JSST2000

International Conference on Modeling, Control and Computation in Simulation

Proceedings of
The JSST 2000 International Conference
on Modeling, Control and Computation in Simulation
Tokyo, 24 ~ 26 October 2000

◆

Edited by

T. KAGAWA

The Precision and Intelligence Laboratory
Tokyo Institute of Technology
Japan

Japan Society for Simulation Technology

Unknown Parameters Identification With Excite Test Signals

Oleg N. Granichin

St. Petersburg State University, Dept.of Mathmatics and Mechanics
Bibliotechnaya pl. 2, 198904, St. Peterrsburg, Russia
e-mail: oleg_granitchin@mail.ru

Key words: stochastic approximation, SPSA, exciting perturbation, consistency estimates, regression function, conditional mean value.

1 Introduction

Recently the electronics development moved close to the step of intelligent control device creating. Even now we have a real opportunity of the effective using of new mathematics algorithms of the dynamic systems unknown parameters identification theories, optimal and adaptive control, mathematical design of experiment theory during the practice problems solving.

For some problems earlier we could only try to get the most probable set, which contains a vector of unknown object parameters (or of functioning dynamic system). But now the opportunity for getting rather exact decisions for such types of problems appears. Solving tasks, which find exact unknown parameters data, was often unavailable owing to calculating difficulties. The exact decision of any problem can be find if there is an exact target setting. But in our real existing world all the connections and the relationships are so difficult and many-sided, that it is practically impossible to give a strict mathematical description of many phenomena. A typical theoretic approach is to chose a mathematical model close to real processes and to "include" into it different noises which from one side are some kind of "roughness" of the mathematical model and from another one are characteristics of unobserved object or system outside perturbations. During the last 40 years in majority of mathematical researches "useful" statistic properties are arrogated to noises to develop solving task algorithms and at this base consistency of an algorithm is proved mathematically. In the engineering practice the algorithms based on "least-squares method" or on "maximum likelihood method" are used very often without enough basis of statistic obstruction properties assumptions. As a matter of fact it is inexpediently to use most that algorithms in the conditions of possible enemy counteraction. It is well known for specialists in the theory of unknown parameters identification that if the observation noises is a determined unknown function (the enemy "jam" the signal) or observation noises are a "dependent" sequence then in that case the getting decisions by many algorithms are wrong. In that cases the theorists say that observation sequence is "singular" and such decisions are not concerned. To "enrich" the information in the observation channel a test exciting noise with the well-known (input) statistic properties is supposed to be included into the system. It can be reached either by adding a trial impact through the system control channel (if it exists) or by random way forming of an observation plan (experiment). The theoretic bases of the corresponding algorithms can be found in a series of works of Granichin O.N., Fomin V.N., Polyak B.T., Tsybakov A.B., Han Fu-Chen, Lei Guo,J.Spall [5-11] and others. A considerable restriction of using these algorithms is an assumption about a "week correlation" ("independence") of adding trail noise into the system, while there are no other assumptions about noises properties. The restriction at the exciting noise trail properties and external noise is natural, if noise is a determined function, or any other foreign noise generated by the enemy, who understood that we try to determine or investigate him. It is possible to elaborate the suggestions about identification methods, which include trail exciting noise, for some tasks solving. In itself this approach has a universal nature and may be used in solving different problems, updating the exist algorithms, which are in use. Further I'll try to describe it to stochastic approximation.

2 Stochastic approximation with exciting perturbation under dependent noises

The main ideas of stochastic approximation was formulated by Robbins and Monro [1]. Let $x \in \mathbf{R}$ is "controllable" variable and for each x we can measure random value $Y(x)$ with distribution $P(Y(x) < y) = F_x(y)$. Let $M(x)$ is regression function $Y(x)$ on x:

$$M(x) = \int_{-\infty}^{+\infty} Y(x) dF_x(y).$$

Robbins and Monro investigated problem of finding θ unique root of regression equation $M(x) = 0$. Let $M(x)$ is increasing function. Consider recurrent sequence

$$x_{n+1} = x_n - a_n Y_n,$$

where $\{a_n\}$ are positive number and Y_n is result of measurement by x_n. Robbins and Monro shown convergence x_n to θ as sure with natural proposals about $\{a_n\}$ and statistical properties of measurement errors. In this algorithm one can considers statistic

$$z_0 = Y(x)$$

for estimate value of regression function $M(x)$. Usually this is consistency estimate of $M(x)$

$$E_x\{z_0\} = M(x),$$

($E_x\{...\}$ conditional mean value).

The problem regression function minimum point estimation was considered by Kiefer and Wolfowits in [2]. The main idea was to solve the equation $M'(x) = 0$. For this purpose one can measure values Y in two points $x + c, x - c$ and for derivation of regression function in point x one can use statistic

$$z_1 = \frac{1}{2c}(Y(x+c) - Y(x-c)).$$

For the estimation of minimum point Kiefer and Wolfowits proposed algorithm

$$x_{n+1} = x_n - a_n z_{1_n},$$

where $\{a_n\}, \{c_n\}$ some sequences of positive numbers.

In many cases from the practical point of view we don't able to know enough information about statistical properties of measurement errors or it can be deterministic function. There are some problem to establish convergence of ordinary Robbins-Monro or Kiefer-Wolfowits algorithms.

The performance of stochastic approximation algorithms depends from accuracy of estimate $M(x)$ or $M'(x)$. If we change statistics z_0 and z_1 for another we can hope to get better performance. For this purpose we need in better approximation $M(x)$ or $M'(x)$, Fabian [3] modified Kiefer and Wolfowits algorithm. They proposed to use difference approximations of high derivations with some weight. If l is number of continuous derivation of regression function $M(x)$, then Fabian's algorithm provides mean square rate of convergence as $O(n^{-\frac{l-1}{2}})$ for odd l. $>$From computational point of view Fabian's algorithm is very complicated.

In this paper for estimation $M^{(0)}(x) = M(x)$ or $M^{(1)}(x) = M'(x)$ in Robbins-Monro or Kiefer-Wolfowits algorithms we propose to use statistics Z_0 и Z_1

$$Z_{r_n} = h_n^{-r} K_r(\zeta_n) Y(x_n + h_n \zeta_n), r = 0, 1,$$

instead z_0 or z_1. Here $\{h_n\}$ is some sequence of positive number, $\{\zeta_n\}$ is sequence of independent random values which are uncorrelated with errors of measurement on step n, $K_0(\zeta), K_1(\zeta)$ are some kernel function on $\mathbf{R}$. For some $\{h_n\}, \{\zeta_n\}, K_r(\zeta), r = 0, 1$ estimates Z_{r_n} are consistency estimates of $M^{(r)}(x)$:

$$E_{x_n}\{K_r(\zeta_n) h_n^{-r} Y(x_n + h_n \zeta_n\} = M^{(r)}(x_n) + o(h_n^{p-r}), r = 0, 1.$$

Here p is some index of regression function $M(x)$ (in particular, $p = l$ if $l \in \mathbf{N}$ and $M(x)$ has $l - 1$ derivations satisfies Lipchits conditions). Such a stochastic approximation algorithm with additive perturbation noises $h_n \zeta_n$ has recently been investigated by Polyak and Thybakov [4] for independent measurement noises $Y_n - M(x_n + h_n \zeta_n)$ and by Granichin [5],[6],[7],[8] for some special cases with dependent measurement noises. Polyak and Tsybakov have shown that this algorithm had an optimum minimax rate of convergence in wide variety of algorithms. This algorithm has mean square rate of convergence $O(n^{-\frac{p-1}{p}})$.

We propose to consider two new algorithms as Robbins-Monro and Kiefer-Wolfowits. This algorithms provide convergence as sure with high rate. It is possible to prove an asymptotic normality of distribution of random values $x_n - \theta$ after some proposal about statistical properties of measurement errors. We calculate the asymptotic variance of $x_n - \theta$. This approach leads to way of optimal choosing of kernel function $K_r(\zeta)$. The last section deal with new adaptive Robbins-Monro algorithm, which every time use $K_0(\zeta)$ and $K_1(\zeta)$ for estimation function regression root θ and $M^{(1)}(\theta)$. This algorithm has optimal perfomance: minimum variance of asymptotic normal distribution $x_n - \theta$.

3 Differentiation kernel

Let $\{p_m(u)\}$ is some system of orthogonal polinoms on some interval $[-\gamma, \gamma]$ with degree below l and weight function $\psi(u) \geq 0, \gamma > 0$. Then

$$\int_{-\gamma}^{\gamma} \psi(u) p_i(u) p_j(u) du = a_i \delta_{i,j}, \int_{-\gamma}^{\gamma} \psi(u) du = 1, \tag{1}$$

for $i, j = 1, ..., l$, where $\delta_{i,j}$ is equal 1, if $i = j$, and 0 if $i \neq j$, $a_i = \int_{-\gamma}^{\gamma} \psi(u) p_i^2(u) du$ is some constants.

Define the functions $K_r(u), r = 0, 1$ on interval $[-\gamma, \gamma]$ as linear combination of polinoms $p_m, m = 1, ..., l$

$$K_r(u) = \sum_{m=0}^{l} \frac{p_m^{(r)}(0)}{a_m} p_m(u). \tag{2}$$

We can see

$$\int_{-\gamma}^{\gamma} \psi(u) K_r(u) u^q du = \delta_{q,r}, \tag{3}$$

for any $q \in \mathbf{Z}, q \leq l$.

Let function f has l times continuous derivations near point x_0 on $\mathbf{R}$. We have

$$f(x_0 + cu) = \sum_{i=0}^{l} \frac{f^{(i)}(x_0)}{i!} (cu)^i + o(u^l).$$

Consider integral representation of function f with kernel K_r

$$\frac{1}{c^2} < f(x_0 + cu), K_r(u) >= \frac{1}{c^2} \int_{-\gamma}^{\gamma} \psi(u) K_r(u) f(x_0 + cu) du \tag{4}$$

We can obtain

$$\frac{1}{c^2} < f(x_0 + cu), K_0(u) >= f(x_0) + \int_{-\gamma}^{\gamma} \psi(u) K_0(u) o(u^l) du, \tag{5}$$

$$\frac{1}{c^2} < f(x_0 + cu), K_1(u) >= f^{(1)}(x_0) + \int_{-\gamma}^{\gamma} \psi(u) \frac{K_1(u)}{c} o(u^l) du. \tag{6}$$

Equations 5 and 6 shows the main idea of new stochastic approximation algorithms listed below.

Note

$$K_0(0) = \int_{-\gamma}^{\gamma} \psi(u) K_0(u)^2 du, \tag{7}$$

$$K_1^{(1)}(0) = \int_{-\gamma}^{\gamma} \psi(u) K_1(u)^2 du. \qquad (8)$$

We can use Legendre's or Chebuchev's polinoms to build kernel functions $K_r(u), r = 0, 1$, for example. The values $K_0(0)$ and $K_1'(0)$ have importance role in calculation variance of asymptotic distribution $x_n - \theta$ We have for Legendre's polinoms

$$K_0(0) = \sum_{m=0}^{[\frac{l+1}{2}]} [\frac{(2m-1)!!}{2m!!}]^2 (4m+1),$$

$$K_1^{(1)}(0) = \frac{1}{\gamma^2} \sum_{m=0}^{[\frac{l-1}{2}]} (4m+3)(1+\frac{1}{2})^2(1+\frac{1}{4})^2 ... (1+\frac{1}{2m})^2,$$

and for Chebuchev's polinoms

$$K_0(0) = 1 + \frac{1}{2}[\frac{l+1}{2}], K_1'(0) = \frac{1}{\gamma^2} 2([\frac{l-1}{2}]+1)^2(\frac{3}{4}[\frac{l-1}{2}]([\frac{l-1}{2}]+2)+1),$$

[...] is entire function.

4 Convergence and asymptotic normality

Let all random values define on some fixed probability space (Ω, F, P).

Let regression function $M(x)$ define on some compact set $\Theta \in \mathbf{R^N}$. It has l times continuous derivations on Θ and $M^l(x)$ which satisfy Hoelder's conditions with some constant $\alpha, 0 < \alpha \leq 1$ so that

$$M(x_0 + t) = \sum_{m=0}^{l} \frac{M^{(m)}(x_0)}{m!} t^m + o(|t|^p|), \qquad (9)$$

where $p = l + \alpha$.

Theorem 1 *Let random sequences $\{x_n\}$ is "own" design of an experiment and $\{Y_n\}$ is result of measurements(or observations), $E\{x_1\} < \infty$, $\{\zeta_n\}$ is exciting perturbation, the sequence of independent random values with same distributions on some interval $[-\gamma, \gamma](0 < \gamma < \infty)$ with distribution density $\psi(u)$, h and a are some positive constants, the real design of an experiment is determined by summa $x_n + \frac{h}{n^{\frac{1}{2p}}}\zeta_n$, if errors of measurement are random values then random values ζ_n and $Y_1 - M(x_1 + \frac{h}{1}\zeta_1), ..., Y_{n-1} - M(x_{n-1} + \frac{h}{(n-1)^{\frac{1}{2p}}}\zeta_{n-1})$ are uncorrelated,$n = 1, 2, ...$ and measurement errors are satisfied*

$$E\{(Y_n - M(x_n + \frac{h}{n^{\frac{1}{2p}}}\zeta_n))^2\} \leq \sigma^2, E\{(Y_n - M(x_n + \frac{h}{n^{\frac{1}{2p}}}\zeta_n))^2\} \to \sigma^2(\theta)$$

as $x_n \to \theta$, there is some positive constant $\lambda > 0$ so that for any $q > 0$

$$E\{(Y_n - M(x_n + \frac{h}{n^{\frac{1}{2p}}}\zeta_n))^2 \mathbf{1}_{\{(Y_n - M(x_n + \frac{h}{n^{\frac{1}{2p}}}\zeta_n))^2 \geq q n^{\lambda}\}}\} \to 0$$

as $n \to \infty$*,(*$\mathbf{1}_{\{...\}}$ *is indicator function).*

For this conditions we have

1) If $l \geq 0$*, regression equation* $M(x) = 0$ *has the unique root on* Θ *in the point* θ*,*

$$M^{(1)}(\theta) > \frac{1}{2a} \tag{10}$$

there is $B > 0, D > 0$ *so that for any* $x \in \mathbf{R}$

$$|M(x)| \leq B + D|x| \tag{11}$$

for any positive ϵ

$$inf_{\epsilon<|x-\theta|<\epsilon^{-1}}\{sign((x-\theta)M(x))\} > 0, \tag{12}$$

then estimates $\{x_n\}$ *which formed by*

$$x_{n+1} = P_\Theta\{x_n - \frac{a}{n}K_0(\zeta_n)Y_n\} \tag{13}$$

($P_\Theta\{...\}$ *is projection operator) satisfy convergence* $x_n \to \theta$ *as sure. With some additional proposals the random value* $(x_n - \theta)n^{\frac{1}{2}}$ *has asymptotically normality distribution with mean value* 0 *and variance*

$$\frac{a^2\sigma^2(\theta)K_0(0)}{2aM^{(1)}(\theta) - 1}, \tag{14}$$

2) If $l \geq 1$*, regression function* $M(x)$ *has the unique minimum point on* Θ *in the point* θ*,*

$$M^{(2)}(\theta) > \frac{p-1}{2pa} \tag{15}$$

there is $B' > 0, D' > 0$ *so that for any* $x', x" \in \mathbf{R}$

$$|M^{(1)}(x') - M^{(1)}(x")| \leq D'|x' - x"|, |M^{(1)}(\theta)| \leq B' \tag{16}$$

then estimates $\{x_n\}$ *which formed by*

$$x_{n+1} = P_\Theta\{x_n - \frac{a}{n}\frac{h}{n^{\frac{1}{2p}}}K_1(\zeta_n)Y_n\} \tag{17}$$

satisfy convergence $x_n \to \theta$ *as sure. In some cases random value* $(x_n - \theta)n^{\frac{p-1}{2p}}$ *has asymptotically normality distribution with mean value* 0 *and variance*

$$\frac{a^2(\sigma^2(\theta) + M(\theta))K_1^{(1)}(0)}{h^2(2aM^{(2)}(\theta) - \frac{p-1}{p})}. \tag{18}$$

5 Adaptive storage of Robbins-Monro algorithm

Theorem 2 *Let all conditions of part 1 theorem 1 are hold.*

If θ is the unique root of regression equation $M(x) = 0$ on Θ and there is two positive constants $s^+ > s^- > 0$ so such

$$s^- \leq M^{(1)}(\theta) \leq s^+, \tag{19}$$

$P_S\{...\}$ is projection operator on set $S = [s^-, s^+]$,

then estimates $\{x_n\}$ which formed by

$$x_{n+1} = P_\Theta\{x_n - \frac{1}{n}\frac{1}{s_n}K_0(\zeta_n)Y_n\} \tag{20}$$

where $\{s_n\}$ is sequence of random values

$$s_{n+1} = P_S\{\frac{1}{n}\sum_{i=1}^{n}\frac{i^{\frac{1}{2p}}}{h}K_1(\zeta_i)Y_i\}, \tag{21}$$

satisfy convergence $x_n \to \theta$ as sure and $s_n \to M^{(1)}(\theta)$ as sure, random value $(x_n - \theta)n^{\frac{1}{2}}$ has asymptotically normality distribution with mean value 0 and variance

$$\frac{\sigma^2(\theta)K_0(0)}{(M^{(1)}(\theta))^2} \tag{22}$$

and random value $(s_n - M^{(1)}(\theta))n^{\frac{p-1}{2p}}$ has asymptotically normality distribution with mean value 0 and variance

$$\frac{p\sigma^2(\theta)K_1^{(1)}(0)}{p+1}. \tag{23}$$

Note, expressions 22 and 23 are minimum of possible in wide range of similar algorithms. In accordance expressions 14(22) and 18(23) we have one way to choice kernels $K_r(u), r = 0, 1$. For example we can calculate variance for Legendre's and Chebushev's polinoms or we can study dependence between γ and variance of asymptotically distribution.

6 Two practical problems

The method of "dividing" material presence detection in the target, which is based on trail signals using.

The initiation of in the time after target radiation treatment by an electron bunch is one of the characteristics, which points at the presence in the target some of "dividing" material. One of the basic theories of suspicious objects inspection is based at this fact. At the same time the conclusion about the type of the "dividing" material can be done by the retarded radiation intensity changes. If the enemy has a counteraction opportunity, then it is easy to determine the beginning of the inspection (to note the electron bunch) by this method. Also, having some resources in the time of retarded radiation appearance, add to the lag neuron flow corresponding "jam" flow to delete an inspection opportunity. The main aim of the new method with the trail signals using is to give a series of radiate electron bunch, the intensity sequence of which is determined by some accidental process with well known static properties. The new algorithm allows fixing the presence and the nature of the lag neuron flow, in spite of high level noise generated.

Weighing Substance Precise Mass Definition Algorithm, Which Use Trail Obstructions.

The basic accuracy restriction during solving the problem of super precise mass definition of reference material, which connects with the systematical error, defined by the "dry friction" in mechanical part of the system, is well known. If scales have the electromagnetic coupling opportunity at the balancing process, then it is possible to start to shake a little bit the system by random sequence the electromagnetic impulses to get precise result of weighting. Watching the system dynamic, which is defined by the linear difference model of the second-order additive noise with the help of the unknown parameters identification algorithm, coefficient model definition, connected to the inertia moment, is possible. After this is possible to precise mass definition of weighting material.

References

[1] Robbins H., Monro S. A stochastic approximation method. *Ann. Math. Statist. 22*,pp. 400-407, 1951.

[2] Kiefer J., Wolfowits J. Stochastic estimation of the maximum of a regression function. *Ann. Math. Statist. 23*,pp. 452-466, 1952.

[3] Fabian V., Stochastic approximation of minima with improved asymptotic speed. *Ann. Math. Statist. 38*,pp. 191-200, 1967.

[4] Polyak B.T., Tsybakov A.B. Stochastic estimation of the maximum of a regression function. *Problems Inform. Transmission 26*, pp. 126-133, 1990.

[5] Granichin O.N. Stochastic Approximation procedure with perturbations in an input and depended observation disturbances *Vestnik Leningrad Univ., vol. 1(4), 22*, pp.27-31,1989.

[6] Granichin O.N. Stochastic Approximation with sample perturbations in the input *Automat. Remote Control 53*, pp. 97-104, 1992.

[7] Granichin O.N. Unknown function minimum point estimation under dependent noises. *Problems Inform. Transmission 28*, pp. 16-20, 1992.

[8] Granichin O.N. Stochastic Approximation under Dependent Noises, Detecting Signals and Adaptive Control *Approximation, Probability, and Related Fields*, pp. 247-271, Plenum, 1994.

[9] Polyak B.T., Tsybakov A.B. On Stochastic Approximation with Arbitrary Noise (the KW-case). *Advances in Soviet Mathematics*,v.12, pp. 107-113, 1992.

[10] Goldenshluger A.V., Polyak B.T. Estimation of regression parameters with arbitrary noise. *Mathematical Methods of Statistics*,v.2, No.1, pp. 18-29, 1993.

[11] Spall J.C., Multivariate Stochastic Approximation Using a Simultaneous Perturbation Gradient Approximation,*IEEE Transactions on Automatic Control*, v. 37, pp. 332-341, 1992.