# STOCHASTIC APPROXIMATION UNDER DEPENDENT NOISES

O.N.Granichin

St.-Petersburg University, 198904

St.-Petersburg-Petrodvoretz, Russia

**ABSTRACT**

The problem of estimating the unknown function minimum point is considered in the situation where function values are measured at some points with random dependent (or independent) errors. The difference from the ordinary stochastic approximation is in the fact that measurement points are chosen with additive random measurable noises - probing noises. The probability one and mean square convergence is proved by the assumption of the probing and measure noises independence. The method may be used for solving the problem of the spotting signal with dependent noises.

## 1. INTRODUCTION

Let $f_\tau(\theta): \Theta' \to R^1$, $\Theta' \subset R^M$ be a family of unknown functions, $\tau \in I \subset R^M$ be random parameter with the distribution $P_\tau$ and mean value $\bar{\tau}$. The function $f(\theta)=f_{\bar{\tau}}(\theta)$ is considered. Let $f(\theta)$ has the unique minimum point $\theta_*$ in $\Theta'$.

We consider the following problem: estimate $\theta_*$ from given measurements $y_n=f_{\tau_n}(x_n)+\xi_n, n=1,2,\ldots,$ where points $\{x_n\}$ are generated by some random source, $\{\xi_n\}$ are unmeasurable noises, $\{\tau_n\}$ are random with the identical distribution $P_\tau$.

The difference from the classical framework of the stochastic approximation is , at first, in the fact that points $\{x_n\}$ doesn't coincide only with estimates generated by the algorithm, at second, usually the minimum fixed function $f(\theta)$ point estimating problem is considered [1-6]. There are many solved problems of the stochastic approximation methods convergence , but almost all of them is considered cases where measure noises $\{\xi_n\}$ are independent or may be hold properties of "wrong perturbation".

This paper is devoted to the general algorithm of the stochastic approximation with additive probing noises. It was proposed and investigated by Polyak and Tsybakov [4] at the cases of independent measure noises and fixed function $f(\theta)$ and by the author of this paper in [5,6] at particular cases with dependent measure noises. In this paper more general problem is considered. The investigated functions dependence on the unknown parameter $\tau$ is promised to get additional practical

Here $P_\Theta$ is the projection operator on $\Theta$, $\theta_n$ is a current approximation to $\theta_*$, $\phi_n(z),\psi_n(z):\mathbb{R}^K\longrightarrow\mathbb{R}^N$, $n=1,2,\dots$ are sequences of vector-functions satisfying for all $\theta\in\Theta$

$$(2)\quad \int\Big(\sum_{|m|\le\ell}\frac{1}{m!}D^m f_\tau(\theta)\phi_n(z)^m\Big)\psi_n(z)d\mathbb{P}_\zeta(z)=C_0 n^{-1}\nabla f_\tau(\theta)$$

and

$$\int\psi_n(z)d\mathbb{P}_\zeta(z)=0,\quad \int(\psi_n(z),\psi_n(z))d\mathbb{P}_\zeta(z)\le C_\psi n^{-2+\frac{1}{\beta}},$$

$$(3)\quad \int(\phi_n(z),\phi_n(z))^\beta d\mathbb{P}_\zeta(z)\le\Delta^{2\beta}n^{-1},\quad |\phi_n(z)|\le A$$

At all time $n$ we may to measure $y_n$ and $\psi_n(\zeta_n)$.

<u>Theorem</u>. Let the conditions 1-5 for a family of functions $f_\tau(\cdot)$ hold, $\beta\ge 2$, measure noises $\{\xi_n\}$ : $\mathbb{E}\xi_n^2<\sigma^2$, for all fixed $n$ $\zeta_n$ and $\xi_1,\dots,\xi_n,\tau_1,\dots,\tau_n$ are independent. Then

$$(4)\qquad \mathbb{E}\{|\theta_n-\theta_*|^2\}n^{(\beta-1)/\beta}<\infty,$$

and $\theta_n\longrightarrow\theta_*$ as $n\longrightarrow\infty$ with probability one.

Notes, conditions (2)-(3) for functions $\phi_n(\cdot),\psi_n(\cdot)$ hold, for example, for functions ([4])

$$(5)\quad \phi_n(z)=\Delta n^{-1/(2\beta)}z,\ \psi_n(z)=\psi n^{-1+1/(2\beta)}K(z),$$

where the vector-function $K(z)$ is determined by orthogonal Legendre's polynoms $p_j(\cdot)$, $j=0,1,\dots,\ell$ [2,4] :

$$K_i(z)=K_0(z_i)\prod_{m\ne i}\tilde K(z_m),\ i=1,\dots,N\ ,$$

$$z=(z_1,\dots,z_N)\in\mathbb{R}^N,$$

$$(6)\quad K_0(u)=\sum_{j=0}^{\ell}\alpha_j p_j(u)\ ,\quad \alpha_j=p_j'(0)/\int p_j^2(u)du\ ,$$

$$\tilde K(u)=\sum_{j=0}^{\ell-1}\beta_j p_j(u)\ ,\quad \beta_j=p_j(0)/\int p_j^2(u)du\ ,$$

and $\zeta_n$, $n=1,2,\dots$ random independent uniformly distributed on $[-1/2,1/2]^N$ vectors.

Proof. Since for the sufficiently large $n$ $\theta_*\in\Theta_n$ then by the projection property we have

$$|\theta_n-\theta_*|^2\le|\theta_{n-1}-\theta_*-y_n\psi_n(\zeta_n)|^2.$$

Applying the conditional expectation of both sides of given $\theta_0,\dots,\theta_{n-1}$ we have

$$\mathbb{E}\{|\theta_n-\theta_*|^2|\theta_0,\dots,\theta_{n-1}\}\le|\theta_{n-1}-\theta_*|^2-$$

$$(7)\qquad -2(\theta_{n-1}-\theta_*\ ,\mathbb{E}\{y_n\psi_n(\zeta_n)|\theta_0,\dots,\theta_{n-1}\})+$$

$$+\mathbb{E}\{|y_n|^2|\psi_n(\zeta_n)|^2|\theta_0,\dots,\theta_{n-1}\}\ .$$

Since conditions 4,5 for $f_\tau(\cdot)$ hold then we have uniformly in $\tau\in A$

$$f_\tau^2(\theta_{n-1}+\phi_n(z))\le 2B^2+4C^2(|\theta_{n-1}-\theta_*|^2+|\phi_n(z)|^2)+$$

$$+8A^2(|\theta_{n-1}-\theta_*|^4+|\phi_n(z)|^4).$$

For the last item in (7) we obtain

$$\mathbb{E}\{|y_n|^2|\psi_n(\zeta_n)|^2|\theta_0,\dots,\theta_{n-1}\}\le$$

$$\le 2\Big(\int\int f_\tau^2(\theta_{n-1}+\phi_n(z))|\psi_n(\zeta_n)|^2d\mathbb{P}_\zeta(z)d\mathbb{P}_\tau+$$

$$(8)\qquad +\mathbb{E}\{(\xi_n\psi_n(\zeta_n))^2|\theta_0,\dots,\theta_{n-1}\}\Big)\le$$

$$\le(8C^2+16A^2)n^{-2+1/\beta}|\theta_{n-1}-\theta_*|^2+$$

$$+2(\mathbb{E}\{\xi_n^2|\theta_0,\dots,\theta_{n-1}\}+(4B^2+8C^2\Delta+16A^2\Delta^2)C_\psi)n^{-2+\frac{1}{\beta}}$$

Consider the second item in the right hand (7). Since $\xi_n$ and $\zeta_n$ are independent then according to conditions (2)-(3) for $\phi_n(\cdot),\psi_n(\cdot)$ and property 1 of a functions $f_\tau(\cdot)$ family we have

$$-2(\theta_{n-1}-\theta_*,\ \mathbb{E}(y_n\psi_n(\zeta_n)|\theta_0,\dots,\theta_{n-1})) =$$

$$=-2(\theta_{n-1}-\theta_*\cdot\iint(f_\tau(\theta_{n-1}+\phi_n(z))\ \psi_n(\zeta_n)dF_\zeta(z)dF_\tau)-$$

$$-2(\theta_{n-1}-\theta_*,\ \mathbb{E}(\xi_n|\theta_0,\dots,\theta_{n-1})\mathbb{E}\psi_n(\zeta_n))\le$$

$$\le-2C_n\int(\theta_{n-1}-\theta_*,\ \nabla f_\tau(\theta_{n-1}))n^{-1}dF_\tau+$$

$$+LC\sqrt{\Delta^{2\beta}C_\psi}\ n^{-\frac{3}{2}+\frac{1}{2\beta}}\ |\theta_{n-1}-\theta_*|.$$

But

$$En^{-1/2}|\theta_{n-1}-\theta_*|\le E^2\varepsilon^{-1}/4+\varepsilon n^{-1}|\theta_{n-1}-\theta_*|^2$$

for any $E,\varepsilon>0$. According to property 4 of a family of functions $f_\tau(\cdot)$ we find

$$-2(\theta_{n-1}-\theta_*,\ \mathbb{E}(y_n\psi_n(\zeta_n)|\theta_0,\dots,\theta_{n-1}))\le \tag{9}$$
$$\le-(2\delta C_n-\varepsilon)n^{-1}|\theta_{n-1}-\theta_*|^2+\frac{1}{8}\varepsilon^{-1}L^2C^2\Delta^{2\beta}C_\psi n^{-2+\frac{1}{\beta}}.$$

Let $\varepsilon=\delta C_n$. It follows from (7)-(9) that

$$\mathbb{E}(|\theta_n-\theta_*|^2|\theta_0,\dots,\theta_{n-1})\le \tag{10}$$
$$\le|\theta_{n-1}-\theta_*|^2(1-\delta C_n n^{-1}+(8C^2+16A^2)n^{-1-\frac{\beta-1}{\beta}})+\gamma_n,$$

where

$$\gamma_n=\Big((4B^2+8C^2\Delta+16A^2\Delta^4)C_\psi+\frac{1}{8}L^2C^2\delta^{-1}C_n^{-1}\Delta^{2\beta}C_\psi+$$
$$+\mathbb{E}(\xi_n^2|\theta_0,\dots,\theta_{n-1})\Big)n^{-2+\frac{1}{\beta}}.$$

Now apply the unconditional expectation from both sides of (10) we find

$$\mathbb{E}(|\theta_n-\theta_*|^2)\le\mathbb{E}(|\theta_{n-1}-\theta_*|^2)(1-\delta C_n n^{-1})+C_\gamma n^{-2+\frac{1}{\beta}},$$

where

$$C_\gamma=(4B^2+8C^2\Delta+16A^2\Delta^4)C_\psi+\frac{1}{8}L^2C^2\delta^{-1}C_n^{-1}\Delta^{2\beta}C_\psi+\sigma^2.$$

From the lemma on numerical sequences ([3], ch.2, Lemma 3) it follows that $\mathbb{E}(|\theta_n-\theta_*|^2)\longrightarrow0$ as $n\longrightarrow\infty$ and $\mathbb{E}(|\theta_n-\theta_*|^2)n^{(\beta-1)/\beta}<\infty$.

Consider the random sequence $\{\nu_n\}$

$$\nu_n=\nu_n(\theta_0,\dots,\theta_{n-1})=$$

$$-\sum_{k=1}^{n-1}(\gamma_k+(8C^2+16A^2)|\theta_k-\theta_*|^2)(k+1)^{-1+(\beta-1)/\beta}.$$

There is a supermartingal since $\mathbb{E}(\nu_n|\nu_1,\dots,\nu_{n-1})\le\nu_{n-1}\le0$. Since $\mathbb{E}|\nu_n|<\infty$ according to famous Doob's theorem [7] it follows that for some $\nu_*$ $\nu_n\longrightarrow\nu_*$ as $n\longrightarrow\infty$ with probability one. For the sequence $\{\eta_n\}$: $\eta_n=|\theta_n-\theta_*|^2-\nu_n$ from (10) we also have

$$\mathbb{E}(\eta_n|\eta_1,\dots,\eta_{n-1})\le$$

$$\le|\theta_{n-1}-\theta_*|^2(1-\delta C_n n^{-1}+(8C^2+16A^2)n^{-1-(\beta-1)/\beta})-$$

$$-\nu_n\le|\theta_{n-1}-\theta_*|^2-\nu_{n-1}=\eta_{n-1}.$$

Since $\mathbb{E}(\eta_n-|\eta_n|)>-\infty$ then from Doob's theorem [7] it follows that $\eta_n\longrightarrow\eta_*$ as $n\longrightarrow\infty$ with probability one. Since $\mathbb{E}(|\theta_n-\theta_*|^2)\longrightarrow0$ as $n\longrightarrow\infty$ we obtain $|\theta_n-\theta_*|\longrightarrow0$ as $n\longrightarrow\infty$ with probability one.

## 4. SPOTTING SIGNAL WITH DEPENDENT NOISE

For example we consider the problem of the spotting signal with dependent noises: a mean value of the random sequence $\{\tau_n\}$ is to be discovered from given measurements

$$y_n=\tau_n\zeta_n+\nu_n,\quad n=1,2,\dots, \tag{11}$$

where the signal $\{\zeta_n\}$ is the sequence of the random measurable independent and uniformly distributed on $[-1/2,1/2]$ values, $\{\nu_n\}$ are measurement noises, $\zeta_n$ and $\nu_n$ are independent, $\{\tau_n\}$ random identically capabilities of the considered algorithm. For example, we may consider the somebody internal structure investigation problem by testing series of some impulses. In some cases the resulting behavior of the detector is determined not only by structure of the body, probing series of impulses and measure noises, but by some probability characteristics of the body behavior.

The last part of the paper deals with the problem of the spotting signal with dependent noises. The measure canal is described by the linear model with additive noises. The algorithm of the stochastic approximation with additive probing noises is used for constructing estimates of the signal unknown parameter. The almost sure and means square convergence is proved.

## 2. PROBLEM STATEMENT

Let $f_\tau(\theta): \Theta' \to \mathbb{R}^1$, $\Theta' \subset \mathbb{R}^N$ be a family of nonlinear functions, $x_1, x_2, \ldots$ be measurement points, $n=1,2,\ldots$ time series, and

$$y_n = f_{\tau_n}(x_n) + \xi_n, \quad n=1,2,\ldots$$

be results of measuring $f_{\tau_n}(\cdot)$ at those points containing the random noise $\xi_n$, $\{\tau_n\}$ be the sequence of random vectors with the identical distribution $\mathbf{P}_\tau$ and mean value $\tau$, $\tau \in T \subset \mathbb{R}^K$, $T$ - some set.

Let us formulate the main assumptions.

1. For all $\tau \in T$ the function $f_\tau(\theta) \in C^\ell$ is the $\ell$ times differentiable function with $\ell$-th derivative satisfying on $\Theta'$ the Hoelder's conditions with constant $\alpha$, $0<\alpha\le 1$

$$\Big|f_\tau(z) - \sum_{|m|\le \ell} \frac{1}{m!} D^m f_\tau(\theta)(z-\theta)^m\Big| \le L|z-\theta|^\beta,$$

where $\beta = \ell + \alpha$, $m=(m_1,\ldots,m_N) \in \mathbb{N}^N$, $m_i \ge 0$, $u^m = u_1^{m_1} \ldots u_N^{m_N}$, $u=(u_1,\ldots,u_N) \in \mathbb{R}^N$, $D^{(m)} = \partial^{(m)}/\partial u_1^{m_1} \ldots \partial u_N^{m_N}$.

2. The function $f(\theta) = f_{\bar\tau}(\theta)$ $(\bar\tau = \mathbb{E}\tau)$ has unique minimum in $\Theta \subset \Theta'$ at the internal point $\theta_*$, where $\Theta$ is the convex closed bounded set: for some $\Delta>0$ $\forall$ $\theta \in \Theta$ $dist(\theta, \partial\Theta') > \Delta$.

3. For all $\theta \in \Theta$

$$\int (\theta-\theta_*, \nabla f_\tau(\theta)) d\mathbf{P}_\tau \ge \delta |\theta-\theta_*|^2.$$

4. For all points $\theta \in \Theta$, $\tau \in T$

$$|\nabla f_\tau(\theta) - \nabla f_\tau(\theta_*)| < A|\theta-\theta_*|.$$

5. For all $\tau \in T$ $|f_\tau(\theta_*)| \le B$, $|\nabla f_\tau(\theta_*)| \le C$. Here and further on $|x|$ denotes the Euclidean norm, and $(\cdot,\cdot)$ - inner product, and $A, B, C, L, \alpha, \beta, (A>\alpha)$ are some positive constants.

The minimum function $f(\theta)$ point $\theta_*$ is to be estimated from given measurements $y_n$, $n=1,2,\ldots$.

## 3. MAIN RESULT

Let $\{\zeta_n\}$, $\zeta_n \in Z \subset \mathbb{R}^K$, $n=1,2,\ldots$ be random, independent (on each other) and identically distributed (i.i.d.) vectors with the distribution $\mathbf{P}_\zeta$ on Z.

The following recursive algorithm is proposed:

$$\theta_0 = b, \quad x_n = \theta_{n-1} + \phi_n(\zeta_n),$$

$$\text{(1)} \qquad \theta_n = \mathcal{P}_\Theta(\theta_{n-1} - y_n \psi_n(\zeta_n)),$$

$$y_n = f_{\zeta_n}(x_n) + \xi_n$$

distributed on segment $[\tau_1,\tau_2]$ values with the distribution $F$ and mean value $\bar{\tau}$, $\zeta_n$ and $\tau_n$ are independent. If $\bar{\tau}=0$ then the signal is absent, otherwise the signal is present.

Consider the problem of estimating $\bar{\tau}$ from given measurements $y_n$, $n=1,2,\ldots$.

Let $\beta\geq 2$ and $Ev_n^2\leq C_v n^{1/\beta}$, $p_j(\cdot)$ be orthogonal Legendre's polynoms of the order $j$, $j=0,\ldots,\ell$, $\alpha_j$, $j=0,\ldots,\ell$ be constants

$$\alpha_j=p_j'(0)/\int p_j^2(u)du.$$

According the theorem of p.3 the algorithm

$$\theta_0=(\tau_1+\tau_2)/2,$$

$$(12)\quad \theta_n=P_{[\tau_1,\tau_2]}(\theta_{n-1}-n^{-1}\sum_{j=0}^{\ell}\alpha_j p_j(\zeta_n)\times$$

$$\times(\frac{1}{2}(n^{1/\beta}\theta_{n-1}+n^{1/(2\beta)}\zeta_n)^2-y_n))$$

provides convergence of the sequence $\{\theta_n\}$ to $\bar{\tau}$.

Let $f_\tau(\theta)=\frac{1}{2}(\theta-\tau)^2$ $:f_\tau(\theta):[\tau_1,\tau_2]\longrightarrow R^1$ be a family of quadratic functions. For this family conditions 1-4 hold by any $\beta\geq 2$. Consider the sequence of points in $R^N$ $\{x_n\}$: $x_n=\theta_{n-1}+n^{-1/(2\beta)}\zeta_n$, $n=1,2,\ldots$ and some sequence of measurable values $\{Y_n\}$: $Y_n=\frac{1}{2}x_n^2-n^{-1/(2\beta)}y_n$. The functions $\phi_n(z)=n^{-1/(2\beta)}z$ and $\psi_n(z)=n^{-1+1/(2\beta)}z$ satisfy (2) and (3). For measurements $Y_n$, $n=1,2,\ldots$ we have

$$Y_n=f_{\tau_n}(x_n)+\xi_n,\ n=1,2,\ldots,$$

where for all $n=1,2,\ldots$ $\xi_n=\frac{1}{2}\tau_n^2-(\theta_{n-1},\tau_n)-v_n n^{-1/(2\beta)}$ and $\zeta_n$ are independent since $\{\zeta_n\}$ and $\{v_n\}$, $\{\tau_n\}$ are independent. Since

$$\theta_n=P_{[\tau_1,\tau_2]}(\theta_{n-1}-n^{-1+1/(2\beta)}K(\zeta_n)Y_n)$$

then the type of algorithm (12) is similar to (1). The point $\bar{\tau}$ is the function $f(\theta)=f_{\bar{\tau}}(\theta)$ minimum. The convergence rate of algorithm (13) is $O(n^{-(\beta-1)/\beta})$.

## REFERENCES


1. Ljung L., Soderstrom T. Theory and Practice of Recursive Identification. MIT Press. Cambridge. 1983. 529p.

2. Katkovnik V.,Ya. Linear Estimates and Stochastic Optimization Problems. Nauka, Moscow, 1984 (in Russian).

3. Polyak B.T. Introduction to Optimization. Optim.Software, New York, 1987.

4. Polyak B.T., Tsybakov A.B. Optimal Orders of Accuracy for Search Algorithms of Stochastic Optimization. Problems Inform. Transmission, v.26(1990), no 2, p.126-133.

5. Granichin O.N. A Stochastic Recursive Procedure With Dependent Noises in the Observation That Uses Sample Perturbations in the Input. Vestnik Leningrad Univ. Math., v.22(1989), no.1, p.27-31.

6. Granichin O.N. Stochastic Approximation With Sample Perturbations in the Input. Autom. and Remote Control no 2, 1992, p.

7. Doob J.L. Stochastic processes. New York. Tonh Wiley & Sons. 1953. 605p.