WORKING PAPER

NON-STANDARD STOCHASTIC APPROXIMATION SCHEME

Pavel Charamza

March 1989
WP-89-023

IIASA
International Institute for Applied Systems Analysis

NON-STANDARD STOCHASTIC APPROXIMATION SCHEME

Pavel Charamza

March 1989
WP-89-023

The author is from the Department of Statistics, Charles University, Prague. He participated in the SDS YSSP in 1988.

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

Foreword

An adaptive stochastic approximation scheme is suggested for solving a system of equations defined by functions whose values are available only through random observations on a given lattice. To improve the performance of the basic stochastic approximation scheme in the presence of outliers, the author uses isotonic and quasiisotonic regression to fit the observed function values. The convergence of the algorithm is proved.

The paper was finished during the author's stay with the SDS YSSP in 1988.

Alexander B. Kurzhanski
Chairman
System and Decision Sciences Program

Contents

1. INTRODUCTION AND MOTIVATION

In this paper we shall concentrate on suggesting a suitable approach for solving the system of equations

$$E\, \mathbf{g}(\mathbf{x},\omega)=\mathbf{0} \quad ,\mathbf{x}\in\mathbf{X}. \qquad /1/$$

where $\mathbf{g}(.,\omega):\mathbb{R}^{m}\to\mathbb{R}^{m}$, $\mathbf{X}\subset\mathbb{R}^{m}$, and ω is some random variable (or random vector) defined on some probability space.

If we consider a stochastic optimization problem

$$\min_{x\in X} Ef(\mathbf{x},\omega) \qquad /2/$$

where the minimum of $Ef(\mathbf{x},\omega)$ is reached at an interior point of $\mathbf{X}$ this problem can often be transformed /e.g.Rockafellar,Wets/ to /1/.

If

$$\mathbf{g}(\mathbf{x},\omega_t) = E\, \mathbf{g}(\mathbf{x},\omega)+e(\mathbf{x},\omega_t) \quad t=1,2,\ldots \ ,\mathbf{x}\in\mathbf{X}$$

where ω_t are independent random variables (or vectors) and the following conditions are fulfilled

a/ ω_t are independent $t=1,2,\ldots$

b/ $Ee(\mathbf{x},\omega_t) = \mathbf{0}$ $t=1,2,\ldots$

c/ $E\,\|e(\mathbf{x},\omega_t)\|^2 \le \text{const.}(1+\|\mathbf{x}\|^2)$ $t=1,2,\ldots$

d/ $E^2\mathbf{g}(\mathbf{x},\omega) \le \text{const.}(1+\|\mathbf{x}\|^2)$,

then the standard stochastic approximation algorithm

$$x_{t+1} = \Pi_X(x_t - \frac{a}{t}\, g(x_t,\omega_t)) \qquad t=1,2,\ldots\ ,(a>0,x_1 \text{ arbitrary}) \qquad /3/$$

may be used for solving /1/, where Π_X denotes projection onto the set X.(See e.g. Ermoliev (1976)).

For our motivation let us consider the case where $X=\mathbb{R}^1$ (i.e. $g(.,\omega):\mathbb{R}^1 \to \mathbb{R}^1$). Let us denote the observation $g(x_t,\omega_t)$ by y_t and suppose that we have observations up to time n . We can fit these observations by a linear function y=cx+b and take as the estimate of the root of equation /1/ the least squares estimate of $-\frac{b}{c}$. (See picture 1).Let us denote this estimate θ_n.If we suppose that the parameter c is known then we easily get

$$\theta_n = -\frac{1}{c}\,\bar{y}_n + \bar{x}_n,$$

where $\bar{y}_n = \frac{1}{n}\sum_1^n y_t$, $\bar{x}_n = \frac{1}{n}\sum_1^n x_t$.For $x_{n+1} = \theta_n$ it follows by induction that

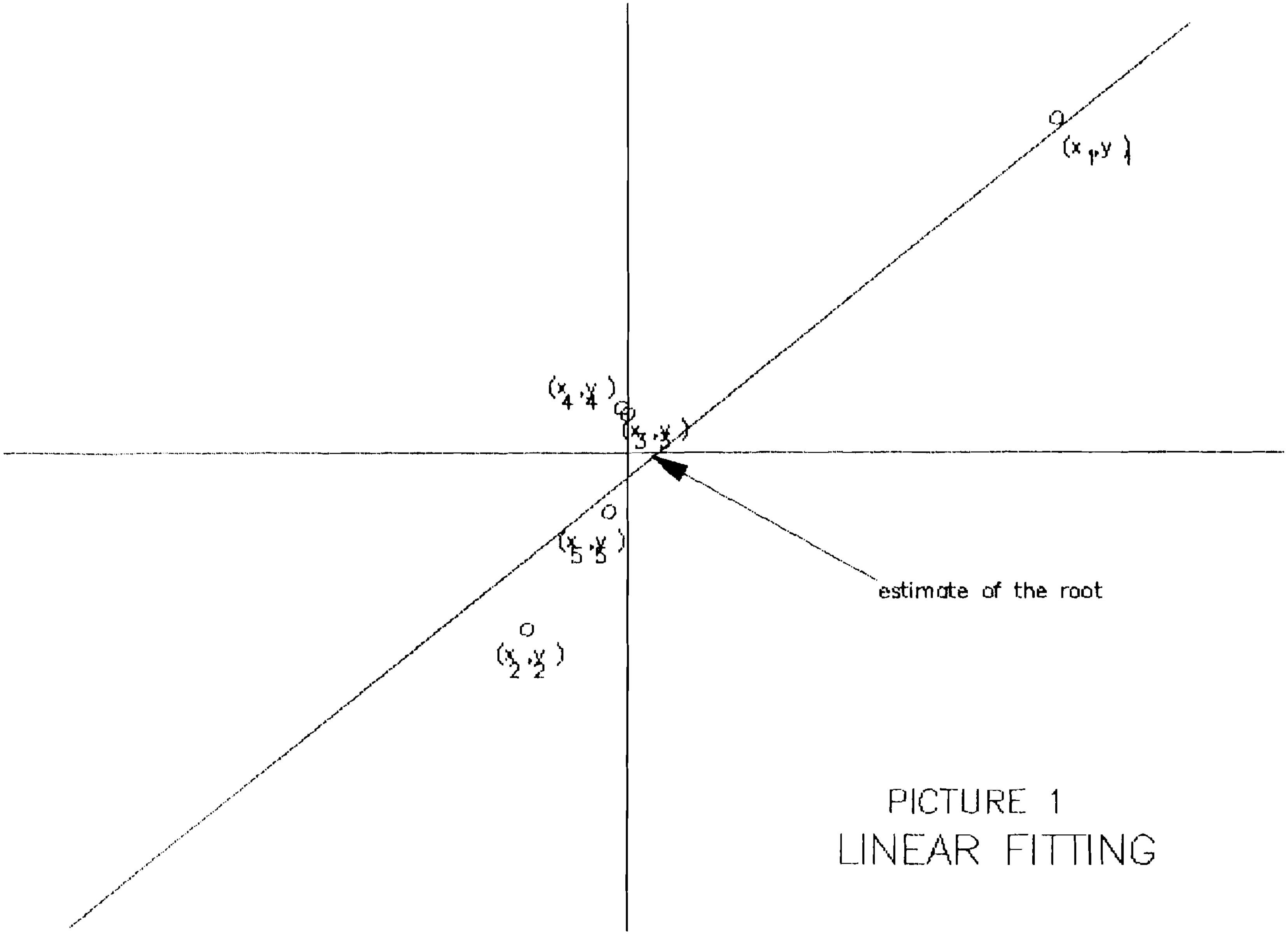
$$x_{n+1} = x_n - \frac{1}{nc}\, y_n$$

which is in the form of /3/.(See e.g. Robbins,Lai for further investigations). Many suggestions were given for adaptive estimate of c. The investigation of Mukerjee (1981) can be interpreted as the attempt to fit the observations at time n by a more suitable function then the linear one. He suggested least squares fitting by isotonic regression function and took x_{n+1} as the root of this fit.His method will be described in a more general way in Chapter 2.

If the observations are really observed and not simulated then we could expect some outliers among them. Algorithm /3/ can totally fail at this moment. (See picture 2). To avoid problems of this kind the robust version of the problem /1/ can be written as

$$m(x) \stackrel{def}{=} \underset{\varphi}{\operatorname{argmin}} \int H(y-\varphi)\, dF_{\mathbf{x}}(y) = 0. \qquad /5/$$

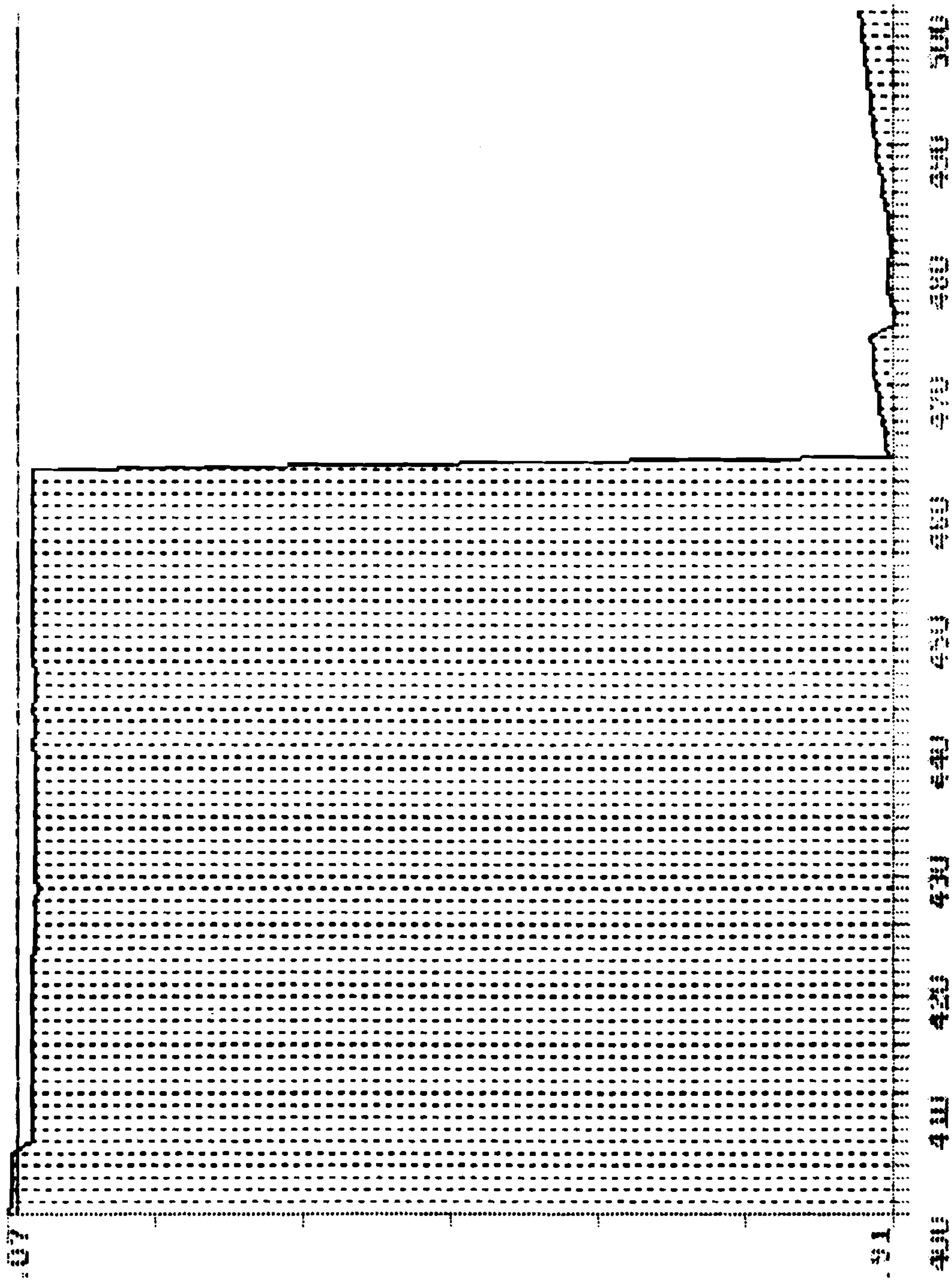
Here $F_{\mathbf{x}}$ is the distribution function of $g(\mathbf{x},\omega)$, $H:\mathbb{R}^1 \dashrightarrow \mathbb{R}^1$. If $H(z) = z^2$, then /5/ is equal to /1/. (See picture 3 for some standard choices of H). In Chapter 2 the procedure analogous to Mukerjee's will be suggested for solving problem /5/. In Chapter 3 some multidimensional extensions of Mukerjee's algorithm will be investigated.

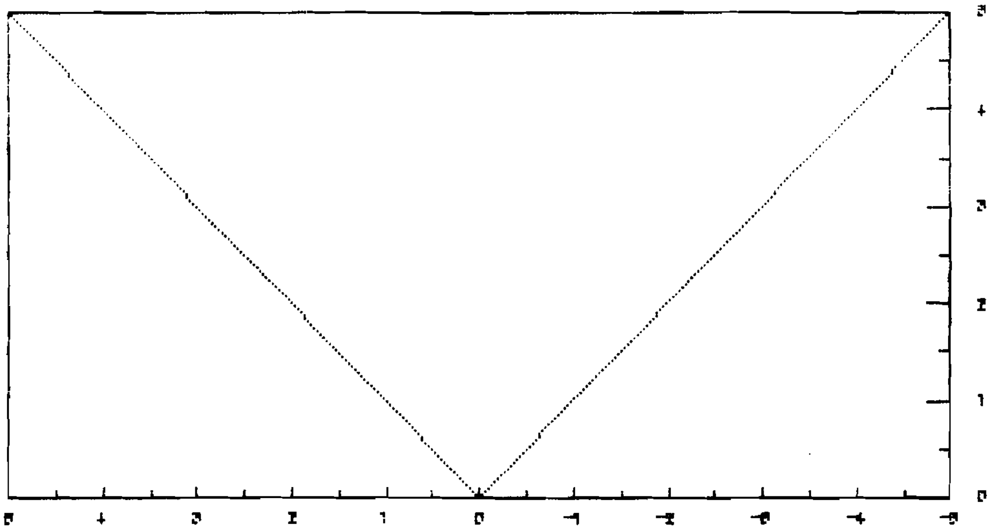


PICTURE 1
LINEAR FITTING

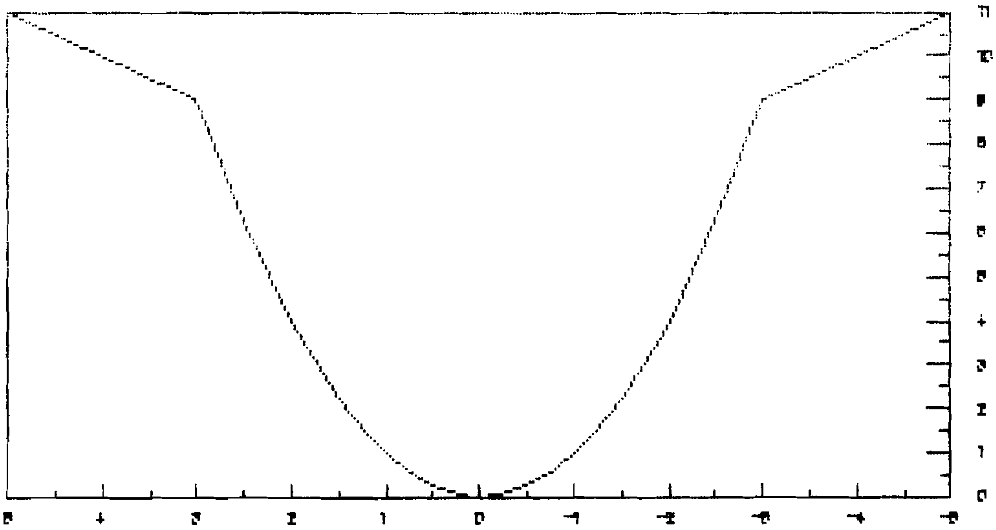
The behaviour of Robbins-Monro procedure in the case of outliers

picture 2





$H(z)=\text{const.}|x|$



$H(z)=cz^4[|z|<K]+[d(|z|-K)+cK^2)]I[|z|>=K]$

PICTURE 3

Choices of H

2. ONE DIMENSIONAL CASE

In this chapter we suggest algorithm for solving /5/. Notice that /5/ is a unconstrained one dimensional problem ($x\in\mathbb{R}^1$). For constrained problem see Remark 1. As already mentioned, in the special case when $H(z)=z^2$, /5/ is equivalent with /1/, where $g(.,\omega):\mathbb{R}^1\to\mathbb{R}^1$, $X=\mathbb{R}^1$.

Notation: 1/ $x^1<\ldots<x^N$ are real numbers,

2/ n_i ($i=1,\ldots N$) denotes the number of observations at the point x^i up to time n,

3/ y_{ij} $i=1,\ldots,N, j=1,\ldots,n_i$ is the value of the j-th observation at the point x^i,

4/ $H:\mathbb{R}^1\dashrightarrow\mathbb{R}^1$ is some measurable function,

5/ $|I|$ denotes the cardinality of the set I,

6/ for $i\in I\subseteq\{1,\ldots,N\}$ let (i) denote the rank of the element i.

Any element of

$$\underset{t_1\leq\ldots\leq t_N}{\operatorname{argmin}} \sum_{i=1}^{N}\sum_{j=1}^{n_i} H(y_{ij}-t_i) \qquad /6/$$

will be called the sample isotonic regression .

For any fixed real number r and positive number d, the set $L=\{r+ld;l\in Z\}$ will be called the lattice with step size d. We denote the set of all integers as Z.

We shall use abbreviations i.o. for "infinitely often",a.s for "almost sure". For example,$P[\, x_n\in A_n$ i.o.$]$ means the probability of the event $[x_n\in A_n$ for an infinite number of n$]$. We shall also write SLLN instead of "strong law of large number".

\- - -

The problem /6/ can be considered as the attempt to fit our observations by a function more suitable than the linear one. In this case we suggest isotonic fitting . For recursive estimation of the root of the equation /5/ we shall use the algorithm 1. In comparison with standard stochastic approximation algorithms it can be useful if we can make observations only at the points of the lattice L. At the end of this chapter we also suggest algorithm 2 for solving /6/ in the case when H is a convex function.

ALGORITHM 1

1/ Select an arbitrary closed interval $[a,b]\subset\mathbb{R}^1$ and set $X_0=[a,b]\cap L$ L. Set n=0 and make an observation at each point of X_0. Go to step 2.

2/ Define $x_{nm}=\min X_n, x_{nM}=\max X_n$. Let N denote the cardinality of X_n $(N=|X_n|)$. Set

$$x^1=x_{nm}, x^2= x_{nm}+d, \dots, x^N=x_{nM}.$$

Solve /6/ and define

$$\theta_{nm}=\max \{x_{nm}-d, \max \{x_i\in X_n; t_i=t(x_i) < 0 \}\}$$

$$\theta_{nM}=\min \{x_{nM}+d, \min \{x_i\in X_n; t_i=t(x_i) > 0 \}\},$$

where we set $\max \emptyset =-\infty, \min \emptyset =\infty$.

Take an arbitrary point $\theta_n\in(\theta_{nm},\theta_{nM})$ and set

$$x_{n+1}=[\theta_n] \overset{def}{=} \max_{\theta\in L, \theta\le\theta_n} \theta$$

$$x'_{n+1}=[\theta_n+d]$$

$$X_{n+1}=X_n \cup \{x_{n+1}\}\cup\{x'_{n+1}\}$$

Make observations at points x_{n+1}, x'_{n+1} ,set n=n+1 and repeat step 2.

- - -

REMARK 1: The interval $[\theta_{nm}, \theta_{nM}]$ can serve as the interval estimate of the root of equation /5/. If we suppose the root of /5/ to be in some set $\mathbf{X} \subset \mathbb{R}^1$, we can take θ_n from the algorithm 1 as any point of $[\theta_{nm}, \theta_{nM}] \cap \mathbf{X}$.

- - -

Now we shall derive some asymptotical properties of the sequence $\{\theta_n\}$.

ASSUMPTIONS:

A0 Let our observation at time t at the point z be $g(z,\omega_t)$, where $g(.,\omega_t)$ is a real valued function $(g(.,\omega_t):\mathbb{R}^1 \to \mathbb{R}^1)$ and ω_t are independent random variables or vectors defined on some probability space $(\Omega, \mathcal{A}, P)$. We suppose that for any fixed $z \in L$ $g(z,\omega_t)$ are identically distributed for all t with distribution function F_z. Let P_z be the probability measure on $\mathbb{R}^1$ corresponding to F_z. Let E_z be the corresponding expectation.

A0' Let H is a real valued continuous function.

Let a be some real valued function such that $\rho(y,\theta) = H(y-\theta) - a(y)$ satisfies conditions A1, A2, A2', A3, A4, A4', A5, A5'.

We define $\gamma_z = \int \rho(y,\theta) dF_z(y) = E_z \rho(y,\theta)$.

A1 There exists a probability measure Q on $\mathbb{R}^1$ such that

$$\forall \theta \in \mathbb{R}^1 \ \exists x_\theta \in \mathbb{R}^1 \ \forall x \geq x_\theta \ \forall z \in L \quad P_z[\,|\rho(y,\theta)| \geq x\,] \leq Q[\,|\rho(y,\theta)| \geq x]$$

and

$$\int |\rho(y,\theta)| \, dQ(y) < \infty.$$

A2 There exist $x_m \in L, \alpha \in \mathbb{R}^1, \varepsilon > 0$ such that

$$\exists \delta > 0 \ \exists \theta_1 \in (-\infty, \alpha-\varepsilon) \ \forall z \in L \cap (-\infty, x_m] \ \forall \theta > \alpha - \frac{\varepsilon}{2} \qquad \gamma_z(\theta) > \gamma_z(\theta_1) + \delta.$$

A2' There exist $x_M \in L$ such that for α and ε from A2

$$\exists \delta' > 0 \ \exists \theta_1' \in (\alpha+\varepsilon, \infty) \ \forall z \in L \cap \langle x_M, \infty) \ \forall \theta < \alpha + \frac{\varepsilon}{2} \qquad \gamma_z(\theta) > \gamma_z(\theta_1') + \delta'$$

A3 There exists a continuous function b such that

$\exists \varkappa > 0 \ \exists C \in \mathbb{R}^1 \ \forall \theta, |\theta| > C, \ \forall z \in L$ the inequality

$$b(\theta) > \frac{\gamma_z(\phi_1) + 2\varkappa}{1 - 2\varkappa}$$

is satisfied, where $\phi_1 = \theta_1, \theta_2$.

A4 $\exists K \ (-\infty, K) \supset (-\infty, \alpha-\varepsilon)$ and $E_z \inf\limits_{\theta \notin (-\infty, K)} \dfrac{\rho(y,\theta)}{b(\theta)} \geq 1 - \varkappa \quad \forall z \in L \cap (-\infty, x_m]$,

where $\varkappa$ is taken from A3.

A4' $\exists K' \ (K', \infty) \supset (\alpha+\varepsilon, \infty)$ and $E_z \inf\limits_{\theta \notin (K', \infty)} \dfrac{\rho(y,\theta)}{b(\theta)} \geq 1 - \varkappa \ \forall z \in L \cap [x_M, \infty)$,

where $\varkappa$ is taken from A3.

A5 For K from A4 there exists a probability distribution Q_1 on $\mathbb{R}^1$ such that

$$P_z\left[\left| \inf_{\theta \notin (-\infty, K)} \frac{\rho(y,\theta)}{b(\theta)} \right| \geq x \right] \leq Q_1\left[\left| \inf_{\theta \notin (-\infty, K)} \frac{\rho(y,\theta)}{b(\theta)} \right| \geq x \right] \ \forall z \in L \cap (-\infty, x_m)$$

and

$$\int \left| \inf_{\theta \notin (-\infty, K)} \frac{\rho(y,\theta)}{b(\theta)} \right| dQ_1(y) < \infty$$

A5' For K' from A4' there exists a probability distribution Q_1' on $\mathbb{R}^1$ such that

$$P_z\left[\left| \inf_{\theta\notin(K',\infty)} \frac{\rho(y,\theta)}{b(\theta)} \right| \geq x \right] \leq Q_1'\left[\left| \inf_{\theta\notin(K',\infty)} \frac{\rho(y,\theta)}{b(\theta)} \right| \geq x \right] \quad \forall z\in L\cap(x_M,\infty).$$

and

$$\int \left| \inf_{\theta\notin(K',\infty)} \frac{\rho(y,\theta)}{b(\theta)} \right| dQ_1'(y)<\infty.$$

A6 $\forall \varkappa_0 \ \forall\theta' \exists U_{\theta'}, \ \exists K \ \forall z$

$$E_z\left\{\left[\inf_{\theta\in U_{\theta'}} \rho(y,\theta) - \rho(y,\theta') \right] I[|y|\geq K \right\} \leq \varkappa_0$$

These rather complicated assumptions are analogous to those of Huber (1967) for the law of large numbers for M-estimates. Here the difference is that we have a set of distributions F_z instead of one distribution as Huber has. If H is a convex function then from the assumptions A2,A2' it follows that for

$$m(z)=\operatorname{argmin}_{\theta}\int \rho(y,\theta)\, dF_z(y) \qquad /5'/$$

the conditions

$$m(z) \leq \alpha-\varepsilon \qquad \forall z\leq x_m$$

$$m(z) \geq \alpha+\varepsilon \qquad \forall z\geq x_M$$

are fulfilled. Assumptions A0 are similar to those of standard stochastic approximation algorithms. Assumptions A1,A5,A5' ensure the validity of SLLN. The other assumptions are technical. Now we give the asymptotical result for the sequence θ_n from algorithm 1.

THEOREM 1: Let H be a convex function. Let a from A0' be continuous.
Then under the assumptions A0, A0', A6, for the sequence $\{\theta_n\}$ from the algorithm 1, the relation

$$P[\theta_n \notin [x_m, x_M] \text{ i.o.}]=0.$$

holds true.

The proof of this theorem will be given in Appendix 1.

Now we give two corollaries for special choices of function H. In these corollaries m(z) is defined by /5'/. The first corollary can be considered as the asymptotical result for estimating the root of regression function by algorithm 1. The second corollary is the asymptotical result for estimating the root of the function of medians by the same algorithm.

COROLLARY 1: (Mukerjee 1981)
Assume the assumption A0 is fulfilled and $H(z)=z^2$. Let there be a probability measure Q such that $P_z[|y|\geq x] \leq Q[|y|\geq x]$ for all x larger than some constant and $\int |y| \, dQ(y)<\infty$. Let there be $\varepsilon>0, x_m \in L, x_M \in L, x_m < x_M$ such that

$$m(x)\leq\alpha-\varepsilon \qquad \forall x\leq x_m$$
$$m(x)\geq\alpha+\varepsilon \qquad \forall x\geq x_M.$$

Then for θ_n from the algorithm 1 the relation

$$P[\theta_n \notin [x_m, x_M] \text{ i.o.}]=0$$

holds true.

- - -

This corollary is proved in Mukerjee (1981) under slightly more general assumptions.

COROLLARY 2:

Set $H(z)=|z|$. Assume the assumption A0 is fulfilled and the following relations hold true

$$\forall\varepsilon_1\ \exists k\ \forall x\in L\quad P_x[|y|\geq k]\leq\varepsilon_1 \qquad /7a/$$

$$\exists\varepsilon_0>0\ \exists\alpha>0\ \exists x_m,x_M\in L, x_m<x_M\ \ (m(x)\leq\alpha-\varepsilon_0\ \forall x\leq x_m)\text{and}(m(x)\geq\alpha+\varepsilon_0\ \forall x\geq x_M). \qquad /7b/$$

For all $x\in L$ let there exist the derivative F'_x of F_x at the point $m(x)$ and

$$\exists\delta_0>0\ \forall x\in L\quad F_x'(m(x))>\delta_0 \qquad /7c/$$

Then for θ_n from the algorithm 1 the relation

$$P[\theta_n\notin[x_m,x_M]\ \text{i.o.}]=0.$$

holds true.

\- - -

For proof see Appendix 2.

Now we shall investigate the problem /6/, i.e.

$$\underset{t_1\leq\ldots\leq t_N}{\operatorname{argmin}}\sum_{i=1}^{N}\sum_{j=1}^{n_i}H(y_{ij}-t_i)$$

The following algorithm can be used for solving /6/.

ALGORITHM 2:

1/ Set $z=1, k_i^z=i$ for $i=1,\ldots N, L^z=N$.

2/ Solve the problems

$$\operatorname*{argmin}_{t} \sum_{i=k_l^z+1}^{k_{l+1}^z} \sum_{j=1}^{n_i} H(y_{ij} - t) \quad l=0,\dots L^z-1. \qquad /8/$$

Any of these solutions are denoted by $t_l^z, l=0,\dots,L^z-1$.

If $t_0^z \le \dots \le t_{L^z-1}^z$ then $t_i^* \overset{def}{=} t_l^z$ $i\in[k_l^z+1,k_{l+1}^z], l=0,\dots,L^z-1$ are the solutions of /6/ and the algorithm is finished. If any of these inequalities is not satisfied then go to step 3.

3/Set $z=z+1, k_0^z=0, k_1^z$ the smallest index such that $t^{z-1}_{k_1^z} < t^{z-1}_{k_1^z+1}$.

k_2^z is the next smallest index and so on. Define L^z as the number of k_l^z defined. Set $k^z_{L^z}=N$ and repeat step 2.

See picture 4 as an example.

\- - -

In the case $H(z)=z^2$ the algorithm 2 was proposed in [1].The solutions of /8/ can be obtained in this case as the weighted averages of the averages at the points $k_l^z+1,\dots, k_{l+1}^z$.The weights are proportional to the number of experiments at each of these points.

The idea of the algorithm is based on the following theorem .

THEOREM 2.Let H be a convex function.For $I_1\subset\{1,\dots,N\}$ let $\tilde{t}=(\tilde{t},..,\tilde{t})$ be the solution of

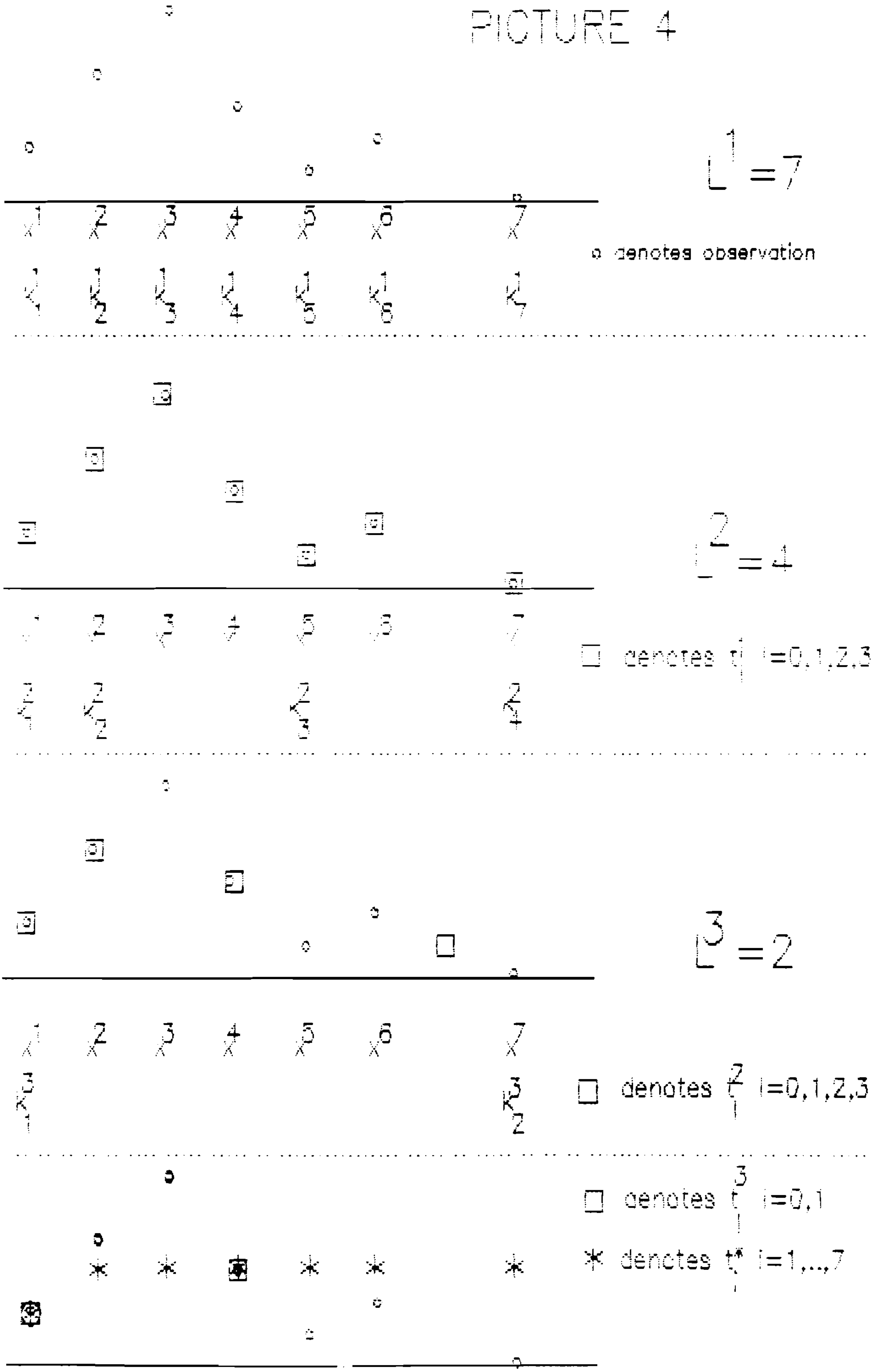
$$\operatorname*{argmin}_{t_1\le..\le t_{|I_1|}} \sum_{i\in I_1} \sum_{j=1}^{n_i} H(y_{ij}-t_{(i)})$$

and for $I_2\subset\{1,\dots,N\}$ such that $\min I_2>\max I_1$ let $\mathbf{t}^*=(t^*,\dots,t^*)$ be the solution of

$$\operatorname*{argmin}_{t_1\le..\le t_{|I_2|}} \sum_{i\in I_2} \sum_{j=1}^{n_i} H(y_{ij}-t_{(i)}) \ .$$

Izotonic regression algorithm for $H(z)=z^2$

PICTURE 4



If $t^* < \tilde{t}$ and $t' \in \underset{t}{\operatorname{argmin}} \sum_{i \in I_1 \cup I_2} \sum_{j=1}^{n_i} H(y_{ij} - t)$ then t' is also the solution of

$$\underset{t_1 \leq \ldots \leq t_{|I_1|+|I_2|}}{\operatorname{argmin}} \sum_{i \in I_1 \cup I_2} \sum_{j=1}^{n_i} H(y_{ij} - t_{(i)}).$$

- - -

Proof of this theorem will be given in Appendix 3.

At the end of this chapter we will recall a representation for the solution of /6/. This representation is very useful for proving the Theorem 1. It can be shown (Nemirovski, Polyak, Cybakov, 1984) that if H is a convex function then any solution $t^* = (t^*_1, \ldots, t^*_N)$ of /6/ can be expressed in the form

$$t^*_i = \max_{k \leq i} \min_{l \geq j} t(k,l), \qquad /9/$$

where

$$t(k,l) \in \underset{t}{\operatorname{argmin}} \sum_{i=k}^{l} \sum_{j=1}^{n_i} H(y_{ij} - t).$$

3. MULTIDIMENSIONAL APPROACH

In this chapter we suggest a multidimensional version of algorithm 1 for solving the problem /1/. We also prove a limit theorem for this multidimensional algorithm. We shall use the notation $\langle .,. \rangle$ for inner product in $\mathbb{R}^m$,

$\|.\|$ for the norm in $\mathbb{R}^m$,

$$x^- \stackrel{def}{=} \max\{0,-x\},$$

X_n is a discrete subset of $\mathbb{R}^m$; its elements are the points at that we made observations up to time n,

$y_j(x)$ is the value of the j-th observation at the point $x \in \mathbb{R}^m$ $(y_j(x) \in \mathbb{R}^m)$,

$n(x)$ is the number of observations at the point x

$\bar{y}_n(x)$ is the average of observations at the point x, i.e.

$$\bar{y}_n(x) = \frac{1}{n(x)} \sum_{j=1}^{n(x)} y_j(x).$$

- - -

We can follow the logic of the one dimensional approach for the multidimensional case of problem /1/ (i.e. $g:\mathbb{R}^m \to \mathbb{R}^m$):

1/ Take the least squares isotonic estimate of the function $Eg(x,\omega)$ as

$$\tilde{m} \in \operatorname*{argmin}_{m \in M} \sum_{x \in X_n} \sum_{j=1}^{n(x)} \|y_j(x) - m(x)\|^2, \qquad /10/$$

where

$$M = \{h:\mathbb{R}^m \to \mathbb{R}^m; \langle h(x)-h(y), x-y \rangle \geq 0 \ \forall \ x,y \in \mathbb{R}^m\}.$$

The set M is the set of multidimensional isotone functions.

2/Take the root of $\tilde{m}$,(i.e. the solution of $\tilde{m}(\theta)=0$) and provide observations near this root.Repeat step 1.

- - -

The solution of /10/ together with step 2/ is complicated. But the problem /10/ can be simplified if we enlarge the set M to the set

$$M'=\{h:\mathbb{R}^m\to\mathbb{R}^m:\exists\theta\in\mathbb{R}^m \text{ such that } \langle h(x),x-\theta\rangle\geq 0\}.$$

The sample quasi-isotonic regression field $m'_n(x), x\in X_n\subset\mathbb{R}^m$,will be defined as an arbitrary solution of

$$\underset{m\in M'}{\operatorname{argmin}} \sum_{x\in X_n}\sum_{j=1}^{n(x)} \|y_j(x)-m(x)\|^2 = \underset{m\in M'}{\operatorname{argmin}} \sum_{x\in X_n} \|\bar{y}_n(x)-m(x)\|^2 n(x)$$

its domain being then restricted to X_n. Any sample regression field has the form

$$m'(x)=\bar{y}_n(x) + \left[\frac{\langle\bar{y}_n(x),x-\theta\rangle^-}{\|x-\theta\|^2}\right](x-\tilde{\theta}) \qquad \text{for } x\neq\tilde{\theta}, x\in X_n,$$

$$m'(\tilde{\theta})=\bar{y}(\tilde{\theta}) \qquad \text{if } \tilde{\theta}\in X_n,$$

where $\tilde{\theta}$ is a solution of

$$\underset{\theta\in\mathbb{R}^m}{\operatorname{argmin}}\sum_{x\in X_n}\left[\frac{\langle\bar{y}_n(x),x-\theta\rangle^-}{\|x-\theta\|}\right]^2 n(x) \qquad /11/$$

REMARK: For the one-dimensional case, a simple algorithm for solving /11/ was suggested by Dupac (1987) and a stochastic approximation scheme analogous to that from Chapter 1 was proven to converge there. A multidimensional version of stochastic approximation algorithm was suggested but without proving a convergence theorem. In the sequel, we shall try to prove the convergence properties for the multidimensional case.

NOTATION: Let $e_i \in \mathbb{R}^m$ be the i-th unit vector. For $i=1,\dots m$ let r_i be an arbitrarily chosen positive number and denote $Z_i = Z \cap [-r_i, r_i]$. For $\delta > 0$, $z \in \mathbb{R}^m$ the set $M_z^\delta = \{z + \sum_1^m \delta e_i l_i ; l_i \in Z_i, i=1,..m\}$ will be called the lattice with the step size δ. For $y \in M_z^\delta$ define the m-dimensional cube with center y as $C_y^\delta = co\{y \pm \delta e_i ; i=1,\dots m\}$. (coA means convex hull of the set A). Finally, let $R_z^\delta = \{\theta \in \mathbb{R}^m ; \exists y \in M_z^\delta$ such that $\theta \in C_y^\delta$ and $y \pm 2\delta e_i \in M_z^\delta$, $i=1,\dots,m\}$.

- - -

ALGORITHM 3

1/Choose the initial point θ_0 arbitrarily from the set R_z^δ. Take a y such that $\theta_0 \in C_y^\delta$ and $y \pm 2\delta e_i \in M_z^\delta$. If there exist more than one such y, choose y randomly among them with equal probabilities, independently on the past. Make observations at the points

$$x_0^i = y + 2\delta e_i \qquad i=1,\dots m,$$

$$x_0^{m+i} = y - 2\delta e_i \qquad i=1,\dots m.$$

Set $n=1$, $X_1 = \bigcup_1^{2m} \{x_0^i\}$.

2/Solve the problem

$$\operatorname*{argmin}_{\theta \in R_z^\delta} \sum_{x \in X_n} \left[\frac{\langle \bar{y}_n(x), x-\theta \rangle^-}{\|x - \theta\|} \right] n(x) \qquad /12/$$

taking 0 instead of $\dfrac{\langle \bar{y}(x), x-\theta \rangle^-}{\|x - \theta\|}$ if $\|x-\theta\|=0$.

3/Let θ_n be an arbitrary solution of /12/ and again choose y such that $\theta_n \in C_y^\delta$ and $y \pm 2\delta e_i \in M_z^\delta$. If there are more than one such y, choose y randomly as in 1/. Make observations at the points

$$x_n^i = y + 2\delta e_i \qquad i=1,\dots m,$$

$$x_n^{m+i} = y - 2\delta e_i \qquad i=1,\dots m,$$

and set $n=n+1$, $X_n = X_{n-1} \cup \left(\bigcup_{1}^{2m} \{x_{n-1}^i\} \right)$. Return to step 2.

\- - -

REMARK: In the algorithm 3 we use as an estimate of θ^* the solution of /12/ which may be easier to solve than /11/. Nevertheless the following theorem remains valid if we replace in this algorithm the problem /12/ by /11/ as Dupac suggested. This fact will become clear from the proof of this theorem.

THEOREM 3:

Let $m(x)$ be a continuous vector field ($m:\mathbb{R}^m \to \mathbb{R}^m$) such that there exist exactly one $\theta^* \in R_z^\delta$ for which $m(\theta^*)=0$ and $\langle m(x), x-\theta^* \rangle > 0$ for every $x \in \mathbb{R}^m, x \neq \theta^*$. Assume that observations $y_i(x)$ can be expressed in the form $y_i(x) = m(x) + f_i(x)$ for $i=1,\dots,$ where $f_i(x)$ are random vectors obeying the SLLN and $Ef_i(x)=0$.

Then for every $\varepsilon>0$ $\exists$ $\delta_0>0$ such that for all $\delta \le \delta_0$ and for all $y \in M_z^\delta$ satisfying

$$(x \in C_y^\delta \Rightarrow \| m(x) \| \ge \varepsilon),$$

the equation $P[\theta_n \in C_y^\delta \text{ i.o.}]=0$ holds, where θ_n are defined by algorithm 3.

\- - -

PROOF: Choose ε arbitrarily. Because of the continuity of m there exists $\delta(\varepsilon)$ such that if $\|x-y\| < \delta(\varepsilon)$ then $\|m(x)-m(y)\| < \varepsilon$. Set $\delta_0 = \delta(\varepsilon/9\sqrt{m})/3$.

Let N_x^ξ denote the random variable such that for all $n \ge N_x^\xi$ the inequality $\|\frac{1}{n}\sum_1^n f_i(x)\| < \xi$ is valid. Due to the SLLN $N_x^\xi < \infty$ a.s.

for all x and $\zeta>0$. There exist constants $D>0$ and $C>0$ such that for all $x\in M_z^\delta, x\neq\theta^*$, for all $n(x)>K \overset{def}{=} \max_{x\in M_z^\delta} N_x^{\frac{C}{D}}$ the following inequalities hold:

$$\langle\bar{y}_n(x), x-\theta^*\rangle = \langle\bar{y}_n(x)-m(x), x-\theta^*\rangle + \langle m(x), x-\theta^*\rangle \geq$$
$$\geq -\|\bar{y}_n(x)-m(x)\|\| x-\theta^*\| + C \geq -D\|\bar{y}_n(x)-m(x)\| + C\geq$$
$$\geq -D\frac{C}{D} + C=0.$$

If $n(x)\leq K$ then $n(x)\langle\bar{y}_n(x), x-\theta^*\rangle^-/\|x-\theta^*\| \leq K \|\bar{y}_n(x)\|\leq K_1$.
From the two previous inequalities it follows that

$$\sum_{\substack{x\in X_n \\ x\neq\theta^*}} \left[\frac{\langle\bar{y}_n(x), x-\theta^*\rangle^-}{\|x - \theta^*\|}\right] n(x) < K_2 \;\forall n, \qquad /13/$$

where K_2 is a random variable finite a.s.

Suppose now that there exists $y\in M_z^\delta$ such that $\theta_n\in C_y^\delta$ i.o. and $\|m(x)\|>\varepsilon$ for each x from C_y^δ. We shall derive in this case that

$$\exists N_0\; \forall n\geq N_0\; \exists z\in\left\{ y \pm 2\delta e_i, i=1,\dots,m\right\} \forall\theta\in C_y^\delta\; \langle\bar{y}_n(z), z-\theta\rangle \leq -\frac{\delta}{3}. \quad /14/$$

From here because of implication ($\theta_n\in C_y$ i.o. $\Rightarrow n(z) \to \infty$) the convergence $\frac{\langle\bar{y}_n(w_n), w_n-\theta\rangle^-}{\|w_n-\theta\|} n(z) \xrightarrow[n\to\infty]{} \infty$ is valid for all $\theta\in C_y^\delta$. This contradicts /13/. Now we shall prove /14/. First we show that for all $x\in C_y^\delta$ there exists $z\in\left\{ y\pm 2\delta e_i, i=1,\dots,m\right\}$ such that for all $\theta\in C_y^\delta$ the inequality $\langle m(x), z-\theta\rangle \leq -\delta\varepsilon/\sqrt{m}$ holds true.

For the arbitrary vector $x\in C_y^\delta$ set $1^*= \operatorname{argmax}_i (|m_i(x)|)$ and take

$z=y-2\delta\ \mathrm{sign}(m_{1^*}(x))\ e_{1^*}$. Now θ can be represented as $\theta=\sum_1^m\left[\lambda_i(y+\delta e_i)+\lambda_{i+m}(y-\delta e_i)\right]$, where λ_i are satisfying $\sum_1^{2m}\lambda_i=1, \lambda_i\geq 0$ for $i=1,\dots,2m$. It implies

$$\langle m(x), y-2\delta\ \mathrm{sign}(m_{1^*}(x))\ e_{1^*}-\theta\rangle=$$

$$=\langle m(x), \sum_1^m(\lambda_{i+m}-\lambda_i)\delta e_i-2\delta\ \mathrm{sign}(m_{1^*}(x))\ e_{1^*}\rangle=$$

$$=\sum_1^m m_i(x)\ (\lambda_{i+m}-\lambda_i)\delta-|m_{1^*}(x)|2\delta\leq\delta|m_{1^*}(x)|(\sum_1^m|\lambda_{i+m}-\lambda_i|-2)\leq$$

$$\leq-\delta|m_{1^*}(x)|\leq-\delta\varepsilon/\sqrt{m}.$$

From the definition of δ it follows that for all $x\in C_y^\delta$ and all $z\in\left\{y\pm 2\delta e_i, i=1,\dots,m\right\}$ the inequality $\|m(x)-m(z)\|\leq\varepsilon/(9\sqrt{m})$ holds true. From SLLN we get $\|\bar{y}_n(z)-m(z)\|\leq\varepsilon/(9\sqrt{m})$ for all $n\geq N_0$, where N_0 is some random variable finite a.s. Because of $n(z)\xrightarrow[n\to\infty]{}\infty$ we get

$$\langle\bar{y}_n(z), z-\theta\rangle=\langle\bar{y}_n(z)-m(z), z-\theta\rangle+\langle m(z)-m(x), z-\theta\rangle+\langle m(x), z-\theta\rangle\leq$$

$$\leq\|\bar{y}_n(z)-m(z)\|\|z-\theta\|+\|m(z)-m(x)\|\|z-\theta\|-\frac{\delta\varepsilon}{\sqrt{m}}\leq$$

$$\frac{\delta\varepsilon}{3\sqrt{m}}+\frac{\delta\varepsilon}{3\sqrt{m}}-\frac{\delta\varepsilon}{\sqrt{m}}\leq-\frac{\delta\varepsilon}{3\sqrt{m}}$$

for all $\theta\in C_y^\delta$ and for all sufficiently large n.

\- - -

APPENDIX 1

Before proving the Theorem 1 we shall prove two lemmas using the following notation:

$F_t=\sigma(x_1,x_2,\dots,x_t)$, σ fields on (Ω,A,P) generated by the indicated random variables,

T_i^B the time of i-th entrance of the sequence x_t into the set $B\subseteq\mathbb{R}^1$,

$F_{T_i}=\sigma\{A;\ A\cap[T_i=t]\in F_t\}$,

$C_n\in \underset{\theta}{\operatorname{argmin}} \sum_i^n \rho(Y_{T_i^B},\theta)$,

U_θ some open neighborhoodof θ.

LEMMA 1: Let assumptions A0,A0' be fulfilled.Then

a/there exists $C_1<\alpha+\varepsilon$ such that for any $B\subseteq[x_M,\infty)$ satisfying $T_i^B<\infty$ $\forall i$), the inequality $C_n\geq C_1$ for all $n>N$ is valid, where N is a.s. finite random variable.

b/there exists $C_2>\alpha-\varepsilon$ such that for any $B\subseteq(-\infty,x_m]$ satisfying ($T_i^B<\infty$ $\forall i$) the inequality $C_n\leq C_2$ is valid for all $n>N$, where N is a.s. finite random variable.

PROOF: We shall use notation T_i instead of T_i^B for simplicity. Using results of Nevelson,Khasminskij (1972) we can derive that

$$E\left[\inf_{\theta<C_1}\frac{\rho(Y_{T_i},\theta)}{b(\theta)}\Big|\ F_{T_i}\right]=E_z\inf_{\theta<C_1}\frac{\rho(y,\theta)}{b(\theta)}\Bigg|_{z=x_{T_i}}.$$

Here be the notation $E_z\inf_{\theta<C_1}\frac{\rho(y,\theta)}{b(\theta)}\Big|_{z=x_{T_i}}$ we denote the value of the function $p(z)\stackrel{def}{=}E_z\inf_{\theta<C_1}\frac{\rho(y,\theta)}{b(\theta)}$ at the point $z=x_{T_i}$.

From A3 follows that there exist $\varkappa>0$, $C<\alpha+\varepsilon$ such that the inequality $b(\theta) > \dfrac{\gamma_z(\theta_1')+2\varkappa}{1-2\varkappa}$ holds true for all $\theta<C$. Due to A4', for this $\varkappa$ there exists some constant K' such that

$$E_z \inf_{\theta\le K'} \frac{\rho(y,\theta)}{b(\theta)} \ge 1-\varkappa \quad \forall z\in L\cap[x_M,\infty). \qquad /L1/$$

We set $C_1=\min(C,K')$. Applying the SLLN for martingale differences (see e.g. Loeve) with the assumption A5' we get

$$\inf_{\theta<C_1} \frac{1}{n}\sum_1^n \frac{\rho(Y_{T_i},\theta)}{b(\theta)} \ge \frac{1}{n}\sum_1^n \inf_{\theta<C_1} \frac{\rho(Y_{T_i},\theta)}{b(\theta)} \ge \frac{1}{n}\sum_1^n E_z \inf_{\theta<C_1} \frac{\rho(y,\theta)}{b(\theta)}\Bigg|_{z=x_{T_i}} - \varkappa \qquad /L2/$$

for all $n\ge N_1$. N_1 is a.s. finite random variable. From /L1/ and /L2/ for any $\theta<C_1$ and for all $z\in[x_M,\infty)$

$$\frac{1}{n}\sum_1^n \rho(Y_{T_i},\theta) \ge (1-2\varkappa)b(\theta) \ge \gamma_z(\theta_1)+2\varkappa.$$

Using A1 we can apply SLLN again, getting

$$\inf_\theta \frac{1}{n}\sum_1^n \rho(Y_{T_i},\theta) \le \frac{1}{n}\sum_1^n \rho(Y_{T_i},\theta_1) \le \frac{1}{n}\sum_1^n \gamma_z(\theta_1)\Bigg|_{z=x_{T_i}} +\varkappa.$$

From this the assertion a/ of Lemma 1 follows. The assertion b/ can be derived similarly.

- - -

LEMMA 2: Assume A0,A0',A6 are valid. Let for any K>0

$$\sup_{|y|<K} \left| \inf_{\theta\in U_{\theta'}} \rho(y,\theta) - \rho(y,\theta') \right| \xrightarrow[U_{\theta'}\to\{\theta'\}]{} 0. \qquad /L3/$$

Then

a/ $C_n\in[\alpha+\varepsilon/2,\infty)$ for any $B\subseteq[x_M,\infty)$ and $n\geq N$, if $T_i^B<\infty$ a.s. for all i. N is a.s. finite random variable.

b/ $C_n\in(-\infty,\alpha-\varepsilon/2]$ for any $B\subseteq(\infty,x_m]$ and $n\geq N$, if $T_i^B<\infty$ a.s. for all i. N is a.s. finite random variable.

PROOF:

From Lemma 1 it follows that there exists a constant $C_1<\alpha+\varepsilon$ such that for all sufficiently large n $C_n\geq C_1>-\infty$. Now from A6 and /L3/ we can derive that

$$E_z\left\{\inf_{\theta\in U_{\theta'}} \rho(y,\theta)-\rho(y,\theta')\right\} \xrightarrow[U_{\theta'}\to\{\theta'\}]{} 0 \text{ uniformly in z.} \qquad /L4/$$

Using A2' we get for any $U\in(C_1,\alpha+\varepsilon/2)$ the inequality

$$\inf_{\theta'\in[C_1,U]} \gamma_z(\theta') > \gamma_z(\theta_1) + \delta'. \qquad /L5/$$

For any $\theta'\in[C_1,U]$ there exists $U_{\theta'}$ such that

$$E_z \inf_{\theta\in U_{\theta'}} \rho(y,\theta) \geq \gamma_z(\theta_1) + 3\frac{\delta'}{4}. \qquad /L6/$$

This follows from /L4/,/L5/. From $\{U_{\theta};\theta\in[C,U], U_{\theta}$ satisfies /L6/$\}$ a finite number S of intervals U_s can be chosen such that $\bigcup_1^S U_s \supset [C_1,U]$.

Using SLLN for martingale differences we get for $n\geq N_s$ (N_s is a.s. finite random variable)

$$\inf_{\theta'\in U_s} \frac{1}{n}\sum_1^n \rho(Y_{T_i},\theta') \geq \frac{1}{n}\sum_1^n \inf_{\theta'\in U_s} \rho(Y_{T_i},\theta') \geq$$

$$\geq \frac{1}{n}\sum_1^n E_z \inf_{\theta'\in U_s} \rho(y,\theta)\Big|_{z=x_{T_i}} - \frac{\delta'}{4} \geq \frac{1}{n}\sum_1^n \gamma_z(\theta_1)\Big|_{z=x_{T_i}} + \frac{\delta'}{2}.$$

We used assumption A1 and the equality

$$E\left\{ \inf_{\theta'\in U_s} \rho(Y_{T_i},\theta) \;\middle|\; F_{T_i} \right\} = E_z \inf_{\theta'\in U_s} \rho(y,\theta)\Big|_{z=x_{T_i}} .$$

On the other hand

$$\inf_{\theta > U} \frac{1}{n}\sum_1^n \rho(Y_{T_i},\theta) \leq \frac{1}{n}\sum_1^n \rho(Y_{T_i},\theta_1) \leq \frac{1}{n}\sum_1^n \gamma_z(\theta_1)\Big|_{z=x_{T_i}} - \frac{\delta'}{4}$$

for $n \geq N_0$ having used SLLN again. Thus for $n \geq \max_{i=0,..,s} N_i$ we get the assertion a/ of Lemma 2. The assertion b/ can be obtained in a similar way.

\- - -

Remark: Lemma 1 and Lemma 2 are formulated and proved in a more general pattern in Charamza 1988.

Proof of Theorem 1: The proof follows the lines of that in Mukerjee (1981). However we use the assertion of Lemma 2 (the SLLN for M-estimators) instead of Mukerjee's Lemma 1,which is in fact the SLLN for martingale differences. We also use general formula /9/ for representation of sample isotonic regression instead of Mukerjee's special case of this representation for $H(z)=z^2$.

APPENDIX 2

Proof of corollary 2:We set $a(y)=|y|$ and check the assumptions of the Theorem 1.

The condition A0 is also the assumption of corollary 2.

The condition A1 is fulfilled because

$P_z[\,||y-\theta|-|y||\geq x\,] = 0$ for all x greater then $|\theta|$.

Set $\theta_1=\alpha-\varepsilon_0/2, \theta_1'=\alpha+\varepsilon_0/2$. Now for any $\theta<\alpha+\varepsilon_0/4$ and any $z\in[x_M,\infty)$

$$\gamma_z(\theta) - \gamma_z(\theta_1') = \int |y-\theta| - |y-\theta_1'|\, dF_z(y) =$$

$$= \int_{-\infty}^{\theta} (\theta - \theta_1')\, dF_z(y) + \int_{\theta}^{\theta_1'} (2y-\theta-\theta_1')\, dF_z(y) + \int_{\theta_1'}^{\infty} (\theta_1'-\theta)\, dF_z(y) \geq$$

$$\geq (\theta_1 - \theta)(1 - 2F_z(\theta_1)) \geq (\theta_1 - \theta)2\delta_1 \geq \varepsilon_0\delta_1/2.$$

The inequalities follow from conditions /7b,c/.Setting $\varepsilon=\varepsilon_0/2$ we get condition A2'. Condition A2 can be obtained similarly.

Setting $b(\theta) = |\theta|+1$ we get from unboundedness of θ_z (that follows from /7a/) easily that assupmtion A3 is fulfilled.

From the assumption /7/ it follows that there exists $K_1>0$ such that $P_z[\ |y|\geq K_1\ ]\leq \frac{\varkappa}{3}$. We find that for any $K'<0$

$$\inf_{\theta<K'} \frac{|y-\theta|-|y|}{|\theta|+1} = \frac{-K'}{-K'+1} I[y>0] + \frac{2y-K'}{-K'+1} I[y\in[K',0>] + \frac{-y}{|y|+1} I[y<K'].$$

Hence $\inf_{\theta<K'} \frac{|y-\theta|-|y|}{|\theta|+1} \geq -1$ and for sufficiently small K'we obtain

$$E_z \inf_{\theta<K'} \frac{|y-\theta|-|y|}{|\theta|+1} \geq E_z \inf_{\theta<K'} \frac{|y-\theta|-|y|}{|\theta|+1} I[|y|<K_1]-E_z I[|y|\geq K_1] \geq$$

$$\geq E_z(1-\frac{\varkappa}{3})I[|y|<K_1]-\frac{\varkappa}{3} \geq 1 - \varkappa.$$

Thus the assumption A4' is fulfilled. The assumption A4 follows in a similar way.

From the unboundedness of the function $\inf\limits_{\theta<K'} \frac{|y-\theta|-|y|}{|\theta|+1}$ the assumption 5A' follows. Assumption 5A is also valid due to the same reason.

Assumption A6 is fulfilled due to the condition /7a/.

- - -

APPENDIX 3:

Here we prove Theorem 3. At first we shall prove two lemmas.

LEMMA 3: Let H be a convex function, $I \subset \{1,\dots,N\}$, and

$$\sum_{i\in I}\sum_{j=1}^{n_i} H(y_{ij} - t_1) \overset{(<)}{\leq} \sum_{i\in I}\sum_{j=1}^{n_i} H(y_{ij} - t_2). \qquad /L7/$$

Then for all $t\in[\min(t_1,t_2),\max(t_1,t_2)]$ the inequality

$$\sum_{i\in I}\sum_{j=1}^{n_i} H(y_{ij} - t) \overset{(<)}{\leq} \sum_{i\in I}\sum_{j=1}^{n_i} H(y_{ij} - t_2)$$

holds true.

PROOF: Choose $t\in(\min(t_1,t_2),\max(t_1,t_2))$ arbitrarily. There exists $\lambda\in(0,1)$ such that

$t=\lambda t_1+(1-\lambda)t_2$. From convexity and assumption /L7/ the following inequalities can be obtained.

$$\sum_{i\in I}\sum_{j=1}^{n_i} H(y_{ij} - t) \leq \sum_{i\in I}\sum_{j=1}^{n_i} \lambda H(y_{ij} - t_1)+(1-\lambda)H(y_{ij} - t_2) \overset{(<)}{\leq}$$

$$\overset{(<)}{\leq} \sum_{i\in I}\sum_{j=1}^{n_i} H(y_{ij} - t_2).$$

- - -

LEMMA 4: Let $t^*=(t^*,\dots,t^*)$ be a solution of /6/. For $t=(t_1,\dots t_N)$ let $t^* \leq t_1 \leq \dots \leq t_N$ (resp. $t_1 \leq \dots \leq t_N \leq t^*$). Then for any

$t'\in[t^*,t_1]$ (resp. $t'\in[t_N,t^*]$) the inequality

$$\sum_{i=1}^{N}\sum_{j=1}^{n_i} H(y_{ij} - t') \le \sum_{i=1}^{N}\sum_{j=1}^{n_i} H(y_{ij} - t_i)$$

holds true.

PROOF: Let us consider the case $t^*\le t_1\le\ldots\le t_N$. The other case can be proved in a similar manner. We shall prove that

$$\sum_{i=1}^{N}\sum_{j=1}^{n_i} H(y_{ij} - t_1) \le \sum_{i=1}^{N}\sum_{j=1}^{n_i} H(y_{ij} - t_i), \qquad /L8/$$

wherefrom the assertion of Lemma 4 follows using Lemma 3.

Let us denote $i_1<\ldots<i_k$ all such indexes i that the inequalities

$$\sum_{j=1}^{n_i} H(y_{ij} - t_i) \le \sum_{j=1}^{n_i} H(y_{ij} - t^*)$$

hold true. If there is no such index then /L8/ follows using Lemma 3. Set $i_0=1, i_{k+1}=N+1$. As the consequence of Lemma 3 we get for any $l\in\{1,\ldots,k\}$ the inequalities

$$\sum_{j=1}^{n_i} H(y_{ij} - t_i) \ge \sum_{j=1}^{n_i} H(y_{ij} - t_{i_l})$$

for all $i\in[i_l,i_{l+1}-1]$. It implies the relations

$$\sum_{i=i_l}^{i_{l+1}-1} \sum_{j=1}^{n_i} H(y_{ij} - t_i) \geq \sum_{i=i_l}^{i_{l+1}-1} \sum_{j=1}^{n_i} H(y_{ij} - t_{i_l}) \qquad /L9/$$

for all $l=1,\dots k$.

Using mathematical induction we prove

$$\sum_{l=z}^{k} \sum_{i=i_l}^{i_{l+1}-1} \sum_{j=1}^{n_i} H(y_{ij} - t_{i_l}) \geq \sum_{i=i_z}^{N} \sum_{j=1}^{n_i} H(y_{ij} - t_{i_z}) \qquad /L10/$$

for all $z \leq k$. For $z=k$ the inequality holds true as equality. Using the induction assumption we get

$$\sum_{l=z-1}^{k} \sum_{i=i_l}^{i_{l+1}-1} \sum_{j=1}^{n_i} H(y_{ij} - t_{i_l}) =$$

$$= \sum_{i=i_{z-1}}^{i_z-1} \sum_{j=1}^{n_i} H(y_{ij} - t_{i_{z-1}}) + \sum_{l=z}^{k} \sum_{i=i_l}^{i_{l+1}-1} \sum_{j=1}^{n_i} H(y_{ij} - t_{i_l}) \geq$$

$$\geq \sum_{i=i_{z-1}}^{i_z-1} \sum_{j=1}^{n_i} H(y_{ij} - t_{i_{z-1}}) + \sum_{i=i_z}^{N} \sum_{j=1}^{n_i} H(y_{ij} - t_{i_z}) \geq$$

$$\geq \sum_{i=i_{z-1}}^{N} \sum_{j=1}^{n_i} H(y_{ij} - t_{i_{z-1}}).$$

The last inequality follows again from Lemma 3 using the relations:a/ $t^* \le t_{i_{z-1}} \le t_{i_z}$

$$b/ \quad \sum_{i=i_z}^{N} \sum_{j=1}^{n_i} H(y_{ij} - t_{i_z}) \ge \sum_{i=i_z}^{N} \sum_{j=1}^{n_i} H(y_{ij} - t^*).$$

The relation b/ can be easily derived from the fact that t^* is a solution of /6/.

Hence L/10/ is proved also for z=0. This fact and /L9/ give /L8/.

- - -

PROOF OF THE THEOREM 2: Let $t=(t_1,\dots,t_N)$ be an arbitrary vector such that $t_1 \le \dots \le t_N$. Let us define $SP=\sup_{i\in I_1} t_{(i)}$, $IN=\inf_{i\in I_2} t_{(i)}$. Now define constant b in the following manner:

(i) if $t^* \le SP \le IN \le \tilde{t}$ then $b = IN$,

(ii) if $SP \le t^* \le IN \le \tilde{t}$ then $b = IN$,

(iii) if $SP \le IN \le t^* \le \tilde{t}$ then $b = t^*$,

(iv) if $t^* \le \tilde{t} \le SP \le IN$ then $b = SP$,

(v) if $t^* \le SP \le \tilde{t} \le IN$ then $b = \tilde{t}$,

(vi) if $SP \le t^* \le \tilde{t} \le IN$ then $b = t^*$.

Using Lemma 4 it can be shown that

$$\sum_{i\in I_1} \sum_{j=1}^{n_i} H(y_{ij} - t_{(i)}) \ge \sum_{i\in I_1} \sum_{j=1}^{n_i} H(y_{ij} - a)$$

$$\sum_{i\in I_2} \sum_{j=1}^{n_i} H(y_{ij} - t_{(i)+|I_1|}) \ge \sum_{i\in I_2} \sum_{j=1}^{n_i} H(y_{ij} - a) .$$

The theorem just follows.

- - -

REFERENCES:

[1]Barlow,R.E.,Bartholomew,D.J.,Bremner,J.M.,and Brunk,H.D.(1972). Statistical Inference Under Order Restrictions.Wiley,London.

[2]Charamza,P.(1989), PhD.thesis.(to appear) Department of Statistics, Charles university,Prague.(In Czech)

[3]Dupac,V.(1987).Quasi-isotonic regression and stochastic approximation.Metrika,vol 34,page 117-123.

[4]Ermoliev,Yu.,M., The Methods of Stochastic Programming, Moscow, Nauka,1976.(In Russian)

[5]Huber,The behavior of MLE under non standard conditions (1967).Proceedings of the Fifth Berkeley Symposium on Math. Statistics and Prob. theory. Univ. of California Press 1967.

[6]Loeve,M.,Probability Theory (1963),New York,Van Nostrand.

[7]Mukerjee,H.G.(1981).A stochastic approximation by observation on a discrete lattice using isotonic regression.The Annals of Math.Statistics,vol 9,n.5,page 1020-1025

[8]Nemirovski,A.S.,Polyak,B.T.,Cybakov A.B.(1984).Signals processing by non parametric methods of maximum likelihood Problemy peredaci informaciji,tom XX,vyp.3.(In Russian)

[9]Nevelson,Khasminskij,Stochastic Approximation and Recurrent Estimation,Moskva,Nauka,1972.(In Russian)

[10]Robbins H.,Lai T.L.(1981),Consistency and asymptotic efficiency of slope estimates in stochastic approximation schemes. Z.Wahrsch.Verw.Gebiete 56,N.3,pp 329-360

[11]Rockafellar,R.T.,Wets,G.,(1982),On the interchange of subdifferentiation and conditional expectation for convex functionals. Journal of Stochastics 1982,7,173-182.