Simulation 2001
Proceedings of the 4th St.Petersburg Workshop on Simulation

St.Petersburg, Russia,
June 18 — 22, 2001

Edited by S.M. Ermakov, Yu.N. Kashtanov and V.B. Melas

NII CHEMISTRY ST.PETERSBURG
UNIVERSITY PUBLISHERS
2001

Regression Parameters Estimation Under Arbitrary Noise

O. N.Granichin[1] St.Petersburg State University[2]

Abstract

The paper is devoted to the linear regression parameters estimation problem in a case where observation noise does not have any good statistical properties. The main asssumption, which will be made, is that the model regressors (inputs) are random. New type of algorithms is studied. They are similar to standart LMS or stochastic approximation algorithms. Optimal rates of convergence are established under some weak assumptions. A typical algorithm behavior under the "bad" noise is illustrated by several numerical examples.

1 Introduction

In linear regression problems the usual assumption as to the observation noises is that noises are considered as realization of some sequence of independent random variables with zero mean. However in applications this assumption is frequently neglected leading to side effects of standard estimations procedures. Therefore it is important to investigate the capability of the regression parameters estimation at a minimum assumptions to statistical properties of observation noises. It may seem surprising but regression parameters can be effectively estimated though not centered, correlated and even nonrandom noises (see [5],[6],[7],[8]). It can be reached under certain conditions when regressors are random. The idea to use random input signals for the removal of displacement effect was put forward by Fisher [1] in the form of the randomized principle of experiment design. Apart from experiment design problem, in which regressors can be randomized by an experimenter, random inputs also arise in many problems of identification, filtration, recognition, etc.

Recurrent algorithms of regression parameters estimation were considered at the random input signals case in works [3],[4] etc. In paper [2], the rate of such algorithms convergence was studied. There was considered the optimum algorithms with best from the possible rates of convergence. In all these papers standard assumptions for noises were made. Namely it were considered that noises are random variables with zero mean, independent or weak

[1]e-mail: oleg_granitchin@mail.ru

[2]Bibliotechnaya sq. 2, St.Petersburg, Russia, 198904.

dependent. The main purpose of this paper is to avoid these assumptions. It will be shown that some estimation algorithms remain to generate consistent estimates under not centered, correlated or even non–random observation noises given input signals are random. Besides, optimum algorithms have the same convergence rate as in the "standard" case.

The problem of linear regression parameters estimation was considered in the [5] and [6] under "unknown, but bounded" deterministic noises. There a more general then standard problem statement was studied, which supposed that the unknown parameters vector can vary in time and it's mean value was estimated in the offered algorithm. Content of this paper in many respects is based on the [7].

2 Problem statement and assumptions

Consider a linear regression model

$$(1) \qquad y_n = \phi_n^T \theta_n^* + v_n, \quad \theta_n^* = \theta_* + \zeta_n, \quad n = 1, 2, \ldots,$$

with output $y_n \in \mathbf{R}^1$, input $\phi_n \in \mathbf{R}^r$ and noise $v_n \in \mathbf{R}^1$, $\zeta_n \in \mathbf{R}^r$. The unknown parameter vector θ_* is to be estimated based on observations y_n, ϕ_n, $n = 1, 2, \ldots$.

Let $\mathcal{F}_n$ be the σ–algebra generated by $\{\phi_1, \ldots, \phi_n, \zeta_1, \ldots, \zeta_n, v_1, \ldots, v_n\}$, $\hat{\mathcal{F}}_{n-1}$ be the σ–algebra generated by $\{\phi_1, \ldots, \phi_{n-1}, \zeta_1, \ldots, \zeta_{n-1}, v_1, \ldots, v_n\}$, и $\tilde{\mathcal{F}}_{n-1}$ be the σ–algebra generated by $\{\phi_1, \ldots, \phi_{n-1}, \zeta_1, \ldots, \zeta_n, v_1, \ldots, v_n\}$, $\mathcal{F}_{n-1} \subset \hat{\mathcal{F}}_{n-1} \subset \tilde{\mathcal{F}}_{n-1} \subset \mathcal{F}_n$.

We make the following assumptions.

(A) Inputs $\{\phi_n\}_{n\geq 1}$ form a sequence of independent random vectors with bounded mean values $\|\mathrm{E}\{\phi_n\}\| \leq M_\phi < \infty$; for all n ϕ_n is independent of $\tilde{\mathcal{F}}_{n-1}$. Random vectors $\Delta_n = \phi_n - \mathrm{E}\{\phi_n\}$ have symmetric distribution functions $\mathbf{P}_n(\cdot)$, i.e. $\mathbf{P}_n(\Omega) = \mathbf{P}_n(-\Omega)$ for any Borel set $\Omega \subset \mathbf{R}^r$; $\mathrm{E}\{\Delta_n \Delta_n^T\} = \mathrm{B_n} > 0$; $\|\mathrm{B}_n\| \leq \sigma_\Delta^2 < \infty$; $\mathrm{E}\{\|\Delta_n\|^4\} \leq M_4 < \infty$.

(B) For all n ζ_n are independent from $\hat{\mathcal{F}}_{n-1}$ and $\mathrm{E}\{\zeta_n\} = 0$. Noises $\{v_n\}_{n\geq 1}$ and $\{\zeta_n\}_{n\geq 1}$ satisfy one of conditions:

(i) $\quad \mathrm{E}\{ v_n^2 | \mathcal{F}_{n-1}\} \leq \sigma_v^2 < \infty$, a.s., $\mathrm{E}\{\|\zeta_n\|^2 | \mathcal{F}_{n-1}\} \leq \sigma_\theta^2 < \infty$ a.s.;

(ii) $\quad \mathrm{E}\{v_n^2\} \leq \sigma_v^2 < \infty, \qquad \mathrm{E}\{\|\zeta_n\|^2\} \leq \sigma_\theta^2 < \infty$;

(iii) $\quad |v_n| \leq C_v < \infty$, a.s., $\qquad \|\zeta_n\| \leq C_\zeta < \infty$ a.s.,

where σ_v, σ_θ, C_v and C_ζ are some constants.

Let's remark that standard assumptions (see [2]) are various in the linear regression parameter estimations problem with random input signals. In particular, that is expressed in the absence of zero-mean condition $\mathrm{E}\{v_n\} = 0$ and assumption, which states that noises $\{v_n\}_{n\geq 1}$ are independent and identically distributed random variables.

3 Estimation algorithms

Firstly we examine the randomized stochastic approximation estimator for (1)

$$\theta_n = \theta_{n-1} - \alpha_n \Gamma \Delta_n (\phi_n^T \theta_{n-1} - y_n), \quad n = 1, 2, \ldots, \tag{2}$$

where $\alpha_n \geq 0$ is a non–random step size and Γ is a positively defined symmetric matrix. We suppose that the initial value θ_0 is an arbitrary non-random vector in$\mathbf{R}^r$.

Theorem 1 *Let Assumption* **(A)** *be fulfilled and* $\sum_{n=1}^{\infty} \alpha_n = \infty,\ \alpha_n \to 0$ *as* $n \to \infty$.

If **(Bi)** *holds and* $\sum_{n=1}^{\infty} \alpha_n^2 < \infty$, **then** *for (2)* $\theta_n \to \theta_*$ *npu* $n \to \infty$ *a.s.*

If **(Bii)** *holds* **then** *for (2)*$\mathrm{E}\{(\theta_n - \theta_*)(\theta_n - \theta_*)^T\} \to 0$ *as* $n \to \infty$. The following theorem establishes the convergence rate for the algorithm (2).

Theorem 2 *Let Assumptions* **(A)** *and* **(Bii)** *be fulfilled,*$\alpha_n = n^{-1}$, $\exists \mathrm{B} > 0 :\ \|\mathrm{B}_n - \mathrm{B}\| = \mathcal{O}(n^{-1})$ *and* $-\Gamma\,\mathrm{B} + \frac{1}{2}\mathrm{I}$ *be a Hurwitz matrix, i.e. all its eigenvalues lie in the left half-plane (* I *is an identity matrix).* **Then**

$$\mathrm{E}\,(\theta_n - \theta_*)(\theta_n - \theta_*)^T \leq n^{-1}\,\mathrm{S} + o(n^{-1})$$

where S *is a solution of matrix equation*

$$\Gamma\,\mathrm{B}\,\mathrm{S} + \mathrm{S}\,\mathrm{B}\,\Gamma - \mathrm{S} = (\sigma_v^2(1 + M_\phi^2 \rho) + M_\phi^2 \sigma_\theta^2)\Gamma \mathrm{B} \Gamma + \sigma_\theta^2 M_4 \Gamma^2,$$

where $\rho > 0$ *is some small positive constant.*

Last equation can be explicitly solved in the case when $\Gamma = \mathrm{B}^{-1}$,$\sigma_\theta = 0$ and $M_\phi = 0$. For the algorithm (2), which has the standart form of

$$\theta_n = \theta_{n-1} - (n\mathrm{B})^{-1} \phi_n (\phi_n^T \theta_{n-1} - y_n), \quad n = 1, 2, \ldots,$$

we have

$$\mathrm{E}\,(\theta_n - \theta_*)(\theta_n - \theta_*)^T \leq n^{-1} \sigma_v^2 \mathrm{B}^{-1} + o(n^{-1}).$$

The same convergence rate holds when v_n are independent zero-mean random variables [2]. Moreover it was shown in the [2] that this choice of α_n and Γ

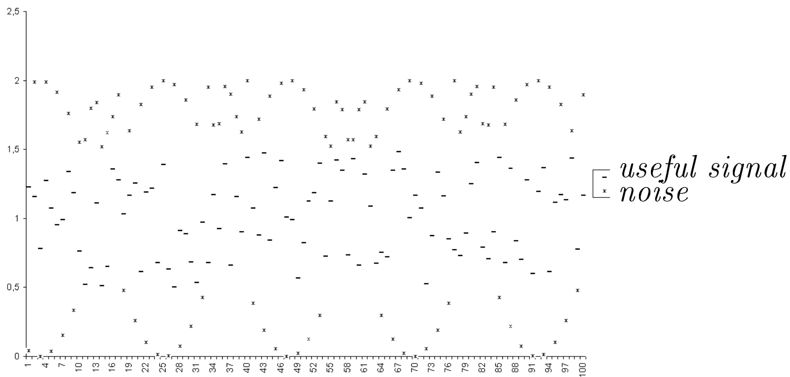

Fig.1 Useful signal and noise

is an optimal for the algorithms (2). **Remark.** In Theorem 2 statement the inequalities in evaluations of the convergence rate can be replaced by equalities in the case when equalities are in the assumption (Bii).

Secondly, for the regression model (1) we consider the regularized least squares estimator with centered regressors:

$$\theta_n \;=\; \Gamma_n \sum_{k=1}^{n} \Delta_k y_k, \qquad \Gamma_n \;=\; (\sum_{k=1}^{n} \Delta_k \Delta_k^T \;+\; \rho \mathrm{I})^{-1},$$

or in the recursive form

$$(3a) \qquad \theta_n = \theta_{n-1} - \Gamma_n \Delta_n (\Delta_n^T \theta_{n-1} - y_n),$$

$$(3b) \qquad \Gamma_n = \Gamma_{n-1} - \frac{\Gamma_{n-1}\Delta_n\Delta_n^T\Gamma_{n-1}}{1 + \Delta_n^T\Gamma_{n-1}\Delta_n},\; \Gamma_0 = \rho^{-1}\mathrm{I},$$

where $\Delta_n = \phi_n - \mathrm{E}\{\phi_n\}$, $\rho > 0$ is a small positive number (regularization parameter), see [3]. Also we assume that the initial value θ_0 is an arbitrary nonrandom vector in $\mathbf{R}^r$.

Theorem 3 *Let* $\Delta_n,\ n = 1, 2, \ldots$ *be identically distributed random vectors and Assumption* **(A)** *be fulfilled.*

If **(Bi)** *holds,* **then** *for algorithms (3)* $\theta_n \to \theta_*$ *as* $n \to \infty$ *a.s.*

If **(Biii)** *holds and* $\|\Delta_n\| \leq C_\Delta < \infty,\ n = 1, 2, \ldots$ *a.s.* **then** *for the algorithm (3) as* $n \to \infty$ $\mathrm{E}\{(\theta_n - \theta_*)(\theta_n - \theta_*)^T\} \to 0$.

The proofs of Theorems 1,2 and 3 in many respects coincide with corresponding proofs in the [7].

4 Experimental results

Let's consider a problem of detecting some "useful" signal $\{\phi_n\}$, which can be present or not in an observation channel. Measurements are made with ad-

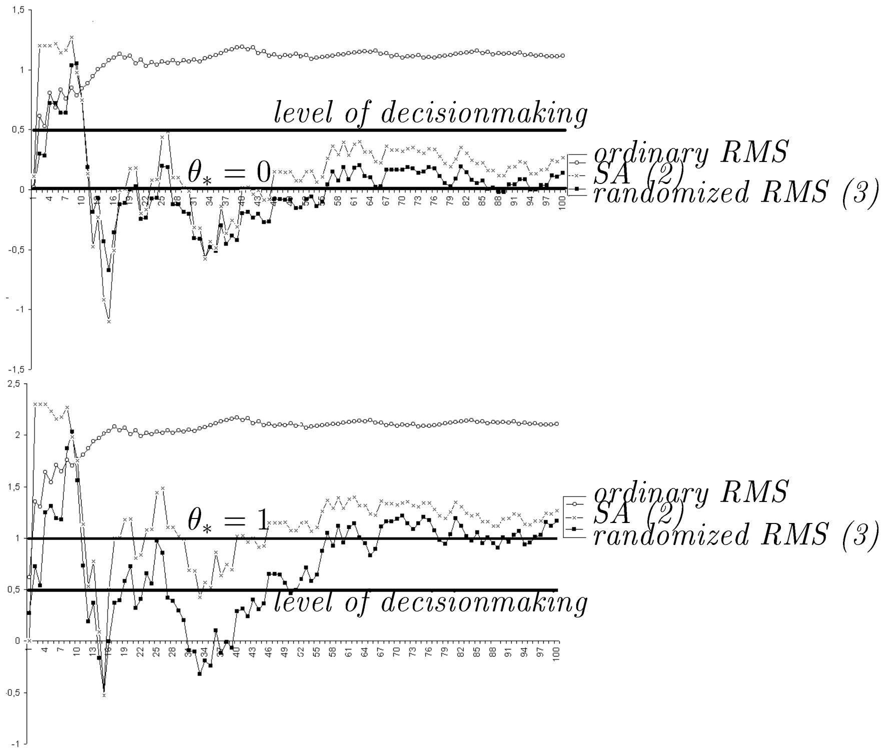

Fig.2 Detecting useful signal

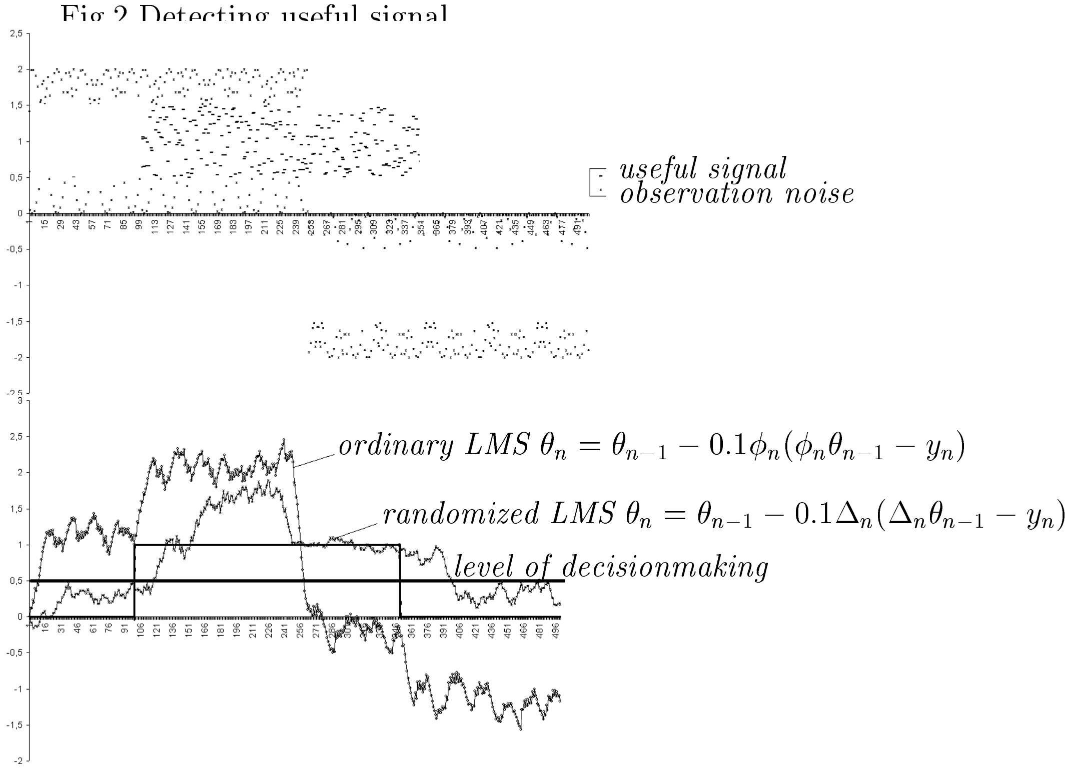

Fig.3 Tracking useful signal

ditive "bad" noises. Assume that case $\{\theta_* = 1\}$ corresponds to the situation when a signal is present in the receiver, and $\{\theta_* = 0\}$ — to absence.

At the computer simulation the useful signal $\{\phi_n\}$ was selected as uniformly distributed on an interval $[0.5, 1.5]$ and was also observed on a hum of noises, which are determined by an "unknown but bounded" deterministic function, $|v_n| \leq C_v = 2$ (see Figure 1). Figure 2 shows the pathes of estimation series for three algorithms in two cases. The level of observation noises is so high that ordinary RMS algorithm estimates exceed the decisionmaking level almost always without the dependence from presence or absence of a signal, while the algorithms (1) and (2) give the correct answers after 50 iterations.

In the following example useful signals $\{\phi_n\}$ and observation noises $\{v_n\}$ satisfy the same conditions as earlier but useful signals "are actuated" in an observation channel temporarily, though its value are accessible to the experimenter during all the observation period. The problem is to design a rule, on which at each moment one could answer the question whether the useful signal acts in the observation channel or there is just noise being registered. Compare with the ordinary LMS the randomized algorithm watches the changes of useful signal parameters more precisely, it gives only 13% of incorrect answers, see Fig.3.

5 Conclusion

In the random input signals case the considered above algorithms require the fulfilment of very weak conditions of observation noises for the proofs of convergence. In particular, the observation noises can be determined by "unknown but bounded" deterministic function. Thus, these algorithms can be useful in many applications. Numerical simulation has demonstrated algorithms effectiveness under diverse noises. For example, in scalar case the experiments were made with a non–random constant, not centered random noises and various non–random sequences of noises. In these experiments pathes of SA and LMS estimates were investigated. The typical pathes behaviour has appeared quite similar to experiments outcomes in the case of random noises with zero mean.

References

[1] Fisher R.A., *The Design of Experiments.* Oliver and Boyd, Edinburgh, 1935.

[2] Polyak B.T., Tsypkin Ya.Z., *Adaptive estimation algorithms (convergence, optimality, stability)* //Automat. Remote Control, v. 40, 1979, pp. 378–389.

[3] Ljung L., Söderström T., *Theory and Practice of Recursive Identification.* MIT Press, Cambridge, MA - London, 1983.

[4] Ermakov S.M., Gigljavskii A.A., *Mathematical theory of disign of an experiment.* M., Nauka, 1987

[5] Granichin O.N., *A Stochastic Appriximation Algorithm with Perturbations in the Input for Identification of Static Nonctationary Plant*// Vestnik Leningr. Univ. Math., v. 21, 1988, No.3, pp. 92–93.

[6] Granichin O.N., *Unknown Function Minimum Point Estimation under Dependent Noise*// Problems Inform.Transmission,v.28,1992,No.2,pp.16–20.

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

[8] Ljung L., Guo L., *The Role of Model Validation for Assessing the Size of the Unmodeled Denamics* //IEEE Trans. on AC, v. 42, 1997, No. 9, pp. 1230–1239.