would be the case if only q is random. Then inequality (3.19) and that of Proposition (3.11) are simply reversed.

The preceding results yields basically a priori bounds, but they can also be exploited in the design of algorithmic procedures using the points generated by the algorithm to construct partitioning schemes,... Other bounds that can be exploited in various situations, a posteriori bounds have been suggested by A. Williams [69], K. Marti [45] and J. Birge [70], cf. [71] for a recent compilation as well as further developments.

ACKNOWLEDGEMENT

The sections on computational procedures and approximation bounds have profited significantly from extended discussions with John Birge (University of Michigan) and Larry Nazareth (IIASA). In writing the introduction to stochastic quasigradient methods, I was greatly aided by the comments of Yuri Ermoliev (Institute of Cybernetics in Kiev, and IIASA) and Georg Pflug (Universities of Vienna and Giessen).

REFERENCES

[1] E.M. Beale, On minimizing a convex function subject to linear inequalities, *J. Royal Stat. Soc.* 17B(1955), 173-184.

[2] G.B. Dantzig, Linear programming under uncertainty, *Management Sci.* 1(1955), 197-206.

[3] G. Tintner, Stochastic linear programming with applications to agricultural economics, in *Proc. Second Symposium in Linear Programming*, ed. H.A. Antosiewicz, Washington, 1955, 197-207.

[4] A. Charnes and W.W. Cooper, Chance-constrained programming, *Management Sci.* 5(1959), 73-79.

[5] R. Grinold, A class of constrained linear control problems with stochastic coefficients, in *Stochastic Programming*, ed. M. Dempster, Academic Press, London, 1980, 97-108.

[6] W. Klein Haneveld, A dual of a dynamic inventory control model: the deterministic and stochastic case, in *Recent Results in Stochastic Programming*, ed. P. Kall and A. Prékopa, Springer Verlag Lecture Notes in Economics and Mathematical Systems, 179(1980), 67-98.

[7] J. Dupačová, Water resource systems using stochastic programming with recourse, in *Recent Results in Stochastic Programming*, ed. P. Kall and A. Prékopa, Springer-Verlag Lecture Notes in Economics and Mathematical Systems, 179(1980), 121-134.

[8] K. Back, Optimality and equilibrium in infinite horizon economies under uncertainty, Ph.D. thesis, University of Kentucky, Lexington, 1982.

[9] R. Everitt and W.T. Ziemba, Two period stochastic programmes with simple recourse, *Operations Research*, 27(1979), 485-502.

[10] M. Dempster, *Stochastic Programming*, Academic Press, London, 1980.

[11] D. Walkup and R.J-B. Wets, Stochastic programs with recourse, *SIAM J. Appl. Math.* 15(1967), 316-339.

[12] P. Olsen, Multistage stochastic programming with recourse: the equivalent deterministic program, *SIAM J. Control Optim.* 14(1976), 495-517.

[13] R. Wets, Stochastic programs with fixed recourse: the equivalent deterministic program, *SIAM Review*, 16(1974), 309-339.

[14] P. Kall and W. Oettli, Measurability theorems for stochastic extremals, *SIAM J. Control Optim.* 13(1975), 994-998.

[15] R. Wets, Induced constraints for stochastic optimization problems, in *Techniques of Optimization*, ed. A. Balakrishnan, Academic Press, London, 1972, 433-443.

[16] A. Prékopa, Logarithmic concave measures with applications to stochastic programming, *Acta Sci. Math.* 32(1971), 301-316.

[17] A. Prékopa, Logarithmic concave measures and related topics, in *Stochastic Programming*, ed. M. Dempster, Academic Press, London, 1980, 63-82.

[18] J.G. Kallberg and W.T. Ziemba, Generalized concave functions in stochastic programming and portfolio theory, in *Generalized Concavity in Optimization and Economics*, ed. S. Schaible and W. Ziemba, Academic Press, New York, 1981, 719-767.

[19] Chr. Borell, Convex set functions in d-spaces, *Period. Math. Hungar.*, 6(1975), 111-136.

[20] C. Van de Panne and W. Popp, Minimum-cost cattle feed under probabilistic protein constraints, *Management Sci.* 9(1963), 405-430.

[21] A. Prékopa, Programming under probabilistic constraints with a random technology matrix, *Math. Operationsforsch. Statist.*, 5(1974), 109-116.

[22] S.M. Sinha, Stochastic Programming, Doc. Thesis, University of California-Berkeley, 1963; see also *Proceed. Symposium on Probability and Statistics*, Ben Hindu University, India.

[23] R.T. Rockafellar, Integrals which are convex functionals II, *Pacific J. Mathematics*, 39(1971), 439-469.

[24] R.T. Rockafellar and R. Wets, On the interchange of sub-differentiation and conditional expectation for convex functionals, *Stochastics* 7(1982), 173-182.

[25] D. Walkup and R. Wets, Stochastic programs with recourse: special forms, in *Proceedings of the Princeton Symposium on Mathematical Programming*, ed. H. Kuhn, Princeton University Press, Princeton, 1970, 139-162.

[26] A. Prékopa, Network planning using two-stage programming under uncertainty, in *Recent Results in Stochastic Programming*, eds. P. Kall and A. Prékopa, Springer Verlag Lecture Notes in Economics and Mathematical Systems, Vol. 179, 1980, 215-237.

[27] A. Prékopa, S. Ganczer, I. Deák and K. Patyi, The STABIL stochastic programming model and its experimental application to the electrical energy sector of the Hungarian economy, in *Stochastic Programming*, ed. M. Dempster, Academic Press, London, 1980, 369-385.

[28] R. Wets, Programming under uncertainty: the equivalent convex program, *SIAM J. Appl. Math.* 14(1966), 89-105.

[29] B. Hansotia, Stochastic linear programs with simple recourse: the equivalent deterministic convex program for the normal, exponential and Erlang cases, *Naval Res. Logist. Quat.* 27(1980), 257-272.

[30] L. Nazareth and R. Wets, Algorithms for stochastic programs: the case of nonstochastic tenders, IIASA Working Paper, Laxenburg, Austria, 1982.

[31] F. Louveaux, A solution method for multistage stochastic programs with recourse with application to an energy investment problem, *Operations Res.* 28(1980), 889-902.

[32] J. Birge, Decomposition and partitioning methods for multistage stochastic linear programs, Tech. Report 82-6, Dept. Industrial and Operations Engineering, University of Michigan, 1982.

[33] B. Strazicky, Some results concerning an algorithm for the discrete recourse problem in *Stochastic Programming*, ed. M. Dempster, Academic Press, London, 1980, 263-274.

[34] P. Kall, Computational methods for solving two-stage stochastic linear programming problems, *Z. Angew. Math. Phys.* 30(1979), 261-271; see also *Large-Scale Linear Programming*, eds. G. Dantzig, M. Dempster and M. Kallio, IIASA Collaborative Proceedings Series, Laxenburg, Austria, 1981, 287-298.

[35] P. Kall and D. Stoyan, Solving stochastic programming problems with recourse including error bounds, *Math. Operationsforsch. Statist. Ser. Optimization* (1982).

[36] G. Dantzig and A. Madansky, On the solution of two-stage linear programs under uncertainty, *Proc. Fourth Berkeley Symposium on Mathematical Statistics and Probability, Vol.1*, University of California Press, Berkeley, 1961, 165-176.

[37] R. Van Slyke and R. Wets, L-shaped linear programs with applications to optimal control and stochastic programming, *SIAM J. Appl. Math.* 17(1969), 638-663.

[38] J. Birge, Solution methods for stochastic dynamic linear programs, Tech. Report SOL 80-29, Systems Optimization Laboratory, Stanford University, 1980.

[39] S. Gartska and D. Ruthenberg, Computation in discrete stochastic programs with recourse, *Operations Res.* 21(1973), 112-122.

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

[41] Y. Ermoliev, Stochastic quasigradient methods and their application in systems optimization, IIASA Working Paper, WP-81-2, Laxenburg, Austria, 1981; to appear in *Stochastics*.

[42] Y. Ermoliev, *Stochastic Programming Methods* (in Russian), in Nauka, Moscow 1976.

[43] Y. Ermoliev and E. Nurminski, Stochastic quasigradient algorithms for minimax problems in stochastic programming, in *Stochastic Programming*, ed. M. Dempster, Academic Press, London, 1980, 275-285.

[44] B. Poljak, Nonlinear programming methods in the presence of noise, *Math. Programming* 14(1978), 87-97.

[45] K. Marti, *Approximationen Stochastischer Optimierungsprobleme*, Verlag Anton Hain, Königstein, 1979.

[46] Y. Ermoliev, G. Leonardi and J. Vira, The stochastic quasigradient method applied to a facility location problem, IIASA Working Paper, WP-81-14, Laxenburg, Austria, 1981.

[47] J. Dodu, M. Goursat, A. Hertz, J.-P. Quadrat and M. Viot, Méthodes de gradient stochastique pour l'optimisation des investissements dans un résau électrique, *E.D.F. Bulletin-Sêrie C*, (1981), No. 2, 133-164.

[48] P. Kall, Approximations to stochastic programs with complete fixed recourse, *Numer.Math.*, 22(1974), 333-339.

[49] P. Olsen, Discretizations of multistage stochastic programming problems, *Mathematical Programming Study* 6(1976), 111-124; also Ph.D. Thesis, Cornell Univ., 1974.

[50] G. Salinetti, Approximations for chance-constrained programming problems, Tech. Report No. 13, Istituto della Probabilitá, Univ. Roma, 1981; to appear in *Stochastics* (1983).

[51] R. Wets, Solving stochastic programs with simple recourse II, in *Proceed. 1975 Conference on Information Sciences and Systems*, The Johns Hopkins University, Baltimore, 1975.

[52] A. Dexter, J. Yu and W. Ziemba, Portfolio selection in a lognormal market when the investor has a power utility function: computational results, in *Stochastic Programming*, ed. M. Dempster, Academic Press, London, 1980, 507-523.

[53] R. Grinold, A new approach to multistage stochastic linear programs, *Mathematical Programming Study* 6(1976), 19-29.

[54] E. Beale, J. Forrest and C. Taylor, Multi-time-period stochastic programming, in *Stochastic Programming*, ed. M. Dempster, Academic Press, London, 1980, 387-402.

[55] M. Queyranne and E. Kao, Aggregation in a two-stage stochastic program for manpower planning in the service sector, Tech. Report, University of Houston, 1981.

[56] E.G. Read, Approaches to stochastic reservoir modelling, Working Paper No. 153, College Business Admin., University of Tennessee, 1982.

[57] K. Marti, Approximationen von Entscheidungsproblemen mit linearer Ergebnisfunktion und positiv homogener, subadditiver Verlustfunktion, *Z. Wahrscheinlichkeitstheorie Verw. Gebiete*, 31(1975), 203-233.

[58] W. Römisch, On discrete approximation in stochastic programming, *Proceed. 13. Jahrestagung Mathematische Optimierung*, Vitte 1981, Humboldt-Univ., Berlin, Sektion Mathematik, Seminarbericht 39, 1981.

[59] B. Van Cutsem, Problems of convergence in stochastic linear programming, in *Techniques of Optimization*, ed. A. Balakrishnan, Academic Press, New York, 1972, 445-454.

[60] H. Attouch and R. Wets, Convergence and approximation in nonlinear optimization, in *Nonlinear Programming 4*, eds. O. Mangasarian, R. Meyer and S. Robinson, Academic Press, New York, 1981, 367-394.

[61] A. Madansky, Inequalities for stochastic linear programming problems, *Management Sci.* 6(1960), 197-204.

[62] R. Hartley, Inequalities in completely convex stochastic programming, *J. Math. Anal. Applic.* 75(1980), 373-384.

[63] C. Huang, W. Ziemba and A. Ben-Tal, Bounds on the expectation of a convex function of a random variable: with application to stochastic programming, *Operations Res.* 25(1977), 315-325.

[64] C. Huang, I. Vertinsky and W. Ziemba, Sharp bounds on the value of perfect information, *Operations Res.* 25(1977), 128-139.

[65] M. Iosifescu and R. Theodorescu, Linear programming under uncertainty, in *Colloquium on Applications of Mathematics to Economics, Budapest, 1963*, ed. A. Prékopa, Publishing House Hungarian Academy of Sciences, Budapest, 1965, 133-140.

[66] J. Žačková, On minimax solutions of stochastic linear programming problems, *Casopis Pest. Mat.* 91(1966), 423-429.

[67] J. Dupačova, On minimax decision rule in stochastic linear programming, *Math. Methods Oper. Res.* 1(1980), 47-60.

[68] J. Dupačova, Minimax approach to stochastic linear programming and the moment problem, (in Czech), *Ekonomickomatematicky Obzor*, 13(1977), 279-307. Summary of results in *ZAMM* 58(1978), T466-T467.

[69] A. Williams, Approximation formulas for stochastic linear programming, *SIAM J. Appl. Math.*, 14(1966), 668-677.

[70] J. Birge, The value of the stochastic solution in stochastic linear programs with fixed recourse, *Math. Programming*, 1983.

[71] J. Birge and R. Wets, On initial solutions and approximation schemes for stochastic programs with recourse, IIASA Worling Paper, Laxenburg, Austria, 1982.