OPTIMIZATION OF FUNCTIONALS WHICH DEPEND ON DISTRIBUTION FUNCTIONS: 1. NONLINEAR FUNCTIONAL AND LINEAR CONSTRAINTS

A. Gaivoronski

November 1983
WP-83-114

INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS
2361 Laxenburg, Austria

PREFACE

The development of optimization techniques for solving complex decision problems under uncertainty is currently a major topic of research in the System and Decision Sciences Area at IIASA, and in the Adaptation and Optimization group in particular.This paper deals with methods for the solution of problems in which the objective function depends on probability distributions. Such problems are common in reliability theory and various other branches of operations research, but methods for dealing with them have only recently begun to emerge.

ANDRZEJ WIERZBICKI
Chairman
System and Decision Sciences

ABSTRACT

The main purpose of this paper is to discuss numerical optimization procedures for problems in which both the objective function and the constraints depend on distribution functions. The objective function is assumed to be nonlinear and to have directional derivatives, while the constraints are taken as linear. The proposed algorithm involves linearization of the objective function at the current point and solution of an auxiliary linear subproblem. This last problem is solved using duality relations and cutting-plane techniques.

OPTIMIZATION OF FUNCTIONALS WHICH DEPEND ON DISTRIBUTION FUNCTIONS: 1. NONLINEAR FUNCTIONAL AND LINEAR CONSTRAINTS

A. Gaivoronski

1. INTRODUCTION

The purpose of this paper is to present numerical methods for solving optimization problems in which both the objective function and the constraints depend on distribution functions. Such problems occur in stochastic programming [1,2], reliability theory and various branches of operations research (surveyed in [3]), and robust statistics [4] , among others .

If both the objective function and the constraints are linear with respect to the distribution functions then the problem can be stated as follows:

$$\min \int_X q(x)\mathrm{d}H(x) \tag{1}$$

$$\int_X f^i(x)\mathrm{d}H(x) \le 0 \ , i=\overline{1,m} \tag{2}$$

$$\int_X \mathrm{d}H(x) = 1 \tag{3}$$

where X is a set in Euclidean space R^n . Some specific cases of problem (1)-(3) arise in the theory of Markovian moments and can be solved analytically

[5]. However, the success of analytical methods is very limited even in the linear case (1)-(3) and therefore numerical algorithms are needed. Such methods began to appear in the middle of the last decade.

One idea is to approximate set X by a sequence of finite subsets $X_1, X_2, \ldots, X_s, \ldots$, where

$$X_s = (x_s^1, \ldots, x_s^{n_s}) \ .$$

This sequence of sets has the property that

$$\sup_{z \in X} \min_{x \in X_s} \| x - z \| \to 0$$

as s tends to infinity, i.e., the greatest distance between points in set X and set X_s tends to zero. Let G_s be the set of all distribution functions which correspond to probabilistic measures concentrated in points from X_s :

$$G_s = \{(x_s^1, p_1), \ldots, (x_s^{n_s}, p_{n_s}) \ , \sum_{i=1}^{n_s} p_i = 1 \ , p_i \geq 0\} \ .$$

If we include one more constraint,

$$H \in G_s \ , \tag{4}$$

problem (1)-(4) becomes a finite-dimensional linear programming problem; this raises the possibility of approximating the original problem (1)-(3) by a sequence of finite-dimensional problems (1)-(4). This idea was explored in [3,6], where it was also used to solve nonlinear and minimax problems involving distribution functions. However, this approach can be used only for sets X of small dimension (in fact not exceeding three) due to the high dimensionality of the associated linear programming problems.

Another possible way of solving (1)-(3) is based on the duality relations between problem (1)-(3) and some finite-dimensional minimax problem. First

proposed in [1], this idea was taken further in [2,7], where methods of the cutting-plane type are given.

The purpose of this paper is to use this last approach to develop solution techniques for nonlinear problems of the kind :

$$\min \Psi(H) \tag{5}$$

subject to

$$H \in G \tag{6}$$

where G is defined in the following way:

$$G=\{H : \int_X f^i(x)\mathrm{d}H(x) \leq 0 \ , i=\overline{1,m} \ ; \ \int_X \mathrm{d}H(x) = 1\} \ . \tag{7}$$

We propose an analogue of the linearization (Frank - Wolfe) method in which we are required to solve subproblems of type (1)-(3). It appears that it is not necessary to solve these subproblems precisely: duality relations make it possible to utilize rough solutions of (1)-(3) so that only a limited number of calculations are needed at each iteration.

2. CHARACTERIZATION OF OPTIMAL DISTRIBUTIONS

We shall use the same letter, say H , to denote both the distribution function and the underlying probabilistic measure, where this will not cause confusion. For a given probabilistic measure H we shall denote by B^+ the collection of all subsets of X with positive measure H , and by dom H the support set of H , i.e.,

$$\text{dom } H = \bigcap_{A \in B^+} A \ .$$

Let us first introduce the class of functionals considered, $\Psi(H)$. What we actually need is some analogue of directional differentiability. Suppose that

$$\Psi(H^*+\alpha(H-H^*)) = \Psi(H^*)+\alpha\int_X f^0(x,H^*)\mathrm{d}(H(x)-H^*(x)) + \tau(\alpha,H^*,H) \qquad (8)$$

where

$$\tau(\alpha,H^*,H)=o(\alpha) \quad .$$

In what follows we assume that functions $f^0(x,H)$, $f^i(x)$ are such that expressions (7) and (8) are meaningful.

The following simple conditions are necessary, and in the convex case also sufficient, for distribution H to be a solution of problem (5)-(7) :

Lemma 1

If $\Psi(H^*) \leq \Psi(H)$ for some $H^* \in G$ and all $H \in G$ then

$$z^* = \inf_{H\in G} \int_X f^0(x,H^*)\mathrm{d}H(x) = \int_X f^0(x,H^*)\mathrm{d}H^*(x) \quad .$$

Proof

The proof is of the traditional type for necessary conditions. Note that

$$z^* \leq \int_X f^0(x,H^*)\mathrm{d}H^*(x)$$

is always true. Suppose, on the contrary, that there exists a $\gamma>0$ such that

$$z^* - \int_X f^0(x,H^*)\mathrm{d}H^*(x) \leq -2\gamma \quad .$$

Then there exists an $\overline{H} \in G$ such that

$$\int_X f^0(x,H^*)\mathrm{d}\overline{H}(x) - \int_X f^0(x,H^*)\mathrm{d}H^*(x) < -\gamma \quad .$$

Consider now distributions H_α :

$$H_\alpha = \alpha\overline{H}+(1-\alpha)H^* \quad .$$

According to (8),

$$\Psi(H_\alpha) = \Psi(H^*)+\alpha\int_X f^0(x,H^*)\mathrm{d}(\overline{H}(x)-H^*(x)) + \tau(\alpha,H^*,\overline{H})$$

and for small α we obtain

$$\Psi(H_\alpha) < \Psi(H^*)$$

which contradicts the optimality of H^* . This completes the proof.

Remark. If , additionally, $\Psi(H)$ is convex, i.e.,

$$\Psi(\alpha_1 H_1 + \alpha_2 H_2) \leq \alpha_1 \Psi(H_1) + \alpha_2 \Psi(H_2) \;\; ; \;\; \alpha_1 + \alpha_2 = 1 \, ,$$
$$\alpha_1 \geq 0 \, , \; \alpha_2 \geq 0$$

then this lemma also gives sufficient conditions for a global minimum.

Lemma 1 implies that to check the necessary conditions for problem (5)-(7) requires solution of a linear problem of the form (1)-(3), where $q(x) \equiv f^0(x,H)$. The solution of problem (1)-(3) can be characterized through the duality relations summarized in the following theorem, which was proved in [7].

Theorem 1

Suppose that the following assumptions are satisfied :

1. Set X is compact and functions $f^i(x)$, $i = \overline{0,m}$ are continuous on X .
2. $\text{co}\, Z \neq \phi$ where

$$Z = \{ z \;\; : z = (f^1(x) , \ldots , f^m(x)) \, , x \in X \} \, .$$

Then

1. A solution of problem (1)-(3) exists and the optimal value of $f^0(x)$ is equal to the optimal value of the following minimax problem:

$$\max_{u \in U^+} \overline{\varphi}(u) \;\; , \;\; \overline{\varphi}(u) = \min_{x \in X} \;\; (f^0(x) + \sum_{i=1}^{m} u_i f^i(x))$$

where $U^+ = \{ u : u \in R^m \;\; , \;\; u_i \geq 0 \}$.

2. For any solution H^* of problem (1)-(3) there exists a $u^* \in U^+$ such that

$$\overline{\varphi}(u^*) = \max_{u \in U^+} \overline{\varphi}(u) \;\; , \;\; \text{dom}\, H^* \subseteq X^*(u^*)$$

where

$$X^*(u^*) = \{ x : x \in X \, , \, \overline{\varphi}(u^*) = f^0(x) + \sum_{i=1}^{m} u_i^* f^i(x) \} \;\; .$$

3. There exists a finite set $X_t = \{ x^1, \ldots, x^t \}$, $t \le m+1$, and a solution H^* of problem (1)-(3) such that dom $H^* = X_t$, i.e., the probabilistic measure H^* can be expressed as a collection of $t \le m+1$ pairs $\{(x^1,\bar{p}_1), \ldots, (x^t,\bar{p}_t)\}$.

The probabilities $\bar{p}_1, \ldots, \bar{p}_t$ are solutions of the following linear programming problem:

$$\min_{p} \sum_{i=1}^{t} p_i f^0(x^i) ,$$

$$\sum_{i=1}^{t} p_i f^j(x^i) \le 0 , \; j = \overline{1,m}$$

$$\sum_{i=1}^{t} p_i = 1 , p_i \ge 0 , i = \overline{1,t} \quad .$$

Combining Theorem 1 and Lemma 1 we obtain the following result:

Theorem 2

Suppose that $\Psi(H^*) \le \Psi(H)$ for all $H \in G$ and the following assumptions are satisfied :

1. Set X is compact and functions $f^0(x,H^*)$, $f^i(x)$, $i = \overline{1,m}$ are continuous on X .

2. co $Z \ne \phi$ where

$$Z = \{ z \; : z = (f^1(x), \ldots, f^m(x)) , x \in X \} .$$

Then

1. We have

$$\int_X f^0(x,H^*) dH^*(x) = \max_{u \in U^+} \varphi(u)$$

where

$$\varphi(u) = \min_{x \in X} \; (f^0(x,H^*) + \sum_{i=1}^{m} u_i f^i(x)) \; .$$

2. There exists a u^*, $\varphi(u^*) = \max_{u \in U^+} \varphi(u)$, $u^* \in U^+$ such that

$$\text{dom } H^* \subseteq X^*(u^*)$$

where

$$X^*(u^*) = \{ x : x \in X, \varphi(u^*) = f^0(x, H^*) + \sum_{i=1}^{m} u_i^* f^i(x) \} .$$

3. NUMERICAL METHOD FOR SOLVING PROBLEMS WITH LINEAR CONSTRAINTS

It is now possible to construct a method which finds points satisfying the necessary conditions of Lemma 1 or, in the convex case, global minima . This method is of the linearization type.

Algorithm 1

1. Begin with an initial distribution H^1.

2. Suppose we have an approximate solution H^s before starting iteration number s . Then at the s-th iteration we do the following :

(i) Find a distribution $\overline{H}^s \in G$ such that

$$\int_X f^0(x, H^s) d\overline{H}^s(x) \leq z_s + \varepsilon_s$$

where

$$z_s = \inf_{H \in G} \int_X f^0(x, H^s) dH(x) \qquad (9)$$

and $\varepsilon_s > 0$ is the accuracy with which problem (9) is solved. It is not necessary to know the value of ε_s , only that $\varepsilon_s \to 0$.

(ii) Check whether functional $\Psi(H)$ decreases in the direction $\overline{H}^s - H^s$. If not, return to step (i) and solve problem (9) with higher accuracy. Otherwise go to step (iii).

(iii) Choose a stepsize ρ_s : $0 < \rho_s \leq 1$ and calculate a new approximation to the optimal solution:

$$H^{s+1} = (1-\rho_s)H^s + \rho_s \overline{H}^s \quad . \tag{10}$$

Then go to step (i).

Remark. The stepsize can be chosen according to a number of different rules:

(a) $\quad \bar{\rho}_s \to 0 \ , \ \sum_{s=0}^{\infty} \bar{\rho}_s = \infty$

$$\rho_s = \begin{cases} 0 \ , \text{if } \Psi(H^s) < \Psi(H^s + \bar{\rho}_s(\overline{H}^s - H^s)) \\ \bar{\rho}_s \ , \text{otherwise} \end{cases} \tag{11}$$

(b) $\quad \rho_s = \arg \min_{\alpha \geq 0} \Psi(H^s + \alpha(\overline{H}^s - H^s)) \tag{12}$

(c) Take a sequence α_s ,where

$$\alpha_s \to 0 \ , \ \sum_{s=0}^{\infty} \alpha_s = \infty$$

and

$$\rho_s = \min\{\arg \min_{\alpha \geq 0} \Psi(H^s + \alpha(\overline{H}^s - H^s)) \ , \ \alpha_s \} \quad . \tag{13}$$

We now introduce topology on set G. We shall use weak (star) convergence topology, which by definition is the weakest topology in the space of all probability measures such that the map

$$H \to \int g(x) \, dH(x)$$

is continuous wherever $g(x)$ is bounded and continuous. Note that in this topology the set of all measures with compact support X is compact. In the discussions that follow we shall use this topology when speaking about the continuity of certain functionals with respect to probabilistic measures.

Let $v(H^1,H^2)$ denote some distance between distributions H^1 and H^2 which will induce weak (star) topology on the set of all distributions - for example, the Levy-Prokhorov distance would do.

We shall now prove the convergence of the algorithm given above.

Theorem 3

Suppose that the following statements are true :

1. $X \subset R^n$ is compact set.

2. Functional $\Psi(H)$ satisfies (8) where

$$\frac{\tau(\alpha, H^*, H)}{\alpha} \to 0$$

uniformly over $H^* \in G$, $H \in G$.

3. Function $f^0(x,H)$ is continuous with respect to $x \in X$ and satisfies the Lipschitz condition with respect to $H \in G$:

$$|f^0(x,H_1) - f^0(x,H_2)| \leq L\, v(H_1,H_2)$$

for $H_1 \in G$, $H_2 \in G$, $x \in X$. Functions $f^i(x)$, $i = \overline{1,m}$ are continuous with respect to $x \in X$.

4. $\varepsilon_s \to 0$.

5. Stepsize ρ_s is chosen according to one of (11)-(13) .

Then

$$\lim_{s \to \infty} \inf_{H \in G} \int_X f^0(x,H^s) d(H(x) - H^s(x)) = 0$$

and for all limit distributions H^* of sequence H^s

$$\inf_{H \in G} \int_X f^0(x,H^s) d(H(x) - H^s(x)) = 0 \ .$$

Proof

1. Note that under assumption 2 of the theorem set G is compact in weak (star) topology and therefore function $f^0(x,H)$ is bounded on $X \times G$. This together with (8) implies the continuity of functional $\Psi(H)$ on G. Therefore $\Psi(H)$ has a minimum on G.

2. The argument given in the proof of Lemma 1 leads to the following inequality:

$$\Psi(H^{s+1}) \le \min\{\Psi(H^s) - \rho_s(\gamma_s - \varepsilon_s) + \tau(\rho_s, H^*, H)\ ,\ \Psi(H^s)\} \le$$
$$\min\{\Psi(H^s) - \rho_s(\gamma_s - \varepsilon_s - \tau_1(\rho_s))\ ,\ \Psi(H^s)\} \qquad (14)$$

where $\tau_1(\rho_s) \to 0$ as $\rho_s \to 0$; $\varepsilon_s \to 0$ due to assumption 1 of the theorem and

$$\gamma_s = \inf_{H \in G} \int_X f^0(x, H^s)\, d(H(x) - H^s(x)) \ .$$

The remaining part of the proof depends on the way in which the stepsize ρ_s is chosen.

2(a) Suppose that ρ_s is chosen according to (12). Define

$$\sigma(\beta) = \sup_{\rho \le 1} \{\rho : o(\rho) \le \beta\rho\} \ .$$

From assumption 1 of the theorem we know that $\sigma(\beta) > 0$ if $\beta > 0$. Taking $\beta = \frac{\gamma_s - \varepsilon_s}{2}$ if $\gamma_s > \varepsilon_s$ and $\rho_s = \sigma(\beta)$ we now get from (14):

$$\Psi(H^{s+1}) \le \begin{cases} \Psi(H^s) & ,\text{if } \gamma_s < \varepsilon_s \\ \Psi(H^s) - \sigma(\frac{\gamma_s - \varepsilon_s}{2})\frac{\gamma_s - \varepsilon_s}{2} & ,\text{otherwise} \ . \end{cases} \qquad (15)$$

Inequality (15) immediately gives $\max\{\gamma_s - \varepsilon_s\ ,\ 0\} \to 0$, which implies that $\gamma_s \to 0$ because $\varepsilon_s \to 0$

2(b) Now let ρ_s be chosen according to (11) . Suppose that there exists an $\bar{s}$ such that for $s > \bar{s}$

$$\gamma_s > \varepsilon_s + \alpha + \tau_1(\rho_s)$$

where $\alpha > 0$. Now from (14) we get :

$$\Psi(H^{s+1}) \le \min\{\Psi(H^s) - \alpha\rho_s\ ,\ \Psi(H^s)\} = \Psi(H^s) - \rho_s \ . \qquad (16)$$

Summing (16) from $s > \bar{s}$ to k we obtain :

$$\Psi(H^k) \le \Psi(H^s) - \alpha \sum_{i=s}^{k-1} \rho_i$$

which contradicts $\Psi(H) > -\infty$ because $\sum_{i=1}^{\infty} \rho_i = \infty$. Therefore there is a subsequence n_k such that

$$\max\{0 , \gamma_{n_k} - \varepsilon_{n_k} - \tau_1(\rho_{n_k})\} \to 0 \quad .$$

Now suppose that there is a subsequence m_k such that

$$\gamma_{m_k} - \varepsilon_{m_k} - \tau_1(\rho_{m_k}) > 2\alpha \quad .$$

We may assume without loss of generality that

$$m_k < n_k < m_{k+1} < n_{k+1} \quad \cdots$$

Let us take a sequence l_k such that

$$m_k \le l_k < n_k \quad ,$$

$$\gamma_i - \varepsilon_i - \tau_1(\rho_i) \ge \alpha \quad \text{for} \quad m_k \le i \le l_k \quad ,$$

$$\gamma_{l_k+1} - \varepsilon_{l_k+1} - \tau_1(\rho_{l_k+1}) < \alpha \quad .$$

We shall now estimate the value of functional $\Psi(H)$ on elements of sequence H^{l_k} . From (14) we can show that

$$\Psi(H^{l_k}) \le \sum_{i=1}^{k} (\Psi(H^{l_i}) - \Psi(H^{m_i})) + \Psi(H^{m_1}) \le$$
$$- \alpha \sum_{i=1}^{k} \sum_{j=m_i}^{l_i-1} \rho_j + \Psi(H^{m_1}) \qquad (17)$$

To proceed with this estimate further it is necessary to estimate $\sum_{j=m_i}^{l_i-1} \rho_j$,

which can be done as follows :

$$\alpha < (\gamma_{m_k} - \varepsilon_{m_k} - \tau_1(\rho_{m_k})) - (\gamma_{l_k+1} - \varepsilon_{l_k+1} - \tau_1(\rho_{l_k+1})) = \gamma_{m_k} - \gamma_{l_k} + \gamma_{l_k} - \gamma_{l_k+1} + \varepsilon^1{}_k \le$$
$$\sum_{i=m_k}^{l_k-1} |\gamma_{i+1} - \gamma_i| + |\gamma_{l_k} - \gamma_{l_k+1}| + \varepsilon^1{}_k \le$$

$$\sum_{i=m_k}^{l_k-1} \sup_{H\in G} \left| \int_X f^0(x,H^{i+1})\mathrm{d}(H(x) - H^{i+1}(x)) - \int_X f^0(x,H^i)\mathrm{d}(H(x) - H^i(x)) \right|$$

$$+ \gamma_{l_k} - \gamma_{l_k+1} + \varepsilon^1{}_k \tag{18}$$

where $\varepsilon^1{}_k \to 0$ as $k \to \infty$.

Now let us estimate the term being summed. From the definitio: of the algorithm we have

$$\upsilon(H^{i+1}, H^i) \le C\rho_i \tag{19}$$

where C is some positive constant. Assumption 2 of the theorem together with (19) yields the following estimate :

$$| f^0(x,H^{i+1}) - f^0(x,H^i) | \le L\upsilon(H^i,H^{i+1}) \le C_1\rho_s$$

leading to

$$\sup_{H\in G} \left| \int_X f^0(x,H^{i+1})\mathrm{d}(H(x) - H^{i+1}(x)) - \int_X f^0(x,H^i)\mathrm{d}(H(x) - H^i(x)) \right| \le$$

$$\sup_{H\in G} \left| \int_X (f^0(x,H^{i+1}) - f^0(x,H^i))\mathrm{d}(H(x) - H^{i+1}(x)) \right| +$$

$$\left| \int_X f^0(x,H^i)\mathrm{d}(H^i(x) - H^{i+1}(x)) \right| \le K_1\rho_i \ . \tag{20}$$

Here we also used the fact that the function $f^0(x,H)$ is bounded on set $X \times G$.

Combining (18) and (20) we get :

$$\alpha < K_1 \sum_{i=m_k}^{l_k-1} \rho_i + K_1\rho_{l_k} + \varepsilon^1{}_k = K_1 \sum_{i=m_k}^{l_k-1} \rho_i + \varepsilon^2{}_k$$

which implies

$$\sum_{i=m_k}^{l_k-1} \rho_i > \frac{\alpha - \varepsilon^2{}_k}{K_1} \ . \tag{21}$$

Substituting (21) into (17) yields

$$\Psi(H^{l_k}) \le \Psi(H^{m_1}) - k\frac{\alpha - \varepsilon^2{}_k}{K_1}$$

which contradicts

$$\inf_{H\in G} \Psi(H) > -\infty$$

since $\varepsilon^2{}_k \to 0$. This again gives $\gamma_s \to 0$.

2(c) Proof for the case when the stepsize is chosen according to (13) is similar to 2(a) and 2(b).

This completes the proof.

Remark. Assumption 2 of the theorem can be easily stated without introducing the notions of weak (star) topology and Levy-Prokhorov distance. One possible way is to assume the following:

$$|f^0(x,H_1) - f^0(x,H_2)| \leq |\int_X \lambda(x,H_1,H_2)\, d(H_1(x) - H_2(x))|$$

where $|\lambda(x,H_1,H_2)| < K < \infty$ for some positive K and $H_1 \in G$, $H_2 \in G$, $x \in X$ and $|f^0(x,H)| < C < \infty$ for some positive C and $H \in G$, $x \in X$.

In order to obtain a practical method from the general framework described in this section, we have to specify ways of performing step 2(i). This is the purpose of the next section.

4. SOLVING THE LINEAR SUBPROBLEM USING CUTTING-PLANE TECHNIQUES

We shall now consider a method for solving linear subproblem (9) which reduces step 2(i) of algorithm 1 to the solution of one finite-dimensional linear programming problem. This method uses some of the same ideas as generalized linear programming [8], cutting-plane algorithms [9], and has much in common with the method proposed in [7] for solution of linear problem (1)-(3). The method is based on the duality relations for problem (1)-(3), which were studied in [7].

Let us assume that the assumptions of Theorems 1 and 3 are fulfilled. Then, according to Theorem 1,

$$\min_{H \in G} \int_X f^0(x,H^s)\, dH(x) = \max_{u \in U^+} \varphi^s(u)$$

where

$$\varphi^s(u) = \min_{x \in X} (f^0(x, H^s) + \sum_{i=1}^{m} u_i f^i(x)) \ .$$

Suppose that distribution H^s is fixed. Then it is possible to solve the problem

$$\max_{u \in U^+} \varphi^s(u) \tag{22}$$

with the help of the following cutting-plane method.

Algorithm 2

1. First select $m+1$ points $x^1, x^2, \ldots, x^{m+1}$ and set $\nu = m+1$. These points are used to approximate function $\varphi^s(u)$ by the function

$$\varphi^s(u,0) = \min_{1 \le j \le \nu} (f^0(x^j, H^s) + \sum_{i=1}^{m} u_i f^i(x^j)) \ .$$

The initial approximation u^0 to the solution of problem (22) maximizes the function $\varphi^s(u,0)$:

$$u^0 = \arg \max_{u \in U^+} \varphi^s(u,0)$$

so that we have to solve a linear programming problem.

2. Suppose that before beginning iteration number k we have ν points $x^1, x^2, \ldots, x^\nu$ and the current estimate of the minimum u^{k-1} . Then iteration number k involves the following stages:

(i) Take $\nu = \nu+1$.

(ii) Find

$$x^\nu = \arg \min_{x \in X} (f^0(x, H^s) + \sum_{i=1}^{m} u_i^{k-1} f^i(x)) \tag{23}$$

(iii) Calculate the next approximation to the optimal solution u^{k+1} :

$$u^k = \arg \max_{u \in U^+} \varphi^s(u,k) \tag{24}$$

where $\varphi^s(u,k)$ is the current approximation of function $\varphi^s(u)$:

$$\varphi^s(u,k) = \min_{1 \le j \le \nu} (f^0(x^j, H^s) + \sum_{i=1}^{m} u_i f^i(x^j)) \ .$$

It should be realized that this is only a general framework for solution - much has already been done to avoid increasing the number of points x^i stored and to implement approximate solutions of problem (23) (for details see [7], [10]). The advantage of this method is that it becomes possible to obtain approximate solutions of the initial problem (9) during the solution of problem (22). These approximations are discrete distributions containing no more than $m+1$ points with positive probabilities:

$$(x^1,\bar{p}_1),(x^2,\bar{p}_2),\ldots,(x^{m+1},\bar{p}_{m+1})$$

where the $\bar{p}_i$ are nonzero solutions of the following linear programming problem :

$$\min_{p} \sum_{i=1}^{\nu} p_i f^0(x^i,H^s),$$

$$\sum_{i=1}^{\nu} p_i f^j(x^i) \le 0,\ j=\overline{1,m}$$

$$\sum_{i=1}^{\nu} p_i = 1,\ p_i \ge 0,\ i=\overline{1,t}$$

and the x^i are the corresponding points. Note that the above problem is actually dual to the linear program equivalent to (24), and therefore both problems can be solved simultaneously.

What we actually need while implementing algorithm 1 is not a precise solution of problem (9) at each step, but rather to track its changing extremum value. The approximate solutions of (9) may be very rough for the first few iterations, gradually increasing in accuracy. It appears that algorithm 2 can be used to follow the extremum value by tracking the changing optimal solution of dual problem (22). It is only necessary to make one iteration of algorithm 2 for each iteration of algorithm 1.

In what follows we shall simplify the notation, writing

$$f^0(x,H^s) = f_s^0(x) \ .$$

We now want an algorithm which allows us to follow the optimal solution of problem (22) as the current distribution H^s changes.

Algorithm 2a

1. First select $m+1$ points $x^1, x^2, \ldots, x^{m+1}$ and set $\nu=m+1$. The initial approximation u^0 to the solution of problem (22) maximizes the function $\varphi^0(u,0)$:

$$u^0 = \arg \max_{u \in U^+} \varphi^0(u,0)$$

$$\varphi^0(u,0) = \min_{1 \le j \le \nu} (f_0^0(x^j) + \sum_{i=1}^{m} u_i f^i(x^j))$$

2. Suppose that before beginning iteration number s we have ν points $x^1, x^2, \ldots, x^\nu$ and the current estimate of the minimum u^{s-1} . Then iteration number s involves the following stages:

(i) Take $\nu = \nu+1$

(ii) Find

$$x^\nu = \arg \min_{x \in X} (f_s^0(x) + \sum_{i=1}^{m} u_i^{s-1} f^i(x))$$

(iii) Calculate the next approximation to the optimal solution u^{s+1} :

$$u^s = \arg \max_{u \in U^+} \varphi^s(u,s)$$

$$\varphi^s(u,s) = \min_{1 \le j \le \nu} (f_s^0(x^j) + \sum_{i=1}^{m} u_i f^i(x^j))$$

The following theorem proves the convergence of this method.

Theorem 4

Assume that:

1. Set X is compact and functions $f^i(x)$, $i = \overline{1,m}$ are continuous.

2. Functions $f_s^0(x)$ are continuous for $x \in X$ uniformly on s and

$$\max_{x \in X} | f_{s+1}^0(x) - f_s^0(x) | \to 0$$

as $s \to \infty$.

3. There exists a $\gamma > 0$ such that for any $u \in U^+$, $\| u \| = 1$ there exists an $i \in \{ 1,...,m+1 \}$ for which

$$\sum_{j=1}^{m} u_j f^j(x^i) < -\gamma \ .$$

Then

$$\max_{u \in U^+} \varphi^s(u) - \varphi^s(u^s) \to 0 \ .$$

Proof

1. We shall first prove that sequence u^s is bounded. Take any point $\bar{u} \in U^+$ and estimate the value of $\varphi^s(u,s)$ at this point. Select $i \leq m+1$ such that

$$\sum_{j=1}^{m} \bar{u}_j f^j(x^i) < -\gamma \| \bar{u} \| \ ,$$

which from assumption 3 of the theorem will always be possible. Then

$$\varphi^s(\bar{u},s) \leq f_s^0(x^i) - \gamma \| \bar{u} \| \ . \qquad (25)$$

The uniform continuity of functions $f_s^0(x)$ implies the existence of a constant C such that $| f_s^0(x) | \leq C$. Combining this with (25) yields

$$\varphi^s(\bar{u},s) \leq C - \gamma \| \bar{u} \|$$

and

$$\varphi^s(0,s) \geq -C \ .$$

These two inequalities lead to

$$\varphi^s(0,s) - \varphi^s(\bar{u},s) \geq \gamma \| \bar{u} \| - 2C$$

which implies that the norm of any point u^* which maximizes $\varphi^s(u,s)$ is bounded, i.e.,

$$||u^*|| \leq \frac{2C}{\gamma}$$

where constants C and γ do not depend on s. This proves that sequence u^s is bounded, because u^s maximizes $\varphi^s(u,s)$.

2. Now suppose, arguing by contradiction, that the theorem is not true and that there exists an $\alpha > 0$ and a sequence s_k such that

$$\max_{u \in U^+} \varphi^{s_k}(u) - \varphi^{s_k}(u^{s_k}) > \alpha \ .$$

From the boundedness of sequence u^s and assumptions 1 and 2 we may assume without loss of generality that

$$u^{s_k} \to u^* \ , \ ||u^{s_k} - u^{s_{k+1}}|| \to 0$$
$$\varphi^{s_k}(u^*) \to \varphi^* \ , \ x^{\nu_k} \to x^* \ , \ f^0{}_{s_k}(x^*) \to f^*$$

where $\nu_k = s_k + 1 + m + 1$.

We shall now estimate the difference

$$\max_{u \in U^+} \varphi^{s_{k+1}}(u) - \varphi^{s_{k+1}}(u^{s_{k+1}}) \ .$$

We have

$$\varphi^{s_{k+1}}(u^{s_{k+1}}) \geq \varphi^{s_{k+1}}(u^{s_k}) - K||u^{s_k} - u^{s_{k+1}}|| \tag{26}$$

for some $K > 0$ from assumption 1. We also know that $\varphi^{s_k}(u^*) \to \varphi^*$, which together with assumptions 1 and 2 implies that

$$\varphi^{s_{k+1}}(u^{s_k}) \to \varphi^* \text{ and } \varphi^{s_k}(u^{s_k}) \to \varphi^* \ .$$

Hence

$$\varphi^{s_{k+1}}(u^{s_k}) \geq \varphi^{s_k}(u^{s_k}) - \varepsilon_1(k) \tag{27}$$

where $\varepsilon_1(k) \to 0$ as $k \to \infty$. We have assumed that

$$\max_{x \in X} |f^0_{s+1}(x) - f^0_s(x)| \to 0$$

as $s \to \infty$ and this gives

$$\varphi^{s_k}(u^{s_k}) \geq \varphi^{s_k+1}(u^{s_k}) - \varepsilon_2(k) \tag{28}$$

where $\varepsilon_2(k) \to 0$ as $k \to \infty$. But from the algorithm we have

$$\varphi^{s_k+1}(u^{s_k}) = f^0_{s_k+1}(x^{\nu_k}) + \sum_{i=1}^{m} u_i{}^{s_k} f^i(x^{\nu_k}) \geq$$

$$f^0_{s_k}(x^{\nu_k}) + \sum_{i=1}^{m} u_i{}^{s_k} f^i(x^{\nu_k}) - \varepsilon_3(k) \geq$$

$$f^0_{s_{k+1}}(x^{\nu_k}) + \sum_{i=1}^{m} u_i{}^{s_k} f^i(x^{\nu_k}) - \varepsilon_4(k) \tag{29}$$

where $\nu_k = s_k + 1 + m + 1$. From estimate (29), we obtain:

$$f^0_{s_{k+1}}(x^{\nu_k}) + \sum_{i=1}^{m} u_i{}^{s_k} f^i(x^{\nu_k}) \geq \min_{1 \leq j \leq \nu_{k+1}-1} [\, f^0_{s_{k+1}}(x^j) + \sum_{i=1}^{m} u_i{}^{s_k} f^i(x^j)] \geq$$

$$\min_{1 \leq j \leq \nu_{k+1}-1} [\, f^0_{s_{k+1}}(x^j) + \sum_{i=1}^{m} u_i{}^{s_{k+1}} f^i(x^j)] -$$

$$|| u^{s_k} - u^{s_{k+1}} || \max_{x \in X,\, 1 \leq i \leq m} | f^i(x) | \geq$$

$$\max_{u \in U^+} \min_{1 \leq j \leq \nu_{k+1}-1} [\, f^0_{s_{k+1}}(x^j) + \sum_{i=1}^{m} u_i f^i(x^j)] -$$

$$K_1 || u^{s_k} - u^{s_{k+1}} || \geq \max_{u \in U^+} \varphi^{s_{k+1}}(u) - K_1 || u^{s_k} - u^{s_{k+1}} || \,. \tag{30}$$

Combining (26)-(30) gives :

$$\max_{u \in U^+} \varphi^{s_{k+1}}(u) - \varphi^{s_{k+1}}(u^{s_{k+1}}) \leq K || u^{s_k} - u^{s_{k+1}} || +$$

$$\varepsilon_1(k) + \varepsilon_2(k) + \varepsilon_4(k) + K_1 || u^{s_k} - u^{s_{k+1}} || \leq \varepsilon_5(k)$$

where $\varepsilon_5(k) \to 0$ as $k \to \infty$.

This contradicts the initial assumption

$$\max_{u \in U^+} \varphi^s(u) - \varphi^s(u^s) > \alpha > 0$$

and thus completes the proof.

5. CUTTING-PLANE VERSION OF ALGORITHM 1

We shall now give an algorithm based on the results obtained in Sections 3 and 4. It is assumed that the conditions of Theorem 3 are met.

Algorithm 1a

1. We begin by choosing an initial distribution H^0 which satisfies assumption 3 of Theorem 4 . Let $x^1, x^2, \ldots, x^{m+1}$ be the $m+1$ points which form the initial distribution H^0. Consider the following linear programming problem:

$$\max_{u \in U^+} u_{m+1} \tag{31}$$

$$\sum_{i=1}^{m} u_i f^i(x^j) - u_{m+1} \geq 0 , \; j = \overline{1,m+1} \tag{32}$$

$$\sum_{i=1}^{m} u_i = 1 . \tag{33}$$

Assumption 3 of Theorem 1 is satisfied if and only if problem (31)-(33) has a solution $\bar{u}_1, \bar{u}_2, \ldots, \bar{u}_{m+1}$ such that $u_{m+1} < 0$. If this is the case, the solution is the same as that of the dual problem

$$\min_{p} p_{m+2} \tag{34}$$

$$\sum_{j=1}^{m+1} p_j f^i(x^j) \leq p_{m+2} , \; i=\overline{1,m} \tag{35}$$

$$\sum_{j=1}^{m+1} p_j = 1 , \; p_j \geq 0 , \; j = \overline{1,m+1} , \tag{36}$$

where the optimal value of p_{m+2} is less then 0 . The distribution $\bar{H}$, where

$$\bar{H} = \{(x^1 , \bar{p}_1) , (x^2 , \bar{p}_2) , \ldots , (x^{m+1} , \bar{p}_{m+1})\}$$

and $\bar{p}_i$ is the optimal solution of (34)-(36), has the property

$$\int_X f^i(x) d\bar{H}(x) < 0 \tag{37}$$

and therefore condition 2 of Theorem 1 is satisfied. The converse is also true (see [7]). Thus if condition 2 of Theorem 1 is fulfilled, it is possible to find $m+1$ points $x_1, \ldots, x_{m+1}$ such that problem (34)-(36) has a solution $\bar{p}_1, \ldots, \bar{p}_{m+2}$ with $\bar{p}_{m+2} < 0$. This guarantees that condition 3 of Theorem 4

is fulfilled. These points, together with the probabilities $\bar{p}_i$, now form the desired initial distribution H^0, which is a solution of the following problem :

$$\min_{H} \mu \tag{38}$$

$$\int_X f^j(x)\, dH(x) \le \mu\,,\ j = \overline{1,m} \tag{39}$$

$$\int_X dH(x) = 1\ . \tag{40}$$

This problem is of the form (1)-(3) and may be solved using algorithm 2 (see [7]). We do not need to solve (38)-(40) exactly - we can stop when the current solution satisfies (37). This will occur after a finite number of iterations of algorithm 2 if condition 2 of Theorem 1 is satisfied. The initial step of algorithm 1a therefore involves the following stages:

(i) Take $\nu_0 = m+1$, where ν_s is the number of points in distribution H^s.

(ii) Obtain the initial distribution H^1, where

$$H^1 = \{(x^1, p_1^1), (x^2, p_2^1), \ldots, (x^{m+1}, p_{m+1}^1)\}\ ,$$

by applying algorithm 2 to problem (38)-(40). This algorithm will produce a sequence of distributions $\bar{H}^s$ which after substitution in problem (38)-(40) gives corresponding μ_s. Take as H^1 the first $\bar{H}^s$ with $\mu_s < 0$.

(iii) Take the initial point $u^1 \in U^+$ for solution of the dual problem.

2. Suppose that before beginning iteration number s we have the current approximation to optimal solution H^s :

$$H^s = \{(x^1, p_1^s), \ldots, (x^{\nu_s}, p_{\nu_s}^s)\}$$

and point u^s. Iteration number s then involves the following operations, where steps (i)-(iii) correspond to steps (i)-(iii) of algorithm 2a and step (i) of algorithm 1, and steps (iv)-(v) correspond to steps (ii)-(iii) of algorithm 1 :

(i) Take $\nu^{s+1} = \nu^s + 1$

(ii) Find a new point $x^{\nu_{s+1}}$, where

$$x^{\nu_{s+1}} = \arg \min_{x \in X} [f^0(x, H^s) + \sum_{i=1}^{m} u_i^s f^i(x)] \tag{41}$$

(iii) Solve the following linear programming problem :

$$\min_{p} \sum_{i=1}^{\nu_{s+1}} p_i f_s^0(x^i) \tag{42}$$

$$\sum_{i=1}^{\nu_{s+1}} p_i f^j(x^i) \leq 0 \ , \ j = \overline{1,m} \tag{43}$$

$$\sum_{i=1}^{\nu_{s+1}} p_i = 1 \ , \ p_i \geq 0 \ , \ i = \overline{1,m+1} \tag{44}$$

together with its dual:

$$\max_{u} u_{m+1} \tag{45}$$

$$f^0(x, H^s) + \sum_{i=1}^{m} u_i f^i(x^j) - u_{m+1} \geq 0 \ , \ j = \overline{1,\nu_{s+1}} \tag{46}$$

$$u_i \geq 0 \ , \ i = \overline{1,m} \ .$$

This will give us the next approximation to the solution of the dual problem, u^{s+1}, and also vector $\bar{p}^{s+1}$:

$$\bar{p}^{s+1} = (\bar{p}_1^{s+1} \ , \ldots , \bar{p}_{\nu_{s+1}}^{s+1})$$

which will have no more than $m+1$ nonzero elements, say $(\bar{p}_{k_1}^{s+1} \ , \ldots , \bar{p}_{k_{m+1}}^{s+1})$.

(iv) Take the family of distributions

$$H^{s+1}(\alpha) = \{(x^1, p_1^{s+1}(\alpha)) \ , \ldots , \ (x^{\nu_{s+1}}, p_{\nu_{s+1}}^{s+1}(\alpha))\}$$

where

$$p_i^{s+1}(\alpha) = \begin{cases} p_i^s & , \text{ if } i \neq k_j \text{ for } j = \overline{1,m+1} \\ p_{k_j}^s (1 - \alpha) + \alpha \bar{p}_{k_j}^{s+1} & , \text{ otherwise } . \end{cases}$$

Then find

$$\alpha_s = \arg \min_{0 \leq \alpha \leq 1} \Psi(H^{s+1}(\alpha))$$

and take $\rho_s = \min\{ \alpha_s, \beta_s \}$, where

$$\beta_s \to 0 \ , \ \sum_{i=1}^{\infty} \beta_i = \infty \ .$$

(v) Take

$$H^{s+1} = H^{s+1}(\rho_s)$$

and go to step 2(i).

Remark. It is not necessary to solve nonlinear programming problem (41) with great accuracy. All we need is a point $x^{\nu_{s+1}}$ such that

$$\lim_{s \to \infty} \{ [f^0(x^{\nu_{s+1}}, H^s) + \sum_{i=1}^{m} u_i^s f^i(x^{\nu_{s+1}})] - \min_{x \in X} [f^0(x, H^s) + \sum_{i=1}^{m} u_i^s f^i(x)] \} = 0 \ .$$

It is also possible to avoid increases in the dimension of linear programming problem (42)-(44) by considering only points x^i which satisfy some additional inequality (see [7]).

Remark. Algorithm 2a adds one additional point to the current approximation of optimal solution H^s at each iteration, which may not be convenient if we have restrictions on the amount of memory available for storing the distribution. In this case measures should be taken to avoid this expansion, perhaps at the expense of accuracy. Some possible ways of achieving this are discussed below.

1. Suppose we want to find the best possible approximation, in no more than N points, to the optimal solution of (5)-(7). (It is assumed that some additional memory is available for storing N further points.) We then proceed as follows:

(i) Run algorithm 2a until the current distribution H^s contains $2N$ points. Arrange these points in order of decreasing probabilities :

$$H^s = \{ (x^{l_1}, p_{l_1}) , \ldots , (x^{l_{2N}}, p_{l_{2N}}) \} \ .$$

(ii) Start algorithm 2a again from the distribution

$$H^1 = \{(x^1 , p_1^1) , \ldots , (x^N , p_N^1)\}$$

$$x^i = x^{l_i} , p_i^1 = \frac{p_{l_i}}{1 - \vartheta} , i = \overline{1,N} , \vartheta = \sum_{j=N+1}^{2N} p_{l_j} .$$

(iii) Continue this process as long as the new $2N$-point distribution has a better value of $\Psi(H)$ than the previous one.

2. Suppose that we want to find an approximation to the optimal solution using at most N points.

(i) Run algorithm 2a until the current distribution contains N points. Let $\bar{p} = \max_{1 \le i \le N} p_i^s$. Divide the set $\{ 1 , \ldots , 2N \}$ into two subsets:

$$I_1 = \{i : p_i^s \ge \chi \bar{p} \} , I_2 = \{i : p_i^s < \chi \bar{p} \}$$

where $\chi > 0$ should be chosen previously.

(ii) Start algorithm 2a again from distribution H^1 :

$$H^1 = \{(x^i , p_i^1) , i \in I_1 , p_i^1 = \frac{p_i^s}{1 - \sum_{j \in I_2} p_j^s} \} .$$

(iii) Continue this process as long as the value of $\Psi(H)$ in consecutive $2N$-point distributions improves and set I_2 is not empty.

3. Another possibility is to use approximation techniques to fit discrete distributions by continuous ones. For example, splines can be used when the dimensions are small or when the distributions H^s have independent components. This approach needs further study.

The convergence of algorithm 2 follows directly from Theorems 3 and 4. The next result may be derived using the remark following Theorem 3.

Theorem 5

Let the following conditions be satisfied:

1. X is a compact set.

2. Functions $f^i(x)$ are continuous for $x \in X$, $i = \overline{1,m}$.

3. $\Psi(H)$ is convex and satisfies (8) and

$$|f^0(x,H_1) - f^0(x,H_2)| \leq |\int_X \lambda(x,H_1,H_2)\, d(H_1 - H_2)|$$

for some summable $\lambda(x,H_1,H_2)$ such that

$$|\lambda(x,H_1,H_2)| < K < \infty$$

for any H_1 , $H_2 \in G$.

4. There exists an $\bar{H}(x)$ such that

$$\int_X f^i(x)dH(x) < 0 \ , \ i = \overline{1,m} \ .$$

Then

$$\lim_{s \to \infty} [\Psi(H^s) - \min_{H \in G} \Psi(H)] = 0 \ .$$

It is interesting to compare this algorithm with the methods for solving stochastic problems with recourse recently proposed by Wets [11]. Although applied to quite different problems, they both use generalized linear programming techniques and on each iteration require solution of one linear programming problem and one nonlinear optimization problem.

More complex problems with nonlinear constraints can be treated in the same way. If the definition of set G of constraints is changed to the following:

$$G=\{\Psi^i(H) \leq 0 \ , \ i=\overline{1,m} \ ; \ \int_X dH(x) = 1\}$$

and functions $\Psi^i(H)$ have directional derivatives of form (8) then analogues of Lemma 1 and Theorem 2 will hold . In this case we can construct a feasible-direction type algorithm, which inherits all of the important characteristics of algorithm 1. The same ideas can be used to solve minimax problems which

depend on distribution functions. This, together with the results of some numerical experiments, will provide the subject of a subsequent paper.

REFERENCES

1. Yu. Ermoliev, "Method for stochastic programming in randomized strategies," *Kibernetica* 1 (1970).

2. Yu. Ermoliev and C. Nedeva, "Stochastic optimization problems with partially known distribution functions", CP-82-60, International Institute for Applied Systems Analysis, Laxenburg, Austria (1982).

3. A. Golodnikov, "Optimization problems with distribution functions", Dissertation, V. Glushkov Institute for Cybernetics, Ukrainian Academy of Sciences, Kiev (1979).

4. P.J. Huber, *Robust Statistics* , John Wiley and Sons (1981).

5. M.G. Krein, "The ideas of P.L. Tchebysheff and A.A. Markov in the theory of limiting values of integrals and their further developments", *Am. Math. Soc. Series* 2, 12(1951), pp. 1-122. (translation).

6. A. Golodnikov, A. Gaivoronski and Le Ding Fung, "A method for solving limit extremum problems applied to the minimization of functionals", *Kiber-*

netica 1 (1980).

7. Yu. Ermoliev, A. Gaivoronski and C. Nedeva, "Stochastic optimization problems with incomplete information on distribution functions", WP-83-xx, International Institute for Applied Systems Analysis, Laxenburg, Austria (1983).

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

9. J.E. Kelley, "The cutting-plane method for solving convex programs", *J. Soc. Indust. Appl. Math.*, 8 (1960), pp. 703-712.

10. B. Curtis Eaves and W.I. Zangwill, "Generalized cutting plane algorithms", *SIAM J. Control*, 9(4), (1971).

11. R. J.-B. Wets, "Stochastic programming: solution techniques and approximation schemes". In *Mathematical Programming. The State of the Art* , Springer-Verlag (1983).