A STOCHASTIC ALGORITHM FOR MINIMAX PROBLEMS

Yu. Ermoliev
A. Gaivoronski

December 1982
CP-82-88

INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS
2361 Laxenburg, Austria

PREFACE

This paper deals with minimax problems in which the "inner" problem of maximization is not concave. A procedure based on the approximation of the inner problem by a stochastic set of elements which can contain only two elements at each iteration is shown to converge with probability 1.

A STOCHASTIC ALGORITHM FOR MINIMAX PROBLEMS

Yu. Ermoliev and A. Gaivoronski

1. Introduction

The main aim of this paper is to prove the convergence of the algorithm proposed in Ermoliev and Nedeva (1982). This algorithm was designed to treat the following problem:

$$\min_{x \in X} \max_{y \in Y} f(x,y) \tag{1}$$

where $f(x,y)$ is a continuous function of (x,y) and a convex function of x for each $y \in Y, X \supset R^n, Y \supset R^m$. A vast amount of work has been done on minimax problems, but virtually all of the existing numerical methods fail if $f(x,y)$ is not a concave function of y. Although

$$F(x) = \max_{y \in Y} f(x,y) \tag{2}$$

is a convex function, to compute a subgradient

$F_x(x) = f_x(x,y(x))$,

$$y(x) = \arg \max_{y \in Y} f(x, y) \quad , \tag{3}$$

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

requires a solution $y(x)$ of nonconcave problem (3). 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 great number of elements. An alternative is to use the procedures described in this paper, which are based on 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, Ermoliev and Gaivoronski, 1982) that, under certain natural assumptions, the sequence of points generated by the rule

$$x^{s+1} = x^s - \rho_s F_x^s(x^s), \; s = 0,1,\ldots \quad , \tag{4}$$

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

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

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

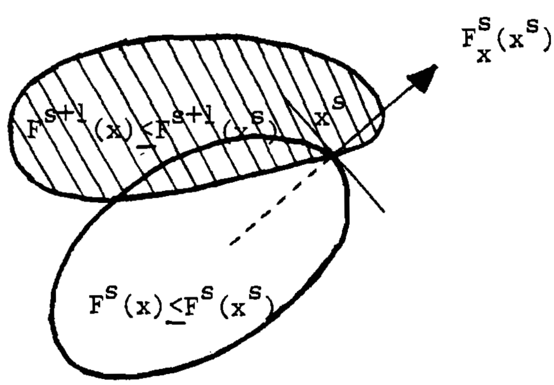
In this paper we will show how Y_s (which depends on x^s) can be chosen so that we obtain the convergence

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

where Y_s contains only a finite number $N_s \geq 2$ of elements. The principal peculiarity of procedure (4) is its nonmonotonicity. Even for differentiable functions $F^s(x)$, there is no guarantee that x^{s+1} will belong to the domain

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

of smaller values of functions $F^{s+1}, F^{s+2}, \ldots$ (see diagram below).



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

2. Algorithm

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:

(i) 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

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

(ii) Take

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

(iii) 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 explain some notation:

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

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

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

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

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

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

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 1. Assume that

(a) X is a convex compact set in R^n and Y is a compact set in R^m;

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

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

(c) Measure P is such that $\gamma(\varepsilon) > 0$ for $\varepsilon > 0$

(d) $\rho_s \to +0, \ \sum_{s=0}^{\infty} \rho_s = \infty$

Then for $s \to \infty$

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

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

$$\sum_{s=0}^{\infty} q^{\tau(s,\varepsilon)} < \infty \ , \tag{5}$$

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 $N_s = N \geq 1$. According to the algorithm

$$f(x^s, y^{s+1}) \geq f(x^s, y^{s,\nu}) \quad , \nu = \overline{0,N}$$

and therefore

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

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

Since there is a constant K such that

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

then

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

We also have

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

or, in particular, for $\nu = 0$

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

Therefore

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

and in the same way

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

etc.

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

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

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

Thus, if

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

then

$$f(x^s, y^s) \geq \max_{y \in Y_{s,\varepsilon}} f(x^s, y) - 2K^2 \varepsilon$$

It is easy to see from this that

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

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

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

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

for $s \to \infty$, 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 (5), $F(x^s) - f(x^s, y^s) \to 0$ with probability 1. It is sufficient to verify that

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

We have

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

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

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

since from assumption (5) the series

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

as $s\to\infty$.

3. We shall prove that $Ew(x^s)\to 0$ as $s\to\infty$, where

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

We have

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

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

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

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

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

$$Ew(x^{s+1})\leq Ew(x^s)-2\rho_sE[F(x^s)-\min_{x\in X}F(x)]+2\rho_s\beta_s+K^2\rho_s^2\quad ,\ (6)$$

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

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

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

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

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

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

where δ is a constant. Then for sufficiently large $s \ge s_1$

$$Ew(x^{s+1}) \le Ew(x^s) - 2\rho_s[\delta - 2\beta_s - K^2\rho_s] \le Ew(x^s) - \delta\rho_s \tag{7}$$

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

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

Summing the inequality (7) from s_1 to k, $k \to \infty$, we obtain from assumption (d) a contradiction of 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

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

where $s_k > s_{k(\alpha)}$. Since, from (6),

$$Ew(x^{s+1}) \le Ew(x^s) + 2\rho_s\beta_s + K^2\rho_s^2$$

there exists a number $s(\alpha)$ such that

$$Ew(x^s) \le 2\alpha \ , \ s \ge s(\alpha) \ .$$

Because α was chosen arbitrarily, this means that $Ew(x^s) \to 0$.

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 (5)

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

with probability 1. If we now assume that

$$w(x^s) > \alpha \, , \; s \geq s_0 \quad ,$$

we will also have

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

etc.

We shall now give some special cases in which condition (5) is satisfied.

Lemma. Assume

$$\rho_s = \frac{a}{s^b} \, , \; a > 0, \, 0 < b \leq 1 \; .$$

Then condition (5) is satisfied.

Proof. Let us evaluate $\tau(s, \varepsilon)$. If

$$\sum_{s=l}^{m-1} \frac{a}{s^b} < \varepsilon \; ,$$

then $m - l \leq \varepsilon m^b / a$ and

$$\max_{l \leq k \leq m} \rho_k = \frac{a}{l^b} \leq a / [m - \frac{\varepsilon m^b}{a}]^b \quad ,$$

$$\frac{\max\limits_{1\le k\le m}\rho_k}{\min\limits_{1\le k\le m}\rho_k} \le \frac{m^b}{[m-\frac{\varepsilon m^b}{a}]^b} = \frac{1}{[1-\frac{\varepsilon}{am^{1-b}}]^b} \le \frac{1}{(1-\frac{\varepsilon}{a})^b} = Q(\varepsilon,b) \quad .$$

Therefore, we have

$$\varepsilon - \rho_k \le \sum_{s=k-\tau(k,\varepsilon)}^{k-1} \rho_s \le \tau(k,\varepsilon)\rho_{k-\tau(k,\varepsilon)} \le Q(\varepsilon,b)\tau(k,\varepsilon)\rho_k$$

and

$$\tau(k,\varepsilon) \ge \frac{\varepsilon-\rho_k}{\rho_k Q(\varepsilon,b)} \ge \frac{\varepsilon}{\rho_k Q(\varepsilon,b)}$$

Then

$$\sum_{s=0}^{\infty} q^{\tau(s,\varepsilon)} \le \sum_{s=0}^{\infty} q^{\frac{\varepsilon}{\rho_s Q(\varepsilon,b)}} = \sum_{s=0}^{\infty}\left[q^{\frac{\varepsilon}{aQ(\varepsilon,b)}}\right]^{s^b} = \sum_{s=0}^{\infty}\mu^{s^b},$$

where $\mu = q^{\frac{\varepsilon}{aQ(\varepsilon,b)}} < 1$.

The convergence of $\sum\limits_{s=0}^{\infty}\mu^{s^b}$ can easily be verified by Raab's test.

3. Maximization with respect to unknown distribution functions

Minimax problems arise frequently in practice, in particular in decision making under uncertainty. An important class of problems of this type was discussed in Ermoliev and Nedeva (1982); this is the class of duals to problems involving maximization of mathematical expectations with respect to distribution functions of unknown parameters. The problem is to find the distribution function $H(z)$ that maximizes (minimizes) the integral

$$\int_Z g^0(z)\,dH(z) \tag{8}$$

subject to

$$\int_Z g^l(z)dH(z) \le 0 \ , \quad l = \overline{1,L} \ , \tag{9}$$

$$\int_Z dH(z) = 1 \ . \tag{10}$$

This can be regarded as the problem of evaluating system reliability, where the integral (8) defines the expected reliability, $H(z)$ is a partially known distribution function of random disturbances z, and constraints (9) and (10) contain known information about H (for instance, its moments or upper and lower bounds to the disturbances).

Maximization and minimization of integral (8) with respect to a distribution function H which possesses properties (9) and (10) then gives us the upper and lower bounds, respectively, of the system reliability.

Under rather general assumptions, for instance, if Z is a compact set and g^ν, $\nu = \overline{0,L}$, are continuous functions, the dual problem to the above is to minimize the convex function

$$\psi(u) = \max_{z \in Z} \, [g^0(z) + \sum_{l=1}^{L} u_l g^l(z)]$$

subject to

$$u \ge 0$$

In this case the "inner" problem of maximization is not concave.

REFERENCES

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

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