An Improved Variable Kernel Width for Maximum Correntropy Criterion Algorithm

Long Shi, Haiquan Zhao, *Senior Member, IEEE,* and Yuriy Zakharov, *Senior Member, IEEE*

***Abstract*—The maximum correntropy criterion (MCC) algorithm has attracted much attention due to its capability of combating impulsive noise. However, its performance depends on choice of the kernel width, which is a hard issue. Several variable kernel width schemes based on various error functions have been proposed to address this problem. Nevertheless, these methods may not provide an optimal kernel width because they do not contain any knowledge of the background noise that actually has influence on the optimization of the kernel width. This paper proposes an improved variable kernel width MCC algorithm, which is derived by minimizing the squared deviation at each iteration. We also design a reset mechanism for the proposed algorithm to improve its tracking capability when the estimated vector encounters a sudden change. Simulations for system identification and echo cancellation scenarios show that the proposed scheme outperforms other variable kernel width algorithms.**

***Index Terms*—maximum correntropy criterion, variable kernel width, squared deviation, reset mechanism.**

I. Introduction

THE least mean square (LMS) algorithm is utilized in many applications including system identification, acoustic echo cancellation, active noise control, channel equalization and so on [1]–[5]. The LMS algorithm can exhibit desirable performance in Gaussian noise since it is established on the second-order moment of the error signal. However, in practice, very often the impulsive noise is present [6]. In such scenarios, the conventional LMS algorithm performs poorly in terms of convergence and steady-state misalignment. To overcome this problem, many methods that can combat impulsive noise have been proposed, such as the sign algorithm [7], least mean p-power algorithm [8], least mean M-estimate algorithm [9], and their modifications [10]–[12].

With the development of information theoretic learning (ITL) [13], the maximum correntropy criterion (MCC) has recently been introduced in the adaptive filtering. The MCC adaptive filters can effectively suppress impulsive noise using more orders of the error signal [14], [15]. Unfortunately, the performance of the MCC algorithm depends on the choice of the kernel width, whereas in practice, it is difficult to select a reliable kernel width to guarantee favorable performance. To eliminate this shortcoming, the scheme using the convex combination of MCC algorithms (CMCC) has been proposed [16], which exhibits faster convergence than the MCC algorithm. However, the CMCC algorithm requires extra computations due to updates in two filters.

Apart from the CMCC algorithm, various time-varying kernel width (TKW) methods have been proposed [17]–[19]. The switch kernel width MCC (SMCC) algorithm is based on adaptively choosing the maximum between an instantaneous value constructed from the error signal and a predetermined kernel width [17]. The adaptive kernel width MCC (AMCC) is derived by dynamically calculating the sum of the squares for the kernel width and instantaneous error [18]. The recently proposed variable kernel width MCC (VKW-MCC) algorithm [19] is obtained by maximizing an error nonlinearity function described by the exponential expression. In the VKW-MCC algorithm, the estimation window strategy is introduced to attenuate the adverse impact of the impulsive interference-corrupted error signal. The VKW-MCC algorithm performance depends on the choice of the length of the estimation window. Only when the length of the window is set to be large enough can the VKW-MCC algorithm provide superior performance, but it will lead to the increasing cost for storing past values of the error signal. In addition, these TKW methods do not contain any knowledge of the background noise that actually has influence on the optimization of the kernel width. Therefore, there is a room for improvement of TKW schemes.

In this paper, we put forward an improved variable kernel width method for MCC (IVKW-MCC) algorithm to further improve the performance. The squared kernel width is obtained from minimizing the squared deviation at each iteration. To implement the proposed IVKW-MCC algorithm, we employ the moving average method to complete the update for the squared kernel width. In addition, a reset mechanism for the IVKW-MCC algorithm is designed to enhance its tracking capability when a sudden change of the estimated vector occurs. We also discuss the complexity of the IVKW-MCC algorithm. In general, our main contributions are twofold: (1) We derive a novel method for updating the kernel width; (2) A reset mechanism is developed to strengthen the tracking capability of the IVKW-MCC algorithm.

The paper is organized as follows. In Section 2, we review the MCC algorithm. In Section 3, the proposed IVKW-MCC algorithm is derived. In Section 4, numerical simulations in the impulsive noise environment are presented. In Section 5, we draw our conclusions.

L. Shi and H. Zhao are with the Key Laboratory of Magnetic Suspension Technology and Maglev Vehicle, Ministry of Education, and also with the School of Electrical Engineering, Southwest Jiaotong University, Chengdu 610031, Peoples Republic of China (e-mail: lshi@my.swjtu.edu.cn; hqzhao_swjtu@126.com), corresponding author: Haiquan Zhao.

Y. Zakharov is with Department of Electronic Engineering, University of York, U.K. (e-mail: yury.zakharov@york.ac.uk).

II. Review of the MCC Algorithm

We consider that the desired signal $d(k)$ arises from the linear model

$$d(k) = \boldsymbol{u}^T(k)\boldsymbol{w}_o + \eta(k), \tag{1}$$

where $\boldsymbol{u}(k) = [u(k), u(k-1), ..., u(k-M+1)]^T$ denotes the input vector with symbol T being transpose, $\boldsymbol{w}_o$ represents the unknown system to be estimated, and $\eta(k)$ stands for the background noise with zero mean and variance σ_η^2. The error signal is given by

$$e(k) = d(k) - \boldsymbol{u}^T(k)\boldsymbol{w}(k), \tag{2}$$

where $\boldsymbol{w}(k)$ is an estimate of $\boldsymbol{w}_o$ at iteration k.

The correntropy measures a local similarity of two random variables, and can be expressed as [14]

$$V_{local}(X,Y) = E[\kappa(X,Y)] = \int \kappa(x,y) dF_{x,y}(X,Y), \tag{3}$$

where $\kappa(x,y)$ denotes a shift-invariant Mercer Kernel, and $F_{x,y}(X,Y)$ is the joint distribution function of (x,y). Among various kernel functions, the Gaussian kernel attracts much popularity [14], [16], [19]

$$\kappa(x,y) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{e^2}{2\sigma^2}\right), \tag{4}$$

where $e = x - y$, and $\sigma > 0$ is the kernel width. The MCC algorithm is based on minimizing the cost function [15]

$$J_{MCC}(\boldsymbol{w}(k)) = E\left[\exp\left(-\frac{e^2(k)}{2\sigma^2}\right)\right]. \tag{5}$$

By virtue of the stochastic gradient method, the MCC algorithm recursion is given by [14]:

$$\boldsymbol{w}(k+1) = \boldsymbol{w}(k) + \mu \exp\left(-\frac{e^2(k)}{2\sigma^2}\right) e(k)\boldsymbol{u}(k), \tag{6}$$

where μ is a step size. The factor $\mu \exp\left(-\frac{e^2(k)}{2\sigma^2}\right)$ can be considered as an overall step size of the LMS algorithm. For a fixed μ, a large (small) σ leads to a large (small) value for the overall step size, which provides a fast (slow) convergence rate along with a high (low) steady-state misalignment. This implies that the performance of the MCC algorithm depends on the choice of the kernel width.

III. Proposed IVKW-MCC Algorithm

A. Design of the Variable Kernel Width

We replace σ in (6) with a time-varying kernel width $\sigma(k)$

$$\boldsymbol{w}(k+1) = \boldsymbol{w}(k) + \mu \exp\left(-\frac{e^2(k)}{2\sigma^2(k)}\right) e(k)\boldsymbol{u}(k). \tag{7}$$

Subtracting (7) from $\boldsymbol{w}_o$ gives rise to

$$\tilde{\boldsymbol{w}}(k+1) = \tilde{\boldsymbol{w}}(k) - \mu \exp\left(-\frac{e^2(k)}{2\sigma^2(k)}\right) e(k)\boldsymbol{u}(k), \tag{8}$$

where $\tilde{\boldsymbol{w}}(k) = \boldsymbol{w}_o - \boldsymbol{w}(k)$ is the weight error vector.

From (8), we arrive at the recursive update of the squared deviation

$$\begin{aligned}\|\tilde{\boldsymbol{w}}(k+1)\|^2 &= \|\tilde{\boldsymbol{w}}(k)\|^2 - 2\mu \left\{ e_a(k)e(k) \exp\left(-\frac{e^2(k)}{2\sigma^2(k)}\right)\right\} \\ &\quad + \mu^2 \left\{ \|\boldsymbol{u}(k)\|^2 e^2(k) \left[\exp\left(-\frac{e^2(k)}{2\sigma^2(k)}\right)\right]^2 \right\} \\ &\triangleq \|\tilde{\boldsymbol{w}}(k)\|^2 - f\left(\sigma(k)\right),\end{aligned} \tag{9}$$

where $e_a(k) = \tilde{\boldsymbol{w}}^T(k)\boldsymbol{u}(k)$ represents noise-free a *priori* error signal, and $f\left(\sigma(k)\right)$ is expressed as

$$\begin{aligned}f\left(\sigma(k)\right) &= 2\mu \left\{ e_a(k)e(k) \exp\left(-\frac{e^2(k)}{2\sigma^2(k)}\right)\right\} \\ &\quad - \mu^2 \left\{ \|\boldsymbol{u}(k)\|^2 e^2(k) \left[\exp\left(-\frac{e^2(k)}{2\sigma^2(k)}\right)\right]^2 \right\}.\end{aligned} \tag{10}$$

Considering that $e(k) = e_a(k) + \eta(k)$, $f\left(\sigma(k)\right)$ can be further written as

$$\begin{aligned}f\left(\sigma(k)\right) &= 2\mu \left\{ [e^2(k) - \eta^2(k) - e_a(k)\eta(k)] \exp\left(-\frac{e^2(k)}{2\sigma^2(k)}\right)\right\} \\ &\quad - \mu^2 \left\{ \|\boldsymbol{u}(k)\|^2 e^2(k) \left[\exp\left(-\frac{e^2(k)}{2\sigma^2(k)}\right)\right]^2 \right\}.\end{aligned} \tag{11}$$

To minimize the squared deviation at iteration k, we need to maximize $f\left(\sigma(k)\right)$. Taking the derivative of (11) with respect to $\sigma(k)$, we arrive at

$$\begin{aligned}\frac{\partial f(\sigma(k))}{\partial \sigma(k)} =& 2\mu \exp\left(-\frac{e^2(k)}{2\sigma^2(k)}\right) \frac{e^2(k)}{\sigma^3(k)} \\ & \left\{ [e^2(k) - \eta^2(k) - e_a(k)\eta(k)] \right. \\ & \left. - \mu \|\boldsymbol{u}(k)\|^2 e^2(k) \exp\left(-\frac{e^2(k)}{2\sigma^2(k)}\right) \right\}.\end{aligned} \tag{12}$$

Setting (12) to zero gives rise to

$$\frac{e^2(k) - \eta^2(k) - e_a(k)\eta(k)}{\mu \|\boldsymbol{u}(k)\|^2 e^2(k)} = \exp\left(-\frac{e^2(k)}{2\sigma^2(k)}\right). \tag{13}$$

The squared kernel width $\sigma^2(k)$ is obtained by taking the logarithm of both sides of (13)

$$\sigma^2(k) = \frac{-e^2(k)}{2\ln\left[\frac{e^2(k) - \eta^2(k) - e_a(k)\eta(k)}{\mu \|\boldsymbol{u}(k)\|^2 e^2(k)}\right]}. \tag{14}$$

We cannot use (14) directly since the noise realization $\eta(k)$ is unavailable. To resolve this problem, we replace $\eta^2(k)$ with the noise variance σ_η^2. We also set the product $e_a(k)\eta(k)$ to zero. The later step is based on a widely used assumption, that is, the noise-free a *priori* error signal $e_a(k)$ is independent of the background noise $\eta(k)$ [20]. Therefore, we have $E\{e_a(k)\eta(k)\} = 0$, and introducing the approximation $e_a(k)\eta(k) \approx 0$ is reasonable because on average it is zero. As a result, we arrive at

$$\sigma^2(k) = \frac{-e^2(k)}{2\ln\left[\frac{e^2(k) - \sigma_\eta^2}{\mu \|\boldsymbol{u}(k)\|^2 e^2(k)}\right]}. \tag{15}$$

To smoothly update the squared kernel width, motivated by the moving average method [21], we employ an alternative manner as follows

$$\sigma^2(k) = \begin{cases} \alpha\sigma^2(k-1)+ \\ \quad (1-\alpha)\min\left(\frac{-e^2(k)}{2\ln[\chi(k)]}, \sigma^2(k-1)\right), & \text{if } 0<\chi(k)<1 \\ \sigma^2(k-1), & \text{otherwise,} \end{cases} \tag{16}$$

where α is the smoothing factor that is close to one, and

$$\chi(k) = \frac{e^2(k) - \sigma_\eta^2}{\mu\|\boldsymbol{u}(k)\|^2 e^2(k)}. \tag{17}$$

Remark 1. As observed in (16), to ensure a positive squared kernel width $\sigma^2(k)$, the proposed kernel width is updated when $0 < \chi(k) < 1$. Moreover, with (16), we not only guarantee the decrease of $\sigma^2(k)$, but also prevent large fluctuations of $\sigma^2(k)$, e.g., due to the impulsive noise. If the variance σ_η^2 of the background noise is not known in advance, it can be estimated using an online learning method. For example, in [22], based on establishing the relationship between the error power and the output-noise power, the variance of the background noise is estimated by employing the moving average strategy.

Remark 2. From (16), the update of the squared kernel width $\sigma^2(k)$ mainly relies on $\chi(k)$. As noted in (17), $\chi(k)$ depends on the instantaneous quantities $e^2(k)$, $\|\boldsymbol{u}(k)\|^2$, and the statistical value σ_η^2. In the case of noise with time-varying characteristics, the learning strategy in [22] can also be used to estimate the time-varying noise variance. Therefore, our algorithm is also applicable in the case of the non-stationary background noise.

B. Practical Considerations

Since the sequence $\{\sigma^2(k)\}$ is monotonically decreasing, when the estimated vector changes suddenly, the IVKW-MCC algorithm loses its tracking capability. To address this problem, a reset mechanism designed for $\sigma^2(k)$ is put forward by learning from [23]. In [23], the reset mechanism is provided for the step size, while in our algorithm, it is made for the kernel width. Fortunately, both of them have the same monotonicity, and therefore it is feasible to replace the step size with the kernel width in the original reset method in [23]. The summary of the proposed reset mechanism is presented in Table I.

In Table I, $mod(\cdot)$ denotes the remainder operator, V_T and V_D are positive integers ($V_T > V_D$), $sort(\cdot)$ represents the ascending order operator, $\boldsymbol{Q} = diag(1, \cdots, 1, 0, \cdots, 0)$ is a diagonal matrix with its first $V_T - V_D$ elements being one, and ξ stands for a threshold value. As pointed in [23], typical values for V_T, V_D and ξ are $V_T = \tau M$ with τ ranging in [1, 3], $V_D = 0.75V_T$ and $\xi = 1$.

Note that when the impulsive noise occurs with high probability, $V_T - V_D$ should be increased to exclude noise samples of high magnitude. The condition $\Delta_k > \xi$ implies that a significant change in the estimated system is detected, and thus requiring to initialize the kernel width $\sigma(k)$ and the weight vector $\boldsymbol{w}(k)$. If the change is minor, the execution statement corresponding to the "elseif" can address the tracking issue without performing the initialization. If no change occurs, the kernel width will be updated by (16) and (17).

TABLE I
THE RESET MECHANISM

if $mod(k, V_T) = 0$
  $\boldsymbol{C} = sort\left(\frac{|e(k)|}{\|\boldsymbol{u}(k)\|^2+\epsilon}, ..., \frac{|e(k-V_T+1)|}{\|\boldsymbol{u}(k-V_T+1)\|^2+\epsilon}\right)^T$
  $ctrl_{new} = \frac{\boldsymbol{C}^T\boldsymbol{Q}\boldsymbol{C}}{V_T - V_D}$
end
$\Delta_k = (ctrl_{new} - ctrl_{old})/\sqrt{\sigma^2(k-1)}$
if $\Delta_k > \xi$
  $\sigma^2(k) = \sigma^2(0)$
  $\boldsymbol{w}(k) = \boldsymbol{0}$
elseif $ctrl_{new} > ctrl_{old}$
  $\sigma^2(k) = \sigma^2(k-1) + (ctrl_{new} - ctrl_{old})$
else
  invoke proposed scheme to update
end
$ctrl_{old} = ctrl_{new}$

C. Complexity

In Table II, we discuss the complexity of some existing algorithms and proposed IVKW-MCC algorithm. Specifically, the number of multiplications, additions, and exponents per iteration are shown. For the proposed IVKW-MCC, we omit the computational cost of the reset mechanism because it is performed only every V_T iteration and therefore, for large V_T, it is small compared to other computations.

TABLE II
COMPLEXITY OF ALGORITHMS PER ITERATION

Algorithms	Multiplications	Additions	Exponents
MCC	$2M+6$	$2M$	1
CMCC	$8M+37$	$7M+8$	6
SMCC	$2M+8$	$2M$	1
AMCC	$2M+6$	$2M+1$	1
VKW-MCC	$2M+9$	$2M+2$	1
IVKW-MCC	$3M+11$	$3M+1$	1

As can be seen, the MCC algorithm consumes the least computational cost, only requiring $2M + 6$ multiplications, $2M$ additions and 1 exponent, whereas the CMCC algorithm is the most complicated owing to the updates for two adaptive filters. As compared with the MCC algorithm, both the SMCC algorithm and the AMCC algorithm have a slight increase of multiplications and additions. The VKW-MCC algorithm requires $2M + 9$ multiplications, $2M + 2$ additions and 1 exponent. Finally, the proposed IVKW-MCC algorithm requires $M + 5$ more multiplications, $M + 1$ more additions than the MCC algorithm, which can be viewed as a moderate increase.

IV. NUMERICAL SIMULATIONS

The proposed IVKW-MCC is measured by simulations for system identification and echo cancellation scenarios. The unknown system and adaptive filters have the same number of taps. The background noise $\eta(k)$ and the impulsive noise $v(k)$ that are independent of each other are added to the signal of interest $\boldsymbol{w}_o^T\boldsymbol{u}(k)$. The background noise is a zero-mean white Gaussian noise sequence. The impulsive noise is modeled as a Bernoulli-Gaussian process, i.e. $v(k) = q(k)h(k)$, where

$q(k)$ is a zero-mean white Gaussian process with variance $\sigma_q^2 = 1000E[(\boldsymbol{w}_o^T\boldsymbol{u}(k))^2]$, and $h(k)$ stands for a Bernoulli process with the probability mass function $P(h(k)=1)=P_r$ and $P(h(k)=0)=1-P_r$ [19]. In the simulation, we set $\alpha = 0.99$ obtained by training for the proposed IVKW-MCC algorithm, $V_T = 3M$, $V_D = 0.75V_T$, and $\xi = 1$. The normalized mean-square deviation (NMSD) defined as $10\log_{10}[\|\boldsymbol{w}_o - \boldsymbol{w}(k)(k)\|^2/|\boldsymbol{w}_o\|^2]$, is used as a measure of the algorithm performance. The NMSD curves are obtained by averaging over 100 simulation trials.

A. System Identification

For all system identification experiments, the unknown vector $\boldsymbol{w}_o$ is randomly generated with $M = 128$ and its taps obey the zero-mean normal distribution. A zero-mean white Gaussian sequence with unit variance is utilized as the input signal.

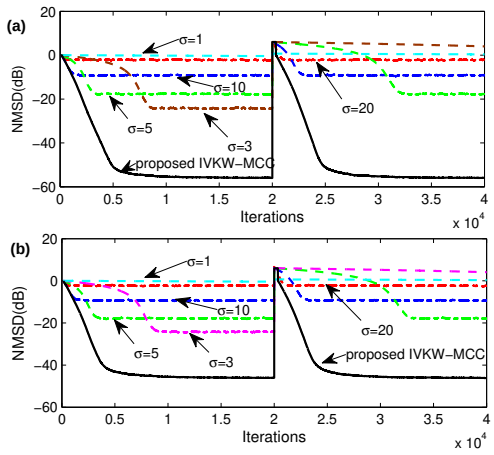

Fig. 1. NMSD curves of the MCC algorithm with different fixed kernel widths and the proposed IVKW-MCC algorithm. The unknown vector $\boldsymbol{w}_o$ changes to $-\boldsymbol{w}_o$ at iteration 2×10^4, and $P_r = 0.5$. For the MCC algorithm: $\mu = 0.01$. For the IVKW-MCC algorithm: $\mu = 0.01, \alpha = 0.99, \sigma(0) = 20$. (a) $\sigma_\eta^2 = 10^{-3}$. (b) $\sigma_\eta^2 = 10^{-2}$.

We firstly implement the comparison between the IVKW-MCC algorithm and the MCC algorithm using different fixed kernel widths σ, as shown in Fig. 1. As can be seen, with the increase of the kernel width σ, the MCC algorithm achieves faster convergence, but it also shows a higher steady-state misalignment, which implies that the MCC algorithm encounters the conflicting requirement between fast convergence rate and low steady-state misalignment. In comparison, the proposed IVKW-MCC algorithm improves both the convergence rate and steady-state misalignment. Moreover, the reset mechanism retains a desirable tracking capability for the IVKW-MCC algorithm.

We then examine the performance of some known MCC-type algorithms and proposed IVKW-MCC algorithm under different probabilities P_r and different variances σ_η^2, depicted in Fig. 2. For the CMCC algorithm, we set $\sigma = 5, \mu_a = 4.5, \beta = 0.8$ [16], for the SMCC and AMCC algorithms, we select $\sigma = 2$ [17], [18], and for the VKW-MCC algorithm, we choose $\sigma_0 = k_\sigma = 20, \alpha = 0.98, N_w = 26$ [19]. The

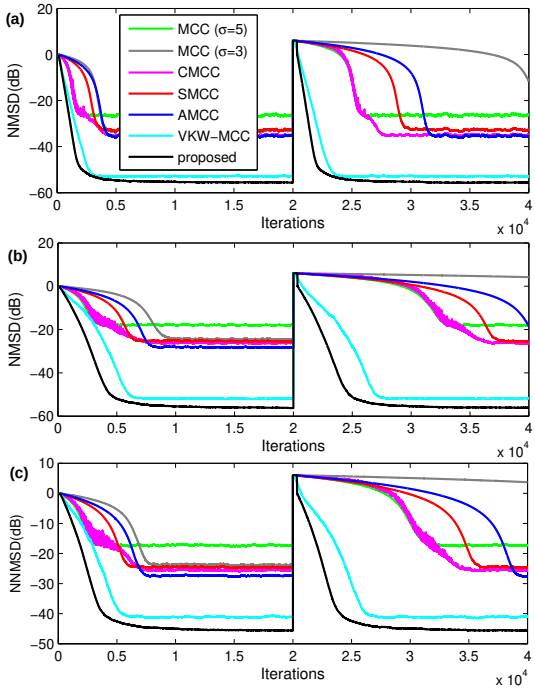

Fig. 2. NMSD curves of some known algorithms and proposed IVKW-MCC algorithm. The unknown vector $\boldsymbol{w}_o$ changes to $-\boldsymbol{w}_o$ at iteration 2×10^4. (a) $P_r = 0.1$, $\sigma_\eta^2 = 10^{-3}$. (b) $P_r = 0.5$, $\sigma_\eta^2 = 10^{-3}$. (c) $P_r = 0.5$, $\sigma_\eta^2 = 10^{-2}$

remaining parameters are set as: $\mu = 0.01$ for the MCC, $\mu_1 = 0.01, \mu_2 = 0.003$ for the CMCC, $\mu = 0.05$ for the SMCC and AMCC algorithms, $\mu = 0.01$ for the VKW-MCC, and $\mu = 0.01, \alpha = 0.99, \sigma(0) = 20$ for the proposed IVKW-MCC algorithm. As can be observed, among the known algorithms, it is quite clear that the CMCC and VKW-MCC algorithms outperform the MCC, SMCC and AMCC algorithms, while the SMCC and AMCC algorithms perform better than the MCC algorithm with $\sigma = 3$. Importantly, the proposed IVKW-MCC algorithm is superior to other algorithms in the convergence rate and steady-state misalignment, which is more obvious when the impulsive noise occurs with a higher probability such as $P_r = 0.5$.

B. Echo Cancellation

We here test the performance of various algorithms in an echo cancellation application. The impulse response of the sparse echo channel with $M = 128$ is depicted in Fig. 3(a), and the speech input sampled at 8kHz is illustrated in Fig. 3(b). The step size of $\mu = 0.01$ is set for the MCC algorithm, $\mu_1 = 0.05, \mu_2 = 0.01$ for the CMCC, and $\mu = 0.05$ for other algorithms including the proposed algorithm. The remaining parameters for all algorithms are consistent with that in the system identification except $\sigma = 2$ is set for the CMCC and $\sigma(0) = 5$ is chosen for the IVKW-MCC algorithm. The probability of the impulsive noise is $P_r = 0.1$, and the signal-to-noise (SNR) of the background noise $\eta(k)$ is 30dB. The learning curves regarding these algorithms are presented in Fig. 4. As can be seen, the proposed IVKW-MCC algorithm provides better performance as compared to other algorithms.

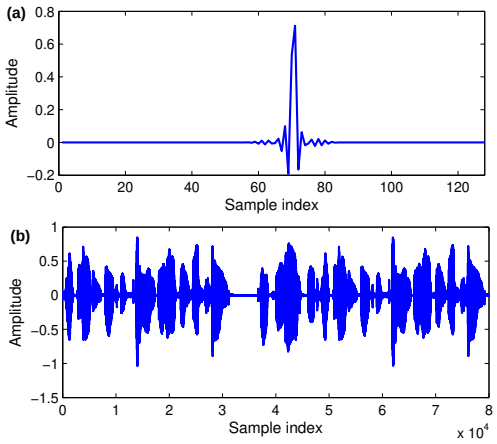

Fig. 3. Impulse response of the sparse echo channel (top) and speech signal (bottom)

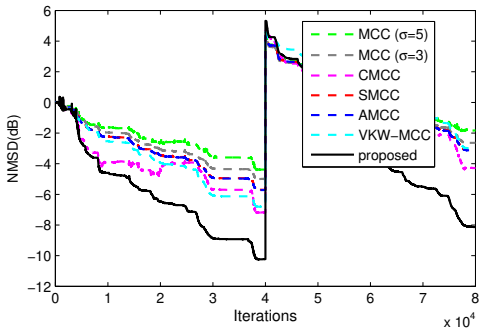

Fig. 4. NMSD curves of some known algorithms and proposed IVKW-MCC algorithm in the acoustic echo cancellation scenario, where the unknown vector $\boldsymbol{w}_o$ changes to $-\boldsymbol{w}_o$ at the middle of iterations.

V. Conclusion

This paper proposes a novel variable kernel width method to overcome the shortcoming on how to choose a reliable kernel width in the MCC algorithm. The squared kernel width is obtained by minimizing the squared deviation using the stochastic gradient method. To implement the proposed scheme, the moving average strategy is utilized to update the squared kernel width. In addition, we design a reset mechanism to improve the tracking capability when the estimated system encounters a sudden change. Simulations conducted in the system identification and echo cancellation have demonstrated that the proposed IVKW-MCC algorithm exhibits robustness against impulsive noise and is superior to other established TKW schemes.

Acknowledgment

The work of L. Shi and H. Zhao was partially supported by National Science Foundation of PR China (Grant: 61571374, 61871461, and 61433011). The work of Y. Zakharov was supported by the U.K. EPSRC (Grants EP/P017975/1 and EP/R003297/1).

References

[1] R. H. Kwong and E. W. Johnston, "A variable step size LMS algorithm," *IEEE Trans Signal Processing*, vol. 40, no. 7, pp. 1633–1642, 1992.

[2] H. Deng and M. Doroslovacki, "Proportionate adaptive algorithms for network echo cancellation," *IEEE Transactions on Signal Processing*, vol. 54, no. 5, pp. 1794–1803, 2006.

[3] S. M. Kuo, S. Mitra, and W. S. Gan, "Active noise control system for headphone applications," *IEEE Transactions on Control Systems Technology*, vol. 14, no. 2, pp. 331–335, 2006.

[4] C. Richard, J. C. M. Bermudez, and P. Honeine, "Online prediction of time series data with kernels," *IEEE Transactions on Signal Processing*, vol. 57, no. 3, pp. 1058–1067, 2009.

[5] Y. Zhang, Y. V. Zakharov, and J. Li, "Soft-decision-driven sparse channel estimation and turbo equalization for mimo underwater acoustic communications," *IEEE Access*, vol. 6, pp. 4955–4973, 2018.

[6] H. Cho, S. Park, K. Lee, and N. Kim, "A novel multiple-channel active noise control approach with neural secondary-path model for interior acoustic noise attenuation of railway train systems," *IET Signal Processing*, vol. 6, no. 8, pp. 772–780, 2012.

[7] V. Mathews and S. Cho, "Improved convergence analysis of stochastic gradient adaptive filters using the sign algorithm," *IEEE Transactions on Acoustics Speech & Signal Processing*, vol. 35, no. 4, pp. 450–454, 1987.

[8] J. A. Chambers, O. Tanrikulu, and A. G. Constantinides, "Least mean mixed-norm adaptive filtering," *Electronics Letters*, vol. 30, no. 19, pp. 1574–1575, 1994.

[9] Y. Zou, S.-C. Chan, and T.-S. Ng, "Least mean m-estimate algorithms for robust adaptive filtering in impulse noise," *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, vol. 47, no. 12, pp. 1564–1569, 2000.

[10] J. Ni, X. Chen, and J. Yang, "Two variants of the sign subband adaptive filter with improved convergence rate ," *Signal Processing*, vol. 96, no. 5, pp. 325–331, 2014.

[11] H. Zayyani, "Continuous mixed -norm adaptive algorithm for system identification," *IEEE Signal Processing Letters*, vol. 21, no. 9, pp. 1108–1110, 2014.

[12] Y. Zhou, S. C. Chan, and K. L. Ho, "New sequential partial-update least mean m-estimate algorithms for robust adaptive system identification in impulsive noise," *IEEE Transactions on Industrial Electronics*, vol. 58, no. 9, pp. 4455–4470, 2011.

[13] J. C. Principe, *Information Theoretic Learning: Renyi's Entropy and Kernel Perspectives*. Springer Publishing Company, Incorporated, 2010.

[14] W. Liu, P. P. Pokharel, and J. C. Príncipe, "Correntropy: Properties and applications in non-gaussian signal processing," *IEEE Transactions on Signal Processing*, vol. 55, no. 11, pp. 5286–5298, 2007.

[15] A. Singh and J. C. Principe, "Using correntropy as a cost function in linear adaptive filters," in *IEEE International Joint Conference on Neural Networks (IJCNN)*, 2009, pp. 2950–2955.

[16] L. Shi and Y. Lin, "Convex combination of adaptive filters under the maximum correntropy criterion in impulsive interference," *IEEE Signal Processing Letters*, vol. 21, no. 11, pp. 1385–1388, 2014.

[17] W. Wang, J. Zhao, H. Qu, B. Chen, and J. C. Principe, "A switch kernel width method of correntropy for channel estimation," in *IEEE International Joint Conference on Neural Networks (IJCNN)*, 2015, pp. 1–7.

[18] ——, "An adaptive kernel width update method of correntropy for channel estimation," in *IEEE International Conference on Digital Signal Processing (DSP)*, 2015, pp. 916–920.

[19] F. Huang, J. Zhang, and S. Zhang, "Adaptive filtering under a variable kernel width maximum correntropy criterion," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 64, no. 10, pp. 1247–1251, 2017.

[20] 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.

[21] J. Shin, J. Yoo, and P. Park, "Variable step-size sign subband adaptive filter," *IEEE Signal Processing Letters*, vol. 20, no. 2, pp. 173–176, 2013.

[22] J. Yoo, J. Shin, and P. Park, "A band-dependent variable step-size sign subband adaptive filter," *Signal Processing*, vol. 104, pp. 407–411, 2014.

[23] L. R. Vega, H. Rey, J. Benesty, and S. Tressens, "A new robust variable step-size NLMS algorithm," *IEEE Transactions on Signal Processing*, vol. 56, no. 5, pp. 1878–1893, 2008.