



Tracking Performance of Improved Convex Combination Adaptive Filter Based on Maximum Correntropy Criterion

Wenjing Wu, Zhonghua Liang^(✉), Qianwen Luo, and Wei Li

School of Information Engineering, Chang'an University, Xi'an 710064, P. R. China
lzhxjd@hotmail.com

Abstract. A convex combination adaptive filter based on maximum correntropy criterion (CMCC) was widely used to solve the contradiction between the step size and the misadjustment in impulsive interference. However, one of the major drawbacks of the CMCC is its poor tracking ability. In order to solve this problem, this paper proposes an improved convex combination based on the maximum correntropy criterion (ICMCC), and investigates its estimation performance for system identification in the presence of non-Gaussian noise. The proposed ICMCC algorithm implements the combination of arbitrary number of maximum correntropy criterion (MCC) based adaptive filters with different adaption steps. Each MCC filter in the ICMCC is capable of tracking a specific change speed, such that the combined filter can track a variety of the change speed of weight vectors. In terms of normalized mean square deviation (NMSD) and tracking speed, the proposed algorithm shows good performance in the system identification for four non-Gaussian noise scenarios.

Keywords: Convex combination · Maximum correntropy criterion (MCC); Non-Gaussian noise; Normalized mean square deviation (NMSD); System identification

1 Introduction

Due to the low computational complexity and ease of implementation of the least mean square Algorithm (LMS), it is widely used in signal processing, system identification, acoustic echo cancellation, blind equalization, and so on [1]. However, the output of the LMS filter is not only sensitive to the amount of scaling of the input [2,3], but also degrades under non-Gaussian noise [4–6].

This work was supported in part by the National Natural Science Foundation of China under Grant No. 61271262, and in part by the Special Fund for Basic Scientific Research of Central Colleges, Chang'an University (310824171004).

Therefore, the least mean fourth (LMF) algorithm [7,8], least mean p-norm algorithm (LMP) [9], and recursive least p-norm algorithm (RLP) [10] based on the gradient algorithm are proposed to improve the performance degradation under non-Gaussian noise. Recently, a more robust adaptive algorithm from the information theoretic (IT) has been proposed by Principle and et al, where the algorithm includes entropy [11], mutual information [12], and correntropy [13]. Due to its simplicity and robustness to non-Gaussian environments, the maximum correntropy criterion (MCC) [14] and the minimum error entropy (MEE) [15] have been paid attention in recent years. Although the performance of MCC and MEE is similar, the computational complexity of MEE is relatively high compared to MCC.

Recently, MCC has been used as an adaptive criterion for non-gaussian signal processing in [4]. At the same time, the tracking analysis and steady-state mean square error analysis of MCC were proposed in [16,17]. The steady-state error of the MCC algorithm depends mainly on the step size, and its convergence speed is mainly based on step length and kernel width. When the step length is fixed, the contradiction between convergence speed and steady-state mean square error can be overcome by changing the kernel width. In [18], Weihua Wang et al. proposed a switch width based on maximum correntropy. In [19], Yicong He et al. proposed a new adaptive algorithm based on generalized correntropy, using generalized Gaussian density instead of traditional width. When the width is certain, the step length is inversely proportional to the misalignment. In [20], Ren Wang et al. proposed a variable step maximum correntropy adaptive filter. In [21], the convex combination is introduced into an MCC-based adaptive filter, so that the combination filter simultaneously gets the fast convergence speed of the filter with large step size as well as the low misadjustment of the filter with small step size. However, in the convex combination of maximum correntropy criterion (CMCC) filter, the tracking and convergence performance of the combined filter are reduced. Therefore, an improved convex combination filter based on maximum correntropy criterion (ICMCC) is presented in this paper. Compared with the CMCC, ICMCC not only has fast convergence speed and low misalignment, but also can track the optimal value fast in any weight coefficient changes.

The rest of the paper is organized as follows. we review briefly MCC-based algorithms in Sect. 2. Then, ICMCC, adding weight transfer algorithm to enhance convergence rate, is proposed in Sect. 3. The simulation results are given in Sect. 4, and the conclusion is presented in Sect. 5.

2 MCC-Based Algorithms

Correntropy is a measure of local similarity between two random variables, and can also be used as a cost function in adaptive filtering [22]. Considered X and Y two random variables, the correntropy is [23]

$$v(X, Y) = E[k(X, Y)] = \int k(X, Y) dF_{XY} F(x, y), \quad (1)$$

where $\kappa(\cdot, \cdot)$ is a shift-invariant Mercer kernel, and F_{XY} denotes the joint distribution function of (X, Y) . The most widely used kernel in correntropy is the Gaussian kernel

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

where σ is the kernel width, and $e = x - y$. The MCC algorithm finds the optimal value by maximizing the correntropy.

According to the stochastic gradient principle of adaptive algorithm, the updating equation of weight coefficient based on maximum correntropy is [21]

$$w[k] = w[k-1] + \lambda \exp\left(\frac{-e[k]^2}{2\sigma^2}\right) e[k] X[k], \quad (3)$$

where λ is the step size and $X[k]$ is the input at the moment k .

3 Improvement of MCC-Based Adaptive Filter

CMCC adaptive filtering algorithm is the latest development based on the maximum correntropy criterion in the contradiction between convergence speed and misalignment. This method combines two adaptive filters by convex combination, thus obtaining the fast convergence speed of large step and the low offset of small step length. But this method has a major challenge, that is, the tracking ability of the combinational filter is reduced. In this part, we first introduce the CMCC algorithm, and then extend the method to arbitrary number adaptive filters through the maximization of the correntropy of the combined filter to improve the tracking ability and convergence speed of the combined filter.

3.1 CMCC Algorithm

The implementation of the CMCC first requires two filters with different step sizes, and then the two filters update their own weights according to their own criteria and errors. However, the update criterion of the mixing coefficient is to maximize the correntropy of combined filter. The combination weight of CMCC can be expressed as [25]

$$w[k] = v[k]w_1[k] + (1 - v[k])w_2[k], \quad (4)$$

where the mixing coefficient $v[k]$ can be denoted as $v[k] = \text{sgm}\alpha[k] = 1/(1 + e^{-\alpha[k]})$. $w_1[k]$ and $w_2[k]$ are the weights of large step and small step respectively, and they are expressed as $w_i[k] = w_i[k-1] + \lambda_i \exp\left(\frac{-e_i[k]^2}{2\sigma^2}\right) e_i[k] X[k]$, $i = 1, 2$. And $e_i[k] = d[k] - y_i[k]$, $i = 1, 2$ is the error incurred by the component adaptive filter. Similarly, the combined filter output can be obtained and expressed as

$$y[k] = v[k]y_1[k] + (1 - v[k])y_2[k], \quad (5)$$

where $y_1[k] = X^T w_1[k]$ and $y_2[k] = X^T w_2[k]$ represent the output of a large step filter and a small step filter, respectively. The parameter $\alpha[k]$ is used to indirectly adjust the mixing coefficient, and updated by gradient algorithm that maximizes the correntropy of the combined filter, that is [21]:

$$\begin{aligned}\alpha[k+1] &= \alpha[k] + \mu_\alpha \sigma^2 \frac{\partial \exp(-\frac{e^2[k]}{2\sigma^2})}{\partial \alpha[k]} \\ &= \alpha[k] + \mu_\alpha v[k](1-v[k])(y_1[k] - y_2[k]) \exp(-\frac{e^2[k]}{2\sigma^2}) e[k],\end{aligned}\quad (6)$$

where μ_α represents the step size of the parameter $\alpha[k]$. In order to ensure that the adaptive speed of the combined filter is faster than that of the large step filter, μ_α must be larger than μ_1 . In order to prevent the $v[k]$ from approaching 0 or 1, the range of $\alpha[k]$ is limited to $[-4, 4]$ [25].

Following [20], when the fast filter is significantly better than the slow filter, we can accelerate the convergence performance of the algorithm by the following formula. The modified small step filter can be expressed as [21]

$$w_2[k] = \beta w_2[k-1] + \lambda_2 \exp(-\frac{e_2^2[k]}{2\sigma^2}) e_2[k] x[k] + (1-\beta) w_1[k], \quad (7)$$

where β is the transfer coefficient.

3.2 ICMCC Algorithm

In order to improve the disadvantage of poor tracking performance of CMCC, this paper makes a convex combination of arbitrary number filters with different steps and obtains the ICMCC algorithm. When the number of L adaptive filters based on MCC is combined, the weight of the combined filter is obtained as follows [25]:

$$w_{eq}[k] = \sum_{i=1}^L v_i[k] w_i[k], \quad (8)$$

where $w_i[k]$ represents the weight of the component filter whose the step size is denoted by μ_i ($\mu_1 > \mu_2 > \dots > \mu_L$). Each filter updates its weight based on its own error and can be expressed as $w_i[k] = w_i[k-1] + \lambda_i \exp(-\frac{e_i^2[k]}{2\sigma^2}) e_i[k] X[k]$, $i = 1, 2, \dots, L$. $v_i[k]$ is the mixing coefficient and satisfies $\sum_{i=1}^L v_i[k] = 1$.

In the case of combining arbitrary number of filters, the L auxiliary parameters $\alpha_i[k]$ were updated with stochastic positive gradient method to adjust the L mixing parameters. We use softmax activation function to define the relationship between $v_i[k]$ and $\alpha_i[k]$ to ensure the stability of $v_i[k]$. $v_i[k]$ can be expressed as [25]

$$v_i[k] = \frac{\exp(\alpha_i[k])}{\sum_{j=1}^L \exp(\alpha_j[k])}, \quad (9)$$

By multiplying both sides of formula (8) by $X^T[k]$, it is concluded that the output of ICMCC filter is a convex combination of all filter outputs.

$$y_{eq}[k] = \sum_{i=1}^L v_i[k]y_i[k], \tag{10}$$

where $y_i[k] = X^T w_i[k]$, $i = 1, 2 \dots L$ is the component filter output.

As for the CMCC filter, the parameter $\alpha_i[k]$ is updated using MCC rules to maximize the overall correntropy

$$\alpha_i[k + 1] = \alpha_i[k] + \mu_\alpha v_i[k](y_i[k] - y_{eq}[k]) \exp\left(\frac{-e^2[k]}{2\sigma^2}\right) e[k]. \tag{11}$$

In (11), We must qualify μ_α larger than the step size of any component filter. To prevent the ICMCC algorithm from stopping, we usually limit $|v_i[k]| < 0.95$. Since $v_i[k]$ is regulated by $\alpha_i[k]$, the range of $\alpha_i[k]$ is $|\alpha_i[k]| \leq 0.5 \ln(19(L - 1))$.

Algorithm 1 ICMCC ALGORITHM

Initialization:

1. Parametres: $\mu_\alpha, \beta, \sigma, L, \gamma, \mu_i, i = 1, \dots L$
2. Initialize $\alpha_i[0] = 0, v_i[0] = 1/L, w_i[0] = 0, i = 1, \dots L$

Update:

for $k = 0, 1, 2, \dots$.

$y_i[k] = w_i^T[k]X[k], i = 1, \dots L$

$e_i[k] = d[k] - y_i[k], i = 1, \dots L$

$y_i[k] = \sum_{i=1}^L w_i^T[k]X[k], i = 1, \dots L$

$e[k] = d[k] - y_{eq}[k]$

$\alpha_i[k + 1] = \alpha_i[k] + \mu_\alpha v_i[k](y_i[k] - y_{eq}[k]) \exp\left(\frac{-e^2[k]}{2\sigma^2}\right) e[k]$

$v_i[k] = \frac{\exp(\alpha_i[k])}{\sum_{j=1}^L \exp(\alpha_j[k])}$

if $\gamma_i \geq \text{cor}(e^2[k])/\text{cor}(e_i^2[k])$

$w_i[k] = \beta w_i[k - 1] + \mu_i \exp\left(\frac{-e_i^2[k]}{2\sigma^2}\right) e_i[k]X[k] + (1 - \beta)w_{eq}[k]$

else

$w_i[k] = w_i[k - 1] + \lambda_i \exp\left(\frac{-e_i[k]^2}{2\sigma^2}\right) e_i[k]X[k]$

end

End

In order to increase the tracking performance of the combined filter, we can determine whether to accelerate by calculating the ratio of the estimated correntropy of each component filter to the estimated correntropy of the combined filter. The estimated correntropy of the component filter and the combined filter are

$$\text{cor}(e_i[k]) = 0.98\text{cor}(e_i[k - 1]) + 0.02 \exp\left(\frac{-e_i^2[k]}{2\sigma^2}\right), \tag{12}$$

$$\text{cor}(e[k]) = 0.98\text{cor}(e[k - 1]) + 0.02 \exp\left(\frac{-e^2[k]}{2\sigma^2}\right). \tag{13}$$

When $\gamma \geq \text{cor}(e^2[k])/\text{cor}(e_i^2[k])$, $\gamma > 1$, we transfer a certain proportion of the combined filter to component filters that are worse than the combined filter. The modified adaption rule for $w_i[k]$ becomes

$$w_i[k] = \beta w_i[k-1] + \mu_i \exp\left(\frac{-e_i^2[k]}{2\sigma^2}\right) e_i[k] x[k] + (1-\beta) w_{eq}[k], \quad (14)$$

where β is the transfer coefficient. The condition for using Eq. (14) is that the large step size filter is significantly better than the small step size. Through a large number of experiments, γ and β were selected to be 2 and 0.8 respectively, to achieve the best transfer effect. The closer the choice of β is to 1, the more likely it is that the convex combination filter does not have a transfer coefficient. The pseudocodes of the proposed ICMCC are presented in Algorithm 1.

4 Simulation Results in System Identification Scenarios

In this section, we simulate the non-stationary system identification under non-Gaussian noise to verify the tracking performance of CMCC and ICMCC, and quantify each algorithm by normalized mean square deviation (NMSD) calculation which is expressed as $NMSD = 10 \log_{10} (\|w_i - w_0\|_2) / (\|w_0\|_2)$.

The length of the unknown system is 10, which is the same length as the adaptive filter, and the input signal is a Gaussian signal with zero-mean and unit power. At the output of the plant we add measurement noise $N[k]$, we give four different distributions for measurement noise including (1) uniform noise, where the uniform noise is distributed over $[-1, 1]$; (2) Laplace noise, where the probability density function of Laplace noise is $p(n) = 1/\pi(1+n^2)$; (3) binary noise, where the binary noise is either -1 or 1 (each with probability 0.5); (4) mixed Gaussian noise, where the mixed Gaussian noise is $N[k] = (1-\theta)N(\zeta_1, \delta_1^2) + \theta N(\zeta_2, \delta_2^2)$ (ζ_i and δ_i^2 represent mean and variance, respectively). In this paper, the parameter is set to (0, 0, 0.001, 10, 0.1).

For the sake of simplicity, $L = 4$ MCC filters with step sizes $\mu_1 = 0.1$, $\mu_2 = 0.03$, $\mu_3 = 0.01$ and $\mu_4 = 0.002$ are considered as component filters. Simultaneously, the two steps in the CMCC algorithm are selected as $\mu_1 = 0.1$ and $\mu_4 = 0.002$. The step size μ_α of the parameter $\alpha[k]$ is fixed at 30. The initial value of unknown system is random values between -1 and 1 , and the random-walk model introduces different rate of change of weight vectors. The random-walk model can express as:

$$w_0[k+1] = w_0[k] + q[k], \quad (15)$$

where $q[k]$ is an i.i.d. vector, with autocorrelation matrix $Q = E\{q[k]q^T[k]\}$. $Tr(Q)$ is a measure of the speed of the weight vector. In this paper, we consider that $q[k]$ is an independent Gaussian distribution.

In order to better embody the tracking performance of ICMCC, we choose $Tr(Q_1) = 10^{-6}$ and $Tr(Q_2) = 10^{-7}$ two different speed of change for weighting coefficients in $5000 < k < 10000$ and $15000 < k < 20000$ respectively. From

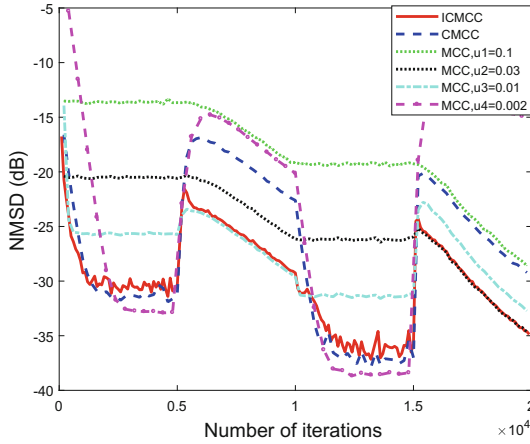


Fig. 1. Comparison of the convergence curve of a ICMCC filter and a CMCC filter with uniform noise.

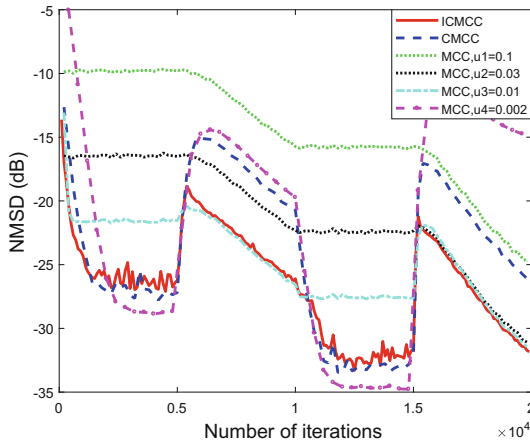


Fig. 2. Comparison of the convergence curve of a ICMCC filter and a CMCC filter with Laplace noise.

Figs. 1, 2, 3 and 4, when $k < 5000$, the adaptive filter of μ_1 has the fastest convergence speed but the highest amount of offset; μ_4 's adaptive filter has the lowest amount of offset, but the slowest convergence rate. Although the CMCC adaptive filter has a fast convergence rate and a slow offset amount, the tracking performance is not as good as that of the ICMCC. When $5000 < k < 10000$ and $15000 < k < 20000$, since the two weight coefficients of $Tr(Q_1)$ and $Tr(Q_2)$ are added to change the speed, the optimal value of the weight coefficient changes. Compared to the CMCC, the ICMCC not only quickly tracks the optimal value,

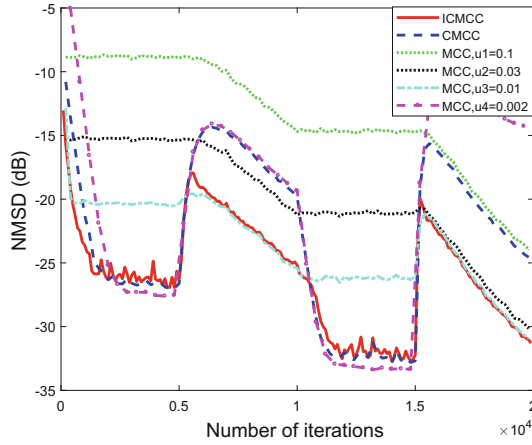


Fig. 3. Comparison of the convergence curve of a ICMCC filter and a CMCC filter with binary noise.

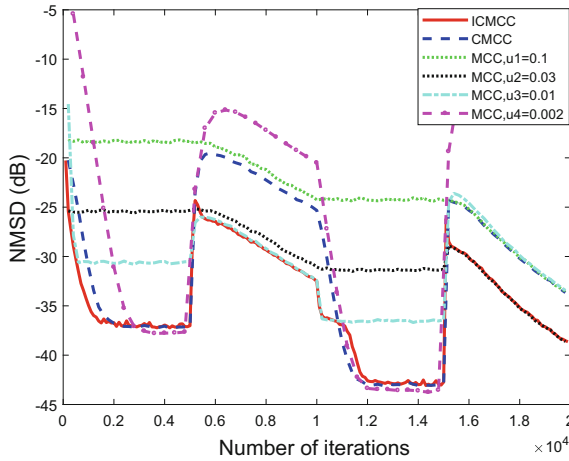


Fig. 4. Comparison of the convergence curve of a ICMCC filter and a CMCC filter with mixed Gaussian noise.

but also maintains a lower amount of misalignment. In summary, ICMCC gathers fast convergence speed, low misalignment and good tracking ability.

Simultaneously, we can study the tracking ability of ICMCC from four variations of the mixing coefficients. As shown in Fig. 5, the change process of the ICMCC four mixed parameters is indicated. We can see that in $5000 < k < 10000$ and $15000 < k < 20000$, ICMCC will use the hybrid coefficient adaptive filter to select the optimal performance as the primary role. Therefore, ICMCC shows the same performance as the optimal partial filter at any given moment. Therefore, ICMCC has good tracking performance and tracks a variety of changes.

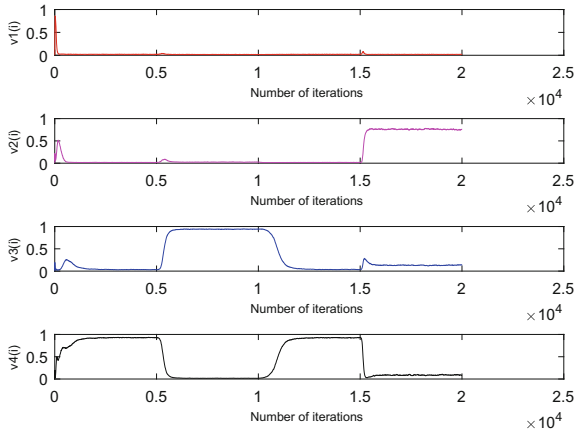


Fig. 5. Evolution of the mixing coefficients $v_1(i)$, $v_2(i)$, $v_3(i)$ and $v_4(i)$

5 Conclusions

In this paper, arbitrary number convex combination technique is employed to improve the tracking performance of CMCC filters. The improved algorithm not only has fast convergence speed and low offset, but also can track a variety of weight vector changes. Compared with the original CMCC algorithm, the improved algorithm is more suitable for system identification scenarios with non-Gaussian noises and abrupt change. The proposed adaptive filter can be applied to signal processing, system identification, noise cancellation, automatic equalization, echo cancellation and antenna array beamforming.

References

1. Sayed, A.H.: *Fundamentals of Adaptive Filtering*. Wiley, Hoboken (2003)
2. Gui, G., Mehdobniya, A., Adachi, F.: Least mean square/fourth algorithm for adaptive sparse channel estimation. In: 2013 IEEE 24th International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC), pp. 296–300 (2013)
3. Li, Y., Wang, Y., Jiang, T.: Sparse-aware set-membership NLMS algorithms and their application for sparse channel estimation and echo cancelation. *AEU-Int. J. Electron. Commun.* **70**, 895–902 (2016)
4. Chen, B., Xing, L., Zhao, H.: Generalized correntropy for robust adaptive filtering. *IEEE Trans. Signal Process.* **64**, 3376–3387 (2016)
5. Li, Y., Wang, Y., Yang, R., Albu, F.: A soft parameter function penalized normalized maximum correntropy criterion algorithm for sparse system identification. *Entropy* **19**, 1–16 (2017)
6. Li, Y., Jin, Z.: Sparse normalized maximum correntropy criterion algorithm with L1-norm penalties for channel estimation. In: 2017 Progress in Electromagnetics Research Symposium, pp. 1677–1682 (2017)

7. Walach, E., Widrow, B.: The least mean fourth (LMF) adaptive algorithm and its family. *IEEE Trans. Inf. Theory* **30**, 275–283 (1984)
8. Hajiabadi, M., Zamiri-Jafarian, H.: Distributed adaptive LMF algorithm for sparse parameter estimation in Gaussian mixture noise. In: 7th International Symposium on Telecommunications, pp. 1046–1049. Tehran (2014)
9. Lu, L., Zhao, H., Wang, W.: Performance analysis of the robust diffusion normalized least mean P-power algorithm. *IEEE Trans. Circuits Syst. II: Express Briefs* 1–1 (2018)
10. Lu, L., Zhao, H., Chen, B.: Improved variable forgetting factor recursive algorithm based on the logarithmic cost for volterra system identification. *IEEE Trans. Circuits Syst. II: Express Briefs* **63**, 588–592 (2016)
11. Principe, J.C.: *Information Theoretic Learning: Renyis Entropy and Kernel Perspectives*. Springer, Berlin (2010)
12. Erdogmus, D., Principe, J.C.: Convergence properties and data efficiency of the minimum error entropy criterion in adaline training. *IEEE Trans. Signal Process.* **51**, 1966–1978 (2003)
13. Chen, B., et al.: Stochastic gradient identification of wiener system with maximum mutual information criterion. *IET Signal Process* **5**, 589–597 (2011)
14. Liu, Y., Fan, Y., Zhou, L.C.: Ensemble correntropy based mooney viscosity prediction model for an industrial rubber mixing process. *Chem. Eng. Technol.* **39**, 1804–1812 (2016)
15. Han, S., Jeong, K.H., Principe, J.: Robust adaptive minimum entropy beam former in impulsive noise. In: *IEEE Workshop on Machine Learning for Signal Processing* (2007)
16. Chen, B., Xing, L., Liang, J.: Steady-state mean-square error analysis for adaptive filtering under the maximum correntropy criterion. *IEEE Signal Process. Lett.* **21**, 880–884 (2014)
17. Khalili, A., Rastegarnia, A.: Steady-state tracking analysis of adaptive filter with maximum correntropy criterion. *Circuits Syst. Signal Process* (2016)
18. Wang, W., Zhao, J., Qu, H., Chen, B.: A switch kernel width method of correntropy for channel estimation. In: *International Joint Conference on Neural Networks*, pp. 1–7 (2015)
19. He, Y., Wang, F.: Kernel adaptive filtering under generalized maximum correntropy criterion. In: *International Joint Conference on Neural Networks*, pp. 1738–1745 (2016)
20. Wang, R., Chen, B.: A variable step-size adaptive algorithm under maximum correntropy criterion. In: *IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP)*, pp. 1–5 (2015)
21. Shi, L.: Convex combination of adaptive filters under the maximum correntropy criterion in impulsive interference. *IEEE Signal Process. Lett.* **21**, 1385–1388 (2014)
22. Liu, W.: Correntropy: properties and applications in non-Gaussian signal processing. *IEEE Trans. Signal Process.* **55**, 5286–5298 (2007)
23. Liu, C., Qi, Y., Ding, W.: The data-reusing MCC-based algorithm and its performance analysis. *Chin. J. Electron.* **25**, 719–725 (2016)
24. Arenas-Garcia, J., Figueiras-Vidal, A.R., Sayed A.H.: Improved adaptive filtering schemes via adaptive combination. In: *2009 Conference Record of the Forty-Third Asilomar Conference on Signals, Systems and Computers*, pp. 1–4 (2009)
25. Arenas-Garca, J., Gmez-Verdejo, V., Figueiras-Vidal, A.R.: Combinations of adaptive filters: performance and convergence properties. *IEEE Signal Process. Mag.* **33**, 120C–140 (2016)