

UDC 004.852
 DOI: 10.15587/1729-4061.2021.225765

DEVELOPING A MULTI-STEP RECURRENT ALGORITHM TO MAXIMIZE THE CRITERIA OF CORRENTROPY

O. Rudenko

Doctor of Technical Sciences, Professor,
 Head of Department*
 E-mail: oleh.rudenko@nure.ua

O. Bezsonov

Doctor of Technical Sciences, Professor*
 E-mail: oleksandr.bezsonov@nure.ua

V. Borysenko

PhD, Associate Professor**
 E-mail: victor.borysenko@nure.ua

T. Borysenko

PhD, Associate Professor**
 E-mail: tetiana.borysenko@nure.ua

S. Lyashenko

Doctor of Technical Sciences, Professor
 Department of Life Safety and Law
 Kharkiv Petro Vasylenko
 National Technical University of Agriculture
 Alchevskih str., 44, Kharkiv, Ukraine, 61002
 E-mail: yashenkosa05@ukr.net

*Department of Computer Intelligent
 Technologies and Systems***

Department of Electronic Computers*

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

This paper considers the task of constructing a linear model of the object studied using a robust criterion. The functionality applied, in this case, is correntropy. That makes it possible to obtain estimates that have robust properties. The evaluation algorithm is a multi-step procedure that employs a limited number of information measurements, that is, it has limited memory. The feature of the algorithm is that the matrices and observation vectors involved in estimate construction are formed in the following way: they include information about the newly arrived measurements and exclude information about the oldest ones. Depending on the way these matrices and vectors are built (new information is added first, and then outdated is excluded, or the outdated is first excluded, and then a new one is added), two estimate forms are possible. The second Lyapunov method is used to study the convergence of the algorithm. The conditions of convergence for a multi-step algorithm have been defined. The analysis of the established regime has revealed that the algorithm ensures that unbiased estimates are obtained.

It should be noted that all the estimates reported in this work depend on the choice of the width of the nucleus, the information weighting factor, and the algorithm memory, the task of determining which remains open. Therefore, these parameters' estimates should be applied for the practical use of such multi-step algorithms.

The estimates obtained in this paper allow the researcher to pre-evaluate the possibilities of identification using a multi-step algorithm, as well as the effectiveness of its application when solving practical tasks

Keywords: correntropy, multi-step algorithm, kernel width, information weighting factor, algorithm memory, permanence

Received date 17.12.2020

Accepted date 05.02.2021

Published date 26.02.2021

Copyright © 2021, O. Rudenko, O. Bezsonov, V. Borysenko, T. Borysenko, S. Lyashenko

This is an open access article under the CC BY license

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

1. Introduction

Many of the tasks related to information processing (identification, management, forecasting, classification, filtering, etc.) [1–15] are reduced to constructing and analyzing a model in the following form

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

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

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

is the vector of input signals $N \times 1$; $\theta^* = (\theta_1^*, \theta_2^*, \dots, \theta_N^*)^T$ is the vector of the desired parameters $N \times 1$; $\xi(k)$ is an interference.

For example, identification and filtering tasks are to determine (estimate) the vector of parameters θ^* included in equation (1). To this end, one uses some functionality, chosen in advance, whose minimization produces the required solution. The functionality type depends on the interference distribution. The application of the most widely-used quadratic functionality for this purpose ensures obtaining the asymptotically optimal estimate of the vector θ^* with minimal variance in the class of unbiased estimates at normal interfer-

ence distributions, that is $\xi(k) \sim \mathcal{N}(0, \sigma_\xi^2)$. If the interference distribution is different from normal, the estimate from a least-square method (LSM) is unstable. The instability of the LSM-estimate in the presence of non-Gaussian interference was the basis for the development of an alternative, robust estimation in statistics, the purpose of which was to eliminate the influence of interference.

It should be noted that having information about interference ξ belonging to a certain class of distributions makes the task much easier. In this case, a maximum likelihood estimate (M-estimate) can be derived by minimizing the criterion that represents the reverse logarithm of the interference distribution function [1]. If such information is not available, a non-quadratic criterion must be applied to assess the vector of parameters θ^* . This ensures that the estimate is robust. One of these criteria is the criterion of maximum correntropy [2, 3].

Existing algorithms that maximize this criterion are different modifications of gradient procedures, characterized by low convergence rate, which is why they are ineffective in evaluating non-stationary parameters. Therefore, it is a relevant task to develop algorithms that would provide for quality estimation and robustness of non-stationary parameters and could demonstrate a greater convergence speed.

2. Literature review and problem statement

Estimation based on the maximization of the criterion of maximum correntropy was the development of the ideas of robust evaluation through the use of non-quadratic criteria. One such criterion, in particular, is modular. The application of this criterion leads to symbolic algorithms. It is shown in [2–6] that they are quite effective when there is pulsed interference. Thus, [2, 3] studied the efficacy of the affinity projection symbolic algorithm. The affinity projection symbolic algorithm with a variable gain factor was used in [4].

It should be noted, however, that the symbolic algorithms, while providing for the robustness of the resulting estimate, demonstrate a low convergence rate. Attempts to speed up the operation of such algorithms, undertaken in [5, 6], require quite a lot of additional information and lead to an increase in their computational complexity.

The classic robust criteria, proposed by Huber [7] and Hampel [8], are a combination of quadratic and modular functionality. As shown in the cited works, this combination provides for the optimal estimates for the Gauss distribution, as well as robustness to distributions with heavy tails (emissions). It should be noted, however, that the effectiveness of the robust estimates obtained depends significantly on the many parameters used in these criteria. Although there are some recommendations for choosing these parameters, in most cases they are chosen based on the experience of the researcher [9]. Some practical recommendations have been developed in [10–12] for the choice of functional parameters for robust neural network training. The more common problem of robust estimation in the presence of interference with asymmetrical distributions was investigated in [13]. However, the issue of choosing the parameters of the functionality remains open.

A simpler approach to building combined functionalities, consisting of both quadratic and modular, and with no specified disadvantage, was developed in [14–18]. A given

criterion, proposed in [14], was used in the cited works to solve the problem of identification in the presence of pulse interference. The issues of the stability of the normalized algorithm were studied in [15]; the issues of convergence of the algorithm were discussed in [17] – the task of choosing the optimal values for algorithm parameters was solved.

The minimum fourth-power criterion was proposed in [18]. The task of increasing the convergence rate of the algorithm to minimize a given criterion by using the optimal setting step parameter was studied in [19, 20]. In order to ensure the robustness and stability of the algorithm, it was proposed in [21] to use a variable step parameter that takes into consideration the energy of the error (in terms of the least-squares). In [22], it was proposed to modify the algorithm of the method of the least fourth power based on the quasi-Newtonian procedure. And, finally, work [23] considered the implementation of a given algorithm using quantum computing

A combined estimation criterion to speed up the identification process, which uses the combination of the quadratic criterion and the fourth-power criterion, is proposed in [24]. In [25], a given approach was used to speed up the identification process in the presence of pulse interference. The properties of the adaptive algorithm to minimize such a combined criterion were studied in [26].

A combined criterion consisting of a fourth-power and modular criterion was proposed in [27, 28]. These works established the asymptotic and non-asymptomatic properties of the identification algorithm and investigated the effect of selecting the mixing parameter value on the properties of estimates.

The criterion of the lowest average excess has been introduced in [29, 30]; a fairly simple identification algorithm has been obtained. As shown in the cited works, a given algorithm is resistant to a wide range of noises (pulse, evenly distributed, and Gaussian). In order to increase the speed of the evaluation process using a given criterion, the kernel-recursive method of the least squares is proposed in [31], and, in [32], the kernel algorithm of affinity projection. These algorithms are modifications of the LSM-type and the recurrent LSM-type algorithms that use a kernel representation [33].

Our analysis of the above works has revealed that the implementation of the algorithms reported in them is associated with problems in choosing the parameters that are included in the minimized functionalities. In addition, theoretical studies of the convergence of these algorithms require the introduction of simplistic assumptions far enough from practice. Therefore, the effectiveness of these methods depends significantly on the experience of the researcher.

Another widely adopted approach is the approach based on signal information characteristics, particularly entropy. The functionality used in this case is an explicit functionality from the probability density function (PDF) and includes all the higher-order statistical properties defined in the PDF. Because entropy measures the average uncertainty contained in a given PDF, minimizing it reduces the error. The concept of information-theoretical learning, which uses Renyi quadratic entropy as a criterion, was introduced in [34, 35], for which a non-parametric estimate based on Parzen's windows with Gauss nuclei is defined directly from the data samples. In the cited works, it was shown that when using Renyi entropy the result of training minimizes the Renyi distance between the conditional probability of the

density function of the desired and actual output signals for the specified input signals.

Numerous studies [36–44] have shown that with non-Gaussian noise in measurements an informational approach is very effective. At the same time, the criterion used should take into consideration the statistics of the error signal not only of the second but also of a higher order.

Correntropy was introduced in [36, 37] as a generalized measure of similarity, maximizing which underlies the production of sufficiently effective robust algorithms. However, while a gradient algorithm was used in [38–42] to maximize functionality, the algorithm of the recurring LSM was used in [43] to maximize functionality. It should be noted that due to the low convergence rate, both algorithms are quite ineffective in assessing non-stationary parameters. Two approaches are used to effectively assess non-stationary parameters. The first is based on the modification of RLSM by applying an information weighting parameter. Another approach suggests the use of some additional signal information in algorithms at a series of previous cycles. This approach is implemented in the algorithms of current regression analysis (CRA). However, a given approach is designed to minimize quadratic functionality, which does not ensure the robustness of estimates. In addition, existing CRA algorithms are not convenient in real time as they require, similar to LSM, a recalculation at each cycle of evaluation of the inverse matrix of observations.

In this regard, it is important to derive the recurrent ratios describing the CRA method to maximize the criterion of correntropy and to study their properties. At the same time, the recurrence that excludes the matrix inversion operation ensures their convenience for real-time implementation and the use of a correntropy criterion – the robustness of estimates.

3. The aim and objectives of the study

The aim of this study is to derive the recurrent ratios of multi-step algorithms to maximize the criterion of correntropy. This would make it possible to obtain and adjust robust estimates while information about the studied process is acquired on-line.

To accomplish the aim, the following tasks have been set:

- to derive the recurrent ratios describing a generalized modified multi-step algorithm that maximizes a correntropy criterion;
- to obtain analytical estimates of the convergence rate of multi-step algorithms maximizing the criterion of correntropy;
- to investigate the established estimation regime under the examined conditions;
- to model the process of evaluating the parameters of a linear object using multi-step algorithms.

4. Using correntropy as a similarity measure

Correntropy, defined as a localized measure of similarity, has proven to be very effective at obtaining robust estimates because it is less sensitive to emissions.

For two random variables X and Y , correntropy is defined as

$$V(X, Y) \triangleq M\{k_\sigma(X, Y)\},$$

where $M\{\bullet\}$ is a symbol of mathematical expectation; $k_\sigma(\bullet)$ are the invariants to the turns of Mercer’s kernel; σ is the kernel width.

The most widely used in the calculation of correntropy are the Gaussian ones, determined from the following formula

$$k_\sigma(X, Y) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left\{-\frac{\|X - Y\|^2}{2\sigma^2}\right\}.$$

In the tasks of identification, filtration, etc., the functionality used is the correntropy between the required output signal d_i and the output signal of the model (real) y_i . Using Gaussian kernels, the optimized functionality takes the form:

$$J_{corr}(n) = \frac{1}{\sqrt{2\pi\sigma}} \frac{1}{N} \sum_{i=n-N+1}^n \exp\left(-\frac{e_i^2}{2\sigma^2}\right),$$

where $e_i = d_i - y_i$ is the identification error (filtering).

Using Taylor’s series expansion for the Gaussian kernel makes it possible to record correntropy as follows:

$$V(X, Y) = \frac{1}{\sqrt{2\pi\sigma}} \sum_{n=0}^{\infty} \frac{(-1)^n}{2^n \sigma^{2n} n!} M\{\|X - Y\|^{2n}\}.$$

The last expression includes all the moments of the even order of random magnitude $\|x_i - y_j\|$.

In work [43], in order to eliminate pulse interference, it was proposed to use a recurrent method of weighted least squares (RLSM), minimizing the following criterion

$$\Psi_{n+1} = \exp\left(-\frac{e_{n+1}^2}{2\sigma^2}\right) \tag{2}$$

and taking the following form

$$c_{n+1} = c_n + \frac{\Psi_{n+1} P_n x_{n+1}}{\lambda + \Psi_{n+1} x_{n+1}^T P_n x_{n+1}} (y_{n+1} - c_n^T x_{n+1}); \tag{3}$$

$$P_{n+1} = \lambda^{-1} \left(P_n - \frac{\Psi_{n+1} P_n x_{n+1} x_{n+1}^T P_n}{\lambda + \Psi_{n+1} x_{n+1}^T P_n x_{n+1}} \right). \tag{4}$$

Here is $0 \leq \lambda < 1$ – the weighting factor.

Thus, the following approximation was used to derive the calculation formula P_{n+1} (4)

$$P_{n+1} = \lambda P_n + \Psi_{n+1} x_{n+1} x_{n+1}^T. \tag{5}$$

One can see from (5) that a given algorithm is some modification of the weighted RLSM.

5. Building recurrent ratios for a modified multi-step algorithm that maximizes correntropy

Current regression analysis algorithm that takes the following form

$$c_{n+1|L} = \left(X_{n+1|L}^T X_{n+1|L} \right)^{-1} X_{n+1|L}^T Y_{n+1|L}, \tag{6}$$

to maximize correntropy was studied in [44]. In (6), $L = \text{const} (L \geq N)$ is the memory of the algorithm.

As already noted, algorithms such as the recurrent current regression analysis (RCRA), the modifications of RLSM that use the finite number of observations, are prom-

ising to evaluate non-stationary parameters. However, while the use of all observations in LSM leads to an unbiased estimate, that is, the accumulation of information is a filter, then when using the finite number of observations, such filtration of interference is not possible. Therefore, in order to give the algorithm additional filtering properties, one can use the idea of exponential smoothing, using in the algorithm the mechanism of weighing the information.

The modified re-current algorithm CRA can be obtained similar to a weighted LSM, which has a mechanism to give large weights to new incoming information.

Denote the number of steps used in the construction of an estimate as L ($L \geq N$). Inclusion in the weight matrix estimate

$$A_L = \begin{pmatrix} \lambda^{L-1} & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots \\ 0 & \dots & \lambda & 0 \\ 0 & \dots & 0 & 1 \end{pmatrix}, \quad (7)$$

the dimensionality $L \times L$, where $0 < \lambda \leq 1$, modifies (6) as follows:

$$c_{n+1|L} = \left(X_{n+1|L}^T A_L X_{n+1|L} \right)^{-1} X_{n+1|L}^T A_L Y_{n+1|L}. \quad (8)$$

The feature of the algorithms with $L = \text{const}$ is that the matrix and observation vectors used in the estimate construction are formed at each estimation step as follows: they include information about newly received measurements and exclude information about the oldest ones. Depending on how these matrices and vectors are built (whether new information is added first and then the outdated is excluded, or whether the outdated is first excluded and then a new one is added), two forms of evaluation are possible.

Receiving new information (adding a new dimension) results in a calculation of the estimate, which, similar to (8), can be recorded as follows:

$$c_{n+1|L+1} = \left(X_{n+1|L+1}^T A_{L+1} X_{n+1|L+1} \right)^{-1} X_{n+1|L+1}^T A_{L+1} Y_{n+1|L+1}, \quad (9)$$

where

$$Y_{n+1|L+1} = \begin{pmatrix} Y_{n|L} \\ \dots \\ y_{n+1} \end{pmatrix} = \begin{pmatrix} y_{n-L+1} \\ \dots \\ Y_{n+1|L} \end{pmatrix} - \text{vector } (L+1) \times 1; \quad (10)$$

$$X_{n+1|L+1} = \begin{pmatrix} X_{n|L} \\ \dots \\ x_{n+1}^T \end{pmatrix} = \begin{pmatrix} x_{n-L+1}^T \\ \dots \\ X_{n+1|L} \end{pmatrix} - \text{matrix } (L+1) \times N; \quad (11)$$

$$A_{L+1} = \begin{pmatrix} \lambda A_L & \vdots & 0 \\ - & - & - \\ 0 & \vdots & 1 \end{pmatrix} = \begin{pmatrix} \lambda^L & \vdots & 0 \\ - & - & - \\ 0 & \vdots & A_L \end{pmatrix} - \text{matrix } (L+1) \times (L+1). \quad (12)$$

Consider the modification of the current regression analysis algorithm used to maximize correntropy that takes the following form

$$c_{n+1|L} = \left(\lambda X_{n|L}^T X_{n|L} + \Psi_{n+1} x_{n+1} x_{n+1}^T - \right)^{-1} \times \left(\Psi_{n-L+1} x_{n-L+1} x_{n-L+1}^T : \Psi_{n+1} x_{n+1} \right) \times \begin{pmatrix} y_{n-L+1} \\ Y_{n|L} \\ \dots \\ y_{n+1} \end{pmatrix}. \quad (13)$$

Introduce the following designations:

$$P_{n+1|L+1}^{-1} = X_{n+1|L+1}^T A_{L+1} X_{n+1|L+1} = \lambda P_{n|L}^{-1} + \Psi_{n+1} x_{n+1} x_{n+1}^T; \quad (14)$$

$$P_{n+1|L}^{-1} = X_{n+1|L}^T A_L X_{n+1|L} = P_{n+1|L+1}^{-1} - \lambda^L \Psi_{n-L+1} x_{n-L+1} x_{n-L+1}^T.$$

Then

$$P_{n+1|L}^{-1} = \lambda P_{n|L}^{-1} + \Psi_{n+1} x_{n+1} x_{n+1}^T - \lambda^L \Psi_{n-L+1} x_{n-L+1} x_{n-L+1}^T. \quad (15)$$

Applying to (14), (15) a lemma about matrix inversion, one can obtain, as already noted, two forms of calculations: one uses first the accumulation of information (includes the newly received signal x_{n+1}), and then resets the outdated information (excluded signal x_{n-L+1}), and vice versa.

Thus, the refinement of estimates and the calculation of the matrix when the outdated information is reset is performed, respectively, according to the following formulas

$$c_{n|L-1} = c_{n|L} - \frac{\lambda^{L-1} \Psi_{n-L+1} P_{n+1|L+1} x_{n-L+1}}{1 - \lambda^{L-1} \Psi_{n-L+1} x_{n-L+1}^T P_{n+1|L+1} x_{n-L+1}} \times \left(y_{n-L+1} - c_{n|L}^T x_{n-L+1} \right); \quad (16)$$

$$P_{n|L-1} = P_{n|L} + \frac{\lambda^{L-1} \Psi_{n-L+1} P_{n+1|L+1} x_{n-L+1} x_{n-L+1}^T P_{n+1|L+1}}{1 - \lambda^{L-1} \Psi_{n-L+1} x_{n-L+1}^T P_{n+1|L+1} x_{n-L+1}}, \quad (17)$$

and the ratios that describe the accumulation of information take the following form

$$c_{n+1|L} = c_{n|L-1} + \frac{\Psi_{n+1} P_{n|L} x_{n+1}}{\lambda + \Psi_{n+1} x_{n+1}^T P_{n|L} x_{n+1}} \left(y_{n+1} - c_{n|L-1}^T x_{n+1} \right); \quad (18)$$

$$P_{n+1|L+1} = \frac{1}{\lambda} \left[P_{n|L} - \frac{\Psi_{n+1} P_{n|L} x_{n+1} x_{n+1}^T P_{n|L}}{\lambda + \Psi_{n+1} x_{n+1}^T P_{n|L} x_{n+1}} \right]. \quad (19)$$

Thus, the recurrent evaluation algorithm, obtained by excluding outdated information and then adding a new one, is described by ratios (16) to (19).

6. Algorithm convergence study

To determine conditions for the convergence of algorithm (16) to (19), we shall introduce a Lyapunov function [44].

$$V_{n+1|L} = \Theta_{n+1|L}^T P_{n+1|L}^{-1} \Theta_{n+1|L}, \quad (20)$$

where $\Theta_{n+1|L} = c_{n+1|L} - c^*$ is an estimated error in the $(n+1)$ -step obtained from L observations.

Subtracting c^* from both parts of (18) considering (10) to (12), write down the algorithm regarding identification errors

$$\Theta_{n+1|L} = \left(I + \frac{\lambda^L \Psi_{n-L+1} P_{n+1|L} x_{n-L+1} x_{n-L+1}^T}{1 - \lambda^L \Psi_{n-L+1} x_{n-L+1}^T P_{n+1|L} x_{n-L+1}} \right) \Theta_{n+1|L+1}, \quad (21)$$

where I is the singular function $N \times N$.

On the other hand, taking into consideration (17), ratio (21) can be rewritten as follows:

$$\Theta_{n+1|L} = P_{n+1|L} P_{n+1|L+1}^{-1} \Theta_{n+1|L+1}, \quad (22)$$

Similarly, writing down (18) regarding estimate errors and considering (19), we obtain

$$\begin{aligned} \Theta_{n+1|L} &= \left(I - \frac{\Psi_{n+1} P_{n|L} x_{n+1} x_{n+1}^T}{\lambda + \Psi_{n+1} x_{n+1}^T P_{n|L} x_{n+1}} \right) \Theta_{n|L-1} = \\ &= P_{n+1|L+1} P_{n|L}^{-1} \Theta_{n|L}. \end{aligned} \quad (23)$$

Fitting (23) in (22) produces

$$\Theta_{n+1|L} = \lambda P_{n+1|L} P_{n|L}^{-1} \Theta_{n|L}. \quad (24)$$

Then

$$V_{n+1|L} = \lambda^2 \Theta_{n|L}^T P_{n|L}^{-1} P_{n+1|L} P_{n|L}^{-1} \Theta_{n|L}, \quad (25)$$

and the increment of the Lyapunov function is equal to

$$\begin{aligned} \Delta V_{n+1|L} &= V_{n+1|L} - V_{n|L} = \\ &= \Theta_{n|L}^T P_{n|L}^{-1} \left[\lambda^2 P_{n+1|L} P_{n|L}^{-1} - I \right] \Theta_{n|L}. \end{aligned} \quad (26)$$

Applying formulae (14), (15), we determine

$$\begin{aligned} P_{n+1|L} P_{n|L}^{-1} &= \\ &= \left[\frac{1}{\lambda} \left[P_{n|L} + \frac{\lambda^{L-1} \Psi_{n+1} P_{n|L} x_{n+1} x_{n+1}^T P_{n|L}}{1 - \lambda^{L-1} \Psi_{n+1} x_{n+1}^T P_{n|L} x_{n+1}} \right] - \right. \\ &\quad \left. \frac{\Psi_{n+1} \frac{1}{\lambda^2} \left(P_{n|L} + \frac{\lambda^{L-1} \Psi_{n+1} P_{n|L} x_{n+1} x_{n+1}^T P_{n|L}}{1 - \lambda^{L-1} \Psi_{n+1} x_{n+1}^T P_{n|L} x_{n+1}} \right) x_{n+1} x_{n+1}^T \times}{1 - \Psi_{n+1} x_{n+1}^T \left(P_{n|L} - \frac{\lambda^{L-1} \Psi_{n+1} P_{n|L} x_{n+1} x_{n+1}^T P_{n|L}}{1 - \lambda^{L-1} \Psi_{n+1} x_{n+1}^T P_{n|L} x_{n+1}} \right) x_{n-L+1}} \right] P_{n|L}^{-1}. \end{aligned} \quad (27)$$

To calculate the third term in the bracket, we introduce the following designations:

$$\begin{aligned} e_n &= \Theta_{n+1|n-L}^T x_{n+1}; \quad e_L = \Theta_{n+1|L}^T x_{n-L+1}; \\ \alpha &= x_{n+1}^T P_{n+1|L} x_{n+1}; \quad \beta = x_{n-L+1}^T P_{n+1|L} x_{n-L+1}; \end{aligned} \quad (28)$$

$$\gamma = x_{n+1}^T P_{n+1|L} x_{n-L+1}; \quad \Psi_L = \Psi_{n-L+1}; \quad \Psi_n = \Psi_{n+1}.$$

After simple transformations, we shall obtain that the numerator and denominator of this component take the following form:

– the numerator

$$\Psi_n (1 - \lambda^{L-1} \Psi_L \beta) e_n^2 + 2\lambda^{L-1} \Psi_n \Psi_L \gamma e_n e_L + \frac{\lambda^{2(L-1)} \Psi_L^2 \Psi_n \gamma^2 e_L^2}{1 - \lambda^{L-1} \Psi_L \beta};$$

– denominator

$$\frac{\lambda (1 - \lambda^{L-1} \Psi_L \beta) + \Psi_n \alpha (1 - \lambda^{L-1} \Psi_L \beta) + \lambda^{L-1} \Psi_n \Psi_L \gamma^2}{1 - \lambda^{L-1} \Psi_L \beta}.$$

Substituting these expressions in (27) and the simple transformations considering the introduced designations (28) produce

$$\begin{aligned} &\frac{1}{\lambda} \frac{\lambda^{L-1} \Psi_L e_L^2}{(1 - \lambda^{L-1} \Psi_L \beta)} + \\ &\frac{\Psi_n (1 - \lambda^{L-1} \Psi_L \beta) e_n^2 - 2\lambda^{L-1} \Psi_n \Psi_L \gamma e_n e_L + \frac{\lambda^{2(L-1)} \Psi_L^2 \Psi_n \gamma^2 e_L^2}{(1 - \lambda^{L-1} \Psi_L \beta)}}{\lambda^{-1} [(1 - \lambda^{L-1} \Psi_L \beta) + \Psi_n (1 - \lambda^{L-1} \Psi_L \beta) + \lambda^{L-1} \Psi_n \Psi_L \gamma^2]} = \\ &= \frac{\Psi_L (e_L^2 + \Psi_n \Psi_L \alpha e_L^2 - \Psi_n (1 - \Psi_L \beta) e_n^2 - 2\Psi_n \Psi_L \gamma e_n e_L)}{(1 + \Psi_n \alpha) - \Psi_L \beta (1 + \Psi_n \alpha) + \Psi_n \Psi_L \gamma^2} = \\ &= \frac{\lambda^{L-1} \Psi_L [\lambda e_L^2 + \Psi_n \alpha e_L^2 + \Psi_n \beta e_n^2 - 2\Psi_n \gamma e_n e_L] - \Psi_n e_n^2}{\lambda^{L-1} [\lambda \Psi_L \beta - \Psi_n \Psi_L \gamma^2 + \Psi_n \Psi_L \alpha \beta] - \lambda - \Psi_n \alpha}. \end{aligned} \quad (29)$$

For the algorithm to converge, one needs to meet the following condition

$$\begin{aligned} V_{n+1|L} - V_{n|L} &= (\lambda - 1) V_{n|L} + \\ &+ \frac{\lambda^{L-1} \Psi_L [\lambda e_L^2 + \Psi_n \alpha e_L^2 + \Psi_n \beta e_n^2 - 2\Psi_n \gamma e_n e_L] - \Psi_n e_n^2}{\lambda^{L-1} [\lambda \Psi_L \beta - \Psi_n \Psi_L \gamma^2 + \Psi_n \Psi_L \alpha \beta] - \lambda - \Psi_n \alpha} \leq 0, \end{aligned} \quad (30)$$

that is, the expression in the right-hand part of (30) should be negative. Because $\lambda \leq 1$, the first term is not positive.

For the second term to be negative, the following condition must be met

$$\begin{aligned} &\left[\lambda^{L-1} \Psi_L \left(\lambda e_L^2 + \Psi_n \Psi_L \alpha e_L^2 + \right. \right. \\ &\quad \left. \left. + \Psi_n \beta e_n^2 - 2\Psi_n \gamma e_n e_L \right) - \Psi_n e_n^2 \right] \times \\ &\times \left[\lambda^{L-1} \left(\lambda \Psi_L \beta - \Psi_n \Psi_L \gamma^2 + \right. \right. \\ &\quad \left. \left. + \Psi_n \Psi_L \alpha \beta \right) - \lambda - \Psi_n \alpha \right] > 0. \end{aligned} \quad (31)$$

For the first term of inequality (31), taking into consideration the introduced designations, we obtain

$$\begin{aligned} &\lambda^{L-1} \Psi_L \left(e_L^2 + \Psi_n \Psi_L \alpha e_L^2 + \right. \\ &\quad \left. + \Psi_n \beta e_n^2 - 2\Psi_n \gamma e_n e_L \right) - \Psi_n e_n^2 = \\ &= \lambda^{L-1} \Psi_L \left(e_L^2 + \Psi_n (x_{n+1} e_L - x_{n-L+1} e_n)^T \times \right. \\ &\quad \left. \times P_{n+1|L} (x_{n+1} e_L - x_{n-L+1} e_n) \right) - \Psi_n e_n^2. \end{aligned}$$

Since $\Psi_L \geq 0$, we obtain

$$\begin{aligned} &\Psi_L (e_L^2 + \Psi_n \Psi_L \alpha e_L^2 + \Psi_n \beta e_n^2 - 2\Psi_n \gamma e_n e_L) = \\ &= \Psi_L \left(e_L^2 + \Psi_n (x_{n+1} e_L - x_{n-L+1} e_n)^T \times \right. \\ &\quad \left. \times P_{n+1|L} (x_{n+1} e_L - x_{n-L+1} e_n) \right) \geq 0. \end{aligned} \quad (32)$$

Denote

$$A = \frac{\Psi_n e_n^2}{\lambda^{L-1} \Psi_L (\lambda e_L^2 + \Psi_n (\alpha e_L^2 - 2\gamma e_n e_L + \beta e_n^2))}. \quad (33)$$

For the second multiplier, one can write

$$\begin{aligned} & \lambda^{L-1} \Psi_n \Psi_L (\alpha \beta - \gamma^2) + \Psi_L \beta - 1 - \Psi_n \alpha = \\ & = \lambda^{L-1} \Psi_L [\Psi_n (\alpha \beta - \gamma^2) + \beta] - 1 - \Psi_n \alpha. \end{aligned}$$

It should be noted that, given that the Cauchy-Bunyakovsky inequality is satisfied, the following holds

$$\alpha \beta - \gamma^2 \geq 0,$$

or

$$\begin{aligned} & x_{n+1}^T P_{n|L} x_{n+1} x_{n-L+1}^T P_{n|L} x_{n-L+1} \geq \\ & \geq x_{n+1}^T P_{n|L} x_{n-L+1} x_{n+1}^T P_{n|L} x_{n-L+1}. \end{aligned}$$

Thus, considering $\Psi_L \geq 0$,

$$\Psi_L [\Psi_n (\alpha \beta - \gamma^2) + \beta] \geq 0. \quad (34)$$

Denote

$$B = \frac{\lambda + \Psi_n \alpha}{\lambda^{L-1} \Psi_L [\Psi_n (\alpha \beta - \gamma^2) + \lambda \beta]}. \quad (35)$$

It should be noted that given that (32) and (34) are met, $A \geq 0$ and $B \geq 0$.

Considering the designations in (33) and (35), condition (31) can be written as follows:

$$(1-A)(1-B) > 0. \quad (36)$$

One can see from (36) that, in order to ensure the convergence of the algorithm, it is required that A and B should exceed unity.

On the other hand, consider the difference $(B-A)$.

Substituting expressions for A and B (33) and (35), after simple calculations, we obtain

$$\begin{aligned} B - A = & \\ = & \frac{(\lambda e_L + \Psi_n \alpha e_L - \Psi_n \gamma e_n)^2}{\lambda^{L-1} \Psi_L [\lambda e_L + \Psi_n \alpha e_L + \Psi_n \beta e_n^2 - 2\Psi_n \gamma e_n e_L] [\Psi_n (\alpha \beta - \gamma^2) + \lambda \beta]}. \end{aligned}$$

And since both the numerator and the denominator of a given expression are not negative, the difference $B - A \geq 0$, that is, $B \geq A$. Thus, to meet the conditions for the convergence of the algorithm, that is,

$$\Theta_{n+1|L}^T P_{n+1|L}^{-1} \Theta_{n+1|L} \leq \Theta_{n|L}^T P_{n|L}^{-1} \Theta_{n|L}, \quad (37)$$

or inequality (36), it is necessary that $A \geq 1$.

Consider the properties of the P^{-1} matrix, which is part of the Lyapunov function. The positive certainty of this matrix is required for the algorithm to converge. Consider a step-by-step change in P^{-1} according to (15).

If, in the n -th step, the matrix $P_{n|L}^{-1}$ is positively defined, then, in the case of the positive certainty of the matrix $\Psi_{n+1} x_{n+1} x_{n+1}^T - \lambda^L \Psi_{n-L+1} x_{n-L+1} x_{n-L+1}^T$, the matrix $P_{n+1|L}^{-1}$ is also positively

defined. Multiplying a given matrix, in the left-hand part, by $\Theta_{n|L}^T$, and, in the right-hand part, by $\Theta_{n|L}$, taking into consideration the designations in (40), consider the scalar value

$$\begin{aligned} & \Theta_{n|L}^T [\Psi_{n+1} x_{n+1} x_{n+1}^T - \lambda^L \Psi_{n-L+1} x_{n-L+1} x_{n-L+1}^T] \Theta_{n|L} = \\ & = \Psi_{n+1} e_{n+1}^2 - \lambda^L \Psi_{n-L+1} e_{n-L+1}^2. \end{aligned} \quad (38)$$

Since

$$\Psi_{n+1} e_{n+1}^2 - \lambda^L \Psi_{n-L+1} e_{n-L+1}^2 \geq \Psi_{n+1} e_{n+1}^2 - A \lambda^L \Psi_{n-L+1} e_{n-L+1}^2,$$

then, by substituting the expression for A (33) in a given inequality, we obtain

$$\begin{aligned} & \Psi_{n+1} e_{n+1}^2 - A \lambda^L \Psi_{n-L+1} e_{n-L+1}^2 = \\ & = \Psi_{n+1} e_{n+1}^2 \left[\frac{\Psi_{n+1} (e_{n+1} x_{n-L+1} - e_{n-L+1} x_{n+1})^T P_{n|L} (e_{n+1} x_{n-L+1} - e_{n-L+1} x_{n+1})}{e_{n-L+1}^2 + \Psi_{n+1} (e_{n+1} x_{n-L+1} - e_{n-L+1} x_{n+1})^T P_{n|L} (e_{n+1} x_{n-L+1} - e_{n-L+1} x_{n+1}) + \lambda e_{n-L+1}^2} \right] \geq 0. \end{aligned}$$

It follows that the matrix $P_{n+1|L}^{-1}$ is non-negatively defined if $P_{n|L}$ is non-negatively defined (and, therefore, $P_{n|L}^{-1}$ is a non-negatively defined matrix as well). This can be achieved by selecting the initial matrix $P_{0|L}$ as non-negatively defined, for example, as in a normal RLSM, $P_{0|L} = \alpha I$, where α is a positive number.

Therefore, the Lyapunov function, if these conditions are met, will be non-negative and limited

$$\Theta_{n+1|L}^T P_{n+1|L}^{-1} \Theta_{n+1|L} \leq \Theta_{n|L}^T P_{n|L}^{-1} \Theta_{n|L} \leq \dots \leq \Theta_{0|L}^T P_{0|L}^{-1} \Theta_{0|L}, \quad (39)$$

that is, it is limited by $V_{0|L} = \Theta_{0|L}^T P_{0|L}^{-1} \Theta_{0|L}$.

It follows from (26) that

$$\begin{aligned} & V_{n+1|L} - V_{n|L} = (\lambda - 1) V_{n|L} + \\ & + \frac{\lambda^{L-1} \Psi_L [\lambda e_L^2 + \Psi_n \alpha e_L^2 + \Psi_n \beta e_n^2 - 2\Psi_n \gamma e_n e_L] - \Psi_n e_n^2}{\lambda + \Psi_n \alpha - \lambda^{L-1} [\lambda \Psi_L \beta - \Psi_n \Psi_L \gamma^2 + \Psi_n \Psi_L \alpha \beta]} \leq 0, \end{aligned}$$

hence, it follows, since $\lambda \neq 0$,

$$\lim_{N \rightarrow \infty} \sum_{n=1}^N \frac{\lambda^{L-1} \Psi_L [\lambda e_L^2 + \Psi_n \alpha e_L^2 + \Psi_n \beta e_n^2 - 2\Psi_n \gamma e_n e_L] - \Psi_n e_n^2}{\lambda + \Psi_n \alpha - \lambda^{L-1} [\lambda \Psi_L \beta - \Psi_n \Psi_L \gamma^2 + \Psi_n \Psi_L \alpha \beta]} < 0. \quad (40)$$

In turn, we obtain from (40)

$$\begin{aligned} & \lim_{n \rightarrow \infty} \frac{\lambda^{L-1} \Psi_L [\lambda e_L^2 + \Psi_n \alpha e_L^2 + \Psi_n \beta e_n^2 - 2\Psi_n \gamma e_n e_L] - \Psi_n e_n^2}{\lambda + \Psi_n \alpha - \lambda^{L-1} [\lambda \Psi_L \beta - \Psi_n \Psi_L \gamma^2 + \Psi_n \Psi_L \alpha \beta]} = \\ & = 0, \end{aligned} \quad (41)$$

that is, the identification error decreases as time increases.

Because

$$\lambda_{\min} [P_{n+1|L}^{-1}] \geq \lambda_{\min} [P_{n|L}^{-1}] \geq \lambda_{\min} [P_0^{-1}],$$

$$\lambda_{\min} [P_{n+1|L}^{-1}] \geq \lambda_{\min} [P_{n|L}^{-1}] \geq \lambda_{\min} [P_0^{-1}],$$

and, on the other hand,

$$\begin{aligned} & \lambda_{\min} [P_0^{-1}] \|\Theta_{n+1}\|^2 \leq \lambda_{\min} [P_{n+1|L}^{-1}] \|\Theta_{n+1}\|^2 \leq \\ & \leq \Theta_{n+1}^T P_{n+1|L}^{-1} \Theta_{n+1} \leq \Theta_0^T P_0^{-1} \Theta_0 \leq \lambda_{\max} [P_0^{-1}] \|\Theta_0\|^2, \end{aligned}$$

then for the magnitude of the evaluation error $\|\theta_{n+1}\|^2$, the following inequality holds

$$\|\theta_{n+1}\|^2 \leq \frac{\lambda_{\max}[P_0^{-1}]}{\lambda_{\min}[P_0^{-1}]} \|\theta_0\|^2. \tag{42}$$

According to the above formulae, the amount of the estimation error $\|\theta_{n+1}\|^2$ decreases over time.

7. Exploring the established regime

At last, consider the established mode when the estimate is no longer adjusted, that is, $c_{n+1L} = c_{nL}$. Then

$$\Theta_{nL} = \lambda P_{n+1L} P_{nL}^{-1} \Theta_{nL}.$$

In other words,

$$(I - \lambda P_{n+1L} P_{nL}^{-1}) \Theta_{nL} = 0.$$

For the convergence of the algorithm, it is necessary that a given equality is satisfied only when $\Theta_{nL} = 0$, and the matrix $I - \lambda P_{n+1L} P_{nL}^{-1}$ is not zero.

Denote the matrix's eigenvalues through $\mu[\bullet]$.

$$\mu[I - \lambda P_{n+1L} P_{nL}^{-1}] \neq 0. \tag{43}$$

Consider the singular numbers of the matrices P_{n+1L} and P_{nL}^{-1} .

As it is known, for the maximum and minimum values of the singular numbers σ_{\max} , σ_{\min} and the eigenvalues μ of the square matrix A, the following ratio holds

$$\sigma_{\min}[A] \leq |\mu[A]| \leq \sigma_{\max}[A].$$

Since the matrices P_{n+1L} and P_{nL}^{-1} are square, and their maximum singular numbers σ_{\max} are no less than the maximum eigenvalues μ_{\max} , we obtain

$$\sigma_{\max}[\lambda P_{n+1L} P_{nL}^{-1}] \geq \mu_{\max}[\lambda P_{n+1L} P_{nL}^{-1}]. \tag{44}$$

It follows from the Cauchy-Bunyakovsky inequality that

$$\sigma_{\max}[\lambda P_{n+1L} P_{nL}^{-1}] \leq \sigma_{\max}[P_{n+1L}] \sigma_{\max}[\lambda P_{nL}^{-1}]. \tag{45}$$

Because, as shown above, the matrix

$$\Psi_{n+1} x_{n+1} x_{n+1}^T - \lambda^L \Psi_{n-L+1} x_{n-L+1} x_{n-L+1}^T$$

is square and non-negatively defined, it follows then from (14)

$$\sigma_{\max}[P_{n+1L}^{-1}] \geq \sigma_{\max}[\lambda P_{nL}^{-1}].$$

Consequently, inequality (45) can be recorded as follows:

$$\begin{aligned} \sigma_{\max}[\lambda P_{n+1L} P_{nL}^{-1}] &\leq \\ &\leq \sigma_{\max}[P_{n+1L}] \sigma_{\max}[\lambda P_{nL}^{-1}] = \frac{\sigma_{\min}[P_{n+1L}^{-1}]}{\sigma_{\max}[P_{n+1L}]} \leq 1, \end{aligned}$$

where $\sigma_{\min}[P_{n+1L}]$ is the minimum singular number of the matrix P_{n+1L} . Taking this into consideration, it follows from (45) that

$$\mu[\lambda P_{n+1L} P_{nL}^{-1}] < 1,$$

that is, (43) does hold.

Thus, under the established mode $\Theta_{nL} = 0$ or $c_{nL} = c^*$, that is, the estimate obtained using the algorithm in question is unbiased.

8. Simulation of the process of evaluating the parameters of a linear object using multi-step algorithms

They conducted 2 experiments. The first considered the task of building a stationary model, described by equation (1), with the following parameters

$$\theta^* = (-0.1, -0.3, -0.5, -0.7, 0.1, 0.2, 0.3, 0.4, 0.7, 0.9)^T.$$

The sequences of normally distributed values $x_n \sim N\{0,1\}$ were chosen as an input signal. Fig. 1 shows changes in the values of an estimation accuracy criterion

$$F = \sum_{i=1}^{10} (c_{i,n}^* - c_{i,0}^*)^2 / \sum_{i=1}^{10} (c_{i,0}^* - c_{i,0}^*)^2$$

after $n=100$ cycles when selecting, in (2), $\sigma=1$ for different values of the algorithm memory L . Fig. 1, a shows the simulation results for $\lambda=1$; Fig. 1, b – for $\lambda=0.5$.

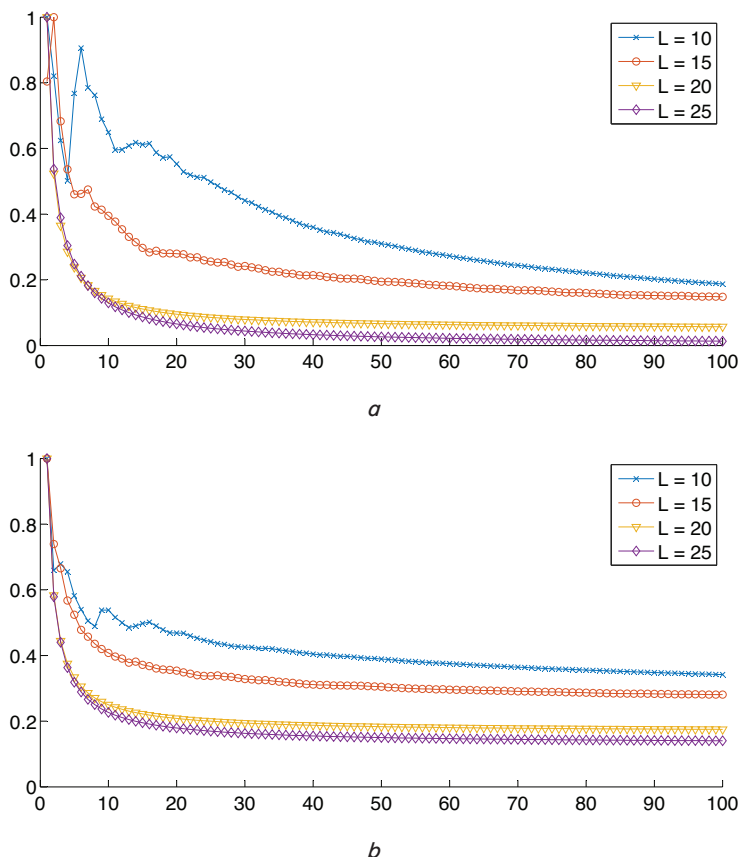


Fig. 1. Changes in the values of the estimation accuracy criterion: a – $\lambda=1$; b – $\lambda=0.5$

In the second experiment, it was assumed that in model (1) the θ_3^* , θ_5^* , θ_8^* parameters changed according to the sinusoidal law while the rest of the parameters remained the same as in the first experiment. The results of a given experiment are shown in Fig. 2.

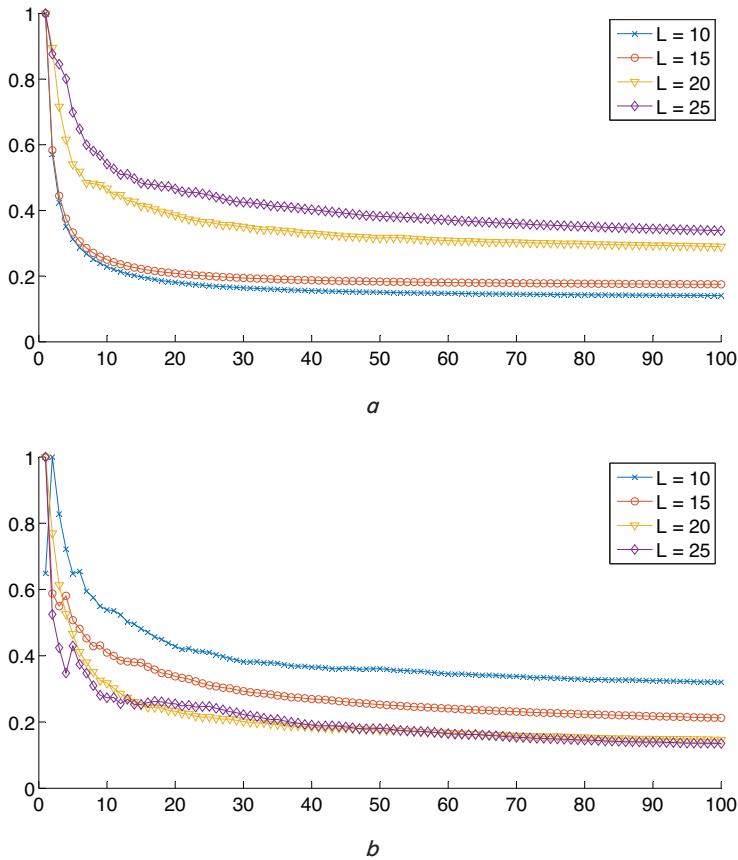


Fig. 2. Changes in the values of the estimation accuracy criterion: $a - \lambda=1$; $b - \lambda=0.5$

The above results demonstrate that when assessing the stationary parameters of model (1), it is advisable to enhance the memory of the multi-step algorithm (approximating it with RLSM). If non-stationary parameters are evaluated, then one should choose the memory whose magnitude is minimally different from the dimensionality of the object (in our case, $N=1$ and $L_{opt}=10$).

9. Discussion of results of studying the convergence of a multi-step algorithm of correntropy maximization

The research presented in this paper is a continuation and development of earlier studies described in [17, 28, 44]. While [17, 28] considered the issues related to obtaining robust estimates based on the minimization of the combined functionalities, then [44] addressed the task of robust training of the Adaline neural network. The results reported in those works were applied in this paper to study the properties of a generalized multi-step algorithm that maximizes the criterion of correntropy.

The developed recurrent ratios, describing a generalized modified multi-step algorithm to maximize the criterion of correntropy (16) to (19), make it possible to evaluate unknown parameters on-line as information becomes available.

According to the above formulae, the implementation of a given algorithm does not cause difficulties. Note that the initial values for the observation matrix should be selected in a similar way to RLSM.

The use of Lyapunov functions has made it possible to define the condition of convergence for a multi-step algorithm, determined from expression (36). Based on the given analytical estimates of the convergence of the multi-step CRA algorithms, we derived expressions to select the optimal values of the algorithm parameters that ensure its maximum convergence rate. The resulting formulae indicate that the value of the estimation error $\|\theta_{n+1}\|^2$, provided by the algorithm in question decreases over time. Expression (42) makes it possible, for any point in time, to estimate the magnitude $\|\theta_{n+1}\|^2$, which is shown to be determined by the ratio of the maximum and minimum singular numbers of the inverse covariation observation matrix $P_{n+1|L}^{-1}$. Therefore, it is necessary to know these numbers in order to estimate the magnitude $\|\theta_{n+1}\|^2$.

Studying the established estimation regime under the considered conditions showed that, given that condition (43) is met, the matrix $I - \lambda P_{n+1|L}^{-1} P_{n|L}^{-1}$ is not zero. Thus, the estimates received are unbiased, that is, $c_{i|L} = c^*$.

It should be noted that the implementation of the resulting recurrent algorithm, described by ratios (16) to (19), does not cause difficulties; it is similar to the implementation of RLSM.

Although all theoretic results have been obtained for the case of stationary parameters estimation, the results of the simulation show the effectiveness of the application of the maximum correntropy functionality for identification and linear non-stationary objects (experiment 2).

Here's what one needs to consider when applying the considered algorithm in practice. Based on our results, when evaluating stationary parameters of model (1), it is advisable to enhance the memory L of the CRA multi-step algorithm. That ensures that it is closer to the RLSM, which is optimal for a stationary case. If non-stationary parameters are evaluated, the simulation shows that, given $L \geq N$, one should select the memory L whose magnitude is minimally different from the dimensionality of the object N .

In addition, it should be noted that the estimates received in this work depend on the parameters σ used in the algorithm (kernel width), $0 < \lambda \leq 1$ (information weighting parameter), and $L = \text{const}$ (algorithm memory), the issue of selecting the values of which remains open. Therefore, when applying a given algorithm in practice, one should use the estimates of these parameters. However, the estimates derived from (40) to (42) allow the researcher to pre-evaluate the limits of a given algorithm and the effectiveness of its application when solving practical tasks.

The limitation of our study is to consider only a regular case (no interference), although it makes it possible to determine the limits of the algorithm. It would therefore seem appropriate to extend a given approach for the case involving interference to obtain appropriate statistical estimates.

In addition, the continuation of research into the dynamic properties of a given algorithm is of undoubted interest. That could make it possible to assess the effectiveness of the multi-step algorithm to evaluate the parameters of a non-stationary

object (1) in the presence and absence of information about the type of non-stationarity.

10. Conclusions

1. The recurrent ratios have been derived describing the generalized modified multi-step algorithm (CRA) that maximizes the criterion of correntropy. The use of given ratios makes it possible to evaluate the unknown parameters of an examined object under on-line mode as information becomes available and ensure that the estimates are robust. According to the derived formulae, the implementation of a given algorithm does not cause difficulties. The initial values for the observation matrix should be selected in a similar way to RLSM.

2. The application of Lyapunov functions has made it possible to determine the condition from the convergence of the multi-step algorithm, described by the expression, reduced to the verification of the ratio between the matrices of transformations A and B . Based on the obtained analytical estimates of the convergence of multi-step CRA algorithms, we defined expressions for the selection of the optimal values of parameters of the algorithm, ensuring its maximum convergence rate. According to the above formulae, the value of the estimation error $\|\theta_{n+1}\|^2$ decreases over time. The resulting expression for $\|\theta_{n+1}\|^2$ makes it possible, for any point in time, to estimate the magnitude of the estimation error, which is shown to be determined by the ratio of the maximal and minimal singular numbers of the inverse covariation observation

matrix $P_{n+1|L}^{-1}$. Therefore, to estimate the magnitude $\|\theta_{n+1}\|^2$, it is necessary to know these numbers.

3. Our study of the established estimation regime under the considered conditions showed that, due to the inequality to zero of the eigen numbers of the observation matrix $I - \lambda P_{n+1|L} P_{n|L}^{-1}$, the estimates received would be unbiased, that is, $c_{n|L} = c$.

4. We have performed the simulation of the process of evaluating the stationary and non-stationary parameters of a linear object at a different selection of the depth of the algorithm's memory. The simulation results form the basis for selecting the parameters of the algorithm when it is implemented. Based on the analysis of the simulation results, the following conclusions can be drawn. First, the application of maximum correntropy functionality is effective enough to identify the linear stationary and non-stationary objects (experiment 2). Second, the choice of the depth of the algorithm's memory $L = \text{const} (L \geq N)$ is different for identifying stationary and non-stationary objects. In the first case, the memory needs to be enhanced, in the second case – reduced to a value that is minimally different from the dimensionality of the object N .

Acknowledgments

The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

References

1. Tsympkin, Ya. Z., Polyak, B. T. (1977) Ogrublenniy metod maksimal'nogo pravdopodobiya. V kn. Dinamika sistem. Gor'kiy, 12, 22–46.
2. Tiange Shao, 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>
3. 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>
4. 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>
5. 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>
6. 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 (03). doi: <https://doi.org/10.22201/icat.16656423.2011.9.03.425>
7. Huber, P. J. (1977). Robust methods of estimation of regression coefficients. *Series Statistics*, 8 (1), 41–53. doi: <https://doi.org/10.1080/02331887708801356>
8. 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>
9. 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>
10. Rudenko, O. G., Bezsonov, O. O. (2011). Robust training of radial basis networks. *Cybernetics and Systems Analysis*, 47 (6), 38–46.
11. Rudenko, O. G., Bezsonov, O. O. (2014). Robust Neuroevolutionary Identification of Nonlinear Nonstationary Objects. *Cybernetics and Systems Analysis*, 50 (1), 17–30. doi: <https://doi.org/10.1007/s10559-014-9589-5>
12. Rudenko, O. G., Bezsonov, O. O., Rudenko, S. O. (2013). Robastnaya identifikatsiya nelineynyh obektov s pomosh'yu evolyutsioniruyushey radial'no-bazisnoy seti. *Cybernetics and Systems Analysis*, 49 (2), 15–26.
13. 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>
14. 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>
15. 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>

16. 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>
17. Rudenko, O., Bezsonov, O., Lebediev, O., Serdiuk, N. (2019). Robust identification of non-stationary objects with nongaussian interference. *Eastern-European Journal of Enterprise Technologies*, 5 (4 (101)), 44–52. doi: <https://doi.org/10.15587/1729-4061.2019.181256>
18. 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>
19. Zhang, S., Zhang, J. (2015). Fast stable normalised least-mean fourth algorithm. *Electronics Letters*, 51 (16), 1276–1277. doi: <https://doi.org/10.1049/el.2015.0421>
20. Guan, S., Meng, C., Biswal, B. (2019). Optimal step-size of least mean absolute fourth algorithm in low SNR. arXiv.org. Available at: <https://arxiv.org/ftp/arxiv/papers/1908/1908.08165.pdf>
21. Asad, S. M., Moinuddin, M., Zerguine, A., Chambers, J. (2019). A robust and stable variable step-size design for the least-mean fourth algorithm using quotient form. *Signal Processing*, 162, 196–210. doi: <https://doi.org/10.1016/j.sigpro.2019.04.021>
22. Bin Mansoor, U., Mayyala, Q., Moinuddin, M., Zerguine, A. (2017). Quasi-Newton least-mean fourth adaptive algorithm. 2017 25th European Signal Processing Conference (EUSIPCO). doi: <https://doi.org/10.23919/eusipco.2017.8081689>
23. Sadiq, A., Usman, M., Khan, S., Naseem, I., Moinuddin, M., Al-Saggaf, U. M. (2019). q-LMF: Quantum Calculus-Based Least Mean Fourth Algorithm. *Fourth International Congress on Information and Communication Technology*, 303–311. doi: https://doi.org/10.1007/978-981-15-0637-6_25
24. Zerguine, A., Cowan, C. F. N., Bettayeb, M. (1996). LMS-LMF adaptive scheme for echo cancellation. *Electronics Letters*, 32 (19), 1776. doi: <https://doi.org/10.1049/el:19961202>
25. Zerguine, A., Aboulnasr, T. (2000). Convergence analysis of the variable weight mixed-norm LMS-LMF adaptive algorithm. *Conference Record of the Thirty-Fourth Asilomar Conference on Signals, Systems and Computers (Cat. No.00CH37154)*. doi: <https://doi.org/10.1109/acssc.2000.910959>
26. Zerguine, A. (2012). A variable-parameter normalized mixed-norm (VPNMN) adaptive algorithm. *EURASIP Journal on Advances in Signal Processing*, 2012 (1). doi: <https://doi.org/10.1186/1687-6180-2012-55>
27. Rudenko, O., Bezsonov, O., Serdiuk, N., Oliinyk, K., Romanyk, O. (2020). Workable identification of objects based on minimization of combined functional. *Information Processing Systems*, 1 (160), 80–88. doi: <https://doi.org/10.30748/soi.2020.160.10>
28. Rudenko, O., Bezsonov, O., Lebediev, O., Lebediev, V., Oliinyk, K. (2020). Studying the properties of a robust algorithm for identifying linear objects, which minimizes a combined functional. *Eastern-European Journal of Enterprise Technologies*, 4 (4 (106)), 37–46. doi: <https://doi.org/10.15587/1729-4061.2020.210129>
29. Tanrikulu, O., Constantinides, A. G. (1994). Least-mean kurtosis: A novel higher-order statistics based adaptive filtering algorithm. *Electronics Letters*, 30 (3), 189–190. doi: <https://doi.org/10.1049/el:19940129>
30. Pazaitis, D. I., Constantinides, A. G. (1999). A novel kurtosis driven variable step-size adaptive algorithm. *IEEE Transactions on Signal Processing*, 47 (3), 864–872. doi: <https://doi.org/10.1109/78.747793>
31. Engel, Y., Mannor, S., Meir, R. (2004). The Kernel Recursive Least-Squares Algorithm. *IEEE Transactions on Signal Processing*, 52 (8), 2275–2285. doi: <https://doi.org/10.1109/tsp.2004.830985>
32. Gil-Cacho, J. M., Signoretto, M., van Waterschoot, T., Moonen, M., Jensen, S. H. (2013). Nonlinear Acoustic Echo Cancellation Based on a Sliding-Window Leaky Kernel Affine Projection Algorithm. *IEEE Transactions on Audio, Speech, and Language Processing*, 21 (9), 1867–1878. doi: <https://doi.org/10.1109/tasl.2013.2260742>
33. Eksioğlu, E. M., Tanc, A. K. (2011). RLS Algorithm With Convex Regularization. *IEEE Signal Processing Letters*, 18 (8), 470–473. doi: <https://doi.org/10.1109/lsp.2011.2159373>
34. Principe, J. C., Xu, D., Zhao, Q., Fisher, J. W. (2000). Learning from examples with information theoretic criteria. *Journal of VLSI signal processing systems for signal, image and video technology*, 26, 61–77. doi: <https://doi.org/10.1023/A:1008143417156>
35. Principe, J. C., Xu, D., Fisher, J. (2000). Information-theoretic learning. *Unsupervised Adaptive Filtering*. Wiley, 265–319.
36. Santamaria, I., Pokharel, P. P., Principe, J. C. (2006). Generalized correlation function: definition, properties, and application to blind equalization. *IEEE Transactions on Signal Processing*, 54 (6), 2187–2197. doi: <https://doi.org/10.1109/tsp.2006.872524>
37. Liu, W., Pokharel, P. P., Principe, J. C. (2007). Correntropy: Properties and Applications in Non-Gaussian Signal Processing. *IEEE Transactions on Signal Processing*, 55 (11), 5286–5298. doi: <https://doi.org/10.1109/tsp.2007.896065>
38. Wang, W., Zhao, J., Qu, H., Chen, B., Principe, J. C. (2015). An adaptive kernel width update method of correntropy for channel estimation. 2015 IEEE International Conference on Digital Signal Processing (DSP). doi: <https://doi.org/10.1109/icdsp.2015.7252010>
39. Ma, W., Qu, H., Gui, G., Xu, L., Zhao, J., Chen, B. (2015). Maximum correntropy criterion based sparse adaptive filtering algorithms for robust channel estimation under non-Gaussian environments. *Journal of the Franklin Institute*, 352 (7), 2708–2727. doi: <https://doi.org/10.1016/j.franklin.2015.03.039>
40. Guo, Y., Ma, B., Li, Y. (2020). A Kernel-Width Adaption Diffusion Maximum Correntropy Algorithm. *IEEE Access*, 8, 33574–33587. doi: <https://doi.org/10.1109/access.2020.2972905>
41. Shi, L., Zhao, H., Zakharov, Y. (2020). An Improved Variable Kernel Width for Maximum Correntropy Criterion Algorithm. *IEEE Transactions on Circuits and Systems II: Express Briefs*, 67 (7), 1339–1343. doi: <https://doi.org/10.1109/tcsii.2018.2880564>
42. Huang, F., Zhang, J., Zhang, S. (2017). Adaptive Filtering Under a Variable Kernel Width Maximum Correntropy Criterion. *IEEE Transactions on Circuits and Systems II: Express Briefs*, 64 (10), 1247–1251. doi: <https://doi.org/10.1109/tcsii.2017.2671339>
43. Lu, L., Zhao, H. (2017). Active impulsive noise control using maximum correntropy with adaptive kernel size. *Mechanical Systems and Signal Processing*, 87, 180–191. doi: <https://doi.org/10.1016/j.ymssp.2016.10.020>
44. Rudenko, O. G., Bezsonov, O. O. (2020). ADALINE Robust Multistep Training Algorithm. *Control Systems and Computers*, 3 (287), 15–27. doi: <https://doi.org/10.15407/csc.2020.03.015>