Signal Processing 226 (2025) 109643

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Decorrelation algorithm based on the information theoretic learning

Xinyan Hou, Haiquan Zhao *, Xiaoqiang Long

Key Laboratory of Magnetic Suspension Technology and Maglev Vehicle, Ministry of Education, School of Electrical Engineering, Southwest Jiaotong University, Chengdu, 610031, China

ARTICLE INFO

Keywords:
Adaptive filtering
Decorrelation
Maximum correntropy
Acoustic echo cancellation
Performance analysis

ABSTRACT

Recently, the Bayesian decorrelation algorithms have gained attention because they can effectively avoid trade-off between learning rate and estimation accuracy by circumventing the problems associated with constant step-size and significantly enhance the convergence speed of adaptive filtering algorithms by decorrelation of signals. However, the present Bayesian decorrelation algorithms are constructed on the basis of the minimum-mean-square-error (MMSE) criterion, leading to deterioration in performance when confronted with non-Gaussian noise environments. Therefore, this paper introduces the maximum correntropy criterion in the decorrelation algorithm and develops a decorrelation recursive maximum correntropy criterion (DRMCC) algorithm based on the observation model to achieve robust performance. The proposed algorithm has a similar iterative equations as the Bayesian decorrelation algorithm. In addition, we present the convergence condition and analyze the transient behavior of the DRMCC algorithm. When the free parameters of the proposed algorithm are unknown, a common method of parameter estimation is provided. Finally, numerical simulations corroborate that the designed DRMCC algorithm can realize outstanding performance under system identification and acoustic echo cancellation (AEC) applications in non-Gaussian noise environments and the derived theoretical model can arrive at good agreement with simulation results for both stationary and nonstationary scenarios.

1. Introduction

Adaptive filtering algorithms (AFAs) have been extensively employed across several fields, including acoustic echo cancellation (AEC), and active noise control [1,2]. Among them, the least-mean-square (LMS) algorithm and its normalized variant (normalized LMS, NLMS) gain popularity owing to their simple implementation and low complexity. However, when dealing with strongly correlated input signals like speech data, the LMS and NLMS algorithms may exhibit reduced convergence speed. To accelerate the algorithm's convergence rate, various frameworks are employed to adaptive filters, including the subband adaptive filter (SAF) [3–5], the affine projection (AP) [6–9], and the decorrelation filter [10,11]. The SAF algorithm can decompose the fullband input and output signals into several mutually independent subbands using the analysis filter bank. This decomposition serves to diminish the correlation among the signals. The AP algorithm effectively mitigates the effects of correlated inputs by combining signals from both current and past time instances. The AP algorithm exhibits high computational complexity because it involves the calculation of the inverse of the input matrix. The decorrelation filter based on observation model utilize different methods to estimate the vector of decorrelation coefficients of the filter, such as decorrelation NLMS (DNLMS) [11]. However, DNLMS algorithm suffers from the contradiction between the requirements of adaption speed and filtering accuracy under constant step-size. To solve this problem, a Bayesian decorrelation LMS (BDLMS) algorithm is proposed by utilizing Bayesian techniques to estimate the coefficients of the system model and the autoregressive model [12].

Unfortunately, the algorithms mentioned above are constructed on the minimum-mean-square-error (MMSE) criterion and encounter issues related to convergence and stability problems when confronted with non-Gaussian noise environments. The widespread occurrence of non-Gaussian noise has sparked significant interest in exploring the non-MMSE criteria to enhance the algorithm's robustness [13]. Over the past years, there has been a notable adoption of information-theoretic learning (ITL) to develop various robust AFAs. This is attributed to its ability to excerpt more statistical information from the original signal [14]. The maximum correntropy criterion (MCC) as a representative of the ITL theory exhibits exceptional performance, especially in impulsive noise environments [15–18].

To strengthen robustness of the BDLMS algorithm, a decorrelation recursive MCC (DRMCC) algorithm is developed in this paper, which is developed via incorporating the observation model and MCC criterion. The proposed algorithm is the same as the BDLMS algorithm when the

* Corresponding author.
E-mail addresses: houxy@my.swjtu.edu.cn (X. Hou), hqzhao_swjtu@126.com (H. Zhao), xiaoqiangL@my.swjtu.edu.cn (X. Long).

https://doi.org/10.1016/j.sigpro.2024.109643
Received 27 March 2024; Received in revised form 30 July 2024; Accepted 31 July 2024
Available online 8 August 2024
0165-1684/© 2024 Elsevier B.V. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

kernel widths tend to infinity. Additionally, the DRMCC algorithm's convergence condition and transient behavior are subjected to analysis. Numerical simulations exhibit that compared to the BDLMS algorithm, the developed DRMCC algorithm retains the advantages of the BDLMS algorithm in processing colored input signals and at the same time exhibits excellent robustness under non-Gaussian noise environments. The contributions of this work are summarized as follows.

(1) The DRMCC algorithm is proposed based on the observation model to cope with the performance degradation of existing decorrelation algorithms in non-Gaussian noise environments.
(2) The convergence conditions and transient behavior of the proposed algorithm are analyzed based on statistical properties.
(3) The experimental results show that the proposed algorithm has excellent performance in system Identification and acoustic echo cancellation (AEC) applications.

2. Robust decorrelation algorithm

Consider the scenario in which the input signal $\boldsymbol{x}_l$ and the desired signal y_l satisfy the linear model

$$y_l = \boldsymbol{x}_l^T \boldsymbol{\omega}_{o,l} + \upsilon_l, \tag{1}$$

where $\boldsymbol{\omega}_{o,l} \in \mathbb{R}^L$ denotes the system weight vector of length L to be estimated, $\boldsymbol{x}_l = [x_l, \ldots, x_{l-L+1}]^T$, and υ_l is the measurement noise with variance σ_υ^2. In this paper, we assume that $\boldsymbol{\omega}_{o,l}$ is a random walk model, i.e.,

$$\boldsymbol{\omega}_{o,l} = \boldsymbol{\omega}_{o,l-1} + \boldsymbol{q}_l, \tag{2}$$

where $\boldsymbol{q}_l$ denotes the process perturbation with covariance matrix $\sigma_q^2 \boldsymbol{I}_L$, and $\boldsymbol{I}_L$ denotes the $L \times L$ identity matrix. We make the assumption that the correlated input signal x_l is produced based on the following autoregressive (AR) model:

$$x_l = \sum_{i=1}^{M} \alpha_i x_{l-i} + z_l = \boldsymbol{x}_{\alpha,l}^T \boldsymbol{\alpha} + z_l, \tag{3}$$

where $\boldsymbol{x}_{\alpha,l} = [x_{l-1}, \ldots, x_{l-M}]^T$, $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_M]^T$, z_l represents a zero-mean noise with variance σ_z^2, and M denotes the order of the autoregressive model. Therefore, Eq. (3) is rewritten as

$$\boldsymbol{x}_l = \sum_{i=1}^{M} \alpha_i \boldsymbol{x}_{l-i} + \boldsymbol{z}_l = \boldsymbol{X}_l \boldsymbol{\alpha} + \boldsymbol{z}_l, \tag{4}$$

where $\boldsymbol{X}_l = [\boldsymbol{x}_{l-1}, \ldots, \boldsymbol{x}_{l-M}]$, and $\boldsymbol{z}_l = [z_l, \ldots, z_{l-L+1}]^T$. To obtain the decorrelated observation model, we represent the estimate of $\boldsymbol{\alpha}$ at moment l as $\hat{\boldsymbol{\alpha}}_l$. The decorrelated input signal vector $\bar{\boldsymbol{x}}_l$ can be defined as the estimation $\hat{\boldsymbol{z}}_l$ of the white noise process $\boldsymbol{z}_l$, which is represented as

$$\bar{\boldsymbol{x}}_l \triangleq \hat{\boldsymbol{z}}_l = \boldsymbol{x}_l - \boldsymbol{X}_l \hat{\boldsymbol{\alpha}}_l. \tag{5}$$

Substituting Eq. (5) into Eq. (1), we have

$$y_l = \left[\boldsymbol{x}_{l-1}^T \boldsymbol{\omega}_{o,l}, \ldots, \boldsymbol{x}_{l-M}^T \boldsymbol{\omega}_{o,l}\right] \hat{\boldsymbol{\alpha}}_l + \bar{\boldsymbol{x}}_l^T \boldsymbol{\omega}_{o,l} + \upsilon_l. \tag{6}$$

Under stationary scenario, it holds that $\boldsymbol{\omega}_o = \boldsymbol{\omega}_{o,l}$, which implies

$$\boldsymbol{x}_{l-i}^T \boldsymbol{\omega}_{o,l} = y_{l-i} - \upsilon_{l-i}. \tag{7}$$

Under the nonstationary scenario, we assume that the unknown weight vector undergoes gradual change and M is relatively small. This permits us to make the following approximation:

$$\boldsymbol{x}_{l-i}^T \boldsymbol{\omega}_{o,l} \approx y_{l-i} - \upsilon_{l-i}. \tag{8}$$

Using Eqs. (7) and (8), Eq. (6) is concluded as

$$\begin{aligned} y_l &\approx \left[y_{l-1} - \upsilon_{l-1}, \ldots, y_{l-M} - \upsilon_{l-M}\right] \hat{\boldsymbol{\alpha}}_l + \bar{\boldsymbol{x}}_l^T \boldsymbol{\omega}_{o,l} + \upsilon_l \\ &= \boldsymbol{y}_l^T \hat{\boldsymbol{\alpha}}_l - \boldsymbol{v}_l^T \hat{\boldsymbol{\alpha}}_l + \bar{\boldsymbol{x}}_l^T \boldsymbol{\omega}_{o,l} + \upsilon_l, \end{aligned} \tag{9}$$

where $\boldsymbol{y}_l = [y_{l-1}, \ldots, y_{l-M}]^T$ and $\boldsymbol{v}_l = [\upsilon_{l-1}, \ldots, \upsilon_{l-M}]^T$. Define the decorrelated observation signal $\bar{y}_l$ as

$$\bar{y}_l \triangleq y_l - \boldsymbol{y}_l^T \hat{\boldsymbol{\alpha}}_l. \tag{10}$$

Thus, the decorrelated observation model is stated as

$$\bar{y}_l = \bar{\boldsymbol{x}}_l^T \boldsymbol{\omega}_{o,l} + \bar{\upsilon}_l. \tag{11}$$

where $\bar{\upsilon}_l = \upsilon_l - \boldsymbol{v}_l^T \hat{\boldsymbol{\alpha}}_l$.

The correntropy is a similarity measure between random variables X and Y [15,16], defined as

$$V(X,Y) = E\left[k_\sigma(X,Y)\right]. \tag{12}$$

where $E[\cdot]$ is the expectation operator and $k_\sigma(X,Y)$ is the kernel function with width σ. The kernel function is usually a Gaussian kernel, i.e.,

$$k_\sigma(X,Y) = \exp\left(-\frac{(X-Y)^2}{2\sigma^2}\right). \tag{13}$$

Using the parzen estimation method, the correntropy estimation based on N samples is expressed as

$$\hat{V}(X,Y) = \frac{1}{N}\sum_{i=1}^{N} k_\sigma(X,Y). \tag{14}$$

For the decorrelated observation model presented in Eq. (11), the cost function is defined as follows through using both the MCC criterion and weighted least squares (WLS) method as follows [19]:

$$\boldsymbol{J} = k_{\sigma_1}\left(\left\|\bar{e}_l\right\|_{r_{\bar{e}}^{-1}}^2\right) + k_{\sigma_1}\left(\left\|\boldsymbol{e}_{\omega,l}\right\|_{\boldsymbol{P}_{\omega,l}^{-1}}^2\right), \tag{15}$$

where $\bar{e}_l = \bar{y}_l - \bar{\boldsymbol{x}}_l^T \boldsymbol{\omega}_l$, $r_{\bar{e}} = E[\bar{\upsilon}_l^2] = \sigma_{\bar{\upsilon}}^2$, $\boldsymbol{P}_{\omega,l} = E[\boldsymbol{e}_{\omega,l}\boldsymbol{e}_{\omega,l}^T]$, and $\boldsymbol{e}_{\omega,l} = \boldsymbol{\omega}_l - \boldsymbol{\omega}_{l-1}$. To derive the optimal estimate $\boldsymbol{\omega}_l$, it is necessary for the gradient of $\boldsymbol{J}$ with respect to $\boldsymbol{\omega}_l$ to be zero, which is expressed as follows:

$$(\boldsymbol{P}_{\omega,l}^{-1} + L_{\omega,l} r_{\bar{e}}^{-1} \bar{\boldsymbol{x}}_l \bar{\boldsymbol{x}}_l^T)\boldsymbol{\omega}_l = \boldsymbol{P}_{\omega,l}^{-1}\boldsymbol{\omega}_{l-1} + L_{\omega,l} r_{\bar{e}}^{-1} \bar{\boldsymbol{x}}_l \bar{y}_l \tag{16}$$

where $L_{\omega,l} = k_{\sigma_1}\left(\left\|\bar{e}_l\right\|_{r_{\bar{e}}^{-1}}^2\right) / k_{\sigma_1}\left(\left\|\boldsymbol{e}_{\omega,l}\right\|_{\boldsymbol{P}_{\omega,l}^{-1}}^2\right)$. To obtain the optimal estimation $\boldsymbol{\omega}_{o,l}$ by adding and subtracting the term $L_{\omega,l} r_{\bar{e}}^{-1} \bar{\boldsymbol{x}}_l \bar{\boldsymbol{x}}_l^T \boldsymbol{\omega}_{l-1}$ to the right-hand side of Eq. (16) yields

$$\begin{aligned} &(\boldsymbol{P}_{\omega,l}^{-1} + L_{\omega,l} r_{\bar{e}}^{-1} \bar{\boldsymbol{x}}_l \bar{\boldsymbol{x}}_l^T)\boldsymbol{\omega}_l = L_{\omega,l} r_{\bar{e}}^{-1} \bar{\boldsymbol{x}}_l (\bar{y}_l - \bar{\boldsymbol{x}}_l^T \boldsymbol{\omega}_{l-1}) \\ &+ (\boldsymbol{P}_{\omega,l}^{-1} + L_{\omega,l} r_{\bar{e}}^{-1} \bar{\boldsymbol{x}}_l \bar{\boldsymbol{x}}_l^T)\boldsymbol{\omega}_{l-1}. \end{aligned} \tag{17}$$

If the term $(\boldsymbol{P}_{\omega,l}^{-1} + L_{\omega,l} r_{\bar{e}}^{-1} \bar{\boldsymbol{x}}_l \bar{\boldsymbol{x}}_l^T)$ is reversible, we have

$$\boldsymbol{\omega}_l = \boldsymbol{\omega}_{l-1} + \boldsymbol{K}_l(\bar{y}_l - \bar{\boldsymbol{x}}_l^T \boldsymbol{\omega}_{l-1}). \tag{18}$$

where $\boldsymbol{K}_l = L_{\omega,l} r_{\bar{e}}^{-1} (\boldsymbol{P}_{\omega,l}^{-1} + L_{\omega,l} r_{\bar{e}}^{-1} \bar{\boldsymbol{x}}_l \bar{\boldsymbol{x}}_l^T)^{-1} \bar{\boldsymbol{x}}_l$ denotes Kalman gain. To simplify the calculation of $\boldsymbol{K}_l$, using the matrix inversion lemma [20, 21] with $\boldsymbol{A} = \boldsymbol{P}_{\omega,l}^{-1}$, $\boldsymbol{B} = \bar{\boldsymbol{x}}_l$, $\boldsymbol{C} = L_{\omega,l} r_{\bar{e}}^{-1} \boldsymbol{I}_L$ and $\boldsymbol{D} = \bar{\boldsymbol{x}}_l^T$ yields

$$\begin{aligned} \boldsymbol{K}_l &= L_{\omega,l} r_{\bar{e}}^{-1} \left(\boldsymbol{P}_{\omega,l} - \boldsymbol{P}_{\omega,l} \bar{\boldsymbol{x}}_l \left(L_{\omega,l}^{-1} r_{\bar{e}} + \bar{\boldsymbol{x}}_l^T \boldsymbol{P}_{\omega,l} \bar{\boldsymbol{x}}_l\right)^{-1} \bar{\boldsymbol{x}}_l^T \boldsymbol{P}_{\omega,l}\right) \bar{\boldsymbol{x}}_l \\ &= L_{\omega,l} r_{\bar{e}}^{-1} \left(\boldsymbol{P}_{\omega,l} \bar{\boldsymbol{x}}_l - \frac{\boldsymbol{P}_{\omega,l} \bar{\boldsymbol{x}}_l \bar{\boldsymbol{x}}_l^T \boldsymbol{P}_{\omega,l} \bar{\boldsymbol{x}}_l}{L_{\omega,l}^{-1} r_{\bar{e}} + \bar{\boldsymbol{x}}_l^T \boldsymbol{P}_{\omega,l} \bar{\boldsymbol{x}}_l}\right) \\ &= L_{\omega,l} r_{\bar{e}}^{-1} \left(\frac{L_{\omega,l}^{-1} r_{\bar{e}} \boldsymbol{P}_{\omega,l} \bar{\boldsymbol{x}}_l + \boldsymbol{P}_{\omega,l} \bar{\boldsymbol{x}}_l \bar{\boldsymbol{x}}_l^T \boldsymbol{P}_{\omega,l} \bar{\boldsymbol{x}}_l - \boldsymbol{P}_{\omega,l} \bar{\boldsymbol{x}}_l \bar{\boldsymbol{x}}_l^T \boldsymbol{P}_{\omega,l} \bar{\boldsymbol{x}}_l}{L_{\omega,l}^{-1} r_{\bar{e}} + \bar{\boldsymbol{x}}_l^T \boldsymbol{P}_{\omega,l} \bar{\boldsymbol{x}}_l}\right) \\ &= \frac{L_{\omega,l} \boldsymbol{P}_{\omega,l} \bar{\boldsymbol{x}}_l}{r_{\bar{e}} + L_{\omega,l} \bar{\boldsymbol{x}}_l^T \boldsymbol{P}_{\omega,l} \bar{\boldsymbol{x}}_l} \end{aligned} \tag{19}$$

In addition, according to the Riccati equation, the weight error covariance can be updated with

$$\boldsymbol{P}_{\omega,l} = [\boldsymbol{I}_L - \boldsymbol{K}_l \bar{\boldsymbol{x}}_l^T][\boldsymbol{P}_{\omega,l-1} + \sigma_q^2 \boldsymbol{I}_L] \tag{20}$$

To further enhance the computational efficiency of the DRMCC algorithm, we configure the covariance matrix $\boldsymbol{P}_{\omega,l}$ as $\sigma_{\omega,l}^2 \boldsymbol{I}_L$, where

$\sigma_{\omega,l}^2 = \mathrm{Tr}(\boldsymbol{P}_{\omega,l})/L$ and $\mathrm{Tr}(\cdot)$ is the trace operation. Hence, the Kalman gain is expressed as

$$\boldsymbol{K}_l = \xi_{\omega,l}\bar{\boldsymbol{x}}_l = L_{\omega,l}\sigma_{\omega,l}^2\left(L_{\omega,l}\sigma_{\omega,l}^2\|\bar{\boldsymbol{x}}_l\|^2 + \sigma_{\bar{v}}^2\right)^{-1}\bar{\boldsymbol{x}}_l, \tag{21}$$

where $\xi_{\omega,l} = L_{\omega,l}\sigma_{\omega,l}^2\left(L_{\omega,l}\sigma_{\omega,l}^2\|\bar{\boldsymbol{x}}_l\|^2 + \sigma_{\bar{v}}^2\right)^{-1}$. Taking Eq. (21) into Eq. (20), performing tracking operation and dividing by L, results in

$$\sigma_{\omega,l}^2 = (1 - \xi_{\omega,l}\|\bar{\boldsymbol{x}}_l\|^2/L)(\sigma_{\omega,l-1}^2 + \sigma_q^2). \tag{22}$$

Algorithm 1: Pseudo code for DRMCC algorithm

Input : Inputs x_l and outputs y_l of the model
Output: Weight vector $\boldsymbol{\omega}_l$ of the filter
/* Initialization */
$l \leftarrow 1$, $\hat{\boldsymbol{\alpha}}_{-1} \leftarrow \mathbf{0}$, $\hat{\boldsymbol{\alpha}}_0 \leftarrow \mathbf{0}$, $\boldsymbol{\omega}_{-1} \leftarrow \mathbf{0}$, $\boldsymbol{\omega}_0 \leftarrow \mathbf{0}$, $\sigma_{\alpha,0}^2 \leftarrow 1$, $\sigma_{\omega,0} \leftarrow 1$
/* Parameter settings */
Set kernel bandwidths σ_1, σ_2. $\sigma_{\bar{v}}^2$, σ_q^2, and σ_z^2 are known or estimated online
/* Computation */
while $l \le S$ **do**
  $\boldsymbol{x}_{\alpha,l} \leftarrow [x_{l-1}, \ldots, x_{l-M}]^T$
  $\boldsymbol{x}_l \leftarrow [x_l, \ldots, x_{l-L+1}]^T$
  $\boldsymbol{X}_l \leftarrow [\boldsymbol{x}_{l-1}, \ldots, \boldsymbol{x}_{l-M}]$
  $\boldsymbol{y}_l \leftarrow [y_{l-1}, \ldots, y_{l-M}]^T$
  $e_{x,l} \leftarrow x_l - \boldsymbol{x}_{\alpha,l}^T\hat{\boldsymbol{\alpha}}_{l-1}$
  $\boldsymbol{e}_{\alpha,l} \leftarrow \hat{\boldsymbol{\alpha}}_{l-1} - \hat{\boldsymbol{\alpha}}_{l-2}$
  $L_{\alpha,l} \leftarrow k_{\sigma_2}\left(\|e_{x,l}\|_{\sigma_z^{-2}}^2\right)/k_{\sigma_2}\left(\|\boldsymbol{e}_{\alpha,l}\|_{\sigma_{\alpha,l-1}^{-2}}^2\right)$
  $\xi_{\alpha,l} \leftarrow L_{\alpha,l}\sigma_{\alpha,l-1}^2\left(L_{\alpha,l}\sigma_{\alpha,l-1}^2\|\boldsymbol{x}_{\alpha,l}\|^2 + \sigma_z^2\right)^{-1}$
  $\sigma_{\alpha,l}^2 \leftarrow \left[1 - \xi_{\alpha,l}\|\boldsymbol{x}_{\alpha,l}\|^2/M\right]\sigma_{\alpha,l-1}^2$
  $\hat{\boldsymbol{\alpha}}_l \leftarrow \hat{\boldsymbol{\alpha}}_{l-1} + \xi_{\alpha,l}e_{x,l}\boldsymbol{x}_{\alpha,l}$
  $\bar{\boldsymbol{x}}_l \leftarrow \boldsymbol{x}_l - \boldsymbol{X}_l\hat{\boldsymbol{\alpha}}_l$
  $\bar{y}_l \leftarrow y_l - \boldsymbol{y}_l^T\hat{\boldsymbol{\alpha}}_l$
  $\bar{e}_l \leftarrow \bar{y}_l - \bar{\boldsymbol{x}}_l^T\boldsymbol{\omega}_{l-1}$
  $\boldsymbol{e}_{\omega,l} \leftarrow \boldsymbol{\omega}_{l-1} - \boldsymbol{\omega}_{l-2}$
  $L_{\omega,l} \leftarrow k_{\sigma_1}\left(\|\bar{e}_l\|_{\sigma_{\bar{v}}^{-2}}^2\right)/k_{\sigma_1}\left(\|\boldsymbol{e}_{\omega,l}\|_{\sigma_{\omega,l-1}^{-2}}^2\right)$
  $\xi_{\omega,l} \leftarrow L_{\omega,l}\sigma_{\omega,l-1}^2\left(L_{\omega,l}\sigma_{\omega,l-1}^2\|\bar{\boldsymbol{x}}_l\|^2 + \sigma_{\bar{v}}^2\right)^{-1}$
  $\sigma_{\omega,l}^2 \leftarrow \left[1 - \xi_{\omega,l}\|\bar{\boldsymbol{x}}_l\|^2/L\right][\sigma_{\omega,l}^2 + \sigma_q^2]$
  $\boldsymbol{\omega}_l \leftarrow \boldsymbol{\omega}_{l-1} + \xi_{\omega,l}\bar{e}_l\bar{\boldsymbol{x}}_l$
  $l \leftarrow l + 1$
end
return $\boldsymbol{\omega}$

Therefore, the DRMCC algorithm is summarized as

$$\xi_{\omega,l} = L_{\omega,l}\sigma_{\omega,l-1}^2\left(L_{\omega,l}\sigma_{\omega,l-1}^2\|\bar{\boldsymbol{x}}_l\|^2 + \sigma_{\bar{v}}^2\right)^{-1}, \tag{23}$$

$$\sigma_{\omega,l}^2 = (1 - \xi_{\omega,l}\|\bar{\boldsymbol{x}}_l\|^2/L)(\sigma_{\omega,l-1}^2 + \sigma_q^2), \tag{24}$$

$$\boldsymbol{\omega}_l = \boldsymbol{\omega}_{l-1} + \xi_{\omega,l}(\bar{y}_l - \bar{\boldsymbol{x}}_l^T\boldsymbol{\omega}_{l-1})\bar{\boldsymbol{x}}_l. \tag{25}$$

Finally, we need to estimate $\boldsymbol{\alpha}$ in the iterative progression. Applying the earlier derivation process, we can effortlessly arrive at an updated equation for estimating the decorrelation coefficient vector $\hat{\boldsymbol{\alpha}}_l$, which is outlined as

$$\xi_{\alpha,l} = L_{\alpha,l}\sigma_{\alpha,l-1}^2\left(L_{\alpha,l}\sigma_{\alpha,l-1}^2\|\boldsymbol{x}_{\alpha,l}\|^2 + \sigma_z^2\right)^{-1}, \tag{26}$$

$$\sigma_{\alpha,l}^2 = \left[1 - \xi_{\alpha,l}\|\boldsymbol{x}_{\alpha,l}\|^2/M\right]\sigma_{\alpha,l-1}^2, \tag{27}$$

$$\hat{\boldsymbol{\alpha}}_l = \hat{\boldsymbol{\alpha}}_{l-1} + \xi_{\alpha,l}(x_l - \boldsymbol{x}_{\alpha,l}^T\hat{\boldsymbol{\alpha}}_{l-1})\boldsymbol{x}_{\alpha,l}, \tag{28}$$

where $L_{\alpha,l} = k_{\sigma_2}\left(\|e_{x,l}\|_{r_x^{-1}}^2\right)/k_{\sigma_2}\left(\|\boldsymbol{e}_{\alpha,l}\|_{\boldsymbol{P}_{\alpha,l}^{-1}}^2\right)$, $e_{x,l} = x_l - \boldsymbol{x}_{\alpha,l}^T\hat{\boldsymbol{\alpha}}_l$, $r_x = E[z_l^2] = \sigma_z^2$, $\boldsymbol{e}_{\alpha,l} = \hat{\boldsymbol{\alpha}}_l - \hat{\boldsymbol{\alpha}}_{l-1}$, and $\boldsymbol{P}_{\alpha,l} = E[\boldsymbol{e}_{\alpha,l}\boldsymbol{e}_{\alpha,l}^T]$.

Remark 1. It can be seen from Eqs. (23) to (28) that the DRMCC algorithm will degrade to the BDLMS algorithm when the kernel widths σ_1 and σ_2 tend to infinity.

Remark 2. As shown in the iterative equations of the DRMCC algorithm, the DRMCC algorithm can be regarded as a variable step-size decorrelation algorithm. Therefore, the proposed algorithm is effective in avoiding the tradeoff caused by constant step size.

Derived from the above analysis, Algorithm 1 provides the pseudo-code for the DRMCC algorithm.

3. Performance analysis

In this section, we focus on the mean stability and mean-square performance of the DRMCC algorithm from a statistical perspective. For subsequent analysis, we define the weight-error vector as

$$\tilde{\boldsymbol{\omega}}_l = \boldsymbol{\omega}_{o,l} - \boldsymbol{\omega}_l. \tag{29}$$

The errors $\bar{e}_l$ and $\boldsymbol{e}_{\omega,l}$ can be written as

$$\bar{e}_l = \bar{y}_l - \bar{\boldsymbol{x}}_l^T\boldsymbol{\omega}_{l-1} = \bar{\boldsymbol{x}}_l^T\tilde{\boldsymbol{\omega}}_{l-1} + \bar{\boldsymbol{x}}_l^T\boldsymbol{q}_l + \bar{v}_l, \tag{30}$$

$$\boldsymbol{e}_{\omega,l} = \boldsymbol{\omega}_l - \boldsymbol{\omega}_{l-1} = \tilde{\boldsymbol{\omega}}_{l-1} + \boldsymbol{q}_l - \tilde{\boldsymbol{\omega}}_l. \tag{31}$$

Therefore, the weight-error update equation is calculated by

$$\begin{aligned}\tilde{\boldsymbol{\omega}}_l &= \tilde{\boldsymbol{\omega}}_{l-1} + \boldsymbol{q}_l - \xi_{\omega,l}[\bar{\boldsymbol{x}}_l\bar{\boldsymbol{x}}_l^T\tilde{\boldsymbol{\omega}}_{l-1} + \bar{\boldsymbol{x}}_l\bar{\boldsymbol{x}}_l^T\boldsymbol{q}_l + \bar{\boldsymbol{x}}_l\bar{v}_l] \\ &= \boldsymbol{A}_l\tilde{\boldsymbol{\omega}}_{l-1} + \boldsymbol{A}_l\boldsymbol{q}_l - \xi_{\omega,l}\bar{\boldsymbol{x}}_l\bar{v}_l,\end{aligned} \tag{32}$$

where $\boldsymbol{A}_l = \boldsymbol{I}_L - \xi_{\omega,l}\bar{\boldsymbol{x}}_l\bar{\boldsymbol{x}}_l^T$. To perform performance analysis, we adopt the following typical assumptions.

- A1 The noise z_l is a zero-mean random variable with variance $E[z_l^2] = \sigma_z^2$.
- A2 The noise v_l is an independent, identically distributed process with zero mean and variance $E\left[v_l^2\right] = \sigma_v^2$ [12].
- A3 The perturbation $\boldsymbol{q}_l$ is a zero-mean random vector with covariance $\boldsymbol{Q} = E[\boldsymbol{q}_l\boldsymbol{q}_l^T] = \sigma_q^2\boldsymbol{I}_L$ [22].
- A4 The decorrelated input vector $\bar{\boldsymbol{x}}_l$, the gain $\xi_{\omega,l}$, the noise v_l, and the perturbation $\boldsymbol{q}_l$ are mutually independent.
- A5 L is long enough to ensure that the fluctuations of $\bar{\boldsymbol{x}}_l$ at adjacent time are negligible.

Substituting Eq. (23) into Eq. (24) yields

$$\sigma_{\omega,l}^2 = \left\{1 - \frac{1}{L}\frac{L_{\omega,l}\sigma_{\omega,l-1}^2\|\bar{\boldsymbol{x}}_l\|^2}{L_{\omega,l}\sigma_{\omega,l-1}^2\|\bar{\boldsymbol{x}}_l\|^2 + \sigma_{\bar{v}}^2}\right\}[\sigma_{\omega,l-1}^2 + \sigma_q^2] \tag{33}$$

Using Assumptions A4 and A5, for sufficiently long L, $\sigma_{\omega,l-1}^2$ is sufficiently small such that $E[\sigma_{\omega,l}^2\|\bar{\boldsymbol{x}}_l\|^2] \approx E[\sigma_{\omega,l}^2]E[\|\bar{\boldsymbol{x}}_l\|^2]$. And the expectation of gain $\xi_{\omega,l}$ can be written as

$$E[\xi_{\omega,l}] = \frac{E[L_{\omega,l}]E[\sigma_{\omega,l-1}^2]}{E[L_{\omega,l}]E[\sigma_{\omega,l-1}^2]E[\|\bar{\boldsymbol{x}}_l\|^2] + \sigma_{\bar{v}}^2}, \tag{34}$$

From Eq. (34), it is clear that the variance of $\xi_{\omega,l}$ is also remarkably small. Consequently, $E[\sigma_{\omega,l}^2]$ is approximated as

$$E[\sigma_{\omega,l}^2] \approx \left[1 - E[\xi_{\omega,l}]E[\|\bar{\boldsymbol{x}}_l\|^2]/L\right][E[\sigma_{\omega,l-1}^2] + \sigma_q^2]. \tag{35}$$

In addition, using the Taylor approximation, $L_{\omega,l}$ can be expressed as

$$L_{\omega,l} \approx \left\{1 - \frac{\|\bar{e}_l\|_{\sigma_{\bar{v}}^{-2}}^2}{2\sigma_1^2}\right\} \Big/ \left\{1 - \frac{\|\boldsymbol{e}_{\omega,l}\|_{\sigma_{\omega,l-1}^{-2}}^2}{2\sigma_1^2}\right\}. \tag{36}$$

Since L is sufficiently long, we can assume that $E\left[\left\|\bar{\boldsymbol{x}}_l^T\tilde{\boldsymbol{\omega}}_{l-1}\right\|^2\right]$ and $E\left[\left\|\tilde{\boldsymbol{\omega}}_l\right\|^2\right]$ is equal to zero. Therefore, the expectation of $L_{\omega,l}$ is rewritten as

$$E\left[L_{\omega,l}\right] \approx \left\{1-\frac{\sigma_q^2\mathrm{Tr}\left(\boldsymbol{R}_{\bar{x}\bar{x}}\right)+\sigma_{\bar{v}}^2}{2\sigma_1^2\sigma_{\bar{v}}^2}\right\} \Big/ \left\{1-\frac{L\sigma_q^2}{2\sigma_1^2\sigma_{\omega,l-1}^2}\right\}, \tag{37}$$

where $\boldsymbol{R}_{\bar{x}\bar{x}}$ denotes the covariance matrix of the decorrelated input vector $\bar{\boldsymbol{x}}_l$. Since $\bar{\boldsymbol{x}}_l$ is decorrelated, we can assume that $\boldsymbol{R}_{\bar{x}\bar{x}}=\sigma_z^2\boldsymbol{I}_L$, which implies that the decorrelated input signals are independent of each other and the covariance matrix of $\bar{\boldsymbol{x}}_l$ is only related to the noise z_l. Therefore, Eq. (37) is rewritten as

$$E\left[L_{\omega,l}\right] \approx \left\{1-\frac{L\sigma_q^2\sigma_z^2+\sigma_{\bar{v}}^2}{2\sigma_1^2\sigma_{\bar{v}}^2}\right\} \Big/ \left\{1-\frac{L\sigma_q^2}{2\sigma_1^2\sigma_{\omega,l-1}^2}\right\}. \tag{38}$$

According to the definition $\bar{v}_l=v_l-\boldsymbol{v}_l^T\hat{\boldsymbol{\alpha}}_l$, the variance of $\bar{v}_l$ can be written as

$$\sigma_{\bar{v}}^2=\sigma_v^2\left\{1+E[\|\hat{\boldsymbol{\alpha}}_l\|^2]\right\}. \tag{39}$$

For analytical convenience, we directly utilize the expectation of $\|\hat{\boldsymbol{\alpha}}_l\|^2$ obtained through ensemble averages.

Using the above assumptions and approximations, we can obtain the expectations of the all variables in Eq. (34), which facilitates our subsequent analysis.

3.1. Mean behavior

Using the assumptions A1–A5, computing the expectation of Eq. (32) yields

$$E\left[\tilde{\boldsymbol{\omega}}_l\right]=E\left[\boldsymbol{A}_l\right]E\left[\tilde{\boldsymbol{\omega}}_{l-1}\right], \tag{40}$$

where $E\left[\boldsymbol{A}_l\right]=\boldsymbol{I}_L-E[\xi_{\omega,l}]\boldsymbol{R}_{\bar{x}\bar{x}}=[1-E[\xi_{\omega,l}]\sigma_z^2]\boldsymbol{I}_L$. Based on Eq. (40), the DRMCC algorithm converges asymptotically for any initial condition if the gain following conditions are satisfied

$$0<E[\xi_{\omega,l}]<\frac{2}{\lambda_{max}(\boldsymbol{R}_{\bar{x}\bar{x}})}, \tag{41}$$

where $\lambda_{\max}(\cdot)$ indicates the maximum eigenvalue of the matrix.

3.2. Mean-square behavior

Firstly, the mean-square deviation (MSD) is described as

$$\mathrm{MSD}_l=E\left[\|\tilde{\boldsymbol{\omega}}_l\|^2\right]. \tag{42}$$

Calculating the expectation of the squared weighted Euclidean norm on both sides of Eq. (32) and adopting Assumptions A1-A5 yields

$$\begin{aligned}E\left[\|\tilde{\boldsymbol{\omega}}_l\|^2\right]=&E\left[\|\tilde{\boldsymbol{\omega}}_{l-1}\|_{\boldsymbol{B}_l}^2\right]+\mathrm{Tr}(\boldsymbol{B}_l\boldsymbol{Q})\\&+\mathrm{Tr}(\sigma_{\bar{v}}^2E[\xi_{\omega,l}]^2\sigma_z^2\boldsymbol{I}_L)\end{aligned} \tag{43}$$

where $\boldsymbol{B}_l=E[\boldsymbol{A}_l^T\boldsymbol{A}_l]$. Using the vectorization operation, Eq. (43) is converted to

$$\begin{aligned}E\left[\|\tilde{\boldsymbol{\omega}}_l\|_{\mathrm{vec}(\boldsymbol{I}_L)}^2\right]=&E\left[\|\tilde{\boldsymbol{\omega}}_{l-1}\|_{\boldsymbol{F}_l\mathrm{vec}(\boldsymbol{I}_L)}^2\right]+\mathrm{vec}(\boldsymbol{Q}^T)^T\boldsymbol{F}_l\mathrm{vec}(\boldsymbol{I}_L)\\&+\sigma_{\bar{v}}^2\sigma_z^2E[\xi_{\omega,l}]^2\mathrm{vec}(\boldsymbol{I}_L)^T\mathrm{vec}(\boldsymbol{I}_L),\end{aligned} \tag{44}$$

where $\boldsymbol{F}_l=E[\boldsymbol{A}_l^T]\otimes E[\boldsymbol{A}_l^T]$, $\otimes$ denotes the Kronecker product, and $\mathrm{vec}(\cdot)$ represents the vectorization operator. Let $l=1$, then Eq. (44) is written as

$$\begin{aligned}E\left[\|\tilde{\boldsymbol{\omega}}_1\|_{\mathrm{vec}(\boldsymbol{I}_L)}^2\right]=&E\left[\|\tilde{\boldsymbol{\omega}}_0\|_{\boldsymbol{F}_1\mathrm{vec}(\boldsymbol{I}_L)}^2\right]+\mathrm{vec}(\boldsymbol{Q}^T)^T\boldsymbol{F}_1\mathrm{vec}(\boldsymbol{I}_L)\\&+\sigma_{\bar{v}}^2\sigma_z^2E[\xi_{\omega,1}]^2\mathrm{vec}(\boldsymbol{I}_L)^T\mathrm{vec}(\boldsymbol{I}_L).\end{aligned} \tag{45}$$

When $l=2$, then Eq. (44) is expressed as

$$\begin{aligned}&E\left[\|\tilde{\boldsymbol{\omega}}_2\|_{\mathrm{vec}(\boldsymbol{I}_L)}^2\right]=E\left[\|\tilde{\boldsymbol{\omega}}_1\|_{\boldsymbol{F}_2\mathrm{vec}(\boldsymbol{I}_L)}^2\right]+\mathrm{vec}(\boldsymbol{Q}^T)^T\boldsymbol{F}_2\mathrm{vec}(\boldsymbol{I}_L)\\&+\sigma_{\bar{v}}^2\sigma_z^2E[\xi_{\omega,2}]^2\mathrm{vec}(\boldsymbol{I}_L)^T\mathrm{vec}(\boldsymbol{I}_L)\\&=E\left[\|\tilde{\boldsymbol{\omega}}_0\|_{\boldsymbol{F}_1\boldsymbol{F}_2\mathrm{vec}(\boldsymbol{I}_L)}^2\right]+\mathrm{vec}(\boldsymbol{Q}^T)^T\boldsymbol{F}_1\boldsymbol{F}_2\mathrm{vec}(\boldsymbol{I}_L)\\&+\sigma_{\bar{v}}^2\sigma_z^2E[\xi_{\omega,1}]^2\mathrm{vec}(\boldsymbol{I}_L)^T\boldsymbol{F}_2\mathrm{vec}(\boldsymbol{I}_L)+\mathrm{vec}(\boldsymbol{Q}^T)^T\boldsymbol{F}_2\mathrm{vec}(\boldsymbol{I}_L)\\&+\sigma_{\bar{v}}^2\sigma_z^2E[\xi_{\omega,2}]^2\mathrm{vec}(\boldsymbol{I}_L)^T\mathrm{vec}(\boldsymbol{I}_L)\\&=E\left[\|\tilde{\boldsymbol{\omega}}_0\|_{\left(\prod_{i=1}^2\boldsymbol{F}_i\right)\mathrm{vec}(\boldsymbol{I}_L)}^2\right]+\mathrm{vec}(\boldsymbol{Q}^T)^T\left[\left(\prod_{j=i}^2\boldsymbol{F}_i\right)+\boldsymbol{F}_2\right]\mathrm{vec}(\boldsymbol{I}_L)\\&+\sigma_{\bar{v}}^2\sigma_z^2\mathrm{vec}(\boldsymbol{I}_L)^T\left\{E[\xi_{\omega,1}]^2\boldsymbol{F}_2+E[\xi_{\omega,2}]^2\right\}\mathrm{vec}(\boldsymbol{I}_L).\end{aligned} \tag{46}$$

Therefore, iterating Eq. (44), we have

$$\begin{aligned}E\left[\|\tilde{\boldsymbol{\omega}}_l\|_{\mathrm{vec}(\boldsymbol{I}_L)}^2\right]=&E\left[\|\tilde{\boldsymbol{\omega}}_0\|_{\left(\prod_{i=1}^l\boldsymbol{F}_i\right)\mathrm{vec}(\boldsymbol{I}_L)}^2\right]\\&+\mathrm{vec}(\boldsymbol{Q}^T)^T\left[\sum_{i=1}^l\left(\prod_{j=i}^l\boldsymbol{F}_i\right)\right]\mathrm{vec}(\boldsymbol{I}_L)\\&+\sigma_{\bar{v}}^2\sigma_z^2\mathrm{vec}(\boldsymbol{I}_L)^T\left\{\sum_{i=1}^{l-1}E[\xi_{\omega,i}]^2\left(\prod_{j=i+1}^l\boldsymbol{F}_j\right)\right\}\mathrm{vec}(\boldsymbol{I}_L)\\&+\sigma_{\bar{v}}^2\sigma_z^2E[\xi_{\omega,l}]^2\mathrm{vec}(\boldsymbol{I}_L)^T\mathrm{vec}(\boldsymbol{I}_L)\end{aligned} \tag{47}$$

Subtracting Eq. (47) at time l from Eq. (47) at time $l-1$ yields

$$\begin{aligned}&E\left[\|\tilde{\boldsymbol{\omega}}_l\|_{\mathrm{vec}(\boldsymbol{I}_L)}^2\right]=E\left[\|\tilde{\boldsymbol{\omega}}_{l-1}\|_{\mathrm{vec}(\boldsymbol{I}_L)}^2\right]+\sigma_{\bar{v}}^2\sigma_z^2E[\xi_{\omega,l}]^2\mathrm{vec}(\boldsymbol{I}_L)^T\mathrm{vec}(\boldsymbol{I}_L)\\&-\sigma_{\bar{v}}^2\sigma_z^2E[\xi_{\omega,l-1}]^2\mathrm{vec}(\boldsymbol{I}_L)^T\mathrm{vec}(\boldsymbol{I}_L)+E\left[\|\tilde{\boldsymbol{\omega}}_0\|_{\left(\prod_{i=1}^l\boldsymbol{F}_i-\prod_{i=1}^{l-1}\boldsymbol{F}_i\right)\mathrm{vec}(\boldsymbol{I}_L)}^2\right]\\&+\mathrm{vec}(\boldsymbol{Q}^T)^T\left[\boldsymbol{F}_l+\sum_{i=1}^{l-1}\left(\prod_{j=i}^{l-1}\boldsymbol{F}_j\right)\boldsymbol{F}_l-\sum_{i=1}^{l-1}\left(\prod_{j=i}^{l-1}\boldsymbol{F}_j\right)\right]\mathrm{vec}(\boldsymbol{I}_L)\\&+\sigma_{\bar{v}}^2\sigma_z^2\mathrm{vec}(\boldsymbol{I}_L)^T\left\{E[\xi_{\omega,l-1}]^2\boldsymbol{F}_l+\sum_{i=1}^{l-2}E[\xi_{\omega,i}]^2\left(\prod_{j=i+1}^{l-1}\boldsymbol{F}_j\right)\boldsymbol{F}_l\right.\\&\left.-\sum_{i=1}^{l-2}E[\xi_{\omega,i}]^2\left(\prod_{j=i+1}^{l-1}\boldsymbol{F}_j\right)\right\}\mathrm{vec}(\boldsymbol{I}_L)\\&=E\left[\|\tilde{\boldsymbol{\omega}}_{l-1}\|_{\mathrm{vec}(\boldsymbol{I}_L)}^2\right]+\sigma_{\bar{v}}^2\sigma_z^2E[\xi_{\omega,l}]^2\mathrm{vec}(\boldsymbol{I}_L)^T\mathrm{vec}(\boldsymbol{I}_L)\\&+E\left[\|\tilde{\boldsymbol{\omega}}_0\|_{\left(\prod_{i=1}^{l-1}\boldsymbol{F}_i\right)\left(\boldsymbol{F}_l-\boldsymbol{I}_{L^2}\right)\mathrm{vec}(\boldsymbol{I}_L)}^2\right]\\&+\mathrm{vec}(\boldsymbol{Q}^T)^T\left[\sum_{i=1}^{l-1}\left(\prod_{j=i}^{l-1}\boldsymbol{F}_j\right)\right]\left(\boldsymbol{F}_l-\boldsymbol{I}_{L^2}\right)\mathrm{vec}(\boldsymbol{I}_L)\\&+\mathrm{vec}(\boldsymbol{Q}^T)^T\boldsymbol{F}_l\mathrm{vec}(\boldsymbol{I}_L)\\&+\sigma_{\bar{v}}^2\sigma_z^2\mathrm{vec}(\boldsymbol{I}_L)^T\left\{\sum_{i=1}^{l-2}E[\xi_{\omega,i}]^2\left(\prod_{j=i+1}^{l-1}\boldsymbol{F}_j\right)\right\}\left(\boldsymbol{F}_l-\boldsymbol{I}_{L^2}\right)\mathrm{vec}(\boldsymbol{I}_L)\\&+\sigma_{\bar{v}}^2\sigma_z^2E[\xi_{\omega,l-1}]^2\mathrm{vec}(\boldsymbol{I}_L)^T\left(\boldsymbol{F}_l-\boldsymbol{I}_{L^2}\right)\mathrm{vec}(\boldsymbol{I}_L).\end{aligned} \tag{48}$$

Utilizing Eq. (48), it becomes straightforward to derive the transient MSD behavior of the DRMCC algorithm.

3.3. Complexity analysis

In this subsection, we compare the computational complexity of the DRMCC algorithm with several existing algorithms. The number of addition, multiplication and exponentiation operations per iteration are summarized in Table 1, where P denotes the projection order.

As can be seen from Table 1, the computational complexity of all the algorithms increases with the filter length L. In addition, the computational complexity of the APMCC algorithm exhibits a cubic relationship

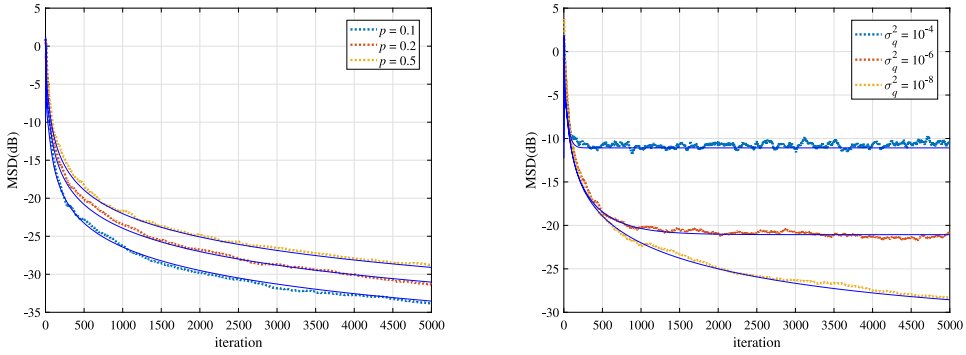

Fig. 1. Simulated and theoretical learning curves under (a) the stationary and (b) the non-stationary scenarios.

Table 1
Comparison of computational complexity.

Algorithms	Addition	Multiplication	Exponentiation
MCC	$2L$	$2L+6$	1
NMCC	$2L+1$	$3L+7$	1
APMCC	$P^3+2LP^2-P^2$ $+LP-P+L$	$P^3+2LP^2+2P^2$ $LP+4P$	P
DRMCC	$LM+7M+6L-1$	$LM+6M+5L+36$	4

with the projection order and has the highest computational complexity among all algorithms. The computational complexity of the DRMCC algorithm is slightly higher than the MCC algorithm and the NMCC algorithm.

4. Parameter estimation

Given that $\sigma_{\bar{v}}^2$, σ_q^2, and σ_z^2 are generally unknown in real-world applications, it becomes necessary to undertake their estimation during operation.

Substituting Eq. (11) into Eq. (25) yields

$$\boldsymbol{\omega}_l = \boldsymbol{\omega}_{l-1} + \xi_{\omega,l}\bar{\boldsymbol{x}}_l\bar{v}_l + \xi_{\omega,l}\bar{\boldsymbol{x}}_l\bar{\boldsymbol{x}}_l^T\left(\boldsymbol{\omega}_{o,l} - \boldsymbol{\omega}_{l-1}\right). \tag{49}$$

Subtracting both sides of Eq. (49) from $\boldsymbol{\omega}_{o,l}$ and pre-multiplying $\bar{\boldsymbol{x}}_l^T$ on both sides of the result yields

$$\bar{\boldsymbol{x}}_l^T\left(\boldsymbol{\omega}_{o,l} - \boldsymbol{\omega}_l\right) = \left(1 - \xi_{\omega,l}\left\|\bar{\boldsymbol{x}}_l\right\|^2\right)\bar{\boldsymbol{x}}_l^T\left(\boldsymbol{\omega}_{o,l} - \boldsymbol{\omega}_{l-1}\right) - \xi_{\omega,l}\left\|\bar{\boldsymbol{x}}_l\right\|^2\bar{v} \tag{50}$$

When close to the steady state in the stationary scenario ($\sigma_q^2 = 0$), $\sigma_{\omega,l-1}^2$ is small enough so that $\xi_{\omega,l} \approx 0$. And in the nonstationary scenario, the parameter estimation of $\sigma_{\bar{v},l}^2$ will be discussed in the estimation of σ_q^2. To eliminate the unknown weight vector $\boldsymbol{\omega}_{o,l}$ in Eq. (50), we assume that $\mu_l = \xi_{\omega,l}\left\|\bar{\boldsymbol{x}}_l\right\|^2$ is small enough. Furthermore, since the performance of the algorithm is determined by the ratio between $\sigma_{\bar{v},l}^2$ and $\sigma_{z,l}^2$ [23], Eq. (50) can be rewritten as

$$\begin{aligned}\mu_l\bar{v} &\approx \bar{\boldsymbol{x}}_l^T\left(\boldsymbol{\omega}_{o,l} - \boldsymbol{\omega}_{l-1}\right) - \bar{\boldsymbol{x}}_l^T\left(\boldsymbol{\omega}_{o,l} - \boldsymbol{\omega}_l\right)\\ &= \bar{\boldsymbol{x}}_l^T\left(\boldsymbol{\omega}_l - \boldsymbol{\omega}_{l-1}\right)\end{aligned} \tag{51}$$

By computing the expectation of the squares of both sides of Eq. (51) [24], we derive

$$\sigma_{\bar{v},l}^2 = \frac{1}{\mu_l^2}E\left[\left[\bar{\boldsymbol{x}}_l^T\left(\boldsymbol{\omega}_l - \boldsymbol{\omega}_{l-1}\right)\right]^2\right]. \tag{52}$$

Using exponential moving average [25], $\sigma_{\bar{v},l}^2$ is represented as

$$\sigma_{\bar{v},l}^2 = \lambda_{\bar{v}}\sigma_{\bar{v},l-1}^2 + \left(1-\lambda_{\bar{v}}\right)\left[\bar{\boldsymbol{x}}_l^T\left(\boldsymbol{\omega}_l - \boldsymbol{\omega}_{l-1}\right)\right]^2 \tag{53}$$

where $\lambda_{\bar{v}}$ denotes the forgetting factor.

Substituting Eq. (3) into Eq. (28) yields

$$\hat{\boldsymbol{\alpha}}_l = \hat{\boldsymbol{\alpha}}_{l-1} + \xi_{\alpha,l}\boldsymbol{x}_{\alpha,l}z_l + \xi_{\alpha,l}\boldsymbol{x}_{\alpha,l}\boldsymbol{x}_{\alpha,l}^T(\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}}_{l-1}). \tag{54}$$

Similarly, an estimate of $\sigma_z(l)^2$ is written as

$$\sigma_{z,l}^2 = \lambda_z\sigma_{z,l-1}^2 + \left(1-\lambda_z\right)\left[\bar{\boldsymbol{x}}_{\alpha,l}^T\left(\hat{\boldsymbol{\alpha}}_l - \hat{\boldsymbol{\alpha}}_{l-1}\right)\right]^2 \tag{55}$$

where λ_z denotes the forgetting factor.

Based on Eq. (2), we arrive at

$$L\sigma_q^2 = E\left[\left\|\boldsymbol{\omega}_{o,l} - \boldsymbol{\omega}_{o,l-1}\right\|^2\right]. \tag{56}$$

Given that $\boldsymbol{\omega}_{o,l}$ and $\boldsymbol{\omega}_{o,l-1}$ are unknown, we replace them with $\boldsymbol{\omega}_l$ and $\boldsymbol{\omega}_{l-1}$ from the iterative process, i.e.,

$$L\sigma_q^2 = E\left[\left\|\boldsymbol{\omega}_l - \boldsymbol{\omega}_{l-1}\right\|^2\right]. \tag{57}$$

Therefore, σ_q^2 is given as

$$\sigma_{\omega,l}^2 = \lambda_q\sigma_{\omega,l-1}^2 + (1-\lambda_q)\left\|\boldsymbol{\omega}_l - \boldsymbol{\omega}_{l-1}\right\|^2, \tag{58}$$

$$\sigma_{q,l}^2 = \sigma_{\omega,l}^2/L, \tag{59}$$

where λ_q represents the forgetting factor. When the filter length L is long enough in a nonstationary scenario, $\sigma_{q,l}^2$ is small enough. Therefore, $\sigma_{\bar{v},l}^2$ can be similarly estimated by Eq. (53).

5. Numerical simulations

In this section, Monte Carlo (MC) simulations are performed to corroborate the theoretical model under both stationary and non-stationary scenarios. In addition, we evaluate the learning behavior of the developed DRMCC contrasted to other considered algorithms under non-Gaussian environments. In the simulation, the unknown weight vector $\boldsymbol{\omega}_{o,l}$ and the correlation indices $\boldsymbol{\alpha}$ are produced randomly. The correlated input signal $\boldsymbol{x}_l$ is produced according to Eq. (3), with both z_l and $\boldsymbol{q}_l$ being zero-mean Gaussian sequence and $\sigma_z^2 = 1$. The filter's weight vector is null vector. Furthermore, the simulation curves are displayed via averaging results from 50 independent MC trials.

5.1. Validation of theoretical model

First, we give the evolution of MSDs for the DRMCC algorithm in the stationary and non-stationary scenarios with $L = 10$ and $M = 2$. The noise v_l obeys the binary distribution over $\{-p, 1-p\}$, where the probability mass function is defined as $Pr(v_l = 1-p) = p$. Fig. 1 gives the transient behavior of the MSDs under the stationary and nonstationary scenarios, where the dotted line refers to the simulation results, and the solid lines depict transient behavior. As shown in Fig. 1, the theoretical curve resulted from Eq. (48) matches very well with the simulation results.

5.2. Performance comparison in system identification

Second, we assess the performance of the DRMCC algorithm against other algorithms in the stationary and nonstationary scenarios. The

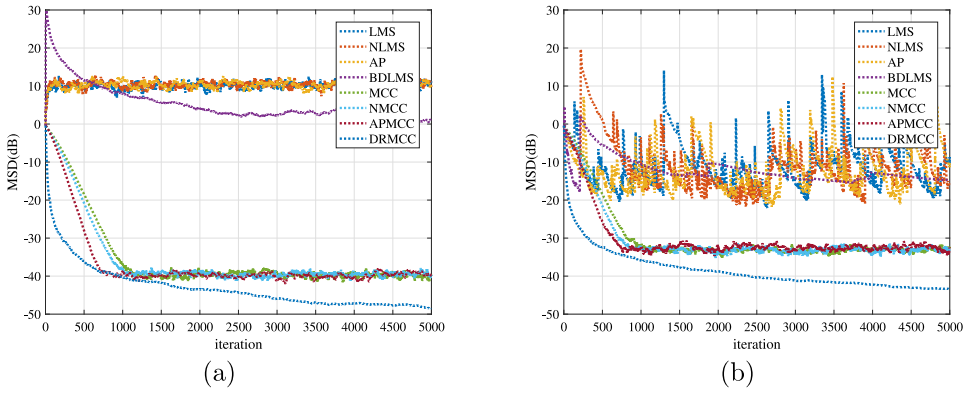

Fig. 2. Performance comparison under stationary scenario in (a) Bernoulli–Gaussian noise and (b) alpha-stable noise environments.

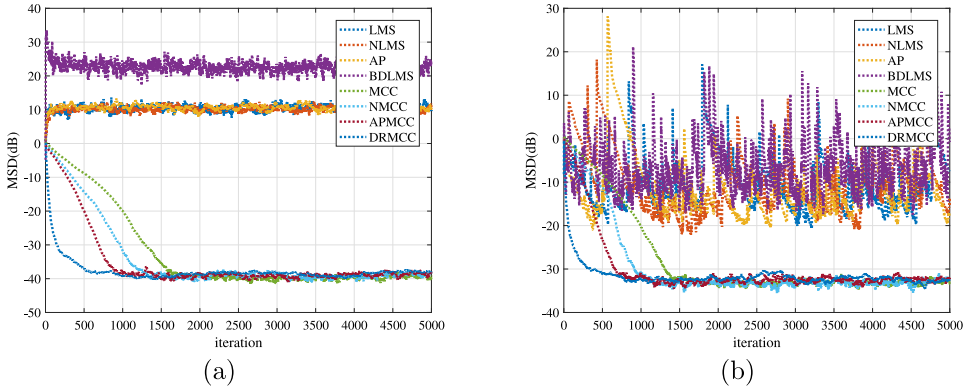

Fig. 3. Performance comparison under nonstationary scenario in (a) Bernoulli–Gaussian noise and (b) alpha-stable noise environments.

Table 2
Parameter selection for the used algorithm in system identification.

Algorithms	Step size	Kernel width	λ_z	λ_q	λ_b
LMS [2]	0.01	–	–	–	–
NLMS [27]	0.3	–	–	–	–
AP [8]	0.18	–	–	–	–
BDLMS [12]	–	–	0.9999	0.999999	0.999999
MCC [17]	0.01	1	–	–	–
NMCC [28]	0.3	1	–	–	–
APMCC [29]	0.18	1	–	–	–
DRMCC	–	$\sigma_1 = 1$, $\sigma_2 = 4$	0.9999	0.999999	0.999999

correlated input signal is produced by $x_l = tx_{l-1} + \sqrt{1-t^2}x_{l-2} + z_l$, where $t = 0.99$. The noise v_l is generated using the Bernoulli–Gaussian and alpha-stable models. The Bernoulli–Gaussian model is expressed as $v_l = b_l\mathcal{N}(0, 0.01) + (1 - b_l)\mathcal{N}(0, 10000)$, where $\mathcal{N}(0, \epsilon^2)$ represents the zero-mean Gaussian distribution with variance ϵ^2, and b_l indicates the binary distribution over $\{0, 1\}$ with the probability mass function defined as $Pr(v_l = 1) = 0.9$ [26]. The alpha-stable model is defined with tail parameter $\alpha = 1.2$ and dispersion parameter $\gamma = 0.1$. In the nonstationary scenario, the perturbation variance is set to $\sigma_q^2 = 10^{-7}$ for the Bernoulli–Gaussian environment and $\sigma_q^2 = 10^{-6}$ for the alpha-stable environment. The parameters of the competing algorithms are adjusted so that they have the same error in the nonstationary scenario. The used algorithm parameters are summarized in Table 2. The regularization parameters for the NLMS, NMCC, AP and APMCC algorithms are chosen as $\lambda = 20\sigma_x^2$, where σ_x^2 represents the variance of the input signal [9]. The projection order of AP and APMCC algorithms is 8.

The MSD convergence curves for the algorithms in various noise environments are illustrated in Figs. 2 and 3. As depicted in Figs. 2 and 3, the algorithms based on the MMSE criterion does not converge in these impulsive noise environments, while the algorithm based on

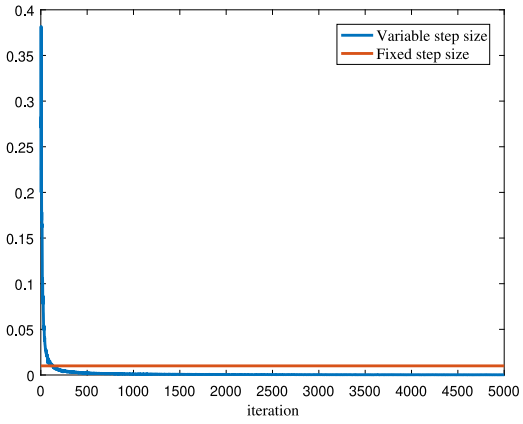

Fig. 4. Variable step-size curve under stationary scenario in Bernoulli–Gaussian noise environment.

the MCC criterion exhibits robust performance against impulsive noise. It is worth noting that the BDLMS algorithm performs worse in nonstationary scenario than the LMS, NLMS, and AP algorithms because the BDLMS algorithm in such scenario needs to estimate the parameter σ_q^2 associated with the filter weights as shown in Eq. (59). The weights of the BDLMS algorithm based on the MMSE criterion are susceptible to impulsive noise leading to inaccurate estimation of σ_q^2, which makes the algorithm perform worse than other algorithms. Since the weights of the DRMCC algorithm are insensitive to the impulsive noise, it still has better performance than the BDLMS algorithm in nonstationary scenario. It can be seen from Fig. 2 that the DRMCC algorithm outperforms the MCC, NMCC and APMCC algorithms in terms of convergence speed

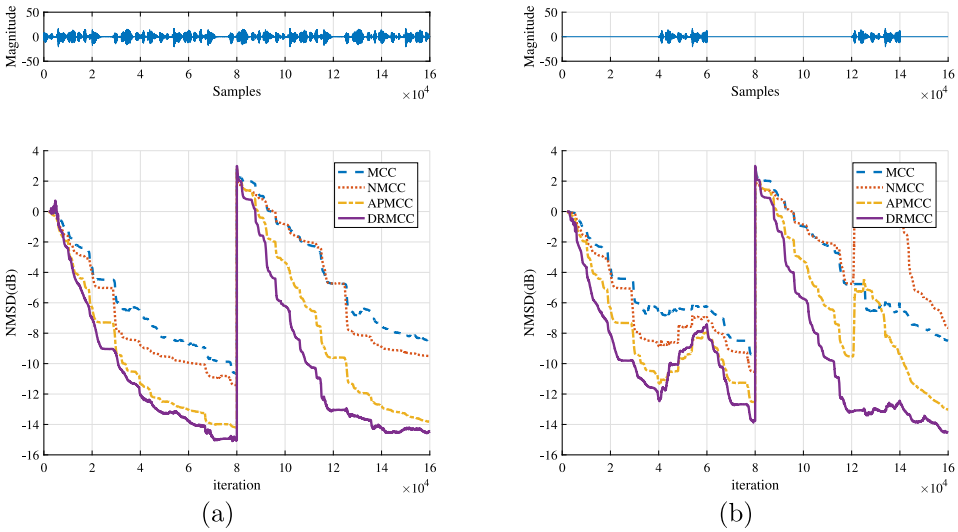

Fig. 5. Comparison of NMSD convergence curves of the algorithms under alpha-stable noise environment in (a) single-talk and (b) double-talk scenarios.

Table 3
Parameter selection for the used algorithm in AEC.

Algorithms	Step size	Kernel width	λ_z	λ_q	$\lambda_{\tilde{v}}$
MCC	0.00008	10	–	–	–
NMCC	0.6	10	–	–	–
APMCC	0.06	10	–	–	–
DRMCC	–	$\sigma_1 = 10,\ \sigma_2 = 10$	0.999996	0.9999	0.999999

and steady state error in the stationary scenario. As shown in Fig. 3, the DRMCC algorithm has a faster convergence rate in nonstationary scenario when the same steady state error is achieved than the MCC, NMCC and APMCC algorithms. In addition, the variable step size $\xi_{\omega,l}$ of the DRMCC algorithm and the value of the step size in the MCC under stationary scenario in Bernoulli–Gaussian noise environment are shown in Fig. 4. As can be seen in Fig. 4, the step size is larger in the initial phase and smaller as the algorithm approaches the steady state. Therefore, the proposed algorithm is effective in avoiding the tradeoff caused by constant step size.

5.3. Acoustic echo cancellation

Finally, we evaluate the DRMCC algorithm's performance under both single-talk and double-talk situations in AEC. We use the common normalized MSD (NMSD) as the evaluation metric, which can be defined as

$$\mathrm{NMSD}_l = 10\log_{10} E\left[\|\tilde{\boldsymbol{\omega}}_l\|^2/\|\boldsymbol{\omega}_o\|^2\right]. \tag{60}$$

The impulse response of the echo path $\boldsymbol{\omega}_o$ with length $L = 2048$ [9]. The sampling frequency and sampling time for both the far-end and near-end speech signals are 8 kHz and 20s, respectively. To evaluate the tracking performance of the algorithms, we simulate abrupt changes in the echo path by right-shifting $\boldsymbol{\omega}_o$ by 12 bits during the iteration process. The used algorithm parameters are summarized in Table 3. The regularization parameters for the AP and APMCC algorithms are chosen as $\lambda = 20\sigma_x^2$. The projection order of APMCC algorithms is 4. The order of the autoregressive model is chosen to be 2. The noise v_l is a dual source, including the zero-mean Gaussian noise with variance of 0.01 and the alpha-stable distribution noise with tail parameter $\alpha = 1.2$ and dispersion parameter $\gamma = 0.01$. The NMSD curves for all the algorithms under both single-talk and double-talk scenarios in alpha-stable noise environments are given in Fig. 5. The Fig. 5 illustrates that the DRMCC algorithm achieves the quicker adaption speed and maintains the higher estimation accuracy contrasted to other algorithms. It is noteworthy that the DRMCC algorithm exhibits insensitivity to double-talk interference.

6. Conclusions

This paper presents an observation model-based DRMCC algorithm for AEC in non-Gaussian noise environments. Furthermore, the corresponding convergence conditions and transient behavior in the mean-square sense are discussed. Finally, the parameter estimation method is given to solve the problem where the noise variance $\sigma_{\tilde{v}}^2$, the perturbation variances σ_q^2 and σ_z^2 are unknown. Simulation results demonstrate a good agreement between the theoretical model and experimental results in a non-Gaussian noise environment. It is noteworthy that the DRMCC algorithm significantly enhance the convergence performance in non-Gaussian noise environments compared with existing adaptive filtering algorithms.

CRediT authorship contribution statement

Xinyan Hou: Writing – review & editing, Writing – original draft, Validation, Methodology. **Haiquan Zhao:** Writing – review & editing, Methodology, Funding acquisition. **Xiaoqiang Long:** Writing – review & editing, Investigation.

Data availability

No data was used for the research described in the article.

Acknowledgments

This work was supported in part by National Natural Science Foundation of China under Grant 62171388, 61871461, 61571374.

References

[1] S. Haykin, Adaptive Filter Theory, Prentice-Hall, Englewood Cliffs, NJ, USA, 2002.
[2] A.H. Sayed, Fundamentals of Adaptive Filtering, Wiley, Hoboken, NJ, USA, 2003.
[3] K.-A. Lee, W.-S. Gan, S.M. Kuo, Subband adaptive filtering: theory and implementation, Wiley, Hoboken, NJ, USA, 2009.
[4] J.J. Jeong, S.H. Kim, G. Koo, S.W. Kim, Mean-square deviation analysis of multiband-structured subband adaptive filter algorithm, IEEE Trans. Signal Process. 64 (4) (2016) 985–994.
[5] H. Zhao, D. Liu, S. Lv, Robust maximum correntropy criterion subband adaptive filter algorithm for impulsive noise and noisy input, IEEE Trans. Circuits Syst. II 69 (2) (2022) 604–608.

[6] P. Park, C.H. Lee, J.W. Ko, Mean-square deviation analysis of affine projection algorithm, IEEE Trans. Signal Process. 59 (12) (2011) 5789–5799.
[7] J. Yoo, J. Shin, P. Park, Variable step-size affine projection sign algorithm, IEEE Trans. Circuits Syst. II 61 (4) (2014) 274–278.
[8] H.-C. Shin, A. Sayed, Mean-square performance of a family of affine projection algorithms, IEEE Trans. Signal Process. 52 (1) (2004) 90–102.
[9] S. Lv, H. Zhao, W. Xu, Robust widely linear affine projection M-estimate adaptive algorithm: Performance analysis and application, IEEE Trans. Signal Process. 71 (2023) 3623–3636.
[10] S. Zhang, H.C. So, W. Mi, H. Han, A family of adaptive decorrelation NLMS algorithms and its diffusion version over adaptive networks, IEEE Trans. Circuits Syst. I. Regul. Pap. 65 (2) (2018) 638–649.
[11] S. Zhang, J. Zhang, H.C. So, Low-complexity decorrelation NLMS algorithms: Performance analysis and AEC application, IEEE Trans. Signal Process. 68 (2020) 6621–6632.
[12] F. Huang, S. Zhang, J. Zhang, H. Chen, A.H. Sayed, Diffusion Bayesian decorrelation algorithms over networks, IEEE Trans. Signal Process. 71 (2023) 571–586.
[13] R. Pogula, T.K. Kumar, F. Albu, Robust sparse normalized LMAT algorithms for adaptive system identification under impulsive noise environments, Circuits Systems Signal Process. 38 (11) (2019) 5103–5134.
[14] J.C. Principe, Information theoretic learning: Renyi's entropy and kernel perspectives, Springer, New York, USA, 2010.
[15] W. Liu, P.P. Pokharel, J.C. Príncipe, Correntropy: Properties and applications in non-Gaussian signal processing, IEEE Trans. Signal Process. 55 (11) (2007) 5286–5298.
[16] H. Song, D. Ding, H. Dong, Q.-L. Han, Distributed maximum correntropy filtering for stochastic nonlinear systems under deception attacks, IEEE Trans. Cybern. 52 (5) (2022) 3733–3744.
[17] X. Hou, H. Zhao, X. Long, An innovative transient analysis of adaptive filter with maximum correntropy criterion, IEEE Signal Process. Lett. 29 (2022) 1689–1693.
[18] Y. Wang, Y. Li, F. Albu, R. Yang, Group-constrained maximum correntropy criterion algorithms for estimating sparse mix-noised channels, Circuits Systems Signal Process. 19 (8) (2017) 432.
[19] R. Izanloo, S.A. Fakoorian, H.S. Yazdi, D. Simon, Kalman filtering based on the maximum correntropy criterion in the presence of non-Gaussian noise, in: 2016 Annual Conference on Information Science and Systems, CISS, 2016, pp. 500–505.
[20] X. Hou, H. Zhao, X. Long, W. Jin, The kernel recursive maximum total correntropy algorithm, IEEE Trans. Circuits Syst. II 69 (12) (2022) 5139–5143.
[21] Y. Engel, S. Mannor, R. Meir, The kernel recursive least-squares algorithm, IEEE Trans. Signal Process. 52 (8) (2004) 2275–2285.
[22] J. Chen, C. Richard, A.H. Sayed, Diffusion LMS over multitask networks, IEEE Trans. Signal Process. 63 (11) (2015) 2733–2748.
[23] S. Akhlaghi, N. Zhou, Z. Huang, Adaptive adjustment of noise covariance in Kalman filter for dynamic state estimation, in: 2017 IEEE Power & Energy Society General Meeting, 2017, pp. 1–5.
[24] J.G. Proakis, Digital Communications, Prentice-Hall, Englewood Cliffs, NJ, USA, 2001.
[25] L. Shi, H. Zhao, W. Wang, L. Lu, Combined regularization parameter for normalized LMS algorithm and its performance analysis, Signal Process. 162 (2019) 75–82.
[26] J. Rong, J. Zhang, H. Duan, Robust sparse Bayesian learning based on the Bernoulli-Gaussian model of impulsive noise, Digit. Signal Process. 136 (2023) 104013.
[27] H.-C. Huang, J. Lee, A new variable step-size NLMS algorithm and its performance analysis, IEEE Trans. Signal Process. 60 (4) (2012) 2055–2060.
[28] W. Ma, D. Zheng, Y. Li, Z. Zhang, B. Chen, Bias-compensated normalized maximum correntropy criterion algorithm for system identification with noisy input, Signal Process. 152 (2018) 160–164.
[29] C. Qiu, G. Qian, S. Wang, Widely linear maximum complex correntropy criterion affine projection algorithm and its performance analysis, IEEE Trans. Signal Process. 70 (2022) 3540–3550.