IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received November 27, 2019, accepted December 12, 2019, date of publication December 16, 2019,
date of current version December 26, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2959919

Affine Projection Algorithm by Employing Maximum Correntropy Criterion for System Identification of Mixed Noise

XIAODING WANG AND JUN HAN
Ocean College, Zhejiang University, Hangzhou 310027, China
Corresponding author: Jun Han (15865582181@163.com)

This work was supported in part by the Postgraduate Project of Zhejiang University.

ABSTRACT The adaptive algorithms have been widely studied in Gaussian environment. However, the impulsive noise and other non-Gaussian noise may largely deteriorate the performance of algorithm in practical applications. To address this problem, in this paper, we propose two novel adaptive algorithms for system identification problem with mixed noise scenarios. Both proposed algorithms are based on the framework of the affine projection (AP) algorithm. The first proposed algorithm, termed as VS-APMCCA, combines variable step-size (VS) strategy and maximum correntropy criterion (MCC) to obtain improved performance. For further performance improvement, the VC-VS-APMCCA is developed, which is based on the variable center (VC) scheme of MCC. The convergence analysis of the VC-VS-APMCCA is conduced. Finally, simulation results demonstrate the superior performance of the VS-APMCCA and VC-VS-APMCCA.

INDEX TERMS Affine projection algorithm, variable step-size, maximum correntropy criterion (MCC), variable center (VC), mixed noise.

I. INTRODUCTION

Adaptive filtering technique has been successfully applied in diverse signal processing fields, such as active noise control, system identification, and adaptive echo cancellation [1]–[3]. The most popular algorithm used in such a technique is the least-mean-square (LMS) algorithm because of its simplicity [4]. The LMS algorithm and its improved version can obtain stable filtering performance in Gaussian noise environment, and they also promote the development of adaptive algorithms [5]–[7].

To speed up the convergence rate of the LMS algorithm, the affine projection (AP) algorithm was proposed [8]. The AP algorithm enjoys improved performance under the colored input signal [9], [10]. In [11], a novel affine projection sign algorithm (APSA) was presented by making use of the L_1-norm algorithm for system identification in impulsive scenario. Following this work, several APSA-based algorithms were proposed by using variable step-size (VS) scheme [12]–[14], convex combination strategy [15] and so on.

The associate editor coordinating the review of this manuscript and approving it for publication was Chuan Li.

The impulsive noise has short time duration and large amplitude, and as such, leads to 'outliers' of the signal [16], [17]. Such outliers can cause parameter drift of adaptive algorithm and even make the filter fails to work [17]–[20]. Moreover, in many applications, the impulsive noise may be mixed with other noises, such as Gaussian and uniform [21]. Therefore, filtering such non-Gaussian noise became an important topic in the field of adaptive signal processing.

During the past two decades, several approaches were developed for performance improvement in the presence of impulsive noise [22]–[24]. The sign algorithm and clipped algorithm were proposed for stable performance when large outliers exist in the system [25], [26]. However, they suffer from slow convergence rate. In [27], the M-estimate method was first proposed which is based on robust estimation. This method is essentially a limiting idea to achieve stable filtering performance under impulsive environment. Besides, this method can be easily extended to other classes of the adaptive filter. Representative methods include recursive M-estimate algorithm [28], and active noise control based on M-estimate algorithm [29], [30]. The fractional lower-order moments (FLOM) strategy is also an alternative

method to deal with impulsive noise [31]. The celebrated least mean pth (LMP) algorithm was derived from this scheme [31], [32]. To further enhance the performance, the normalization version of the LMP and some variants were also developed in [33], [34] and references therein. The main bottleneck of the FLOM strategy is: it needs the *a prior* information of the impulsive noise, and then selects a suitable value of p to achieve improved performance [32]. Such a drawback largely prohibits its practical applications.

Compared to conventional mean square error (MSE) criterion, the information theoretic learning (ITL) avenue can extract more information from the original signal and as such, it has attracted much attention during the past decade [35]–[37]. According to the theory of ITL, two types of algorithms were extensively investigated in the previous studies, namely minimum error entropy criterion (MEEC) [38], [39] and maximum correntropy criterion (MCC) [40], [41]. Among these, the MEEC has proven to be an improved performance when both input and output signals are contaminated by noise [38]. The MCC has a wider range of applications as compared with the MEEC. Chen et al. proved that the MCC is a smoothed maximum a posteriori estimation and has robust performance in impulsive noise [35]. Owing to these good properties, it has been applied to active noise control [42], system identification [41], and sparse subspace learning [43].

The well known MCC method can effectively combat impulsive noise and a great many of researches were performed to improve the performance of MCC. In [44], a risk-sensitive loss in kernel space was proposed, motivated by risk-sensitive criterion. This new cost function is benefited by the risk-sensitive criterion and kernel trick, and can lead to a faster initial convergence rate. Consider the filtering task with complex-valued data, some complex MCC algorithms were introduced, which provide a new solution method for complex signal processing [45], [46]. In 2016, generalized MCC (GMCC) and its corresponding applications were proposed by Chen *et al.* [47]. The GMCC can be interpreted as a generalized form of the classical MCC, and it was also proposed as an optimality criterion in estimation related problems [47].

Very recently, the MCC with variable center (VC) was proposed for enhancing performance of MCC [48]. In conventional MCC, the Gaussian kernel function with zero-mean was employed. However, such function may become sub-optimal in certain circumstance. The MCC with VC scheme exploits a more generalized form of MCC and offers improved performance with slightly increased computational complexity. Moreover, such a method was applied to online learning for refining the kernel width and center location [49].

Motivated by [50], we propose a novel AP algorithm by using MCC approach and develop a new VS scheme, resulting in VS-APMCCA. In comparison with the FLOM method, the VS-APMCCA uses the high-order error power criterion (HOEP) and can obtain refined performance in mixed noise. Before that, a novel AP algorithm based on MCC has been developed [51]. The difference between our proposed VS-APMCCA and the algorithm in [51] is that the proposed algorithm is designed for mixed noise and the latter algorithm is for sparse system. The algorithm [51] has inverse operation. Therefore, it has increased computational burden for system identification. In contrast, the VS-APMCCA incorporates the VS strategy to overcome the conflict requirement under the fixed step size. While using the variable step size strategy, we also consider reducing the computational complexity of the algorithm. In the previous studies, the convex combination strategy has been proposed to address the conflicting requirement between fast convergence rate and small misadjustment [52]. In general, the convex combination scheme requires more than twice the complexity of a single filter. This drawback may hinder its practical application. The proposed algorithm based on VS scheme can significantly reduce the computational complexity as compared with the convex combination algorithm. Its complexity is even close to the conventional MCC algorithm [41] and as such, it can be easy to implement in engineering. For performance improvement, a VC-VS-APMCCA is proposed based on VC scheme to further enhance the identification accuracy of the adaptive algorithm. In summarize, our contributions can be listed as follows.

1) A novel VS-APMCCA is proposed by using HOEP and VS scheme, which enables the proposed VS-APMCCA to obtain anti-impulsive ability of MCC algorithm and the fast convergence rate of the APA;

2) A novel VC-VS-APMCCA is developed as an added contribution, with an adjustable center position;

3) Convergence behavior of the VC-VS-APMCCA is analyzed;

4) Simulations are performed to demonstrate the effectiveness of the VS-APMCCA and VC-VS-APMCCA.

The content of this work is organized in the following. In Section II, the system identification model used in this paper is described. In Section III, the derivations of the VS-APMCCA and VC-VS-APMCCA are given in detail. In Section IV, the convergence behavior of the VC-VS-APMCCA is provided. Simulation results are performed in Section V, and some conclusions are drawn in Section VI.

II. PROBLEM FORMULATION

A. SYSTEM MODEL

System identification is an art of establishing the relationship of input-output. The system identification problem can be solved by the least-square method, the evolutionary algorithm, and adaptive filter [53]–[55]. Among these methods, the adaptive filter can provide more flexible modeling capability and it has been extensively studied [56]. Fig. 1 shows the block diagram of the system identification problem with adaptive filter, where $x(k)$ denotes the input signal at time instant k, $y(k)$ denotes the output signal which is used to approximate the desired signal $d(k)$, $\xi(k)$ denotes the

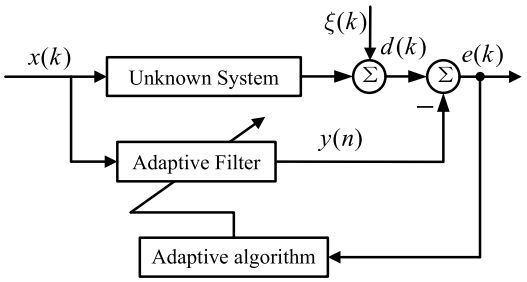

FIGURE 1. System identification problem by using adaptive filter.

noise signal, and $e(k)$ denotes the error signal. Considering the above system model, the desired signal $d(k)$ can be expressed as

$$d(k) = \boldsymbol{x}^T(k)\boldsymbol{w}_o + \xi(k) \tag{1}$$

where $\boldsymbol{x}(k) = [x(k), x(k-1), \ldots, x(k-M+1)]^T$, M stands for the filter length, $\boldsymbol{w}_o$ represents the unknown parameter vector with the size M, and T stands for the transpose operation.

At time instant k, the error signal $e(k)$ can be given by

$$\begin{aligned} e(k) &= d(k) - y(k) \\ &= d(k) - \boldsymbol{x}^T(k)\boldsymbol{w}(k) \end{aligned} \tag{2}$$

where $\boldsymbol{w}(k)$ is the weight vector of adaptive algorithm at time instant k. In this expression, the output signal can be computed by $y(k) = \boldsymbol{x}^T(k)\boldsymbol{w}(k)$.

B. A REVIEW OF AP ALGORITHM

The AP algorithm can effectively accelerate the convergence rate as compared with basic LMS algorithm. The AP algorithm needs to define the *a posteriori* error, which can be defined by

$$e_p(k) = d(k) - \boldsymbol{x}^T(k)\boldsymbol{w}(k+1). \tag{3}$$

At every iteration k, it minimizes the following optimization problem [56]

$$\begin{aligned} &\underset{\boldsymbol{w}(k+1)}{\text{minimize}} \quad \|\boldsymbol{w}(k+1) - \boldsymbol{w}(k)\|_2^2 \\ &\text{subject to} \quad \boldsymbol{d}(k) - \boldsymbol{X}^T(k)\boldsymbol{w}(k+1) = 0 \end{aligned} \tag{4}$$

where $\|\cdot\|_2$ denotes the L_2-norm, $\boldsymbol{d}(k) = [d(k), d(k-1), \ldots, d(k-P+1)]^T$ denotes the desired signal vector which combines the past P desired signal from past times, and $\boldsymbol{X}(k) = [\boldsymbol{x}(k), \boldsymbol{x}(k-1), \ldots, \boldsymbol{x}(k-P+1)]$ denotes the input signal matrix with the size of $M \times P$, where P represents the projection order. The update equation of basic AP algorithm is expressed as

$$\boldsymbol{w}(k+1) = \boldsymbol{w}(k) + \mu\boldsymbol{X}(k)\left[\boldsymbol{X}^T(k)\boldsymbol{X}(k)\right]^{-1}\boldsymbol{e}(k) \tag{5}$$

where μ is the learning rate and $\boldsymbol{e}(k) = [e(k), e(k-1), \ldots e(k-P+1)]^T$.

C. A REVIEW OF MCC

Considering the definition of the Gaussian kernel function in MCC

$$\kappa(D, Y) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{|D-Y|^2}{2\sigma^2}\right) \tag{6}$$

where D and Y are random variables with the same dimensions, and σ denotes the kernel bandwidth (kernel size). Fig. 2 depicts the $J(k) = \kappa(d, y)$ in the joint space of d and y.

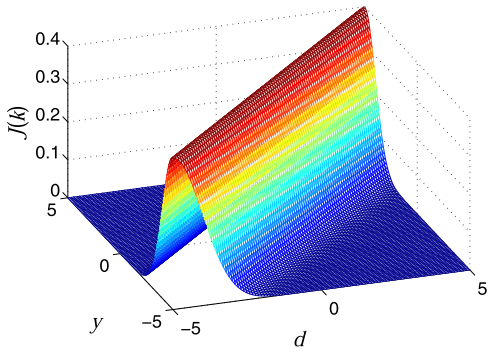

FIGURE 2. Gaussian kernel function in the joint space.

The MCC has the ability of dividing space, and it can be derived according to the following cost function [41]:

$$\begin{aligned} J(k) &= E\left[\exp\left(-\frac{e^2(k)}{2\sigma^2}\right)\right] \\ &\approx \exp\left(-\frac{e^2(k)}{2\sigma^2}\right) \end{aligned} \tag{7}$$

where $E[\cdot]$ denotes the expectation operation. Using the stochastic gradient method, one can obtain the gradient of (7)

$$\frac{\partial J(k)}{\partial \boldsymbol{w}(k)} = -\exp\left(-\frac{e^2(k)}{2\sigma^2}\right)\boldsymbol{x}(k)e(k). \tag{8}$$

Thus, the update rule of the MCC is

$$\boldsymbol{w}(k+1) = \boldsymbol{w}(k) + \mu\exp\left(-\frac{e^2(k)}{2\sigma^2}\right)\boldsymbol{x}(k)e(k). \tag{9}$$

As we noticed, the AP algorithm is based on MSE criterion to obtain a refined performance. Hence, it is natural to take HOEP criterion into account. Unlike the MSE rule, HOEP uses high-order errors to define cost functions, such as $E[e^3(k)]$, $E[e^4(k)]$, and so on. Since the Taylor expansion of $\exp(e^2(k))$ can be expressed as $1 + \frac{1}{1!}e^2(k) + \frac{1}{2!}e^4(k) + \ldots + h.o.t.$, where h.o.t. denotes the higher-order terms of the remainder of the Taylor series expansion. Therefore, the cost function of the form $\exp(e^2(k))$ can be regarded as a kind of HOEP criterion. Pei and Tseng have demonstrated that the HOEP criterion is preferred when the unknown system is contaminated with impulsive noise [57]. Considering the good suppression performance of the MCC, we utilize the MCC as the HOEP criterion. In the next section, we propose novel VS-APMCCA and VC-VS-APMCCA for performance improvement for mixed-noise scenarios.

III. PROPOSED ALGORITHMS

Due to the good suppression performance of the MCC, we utilize the MCC as the HOEP criterion. A class of APs based on MCC for performance improvement under mixed-noise scenarios are proposed in this section.

A. PROPOSED VS-APMCCA

Considering affine projection space, based on the HOEP criterion, we re-define the optimization problem in (4) as

$$\begin{aligned} &\underset{\boldsymbol{w}(k+1)}{\text{minimize}} \quad \exp\Big(-\frac{\|\boldsymbol{e}_p(k)\|_2^2}{2\sigma^2}\Big) \\ &\text{subject to} \quad \|\boldsymbol{w}(k+1)-\boldsymbol{w}(k)\|_2^2 \le \rho^2 \end{aligned} \tag{10}$$

where ρ^2 is a threshold and $\boldsymbol{e}_p(k) = [e_p(k), e_p(k-1), \ldots, e_p(k-P+1)]^T$. In our derivation, the threshold will be absorbed into the step size. Note that such threshold is similar to the set-membership algorithms and such threshold can be selected according to [58], [59]. To show the cost function more intuitively, Fig. 3 shows the cost function of MCC and its corresponding gradient value. As can be seen, the large value of kernel size leads to a flatter cost function shape. The increase in amplitude of the error signal can lead to a small value of gradient. This property can offer stable and robust performance in 'outliers' environment. Utilizing the constraint parameter ζ, we draw the cost function as

$$J(k)=\exp\Big(-\frac{\|\boldsymbol{e}_p(k)\|_2^2}{2\sigma^2}\Big)+\zeta\Big(\|\boldsymbol{w}(k+1)-\boldsymbol{w}(k)\|_2^2-\rho^2\Big). \tag{11}$$

By taking gradient of (11) with respect to $\boldsymbol{w}(k+1)$, thus yielding

$$\begin{aligned} \frac{\partial J(k)}{\partial \boldsymbol{w}(k+1)} &= 2\zeta\Big(\boldsymbol{w}(k+1)-\boldsymbol{w}(k)\Big) \\ &\quad -\frac{2}{2\sigma^2}\sum_{m=0}^{P-1}\exp\Big(-\frac{\|\boldsymbol{e}_p(k-m)\|_2^2}{2\sigma^2}\Big) \\ &\quad \times \boldsymbol{x}(k-m)e_p(k-m) \\ &= 2\zeta\Big(\boldsymbol{w}(k+1)-\boldsymbol{w}(k)\Big)-\frac{1}{\sigma^2}\boldsymbol{X}(k)\boldsymbol{\epsilon}(k) \end{aligned} \tag{12}$$

where

$$\boldsymbol{\epsilon}(k)=\exp\Big(-\frac{\|\boldsymbol{e}_p(k)\|_2^2}{2\sigma^2}\Big)\boldsymbol{e}_p(k). \tag{13}$$

Letting (12) equals to zero, we have

$$\boldsymbol{w}(k+1)=\boldsymbol{w}(k)+\frac{1}{2\zeta\sigma^2}\boldsymbol{X}(k)\boldsymbol{\epsilon}(k). \tag{14}$$

Substituting (14) into (10) results in

$$\boldsymbol{w}(k+1)=\boldsymbol{w}(k)+\mu_f\frac{\exp\Big(-\frac{\|\boldsymbol{e}_p(k)\|_2^2}{2\sigma^2}\Big)}{2\zeta\sigma^2}\boldsymbol{X}(k)\boldsymbol{e}_p(k) \tag{15}$$

where $\mu_f > 0$ is the scalar parameter. Then, we can combine constant terms σ^2, ζ, and μ_f, at below

$$\frac{\mu_f}{2\zeta\sigma^2}=\frac{\|\boldsymbol{e}_p(k)\|_2^2}{\sigma^2}. \tag{16}$$

In this expression, we take $\frac{\mu_f}{\zeta}$ as the L_2-norms of the error signal. This equation can be regarded as a VS strategy. When the error is large, the step size of the algorithm becomes large, and the convergence is accelerated. The step size of the small algorithm becomes smaller, resulting in small steady-state error. Therefore, the adaptation of the VS-APMCCA can be expressed as follow:

$$\boldsymbol{w}(k+1)=\boldsymbol{w}(k)+\mu(k)\boldsymbol{X}(k)\boldsymbol{e}_p(k) \tag{17}$$

where

$$\mu(k)=\frac{\|\boldsymbol{e}_p(k)\|_2^2}{\sigma^2}\exp\Big(-\frac{\|\boldsymbol{e}_p(k)\|_2^2}{2\sigma^2}\Big). \tag{18}$$

We can observe from (17) and (18), the proposed VS-APMCCA can be regarded as the algorithm based on the VS scheme. Note that, at each iteration k, the *a posteriori* error $\boldsymbol{e}_p(k)$ can be calculated via the value of weight vector $\boldsymbol{w}(k+1)$. However, $\boldsymbol{e}_p(k)$ is unavailable in the current iteration k. To overcome this problem, the *a priori* error $\boldsymbol{e}(k)$ is employed to replace the *a posteriori* error $\boldsymbol{e}_p(k)$ during weight adaptation. This approximation is widely applied to

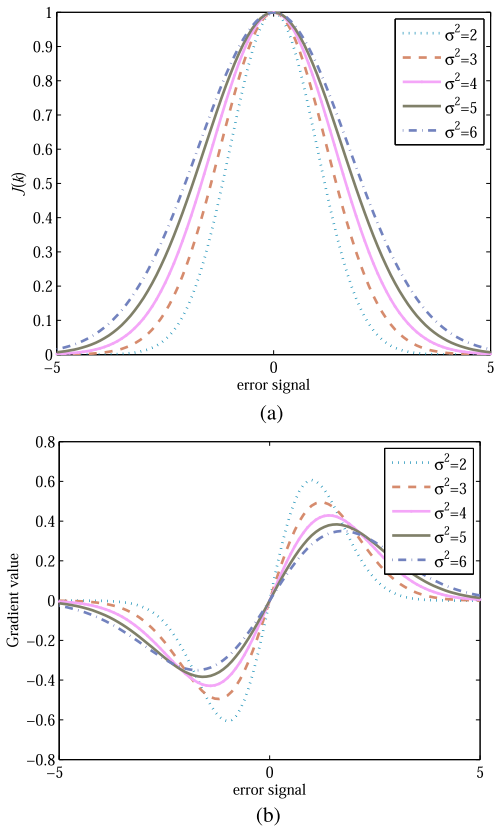

FIGURE 3. Cost function of the MCC and its gradient. (a) MCC cost function with different selections of σ^2, (b) Gradient of MCC cost function with different selections of σ^2.

the derivation of the APA-based algorithms, and the similar method can be found in [11]. Finally, we arrive at

$$\boldsymbol{w}(k+1) = \boldsymbol{w}(k) + \mu(k)\boldsymbol{X}(k)\boldsymbol{e}(k) \tag{19}$$

where

$$\mu(k) = \frac{\|\boldsymbol{e}(k)\|_2^2}{\sigma^2} \exp\Big(-\frac{\|\boldsymbol{e}(k)\|_2^2}{2\sigma^2}\Big). \tag{20}$$

In summary, the detailed VS-APMCCA is outlined in Table 1.

B. PROPOSED VC-VS-APMCCA

To further enhance the performance of the VS-APMCCA, in this section, we develop a novel VC-VS-APMCCA by using VC strategy. The previous studies have demonstrated that the VC strategy can maintain the convergence rate of the algorithm and can provide an accelerated convergence rate. To illustrate the cost function of MCC with VC, Fig. 4 depicts the cost function and its correspondingly gradient value. By using VC scheme, the cost function and gradient value translate $\boldsymbol{c}$ units to the right. They still hold the property of conventional MCC. Fig. 5 further plots a comparison of the LMS, MCC, and MCC with VC. As can be seen from this figure, the MCC and MCC with VC can provide a stable performance when the error signal becomes large.

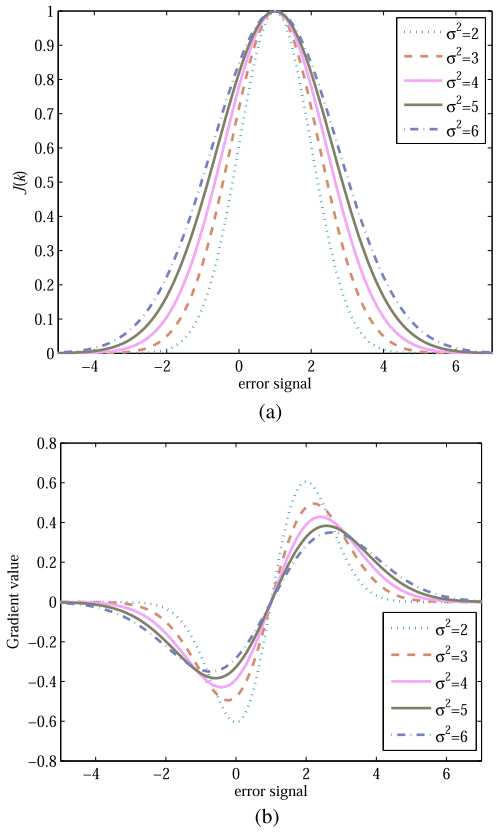

FIGURE 4. Cost function of the MCC with variable center and its gradient ($c = 1$). (a) MCC with variable center versus σ, (b) Gradient of MCC with variable center cost function versus σ.

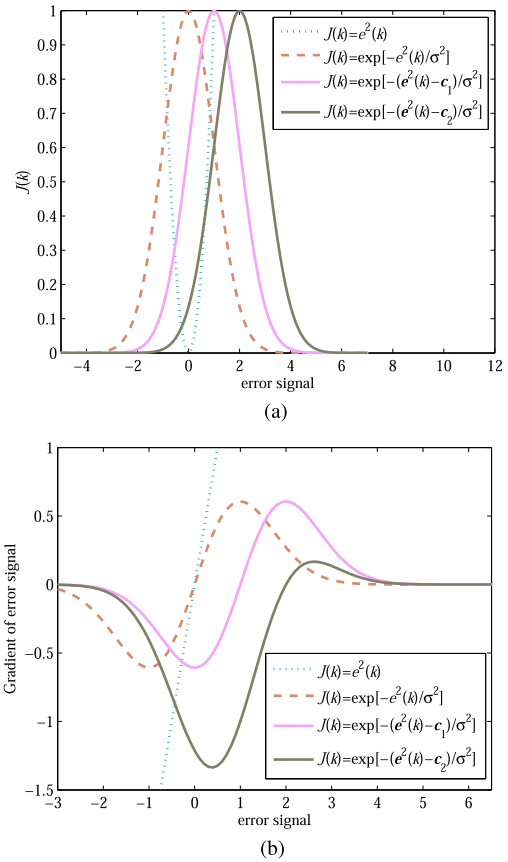

FIGURE 5. Cost function of the LMS, MCC and MCC with variable center and their gradient ($c_1 = [1\ 1\ 1\ 1\ 1\ 1\ 1\ 1]$ and $c_2 = [2\ 2\ 2\ 2\ 2\ 2\ 2\ 2]$). (a) MCC with variable center versus σ, (b) Gradient of MCC with variable center cost function versus σ.

Hence, the updated formulation of the VC-VS-APMCCA can be given by

$$\boldsymbol{w}(k+1) = \boldsymbol{w}(k) + \mu(k)\boldsymbol{X}(k)\boldsymbol{e}(k) \tag{21}$$

where

$$\mu(k) = \frac{\|\boldsymbol{e}(k)\|_2^2}{\sigma^2} \exp\Big(-\frac{\|\boldsymbol{e}(k)-\boldsymbol{c}\|_2^2}{2\sigma^2}\Big) \tag{22}$$

where $\boldsymbol{c}$ is the vector. Note that $\frac{\|\boldsymbol{e}(k)\|_2^2}{\sigma^2}$ can be interpreted as an estimation of step size, and VC scheme is only applied to Gaussian kernel. In the previous studies [48], [49], the VC strategy is only integrated into Gaussian kernel. The VC-VS-APMCCA is listed in Table 1.

C. ANALYSIS OF COMPUTATIONAL COMPLEXITY

In this subsection, we analyze the computational complexity of the proposed algorithms along with the existing algorithms. The MCC [41], least-mean mixed-norm (LMMN) [60], [61], and AP [8] algorithms are used as the comparison algorithms. We compare the additions, multiplications, divisions, and exponential operations of the five algorithms, as shown in Table 2. We can observe from this table that the LMMN algorithm, which benefits from the

TABLE 1. Summary of the VS-APMCCA and VC-VS-APMCCA.

Initialization	$\boldsymbol{w}(0) = \mathbf{0}$
Parameter	step-size μ, projection order P, filter length M, kernel size σ parameter $\boldsymbol{c}$
Loop	**For** $k = 1, 2, 3, ...$ Input $\boldsymbol{X}(k)$, $\boldsymbol{X}(k) = [\boldsymbol{x}(k), \boldsymbol{x}(k-1), \ldots, \boldsymbol{x}(k-P+1)]$ $\boldsymbol{e}(k) \triangleq \boldsymbol{y}(k) - \boldsymbol{X}^T(k)\boldsymbol{w}(k)$ $\mu(k) = \frac{\|\boldsymbol{e}(k)\|_2^2}{\sigma^2} \exp\left(-\frac{\|\boldsymbol{e}(k)\|_2^2}{2\sigma^2}\right)$ (VS-APMCCA) $\mu(k) = \frac{\|\boldsymbol{e}(k)\|_2^2}{\sigma^2} \exp\left(-\frac{\|\boldsymbol{e}(k)-\boldsymbol{c}\|_2^2}{2\sigma^2}\right)$ (VC-VS-APMCCA) $\boldsymbol{w}(k+1) = \boldsymbol{w}(k) + \mu(k)\boldsymbol{X}(k)\boldsymbol{e}(k)$ (VS-APMCCA / VC-VS-APMCCA) **End For**

TABLE 2. Comparsion of the computational complexity of five algorithms.

Algorithms	+	×	÷	Exponential operation
MCC [41]	$2M$	$2M+5$	1	1
LMMN [60]	$2M+2$	$2M+4$	0	0
AP [8]	$2P^2M + PM - P^2$	$2P^2 + 2PM + M$	0	0
VS-APMCCA	$2M$	$PM + P + 6$	2	1
VC-VS-APMCCA	$2M + P$	$PM + P + 6$	2	1

LMS and LMF algorithms, has the lowest computational complexity. It needs $2M + 2$ additions and $2M + 4$ multiplications. The MCC algorithm, based on the MCC criterion, slightly increases the computational complexity as compared with the LMMN algorithm. It requires $2M$ additions, $2M + 5$ multiplications, 1 division, and 1 exponential operation. The AP algorithm uses the past information of the input signal, and it requires $2P^2 + PM - P^2$ additions and $2P^2 + 2PM + M$ multiplications. The proposed VS-APMCCA has reduced computational complexity since it does not need the inverse operation of matrix. It requires $2M$ additions, $PM+P+6$ multiplication and 2 divisions. Because it utilizes the MCC, this algorithm needs additional 1 exponential operation per iteration. The VC-VS-APMCCA based on VC scheme, slightly increases the computational complexity as compared with the VS-APMCCA. It requires $2M+P$ additions, $PM+P+6$ multiplications, 2 divisions, and 1 exponential operation. It still has reduced computational cost as compared with celebrated AP algorithm.

IV. PERFORMANCE ANALYSIS

To begin with the analysis, we define the weight-error vector as follow:

$$\widetilde{\boldsymbol{w}}(k) \triangleq \boldsymbol{w}_o - \boldsymbol{w}(k). \tag{23}$$

Combining (23) and (21), we can utilize the weight-error vector to rewrite the updated equation as

$$\widetilde{\boldsymbol{w}}(k+1) = \widetilde{\boldsymbol{w}}(k) - \mu(k)\boldsymbol{X}(k)\boldsymbol{e}(k). \tag{24}$$

Then, defining the weighted *a priori* and *a posteriori* errors by $\boldsymbol{\epsilon}_a(k) = \boldsymbol{X}^T(k)\boldsymbol{\Sigma}\widetilde{\boldsymbol{w}}(k)$ and $\boldsymbol{\epsilon}_p(k) = \boldsymbol{X}^T(k)\boldsymbol{\Sigma}\widetilde{\boldsymbol{w}}(k+1)$, where $\boldsymbol{\Sigma}$ is the Hermitian positive-definite matrix. Multiplying both sides of (24) by $\boldsymbol{X}^T(k)\boldsymbol{\Sigma}$ from the left, we can gain a relation between the estimation errors $\boldsymbol{\epsilon}_a(k)$ and $\boldsymbol{\epsilon}_p(k)$

$$\boldsymbol{\epsilon}_p(k) = \boldsymbol{\epsilon}_a(k) - \mu(k)\boldsymbol{X}^{\mathrm{T}}(k)\boldsymbol{\Sigma}\boldsymbol{X}(k)\boldsymbol{e}(k). \tag{25}$$

Supposing that $\boldsymbol{X}^T(k)\boldsymbol{\Sigma}\boldsymbol{X}(k)$ is invertible, (25) is expressed as

$$\boldsymbol{e}(k) = \frac{1}{\mu}\left(\boldsymbol{X}^{\mathrm{T}}(k)\boldsymbol{\Sigma}\boldsymbol{X}(k)\right)^{-1}\left(\boldsymbol{\epsilon}_a(k) - \boldsymbol{\epsilon}_p(k)\right). \tag{26}$$

Inserting (26) into (24) and rearranging terms, and taking the weighted Euclidean norms on both sides, the energy conservation relation of the VC-VS-APMCCA can be established as follow:

$$\begin{aligned}\|\widetilde{\boldsymbol{w}}(k+1)\|_{\boldsymbol{\Sigma}}^2 + \boldsymbol{\epsilon}_a^T(k)\left(\boldsymbol{X}^T(k)\boldsymbol{\Sigma}\boldsymbol{X}(k)\right)^{-1}\boldsymbol{\epsilon}_a(k) \\ = \|\widetilde{\boldsymbol{w}}(k)\|_{\boldsymbol{\Sigma}}^2 + \boldsymbol{\epsilon}_p^T(k)\left(\boldsymbol{X}^T(k)\boldsymbol{\Sigma}\boldsymbol{X}(k)\right)^{-1}\boldsymbol{\epsilon}_p(k)\end{aligned} \tag{27}$$

where $\|\widetilde{\boldsymbol{w}}(k+1)\|_{\boldsymbol{\Sigma}}^2$ denotes the weighted squared norm $\widetilde{\boldsymbol{w}}^T(k+1)\boldsymbol{\Sigma}\widetilde{\boldsymbol{w}}(k+1)$. By combining the expressions (25) and (27) yields

$$\begin{aligned}\|\widetilde{\boldsymbol{w}}(k+1)\|_{\boldsymbol{\Sigma}}^2 = \|\widetilde{\boldsymbol{w}}(k)\|_{\boldsymbol{\Sigma}}^2 - \mu\boldsymbol{\epsilon}_a^T(k)\boldsymbol{e}(k) \\ - \mu\boldsymbol{e}^T(k)\boldsymbol{\epsilon}_a(k) \\ + \mu^2\boldsymbol{e}^T(k)\boldsymbol{X}^T(k)\boldsymbol{\Sigma}\boldsymbol{X}(k)\boldsymbol{e}(k).\end{aligned} \tag{28}$$

Defining $\boldsymbol{e}_a(k) \triangleq \boldsymbol{X}^T(k)\widetilde{\boldsymbol{w}}(k)$. Since $\boldsymbol{e}_a(k)$ would not affected by the noise signal, it is reasonable to approximate $\boldsymbol{e}(k)$ with $\boldsymbol{e}_a(k)$. Eq. (1) gives

$$\boldsymbol{e}(k) = \boldsymbol{e}_a(k) + \boldsymbol{\xi}(k) \tag{29}$$

where $\boldsymbol{\xi}(k) = [\xi(k), \xi(k-1), \ldots, \xi(k-P+1)]$. To simplify the analysis, the following two assumptions will be used throughout our analysis:

(I) *The noise signal* $\boldsymbol{\xi}(k)$ *is independent and identically distributed (i.i.d.) and statistically independent of the regression matrix* $\boldsymbol{X}(k)$.

(II) *The matrix sequence* $\boldsymbol{X}(k)$ *is i.i.d. and* $\widetilde{\boldsymbol{w}}(k)$ *is independent of* $\boldsymbol{X}(k)\boldsymbol{X}^T(k)$.

By using assumption I and taking the expectation on both sides of (28) yields

$$E\{\|\widetilde{\boldsymbol{w}}(k+1)\|_{\boldsymbol{\Sigma}}^2\} = E\{\|\widetilde{\boldsymbol{w}}(k)\|_{\boldsymbol{\Sigma}'}^2\} + \mu^2 E\{\boldsymbol{\xi}^T(k)\boldsymbol{X}^T(k)\boldsymbol{\Sigma}\boldsymbol{X}(k)\boldsymbol{\xi}(k)\} \quad (30)$$

where

$$\boldsymbol{\Sigma}' = \boldsymbol{\Sigma} - \mu(k)\boldsymbol{\Sigma}\boldsymbol{X}(k)\boldsymbol{X}^T(k) - \mu(k)\boldsymbol{X}(k)\boldsymbol{X}^T(k)\boldsymbol{\Sigma} + \mu^2(k)\boldsymbol{X}(k)\boldsymbol{\xi}(k)\boldsymbol{X}^T(k). \quad (31)$$

Now, define $\boldsymbol{\sigma} \triangleq \text{vec}(\boldsymbol{\Sigma})$ and $\boldsymbol{\Sigma} \triangleq \text{vec}^{-1}(\boldsymbol{\sigma})$, where vec$(\cdot)$ denotes stack the columns of its matrix argument on top of each other. Then, the notation $\|\widetilde{\boldsymbol{w}}(k)\|_{\text{vec}(\boldsymbol{\sigma})}^2$ is used to denote $\|\widetilde{\boldsymbol{w}}(k)\|_{\boldsymbol{\Sigma}}^2$. Under the assumption II, (30) can be expressed as

$$E\{\|\widetilde{\boldsymbol{w}}(k+1)\|_{\text{vec}(\boldsymbol{\sigma})}^2\} = E\{\|\widetilde{\boldsymbol{w}}(k)\|_{\text{vec}(\boldsymbol{\sigma}')}^2\} + \mu^2(k)\boldsymbol{\psi}^2(k)\boldsymbol{\kappa}^T\boldsymbol{\sigma} \quad (32)$$

where

$$\boldsymbol{\sigma}' = \boldsymbol{\Xi}\boldsymbol{\sigma} \quad (33a)$$

$$\boldsymbol{\Xi} = \boldsymbol{I} - \mu(k)\left(E\left\{\boldsymbol{Q}^T(k)\right\} \otimes \boldsymbol{I} + \boldsymbol{I} \otimes E\left\{\boldsymbol{Q}^T(k)\right\}\right) + \mu^2(k)\text{E}\left\{\boldsymbol{Q}^T(k) \otimes \boldsymbol{Q}(k)\right\} \quad (33b)$$

$$\boldsymbol{Q}(k) = \boldsymbol{X}(k)\boldsymbol{X}^T(k) \quad (33c)$$

$$\boldsymbol{\psi}^2(k) = E\left\{\boldsymbol{\xi}^T(k)\boldsymbol{\xi}(k)\right\} \quad (33d)$$

$$\boldsymbol{\kappa} = \text{vec}\left(E\left\{\boldsymbol{X}(k)\boldsymbol{X}^T(k)\right\}\right) \quad (33e)$$

and $\otimes$ stands for the Kronecker product. Note that, the recursion (32) is stable if, and only if, the matrix $\boldsymbol{\Xi}$ is stable [50]. Therefore, the convergence of the VC-VS-APMCCA in the mean-square sense is guaranteed for any $\mu(k)$ in the range

$$0 < \mu(k) < \min\left\{\frac{1}{\lambda_{\max}\left(\boldsymbol{P}^{-1}\boldsymbol{R}\right)}, \frac{1}{\max\left(\lambda\left(\boldsymbol{H} \in \mathbb{R}^+\right)\right)}\right\} \quad (34)$$

where $\lambda_{\max}(\cdot)$ denotes maximum eigenvalue of a matrix, $\max\left(\lambda\left(\boldsymbol{H} \in \mathbb{R}^+\right)\right)$ denotes the largest positive eigenvalue of $\boldsymbol{H}$ when it exists,

$$\boldsymbol{P} = E\left\{\boldsymbol{Q}^T(k)\right\} \otimes \boldsymbol{I} + \boldsymbol{I} \otimes E\left\{\boldsymbol{Q}(k)\right\}, \quad (35)$$

$$\boldsymbol{R} = E\left\{\boldsymbol{Q}^T(k) \otimes \boldsymbol{Q}(k)\right\} \quad (36)$$

and

$$\boldsymbol{H} = \begin{pmatrix} \frac{1}{2}\boldsymbol{P} & \frac{1}{2}\boldsymbol{R} \\ \boldsymbol{I} & \boldsymbol{0} \end{pmatrix}. \quad (37)$$

V. SIMULATIONS

In this section, extensive simulations are conducted to demonstrate the effectiveness of the proposed algorithms. We have measured the performance of the algorithms on a 2.2-GHz Intel Core i7 processor with 12Gb of RAM, running Matlab R2017b on Windows 7. The following simulations are consisted of three examples. In the first example, the Gaussian and impulsive noises are considered as the noise signal.

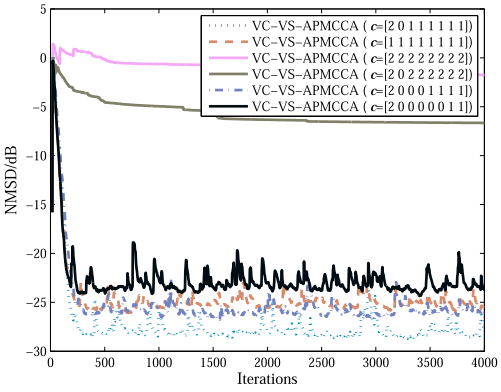

FIGURE 6. NMSDs of the algorithms versus different c under Gaussian noise and impulsive noise.

Note that this case widely exists in ANC system [21]. In the second example, the uniform noise and impulsive noise are used as the noise signal, and in the third example, the exponential and impulsive noises are employed as the noise signal. The impulsive noise is generated by Bernoulli-Gaussian (BG) distribution, which contains a Bernoulli process and Gaussian process [11], [62]

$$\zeta(k) = B_e(k)G_a(k) \quad (38)$$

where B_e denotes a Bernoulli process with the probability P_r, which can be defined by $P(B_e) = 1 - P_r$ with $B_e = 0$, and $P(B_e) = P_r$ with $B_e = 1$, and $G_a(k)$ stands for a white Gaussian process with zero mean and variance σ_ζ^2.

To qualify the identification performance of the algorithm, both excess MSE (EMSE) and normalized MSD (NMSD) are adopted in the simulations, which can be respectively defined as

$$\text{EMSE, db} = 10\log_{10}\left\{e_a^2(k)\right\} \quad (39)$$

and

$$\text{NMSD, db} = 10\log_{10}\left\{\frac{\|\boldsymbol{w}_o - \boldsymbol{w}(k)\|_2^2}{\|\boldsymbol{w}_o\|_2^2}\right\}. \quad (40)$$

In EMSE and NMSD, we utilize y-coordinates of the logarithm to make their performance more clear. The length of unknown system is set to $M = 10$, and $\boldsymbol{w}_o$ is generated by random. For a fair comparison, the MCC algorithm [41], the LMMN algorithm [60], and the AP algorithm [8] are used as the benchmark algorithms. In the following simulations, all the learning curves are obtained by over 100 independent trials.

A. GAUSSIAN NOISE AND IMPULSIVE NOISE

In the first case, the white Gaussian noise with zero-mean and unit variance is used as the input signal. The mixed noise is modeled by Gaussian noise and impulsive noise. The Gaussian noise is a zero-mean with signal-to-noise ratio (SNR) = 10dB, and the impulsive noise with $Pr = 0.001$ and $\sigma_\zeta^2 = 30$.

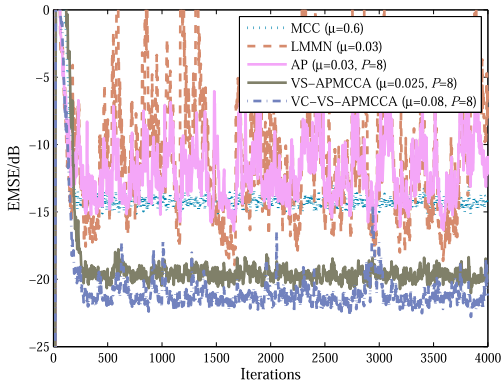

FIGURE 7. The EMSEs of algorithms under Gaussian noise and impulsive noise.

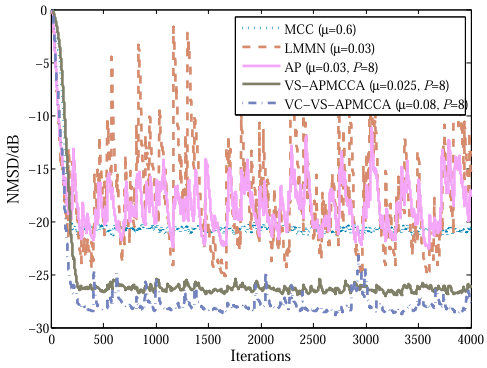

FIGURE 8. The NMSDs of algorithms under Gaussian noise and impulsive noise.

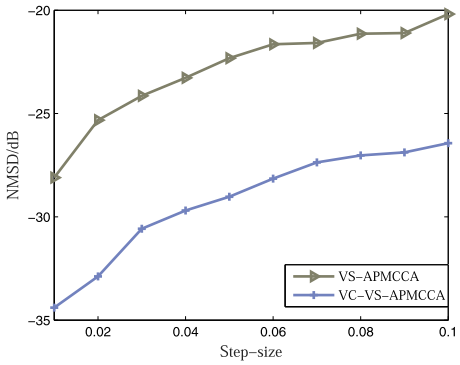

FIGURE 9. Steady-state NMSDs of the algorithms versus step-size for Example 1.

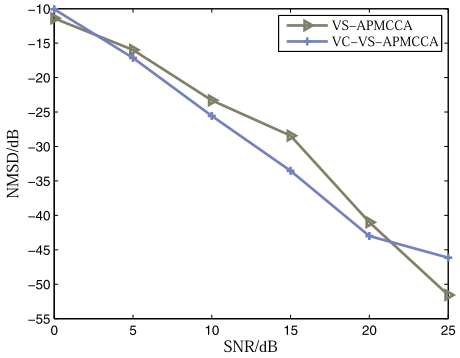

FIGURE 10. Steady-state NMSDs of the algorithms versus SNR for Example 1.

To demonstrate the effect of $\boldsymbol{c}$ on the performance of the VC-VS-APMCCA, Fig. 6 shows the NMSDs of the algorithm versus different $\boldsymbol{c}$. We observe that the VC-VS-APMCCA with $\boldsymbol{c} = [2\ 0\ 1\ 1\ 1\ 1\ 1\ 1]$ has the best performance among these cases. Therefore, we fix this choice in the following studies. The EMSE performance of five algorithms is tested, as shown in Fig 7. In the MCC algorithm, the step-size μ is set to 0.6 and the kernel size $\sigma = 1$. In the LMMN algorithm, the step-size $\mu = 0.03$, the mixing parameter $\lambda = 0.04$. For the conventional AP algorithm, we fix the step-size $\mu = 0.025$ and the projection order $P = 8$ for a fair comparison. The proposed two algorithms, VS-APMCCA and VC-VS-APMCCA, we set $\mu = 0.025$, $\mu = 0.08$, $\sigma = 2$, and $P = 8$. Specifically, for the vector $\boldsymbol{c}$, we select $\boldsymbol{c} = [2, 0, 1, 1, 1, 1, 1, 1]^T$. All the parameter settings are selected to ensure the fast convergence rate and small final EMSE. As can be seen from this figure, the AP and LMMN algorithms have large fluctuations during adaptation. The MCC algorithm has a stable performance in the presence of mixed noise, it reaches -15dB for system identification. Owing to using MCC and HOEP criterion, the proposed two algorithms significantly enhance the accuracy of system identification with similar initial convergence rate. Fig. 8 shows the corresponding NMSD learning curves of the algorithms for Example 1. It can be observed from this figure that all the NMSDs have the same trends as compared with Fig. 7. With the similar initial convergence rate, the VS-APMCCA achieves improved misadjustment in comparison with the state-of-the-art algorithms. Because of VC strategy, the VC-VS-APMCCA can achieve smaller misadjustment as compared with the VS-APMCCA. To further demonstrate the performance, Fig. 9 and Fig. 10 illustrate the steady-state NMSDs versus step-size and SNR, respectively. We can see from Fig. 9 that the VC-VS-APMCCA achieves reduced NMSD(about 5dB) with different μ. For different SNRs, the VC-VS-APMCCA has improved performance when SNR ranges from 5dB to 20dB (See Fig. 10). When SNR = 0dB and 25dB, the VS-APMCCA owns enhanced steady-state NMSD.

B. UNIFORM NOISE AND IMPULSIVE NOISE

In the second example, the performance of algorithms under uniform and impulsive noises is investigated. The uniform noise is modeled by uniform distribution U(-0.5,0.5), and the impulsive noise used in this example is the same as Example 1. The white Gaussian noise with zero-mean and unit variance is utilized as the input signal. Fig. 11 plots he NMSD learning curves of the VC-VS-APMCCA with difference selections of $\boldsymbol{c}$. It can be observed that the VC-VS-APMCCA with $\boldsymbol{c} = [2\ 0\ 1\ 1\ 1\ 1\ 1\ 1]$, $\boldsymbol{c} = [1\ 1\ 1\ 1\ 1\ 1\ 1\ 1]$

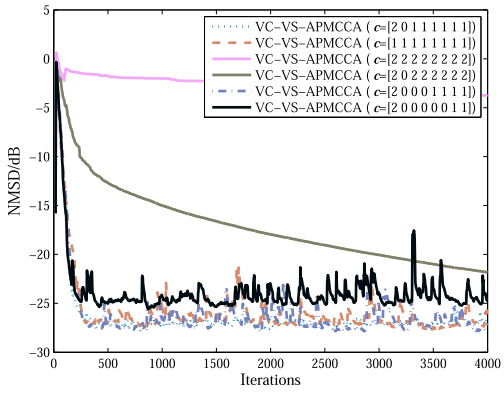

FIGURE 11. NMSDs of the algorithms versus different c under uniform noise and impulsive noise.

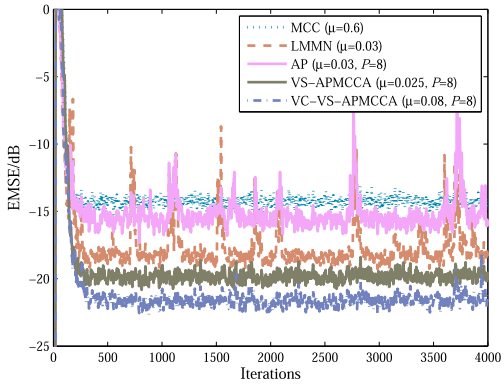

FIGURE 12. The EMSEs of algorithms under uniform noise and impulsive noise.

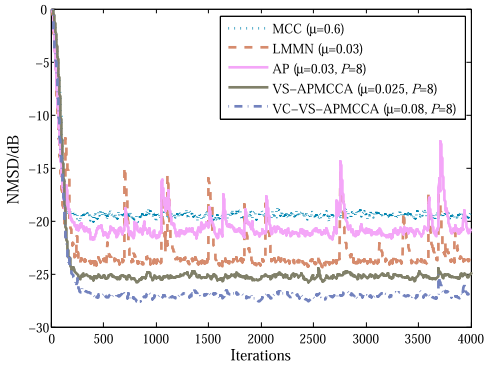

FIGURE 13. The NMSDs of algorithms under uniform noise and impulsive noise.

and $\boldsymbol{c} = [2\ 0\ 0\ 0\ 1\ 1\ 1\ 1]$ achieve similar performance. Since the VC-VS-APMCCA with $\boldsymbol{c} = [2\ 0\ 1\ 1\ 1\ 1\ 1\ 1]$ has slightly faster initial convergence rate as compared with other selections, we choose $\boldsymbol{c} = [2\ 0\ 1\ 1\ 1\ 1\ 1\ 1]$ in this case. Fig 12 shows the EMSEs of the MCC, LMMN, AP, VS-APMCCA, and VC-VS-APMCCA in the presence of uniform noise and impulsive noise. All the parameter settings are the same as Example 1. As can be seen, the AP and LMMN algorithms suffer from stability problem under this scenario since these algorithms are based on MSE criterion.

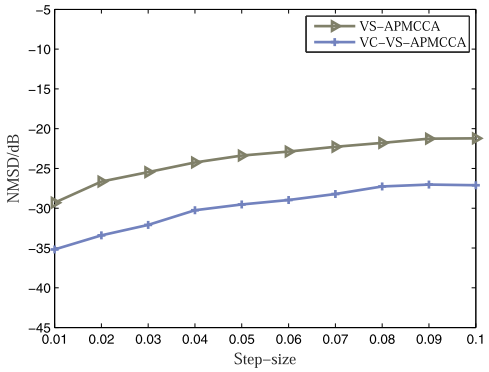

FIGURE 14. Steady-state NMSDs of the algorithms versus step-size for Example 2.

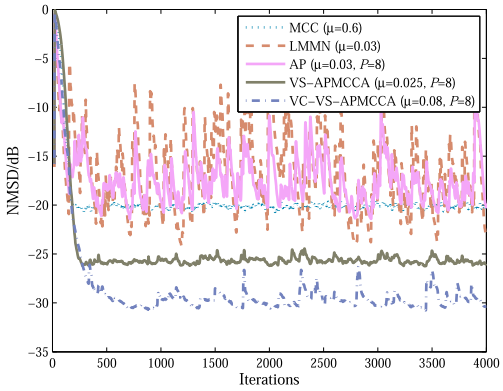

FIGURE 15. NMSDs of the algorithms versus with colored input.

In contrast, the MCC algorithm can obtain robust performance. The VS-APMCCA and VC-VS-APMCCA converge as fast as the MCC algorithm, and can achieve smaller NMSD. To further demonstrate the identification performance, we plot the NMSD learning curves of five algorithms in Fig. 13. One can see that all the algorithms share very close initial convergence rate and they reach steady-state at iteration about 250. The VS-APMCCA reaches final NMSD about -25dB and the VC-VS-APMCCA achieves reduced NMSD about 2dB. Finally, the performance of steady-state NMSD versus step-size is tested in Fig. 14. It can be observed from this figure that the VS-APMCCA has smaller NMSD. The smaller value of step-size, the smaller steady-state NMSD is. As a modified version of VS-APMCCA, the VC-VS-APMCCA achieves more accuracy identification performance, it reduces final NMSD performance of the VS-APMCCA about 5dB. We can also conclude from Fig. 14 and Fig. 9, the VC-VS-APMCCA can obtain great performance enhancement in the case of Example 2. Finally, we investigate the proposed algorithm with colored input in Fig. 15. The colored input is a first-order autoregressive (AR(1)) signal with a pole at 0.2. Other simulation conditions are the same as Fig. 14. One can observe from this figure that the VS-APMCCA and VC-VS-MCCA can obtain smaller steady-state NMSD under the colored input.

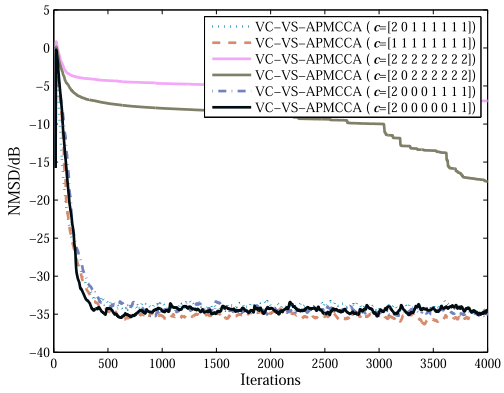

FIGURE 16. NMSDs of the algorithms versus different c under exponentiated noise and impulsive noise.

C. EXPONENTIATED NOISE AND IMPULSIVE NOISE

In the last example, the unknown system is corrupted by exponentiated and impulsive noises, where the exponentiated noise is modeled by exponential distribution with mean parameter (exponential parameter) 0.1 and the impulsive noise is generated by BG distribution with $P_r = 0.001$ and $\sigma_\zeta^2 = 30$. Fig. 16 shows NMSDs of the VC-VS-APMCCA with six different selections. As can be seen from this figure, $\boldsymbol{c} = [2\ 0\ 1\ 1\ 1\ 1\ 1\ 1]$ slightly enhances the initial convergence rate. Fig. 17 shows the EMSEs of existing algorithms and the proposed algorithms during adaptation. We repeat the parameter settings of Example 2. We can observe from this figure that all the algorithms achieve stable filtering performance in this case. The MCC and AP algorithms have quite close performance in this scenario, and both algorithms reach -20dB final EMSE. The LMMN algorithm, which combines the LMS and least mean fourth (LMF) algorithms, has improved steady-state error as compared with the MCC and AP algorithms. The proposed VS-APMCCA and VC-VS-APMCCA further reduce the EMSE, and the VC-VS-APMCCA has slightly smaller EMSE when compared to the VS-APMCCA. Correspondingly, the NMSDs of five algorithms are depicted in Fig. 18. Again, the

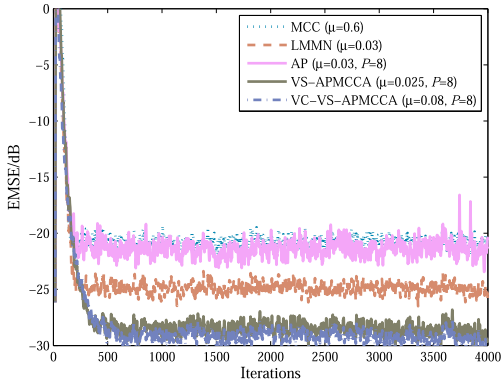

FIGURE 17. The EMSEs of algorithms under exponentiated noise and impulsive noise.

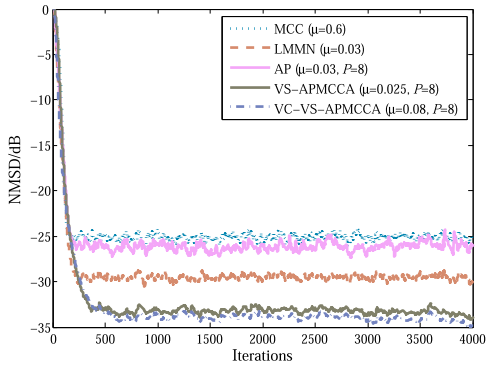

FIGURE 18. The NMSDs of algorithms under exponentiated noise and impulsive noise.

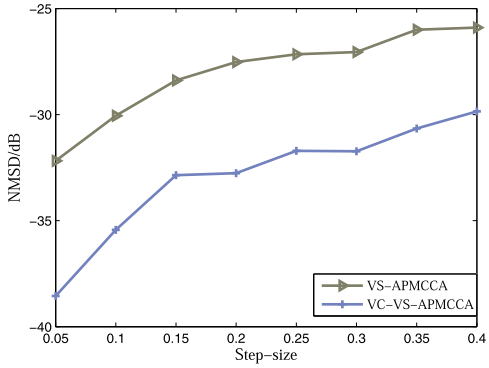

FIGURE 19. Steady-state NMSDs of the algorithms versus step-size for Example 3.

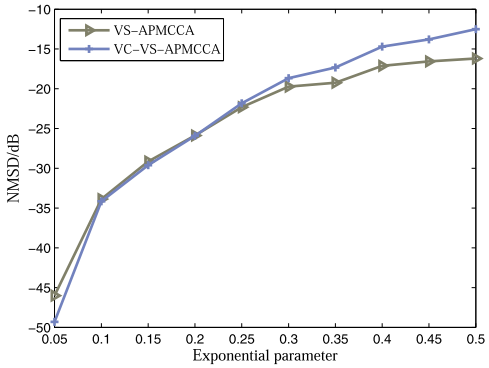

FIGURE 20. Steady-state NMSDs of the algorithms versus exponential parameter for Example 3.

proposed algorithms achieve improved performance and the VC-VS-APMCCA has the best performance. Fig. 19 plots the steady-state NMSDs with different selections of μ. It can be found that the small value of step-size gains smaller NMSD, as we expected. The VC-VS-APMCCA reduces the steady-state NMSD as compared with the VS-APMCCA. In particular, we discuss the effect of exponential parameter on the steady-state NMSD performance. As shown in Fig. 20, two proposed algorithms have quite close identification performance in most cases. Moreover, the VC-VS-APMCCA

has improved performance when the exponential parameter is set to 0.05, and the VS-APMCCA enjoys smaller NMSD when the exponential parameter is in the range of 0.3 to 0.5.

VI. CONCLUSION

In this paper, we have proposed two novel AP algorithms based on MCC and VC strategy for performance improvement in system identification with mixed noise. The proposed VS-APMCCA combined the benefits of AP and MCC algorithms, which can significantly reduce steady-state error with moderate computational complexity in various environments. To further enhance its performance, the VC scheme has been introduced to VS-APMCCA, resulting in VC-VS-APMCCA. Moreover, the convergence behavior of the VC-VS-APMCCA has been analyzed. Extensive simulation results validate the effectiveness of the proposed algorithms as compared with existing algorithms for system identification problem. It should be stressed that the selection of variable center $\boldsymbol{c}$ is based on simulations, which may prohibit its engineering applications. Hence, it is reasonable to employ the method in [63] to the proposed algorithm. In the future, we will conduct research in this approach.

REFERENCES

[1] W. Zhu, L. Luo, A. Xie, and J. Sun, "A novel FELMS-based narrowband active noise control system and its convergence analysis," *Appl. Acoust.*, vol. 156, pp. 229–245, Dec. 2019.

[2] L. Lu, Y. Yu, X. Yang, and W. Wu, "Time delay Chebyshev functional link artificial neural network," *Neurocomputing*, vol. 329, pp. 153–164, Feb. 2019.

[3] Y. Li, Z. Jiang, W. Shi, X. Han, and B. Chen, "Blocked maximum correntropy criterion algorithm for cluster-sparse system identifications," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 66, no. 11, pp. 1915–1919, Nov. 2019.

[4] N. J. Bershad, E. Eweda, and J. C. M. Bermudez, "Performance of soft limiters in the LMS algorithm for cyclostationary white Gaussian inputs," *Signal Process.*, vol. 152, pp. 197–205, Nov. 2018.

[5] B. Chen, S. Zhao, S. Seth, and J. C. Principe, "Online efficient learning with quantized KLMS and l_1 regularization," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, 2012, pp. 1–6.

[6] Y. Li, Y. Wang, and T. Jiang, "Sparse-aware set-membership NLMS algorithms and their application for sparse channel estimation and echo cancelation," *AEU-Int. J. Electron. Commun.*, vol. 70, no. 7, pp. 895–902, 2016.

[7] C. C. D. Santos, J. F. Galdino, and E. L. Pinto, "On the design of LMS-based channel estimators using the Doppler spread parameter," *Digit. Signal Process.*, vol. 23, no. 1, pp. 281–288, 2013.

[8] K. Ozeki and T. Umeda, "An adaptive filtering algorithm using an orthogonal projection to an affine subspace and its properties," *Electron. Commun. Jpn.*, vol. 67-A, no. 5, pp. 19–27, 1984.

[9] Y. Li, Z. Jiang, O. M. Osman, X. Han, and J. Yin, "Mixed norm constrained sparse APA algorithm for satellite and network echo channel estimation," *IEEE Access*, vol. 6, pp. 65901–65908, 2018.

[10] B. Chen, X. Yang, H. Ji, H. Qu, N. Zheng, and J. C. Principe, "Trimmed affine projection algorithms," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, 2014, pp. 1923–1928.

[11] T. Shao, Y. R. Zheng, and J. Benesty, "An affine projection sign algorithm robust against impulsive interferences," *IEEE Signal Process. Lett.*, vol. 17. no. 4, pp. 327–330, Apr. 2010.

[12] S. H. Kim, J. J. Jeong, J. Choi, and S. W. Kim, "Variable step-size affine projection sign algorithm using selective input vectors," *Signal Process.*, vol. 115, pp. 151–156, Oct. 2015.

[13] J. Yoo, J. Shin, and P. Park, "Variable step-size affine projection sign algorithm," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 61, no. 4, pp. 274–278, Apr. 2014.

[14] W. Wang, J. Zhao, H. Qu, and B. Chen, "A correntropy inspired variable step-size sign algorithm against impulsive noises," *Signal Process.*, vol. 141, pp. 168–175, Dec. 2017.

[15] L. Shi, Y. Lin, and X. Xie, "Combination of affine projection sign algorithms for robust adaptive filtering in non-Gaussian impulsive interference," *Electron. Lett.*, vol. 50, no. 6, pp. 466–467, Mar. 2014.

[16] Z. Yang, Y. R. Zheng, and S. L. Grant, "Proportionate affine projection sign algorithms for network echo cancellation," *IEEE Trans. Audio, Speech, Language Process.*, vol. 19, no. 8, pp. 2273–2284, Nov. 2011.

[17] M. T. Akhtar and W. Mitsuhashi, "Improving performance of FxLMS algorithm for active noise control of impulsive noise," *J. Sound Vib.*, vol. 327, no. 3, pp. 647–656, 2009.

[18] Y. Li, T. Lee, and B. Wu, "A variable step-size sign algorithm for channel estimation," *Signal Process.*, vol. 102, pp. 304–312, Sep. 2014.

[19] E. V. Papoulis and T. Stathaki, "A normalized robust mixed-norm adaptive algorithm for system identification," *IEEE Signal Process. Lett.*, vol. 11, no. 1, pp. 56–59, Jan. 2004.

[20] Z. Zheng, Z. Liu, and L. Lu, "Bias-compensated robust set-membership NLMS algorithm against impulsive noises and noisy inputs," *Electron. Lett.*, vol. 53, no. 16, pp. 1100–1102, 2017.

[21] L. Wu, X. Qiu, I. S. Burnett, and Y. Guo, "A recursive least square algorithm for active control of mixed noise," *J. Sound Vib.*, vol. 339, pp. 1–10, Mar. 2015.

[22] J. Ni and F. Li, "Efficient implementation of the affine projection sign algorithm," *IEEE Signal Process. Lett.*, vol. 19, no. 1, pp. 24–26, Jan. 2012.

[23] J. Ni and F. Li, "Variable regularisation parameter sign subband adaptive filter," *Electron. Lett.*, vol. 46, no. 24, pp. 1605–1607, 2010.

[24] L. Lu, W. Wang, X. Yang, W. Wu, and G. Zhu, "Recursive Geman–McClure estimator for implementing second-order Volterra filter," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 66, no. 7, pp. 1272–1276, Jul. 2019.

[25] J. Yoo, J. Shin, and P. Park, "Variable step-size sign algorithm against impulsive noises," *IET Signal Process.*, vol. 9, no. 6, pp. 506–510, 2015.

[26] M. Bekrani and M. Lotfizad, "Clipped LMS/RLS adaptive algorithms: Analytical evaluation and performance comparison with low-complexity counterparts," *Circuits Syst. Signal Process.*, vol. 34, no. 5, pp. 1655–1682, 2015.

[27] S.-C. Chan and Y.-X. Zou, "A recursive least M-estimate algorithm for robust adaptive filtering in impulsive noise: Fast algorithm and convergence performance analysis," *IEEE Trans. Signal Process.*, vol. 52, no. 4, pp. 975–991, Apr. 2004.

[28] Y. Zou, S. C. Chan, and T. S. Ng, "Robust M-estimate adaptive filtering," *IEE Proc.-Vis., Image Signal Process.*, vol. 148, no. 4, pp. 289–294, Aug. 2001.

[29] L. Wu and X. Qiu, "An M-estimator based algorithm for active impulse-like noise control," *Appl. Acoust.*, vol. 74, no. 3, pp. 407–412, 2013.

[30] G. Sun, M. Li, and T. C. Lim, "Enhanced filtered-x least mean M-estimate algorithm for active impulsive noise control," *Appl. Acoust.*, vol. 90, pp. 31–41, Apr. 2015.

[31] M. Shao and C. L. Nikias, "Signal processing with fractional lower order moments: Stable processes and their applications," *Proc. IEEE*, vol. 81, no. 7, pp. 986–1010, Jul. 1993.

[32] B. Weng and K. E. Barner, "Nonlinear system identification in impulsive environments," *IEEE Trans. Signal Process.*, vol. 53, no. 7, pp. 2588–2594, Jul. 2005.

[33] O. Arikan, A. E. Cetin, and E. Erzin, "Adaptive filtering for non-Gaussian stable processes," *IEEE Signal Process. Lett.*, vol. 1, no. 11, pp. 163–165, Nov. 1994.

[34] B. Chen, L. Xing, Z. Wu, J. Liang, J. C. Principe, and N. Zheng, "Smoothed least mean p-power error criterion for adaptive filtering," *Digit. Signal Process.*, vol. 40, pp. 154–163, May 2015.

[35] B. Chen and J. C. Principe, "Maximum correntropy estimation is a smoothed MAP estimation," *IEEE Signal Process. Lett.*, vol. 19, no. 8, pp. 491–494, Aug. 2012.

[36] B. Chen, J. Hu, L. Pu, and Z. Sun, "Stochastic gradient algorithm under (h, ϕ)-entropy criterion," *Circuits Syst. Signal Process.*, vol. 26, no. 6, pp. 941–960, 2007.

[37] B. Chen, P. Zhu, and J. C. Principe, "Survival information potential: A new criterion for adaptive system training," *IEEE Trans. Signal Process.*, vol. 60, no. 3, pp. 1184–1194, Mar. 2011.

[38] B. Chen, Y. Zhu, J. Hu, and M. Zhang, "A new interpretation on the MMSE as a robust MEE criterion," *Signal Process.*, vol. 90, no. 12, pp. 3313–3316, 2010.

[39] S. Peng, W. Ser, B. Chen, L. Sun, and Z. Lin, "Robust constrained adaptive filtering under minimum error entropy criterion," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 65, no. 8, pp. 1119–1123, Aug. 2018.
[40] G. Qian, S. Wang, L. Wang, and S. Duan, "Convergence analysis of a fixed point algorithm under maximum complex correntropy criterion," *IEEE Signal Process. Lett.*, vol. 25, no. 12, pp. 1830–1834, Dec. 2018.
[41] B. Chen, L. Xing, J. Liang, N. Zheng, and J. C. Principe, "Steady-state mean-square error analysis for adaptive filtering under the maximum correntropy criterion," *IEEE Signal Process. Lett.*, vol. 21, no. 7, pp. 880–884, Jul. 2014.
[42] N. C. Kurian, K. Patel, and N. V. George, "Robust active noise control: An information theoretic learning approach," *Appl. Acoust.*, vol. 117, pp. 180–184, Feb. 2017.
[43] N. Zhou, Y. Xu, H. Cheng, Z. Yuan, and B. Chen, "Maximum correntropy criterion-based sparse subspace learning for unsupervised feature selection," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 2, pp. 404–417, Feb. 2019.
[44] B. Chen, L. Xing, B. Xu, H. Zhao, N. Zheng, and J. C. Principe, "Kernel risk-sensitive loss: Definition, properties and application to robust adaptive filtering," *IEEE Trans. Signal Process.*, vol. 65, no. 11, pp. 2888–2901, Jun. 2017.
[45] J. P. F. Guimarães, A. I. R. Fontes, J. B. A. Rego, A. de M. Martins, and J. C. Principe, "Complex correntropy: Probabilistic interpretation and application to complex-valued data," *IEEE Signal Process. Lett.*, vol. 24, no. 1, pp. 42–45, Jan. 2017.
[46] G. Qian and S. Wang, "Generalized complex correntropy: Application to adaptive filtering of complex data," *IEEE Access*, vol. 6, pp. 19113–19120, 2018.
[47] B. Chen, L. Xing, H. Zhao, N. Zheng, and J. C. Principe, "Generalized correntropy for robust adaptive filtering," *IEEE Trans. Signal Process.*, vol. 64, no. 13, pp. 3376–3387, Jul. 2016.
[48] B. Chen, X. Wang, Y. Li, and J. C. Principe, "Maximum correntropy criterion with variable center," *IEEE Signal Process. Lett.*, vol. 26, no. 5, pp. 1212–1216, Aug. 2019.
[49] L. Zhu, C. Song, L. Pan, and J. Li, "Adaptive filtering under the maximum correntropy criterion with variable center," *IEEE Access*, vol. 7, pp. 105902–105908, 2019.
[50] L. Lu, G. Zhu, X. Yang, K. Zhou, Z. Liu, and W. Wu, "Affine projection algorithm based high-order error power for partial discharge denoising in power cables," *IEEE Trans. Instrum. Meas.*, to be published.
[51] Z. Jiang, Y. Li, and X. Huang, "A correntropy-based proportionate affine projection algorithm for estimating sparse channels with impulsive noise," *Entropy*, vol. 21, no. 6, p. 555, 2019.
[52] M. T. M. Silva and V. H. Nascimento, "Improving the tracking capability of adaptive filters via convex combination," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 3137–3149, Jul. 2008.
[53] H. Modares, A. Alfi, and M. B. N. Sistani, "Parameter estimation of bilinear systems based on an adaptive particle swarm optimization," *Eng. Appl. Artif. Intell.*, vol. 23, no. 7, pp. 1105–1111, 2010.
[54] L. Lu, Z. Zheng, B. Champagne, X. Yang, and W. Wu, "Self-regularized nonlinear diffusion algorithm based on levenberg gradient descent," *Signal Process.*, vol. 163, pp. 107–114, Oct. 2019.
[55] E. Bai, "A random least-trimmed-squares identification algorithm," *Automatica*, vol. 39, no. 9, pp. 1651–1659, 2003.
[56] A. H. Sayed, *Fundamentals of Adaptive Filtering*. Hoboken, NJ, USA: Wiley, 2003.
[57] S.-C. Pei and C.-C. Tseng, "Least mean *p*-power error criterion for adaptive FIR filter," *IEEE J. Sel. Areas Commun.*, vol. 12, no. 9, pp. 1540–1547, Dec. 1994.
[58] H. Yazdanpanah, M. V. S. Lima, and P. S. R. Diniz, "On the robustness of set-membership adaptive filtering algorithms," *EURASIP J. Adv. Signal Process.*, vol. 2017, no. 1, p. 72, 2017.
[59] P. S. R. Diniz, "On data-selective adaptive filtering," *IEEE Trans. Signal Process.*, vol. 66, no. 16, pp. 4239–4252, Aug. 2018.
[60] J. A. Chambers, O. Tanrikulu, and A. G. Constantinides, "Least mean mixed-norm adaptive filtering," *Electron. Lett.*, vol. 30, no. 19, pp. 1574–1575, Sep. 1994.
[61] Y. Li, Y. Wang, and F. Albu, "Sparse channel estimation based on a reweighted least-mean mixed-norm adaptive filter algorithm," in *Proc. Eur. Signal Process. Conf.*, 2016, pp. 2380–2384.
[62] Y. Yu, L. Lu, Z. Zheng, W. Wang, Y. Zakharov, and R. C. de Lamare, "DCD-based recursive adaptive algorithms robust against impulsive noise," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, to be published.
[63] W. Wang, J. Zhao, H. Qu, B. Chen, and J. C. Principe, "Convergence performance analysis of an adaptive kernel width MCC algorithm," *AEU-Int. J. Electron. Commun.*, vol. 76, pp. 71–76, Jun. 2017.

XIAODING WANG received the B.S. and M.S. degrees from Chongqing University, Chongqing, China, in 2010 and 2013, respectively. He is currently pursuing the Ph.D. degree in wireless power transfer with the Institute of Marine Information Engineering, Zhejiang University. His major research interests include wireless power transfer and underwater equipment.

JUN HAN received the B.S. and Ph.D. degrees in marine science and technology from the Tokyo University of Fisheries, Tokyo, Japan, in 1992 and 1995, respectively. He is currently a Professor with the Institute of Marine Information Engineering, Zhejiang University. His major research interests include autonomous underwater robot and its applications, and underwater acoustic system engineering.