International Institute for
Applied Systems Analysis
IIASA www.iiasa.ac.at

Stochastic Optimization Problems with Incomplete Information on Distribution Functions

Ermoliev, Y.M., Gaivoronski, A.A. & Nedeva, C.

IIASA Working Paper

WP-83-113

November 1983

STOCHASTIC OPTIMIZATION PROBLEMS WITH INCOMPLETE INFORMATION ON DISTRIBUTION FUNCTIONS

Yu. Ermoliev
A. Gaivoronski
C. Nedeva

November 1983
WP-83-113

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 a new approach to optimization under uncertainty which tackles the problems caused by incomplete information about the probabilistic behavior of random variables. In contrast to the usual approach, which is to assume that probability distributions are known, the authors consider the more common case which arises when the information available is sufficient only to single out a set of possible distributions. The resulting optimization algorithms can be used in reliability analysis, mathematical statistics and many other fields.

ANDRZEJ WIERZBICKI
Chairman
System and Decision Sciences

ABSTRACT

The main purpose of this paper is to discuss numerical optimization procedures, based on duality theory, for problems in which the distribution function is only partially known. The dual problem is formulated as a minimax-type problem in which the "inner" problem of maximization is not concave. Numerical procedures that avoid the difficulties associated with solving the "inner" problem are proposed.

STOCHASTIC OPTIMIZATION PROBLEMS WITH INCOMPLETE INFORMATION ON DISTRIBUTION FUNCTIONS

Yu. Ermoliev, A. Gaivoronski and C. Nedeva

1. INTRODUCTION

A conventional stochastic programming problem may be formulated with some generality as minimization of the function

$$T(x) = E_y v(x,y) = \int v(x,y) \mathrm{d}H(y) \tag{1}$$

subject to

$$x \in X \subset R^n \quad , \tag{2}$$

where $y \in Y \subset R^m$ is a vector of random parameters, $H(y)$ is a given distribution function and $v(x, \cdot)$ is a random function possessing all the properties necessary for expression (1) to be meaningful [1].

In practice, we often do not have full information on $H(y)$; we sometimes only have some of its characteristics, in particular bounds for the mean value or other moments. Such information can often be written in terms of constraints

$$Q^k(H) = E_y q^k(y) = \int_Y q^k(y) \mathrm{d}H(y) \leq 0, \quad k = \overline{1,l} \tag{3}$$

$$\int_Y \mathrm{d}H(y) = 1 \quad , \tag{4}$$

where the $q^k(y)$, $k = \overline{1,l}$, are known functions. We could, for example, have the following constraints on joint moments:

$$c_{\tau_1,\tau_2,\dots,\tau_m} \leq Ey_1^{\tau_1} \cdots y_m^{\tau_m} \leq C_{\tau_1,\tau_2,\dots,\tau_m} \quad , \tag{5}$$

where $C_{\tau_1,\tau_2,\dots,\tau_m}$, $c_{\tau_1,\tau_2,\dots,\tau_m}$ are given constants.

Consider the following problem: find a vector x which minimizes

$$T(x) = \max_{H \in K} \int v(x,y) \mathrm{d}H(y) \quad , \tag{6}$$

subject to constraints (2), where K is the set of functions H satisfying constraints (3) and (4).

Special cases of this problem have been studied in [2], [3]. Under certain assumptions concerning the family K and the function $v(\cdot)$, the solution of the "inner" problem has a simple analytical form and hence (6) is reduced to a conventional nonlinear programming problem. The main purpose of this paper is to discuss numerical methods for the solution of problem (6) in the more general case. Sections 2, 3, and 4 deal with the reduction of this problem to minimax-type problems without randomized strategies and describe numerical methods based on some of the same ideas as generalized linear programming. A quite general method for solving the resulting minimax-type problems, in which the inner problem of maximization is not concave, is considered in Section 5.

2. OPTIMIZATION WITH RESPECT TO DISTRIBUTION FUNCTIONS

The possible methods of minimizing $T(x)$ depend on solution procedures for the following "inner" maximization problem: find a distribution function H that maximizes

$$Q^0(H) = Eq^0(y) = \int q^0(y)\mathrm{d}H(y) \tag{7}$$

subject to

$$Q^k(H) = Eq^k(y) = \int q^k(y)\mathrm{d}H(y) \le 0,\ k = \overline{1,l} \tag{8}$$

$$\int_Y \mathrm{d}H(y) = 1 \quad , \tag{9}$$

where q^v , $v = \overline{0,l}$, are given functions $R^m \to R^1$. This is a generalization of the known moments problem (see, for instance, [4–6]). It can also be regarded as a generalization of the nonlinear programming problem

$$\max \{q^0(y):q^k(y) \le 0 \, , \, y \in Y \, , \, k = \overline{1,l}\}$$

to an optimization problem involving randomized strategies [7–9].

It appears possible to solve problem (7)–(9) by means of a modification of the revised simplex method [8,10]. This modification is based on Krein's "geometrical approach" to the theory of moments [1,5,6]. Consider the set

$$Z = \{z : z = (q^0(y) \, , \, q^1(y), \ldots, q^l(y)) \, , \, y \in Y\}$$

and suppose that Z is compact. This will be true, for instance, if Y is compact and functions q^v , $v = \overline{0,l}$, are continuous. Consider also the convex hull of Z:

$$\mathrm{co}\, Z = \{z : z = \sum_{t=1}^{N} p_t z^t \, , \, z^t \in Z \, , \, \sum_{t=1}^{N} p_t = 1 \, , \, p_t \ge 0 \, , \, t = \overline{1,N}\} \quad ,$$

where N is an arbitrary finite number. Then general results from convex analysis lead to

$$\mathrm{co}\, Z = \{Q = (Q^0(H) \, , \, Q^1(H), \ldots, Q^l(H) \,|\, H \ge 0 \, , \, \int_Y \mathrm{d}H = 1\} \quad . \tag{10}$$

Therefore problem (7)–(9) is equivalent to maximizing z_0 subject to

$$z = (z_0, z_1, \ldots, z_l) \in \text{co } Z \, , \; z_k \le 0 \, , \; k = \overline{1,l} \quad .$$

According to the Caratheodory theorem, each point on the boundary of co Z can be represented as a convex combination of at most $l + 1$ points from Z:

$$\text{co } Z = \{ z : z_v = \sum_{j=1}^{l+1} q^v (y^j) p_j \, , \; v = \overline{0,l} \, , \; p_j \ge 0 \, , \; \sum_{j=1}^{l+1} p_j = 1 \, , \; y^j \in Y \} \quad .$$

Thus problem (7)–(9) is equivalent to the following generalized linear programming problem [11]: find points $y^j \in Y$, $j = \overline{1,t}$, $t \le l + 1$ and real numbers p_j, $j = \overline{1,t}$, such that

$$\sum_{j=1}^{t} q^0(y^j) p_j = \max \tag{11}$$

subject to

$$\sum_{j=1}^{t} q^k(y^j) p_j \le 0, \; k = \overline{1,l} \quad , \tag{12}$$

$$\sum_{j=1}^{t} p_j = 1 \, , \; p_j \ge 0 \, , \; j = \overline{1,t} \quad . \tag{13}$$

Consider arbitrary points $\bar{y}^j$, $j = \overline{1,l+1}$ (setting $t = l + 1$), and for the fixed set $\{\bar{y}^1, \bar{y}^2, \ldots, \bar{y}^{l+1}\}$ find a solution $\bar{p} = (\bar{p}_1, \bar{p}_2, \ldots, \bar{p}_{l+1})$ of problem (11)–(13) with respect to p. Assume that $\bar{p}$ exists and that $(\bar{u}_1, \bar{u}_2, \ldots, \bar{u}_{l+1})$ are the corresponding dual variables, i.e., solve the problem

$$\min u_{l+1} \tag{14}$$

subject to

$$q^0(\bar{y}^j) - \sum_{k=1}^{l} u_k q^k(\bar{y}^j) - u_{l+1} \le 0 \, , \; j = \overline{1,l+1} \tag{15}$$

$$u_k \ge 0 \, , \; k = \overline{1,l} \quad . \tag{16}$$

Now let y be an arbitrary point of Y. Consider the following augmented problem of maximization with respect to $(p_1,p_2,\ldots,p_{l+1},p)$: maximize

$$\sum_{j=1}^{l+1} q^0(\bar{y}^j)p_j + q^0(y)p \tag{17}$$

subject to

$$\sum_{j=1}^{l+1} q^k(\bar{y}^j)p_j + q^k(y)p \leq 0,\ k = \overline{1,l} \quad , \tag{18}$$

$$\sum_{j=1}^{l+1} p_j + p = 1 \quad . \tag{19}$$

It is clear that if there exists a point y^* such that

$$q^0(y^*) - \sum_{k=1}^{l} \bar{u}_k q^k(y^*) - \bar{u}_{l+1} > 0$$

then the solution $\bar{p}$ could be improved by dropping one of the columns $(q^0(\bar{y}^j)\,,\, q^1(\bar{y}^j)\,,\,\ldots\,,q^l(\bar{y}^j),\,1),\ j = \overline{1,l+1}$, from the basis and replacing it by the column $(q^0(y^*),q^1(y^*),\ldots,q^l(y^*),1),\ j = \overline{1,l+1}$, following the revised simplex method. Point y^* could be defined as

$$y^* = \arg\max_{y\in Y}\,[q^0(y) - \sum_{k=1}^{l} \bar{u}_k q^k(y)] \quad . \tag{20}$$

Then a new solution p^* of (11)–(13) with fixed $y = y^*$ can be determined in the same way as $\bar{p}$, together with the dual variables u^*. This method gives us a conceptual framework for solving not only (6) but also some more general classes of problems.

If $y(x) = (y^1(x)\,,\,y^2(x),\ldots,y^{l+1}(x))\,,\,p(x) = (p_1(x)\,,\,p_2(x),\ldots,p_{l+1}(x))$ is a solution of the inner optimization problem for fixed x, then the function (6) may be nondifferentiable with subgradient

$$T_x(x) = \sum_{j=1}^{l+1} v_x(x,y^j(x))p_j(x) \quad ,$$

where v_x is a subgradient of function $v(\cdot\,, y)$. Nondifferentiable optimization techniques could therefore be used to minimize $T(x)$. The main difficulty of such an approach would be to obtain a solution of (20) and exact values of $y(x)$, $p(x)$ at each current point x^s for iterations $s = 0,1,...$ This last difficulty can sometimes be avoided by dealing with approximate solutions rather than precise values $y(x)$, $p(x)$, and using ε-subgradient methods (see [12],[13]). Generalized linear programming methods which do not require exact solutions of subproblem (20) are studied in Section 4.

3. DUALITY RELATIONS

The duality relations for problem (7)–(9) enable us to find a more general approach to the solution of problem (6). Consider the following problem:

$$\min_{u \in U^+} \max_{y \in Y} [q^0(y) - \sum_{k=1}^{l} u_k q^k(y)] \quad , \tag{21}$$

where

$$U^+ = \{u : u = (u_1, u_2, \ldots, u_l) \, , \, u_i \geq 0 \, , \, i = \overline{1,m}\} \quad .$$

This problem can be regarded as dual to (7)–(9) or (11)–(13), but to explain this we must introduce some more definitions.

In what follows we shall use the same letter, say H, for both the distribution function and the underlying probabilistic measure, where this will not cause confusion. We shall denote by $Y^+(H)$ the collection of all subsets of Y which have positive measure H, and by dom H the support set of distribution H, i.e.,

$$\operatorname{dom} H = \bigcap_{A \in Y^+(H)} A \quad .$$

Set

$$\psi(u) = \max_{y \in Y} [q^0(y) - \sum_{k=1}^{l} u_k q^k(y)]$$

$$U^* = \{ u^* : u^* \in U^+ , \psi(u^*) = \min_{u \in U^+} \psi(u) \}$$

$$Y(u) = \{ y : y \in Y , \psi(u) = q^0(y) - \sum_{k=1}^{l} u_k q^k(y) \} \quad .$$

Then the following generalization of the results given in [14] holds.

Theorem 1. *Assume that*

1. Y is compact and $q^v(y)$, $v = \overline{0,l}$, are continuous.

2. $\operatorname{int} \operatorname{co} Z \neq \phi$

Then

1. Solutions to both problem (7)–(9) and problem (21) exist, and the optimal values of the objective functions of both problems are equal.

2. For any solution H^* of problem (7)–(9) there exists a $u^* \in U^*$ such that

$$\operatorname{dom} H^* \subseteq Y(u^*) \quad .$$

In other words, the duality gap vanishes in nonlinear programs with randomized strategies. A proof of this theorem can be derived from general duality results [12,15] and the theory of moments [5]. The proof given below is close to the ideas expressed in [12] and illustrates certain connections with results from conventional nonlinear programming.

Proof. From (10), problem (7)–(9) is equivalent to

$$\max \{ z_0 : z = (z_1, z_2, \ldots, z_l) \in \operatorname{co} Z , z_k \leq 0 , k = \overline{1,l} \} \quad , \tag{22}$$

where $Z = \{ z : z = (q^0(y), q^1(y), \ldots, q^l(y)) , y \in Y \}$. From assumption 1 of Theorem 1, $\operatorname{co} Z$ is a convex compact set and therefore a solution $z^* = (z_0^*, z_1^*, \ldots, z_l^*)$ to problem (22) exists. Let $L(u,z)$ be a Lagrange function for (22):

$$L(u,z) = z_0 - \sum_{k=1}^{l} u_k z_k \quad .$$

From assumption 2,

$$z_0^* = \max_{z \in \mathrm{co}\, Z} \min_{u \in U^+} L(u,z) = \min_{u \in U^+} \max_{z \in \mathrm{co}\, Z} L(u,z) \quad .$$

According to (10), there exists for any $z \in \mathrm{co}\ Z$ a distribution H such that

$$z_v = \int_Y q^v(y)\mathrm{d}H(y) \,,\; \int_Y \mathrm{d}H(y) = 1 \,,\; v = \overline{0,l} \quad .$$

We therefore have

$$L(u,z) = \bar{L}(u,H) = \int_Y [q^0(y) - \sum_{k=1}^{l} u_k q^k(y)]\mathrm{d}H(y)$$

and

$$\max_{z \in \mathrm{co}\, Z} L(u,z) = \max \{\bar{L}(u,H) \mid H \geq 0 \,,\; \int_Y \mathrm{d}H(y) = 1\} \quad .$$

Obviously

$$\max \{\int_Y [q^0(y) - \sum_{k=1}^{l} u_k q^k(y)]\mathrm{d}H(y) \mid H \geq 0 \,,\; \int_Y \mathrm{d}H(y) = 1\}$$

$$= \max_{y \in Y} [q^0(y) - \sum_{k=1}^{l} u_k q^k(y)] \quad ,$$

which proves the first part of the theorem.

Under the assumptions of the theorem we know that for any solution $(z_0^*, \ldots, z_l^*)$ there exists a $u \in U^*$ such that

$$z_0^* = \max_{z \in \mathrm{co}\, Z} L(u^*,z) \quad .$$

Thus, for any optimal distribution H^* we have

$$\int_Y q^0(y)\mathrm{d}H^*(y) = \max \{\int_Y [q^0(y) - \sum_{k=1}^{l} u_k^* q^k(y)]\mathrm{d}H(y) \mid H \geq 0 \,,\; \int_Y \mathrm{d}H(y) = 1 \} =$$
$$\max_{y \in Y} [q^0(y) - \sum_{k=1}^{l} u_k^* q^k(y)]$$

which proves the second part of the theorem.

Remark 1. From the duality theorem above we have

$$\max_{H \in K} \int v(x,y) \mathrm{d}H(y) = \min_{u \in U^+} \max_{y \in Y} [v(x,y) - \sum_{k=1}^{m} u_k q^k(y)] \quad .$$

for each fixed $x \in X$, where $v(x,\cdot)$ is a continuous function. Problem (6) can then be reduced to a minimax-type problem as follows: minimize the function

$$\gamma(x,u) = \max_{y \in Y} [v(x,y) - \sum_{k=1}^{m} u_k q^k(y)]$$

with respect to $x \in X$, $u \geq 0$.

Remark 2. Theorem 1 can be used to characterize optimal distributions for a variety of nonlinear optimization problems with distribution functions. The approach is, first, to state necessary optimality conditions through linearization and then to apply Theorem 1. This is illustrated in the following example.

Consider the optimization problem

$$\max g^0(H) \tag{7a}$$

$$g^i(H) \leq 0 \quad , \quad i = \overline{1,m} \tag{8a}$$

$$\int_Y \mathrm{d}H(y) = 1 \quad , \tag{9a}$$

where $g^i(H)$, $i = \overline{1,m}$, are nonlinear functionals depending on distribution functions H with support set Y.

Theorem 1a. *Assume that the following statements are true:*

1. Set Y is a compact subset of Euclidean space R^n.

2. For any distributions $H_1 = H_2$ such that $\mathrm{dom}\, H_1 \subseteq Y_1$, $\mathrm{dom}\, H_2 \subseteq Y_2$ *we have*

$$g^i(H_1 + \alpha(H_2 - H_1)) = g^i(H_1) + \alpha \int_Y q^i(y, H_1) \mathrm{d}(H_2 - H_1) + \varepsilon(\alpha, H_1, H_2) \quad ,$$

where $i = \overline{1,m}$, $\alpha \in [0,1]$ *and* $\dfrac{\varepsilon(\alpha, H_1, H_2)}{\alpha} \to 0$ *as* $\alpha \to 0$.

3. Functions $q^i(y,H)$ *are continuous in* y *for every* H *such that* dom $H \subseteq Y$ *; for any* H_1 , H_2 *such that* dom $H_1 \subseteq Y$, dom $H_2 \subseteq Y$ *we have*

$$|q^i(y,H_1) - q^i(y,H_2)| \le |\int_Y \lambda_i(y,H_1,H_2)\mathrm{d}(H_1 - H_2)| \quad ,$$

where $|\lambda_i(y,H_1,H_2)| \le K < \infty$ *for some* K *which does not depend on* H_1 , H_2 .

4. Functions $g^i(H)$, $i = \overline{1,m}$, *are convex, i.e.,*

$$g^i(\alpha_1 H_1 + \alpha_2 H_2) \le \alpha_1 g(H_1) + \alpha_2 g(H_2) \quad ,$$
$$\alpha_1 \ge 0 \ , \ \alpha_2 \ge 0 \ , \ \alpha_1 + \alpha_2 = 1 \quad .$$

5. There exists an $\overline{H}$ *such that* dom $\overline{H} \subseteq Y$ *and* $g^i(\overline{H}) < 0$ *for* $i = \overline{1,m}$.

Then :

1. A solution of problem (7a)–(9a) exists.

2. For any such solution H^* *we have*

$$\int_Y q^0(y,H^*)\mathrm{d}H^* = \min_{u \in U^+} \varphi(u,H^*) \quad ,$$

where

$$\varphi(u,H^*) = \max_{y \in Y} [\, q^0(y,H^*) - \sum_{i=1}^m u_i q^i(y,H^*)] \quad .$$

3. If H^* *is a solution of (7a)–(9a) then for some* u^* *we have* dom $H^* \subseteq Y^*(u^*,H^*)$, *where*

$$\varphi(u^*,H^*) = \min_{u \in U^+} \varphi(u,H^*)$$
$$Y^*(u,H) = \{\, y : y \in Y \, , \, \varphi(u,H) = q^0(y,H) - \sum_{i=1}^m u_i q^i(y,H)\} \quad .$$

Thus, the main assumptions of this theorem are the existence, continuity (in some sense) and boundedness of the directional derivatives of functions $g^i(H)$.

The following theorem is analogous to known results in linear programming and provides a useful stopping rule for methods of the type described in Section 2 (see also Section 4).

Theorem 2. *(Optimality condition) Let the assumptions of Theorem 1 hold and let $\bar{p}$ be a solution of problem (11)–(13) for fixed $\bar{y} = (\bar{y}^1, \bar{y}^2, \ldots, \bar{y}^t)$, $\bar{y} \in R^{t\times m}$. Then the pair $\bar{y}, \bar{p}$ is an optimal solution of problem (11)–(13) if and only if for given $\bar{y}$ there exists a solution $(\bar{u}_1, \bar{u}_2, \ldots, \bar{u}_{l+1})$ of problem (14)–(16) such that*

$$q^0(y) - \sum_{k=1}^{l} \bar{u}_k q^k(y) - \bar{u}_{l+1} \le 0 \text{ for all } y \in Y \quad .$$

Proof

1. Suppose that $\bar{y}, \bar{p}$ is an optimal solution of problem (11)–(13), that $(u_1^*, u_2^*, \ldots, u_{l+1}^*)$ is a solution of problem (14)–(16) for given $\bar{y}$, and that

$$\bar{u}_{l+1} = \min_{u \in U^+} \max_{y \in Y} [q^0(y) - \sum_{k=1}^{l} u_k q^k(y)] = \max_{y \in Y} [q^0(y) - \sum_{k=1}^{l} \bar{u}_k q^k(y)] \quad .$$

We shall show that $\bar{u}_1, \bar{u}_2, \ldots, \bar{u}_{l+1}$ is a solution of problem (14)–(16). Consider the two functions:

$$\psi(u) = \max_{y \in Y} [q^0(y) - \sum_{k=1}^{l} u_k q^k(y)] \quad ,$$

$$\psi_1(u) = \max_{1 \le j \le t} [q^0(\bar{y}^j) - \sum_{k=1}^{l} u_k q^k(\bar{y})] \quad .$$

According to Theorem 1

$$\sum_{j=1}^{t} q^0(\bar{y}^j)\bar{p}_j = \min_{u \in U^+} \psi(u) \quad .$$

Since problem (11)–(13) is dual to problem (14)–(16) for given $\bar{y}$, then

$$\sum_{j=1}^{t} q^0(\bar{y}^j)\bar{p}_j = \min_{u \in U^+} \psi_1(u) \quad .$$

Therefore

$$\psi(\bar{u}) = \min_{u \in U^+} \psi(u) = \min_{u \in U^+} \psi_1(u) = \psi_1(u^*) = u^*_{l+1} \quad , \tag{23}$$

where

$$u^* = (u^*_1, u^*_2, \ldots, u^*_l) \quad .$$

Since $\bar{y}^j \in Y$, $j = \overline{1,t}$, then $\psi_1(u) \leq \psi(u)$ for $u \in U^+$. In particular, $\psi_1(\bar{u}) \leq \psi(\bar{u})$. But (23) implies

$$\psi_1(\bar{u}) \geq \min_{u \in U^+} \psi_1(u) = \psi(\bar{u}) \quad ,$$

and this gives $\psi_1(\bar{u}) = \psi(\bar{u}) = \psi_1(u^*)$. Hence $(\bar{u}_1, \bar{u}_2, \ldots, \bar{u}_{l+1})$ is a solution of problem (14)–(16).

2. Suppose now that for given $\bar{y}$ there exists a solution $(\bar{u}_1, \bar{u}_2, \ldots, \bar{u}_{l+1})$ of problem (14)–(16) such that

$$q^0(y) - \sum_{k=1}^{l} \bar{u}_k q^k(y) - \bar{u}_{l+1} \leq 0 \, , \, y \in Y \quad .$$

From the duality between problems (11)–(13) and (14)–(16) we have

$$\sum_{j=1}^{t} q^0(\bar{y}^j)\bar{p}_j \geq \psi(\bar{u}) \quad ,$$

where $\bar{p} = (\bar{p}_1, \bar{p}_2, \ldots, \bar{p}_t)$ is a solution of problem (11)–(13) for given $\bar{y}$. On the other hand, the duality between problems (21) and (11)–(13) leads to the inequality

$$\sum_{j=1}^{t} q^0(y^j)p_j \leq \psi(\bar{u}) \quad ,$$

for any $\{y^1, y^2, \ldots, y^t\}, p$ satisfying (12)–(13). In other words,

$$\sum_{j=1}^{t} q^0(\bar{y}^j)\bar{p}_j \geq \sum_{j=1}^{t} q^0(y^j)p_j$$

and this completes the proof.

The next theorem provides a means of deriving a solution to the initial problem (7)–(9) from a solution of problem (21), and is complementary to Theorem 1.

Theorem 3. *Assume that the assumptions of Theorem 1 are satisfied and that* $\psi(\bar{u}) = \min\{\psi(u) \mid u \in U^+\}$. *Let* $\bar{y} = (\bar{y}^1, \bar{y}^2, \ldots, \bar{y}^t)$, *where* $\bar{y} \in R^{t \times m}$ *and* $\bar{y}^i \in Y(\bar{u})$, *and let* $\bar{p}$ *be a solution of problem (11)–(13) for given* $\bar{y}$. *Suppose also that there is a solution* p^* *to the inequalities (12)–(13) for* $y^j = \bar{y}^j$ *such that*

$$\sum_{j=1}^{t} q^k(\bar{y}^j) p_j^* \leq 0 \, , \; k \in I_0 \tag{24}$$

$$\sum_{j=1}^{t} q^k(\bar{y}^j) p_j^* = 0 \, , \; k \in I_+ \quad , \tag{25}$$

where $I_+ = \{k \mid \bar{u}_k > 0\}$, $I_0 = \{k \mid \bar{u}_k = 0\}$.

Then the pair $\bar{y}, \bar{p}$ *is an optimal solution of problem (11)–(13).*

Proof. The vectors

$$q(\bar{y}^j) = (-q^1(\bar{y}^j) \, , -q^2(\bar{y}^j), \ldots, -q^l(\bar{y}^j)) \, , \; j = \overline{1,t}$$

are subgradients of the convex function

$$\psi_1(u) = \max_{1 \leq j \leq t} [q^0(\bar{y}^j) - \sum_{k=1}^{l} u_k q^k(\bar{y}^j)]$$

at a point $\bar{u}$. Therefore condition (25) is necessary and sufficient for point $\bar{u}$ to be an optimal solution, i.e., so that

$$\psi_1(\bar{u}) = \min\{\psi_1(u) \mid u \in U^+\} \quad .$$

Then, from (24),

$$\min\{\psi_1(u) \mid u \in U^+\} = \psi(\bar{u}) = \min\{\psi(u) \mid u \in U^+\} \quad . \tag{26}$$

The minimization of $\psi_1(u)$, $u \in U^+$, is equivalent to problem (14)–(16). Hence $\bar{u} = (\bar{u}_1, \bar{u}_2, \ldots, \bar{u}_l)$ together with $\bar{u}_{l+1} = \psi_1(\bar{u})$ give a solution of problem (14)–(16). Since problem (14)–(16) is dual to problem (11)–(13), then problem (11)–(13) has a solution, say $\bar{p} = (\bar{p}_1, \bar{p}_2, \ldots, \bar{p}_t)$, and

$$\bar{u}_{l+1} = \sum_{j=1}^{t} q^0(\bar{y}^j)\bar{p}_j \quad .$$

This together with (26) yields

$$\sum_{j=1}^{t} q^0(\bar{y}^j)\bar{p}_j = \min\{\psi(u) \,|\, u \in U^+\} \quad ,$$

and this completes the proof.

4. ALGORITHMS

Theorems 2 and 3 justify a dual approach to problem (7)–(9) which may involve simultaneous approximation of both primal and dual variables subject to (24)–(25). In this section we consider several versions of generalized-linear-programming-based method discussed briefly in Section 2. In all cases the current estimate of optimal solution satisfies (24)–(25) at each iteration. The convergence of such algorithms has been investigated in a number of papers [16], [11] , under the assumption that the initial column entries for all previous iterations of subproblem (24) and the exact solutions at each iteration are stored in the memory. There are various ways of avoiding this expansion of the memory, mainly through selective deletion of these columns [17–19]. The aim of this section is to discuss a way of avoiding not only the expansion of the memory, but also the need to have a precise solution of (20). The last is important in connection with initial problem (6), as mentioned in Section 2.

Description of Algorithm 1

Fix points $y^{0,1}, y^{0,2}, \ldots, y^{0,l+1}$ and solve problem (11)–(13) with respect to p for $y^j = y^{0,j}$, $j = \overline{1, l+1}$. Suppose that a solution $p^0 = (p_1^0, p_2^0, \ldots, p_{l+1}^0)$ to this problem exists. Let $u^0 = (u_1^0, u_2^0, \ldots, u_{l+1}^0)$ be a solution of the dual problem (14)–(16) with respect to u. The vector u^0 satisfies the following constraints for $y \in \{y^{0,1}, y^{0,2}, \ldots, y^{0,l+1}\}$:

$$q^0(y) - \sum_{k=1}^{l} u_k^0 q^k(y) - u_{l+1} \leq 0 \, , \, u_k^0 \geq 0 \quad . \tag{27}$$

If u^0 satisfies condition (27) for all $y \in Y$, then the pair $\{y^{0,1}, y^{0,2}, \ldots, y^{0,l+1}\}$, p^0 is a solution of the original problem (11)–(13). If this is not the case, consider a new point y^0 such that

$$\Delta(y^0, u^0) = q^0(y^0) - \sum_{k=1}^{l} u_k^0 q^k(y^0) - u_{l+1}^0 > 0$$

and

$$q^0(y^0) - \sum_{k=1}^{l} u_k^0 q^k(y^0) \geq \max_{y \in Y} [q^0(y) - \sum_{k=1}^{l} u_k^0 q^k(y)] - \varepsilon_0$$

for some $\varepsilon_0 > 0$.

Denote by $p^1 = (p_1^1, p_2^1, \ldots, p_{l+1}^1)$ a solution of the augmented problem (17)–(19) with respect to p for fixed $\bar{y}^j = y^{0,j}$, $y = y^0$. We shall use $y^{1,1}, y^{1,2}, \ldots, y^{1,l+1}$ to denote those points $y^{0,1}, \ldots, y^{0,l+1}, y^0$ that correspond to the basic variables of solution p^1.

Thus, the first step of the algorithm is terminated and we pass to the next step: determination of u^1, y^1, etc. In general, after the s-th iteration we have points $y^{s,1}, y^{s,2}, \ldots, y^{s,l+1}$, a solution $p^s = (p_1^s, p_2^s, \ldots, p_{l+1}^s)$ and the corresponding solution $u^s = (u_1^s, u_2^s, \ldots, u_{l+1}^s)$ to the dual problem (14)–(16). For an $\varepsilon_s > 0$, find y^s such that

$$\Delta(y^s, u^s) = q^0(y^s) - \sum_{k=1}^{l} u_k^s q^k(y^s) - u_{l+1}^s > 0$$

and

$$q^0(y^s) - \sum_{k=1}^{l} u_k^s q^k(y^s) \geq \max_{y \in Y}[q^0(y) - \sum_{k=1}^{l} u_k^s q^k(y)] - \varepsilon_s \quad .$$

If we do not obtain $\Delta(y^s, u^s) > 0$ for decreasing values of ε_s we arrive at an optimal solution; otherwise we have to solve the augmented problem (17)–(19) for $\bar{y}^j = y^{s,j}$, $y = y^s$.

Denote by $y^{s+1,1}, y^{s+1,2}, \ldots, y^{s+1,l+1}$ those points from $\{y^{s,1}, y^{s,2}, \ldots, y^{s,l+1}\} \cup y^0$ that correspond to the basic variables of the solution p^{s+1}. The pair $\{y^{s+1,1}, y^{s+1,2}, \ldots, y^{s+1,l+1}\}$, p^{s+1} is the new approximate solution to the original problem, and so if $\Delta(y^s, u^s) \leq 0$, then (according to Theorem 2) the pair $\{y^{s,1}, y^{s,2}, \ldots, y^{s,l+1}\}$, p^s is the desired solution. Define

$$I_s^0 = \{k : u_k^s = 0\} \ , \ I_s^1 = \{k : u_k^s > 0\}$$

and

$$A_s = \{e : e = (e_1, e_2, \ldots, e_l) \ , \ ||e|| = 1 \ , \ e_k \geq 0 \ \text{for} \ k \in I_s^0 \ \text{and}$$
$$\text{arbitrary} \ e_k \ \text{ for } \ k \in I_s^1\} \quad .$$

A_s is actually a set of feasible directions for set U^+ at point u^s. Let

$$\gamma_s = \max_{e \in A_s} \min_{j : p_j^s > 0} \sum_{k=1}^{l} q^k(y^{s,j}) e_k \quad .$$

Note that γ_s is always less than zero because

$$\text{co} \{(q^1(y^{s,j}), q^2(y^{s,j}), \ldots, q^l(y^{s,j})) \ , \ \forall j : p_j^s > 0\}$$

is a set of subgradients of the function

$$\max_{j : p_j^s > 0} \{q^0(y^{s,j}) - \sum_{k=1}^{l} u_k q^k(y^{s,j})\}$$

at point u^s, and this function has a minimum at u^s.

In order to prove that this method is convergent we require, broadly speaking, that $\gamma_s < 0$ and tends to zero only as we approach the optimal solution.

Consider the functions

$$\psi^s(u) = \max_{1 \le j \le l+1} [q^0(y^{s,j}) - \sum_{k=1}^{l} u_k q^k(y^{s,j})]$$

Theorem 4. *Let the conditions of Theorem 1 be satisfied, and the following additional conditions hold:*

1. There exists a nondecreasing function $\tau(t)$, $t \in [0,\infty)$, $\tau(0) = 0$, $\tau(t) > 0$ *for* $t > 0$, *and*

$$\gamma_s \le -\tau(\psi(u^s) - \psi^s(u^s)) \quad . \tag{28}$$

2. $\varepsilon_s > 0$, $\varepsilon_s \to 0$ for $s \to \infty$.

Then any convergent subsequence of sequence $\{y^{s,1}, y^{s,2}, \dots, y^{s,l+1}\}$, p^s *converges to a solution of problem (7)–(9).*

Proof

1. First let us prove that the sequence $\{u^s\}$ is bounded. Suppose, arguing by contradiction, that there exists a subsequence $\{u^{s_r}\}$ such that $||u^{s_r}|| \to \infty$ as $r \to \infty$. Assumption 2 of Theorem 1 implies that $\psi(u^{s_r}) \to \infty$ and therefore that $\psi(u^{s_r}) - \psi^{s_r}(u^{s_r}) \to \infty$, since $\psi^{s_r}(u^{s_r}) \le \min_{u \in U^+} \psi(u)$. Hence, there exist $\bar{r}$ and $\delta > 0$ such that for $r > \bar{r}$,

$$\gamma_{s_r} \le -\delta \quad .$$

Now let us fix an arbitrary point $\bar{u} \in U^+$ and estimate $\psi(\bar{u})$. We obtain

$$\psi(\bar{u}) \ge \psi^{s_r}(\bar{u}) \ge \psi^{s_r}(u^{s_r}) + \sup_{g \in G^{s_r}} (g \, , \bar{u} - u^{s_r}) \quad ,$$

where G^{s_r} is a set of subgradients of function ψ^{s_r} at point u^{s_r}. The definition of ψ^s implies that

$$G_{s_r} \supseteq \text{co}\,\{-q^1(y^{s_r,i}), -q^2(y^{s_r,i}), \ldots, -q^l(y^{s_r,i})\,,\ \forall i{:}p_i^{s_r} > 0\} = \tilde{G}_{s_r}$$

and therefore (28) leads to

$$\psi(\bar{u}) \geq \psi^{s_r}(u^{s_r}) + \sup_{g \in \tilde{G}_{s_r}} (g\ ,\bar{u} - u^{s_r}) \geq$$

$$\geq \psi^{s_r}(u^{s_r}) + ||\bar{u} - u^{s_r}||\min_{e \in A_{s_r}} \max_{i:p_i^{s_r} > 0} (-\sum_{k=1}^{l} e_k q^k(y^{s_r,i}))$$

$$= \psi^{s_r}(u^{s_r}) - ||\bar{u} - u^{s_r}||\max_{e \in A_{s_r}} \min_{i:p_i^{s_r} > 0} \sum_{k=1}^{l} e_k q^k(y^{s_r,i})$$

$$\geq \psi^{s_r}(u^{s_r}) + \delta||\bar{u} - u^{s_r}|| \quad .$$

This last inequality yields $\psi(\bar{u}) = \infty$ if $||u^{s_r}|| \to \infty$, and therefore sequence $\{u^s\}$ is bounded.

2. We shall now estimate the evolution of the quantity $w_s = \psi^s(u^s)$, where

$$u^s = \arg\min_{u \in U^+} \psi^s(u) \quad .$$

Using the same argument as in part 1 of the proof we obtain:

$$\begin{aligned}
w_{s+1} &= \psi^{s+1}(u^{s+1}) = \psi^s(u^{s+1}) \\
&\geq \psi^s(u^s) + \sup_{g \in \tilde{G}_s}(g\ , u^{s+1} - u^s) \\
&\geq \psi^s(u^s) + ||u^{s+1} - u^s||\min_{e \in A_s} \max_{i:p_i^s > 0}(-\sum_{k=1}^{l} e_k q^k(y^{s,i})) \\
&= \psi^s(u^s) - ||u^{s+1} - u^s||\max_{e \in A_s} \min_{i:p_i^s > 0} \sum_{k=1}^{l} e_k q^k(y^{s,i}) \\
&\geq \psi^s(u^s) + \tau(\psi(u^s) - \psi^s(u^s))||u^{s+1} - u^s|| \quad .
\end{aligned}$$

Sequence $\{u^s\}_{s=0}^{\infty}$ is bounded and so $\psi(u^s) = \sup_{y \in Y}(q^0(y) - \sum_{i=1}^{l} u_i^s q^i(y))$ must also be bounded; thus $\psi^s(u^s)$ is bounded since $\psi^s(u^s) \leq \psi(u^s)$. This together with the previous inequality immediately gives

$$\min \{\psi(u^s) - \psi^s(u^s) \, , \, ||u^{s+1} - u^s||\} \to 0 \quad . \tag{29}$$

Now consider any convergent subsequence $\{u^{s_r}\}$ of sequence $\{u^s\}$. We can assume from (29) that either $||u^{s_r} - u^{s_r+1}|| \to 0$ or $\psi(u^{s_r}) - \psi^{s_r}(u^{s_r}) \to 0$. In the latter case we get $\psi^{s_r}(u^{s_r}) \to \min_{u \in U^+} \psi(u) = \psi^*$ because $\psi(u^{s_r}) \geq \psi^*$ and $\psi^{s_r}(u^{s_r}) \leq \psi^*$. In the case $||u^{s_r} - u^{s_r+1}|| \to 0$ we get the following:

$$\psi(u^{s_r}) - \psi^{s_r}(u^{s_r}) = \psi(u^{s_r}) - \psi^{s_r+1}(u^{s_r}) + \psi^{s_r+1}(u^{s_r})$$

$$- \psi^{s_r+1}(u^{s_r+1}) + \psi^{s_r+1}(u^{s_r+1}) - \psi^{s_r+1}(u^{s_r})$$

$$\leq \varepsilon_{s_r} + ||u^{s_r} - u^{s_r+1}|| \sup_{g \in G_{s_r+1}} ||g||$$

so that once again $\psi(u^{s_r}) - \psi^{s_r}(u^{s_r}) \to 0$ and we obtain $\psi^{s_r}(u^{s_r}) \to \min_{u \in U^+} \psi(u)$. However, according to Theorem 1 $\min_{u \in U^+} \psi(u)$ is the optimal solution of the initial problem; $\min_{u \in U^+} \psi^s(u)$ is the optimal solution of problem (11)–(13). Therefore the solution of (11)–(13) tends to the solution of the initial problem, and any convergent subsequence of sequence $\{y^{s,1}, y^{s,2}, \ldots, y^{s,l+1}\}$, p^s , where $s = 0,1,\ldots$ converges to the optimal solution of the initial problem.

This method can be viewed as the dual of a cutting-plane method applied to problem (7)–(9) [12,16,20]. It drops all points $y^{s,i}$ which do not correspond to basic variables. Theorem 4 shows that in some (rare) cases this method does not converge; however, this is not surprising because in certain cases the simplex method does not converge either. It may be possible to modify the algorithm in different ways to ensure convergence.

If we keep all previous points $y^{0,1}, y^{0,2}, \ldots, y^{0,l+1}, y^0, y^1, \ldots$ and solve problem (14)–(16) with an increasing number of corresponding columns, then the method appears to be a form of Kelley's method for minimizing function $\psi(u)$, which converges under the assumptions of Theorem 1. However, it is impossible to allow the set of points to increase *ad infinitum* in practical computations.

In the following modification of the algorithm presented above some non-basic columns are dropped when an additional inequality is satisfied.

Description of Algorithm 2

1. We first choose a sequence of positive real numbers $\{\mu_s\}_{s=0}^{\infty}$, take $r_0 = 0$ and select initial points $\{y^{0,1}, y^{0,2}, \ldots, y^{0,l+1}\}$ such that problem (11)–(13) has a solution with respect to p for $y^j = y^{0,j}$, $j = \overline{1,l+1}$. Let p^0 be a solution of this problem and u^0 be the corresponding dual variables. We then have to find y^0 such that

$$q^0(y^0) - \sum_{k=1}^{l} u_k^0 q^k(y^0) \geq \psi(u^0) - \varepsilon_0 \quad ,$$

where ε_0 is a positive number. If for any ε_0 and the corresponding y^0 we have

$$\Delta(y^0, u^0) = q^0(y^0) - \sum_{k=1}^{l} u_k^0 q^k(y^0) - u_{l+1}^0 \leq 0 \quad ,$$

then the pair $\{y^{0,1}, y^{0,2}, \ldots, y^{0,l+1}\}$, p^0 is an optimal solution of problem (7)–(9). Otherwise it is necessary to select ε_0, y^0 such that $\Delta(y^0, u^0) > 0$ and take $\Delta_0 = \Delta(y^0, u^0)$.

Suppose that after the s-th iteration we have points $y^{s,j}$, $j = \overline{1,l_s}$, a solution $p^s = (p_1^s, p_2^s, \ldots, p_{l_s}^s)$ of problem (11)–(13) for $y^j = y^{s,j}$, $j = \overline{1,l_s}$, $t = l_s$, a corresponding solution $u^s = (u_1^s, u_2^s, \ldots, u_{l+1}^s)$ of the dual problem (14)–(16), a positive integer number r_s and a positive number Δ_s.

2. Find an approximate solution y_s such that

$$(q^0(y^s) - \sum_{k=1}^{l} u_k^s q^k(y)) > \psi(u^s) - \varepsilon_s$$

and

$$q^0(y^s) - \sum_{k=1}^{l} u_k^s q^k(y^s) = \Delta(y^s, u^s) > 0 \quad .$$

If this is not possible then we have arrived at a solution. Otherwise consider the following two cases:

(a) $\Delta(y^s, u^s) \le (1 - \mu_{r_s})\, \Delta_s$

In this case take $\Delta_{s+1} = \Delta(y^s, u^s)$, $l_{s+1} = l + 1$, $r_{s+1} = r_s + 1$ and denote by $y^{s+1,1}, y^{s+1,2}, \ldots, y^{s+1,l+1}$ those points from $\{y^{s,1}, y^{s,2}, \ldots, y^{s,l+1}\} \cup y^s$ that correspond to the basic variables of the solution p^{s+1}.

(b) $\Delta(y^s, u^s) > (1 - \mu_{r_s})\Delta_s$

In this case take

$\Delta_{s+1} = \Delta_s$, $l_{s+1} = l_s + 1$, $r_{s+1} = r_s$, $y^{s+1,j} = y^{s,j}$, $j = \overline{1, l_s}$, $y^{s+1, l_{s+1}} = y^s$.

Find a solution of problem (11)–(13) for $t = l_{s+1}$, $y^j = y^{s+1,j}$, $j = \overline{1, l_{s+1}}$ and the corresponding dual variables u^{s+1}, and proceed to the next iteration.

Theorem 5. *Suppose that the conditions of Theorem 1 are satisfied and the following additional conditions hold:*

1. $\varepsilon_s > 0$, $\varepsilon_s \to 0$, $\sum_{s=0}^{\infty} \mu_s = \infty$, $\mu_s \ge 0$

2. $\varepsilon_s / \mu_s \to 0$

Then $\sum_{i=1}^{l_s} p_i^s q_0(y^{s,i}) = u_s$ *tends to the optimal value of problem (7)–(9).*

Proof

1. Suppose that the inequality $\Delta(y^s, u^s) \le (1 - \mu_{T_s})\Delta_s$ is satisfied only a finite number of times. This implies the existence of a number s_0 such that for $s \ge s_0$ the method turns into Kelley's cutting-plane algorithm for minimization of the convex function $\psi(u)$, where the values of $\psi(u)$ are calculated with an error that tends to zero. From assumption 2 of Theorem 1, $\psi(u)$ has a bounded set of minimum points and the initial approximating function $\psi^{s_0}(u)$ has a minimum. Thus such a method would converge to the minimum of function $\psi(u)$ and $\Delta(y^s, u^s) \le \psi(u^s) - \psi^s(u^s) \to 0$, which contradicts the assumption that for $s \ge s_0$ the inequality $\Delta(y^s, u^s) \le (1 - \mu_{T_s})\Delta_s$ is not satisfied.

2. There exists a subsequence s_k such that

$$\Delta(y^{s_k}, u^{s_k}) \le (1 - \mu_k)\Delta_{s_k} = (1 - \mu_k)\Delta(y^{s_{k-1}}, u^{s_{k-1}}) \quad .$$

From the definition of the algorithm we have

$$\psi(u^{s_k}) - \psi^{s_k}(u^{s_k}) - \varepsilon_{s_k} \le \Delta(y^{s_k}, u^{s_k}) \le \psi(u^{s_k}) - \psi^{s_k}(u^{s_k})$$

and therefore

$$\psi(u^{s_k}) - \psi^{s_k}(u^{s_k}) - \varepsilon_{s_k} \le (1 - \mu_k)(\psi(u^{s_{k-1}}) - \psi^{s_{k-1}}(u^{s_{k-1}})) \quad .$$

Making the substitution $\psi(u^s) - \psi^s(u^s) = w^s$ we obtain

$$w^{s_k} \le (1 - \mu_k)w^{s_{k-1}} + \varepsilon_{s_k} \quad .$$

This inequality together with the assumption $\varepsilon_{s_k}/\mu_k \to 0$ gives $w^{s_k} \to 0$ and therefore $\psi^{s_k}(u^{s_k}) \to \min_{u \in U^+} \psi(u)$ because $\psi^{s_k}(u^{s_k}) \le \psi(u)\ \forall u \in U^+$. But for any s we have $\psi^{s+1}(u^{s+1}) \ge \psi^s(u^s)$, which together with $\psi^{s_k}(u^{s_k}) \to \min_{u \in U^+} \psi(u)$ leads to $\psi^s(u^s) \to \min_{u \in U^+} \psi(u)$. However, $\psi^s(u^s) = \sum_{i=1}^{l_s} p_i^s q^0(y^{s,i})$ and $\min_{u \in U^+} \psi(u)$ is an optimal value of problem (7)–(9) due to Theorem 1. This completes the proof.

Various ways of dropping the cuts in cutting-plane methods have been suggested in [11,20]. The following method, which keeps only $l+1$ points at each iteration, was put forward in [20].

Instead of problem (14)–(16), solve the following problem at each iteration:

$$\min\,(u_{l+1} + \varepsilon||u^s - u||^2)$$

$$q^0(\bar{y}^j) - \sum_{k=1}^{l} q^k(\bar{y}^j)u_k = u_{l+1} \le 0 \quad , j = \overline{1,l+1}$$

$$u_k \ge 0 \quad ,$$

where $u^s = \arg\min_{u \in U^+} \psi^s(u)$. That this modified version converges can be proved in a similar way to Theorem 5.

5. STOCHASTIC PROCEDURE

By a corollary of Theorem 1, problem (6) is reduced to a minimax problem with a nonconcave inner problem of maximization and a convex final problem of minimization. A vast amount of work has been done on minimax problems but virtually all of the existing numerical methods fail if the inner problem is nonconcave. To overcome this difficulty we adopt an approach based on stochastic optimization techniques.

Consider the fairly general minimax problem

$$\min_{x \in X} \max_{y \in Y} f(x,y) \quad , \tag{30}$$

where $f(x,y)$ is a continuous function of (x,y) and a convex function of x for each $y \in Y$, $X \subset R^n$, $Y \subset R^m$. Although

$$F(x) = \max_{y \in Y} f(x,y) \tag{31}$$

is a convex function, to compute a subgradient

$$F_x(x) = f_x(x,y)|_{y=y(x)}$$

$$y(x) = \arg\max_{y \in Y} f(x,y) \tag{32}$$

$$f_x(x,y) \in \partial_x f(x,y) = \{g \mid f(z,y) - f(x,y) \ge \langle g, z - x \rangle, \forall z \in X\}$$

requires a solution $y(x)$ of nonconcave problem (32). In order to avoid the difficulties involved in computing $y(x)$ one could try to approximate Y by an ε-set Y_ε and consider

$$y^\varepsilon(x) = \arg\max_{y \in Y_\varepsilon} f(x,y)$$

instead of $y(x)$. But, in general, this would require a set Y_ε containing a very large number of elements. An alternative is to use the following ideas. Consider a sequence of sets Y_s, $s = 0,1,\ldots$ and the sequence of functions

$$F^s(x) = \max_{y \in Y_s} f(x,y) \quad .$$

It can be proved (see, for instance, [21]) that, under certain assumptions concerning the behavior of sequence F^s, the sequence of points generated by the rule

$$x^{s+1} = x^s - \rho_s F_x^s(x^s), \; s = 0,1,\ldots \quad , \tag{33}$$

$$F_x^s(x^s) \in \partial F^s(x^s) = \{g \mid F^s(x) - F^s(x^s) \ge \langle g, x - x^s \rangle, \forall x\}$$

(where the step size ρ_s satisfies assumptions such as $\rho_s \ge 0$, $\rho_s \to 0$, $\sum_{s=0}^{\infty} \rho_s = \infty$) tends, in some sense, to follow the time-path of optimal solutions: for $s \to \infty$

$$\lim \, [F^s(x^s) - \min F^s(x)] = 0 \quad .$$

We will show below how Y_s (which depends on x^s) can be chosen so that we obtain the convergence

$$\min F^s(x) \to \min F(x) \quad ,$$

where Y_s contains only a finite number $N_s \geq 2$ of random elements.

The principal peculiarity of procedure (33) is its nonmonotonicity. Even for differentiable functions $F^s(x)$, there is no guarantee that x^{s+1} will belong to the domain

$$\{x \mid F^t(x) < F^t(x^s)\} \text{ , } t \geq s + 1$$

of smaller values of functions F^{s+1} , F^{s+2},... (see Figure 1).

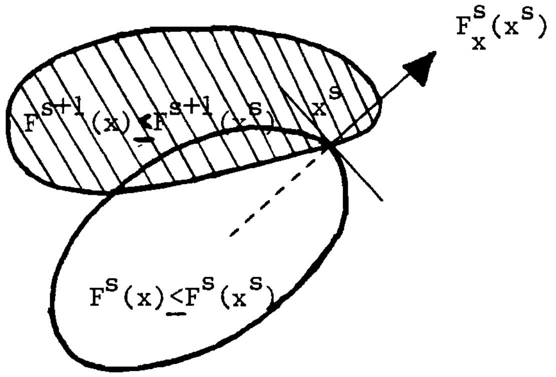

Figure 1

Various devices can be used to prevent the sequence $\{x^s\}_{s=0}^{\infty}$ from leaving the feasible set X.

The procedure adopted here is the following (see [22]).

We start by choosing initial points x^0, y^0, a probabilistic measure P on set Y and an integer $N_0 \geq 1$. Suppose that after the s-th iteration we have arrived at points x^s, y^s. The next approximations x^{s+1}, y^{s+1} are then constructed in the following way. Choose $N_s \geq 1$ points

$$y^{s,1}, y^{s,2}, \ldots, y^{s,N_s}$$

according to measure P, and determine the set

$$Y_s = \{y^{s,1}, y^{s,2}, \ldots, y^{s,N_s}\} \cup y^{s,0},$$

where $y^{s,0} = y^s$. Take

$$y^{s+1} = \arg\max_{y \in Y_s} f(x^s, y)$$

and compute

$$x^{s+1} = \pi[x^s - \rho_s f_x(x^s, y^{s+1})], \ s = 0,1,\ldots$$

where ρ_s is the step size and π is the result of a projection operation on X.

Before studying the convergence of this algorithm, we should first clarify the notation used:

$P(A)$ is a probabilistic measure of set $A \supseteq Y$,

$X^* = \arg\min_{x \in X} F(x)$,

$Y_\varepsilon^*(x) = \{y \mid y \in Y, f(x,y) \geq F(x) - \varepsilon\}, \ \varepsilon > 0$,

$p(\varepsilon, x) = P\{Y_\varepsilon^*(x)\}$,

$\gamma(\varepsilon) = \inf_{x \in X} p(\varepsilon, x)$,

$\tau(k, \varepsilon) = \max\{\tau \mid \sum_{s=k-\tau}^{k-1} \rho_s \leq \varepsilon, \ \tau \leq k\}$,

i.e., $\tau(k,\varepsilon)$ is the largest number of steps preceding step k for which the sum of step sizes does not exceed ε.

Theorem 6. *Assume that*

1. *X is a convex compact set in R^n and Y is a compact set in R^m.*

2. *$f(x,y)$ is a continuous function of (x,y) and a convex function of x for any $y \in Y$,*

$$\sup_{\substack{x \in X \\ y \in Y}} \| f_x(x,y) \| = C < \infty \quad .$$

3. *Measure P is such that $\gamma(\varepsilon) > 0$ for $\varepsilon > 0$.*

4. *$\rho_s \to +0$, $\sum_{s=0}^{\infty} \rho_s = \infty$.*

Then for $s \to \infty$

$$E \min_{x \in X^*} \| x^s - z \| \to 0 \quad .$$

If, in addition, there exists an $\varepsilon_0 > 0$ such that for all $\varepsilon \le \varepsilon_0$ and each $0 < q < 1$

$$\sum_{s=0}^{\infty} q^{\tau(s,\varepsilon)} < \infty \quad , \tag{34}$$

then, as $s \to \infty$,

$$\min \{ \| x^s - z \| \mid z \in X^* \} \to 0$$

with probability 1.

Proof

1. First of all let us prove that

$$F(x^s) - f(x^s, y^s) \to 0$$

in the mean. To simplify the notation we shall assume that $N_s = N \ge 1$. According to the algorithm

$$f(x^s, y^{s+1}) \ge f(x^s, y^{s,v}) \; , \; v = \overline{0,N}$$

and therefore

$$f(x^{s+1}, y^{s+1}) - f(x^{s+1}, y^{s,v}) \ge [f(x^{s+1}, y^{s+1}) - f(x^s, y^{s+1})]$$

$$+ [f(x^s, y^{s,v}) - f(x^{s+1}, y^{s,v})] \quad .$$

Since there is a constant K such that

$$|f(x^{s+1}, y) - f(x^s, y)| \le K||x^{s+1} - x^s|| \le K^2 \rho_s \quad ,$$

then

$$f(x^{s+1}, y^{s+1}) \ge f(x^{s+1}, y^{s,v}) - 2K^2 \rho_s \quad .$$

We also have

$$f(x^{s+1}, y^{s+2}) \ge f(x^{s+1}, y^{s+1,v}) \, , \; v = \overline{0,N} \quad ,$$

or, in particular, for $v = 0$

$$f(x^{s+1}, y^{s+2}) \ge f(x^{s+1}, y^{s+1}) \quad .$$

Therefore

$$f(x^{s+1}, y^{s+2}) \ge f(x^{s+1}, y^{k,v}) - 2K^2 \rho_s \, , \; k = s, s+1 \, , \; v = \overline{0,N} \quad ,$$

and in the same way

$$f(x^{s+2}, y^{s+2}) \ge f(x^{s+2}, y^{k,v}) - 2K^2(\rho_s + \rho_{s+1}) \, , \; k = s, s+1 \, , \; v = \overline{0,N}$$

etc.

Continuing this chain of inequalities, we arrive at the following conclusion:

$$f(x^s, y^s) \ge f(x^s, y^{k,v}) - 2K^2 \sum_{l=s-\tau(s,\varepsilon)}^{s-1} \rho_l$$

$$k = \overline{s - \tau(s,\varepsilon) \, , \, s-1} \, , \; v = \overline{0,N} \quad .$$

Thus, if

$$Y_{s,\varepsilon} = \{y^{k,v} \, , \; v = \overline{0,N} \, , \; k = \overline{s - \tau(s,\varepsilon) \, , \, s-1}\}$$

then

$$f(x^s, y^s) \ge \max_{y \in Y_{s,\tau}} f(x^s, y) - 2K^2 \varepsilon \quad .$$

It is easy to see from this that

$$P\{F(x^s) - f(x^s, y^s) > (1 + 2K^2)\varepsilon\} \le$$

$$P\{F(x^s) - \max_{y \in Y_{s,\tau}} f(x^s, y) > \varepsilon\} \le [1 - \gamma(\varepsilon)]^{N\tau(s,\varepsilon)} \quad .$$

Since $\rho_s \to 0$, then $\tau(s, \varepsilon) \to \infty$ as $s \to \infty$. Hence

$$[1 - \gamma(\varepsilon)]^{N\tau(s,\varepsilon)} \to 0$$

as $s \to \infty$, and this proves the mean convergence of $F(x^s) - f(x^s, y^s)$ to 0.

2. We shall now show that, under assumption (34), $F(x^s) - f(x^s, y^s) \to 0$ with probability 1. It is sufficient to verify that

$$P\{\sup_{k \ge s} [F(x^k) - f(x^k, y^k)] > (1 + 2K^2)\varepsilon\} \to 0 \quad .$$

We have

$$P\{\sup_{k \ge s} [F(x^k) - f(x^k, y^k)] > (1 + 2K^2)\varepsilon\} \le$$

$$P\{\sup_{k \ge s} [F(x^k) - \max_{y \in Y_{k,\tau}} f(x^k, y)] > \varepsilon\} \le$$

$$\sum_{k=s}^{\infty} P\{F(x^k) - \max_{y \in Y_{k,\tau}} f(x^k, y) > \varepsilon\} \le \sum_{k=s}^{\infty} [1 - \gamma(\varepsilon)]^{N\tau(k,\varepsilon)} \to 0 \quad ,$$

since from assumption (34) the series

$$\sum_{k=s}^{\infty} [1 - \gamma(\varepsilon)]^{N\tau(k,\varepsilon)} \to 0$$

as $s \to \infty$.

3. Let us now prove that $Ew(x^s) \to 0$ as $s \to \infty$, where

$$w(x) = \min_{z \in X^*} ||x - z||^2 \quad .$$

We have

$$w(x^{s+1}) = ||x^{s+1} - x_s^*||^2 \le w(x^s) - 2\rho_s < f_x(x^s,y^s) \, , x^s - x_s^* > + \rho_s^2 ||f_x(x^s,y^s)||^2$$

$$\le w(x^s) - 2\rho_s[f(x^s,y^s) - f(x_s^*,y^s)] + K^2\rho_s^2$$

$$\le w(x^s) - 2\rho_s[f(x^s,y^s) - \min_{x \in X} F(x)] + K^2\rho_s^2$$

$$\le w(x^s) - 2\rho_s[F(x^s) - \min_{x \in X} F(x)] + 2\rho_s[F(x^s) - f(x^s,y^s)] + K^2\rho_s^2 \quad .$$

Taking the mathematical expectation of both sides of this inequality leads to

$$Ew(x^{s+1}) \le Ew(x^s) - 2\rho_s E[F(x^s) - \min_{x \in X} F(x)] + 2\rho_s\beta_s + K^2\rho_s^2 \quad , \tag{35}$$

where $\beta_s \to 0$ as $s \to \infty$ since it has already been proved that

$$E[F(x^s) - f(x^s,y^s)] \to 0 \text{ for } s \to \infty \quad .$$

Now let us suppose, contrary to our original assumption, that

$$Ew(x^s) > \alpha > 0 \, , s \ge s_0 \quad .$$

It is easy to see that in this case we also have

$$E[F(x^s) - \min_{x \in X} F(x)] > \delta > 0 \quad ,$$

where $\delta = \delta(\alpha)$ is a constant which depends on α. Then for sufficiently large $s \ge s_1$

$$Ew(x^{s+1}) \le Ew(x^s) - 2\rho_s[\delta - 2\beta_s - K^2\rho_s] \le Ew(x^s) - \delta\rho_s \tag{36}$$

since $\rho_s \to 0$, $\beta_s \to 0$ and therefore we can suppose that

$$\delta - 2\beta_s - K^2\rho_s > \delta/2 \, , s \ge s_1 \quad .$$

Summing the inequality (36) from s_1 to k, $k \to \infty$, we obtain from assumption (4) a contradiction to the non-negativeness of $Ew(x^s)$. Hence, a subsequence $\{x^{s_k}\}$ exists such that

$$Ew(x^{s_k}) \to 0$$

as $k \to \infty$. Therefore for a given $\alpha > 0$ a number $k(\alpha)$ exists such that

$$Ew(x^{s_k}) < \alpha \quad ,$$

where $s_k > s_{k(\alpha)}$. Let r be such that $s_k \leq r \leq s_{k+1}$ and $Ew(x^r) > \alpha$. Take l such that

$$l = \min_{s_k < i \leq r} \{i : Ew(x^j) > \alpha \quad \text{for } i \leq j \leq r\} \quad .$$

Since $\rho_s \to 0$ and $\beta_s \to 0$, we may assume that $2\beta_s + K^2\rho_s < \delta(\alpha)$ for $s > s_{k(\alpha)}$. This and (36) together imply that $Ew(x^r) \leq Ew(x^l)$. Now from (35) and the definition of l we get

$$Ew(x^l) \leq Ew(x^{l-1}) + 2\rho_l\beta_l + K^2\rho_l^2 \leq \alpha + 2\rho_l\beta_l + K^2\rho_l^2 \quad .$$

Thus $Ew(x^s) \to 0$, because α was chosen arbitrarily and $\rho_l \to 0$.

4. It can be proved that $w(x^s)$ converges to 0 with probability 1 in the same way that we have already proved mean convergence. We have the inequality

$$w(x^{s+1}) \leq w(x^s) - 2\rho_s[F(x^s) - \min_{x \in X} F(x)] + 2\rho_s\gamma_s + K^2\rho_s^2 \quad ,$$

where $\gamma_s \to 0$ with probability 1 because it has already been shown that under assumption (34)

$$F(x^s) - f(x^s, y^s) \to 0 \text{ as } s \to \infty$$

with probability 1. If we now assume that

$$w(x^s) > \alpha \text{ , } s \geq s_0$$

for some element of probabilistic space we will also have

$$F(x^s) - \min_{x \in X} F(x) > \delta > 0$$

etc.

We shall now give a special case in which condition (34) is satisfied.

Example. Assume that $\rho_s = a/s^b$, $a > 0$, $0 < b \leq 1$. Then Raab's test for series convergence shows that condition (34) is satisfied.

REFERENCES

1. Yu. Ermoliev, *Methods of Stochastic Programming* (in Russian), Nauka, Moscow (1976).
2. J. Žačkova, "On minimax solutions of stochastic linear programming problems", *Čas. Pěst. Mat.* 91(1966), pp. 423–430.
3. J. Dupacova, "Minimax stochastic programs with nonseparable penalties", Proceedings of the 9th IFIP Conference, Warsaw 1979. Part 1. *Lecture Notes in Control and Information Sciences* 22, Springer-Verlag, Berlin 1980, pp. 157–163.
4. J.H.B. Kemperman, "On a class of moment problems", *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability* 2 (1972), pp. 101–126.
5. J.H.B. Kemperman, "The general moment problem: a geometric approach", *Ann. Math. Statist.* 39 (1968), pp. 93–122.
6. 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).

7. S. Fromovitz, "Nonlinear programming with randomization", *Management Science* 9 (1965).

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

9. A.I. Kaplinski and A.I. Propoi, "Stochastic approach to nonlinear programming", *Automatika i Telemekhanika* 3 (1970) (in Russian).

10. A.N. Golodnikov, *Finding of Optimal Distribution Function in Stochastic Programming Problems* (Dissertation abstract), Institute of Cybernetics Press, Kiev (1979).

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

12. R.T. Rockafeller, *Convex Analysis*, Princeton University Press, 1970.

13. E. Nurminski and A. Zelikhovski, "ε-Quasigradient method for solving nonsmooth extremal problems", *Kibernetica* 5 (1977).

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

15. Ky Fan, *Convex Sets and Their Applications*, Argonne National Laboratory (1959).

16. R. Wets, "Grundlagen Konvexer Optimierung", *Lecture Notes in Economics and Mathematical Systems* 137, Springer-Verlag, Berlin, Heidelberg and New York, 1976.

17. L. Nazareth and R. Wets, "Algorithms for stochastic programs: The case of nonstochastic tenders", WP-83-5, International Institute for Applied

Systems Analysis, Laxenburg, Austria.

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

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

20. V.F. Demianov, *Nondifferentiable Optimization*, Nauka, Moscow (1982).

21. Yu. Ermoliev and A. Gaivoronski, "Simultaneous nonstationary optimization, estimation and approximation procedures", CP-82-16, International Institute for Applied Systems Analysis, Laxenburg, Austria (1982).

22. Yu. Ermoliev and A. Gaivoronsky, "A stochastic algorithm for minimax problems", CP-82-88, International Institute for Applied Systems Analysis, Laxenburg, Austria.