Chinese Journal of Electronics
Vol.25, No.4, July 2016

The Data-Reusing MCC-Based Algorithm and Its Performance Analysis*

LIU Chunhui[1,2], QI Yue[2] and DING Wenrui[1]

(1. *Research Institute of Unmanned Aerial Vehicle, Beihang University, Beijing 100191, China*)

(2. *State Key Laboratory of Virtual Reality Technology and System, Beihang University, Beijing 100191, China*)

Abstract — Maximum correntropy criterion (MCC) provides a robust optimality criterion for non-Gaussian signal processing. In this paper, the weight update equation of the conventional MCC-based adaptive filtering algorithm is modified by reusing the past K input vectors, forming a class of data-reusing MCC-based algorithm, called DR-MCC algorithm. Comparing with the conventional MCC-based algorithm, the DR-MCC algorithm provides a much better convergence performance when the input data is correlated. The mean-square stability bound of the DR-MCC algorithm has been studied theoretically. For both Gaussian noise case and non-Gaussian noise case, the expressions for the steady-state Excess mean square error (EMSE) of DR-MCC algorithm have been derived. The relationship between the data-reusing order and the steady-state EMSEs is also analyzed. Simulation results are in agreement with the theoretical analysis.

Key words — Adaptive filtering, Maximum correntropy criterion (MCC), Data-reusing, Mean-square stability, Steady-state excess mean square error.

I. Introduction

Adaptive filters are used in a wide range of signal processing applications such as channel equalization, noise cancellation, system modeling, *etc*[1]. The Least mean square (LMS) algorithm[2] and its variants[3,4] are the most famous adaptation technique, which is based on the Minimum mean square error (MMSE) criterion. Recently, the Maximum correntropy criterion (MCC), which is a robust optimality criterion for non-Gaussian signal processing, has been successfully applied in adaptive filtering[5,6]. The MCC-based algorithm achieves a much better performance than MMSE-based algorithm in non-Gaussian environments, especially when the data are disturbed by impulsive nosie. Similar to the MMSE-based algorithm, the steady-state Excess mean square error (EMSE) of the MCC-based adaptive filtering algorithm is highly dependent on the trace of the input covariance matrix[7]. Hence the convergence performance of MCC-based algorithm degrades with data correlation.

One available scheme to de-correlate the input data is the data-reusing strategy[8], which has been proven effective in improving the convergence performance of the LMS-type algorithms. The representative data-reusing algorithms is the MMSE-based Affine projection algorithm (APA)[9−12], which accelerates the convergence speed based on multiple input vectors. Inspired by the APA, we modify the weight update equation of the MCC-based adaptive filter by reusing the input data and propose a data-reusing version of MCC-based algorithm, named DR-MCC algorithm. The proposed algorithm provides a better convergence performance especially when the input data is correlated. The relationship between the conventional MCC-based algorithm and the DR-MCC algorithm is discussed. The mean-square stability, the steady state behavior and the computational complexity of the DR-MCC algorithm are addressed from all around. Simulation results are provided to support the analysis. Comparing to the MMSE-based algorithms and the conventional MCC-based algorithm, the proposed DR-MCC algorithm is superior in terms of convergence speed and misadjustment.

II. Data-Reusing MCC-Based Algorithm

1. Conventional MCC-based algorithm

Consider the desired signal d_k that arises from a system identification model $d_k = \boldsymbol{x}_k^{\mathrm{T}}\boldsymbol{h} + v_k$, where $\boldsymbol{h} \in \Re^{M\times 1}$ is the impulse response of the unknown system, $\boldsymbol{x}_k \in \Re^{M\times 1}$ denotes the input-signal vector $\boldsymbol{x}_k = [x_k, x_{k-1}, \cdots, x_{k-L+1}]^{\mathrm{T}}$, v_k is a stationary, zero-mean, and independent noise sequence, which is independent of the input signal $\boldsymbol{x}_k$. The error signal is defined

*Manuscript Received April 7, 2015; Accepted August 7, 2015. This work is supported by the National Natural Science Foundation of China (No.61450008, No.61272348, No.61572054).

 DOI:10.1049/cje.2016.06.019

as $e_k = d_k - \boldsymbol{x}_k^{\mathrm{T}} \mathrm{w}_k$, where w_k is the estimate of 4 at iteration k. The correntropy, a local similarity measure between two random sequences $\boldsymbol{X}$ and $\boldsymbol{Y}$, is defined as follows:

$$V_{local}(\boldsymbol{X}, \boldsymbol{Y}) = E\left[\kappa(\boldsymbol{X}, \boldsymbol{Y})\right] = \int \kappa(x, y) dF_{x,y}(x, y) \quad (1)$$

where x and y denote the elements in sequences $\boldsymbol{X}$ and $\boldsymbol{Y}$, respectively. $\kappa(\cdot, \cdot)$ is a shift-invariant Mercer kKernel and $F_{x,y}(x, y)$ is the joint distribution function of (x, y). A widely used kernel is the Gaussian kernel:

$$\kappa(x, y) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{e^2}{2\sigma^2}\right) \quad (2)$$

where $e = x - y$ denotes element in the prediction error sequence $\boldsymbol{E} = \boldsymbol{X} - \boldsymbol{Y}$, and σ is the kernel width. The cost function of the MCC-based algorithm can be expressed as follows[6]:

$$J_{MCC}(\boldsymbol{w}_k) = \frac{1}{N} \sum_{i=k-N+1}^{k} \exp\left(-\frac{e_i^2}{2\sigma^2}\right) \quad (3)$$

where $e_i = d_i - \boldsymbol{x}_i^{\mathrm{T}} \boldsymbol{w}_i$. Using the stochastic gradient method, the MCC-based adaptive algorithm can be expressed as[6]:

$$\boldsymbol{w}_{k+1} = \boldsymbol{w}_k + \mu f(e_k) \boldsymbol{x}_k \quad (4)$$

where μ is the step-size, $f(e_k)$ is the error nonlinearity:

$$f(e_k) = \exp\left(-\frac{e_k^2}{2\sigma^2}\right) e_k \quad (5)$$

As $\sigma \to \infty$, the MCC-based algorithm will reduce to the LMS algorithm, which is MMSE-based:

$$\boldsymbol{w}_{k+1} = \boldsymbol{w}_k + \mu e_k \boldsymbol{x}_k \quad (6)$$

When the prediction error sequence follows a strong non-Gaussian distribution, especially a impulse distribution case, the probability of large-value error in the error sequence will be high. In this case, due to the threshold effect of the kernel width σ and the negative exponential function in Eq.(5), the contribution of the large-value error to the error nonlinearity is greatly weakened. By comparing Eq.(4) with Eq.(6), it can be concluded that the MCC-based adaptive algorithm helps the weight vector update in a more gentle way than the MMSE-based algorithm when large-value prediction error occurs. That is why the maximum correntropy criterion is more robust than MMSE criterion in the non-Gaussian environments.

2. Data-reusing MCC-based Algorithm

In fact, the weight update equation Eq.(4) is obtained by approximating the sum in Eq.(3) by the current value ($N = 1$). If we reuse the last K input ($N = K$) in Eq.(3), the MCC-based adaptive algorithm can be modified as:

$$\boldsymbol{w}_{k+1} = \boldsymbol{w}_k + \mu \frac{1}{K} \sum_{i=k-K+1}^{k} f(e_i) \boldsymbol{x}_i \quad (7)$$

where K can be viewed as the data-reusing order[8]. We can take the matrix formulation of the Eq.(6) as follow:

$$\boldsymbol{w}_{k+1} = \boldsymbol{w}_k + \frac{\mu}{K} \boldsymbol{F}_k \boldsymbol{X}_k \quad (8)$$

where $\boldsymbol{F}_k = [f(e_k), f(e_{k-1}), \cdots, f(e_{k-K+1})]$, $\boldsymbol{X}_k = [\boldsymbol{x}_k, \boldsymbol{x}_{k-1}, \cdots, \boldsymbol{x}_{k-K+1}]^{\mathrm{T}}$, $e_i = d_i - \mathrm{u}_i^{\mathrm{T}} \boldsymbol{w}_k$, $i = k - K + 1, k - K + 2, \cdots k$.

The variant e_i denotes the i-th error signal produced by the current estimated weight vector $\boldsymbol{w}_k$. The new weight update equation Eq.(7) can be named as the data-reusing MCC-based (DR-MCC) algorithm. By setting $K = 1$, Eq.(7) will return to the Eq.(4).

Comparing Eq.(7) with Eq.(4), it can be seen that the singular feature of the DR-MCC algorithm is the use of the data pairs from previous iterations instead of using the current data pair to generate the new gradient estimate, which are in turn used to update the adaptive weight vector. Since each input data can update the weight vector in different direction, along the average of the last K input data, a more accurate gradient estimates can be expected.

III. Performance Analysis

We now examine the performance of the DR-MCC algorithm. The analysis is based on the energy conservation relation[13−15] in the context of robustness analysis of adaptive filters.

1. Mean-square stability

The Eq.(8) can be written in terms of the weight-error vector, $\widetilde{\boldsymbol{w}}_k = \boldsymbol{h} - \boldsymbol{w}_k$, as

$$\widetilde{\boldsymbol{w}}_{k+1} = \widetilde{\boldsymbol{w}}_k - \frac{\mu}{K} \boldsymbol{F}_k \boldsymbol{X}_k \quad (9)$$

Squaring both sides and taking expectations of Eq.(8), we can find that

$$\begin{aligned} E\|\widetilde{\boldsymbol{w}}_{k+1}\|^2 = & E\|\widetilde{\boldsymbol{w}}_k\|^2 - \frac{2\mu}{K} E\left[(\boldsymbol{F}_k \boldsymbol{X}_k)^{\mathrm{T}} \widetilde{\boldsymbol{w}}_k\right] \\ & + \frac{\mu^2}{K^2} E\left[\boldsymbol{X}_k^{\mathrm{T}} \boldsymbol{F}_k^{\mathrm{T}} \boldsymbol{F}_k \boldsymbol{X}_k\right] \end{aligned} \quad (10)$$

In the mean-square, the convergence of the DR-MCC algorithm is guaranteed by

$$E\|\widetilde{\boldsymbol{w}}_{k+1}\|^2 < E\|\widetilde{\boldsymbol{w}}_k\|^2 \quad (11)$$

Then the step-size μ should satisfy

$$0 < \mu < \frac{2K \cdot E\left[(\boldsymbol{F}_k \boldsymbol{X}_k)^{\mathrm{T}} \widetilde{\boldsymbol{w}}_k\right]}{E\left[\boldsymbol{X}_k^{\mathrm{T}} \boldsymbol{F}_k^{\mathrm{T}} \boldsymbol{F}_k \boldsymbol{X}_k\right]} \quad (12)$$

When the adaptive filter is in a transient state, the following approximation can be made[14]:

$$\boldsymbol{x}_i \widetilde{\boldsymbol{w}}_k \approx e_i \tag{13}$$

Then we can obtain the following equations,

$$\begin{aligned} E\left[(\boldsymbol{F}_k \boldsymbol{X}_k)^{\mathrm{T}} \widetilde{\boldsymbol{w}}_k\right] &= E\left[\sum_{i=k-K+1}^{k} f(e_i) \boldsymbol{x}_i^{\mathrm{T}} \widetilde{\boldsymbol{w}}_k\right] \\ &\approx E\left[\sum_{i=k-K+1}^{k} f(e_i) e_i\right] \\ &= K \cdot E\left[f(e_i) e_i\right] \end{aligned} \tag{14}$$

Using the assumption that the adaptive filter is long enough such that e_i is Gaussian, and x_i is asymptotically uncorrelated with $f(e_i)$, the following equation can be obtained:

$$\begin{aligned} E\left[\boldsymbol{X}_k^{\mathrm{T}} \boldsymbol{F}_k^{\mathrm{T}} \boldsymbol{F}_k \boldsymbol{X}_k\right] &= E\left[\sum_{i=k-K+1}^{k} f^2(e_i) \boldsymbol{x}_i^{\mathrm{T}} \boldsymbol{x}_i\right] \\ &= K \cdot E\left[\boldsymbol{x}_i^{\mathrm{T}} \boldsymbol{x}_i\right] \cdot E\left[f^2(e_i)\right] \\ &= K \cdot Tr(\boldsymbol{R}_{\boldsymbol{X}}) \cdot E\left[f^2(e_i)\right] \end{aligned} \tag{15}$$

where $\boldsymbol{R}_{\boldsymbol{X}} = E\left[\boldsymbol{x}_i \boldsymbol{x}_i^{\mathrm{T}}\right]$ is the covariance matrix of the input vector, and the $Tr(\cdot)$ is the trace operator.

Substituting Eqs.(14) and (15) into Eq.(12) yields

$$0 < \mu < \frac{2K \cdot E\left[f(e_i) e_i\right]}{Tr(\boldsymbol{R}_{\boldsymbol{X}}) \cdot E\left[f^2(e_i)\right]} \tag{16}$$

It is obvious that $f(e_k) \leq e_k$, then Eq.(16) satisfies the following bound:

$$0 < \mu < \frac{2K}{Tr(\boldsymbol{R}_{\boldsymbol{X}})} \leq \frac{2K \cdot E\left[f(e_i) e_i\right]}{Tr(\boldsymbol{R}_{\boldsymbol{X}}) \cdot E\left[f^2(e_i)\right]} \tag{17}$$

It should be noted that the LMS algorithm is convergent in the mean if the step-size μ satisfies[2]

$$0 < \mu < \frac{2}{Tr(\boldsymbol{R}_{\boldsymbol{X}})} \tag{18}$$

Comparing Eq.(17) with Eq.(18), it can be concluded that the upper bound of the step-size μ of conventional MCC-based algorithm is similar to that of LMS algorithm. It is practical to use the same upper bound of the LMS algorithm for the MCC-based algorithm. For DR-MCC algorithm, as the data-reusing order K increases, the upper bound of the step-size μ increases manyfold.

2. Steady-state behavior

Based on the weight-error vector $\widetilde{\boldsymbol{w}}_k$, the error signal can be written as

$$e_k = e_{a,k} + v_k \tag{19}$$

where $e_{a,k} = \boldsymbol{x}_k^{\mathrm{T}} \widetilde{\boldsymbol{w}}_k$ is the a priori error, which is assumed zero-mean and independent of the noise v_k. The mean-square a priori error $E\left[e_{a,k}^2\right]$ is the so-called Excess mean square error (EMSE), which is a popular measure of performance[7]. The limit of the EMSE is the so-called steady-state EMSE, which is defined as following

$$S = \lim_{k \to \infty} E\left[e_{a,k}^2\right] \tag{20}$$

The steady-state EMSE of the conventional MCC-based algorithm is derived and verified in literature[7]. For DR-MCC algorithm, we focus on the relationship between the data-reusing order and the steady-state EMSE.

As $k \to \infty$, the filter is stable and achieves the steady-state value such that the weight error power

$$\lim_{k \to \infty} E\|\widetilde{\boldsymbol{w}}_{k+1}\|^2 = \lim_{k \to \infty} E\|\widetilde{\boldsymbol{w}}_k\|^2 \tag{21}$$

Based on Eq.(19), Eq.(10) becomes

$$2K \cdot \lim_{k \to \infty} E\left[f(e_k) e_{a,k}\right] = \mu \cdot Tr(\boldsymbol{R}_{\boldsymbol{X}}) \cdot \lim_{k \to \infty} E\left[f^2(e_k)\right] \tag{22}$$

Following the Ref.[7], we also consider the Gaussian noise case and non-Gaussian case to derive the theoretical steady-state EMSE.

1) Gaussian noise case:

Assume that the noise v_k is zero-mean Gaussian distributed, with a variance σ_v^2. Using the independence assumption, the variance of the error e_k is defined as σ_e^2, where $\sigma_e^2 = E\left[e_{a,k}^2\right] + \sigma_v^2$. The following equations can be derived[7]

$$\lim_{k \to \infty} E\left[f(e_k) e_{a,k}\right] = \frac{\sigma^3 S}{\left(\sigma^2 + \sigma_v^2 + S\right)^{3/2}} \tag{23}$$

$$\lim_{k \to \infty} E\left[f^2(e_k)\right] = \frac{\sigma^3 (S + \sigma_v^2)}{\left(2S + 2\sigma_v^2 + \sigma^2\right)^{3/2}} \tag{24}$$

Substituting Eq.(23) and Eq.(24) into Eq.(22) yields

$$\frac{2K \cdot \sigma^3 S}{\left(\sigma^2 + \sigma_v^2 + S\right)^{3/2}} = \mu \cdot Tr\left(\boldsymbol{R}_{\boldsymbol{X}}\right) \cdot \frac{\sigma^3 (S + \sigma_v^2)}{\left(2S + 2\sigma_v^2 + \sigma^2\right)^{3/2}} \tag{25}$$

Then S can be obtained by solving the following fixed-point equation:

$$S = \frac{\mu}{2K} \cdot Tr(\boldsymbol{R}_{\boldsymbol{X}}) \cdot \frac{\left(\sigma^2 + \sigma_v^2 + S\right)^{3/2} (S + \sigma_v^2)}{\left(2S + 2\sigma_v^2 + \sigma^2\right)^{3/2}} \tag{26}$$

From Eq.(25), one can derive that

$$S = \frac{\lambda \mu Tr(\boldsymbol{R}_{\boldsymbol{X}}) \sigma_v^2}{2K - \lambda \mu Tr(\boldsymbol{R}_{\boldsymbol{X}})} \tag{27}$$

where $\lambda = (\frac{\sigma^2 + \sigma_v^2 + S}{2S + 2\sigma_v^2 + \sigma^2})^{3/2}$, $\lambda < 1$.

2) Non-Gaussian noise case:

In the non-Gaussian noise case, the steady-state EMSE can be derived based on Taylor series expansion[7] of the function $f(\cdot)$ as defined in Eq.(5). At steady-state, the distributions of $e_{a,k}$ and e_k are independent of k, hence the time index k can be omitted for brevity and rewrite Eq.(22) as

$$2K \cdot E\left[f(e)e_a\right] = \mu \cdot Tr(\boldsymbol{R_X}) \cdot E\left[f^2(e)\right] \tag{28}$$

Taking the Taylor expansion of $f(e) = f(e_a + v)$ with respect to e_a around v yields

$$\begin{aligned} f(e) &= f(e_a + v) \\ &= f(v) + f'(v)e_a + \frac{1}{2}f''(v)e_a^2 + o(e_a^2) \end{aligned} \tag{29}$$

where $o(e_a^2)$ denotes the third and higher-order terms, and

$$\begin{aligned} f'(v) &= \exp(-\frac{v^2}{2\sigma^2})(1 - \frac{v^2}{\sigma^2}) \\ f''(v) &= \exp(-\frac{v^2}{2\sigma^2})(\frac{v^3}{\sigma^4} - \frac{3v}{\sigma^2}) \end{aligned} \tag{30}$$

Then the following equations can be derived[7]:

$$E\left[f(e)e_a\right] \approx E\left[f'(v)\right]S \tag{31}$$

$$E\left[f^2(e)\right] \approx E\left[f^2(v)\right] + E\left[f(v)f''(v) + |f'(v)|^2\right]S \tag{32}$$

Substituting Eq.(31) and Eq.(32) into Eq.(28) yields

$$\begin{aligned} S =& \mu \cdot Tr(\boldsymbol{R_X}) \cdot E\left[f^2(v)\right] / \{2K \cdot E\left[f'(v)\right] \\ &- \mu \cdot Tr(\boldsymbol{R_X}) \cdot E\left[f(v)f''(v) + |f'(v)|^2\right]\} \end{aligned} \tag{33}$$

Based on Eqs.(30), the following equation can be obtained

$$\begin{aligned} S =& \mu \cdot Tr(\boldsymbol{R_X}) \cdot E\left[\exp(-\frac{v^2}{2\sigma^2})v^2\right] / \\ &\left\{2K \cdot E\left[\exp(-\frac{v^2}{2\sigma^2})(1 - \frac{v^2}{\sigma^2})\right] - \right. \\ &\left. \mu \cdot Tr(\boldsymbol{R_X}) \cdot E\left[\exp(-\frac{v^2}{2\sigma^2})(1 + \frac{2v^4}{\sigma^4} - \frac{5v^2}{\sigma^2})\right]\right\} \end{aligned} \tag{34}$$

Note that the steady-state EMSE of Eq.(34) is valid only when the a priori error e_a is small such that the term $E\left[o(e_a^2)\right]$ is negligible. This implies that a larger step-size μ or noise power will result in a larger a priori error.

In brief, it can be concluded from Eq.(27) and Eq.(34) that, when the step-size is small enough, as the data-reusing order increases, the steady-state EMSE of the DR-MCC algorithm approximately decreases proportionally both in Gaussian noise and non-Gaussian noise environment.

3. Computational complexity

We also evaluate the computational complexity of the DR-MCC algorithm. Since the number of additions is, in all cases, of the same order of magnitude than the multiplications, we will just consider the number of multiplications.

It is well known that the LMS algorithm requires $2M + 1$ multiplications[17] to calculate the output of the filter and to update its weights (see Eq.(6)), where M is the tap length of the adaptive filter. Comparing with LMS algorithm, the error nonlinearity $f(e_k)$ is used in the MCC-based algorithm (see Eq.(4)) and it steps up the computational burden. Due to that the exponential function can be implemented by the look-up table[18], the MCC-based algorithm needs 3 more $(2M + 4)$ multiplications than the LMS algorithm to calculate the output.

Regarding the proposed DR-MCC algorithm, it requires $K(M + 3) + 1$ multiplications to update its weights and M more multiplications to calculate the output. Although the DR-MCC algorithm is based on reusing K input data, it only requires $(K + 1)M + 3K + 1$ multiplications, much less than K single MCC-based algorithm. As the data-reusing order K increases, the computational complexity of the DR-MCC algorithm increases almost linearly.

IV. Performance Analysis

In this section, we illustrate simulation results to confirm the theoretical analysis of the DR-MCC algorithm in a system identification scenario. The unknown system has 20 taps and is randomly generated. The adaptive filter and the unknown system are assumed to have the same number of taps. The input signal $\boldsymbol{x}_k$ is obtained by filtering a white, zero-mean, Gaussian random sequence through the system[13]

$$G(z) = \frac{1 + 0.5z^{-1} + 0.81z^{-2}}{1 - 0.59z^{-1} + 0.4z^{-2}} \tag{35}$$

Then, a highly correlated Gaussian signal of which the trace of the input covariance matrix is around 1227 is generated.

1. Mean-square stability

Firstly, we evaluate the stability bounds from Eq.(17) for DR-MCC algorithm with different data-reusing order in Table 1.

Table 1. Mean-square stability bound

Data-reusing order	$\frac{2K}{Tr(\boldsymbol{R_X})}$	$\mu_{\max}$
$K = 1$	0.0016	0.0016
$K = 2$	0.0033	0.0033
$K = 4$	0.0065	0.0065
$K = 8$	0.013	0.013

A Gaussian noise v_k is added to the output such that the Signal-to-noise ratio (SNR) is 30 dB. The simulated MSEs are the average of the ensemble 10^4 iterations, hence it reflects both the transient and steady states. The stability bounds are numerically verified in Fig.1, where the simulated MSEs of DR-MCC are plotted as a function of the step size.

As shown in Fig.1, a higher value of K causes a lower value of MSE. When the step sizes of the DR-MCC algorithm are small, the MSEs of all values of K are low. However, when the step sizes become large, the MSEs of all value of K increase and a lower value of K causes a faster ascending speed. When the step sizes approach the theoretical stability bounds, the MSEs of all values of K come to a large value even exceed 0dB. Though some deviation produced by approximation of Eqs.(14) and (17) may exist, the simulation results are generally in agreement with the theoretical analysis.

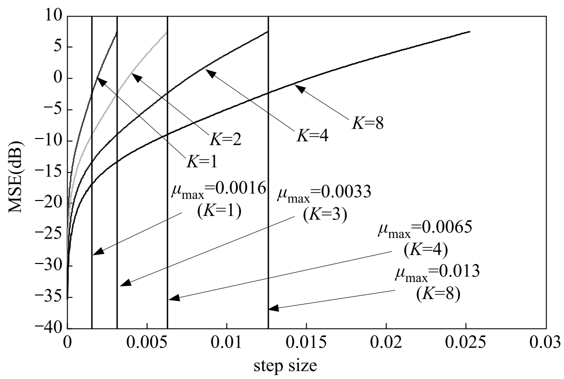

Fig. 1. Simulated MSE versus step-size μ

2. Steady-state behavior

Next, we evaluate the steady-state behavior of the DR-MCC algorithm. Based on the same input source above, we verify the theoretical EMSE versus the step-size and noise variance in Gaussian noise case and non-Gaussian noise case, respectively. The simulated EMSEs are computed as an average over 100 independent Monte Carlo simulations and in each Monte Carlo simulation, 10^5 iterations were run to ensure the algorithm to reach the steady state, and the EMSE was obtained as the average over the last 2×10^4 iterations. Fig.2 and Fig.3 show the theoretical and simulated steady-state EMSEs in Gaussian noise, versus the step-size and noise variance.

As shown in Fig.2 and Fig.3, a higher value of K causes a lower value of steady-state EMSE, both in simulation and theoretical analysis. Both the Fig.2 and Fig.3 illustrate that the simulated steady-state EMSEs versus the step-size match well with those calculated by theory.

In Fig.2, as the step-size increases, the differences between the steady-state EMSEs obtained by different K become large. When the step-size approaches to the mean-square stability bound, the steady-state EMSE become very large (As shown in Fig.2, when the step-size μ reach to 0.84 in $K = 1$, the value of steady-state EMSE is already very high).

In Fig.3, for the same step-size, as the noise variance increases, the differences between the steady-state EMSEs obtained by different K become small. This phenomenon can be interpreted as the damaging effect of the high power noise to the performance of the DR-MCC algorithm.

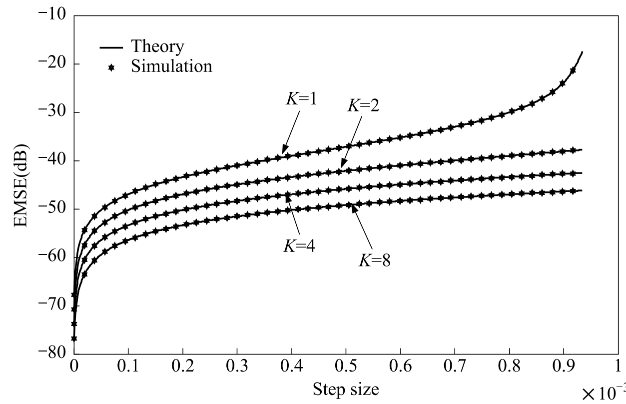

Fig. 2. EMSEs versus step-size μ in Gaussian noise ($\sigma^2 = 3$, $\sigma_v^2 = 5 \times 10^{-5}$)

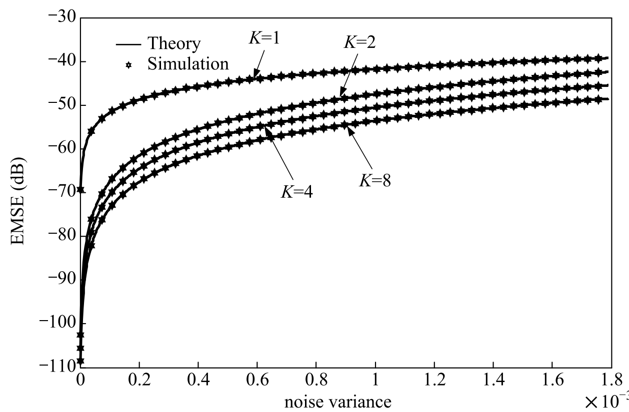

Fig. 3. EMSEs versus noise variance σ_v^2 in Gaussian noise ($\sigma^2 = 3$, $\mu = 0.001$)

Table 2 presents the theoretical and simulated EMSEs in several non-Gaussian noise sequences, where the uniform noise is distributed over [-1,1], binary noise is either -1 or 1(each with probability 0.5), Laplace noise is zero-mean with standard deviation 1, and the Cauchy noise is distributed with PDF: $p(v) = 1/\pi(1 + v^2)$. Again, the simulation results are in agreement with the theoretical calculations.

It should be pointed out that the step size and the noise variance in Table 2 are chosen as small value to guarantee the rationality of Eq.(34). However, as mentioned previously, when the step-size or the noise power is large, the derived value in Eq.(34) will not accurately enough characterize the performance of the DR-MCC algorithm. This expectation is confirmed in Fig.4 and Fig.5, where the theoretical and simulated EMSEs are plotted in uniform noise, versus the step-size and noise variance respectively.

Table 2. Theoretical and simulated EMSEs in several non-Gaussian noises (μ=0.001,σ^2=3)

Distribution	Data-reusing order	Theory	Simulation
Uniform	$K=1$	0.3381	0.3381 ± 0.1037
	$K=2$	0.1317	0.1317 ± 0.0923
	$K=4$	0.0593	0.0593 ± 0.0128
	$K=8$	0.0283	0.0283 ± 0.0021
Binary	$K=1$	0.6440	0.6440 ± 0.1278
	$K=2$	0.3758	0.3758 ± 0.0912
	$K=4$	0.2050	0.2050 ± 0.0638
	$K=8$	0.1074	0.1074 ± 0.0238
Laplace	$K=1$	0.6121	0.6121 ± 0.1790
	$K=2$	0.3396	0.3396 ± 0.0846
	$K=4$	0.1796	0.1796 ± 0.0283
	$K=8$	0.0925	0.0925 ± 0.0098
Cauchy	$K=1$	0.3497	0.3497 ± 0.1294
	$K=2$	0.1586	0.1586 ± 0.0582
	$K=4$	0.0453	0.0453 ± 0.0129
	$K=8$	0.0207	0.0207 ± 0.0079

As shown in Fig.4 and Fig.5, both the theoretical and simulated EMSEs versus the step-size and noise variance decrease in uniform noise as the data-reusing order K increases.

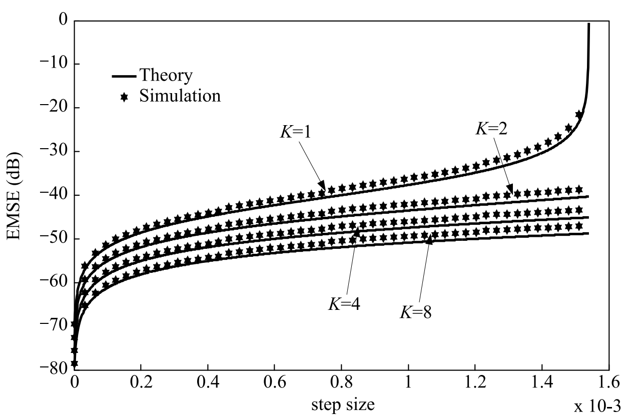

Fig. 4. EMSEs versus step-size μ in uniform noise ($\sigma^2 = 3$, $\sigma_v^2 = 5 \times 10^{-5}$)

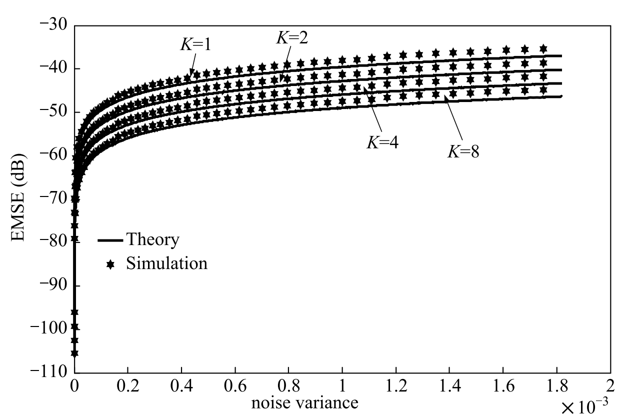

Fig. 5. EMSEs versus noise variance σ_v^2 in uniform noise ($\sigma^2 = 3$, $\mu = 1.5 \times 10^{-4}$)

In Fig.4, when the step-size reaches to a certain value ($\mu = 1.53 \times 10^{-3}, K = 1$), nearly the mean-square stability bound in Table 2, the steady-state EMSE has already reached a very high value.

In Fig.5, due to the effect of the high power noise, the steady-state EMSEs obtained by different K become close to each other as the noise variance increases. These trends are in accord with that in Gaussian noise.

However, as can be seen in Fig.4 and Fig.5, when the value of step-size or the noise variance is low, the simulated steady-state EMSEs match the theoretical value well. But when the value of step-size or the noise variance becomes large, the simulated steady-state EMSEs gradually deviate from the theoretical values. These results confirm the previous analysis for Eq.(34).

3. Performance comparison

Finally, we compare the performance of the proposed DR-MCC algorithm with that of the Normalized least mean square (NLMS)[3], APA[9], and conventional MCC-based Algorithm[6] in a impulsive noise environment with the same system identification application as above.

The Mean-square-deviation (MSD) learning curves[16] are plotted to measure the performance of the four algorithms. The MSD is computed as $20\log_{10}\|4 - w_k\|^2/\|4\|^2$ averaged 100 independent trials in each algorithm. The weight vectors are initialized as zero vectors for all the algorithms. The impulsive noise is generated as $b_k A_k$, where b_k is a Bernoulli process with a probability of $P[b_k = 1] = 0.01$. The power of A_k is $\sigma_A^2 = 0.1\sigma_y^2$, where σ_y^2 is the power of system output: $y_k = x_k^{\mathrm{T}} 4$. The data-reusing order K is set to 4 for the DR-MCC algorithm and APA, and the step-size μ is set to 10^{-3} for all the algorithms. The MSD learning curves obtained by the four algorithms above are illustrated in Fig.6.

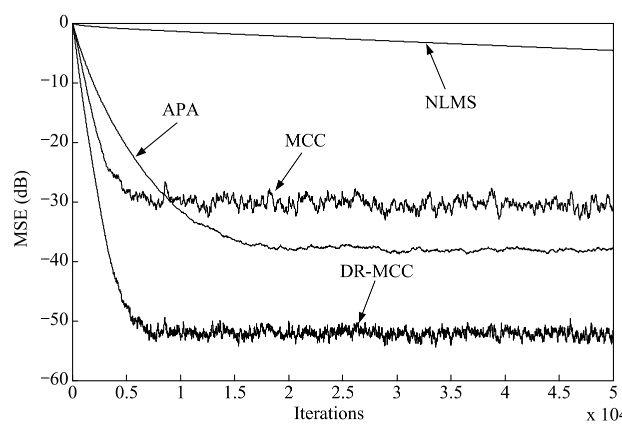

Fig. 6. The MSD learning curves of the four algorithms

As can be seen in Fig.6, the APA yields much better convergence performance than the NLMS algorithm due to the increased projection order to de-correlate the colored input. The conventional MCC-based algorithm yields better convergence speed than the two MMSE-based algorithms due to the robustness of the maximum correntropy criterion in the impulsive noise environment. The proposed DR-MCC algorithm yields both a significantly fast convergence speed and a reduced steady-state misalignment as compared to the other three competing algorithms.

V. Conclusion

In this paper, a class of data-reusing MCC-based algorithm, named DR-MCC algorithm, is proposed. By reusing the last K input data, the DR-MCC algorithm provides a more accurate gradient estimates and obtains a much better convergence performance than the conventional MCC-based algorithm. A thorough performance analysis for the DR-MCC algorithm is addressed. Simulation illustrates that experiment results are in agreement with the theoretical calculations and analysis. And it also shows that the proposed DR-MCC algorithm outclasses both the MMSE-based algorithms and the conventional MCC-based in terms of convergence speed and misadjustment.

References

[1] C. Wang, M. Ma, R.D. Ying, *et al.*, "Narrowband interference mitigation in DS-UWB Systems", *IEEE Signal Processing Letters*, Vol.12, No.12, pp.429–434, 2010.

[2] Y. Tan, Z.Q. He and B.Y. Tian, "A Novel Generalization of Modified LMS Algorithm to Fractional Order", *IEEE Signal Processing Letters*, Vol.22, No.9, pp.1244–1248, 2015.

[3] D.T.M. Slock, "On the convergence behavior of the LMS and the normalized LMS algorithms", *IEEE Transactions on Signal Processing*, Vol.41, No.9, pp.2811–2825, 1993.

[4] H. Shin, A.H. Sayed and W. Song, "Variable step-size NLMS and affine projection algorithms", *IEEE Signal Processing Letters*, Vol.11, No.2, pp.132–135, 2004.

[5] W. Liu, P. Pokharel and J. Principe, "Correntropy: Properties and applications in non-gaussian signal processing", *IEEE Transactions on Signal Processing*, Vol.55, No.11, pp.5286–5298, 2007.

[6] A. Singh and J.C. Principe, "Using correntropy as a cost function in linear adaptive filters", *Proc. of International Joint Conference on Neural Networks*, Atlanta, Georgia, USA, pp.2950–2955, 2009.

[7] B. Chen, X. Lei, J. Liang, *et al.*, "Steady-state mean-square error analysis for adaptive filtering under the maximum correntropy criterion", *IEEE Signal Processing Letters*, Vol.21, No.7, pp.880–884, 2014.

[8] J.A. Apolinrio, M.L.R. de Campos and P.S.R. Diniz, "Convergence analysis of the binormalized data-reusing LMS algorithm", *IEEE Transactions on Signal Processing*, Vol.48, No.11, pp.3235–3242, 2000.

[9] T. Shao, Y.R. Zheng and J. Benesty, "An affine projection sign algorithm robust against impulsive interferences", *IEEE Signal Processing Letters*, Vol.17, No.4, pp.327–330, 2010.

[10] S. Werner and P.S.R. Diniz, "Set-membership affine projection algorithm", *IEEE Signal Processing Letters*, Vol.8, No.8, pp.231–235, 2001.

[11] S.J. Ban and S. Kim, "Pseudo affine projection sign algorithm for robust system identification", *Electronics Letters*, Vol.46, No.12, pp.865–866, 2010.

[12] M.Z.A. Bhotto, M.O. Ahmad and M.N.S. Swamy, "Robust shrinkage affine-projection sign adaptive-filtering algorithms for impulsive noise environments", *IEEE Transactions on Signal Processing*, Vol.62, No.13, pp.3349–3359, 2014.

[13] H.C. Shin, W.J. Song and A.H. Sayed, "Mean-square performance of data-reusing adaptive algorithms", *IEEE Signal Processing Letters*, Vol.12, No.12, pp.851–854, 2005.

[14] N.R. Yousef and A.H. Sayed, "A unified approach to the steady-state and tracking analysis of adaptive filters", *IEEE Transactions on Signal Processing*, Vol.49, No.2, pp.314–324, 2001.

[15] B. Lin, R. He, X. Wang, *et al.*, "The steady-state mean-square error analysis for least mean p-order algorithm", *IEEE Signal Processing Letters*, Vol.16, No.3, pp.176–179, 2009.

[16] M.Z.A. Bhotto and A. Antoniou, "A family of shrinkage adaptive-filtering algorithms", *IEEE Transactions on Signal Processing*, Vol.61, No.7, pp.1689–1697, 2013.

[17] S. Haykin, *Adaptive Filter Theory*, NJ: Prentice Hall Press, Englewood Cliffs, 1996.

[18] J. Arenas-García, V. Gómez-Verdejo and A.R.Figueiras-Vidal, "New algorithms for improved adaptive convex combination of LMS transversal filters", *IEEE Trans. on Instrumentation and Measurement*, Vol.54, No.6, pp.2239–2249, 2005.

LIU Chunhui was born in Tianjin, China, in 1984. He received the B.S. degree and the M.S. degree from Beihang University in 2006 and 2009 respectively. From 2009 to now, He was an engineer in Research Institute of Unmanned Aerial Vehicle, Beihang University. He is now pursuing Ph.D. degree in Beihang University. His research interests include signal processing and image processing. (Email: liuchunhui2134@126.com)

QI Yue received the B.S. degree and the Ph.D. from National University of Defense Technology in 1991 and 2001 respectively, the M.S. degree from the University of Science and Technology of China in 1995. He is a professor in the State Key Laboratory of Virtual Reality Technology and Systems at Beihang University. His research interests include virtual reality, computer vision and computer graphics. (Email: qy@buaa.edu.cn)

DING Wenrui received the B.S. degree in computer science from Beihang University in 1994. From 1994 to now, she was a teacher of Beihang University where she is a now a professor, deputy general designer of UAV system. Her research interests include data link, computer vision, and artificial intelligence. (Email: ding@buaa.edu.cn)