Identification process in stationary and non-stationary control objects has been investigated in this study. The task addressed is to construct mathematical models, to devise methods and procedures, and to develop programs focused on solving object identification problems.

This paper tackles the issue of robust identification of control objects under the action of additive random noise of various statistical nature, in particular Gaussian and non-Gaussian noise. An approach to building identification algorithms based on the mixed correntropy criterion has been proposed, which combines the advantages of classical mean-square methods and information-theoretic optimality criteria.

The use of Price's theorem made it possible to define convergence conditions for the robust identification algorithm in both stationary and non-stationary cases in the presence of Gaussian and non-Gaussian noise. The influence of algorithm parameters and noise characteristics on its dynamic properties has been established. Expressions for determining the optimal values of the algorithm's convergence parameter, which ensure the maximum convergence rate, have been derived.

To confirm theoretical findings, simulation modeling was carried out, the results of which confirm the effectiveness of the proposed approach and its advantages compared to conventional identification methods, especially under conditions of non-Gaussian noise and nonstationarity, which indicates the feasibility of its use in adaptive and robust control systems.

However, the resulting estimates are rather general and depend both on the degree of nonstationarity of the object and on the statistical characteristics of usable signals and disturbances, which are often unknown. Therefore, the results could be applied in practice if such information is available or when estimates of these characteristics are used

Keywords: robust resilience; correntropy; kernel; algorithm convergence; steady state; simulation modeling

UDC 004.852
DOI: 10.15587/1729-4061.2026.353218

DETERMINING THE STATISTICAL PROPERTIES OF A ROBUST CONTROL OBJECT IDENTIFICATION ALGORITHM USING MIXED CORRENTROPY

Oleksandr Bezsonov
Corresponding author
Doctor of Technical Sciences*
E-mail: oleksandr.bezsonov@nure.ua
ORCID: https://orcid.org/0000-0001-6104-4275

Serhii Liashenko
Doctor of Technical Sciences
Department of Mechatronics, Life Safety and Quality Management
State Biotechnological University
Alchevskykh str., 44, Kharkiv, Ukraine, 61002
ORCID: https://orcid.org/0000-0001-8304-9309

Oleg Rudenko
Doctor of Technical Sciences*
ORCID: https://orcid.org/0000-0003-0859-2015

Serhii Rudenko
PhD Student
Department of Informatics**
ORCID: https://orcid.org/0009-0001-0912-2106

Kyrylo Oliinyk
Assistant*
ORCID: https://orcid.org/0000-0001-8536-5217

*Department of Computer Intelligent Technologies and Systems**
**Kharkiv National University of Radio Electronics
Nauky ave., 14, Kharkiv, Ukraine, 61166

Received 08.12.2025
Received in revised form 27.01.2026
Accepted date 12.02.2026
Published date 27.02.2026

***How to Cite:** Bezsonov, O., Liashenko, S., Rudenko, O., Rudenko, S., Oliinyk, K. (2026). Determining the statistical properties of a robust control object identification algorithm using mixed correntropy. Eastern-European Journal of Enterprise Technologies, 1 (4 (139)), 36–47. https://doi.org/10.15587/1729-4061.2026.353218*

1. Introduction

Many tasks related to control [1], forecasting [2], pattern recognition [3], etc. are associated with building a model of the following form

$$y(i) = W^{*\mathrm{T}}X(i) + \xi(i), \tag{1}$$

where $y(i)$ is the observed output signal; $X(i) = (x_1(i), x_2(i), ..., x_N(i))^{\mathrm{T}}$ is the vector of input signals $N\times 1$; $W^* = (w_1, w_2, ..., w_N)^{\mathrm{T}}$ is the vector of unknown parameters $N\times 1$; $\xi(i)$ is the output noise; i is the discrete time, and are reduced to the minimization of some previously selected quality functional (identification criterion)

$$F\left[e(i)\right] = \sum_{i=1}^{n} \rho\left(e(i)\right), \tag{2}$$

where $e(i) = y(i) - \hat{y}(i)$; $\hat{y}(i) = W^{\mathrm{T}}(i-1)x(i)$ is the output signal of the model; W is the vector estimate W^*; $\rho(e(i))$ is some differentiable loss function satisfying the conditions:

1) $\rho(e(i)) \geq 0$;
2) $\rho(0) = 0$;
3) $\rho(e(i)) = \rho(-e_i)$;
4) $\rho(e(i)) \geq \rho(e_j)$ for $|e(i)| \geq |e(j)|$.

The identification task is to find the estimate W defined as a solution to the extreme minimum problem

$$F(\theta) = \min, \tag{3}$$

or as a solution to a system of equations

$$\frac{\partial F(e)}{\partial \theta_j} = \sum_{i=1}^{n} \rho'(e_i) \frac{\partial e_i}{\partial \theta_j} = 0, \tag{4}$$

where $\rho'(e_i)\frac{\partial e_i}{\partial \theta_j}$ is the function of influence.

When choosing $\rho(e(i)) = 0,5e^2(i)$ the influence function is equal to $\rho'(e(i)) = e(i)$, that is, it increases linearly with increasing $e(i)$, which explains the instability of the least squares method (LSM) estimate to outliers and noise, the distributions of which have large tails. Such an estimate cannot be used in practice. In this regard, the problem of constructing estimates arises, which, unlike LSM, are stable in the presence of non-Gaussian noise. Among such estimates, the most popular is the M-estimate, which is defined as a solution to extreme problem (2) when choosing a loss function other than quadratic.

However, existing robust algorithms for constructing M-estimate mainly use second-order statistics and require for their implementation sufficiently complete a priori information about the properties of objects and noise. In addition, they are inconvenient for use in real time.

Therefore, research into the development of convenient robust identification algorithms that use higher-order statistics is relevant. Such algorithms could make it possible not only to expand the range of problems being solved but also to significantly improve the results in solving existing practical tasks.

2. Literature review and problem statement

The classical robust criteria proposed in [4, 5] are a combination of quadratic and modular functionals. As shown in these works, such a combination ensures optimality of estimates for the Gaussian distribution and robustness to distributions with heavy "tails" (outliers). It should be noted, however, that the efficiency of the robust estimates obtained in those papers significantly depends on the numerous parameters that are in these criteria. Although there are currently some recommendations for the choice of these parameters, they are mostly chosen taking into account the experience of the researcher.

The combined functionals proposed in [6, 7] expand the possibilities of robust identification, but require the specification of a mixing parameter. This approach was used in [8] to solve the identification problem, but this parameter was not specified in that work either.

In [9], the criterion of the smallest average kurtosis was introduced, and a fairly simple identification algorithm was obtained. As shown in the work, this algorithm is resistant to a wide range of noise (impulse, uniformly distributed and Gaussian).

To solve the tasks of machine learning and signal processing (control, forecasting, etc.), the use of the information approach is becoming increasingly widespread [10]. This is due to the fact that one of the basic operations of statistics, machine learning and signal processing is correlation, which makes it possible to quantitatively determine the similarity of random variables (processes) using their second-order statistics. On the one hand, it is very convenient to use, and on the other hand, its optimality is limited by the Gaussian data distribution.

During training, parameter tuning, etc., the most widely used criterion MSE also employs the second-order statistics of the error signal and is optimal under the assumptions of linearity and Gaussianity of the signals. If non-Gaussian, in particular, impulse noise is present in the measurements, a criterion that takes into account all higher-order statistics of the error signal will be more effective.

In [11], the concept of information-theoretic learning (ITL) was introduced, using the quadratic Renyi entropy as a criterion, for which a nonparametric estimate based on Parzen windows with Gaussian kernels is determined directly from the data samples. In [12], it was proven that when using Renyi entropy, the training results in minimizing the Renyi distance between the conditional probability density function of the desired and actual output signals for given input signals.

The results of numerous studies [13–15] indicate that in the presence of non-Gaussian, in particular, impulse noise, in measurements, the approach based on the information characteristics of signals is effective. Thus, in work [13], that was demonstrated on the problem of equalizer synthesis. In work [14] – on the problem of processing non-Gaussian signals, and in [15] – on the task of adaptive control. However, a criterion that takes into account all the statistics of the higher-order error signal turns out to be more appropriate.

Recently, the approach based on maximization of the correntropy criterion has gained the greatest development in the implementation of robust Kalman filters [16, 17]. In identification, filtering, etc. problems, the correntropy between the required output signal and the model output signal (real) is used as a functional. In [18], a multi-step identification algorithm based on maximization of the correntropy criterion was studied. However, when implementing this algorithm, the question arises of choosing the memory depth of the algorithm, the solution of which requires additional research.

Correntropy contains rotation-invariant Mercer kernels, among which the most widely used are Gaussian ones. However, the center of the Gaussian kernel in correntropy is always located at zero, which may not be the best choice for many situations with non-zero average noise. In this regard, the maximum correntropy criterion with a variable center was proposed to improve convergence in more common situations [19]. This approach was used in [20] to solve the identification problem. However, [19, 20] do not indicate how to choose the parameters of the algorithms.

In addition, it was proposed to use Laplace kernels [21] and Cauchy kernels [22] instead of Gaussian kernels.

A further development in this direction was the use of mixed criteria, the first of which was proposed in [23]. In [23–26], a mixed criterion based on Gaussian kernels was considered. In particular, in [23], the task of robust learning was considered. In that work, the concept of multi-kernel correntropy (MKC) was proposed, in which each component of a mixed Gaussian kernel can be concentrated at a different point. In [24], the task of time series forecasting was considered, and in [25], the problem of Kalman filter synthesis was tackled. In [26], the use of generalized maximum correntropy with a mixed norm for robust adaptive filtering was considered. In [27], the mixture correntropy criterion was proposed, where two Laplace kernel functions are combined as a kernel function. In [28], a criterion representing a double Cauchy mixture was studied. However, those papers have neglected the issue of choosing the optimal algorithm parameters.

Work [29] reports a study on convex and non-convex kernel functions from the point of view of robustness and stability. To improve the ability of stable filters to high levels of non-Gaussian observation noise, a mixed strategy of convex and non-convex stable functions is presented. The simulation results show that under different levels of severe non-Gaussian noise, the mixed strategy can avoid the local minimum by applying a single non-convex function and further improve the accuracy of filtering based on the convex function. Therefore, the mixed strategy can comprehensively improve the efficiency of the generalized M-estimator based on fixed-point iterations.

It should be noted that algorithms that maximize correntropy are easy to implement and efficient. However, the task of determining the statistical properties of such algorithms remains unsolved at present.

All this gives grounds to argue that it is advisable to conduct a study aimed at determining the statistical properties of gradient algorithms for identifying stationary and non-stationary objects in the presence of Gaussian and non-Gaussian noise.

3. The aim and objectives of the study

The aim of our work is to determine the statistical properties of gradient algorithms for identifying stationary and non-stationary parameters of control objects, using mixed correntropy. This will make it possible to obtain and adjust robust estimates as information about the studied process is received online.

To achieve this goal, the following tasks were set:

– to consider correntropy as a measure of similarity;

– to investigate the issue of convergence of the algorithm and the steady state of the estimation process in the stationary case in the presence of Gaussian and non-Gaussian noise;

– to investigate the issue of convergence of the algorithm and the steady state of the estimation process in the non-stationary case in the presence of Gaussian and non-Gaussian noise;

– to simulate the process of estimating the parameters of a linear object using robust algorithms.

4. Materials and methods

The object of our study is the process of identification of stationary and non-stationary control objects. As a hypothesis, it is assumed that the object under study is linear but can be stationary or non-stationary. Due to the fact that the presence of Gaussian or non-Gaussian noise in the measurements was assumed, an information approach based on the concept of mixed correntropy was applied to obtain robust estimates. In this case, it is assumed that mixed correntropy uses a mixture of two Gaussian functions as a kernel function. To simplify the study of the case of a non-stationary object, it is assumed that its parameters are described by a first-order Markov model.

To verify the effectiveness of the proposed algorithms, their robust properties, and also to study the speed of their convergence, a series of experiments were conducted in the MATLAB environment. To solve the identification problem, corresponding samples of input and output signals of stationary and non-stationary linear objects were generated. Independent noise distributed according to Rayleigh, Laplace, and Gaussian laws was added to the original signals of the object.

The generated data was divided into training, validation, and test samples. Before the identification process, data pre-processing was performed, which included normalization and removal of outliers.

5. Results of investigating statistical properties of the identification algorithm in the stationary case

5. 1. Correntropy as a measure of similarity

Correntropy is a local similarity function, which is defined as a generalized correlation in kernel space. It is closely related to the cross-information potential (CIP) in information theory of learning (ITL) [2]. It shows very promising results in nonlinear non-Gaussian signal processing.

Correntropy is used to represent the similarity between two random variables X and Y.

Let $k_\sigma(\cdot,\cdot)$ be a Mercer kernel function with kernel bandwidth σ. Then correntropy can be defined as

$$V(X,Y) = E\{k_\sigma(X,Y)\}. \tag{5}$$

Here $E\{\cdot\}$ is the mathematical expectation.

In general, the Gaussian kernel is the most widely used kernel in correntropy, and it looks like this

$$k_\sigma(X,Y) = G_\sigma(e) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{e^2}{2\sigma^2}\right), \tag{6}$$

where $e = X - Y$ is the error value.

Here, the nonlinear mapping $\varphi(\cdot)$ is used by the kernel function to map the input space U into the multidimensional space F, and it satisfies condition $\langle\varphi(x),\varphi(y)\rangle = k_\sigma(X,Y)$. Then (5) is rewritten as

$$V(X,Y) = E\{\langle\varphi(X),\varphi(Y)\rangle\}.$$

One of the key parameters in correntropy is the kernel bandwidth. Usually, a small kernel bandwidth makes the algorithm more robust to outliers, but it leads to slow convergence and low accuracy. On the other hand, when the kernel bandwidth increases, robustness will be significantly reduced in the case of anomalous values. To achieve better performance, a new similarity measure, mixture correntropy, has been proposed [23]. It can achieve fast convergence speed and higher filtering accuracy while maintaining robustness to outliers. Mixed correntropy uses a mixture of two Gaussian functions as the kernel function as follows

$$M(X,Y) = E\left[\alpha G_{\sigma_1}(e) + (1-\alpha)G_{\sigma_2}(e)\right],$$

where $0 \le \alpha \le 1$ is the mixture coefficient; σ_1 and σ_2 are the kernel bandwidths of the Gaussian functions $G_{\sigma_1}(\cdot)$ and $G_{\sigma_2}(\cdot)$ respectively. When the mixture coefficient α takes an appropriate value, the performance of MCC can be better than that of the original correntropy criterion, so the mixture correntropy is a more flexible similarity measure.

Generally, the empirical loss of the mixture correntropy can be expressed as $\hat{L}(X,Y)$ or $\hat{L}(e)$, where $X = [x_1, x_2, \ldots, x_N]^T$, $Y = [y_1, y_2, \ldots, y_N]^T$, та $e = [e_1, e_2, \ldots, e_N]^T$. It is defined as follows

$$\hat{L}(X,Y) = 1 - \hat{M}(X,Y) =$$
$$= 1 - \frac{1}{N}\sum_{i=1}^{N}\left[\alpha G_{\sigma_1}(e) + (1-\alpha)G_{\sigma_2}(e)\right],$$

where $e_i = x_i - y_i$.

Then (5) can be written as follows

$$M(X,Y)=\frac{\alpha}{N}\sum_{i=1}^{N}k_{\sigma_1}(x_i,y_i)+\frac{1-\alpha}{N}\sum_{i=1}^{N}k_{\sigma_2}(x_i,y_i)= =\frac{\alpha}{N}\sum_{i=1}^{N}\exp\left(\frac{e_i}{2\sigma_1^2}\right)+\frac{1-\alpha}{N}\sum_{i=1}^{N}\exp\left(\frac{e_i}{2\sigma_2^2}\right).$$

Using (6), we obtained

$$V_{\sigma,c}(X,Y)=\frac{1}{\sqrt{2\pi}\sigma}\sum_{n=0}^{\infty}\frac{(-1)^n}{2^n n!}E\left\{\left(\frac{(e-c)^{2n}}{\sigma^{2n}}\right)\right\}.$$

As σ increases, the higher-order moments with respect to the center will decrease faster, so the second-order moment will dominate in the value of $V_{\sigma,c}(X, Y)$. In particular, for $c=E\{e\}$ and $\sigma\to\infty$, maximizing the correntropy of the center is equivalent to minimizing the error variance.

5. 2. Studying the statistical properties of the identification algorithm in the stationary case

Hereinafter, the algorithm that maximizes the mixed correntropy criterion is denoted as MCC, and the algorithm that is an MCC algorithm with a variable center is denoted as MCC-VC. The weight vector is determined by maximizing

$$J_{MCC-VC}=E\{G_\sigma(e-c)\}= =E\left\{\begin{aligned}&\alpha\frac{1}{\sqrt{2\pi}\sigma_1}\exp\left(-\frac{(e-c)^2}{2\sigma_1^2}\right)+\\&+(1-\alpha)\frac{1}{\sqrt{2\pi}\sigma_2}\exp\left(-\frac{(e-c)^2}{2\sigma_2^2}\right)\end{aligned}\right\},$$

where $c\in R$ is the location of the center.

When $c=0$, the above cost function reduces to the original MCC algorithm. The MCC-VC algorithm, which is an adaptive algorithm based on stochastic gradient, takes the following form [23]

$$\begin{aligned}W(i)&=W(i-1)+\\&+\alpha\eta_1\exp\left(-\frac{(e-c)^2}{2\sigma_1^2}\right)(e(i)-c)X(i)+\\&+(1-\alpha)\eta_2\exp\left(-\frac{(e-c)^2}{2\sigma_2^2}\right)(e(i)-c)X(i),\end{aligned}\qquad(7)$$

where $W(i)$ denotes the vector of weights at the iteration; η_1, $\eta_2>0$ is the step size; $e(i)$ is the prediction error, which can be expressed as

$$e(i)=d(i)-W^T(i-1)X(i).\qquad(8)$$

If $\sigma\to\infty$ and $c=0$, the MCC-VC algorithm is reduced to the LMS algorithm

$$W(i)=W(i-1)+\eta e(i)X(i).\qquad(9)$$

According to Parzen's window theory, the following ratio is true

$$E\{G_\sigma(e(k)-c)\}\approx\frac{1}{L}\sum_{i=k-L}^{k}G(e(i)-c),\qquad(10)$$

where L is the sliding length of the data: $\{e(k)\}_{i=k-L}^{L}$ denotes L number of error samples (L must be large enough to ensure that the error curve is suitable for parameter estimation). Then for online parameter adaptation, the following holds:

$$(\sigma_1(k),\sigma_2(k),c(k))= =\underset{\sigma\in M,c\in C}{\arg\min}\left\{\begin{aligned}&\frac{1}{2\sqrt{\pi}\sigma}-\\&-E\left\{\begin{aligned}&\alpha\frac{1}{\sqrt{2\pi}\sigma_1}\exp\left(-\frac{(e-c)^2}{2\sigma_1^2}\right)+\\&+(1-\alpha)\frac{1}{\sqrt{2\pi}\sigma_2}\exp\left(-\frac{(e-c)^2}{2\sigma_2^2}\right)\end{aligned}\right\}\end{aligned}\right\},\qquad(11)$$

where M and C denote the admissible parameter sets, $\sigma_1(k)$, $\sigma_2(k)$, c and $c(k)$ denote the adapted parameters at iteration time k.

The MCC-VC algorithm can be rewritten as an adaptive filter with nonlinear error

$$W(i)=W(i-1)+\eta f(e(i))X(i),\qquad(12)$$

where $f(e(i))$ – scalar error function $e(i)$. For the MCC-VC algorithm, we obtain

$$\begin{aligned}f(i)&=\alpha\exp\left(-\frac{(e-c)^2}{2\sigma_1^2}\right)(e(i)-c)+\\&+(1-\alpha)\exp\left(-\frac{(e-c)^2}{2\sigma_2^2}\right)(e(i)-c).\end{aligned}\qquad(13)$$

Since the desired signal $y(i)$ is expressed by equation (1), we obtain

$$e(i)=\tilde{W}^T(i-1)X(i)+v(i)=e^a(i)+\xi(i),\qquad(14)$$

where $\tilde{W}^T(i-1)=W^*-W(i-1)$ is the estimation error vector at iteration $(i-1)$; $e^a(i)=\tilde{W}^T(i-1)X(i)$ is the prior error.

After writing the algorithm (12) with respect to errors $\tilde{W}$, multiplying both parts on the left by $\tilde{W}^T$ for the mathematical expectation we obtain

$$\begin{aligned}E\left\{\|\tilde{W}(i)\|^2\right\}&=E\left\{\|\tilde{W}(i-1)\|^2\right\}-\\&-2\eta E\{e^a(k)f(e(k))\}+\mu^2E\left\{\|X(i)\|^2f^2(e(k))\right\}.\end{aligned}\qquad(15)$$

To guarantee the convergence of the solution, it is necessary to gradually decrease the value of $E\left\{\|\tilde{W}(i)\|^2\right\}$.

Thus

$$\begin{aligned}&E\left\{\|\tilde{W}(i)\|^2\right\}\le E\left\{\|\tilde{W}(i-1)\|^2\right\}\Leftrightarrow-2\eta E\{e^a(k)f(e(k))\}+\\&+\mu^2E\left\{\|X(i)\|^2f^2(e(k))\right\}\le0\Leftrightarrow0<\eta\le\frac{2E\{e^a(k)f(e(k))\}}{E\left\{\|X(i)\|^2f^2(e(k))\right\}}.\end{aligned}\qquad(16)$$

As long as the MCC-VC algorithm step satisfies (16), the sequence $E\left\{\|\tilde{W}(i)\|^2\right\}$ decreases and converges, and the learning process is stable.

There is an exact measure of convergence efficiency called the excess mean square error (EMSE), which is the mean square prior error $M\left\{e^{2a}\left(k\right)\right\}$.

The following assumptions are adopted to study convergence:

1) the noise $\xi(i)$ is independent, identically distributed, and independent of the input signal $X(i)$;

2) the a priori error $e^a(i)$ is independent of the noise $\xi(i)$;

3) the filter is long enough so that $e^a(i)$ is Gaussian and $\left\|X\left(i\right)\right\|^2$ asymptotically uncorrelated with $f^2(e(i))$, i.e.

$$\lim_{i\to\infty} E\left\{\left\|X\left(i\right)\right\|^2 f^2\left(e\left(k\right)\right)\right\} = Tr\left(R_X\right)\lim_{i\to\infty} E\left\{f^2\left(e\left(k\right)\right)\right\}, \quad (17)$$

where $R_X = E\left\{X\left(i\right)X^T\left(i\right)\right\}$ is the covariance matrix of the input vector, and $Tr(\cdot)$ is the trace operator.

Assuming that the estimation process is stable and in a steady state, the following holds

$$\lim_{i\to\infty} E\left\{\left\|\tilde{W}\left(i\right)\right\|^2\right\} = \lim_{i\to\infty} E\left\{\left\|\tilde{W}\left(i-1\right)\right\|^2\right\}. \quad (18)$$

Then boundary (15) at $i \to \infty$ produces

$$2\lim_{i\to\infty} M\left\{e^a\left(i\right)f\left(e\left(i\right)\right)\right\} = \eta Tr \lim_{i\to\infty} M\left\{f^2\left(e\left(i\right)\right)\right\}. \quad (19)$$

For the case of Gaussian noise, under the above assumptions, we obtain the noise $\xi(i)$ with variance σ_ξ^2. According to the Gaussian assumptions and Price's theorem [30], the following holds:

$$\lim_{i\to\infty} E\left\{e^a\left(i\right)f\left(e\left(i\right)\right)\right\} = \lim_{i\to\infty} E\left\{e^a\left(i\right)f\left(e\left(i\right)+\xi\left(i\right)\right)\right\} =$$
$$= \lim_{i\to\infty} E\left\{e^{a2}\left(i\right)\right\}E\left\{f'\left(e\left(i\right)\right)\right\} =$$
$$= \lim_{i\to\infty} S\left[\begin{aligned}&E\left\{\exp\left(-\frac{\left(e\left(i\right)-c\right)^2}{2\sigma_1^2}\right)\left(1-\frac{\left(e\left(i\right)-c\right)^2}{\sigma_1^2}\right)\right\}+\\&+E\left\{\exp\left(-\frac{\left(e\left(i\right)-c\right)^2}{2\sigma_2^2}\right)\left(1-\frac{\left(e\left(i\right)-c\right)^2}{\sigma_2^2}\right)\right\}\end{aligned}\right] =$$
$$= \frac{S}{\sqrt{2\pi\sigma}}\lim_{i\to\infty}\int_{-\infty}^{\infty}\left[\begin{aligned}&E\left\{\exp\left(-\frac{\left(e\left(i\right)-c\right)^2}{2\sigma_1^2}\right)\left(1-\frac{\left(e\left(i\right)-c\right)^2}{\sigma_1^2}\right)\right\}+\\&+E\left\{\exp\left(-\frac{\left(e\left(i\right)-c\right)^2}{2\sigma_2^2}\right)\left(1-\frac{\left(e\left(i\right)-c\right)^2}{\sigma_2^2}\right)\right\}\end{aligned}\right]*$$
$$*\exp\left(-\frac{\left(e\left(i\right)-c_e\right)^2}{2\sigma_e^2}\right)de\left(i\right), \quad (20)$$

where $\sigma_e^2 = E\left\{e^{a2}\left(i\right)\right\}+\sigma_\xi^2$ denotes the variance of the error $e(i)$; $S = \lim_{i\to\infty} E\left\{e^{a2}\left(i\right)\right\}$ EMSE; c_e is the mean and central position of the Gaussian error e(i).

Similarly, we obtain

$$\lim_{i\to\infty} E\left\{f^2\left(e\left(i\right)\right)\right\} = \lim_{i\to\infty}\left[E\left\{\alpha\exp\left(-\frac{\left(e\left(i\right)-c\right)^2}{2\sigma_1^2}\right)\left(e\left(i\right)-c\right)^2\right\}+\left(1-\alpha\right)\exp\left(-\frac{\left(e\left(i\right)-c\right)^2}{2\sigma_2^2}\right)\left(e\left(i\right)-c\right)^2\right] =$$
$$= \frac{S}{\sqrt{2\pi\sigma_e}}\lim_{i\to\infty}\int_{-\infty}^{\infty}\left[\alpha\exp\left(-\frac{\left(e\left(i\right)-c\right)^2}{2\sigma_1^2}\right)\left(e\left(i\right)-c\right)^2+\left(1-\alpha\right)\exp\left(-\frac{\left(e\left(i\right)-c\right)^2}{2\sigma_2^2}\right)\left(e\left(i\right)-c\right)^2\right]\exp\left(-\frac{\left(e\left(i\right)-c_e\right)^2}{2\sigma_e^2}\right)\mathrm{d}e\left(i\right). \quad (21)$$

The stationary EMSE of the MCC-VC algorithm is actually a monotonically increasing function of the kernel width.

Compared to LMS, MCC-VC has adaptive step parameters:

$$\eta_1\left(i\right) = \eta\exp\left(\left(-\left(e\left(i\right)\right)-c\right)^2/2\sigma_1^2\right);$$
$$\eta_2\left(i\right) = \eta\exp\left(\left(-\left(e\left(i\right)\right)-c\right)^2/2\sigma_2^2\right).$$

Since $\eta_1\left(i\right)\le\eta,\ \eta_2\left(i\right)\le\eta$ MCC-VC can achieve a lower steady-state EMSE than LMS, but also a correspondingly slower convergence rate.

It is obvious that for the case of Gaussian noise, the EMSE value will be the smallest when $c = c_e$. This means that the noise center is found, and the MCC-VC algorithm reduces to the MCC algorithm. This extends the application of the adaptive filter to the case where the mean value of the noise is not zero.

In the case of non-Gaussian noise, a Taylor series expansion is used to obtain the steady-state EMSE. In the steady-state state, the iterative changes in the parameter values become insignificant, so we can rewrite (19) as

$$2E\left\{e^a f\left(e\right)\right\} = \eta TrE\left\{f^2\left(e\right)\right\}. \quad (22)$$

When taking the Taylor expansion $f(e)$ with respect to e^a around ξ, it is valid to have

$$f\left(e\right) = f\left(e^a+\xi\right) = f\left(\xi\right)+f'\left(\xi\right)e^a+$$
$$+\frac{1}{2}f''\left(\xi\right)e^a+o\left(e^{a2}\right), \quad (23)$$

where $o\left(e^{a2}\right)$ denotes the third and higher order terms, and according to (11) we obtain:

$$f'\left(v\right) = \exp\left(-\frac{\left(\xi-c\right)^2}{2\sigma_k^2}\right)\left(1-\frac{\left(\xi-c\right)^2}{2\sigma_k^2}\right); \quad (24)$$

$$f''\left(v\right) = \exp\left(-\frac{\left(\xi-c\right)^2}{2\sigma_k^2}\right)\left(\frac{\xi^3}{\sigma_k^4}-\frac{3\xi}{\sigma_k^2}\right),\quad k=1,2. \quad (25)$$

Based on assumptions 1 and 2, if $E\left\{o\left(e^{a2}\right)\right\}$ is sufficiently small, the following relations are obtained:

$$E\left\{e^a f\left(e\right)\right\} = E\left\{e^a f\left(\xi\right)\right\}+f'\left(\xi\right)e^{a2}+$$
$$+o\left(e^{a2}\right)\approx E\left\{f'\left(\xi\right)\right\}S; \quad (26)$$

$$E\left\{f^2\left(e\right)\right\}\approx E\left\{f^2\left(\xi\right)\right\}+E\left\{f\left(v\right)f''\left(\xi\right)+|f'\left(\xi\right)|^2\right\}. \quad (27)$$

Substitution (26) and (27) in (22), allowed us to obtain

$$S = \frac{\eta Tr\left(R_X\right)E\left\{f^2\left(\xi\right)\right\}}{2E\left\{f'\left(\xi\right)\right\}-\eta Tr\left(R_X\right)E\left\{f\left(\xi\right)f'\left(\xi\right)+\left|f'\left(\xi\right)\right|^2\right\}}. \quad (28)$$

Further, after substituting (24) and (25) in (28), the following ratio is obtained:

$$S=\frac{\eta A}{B-\eta C}, \qquad (29)$$

where:

$$A=Tr\left(R_X\right)E\left\{\left[\begin{array}{l}\alpha\exp\left(-\frac{(\xi-c)^2}{\sigma_1^2}\right)+\\+(1-\alpha)\exp\left(-\frac{(\xi-c)^2}{\sigma_2^2}\right)\end{array}\right](\xi-c)^2\right\};$$

$$B=2E\left\{\begin{array}{l}\alpha\exp\left(-\frac{(\xi-c)^2}{\sigma_1^2}\right)\left(1-\frac{(\xi-c)^2}{\sigma_1^2}\right)+\\+(1-\alpha)\exp\left(-\frac{(\xi-c)^2}{\sigma_2^2}\right)\left(1-\frac{(\xi-c)^2}{\sigma_2^2}\right)\end{array}\right\};$$

$$C=Tr\left(R_X\right)E\left\{\begin{array}{l}\alpha\exp\left(-\frac{(\xi-c)^2}{\sigma_1^2}\right)\times\\\times\left(1+\frac{2(\xi-c)^4}{\sigma_1^4}-\frac{5(\xi-c)^2}{\sigma_1^2}\right)+\\+(1-\alpha)\exp\left(-\frac{(\xi-c)^2}{\sigma_2^2}\right)\times\\\times\left(1+\frac{2(\xi-c)^4}{\sigma_2^4}-\frac{5(\xi-c)^2}{\sigma_2^2}\right)\end{array}\right\}.$$

From the above equation, it is clear that $S\to S_{LMS}$ as $\sigma\to\infty$.

Furthermore, when the η step size is small enough, (29) can be simplified to

$$S=\frac{\eta Tr\left(R_X\right)A}{2B},$$

where

$$A=E\left\{\left[\begin{array}{l}\alpha\exp\left(-\frac{(\xi-c)^2}{\sigma_1^2}\right)+\\+(1-\alpha)\left(-\frac{(\xi-c)^2}{\sigma_2^2}\right)\end{array}\right](\xi-c)^2\right\};$$

$$B=E\left\{\begin{array}{l}\alpha\exp\left(-\frac{(\xi-c)^2}{2\sigma_1^2}\right)\left(1-\frac{(\xi-c)^2}{\sigma_1^2}\right)+\\+(1-\alpha)\exp\left(-\frac{(\xi-c)^2}{2\sigma_2^2}\right)\left(1-\frac{(\xi-c)^2}{\sigma_2^2}\right)\end{array}\right\}.$$

The steady-state error (EMSE), which is defined by the above equation, can be accurate and efficient only when the a priori error e^a is small enough to make the term $E\left\{o\left(e^{a2}\right)\right\}$ negligible.

This means that if the step size or noise power is too large and the central position of the kernel function deviates from the noise mean, there will be a large discrepancy between the predicted and actual EMSE values.

5. 3. Results of investigating the statistical properties of the identification algorithm in the non-stationary case

Next, it is assumed that the parameters to be estimated are non-stationary, that is, described by a first-order Markov model

$$c^*(i)=c^*(i)+\Delta c^*, \qquad (30)$$

where $\Delta c^*=\left(\Delta c_1^*,\Delta c_2^*,...,\Delta c_N^*\right)^T$ is a vector of a random sequence $N\times 1$, the components of which have zero mathematical expectation, the correlation matrix of which is equal to $R_c=E\left\{c^*c^{*T}\right\}$.

For the error vector $\tilde{W}(i)=c(i)-c^*(i)$, the following is true

$$\begin{aligned}&\tilde{W}(i)=\tilde{W}(i-1)-\\&-c^*(i)+\eta f\left(e(i)\right)x(i)=\\&=\tilde{W}(i-1)-\Delta c^*+\\&+\eta f\left(e(i)\right)x(i).\end{aligned} \qquad (31)$$

Multiplying both parts of (31) on the left by $\tilde{W}^T(i)$ and calculating the mathematical expectation allowed us to obtain

$$\begin{aligned}&E\left\{\left\|\tilde{W}(i)\right\|^2\right\}=\\&=E\left\{\left\|\tilde{W}(i-1)\right\|^2\right\}-\\&-2\eta E\left\{x^T(i)\tilde{W}(i-1)f\left(e(i)\right)\right\}+\\&+\eta^2E\left\{f^2\left(e(i)\right)\left\|x(i)\right\|^2\right\}+\\&+E\left\{\left\|\Delta c^*\right\|^2\right\}+\\&+E\left\{x^T(i)\Delta c^*\right\}+\\&+E\left\{\Delta c^{*T}x(i)\right\}-\\&-2\eta E\left\{x^T(i)\Delta c^*f\left(e(i)\right)\right\}.\end{aligned} \qquad (32)$$

Taking into account the statistical properties of signals and noise made it possible to obtain

$$\begin{aligned}&E\left\{\left\|\tilde{W}(i)\right\|^2\right\}=\\&=E\left\{\left\|\tilde{W}(i-1)\right\|^2\right\}-\\&-2\eta E\left\{e^a(i)f\left(e(i)\right)\right\}+\\&+\eta^2E\left\{f^2\left(e(i)\right)\left\|x(i)\right\|^2\right\}+\\&+E\left\{\left\|\Delta c^*\right\|^2\right\}.\end{aligned}$$

Thus, the magnitude of the error depends not only on the properties of signals and interference, but also on the non-stationarity of the object.

For the Gaussian interference, the use of Price's theorem [30] made it possible to obtain:

$$\lim_{i\to\infty} E\left\{e^a(i) f\left(e(i)\right)\right\} = \lim_{i\to\infty} E\left\{e^a(i) f\left(e^a(i)+\xi_{n+1}\right)\right\} = \lim_{i\to\infty} E\left\{\left(e^a(i)\right)^2\right\} E\left\{f'\left(e(i)\right)\right\} =$$

$$= \lim_{i\to\infty} SE\left\{\alpha \exp\left(-\frac{\left(e(i)-c\right)^2}{2\sigma_1^2}\right)\left(1-\frac{\left(e(i)-c\right)^2}{\sigma_1^2}\right)+(1-\alpha)\exp\left(-\frac{\left(e(i)-c\right)^2}{2\sigma_2^2}\right)\left(1-\frac{\left(e(i)-c\right)^2}{\sigma_2^2}\right)\right\}=$$

$$=\frac{S}{\sqrt{2\pi}\sigma_e}\lim_{i\to\infty}\int_{-\infty}^{+\infty}\left[\left\{\alpha \exp\left(-\frac{\left(e(i)-c\right)^2}{2\sigma_1^2}\right)\left(1-\frac{\left(e(i)-c\right)^2}{\sigma_1^2}\right)+(1-\alpha)\exp\left(-\frac{\left(e(i)-c\right)^2}{2\sigma_2^2}\right)\left(1-\frac{\left(e(i)-c\right)^2}{\sigma_2^2}\right)\right\}\right]\exp\left(-\frac{\left(e(i)-c_e\right)^2}{2\sigma_e^2}\right)\mathrm{d}e(i)=$$

$$=\alpha\frac{S\sigma_1^3}{\left(\sigma_1^2+\sigma_\xi^2+S\right)^{\frac{3}{2}}}+(1-\alpha)\frac{S\sigma_2^3}{\left(\sigma_2^2+\sigma_\xi^2+S\right)^{\frac{3}{2}}};$$

$$E\left\{f^2\left(e(i)\right)\right\}=\lim_{i\to\infty} E\left\{\alpha \exp\left(-\frac{\left(e(i)-c\right)^2}{2\sigma_1^2}\right)\left(1-\frac{\left(e(i)-c\right)^2}{\sigma_1^2}\right)+(1-\alpha)\exp\left(-\frac{\left(e(i)-c\right)^2}{2\sigma_2^2}\right)\left(1-\frac{\left(e(i)-c\right)^2}{\sigma_2^2}\right)\right\}=$$

$$=\frac{1}{\sqrt{2\pi}\sigma_e}\lim_{i\to\infty}\int_{-\infty}^{+\infty}\left[\alpha \exp\left(-\frac{\left(e(i)-c\right)^2}{2\sigma^2}\right)+(1-\alpha)\exp\left(-\frac{\left(e(i)-c\right)^2}{2\sigma^2}\right)\right]\left(e(i)-c\right)^2\times\exp\left(-\frac{\left(e(i)-c_e\right)^2}{2\sigma_e^2}\right)\mathrm{d}e(i)=\alpha\frac{\sigma_1^2\left(S+\sigma_\xi^2\right)}{\left(2\sigma_\xi^2+\sigma_1^2+2S\right)^{\frac{3}{2}}}+$$

$$+(1-\alpha)\frac{\sigma_2^2\left(S+\sigma_\xi^2\right)}{\left(2\sigma_\xi^2+\sigma_2^2+2S\right)^{\frac{3}{2}}}.$$

For the steady state, when

$$\lim_{i\to\infty} E\left\{\left\|\tilde{W}(i)\right\|^2\right\}=\lim_{i\to\infty} E\left\{\left\|\tilde{W}(i-1)\right\|^2\right\},$$

taking into account the fact that

$$E\left\{\left\|\Delta c^*\right\|^2\right\}=E\left\{\Delta c^*\Delta c^{*T}\right\}=trR_c,$$

the following relation is obtained from expression (32)

$$\frac{2\alpha S}{\left(\sigma_1^2+\sigma_\xi^2+S\right)^{\frac{3}{2}}}+\frac{2(1-\alpha)S}{\left(\sigma_2^2+\sigma_\xi^2+S\right)^{\frac{3}{2}}}=$$

$$=\frac{\gamma trR_x\left(\sigma_\xi^2+S\right)}{\left(\sigma_1^2+2\sigma_\xi^2+2S\right)^{\frac{3}{2}}}+\frac{trR_c}{\eta\sigma_1^3}+\frac{\gamma trR_x\left(\sigma_\xi^2+S\right)}{\left(\sigma_2^2+2\sigma_\xi^2+2S\right)^{\frac{3}{2}}}+\frac{trR_c}{\eta\sigma_2^3}.$$

This ratio made it possible to determine the following value of S

$$S=\frac{\eta trR_x\left(\sigma_\xi^2+S\right)\left[\left(\sigma_2^2+\sigma_\xi^2+S\right)^{\frac{3}{2}}+\left(\sigma_1^2+\sigma_\xi^2+2S\right)^{\frac{3}{2}}\right]}{\alpha\left(\sigma_2^2+2\sigma_\xi^2+S\right)^{\frac{3}{2}}+(1-\alpha)\left(\sigma_1^2+2\sigma_\xi^2+S\right)^{\frac{3}{2}}}+$$

$$+\frac{trR_c\left(\sigma_1^3+\sigma_2^3\right)\left(\sigma_1^2+\sigma_\xi^2+S\right)^{\frac{3}{2}}\left(\sigma_2^2+\sigma_\xi^2+2S\right)^{\frac{3}{2}}}{2\eta\left(\sigma_1^3\sigma_2^3\right)}. \quad (33)$$

Thus, the magnitude of S significantly depends on the non-stationarity of the object.

Consider the case of non-Gaussian noise.

In the case of non-Gaussian noise, the following is true:

$$E\left\{e^a(i) f\left(e(i)\right)\right\}\approx E\left\{\begin{array}{l}e^a(i) f\left(\xi(i)\right)+\\+e^a(i) f'\left(\xi(i)\right)\end{array}\right\}\approx SE\left\{f'\left(\xi(i)\right)\right\}, \quad (34)$$

$$E\left\{f^2\left(e(i)\right)\right\}\approx E\left\{\left(\begin{array}{l}f\left(\xi(i)\right)+e^a(i) f'\left(\xi(i)\right)+\\+0,5 f''\left(\xi(i)\right)e^{a2}(i)\end{array}\right)^2\right\}\approx$$

$$\approx E\left\{f^2\left(\xi(i)\right)\right\}+$$

$$+SE\left\{\left(f\left(\xi(i)\right)f''\left(\xi(i)\right)+\left(f'\left(\xi(i)\right)\right)^2\right)\right\}, \quad (35)$$

where:

$$E\left\{e^a(i) f\left(e(i)\right)\right\}\approx E\left\{\begin{array}{l}e^a(i) f\left(\xi(i)\right)+\\+e^a(i) f'\left(\xi(i)\right)\end{array}\right\}\approx SE\left\{f'\left(\xi(i)\right)\right\}. \quad (36)$$

$$f'\left(\xi(i)\right)=\exp\left(-\frac{\left(\xi-c\right)^2}{2\sigma_k^2}\right)\left(1-\frac{\left(\xi-c\right)^2}{\sigma_k^2}\right),\quad k=1,2; \quad (37)$$

$$f''\left(\xi(i)\right)=\exp\left(-\frac{\left(\xi(i)-c\right)^2}{2\sigma_k^2}\right)\left(\frac{\xi^3(i)}{\sigma^4(i)}-\frac{3\xi(i)}{\sigma_k^2}\right),\ k=1,2. \quad (38)$$

Substituting (37) and (38) into (32), after simple transformations, the following expression for S is obtained:

$$S=\frac{\eta A+\eta^{-1}B}{C-\eta D}, \quad (39)$$

where

$$A=trR_x E\left\{\left(\xi(i)-c\right)^2\left[\begin{array}{l}\alpha\exp\left(-\frac{\left(\xi(i)-c\right)^2}{\sigma_1^2}\right)+\\+(1-\alpha)\exp\left(-\frac{\left(\xi(i)-c\right)^2}{\sigma_2^2}\right)\end{array}\right]\right\};$$

$$B=trR_c;$$

$$C=2E\left\{\begin{aligned}&\alpha\left(1-\frac{\left(\xi(i)-c\right)^2}{2\sigma_1^2}\right)\exp\left(-\frac{\left(\xi(i)-c\right)^2}{\sigma_1^2}\right)+\\&+(1-\alpha)\left(1-\frac{\left(\xi(i)-c\right)^2}{2\sigma_2^2}\right)\exp\left(-\frac{\left(\xi(i)-c\right)^2}{\sigma_2^2}\right)\end{aligned}\right\};$$

$$D=trR_xE\{F\},$$

$$F=\left[\begin{aligned}&\alpha\left(1+\frac{2\left(\xi(i)-c\right)^4}{\sigma_1^4}-\frac{5\left(\xi(i)-c\right)^2}{\sigma_1^2}\right)\times\\&\times\exp\left(-\frac{\left(\xi(i)-c\right)^2}{\sigma_1^2}\right)+\\&+(1-\alpha)\left(1+\frac{2\left(\xi(i)-c\right)^4}{\sigma_2^4}-\frac{5\left(\xi(i)-c\right)^2}{\sigma_2^2}\right)\times\\&\times\exp\left(-\frac{\left(\xi(i)-c\right)^2}{\sigma^2}\right)\end{aligned}\right],$$

i.e., S is a monotonically non-increasing function of parameter η.

From condition $\partial S/\partial\eta=0$, an equation is obtained to determine the optimal value of the parameter , which provides the minimum value of S.

As can be seen from the results given in this section, the stationary value of the estimation error depends not only on the type of noise distribution but also on the degree of non-stationarity (Δc^*) of the estimated parameters

$$AC\eta^2+BD\eta-BC=0. \qquad (40)$$

Solving equation (40), we get the value of the desired parameter η.

5. 4. Results of modeling the process of estimating model parameters in the presence of non-Gaussian interference

Experiment 1. The task of identifying a stationary linear object was considered, which is described by equation (1) with the following parameters

$$W^*=\begin{pmatrix}-128;-96;-85;-64;-57;-48;-31;\\-21;0;2;20;32;62;97;108;127\end{pmatrix}^T.$$

As an input signal, a sequence of normally distributed quantities x(i) ~ $N(0,1)$ was chosen. When checking the robustness of the algorithms, independent noise distributed according to the Rayleigh law with $\sigma=1$, Laplace with $\sigma=3$, and Gaussian with $\sigma=3$, was added to the output signal of the object. A comparative analysis was performed to compare the robust properties of the LSM algorithm and algorithm (31) with different c, σ_1, σ_2, α parameters. To generate a mixed Gaussian correntropy, and where σ_1 and σ_2 represent specific kernel parameters, and $\sigma_1>\sigma_2$, $G_{\sigma_1}(e)$ and $G_{\sigma_2}(e)$, were mixed. The simulation results are shown in Fig. 1, which demonstrates the change in the value

$$RMSE=\sqrt{\frac{1}{2}\left\|W(i)-W^*\right\|^2},$$

where $W(i)$ and W^* denote the vectors of estimated and target parameters, respectively.

In Fig. 1, *a–f*, the curve with triangles corresponds to the LSM algorithm, the curve with crosses to algorithm (31) with $c=0.5$, $\sigma_1=5$, $\sigma_2=2$, the curve with squares to algorithm (31) with $c=1$, $\sigma_1=7$, $\sigma_2=3$, the curve with circles to algorithm (31) with $c=1.5$, $\sigma_1=7$, $\sigma_2=3$.

As can be seen from Fig. 1, *a–f*, the LMS algorithm has the highest convergence rate for the case of interference with a Gaussian distribution. If the interference has a different distribution, the LMS algorithm, unlike algorithm (31), does not converge.

Experiment 2. The task of identifying a non-stationary linear object described by equation (1), the parameters of which vary according to (30), was considered. As in Experiment 1, independent noise with the same distributions was added to the output signal of the object.

The results of the experiment are given in Tables 1, 2, where the value of the normalized error after 100 iterations of each simulation algorithm is presented.

Table 1

Results of experiment 2 at $\alpha=0.5$

Algorithm/Noise	Rayleigh, $\sigma=1$	Laplace, $\sigma=3$	Normal, $\sigma=3$
LMS	0.36	0.29	0.21
(31) $c=0$, $\sigma_1=5$, $\sigma_2=2$	0.12	0.08	0.054
(31) $c=0$, $\sigma_1=7$, $\sigma_2=3$	0.10	0.067	0.043
(31) $c=0$, $\sigma_1=9$, $\sigma_2=5$	0.09	0.059	0.032
(31) $c=0.5$, $\sigma_1=5$, $\sigma_2=2$	0.08	0.05	0.025
(31) $c=0.5$, $\sigma_1=7$, $\sigma_2=3$	0.075	0.034	0.018
(31) $c=0.5$, $\sigma_1=9$, $\sigma_2=5$	0.068	0.025	0.009
(31) $c=1$, $\sigma_1=5$, $\sigma_2=2$	0.015	0.009	0.006
(31) $c=1$, $\sigma_1=7$, $\sigma_2=3$	0.011	0.008	0.0057
(31) $c=1$, $\sigma_1=9$, $\sigma_2=5$	0.010	0.0076	0.0041
(31) $c=1.5$, $\sigma_1=5$, $\sigma_2=2$	0.012	0.007	0.005
(31) $c=1.5$, $\sigma_1=7$, $\sigma_2=3$	0.010	0.0063	0.0032
(31) $c=1.5$, $\sigma_1=9$, $\sigma_2=5$	0.007	0.0054	0.0026

Table 2

Results of experiment 2 at $\alpha=0.8$

Algorithm/Noise	Rayleigh, $\sigma=1$	Laplace, $\sigma=3$	Normal, $\sigma=3$
LMS	0.36	0.29	0.21
(31) $c=0$, $\sigma_1=5$, $\sigma_2=2$	0.11	0.072	0.044
(31) $c=0$, $\sigma_1=7$, $\sigma_2=3$	0.09	0.061	0.037
(31) $c=0$, $\sigma_1=9$, $\sigma_2=5$	0.08	0.052	0.026
(31) $c=0.5$, $\sigma_1=5$, $\sigma_2=2$	0.075	0.042	0.022
(31) $c=0.5$, $\sigma_1=7$, $\sigma_2=3$	0.069	0.029	0.012
(31) $c=0.5$, $\sigma_1=9$, $\sigma_2=5$	0.054	0.018	0.007
(31) $c=1$, $\sigma_1=5$, $\sigma_2=2$	0.010	0.007	0.0052
(31) $c=1$, $\sigma_1=7$, $\sigma_2=3$	0.0092	0.0054	0.0046
(31) $c=1$, $\sigma_1=9$, $\sigma_2=5$	0.00081	0.0048	0.0038
(31) $c=1.5$, $\sigma_1=5$, $\sigma_2=2$	0.0073	0.006	0.0041
(31) $c=1.5$, $\sigma_1=7$, $\sigma_2=3$	0.0067	0.0057	0.0026
(31) $c=1.5$, $\sigma_1=9$, $\sigma_2=5$	0.0056	0.0034	0.0017

As can be seen from Tables 1, 2, algorithm (31) with different c, σ_1, σ_2, α parameters works well for all considered noise types. In a certain range, the estimation accuracy improved with increasing σ_1 since this provided better compatibility with Gaussian noise for the estimation algorithm.

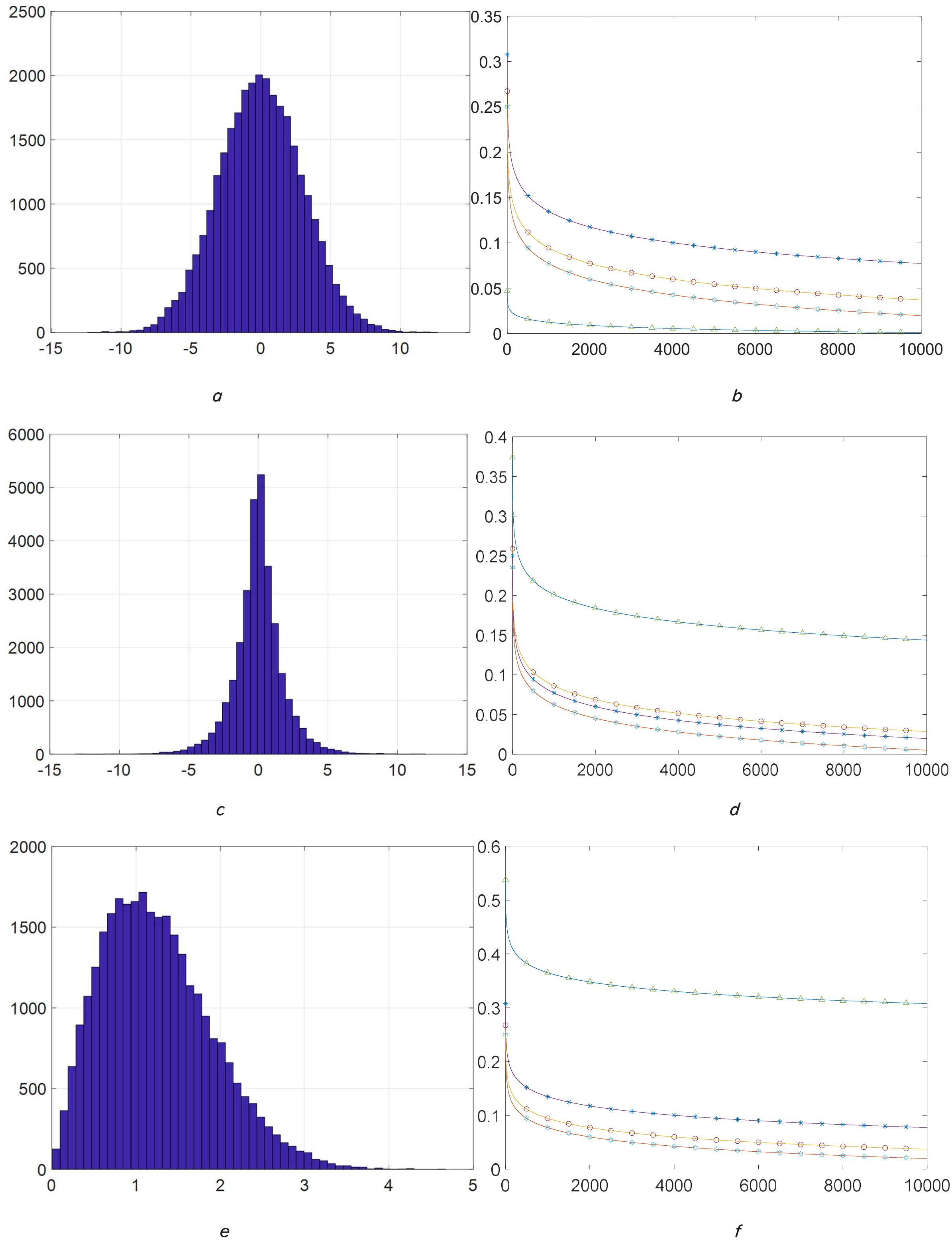

Fig. 1. Histograms of noise (left) and training error (right) for the distributions: *a, b* – Gaussian; *c, d* – Laplace; *e, f* – Rayleigh

6. Discussion of results based on investigating the properties of the algorithm for robust identification of control objects using mixed correntropy

The research reported in this work is a continuation and development of previous studies described in [8, 18, 20]. The results from those works were applied in this paper to determine the properties of the identification algorithm, which maximizes, in contrast to the criteria used in those papers, the mixed correntropy criterion. The results from the identification of stationary and non-stationary objects are more general than the existing ones. Thus, the identification algorithm, which takes the form of (7), consists of two parts, which correspond to two different Gaussian kernels.

Unlike the algorithms considered in [25, 27], which require the calculation of the inverse covariance matrix of observations, which has, as a rule, a large dimensionality, the algorithm (12), (13) uses information only about the last observation. Thus, the implementation of the given algorithm does not cause difficulties.

The use of Lyapunov functions allowed us to determine the convergence condition for algorithm (7), which is determined from expression (16). To determine an exact measure of convergence efficiency, called the excess mean square error (EMSE), which is the mean square prior error $M\{e^{2a}(k)\}$, it is assumed that:

1) the noise $\xi(i)$ is independent, identically distributed, and independent of the input signal $X(i)$;

2) the a priori error $e^a(i)$ is independent of the noise $\xi(i)$;

3) the filter is long enough so that $e^a(i)$ is Gaussian and $||X(i)||^2$ is asymptotically uncorrelated with $f^2(e(i))$.

These assumptions are a significant simplification, which provided the advantages of this study, which are to obtain clear and transparent results.

The resulting expression (21) indicates that in the case of Gaussian noise, the investigated algorithm for identifying stationary objects can achieve a lower stationary EMSE than LMS, but also a correspondingly slower convergence rate. In the case of non-Gaussian noise, Taylor series expansion was used to obtain the stationary EMSE in the stationary case. It should be noted that the obtained expression (29) for the stationary error (EMSE) can be accurate and effective only when the a priori error e^a is small enough. This means that if the step size or noise power is too large, and the central position of the kernel function deviates from the noise mean, there will be a large discrepancy between the predicted and actual EMSE values.

In the study of non-stationary objects described by the first-order Markov model (30), expressions for the value S for Gaussian (33) and non-Gaussian (39) disturbances were derived, which significantly depends on the non-stationarity of the object. These expressions show that S is a monotonically non-increasing function of parameter η. In our work, equation (40) was derived, solving which allows finding the optimal value of the desired parameter η that provides the minimum value of S.

The simulation results confirm the effectiveness of using the mixed correntropy functional for identifying stationary and non-stationary objects.

Our results, represented by expressions (16), (21), (29), (33), (39), allow the researcher to preliminarily assess the capabilities of a particular algorithm and the effectiveness of its application when solving practical tasks. In addition, these results can be an inspiration for improving and optimizing robust identification algorithms.

The significance of our results relates to the fact that they make it possible to assess the capabilities of a robust algorithm when solving practical problems under conditions of a priori and current uncertainty regarding the properties of objects, signals and interference.

The results of our study, shown in Fig. 1 and given in Tables 1, 2, indicate the possibility of improving existing data processing technologies based on robust estimates. In addition, it is possible to accelerate the development of new applications for real-time calculations in continuous technological process control systems, time series forecasting, etc. Their implementation does not impose special requirements on computing equipment.

The limitation of the study is that the obtained estimates are quite general and depend both on the degree of non-stationarity of the object and on the statistical characteristics of the usable signals and interference, which are often unknown.

Due to the fact that the issue of the optimal choice of parameter values remains open, it seems important and advisable to conduct research into the following:

1) determining the effectiveness of the devised approach in learning in the non-stationary case, when a model other than the first-order Markov model is used to describe non-stationarity;

2) establishing the dependence of the speed of the learning algorithm on the degree of non-stationarity of the studied object, i.e., Δc^*;

3) compiling recommendations for choosing optimal parameter values and/or rules for their correction;

4) defining the effectiveness of the devised approach for identifying nonlinear objects.

7. Conclusions

1. The issues of algorithm convergence and steady state of the estimation process in the stationary case in the presence of Gaussian and non-Gaussian noise have been investigated. Our results indicate that in the case of Gaussian noise, the investigated algorithm for stationary object identification can achieve a lower stationary EMSE than LMS. However, the convergence rate becomes slower. In the case of non-Gaussian noise, the obtained expression for the EMSE (S) value can be accurate and effective only when the a priori error e^a is sufficiently small. This means that if the step size or noise power is too large and the central position of the kernel function deviates from the noise mean, there will be a large discrepancy between the predicted and actual EMSE values.

2. The issue of algorithm convergence and steady state of the estimation process in the non-stationary case in the presence of Gaussian and non-Gaussian noise is investigated. In this case, non-stationarity is considered in the form of a first-order Markov model. Expressions for the value of S for Gaussian and non-Gaussian noise were derived.

3. A second-order equation has been built, solving which allows finding the optimal value of the desired algorithm parameter η, which ensures its maximum convergence speed when identifying non-stationary objects and measurement noise. This is explained by the fact that this equation is obtained under the conditions of minimizing the stationary EMSE.

4. The process of estimating stationary and non-stationary parameters of a linear object was simulated. A sequence of normally distributed quantities $x(i) \sim N(0,1)$ was selected as the input signal. When checking the robustness of the algorithms, independent noise distributed according to the Rayleigh law with $\sigma = 1$, Laplace with $\sigma = 3$, and Gaussian with $\sigma = 3$, was added to the output signal of the object. A comparative analysis was performed to compare the robust properties of the LSM algorithm and the proposed algorithm with different c, σ_1, σ_2, α parameters.

The simulation results lay the foundation for selecting algorithm parameters during its implementation. Based on the analysis of the simulation results, the following conclusions can be drawn.

First, the use of the maximum correntropy functional is quite effective for identifying linear stationary and non-stationary objects.

As can be seen from the simulation results, the LMS algorithm has the highest convergence rate for the case of interference with a Gaussian distribution.

If the interference has a different distribution, the LMS algorithm, unlike the proposed algorithm, does not converge, which is explained by the fact that the influence function of the algorithm increases linearly with increasing $e(i)$.

Second, the simulation results show that the proposed algorithm with different c, σ_1, σ_2, α parameters works well for all considered noise types. In a certain range, the estimation accuracy improved with increasing σ_1. This is due to the better compatibility with Gaussian noises for the estimation algorithm.

Conflicts of interest

The authors declare that they have no conflicts of interest in relation to the current study, including financial, personal, authorship, or any other, that could affect the study, as well as the results reported in this paper.

Funding

Results in this study were obtained under the international research project INITIATE under grant No. 101136775-HORIZON-WIDERA-2023-ACCESS-03.

Data availability

All data are available, either in numerical or graphical form, in the main text of the manuscript.

Use of artificial intelligence

The authors confirm that they did not use artificial intelligence technologies when creating the current work.

Author contributions

Oleksandr Bezsonov: conceptualization, methodology; **Serhii Liashenko:** formal analysis; **Oleg Rudenko:** methodology, writing – original draft; **Serhii Rudenko:** software; **Kyrylo Oliinyk:** visualization.

References

1. Frank, S. A. (2018). Control Theory Tutorial. In SpringerBriefs in Applied Sciences and Technology. Springer International Publishing. https://doi.org/10.1007/978-3-319-91707-8
2. Petropoulos, F., Apiletti, D., Assimakopoulos, V., Babai, M. Z., Barrow, D. K., Ben Taieb, S. et al. (2022). Forecasting: theory and practice. International Journal of Forecasting, 38 (3), 705–871. https://doi.org/10.1016/j.ijforecast.2021.11.001
3. Nasrabad, N. M. (2006). Pattern Recognition and Machine Learning. Journal of Electronic Imaging, 16 (4). https://doi.org/10.1117/1.2819119
4. Huber, P. J., Ronchetti, E. M. (2009). Robust Statistics. Wiley Series in Probability and Statistics. https://doi.org/10.1002/9780470434697
5. Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., Stahel, W. A. (2005). Robust Statistics. Wiley Series in Probability and Statistics. https://doi.org/10.1002/9781118186435
6. Chambers, J. A., Tanrikulu, O., Constantinides, A. G. (1994). Least mean mixed-norm adaptive filtering. Electronics Letters, 30 (19), 1574–1575. https://doi.org/10.1049/el:19941060
7. Arenas-García, J., Figueiras-Vidal, A. R. (2005). Adaptive combination of normalised filters for robust system identification. Electronics Letters, 41 (15), 874–875. https://doi.org/10.1049/el:20051936
8. 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. https://doi.org/10.15587/1729-4061.2019.181256
9. 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. https://doi.org/10.1109/78.747793
10. Principe, J. C. (2010). Information Theoretic Learning. In Information Science and Statistics. Springer New York. https://doi.org/10.1007/978-1-4419-1570-2
11. 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 (1-2), 61–77. https://doi.org/10.1023/a:1008143417156
12. Chen, B., Zhu, Y., Hu, J., Principe, J. C. (2013). System Parameter Identification: Information Criteria and Algorithms. Elsevier Inc., 249. https://doi.org/10.1016/c2012-0-01233-1
13. 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. https://doi.org/10.1109/tsp.2006.872524
14. 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. https://doi.org/10.1109/tsp.2007.896065
15. Flores, T. K. S., Villanueva, J. M. M., Gomes, H. P., Catunda, S. Y. C. (2021). Adaptive Pressure Control System Based on the Maximum Correntropy Criterion. Sensors, 21 (15), 5156. https://doi.org/10.3390/s21155156
16. Leite, G. R., Araújo, Í. B. Q. de, Martins, A. de M. (2023). Regularized Maximum Correntropy Criterion Kalman Filter for Uncalibrated Visual Servoing in the Presence of Non-Gaussian Feature Tracking Noise. Sensors, 23 (20), 8518. https://doi.org/10.3390/s23208518

17. Wu, C., Lin, D., Zheng, Y., He, F., Wang, S. (2026). Maximum correntropy criterion-based Kalman filter for replay attack in non-Gaussian noises. Signal Processing, 238, 110098. https://doi.org/10.1016/j.sigpro.2025.110098
18. Rudenko, O., Bezsonov, O., Borysenko, V., Borysenko, T., Lyashenko, S. (2021). Developing a multi-step recurrent algorithm to maximize the criteria of correntropy. Eastern-European Journal of Enterprise Technologies, 1 (4 (109)), 54–63. https://doi.org/10.15587/1729-4061.2021.225765
19. Chen, B., Wang, X., Li, Y., Principe, J. C. (2019). Maximum Correntropy Criterion With Variable Center. IEEE Signal Processing Letters, 26 (8), 1212–1216. https://doi.org/10.1109/lsp.2019.2925692
20. Rudenko, O., Bezsonov, O. (2022). Adaptive identification under the maximum correntropy criterion with variable center. RADIO-ELECTRONIC AND COMPUTER SYSTEMS, 1, 216–228. https://doi.org/10.32620/reks.2022.1.17
21. Hu, C., Wang, G., Ho, K. C., Liang, J. (2021). Robust Ellipse Fitting With Laplacian Kernel Based Maximum Correntropy Criterion. IEEE Transactions on Image Processing, 30, 3127–3141. https://doi.org/10.1109/tip.2021.3058785
22. Ye, X., Lu, S., Wang, J., Wu, D., Zhang, Y. (2023). Robust State Estimation Using the Maximum Correntropy Cubature Kalman Filter with Adaptive Cauchy-Kernel Size. Electronics, 13 (1), 114. https://doi.org/10.3390/electronics13010114
23. Chen, B., Xie, Y., Wang, X., Yuan, Z., Ren, P., Qin, J. (2022). Multikernel Correntropy for Robust Learning. IEEE Transactions on Cybernetics, 52 (12), 13500–13511. https://doi.org/10.1109/tcyb.2021.3110732
24. Xue, N., Luo, X., Gao, Y., Wang, W., Wang, L., Huang, C., Zhao, W. (2019). Kernel Mixture Correntropy Conjugate Gradient Algorithm for Time Series Prediction. Entropy, 21 (8), 785. https://doi.org/10.3390/e21080785
25. Li, X., Guo, Y., Meng, Q. (2022). Variational Bayesian-Based Improved Maximum Mixture Correntropy Kalman Filter for Non-Gaussian Noise. Entropy, 24 (1), 117. https://doi.org/10.3390/e24010117
26. Li, G., Zhang, H., Wang, S., Wang, G., Zhao, J. (2025). Generalized mixed-norm maximum correntropy for robust adaptive filtering. Applied Acoustics, 228, 110382. https://doi.org/10.1016/j.apacoust.2024.110382
27. Yuan, C., Zhou, C., Peng, J., Li, H. (2024). Mixture correntropy-based robust distance metric learning for classification. Knowledge-Based Systems, 295, 111791. https://doi.org/10.1016/j.knosys.2024.111791
28. Bai, X., Ge, Q., Zeng, P. (2025). DCM_MCCKF: A non-Gaussian state estimator with adaptive kernel size based on CS divergence. Neurocomputing, 617, 128809. https://doi.org/10.1016/j.neucom.2024.128809
29. Li, S., Xu, S., Jin, X., Tan, P. (2025). Comparison of Kernel Functions in Generalized M-estimation Using Fixed-Point Iteration. Advances in Guidance, Navigation and Control, 596–605. https://doi.org/10.1007/978-981-96-2216-0_57
30. Price, R. (1958). A useful theorem for nonlinear devices having Gaussian inputs. IEEE Transactions on Information Theory, 4 (2), 69–72. https://doi.org/10.1109/tit.1958.1057444