

Розглядається задача ідентифікації нестационарних параметрів лінійного об'єкта, які можна описати Марківською моделлю першого порядку, за наявності негаусівських завад. Алгоритм ідентифікації є градієнтною процедурою мінімізації комбінованого функціоналу. Комбінований функціонал, в свою чергу, складається з квадратичного і модульного функціоналів, ваги яких встановлюються за допомогою параметра змішування. Така комбінація функціоналів дозволяє отримати оцінки, що володіють робастними властивостями. Алгоритм ідентифікації не вимагає знання інформації про ступінь нестационарності досліджуваного об'єкта. Він є найбільш простим, тому що використовує при побудові моделі інформацію тільки про один такт (крок) вимірювань. Застосування Марківської моделі є досить ефективним, оскільки дає можливість отримати аналітичні оцінки властивостей алгоритму. Визначено умови збіжності градієнтного алгоритму при оцінці нестационарних параметрів в середньому і середньоквадратичному за наявності негаусівських завад вимірів.

Отримані оцінки є досить загальними і залежать як від ступеня нестационарності об'єкту, так і від статистичних характеристик корисних сигналів і завад. Крім того, визначено вирази для асимптотичних значень помилки оцінювання параметрів і асимптотичної точності ідентифікації. У зв'язку з тим, що дані вирази містять ряд невідомих параметрів (значення дисперсій сигналів і завад, дисперсії, що характеризує нестационарність), для їх практичного застосування слід використовувати оцінки цих параметрів. З цією метою слід скористатися будь-якою рекурентною процедурою оцінки невідомих параметрів і використовувати одержувані оцінки для уточнення тих параметрів, які входять в алгоритми. Крім того, асимптотичні значення помилки оцінювання і точності ідентифікації залежать від вибору параметра змішування

Ключові слова: Марківська модель, градієнтний алгоритм, параметр зважування, рекурентна процедура, асимптотична оцінка, точність ідентифікації

Received date 04.09.2019

Accepted date 17.10.2019

Published date 31.10.2019

1. Introduction

The identification problem is not only of interest in itself, but is also an integral part of the general optimization problem. Many problems of management, forecasting, pattern recognition, etc., are associated with the construction of a model of the following form:

$$y(k) = \theta^{*T} x(k) + \xi(k), \quad (1)$$

where $y(k)$ is the observed output signal;

$$x(k) = (x_1(k), x_2(k), \dots, x_N(k))^T$$

is the input vector $N \times 1$;

$$\theta^* = (\theta_1^*, \theta_2^*, \dots, \theta_M^*)^T$$

is the vector of unknowns $M \times 1$; $\xi(k)$ is interference, and are reduced to minimizing some pre-selected quality functional (identification criterion).

UDC 004.852

DOI: 10.15587/1729-4061.2019.181256

ROBUST IDENTIFICATION OF NON-STATIONARY OBJECTS WITH NONGAUSSIAN INTERFERENCE

O. Rudenko

Doctor of Technical Sciences,
Professor, Head of Department*

E-mail: oleh.rudenko@nure.ua

O. Bezsonov

Doctor of Technical Sciences, Associate Professor*

E-mail: oleksandr.bezsonov@hneu.net

O. Lebediev

PhD, Associate Professor
Department of Electronic Computers**

E-mail: oleh.lebediev@nure.ua

N. Serdiuk

PhD, Associate Professor*
E-mail: nataliya.serdyuk@nure.ua

*Department of

Computer Intelligent Technologies and Systems**

**Kharkiv National University of Radio Electronics
Nauky ave., 14, Kharkiv, Ukraine, 61166

Copyright © 2019, O. Rudenko, O. Bezsonov, O. Lebediev, N. Serdiuk

This is an open access article under the CC BY license

(<http://creativecommons.org/licenses/by/4.0>)

However, the identification problem is significantly complicated if the parameters θ change (drift) over time, i. e. $\theta^*(k) = \text{var}$.

The quadratic functional most widely used in practice leads to various identification algorithms that allow obtaining estimates of the sought vector θ^* with normal interference distributions, i. e. $\xi(k) \sim N(0, \sigma_\xi^2)$.

Most of available identification methods are based on the use of strict and difficult to test conditions associated with the hypothesis of normality of the interference distribution law and justified by references to the central limit theorem. As is known, the normal law of distribution density describes interference present in measurements carried out under absolute stability of measurement conditions, the Laplace's law having longer "tails" – interference occurring under maximum instability of conditions. Accordingly, identification algorithms in the case of Gaussian interference are based on the least squares (LS) method, and in the case of interference distributed according to Laplace's law, they are based on the least absolute deviations (LAD) method. Both of these methods are optimal in their conditions and

the solutions obtained with their help may vary greatly. Furthermore, since in practice these extreme cases are very rarely implemented, neither the Gauss’s law nor the Laplace’s law are usually fulfilled.

In this regard, it seems very relevant to develop an approach to robust estimation of non-stationary parameters using some combined functional, which allows combining LS and LAD.

2. Literature review and problem statement

To estimate non-stationary parameters, modified LS algorithms (use of a sliding window or exponential smoothing, etc.), Kaczmarz algorithm proposed in [1] and its modifications, dynamic adaptation algorithms, etc. are commonly used. In particular, the Kaczmarz algorithm has been studied in sufficient detail in [2–6]. Modifications of this algorithm are associated with an increase in computational stability and improvement of dynamic properties. So, in [2], the modified (regularized) Kaczmarz algorithm was studied, in [3] its multi-step modification was considered, in [4] weighting of estimates to increase the speed of this algorithm was proposed. In [5], analytical expressions for asymptotic and non-asymptotic estimates were obtained and expressions for optimal values of the relaxation parameter of the Kaczmarz algorithm providing its maximum convergence rate were determined. In [6], a randomized version of the Kaczmarz method for consistent, overdetermined linear systems was proposed and it was proved that it converges at the expected exponential rate. In [7, 8], the efficiency of the Kaczmarz algorithm in estimating the non-stationary parameters described by the first-order Markov model was studied.

It should be noted that both the algorithm and all the modifications mentioned are based on the use of quadratic identification (estimation) criterion, i. e., they are varieties of the least squares (LS) method. Being the optimal estimation method with Gaussian interference, LS is not stable with non-Gaussian interference. This is because in this case the objective function can grow to infinity and outliers can become dominant dimensions that actually test the real model. Alternatively, to ensure robustness, the objective function is modified to limit the influence of the largest measurements. The main consequence of this is generally a lower convergence rate of optimization algorithms. This is due to the fact that distinguishing between outliers and useful measurements for the first time is very difficult. In this regard, some outliers can be filtered out, leading to a decrease in convergence rate. In the most difficult case, small but biased measurements move the minimum of the objective function.

If information about the affiliation of interference ξ to a certain class of distributions is known, then by minimizing the optimal criterion, which is the inverse logarithm of the interference distribution function, the maximum likelihood estimate (M -estimate) can be obtained. If there is no such information, then to evaluate the desired parameter vector θ , one should apply some non-quadratic criterion that ensures the robustness of the resulting estimate. One of the criteria is modular criterion leading to a sign algorithm. Application of this criterion in the problem of object identification with impulse interference was considered in [9–12]. So, [9] studied the efficiency of the affine projection algorithm, [10] used the variable-gain affine projection algorithm. It should be

noted that sign algorithms, providing robustness of the obtained estimates, have a low convergence rate. Therefore, in order to accelerate the estimation process, a normalized sign identification algorithm was proposed and studied in [11]. [12] studies an easy-to-implement algorithm, which uses RMS error and estimated interference power to correct the step size. The theoretical results of its stationary behavior, obtained for the case of Gaussian input signals, are in good agreement with the experimental results. Similarly, [13] considers the Kaczmarz algorithm with variable gain depending on the squared cross-correlation between the squared output error and adaptive model output and shows its effectiveness in solving some noise reduction problems.

There are a fairly large number of functionals that provide robust M -estimates. The most common are combined functionals proposed by Huber [14, 15] and Hempel [16, 17]. They consist of a quadratic functional that ensures the optimality of estimates for the Gaussian distribution and modular one that allows obtaining a more robust estimate for distributions with heavy “tails” (outliers).

These functionals (ρ) and their influence functions (ψ) have the following form:

$$\begin{aligned} \rho_1(e) &= \begin{cases} \frac{e^2}{2}, & |e| \leq c; \\ c|e| - \frac{c^2}{2}, & c < |e|; \end{cases} \\ \psi_1(e) &= \begin{cases} e, & |e| \leq c; \\ c \operatorname{sign}(e), & c < |e|; \end{cases} \end{aligned} \quad (2)$$

$$\begin{aligned} \rho_2(e) &= \begin{cases} \frac{e^2}{2}, & 0 \leq |e| < b; \\ b|e| - \frac{b^2}{2}, & b \leq |e| < c; \\ \frac{b}{2}(c+d) - \frac{b}{2} \left(b - \frac{(|e|-d)^2}{c-d} \right), & c \leq |e| < d; \\ \frac{b}{2}(c+d) - \frac{b^2}{2}, & d \leq |e|; \end{cases} \\ \psi_2(e) &= \begin{cases} e, & 0 \leq |e| < b; \\ b \operatorname{sign}(e), & b \leq |e| < c; \\ \frac{b(|e|-d)}{c-d} \operatorname{sign}(e), & c \leq |e| < d; \\ 0, & d \leq |e|, \end{cases} \end{aligned} \quad (3)$$

where e is the estimation error.

It should be noted that M -estimates are usually described by setting an influence function rather than a minimized functional.

The Huber function ψ is monotonic, and the Hempel function ψ is nonmonotonic. As noted in [18], with heavy-tailed distribution, the use of nonmonotonic ψ functions improves the estimation results.

The effectiveness of these functionals depends on how well the constants a , b , c and d included in them are chosen, which determine the degree of interference immunity. In the above studies, it is recommended to choose the values a from the interval $[\sigma, 2\sigma]$, where σ is the standard deviation

of observation x , and set the values b , c and d to 1.5, 3.5, and 8, respectively.

A modification of Huber estimates are the Mallows estimates [19]. Along with weighing of residuals ϵ , weighing of factors is performed, which makes it possible to reduce the influence of points that stand out sharply in the space of independent variables. To find the Mallows estimates, it is necessary to solve $\rho+1$ equations

$$\psi(\epsilon) = \begin{cases} \sum_{i=1}^n \psi(\epsilon[i]); \\ \sum_{i=1}^n x_m[i] \psi^x(x[i]) \psi(\epsilon[i]), m = 0, 1, \dots, \rho, \end{cases}$$

where ψ is determined by Huber (2), and

$$\psi^x(x[i]) = \prod_{j=1}^{\rho} \psi^{(j)}(x[i]).$$

Due to the fact that the modular criterion provides an estimate that is less sensitive to interference distribution tails than the LS estimate, such a kind of modular criterion as A . Forsythe functional is of interest [20]

$$F[e(k)] = |e(k)|^\lambda, \tag{4}$$

where $0 < \lambda < 2$.

Forsythe estimates (4) are close to Huber ones, yet lack such a convincing theoretical justification. It is empirically shown that $\lambda=1.5$ is acceptable. For $\lambda=2$, the Forsythe estimates coincide with the LS estimates, and for $\lambda=1$ we obtain the least absolute deviations (modules) (LDA) method, which minimizes the functional

$$Q = \sum_{i=1}^k |\epsilon(i)|. \tag{5}$$

To obtain Merrill-Schweppe estimates, the functional [20] is used

$$\rho(\epsilon) = \begin{cases} \gamma_1 \epsilon^2, & |\epsilon| \leq k, \\ \gamma_2 \sqrt{\epsilon}, & |\epsilon| > k, \end{cases}$$

where $\gamma_1 > 0, \gamma_2 > 0, 0 \leq k$.

As follows from the above formulas and as noted above, the effectiveness of the obtained robust estimates substantially depends on the numerous parameters used in the criteria and selected based on the researcher's experience.

The practical application of the considered functionals for solving the identification problem was considered in many works. In particular, in [21–24] the robust approach was applied to the identification of nonlinear systems. For this purpose, radial basis function networks [21, 22], evolving networks [23], and evolving radial basis function networks [24] were used. Learning of these networks was carried out on the basis of minimizing the robust functionals considered above.

Another approach to obtaining robust estimates devoid of this drawback is the use of a combined criterion.

A combined estimation criterion to accelerate the identification process using a combination of quadratic criterion and fourth degree criterion, proposed and studied in [25], was developed in [26–30]. [26] investigated the stability of the

algorithm under Gaussian input signals. In [27], normalization of the least mean fourth algorithm was proposed, which protects the algorithm from divergence when the input signal power increases and an approximate stability boundary of this algorithm was obtained. In [28], the problem of stability of the adaptive least mean algorithm was considered and normalization of the update term of the weight vector using the fourth order in the regressor and the second order in the estimation error was proposed. This allows increasing the stability of the algorithm with increasing the dispersion of the input signal and the type of distribution of the input signal. [29] also studied the problem of increasing the stability of the least mean algorithm in the context of adaptive interference reduction and showed under what conditions the algorithm minimizing the fourth degree criterion is superior to the Kaczmarz algorithm. In [30], the mean-square convergence of the least mean fourth algorithm for various cases, including non-Gaussian interference distributions, was analyzed. However, the analysis assumes the presence of a reference zero-mean Gaussian signal, which is not always possible.

In [31], the fourth degree criterion was replaced with the least absolute deviations criterion, which made it possible to ensure the robustness of the obtained estimates under impulse interference conditions. The normalized modification of the identification algorithm considered in [32] was studied in [33, 34], where the presence of impulse interference was also taken into account.

In [35], the use of an adaptive combination of two normalized filters to obtain robust estimates in the identification problem was studied. It should be noted that this criterion proved to be very effective and much easier to implement in the identification procedure.

3. The aim and objectives of the study

The aim of the work is to study the convergence of gradient algorithms of identification of non-stationary parameters described by the first-order Markov model under non-Gaussian interference and to determine parameters of the algorithms ensuring their maximum convergence rate.

To achieve the aim, the following objectives were set:

- to obtain analytical estimates of mean and mean-square convergence of the gradient minimization algorithm of the combined functional;
- to determine the maximum attainable (asymptotic) values of parameter estimation errors and identification errors in the considered conditions.

4. Obtaining analytical estimates of convergence of robust identification procedure

Note that to ensure robustness of the obtained estimates, it is quite effective to use a combined learning functional [31, 32]

$$F[e(k)] = \lambda M\{e^2(k)\} + (1-\lambda)M\{|e(k)|\}, \tag{6}$$

where

$$\begin{aligned} e(k+1) &= y(k+1) - \hat{y}(k+1) = \\ &= y(k+1) - \hat{\theta}^T(k)x(k+1), \end{aligned}$$

$$\hat{y}(k+1)$$

is the output signal of the model;

$$\hat{\theta}(k) = (\hat{\theta}_1(k), \hat{\theta}_2(k), \dots, \hat{\theta}_N(k))^T$$

is the vector of estimated parameters $N \times 1$; γ is the parameter affecting the convergence rate of the algorithm; $\lambda \in [0, 1]$ is the mixing parameter.

When using the combined criterion (6), the gradient minimization procedure has the following form

$$\begin{aligned} \theta(k) &= \theta(k-1) + \\ &+ \gamma(k) [\lambda 2e(k) + (1-\lambda) \text{sign } e(k)] x(k). \end{aligned} \quad (7)$$

This procedure combines the properties of LS with those of LAD, since when $\lambda=1$ (7) implies the LS algorithm, and when $\lambda=0$ – LAD algorithm (5), and allows dealing with impulse interference. By varying the parameter λ , one can change the properties of the algorithm.

To obtain analytical estimates in the non-stationary case, it was assumed, like in [7, 8], that the non-stationary parameters of an object can be represented by a first-order Markov model

$$\theta(k) = \theta(k-1) + S(k), \quad (8)$$

where

$$S(k) = (S_1(k), S_2(k), \dots, S_N(k))^T$$

is the vector of random sequence $N \times 1$; $S_i \sim N(0, \sigma_s^2)$.

We introduce the estimation error

$$\tilde{\theta}(k) = \theta(k) - \hat{\theta}(k). \quad (9)$$

It is assumed that components of the estimation error vector $\tilde{\theta}(k)$ obey the normal distribution law with

$$\tilde{\theta}_i(k) \triangleq M \{ \tilde{\theta}_i(k) \}$$

and dispersion $\sigma_i^2(k)$ [36], that is, all components of the estimation vector $\hat{\theta}_i(k)$ are distributed according to the normal law

$$\hat{\theta}_i(k) \sim N(m_i(k), \sigma_i^2(k)),$$

with the probability density function

$$\begin{aligned} f(\hat{\theta}_i(k)) &= \frac{1}{\sqrt{2\pi\sigma_i^2(k)}} \times \\ &\times \left[e^{-\frac{(\hat{\theta}_i(k)-m_i(k))^2}{2\sigma_i^2(k)}} + e^{-\frac{(\hat{\theta}_i(k)+m_i(k))^2}{2\sigma_i^2(k)}} \right] U(\hat{\theta}_i(k)), \end{aligned}$$

where

$$m_i(k) = \theta_i(k) - \tilde{\theta}_i(k);$$

$$\sigma_i^2(k) \triangleq M \{ \tilde{\theta}_i^2(k) \} - M^2 \{ \tilde{\theta}_i(k) \};$$

$$U(\hat{\theta}_i(k)) = \begin{cases} 0, & \hat{\theta}_i(k) < 0, \\ 1, & \hat{\theta}_i(k) \geq 0. \end{cases}$$

The mean of this distribution is determined by the formula

$$\begin{aligned} M \{ \hat{\theta}(k) \} &= \int_{-\infty}^{\infty} \hat{\theta}(k) f(\hat{\theta}(k)) d\hat{\theta}(k) = \\ &= m_i(k) \text{erf} \left(\frac{m_i(k)}{\sqrt{2\sigma_i^2(k)}} \right) + \sqrt{\frac{2}{\pi}} \sigma_i(k) e^{-\frac{m_i^2(k)}{2\sigma_i^2(k)}}, \end{aligned}$$

where

$$\text{erf}(x) \triangleq \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$$

is the Gaussian error function.

In view of (8), the expression for $e(k)$ can be written as follows:

$$e(k) = \tilde{\theta}^T(k-1)x(k) + S^T(k)x(k) + \xi(k). \quad (10)$$

Since it is assumed that $\xi(k) \sim N(0, \sigma_\xi^2)$, we have

$$M \{ e^2(k) \} = \sigma_\xi^2 + \sigma_x^2 M \{ \|\tilde{\theta}(k)\|^2 \}, \quad (11)$$

where M is the symbol of mathematical expectation; $\|\bullet\|$ is the Euclidean norm.

Consider the convergence of the procedure (7). To this end, we introduce the Lyapunov function $\|\tilde{\theta}(k)\|^2$.

Given the accepted form of non-stationarity (3), the algorithm of identification errors $\tilde{\theta}(i)$ can be written as follows

$$\tilde{\theta}(k) = \tilde{\theta}(k-1) - \gamma [2\lambda e(k) + (1-\lambda) \text{sign } e(k)] x(k). \quad (12)$$

Multiplying both sides of (12) on the left by $\tilde{\theta}^T(k)$, taking into account that $e(k) = \tilde{\theta}^T(k-1)x(k)$, we have

$$\begin{aligned} \|\tilde{\theta}(k)\|^2 &= \|\tilde{\theta}(k-1)\|^2 - 4\gamma\lambda e^2(k) - \\ &- 2\gamma(1-\lambda)e(k) \text{sign } e(k) + 4\gamma^2\lambda^2 e^2(k) \|x(k)\|^2 + \\ &+ 4\gamma^2\lambda(1-\lambda)e(k) \text{sign } e(k) \|x(k)\|^2 + \\ &+ \gamma^2(1-\lambda)^2 \|x(k)\|^2. \end{aligned} \quad (13)$$

From (13) it can be seen that since $\gamma > 0$, the increment of the Lyapunov function

$$\Delta V(k) = \|\tilde{\theta}(k)\|^2 - \|\tilde{\theta}(k-1)\|^2$$

will be negative if

$$\begin{aligned} 2(2\lambda e^2(k) + (1-\lambda)|e(k)|) &> \\ > \gamma(2\lambda e(k) + (1-\lambda)|e(k)|)^2 \|x(k)\|^2. \end{aligned} \quad (14)$$

Thus, the convergence condition of the procedure (7) will hold if the parameter γ satisfies the inequality

$$0 < \gamma < \frac{2e(k)}{(\lambda e(k) + (1-\lambda) \text{sign } e(k)) \|x(k)\|^2}. \quad (15)$$

The optimal value of the parameter γ is determined from the equation obtained by differentiating (13) by γ and equating the derivative to zero. Thus

$$\gamma^{omn} = \frac{e_n}{(\lambda e_n + (1-\lambda)\text{sign} e_n)\|x_n\|^2}. \tag{16}$$

We examine the statistical properties of the learning procedure (7) with measurement interference, i. e.

$$y^*(k) = \theta^T x(k) + \xi(k), \quad \xi(k) \sim N(0, \sigma^2).$$

Suppose that interference is not correlated with useful signals. Writing (7) regarding learning errors, we have (12).

Consider the expectation $M\{\tilde{\theta}(k)\}$. Averaging both sides of (12), we obtain

$$\begin{aligned} M\{\tilde{\theta}(k)\} &= \\ &= M\{\tilde{\theta}(k-1) - 2\gamma\lambda(\tilde{\theta}^T(k-1)x(k) + \xi(k))x(k) - \\ &\quad - \gamma(1-\lambda)\text{sign}(\tilde{\theta}^T(k-1)x(k))x(k)\}. \end{aligned} \tag{17}$$

It's easy to see that

$$\begin{aligned} M\{\tilde{\theta}^T(k-1)x(k)x(k)\} &= M\{x(k)x^T(k)\tilde{\theta}(k-1)\} = \\ &= M\{x(k)x^T(k)\}M\{\tilde{\theta}(k-1)\} = R_{xx}M\{\tilde{\theta}(k-1)\}, \end{aligned} \tag{18}$$

where R_{xx} is the correlation matrix of the input signal.

We consider in detail the expression

$$M\{x(k)\text{sign}(\tilde{\theta}^T(k-1)x(k))x(k)\}.$$

In this case, if the signal $x(k) \sim N(0, \sigma_x^2)$, we have

$$\begin{aligned} M\{x(k)\text{sign}(\tilde{\theta}^T(k-1)x(k))\} &= \\ &= M\{M\{x(k)\text{sign}(\tilde{\theta}^T(k-1)x(k))|\tilde{\theta}(k-1)\}\} = \\ &= M\left\{\sqrt{\frac{2}{\pi}}\frac{1}{\sigma_{e\tilde{\theta}(k-1)}}M\{x(k)(\theta^T(k-1)x(k)|\theta(k-1))\}\right\} = \\ &= M\left\{\sqrt{\frac{2}{\pi\sigma_e^2}}M\{x(k)x^T(k)\theta(k-1)|\theta(k-1)\}\right\} = \\ &= \sqrt{\frac{2}{\pi\sigma_e^2}}M\{x(k)x^T(k)\theta(k-1)\} = \sqrt{\frac{2}{\pi\sigma_e^2}}R_{xx}M\{\theta(k-1)\}, \end{aligned} \tag{19}$$

where σ_e is the RMS value of the error $e(k)$.

The expression (14) was obtained using Price's theorem [37], according to which for two random Gaussian quantities x and y with zero expectations, the following is true

$$M\{x \text{sign} y\} = \sqrt{\frac{2}{\pi}}\frac{1}{\sigma_y}M\{xy\},$$

where σ_y is the RMS value of y .

Taking into account the properties of interference

$$M\{\xi(k)x(k)\} = 0$$

and expressions (13), (14), we have

$$M\{\tilde{\theta}(k)\} = \{I - 2\gamma\lambda R_{xx} - \gamma(1-\lambda)\beta R_{xx}\}M\{\tilde{\theta}(k-1)\}, \tag{20}$$

whence it follows that the procedure (7) will converge in mean if the parameter γ satisfies the inequality

$$0 < \gamma < \frac{2}{(2\lambda + (1-\lambda)\beta)\text{tr}R_{xx}}. \tag{21}$$

Here

$$\beta = \sqrt{\frac{2}{\pi\sigma_e^2}};$$

$\text{tr}R_{xx}$ is the trace of the matrix R_{xx} .

Given the properties of the input signals, we can write the condition (21) as follows:

$$0 < \gamma < \frac{2}{(2\lambda + (1-\lambda)\beta)N\sigma_x^2}. \tag{22}$$

Consider the Lyapunov function $M\{\|\tilde{\theta}(k)\|^2\}$.

Taking the mathematical expectation from both sides of (13), we have

$$\begin{aligned} M\{\|\tilde{\theta}(k)\|^2\} &= M\{\|\tilde{\theta}(k-1)\|^2 - 4\gamma\lambda(\tilde{\theta}^T(k-1)x(k))^2 - \\ &\quad - 2\gamma(1-\lambda)\tilde{\theta}^T(k-1)x(k)\text{sign}(\tilde{\theta}^T(k-1)x(k)) + \\ &\quad + 4\gamma^2\lambda^2(\tilde{\theta}^T(k-1)x(k))^2\|x(k)\|^2 + \\ &\quad + 4\gamma^2\lambda(1-\lambda)(\tilde{\theta}^T(k-1)x(k))\text{sign}(\tilde{\theta}^T(k-1)x(k))\|x(k)\|^2 + \\ &\quad + \gamma^2(1-\lambda)^2\|x(k)\|^2\}. \end{aligned} \tag{23}$$

Consider each term on the right-hand side of (23) taking into account the statistical properties of signals and interference. In our case

$$M\{(\tilde{\theta}^T(k-1)x(k))^2\} = \sigma_\xi^2 + \sigma_x^2 M\{\|\theta(k-1)\|^2\}; \tag{24}$$

$$\begin{aligned} M\{(\tilde{\theta}^T(k-1)x(k))^2\|x(k)\|^2\} &= \\ &= 3\sigma_x^4 M\{\|\tilde{\theta}(k-1)\|^2\} + N\sigma_x^2\sigma_\xi^4; \end{aligned} \tag{25}$$

$$M\{\|x(k)\|^2\} = N\sigma_x^2. \tag{26}$$

The expression for

$$M\{\tilde{\theta}^T(k-1)x(k)\text{sign}(\theta^T(k-1)x(k))\}$$

is analyzed by analogy with (14)

$$\begin{aligned} M\{\tilde{\theta}^T(k-1)x(k)\text{sign}(\tilde{\theta}^T(k-1)x(k))\} &= \\ &= M\{M\{\tilde{\theta}^T(k-1)x(k)\text{sign}(\tilde{\theta}^T(k-1)x(k))|\tilde{\theta}(k-1)\}\} = \\ &= M\left\{\tilde{\theta}^T(k-1)\sqrt{\frac{2}{\pi\sigma_{e\tilde{\theta}(k-1)}}}M\{x(k)x^T(k)\tilde{\theta}(k-1)|\tilde{\theta}(k-1)\}\right\} \approx \\ &\approx M\{\tilde{\theta}^T(k-1)\beta\tilde{\theta}(k-1)\}R_{xx} = \\ &= \beta\text{tr}R_{xx}M\{\|\tilde{\theta}(k-1)\|^2\} = \beta\sigma_x^2 M\{\|\tilde{\theta}(k-1)\|^2\}. \end{aligned} \tag{27}$$

Similarly, we obtain

$$M\left\{\tilde{\theta}^T(k-1)x(k)\text{sign}(\tilde{\theta}^T(k-1)x(k))\|x(k)\|^2\right\} = 3\beta\sigma_x^4 M\left\{\|\tilde{\theta}(k-1)\|^2\right\}. \quad (28)$$

Substituting the expressions (19), (24)–(28) in (23) and simple transformations yield

$$M\left\{\|\tilde{\theta}(k)\|^2\right\} = \left(\frac{1-4\gamma\lambda\sigma_x^2 + 2\gamma(1-\lambda)\beta\sigma_x^2 +}{+12\gamma^2\lambda^2\sigma_x^4 + 12\gamma^2(1-\lambda)\sigma_x^4} \right) \times M\left\{\|\tilde{\theta}(k)\|^2\right\} + \gamma^2(1-\lambda)^2 N\sigma_\xi^2\sigma_x^2. \quad (29)$$

It follows from (29) that the procedure (7) will converge in mean-square (the increment of the Lyapunov function will be negative) if the following condition holds

$$\left| \frac{1-4\gamma\lambda\sigma_x^2 - 2\gamma(1-\lambda)\beta\sigma_x^2 +}{+12\gamma^2\lambda^2\sigma_x^4 + 12\gamma^2\lambda^2(\lambda+(1-\lambda)\beta)\sigma_x^4} \right| < 1. \quad (30)$$

5. Determination of maximum attainable (asymptotic) values of parameter estimation errors and identification errors

The above relation (29) allows obtaining an expression for the asymptotic estimation error

$$M\left\{\|\tilde{\theta}(\infty)\|^2\right\} = \frac{4\gamma^2\lambda^2 N\sigma_x^2\sigma_\xi^2 + \gamma^2(1-\lambda)^2 N(1-\varepsilon)\sigma_x^2 + (1-A)\sigma_s^2}{A}, \quad (31)$$

where

$$A = 4\gamma\lambda\sigma_x^2 + 2\gamma(1-\lambda)\beta\sigma_x^2 - 12\gamma^2\lambda N\sigma_x^4\sigma_\xi^2[\lambda+(1-\lambda)\beta].$$

It is assumed that in a stationary state

$$M\left\{\|\tilde{\theta}(k)\|^2\right\} = M\left\{\|\tilde{\theta}(k-1)\|^2\right\} = M\left\{\|\tilde{\theta}(\infty)\|^2\right\}.$$

Substitution of (30) into (11) yields an expression for the asymptotic identification error.

The relations obtained give non-asymptotic and asymptotic estimates for the gradient least squares algorithm ($\lambda=1$) and gradient least absolute deviations algorithm ($\lambda=0$). Thus, the conditions of mean-square convergence (29) of the gradient least squares algorithm take the following form

$$|1-4\gamma\lambda\sigma_x^2 + 12\gamma^2\beta\sigma_x^4| < 1.$$

Where we get the inequality for the parameter γ

$$0 < \gamma < \frac{1}{3\sigma_x^2},$$

which coincides with the result obtained in [38].

Moreover, the asymptotic error is determined by the relation

$$M\left\{\|\tilde{\theta}(\infty)\|^2\right\} = \frac{4\gamma^2 N\sigma_x^2\sigma_\xi^2 + (1-4\gamma\sigma_x^2 + 12\gamma^2\sigma_x^4)\sigma_s^2}{4\gamma\sigma_x^2(1-3\gamma\sigma_x^2)}.$$

It follows that when identifying a stationary object ($\sigma_s^2=0$) using this algorithm, the asymptotic estimation error is determined as follows:

$$M\left\{\|\tilde{\theta}(\infty)\|^2\right\} = \frac{\gamma N\sigma_\xi^2}{1-3\gamma\sigma_x^2}.$$

From the resulting expression, it can be seen that, to ensure

$$\lim_{k \rightarrow \infty} M\left\{\|\tilde{\theta}(k-1)\|^2\right\} = 0$$

the parameter γ should be selected as variable and tend to zero with increasing k , that is, satisfy the Dvoretzky conditions [39].

For the gradient least absolute deviations algorithm obtained from (30) with ($\lambda=0$), the inequality for γ has the following form

$$|1-2\gamma\beta\sigma_x^2| < 1,$$

where

$$0 < \gamma < \frac{1}{\beta\sigma_x^2}.$$

Substitution of ($\lambda=0$) in (31) leads to the following expression for the asymptotic error:

$$M\left\{\|\tilde{\theta}(\infty)\|^2\right\} = \frac{\gamma^2 N\sigma_x^2 + (1-2\gamma\beta\sigma_x^2)\sigma_s^2}{2\gamma\beta\sigma_x^2}.$$

From the last expression we obtain that for the asymptotic error in the stationary case ($\sigma_s^2=0$), the following relation is valid

$$M\left\{\|\tilde{\theta}(\infty)\|^2\right\} = \frac{\gamma N}{2\beta},$$

i. e., to ensure

$$\lim_{k \rightarrow \infty} M\left\{\|\tilde{\theta}(k-1)\|^2\right\} = 0$$

the parameter γ should also be selected as variable and tend to zero with increasing n , i. e., satisfy the Dvoretzky conditions.

We note that in [23, 24] the question of choosing the variable mixing parameter λ , that is $\lambda=\lambda(k)$, was considered for the case of symmetric distribution $y(k)$ ($y(k) \sim N(0, \sigma_y^2)$)

$$\lambda(k) = Prob\{y(k) > y_0 \cup y(k) < -y_0\} = 2erfc\left[\frac{|y(k)|}{\hat{\sigma}_y}\right] = 2\left[\frac{1}{2} - \int_0^{|y(k)|/\sigma_y} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx\right],$$

where $\hat{\sigma}_y$ is the standard deviation $y(k)$ in the absence of interference; y_0 is some positive number. Analysis of

this formula indicates that it is quite problematic to use in practice, since it contains a number of unknown quantities $(y(k), y_0, \sigma_y^2)$.

6. Simulation of capabilities of investigated algorithms

For the experimental study of the capabilities of the algorithm (7), identification of a linear object (FIR filter [25, 26]) was performed described by equation (1) with

$$\theta^* = (1, 2, 3, 4, 5, 4, 3, 2, 1)^T.$$

When testing the robustness of the algorithms, independent Gaussian interference with a much larger amplitude was added to the output signal of the object to simulate “outliers” in the system (impulse interference). An example of such interference is shown in Fig. 1. The simulation results are presented in Fig. 2–5. At the same time, graphs of the model parameters adjustment are shown on the left, and identification error changes on the right. Fig. 2 reflects the result of identification of a linear system with $\lambda=1$ in (7) and no measurement interference. Fig. 3–5 show the results of identification with impulse interference using the algorithm (7) for $\lambda=1, 0.8, 0.6$, respectively. When $\lambda=0$, i. e., when using the sign algorithm, the identification process diverges.

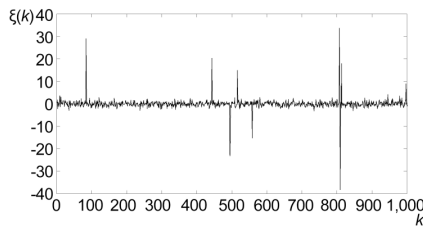


Fig. 1. Example of interference

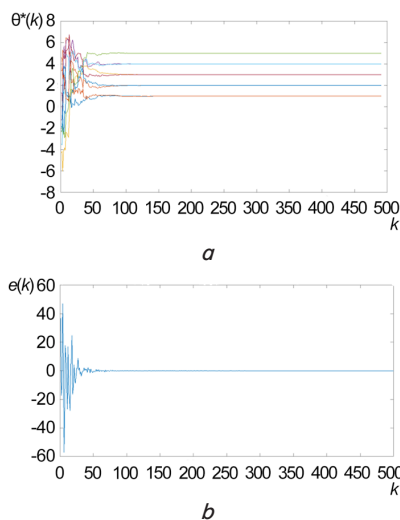


Fig. 2. Simulation results with $\lambda=1$ in (7) and no measurement interference: *a* – result of parameters adjustment; *b* – identification error

From Fig. 1–5 it is seen that the use of the combined functional (6) allows obtaining estimates that are robust with respect to non-Gaussian (impulse interference). However, the question of choosing the optimal weighting parameter λ remains open.

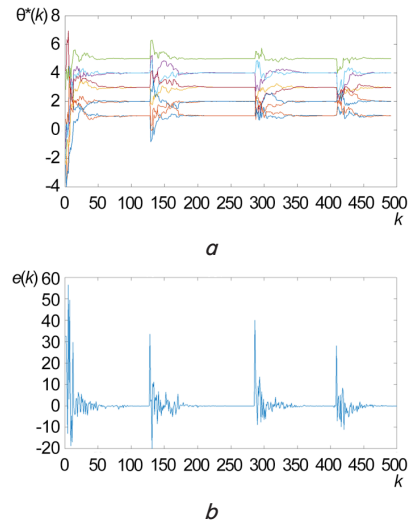


Fig. 3. Simulation results with $\lambda=1$ in (7): *a* – result of parameters adjustment; *b* – identification error

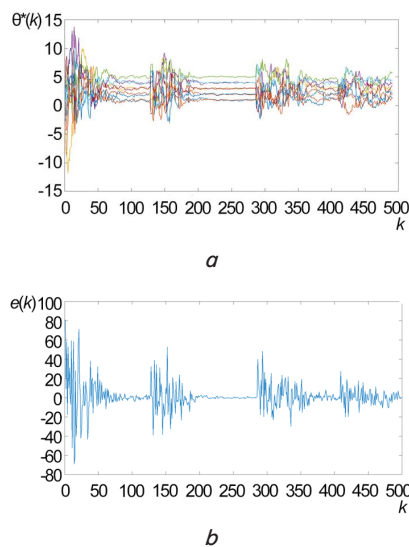


Fig. 4. Simulation results with $\lambda=0.8$ in (7): *a* – result of parameters adjustment; *b* – identification error

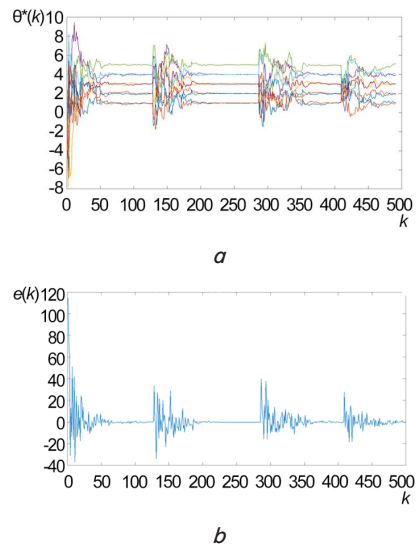


Fig. 5. Simulation results with $\lambda=0.6$ in (7): *a* – result of parameters adjustment; *b* – identification error

7. Discussion of the results of the study of the identification algorithm of a non-stationary object with non-Gaussian interference

The studies presented in the paper are a continuation and development of previous studies described in [5–7]. In these works, rigorous results were obtained for the Kaczmarz and Nagumo-Noda algorithms in the identification of non-stationary objects under Gaussian interference. These results were used to study the properties of the gradient algorithm of identification of a non-stationary object with non-Gaussian measurement interference.

As the research results showed, the use of the combined functional for object identification under non-Gaussian interference allows obtaining robust estimates. Conditions of convergence of the gradient algorithm of identification of a non-stationary object with non-Gaussian measurement interference determined by the expression (30) were obtained. In addition, the relation for the maximum attainable (asymptotic) estimation error is derived as the formula (31). As shown in the paper, the obtained relations yield the non-asymptotic and asymptotic estimates for the gradient least squares algorithm ($\lambda=1$) and gradient least absolute deviations algorithm ($\lambda=0$). The experimental results shown in Fig. 1–5 indicate the effectiveness of the developed approach and feasibility of using the combined functional in solving practical problems.

The obtained estimates are fairly general and depend both on the degree of object non-stationarity and on the statistical characteristics of useful signals and interference. In addition, expressions for the asymptotic values of the parameter estimation error and asymptotic accuracy of identification are determined. Since these expressions contain a number of unknown parameters (dispersion of signals σ_x^2 and interference σ_ξ^2 and degree of object non-stationarity σ_s^2), estimates of these parameters should be used for their practical application. So, in on-line identification, it is possible to apply any recurrent procedure for estimating them and use the resulting estimates to refine the parameters included in the algorithms. In off-line identification, the result obtained after all calculations should be corrected. In addition, the asymptotic values of the estimation error and identification accuracy depend on the choice of mixing parameter $\lambda \in [0, 1]$.

It should be noted that the estimates obtained in this paper allow the researcher, when solving practical problems, to preliminarily evaluate the capabilities and efficiency of this algorithm. However, the question remains of the optimal choice of the value of mixing parameter $\lambda \in [0, 1]$. Therefore, it seems important to conduct research in the direction of:

1) studying the effectiveness of the developed approach in identifying non-stationary objects using a model different from the first-order Markov model to describe non-stationarity;

2) establishing the dependence of the speed of the identification algorithm on the degree of non-stationarity of the investigated object;

3) developing recommendations for choosing the optimal value of mixing parameter λ or rules for its correction;

4) studying the effectiveness of such an approach in identifying objects in conditions of not only impulse interference, but also interference, having, for example, asymmetric distributions.

8. Conclusions

1 Conditions are determined and analytical estimates of mean and mean-square convergence rate of the gradient algorithm of identification of the parameters described by the first-order Markov model with non-Gaussian measurement interference are obtained.

2. Non-asymptotic and asymptotic estimates of estimation accuracy and identification error values are obtained, which allow determining the maximum attainable properties of the algorithm. These estimates are quite general and depend both on the degree of object non-stationarity σ_s^2 , and statistical characteristics of the signals σ_x^2 of interference σ_ξ^2 .

3. It is shown that the asymptotic values of the estimation error and identification accuracy depend on the choice of mixing parameter $\lambda \in [0, 1]$. It is intuitively clear that the efficiency of choosing this parameter depends on the characteristics of the problem to be solved. Therefore, it is not possible to develop general recommendations regarding the choice. However, it seems advisable to develop recommendations for choosing this parameter for a number of distribution classes.

References

1. Kaczmarz, S. (1993). Approximate solution of systems of linear equations†. *International Journal of Control*, 57 (6), 1269–1271. doi: <https://doi.org/10.1080/00207179308934446>
2. Raybman, N. S., Chadeev, V. M. (1966). *Adaptivnye modeli v sistemah upravleniya*. Moscow: Sovetskoe radio, 156.
3. Aved'yan, E. D. (1978). Modified Kaczmarz algorithms for estimating the parameters of linear plants. *Avtomatika i telemekhanika*, 5, 64–72.
4. Liberol', B. D., Rudenko, O. G., Timofeev, V. A. (1995). Modifitsirovannyi algoritm Kachmazha dlya otsenivaniya parametrov nestatsionarnykh obektov. *Problemy upravleniya i informatiki*, 3, 81–89.
5. Liberol', B. D., Rudenko, O. G., Bessonov, A. A. (2018). Issledovanie shodimosti odnoshagovykh adaptivnykh algoritmov identifikatsii. *Problemy upravleniya i informatiki*, 5, 19–32.
6. Strohmer, T., Vershynin, R. (2009). Comments on the Randomized Kaczmarz Method. *Journal of Fourier Analysis and Applications*, 15 (4), 437–440. doi: <https://doi.org/10.1007/s00041-009-9082-0>
7. Rudenko, O., Bezsonov, O., Romanyk, O., Lebediev, V. (2019). Analysis of convergence of adaptive singlestep algorithms for the identification of nonstationary objects. *Eastern-European Journal of Enterprise Technologies*, 1 (4 (97)), 6–14. doi: <https://doi.org/10.15587/1729-4061.2019.157288>
8. Rudenko, O. G., Bezsonov, O. O. (2019). The Regularized Adaline Learning Algorithm for the Problem of Evaluation of Non-Stationary Parameters. *Control Systems and Computers*, 1, 22–30. doi: <https://doi.org/10.15407/usim.2019.01.022>
9. Shao, T., Zheng, Y. R., Benesty, J. (2010). An Affine Projection Sign Algorithm Robust Against Impulsive Interferences. *IEEE Signal Processing Letters*, 17 (4), 327–330. doi: <https://doi.org/10.1109/lsp.2010.2040203>

10. Shin, J., Yoo, J., Park, P. (2012). Variable step-size affine projection sign algorithm. *Electronics Letters*, 48 (9), 483. doi: <https://doi.org/10.1049/el.2012.0751>
11. Lu, L., Zhao, H., Li, K., Chen, B. (2015). A Novel Normalized Sign Algorithm for System Identification Under Impulsive Noise Interference. *Circuits, Systems, and Signal Processing*, 35 (9), 3244–3265. doi: <https://doi.org/10.1007/s00034-015-0195-1>
12. Huang, H.-C., Lee, J. (2012). A New Variable Step-Size NLMS Algorithm and Its Performance Analysis. *IEEE Transactions on Signal Processing*, 60 (4), 2055–2060. doi: <https://doi.org/10.1109/tsp.2011.2181505>
13. Casco-Sánchez, F. M., Medina-Ramírez, R. C., López-Guerrero, M. (2011). A New Variable Step-Size NLMS Algorithm and its Performance Evaluation in Echo Cancelling Applications. *Journal of Applied Research and Technology*, 9 (3), 302–313.
14. Huber, P. J. (1977). Robust methods of estimation of regression coefficients. *Series Statistics*, 8 (1), 41–53. doi: <https://doi.org/10.1080/02331887708801356>
15. Huber, P. J. (1964). Robust Estimation of a Location Parameter. *The Annals of Mathematical Statistics*, 35 (1), 73–101. doi: <https://doi.org/10.1214/aoms/1177703732>
16. Hampel, F. R. (1974). The Influence Curve and its Role in Robust Estimation. *Journal of the American Statistical Association*, 69 (346), 383–393. doi: <https://doi.org/10.1080/01621459.1974.10482962>
17. Adamczyk, T. (2017). Application of the Huber and Hampel M-estimation in real estate value modeling. *Geomatics and Environmental Engineering*, 11 (1), 15. doi: <https://doi.org/10.7494/geom.2017.11.1.15>
18. Andrews, D. F. (1974). A Robust Method for Multiple Linear Regression. *Technometrics*, 16 (4), 523. doi: <https://doi.org/10.2307/1267603>
19. Yohai, V. J. (1987). High Breakdown-Point and High Efficiency Robust Estimates for Regression. *The Annals of Statistics*, 15 (2), 642–656. doi: <https://doi.org/10.1214/aos/1176350366>
20. Croux, C., Rousseeuw, P. J., Hossjer, O. (1994). Generalized S-Estimators. *Journal of the American Statistical Association*, 89 (428), 1271. doi: <https://doi.org/10.2307/2290990>
21. Rudenko, O. G., Bessonov, A. A. (2011). Robastnoe obuchenie radial'no-bazisnyh setey. *Kibernetika i sistemniy analiz*, 6, 38–46.
22. Rudenko, O. G., Bessonov, A. A. (2014). Robust neuroevolutionary identification of nonlinear nonstationary objects. *Kibernetika i sistemniy analiz*, 50 (1), 21–36.
23. Rudenko, O. G., Bessonov, A. A., Rudenko, C. O. (2013). Robastnaya identifikatsiya nelineynyh obektov s pomoshch'yu evolyutsioniruyushchey radial'no-bazisnoy seti. *Kibernetika i sistemniy analiz*, 2, 15–26.
24. Rudenko, O., Bezsonov, O. (2011). Function Approximation Using Robust Radial Basis Function Networks. *Journal of Intelligent Learning Systems and Applications*, 03 (01), 17–25. doi: <https://doi.org/10.4236/jilsa.2011.31003>
25. Walach, E., Widrow, B. (1984). The least mean fourth (LMF) adaptive algorithm and its family. *IEEE Transactions on Information Theory*, 30 (2), 275–283. doi: <https://doi.org/10.1109/tit.1984.1056886>
26. Bershad, N. J., Bermudez, J. C. M. (2011). Mean-square stability of the Normalized Least-Mean Fourth algorithm for white Gaussian inputs. *Digital Signal Processing*, 21 (6), 694–700. doi: <https://doi.org/10.1016/j.dsp.2011.06.002>
27. Eweda, E., Zerguine, A. (2011). New insights into the normalization of the least mean fourth algorithm. *Signal, Image and Video Processing*, 7 (2), 255–262. doi: <https://doi.org/10.1007/s11760-011-0231-y>
28. Eweda, E. (2012). Global Stabilization of the Least Mean Fourth Algorithm. *IEEE Transactions on Signal Processing*, 60 (3), 1473–1477. doi: <https://doi.org/10.1109/tsp.2011.2177976>
29. Eweda, E., Bershad, N. J. (2012). Stochastic Analysis of a Stable Normalized Least Mean Fourth Algorithm for Adaptive Noise Canceling With a White Gaussian Reference. *IEEE Transactions on Signal Processing*, 60 (12), 6235–6244. doi: <https://doi.org/10.1109/tsp.2012.2215607>
30. Hubscher, P. I., Bermudez, J. C. M., Nascimento, Vi. H. (2007). A Mean-Square Stability Analysis of the Least Mean Fourth Adaptive Algorithm. *IEEE Transactions on Signal Processing*, 55 (8), 4018–4028. doi: <https://doi.org/10.1109/tsp.2007.894423>
31. Chambers, J. A., Tanrikulu, O., Constantinides, A. G. (1994). Least mean mixed-norm adaptive filtering. *Electronics Letters*, 30 (19), 1574–1575. doi: <https://doi.org/10.1049/el:19941060>
32. Rakesh, P., Kumar, T. K., Albu, F. (2019). Modified Least-Mean Mixed-Norm Algorithms For Adaptive Sparse System Identification Under Impulsive Noise Environment. 2019 42nd International Conference on Telecommunications and Signal Processing (TSP). doi: <https://doi.org/10.1109/tsp.2019.8768813>
33. Chambers, J., Avlonitis, A. (1997). A robust mixed-norm adaptive filter algorithm. *IEEE Signal Processing Letters*, 4 (2), 46–48. doi: <https://doi.org/10.1109/97.554469>
34. Papoulis, E. V., Stathaki, T. (2004). A Normalized Robust Mixed-Norm Adaptive Algorithm for System Identification. *IEEE Signal Processing Letters*, 11 (1), 56–59. doi: <https://doi.org/10.1109/lsp.2003.819353>
35. Arenas-García, J., Figueiras-Vidal, A. R. (2005). Adaptive combination of normalised filters for robust system identification. *Electronics Letters*, 41 (15), 874. doi: <https://doi.org/10.1049/el:20051936>
36. Wagner, K. T., Doroslovacki, M. I. (2008). Towards analytical convergence analysis of proportionate-type nlms algorithms. 2008 IEEE International Conference on Acoustics, Speech and Signal Processing. doi: <https://doi.org/10.1109/icassp.2008.4518487>
37. Price, R. (1958). A useful theorem for nonlinear devices having Gaussian inputs. *IEEE Transactions on Information Theory*, 4 (2), 69–72. doi: <https://doi.org/10.1109/tit.1958.1057444>
38. Feuer, A., Weinstein, E. (1985). Convergence analysis of LMS filters with uncorrelated Gaussian data. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 33 (1), 222–230. doi: <https://doi.org/10.1109/tassp.1985.1164493>
39. Gladyshev, E. G. (1965). On Stochastic Approximation. *Theory of Probability & Its Applications*, 10 (2), 275–278. doi: <https://doi.org/10.1137/1110031>