IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received May 23, 2017, accepted June 20, 2017, date of publication July 3, 2017, date of current version August 14, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2722464

A Family of Robust M-Shaped Error Weighted Least Mean Square Algorithms: Performance Analysis and Echo Cancellation Application

SHENG ZHANG[1,2], WEI XING ZHENG[2], (Fellow, IEEE), JIASHU ZHANG[3], AND HONGYU HAN[4]

[1]School of Information Science and Technology, Southwest Jiaotong University, Chengdu 611756, China
[2]School of Computing, Engineering and Mathematics, Western Sydney University, Sydney, NSW 2751, Australia
[3]Sichuan Province Key Laboratory of Signal and Information Processing, Southwest Jiaotong University, Chengdu 610031, China
[4]School of Computer Science, Sichuan Normal University, Chengdu 610068, China

Corresponding author: Wei Xing Zheng (w.zheng@westernsydney.edu.au)

This work was supported by the Australian Research Council under Grant DP120104986.

ABSTRACT Due to the good filter performance for the non-Gaussian noise, the adaptive filters with error nonlinearities have received increasing attention recently. From the viewpoint of the weighted function, in this paper, the existing least mean square (LMS)-based adaptive algorithms with error nonlinearities are divided into three types, i.e., V-shaped, Λ-shaped, and M-shaped algorithms. Then, to obtain the merits of the V-shaped and Λ-shaped algorithms, a new family of robust M-shaped error weighted LMS algorithms is proposed. Their steady-state mean square deviation (MSD) analyses are made, which reveal the learning abilities of error nonlinearities: 1) for the V-shaped algorithm, it can achieve smaller steady state MSD for sub-Gaussian noise than that for super-Gaussian noise; 2) the Λ-shaped algorithm can be used more effectively for super-Gaussian noise than that for sub-Gaussian noise; and 3) the M-shaped algorithm combines the characteristics of the V-shaped and Λ-shaped algorithms. Furthermore, based on the proposed robust M-shaped function, a proportionate normalized robust M-shaped algorithm is presented for echo cancellation application. Finally, Monte Carlo simulations are conducted to verify the theoretical results and to demonstrate the efficiency of the proposed algorithms in different environments.

INDEX TERMS Adaptive filter, weighted function, steady-state analysis.

I. INTRODUCTION

The least mean square (LMS) algorithm has been widely used in system identification, active noise control (ANC), echo cancellation, adaptive equalization, blind source separation (BSS), beamforming, image processing and so on [1]–[5]. It is well-known that the LMS is based on the second-order moment of the error signal, which is commonly recognized as the natural tool in the presence of Gaussian noise. However, real-world signals often exhibit non-Gaussian properties. For example, in the beamforming, sub-Gaussian (light-tailed) signals are frequently encountered [6]. In some ANC applications, mechanical friction, vibration noise, and speech signal are super-Gaussian/impulsive (heavy-tailed) signals [7]. In the BSS, adaptive receivers with multiple antennas, and image denoising, the signal may be the mixed sub-Gaussian and super-Gaussian/impulsive signal [8]–[15]. To improve the filter performance in different non-Gaussian noise environments, several LMS-based algorithms with error nonlinearities have been proposed [16], [17].

When the system output is disturbed by sub-Gaussian noise, a higher-order moment of the error signal has been employed as a general tool to design an adaptive algorithm. By use of even powers of the error signal, the least mean fourth (LMF) algorithm and its family have been proposed in [18]. Theoretical results show that the LMF family achieves lower steady-state mean square deviation (MSD) for sub-Gaussian (uniform, binary, etc) noise than that for super-Gaussian (Laplace, etc) noise [18], [19]. Moreover, for sub-Gaussian noise, the steady-state error of the LMF outperforms the LMS at the same convergence rate. However, the stability of the LMF is dependent on the input signal power, the initialization of adaptive weights, and the non-stationarity of target weights [20]. In order to utilize the higher-order moment of the noise and improve the convergence performance of the LMF, the least mean absolute third (LMAT) [21], mixed LMS and LMF (mixed LMS-LMF) [22], least mean logarithmic square (LMLS) [23], and several other modified versions [24]–[29] have been introduced, where the LMLS and globally stable normalized LMF (GSNLMF) algorithms can be derived from the logarithmic cost function.

Super-Gaussian/impulsive noise generally has more outliers than sub-Gaussian noise [3]. Thus, in the presence of super-Gaussian/impulsive noise interference, the LMS-based algorithm derived from the higher-order moment of the error will suffer from convergence and stability problems. To achieve a good robustness against outliers, the lower-order moment of the error has been widely utilized to design robust adaptive algorithms [30]. Such algorithms include the sign algorithm (SA) [31], proportion-sign algorithm (PSA) [32], least l_p-norm (LMP) algorithm [30], mixed LMS and sign algorithm (mixed LMS-SA) [33], least logarithmic absolute difference (LLAD) algorithm [23], and several other similar versions [34]–[36]. Though these algorithms achieve good robustness against the outliers, the high-order moment information of the noise is not utilized. Moreover, to the best of the authors' knowledge, the steady-state MSD performance of these algorithms for different noise distributions has not yet been well studied in the literature.

It is well-known that the above-mentioned LMS-based adaptive algorithms with error nonlinearities are derived from the corresponding cost functions. Different from the traditional view, based on the difference of the weighted function of the error, the LMS-based adaptive algorithms with error nonlinearities are differentiated into the V–shaped, Λ–shaped, and M–shaped algorithms in this paper. Then, combining the excellent characteristics of the V–shaped and Λ–shaped weighted functions, a new family of robust M-shaped error weighted LMS (NFRMS-LMS) algorithms is introduced. Using the individual weight error variance analysis method [37], [38] and Taylor series expansion, the analytical expression of steady-state MSD for different weighted functions is derived, which reveals the effects of different noise distributions on the different algorithms. In details, the contributions of this paper for the LMS-based adaptive filters with error nonlinearities are summarized as follows:

(i) The view of the weighted function of the error is firstly proposed. Based on this view, the existing LMS-based adaptive algorithms with error nonlinearities are differentiated into the V–shaped, Λ–shaped, and M-shaped algorithms.

(ii) When sub-Gaussian and impulsive noises appear simultaneously, the NFRMS-LMS algorithms are proposed, in which the higher-order and lower-order moment information of the noise can be jointly used.

(iii) A relationship is revealed between the type of weighted function shapes providing the minimum steady-state MSD and the type of noise.

(iv) Combining the proportionate method and the proposed robust M-shaped function, the proportionate normalized robust M-shaped (PNRMS) algorithm is presented for the acoustic echo cancellation (AEC) application.

The rest of this paper is organized as follows. Section II introduces the concept of the weighted function for LMS-based adaptive filters with error nonlinearities. In Section III, the proposed three categories of the error nonlinearities are presented. In Section IV, the convergence performance is analyzed, followed by the application to acoustic echo cancellation in Section V. Section VI provides simulations. Finally, conclusions are drawn in Section VII.

II. REVIEW OF THE ERROR NONLINEARITIES

Consider the system identification problem. The unknown coefficients and input vector at time instant n are denoted by $\mathbf{w}_{opt} = [w_0, w_1, \ldots, w_{L-1}]^T$ and $\mathbf{x}_n = [x_n, x_{n-1}, \ldots, x_{n-L+1}]^T$, respectively, where L is the filter length. The observed signal is assumed to be

$$y_n = \mathbf{w}_{opt}^T \mathbf{x}_n + \upsilon_n \tag{1}$$

where υ_n is the additive background noise. The estimated error between the outputs of the unknown system and the adaptive filter is $e_n = y_n - \mathbf{x}_n^T \hat{\mathbf{w}}_n$, where $\hat{\mathbf{w}}_n = [\hat{w}_{1,n}, \hat{w}_{2,n}, \ldots, \hat{w}_{L,n}]^T$ denotes the adaptive filter tap-weights at time n.

The weight update equation of LMS-based adaptive algorithms with error nonlinearities is written in a unified form as

$$\hat{\mathbf{w}}_{n+1} = \hat{\mathbf{w}}_n + \mu f\,(e_n)\, e_n \mathbf{x}_n \tag{2}$$

where μ is a step-size, $f(e_n)$ is a nonlinear mapping about e_n. Compared to the LMS algorithm, $f(e_n)$ in (2) can be seen as the weighted function of the error. Table 1 lists a number of adaptive algorithms and their corresponding weighted functions. Apparently, the weighted function $f(e_n)$ should be an even function of the error e_n.

III. THE PROPOSED THREE CATEGORIES OF ERROR NONLINEARITIES ALGORITHMS

In this section, according to the characteristic of the first-order derivative of the weighted function, the existing LMS-based adaptive algorithms with error nonlinearities will be firstly differentiated into three categories, i.e., V–shaped, Λ–shaped, and M–shaped algorithms. Then, a new robust M-shaped algorithm is proposed, which combines the superior characteristics of the V–shaped and Λ–shaped algorithms.

A. FIRST CLASS: V–SHAPED ALGORITHM

The V–shaped algorithm is the LMS-based adaptive algorithm using the V–shaped weighted function, which satisfies the condition

$$f'\,(x) \geq 0, \quad x > 0 \tag{3}$$

where f' denotes the first-order derivative of the function f. See Fig. 1 for illustration of the V–shaped weighted function.

As shown in Table 1, the V–shaped algorithms are the LMS-based algorithms mainly derived from the higher-order moment of the error, which can offer better steady-state performance for sub-Gaussian (uniform) noise than that for super-Gaussian (Laplace) noise [6], [18], [22], [39].

TABLE 1. Examples of $f(e_n)$, where $\varepsilon > 0$, $0 < \delta < 1$, $0 \ll \beta < 1$.

Types of Algorithm	The literature	Algorithm	$f(e_n)$	+	×
V–shaped Algorithms	[18]	LMF family	$f(e_n)=\lvert e_n\rvert^p, p>0$		1 (p=2)
	[22]	mixed LMS-LMF	$f(e_n)=\delta+(1-\delta)\lvert e_n\rvert^2$	1	2
	[48]	LMK	$f(e_n)=3\beta\sigma_{e,n-1}^2+2\lvert e_n\rvert^2$ $\sigma_{e,n}^2=\beta\sigma_{e,n-1}^2+\lvert e_n\rvert^2$	2	4
	[23]	LMLS	$f(e_n)=\lvert e_n\rvert^2/(\varepsilon+\lvert e_n\rvert^2)$	1	3
Λ–shaped Algorithms	[32]	SA	$f(e_n)=1/\lvert e_n\rvert$		1
	[30]	LMP	$f(e_n)=\lvert e_n\rvert^p, -1<p<0$		
	[40],[50]	Error-saturation LMS	$f(e_n)=erf(e_n)/e_n$		
	[23]	LLAD	$f(e_n)=(\varepsilon+\lvert e_n\rvert)^{-1}$	1	1
	[33]	Mixed LMS-SA	$f(e_n)=\delta+(1-\delta)/\lvert e_n\rvert$	1	2
	[41]	Step-size scaler LMS without normalized term	$f(e_n)=(1+\gamma\lvert e_n\rvert^2)^{-1}, \gamma>0$	1	3
	[34], [35]	robust FsLMS	$f(e_n)=\dfrac{1}{\lvert e_n\rvert^2+2\sigma^2}, \sigma>0$	1	2
	[36]	logLMP	$f(e_n)=\dfrac{1}{\lvert e_n\rvert^{2-p}+\lvert e_n\rvert^2}, 0<p\le 2$	1 (p=1)	2 (p=1)
		logCLMP	$f(e_n)=-\dfrac{\log(\lvert e_n\rvert+1)-\log(\lvert e_n\rvert^2+1)}{\log(\lvert e_n\rvert)\lvert e_n\rvert^2}$		
	[47]	GMCC	$f(e_n)=\dfrac{\lvert e_n\rvert^{\alpha-2}}{\exp(\lambda\lvert e_n\rvert^\alpha)}, 0<\alpha\le 2, \lambda>0$		
M–shaped Algorithms	[47]	GMCC	$f(e_n)=\dfrac{\lvert e_n\rvert^{\alpha-2}}{\exp(\lambda\lvert e_n\rvert^\alpha)}, \alpha>2, \lambda>0$		
	This paper	**Proposed robust M-shaped**	$f(e_n)=\dfrac{\lvert e_n\rvert^p}{\varepsilon+\lvert e_n\rvert^{p+1}}, p>0, \varepsilon>0$	1 (p=2)	5 (p=2)

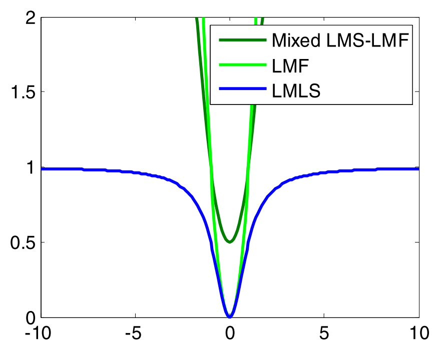


FIGURE 1. The V–shaped functions $f(x)$ with respect to x, where $\delta = 0.5$, $\varepsilon = 1$.

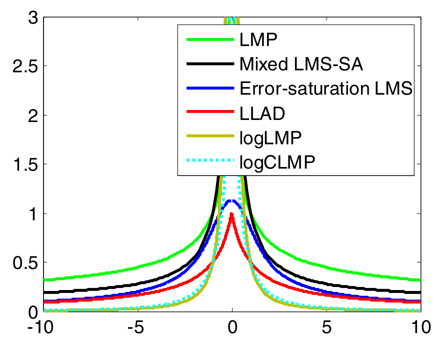


FIGURE 2. The Λ–shaped functions $f(x)$ with respect to x, where $p = -0.5$, $\varepsilon = 1$, $\delta = 0.1$.

B. SECOND CLASS: Λ–SHAPED ALGORITHM

The Λ–shaped algorithm is the LMS-based adaptive algorithm using the Λ–shaped weighted function, which satisfies the condition

$$f'(x) \le 0, \quad x > 0 \tag{4}$$

See Fig. 2 for illustration of the Λ–shaped weighted function. The Λ-shaped weighted functions with corresponding adaptive algorithms are summarized in Table 1. Shown in [30]–[33], [40]–[46], the main purpose of the Λ–shaped algorithm is to lessen the adverse effect of impulsive noise and outliers, in which the lower-order moment information of the error is only utilized.

C. THIRD CLASS: M–SHAPED ALGORITHM

Unlike the V-shaped and Λ–shaped algorithms, the third (M-shaped) category algorithm has the mixed characteristics of the V–shaped and Λ–shaped functions. In other words, it will simultaneously possess the V–shaped and Λ–shaped parts. Thus, the M–shaped algorithm can simultaneously use the higher-order and lower-order information of the noise.

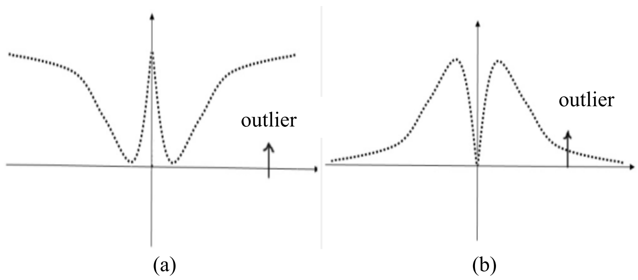


FIGURE 3. Two typical M-shaped functions. (a) Invertible M-shaped. (b) Non-invertible M-shaped.

Without loss of generality, we firstly list two typical M-shaped curves shown in Fig. 3. One type is the invertible M-shaped and the other is the non-invertible M-shaped. It can be seen that the invertible M-shaped algorithm uses the V–shaped part for the large error value (i.e., outliers). Thus, this invertible M-shaped algorithm is not robust against impulsive noise and outliers. Compared with the invertible M-shaped algorithm, the non-invertible M-shaped algorithm uses its Λ–shaped part for the outlier to achieve good robust performance. Due to the good robustness of the non-invertible M-shaped function, we mainly focus on this type M–shaped algorithm, which is also named as robust M–shaped algorithm throughout this paper.

As a robust M-shaped algorithm, we wish that the algorithm works as the V-shaped and Λ–shaped algorithms for the sub-Gaussian noise and super-Gaussian/impulsive noise, respectively. To satisfy the performance, we propose a new family of robust M–shaped (NFRMS) functions, where its weighted function is

$$f(e_n) = \frac{|e_n|^p}{\varepsilon + |e_n|^{p+1}}, \quad p > 0, \qquad \text{(NFRMS)}$$

where $\varepsilon > 0$ is the design parameter. See Fig. 4 for illustration of the M–shaped algorithms.

In the proposed robust M–shaped function, its denominator is with an order higher than the numerator. Using the gradient descent, this algorithm essentially solves the stochastic cost function

$$J(e_n) = \int_0^{e_n} \frac{x\,|x|^p}{\varepsilon + |x|^{p+1}} dx \tag{5}$$

and the Hessian matrix of (5) is given by

$$\mathbf{H} = \frac{(\varepsilon(1+p)\,|e_n|^p)}{\varepsilon + |e_n|^{p+1}} \mathbf{x}_n \mathbf{x}_n^T \tag{6}$$

which is positive semi-definite for $\varepsilon > 0, p > -1$. Thus, the proposed robust M–shaped algorithm will not suffer from the local minimum problem.

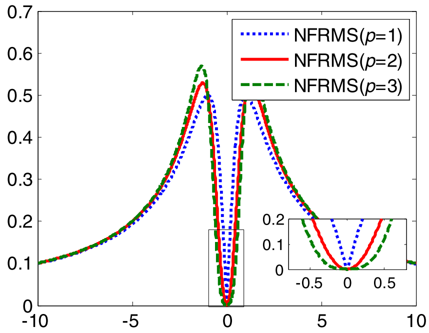


FIGURE 4. The robust M–shaped functions $f(x)$ with respect to x, where $\varepsilon = 1$.

Remark 1: (i) From the generalized maximum correntropy criterion (GMCC), *Chen* proposed the GMCC algorithm in [47]. According to our classification in this paper, this GMCC algorithm with $\alpha > 2$ should belong to the robust M-shaped function. Using the stochastic gradient descent method, this GMCC algorithm can be also derived from the following cost function:

$$J_{GMCC} = 1 - \exp\left(-\lambda\,|e_n|^\alpha\right), \quad \lambda > 0,\ \alpha > 0 \tag{7}$$

where its Hessian matrix is

$$\mathbf{H}_{GMCC} = 4\lambda\,|e_n|^4 \exp\left(-\lambda\,|e_n|^\alpha\right)\left(3 - 4\lambda\,|e_n|^2\right)\mathbf{x}_n\mathbf{x}_n^T \tag{8}$$

Obviously, $\mathbf{H}_{GMCC}$ is positive semi-definite only when $|e_n|^2 < 3/(4\lambda)$. Thus, the cost function (7) is not convex for all λ. When using the GMCC algorithm in practice, the parameter λshould be chosen carefully to guarantee the positive semi-definite of $\mathbf{H}_{GMCC}$, the upper bound of which should depend on the model error, input signal $\mathbf{x}_n$, and noise level.

(ii) Though the proposed robust M-shaped algorithm combines the performances of the V–shaped and Λ–shaped algorithms, it is different from the adaptive combinations [49]. The adaptive combinations use two filters with different parameters (step-sizes). However, the M-shaped algorithm only uses one filter to achieve the mixed performance of V–shaped and Λ–shaped algorithms. The last column of Table 1 describes the computational complexity of the error weighted function in terms of number of multiplications and additions. It can be seen that the LMS type algorithm has similar computational complexity.

(iii) When choosing $-1 < p < 0$ and $\varepsilon \gg |e_n|^{p+1}$ for all n, the weighted function of the NFRMS-LMS algorithm will be

$$f(e_n) = \frac{|e_n|^p}{\varepsilon}, \quad -1 < p < 0$$

which is the weighted function of the LMP. At this time, its performance will be similar to the LMP algorithm.

IV. PERFORMANCE ANALYSIS OF LMS-BASED ADAPTIVE ALGORITHMS WITH ERROR NONLINEARITIES

In this section, the convergence and steady-state properties for different shaped functions will be discussed for different noise distributions. The following assumptions are adopted.

Assumption i: The stationary input signal x_n is a zero-mean random sequence with covariance matrix $\mathbf{R}$.

Assumption ii: $\{\upsilon_n\}$ is independent and identically distributed (*i.i.d.*) with zero-mean and variance σ_0^2, and is independent of $\mathbf{x}_j$ for all n and j.

Assumption iii: $\hat{\mathbf{w}}_n$ is independent of $\mathbf{x}_n$. It is the independent assumption, which is introduced to simplify the analysis in the literatures [50]–[53].

A. CONVERGENCE CONDITION

In this subsection, the convergence condition of LMS-based adaptive algorithms with error nonlinearities is discussed. Let $\mathbf{v}_n = \mathbf{w}_{opt} - \hat{\mathbf{w}}_n$. Then the weight deviation vector recursion of (2) can be written as

$$\mathbf{v}_{n+1} = \mathbf{v}_n - \mu f(e_n)\,\mathbf{x}_n e_n \tag{9}$$

After some simple manipulation, we obtain

$$E\left\{\|\mathbf{v}_{n+1}\|^2\right\} = E\left\{\|\mathbf{v}_n\|^2\right\} - 2\mu E\left\{f(e_n)\,e_n\mathbf{x}_n^T\mathbf{v}_n\right\} + \mu^2 E\left\{f^2(e_n)\,e_n^2\,\|\mathbf{x}_n\|^2\right\} \tag{10}$$

For guaranteeing the stability of the recursion, i.e., $E\left\{\|\mathbf{v}_{n+1}\|^2\right\} \le E\left\{\|\mathbf{v}_n\|^2\right\}$, the step-size μ should satisfy

$$\mu \le 2\frac{E\left\{f(e_n)\,e_n\mathbf{x}_n^T\mathbf{v}_n\right\}}{E\left\{f^2(e_n)\,e_n^2\,\|\mathbf{x}_n\|^2\right\}} \quad \text{for all } n \tag{11}$$

Remark 2: Using the method in [16] and the Gaussianity assumption of $\mathbf{x}_n^T\mathbf{v}_n$ for sufficiently long filters, the bound in (11) can be evaluated for general noise as follows:

$$\mu \le 2\frac{1}{E\left\{\|\mathbf{x}_n\|^2\right\}} \inf_{\lambda \le E\{e_{a,n}^2\}} \frac{h_g\left[E\left\{e_{a,n}^2\right\}\right]}{h_c\left[E\left\{e_{a,n}^2\right\}\right]} \tag{12}$$

where λ is the Cramer-Rao bound associated with the underlying estimation process, i.e., using $\mathbf{x}_n^T\hat{\mathbf{w}}_n$ to estimate the random quantity $\mathbf{x}_n^T\mathbf{w}_{opt}$, and

$$\begin{aligned} h_c\left[E\left\{e_{a,n}^2\right\}\right] &= E\left\{f^2\left(\upsilon_n + e_{a,n}\right)\left(\upsilon_n + e_{a,n}\right)^2\right\} \\ h_g\left[E\left\{e_{a,n}^2\right\}\right] &= E\left\{f\left(e_{a,n}\right)\left(\upsilon_n + e_{a,n}\right)e_{a,n}\right\} \end{aligned} \tag{13}$$

with $e_{a,n} = \mathbf{x}_n^T\mathbf{v}_n$.

B. STEADY-STATE ANALYSIS

In this subsection, we assume that the adaptive algorithm has converged. Applying the property $\mathrm{Tr}(\mathbf{XY}) = (vec(\mathbf{X}^T))^T vec(\mathbf{Y})$, the steady-state excess mean square error (EMSE) and MSD are given respectively by

$$\text{EMSE} = \left(vec\left(E\left(\mathbf{v}_\infty\mathbf{v}_\infty^T\right)\right)\right)^T vec(\mathbf{R}) \tag{14}$$

$$\text{MSD} = \left(vec\left(E\left(\mathbf{v}_\infty\mathbf{v}_\infty^T\right)\right)\right)^T vec(\mathbf{I}) \tag{15}$$

where the independent assumption is used in (14), $vec(\bullet)$, $\mathrm{Tr}(\cdot)$, and $\mathbf{I}$ denote stacking the columns, the trace of a matrix, and the identity matrix of appropriate dimension, respectively. From (14) and (15), we know that $vec\left(E\left(\mathbf{v}_\infty\mathbf{v}_\infty^T\right)\right)$ is an important performance measure for the adaptive filtering algorithms.

Firstly, the series expansion of the weighted function $f(e_n)$ with respect to $\mathbf{x}_n^T\mathbf{v}_n$ around υ_n can be written as [17]

$$f(e_n) = f(\upsilon_n) + f'(\upsilon_n)\,\mathbf{x}_n^T\mathbf{v}_n + \frac{1}{2}f''(\upsilon_n)\left(\mathbf{x}_n^T\mathbf{v}_n\right)^2 + O\left(\left(\mathbf{x}_n^T\mathbf{v}_n\right)^3\right) \tag{16}$$

where f'' represents the second-order derivative of the function f, and $O\left(\left(\mathbf{x}_n^T\mathbf{v}_n\right)^3\right)$ denotes the third and higher-power terms of $\mathbf{x}_n^T\mathbf{v}_n$. Though this Taylor series representation limits the weighted function $f(\cdot)$ to a restricted class of functions that are differentiable up to second order, this is easily satisfied by approximating it with a suitably smoothed version.

Now we examine the behavior of (2) in the case where the weight error vector is small such that the term $O\left(\left(\mathbf{x}_n^T\mathbf{v}_n\right)^3\right)$ in (16) may be neglected relative to the lower-order terms, especially for unbiased algorithms[1]. Then, the weight deviation vector recursion can be written as

$$\mathbf{v}_{n+1} = \mathbf{v}_n - \mu\left(f(\upsilon_n) + f'(\upsilon_n)\,\mathbf{x}_n^T\mathbf{v}_n + \frac{1}{2}f''(\upsilon_n)\left(\mathbf{x}_n^T\mathbf{v}_n\right)^2\right)\left(\mathbf{x}_n^T\mathbf{v}_n + \upsilon_n\right)\mathbf{x}_n \tag{17}$$

By multiplying (17) by $\mathbf{v}_{n+1}^T$ from the right, we get that

$$\begin{aligned} \mathbf{v}_{n+1}\mathbf{v}_{n+1}^T &= \mathbf{v}_n\mathbf{v}_n^T - \mu\left(f(\upsilon_n) + f'(\upsilon_n)\,\upsilon_n\right) \\ &\quad \times \left(\mathbf{x}_n\mathbf{x}_n^T\mathbf{v}_n\mathbf{v}_n^T + \mathbf{v}_n\mathbf{v}_n^T\mathbf{x}_n\mathbf{x}_n^T\right) \\ &\quad + \mu^2\left(\left(f(\upsilon_n) + f'(\upsilon_n)\,\upsilon_n\right)^2 \right. \\ &\qquad \left. + \upsilon_n f(\upsilon_n)\left(2f'(\upsilon_n) + f''(\upsilon_n)\,\upsilon_n\right)\right) \\ &\quad \mathbf{x}_n\mathbf{x}_n^T\mathbf{v}_n\mathbf{v}_n^T\mathbf{x}_n\mathbf{x}_n^T + \mu^2\left(f(\upsilon_n)\,\upsilon_n\right)^2\mathbf{x}_n\mathbf{x}_n^T + \upsilon_n(\cdot) \end{aligned} \tag{18}$$

where $\upsilon_n(\cdot)$ is the odd function term with respect to the noise.

By using the Kronecker product notation, and the following property [38] that for arbitrary matrices $\{\mathbf{X},\ \mathbf{Y},\ \mathbf{Z}\}$ of compatible dimensions,

$$vec(\mathbf{XYZ}) = \left(\mathbf{Z}^T \otimes \mathbf{X}\right) vec(\mathbf{Y}), \tag{19}$$

[1] Under the Assumptions i-iii, the weight update equation (2) is unbiased when it is stable [17], [31], [40], [51], [53]. In this case, when using the small step-size the third- and higher-order terms are enough small such that these terms can be ignored at steady-state. For the study of the biased solutions, it could not accommodated in this work, as is left for the next work.

Eq. (18) can be converted into

$$\begin{aligned}
&vec\left(\mathbf{v}_{n+1}\mathbf{v}_{n+1}^T\right)\\
&= vec\left(\mathbf{v}_n\mathbf{v}_n^T\right) - \mu\left(f\left(\upsilon_n\right) + f'\left(\upsilon_n\right)\upsilon_n\right)\\
&\quad\times\left(\mathbf{I}\otimes\mathbf{x}_n\mathbf{x}_n^T + \mathbf{x}_n\mathbf{x}_n^T\otimes\mathbf{I}\right) vec\left(\mathbf{v}_n\mathbf{v}_n^T\right)\\
&\quad+\mu^2\Big(\left(f\left(\upsilon_n\right) + f'\left(\upsilon_n\right)\upsilon_n\right)^2 + f\left(\upsilon_n\right)\upsilon_n\left(2f'\left(\upsilon_n\right)\right.\\
&\qquad\qquad\left.+f''\left(\upsilon_n\right)\upsilon_n\right)\Big)\\
&\quad\times\left(\mathbf{x}_n\mathbf{x}_n^T\otimes\mathbf{x}_n\mathbf{x}_n^T\right)vec\left(\mathbf{v}_n\mathbf{v}_n^T\right)+\mu^2 f^2(\upsilon_n)\upsilon_n^2 vec\left(\mathbf{x}_n\mathbf{x}_n^T\right)\\
&\quad+\upsilon_n vec\left(\cdot\right)
\end{aligned}\tag{20}$$

Taking expectations of both sides of (20), and invoking Assumptions i-iii, we have

$$\begin{aligned}
vec\left(\mathbf{V}_{n+1}\right) &= vec\left(\mathbf{V}_n\right) - \mu f_a\mathbf{A}vec\left(\mathbf{V}_n\right)\\
&\quad+\mu^2 f_b\mathbf{B}vec\left(\mathbf{V}_n\right) + \mu^2 f_r vec\left(\mathbf{C}\right)
\end{aligned}\tag{21}$$

where

$$\mathbf{V}_n = E\left\{\mathbf{v}_n\mathbf{v}_n^T\right\}\tag{22}$$

$$\mathbf{A} = E\left\{\mathbf{I}\otimes\mathbf{x}_n\mathbf{x}_n^T + \mathbf{x}_n\mathbf{x}_n^T\otimes\mathbf{I}\right\}\tag{23}$$

$$\mathbf{B} = E\left\{\mathbf{x}_n\mathbf{x}_n^T\otimes\mathbf{x}_n\mathbf{x}_n^T\right\}\tag{24}$$

$$\mathbf{C} = E\left\{\mathbf{x}_n\mathbf{x}_n^T\right\}\tag{25}$$

$$f_r = E\left\{f^2\left(\upsilon_n\right)\upsilon_n^2\right\}\tag{26}$$

$$f_a = E\left\{f\left(\upsilon_n\right) + f'\left(\upsilon_n\right)\upsilon_n\right\}\tag{27}$$

$$\begin{aligned}
f_b &= E\left\{\left(f\left(\upsilon_n\right) + f'\left(\upsilon_n\right)\upsilon_n\right)^2\right\}\\
&\quad+E\left\{f\left(\upsilon_n\right)\upsilon_n\left(2f'\left(\upsilon_n\right) + f''\left(\upsilon_n\right)\upsilon_n\right)\right\}
\end{aligned}\tag{28}$$

The resulting steady-state solution of (21) can be found by setting $vec\left(\mathbf{V}_{n+1}\right) = vec\left(\mathbf{V}_n\right)$ as $n\to\infty$, that is,

$$vec\left(\mathbf{V}_\infty\right) = \mu f_r\left(f_a\mathbf{A} - \mu f_b\mathbf{B}\right)^{-1} vec\left(\mathbf{R}\right)\tag{29}$$

For values of μ that are small enough so that the term $\mu f_b\mathbf{B}$ could be ignored in (29), we obtain

$$vec\left(\mathbf{V}_\infty\right) = \mu\frac{f_r}{f_a}\mathbf{A}^{-1}vec\left(\mathbf{C}\right)\tag{30}$$

Substituting it into (15) yields

$$MSD = \mu\frac{f_r}{f_a}\psi\tag{31}$$

where

$$\psi = vec\left(\mathbf{I}\right)^T\mathbf{A}^{-1}vec\left(\mathbf{C}\right)\tag{32}$$

For white input signal with $\mathbf{R} = \sigma_x^2\mathbf{I}$, the steady-state MSD is

$$\mathrm{MSD} = \frac{E\left\{f^2\left(\upsilon_n\right)\upsilon_n^2\right\}}{E\left\{f\left(\upsilon_n\right) + f'\left(\upsilon_n\right)\upsilon_n\right\}}\frac{\mu L}{2}\quad(\text{small }\mu)\tag{33}$$

According to (31) and (33), the steady-state MSD of the LMS-based algorithm with error nonlinearity is not only controlled by the step-size μ and the filter length L, but is also dependent on the distribution of the noise signal, the weighted function $f(\cdot)$ and its first-order derivative $f'(\cdot)$. The *Appendix* discusses the effect of the first-order derivative of the weighted function for binary noise signal. Substituting $f = 1$ and $f' = 0$ for the LMS algorithm with small step-size condition into (33) yields $MSD = \mu L\sigma_0^2/2$ [2]. Obviously, it is a function with respect to the variance of the noise, which is independent of the noise distribution. Thus, it is different from the performance of the LMS-based algorithm with error nonlinearity.

In the sequel, the steady-state MSD for the NFRMS-LMS algorithm is presented. Similarly, the final steady-state MSD expression for other weighted functions can be derived, which is not reported here.

1) FOR THE PROPOSED NFRMS-LMS ALGORITHM

For the proposed NFRMS-LMS algorithm, we have

Substituting the following (34a) and (34b):

$$f\left(\upsilon_n\right) = \frac{\left|\upsilon_n\right|^p}{\varepsilon + \left|\upsilon_n\right|^{p+1}}\tag{34a}$$

$$f'\left(\upsilon_n\right) = \frac{\varepsilon p sign\left(\upsilon_n\right)\left|\upsilon_n\right|^{p-1} - sign\left(\upsilon_n\right)\left|\upsilon_n\right|^{2p}}{\left(\varepsilon + \left|\upsilon_n\right|^{p+1}\right)^2}\tag{34b}$$

into (33) yields the steady-state MSD as shown in (35), as shown at the bottom of this page.

Remark 3 (For the NFRMS-LMS Algorithm With p=2): (i) If the term $\left|\upsilon_n\right|^3$ is sufficiently small so that $\varepsilon \gg \left|\upsilon_n\right|^3$, then the steady-state MSD (35) implies that

$$MSD = \mu LE\left\{\left|\upsilon_n\right|^6\right\}\Big/\left(6\varepsilon E\left\{\left|\upsilon_n\right|^2\right\}\right)\tag{36}$$

When $\varepsilon = 1$, it is the steady-state MSD of the LMF [18]. If the noise term $\left|\upsilon_n\right|^3$ is sufficiently large such that $\varepsilon << \left|\upsilon_n\right|^3$, then the steady-state MSD (35) becomes

$$MSD = \mu L\left(E\left\{\left|\upsilon_n\right|^{-2}\right\} + 2\varepsilon E\left\{\left|\upsilon_n\right|^{-4}\right\}\right)^{-1}\tag{37}$$

From (36) and (37), we observe that the steady-state MSD of the proposed NFRMS-LMS depends on the magnitude of the noise. In other words, for large noise, the steady-state

$$MSD_{NFRMS-LMS} = \frac{E\left\{\left|\upsilon_n\right|^{2p+2}/\left(\varepsilon + \left|\upsilon_n\right|^{p+1}\right)^2\right\}}{E\left\{\left|\upsilon_n\right|^p/\left(\varepsilon + \left|\upsilon_n\right|^{p+1}\right)\right\} + E\left\{\left(\varepsilon p\left|\upsilon_n\right|^p - \left|\upsilon_n\right|^{2p+1}\right)/\left(\varepsilon + \left|\upsilon_n\right|^{p+1}\right)^2\right\}}\frac{\mu L}{2}\tag{35}$$

MSD depends on the negative order moments of the noise due to the fact that the Λ–shaped part works. However, for small noise, the steady-state MSD is related to the higher-order moments due to the fact that the V–shaped part works.

(ii) In the proposed NFRMS-LMS, it contains two parameters, ε, p, which increases the freedom of algorithm design. The choice of the parameter p determines the used order moments of the noise in the steady-state MSD (35). The value of the parameter ε determines the proportion of the V-shaped and Λ–shaped parts in the proposed NFRMS-LMS algorithm. As an example, for $\varepsilon = 0$, it is a Λ–shaped function. When $\varepsilon \to \infty$, it will be a V-shaped function. Moreover, through the optimization of p and ε, we can further improve the convergence performance of the NFRMS-LMS. Theoretically speaking, the optimal values of p and ε should be derived by minimizing Eq. (35) with prior noise information about the environment, which is a complicated two-dimensional nonlinear optimization problem. This topic has been beyond the scope of this article, which is left for our future research work.

C. THEORETICAL STEADY-STATE MSD CURVES OF DIFFERENT SHAPED ALGORITHMS

Though the theoretical steady-state MSD expression for different weighted functions has been derived in the preceding subsection, we still do not know the steady-state behaviors of different shaped algorithms for different noise distributions except for Binary distribution (See *the Appendix*). Here, the theoretical steady-state MSD curves (33) with respect to the powers of the noise will be drawn by doing the numerical calculation. See Figs. 5-7 for illustration of the theoretical steady-state MSDs of the V–shaped, Λ–shaped and M–shaped algorithms for different noise distributions (uniform, Gaussian, Laplace). For comparison, the simulated steady-state MSD curves are drawn in these figures, where the variance of the inputs is 1 and $L = 16$.

It is shown that the LMS-based adaptive algorithms using different shapes of the weighted function bring about different steady-state characteristics for different noise distributions. The same shaped algorithm has the same steady-state characteristic for different noise distributions, because it uses the same type (higher-order or lower-order) moment of the noise. For the V–shaped algorithm, it can offer lower steady-state MSD for uniform (sub-Gaussian) noise than that for Laplace (super-Gaussian) noise. On the contrary, the steady-state MSD of the Λ–shaped algorithm for super-Gaussian noise can outperform that for sub-Gaussian noise. Equipped with both the V–shaped and Λ–shaped parts, the M–shaped algorithm has the mixed steady-state characteristics. In the mixed sub-Gaussian and super-Gaussian/impulsive noise environment, the robust M–shaped algorithm can be a best choice. For instance, by using the V–shaped and Λ–shaped parts for uniform and impulsive noises, respectively, the proposed NFRMS-LMS algorithm can offer small steady-state MSD in the presence of the mixture of uniform and impulsive noises.

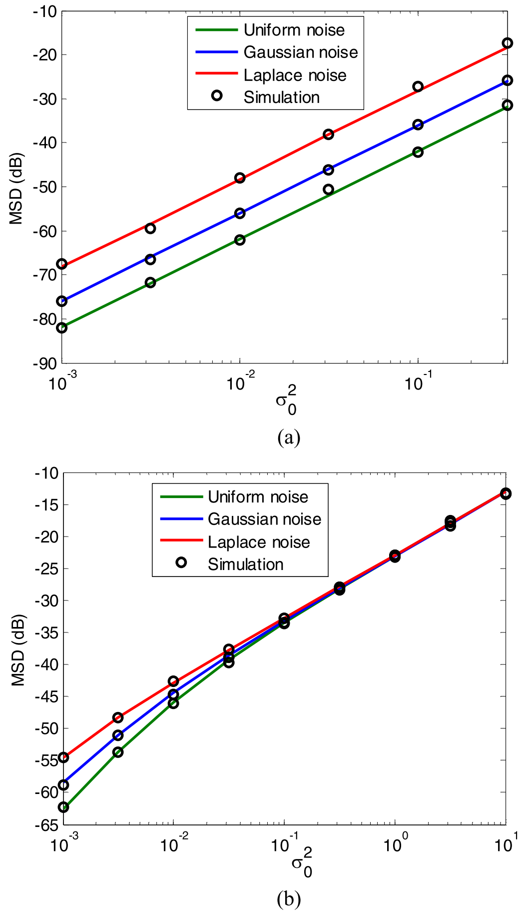


FIGURE 5. Theoretical steady-state MSD curves of the V–shaped algorithms with respect to the powers of the noise, where $\mu = 0.01/L$. (a) LMF. (b) LMLS with $\varepsilon = 0.01$.

Finally, from the perspective of steady-state MSD for different noises, the applicability of the V–shaped, Λ–shaped, and M–shaped algorithms is summarized in Table 2.

V. APPLICATION TO ACOUSTIC ECHO CANCELLATION

In the echo cancellation framework [54]–[56], the sparse channel and the influence near-end speech are the two key issues affecting the performance of the adaptive algorithm. In this section, based on the stochastic cost function (5), the PNRMS algorithm will be applied to the AEC.

Defining the *a posteriori error* as

$$e_{p,n} = \mathbf{x}_n^T \hat{\mathbf{w}}_{n+1}$$

Then, we introduce a normalized error cost function:

$$\min J_{AC} = \left\| \hat{\mathbf{w}}_{n+1} - \hat{\mathbf{w}}_n \right\|_{G_n}^2 + \mu \int_0^{\frac{e_{p,n}}{\|\mathbf{x}_n\|_{G_n}}} \frac{x\,|x|^p}{\varepsilon + |x|^{p+1}} dx \tag{38}$$

where $\mu > 0$ is a small positive parameter which works as a step-size in the proposed PNRMS algorithm, G_n is a

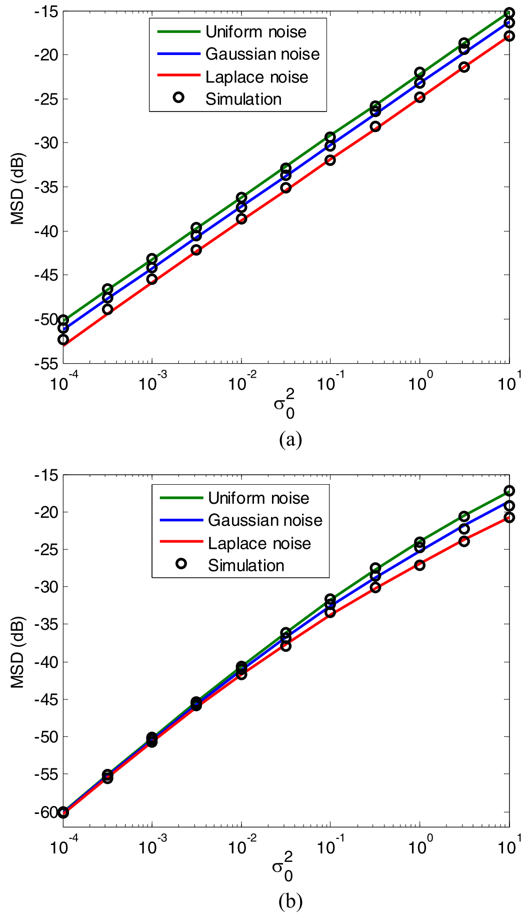


FIGURE 6. **Theoretical steady-state MSD curves of the Λ–shaped algorithms with respect to the powers of the noise, where $\mu = 0.01/L$. (a) LMP with $p = -0.6$. (b) LLAD with $\varepsilon = 0.5$.**

TABLE 2. **Applicability among three shaped algorithms in the stationary environment.**

Types of Algorithm	sub-Gaussian noise	super-Gaussian/ impulsive noise	mixed sub- and super-Gaussian/ impulsive noise
V–shaped	√		
Λ–shaped		√	
Robust M–shaped			√

proportionate matrix and its diagonal elements are chosen as [55]

$$g_{i,n} = \frac{1-\rho}{2L} + (1+\rho)\frac{\left|\hat{w}_{i,n}\right|}{2\left\|\hat{\mathbf{w}}_{i,n}\right\|_1 + \delta} \tag{39}$$

where $i = 1, 2, \cdots, L$, $-1 \leq \rho \leq 1$, and δ is a small positive number that is used to avoid divisions by zero.

Let the derivative of (38) with respect to $\hat{\mathbf{w}}_{n+1}$ equal to zero, we have

$$\hat{\mathbf{w}}_{n+1} = \hat{\mathbf{w}}_n + \mu\frac{\left|e_{p,n}\right|^p}{\varepsilon\left\|\mathbf{x}_n\right\|_{G_n}^{p+1} + \left|e_{p,n}\right|^{p+1}}\frac{e_{p,n}G_n\mathbf{x}_n}{\left\|\mathbf{x}_n\right\|_{G_n}} \tag{40}$$

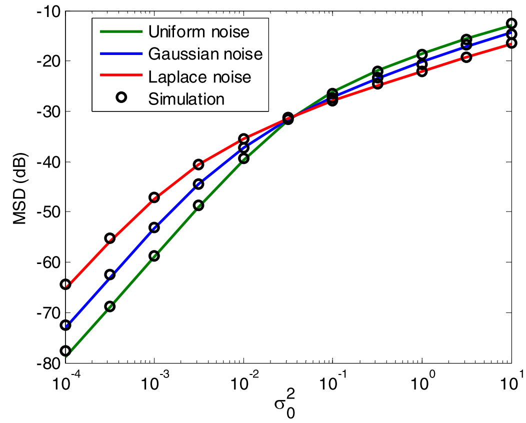


FIGURE 7. **Theoretical steady-state MSD curves of the M–shaped algorithms with respect to the powers of the noise, where $\mu = 0.02/L$, $p = 2$, $\varepsilon = 0.01$.**

In (40), it needs the *a posteriori error*, which is unknown at time n. A reliable approximation for $e_p(n)$ in (40) is $e_{p,n} \approx e_n$ for small step-size μ. Therefore, we finally get the update equation of the PNRMS algorithm:

$$\hat{\mathbf{w}}_{n+1} = \hat{\mathbf{w}}_n + \mu\frac{\left|e_n\right|^p}{\varepsilon\left\|\mathbf{x}_n\right\|_{G_n}^{p+1} + \left|e_n\right|^{p+1}}\frac{e_nG_n\mathbf{x}_n}{\left\|\mathbf{x}_n\right\|_{G_n}} \tag{41}$$

VI. SIMULATIONS

In this section, Monte-Carlo simulations of the proposed NFRMS-LMS and PNRMS algorithms are carried out for the system identification and echo cancellation problems. The unknown system to be estimated in all experiments is a finite-impulse-response filter, and its weight vector is randomly generated and normalized to unit energy. The length of the adaptive filter is equal to that of the corresponding unknown channels. The initial weight vector of the adaptive filter is an all zero vector. All the learning curves are obtained by averaging the results of 100 independent runs. The signal-to-noise ratio (SNR) at the system output is defined as

$$SNR = 10\log_{10}\left(\frac{E\left\{d_n^2\right\}}{E\left\{\upsilon_n^2\right\}}\right)$$

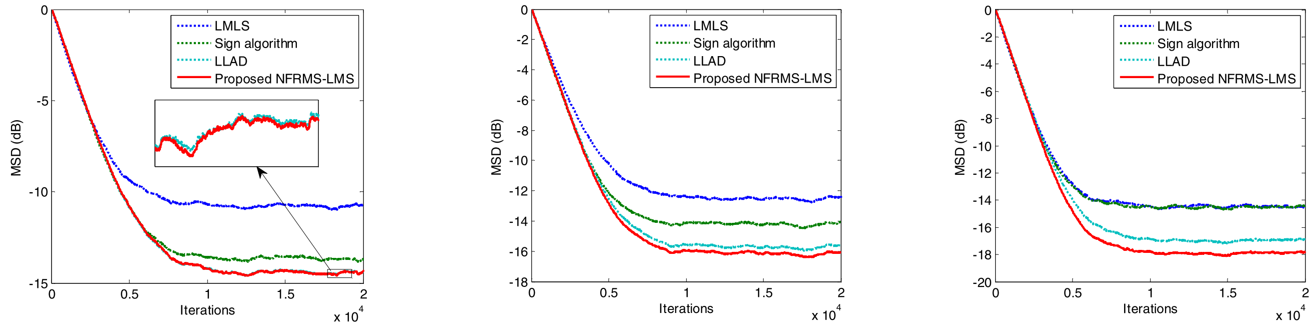
where d_n is the uncorrupted output. However, for the symmetric $\alpha - S$ noise, the fractional-order SNR (FSNR) with $\tau < 2$ is used [46]

$$FSNR = 10\log_{10}\left(\frac{E\left\{\left|d_n\right|^\tau\right\}}{E\left\{\left|\upsilon_n\right|^\tau\right\}}\right)$$

A. TEST OF THE NFRMS-LMS FOR THE SYSTEM IDENTIFICATION

1) IN THE MIXED SUB-GAUSSIAN AND SUPER-GAUSSIAN NOISES ENVIRONMENT

In this experiment, the input is a zero-mean white Gaussian noise with unit variance. To model the mixed sub-Gaussian and super-Gaussian noises, we use the mixed uniform and Laplace distributed noises in Fig. 8, where the uniform and Laplace noise signals are both zero-mean *i.i.d.* noise signals with the variances 1 and 10, respectively. To achieve the



(a) (b) (c)

FIGURE 8. Simulated MSD curves for the proposed algorithm for the mixed uniform and Laplace distributed noises, where $L = 48$. (a) Uniform (20%) and Laplace (80%). (b) Uniform (50%) and Laplace (50%). (c) Uniform (80%) and Laplace (20%).

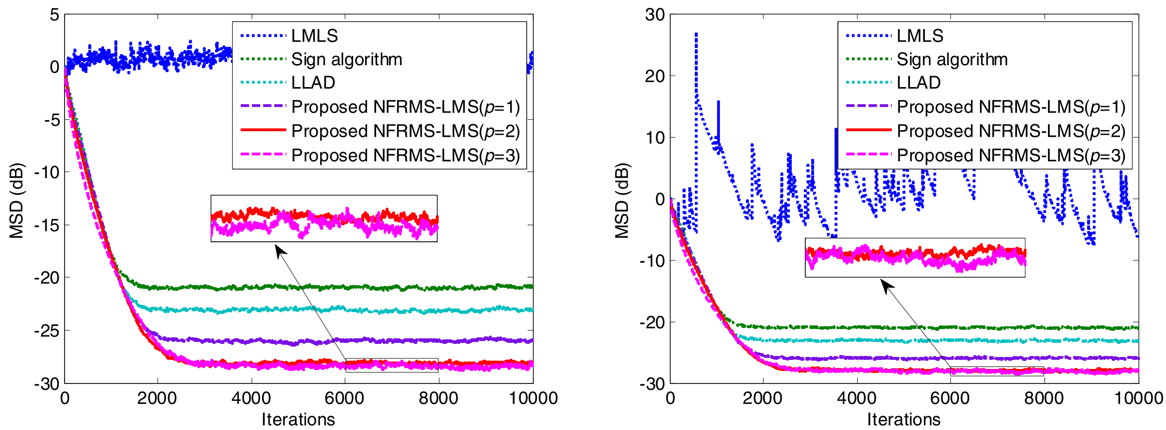


FIGURE 9. Simulated MSD curves for the proposed algorithm. (a) Impulsive noise model-1. (b) Impulsive noise model-2.

same convergence speed for all algorithms, the parameters are chosen as $\mu_{\text{LLAD}} = 0.0012$, $\mu_{\text{SA}} = 0.0008$, $\mu_{\text{LMLS}} = 0.0004$, $\mu_{\text{NFRMS-LMS}} = 0.001$, $\varepsilon = 1$, and $p = 2$. As can be seen, in the mixed uniform and Laplace noises environment the proposed algorithm can offer good steady-state MSD, especially for the uniform noise with large percentage in the mixed noises shown in Fig. 8 (c). This is because the robust M-shaped algorithm utilized the higher-order and lower-order moment information of the mixed noise.

2) IN THE MIXED SUB-GAUSSIAN AND IMPULSIVE NOISE ENVIRONMENT

In this experiment, we test the performance of the proposed algorithm in the mixed sub-Gaussian and impulsive noise environment in Fig. 9, where $L = 16$. The input is a colored signal, which is generated by filtering a white Gaussian noise through a first-order system with a pole at 0.8. The sub-Gaussian noise is a uniform distributed with zero-mean and SNR=20 dB. For the impulsive noise, two different impulsive noise models are used.

(Impulsive Model-1): $\varpi_n = b_n \upsilon_{b,n}$, where b_n is an *i.i.d.* Bernoulli random sequence whose value at any time instant n is either zero or one, with occurrence probability $P(b_n = 1) = 0.005$ and $\upsilon_{b,n}$ is an *i.i.d.* zero-mean Gaussian sequence with variance $G = 5000$.

(Impulsive Model-2): Symmetric $\alpha - S$ noise, $\tau = 0.7$, $\alpha = 0.8$, FSNR=15 dB.

The parameters are set to $\mu_{\text{LLAD}} = 0.003$, $\mu_{\text{SA}} = 0.0025$, $\mu_{\text{LMLS}} = 0.005$, $\mu_{\text{NFRMS-LMS}} = 0.0035$ ($p = 1$), $\mu_{\text{NFRMS-LMS}} = 0.0047$ ($p = 2$), $\mu_{\text{NFRMS-LMS}} = 0.007$ ($p = 3$), and $\varepsilon = 0.1$. Through Fig. 9, it has been shown that the proposed algorithm can provide better steady-state MSD relative to the other algorithms. In addition, it can be noticed that the algorithm with $p = 2$ can offer better steady-state MSD than that for $p = 1$. However, the proposed algorithm for $p = 3$ has the similar convergence performance with $p = 2$.

B. TEST OF THE PNRMS FOR ECHO CANCELLATION

In this experiment, we compare the performances of the NLLAD, stable-PNLMF, and proposed PNRMS algorithms in the double-talk simulation. The impulse response of the network acoustic channel [57] and speech signals (male mandarin) are seen in Fig. 10. The measurement noise is assumed to the zero-mean Gaussian signal with SNR=5 dB. A classical Geigel double-talk detector is used for all algorithms, where the detection threshold is set to 1.4. The step-sizes are set to $\mu_{\text{LLAD}} = 0.03$, $\mu_{\text{stable-PNLMF}} = 0.08$, $\mu_{\text{PNRMS}} = 0.06$ ($p = 2$), and $\varepsilon = 0.1$. As can be seen from Fig. 11, at the same

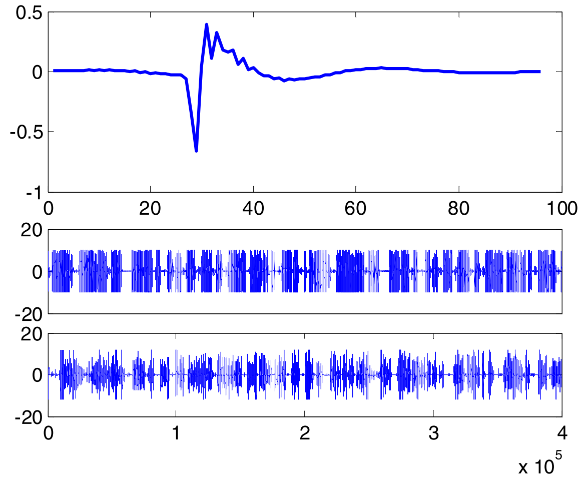


FIGURE 10. **The echo path (*Top*), far-end speech signal (*Middle*) and near-end speech signal (*Bottom*).**

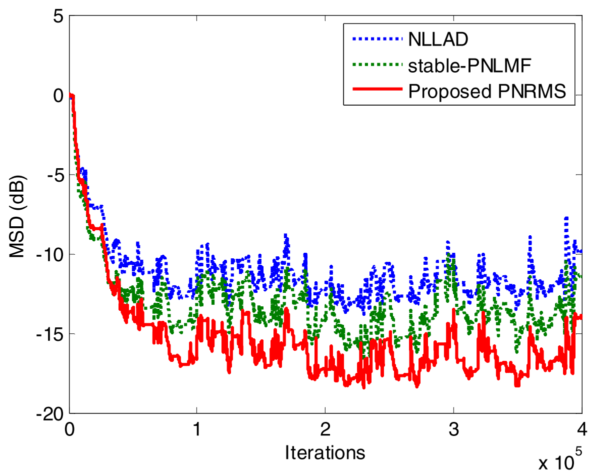


FIGURE 11. **MSD curves of the proposed proportionate normalized algorithm in the double-talk scenario, where the inputs are speech signals.**

initial convergence speed, the PNRMS algorithm reaches a lower steady-state MSD than the NLLAD [23] and stable-PNLMF [58] algorithms.

VII. CONCLUSION

From the view of the weighted function in this paper, the LMS-based algorithms with error nonlinearities have been grouped into three shaped algorithms, i.e., V–shaped, Λ–shaped, and M–shaped. The NFRMS-LMS and PNRMS algorithms have been proposed, which provide good filter performance in the mixed sub-Gaussian and super-Gaussian/impulsive noise environment. Also, the theoretical MSD expressions for different shaped algorithms have been derived. It has been found that the V–shaped and Λ–shaped algorithms can be used effectively for sub-Gaussian and super-Gaussian noise, respectively. For the robust M–shaped algorithm, it may be the good choice for the mixed sub-Gaussian and super-Gaussian/impulsive noise, because it can use higher-order and lower-order information of the noise. Monte-Carlo simulations of three shaped algorithms have provided strong support for these theoretical results. It should be emphasized that these obtained results can lead to a deep understanding and useful insights on the error nonlinearity for different noise distributions. The results will also generate effective design mechanisms to design the different weighted functions in different noise environments.

APPENDIX

Here, we explain the effect of the first-order derivative of the weighted function in the LMS-based adaptive algorithm. Assume that the noise is binary noise with variance σ_0^2 (i.e. $\upsilon(n) = \pm\sigma_0$ with equal probability). Then, the steady-state MSD can be written as

$$\mathrm{MSD} = \frac{\frac{1}{2}f^2(\sigma_0)\,\sigma_0^2 + \frac{1}{2}f^2(-\sigma_0)\,\sigma_0^2}{E\{f(n_0)\} + \left\{\frac{1}{2}f'(\sigma_0)\,\sigma_0 - \frac{1}{2}\sigma_0 f'(-\sigma_0)\right\}}\,\frac{\mu L}{2} \tag{42}$$

In order to obtain the effect of $f'(\cdot)$, we choose

$$f_1(\pm\sigma_0) = f_2(\pm\sigma_0) = 1 \tag{43}$$

and

$$f_1'(\sigma_0) \in R^+, \quad f_2'(\sigma_0) \in R^- \tag{44}$$

Apparently, f_1 belongs to the V–shaped weighted function and f_2 is the Λ–shaped weighted function. Inserting (43) in (42) yields

$$MSD = \frac{\sigma_0^2}{1 + \sigma_0 f'(\sigma_0)}\,\frac{\mu L}{2} \tag{45}$$

Then, (44) and (45) imply that

$$MSD_1 < MSD_{LMS} < MSD_2$$

where MSD_{LMS} is the steady-state MSD of the LMS. The example of f_1 and f_2 is shown as follows.

Example: Let f_1 and f_2 be

$$f_1(x) = \frac{1}{\sigma_0^2}x^2, \quad f_2(x) = \frac{1}{1 - \sigma_0 + |x|}. \tag{46}$$

We obtain the steady-state MSD (45) that for binary noise with variance σ_0^2

$$MSD_1 = \frac{\sigma_0^2}{3}\,\frac{\mu L}{2}, \quad MSD_2 = \frac{\sigma_0^2}{1 - |\sigma_0|}\,\frac{\mu L}{2} \tag{47}$$

REFERENCES

[1] C. F. N. Cowan and P. M. Grant, *Adaptive Filters*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1985.
[2] S. Haykin, *Adaptive Filter Theory*, 4th ed. Upper Saddle River, NJ, USA: Prentice-Hall, 2002.
[3] P. Huber and E. Ronchetti, *Robust Statistics*. Hoboken, NJ, USA: Wiley, 2009.
[4] W. Ma, J. Duan, W. Man, J. Liang, and B. Chen, "General mixed norm-based diffusion adaptive filtering algorithm for distributed estimation over network," *IEEE Access*, vol. 5, pp. 1090–1102, 2017.

[5] C. Song, K. J. Lee, and I. Lee, "Designs of MIMO amplify-and-forward wireless relaying networks: Challenges and solutions," *IEEE Access*, to be published, doi: 10.1109/ACCESS.2017.2703125.

[6] X. Jiang, W.-J. Zeng, A. Yasotharan, H. C. So, and T. Kirubarajan, "Robust beamforming by linear programming," *IEEE Trans. Signal Process.*, vol. 62, no. 7, pp. 1834–1849, Apr. 2014.

[7] H. C. Cho, S. W. Park, K. S. Lee, and N. H. 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 Process.*, vol. 6, no. 8, pp. 772–780, Oct. 2012.

[8] M. Kawamoto, K. Matsuoka, and N. Ohnishi, "A method of blind separation for convolved non-stationary signals," *Neurocomputing*, vol. 22, no. 1, pp. 157–171, 1998.

[9] T. W. Lee, M. Girolami, and T. Sejnowski, "Independent component analysis using an extended infomax algorithm for mixed sub-Gaussian and super-Gaussian sources," *Neural Comput.*, vol. 11, no. 2, pp. 409–433, 1999.

[10] G. A. Tsihrintzis and C. L. Nikias, "Data-adaptive algorithms for signal detection in sub-Gaussian impulsive interference," *IEEE Trans. Signal Process.*, vol. 45, no. 7, pp. 1873–1878, Jul. 1997.

[11] W. Ying, Y. Jiang, Y. Liu, and P. Li, "A blind receiver with multiple antennas in impulsive noise modeled as the sub-Gaussian distribution via the MCMC algorithm," *IEEE Trans. Veh. Technol.*, vol. 62, no. 7, pp. 3492–3497, Sep. 2013.

[12] F. Zhong and J. Zhang, "Linear discriminant analysis based on L1-norm maximization," *IEEE Trans. Image Process.*, vol. 22, no. 8, pp. 3018–3027, Aug. 2013.

[13] L. N. Sharma, S. Dandapat, and A. Mahanta, "ECG signal denoising using higher order statistics in wavelet subbands," *Biomed. Signal Process. Control*, vol. 5, no. 3, pp. 214–222, 2010.

[14] J. Jin, B. Yang, K. Liang, and X. Wang, "General image denoising framework based on compressive sensing theory," *Comput. Graph.*, vol. 38, pp. 382–391, Feb. 2014.

[15] T.-W. Lee and M. S. Lewicki, "Unsupervised image classification, segmentation, and enhancement using ICA mixture models," *IEEE Trans. Image Process.*, vol. 11, no. 3, pp. 270–279, Mar. 2002.

[16] T. Y. Al-Naffouri and A. H. Sayed, "Adaptive filters with error nonlinearities: Mean-square analysis and optimum design," *EURASIP J. Appl. Signal Process.*, vol. 200, no. 4, pp. 192–205, 2001.

[17] S. C. Douglas and T. H. Meng, "Stochastic gradient adaptation under general error criteria," *IEEE Trans. Signal Process.*, vol. 42, no. 6, pp. 1335–1351, Jun. 1994.

[18] E. Walach and B. Widrow, "The least mean fourth (LMF) adaptive algorithm and its family," *IEEE Trans. Inf. Theory*, vol. 30, no. 2, pp. 275–283, Mar. 1984.

[19] N. R. Yousef and A. H. Sayed, "A unified approach to the steady-state and tracking analyses of adaptive filters," *IEEE Trans. Signal Process.*, vol. 49, no. 2, pp. 314–324, Feb. 2001.

[20] E. Eweda, "Dependence of the stability of the least mean fourth algorithm on target weights non-stationarity," *IEEE Trans. Signal Process.*, vol. 62, no. 7, pp. 1634–1643, Apr. 2014.

[21] Y. H. Lee, J. D. Mok, S. D. Kim, and S. H. Cho, "Performance of least mean absolute third (LMAT) adaptive algorithm in various noise environments," *Electron. Lett.*, vol. 34, no. 3, pp. 241–243, Feb. 1998.

[22] 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.

[23] M. O. Sayin, N. D. Vanli, and S. S. Kozat, "A novel family of adaptive filtering algorithms based on the logarithmic cost," *IEEE Trans. Signal Process.*, vol. 62, no. 17, pp. 4411–4424, Sep. 2014.

[24] A. Zerguine, M. K. Chanb, T. Y. Al-Naffouria, M. Moinuddina, and C. F. N. Cowanc, "Convergence and tracking analysis of a variable normalised LMF (XE-NLMF) algorithm," *Signal Process.*, vol. 89, no. 5, pp. 778–790, 2009.

[25] E. Eweda, "Global stabilization of the least mean fourth algorithm," *IEEE Trans. Signal Process.*, vol. 60, no. 3, pp. 1473–1477, Mar. 2012.

[26] S. Zhang and J. Zhang, "Fast stable normalised least-mean fourth algorithm," *Electron. Lett.*, vol. 51, no. 16, pp. 1276–1277, Aug. 2015.

[27] H. Zhao, Y. Yu, S. Gao, X. Zeng, and Z. He, "A new normalized LMAT algorithm and its performance analysis," *Signal Process.*, vol. 105, pp. 399–409, Dec. 2014.

[28] E. Eweda and N. J. Bershad, "Stochastic analysis of a stable normalized least mean fourth algorithm for adaptive noise canceling with a white Gaussian reference," *IEEE Trans. Signal Process.*, vol. 60, no. 12, pp. 6235–6244, Dec. 2012.

[29] E. Eweda, "Mean-square stability analysis of a normalized least mean fourth algorithm for a Markov plant," *IEEE Trans. Signal Process.*, vol. 62, no. 24, pp. 6545–6553, Dec. 2014.

[30] 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.

[31] S. C. Bang and S. Ann, "A robust algorithm for adaptive FIR filtering and its performance analysis with additive contaminated-Gaussian noise," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 43, no. 5, pp. 361–369, May 1996.

[32] E. Eweda, "Tracking analysis of the sign algorithm without the Gaussian constraint," *IEEE Trans. Circuits Syst. II, Analog Digit. Signal Process.*, vol. 45, no. 1, pp. 115–122, Jan. 1998.

[33] J. Chambers and A. Avlonitis, "A robust mixed-norm adaptive filter algorithm," *IEEE Signal Process. Lett.*, vol. 4, no. 2, pp. 46–48, Feb. 1997.

[34] M. J. Black, G. Sapiro, D. H. Marimont, and D. Heeger, "Robust anisotropic diffusion," *IEEE Trans. Image Process.*, vol. 7, no. 3, pp. 421–432, Mar. 1998.

[35] N. V. George and G. Panda, "A robust filtered-s LMS algorithm for nonlinear active noise control," *Appl. Acoust.*, vol. 73, no. 8, pp. 836–841, 2012.

[36] L. Lu and H. Zhao, "Adaptive Volterra filter with continuous lp-norm using a logarithmic cost for nonlinear active noise control," *J. Sound Vibrat.*, vol. 364, pp. 14–29, Mar. 2016.

[37] S. Zhang and J. Zhang, "Transient analysis of zero attracting NLMS algorithms without Gaussian input signal," *Signal Process.*, vol. 97, pp. 100–109, Apr. 2014.

[38] S. Zhang and J. Zhang, "New steady-state analysis results of variable step-size LMS algorithm with different noise distributions," *IEEE Signal Process. Lett.*, vol. 21, no. 6, pp. 653–657, Jun. 2014.

[39] X. Jiang, W.-J. Zeng, A. Yasotharan, H. C. So, and T. Kirubarajan, "Minimum dispersion beamforming for non-Gaussian signals," *IEEE Trans. Signal Process.*, vol. 62, no. 7, pp. 1879–1893, Apr. 2014.

[40] N. J. Bershad, "On error saturation nonlinearities for LMS adaptation in impulsive noise," *IEEE Trans. Signal Process.*, vol. 56, no. 9, pp. 4526–4529, May 2008.

[41] I. Song, P. Park, and R. W. Newcomb, "A normalized least mean squares algorithm with a step-size scaler against impulsive measurement noise," *IEEE Trans. Circuits Syst. II, Express Briefs*, vol. 60, no. 7, pp. 442–445, Jul. 2013.

[42] W. J. Zeng, H. C. So, and A. M. Zoubir, "An ℓ_p-norm minimization approach to time delay estimation in impulsive noise," *Digit. Signal Process.*, vol. 23, no. 4, pp. 1247–1254, Jul. 2013.

[43] W. J. Zeng, H. C. So, and L. Huang, "ℓ_p-MUSIC: Robust direction-of-arrival estimator for impulsive noise environments," *IEEE Trans. Signal Process.*, vol. 61, no. 17, pp. 4296–4308, Sep. 2013.

[44] T. Shao and Y. R. Zheng, "A new variable step-size fractional lower order moment algorithm for non-Gaussian interference environments," in *Proc. IEEE ISCAS*, Taiwan, Taipei, May 2009, pp. 2065–2068.

[45] Y. R. Zheng and V. H. Nascimento, "Two variable step-size adaptive algorithms for non-Gaussian interference environment using fractionally lower-order moment minimization," *Digit. Signal Process.*, vol. 23, pp. 831–844, May 2013.

[46] S. Zhang, J. Zhang, and H. Han, "Robust shrinkage normalized sign algorithm in impulsive noise environment," *IEEE Trans. Circuits Syst. II, Express Briefs*, vol. 64, no. 1, pp. 91–95, Jan. 2017.

[47] B. Chen, L. Xing, H. Zhao, N. Zheng, and J. C. Príncipe, "Generalized correntropy for robust adaptive filtering," *IEEE Trans. Signal Process.*, vol. 64, no. 13, pp. 3376–3387, Jul. 2016.

[48] O. Tanrikulu and A. G. Constantinides, "Least mean kurtosis: A novel higher-order statistics based adaptive filtering algorithm," *Electron. Lett.*, vol. 30, no. 3, pp. 189–190, Feb. 1994.

[49] J. Arenas-Garcia, L. A. Azpicueta-Ruiz, M. T. M. Silva, V. H. Nascimento, and A. H. Sayed, "Combinations of adaptive filters," *IEEE Signal Process. Mag.*, vol. 33, no. 1, pp. 120–140, Jan. 2016.

[50] N. J. Bershad, "On error-saturation nonlinearities in LMS adaptation," *IEEE Trans. Acoust., Speech Signal Process.*, vol. 36, no. 4, pp. 440–452, Apr. 1988.

[51] T. Y. Al-Naffouri and A. H. Sayed, "Transient analysis of adaptive filters with error nonlinearities," *IEEE Trans. Signal Process.*, vol. 51, no. 3, pp. 653–663, Mar. 2003.
[52] B. Lin, R. X. He, L. M. Song, and B. S. Wang, "Steady-state performance analysis for adaptive filters with error nonlinearities," in *Proc. IEEE ICASSP*, Apr. 2009, pp. 3093–3096.
[53] N. J. Bershad and J. C. M. Bermudez, "Stochastic analysis of the least mean kurtosis algorithm for Gaussian inputs," *Digit. Signal Process.*, vol. 54, pp. 35–45, Jul. 2016.
[54] D. L. Duttweiler, "Proportionate normalized least-mean-squares adaptation in echo cancelers," *IEEE Trans. Speech Audio Process.*, vol. 8, no. 5, pp. 508–518, Sep. 2000.
[55] J. Benesty and S. L. Gay, "An improved PNLMS algorithm," in *Proc. IEEE ICASSP*, Orlando, FL, USA, May 2002, pp. 1881–1884.
[56] S. Zhang, J. Zhang, and H. Han, "Robust variable step-size decorrelation normalized least-mean-square algorithm and its application to acoustic echo cancellation," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 24, no. 12, pp. 2368–2376, Dec. 2016.
[57] *Digital Network Echo Cancellers*, document ITU-T Rec. G.168, 2002.
[58] M. O. Sayin, Y. Yilmaz, A. Demir, and S. S. Kozat, "The Krylov-proportionate normalized least mean fourth approach: Formulation and performance analysis," *Signal Process.*, vol. 109, pp. 1–13, Apr. 2015.

SHENG ZHANG received the B.S. degree from the College of Mathematics and Information Sciences, North China University of Water Resources and Electric Power, Zhengzhou, China, in 2010, and the Ph.D. degree in signal and information processing from Southwest Jiaotong University, Chengdu, China, in 2016. From 2016 to 2017, he was a Research Fellow with the School of Computing, Engineering and Mathematics, Western Sydney University, Sydney, Australia. In 2017, he joined the School of Information Science and Technology, Southwest Jiaotong University, Chengdu. His research interests include adaptive filtering algorithms.

WEI XING ZHENG (M'93–SM'98–F'14) received the Ph.D. degree in electrical engineering from Southeast University, Nanjing, China, in 1989. Over the years, he has held various faculty/research/visiting positions with Southeast University, the Imperial College of Science, Technology and Medicine, London, U.K., The University of Western Australia, Perth, Australia, the Curtin University of Technology, Perth, the Munich University of Technology, Munich, Germany, the University of Virginia, Charlottesville, VA, USA, and the University of California at Davis, Davis, CA, USA. He currently holds the rank of Full Professor with Western Sydney University, Sydney, Australia.

Dr. Zheng was named as a Thomson Reuters Highly Cited Researcher in 2015 and 2016. He also served as the Chair of the IEEE Circuits and Systems Society's Technical Committee on Neural Systems and Applications and the IEEE Circuits and Systems Society's Technical Committee on Blind Signal Processing. He served as an Associate Editor of the IEEE Transactions on Circuits and Systems-i: Fundamental Theory and Applications, the IEEE Transactions on Automatic Control, the IEEE Signal Processing Letters, the IEEE Transactions on Circuits and Systems-ii: Express Briefs, and the IEEE Transactions on Fuzzy systems, and as a Guest Editor of the IEEE Transactions on Circuits and Systems-i: Regular Papers. He is currently an Associate Editor of the *Automatica*, the IEEE Transactions on Automatic control (the second term), the IEEE Transactions on Cybernetics, the IEEE Transactions on Neural Networks and Learning Systems, the IEEE Transactions on Control of Network Systems, and other scholarly journals.

JIASHU ZHANG received the B.S. degree from the University of Electronic Science and Technology of China, Chengdu, China, in 1987, the M.S. degree from Chongqing University, Chongqing, China, in 1990, and the Ph.D. degree from the University of Electronic Science and Technology of China, in 2001. In 2001, he joined the School of Information Science and Technology, Southwest Jiaotong University, Chengdu. He is currently a Professor and the Director of the Sichuan Province Key Laboratory of Signal and Information Processing, Southwest Jiaotong University. He has authored over 200 international journal and conference papers. His research interests include biometric security and privacy, information forensic and data hiding, and nonlinear and adaptive signal processing.

HONGYU HAN received the B.S. degree in information security and the Ph.D. degree in communication and information system from Southwest Jiaotong University, Chengdu, China, in 2010 and 2016, respectively. From 2014 to 2015, she was a Visiting Student with the Department of Computing and Information Systems, University of Melbourne, Australia. Since 2016, she has been with the School of Computer Science, Sichuan Normal University, Chengdu. Her research interests include robust adaptive signal processing, sequence design, and coding theory.