Adaptive ADALINE Robust Training Algorithm Under the Maximum Correntropy Criterion With Variable Center

Oleg G. Rudenko [1], Oleksandr O. Bezsonov[1] and Andrzej Szajna[2]

[1] Kharkiv National University of Radio Electronics, Nauky Ave. 14, Kharkiv, 61166, Ukraine
[2] Uniwersytet Zielonogórski, ul. Licealna 9, Zielona Gora, 65-417, Poland

Abstract

The problem of training ADALINA in the presence of non-Gaussian interference is considered. The learning algorithm is a gradient procedure for maximizing the functional. In contrast to the commonly used Gaussian kernels, the centers of which are at zero and effective for distributions with zero mean, the paper considers a modification of the criterion suitable for distributions with nonzero mean. The modification is to use correntropy with a variable center. The use of Gaussian kernels with a variable center will allow us to estimate unknown parameters under Gaussian and non-Gaussian noises with zero and non-zero mean distributions. The properties of its convergence in the stationary and non-stationary cases in conditions of Gaussian and non-Gaussian noises are investigated.

Keywords

Correntropy, maximization, functional, gradient algorithm, asymptotic estimation, convergence, non-stationary, steady state

1. Introduction

Adaptive linear element (ADALINE) was the first linear neural network proposed by Widrow B. and Hoff M., and became an alternative to the perceptron [1]. Subsequently, this element and its learning algorithm are being very commonly used in problems of identification, control, filtering, etc. The learning algorithm of Widrow-Hoff is the Kaczmarz algorithm for solving systems of linear algebraic equations [2]. Properties of this algorithm dealt with the solution of the identification problem is sufficiently described in [3].

2. The problem of ADALINE training

ADALINE is described by the equation

$$y_{n+1} = c^{*T} x_{n+1} + \xi_{n+1}, \quad (1)$$

where y_{n+1} is the observed output signal; $x_{n+1} = (x_{1,n+1}, x_{2,n+1}, .. x_{N,n+1})^T$ is the vector of output signals $N \times 1$; $c^* = (c_1^*, c_2^*, .. c_N^*)^T$ is the vector of desired parameters $N \times 1$; ξ_{n+1} is the noise; n is the discrete time.

The task of its learning consists in the definition (estimation) of the vector of parameters c^* and is reduced to minimize some of the chosen in advance performance functional (identification criterion)

$$F[e_n] = \sum_{i=1}^{n} \rho(e_i), \quad (2)$$

COLINS-2022: 6th International Conference on Computational Linguistics and Intelligent Systems, May 12–13, 2022, Gliwice, Poland
EMAIL: oleh.rudenko@nure.ua (O. Rudenko); oleksandr.bezsonov@nure.ua (O. Bezsonov); a.szajna@dtpoland.com (A. Szajna)
ORCID: 0000-0003-0859-2015 (O. Rudenko); 0000-0001-6104-4275 (O. Bezsonov); 0000-0003-4975-0534 (A. Szajna)

where $e_i = y_i - \hat{y}_i$; $\hat{y}_i = c_{i-1}^T x_i$ is the output model signal; c is the vector estimation c^*; $\rho(e_i)$ – some differential loss function satisfying the conditions:

$$\rho(e_i) \geq 0;\ \rho(0) = 0;\ \rho(e_i) = \rho(-e_i);\ \rho(e_i) \geq \rho(e_j) \qquad for \quad |e_i| \geq |e_j|.$$

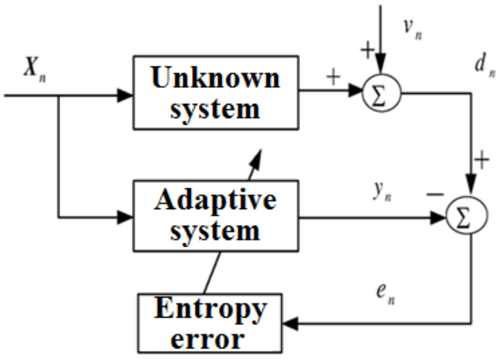

Figure 1: ADALINE

The training objective is to search for estimate c defined as the solution of a minimum extreme problem

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

or as solving equation system

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

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

If we introduce the weigh function $\omega(e) = \rho'(e)/e$, the system of equations (4) may be put as following:

$$\sum_{i=1}^{n} \omega(e_i) e_i \frac{\partial e_i}{\partial c_j} = 0, \tag{5}$$

while functional minimization (2) will be equivalent to minimizing a weighted quadratic functional, most often seen in practice

$$\min \sum_{i=1}^{n} \omega(e_i) e_i^2. \tag{6}$$

A quadratic functional the most widely used in estimating the parameters uses the second order statistics of the error signal and is quite optimal in assuming linearity and Gauss nature of signals. Indeed, when choosing $\rho(e_i) = 0.5e_i^2$ the influence function $\rho'(e_i) = e_i$, i.e. grows linearly with the increase of e_i, that explains the volatility of the least squares method valuation to outliers and distortions with big distribution "tails".

Stable M-estimation is also estimation c, defined as solving an extremal problem (3) or solving a system of equations (4), however loss function $\rho(e_i)$ is chosen as different from the quadratic one.

There are quite a number of functionals that provide the robust M-estimates but the most common are combined functionals proposed by Huber [4] and Hampel [5] consisting of quadratic, that ensures optimal estimates for the Gaussian distribution, and modular, that allows to get an estimate that is more robust to distributions with heavy "tails" (outliers). However, the effectiveness of the resulting robust estimations depends significantly on many parameters used in these criteria and chosen depending on the experience of the researcher.

The practical application of these functionals for solving the identification problem was considered in many works, in [6, 7], in particular.

Another approach to obtain robust estimates, devoid of this drawback, is the use of the fourth degree criterion [8], combined criteria using a combination of the quadratic criterion and the criterion of smallest moduli [9–11], the quadratic criterion and the fourth degree criterion [12], the fourth degree criterion and the criterion of smallest moduli [13]. It should be noted that the use of the combined criterion turned out to be very effective and much simpler when implementing the identification procedure.

One more approach that is currently widely used is the approach based on information characteristics of signals, entropy, in particular. The functional used in this case is an explicit functional of the probability density function (PDF) and includes all the higher-order statistical properties defined in PDF. Since entropy measures the mean uncertainty contained in a given PDF, minimizing it provides a reduction in error. In [14, 15], the concept of information theoretic learning (ITL) was introduced, using as a criterion the Rényi quadratic entropy, for which a nonparametric estimate based on Parzen windows with Gauss kernels is determined directly from data samples. In these works, it was proved that when using the Rényi entropy, as a result of training, the Rényi distance between the conditional probability of the density function of the desired and actual output signals for the given input signals is minimized.

The results of numerous studies indicate that in the presence of non-Gaussian, in particular, impulse noise, in measurements, an approach based on information characteristics of signals is very effective, while a criterion that considers all statistics of a higher-order error signal turns out to be more appropriate. Correntropy was introduced in [16] as a generalized measure of similarity, the maximization of which underlies the development of sufficiently simple and efficient robust algorithms.

3. Correntropy as a measure of similarity

Correntropy, defined as a localized measure of similarity, has proven to be very efficient for obtaining robust estimates due to its less sensitivity to outliers. Its name emphasizes the relationship with correlation, and also indicates the fact that its average value over time or measurements is associated with entropy, more precisely, with the argument of the logarithm in the quadratic Rényi entropy, estimated with the help of Parzen windows [17].

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

$$V(X,Y) = M\{k_\sigma(X,Y)\}, \tag{7}$$

where $M\{\bullet\}$ – is the expectation symbol; $k_\sigma(\bullet)$ – rotation invariant Mercer kernels; σ – kernel width.

The most widely used in calculating the correntropy are Gaussian ones, defined by the formula

$$k_\sigma(X,Y) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{\|X-Y\|^2}{2\sigma^2}\right\}. \tag{8}$$

When calculating the correntropy, it is necessary to know the joint distribution of random variables *X* and *Y*, which, as a rule, is not known. Since in practice there are usually a finite number of samples $\{x_i, y_i\}, i = 1,2,..., N$, the most simple estimate of the correntopy is calculated as follows:

$$\hat{V}(X,Y) = \frac{1}{N}\sum_{i=1}^{N} k_\sigma(x_i - y_i). \tag{9}$$

In tasks of identification, filtering, etc. as a functional, the correntropy between the required output signal d_i and the model output signal (real) y_i is used. When using Gaussian kernels, the optimized functional takes the form

$$J_{corr}(n) = \frac{1}{\sqrt{2\pi}\sigma}\frac{1}{N}\sum_{i=n-N+1}^{N}\exp\left(-\frac{e_i^2}{2\sigma^2}\right), \tag{10}$$

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

The use of the Taylor series expansion for the Gaussian kernel makes it possible to write the correntropy as follows:

$$V(X,Y) = \frac{1}{\sqrt{2\pi}\sigma}\sum_{n=0}^{\infty}\frac{(-1)^n}{2^n\sigma^{2n}n!}M\left\{\|X-Y\|^{2n}\right\} \tag{11}$$

4. Correntropy maximization algorithms

The gradient optimization algorithm (10) at $N = 1$ will have the form [18, 19]

$$w_{n+1} = w_n + \gamma \exp\left(-\frac{e_{n+1}^2}{2\sigma^2}\right) e_n x_{n+1}, \quad (12)$$

where γ is the parameter affecting the rate of convergence.

A significant drawback of this algorithm is the low convergence rate, which significantly limits the possibility of its use in identifying nonstationary objects. It should be noted that finding the optimal value of the parameter γ, that provides the maximum convergence rate of the algorithm, equal, as it is easy to show,

$$\gamma_{n+1} = \left(\psi_{n+1}\|x_{n+1}\|^2\right)^{-1}, \quad (13)$$

where $\psi_{n+1} = \exp\left(-\frac{e_{n+1}^2}{2\sigma^2}\right)$, leads to an analogue of Kaczmarz algorithm (Widrow–Hoff's).

In [20–23], to reduce impulse noise, a recurrent weighted least squares (RWLS) method was proposed, which minimizes the criterion

$$\psi_{n+1} = \exp\left(-\frac{e_{n+1}^2}{2\sigma^2}\right) \quad (14)$$

and having the form

$$c_{n+1} = c_n + \frac{\psi_{n+1} P_n x_{n+1}}{\lambda + \psi_{n+1} x_{n+1}^T P_n x_{n+1}} (y_{n+1} - c_n^T x_{n+1}), \quad (15)$$

$$P_{n+1} = \lambda^{-1}\left(P_n - \frac{\psi_{n+1} P_n x_{n+1} x_{n+1}^T P_n}{\lambda + \psi_{n+1} x_{n+1}^T P_n x_{n+1}} \right), \quad (16)$$

where $0 \le \lambda < 1$ is the weighing factor.

Thus, when deriving the formula for calculating P_{n+1} (16), the approximation was used

$$P_{n+1} = \lambda P_n + \psi_{n+1} x_{n+1} x_{n+1}^T. \quad (17)$$

As known, introducing a parameter λ into an algorithm is advisable when identifying nonstationary parameters.

Since a function $G_\sigma(e)$ is a local function of error e, correntropy can be used as an indicator of error in information processing and machine learning problems

$$G_\sigma(e) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{e^2}{2\sigma^2}\right). \quad (18)$$

It can be seen from (18) that the center of the Gaussian nucleus is at zero. This circumstance can lead to the fact that if the distribution of errors (noise) has a nonzero mean, function (18) will not correspond to this distribution. Therefore, the problem arises of choosing such a correntropy function that would be effective for noises having a nonzero mean.

One of the approaches to solving this problem is the use of correntropy with a variable center [24-26]

$$V_{\sigma,c}(T,Y) = M\{G_{\sigma,c}(e)\} G_\sigma(x,y) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\{e-c\}^2}{2\sigma^2}\right), \quad (19)$$

where $c \in R$ is the center.

In this case

$$V_{\sigma,c}(T,Y) = \frac{1}{\sqrt{2\pi}\sigma} \sum_{n=0}^{\infty} \frac{(-1)^n}{2^n n!} M\left(\frac{(e-c)^{2n}}{\sigma^{2n}}\right). \quad (20)$$

When σ increasing, the moments of higher orders relative to the center will decrease faster, therefore, the moment of the second order will prevail in the value $V_{\sigma,c}(T,Y)$. In particular, for $c = M\{e\}$ and $\sigma \to \infty$, maximizing the correntropy whih the center c is equivalent to minimizing the error variance.

In [27], it was proposed complex correntropy with variable center, in [28] was introduced generalized correntropy criterion. In [29] was considered maximum mixture correntropy criterion.

The solution of practical problems based on the minimization of the corresponding criteria was considered in [30–33].

Sparsity Constrained Recursive Generalized maximum correntropy criterion (MCC) with variable center algorithm was studied in [34]. Work [35], is interested in distributed MCC algorithms, based on a divide-and-conquer strategy.

Minimizing functional (19) with respect to the parameters of the model, we obtain

$$\frac{\partial E_{n+1}}{\partial w} = -\exp\left(\frac{(e_{n+1}-c)^2}{2\sigma^2}\right)\frac{(e_{n+1}-c)}{2\sigma^2}x_{n+1}; \qquad (21)$$

$$\frac{\partial E_{n+1}}{\partial c} = w\exp\left(-\frac{(e_{n+1}-c)^2}{2\sigma^2}\right)\frac{(e_{n+1}-c)}{\sigma^2}; \qquad (22)$$

$$\frac{\partial E_{n+1}}{\partial \sigma^2} = -w\exp\left(-\frac{(e_{n+1}-c)^2}{2\sigma^2}\right)\frac{(e_{n+1}-c)^2}{\sigma^3}. \qquad (23)$$

Taking these expressions into account, the algorithms for correcting the network parameters will have the form

$$w_{n+1} = w_n + \gamma_w \exp\left(-\frac{(e_{n+1}-c_{n+1})^2}{2\sigma_{n+1}^2}\right)(e_{n+1}-c_{n+1})x_{n+1}, \qquad (24)$$

$$c_{n+1} = c_n + \gamma_c \exp\left(-\frac{(e_{n+1}-c_{n+1})^2}{2\sigma_{n+1}^2}\right)(e_{n+1}-c_n); \qquad (25)$$

$$\sigma_{n+1}^2 = \sigma_n^2 - \gamma_\sigma w_{n+1} \exp\left(-\frac{(e_{n+1}-c_{n+1})^2}{2\sigma_n^2}\right)\frac{(e_{n+1}-c_{n+1})^2}{\sigma_n^3}, \qquad (26)$$

where γ_w, γ_c, γ_σ are the parameters of the algorithm that regulate the step size and affect the rate of its convergence.

4.1. Multidimensional object

If the object under study has several outputs, then the output signal will be a vector signal and the error will also be a vector value, and the learning algorithm will have the form

$$w_{n+1} = w_n + \gamma \exp\left(-\|e_{n+1}-c\|_{R^{-1}}^2\right)e_{n+1}x_{n+1}, \qquad (27)$$

where $\|e_{n+1}-c\|_{R^{-1}}^2 = (e_{n+1}-c)^T R^{-1}(e_{n+1}-c)$; R^{-1} is the covariance matrix of the input vector

$$R_{n+1}^{-1} = R_n^{-1} - \gamma_R w_{n+1} \exp\left(-\|e_{n+1}-c_{n+1}\|_{R_n^{-1}}^2\right)(e_{n+1}-c_{n+1})(e_{n+1}-c_{n+1})^T. \qquad (28)$$

4.2. Investigation of the issues of convergence of the algorithm.

Consider the estimation error

$$\Theta_{n+1} = c_{n+1} - c^*. \qquad (29)$$

Then

$$e_{n+1} = \Theta_{n+1}^T x_{n+1} + \xi_{n+1} = e_{n+1}^a + \xi_{n+1}, \qquad (30)$$

where $e_{n+1}^a = \Theta_{n+1}^T x_{n+1}$ is a priori error.

In this case, the estimation algorithm can be written as

$$w_{n+1} = w_n + \gamma f(e_{n+1}) x_{n+1}, \tag{31}$$

where $f(e_{n+1}) = \exp\left(\frac{(e_{n+1}-c)^2}{2\sigma^2}\right)(e_{n+1}-c).$

Writing down algorithm (31) with respect to estimation errors, we have

$$\theta_{n+1} = \theta_n - \gamma f(e_{n+1}) x_{n+1}.$$

Multiplying both sides of the given expression on the left by θ_{n+1}^T, we get

$$\|\theta_{n+1}\|^2 = \|\theta_n\|^2 - 2\gamma f(e_{n+1}) e_{n+1}^a + \gamma^2 f^2(e_{n+1}) \|x_{n+1}\|^2. \tag{32}$$

Averaging both sides of (32), i.e.

$$M\left\{\|\theta_{n+1}\|^2\right\} = M\left\{\|\theta_n\|^2\right\} - 2\gamma M\left\{f(e_{n+1}) e_{n+1}^a\right\} + \gamma^2 M\left\{f^2(e_{n+1}) \|x_{n+1}\|^2\right\}, \tag{33}$$

we obtain the condition for the convergence of algorithm (31) in the mean square

$$0 < \gamma \le \frac{2M\left\{f(e_{n+1}) e_{n+1}^a\right\}}{M\left\{f^2(e_{n+1}) \|x_{n+1}\|^2\right\}}.$$

Consider a steady state. Since in steady state

$$\lim_{n\to\infty} M\left\{\|\theta_{n+1}\|^2\right\} = \lim_{n\to\infty} M\left\{\|\theta_n\|^2\right\},$$

it follows from (33) that

$$2\lim_{n\to\infty} M\left\{e_{n+1}^a f(e_{n+1})\right\} = \gamma tr R_x \lim_{n\to\infty} M\left\{f^2(e_{n+1})\right\}, \tag{34}$$

where tr denotes the trace operator.

To calculate the steady-state value of the estimation error, we define $M\left\{f^2(e_{n+1}) \|x_{n+1}\|^2\right\}$ and $M\left\{f^2(e_{n+1})\right\}$

Consider the case of Gaussian noise $\xi \sim N\left(0, \sigma_\xi^2\right)$. Using Price's theorem [36], we obtain

$$\lim_{n\to\infty} M\left\{e_{n+1}^a f(e_{n+1})\right\} = \lim_{n\to\infty} M\left\{e_{n+1}^a f(e_{n+1}^a + \xi_{n+1})\right\} = \lim_{n\to\infty} M\left\{\left(e_{n+1}^a\right)^2\right\} M\left\{f'(e_{n+1})\right\} =$$
$$= \lim_{n\to\infty} SM\left\{\exp\left(-\frac{(e_{n+1}-c)^2}{2\sigma^2}\right)\left(1-\frac{(e_{n+1}-c)^2}{\sigma^2}\right)\right\} = \frac{S}{\sqrt{2\pi}\sigma_e} \lim_{n\to\infty} \int_{-\infty}^{+\infty} \exp\left(-\frac{(e_{n+1}-c)^2}{2\sigma^2}\right)\left(1-\frac{(e_{n+1}-c)^2}{\sigma^2}\right) * \tag{35}$$
$$* \exp\left(-\frac{(e_{n+1}-c_e)^2}{2\sigma_e^2}\right) de_{n+1},$$

where $\sigma_e^2 = M\left\{\left(e_{n+1}^a\right)^2 + \sigma_\xi^2\right\}$; $S = \lim_{n\to\infty} M\left\{\left(a_{n+1}^a\right)^2\right\}$; c_e is the center of gaussian error e_{n+1}.

Similarly, we define

$$M\left\{f^2(e_{n+1})\right\} = \lim_{n\to\infty} M\left\{\exp\left(-\frac{(e_{n+1}-c)^2}{2\sigma^2}\right)(e_{n+1}-c)^2\right\} =$$
$$= \frac{1}{\sqrt{2\pi}\sigma_e} \lim_{n\to\infty} \int_{-\infty}^{+\infty} \exp\left(-\frac{(e_{n+1}-c)^2}{2\sigma^2}\right)(e_{n+1}-c)^2 \exp\left(-\frac{(e_{n+1}-c_e)^2}{2\sigma_e^2}\right) de_{n+1}. \tag{36}$$

Substitution of (35) and (36) into (34) gives the expression for the steady-state error $\lim_{n\to\infty} M\left\{\left(e_{n+1}^a\right)^2\right\}$

$$\lim_{n\to\infty} M\left\{\left(e_{n+1}^a\right)^2\right\} = \frac{A}{2B}, \tag{37}$$

where

$$A=\gamma trR_x \lim_{n\to\infty}\int_{-\infty}^{+\infty}\exp\left(-\frac{(e_{n+1}-c)^2}{2\sigma^2}\right)(e_{n+1}-c)^2\exp\left(-\frac{(e_{n+1}-c_e)^2}{2\sigma_e^2}\right)de_{n+1};$$

$$B=\lim_{n\to\infty}\int_{-\infty}^{+\infty}\exp\left(-\frac{(e_{n+1}-c)^2}{2\sigma^2}\right)\left(1-\frac{(e_{n+1}-c)^2}{\sigma^2}\right)\exp\left(-\frac{(e_{n+1}-c_e)^2}{2\sigma_e^2}\right)de_{n+1};$$

or

$$\lim_{n\to\infty} M\left\{\left(e_{n+1}^a\right)^2\right\}= \\ =\frac{\gamma trR_x \lim_{n\to\infty}\int_{-\infty}^{+\infty}\exp\left(-\frac{(e_{n+1}-c)^2}{2\sigma^2}\right)(e_{n+1}-c)^2\exp\left(-\frac{(e_{n+1}-c_e)^2}{2\sigma_e^2}\right)de_{n+1}}{\lim_{n\to\infty}\int_{-\infty}^{+\infty}\exp\left(-\frac{(e_{n+1}-c)^2}{2\sigma^2}\right)\left(1-\frac{(e_{n+1}-c)^2}{\sigma^2}\right)\exp\left(-\frac{(e_{n+1}-c_e)^2}{2\sigma_e^2}\right)de_{n+1}}. \quad (38)$$

This expression shows that $\lim_{n\to\infty} M\left\{\left(e_{n+1}^a\right)^2\right\}=0$ when choosing $\gamma\to 0$.

Consider the case of non-Gaussian interference. In this case, we use the Taylor series expansion. In the steady state, the estimated parameters change (are corrected) insignificantly. Therefore, we can rewrite (34) as follows:

$$2M\left\{e^a f(e)\right\}=\gamma trR_x M\left\{f^2(e)\right\} \quad (39)$$

We expand the function $f(e)$ in a Taylor series, limiting ourselves to terms of the second order of smallness

$$f(e)=f(e^a+\xi)=f(\xi)+f'(\xi)e^a+0.5f''(\xi)\left(e^a\right)^2+o\left(\left(e^a\right)^2\right), \quad (40)$$

where

$$f'(\xi)=\exp\left(-\frac{(\xi-c)^2}{2\sigma^2}\right)\left(1-\frac{(\xi-c)^2}{\sigma^2}\right); \quad (41)$$

$$f''(\xi)=\exp\left(-\frac{(\xi-c)^2}{2\sigma^2}\right)\left(\frac{(\xi-c)^3}{\sigma^4}-\frac{3(\xi-c)}{\sigma^2}\right). \quad (42)$$

Assuming that the interference does not correlate with the signals and the prior error e^a, we can write

$$M\left\{e^a f(e)\right\}=M\left\{e^a f(\xi)+f'(\xi)\left(e^a\right)^2+o\left(e^a\right)^2\right\}\approx SM\left\{f'(\xi)\right\}; \quad (43)$$

$$M\left\{f^2(e)\right\}\approx M\left\{f^2(\xi)\right\}+SM\left\{f(\xi)f''(\xi)+\left|f'(\xi)\right|^2\right\} \quad (44)$$

Substituting (43) and (44) into (40), we have

$$S=\frac{\gamma trR_x M\left\{f^2(\xi-c)\right\}}{2M\left\{f'(\xi-c)\right\}-\gamma trR_x M\left\{f(\xi-c)f''(\xi-c)+\left|f'(\xi-c)\right|^2\right\}}. \quad (45)$$

Substitution of (41), (42) into (45) gives

$$S=\frac{\gamma trR_x M\left\{K(\xi-c)^2\right\}}{2M\left\{K'\left(1-\frac{(\xi-c)^2}{\sigma^2}\right)\right\}-\gamma trR_x M\left\{K\left(1+\frac{2(\xi-c)^4}{\sigma^4}-\frac{5(\xi-c)^2}{\sigma^2}\right)\right\}}, \quad (46)$$

where

$$K=\exp\left(-\frac{(\xi-c)^2}{\sigma^2}\right);\quad K'=\exp\left(-\frac{(\xi-c)^2}{2\sigma^2}\right).$$

4.3. Non-stationary case

Let us assume that the estimated parameters are non-stationary, i.e.

$$c_{n+1}^{*}=c_{n}^{*}+\Delta c^{*}, \tag{47}$$

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

Consider the error vector $\theta_{n+1}=c_{n+1}-c_{n+1}^{*}$.

Then, taking into account (30), the estimation algorithm can be written as

$$\theta_{n+1}=\theta_n-c_{n+1}^{*}+\gamma f(e_{n+1})x_{n+1}=\theta_n-\Delta c^{*}+\gamma f(e_{n+1})x_{n+1}, \tag{48}$$

Multiplying both sides of (48) on the left by θ_{n+1}^T and calculating the mathematical expectation, we get

$$M\left\{\|\theta_{n+1}\|^2\right\}=M\left\{\|\theta_n\|^2\right\}-2\gamma M\left\{x_{n+1}^T\theta_n f(e_{n+1})\right\}+\gamma^2 M\left\{f^2(e_{n+1})\|x_{n+1}\|^2\right\}+$$
$$+M\left\{\left\|\Delta c^{*}\right\|^2\right\}+M\left\{x_{n+1}^T\Delta c^{*}\right\}+M\left\{\Delta c^{*T}x_{n+1}\right\}-2\gamma M\left\{x_{n+1}^T\Delta c^{*}f(e_{n+1})\right\},$$

Taking into account the statistical properties of signals and noise, we have

$$M\left\{\|\theta_{n+1}\|^2\right\}=M\left\{\|\theta_n\|^2\right\}-2\gamma M\left\{e_{n+1}^a f(e_{n+1})\right\}+\gamma^2 M\left\{f^2(e_{n+1})\|x_{n+1}\|^2\right\}+M\left\{\left\|\Delta c^{*}\right\|^2\right\}. \tag{49}$$

For Gaussian interference, using Price's theorem gives

$$\lim_{n\to\infty}M\left\{e_{n+1}^a f(e_{n+1})\right\}=\lim_{n\to\infty}M\left\{e_{n+1}^a f(e_{n+1}^a+\xi_{n+1})\right\}=\lim_{n\to\infty}M\left\{\left(e_{n+1}^a\right)^2\right\}M\left\{f'(e_{n+1})\right\}=$$
$$=\lim_{n\to\infty}SM\left\{\exp\left(-\frac{(e_{n+1}-c)^2}{2\sigma^2}\right)\left(1-\frac{(e_{n+1}-c)^2}{\sigma^2}\right)\right\}= \tag{50}$$
$$=\frac{S}{\sqrt{2\pi}\sigma_e}\lim_{n\to\infty}\int_{-\infty}^{+\infty}\exp\left(-\frac{(e_{n+1}-c)^2}{2\sigma^2}\right)\left(1-\frac{(e_{n+1}-c)^2}{\sigma^2}\right)\exp\left(-\frac{(e_{n+1}-c_e)^2}{2\sigma_e^2}\right)de_{n+1}=\frac{S\sigma^3}{\left(\sigma^2+\sigma_\xi^2+S\right)^{\frac{3}{2}}};$$

$$M\left\{f^2(e_{n+1})\right\}=\lim_{n\to\infty}M\left\{\exp\left(-\frac{(e_{n+1}-c)^2}{2\sigma^2}\right)(e_{n+1}-c)^2\right\}=$$
$$=\frac{1}{\sqrt{2\pi}\sigma_e}\lim_{n\to\infty}\int_{-\infty}^{+\infty}\exp\left(-\frac{(e_{n+1}-c)^2}{2\sigma^2}\right)(e_{n+1}-c)^2\exp\left(-\frac{(e_{n+1}-c_e)^2}{2\sigma_e^2}\right)de_{n+1}=\frac{\sigma^3\left(S+\sigma_\xi^2\right)}{\left(2\sigma_\xi^2+\sigma^2+2S\right)^{\frac{3}{2}}}. \tag{51}$$

Considering that

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

for steady state when $\lim_{n\to\infty}M\left\{\|\theta_{n+1}\|^2\right\}=\lim_{n\to\infty}M\left\{\|\theta_n\|^2\right\}$,

from expression (49) we obtain

$$\frac{2S}{\left(\sigma^2+\sigma_\xi^2+S\right)^{\frac{3}{2}}}=\frac{\gamma trR_x\left(\sigma_\xi^2+S\right)}{\left(\sigma^2+2\sigma_\xi^2+2S\right)^{\frac{3}{2}}}+\frac{trR_c}{\gamma\sigma^3}. \tag{52}$$

From this ratio, we can determine the value S

$$S=\frac{\gamma trR_x\left(\sigma_\xi^2+S\right)\left(\sigma^2+\sigma_\xi^2+S\right)^{\frac{3}{2}}}{\left(\sigma^2+2\sigma_\xi^2+2S\right)^{\frac{3}{2}}}+\frac{trR_c\left(\sigma^2+\sigma_\xi^2+S\right)^{\frac{3}{2}}}{2\gamma\sigma^3}. \quad (53)$$

For $\sigma^2\to\infty$, we have the value of S for the least squares

$$\lim_{\sigma\to\infty} S=\frac{\gamma trR_x\sigma_\xi^2+\gamma^{-1}trR_c}{2-\gamma trR_x}.$$

In the case of non-Gaussian noise, we have

$$M\left\{e_{n+1}^a f\left(e_{n+1}\right)\right\}\approx M\left\{e_{n+1}^a f\left(\xi_{n+1}\right)+e_{n+1}^a f'\left(\xi_{n+1}\right)\right\}\approx SM\left\{f'\left(\xi_{n+1}\right)\right\}. \quad (54)$$

$$M\left\{f^2\left(e_{n+1}\right)\right\}\approx M\left\{\left(f\left(\xi_{n+1}\right)+e_{n+1}^a f'\left(\xi_{n+1}\right)+0{,}5f''\left(\xi_{n+1}\right)e_{n+1}^{a2}\right)^2\right\}\approx$$
$$\approx M\left\{f^2\left(\xi_{n+1}\right)\right\}+SM\left\{\left(f\left(\xi_{n+1}\right)f''\left(\xi_{n+1}\right)+\left(f'\left(\xi_{n+1}\right)\right)^2\right)\right\}, \quad (55)$$

where

$$f'(\xi_{n+1})=\exp\left(-\frac{(\xi-c)^2}{2\sigma^2}\right)\left(1-\frac{(\xi-c)^2}{\sigma^2}\right);$$

$$f''(\xi_{n+1})=\exp\left(-\frac{(\xi_{n+1}-c)^2}{2\sigma^2}\right)\left(\frac{\xi_{n+1}^3}{\xi_{n+1}^4}-\frac{3\xi_{n+1}}{\sigma^2}\right).$$

Substituting (54) and (55) into (49), after simple transformations we obtain

$$S=\frac{\gamma A+\gamma^{-1}B}{C-\gamma D}, \quad (56)$$

where

$$A=trR_xM\left\{(\xi_{n+1}-c)^2\exp\left(-\frac{(\xi_{n+1}-c)^2}{\sigma^2}\right)\right\};$$

$$B=trR_c;$$

$$C=2M\left\{\left(1-\frac{(\xi_{n+1}-c)^2}{2\sigma^2}\right)\exp\left(-\frac{(\xi_{n+1}-c)^2}{\sigma^2}\right)\right\};$$

$$D=trR_xM\left\{\left(1+\frac{2(\xi_{n+1}-c)^4}{\sigma^4}-\frac{5(\xi_{n+1}-c)^2}{\sigma^2}\right)\exp\left(-\frac{(\xi_{n+1}-c)^2}{\sigma^2}\right)\right\}$$

This expression shows that S is a monotonically non-increasing function of the parameter γ.

From the condition $\partial S/\partial\gamma=0$, an equation can be obtained to determine the optimal value of the parameter γ that provides the minimum value S

$$AC\gamma^2+BD\gamma-BC=0.$$

5. Numerical experiments

The problem of ADALINE parameters adjustment was considered. Sequences of normally distributed quantities $x(k)\sim N(0;1)$ were chosen as the input signal $x(k)$. When testing the robustness of the algorithms, an independent noise distributed according to the Rayleigh law with $\sigma=1$ was added to the output signal of the object. The histogram of such noise is shown in fig. 2. The simulation results for various values of the parameter are shown in fig. 3. In fig. 4 shows the graphs of changes in the error when choosing the RWLS algorithm (15)-(16) and algorithm (31) respectively, here

$$RMSE = \sqrt{\frac{1}{2}\left\| c_n - c^* \right\|^2},$$

where c_n and c^* denote estimated and target parameters vectors respectively.

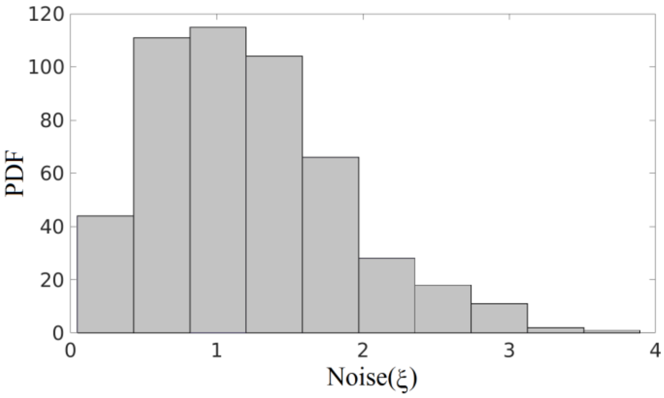

Figure 2: Noise distribution

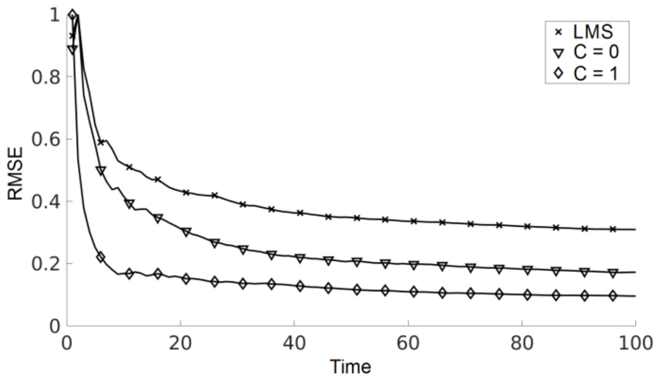

Figure 3: Different algorithms results

6. Conclusion

The work considered an adaptive robust learning algorithm for ADALINE when using the information criterion of correntropy with variable center as a learning criterion.

The properties of its convergence in the stationary and non-stationary cases in conditions of non-Gaussian noises are investigated.

The importance of choosing the width of the Gaussian kernel, which affects the rate of convergence of estimation algorithms and the error in the steady state, is noted, and the expediency of developing procedures for adaptive correction of the kernel width is indicated.

The estimates obtained are quite general and depend both on the degree of nonstationarity of the object and on the statistical characteristics of useful signals and interference.

7. References

[1] Widrow, B., Hoff, M. Adaptive switching circuits, IRE WESCON Convention Record. Part 4. New York: Institute of Radio Engineers, pp. 96–104.(1960)

[2] Kaczmarz, S. Angenäherle Auflösung von Systemen linearer Gleichungen, Bull. Int. Acad. Polon. Sci. Lett., C 1, Sci. Math. Nat. Ser. A. S. 355-357.(1937)

English translation: Kaczmarz S. Approximate solution of systems of linear equations, Int. J. of Control, 1993, 57, p. 1269-1271.
[3] Либероль, Б.Д., Руденко. О.Г, Бессонов, А.А. Исследование сходимости одношаговых адаптивных алгоритмов идентификации. Проблемы управления и информатики, 5, с.19-32(2018).
[4] Хьюбер, П. Робастность в статистике. М.:Мир. 304 с.(1984)
[5] Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J., Stahel, W.A. Robust Statistics. The Approach Based on Influence Functions. N.Y.: John Wiley and Sons, 526 p.(1986)
[6] Rudenko, O., Bezsonov, O. Function approximation using robust radial basis function networks. J. of Intelligent Learning Systems and Applications, 3, pp.17-25 (2011).
[7] Руденко, О.Г., Бессонов, А.А. М-обучение радиально-базисных сетей с использованием асимметричных функций влияния. Проблемы управления и информатики, 1, с. 79-93 (2012).
[8] Walach, E. Widrow D. The least mean fourth (LMF) adaptive algorithm and its family, IEEE Trans, IT 30, pp. 275–283 (1984).
[9] Chambers, J., Avlonitis, A. A Robust Mixed-Norm Adaptive Filter Algorithm. IEEE Signal Processing Letters, 4, 2, pp. 46-48 (1997).
[10] Papoulis, E.V., Stathaki T. A Normalized Robust Mixed-Norm Adaptive Algorithm for System Identification, IEEE Signal Processing Letters, 2004, 11,1, p. 56-59
[11] Chambers, J., Tanrikulu, O., Constantinides, A.G. Least mean mixed-norm adaptive filtering, Electronics letters, 30, 19, pp. 1574-1575 (1984).
[12] Zerguine, A. A variable-parameter normalized mixed-norm (VPNMN) adaptive algorithm. EURASIP Journal on Advances in Signal Processing, 55, 13 p. (2012)
[13] Руденко, О.Г., Безсонов, О.О., Сердюк, Н.М. Олійник, К.О. Романюк, О.С. Робастна ідентифікація об'єктів на основі мінімізації комбінованого функціоналу, Системи обробки інформації, 1 (160), с.80-88 (2020).
[14] Principe, J.C., Xu, D., Zhao, Q., Fisher, J. W. Learning from examples with information theoretic criteria. J. VLSI Signal Process. Syst., 26, 1–2, p. 61–77 (2000).
[15] Principe, J.C., Xu, D., Fisher, J. Information theoretic learning / In: S. Haykin (Ed.), Unsupervised Adaptive Filtering. New York: Wiley, pp. 265–319 (2000).
[16] Liu, W., Pokharel, P.P, Principe, J.C. Correntropy: Properties and Applications in Non-Gaussian Signal Processing. IEEE Trans. on Signal Processing, 1, pp. 5286-5298.(2007)
[17] Wang, W., Zhao, J., Qu, H., Chen, B., Principe, J.C. An adaptive kernel width update method of correntropy for channel estimation. IEEE International Conference on Digital Signal Processing (DSP), pp. 916–920 (2015).
[18] Chen, B., Xing, L., Liang, J., Zheng, N, Principe, J.C. Steady-state mean-square error analysis for adaptive filtering under the maximum correntropy criterion. Signal Process. Lett. IEEE, 21(7), pp. 880–884 (2014).
[19] Ma, W., Qua, H., Guib, G., Li, L., Zhao, J., Chen, B. Maximum correntropy criterion based sparse adaptive filtering algorithms for robust channel estimation under non-Gaussian environments. J. of the Franklin Institute, 352,2, pp. 2708-2727 (2015)
[20] Xiong, W., Schindelhauer, C., So, H, Wang, Z. Maximum Correntropy Criterion for Robust TOA-Based Localization in NLOS Environments. Circuits, Systems, and Signal Processing,40, pp. 6325–6339 (2021).
[21] Hu, C., Wang,G,. Ho, K.C., Liang J. Robust Ellipse Fitting With Laplacian Kernel Based Maximum Correntropy Criterion. IEEE Transactions on Image Processing,30, pp. 3127 - 3141 (2021).
[22] Flores, T.K.S., Villanueva, J.M.M., Gomes, H.P., Catunda, S.Y.C. Adaptive Pressure Control System Based on the Maximum Correntropy Criterion. *Sensors*, *21*(15), 5156 (2021).
[23] Hu, T. Kernel-based maximum correntropy criterion with gradient method, CPAA, 19(8), 4159-4177 (2020).
[24] Chen, B., Wang, X., Li, Y., Principe, J.C Maximum Correntropy Criterion with Variable Center. IEEE Signal Process. Letter,26.(5), pp.1212–1216 (2019)

[25] Zhu, L., Song, C., Pan, L., Li, J. Adaptive Filtering Under the Maximum Correntropy Criterion With Variable Center. IEEE Acces, pp. 105902–105908 (2019).
[26] Wang, X., Han, J. Affine Projection Algorithm by Employing Maximum Correntropy Criterion for System Identification of Mixed Noise.IEEE Acces, 7, pp. 182515– 182526 (2019)
[27] Dong,F., Qian,G., Wang, S. Complex Correntropy with Variable Center: Definition, Properties, and Application to Adaptive Filtering. Entropy (Basel). 22(1): 70 p.(2020)
[28] Yang, J., Cao, J., Xue, A. Robust Maximum Mixture Correntropy Criterion-Based Semi-Supervised ELM With Variable Center. IEEE Transactions on Circuits and Systems II: Express Briefs, 67, 12, pp. 3572–3576 (2020)
[29] Zhang, J., Huang, G., Zhan, L. Generalized Correntropy Criterion-Based Performance Assessment for Non-Gaussian Stochastic Systems. Entropy, 23, 764 (2021).
[30] 30. Li, Y., Wang, Y., Sun, L. A Proportionate Normalized Maximum Correntropy Criterion Algorithm with Correntropy Induced Metric Constraint for Identifying Sparse Systems. Symmetry, 10, 683 (2018).
[31] 31. Zhang, J., Huang, G., Zhan, L. Generalized Correntropy Criterion-Based Performance Assessment for Non-Gaussian Stochastic Systems. Entropy, 23, 764 (2021).
[32] 32. Wang, X., Han, J. Affine Projection Algorithm by Employing Maximum Correntropy Criterion for System Identification of Mixed Noise. IEEE Acces, 7, pp. 182515–182526 (2019).
[33] 33. Sun, Q., Zhang, H., Wang, X., Ma, W., Chen, B. Sparsity Constrained Recursive Generalized Maximum Correntropy Criterion With Variable Center Algorithm. IEEE Transactions on Circuits and Systems II: Express Briefs, 67, 12, pp. 3517–3521 (2020).
[34] Xie, F., Hu,T., Wang, S., Wang, D. Maximum Correntropy Criterion with Distributed Method. Mathematics, 10, 304, 17 p. (2022).
[35] Wang, X., Han, J. Affine Projection Algorithm by Employing Maximum Correntropy Criterion for System Identification of Mixed Noise. IEEE Acces, 7, pp. 182515–182526 (2019).
[36] Price, R. A useful theorem for nonlinear devices having Gaussian inputs. IEEE Transactions on Information Theory, 4 (2), pp. 69–72. (1958).