SHORT PAPER

Adaptive Filtering Based on Minimum Error Entropy Conjugate Gradient

Guoliang Li[1] · **Qi Sun[1]** · **Ying Zhang[2]** · **Gang Wang[1]** · **Xinyue Yang[1]** · **Linqiang Ouyang[1]**

Abstract

Based on the criterion of minimum error entropy, this paper proposes a novel conjugate gradient algorithm, called MEE-CG. This algorithm has robust performance under non-Gaussian interference. Theoretical analysis and experimental results demonstrate that the proposed algorithm displays more robust performance than the conventional conjugate gradient methods on the basis of the mean square error and the maximum correntropy criterion. Compared with the stochastic gradient minimum error entropy algorithm and the recursive minimum error entropy algorithm, the proposed algorithm provides a trade-off between computational complexity and convergence speed.

✉ Ying Zhang
zhangyingxf@uestc.edu.cn

Guoliang Li
lgl-flash@qq.com

Qi Sun
sunqi_uestc@126.com

Gang Wang
wanggang_hld@uestc.edu.cn

Xinyue Yang
919051369@qq.com

Linqiang Ouyang
787977244@qq.com

[1] School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China

[2] School of Resources and Environment, University of Electronic Science and Technology of China, Chengdu 611731, China

1 Introduction

Over the past decades, the rapid development in adaptive filtering (AF) algorithms is regarded as a significance for practical applications [20]. The AF algorithms have been widely adopted in the field of signal processing including noise cancellation [11], channel equalization [8] and system identification [30]. The fundamental AF algorithms cover the famous least mean squares (LMS) [9], recursive least squares (RLS) [22] and conjugate gradient (CG) algorithm [29]. The LMS with the fixed step size always presents slow convergent speed [19]. Although the conventional RLS algorithm has a faster convergent speed, it suffers from higher computation complexity and numerical instability [25]. Comparing with the well-known RLS and LMS algorithms, the CG algorithm provides a trade-off between computational complexity and convergence speed [31]. Because of the variable step length according to the real-time input and the avoidance for calculating matrix inversion during derivation, the CG algorithm exhibits comparable performance in convergence with a smaller burden in computational complexity than the RLS algorithm [7].

The mean square error (MSE) acts as one of the most popular criteria in AF algorithms, which works well under Gaussian noises. However, signals will be most likely contaminated by non-Gaussian noise in various physical applications, including wireless channel tracking [18], multipath estimation [12] and acoustic echo cancellation [15]. Under these circumstances, the performance of the adaptive filter under the MSE criterion may work poorly [3, 5]. Therefore, combining with the Information Theoretic Learning (ITL) [23], various AF algorithms have been proposed to cope with inference from the non-Gaussian noise. The maximum correntropy criterion (MCC) [17, 24, 27] and minimum error entropy (MEE) [2, 4, 14], acting as the two typical examples for ITL, are insensitive to larger outliers, and improve the performance under impulsive noise. Several algorithms based on the MCC criterion have been proposed, such as the LMS-type algorithm [16, 21, 26], the RLS-type algorithm named the recursive maximum correntropy (RMC) [32] and the CG-type algorithm called the q-Gaussian MCC-CG [6]. Recent research has shown that the MEE criterion presents more robust performance than the MCC [10], and has been effectively utilized in the LMS-type algorithm [26], RLS-type algorithm named the recursive minimum error entropy (RMEE) [23] and the Kalman filter called the MEE-KF [2]. To the best of our knowledge, the CG has not been applied to the MEE criterion.

Under the MEE criterion, the LMS-type algorithm has slow convergent speed or poor steady-state error [14], and the RLS-type algorithm, or the RMEE, has improved both the convergent speed and the steady state error at the cost of a high computational complexity [23]. In this paper, we derive an CG-typed algorithm based on the MEE criterion, called MEE-CG. It can be expected that the MEE-CG algorithm will achieve a trade-off between computational complexity and convergence speed, compared with the LMS-type and RLS-type algorithms based on the MEE criterion.

For comparison, the LMS-type, CG-type [6] and RLS-type [32] algorithm under the MCC criterion are called MCC-LMS, MCC-CG and MCC-RLS, respectively. The LMS-type, CG-type and RLS-type [23] algorithm under the MEE criterion are called MEE-LMS, MEE-CG (proposed in this paper) and MEE-RLS, respectively.

The organization of the paper is structured as follows: in Sect. 2, the problem is addressed; in the following Sect. 3, we derive the MEE-CG algorithm; the analyses on the convergence and computational complexity are given in Sect. 4 and then experimental simulations for system identification are provided in Sect. 5; in the final Sect. 6, we summarize the paper and obtain the conclusion.

2 Problem Statement

2.1 Adaptive Filtering Theory

The block diagram of the basic adaptive filter is shown in Fig. 1. When taking the adaptive filtering theory into consideration, the desired response $d \in \mathbf{R}^1$ are generated via an input $\mathbf{u} \in \mathbf{R}^M$ at instant n

$$d(n) = \mathbf{w}_o^T \mathbf{u}(n) + v(n), \tag{1}$$

where $\mathbf{w}_o \in \mathbf{R}^M$ denotes the unknown coefficient, and $v(n)$ represents the zero-mean observation noise with variance σ_v^2. The estimation error can be represented as

$$\begin{aligned} e(n) &= d(n) - \mathbf{w}^T(n-1)\mathbf{u}(n) \\ &= d(n) - y(n), \end{aligned} \tag{2}$$

where $y(n) = \mathbf{w}^T(n-1)\mathbf{u}(n)$, and $\mathbf{w}(n-1)$ accounts for the estimation of $\mathbf{w}_o$ at instant $n-1$. For simplification, we make the assumptions as follows: 1) The additive noise is white, and we have

$$E\{v(m)v(n)\} = 0, m \neq n. \tag{3}$$

2) The inputs $\mathbf{u}(n)$ is composed of a zero-mean white sequence

$$E\left\{\mathbf{u}^T(m)\mathbf{u}(n)\right\} = E\left\{\mathbf{u}(m)\mathbf{u}^T(n)\right\} = 0, m \neq n. \tag{4}$$
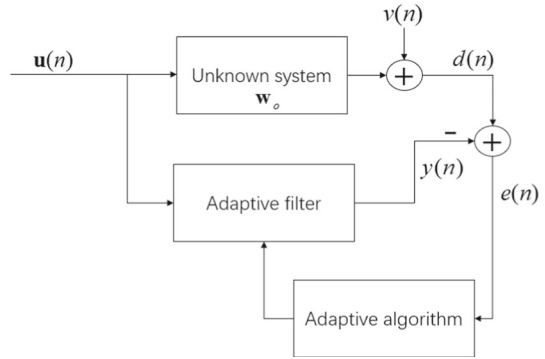
3) The inputs are uncorrelated with the additive noise at moments (m, n)

$$E\left\{\mathbf{u}^H(m)v(n)\right\} = 0, m \neq n. \tag{5}$$

2.2 Minimum Error Entropy (MEE) Criterion

From the Information Theoretic Learning (ITL), we can obtain the empirical version of the quadratic information potential [14, 26] for filters with sliding window in length L

Fig. 1 Block diagram of the basic adaptive filter

$$V(e) = \frac{1}{L^2} \sum_{i=1}^{L} \sum_{j=1}^{L} \lambda^{i+j} G_\sigma \left(e(i) - e(j) \right), \tag{6}$$

where G_σ stands for the Gaussian kernel function with bandwidth σ, and λ $(0 < \lambda \leq 1)$ denotes the forgetting factor

$$G_\sigma(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{x^2}{2\sigma^2} \right). \tag{7}$$

Assume that the error is a random variable with probability density function and an estimator of Renyi's quadratic entropy [14] for the error can be written as

$$R_q(e) = \log \frac{1}{V(e)}. \tag{8}$$

According to ITL, minimizing the error entropy is equivalent to maximizing the formula (8). Therefore, based on the MEE [2, 4, 14], the cost function can be obtained as

$$J_{MEE}(\mathbf{w}) = \frac{1}{L^2} \sum_{i=1}^{L} \sum_{j=1}^{L} \lambda^{i+j} G_\sigma \left(e(i) - e(j) \right). \tag{9}$$

We get the gradient of (9) as

$$\nabla J_{MEE}(\mathbf{w})$$

$$= \frac{1}{L^2\sigma^2} \sum_{i=1}^{L} \sum_{j=1}^{L} \lambda^{i+j} \left(\mathbf{u}(i) - \mathbf{u}(j) \right) G_\sigma \left(e(i) - e(j) \right) \left(e(i) - e(j) \right). \tag{10}$$

For each instant n, we simplify the following expressions

$$
\begin{cases}
\mathbf{u}_i = \mathbf{u}\,(n - i)\,, \\
d_i = d\,(n - i)\,, \\
e_i = e\,(n - i)\,, i = 0, 1, \cdots, L - 1.
\end{cases}
\tag{11}
$$

Thus, we are able to change the form of the following expressions

$$
\begin{cases}
\mathbf{U}_L = \begin{bmatrix} \mathbf{u}_0, \mathbf{u}_1, \ldots, \mathbf{u}_{L-1} \end{bmatrix} = \begin{bmatrix} \mathbf{u}_0, \mathbf{U}_{L-1} \end{bmatrix}, \\
\mathbf{D}_L = \begin{bmatrix} d_0, d_1, \ldots, d_{L-1} \end{bmatrix}^T = \begin{bmatrix} d_0, \mathbf{D}_{L-1} \end{bmatrix}^T, \\
\varepsilon_L = \begin{bmatrix} e_0, e_1, \ldots, e_{L-1} \end{bmatrix}^T = \begin{bmatrix} e_0, \varepsilon_{L-1} \end{bmatrix}^T.
\end{cases}
\tag{12}
$$

So (10) can be rewritten as

$$
\begin{aligned}
&\nabla J_{MEE}\,(\mathbf{w}) \\
&= \frac{1}{L^2 \sigma^2} \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} \lambda^{i+j}\,(\mathbf{u}_i - \mathbf{u}_j)\,G_\sigma\,(e_i - e_j)\,(e_i - e_j) \\
&= \frac{2}{L^2 \sigma^2} \mathbf{U}_L (\mathbf{P}_L - \mathbf{Q}_L) \varepsilon_L \\
&= \frac{2}{L^2 \sigma^2} \mathbf{U}_L \Phi_L \varepsilon_L.
\end{aligned}
\tag{13}
$$

Then we obtain

$$
\begin{cases}
[\mathbf{Q}]_{(i+1)(j+1)} = \lambda^{i+j} G_\sigma\,(e_i - e_j)\,, i, j = 0, 1, \ldots, L - 1 \\
[\mathbf{P}]_{(i+1)(j+1)} = \begin{cases} \displaystyle\sum_{k=0}^{L-1} \lambda^{i+k} G_\sigma\,(e_i - e_k), i = j \\ 0, i \neq j \end{cases} \\
[\varphi]_i = \lambda^i G_\sigma\,(e_i - e_0)\,, i = 1, \ldots, L - 1 \\
\phi_0 = \displaystyle\sum_{k=1}^{L-1} \lambda^k G_\sigma\,[e_0 - e_k].
\end{cases}
\tag{14}
$$

Through expressions (12–14), we can obtain the Φ_L (see the formula (11) in [26] for details) of the objective function under the MEE criterion and its recursive method.

3 Proposed Minimum Error Entropy Conjugate Gradient

Through the derivations in Sect. 2, we have

$$
\begin{cases}
\mathbf{R}_L = \mathbf{U}_L \Phi_L \mathbf{U}_L^T, \\
\mathbf{r}_L = \mathbf{U}_L \Phi_L \mathbf{D}_L.
\end{cases}
\tag{15}
$$

From (12), (14) and (15), we have

$$
\begin{cases}
\mathbf{R}_L = \lambda^2 \mathbf{U}_{L-1} \Phi_{L-1} \mathbf{U}_{L-1}^T \\
\quad + \left[\mathbf{u}_0 \phi_0 \mathbf{u}_0^T + \mathbf{U}_{L-1} \varphi \mathbf{u}_0^T + \mathbf{u}_0 \varphi^T \mathbf{U}_{L-1}^T \right], \\
\mathbf{r}_L = \lambda^2 \mathbf{U}_{L-1} \Phi_{L-1} \mathbf{D}_{L-1}^T \\
\quad + \left[\mathbf{u}_0 \phi_0 d_0^T + \mathbf{U}_{L-1} \varphi d_0^T + \mathbf{u}_0 \varphi^T \mathbf{D}_{L-1}^T \right].
\end{cases}
\tag{16}
$$

After applying the attenuation window to the correlation function for the data matrix in the CG, we achieve the same expressions as the RLS-type algorithm [1, 23]

$$
\begin{cases}
\mathbf{R}_L = \lambda^2 \mathbf{R}_{L-1} + \mathbf{u}_0 \phi_0 \mathbf{u}_0^T, \\
\mathbf{r}_L = \lambda^2 \mathbf{r}_{L-1} + \mathbf{u}_0 \phi_0 d_0.
\end{cases}
\tag{17}
$$

The online CG method aims to minimize the following cost function [7]

$$
\min F\left(\mathbf{w}_L \right) = \min \left(\frac{1}{2} \mathbf{w}_L^T \mathbf{R}_L \mathbf{w}_L - \mathbf{r}_L^T \mathbf{w}_L \right),
\tag{18}
$$

and the solution of (18) is

$$
\mathbf{R}_L \mathbf{w}_L = \mathbf{r}_L.
\tag{19}
$$

The weight vector $\mathbf{w}_L$ and the direction vector $\mathbf{p}_L$ can be updated as

$$
\mathbf{w}_L = \mathbf{w}_{L-1} + \alpha_L \mathbf{p}_L,
\tag{20}
$$

$$
\mathbf{p}_{L+1} = \mathbf{g}_L + \beta_L \mathbf{p}_L,
\tag{21}
$$

where step factor α_L and direction factor β_L will update for each iteration and $\mathbf{g}_L$ denotes residual vector.

Substituting (17) into (20) and (21), the residual vector $\mathbf{g}_L$ for the online CG method is

$$
\begin{aligned}
\mathbf{g}_L &= \mathbf{r}_L - \mathbf{R}_L \mathbf{w}_L \\
&= \lambda^2 \mathbf{g}_{L-1} - \alpha_L \mathbf{R}_L \mathbf{p}_L + \mathbf{u}_0 \phi_0 (d_0 - \mathbf{u}_0^H \mathbf{w}_{L-1}).
\end{aligned}
\tag{22}
$$

According to the conjugate properties for $\mathbf{p}_L$ of the CG, we can get

$$
\begin{aligned}
\mathbf{p}_{L+1}^T \mathbf{R}_L \mathbf{p}_L &= (\mathbf{g}_L + \beta_L \mathbf{p}_L)^T \mathbf{R}_L \mathbf{p}_L \\
&= \mathbf{g}_L^T \mathbf{R}_L \mathbf{p}_L + \beta_L \mathbf{p}_L^T \mathbf{R}_L \mathbf{p}_L \\
&= 0,
\end{aligned}
\tag{23}
$$

and then the expression of β_L can be derived as [1]

$$\beta_L = -\frac{\mathbf{g}_L^T \mathbf{R}_L \mathbf{p}_L}{\mathbf{p}_L^T \mathbf{R}_L \mathbf{p}_L}. \tag{24}$$

According to the definition of the step factor α_L in (20), substituting (20) into the objective function $F(\mathbf{w}_L)$ in (18) yields

$$F(\mathbf{w}_L) = F(\mathbf{w}_{L-1} + \alpha_L \mathbf{p}_L). \tag{25}$$

Setting the derivative of (18) to zero, α_L in (25) becomes

$$\alpha_L = \frac{\mathbf{r}_L^T \mathbf{p}_L - \mathbf{w}_{L-1}^T \mathbf{R}_L \mathbf{p}_L}{\mathbf{p}_L^T \mathbf{R}_L \mathbf{p}_L}. \tag{26}$$

According to the derivations above, the detailed description for the MEE-CG algorithm is given in **Algorithm 1**.

Algorithm 1 Proposed MEE-CG algorithm.

Initialize: $\mathbf{w}(0) = \mathbf{0}$, $\mathbf{p}(1) = \mathbf{g}(0) = \mathbf{r}(0) = \mathbf{u}(1)\phi(1)d(1)$.
Iterate:
1: $e(n) = d(n) - \mathbf{w}^T(n-1)\mathbf{u}(n)$,
2: $\phi(n) = \sum\limits_{k=1}^{L-1} \lambda^k G_\sigma[e(n) - e(n-k)]$,
3: $\mathbf{R}(n) = \lambda^2 \mathbf{R}(n-1) + \mathbf{u}(n)\phi(n)\mathbf{u}(n)^T$,
4: $\mathbf{r}(n) = \lambda^2 \mathbf{r}(n-1) + \mathbf{u}(n)\phi(n)d(n)$,
5: $\mathbf{q}(n) = \mathbf{R}(n)\mathbf{p}(n)$,
6: $\alpha(n) = \left(\mathbf{r}(n)^T \mathbf{p}(n) - \mathbf{w}(n-1)^T \mathbf{q}(n)\right) \Big/ \mathbf{p}(n)^T \mathbf{q}(n)$,
7: $\mathbf{w}(n) = \mathbf{w}(n-1) + \alpha(n)\mathbf{p}(n)$,
8: $\mathbf{g}(n) = \lambda^2 \mathbf{g}(n-1) - \alpha(n)\mathbf{q}(n) + \mathbf{u}(n)\phi(n)e(n)$,
9: $\beta(n) = -(\mathbf{g}(n)^T \mathbf{q}(n))/(\mathbf{p}(n)^T \mathbf{q}(n))$,
10: $\mathbf{p}(n+1) = \mathbf{g}(n) + \beta(n)\mathbf{p}(n)$.

---

4 Performance Analyses

4.1 Mean Value Behavior

Assuming the input signal to be ergodic and wide-sense stationary, $\alpha(n)$, $\beta(n)$, $p(n)$, $\mathbf{w}(n)$, $\mathbf{g}(n)$, $\mathbf{r}(n)$ and $\mathbf{R}(n)$ are independent with each other, and then we have $E[\alpha(n)] = \bar{\alpha}$, $E[\beta(n)] = \bar{\beta}$, $E[\mathbf{r}(n)] = r$, and $E[\mathbf{R}(n)] = R$ for simplicity. After applying Z-transform to the proposed MEE-CG, to ensure the stability of the

described system, we obtain the same conclusion about the range of α_L and $\bar{\beta}$ in [1]

$$\begin{cases} -1 \le \bar{\beta} + \bar{\alpha}\lambda_i - \bar{\beta} - 1, \\ -1 \le \bar{\beta} - \bar{\alpha}\lambda_i + \bar{\beta} + 1, \Rightarrow 0 \le \bar{\alpha} \le \dfrac{2\bar{\beta} + 2}{\lambda_{\max}}. \\ -1 \le \bar{\beta} \le 1. \end{cases} \tag{27}$$

where λ_i is the ith eigenvalue of R. When $\bar{\beta} \to 0$, we have $0 \le \bar{\alpha} \le 2\lambda_{\max}^{-1}$, which also supports the steepest descent algorithm in [9].

Besides, after M iterations, we have $\tilde{\mathbf{R}}(n)\tilde{\mathbf{w}}(n) \approx \tilde{\mathbf{r}}(n)$, where $\tilde{\mathbf{x}}$ represents the estimation for $\mathbf{x}$. Hence, the norm of $\mathbf{g}(M)$ satisfies

$$\| \mathbf{g}(M) \| = \| \tilde{\mathbf{r}}(n) - \tilde{\mathbf{R}}(n)\tilde{\mathbf{w}}(M) \| < \varepsilon, \tag{28}$$

where ε can be an arbitrary small value. According to [1], we get the bound for the norm of $\| \mathbf{w}_o - \mathbf{w}(n) \|$ as follows

$$\| \mathbf{w}_o - \mathbf{w}(n) \|_{\mathbf{R}_1} \le 2 \| \mathbf{w}_o - \mathbf{w}(0) \|_{\mathbf{R}_1} \left(\frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1} \right)^n \tag{29}$$

where $\mathbf{R}_1 = \tilde{\mathbf{R}}(n)$ is positive define, $\|\mathbf{w}\|_{\mathbf{R}_1} = \sqrt{\mathbf{w}^T \mathbf{R}_1 \mathbf{w}}$, and the condition number is defined as $\kappa = \|\mathbf{R}\|_2 \|\mathbf{R}^{-1}\|_2$. After taking the expectations on both side of (29), we can reach the conclusion that MEE-CG is convergent in mean value.

4.2 Mean Square Behavior

After reaching the steady state, we have $\mathbf{w}(n) \approx \mathbf{w}(n-1)$. From (20), we have $\alpha(n)\mathbf{p}(n) \approx \mathbf{0}$. Then multiplying $\alpha(n)$ with (26), we can have $\alpha(n) \approx 0$, and thus $\mathbf{g}(n) \approx \mathbf{g}(n-1)$. According to (24), $\beta(n) \approx 0$, we can obtain $\mathbf{p}(n+1) \approx \mathbf{g}(n)$ from (21). Taking this equation into (26), one can obtain $\mathbf{g}(n) \approx \mathbf{0}$. Since $\mathbf{g}(n) = \mathbf{r}(n) - \mathbf{R}(n)\mathbf{w}(n) \approx \mathbf{r}(n) - \mathbf{R}(n)\mathbf{w}(n-1)$, we can easily get $\mathbf{w}(n) \approx \mathbf{R}(n)^{-1}\mathbf{r}(n)$, which is the same as the MEE-RLS algorithm. Therefore, it can be concluded that the steady state behavior for both CG and RLS algorithms are equivalent [7]. Then, according to [23], the total weighted error in steady state for the proposed MEE-CG is the same as the RMEE algorithm

$$\lim_{n \to \infty} E \left\| \tilde{\mathbf{w}}(n) \right\|_2^2 \approx \frac{1 - \lambda^2}{1 + \lambda^2} M \sigma_u^{-2} E \left\{ \mathbf{v}^2(n)\phi(n)^2 \right\} E^{-2} \left\{ \phi(n) \right\}. \tag{30}$$

The following numerical simulations also verify the correctness of the theoretical values.

4.3 Computational Complexity

Table 1 displays the computational complexities of the MSE-RLS, MSE-CG, MCC-RLS, MCC-CG, MEE-RLS and MEE-CG per iteration. Table 1 shows that the

Table 1 Computational complexities of the MSE-RLS,MSE-CG, MCC-RLS, MCC-CG, MEE-RLS and MEE-CG per iteration

Algorithms	$\times/\div$	$\pm$	Exponents
MSE-RLS	$4M^2 + 3M + 2$	$4M^2$	0
MSE-CG	$3M^2 + 13M + 2$	$2M^2 + 10M - 4$	0
MCC-RLS	$4M^2 + 3M + 5$	$4M^2$	1
MCC-CG	$3M^2 + 15M + 2$	$2M^2 + 10M - 4$	1
MEE-RLS	$4M^2 + 3M + 5$ $+(L+4)(L-1)/2$	$4M^2 + 2L - 3$	$L - 1$
MEE-CG	$3M^2 + 15M + 2$ $+(L+4)(L-1)/2$	$2M^2 + 10M$ $+2L - 7$	$L - 1$

Fig. 2 Transient MSDs (dB) of the MCC-CG for different parameters values of λ^2 when $\sigma = 1$
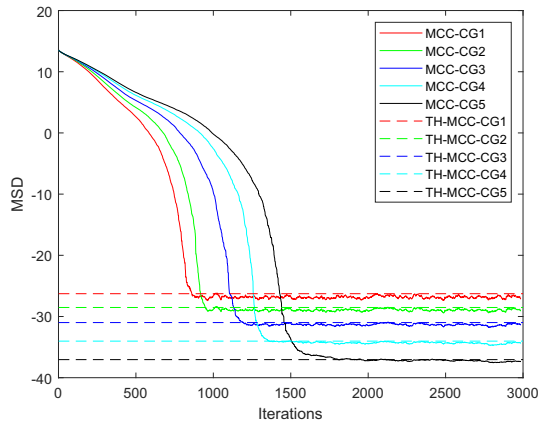
computational complexity of the MSE-CG, MCC-CG, and MEE-CG is lower than that of the MSE-RLS, MCC-RLS, and MEE-RLS, respectively.

Sections 4.2 and 4.3 show that the proposed MEE-CG algorithm are capable of achieving the same steady error as the MEE-RLS with lower burden in computation than the MEE-RLS.

5 Experimental Results

In this section, the simulations will be carried out to prove the theoretical analysis and verify the superiority of the proposed MEE-CG algorithm. We gain all the simulations from the average of independent 100 Monte Carlo trials and the performances of the algorithms are evaluated by the steady-state mean-square deviation (MSD)

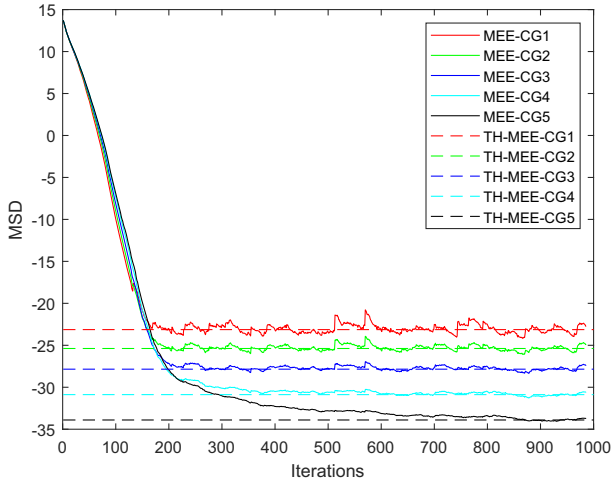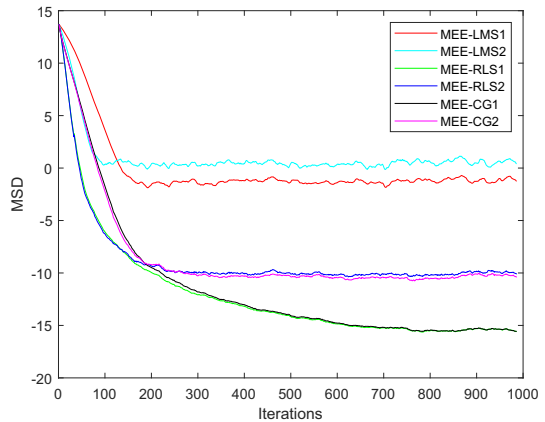$$\text{MSD} = \lim_{n \to \infty} E\left\{\|\mathbf{w}_o - \mathbf{w}(n)\|_2^2\right\}. \tag{31}$$

Fig. 3 Transient MSDs (dB) of the MEE-CG algorithm for different parameters values of λ^2 when $\sigma = 1$

Fig. 4 Transient MSDs (dB) of MEE-CG and other MEE-based algorithms

In the simulations, the unknown parameter $\mathbf{w}_o$ is chosen to be a vector with the size 12×1, and we set the sliding window in the length of $L=15$. We choose the white Gaussian random sequences as inputs $\mathbf{u}$ where the covariance matrix $E\{\mathbf{u}\mathbf{u}^T\}=\mathbf{I}_5$ and $E\{\mathbf{u}^T\mathbf{u}\}=5$. The additive impulsive noise is $\upsilon(n) \sim 0.94N(0, 0.01) + 0.06N(0, 225)$. Figure 1 compares the proposed MEE-CG under $\sigma=1$. The λ^2 parameters for MEE-CG1, MEE-CG2, MEE-CG3, MEE-CG4 and MEE-CG5 denote 0.942, 0.965, 0.980, 0.990 and 0.995 respectively. The corresponding theoretical values (30) are TH-MEE-CG1, TH-MEE-CG2, TH-MEE-CG3, TH-MEE-CG4 and TH-MEE-CG5 respectively.

Figure 2 shows that the convergence rate of the MCC-CG algorithm of these five sets of parameters has proportional to the value of λ, and their steady-state errors are inversely proportional to the value of λ. The theoretical values for steady-state error

will be increased under the decrease of the forgetting factor λ as well as the increase of the bandwidth σ.

Figure 3 shows that the convergence rate of the MEE-CG algorithm of these five sets of parameters has proportional to the value of λ, and their steady-state errors are inversely proportional to the value of λ. The theoretical values for steady-state error will be increased when decreasing the forgetting factor λ or increasing the bandwidth σ. And, the convergence rates of the MCC-CG and MEE-CG can be found from Figs. 2 and 3, respectively. One can see that the MEE-CG1 reaches the same MSD for about 800 iterations than the MCC-CG1 and the MEE-CG4 is 400 iterations faster than the MCC-CG4 in Figs. 2 and 3. Thus, the proposed MEE-CG algorithm achieves faster convergence speed than the MCC-CG. When the kernel width sets within an appropriate range through the experimental results and related researches, the MEE-type algorithms can have good filtering performance and the kernel width can be chosen by cross-validation or trial and error methods in practical applications [2, 4, 14, 26]. In this paper, $\sigma = 1$ is a good choice to ensure the performance of the proposed algorithm by trial and error methods.

Then, the comparison between MEE-LMS1 ($\mu = 0.2$), MEE-LMS2 ($\mu = 0.4$), MEE-RLS1 [23] $\left(\lambda^2 = 0.990\right)$, MEE-RLS2 [23] $\left(\lambda^2 = 0.965\right)$, MEE-CG1 $\left(\lambda^2 = 0.990\right)$ and MEE-CG2 $\left(\lambda^2 = 0.965\right)$ are given in Fig. 4. The proposed MEE-CG has a smaller MSD value and a faster convergence rate than MEE-LMS algorithm. Under the same λ^2 values, it has the same MSD as MEE-RLS, and the convergence rate of MEE-CG becomes slightly slower than the MEE-RLS.

6 Conclusion

We have proposed a new MEE-CG method based on MEE criterion, which has the same theoretical steady state error as that of the MEE-RLS. The numerical simulations reveal that when coping with non-Gaussian noise, the MEE-CG has faster convergence speed than MEE-LMS algorithm and smaller steady state error than MCC-CG algorithm. In addition, the MEE-CG has comparable performance with the MEE-RLS, and its computational complexity is lower than the MEE-RLS.

Declarations

Conflict of interest The author declares no Conflict of interest.

References

1. P.S. Chang, A.N. Willson, Analysis of conjugate gradient algorithms for adaptive filtering. IEEE Trans. Signal Process. **48**(2), 409–418 (2000)
2. B. Chen, L. Dang, Y. Gu, N. Zheng, J.C. Principe, Minimum error entropy Kalman filter. IEEE Trans. Syst. Man. Cybern. Syst. **51**(9), 5819–5829 (2021)
3. B. Chen, J.S. Zhao, P. Zhu, J.C. Principe, Quantized Kernel least mean square algorithm. IEEE Trans. Neural Networks Learn. Syst. **23**(1), 22–32 (2012)
4. B. Chen, Y. Zhu, J. Hu, Mean−square convergence analysis of Adaline training with minimum error entropy criterion. IEEE Trans. Neural Netw. **21**(7), 1168–1179 (2010)
5. B. Chen, Y. Zhu, J. Hu, J.C. Principe, *System Parameter Identification: Information Criteria and Algorithms* (Newnes, Oxford, 2013)
6. T. Chen, K. Xiong, H. Xie, G. Li, S. Wang, q-Gaussian Maximum Correntropy Adaptive Filter with Conjugate Gradient Method. In: *2019 Chinese Automation Congress (CAC)* (IEEE, 2019), pp. 1164–1168
7. P.S.R. Diniz, M.O.K. Mendonca, J.O. Ferreira, T.N. Ferreira, Data-Selective Conjugate Gradient Algorithm. In: *2018 26th European Signal Processing Conference (EUSIPCO)* (IEEE, 2018), pp. 707–711
8. Y. Ge, W. Zhang, H. Wang, On adaptive length of temporal filter for space-time equalization with cochannel interference. IEEE Signal Process. Lett. **25**(7), 999–1003 (2018)
9. S. Haykin, *Adaptive Filter Theory* (Prentice-Hall, Hoboken, 2020)
10. A.R. Heravi, G.A. Hodtani, Where does minimum error entropy outperform minimum mean square error? A new and closer look. IEEE Access **6**, 5856–5864 (2018)
11. E.P. Jayakumar, P.S. Sathidevi, An integrated acoustic echo and noise cancellation system using cross-band adaptive filters and wavelet thresholding of multitaper spectrum. Appl. Acoust. **141**, 9–18 (2018)
12. C. Lan, H. Yue, Y. Xing, M. Ren, Multipath estimation based on modified epsilon-constrained rank-based differential evolution with minimum error entropy. IEEE Access **6**, 61569–61582 (2018)
13. L. Li, X. Zhao, Variable step-size LMS algorithm based on hyperbolic tangent function. Circuits Syst. Signal Process. **42**, 4415–4431 (2023)
14. Z. Li, L. Xing, B. Chen, Adaptive filtering with quantized minimum error entropy criterion. Signal Process. **172**, 107534 (2020)
15. W. Liu, I. Park, J.C. Principe, An information theoretic approach of designing sparse Kernel adaptive filters. IEEE Trans. Neural Networks **20**(12), 1950–1961 (2009)
16. W. Liu, P. Pokharel, J.C. Principe, Correntropy: properties and applications in non-Gaussian signal processing. IEEE Trans. Signal Process. **55**(11), 5286–5298 (2007)
17. X. Liu, B. Chen, H. Zhao, J. Qin, J. Cao, Maximum correntropy Kalman filter with state constraints. IEEE Access **5**, 25846–25853 (2017)
18. R. Pogula, T.K. Kumar, F. Albu, Robust sparse normalized LMAT algorithms for adaptive system identification under impulsive noise environments. Circuits Syst. Signal Process. **38**(11), 5103–5134 (2019)
19. H. Radmanesh, M. Hajiabadi, Recursive maximum correntropy learning algorithm with adaptive Kernel size. Circuits and Syst. II: IEEE Trans. Exp. Briefs **65**(7), 958–962 (2018)
20. A.H. Sayed, Adaptive networks. Proc. IEEE **102**(4), 460–497 (2014)
21. A. Singh, J.C Principe, Using Correntropy as a Cost Function in Linear Adaptive Filters. In: *2009 International Joint Conference on Neural Networks (IJCNN)* (IEEE, 2009), pp. 2950–2955
22. S.D. Stearns, *Adaptive Signal Processing* (Prentice-Hall, Hoboken, 1985)
23. G. Wang, B. Peng, Z. Feng, X. Yang, J. Deng, N. Wang, Adaptive filtering based on recursive minimum error entropy criterion. Signal Process. **179**, 107836 (2021)
24. G. Wang, R. Xue, J. Wang, A distributed maximum correntropy Kalman filter. Signal Process. **160**, 247–251 (2019)
25. S. Wang, W. Wang, S. Duan, L. Wang, Kernel recursive least squares with multiple feedback and its convergence analysis. IEEE Trans. Circuits Syst. II Exp. Briefs **64**(10), 1237–1241 (2017)
26. Z. Wu, S. Peng, W. Ma, B. Chen, C. Jose, Principe, minimum error entropy algorithms with sparsity penalty constraints. Entropy **17**(5), 3419–3437 (2015)
27. Z. Wu, J. Shi, X. Zhang, W. Ma, B. Chen, Kernel recursive maximum correntropy. Signal Process. **117**, 11–16 (2015)
28. H. Xia, W. Ren, M. Han, Kernel Generalized half-quadratic correntropy conjugate gradient algorithm for online prediction of chaotic time series. Circuits Syst. Signal Process. **42**, 2698–2722 (2023)

29. K. Xiong, S. Wang, The online random Fourier features conjugate gradient algorithm. IEEE Signal Process. Lett. **26**(5), 740–744 (2019)
30. L. Yang, J. Liu, R. Yan, X. Chen, Spline adaptive filter with arctangent-momentum strategy for nonlinear system identification. Signal Process. **164**, 99–109 (2019)
31. M. Zhang, X. Wang, X. Chen, A. Zhang, The Kernel conjugate gradient algorithms. IEEE Trans. Signal Process. **66**(16), 4377–4387 (2018)
32. X. Zhang, K. Li, Z. Wu, Y. Fu, H. Zhao, B. Chen, Convex regularized recursive maximum correntropy algorithm. Signal Process. **129**, 12–16 (2016)