A Normalized Least Mean Squares Algorithm With a Step-Size Scaler Against Impulsive Measurement Noise

Insun Song, PooGyeon Park, *Member, IEEE*, and Robert W. Newcomb, *Life Fellow, IEEE*

***Abstract*—This brief introduces the concept of a step-size scaler by investigating and modifying the tanh cost function for adaptive filtering with impulsive measurement noise. The step-size scaler instantly scales down the step size of gradient-based adaptive algorithms whenever impulsive measurement noise appears, which eliminates a possibility of updating weight vector estimates based on wrong information due to impulsive noise. The most attractive feature of the step-size scaler is that this is easily applicable to various gradient-based adaptive algorithms. Several representative gradient-based adaptive algorithms are performed without or with the step-size scaler in impulsive-noise environments, which shows the improvement of robustness against impulsive noise.**

***Index Terms*—Adaptive filters, impulsive measurement noise, robust filtering, step-size scaler.**

I. Introduction

ADAPTIVE algorithms have played major roles in various signal processing applications, such as echo cancelation, channel equalization, and control in communication networks [1]–[4]. For this, the normalized least mean squares (NLMS) algorithm has been widely used in many applications due to its simplicity and robustness [1], [2]. However, in most real environments, adaptive filtering applications can be influenced by various outliers including impulsive measurement noise, and they cause the performance degradation of many adaptive filters. To overcome this drawback, many different schemes have been proposed [5]–[15].

When using gradient-based adaptive filtering, the next weight estimate is determined to reduce the cost function associated with the *a posteriori* output errors by using the gradient of the cost function with respect to the current weight vector, where the step size in (0, 1] is usually designed to enhance the rate of convergence. Therefore, the cost function and the step size play an extremely important role in gradient-based adaptive filtering. The most common idea to handle impulsive measurement noise has been considered to design the appropriate cost function, which directly determines the performance of the algorithms.

Various robust cost functions using the $\mathcal{L}_1$ norm have been developed for impulsive-noise environments [5]–[11]. A normalized least mean absolute deviation (NLMAD) algorithm [5], a dual sign algorithm [6], variable-step-size (VSS) sign algorithms [7], [8], and an affine projection sign algorithm [9] are based on the $\mathcal{L}_1$-norm minimization; thus, the algorithms use the sign of the output error to update the next weight estimate. A robust mixed-norm adaptive algorithm is based on a robust cost function using the convex combination of the $\mathcal{L}_1$ and $\mathcal{L}_2$ norms [10]. A robust VSS NLMS algorithm uses a robust cost function which switches between the $\mathcal{L}_1$ and $\mathcal{L}_2$ norms [11]. It is associated with how the algorithm constrains the energy of the filter update at each iteration. Various adaptive algorithms use other robust cost functions for robustness against impulsive measurement noise [12]–[14]. When the magnitude of the output error is larger than a threshold, the Huber mixed-norm M-estimate cost function uses the $\mathcal{L}_1$ norm minimization [12]; on the other hand, the Hampel three-part redescending M-estimate cost function sets the error signal as a constant value [13], [14].

Since all these gradient-based adaptive algorithms [5]–[14] are based on their own cost function robust over impulsive noise, they also robustly perform over impulsive noise. As mentioned earlier, there is another factor, besides the cost function, that determines the performance in the gradient-based adaptive algorithms: the step size. So far, researchers mainly focus on the cost function, rather than the step size, to handle impulsive noise. This brief will focus on the step size and its effect against impulsive noise.

This brief proposes a method to improve the robustness against impulsive measurement noise by scaling the step-size in gradient-based adaptive algorithms. The proposed method uses the concept of a step-size scaler, which instantly scales down the step size whenever impulsive measurement noise appears. This eliminates a possibility of updating weight estimates based on wrong information due to impulsive noise. The most attractive feature of the step-size scaler is that it is readily applied to various gradient-based adaptive algorithms whether the cost functions of the algorithms are adapted in impulsive-noise environments or not. Moreover, the step-size scaler improves the robustness against impulsive noise of high probability. The

Manuscript received July 23, 2012; revised September 23, 2012, November 18, 2012, and February 8, 2013; accepted March 24, 2013. Date of publication April 23, 2013; date of current version July 13, 2013. This work was supported by World Class University program funded by the Ministry of Education, Science and Technology through the National Research Foundation of Korea under Grant R31-10100. The work of P. Park was supported by the LG YONAM Foundation through the Professors overseas research program for sabbatical research leave at the University of Maryland. This brief was recommended by Associate Editor Y. Miyanaga.

I. Song and P. Park are with the Division of IT Convergence Engineering and the Department of Electrical and Computer Engineering, Pohang University of Science and Technology, Pohang 790-784, Korea (e-mail: weedsis@postech.ac.kr; ppg@postech.ac.kr).

R. W. Newcomb is with the Department of Electrical and Computer Engineering, University of Maryland, College Park, MD 20742 USA (e-mail: newcomb@umd.edu).

Digital Object Identifier 10.1109/TCSII.2013.2258266

concept of the step-size scaler is derived from a gradient-based adaptive algorithm based on a new cost function which is developed by investigating and modifying the $\tanh$ cost function. We apply the proposed method to NLMS, the VSS NLMS [18], the NLMAD [5], and the robust VSS NLMS [11] and compare the performance with colored input in impulsive-noise environments.

II. Step-Size Scaler

In a model for system identification, the desired signal d_i is represented as

$$d_i = \mathbf{u}_i^{\mathrm{T}} \mathbf{w} + v_i. \tag{1}$$

Here, $\mathbf{w} \in \Re^n$ is a coefficient vector of the unknown system, $\mathbf{u}_i \in \Re^n$ denotes an input vector such as $[u_i, \ldots, u_{i-n+1}]^{\mathrm{T}}$, and a scalar variable $v_i = b_i + \eta_i$, where a scalar variable u_i is colored with variance σ_u^2, b_i denotes the white Gaussian measurement noise with $N(0, \sigma_b^2)$, and η_i is the impulsive measurement noise. Let $\widehat{\mathbf{w}}_i$ be an estimate of $\mathbf{w}$ at the i_{th} iteration and e_i be an *a priori* measurement error, defined as

$$e_i = d_i - \mathbf{u}_i^{\mathrm{T}} \widehat{\mathbf{w}}_i. \tag{2}$$

Hampel introduced the concept of the $\tanh$ cost function for M-estimators in robust estimation [16]. Using the concept, Wang developed a robust cost function using the square value of the output error [17]. Similar to this function, we shall suggest a new cost function using the square value of the normalized instant output error with respect to the input vector as follows:

$$J(\widehat{\mathbf{w}}_i) = \frac{1}{\beta} \tanh\left(\frac{\beta}{2} \left(e_i/\|\mathbf{u}_i\|\right)^2\right) \tag{3}$$

$$= \frac{1}{\beta} \frac{1 - \exp\left(-\beta \left(e_i/\|\mathbf{u}_i\|\right)^2\right)}{1 + \exp\left(-\beta \left(e_i/\|\mathbf{u}_i\|\right)^2\right)} \tag{4}$$

where $\beta > 0$ determines the sharpness of the shape. When $(e_i/\|\mathbf{u}_i\|)$ is small, Taylor expansions provide the approximation as follows:

$$J(\widehat{\mathbf{w}}_i) \cong \left(e_i/\|\mathbf{u}_i\|\right)^2 \tag{5}$$

from which this function works like a $\mathcal{L}_2$ square mean and the associated algorithm performs like an NLMS algorithm. If there happens an impulsive noise, then $(e_i/\|\mathbf{u}_i\|)$ becomes very large, which sends the cost function into $1/\beta$, i.e., $J(\widehat{\mathbf{w}}_i) \cong 1/\beta$, and, furthermore, its derivative into zero, which plays a crucial role in achieving robustness over impulsive noise.

Let us develop the adaptive gradient algorithm based on this cost function

$$\widehat{\mathbf{w}}_{i+1} = \widehat{\mathbf{w}}_i - \mu_i \nabla_{\widehat{\mathbf{w}}_i} J(\widehat{\mathbf{w}}_i) \tag{6}$$

where μ_i is a step size at time i and $\nabla_{\widehat{\mathbf{w}}_i} J(\widehat{\mathbf{w}}_i)$ denotes the gradient of $J(\widehat{\mathbf{w}}_i)$ with respect to $\widehat{\mathbf{w}}_i$ as follows:

$$\nabla_{\widehat{\mathbf{w}}_i} J(\widehat{\mathbf{w}}_i) = -s\left(\beta, e_i/\|\mathbf{u}_i\|\right) \frac{\mathbf{u}_i}{\|\mathbf{u}_i\|^2} e_i \tag{7}$$

$$s\left(\beta, e_i/\|\mathbf{u}_i\|\right) \triangleq \frac{4 \exp\left(-\beta \left(e_i/\|\mathbf{u}_i\|\right)^2\right)}{\left(1 + \exp\left(-\beta \left(e_i/\|\mathbf{u}_i\|\right)^2\right)\right)^2}. \tag{8}$$

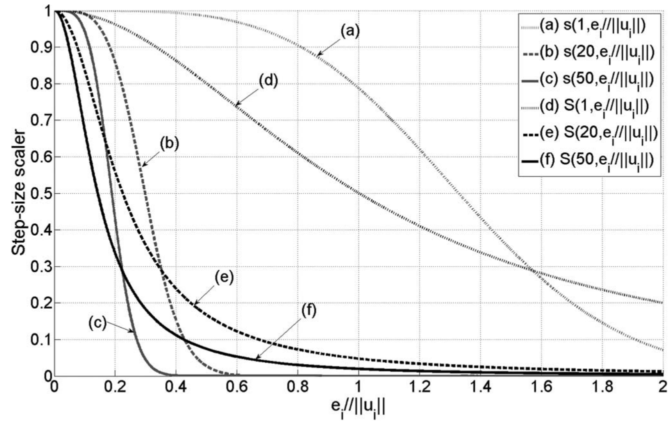

Fig. 1. Step-size scalers.

The resulting algorithm can be summarized as

$$\widehat{\mathbf{w}}_{i+1} = \widehat{\mathbf{w}}_i + \mu_i s\left(\beta, e_i/\|\mathbf{u}_i\|\right) \frac{\mathbf{u}_i}{\|\mathbf{u}_i\|^2} e_i \tag{9}$$

which has exactly the same form as the NLMS algorithm with the step size μ_i except the term $s(\beta, e_i/\|\mathbf{u}_i\|)$. When the magnitude of the normalized error is small, $s(\beta, e_i/\|\mathbf{u}_i\|) \cong 1$, and thus, the adaptive algorithm operates like the conventional NLMS algorithm with the step size μ_i. The larger magnitude of the normalized error leads to the smaller $s(\beta, e_i/\|\mathbf{u}_i\|)$, which means $s(\beta, e_i/\|\mathbf{u}_i\|)$ performs a role of scaling the step size μ_i to eliminate possible malfunctions due to large impulsive noise. Therefore, we shall henceforth call $s(\beta, e_i/\|\mathbf{u}_i\|)$ as *a step-size scaler*.

Remark 1: $s(\beta, e_i/\|\mathbf{u}_i\|)$ needs to use a nonlinear function $\tanh(\cdot)$ which is a major drawback. To reduce the computational complexity, we find a step-size scalers

$$S\left(\gamma, e_i/\|\mathbf{u}_i\|\right) \triangleq 1/\left(1 + \gamma \left|e_i/\|\mathbf{u}_i\|\right|^2\right)$$

which is derived form a new cost function

$$\overline{J}(\widehat{\mathbf{w}}_i) \triangleq \frac{1}{2\gamma} \ln\left(1 + \gamma \left|e_i/\|\mathbf{u}_i\|\right|^2\right).$$

$\overline{J}(\widehat{\mathbf{w}}_i)$ is even, nonnegative, and differentiable for all $e_i/\|\mathbf{u}_i\|$; thus, it can be a cost function. The adaptive gradient algorithm based on the cost function $\overline{J}(\widehat{\mathbf{w}}_i)$ can be summarized as

$$\widehat{\mathbf{w}}_{i+1} = \widehat{\mathbf{w}}_i + \mu_i S\left(\gamma, e_i/\|\mathbf{u}_i\|\right) \frac{\mathbf{u}_i}{\|\mathbf{u}_i\|^2} e_i.$$

This step-size scaler also improves the robustness against impulsive measurement noise by scaling the step-size in gradient-based adaptive algorithms. Note that various step-size scalers can be existed and they can similarly perform like the step-size scaler $s(\beta, e_i/\|\mathbf{u}_i\|)$.

Fig. 1 shows the step-size scalers $s(\beta, e_i/\|\mathbf{u}_i\|)$ and $S(\gamma, e_i/\|\mathbf{u}_i\|)$ for various β's and γ's. Whenever impulsive measurement noise appears, the magnitude of the normalized error becomes very large; then, the step-size scalers scale down the step size. This eliminates a possibility of updating weight estimates based on wrong information due to impulsive noise. However, if impulsive measurement noise does not appear, then

TABLE I
ADAPTIVE ALGORITHMS USING THE PROPOSED STEP-SIZE SCALER

Adaptive algorithms using $s\left(\beta, e_i/\left\|\mathbf{u}_i\right\|\right)$
NLMS using the step-size scaler : $$\widehat{\mathbf{w}}_{i+1}=\widehat{\mathbf{w}}_i+s\left(\beta, e_i/\left\|\mathbf{u}_i\right\|\right)\mu\frac{\mathbf{u}_i}{\left\|\mathbf{u}_i\right\|^2}e_i$$
VSS NLMS [18] using the step-size scaler : $$\widehat{\mathbf{w}}_{i+1}=\widehat{\mathbf{w}}_i+s\left(\beta, e_i/\left\|\mathbf{u}_i\right\|\right)\mu_i\frac{\mathbf{u}_i}{\left\|\mathbf{u}_i\right\|^2}e_i,$$ $$\mu'_{i+1}=\alpha\mu_i+\gamma e_i^2$$ $$\mu_{i+1}=\begin{cases}\mu_{\max}, & \text{if } \mu'_{i+1}>\mu_{\max}\\ \mu_{\min}, & \text{if } \mu'_{i+1}<\mu_{\min}\\ \mu'_{i+1}, & \text{otherwise}\end{cases}$$ $$\mu_1=\mu_{\max}=1,\ \mu_{\min}=10^{(-5)},\ \alpha=0.999,\ \gamma=0.1$$
NLMAD [5] using the step-size scaler : $$\widehat{\mathbf{w}}_{i+1}=\widehat{\mathbf{w}}_i+s\left(\beta, e_i/\left\|\mathbf{u}_i\right\|\right)\mu\ \mathrm{sign}\left(e_i\right)\frac{\mathbf{u}_i}{\left\|\mathbf{u}_i\right\|_1}$$
Robust VSS NLMS [11] using the step-size scaler : $$\widehat{\mathbf{w}}_{i+1}=\widehat{\mathbf{w}}_i+s\left(\beta, e_i/\left\|\mathbf{u}_i\right\|\right)\mu_i\ \mathrm{sign}\left(e_i\right)\frac{\mathbf{u}_i}{\left\|\mathbf{u}_i\right\|}$$ $$\mu_i=\min\left[\frac{\left\|e_i\right\|}{\left\|\mathbf{u}_i\right\|},\ \sqrt{\delta_i}\right]$$ $$\delta_{i+1}=\alpha\delta_i+(1-\alpha)\min\left[\frac{\left\|e_i^2\right\|}{\left\|\mathbf{u}_i\right\|^2},\ \delta_i\right]$$

the magnitude of the normalized error is very small, and thus, the step-size scalers are close to one. It means that the step-size scalers operate the role only when impulsive noise is present.

III. ADAPTIVE ALGORITHMS USING THE STEP-SIZE SCALER

As mentioned in the previous section, the proposed step-size scaler $s(\beta, e_i/\|\mathbf{u}_i\|)$ improves the robustness against impulsive noise in any gradient-based adaptive algorithms by scaling the step size. We apply the proposed method to NLMS, the VSS NLMS [18], the NLMAD [5], and the robust VSS NLMS [11] algorithms. The concept of the robust VSS NLMS is to constrain the energy of the filter update at each iteration; it leads to the best robustness in VSS NLMS algorithms. The adaptive algorithms used to compare the performance of $s(\beta, e_i/\|\mathbf{u}_i\|)$ are summarized in Table I. The proposed step-size scaler is just multiplied by the each step size of the adaptive algorithms, and it means that the proposed method can be easily applied to various adaptive algorithms.

In the simulation of channel identification, the order of the adaptive filter is 32, which is equal to that of the corresponding unknown channel, and $\mathbf{w}$ is randomly generated as a unit vector. The input sequence is an AR1 with pole in 0.8, and simulation results are obtained from 100 independent trials. The measurement noise b_i is added to $y_i=\mathbf{u}_i{}^{\mathrm{T}}\mathbf{w}$ with 30-dB signal-to-noise ratio (SNR), and the SNR is calculated by

$$\mathrm{SNR}=10\log_{10}\left(\frac{\mathbf{E}\left(y_i^2\right)}{\mathbf{E}\left(b_i^2\right)}\right).$$

The parameter α and the initial value of δ_i in the robust VSS NLMS and the impulsive measurement noise η_i are assigned as

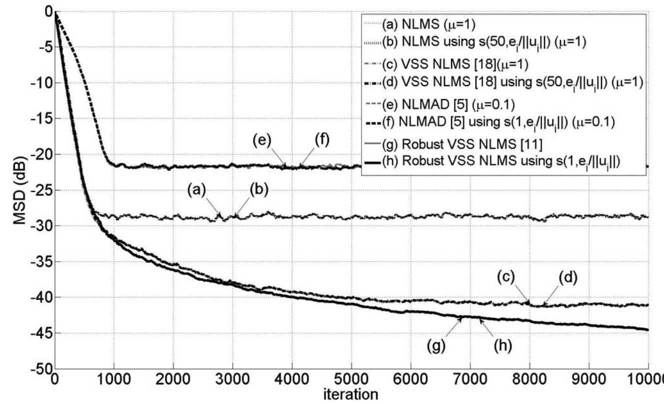

Fig. 2. Mean square deviation (MSD) learning curves (no impulsive measurement noise, $p=0$).

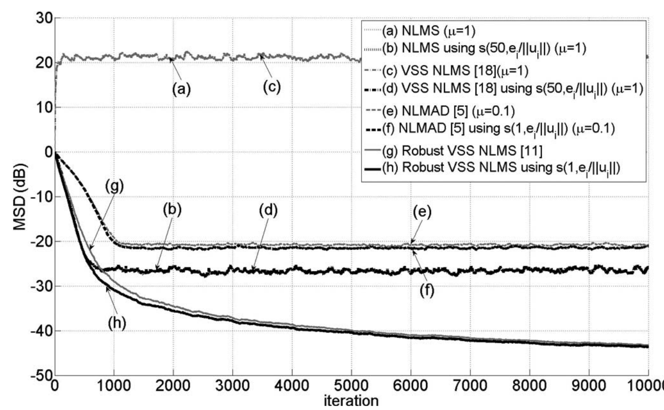

Fig. 3. MSD learning curves with impulsive measurement noise ($p=0.1$).

in [11]. η_i is generated as $\eta_i=\omega_i N_i$, where ω_i is Bernoulli process with $\Pr(\omega_i=1)=p$ and N_i is zero-mean Gaussian with power $\sigma_N^2=1000\sigma_y^2$.

First, we study when impulsive measurement noise does not appear. In Fig. 2, the performance of each adaptive algorithm is similar to that of each adaptive algorithm using $s(\beta, e_i/\|\mathbf{u}_i\|)$. Since impulsive measurement noise does not appear, $(e_i/\|\mathbf{u}_i\|)$ is very small, and thus, $s(\beta, e_i/\|\mathbf{u}_i\|)$ is close to 1, and it dose not operate the role of the step-size scaler.

Next, we consider the impulsive measurement noise under two cases: $p=0.1$ and 0.5. In Fig. 3, NLMS and the VSS NLMS algorithms are seen to not perform as adaptive filters; however, the step-size scaler makes them robust over impulsive noise. The performance of the NLMAD and the robust VSS NLMS algorithms is similar to those of the cases of using the step-size scaler when $p=0.1$. However, Fig. 4 shows that the step-size scaler improves the robustness against impulsive noise of high probability in robust adaptive algorithms $(p=0.5)$, although the probability is very inadequate condition in the signal processing applications.

In the simulations, we set $\beta=50$ in NLMS and VSS NLMS and $\beta=1$ in NLMAD and robust VSS NLMS; it is just rough value not optimal value. The tuning parameter β can be set roughly, and it is changed according to signal-to-impulsive-noise ratio and the adaptive algorithms which the step-size scalers are adapted in. Although the step-size scalers need a turning parameter, the simulation results show the step-size scalers improve the performance of various gradient-based

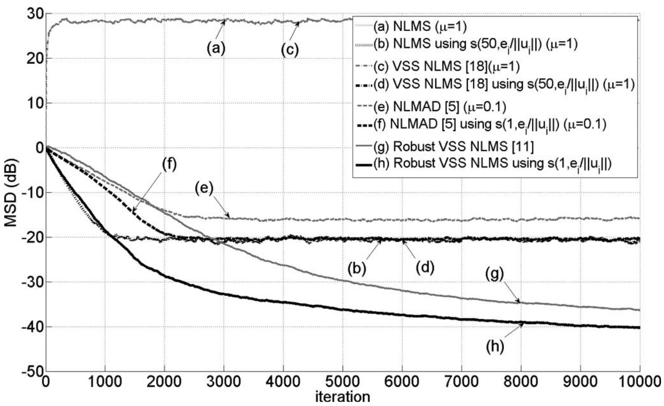

Fig. 4. MSD learning curves with impulsive measurement noise ($p = 0.5$).

adaptive algorithms whether the cost functions of the gradient-based adaptive algorithms are adapted in impulsive-noise environments or not.

IV. Conclusion

This brief has presented the concept of a step-size scaler by investigating and modifying the $\tanh$ cost function for adaptive filtering in impulsive measurement noise. The step-size scaler improves the robustness of various gradient-based adaptive algorithms by scaling the step size whenever impulsive noise appears. The step-size scaler is easily applicable to various gradient-based adaptive algorithms whether the cost functions of the gradient-based adaptive algorithms are adapted in impulsive-noise environments or not. This brief has applied the step-size scaler to NLMS, the VSS NLMS [18], the NLMAD [5], and the robust VSS NLMS [11] algorithms. Simulations have shown that the proposed step-size scaler improves the robustness against impulsive noise in various gradient-based adaptive algorithms.

References

[1] B. Widrow and S. D. Stearns, *Adaptive Signal Processing*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1985.

[2] S. Haykin, *Adaptive Filter Theory*, 4th ed. Englewood Cliffs, NJ, USA: Prentice-Hall, 2002.

[3] J. Benesty and Y. Huang, *Adaptive Signal Processing-Applications to Real World Problems*. Berlin, Germany: Springer-Verlag, 2003.

[4] Y. Huang, J. Benesty, and J. Chen, *Acoustic MIMO Signal Processing*. Boston, MA, USA: Springer-Verlag, 2006.

[5] 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.

[6] C. Kwong, “Dual sign algorithm for adaptive filtering,” *IEEE Trans. Commun.*, vol. COM-34, no. 12, pp. 1272–1275, Dec. 1986.

[7] 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*, May 2009, pp. 2065–2068.

[8] Y. R. Zheng and T. Shao, “A variable step-size LMP algorithm for heavy-tailed interference suppression in phased array radar,” in *Proc. IEEE Aerosp. Conf.*, Mar. 2009, pp. 1–6.

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

[10] 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.

[11] L. R. Vega, H. Rey, J. Benesty, and S. Tressens, “A new robust variable step-size NLMS algorithm,” *IEEE Trans. Signal Process.*, vol. 56, no. 5, pp. 1878–1893, May 2008.

[12] P. Petrus, “Robust Huber adaptive filter,” *IEEE Trans. Signal Process.*, vol. 47, no. 4, pp. 1129–1133, Apr. 1999.

[13] Y. Zou, S. C. Chan, and T. S. Ng, “A Recursive Least M-estimate (RLM) adaptive filter for robust filtering in impulse noise,” *IEEE Signal Process. Lett.*, vol. 7, no. 11, pp. 324–326, Nov. 2000.

[14] Y. Zou and S. C. Chan, “A robust quasi-Newton adaptive filtering algorithm for impulse noise suppression,” in *Proc. IEEE Int. Symp. Circuits Syst.*, May 2001, pp. 677–680.

[15] M. Z. A. Bhotto and A. Antoniou, “Robust quasi-Newton adaptive filtering algorithms,” *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 58, no. 8, pp. 537–541, Aug. 2011.

[16] F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel, *Robust Statistics: The Approach Based on Influence Functions*. New York, NY, USA: Wiley, 1986.

[17] C. Wang, H. C. Wu, and J. C. Principe, “A cost function for robust estimation of PCA,” in *Proc. SPIE*, 1996, vol. 2760, pp. 120–127.

[18] R. H. Kwong and E. W. Johnston, “A variable step size LMS algorithm,” *IEEE Trans. Signal Process.*, vol. 40, no. 7, pp. 1633–1642, Jul. 1992.