M-estimate Based Subband Adaptive Filter Algorithm: Performance Analysis and Improvements

Yi Yu, *Member, IEEE*, HongSen He, *Member, IEEE*, Badong Chen, *Senior Member, IEEE*, Jianghui Li,
Youwen Zhang and Lu Lu, *Member, IEEE*

Abstract—This paper studies the mean and mean-square behaviors of the M-estimate based normalized subband adaptive filter algorithm (M-NSAF) with robustness against impulsive noise. Based on the contaminated-Gaussian noise model, the stability condition, transient and steady-state results for the algorithm are formulated analytically. The analysis results help us to better understand the M-NSAF performance in impulsive noise. To further obtain fast convergence and low steady-state estimation error, we derive a variable step size (VSS) M-NSAF algorithm. This VSS scheme is also generalized to proportionate M-NSAF variant for sparse systems. Computer simulations on system identification in impulsive noise and acoustic echo cancellation with double-talk are performed to demonstrate our theoretical analysis and the effectiveness of the proposed algorithms.

Index Terms—Acoustic echo cancellation, impulsive noise, M-estimate, subband adaptive filter, variable step size

I. INTRODUCTION

ADAPTIVE filter has been found in a variety of applications such as system identification, channel equalization, active noise control, and acoustic/network echo cancellation [1]–[4]. One of the most popular adaptive algorithms is the normalized least mean squares (NLMS) because of its low computational complexity per input sample and stability on the power of input signals. The shortcoming of the NLMS algorithm is slow convergence when the input signal is highly correlated (e.g., autoregressive and speech signals). The family of affine projection (AP) algorithms can speed up the convergence by using the recent multiple input vectors at each update of the filter weight vector [5]–[7], while at the cost of

Y. Yu and H. He are with the School of Information Engineering, Robot Technology Used for Special Environment Key Laboratory of Sichuan Province, Southwest University of Science and Technology, Mianyang, 621010, China (e-mail: yuyi_xyuan@163.com, hongsenhe@gmail.com).

B. Chen is School of Electronic and Information Engineering, Xi'an Jiaotong University, Xian, China (e-mail: chenbd@mail.xjtu.edu.cn).

J. Li is with the institute of Sound and Vibration Research, University of Southampton, Southampton SO17 1BJ, U.K. (e-mail: J.Li@soton.ac.uk).

Y. Zhang is with the Acoustic Science and Technology Laboratory and the College of Underwater Acoustic Engineering, Harbin Engineering University, Harbin, Heilongjiang, 150001, China (e-mail: zhangyouwen@hrbeu.edu.cn).

L. Lu is with School of Electronics and Information Engineering, Sichuan University, Chengdu, China (e-mail: lulu19900303@126.com).

increasing computational complexity due to the matrix inverse. A number of fast implementations of AP algorithm have been developed. However, each fast implementation also brings about own problem, e.g., poor numerical stability, degraded convergence performance, or difficulty of selecting the number of iterations, summarized in detail in [8]

Subband adaptive filter (SAF) is another efficient method to improve the convergence rate [2]. The SAF divides both the input and desired signals into multiple subbands through the analysis filters, and then uses the decimated subband signals to adjust its weight vector. Since the multiband structure of SAF avoids the aliasing and band edge effects by updating the fullband filter weight vector rather than multiple subfilters in the conventional structure, which makes it more attractive [2], [9], [10]. Based on this, Lee and Gan [10] developed the normalized SAF (NSAF) algorithm from the principle of minimum disturbance. The NSAF algorithm obtains significant convergence improvement for correlated input signals in comparison with the NLMS algorithm. The increase of the NSAF algorithm in the computational complexity is insignificant, especially in a high-order adaptive filter application (e.g., echo cancellation). The NSAF algorithm is considered as a generalization of the NLMS algorithm in the subband domain, where for a special case of one subband, the NSAF algorithm reduces to the NLMS algorithm. To avoid the signal delay problem from the input to the output of adaptive system due to the adopted analysis and synthesis filter banks, two delayless configurations for the NSAF algorithm that use an auxiliary loop to compute the output of adaptive system were developed in [11], and thus they are more suited for practical applications. The step size of the NSAF algorithm controls a compromise between fast convergence speed and low steady-state estimation error. To address this problem, various variable step size (VSS) schemes [12]–[14] and combination schemes [15], [16] have presented.

Regrettably, the performance of the NSAF algorithm from the l_2-norm minimization of subband errors is seriously deteriorated by the system output noise that includes impulsive noise. The impulsive noise has a typical feature that is, its realizations have large amplitude and appear with a small probability or a short duration. Such noise scenario is also often encountered in practical applications such as echo cancellation, underwater acoustics, audio processing, communications, and prediction of time-series [17]–[21]. To obtain the stable convergence in impulsive noise, the sign algorithm was firstly proposed by minimizing the absolute value of error signal, but with slow convergence rate [22].

The correntropy function can significantly compress the data with large amplitude, the maximum correntropy criterion (MCC) was frequently used for designing robust LMS-like algorithms [23]–[25], while this requires properly choosing the kernel width parameter. By inserting an upper bound on the squared error into the weights update to suppress the impulsive noise, the normalized least mean absolute third algorithm and its improvements were proposed [26], [27]. Similar to NLMS, these robust algorithms also have no decorrelation capability for correlated input signals. Also considering the decorrelation capability, reference [28] presented a sign SAF algorithm robust against impulsive noise by incorporating the sign strategy into the multiband-based SAF structure. By taking full advantage of the decorrelation feature of SAF for the input signal, the SSAF with individual-weighting-factors (IWF-SSAF) algorithm presented in [29] exhibits faster convergence than the SSAF algorithm. For both the SSAF and IWF-SSAF algorithms, many VSS variants were also proposed to improve the performance in terms of convergence rate and steady-state estimation error [30]–[33]; however, these VSS variants lose the tracking capability for abrupt changes of unknown system.

In some applications, the unknown system is usually sparse, i.e., only a fraction of entries in its impulse response are significantly different from zero; for instance, most of echo channels for both network/acoustic echo cancellation scenarios [34], [35]. It has been reported in the literature [6], [36]–[38] that the proportionate algorithms can improve the convergence in contrast with the non-proportionate counterparts for sparse systems, without the steady-state performance loss. To favor the sparsity of the unknown system, alternative approach is to add a penalty function based on the l_p-norm of the filter weights to the original cost function, where $p = 0$, 1, or $0 < p < 1$. This approach was firstly used in improving the LMS performance when identifying sparse systems [39], and now has been extended to various adaptive algorithms (see [40], [41] and references therein). Recently, these two sparse approaches were also combined to improve the algorithm's performance, e.g., the proportionate NLMS algorithm with the l_1-norm penalty [42] and the proportionate normalized MCC algorithm with the l_p-norm penalty [43]. In comparison, the main drawback of the second sparse approach is that appropriately sets the intensity parameter associated with the penalty term. Also because how to exploit the sparsity is not main issue of this paper, we only focus on the proportionate algorithm. Aiming at such systems in impulsive noise environments, the proportionate SSAF and IWF-SSAF algorithms were developed in [44] and [29], respectively. In [44], Ni *et al.* also proposed the AP-SSAF algorithm by extending the AP concept to the SSAF, further accelerating the convergence. The M-estimator can remove outliers with large amplitude from dataset and it has been used to develop robust adaptive algorithms against impulsive interferences [45]–[47]. On the M-estimate based algorithms, the AP M-estimate SAF (APM-SAF) algorithm was recently proposed [47] by introducing the Hampel's three-part redescending (HTPR) M-estimate function and the AP concept to the NSAF. The APM-SAF algorithm shows better convergence than the the

AP-SSAF algorithm, while it has also higher computational complexity. Additionally, the proportionate APM-SAF [47] was also proposed for sparse systems.

In this paper, we study the M-estimate based NSAF (M-NSAF) algorithm by using the modified Huber (MH) function. From the perspective of the algorithm framework, the M-NSAF algorithm is a special case of the APM-SAF algorithm with the AP order L being 1. Although the APM-SAF algorithm for $L > 1$ introduces the decorrelation property of AP in the time domain to accelerate the convergence for highly correlated input signals, it also loses the low-complexity merit of SAF. Naturally, its complexity is higher than that of the AP algorithm [48]. In addition, even for the M-NSAF algorithm, the theoretical analysis has been not reported yet. Therefore, studying the M-NSAF algorithm is still interesting in detail. The main contributions of this paper are as follows:

1) We analyze the stability condition, transient and steady-state behaviors of the M-NSAF algorithm in impulsive noise in detail, based on the vectorization and Kronecker product operations. The analysis results are supported by simulations.

2) To improve the M-NSAF performance in both the convergence rate and the steady-state estimation error, a VSS scheme is developed by minimizing the squared *a posteriori* subband errors. This VSS is also applied into the proportionate-type M-NSAF algorithm to improve the performance on estimating sparse systems.

3) The proposed algorithms are verified on system identification in α-stable noise and acoustic echo cancellation (AEC) with double-talk scenarios. In AEC, we apply the proposed algorithms into a delayless SAF structure.

The remaining part of this paper is organized as follows. Section II introduces the delayless SAF model and the M-NSAF algorithm. In Section III, we analyze the mean and mean-square theoretical models of the M-NSAF algorithm in impulsive noise. In Section IV, we design a VSS for improving both the M-NSAF algorithm and its proportionate variant. Extensive simulations are presented in Section V. In Section VI, conclusions are drawn.

II. A DELAYLESS SAF STRUCTURE AND M-NSAF ALGORITHM

At time index n, the available desired signal $d(n)$ and the input signal $u(n)$ follow the linear relation:

$$d(n) = \boldsymbol{u}^{\mathrm{T}}(n)\boldsymbol{w}^o + \nu(n) \tag{1}$$

where an $L \times 1$ vector $\boldsymbol{w}^o$ is the impulse response of the unknown system that be estimated, $\boldsymbol{u}(n) = [u(n), u(n - 1), ..., u(n - L + 1)]^{\mathrm{T}}$ is the $L \times 1$ input vector, $v(n)$ is the additive noise independent of $u(n)$, and $(\cdot)^{\mathrm{T}}$ denotes the transpose. In practice, in addition to Gaussian noise, $\nu(n)$ may also contain the impulsive noise with large amplitude.

A delayless multiband-structured SAF with N subbands is shown in Fig. 1 [11], [33]. By decomposing the desired signal $d(n)$ and the input signal $u(n)$ through the analysis filters $H_i(z)$, $i = 0, ..., N - 1$, we obtain band-dependent signals $d_i(n)$ and $u_i(n)$, respectively. Feeding $u_i(n)$ into an adaptive filter with the weight vector $\boldsymbol{w}(k)$ yields the subband output
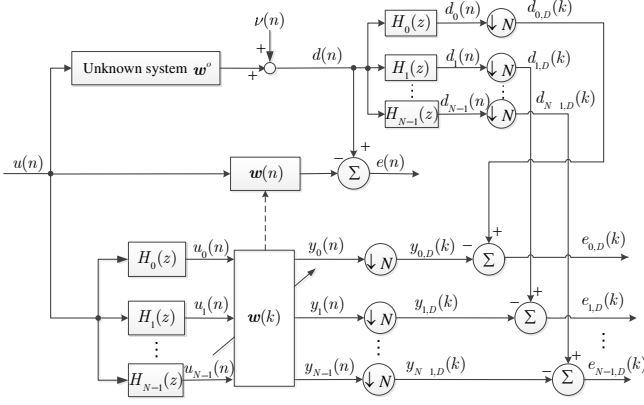
Fig. 1. Delayless multiband-structured SAF.

signal $y_i(n)$. For each subband i, $d_i(n)$ and $y_i(n)$ are critically decimated to lower sampled rate sequences $d_{i,D}(k)$ and $y_{i,D}(k)$, respectively, i.e., $d_{i,D}(k) = d_i(kN)$ and $y_{i,D}(k) = \boldsymbol{u}_i^{\mathrm{T}}(k)\boldsymbol{w}(k)$, where $\boldsymbol{u}_i(k) = [u_i(kN), u_i(kN), ..., u_i(kN - L+1)]^{\mathrm{T}}$. Note that the original sequences are denoted by the variable n and the decimated sequences are denoted by the variable k. The decimated subband error signals are given by

$$e_{i,D}(k) = d_{i,D}(k) - y_{i,D}(k) \qquad (2)$$
$$= d_{i,D}(k) - \boldsymbol{u}_i^{\mathrm{T}}(k)\boldsymbol{w}(k),$$

for $i = 0, ..., N-1$ and they are used to update the weight vector $\boldsymbol{w}(k)$ as an estimate of $\boldsymbol{w}^o$ at iteration k. This is the
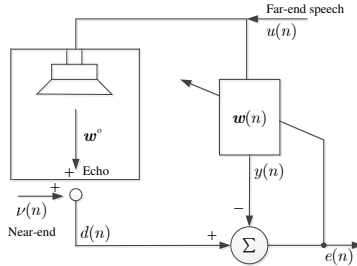

Fig. 2. A simple block diagram of AEC.

system identification problem.

However, when using the SAF in some applications, one finally hope to cancel the output error $e(n)$ in the original domain n. For instance, the AEC in the hands-free telephone system is shown in Fig. 2 [49], where $\boldsymbol{w}^o$ denotes the acoustic echo channel between loudspeaker and microphone. The far-end speech $u(n)$ played at the loudspeaker goes through $\boldsymbol{w}^o$ to produce the signal $\bar{d}(n) = \boldsymbol{u}^{\mathrm{T}}(n)\boldsymbol{w}^o$, and then the microphone pick up it and send to the far-end. Thus, the signal $\bar{d}(n)$ is unwanted echo for speaker at the far-end, which needs to be canceled. The adaptive filter by adjusting $\boldsymbol{w}(n)$ is to estimate $\boldsymbol{w}^o$, and its output $y(n) = \boldsymbol{u}^{\mathrm{T}}(n)\boldsymbol{w}(n)$ by filtering the far-end speech $u(n)$ is the replica of the echo. Since $\boldsymbol{w}(n) \to \boldsymbol{w}^o$ so that $y(n) \to \bar{d}(n)$, the output error $e(n)$ by subtracting $y(n)$ from $d(n)$ will be free of echo and sent to the far-end, thereby improving the call quality. In the SAF structure [10], the output error $e(n)$ can be reconstructed from the decimated subband

error signals $\{e_{i,D}(k)\}_{i=0}^{N-1}$ by means of the synthesis filters and the interpolation operator. However, this gives rise to an inherent signal path delay. To avoid such delay, in Fig. 1 we copy $\boldsymbol{w}(k)$ to $\boldsymbol{w}(n)$ once for every N input samples (i.e., when $n = kN$), then eliminating the echo by performing $e(n) = d(n) - y(n)$ in an auxiliary loop.

For estimating $\boldsymbol{w}^o$ in the decimated domain, the M-NSAF algorithm is formulated as

$$\boldsymbol{w}(k+1) = \boldsymbol{w}(k) + \mu \sum_{i=0}^{N-1} \frac{\varphi'(e_{i,D}(k))\boldsymbol{u}_i(k)}{||\boldsymbol{u}_i(k)||_2^2 + \delta} \qquad (3)$$

where $\mu > 0$ is the step size, and $||\cdot||_2$ denotes the l_2-norm of a vector. The score function $\varphi'(x) = \partial\varphi(x)/\partial x$ in (3) originates from the M-estimator $\varphi(x)$, with robustness against outliers. It should be stressed that δ is man-made regularization parameter, to avert the numerical divergence problem of (3) when $||\boldsymbol{u}_i(k)||_2^2$ is zero such as during the mute period of the far-end speech signal $u(n)$ in echo cancellation; usually, it is a small positive constant.

Different from the Hampel's three-part redescending (HTPR) function presented in [47], here we employ the modified Huber (MH) function for $\varphi(x)$:

$$\varphi(e_{i,D}(k)) = \begin{cases} e_{i,D}^2(k)/2, & \text{if } |e_{i,D}(k)| < \xi \\ \xi^2/2, & \text{if } |e_{i,D}(k)| \geq \xi, \end{cases} \qquad (4)$$

which corresponds to the score function:

$$\varphi'(e_{i,D}(k)) = \begin{cases} e_{i,D}(k), & \text{if } |e_{i,D}(k)| < \xi \\ 0, & \text{if } |e_{i,D}(k)| \geq \xi, \end{cases} \qquad (5)$$

where ξ is a threshold parameter. The main reason for choosing the MH function is the simplicity, which conveniently analyzes the performance of the M-NSAF algorithm in the sequel. Moreover, as can be seen in Fig. 7, the M-NSAF algorithm using the MH function has very close performance to that using the HTPR function in impulsive noise environments.

It can be seen from (3) and (5) that the M-NSAF algorithm is identical to the conventional NSAF algorithm when values of $e_{i,D}(k)$ are smaller than ξ. When $|e_{i,D}(k)| \geq \xi$, $\varphi'(e_{i,D}(k))$ equals to zero and the M-NSAF algorithm stops the adaptation to avoid impulsive interferences. In an ideal case, $|e_{i,D}(k)| \geq \xi$ is only required occurring for the appearance of impulsive interferences. To this end, the threshold ξ for each subband i is chosen automatically by $\xi_i = \kappa\sigma_{e,i}(k)$, where $\sigma_{e,i}^2(k)$ is the variance of $e_{i,D}(k)$ excluding impulsive noises and it can be estimated in a recursion way:

$$\hat{\sigma}_{e,i}^2(k) = \tau\hat{\sigma}_{e,i}^2(k-1) + c_\sigma(1-\tau)\mathrm{med}(\boldsymbol{a}_{e,i}). \qquad (6)$$

In (6), $0 < \tau < 1$ is the weighting factor and usually chosen by $\tau = 1 - N/(\theta_\tau L)$ with $\theta_\tau \geq 1$ (but $\tau = 0$ for the starting point of iterations), $\mathrm{med}(\cdot)$ denotes the median operator whose role is to remove the outliers in the error data set $\boldsymbol{a}_{e,i} = [e_{i,D}^2(k), e_{i,D}^2(k-1), ..., e_{i,D}^2(k-N_w+1)]$ with length of N_w, and $c_\sigma = 1.483(1+5/(N_w-1))$ is the correction factor [51, p.44]. It is noticed that the value of κ should not be too small and large to guarantee fast convergence and robust performance in resisting impulsive noise, respectively. For this

purpose, we choose the typical value $\kappa = 2.576$ [52]. This value means that, assuming that $e_{i,D}(k)$ is Gaussian distributed except when being contaminated accidentally by impulsive noise, the confidence level of halting the weights update (3) when $|e_{i,D}(k)| \geq \xi_i$ is 99%. In Appendix A, it is seen how c_σ and κ affect the performance of the M-NSAF algorithm.

III. PERFORMANCE ANALYSIS

In this section, we will discuss the convergence behavior of the M-NSAF algorithm in the presence of impulsive noise.

Subtracting both sides of (3) from $\boldsymbol{w}^o$, we have

$$\widetilde{\boldsymbol{w}}(k+1) = \widetilde{\boldsymbol{w}}(k) - \mu \sum_{i=0}^{N-1} \frac{\varphi'(e_{i,D}(k))\boldsymbol{u}_i(k)}{||\boldsymbol{u}_i(k)||_2^2} \tag{7}$$

where $\widetilde{\boldsymbol{w}}(k) \triangleq \boldsymbol{w}^o - \boldsymbol{w}(k)$ denotes the weights error vector. For convenience of analysis, here we also neglect the regularization δ in (3) as its value is small.

Let us define the probability of the event $|e_{i,D}(k)| < \xi_i$ at iteration k as

$$P_{e,i}(k) = P\{|e_{i,D}(k)| < \xi_i\}. \tag{8}$$

Accordingly, we can equivalently express (5) in the statistical sense as

$$\varphi'(e_{i,D}(k)) = P_{e,i}(k)e_{i,D}(k), \tag{9}$$

which further changes (7) to

$$\widetilde{\boldsymbol{w}}(k+1) = \widetilde{\boldsymbol{w}}(k) - \mu \sum_{i=0}^{N-1} \frac{P_{e,i}(k)e_{i,D}(k)\boldsymbol{u}_i(k)}{||\boldsymbol{u}_i(k)||_2^2}. \tag{10}$$

Equation (10) overcomes the nonlinearity of the score function $\varphi'(x)$ in (7), which contributes to the analysis simplification[1]. Before proceeding any further, we require the following assumptions.

Assumption 1: The input signal $u(n)$ is a zero-mean stationary random process with positive definite autocorrelation matrix $\boldsymbol{R}_u = \mathrm{E}\{\boldsymbol{u}(n)\boldsymbol{u}^{\mathrm{T}}(n)\}$.

Assumption 2: The additive noise $\nu(n)$ is modeled by the contaminated-Gaussian (CG) process. This is a frequently used model in studying adaptive algorithms robust against impulsive noise [45], [46], [53]. In detail, it includes the background noise $\nu_g(n)$ plus the impulsive noise $\nu_{im}(n)$, i.e., $\nu(n) = \nu_g(n) + \nu_{im}(n)$. The background noise $\nu_g(n)$ is drawn from a zero-mean white Gaussian random process with variance σ_g^2. For the impulsive noise, $\nu_{im}(n) = b(n)\eta(n)$, where $b(n)$ is a Bernoulli random process with the probability distribution function being $P\{b(n)=1\} = p_r$ and $P\{b(n)=0\} = 1-p_r$, and $\eta(n)$ is also a zero-mean white Gaussian random process but variance $\sigma_\eta^2 = \hbar\sigma_g^2$, $\hbar \gg 1$. It is clear that $\nu(n)$ is non-Gaussian except special cases of $p_r = 0$ or 1 [54].

Assumption 3: $\widetilde{\boldsymbol{w}}(k)$ is statistically independent of $\boldsymbol{u}_i(k)$ for $i = 0, ..., N-1$.

Note that Assumption 3 is the widely used *independence assumption* in the performance analysis of adaptive algorithms, e.g., references [55]–[57] for analyzing the NSAF algorithm.

Rewrite the decimated subband desired signals as

$$d_{i,D}(k) = \boldsymbol{u}_i^{\mathrm{T}}(k)\boldsymbol{w}^o + \nu_{i,D}(k), \tag{11}$$

where $\nu_{i,D}(k)$, $i = 0, ..., N-1$ denote the decimated subband noises. Giving that $\boldsymbol{h}_i$ is the impulse response of the i-th analysis filter with length of J, we have the relations:

$$\begin{aligned}\boldsymbol{u}_i(k) &= [\boldsymbol{u}(kN), ...\boldsymbol{u}(kN-J+1)]\boldsymbol{h}_i \\ \nu_{i,D}(k) &= \boldsymbol{h}_i^{\mathrm{T}}[\nu(kN), ..., \nu(kN-J+1)]^{\mathrm{T}}.\end{aligned} \tag{12}$$

Thus, based on the use of $P_{e,i}(k)$ and Assumption 2, we can change $e_{i,D}(k)$ in (10) to

$$e_{i,D}(k) = \boldsymbol{u}_i^{\mathrm{T}}(k)\widetilde{\boldsymbol{w}}(k) + \nu_{\mathrm{g},i,D}(k), \tag{13}$$

where $\nu_{\mathrm{g},i,D}(k)$ is obtained from $\nu_{\mathrm{g}}(n)$ according to the same operation as $\nu_{i,D}(k)$ in (12). Since $\boldsymbol{h}_i$ is deterministic, $\nu_{\mathrm{g},i,D}(k)$ is zero-mean white Gaussian with variance $\sigma_{\mathrm{g},i}^2 = ||\boldsymbol{h}_i||_2^2\sigma_{\mathrm{g}}^2$ [2].

Plugging (13) into (10), we obtain

$$\begin{aligned}\widetilde{\boldsymbol{w}}(k+1) = {}& \left(\boldsymbol{I}_L - \mu \sum_{i=0}^{N-1} P_{e,i}(k)\boldsymbol{A}_i(k)\right)\widetilde{\boldsymbol{w}}(k) - \\ & \mu \sum_{i=0}^{N-1} P_{e,i}(k)\nu_{\mathrm{g},i,D}(k)\boldsymbol{q}_i(k),\end{aligned} \tag{14}$$

where $\boldsymbol{A}_i(k) = \frac{\boldsymbol{u}_i(k)\boldsymbol{u}_i^{\mathrm{T}}(k)}{||\boldsymbol{u}_i(k)||_2^2}$, $\boldsymbol{q}_i(k) = \frac{\boldsymbol{u}_i(k)}{||\boldsymbol{u}_i(k)||_2^2}$, and $\boldsymbol{I}_L$ is the identity matrix of size $L \times L$. Equation (14) is the starting point to study the mean and mean-square behaviors of the M-NSAF algorithm in the sequel.

A. Mean behavior

Taking the expectation of both sides of (14) under Assumption 2, it is established that

$$\mathrm{E}\{\widetilde{\boldsymbol{w}}(k+1)\} = \left(\boldsymbol{I}_L - \mu \sum_{i=0}^{N-1} P_{e,i}(k)\mathrm{E}\{\boldsymbol{A}_i(k)\}\right)\mathrm{E}\{\widetilde{\boldsymbol{w}}(k)\}, \tag{15}$$

To ensure the convergence of $\mathrm{E}\{\widetilde{\boldsymbol{w}}(k)\}$ over iterations, the spectral radius of the matrix $\left(\boldsymbol{I}_L - \mu \sum_{i=0}^{N-1} P_{e,i}(k)\mathrm{E}\{\boldsymbol{A}_i(k)\}\right)$ is required to be less than 1. Consequently, we obtain the step size range for the mean stability of the algorithm:

$$0 < \mu < \frac{2}{P_{\max}(k) \sum_{i=0}^{N-1} \lambda_{\max}(\mathrm{E}\{\boldsymbol{A}_i(k)\})}, \tag{16}$$

where $P_{\max}(k) = \max\{P_{e,i}(k), i = 0, ..., N-1\}$, and $\lambda_{\max}(\cdot)$ denotes the maximum eigenvalue of the matrix.

When the algorithm reaches the steady-state, i.e., $k \to \infty$, from (15) we will obtain:

$$\mathrm{E}\{\widetilde{\boldsymbol{w}}(\infty)\} = \boldsymbol{0}. \tag{17}$$

This relation reveals that the M-NSAF algorithm is unbiased for estimating $\boldsymbol{w}^o$.

---

[1] Similar operation can also be used to simplify the existing analyses in [45], [46] for the MH-based LMS/NLMS algorithms.

[2] If the analysis filter bank is assume to be identical and paraunitary, we have $||\boldsymbol{h}_i||_2^2 = 1/N$, which is commonly used in the performance analysis of NSAF-type algorithm [55], [57].

B. Mean-square behavior

Post-multiplying both sides of (14) by $\widetilde{\boldsymbol{w}}^{\mathrm{T}}(k+1)$, and then taking the expectation of its both sides, the following recursion can be found:

$$\tilde{\mathcal{W}}(k+1) = \boldsymbol{I}_L - \mu \sum_{i=0}^{N-1} P_{e,i}(k)\mathrm{E}\{\boldsymbol{A}_i(k)\}\tilde{\mathcal{W}}(k) -$$

$$\mu\tilde{\mathcal{W}}(k) \sum_{i=0}^{N-1} P_{e,i}(k)\mathrm{E}\{\boldsymbol{A}_i(k)\} +$$

$$\mu^2 \mathrm{E}\left\{\sum_{i=0}^{N-1} P_{e,i}(k)\boldsymbol{A}_i(k)\tilde{\mathcal{W}}(k)\sum_{i=0}^{N-1} P_{e,i}(k)\boldsymbol{A}_i(k)\right\} +$$

$$\mu^2 \sum_{i=0}^{N-1} P_{e,i}^2(k)\sigma_{\mathrm{g},i}^2\mathrm{E}\{\boldsymbol{q}_i(k)\boldsymbol{q}_i^{\mathrm{T}}(k)\},$$

(18)

where $\tilde{\mathcal{W}}(k) \triangleq \mathrm{E}\{\widetilde{\boldsymbol{w}}(k)\widetilde{\boldsymbol{w}}^{\mathrm{T}}(k)\}$ defines the autocorrelation matrix of the weights error vector. Note that, we apply Assumptions 2 and 3 for deriving (18) which makes that the expectation of the cross-terms involving $\boldsymbol{u}_i(k)$ and $\nu_{\mathrm{g},i,D}(k)$ is zero. We also use $\mathrm{E}\{\boldsymbol{A}_i(k)\} = \mathrm{E}\{\boldsymbol{A}_i^{\mathrm{T}}(k)\}$ due to its symmetry.

To mainly separate $\tilde{\mathcal{W}}(k)$ from the third term at the right side of (18), we will resort to the vectorization operation and the Kronecker product [58]. Let $\mathrm{vec}(\cdot)$ represent a transformation from an $L \times L$ matrix to an $L^2 \times 1$ vector, formed by successively stacking the columns of the matrix, and $\mathrm{vec}^{-1}(\cdot)$ be the inverse operation. Giving three matrices $\boldsymbol{X}$, $\boldsymbol{Y}$, and $\boldsymbol{\Sigma}$ with compatible sizes, there is $\mathrm{vec}(\boldsymbol{X}\boldsymbol{\Sigma}\boldsymbol{Y}) = (\boldsymbol{Y}^{\mathrm{T}} \otimes \boldsymbol{X})\mathrm{vec}(\boldsymbol{\Sigma})$, where $\otimes$ denotes the Kronecker product. Therefore, by imposing the vectorization operation on both sides of the recursion (18), we arrive at:

$$\mathrm{vec}(\tilde{\mathcal{W}}(k+1)) = \mathcal{F}_k\mathrm{vec}(\tilde{\mathcal{W}}(k)) +$$
$$\mu^2 \sum_{i=0}^{N-1} P_{e,i}^2(k)\sigma_{\mathrm{g},i}^2\mathrm{vec}(\mathrm{E}\{\boldsymbol{q}_i(k)\boldsymbol{q}_i^{\mathrm{T}}(k)\}),$$

(19)

where

$$\mathcal{F}_k = \boldsymbol{I}_{L^2} - \mu\boldsymbol{I}_L \otimes \sum_{i=0}^{N-1} P_{e,i}(k)\mathrm{E}\{\boldsymbol{A}_i(k)\} -$$

$$\mu \sum_{i=0}^{N-1} P_{e,i}(k)\mathrm{E}\{\boldsymbol{A}_i(k)\} \otimes \boldsymbol{I}_L +$$

$$\mu^2 \underbrace{\mathrm{E}\left\{\sum_{i=0}^{N-1} P_{e,i}(k)\boldsymbol{A}_i(k) \otimes \sum_{i=0}^{N-1} P_{e,i}(k)\boldsymbol{A}_i(k)\right\}}_{(a)}.$$

(20)

Furthermore, to simplify the term (a) in (20), we may make a reasonable approximation:

$$\mathrm{E}\left\{\sum_{i=0}^{N-1} P_{e,i}(k)\boldsymbol{A}_i(k) \otimes \sum_{i=0}^{N-1} P_{e,i}(k)\boldsymbol{A}_i(k)\right\} \approx$$

$$\left(\frac{1}{N}\sum_{i=0}^{N-1} P_{e,i}(k)\right)^2 \mathrm{E}\left\{\sum_{i=0}^{N-1}\boldsymbol{A}_i(k) \otimes \sum_{i=0}^{N-1}\boldsymbol{A}_i(k)\right\},$$

(21)

due to the fact that the difference between $P_{e,i}(k)$ at different subbands is small as can be seen in Appendix B.

In the light of the definition of the mean square deviation (MSD), i.e., $\mathrm{MSD}(k) \triangleq \mathrm{E}\{\widetilde{\boldsymbol{w}}^{\mathrm{T}}(k)\widetilde{\boldsymbol{w}}(k)\} = \mathrm{Tr}(\tilde{\mathcal{W}}(k))$, we can use (19) to describe the transient MSD behavior of the M-NSAF algorithm, where $\mathrm{Tr}(\cdot)$ indicates the trace of a matrix. Moreover, the excess mean-square error (EMSE) of the algorithm can also be computed as $\mathrm{EMSE}(n) = \mathrm{Tr}(\tilde{\mathcal{W}}(n)\boldsymbol{R}_u)$.

By assuming the existence of $(\boldsymbol{I}_{L^2} - \mathcal{F}_\infty)^{-1}$ in the steady-state, it is derived from (19) that

$$\mathrm{vec}(\tilde{\mathcal{W}}(\infty)) = \mu^2(\boldsymbol{I}_{L^2} - \mathcal{F}_\infty)^{-1} \times$$
$$\sum_{i=0}^{N-1} P_{e,i}^2(\infty)\sigma_{\mathrm{g},i}^2\mathrm{vec}(\mathrm{E}\{\boldsymbol{q}_i(k)\boldsymbol{q}_i^{\mathrm{T}}(k)\}).$$

(22)

Based on the property $\mathrm{Tr}(\boldsymbol{X}\boldsymbol{Y}) = \mathrm{vec}^{\mathrm{T}}(\boldsymbol{X}^{\mathrm{T}})\mathrm{vec}(\boldsymbol{Y})$, the steady-state MSD of the M-NSAF algorithm is obtained as

$$\mathrm{MSD}(\infty) = \mu^2\mathrm{vec}^{\mathrm{T}}(\boldsymbol{I}_L)(\boldsymbol{I}_{L^2} - \mathcal{F}_\infty)^{-1} \times$$
$$\sum_{i=0}^{N-1} P_{e,i}^2(\infty)\sigma_{\mathrm{g},i}^2\mathrm{vec}(\mathrm{E}\{\boldsymbol{q}_i(k)\boldsymbol{q}_i^{\mathrm{T}}(k)\}).$$

(23)

Analogy to the Appendix A in [59], we can establish from (19) that the M-NSAF algorithm will be mean-square stable when all the eigenvalues of $\mathcal{F}_k$ are in $(-1, 1)$. Nevertheless, the resulting convergence condition is slightly complicated. Hence, an alternative method will be given. By taking the squared l_2-norm and the expectation for both sides of (14) leads to

$$\mathrm{E}\{\|\widetilde{\boldsymbol{w}}(k+1)\|_2^2\} = \underbrace{\mathrm{E}\{\|\widetilde{\boldsymbol{w}}(k)\|_{\boldsymbol{\Xi}_k}^2\}}_{(b)} +$$
$$\mu^2 \sum_{i=0}^{N-1} P_{e,i}^2(k)\sigma_{\mathrm{g},i}^2\mathrm{E}\left\{\frac{1}{\|\boldsymbol{u}_i(k)\|_2^2}\right\},$$

(24)

where

$$\boldsymbol{\Xi}_k = \boldsymbol{I}_L - \sum_{i=0}^{N-1} \mu(2 - \mu P_{e,i}(k))P_{e,i}(k)\mathrm{E}\{\boldsymbol{A}_i(k)\}, \quad (25)$$

and $\|\widetilde{\boldsymbol{w}}(k)\|_{\boldsymbol{\Xi}_k}^2 \triangleq \widetilde{\boldsymbol{w}}^{\mathrm{T}}(k)\boldsymbol{\Xi}_k\widetilde{\boldsymbol{w}}(k)$. For deriving (24), we also employ the fact that the decimated input vectors at different subbands are orthogonal, i.e., $\boldsymbol{u}_i^{\mathrm{T}}(k)\boldsymbol{u}_j(k) = 0$ when $i \neq j$ [10].

Based on Assumption 1, we take advantage of Lemma 1 in [55] to yield

$$\boldsymbol{0} \preceq \lambda_{\min}(\mathrm{E}\{\boldsymbol{A}_i(k)\})\boldsymbol{I}_L \preceq \mathrm{E}\{\boldsymbol{A}_i(k)\}$$
$$\preceq \lambda_{\max}(\mathrm{E}\{\boldsymbol{A}_i(k)\})\boldsymbol{I}_L,$$

(26)

where $\lambda_{\min}(\cdot)$ is the minimum eigenvalue of the matrix, and $\boldsymbol{X} \preceq \boldsymbol{Y}$ denotes the matrix $(\boldsymbol{Y} - \boldsymbol{X})$ is semi-positive definite. Thus, assuming $0 < \mu < 2/P_{e,i}(k)$ (which will be given in (30) below), we can get

$$\beta_1\boldsymbol{I}_L \preceq \boldsymbol{\Xi}_k \preceq \beta_2\boldsymbol{I}_L, \quad (27)$$

where

$$\beta_1 = 1 - \sum_{i=0}^{N-1} \mu(2 - \mu P_{e,i}(k))P_{e,i}(k)\lambda_{\max}(\mathrm{E}\{\boldsymbol{A}_i(k)\}),$$

$$\beta_2 = 1 - \sum_{i=0}^{N-1} \mu(2 - \mu P_{e,i}(k))P_{e,i}(k)\lambda_{\min}(\mathrm{E}\{\boldsymbol{A}_i(k)\}).$$

(28)

Based on (27), the term (b) in (24) fulfills the following inequality:

$$0 < \beta_1 \mathrm{E}\{||\widetilde{\boldsymbol{w}}(k)||_2^2\} \leq \mathrm{E}\{||\widetilde{\boldsymbol{w}}(k)||_{\Xi_k}^2\} \\ \leq \beta_2 \mathrm{E}\{||\widetilde{\boldsymbol{w}}(k)||_2^2\}. \quad (29)$$

It is concluded from (24) and (29) that the algorithm is mean-square stable if and only if $0 < \beta_1 < \beta_2 < 1$, which further results in the step size range:

$$0 < \mu < 2/P_{\max}(k). \quad (30)$$

Remark 1: Combing (16) and (30), we unify the step size conditions to guarantee the stability of the M-NSAF algorithm:

$$0 < \mu < \frac{1}{P_{\max}(k)} \min\left(\frac{2}{\sum_{i=0}^{N-1} \lambda_{\max}(\mathrm{E}\{\boldsymbol{A}_i(k)\})}, 2 \right). \quad (31)$$

Assuming $P_{\max}(k) = 1$, (31) is the convergence condition of the NSAF algorithm. In comparison, the M-NSAF algorithm has a slightly wider convergence condition due to $P_{\max}(k) < 1$ (see the following subsection C or Appendix B).

Remark 2: As can be seen in (24), the term (b) is determining the convergence behavior of the M-NSAF algorithm. Further the matrix Ξ_k influences the convergence rate, because it is positive definite under the convergence condition (30). According to (27) and (29), the fast convergence speed occurs when the quadratic functions β_1 and β_2 on the step size μ are minimum. Taking the derivative of β_1 and β_2 with respect to μ, we further obtain the optimal μ for the fast convergence:

$$\mu_{fast} = \frac{\sum_{i=0}^{N-1} P_{e,i}(k)}{\sum_{i=0}^{N-1} P_{e,i}^2(k)} \leqslant \frac{1}{P_{\max}(k)}. \quad (32)$$

That means that, the M-NSAF algorithm with larger step size achieves faster convergence speed for $\mu \in (0, \mu_{fast}]$. After μ is larger than μ_{fast}, increasing the step size will lead to slow convergence of the M-NSAF algorithm. For the case of $P_{\max}(k) = 1$, $\mu_{fast} = 1$ is the optimal step size for the convergence of the NSAF algorithm [2]. Since also $P_{\max}(k) < 1$ and close to 1, the optimal step size of the M-NSAF algorithm is slightly larger than 1 for the best convergence. It follows that a practical range on the step size is from 0 to 1 for the M-NSAF algorithm. In addition, the second term at the right side of (24) controls the steady-state performance of the algorithm. Obviously, the smallest steady-state estimation error will be obtained when the step size μ is zero.

C. Calculation of $P_{e,i}(k)$

Now we continue to derive the probability $P_{e,i}(k)$ for implementing the theoretical models (15) and (19). Under Assumption 2, we use the law of total probability to recast (8) as

$$P_{e,i}(k) = P\{b(k) = 1\} P\{|e_{\nu,i,D}(k)| < \xi_i\} + \\ P\{b(k) = 0\} P\{|e_{g,i,D}(k)| < \xi_i\}, \\ = p_r P\{|e_{\nu,i,D}(k)| < \xi_i\} + \\ (1 - p_r) P\{|e_{g,i,D}(k)| < \xi_i\}, \quad (33)$$

where $e_{\nu,i,D}(k) = \boldsymbol{u}_i^{\mathrm{T}}(k)\widetilde{\boldsymbol{w}}(k) + \nu_{g,i,D}(k) + \nu_{\eta,i,D}(k)$, $e_{g,i,D}(k) = \boldsymbol{u}_i^{\mathrm{T}}(k)\widetilde{\boldsymbol{w}}(k) + \nu_{g,i,D}(k)$, and $\nu_{\eta,i,D}(k)$ comes from $\eta(n)$ by the subband decomposition.

Since both $\nu_{g,i,D}(k)$ and $\nu_{\eta,i,D}(k)$ are Gaussian random variables, $e_{\nu,i,D}(k)$ and $e_{g,i,D}(k)$ can be assumed to be zero mean Gaussian variables for a long adaptive filter [59]; accordingly, we obtain

$$P\{|e_\nu| < \xi\} \triangleq \frac{1}{\sqrt{2\pi}\sigma_{e,\nu}} \int_{-\xi}^{\xi} \exp(-\frac{e_\nu^2}{2\sigma_{e,\nu}^2}) de_\nu \\ = \mathrm{erf}(\xi/\sqrt{2}\sigma_{e,\nu}), \quad (34)$$

where $\mathrm{erf}(x) \triangleq \frac{2}{\sqrt{\pi}} \int_0^x \exp(-t^2) dt$, and $\sigma_{e,\nu}^2$ is the variance of e_ν. In (34), we omit the notations i, k and D for convenience. Then, plugging (34) into (33) to yield

$$P_{e,i}(k) = p_r \mathrm{erf}\left(\frac{\xi_i}{\sqrt{2}\sigma_{e_{\nu,i}}(k)} \right) + (1 - p_r) \mathrm{erf}\left(\frac{\xi_i}{\sqrt{2}\sigma_{e_{g,i}}(k)} \right), \quad (35)$$

where $\sigma_{e_{\nu,i}}^2(k) = \mathrm{Tr}(\tilde{\mathcal{W}}(k)\boldsymbol{R}_{u,i}) + (1+\hbar)||\boldsymbol{h}_i||_2^2\sigma_g^2$, $\sigma_{e_{g,i}}^2(k) = \mathrm{Tr}(\tilde{\mathcal{W}}(k)\boldsymbol{R}_{u,i}) + ||\boldsymbol{h}_i||_2^2\sigma_g^2$, and $\boldsymbol{R}_{u,i} = \mathrm{E}\{\boldsymbol{u}_i(k)\boldsymbol{u}_i^{\mathrm{T}}(k)\}$. Moreover, ξ_i is given by

$$\xi_i = \kappa \sigma_{e_{g,i}}(k). \quad (36)$$

It is noticed that because $\sigma_{e_{g,i}}^2(k) \ll \sigma_{e_{\nu,i}}^2(k)$, $P_{e,i}(k)$ computed by (35) is less than 1. In the steady-state, by assuming that $\mathrm{Tr}(\tilde{\mathcal{W}}(k)\boldsymbol{R}_{u,i})$ is negligible in contrast with $||\boldsymbol{h}_i||_2^2\sigma_g^2$, then $P_{e,i}(\infty)$ can be approximated by

$$P_{e,i}(\infty) = p_r \mathrm{erf}\left(\frac{\xi_i}{\sqrt{2(1+\hbar)||\boldsymbol{h}_i||_2^2\sigma_g^2}} \right) + \\ (1 - p_r) \mathrm{erf}\left(\frac{\xi_i}{\sqrt{2||\boldsymbol{h}_i||_2^2\sigma_g^2}} \right), \quad (37)$$

where $\xi_i = \kappa\sqrt{||\boldsymbol{h}_i||_2^2\sigma_g^2}$.

IV. PERFORMANCE IMPROVEMENTS

A. VSS Design

As claimed in Remark 2, in the M-NSAF algorithm there is a trade-off between fast convergence and low steady-state estimation error on choosing the step size[3]. To overcome such issue, the VSS [12], the convex combination of two independently run filters [15], and the combined step sizes [16] are three types of efficient techniques. In this section, we design a VSS scheme to improve the M-NSAF performance.

---

[3]It is a common issue in the step size based adaptive algorithms.

For moving forward, we replace the fixed step size μ in (10) with the time-dependent step sizes $\mu_i(k)$, $i = 1, ..., N$:

$$\widetilde{\boldsymbol{w}}(k+1) = \widetilde{\boldsymbol{w}}(k) - \sum_{i=0}^{N-1} \mu_i(k) \frac{P_{e,i}(k)e_{i,D}(k)\boldsymbol{u}_i(k)}{||\boldsymbol{u}_i(k)||_2^2}. \tag{38}$$

Recalling the orthogonality of decimated input vectors at different subbands [10] and pre-multiplying $\boldsymbol{u}_i^{\mathrm{T}}(k)$ both sides of (38), we arrive at

$$e_{i,p}(k) = e_{i,a}(k) - \mu_i(k)P_{e,i}(k)e_{i,D}(k)$$
$$\overset{(11)}{=} [1 - \mu_i(k)P_{e,i}(k)]e_{i,a}(k) + \mu_i(k)P_{e,i}(k)\nu_{\mathrm{g},i,D}(k), \tag{39}$$

where $e_{i,p}(k) \triangleq \boldsymbol{u}_i^{\mathrm{T}}(k)\widetilde{\boldsymbol{w}}(k+1)$ and $e_{i,a}(k) \triangleq \boldsymbol{u}_i^{\mathrm{T}}(k)\widetilde{\boldsymbol{w}}(k)$ denote the *a posteriori* and the *a priori* decimated subband errors, respectively.

Under Assumption 3, we enforce the expectation to the square of both sides of (39),

$$\mathrm{E}\{e_{i,p}^2(k)\} = [1 - \mu_i(k)P_{e,i}(k)]^2\mathrm{E}\{e_{i,a}^2(k)\} + \mu_i^2(k)P_{e,i}^2(k)\sigma_{g,i}^2. \tag{40}$$

At each iteration k, the optimum step size $\mu_i(k)$ will be obtained when $\mathrm{E}\{e_{i,p}^2(k)\}$ is minimum. Consequently, setting the derivative of $\mathrm{E}\{e_{i,p}^2(k)\}$ with respect to $\mu_i(k)$ to zero, we obtain $\mu_i(k)$ for each subband i:

$$\mu_i(k) = \frac{1}{P_{e,i}(k)} \frac{\mathrm{E}\{e_{i,a}^2(k)\}}{\mathrm{E}\{e_{i,a}^2(k)\} + \sigma_{g,i}^2}. \tag{41}$$

According to the definition of $P_{e,i}(k)$ and its property that $P_{\max}(k) < 1$ and close to 1, an efficient VSS scheme can be expressed as

$$\mu_i(k) = \begin{cases} \dfrac{\mathrm{E}\{e_{i,a}^2(k)\}}{\mathrm{E}\{e_{i,a}^2(k)\} + \sigma_{g,i}^2}, & \text{if } |e_{i,D}(k)| < \xi_i \\ 0, & \text{if } |e_{i,D}(k)| \geq \xi_i. \end{cases} \tag{42}$$

The equation above illustrates that the adaptation of the step size is frozen when $|e_{i,D}(k)| \geq \xi_i$, providing the robustness against impulsive noise.

We have $\mathrm{E}\{e_{i,D}^2(k)\} = \mathrm{E}\{e_{i,a}^2(k)\} + \sigma_{g,i}^2$ under Assumption 3, thus (42) is further changed as

$$\mu_i(k) = \begin{cases} \dfrac{\mathrm{E}\{e_{i,a}^2(k)\}}{\mathrm{E}\{e_{i,D}^2(k)\}}, & \text{if } |e_{i,D}(k)| < \xi_i \\ 0, & \text{if } |e_{i,D}(k)| \geq \xi_i. \end{cases} \tag{43}$$

One can see that the implementation of (43) depends on the second-order moments $\mathrm{E}\{e_{i,D}^2(k)\}$ and $\mathrm{E}\{e_{i,a}^2(k)\}$. Generally, the exponential window method [12], [49] is a simple and efficient method to estimate them. Specifically, $\mathrm{E}\{e_{i,D}^2(k)\}$ is estimated by $\hat{\sigma}_{e,i}^2(k)$,

$$\hat{\sigma}_{e,i}^2(k) = \chi\sigma_{e,i}^2(k-1) + (1-\chi)e_{i,D}^2(k), \tag{44}$$

where χ is a forgetting factor and chosen via $\chi = 1 - 1/(\theta_\chi L)$ with $\theta_\chi \geq 1$. As in [12], $\mathrm{E}\{e_{i,a}^2(k)\}$ is estimated by $\hat{\sigma}_{e,a,i}^2(k)$,

$$\hat{\sigma}_{e,a,i}^2(k) = \frac{||\hat{\boldsymbol{r}}_i(k)||_2^2}{\hat{\sigma}_{u,i}^2(k) + \epsilon_1} \tag{45}$$

where

$$\hat{\sigma}_{u,i}^2(k) = \chi\sigma_{u,i}^2(k-1) + (1-\chi)u_i^2(kN),$$
$$\hat{\boldsymbol{r}}_i(k) = \chi\hat{\boldsymbol{r}}_i(k-1) + (1-\chi)\boldsymbol{u}_i(k)e_{i,D}(k), \tag{46}$$

and ϵ_1 is a small positive number to avoid the division by zero. It is worth mentioning that both (44) and (46) are performed only when $|e_{i,D}(k)| < \xi_i$ to prevent the step sizes from impulsive interferences.

Considering the fact that the estimates $\hat{\sigma}_{e,i}^2(k)$ and $\hat{\sigma}_{e,a,i}^2(k)$ at the early stage of performing the algorithm have relatively large error compared to their true values, as such we set the step size to 1 when $k \leq L$, to guarantee fast initial convergence of the algorithm. Accordingly, for each subband i we reformulate (43) as

$$\begin{aligned} &\text{if } |e_{i,D}(k)| < \xi_i \\ &\quad \text{if } k \leq L \\ &\qquad \mu_i(k) = 1 \\ &\quad \text{else} \\ &\qquad \mu_i(k) = \frac{||\hat{\boldsymbol{r}}_i(k)||_2^2}{\hat{\sigma}_{e,i}^2(k)(\hat{\sigma}_{u,i}^2(k) + \epsilon_1)} \\ &\quad \text{end} \\ &\text{end} \end{aligned} \tag{47}$$

Equation (43) shows that the VSS cannot exceed 1 so that the proposed VSS-M-NSAF algorithm is stable. However, the implementation (47) of VSS may be larger than 1 or even 2 at some iterations due to replacing the second-order moments in (43) with their estimates. According to Remark 2, therefore, we impose a limitation on the step size that $\mu_i(k) = \min\{\mu_i(k), 1\}$ for ensuring good convergence performance. Following (47), the VSS-M-NSAF algorithm performs the weights update:

$$\boldsymbol{w}(k+1) = \boldsymbol{w}(k) + \sum_{i=0}^{N-1} \mu_i(k)\frac{\varphi'(e_{i,D}(k))\boldsymbol{u}_i(k)}{\boldsymbol{u}_i^{\mathrm{T}}(k)\boldsymbol{u}_i(k)}. \tag{48}$$

B. M-PNSAF Algorithm

Motivated by the proportionate rule [6], [36]–[38], the proportionate M-NSAF (M-PNSAF) algorithm for updating the filter's weights is obtained in a straightforward way as

$$\boldsymbol{w}(k+1) = \boldsymbol{w}(k) + \mu \sum_{i=0}^{N-1} \frac{\varphi'(e_{i,D}(k))\boldsymbol{G}(k)\boldsymbol{u}_i(k)}{\boldsymbol{u}_i^{\mathrm{T}}(k)\boldsymbol{G}(k)\boldsymbol{u}_i(k)}, \tag{49}$$

where $\boldsymbol{G}(k)$ is a diagonal matrix of size $L \times L$, whose diagonal elements $g_m(k)$, $m = 1, ..., L-1$ are individual gain coefficients assigned to every weight $w_m(k)$ in $\boldsymbol{w}(k)$. Also, $g_m(k)$ is in proportion to the amplitude of $w_m(k)$, thereby making use of the underlying sparsity of the unknown system. By comparing (3) and (49), it concludes that the M-PNSAF update will reduce to the M-NSAF update when $\boldsymbol{G}(k)$ is the identity matrix. In the existing literature, several strategies of computing $\boldsymbol{G}(k)$ have been developed, summarized in [60] and references therein. Among them, one of the most attractive

rules for computing $g_m(k)$, $m = 1, ..., L - 1$ is formulated as [37]

$$g_m(k) = \frac{1 - \zeta}{2L} + (1 + \zeta)\frac{|w_m(k)|}{2\sum\limits_{m=1}^{L}|w_m(k)| + \epsilon_2}, \quad (50)$$

which is robust to the unknown systems with different levels of sparsity. For most applications such as echo cancellation, good choices of ζ are 0 or -0.5 [37].

Analogously, we apply the VSS (47) to further improve the M-PNSAF performance in both convergence rate and steady-state estimation error and propose the VSS-M-PNSAF algorithm. Actually, this VSS (47) can also be derived from (49) by following the same procedures as in the previous subsection. The difference is that we use another orthogonal assumption $\boldsymbol{u}_i^T(k)\boldsymbol{G}(k)\boldsymbol{u}_j(k) = 0$ at different subbands i, j due mainly to slow variation of $\boldsymbol{G}(k)$ and $\text{Tr}\{\boldsymbol{G}(k)\} = 1$ [61].

C. Computational complexity

Table I summarizes the computational complexity of the NSAF, IWF-SSAF, BDVSS (i.e., IWF-SSAF with band-dependent VSS developed in [33]), M-NSAF, M-PNSAF, VSS-M-NSAF, and VSS-M-PNSAF algorithms in the context of system identification, in terms of the total number of additions, multiplications, divisions, and square-roots per input sample index n. These SAF algorithms have inherent complexity for partitioning the input signal $u(n)$ and the desired signal $d(n)$, which requires $2(J - 1)N$ additions and $2JN$ multiplications. For larger N (number of subbands), the length J of the analysis filter is larger, thereby the SAF's complexity is also higher. In AEC applications, L would be larger than the product JN in most cases; as such, the NSAF algorithm has only slight increase in the complexity relative to the fullband NLMS algorithm [10]. Compared with the NSAF algorithm, the additional complexity of the M-NSAF algorithm stems mainly from (6), i.e., $N_w \log_2 N_w + 1$ additions, 4 multiplications, and 1 square-root, where the comparisons required in the algorithm are counted as additions. Due to the proposed VSS formulas (44), (46), and (47), the VSS-M-NSAF (or VSS-M-PNSAF) algorithm needs more $L + 5$ additions, $2L + 7$ multiplications, and 1 division than the M-NSAF (or M-PNSAF) algorithm. As we shall see, the increase in the complexity leads to the algorithm's performance improvement.

V. SIMULATION RESULTS

In this section, we present several simulations to evaluate the previous theoretical results and the proposed algorithms. Both the adaptive filter and the unknown system have the same length. Cosine modulated analysis filter banks with N subbands are used for the subband structure. All curves are obtained by averaging the results over 200 independent trials, unless otherwise specified.

A. Verification of Analysis Results

The unknown system $\boldsymbol{w}^o$ with $L = 32$ taps is generated from a uniform distribution of $[-0.5, 0.5]$ and then normalized

by $||\boldsymbol{w}^o||_2 = 1$. The input signal $u(n)$ is obtained by filtering a zero-mean white Gaussian noise with unit variance through the first-order autoregressive system with a pole at 0.9; thus, the input vector $\boldsymbol{u}(n)$ is highly correlated with the eigenvalue spread of 263 as compared to the white signal with the eigenvalue spread of 1. As stated in Assumption 2, the additive noise $\nu(n)$ added to the unknown system output is drawn from a CG process. The variance σ_g^2 of the background noise component corresponds to a signal-to-noise ratio (SNR) defined as $10\log_{10}(\sigma_{\bar{d}}^2/\sigma_g^2)$, where $\sigma_{\bar{d}}^2 = \text{E}\{\bar{d}^2(n)\}$ is the power of the output signal of the unknown system. We set the impulsive noise parameters to $\hbar = 300000$ and $p_r = 0.001$. Correspondingly, the ratio $r_{im-g} = \sigma_{im}^2/\sigma_g^2 = p_r\hbar$ [45] measuring the impulsive characteristic of the CG noise equals to 300. The expectations on subband inputs in the theoretical expressions are obtained by the ensemble average.

Fig. 3 shows the MSD results of the NSAF algorithm and the M-NSAF algorithm using different step sizes μ in CG noise. It is clear to see, the convergence of the NSAF algorithm is poor in the presence of impulsive noise, while the M-NSAF algorithm is stable for step sizes in the range of $(0, 2)$. Also, there is an optimal step size, approximately $\mu = 1$, so that the M-NSAF algorithm obtains the fastest convergence. These results are in agreement with the statements in Remarks 1 and 2.
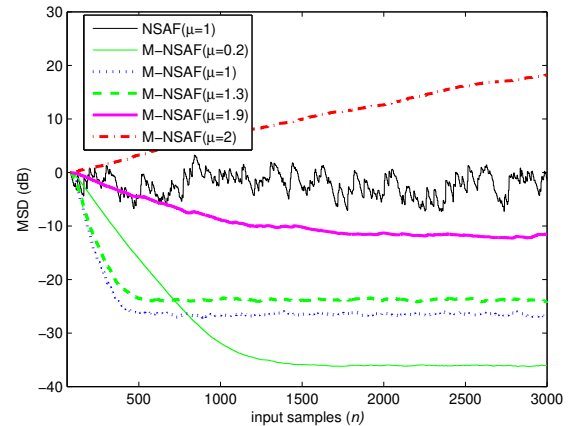

Fig. 3. MSD curves of the NSAF and M-NSAF algorithms in the presence of impulsive noise. [SNR=30 dB, $N = 4$]. Parameters of the M-NSAF algorithm are set to $N_w = 20$ and $\theta_\tau = 1$.

Figs. 4 and 5 show the transient MSD curves of the M-NSAF algorithm versus number of subbands and step sizes in impulsive noise, respectively. As can be seen, the theoretical results calculated by (19) match well with the simulation results. In Fig. 4, increasing the number of subbands N can speed up the convergence of the M-NSAF algorithm, with only slight loss in the steady-state performance, due to the fact that the decimated subband input signals are closer to white. However, for input signals with specified eigenvalue spread, after N is larger than a certain value (e.g., $N = 4$ in this case), the convergence improvement will not be obvious. It is to see from Fig. 5 that large step size accelerates the convergence of the algorithm but increases the steady-state MSD; conversely,

TABLE I
NUMBER OF OF ARITHMETICAL OPERATIONS PER INPUT SAMPLE.

Algorithms	Additions	Multiplications	Divisions	Square-roots
NLMS	$2L+2$	$2L+3$	1	0
NSAF	$2L+2N+2(J-1)N$	$2L+2N+2JN+1$	1	0
IWF-SSAF	$2L+2N+2(J-1)N+1$	$2L+2N+2JN$	1	1
BDVSS	$2L+2N+2(J-1)N+7$	$2L+2N+2JN+2$	3	1
M-NSAF	$2L+2N+2(J-1)N+N_w\log_2 N_w+1$	$2L+2N+2JN+5$	1	1
VSS-M-NSAF	$3L+2N+2(J-1)N+N_w\log_2 N_w+6$	$4L+2N+2JN+12$	2	1
M-PNSAF	$3L+2L/N+2(J-1)N+N_w\log_2 N_w+1$	$4L+(L+1)/N+2JN+5$	$1+1/N$	1
VSS-M-PNSAF	$4L+2L/N+2(J-1)N+N_w\log_2 N_w+6$	$6L+(L+1)/N+2JN+12$	$2+1/N$	1

small step size makes the algorithm keeping good steady-state performance but slows the convergence rate. It follows that the step size should be properly chosen when using the M-NSAF algorithm.
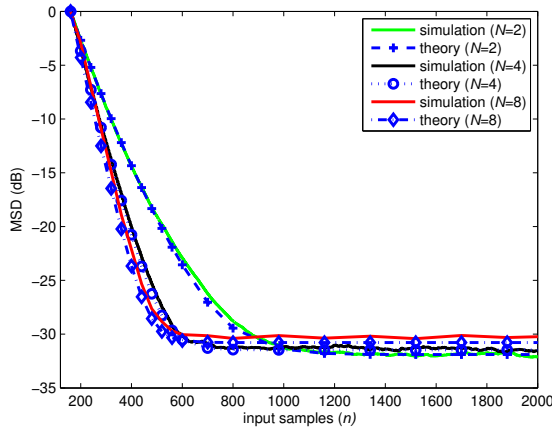
Fig. 4. MSD curves of the M-NSAF algorithm with $N = 2, 4,$ and 8 subbands. [SNR=30 dB, $\mu = 0.5$].

In Fig. 6, the theoretical steady-state MSDs obtained from (23) are compared with the simulated MSD, where the step size μ is from 0.1 to 1. The simulation results are obtained by averaging the last 300 MSDs in the steady-state. It is seen in the figure that the theoretical and simulation MSDs have a good match in the steady-state.

In Fig. 7, we examine the performance of different M-estimate functions (i.e., Huber, MH, and HTPR) for the M-NSAF algorithm. It is observed that the M-NSAF algorithm using the MH has very close performance to that using the HTPR, and it outperforms that using the Huber. However, the MH function is simpler than the HTPR function.

B. Comparison of Algorithms

In this subsection, the unknown vector $\boldsymbol{w}^o$ is a sparse acoustic echo channel with L=512 taps, shown in Fig. 8. Tracking performance is also important for adaptive algorithms, thus an abrupt change of the echo channel is introduced at the input sample index $n = 80001$ by shifting the impulse response to the right by 12 taps [6].

Scenario 1: System Identification
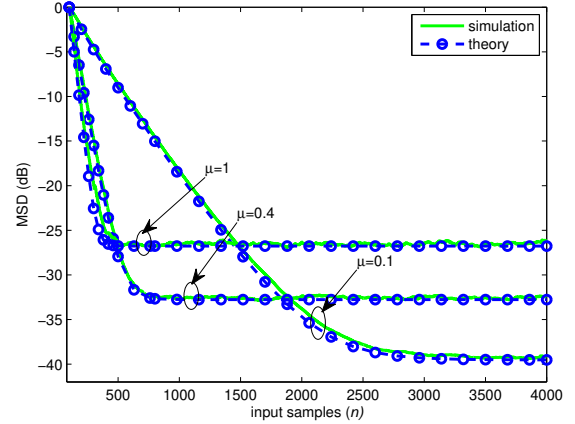
(a)
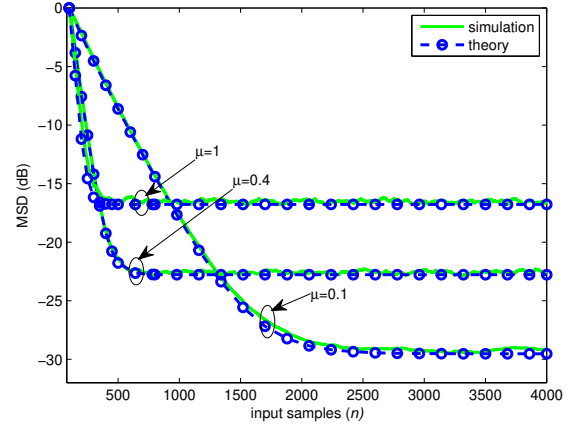
(b)

Fig. 5. MSD curves of the M-NSAF algorithm versus step sizes. (a) SNR=30 dB, (b) SNR=20 dB. [$N = 4$].

The additive noise with impulsive behavior, v_n, is generated from the (symmetric) α-stable process, i.e., the α-stable noise. Its characteristic function is given by [17], [62]

$$\phi(t) = \exp(-\gamma|t|^\alpha), \tag{51}$$

where $\alpha \in (0, 2]$ is the characteristic exponent which controls the impulsiveness of the noise (for lower values of α, the noise has more impulsive behavior), and $\gamma > 0$ represents the dispersion degree of the noise. The performance measure is the normalized MSD, i.e., $\mathrm{NMSD}(n) = 10\log_{10}(\mathrm{MSD}(n)/\|\boldsymbol{w}^o\|_2^2)$. The α-stable distribution includes two special cases: the Gaus-
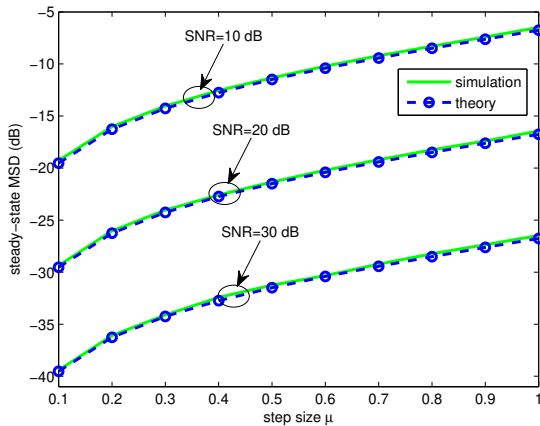
Fig. 6. Steady-state MSDs of the M-NSAF algorithm versus step sizes. [$N = 4$].
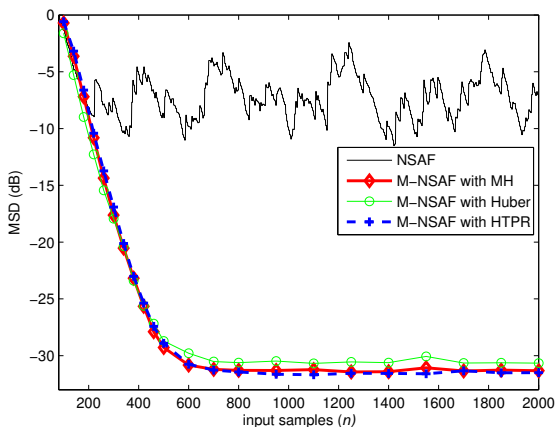

Fig. 7. Performance of the M-NSAF algorithm on different M-estimate functions. [$N = 4$, $\mu = 0.5$].

sian distribution and the Cauchy distribution for $\alpha = 2$ and $\alpha = 1$ respectively. This scenario sets $\alpha = 1.2$ and $\gamma = 1/30$.

Fig. 9 compares the performance of the NSAF and BDVSS algorithms with that of the proposed algorithms. It is seen that the NSAF among these algorithms is sensitive to α-stable noise. By using the proposed VSS, the VSS-M-NSAF algorithm obtains fast convergence of the M-NSAF with a large step size and low steady-state NMSD of that with a small step size. The BDVSS algorithm converges slowly relative to the VSS-M-NSAF algorithm. By decreasing the lower bound μ_L of the step size, the BDVSS algorithm can show have lower steady-state NMSD than the VSS-M-NSAF algorithm, but the tracking capability of the BDVSS algorithm from Fig. 9(b) is poor. Benefited from the proportionate rule, the VSS-M-PNSAF algorithm has faster convergence for reaching the same NMSD (e.g., -25 dB) than the VSS-M-NSAF algorithm when identifying a sparse system.

Scenario 2: AEC

In AEC, all SAF algorithms use the delayless structure shown in Fig. 1. All curves are obtained from the single trial. The far-end input signal $u(n)$ is a speech. Fig. 10 plots the
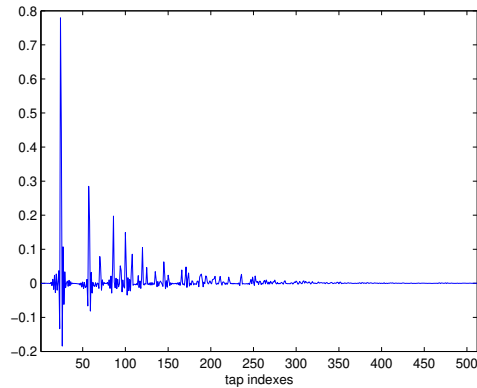

Fig. 8. Sparse acoustic echo channel.

NMSD curves of the algorithms (with $N = 4$ and $N = 8$ subbands respectively) for an AEC application in the α-stable noise with $\alpha = 1.5$. The results from Fig. 10 are similar to those in Fig. 9, which illustrates that the VSS-M-PNSAF among these algorithms is the best for AEC with sparse case.

In Fig. 11, the NMSD performances of the algorithms are also examined in an AEC with double-talk scenario, where the near-end speech signals happen in the periods with input sample indexes $[4, 5] \times 10^4$ and $[12, 13] \times 10^4$. In such a scenario, the near-end speech can be considered as an impulsive component $\nu_{im}(n)$ with clusters of impulses rather than random impulses, all the NSAF-type algorithms are equipped with a double-talk detector (DTD). Existing literature has reported numerous DTD methods [63]–[67], which have the common principle that stops updating the filter weights once the double-talk is detected. Among them, the Geigel DTD [63] is simple and has been widely used, namely, the double-talk is declared if

$$d_i(k) \geq T_c \max(|u_i(kN)|, |u_i(kN-1)|, ... \\ |u_i(kN-L+1)|), \tag{52}$$

where T_c is the detector threshold, and its typical value is 0.5 for the hybrid attenuation of 6 dB. In addition, there is a hangover time including T_{hold} samples that keeps prohibiting the weights adaptation after the double-talk is declared. Here, for the M-estimate based algorithms, we only employ the DTD to amend the vector $\boldsymbol{a}_{e,i}$ in (6) rather than stopping the weights adaptation, i.e., at each iteration k:
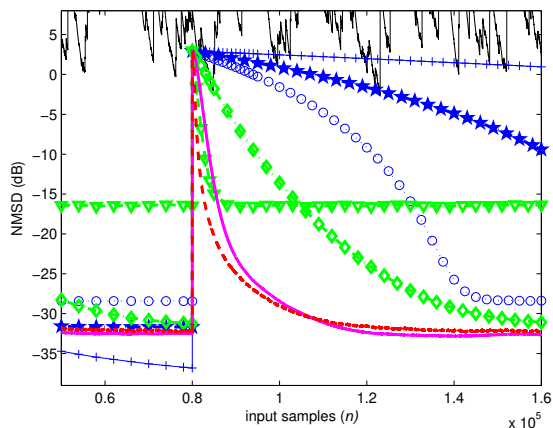
if the double-talk is detected and during T_{hold}

$\quad r_i(k) = 0;$

else

$$\qquad\qquad r_i(k) = e_{i,D}^2(k); \tag{53}$$

end

$$\boldsymbol{a}_{e,i} = [r_i(k), r_i(k-1), ..., r_i(k-N_w+1)].$$

The SNR for the background Gaussian noise is 30 dB. Moreover, the echo return loss enhancement (ERLE) is also a performance measure which is defined as [61], [68]

$$\text{ERLE}(n) = 10\log_{10}(\text{avg}\{d^2(n)\}/\text{avg}\{e^2(n)\}), \tag{54}$$
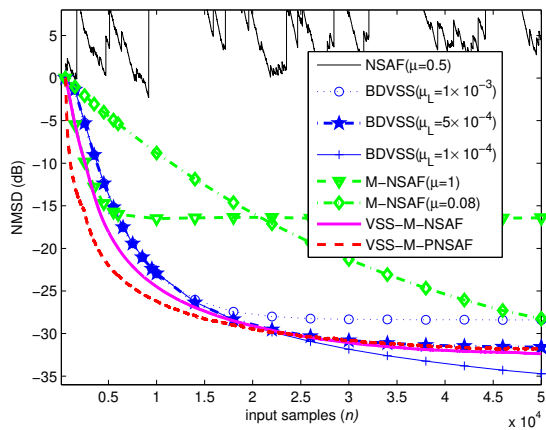
Fig. 9. NMSD curves of various SAF algorithms in α-stable noise. (a) before a sudden change of $\boldsymbol{w}^o$, (b) after a sudden change of $\boldsymbol{w}^o$. [$N = 4$]. Parameters of the algorithms are set to: $\mu_i(0) = \mu_U = 0.1$, $\delta = 0$, $\kappa = 1$ for the BDVSS; using the previous values for the M-estimate parameters; $\theta_\chi = 5$, $\epsilon_1 = 10^{-6}$ for the VSS parameters; $\zeta = 0$, $\epsilon_2 = 10^{-4}$ for the proportionate parameters.

where avg$(\cdot)$ is a smooth processing as the form in (44) with $\chi = 0.999$. The ERLE results are shown in Fig. 12. As can be seen from Figs. 11 and 12, by incorporating the Geigel DTD, the NSAF algorithm can also work in double-talk but with a slow convergence, since its adaptation is frozen when the double-talk is detected. In comparison, the M-NSAF algorithm converges faster, due to the fact that the weights adaptation is not stopped in the period of double-talk. Although the BDVSS algorithm does not require the double-talk detector, but its steady-state and tracking performance are worse than the VSS-M-NSAF algorithm. For a sparse echo channel, the the VSS-M-PNSAF algorithm outperforms the VSS-M-NSAF algorithm in both NMSD and ERLR performance.

Under the above double-talk scenario, we also investigate the performance of the VSS-M-NSAF algorithm (without DTD, with Geigel DTD, and with the signal envelope (SE) based DTD [65]), and compare with that of the published fullband VSS NLMS algorithms including the new nonpara-
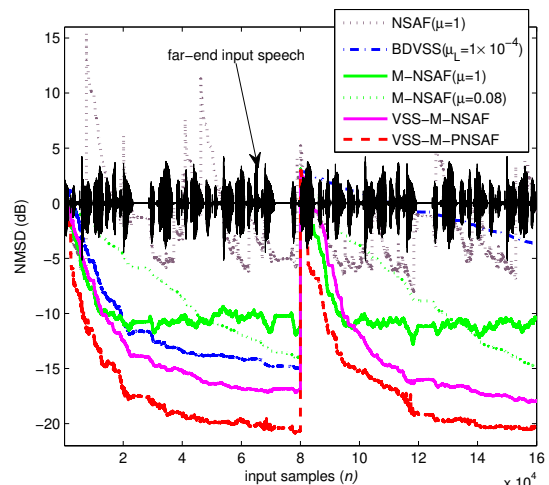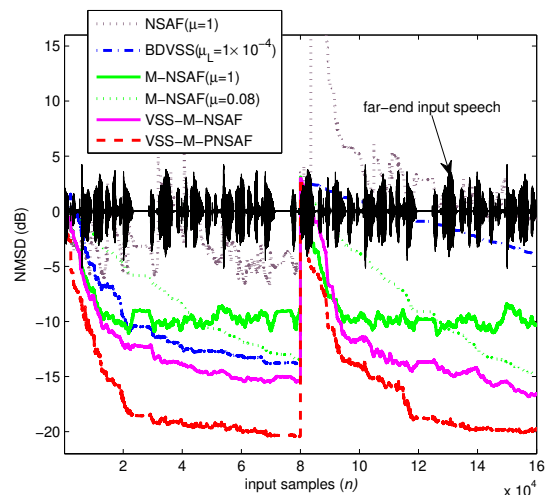
Fig. 10. NMSD curves of various SAF algorithms for speech input. (a) $N = 4$, (b) $N = 8$. Using the same values as Fig. 9 for parameters of the algorithms, except that $\delta = 0.1$, $\kappa = 2$ for the BDVSS. In addition, to avoid the numerical divergence of the algorithms caused by the mute period of the far-end speech signal, we add an addend $\delta = (20/N)\sigma_u^2$ in the denominator of the weights update for the NSAF, M-NSAF and VSS-M-NSAF algorithms, like the form (3), and $\delta = (20/L)\sigma_u^2$ [50] for the VSS-M-PNSAF algorithm.

metric VSS NLMS (NEW-NPVSS-NLMS) [69] and the joint-optimized NLMS (JO-NLMS) [70]. The VSS-M-NLMS algorithm[4] is the fullband version of the VSS-M-NSAF algorithm when $N = 1$. The NMSD results of these algorithms are drawn in Fig. 13. The NEW-NPVSS-NLMS and JO-NLMS algorithms exploit the same approach to estimate the power of near-end signal plus background noise, so the DTD is unnecessary for them. The notation "JO-NLMS with Geigel DTD" is to run the standard JO-NLMS algorithm with the known background noise power. As expected, the SAF structure based algorithm acquires a vast improvement of convergence over the NLMS-type when working the AEC application, thanks to the

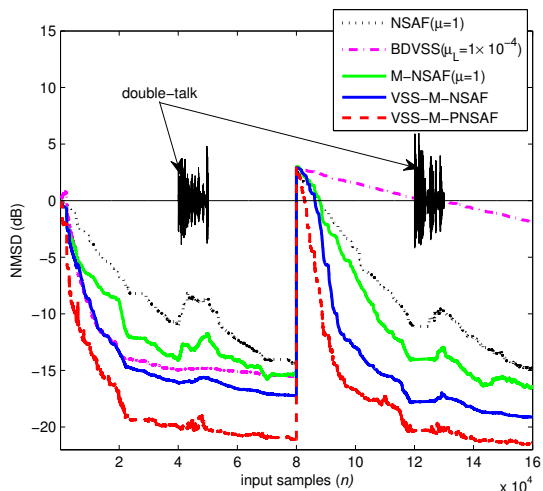[4]It is also a novel VSS modification of the normalized least mean M-etimate algorithm in [46].

Fig. 11. NMSD curves of various SAF algorithms in double-talk. [$N = 4$]. Using the same values as Fig. 10 for parameters of the algorithms, except $\theta_\tau = 6$ for the M-estimate. Also, we set the Geigel DTD parameter $T_{hold} = 50$ used in the NSAF and M-NSAF algorithms, and $T_{hold} = 3$ in the VSS-M-NSAF and VSS-M-PNSAF algorithms. Herein both the NSAF and M-NSAF algorithms use the large step size $\mu = 1$ rather than the small or diminishing step size, thus they require a relatively large hangover time T_{hold} to achieve good immunity during the double-talk. That is to say, when the algorithm uses the small step size, the corresponding T_{hold} can also be decreased.
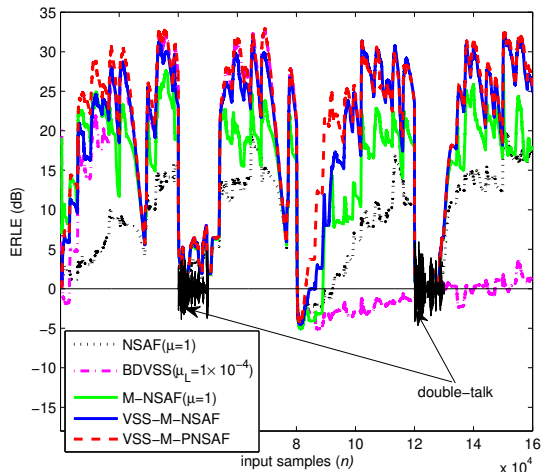

Fig. 12. ERLE curves of various SAF algorithms in double-talk.

fact that the input speech is highly correlated. The VSS-M-NSAF algorithm is not robust enough for double-talk, since the near-end speech signal bursts continuously unlike the α-stable noise with randomness. By equipping with the DTD method, the VSS-M-NSAF algorithm can work well in double-talk. Moreover, the algorithm performance with the SE DTD is better than that with the Geigel DTD.

VI. CONCLUSION

In this work, we have made a detailed analysis on the mean and mean-square behaviors of the M-NSAF algorithm by using the probability event to simplify the MH-based nonlinearity operation in the weights update. The theoretical expressions
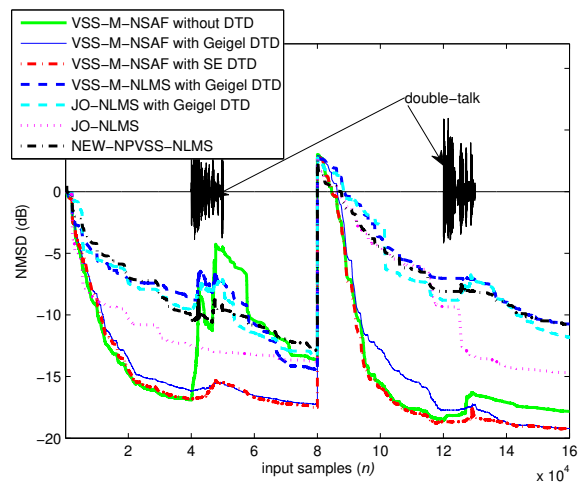

Fig. 13. Performance comparison of the VSS-M-NSAF and the fullband VSS NLMS algorithms. Parameters setting of the algorithms is as follows: using the same values as Fig. 11 for the VSS-M-NSAF; $\alpha = 0.99$, $\gamma = 0.05$ and $\beta = 0.25$ for the SE DTD; $\lambda = 0.9996$ and $\epsilon = 0.25$ for the NEW-NPVSS-NLMS; $\lambda = 0.999$ and $\epsilon = 10$ for the JO-NLMS; $T_{hold} = 50$ for the JO-NLMS with Geigel DTD.

have been given to characterize the stability condition, transient and steady-state results of the M-NSAF algorithm in the CG-based impulsive noise and verified using simulations. We also developed a VSS scheme to further improve the M-NSAF performance in terms of convergence rate and steady-state estimation error. Moreover, this VSS scheme is also suitable for improving the M-PNSAF performance in sparse systems. Simulation results in both the system identification in the α-stable noise and the AEC with double-talk scenarios have demonstrated that the proposed algorithms outperform some previously reported algorithms.

APPENDIX A

Based on the simulation setting in Section V. A, Fig. 14(a) and (b) check the M-NSAF performance with respect to different values of c_σ and κ, respectively. Sure enough, using the typical values $c_\sigma = 1.483(1 + 5/(N_w - 1))$ [51, p.44] and $\kappa = 2.576$ [52], the M-NSAF algorithm has good balance in terms of convergence rate, steady-state MSD, and stability.
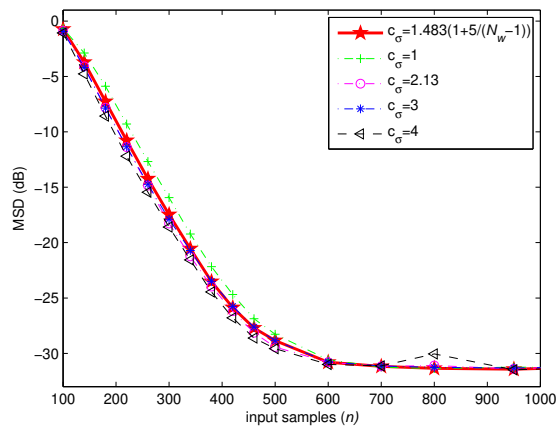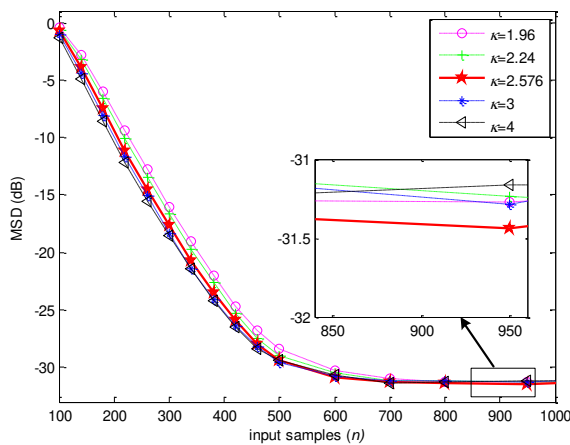
APPENDIX B
VERIFICATION OF (21)

Setting the same parameters as Fig. 5(b) for the M-NSAF algorithm, Fig. 15 depicts the probability of every subband participating in the weights update. The theoretical values of $P_{e,i}(k), i = 1, 2, 3, 4$ are calculated by (35). As one can see, at every iteration k, values of $P_{e,i}(k)$ for different subbands are close to each other. As such, we can make the replacement of every $P_{e,i}(k)$ with $\frac{1}{N}\sum_{i=0}^{N-1} P_{e,i}(k)$ to obtain (21).

REFERENCES

[1] A. H. Sayed, *Fundamentals of adaptive filtering*. John Wiley & Sons, 2003.

(a)

(b)

Fig. 14. MSD curves of the M-NSAF algorithm versus parameters: (a) c_σ and (b) κ.

Fig. 15. Values of $P_{e,i}(k)$ with iterations. [$\mu = 1$, $N = 4$ and SNR=30 dB].

[2] K.-A. Lee, W.-S. Gan, and S. M. Kuo, *Subband adaptive filtering: theory and implementation*. John Wiley & Sons, 2009.

[3] Y. V. Zakharov and J. Li, "Sliding-window homotopy adaptive filter for estimation of sparse UWA channels," in *2016 IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)*. IEEE, 2016, pp. 1–4.

[4] M. Yukawa, R. C. De Lamare, and R. Sampaio-Neto, "Efficient acoustic echo cancellation with reduced-rank adaptive filtering based on selective decimation and adaptive interpolation," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 16, no. 4, pp. 696–710, 2008.

[5] K. Ozeki and T. Umeda, "An adaptive filtering algorithm using an orthogonal projection to an affine subspace and its properties," *Electronics and Communications in Japan (Part I: Communications)*, vol. 67, no. 5, pp. 19–27, 1984.

[6] C. Paleologu, S. Ciochina, and J. Benesty, "An efficient proportionate affine projection algorithm for echo cancellation," *IEEE Signal Processing Letters*, vol. 17, no. 2, pp. 165–168, 2010.

[7] C. H. Lee and P. Park, "Scheduled-step-size affine projection algorithm," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 59, no. 9, pp. 2034–2043, 2012.

[8] F. Yang and J. Yang, "A comparative survey of fast affine projection algorithms," *Digital Signal Processing*, vol. 83, pp. 297–322, 2018.

[9] S. S. Pradham and V. Reddy, "A new approach to subband adaptive filtering," *IEEE Transactions on Signal Processing*, vol. 47, no. 3, pp. 655–664, 1999.

[10] K.-A. Lee and W.-S. Gan, "Improving convergence of the NLMS algorithm using constrained subband updates," *IEEE Signal Processing Letters*, vol. 11, no. 9, pp. 736–739, 2004.

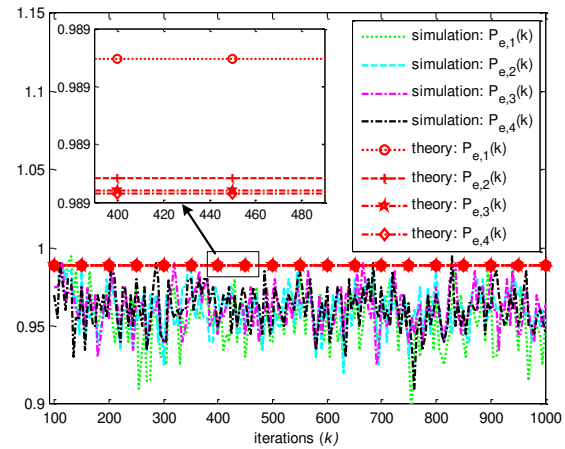[11] K.-A. Lee and W.-S. Gan, "On delayless architecture for the normalized subband adaptive filter," in *2007 IEEE International Conference on Multimedia and Expo*. IEEE, 2007, pp. 1595–1598.

[12] J. Ni and F. Li, "A variable step-size matrix normalized subband adaptive filter," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 6, pp. 1290–1299, 2010.

[13] Y. Yu, H. Zhao, and B. Chen, "A new normalized subband adaptive filter algorithm with individual variable step sizes," *Circuits, Systems, and Signal Processing*, vol. 35, no. 4, pp. 1407–1418, 2016.

[14] J. J. Jeong, K. Koo, G. T. Choi, and S. W. Kim, "A variable step size for normalized subband adaptive filters," *IEEE Signal Processing Letters*, vol. 19, no. 12, pp. 906–909, 2012.

[15] J. Ni and F. Li, "Adaptive combination of subband adaptive filters for acoustic echo cancellation," *IEEE Transactions on Consumer Electronics*, vol. 56, no. 3, pp. 1549–1555, 2010.

[16] Z. Shen, Y. Yu, and T. Huang, "Normalized subband adaptive filter algorithm with combined step size for acoustic echo cancellation," *Circuits, Systems, and Signal Processing*, vol. 36, no. 7, pp. 2991–3003, 2017.

[17] C. L. Nikias and M. Shao, *Signal processing with alpha-stable distributions and applications*. Wiley-Interscience, 1995.

[18] M. Zimmermann and K. Dostert, "Analysis and modeling of impulsive noise in broad-band powerline communications," *IEEE Transactions on Electromagnetic Compatibility*, vol. 44, no. 1, pp. 249–258, 2002.

[19] P. G. Georgiou, P. Tsakalides, and C. Kyriakakis, "Alpha-stable modeling of noise and robust time-delay estimation in the presence of impulsive noise," *IEEE Transactions on Multimedia*, vol. 1, no. 3, pp. 291–301, 1999.

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

[21] L. Dang, B. Chen, S. Wang, Y. Gu, and J. C. Prncipe, "Kernel kalman filtering with conditional embedding and maximum correntropy criterion," *IEEE Transactions on Circuits and Systems I: Regular Papers*, pp. 1–13, 2019.

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

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

[24] B. Chen, L. Xing, B. Xu, H. Zhao, N. Zheng, and J. C. Príncipe, "Kernel risk-sensitive loss: definition, properties and application to robust adaptive filtering," *IEEE Transactions on Signal Processing*, vol. 65, no. 11, pp. 2888–2901, 2017.

[25] 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 Processing Letters*, vol. 21, no. 7, pp. 880–884, 2014.

[26] R. Pogula, T. K. Kumar, and F. Albu, "Robust sparse normalized LMAT algorithms for adaptive system identification under impulsive noise environments," *Circuits, Systems, and Signal Processing*, pp. 1–32, 2019.

[27] K. Xiong, S. Wang, and B. Chen, "Robust normalized least mean absolute third algorithms," *IEEE Access*, vol. 7, pp. 10 318–10 330, 2019.

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

[29] Y. Yu and H. Zhao, "Novel sign subband adaptive filter algorithms with individual weighting factors," *Signal Processing*, vol. 122, pp. 14–23, 2016.

[30] J. Kim, J.-H. Chang, and S. Nam, "Sign subband adaptive filter with l_1-norm minimisation-based variable step-size," *Electronics Letters*, vol. 49, no. 21, pp. 1325–1326, 2013.

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

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

[33] J.-H. Kim, J. Kim, J. H. Jeon, and S. W. Nam, "Delayless individual-weighting-factors sign subband adaptive filter with band-dependent variable step-sizes," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 25, no. 7, pp. 1526–1534, 2017.

[34] J. Radecki, Z. Zilic, and K. Radecka, "Echo cancellation in ip networks," in *The 2002 45th Midwest Symposium on Circuits and Systems, 2002. MWSCAS-2002.*, vol. 2. IEEE, 2002, pp. II–II.

[35] E. Hänsler and G. Schmidt, *Topics in acoustic echo and noise control*. Springer-Verlag, 2006.

[36] D. L. Duttweiler, "Proportionate normalized least-mean-squares adaptation in echo cancelers," *IEEE Transactions on Speech and Audio Processing*, vol. 8, no. 5, pp. 508–518, 2000.

[37] J. Benesty and S. L. Gay, "An improved PNLMS algorithm," in *IEEE International conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2002, pp. I–1881–II–1884.

[38] P. Loganathan, A. W. Khong, and P. A. Naylor, "A class of sparseness-controlled algorithms for echo cancellation," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 17, no. 8, pp. 1591–1601, 2009.

[39] Y. Gu, J. Jin, and S. Mei, "l_0 norm constraint LMS algorithm for sparse system identification," *IEEE Signal Processing Letters*, vol. 16, no. 9, pp. 774–777, 2009.

[40] Y. Li, Y. Wang, and T. Jiang, "Norm-adaption penalized least mean square/fourth algorithm for sparse channel estimation," *Signal Processing*, vol. 128, pp. 243–251, 2016.

[41] W. Wang, H. Zhao, and B. Chen, "Bias compensated zero attracting normalized least mean square adaptive filter and its performance analysis," *Signal Processing*, vol. 143, pp. 94–105, 2018.

[42] R. L. Das and M. Chakraborty, "Improving the performance of the PNLMS algorithm using l_1 norm regularization," *IEEE/ACM Transactions on Audio, Speech and Language Processing*, vol. 24, no. 7, pp. 1280–1290, 2016.

[43] Y. Wang, Y. Li, J. C. M. Bermudez, and X. Han, "An adaptive combination constrained proportionate normalized maximum correntropy criterion algorithm for sparse channel estimations," *EURASIP Journal on Advances in Signal Processing*, vol. 2018, no. 1, p. 58, 2018.

[44] J. Ni, X. Chen, and J. Yang, "Two variants of the sign subband adaptive filter with improved convergence rate," *Signal Processing*, vol. 96, pp. 325–331, 2014.

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

[46] S. Chan and Y. Zhou, "On the performance analysis of the least mean M-estimate and normalized least mean M-estimate algorithms with gaussian inputs and additive gaussian and contaminated gaussian noises," *Journal of Signal Processing Systems*, vol. 60, no. 1, pp. 81–103, 2010.

[47] Z. Zheng and H. Zhao, "Affine projection M-estimate subband adaptive filters for robust adaptive filtering in impulsive noise," *Signal Processing*, vol. 120, pp. 64–70, 2016.

[48] F. Yang, M. Wu, P. Ji, and J. Yang, "An improved multiband-structured subband adaptive filter algorithm," *IEEE Signal Processing Letters*, vol. 19, no. 10, pp. 647–650, 2012.

[49] C. Paleologu, J. Benesty, and S. Ciochina, "A variable step-size affine projection algorithm designed for acoustic echo cancellation," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 16, no. 8, pp. 1466–1478, 2008.

[50] J. Benesty, C. Paleologu, and S. Ciochina, "On regularization in adaptive filtering," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 6, pp. 1734–1742, 2010.

[51] P. J. Rousseeuw and A. M. Leroy, *Robust Regression and Outlier Detection*. New York: Wiley, 1987.

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

[53] Y. Yu, H. Zhao, R. C. de Lamare, Y. Zakharov, and L. Lu, "Robust distributed diffusion recursive least squares algorithms with side information for adaptive networks," *IEEE Transactions on Signal Processing*, vol. 67, no. 6, pp. 1566–1581, 2019.

[54] N. J. Bershad, "On error saturation nonlinearities for LMS adaptation in impulsive noise," *IEEE Transactions on Signal Processing*, vol. 56, no. 9, pp. 4526–4530, 2008.

[55] J. J. Jeong, S. H. Kim, G. Koo, and S. W. Kim, "Mean-square deviation analysis of multiband-structured subband adaptive filter algorithm," *IEEE Transactions on Signal Processing*, vol. 64, no. 4, pp. 985–994, 2016.

[56] S. Zhang and W. X. Zheng, "Mean-square analysis of multi-sampled multiband-structured subband filtering algorithm," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 66, no. 3, pp. 1051–1062, 2019.

[57] W. Yin and A. S. Mehr, "Stochastic analysis of the normalized subband adaptive filter algorithm," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 58, no. 5, pp. 1020–1033, 2011.

[58] A. Graham, *Kronecker products and matrix calculus with applications*. Courier Dover Publications, 2018.

[59] T. Y. Al-Naffouri and A. H. Sayed, "Transient analysis of data-normalized adaptive filters," *IEEE Transactions on Signal Processing*, vol. 51, no. 3, pp. 639–652, 2003.

[60] M. S. E. Abadi and S. Kadkhodazadeh, "A family of proportionate normalized subband adaptive filter algorithms," *Journal of the Franklin Institute*, vol. 348, no. 2, pp. 212–238, 2011.

[61] Y. Yu and H. Zhao, "A band-independent variable step size proportionate normalized subband adaptive filter algorithm," *AEU-International Journal of Electronics and Communications*, vol. 70, no. 9, pp. 1179–1186, 2016.

[62] Y. Yu, H. Zhao, W. Wang, and L. Lu, "Robust diffusion Huber-based normalized least mean square algorithm with adjustable thresholds," *Circuits, Systems, and Signal Processing*, pp. 1–29, 2019. doi: 10.1007/s00034-019-01244-5.

[63] D. Duttweiler, "A twelve-channel digital echo canceler," *IEEE Transactions on Communications*, vol. 26, no. 5, pp. 647–653, 1978.

[64] H. K. Jung, N. S. Kim, and T. Kim, "A new double-talk detector using echo path estimation," *Speech communication*, vol. 45, no. 1, pp. 41–48, 2005.

[65] G. Szwoch, A. Czyżewski, and M. Kulesza, "A low complexity double-talk detector based on the signal envelope," *Signal Processing*, vol. 88, no. 11, pp. 2856–2862, 2008.

[66] M. Hamidia and A. Amrouche, "A new robust double-talk detector based on the stockwell transform for acoustic echo cancellation," *Digital Signal Processing*, vol. 60, pp. 99–112, 2017.

[67] M. A. Iqbal, J. W. Stokes, and S. L. Grant, "Normalized double-talk detection based on microphone and AEC error cross-correlation," in *2007 IEEE International Conference on Multimedia and Expo*. IEEE, 2007, pp. 360–363.

[68] E. Hänsler and G. Schmidt, *Acoustic echo and noise control: a practical approach*. John Wiley & Sons, 2005, vol. 40.

[69] M. A. Iqbal and S. L. Grant, "Novel variable step size NLMS algorithms for echo cancellation," in *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2008, pp. 241–244.

[70] C. Paleologu, S. Ciochină, J. Benesty, and S. L. Grant, "An overview on optimized NLMS algorithms for acoustic echo cancellation," *EURASIP Journal on Advances in Signal Processing*, vol. 2015, no. 1, p. 97, 2015.

Yi Yu (S'17-M'18) received the B.E. degree at XiHua University, Chengdu, China, in 2011, and the M.S. degree and Ph.D degree at Southwest Jiaotong University, Chengdu, China, in 2014 and 2018, respectively. From Dec. 2016 to Dec. 2017, he was a visiting Ph.D student with the Department of Electronic Engineering, University of York, United Kingdom. He is currently an Associate Professor with the School of Information Engineering, Southwest University of Science and Technology, Mianyang, China. His research interests include adaptive signal processing, distributed estimation, and compressive sensing.

Jianghui Li received the B.S. degree in communications engineering from Huazhong University of Science and Technology, Wuhan, China in 2011, the M.Sc. degree in communications engineering, and the Ph.D. degree with receiving the K. M. Stott Prize for excellent research in electronics engineering from the University of York, U.K., in 2013 and 2017, respectively. He has been the first researcher receiving the IEEE OES scholarship in U.K. From 2011 to 2012, he served as a Research Assistant with the Chinese Academy of Sciences, Beijing, China. Since 2017, he has been a Research Fellow with the University of Southampton, U.K. His current research interests include adaptive signal processing, wireless communications, underwater acoustics, and ocean engineering.

Hongsen He received the B.Eng. degree in automation from Southwest University of Science and Technology (SWUST), Mianyang, China, in June 2000, and the Ph.D. degree in signal and information processing from Nanjing University (NJU), Nanjing, China, in June 2013. He joined the School of Information Engineering, SWUST, as a member of faculty in July 2000. From September 2007 to May 2013, he was a Research Assistant at the Institute of Acoustics, NJU, and engaged in research on adaptive signal processing and microphone array signal processing. He is currently an Associate Professor at SWUST. Dr. He is a Member of IEEE and Acoustical Society of America, and serves as a reviewer of several international journals and conferences. His current research interests focus on signal processing and acoustic signal processing, including adaptive filtering, robust signal processing, sparse signal processing, multimicrophone signal processing, acoustic source localization, system identification, dereverberation, adaptive noise cancellation, and machine learning. Dr. He was selected as one of the Distinguished Youth Academic and Technical Pacemakers of Sichuan Province in 2014 (Incubation).

Youwen Zhang received his B.S. degree in 1997, from Wuhan University of Economics and Business, Wuhan, Hubei China, in Computer Application Technology. The M.Sc. degree in 2004 in Signal and Information Processing, The Ph.D. degree in 2005 in Underwater Acoustical Engineering, both from the Harbin Engineering University, Harbin, Heilongjiang China. He has been an assistant professor with the College of Underwater Acoustic Engineering at Harbin Engineering University, Harbin, during 2005-2010, as an associate professor since 2010. His research interests lie in the areas of underwater acoustic communications and networking, and the array signal processing in underwater acoustic environments.

Badong Chen (M'10 - SM'13) received the B.S. and M.S. degrees in control theory and engineering from Chongqing University, in 1997 and 2003, respectively, and the Ph.D. degree in computer science and technology from Tsinghua University in 2008. He was a Postdoctoral Researcher with Tsinghua University from 2008 to 2010, and a Postdoctoral Associate at the University of Florida Computational NeuroEngineering Laboratory (CNEL) during the period October, 2010 to September, 2012. During July to August 2015, he visited the Nanyang Technological University (NTU) as a visiting research scientist. He also served as a senior research fellow with The Hong Kong Polytechnic University from August to November in 2017. Currently he is a professor at the Institute of Artificial Intelligence and Robotics (IAIR), Xian Jiaotong University. His research interests are in signal processing, machine learning, artificial intelligence, neural engineering and robotics. He has published 2 books, 4 chapters, and over 200 papers in various journals and conference proceedings. Dr. Chen is an IEEE Senior Member, a Technical Committee Member of IEEE SPS Machine Learning for Signal Processing (MLSP) and IEEE CIS Cognitive and Developmental Systems (CDS), and an associate editor of IEEE Transactions on Cognitive and Developmental Systems, IEEE Transactions on Neural Networks and Learning Systems and Journal of The Franklin Institute, and has been on the editorial board of Entropy.

Lu Lu was born in Chengdu, China, in 1990. He received the Ph.D. degree in the field of signal and information processing at the School of Electrical Engineering, Southwest Jiaotong University, Chengdu, China, in 2018. From 2017 to 2018, he was a visiting Ph.D. student with the Electrical and Computer Engineering at McGill University, Montreal, QC, Canada. He is currently a Postdoctoral Fellow with the College of Electronics and Information Engineering, Sichuan University, Chengdu, China. His research interests include adaptive filtering, kernel methods and distributed estimation.