An Oracle Normalized Least Mean Square (NLMS) and a Simple Bayesian Detection NLMS Algorithm Robust to Impulse Noise

Hadi Zayyani[1], Yaser Attar[2]
Department of electrical and computer engineering Qom University of Technology Qom, Iran
zayyani@qut.ac.ir

Department of electrical and computer engineering Qom University of Technology Qom, Iran
attarizi@qut.ac.ir

Received : December 2015 Revised : March 2016 Accepted : June 2016

ABSTRACT:
This paper suggests an oracle NLMS algorithm and a simple Bayesian impulse noise detection normalized least-mean-square (NLMS) algorithm as an effective adaptive algorithm against impulsive noises. Initially, to have a fast algorithm, an optimization problem is introduced and then an oracle NLMS algorithm is devised. It has the largest reduction in misalignment error at each iteration with respect to the previous iteration. Oracle NLMS algorithm needs the values of impulse noises and hence is not practical. To have a practical variant of oracle NLMS algorithm, a simple Bayesian impulse noise detection NLMS algorithm is proposed. It is based on a MAP detection criterion and the impulse noise detection rule is proved to be a comparison of the absolute value of the error of the adaptive filter with a threshold and hence is very simple. Also, by assuming the sparsity of the impulse noises, the value of threshold is obtained via a simple statistical estimation. The simulation results in both dispersive and sparse system cases show the effectiveness of the suggested algorithm in terms of convergence rate and complexity.
KEYWORDS: Adaptive filter; NLMS algorithm; impulsive noise

1. INTRODUCTION

Adaptive filters have a wide range of application in signal processing including system identification, channel estimation, and echo cancelation [1]. Among various algorithms, NLMS algorithm is the most widely used algorithm in such applications due to its simplicity and robustness [1]. Unfortunately, NLMS and other adaptive algorithms suffer from impulsive noises which naturally exist in real world applications [2]. Hence, many different schemes are suggested to combat the impulsive noises.

The common trick to make the adaptive filters robust against impulsive noise is to use an appropriate cost function. Then, a gradient based adaptive algorithm could be used to minimize this cost function. Two main category of the cost functions are l^1-norm and l^2-norm. l^1-norm based adaptive algorithms offer robust performance with respect to impulsive noise in comparison to l^2-norm based adaptive algorithms. For the first time, l^1-norm and l^p-norm has been used for this purpose in [3] to combat against non-Gaussian stable processes. So, a normalized least mean absolute deviation (NLMAD) algorithm and a normalized least mean p-norm (NLMP) algorithm is proposed in [3], which uses l^1-norm and l^p-norm, respectively. Also, robust mixed norm (RMN) algorithm used a mixed l^1 and l^2-norm of the adaptive filter error [4]. In sequel, a normalized robust mixed norm (NRMN) adaptive algorithm was suggested for system identification [5]. In addition, a class of adaptive algorithms employing order statistic filtering was presented [6]. Also, a robust M-estimate adaptive filtering was suggested in [7] in which a new cost function based on an M-estimator is used to suppress the effect of noise. Besides, a family of adaptive algorithms robust to impulse noise was introduced which can be considered as a sign-error variant of the LMS algorithm [8]. Additionally, an adaptive threshold nonlinear algorithm is proposed in [9] for adaptive filters to be robust against impulse noise. Moreover, an affine projection sign algorithm (APSA) was suggested in [10] which uses an l^1-norm of the error with a constraint on the filter coefficients. After that, some other variants of it was proposed which proportionate affine projection sign algorithm for the application of network echo cancelation [11], kernel affine projection sign algorithm [12] and robust shrinkage affine projection sign algorithm [13] to name a few. Besides, a robust variable step-size NLMS algorithm was suggested which switches between l^1-norm and l^2-norm [14]. In addition, a new cost function based on hyperbolic tangent function is introduced in [15] which results in a robust algorithm with a step-size scalar. Recently, a continuous mixed p-norm (CMPN) adaptive algorithm has been proposed which continuously combines all l^p-norms for $1 \le p \le 2$ [16]. Recently, an exact NLMS algorithm with l^p-norm constraint is also suggested [17].

In this paper, we assume a general cost function based on the error value of the adaptive filter. Then, we consider all the gradient-based adaptive algorithms based on this general cost function. This can be regarded as an adaptive filter with a general variable step-size. To calculate the general step-size, and to have a fast convergence rate, an optimization problem is introduced which aims to have the largest reduction in misalignment error with respect to the previous iteration. This leads to a variant of NLMS algorithm which is called oracle NLMS. It is impractical since it needs the value of impulse noise. So, a Bayesian impulse noise detection NLMS (ID-NLMS) is proposed where impulse noises are detected by the algorithm. Our algorithm is different from a detection guided NLMS algorithm [18] which detects the activity of the taps of the sparse filter. The final criterion for the impulse noise detection which we derived is similar to those suggested in [9]. But, they propose an adaptive threshold heuristically, while we derive the threshold mathematically and from a Bayesian point of view. Simulation results show the effectiveness of the ID-NLMS algorithm in terms of speed of convergence and complexity in comparison to some other algorithms.

The organization of the paper is as follows. After introduction, we explain our problem in section 2. In section 3, we introduce an oracle NLMS adaptive algorithm. Section 4 suggest a Bayesian impulse noise detection NLMS which is a practical variant of oracle NLMS. Simulation results are presented in section 5, and the paper finally concludes with a conclusion.

2. PROBLEM FORMULATION

An unknown system is assumed to have an FIR impulse response which is $\mathbf{w}_o = [w_{o,0}, w_{o,1}, ..., w_{o,N-1}]^T$ where N is the length of the impulse response. The input vector of the unknown system at time index k is $\mathbf{x}(k) = [x(k), x(k-1), ..., x(k-N+1)]^T$. An adaptive filter with weight vector $\mathbf{w}(k) = [w_0(k), w_1(k), ..., w_{N-1}(k)]^T$ is used to iteratively estimate the unknown impulse response based on the output error of the adaptive filter. The output error is $e(k) = d(k) - \mathbf{w}^T(k)\mathbf{x}(k)$, where the desired signal $d(k)$ is comprised of the output $y(k) = \mathbf{w}_o^T\mathbf{x}(k)$ of the unknown system $\mathbf{w}_o$, and of impulsive noise $I(k)$ and background noise $v(k)$. So, the desired signal is $d(k) = \mathbf{w}_o^T\mathbf{x}(k) + v(k) + I(k)$. The goal of the adaptive filter is to update the weight vector iteratively to estimate the unknown impulse response.

3. THE ORACLE NLMS

To combat impulsive noises, various adaptive filtering algorithms use different cost functions based on the output error of the adaptive filter $e(k)$, where the l^1-norm and l^2-norm are two most well-known cost functions. Here, we assume a general cost function based solely on the present error term which is assumed to be $J(k) = f(e(k))$. Therefore, the gradient based adaptive algorithm based on this cost function is

$$\mathbf{w}(k+1) = \mathbf{w}(k) - \mu\nabla_{\mathbf{w}(k)} J(k) \tag{1}$$

which results in

$$\mathbf{w}(k+1) = \mathbf{w}(k) - \mu f'(e(k))\mathbf{x}(k) \tag{2}$$

and

$$\mathbf{w}(k+1) = \mathbf{w}(k) - \mu(k)\mathbf{x}(k) \tag{3}$$

Where f' is the derivative of the function f and the term $\mu(k) = \mu f'(e(k))$ is the general step-size. By this notation, devising a gradient based adaptive algorithm is equivalent to obtain the step-size $\mu(k)$. To have a fast converging gradient based algorithm, we suggest the following optimization problem:

$$Min \quad \| \mathbf{w}_o - \mathbf{w}(k+1) \|_2^2 - \| \mathbf{w}_o - \mathbf{w}(k) \|_2^2 \tag{4}$$

where

$$g(e(k)) = \| \mathbf{w}_o - \mathbf{w}(k+1) \|_2^2 - \| \mathbf{w}_o - \mathbf{w}(k) \|_2^2$$

and (4) aims to obtain the largest reduction in misalignment error $\| w_o - w(k+1) \|_2^2$ with respect to previous iteration. Substituting (3) into (4), and after some calculations, we have:

$$g(e(k)) = 2\mu(k)(\mathbf{w}(k) - \mathbf{w}_o) + \mu^2(k)\mathbf{x}^T(k)\mathbf{x}(k) \tag{5}$$

To minimize $g(e(k))$ with respect to $\mu(k)$, the step-size should be equal to:

$$\mu(k) = \frac{-(\mathbf{w}(k) - \mathbf{w}_o)^T \mathbf{x}(k)}{\mathbf{x}^T(k)\mathbf{x}(k)} \tag{6}$$

Some simple manipulations show that:

$$e(k) = (\mathbf{w}(k) - \mathbf{w}_o)^T \mathbf{x}(k) + \eta(k) \tag{7}$$

where $\eta(k) = I(k) + v(k)$ is the noise term which is the sum of the background noise $v(k)$ and impulsive noise $I(k)$. Replacing (7) into (6), we have

$$\mu(k) = \frac{\eta(k) - e(k)}{\mathbf{x}^T(k)\mathbf{x}(k)} \tag{8}$$

Hence, the final gradient based adaptive algorithm is

$$\mathbf{w}(k+1)=\mathbf{w}(k)-\frac{\eta(k)-e(k)}{\mathbf{x}^T(k)\mathbf{x}(k)}\mathbf{x}(k) \qquad (9)$$

We call the adaptive algorithm in (9) as oracle NLMS because it needs the value of the noise term $\eta(k)$. When impulse noise is large in comparison to the error $(\mathbf{w}(k)-\mathbf{w}_o)^T\mathbf{x}(k)$, then $e(k)\approx\eta(k)$ and hence the step-size equals $\mu(k)\approx 0$. So, the adaptive algorithm does not update the weight vector of the adaptive filter when the impulse noise is large. However, oracle NLMS is the optimum solution of the optimization problem in (4), but it is impractical as an adaptive filter since it needs the value of the impulse noise $\eta(k)$. In the next section, we propose a simple practical NLMS algorithm to overcome this drawback.

4. BAYESIAN IMPULSIVE NOISE DETECTION NLMS

There are some statistical models for impulsive noise [2, chapter 13]. Among them, for simplicity, we suppose the impulse noise $I(k)=q(k)a(k)$ has a Bernoulli-Gaussian distribution [2,4] which $q(k)=0,1$ is a Bernoulli random variable with probabilities $p,1-p$, and $a(k)$ is the amplitude of impulse noise when the impulse is present and is assumed to be Gaussian. When the large impulse occurs ($q(k)=1$), then $\eta(k)\approx e(k)$ assuming that the error term $\varepsilon(k)=(\mathbf{w}(k)-\mathbf{w}_o)^T\mathbf{x}(k)$ is much smaller than $\eta(k)=a(k)$. So, $\mu(k)$ would be zero and the filter weight vector is not updated. On the other hand, when impulse is not present, then $\eta(k)\approx 0$, So, $\mu(k)=\frac{-e(k)}{\mathbf{x}^T(k)\mathbf{x}(k)}$ and hence the recursion is an NLMS update formula with a step size equal to one. Therefore, the simple impulsive noise detection NLMS algorithm is:

$$\mathbf{w}(k+1)=\begin{cases}\mathbf{w}(k)+\dfrac{e(k)}{\mathbf{x}^T(k)\mathbf{x}(k)}\mathbf{x}(k) & \hat{q}(k)=1\\ w(k) & \hat{q}(k)=0\end{cases} \qquad (10)$$

where $\hat{q}(k)$ determines the presence of impulse noise. To determine $q(k)$, a maximum a posteriori (MAP) detection rule is used. Due to the central limit theorem (CLT), we assume a Gaussian distribution for $\varepsilon(k)$ with zero mean and variance σ_e^2. The MAP detection criterion is

$$\hat{q}(k)=\begin{cases}1 & p(e(k)\,|\,q(k)=1)\geq p(e(k)\,|\,q(k)=0)\\ 0 & Otherwise\end{cases} \qquad (11)$$

where the posterior probabilities $p(e(k)\,|\,q(k)=1)$ and $p(e(k)\,|\,q(k)=0)$ should be calculated. Based on (7), we have $e(k)=\varepsilon(k)+\eta(k)$. $\eta(k)$ is a Bernoulli-Gaussian random variable with the following probability density function (PDF):

$$p(\eta(k))=(1-p)\delta(\eta(k))+pN(0,\sigma_\eta^2) \qquad (12)$$

where p is the probability of occurring impulse noises, σ_η^2 is the variance of impulses, $\delta(.)$ is the Dirac impulse function, and $N(m,\sigma^2)$ is the Gaussian PDF with mean m and variance σ^2. Hence, $e(k)$ would be a mixture of Gaussian random variable

$$p(e(k))=(1-p)N(0,\sigma_1^2)+pN(0,\sigma_2^2)$$

where $\sigma_1^2=\sigma_\eta^2+\sigma_e^2$ and $\sigma_2^2=\sigma_e^2$. Therefore, $p(e(k)\,|\,q(k)=1)=pN(0,\sigma_1^2)$ and $p(e(k)\,|\,q(k)=0)=(1-p)N(0,\sigma_2^2)$. Some manipulations on the detection rule in (11) leads to the final criterion for presence of the impulse noise:

$$|\,e(k)\,|>Th \qquad (13)$$

where Th is the optimum threshold in MAP sense:

$$Th=\sqrt{\frac{2\ln(\frac{1-p}{p}\frac{\sigma_1}{\sigma_2})}{\frac{1}{\sigma_2^2}-\frac{1}{\sigma_1^2}}} \qquad (14)$$

If impulse noises are large or the variance σ_η^2 is much larger than error variance σ_e^2, then $\sigma_1>>\sigma_2$ and the optimum threshold can be approximated as

$$Th\approx\sqrt{2\sigma_2^2\ln(\frac{1-p}{p}\frac{\sigma_1}{\sigma_2})}=\alpha\sigma_2 \qquad (15)$$

where $\alpha=\sqrt{2\ln(\frac{1-p}{p}\frac{\sigma_1}{\sigma_2})}$ is a parameter that control the detection probability and false alarm probability of impulse noises [19]. To calculate this threshold, we use the following formula:

$$E\{e^2(k)\}=p\sigma_1^2+(1-p)\sigma_2^2 \qquad (16)$$

Assuming that the presence of impulse noise is sparse $p<<1$ and $p\sigma_1^2<<(1-p)\sigma_2^2$, the variance of σ_e^2 can be estimated as

$\sigma_e^2 \approx E\{e^2(k)\}$. $E\{e^2(k)\}$ can be simply calculated by the following recursion:

$$\sigma_2^2(k+1) = \gamma\sigma_2^2(k) + (1-\gamma)e^2(k) \qquad (17)$$

where γ is a forgetting factor close to one.

The impulse noise detection criterion which we obtained is similar to those suggested in [9]. The difference is that we derived the final impulse noise detection criterion mathematically and from a Bayesian point of view, while [9] proposes an adaptive threshold heuristically but analyzed mathematically.

5. SIMULATION RESULTS

Experiments were performed in a system identification application in presence of the both impulsive and background noise. Background noise $v(k)$ was modeled by an independent white Gaussian noise with a 20 dB signal-to-noise ratio (SNR). In addition, an impulse noise with a Bernoulli-Gaussian (BG) distribution was added to the system output. The variance of impulse noise is equal to $\sigma_\eta^2 = 100$, and the probability of being impulsive is $p = 0.02$. The unknown system $\mathbf{w}_o$ was assumed to be a 100-tap impulse response. Two cases of dispersive and sparse impulse response were examined. The input signal $x(k)$ was an AR(2) signal with the recursion $x(k) = 0.4x(k-1) - 0.4x(k-2) + z(k)$, where $z(k)$ was a white Gaussian noise with variance $\sigma_z^2 = 1$.

The simulations were repeated 100 times with new unknown system and new input signal. The performances of the algorithms were measured by a normalized misalignment error defined as $\Omega(k) = 20\log_{10}(\frac{\| w_o - w(k) \|_2}{\| w_o \|_2})$ which were averaged on 50 independent trials.

At first experiment, a dispersive impulse response was used. The elements of the impulse response are generated by a Gaussian noise with zero mean and unit variance. We compared ID-NLMS with NLMS. To investigate the effect of parameter α, three values of $\alpha = 1,3,5$ were examined. To fairly compare the algorithms, the step sizes of various algorithms were selected in such a way that all algorithms had a same final misalignment error. So, the step size μ was selected as 0.1, 0.5, 0.8, and 1 for NLMS, ID-NLMS with $\alpha = 5$, ID-NLMS with $\alpha = 3$, and ID-NLMS with $\alpha = 1$, respectively. The misalignment error curves for algorithms are shown in Fig. 1. It shows the superiority of the proposed ID-NLMS over NLMS. Also, it shows that the best value for the parameter α with respect to convergence speed is $\alpha = 3$. So, in the following experiments we used this value for α.

At the second experiment, similar to first experiment, a dispersive impulse response was used. ID-NLMS is compared to oracle-NLMS, NLMS, step-size scalar NLMS (SS-NLMS) [15], robust VSS-NLMS [14] and CMPN with uniform weighting function [16]. The step sizes are equal to 0.4, 1, 0.55 and 0.004 for NLMS, ID-NLMS, SS-NLMS and CMPN. For SS-NLMS, the parameter β is selected as 1 for the best performance with respect to the rate of convergence. The memory factor parameter in robust VSS-NLMS [14] is assumed to be $\alpha = 0.99$ for the best performance in terms of speed of convergence. The misalignment error curves for algorithms are shown in Fig. 2. It confirms that the proposed ID-NLMS has the fastest convergence rate among these algorithms because it is derived from the oracle-NLMS which is devised to have a fast convergence rate. Among the algorithms, robust VSS-NLMS has the minimum final misalignment error at the expense of later convergence.

At the third experiment, a sparse system was examined. The sparse impulse response elements are drawn from a BG distribution with the activity probability equal to 0.1 which means that 10% of the elements are non-zero. The non-zero elements are Gaussian with zero mean and with unit variance. Again, ID-NLMS is compared to the aforementioned algorithms. The step sizes are equal to 0.13, 0.9, 0.35 and 0.003 for NLMS, ID-NLMS, SS-NLMS and CMPN to have the same final misalignment error around -18dB. The other parameters are the same as the second experiment. The misalignment error curves for algorithms are shown in Fig. 3. It confirms again that the proposed ID-NLMS has the fastest convergence rate among these algorithms. So, the results in the case of sparse system are very similar to the case of dispersive system. We also observed that in the case of sparse system, the performance of the NLMS is poorer than the case of dispersive system.

We used the CPU time as a measure of complexity. Although, the CPU time is not an exact measure, it can give us a rough estimation of the complexity for comparing our algorithm with other algorithms. Our simulations were performed in MATLAB7.0 environment using an Intel 3.00 GHz processor with 4 GB of RAM and under windows 7 Microsoft operating system. The results are shown in Table I. It shows that ID-NLMS is slightly more complex than NLMS.

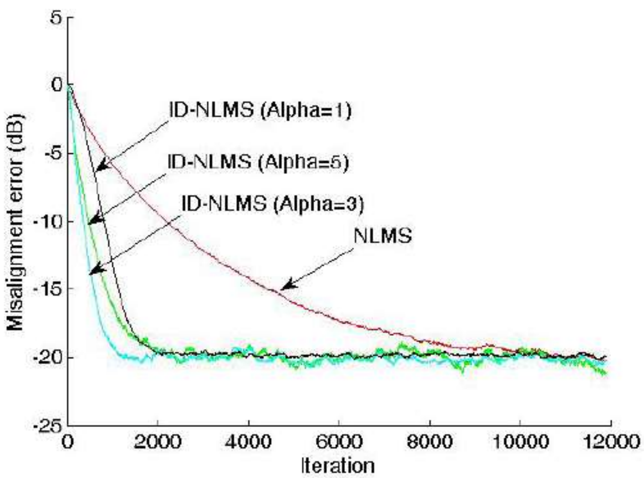

Fig.1, Averaged normalized misalignment error of NLMS, ID-NLMS with $\alpha = 1$, ID-NLMS with $\alpha = 3$ and ID-NLMS with $\alpha = 5$.

6. CONCLUSION

In this paper, a simple Bayesian impulse noise detection NLMS algorithm was proposed for adaptive filtering which is deduced from an oracle-NLMS algorithm. The oracle-NLMS itself is derived from a suggested optimization problem which is proposed to have the largest reduction in misalignment error at each iteration. Simulation results in both dispersive and sparse system identification application show that the suggested algorithm is one of the fastest algorithms among some examined algorithms while it is robust against impulsive noise and has slightly more complexity than the standard NLMS.

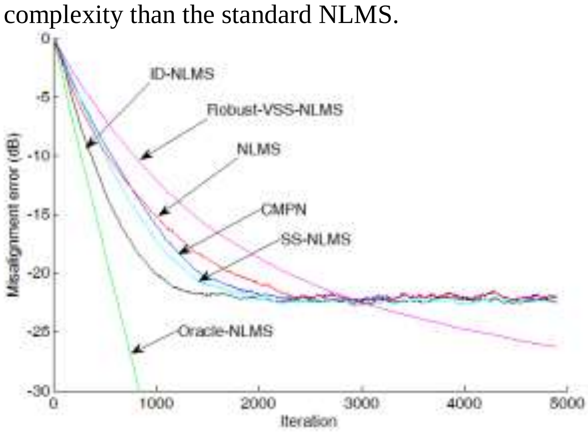

Fig. 2. Averaged normalized misalignment error of oracle-NLMS, ID-NLMS, NLMS, Robust-VSS-NLMS, CMPN and SS-NLMS algorithms in a system identification application with a dispersive system.

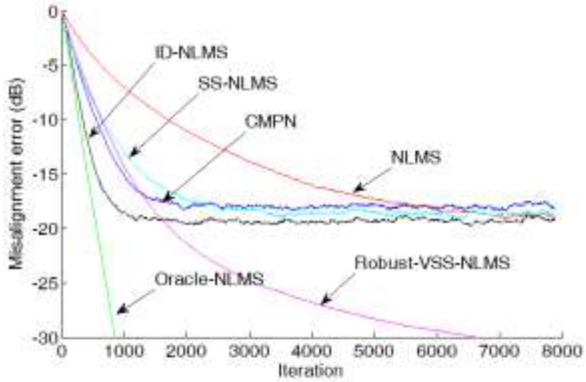

Fig. 3. Averaged normalized misalignment error of oracle-NLMS, ID-NLMS, NLMS, Robust-VSS-NLMS, CMPN and SS-NLMS algorithms in a system identification application with a sparse system.

Table 1. Average Run time of algorithms

Algorithm	Average Run time (Sec)
NLMS	0.3127
Oracle-NLMS	0.3551
Robust-VSS-NLMS	0.3544
CMPN	0.3808
SS-NLMS	0.4846
ID-NLMS	0.3522

7. REFERENCES

[1] S. Haykin, **Adaptive Filter Theory**. Prentice Hall,4th edition, 2002.

[2] S. V. Vaseghi, **Advanced digital signal processing and noise reduction**. John Willey, 4th edition, 2008.

[3] Arikan, Orhan, A. Enis Cetin, and Engin Erzin. "**Adaptive filtering for non-Gaussian stable processes**." *IEEE Signal Processing Letters* 1.11 (1994): 163-165.

[4] Chambers, Jonathon, and Apostolos Avlonitis. "**A robust mixed-norm adaptive filter algorithm**." *IEEE Signal Processing Letter" ,* 4.2 (1997): 46-48.

[5] Papoulis, Eftychios V., and Tania Stathaki. "**A normalized robust mixed-norm adaptive algorithm for system identification**." *IEEE Signal Processing Letters* 11.1 (2004): 56-59.

[6] Haweel, Tarek I., and Peter M. Clarkson. "**A class of order statistic LMS algorithms**." *IEEE Transactions on Signal Processing*, 40.1 (1992): 44-53.

[7] Zou, Yuexian, S. C. Chan, and T. S. Ng. "**Robust M-estimate adaptive filtering**." *IEE Proceedings-Vision, Image and Signal Processing*, 148.4 (2001): 289-294.

[8] Soria, Emilio, et al. "**A Family of Adaptive Algorithms Robust to Impulsive Noise**." 2007 *15th International Conference on Digital Signal Processing. IEEE*, 2007.

[9] Koike, Shiin'ichi. "**Adaptive threshold**

nonlinear algorithm for adaptive filters with robustness against impulse noise." *International Acoustics, Speech, and Signal Processing,*. IEEE, 1996.

[10] Shao, Tiange, Yahong Rosa Zheng, and Jacob Benesty. "**An affine projection sign algorithm robust against impulsive interferences**." *IEEE Signal Processing Letters* 17.4 (2010): 327-330.

[11] Yang, Zengli, Yahong Rosa Zheng, and Steven L. Grant. "**Proportionate affine projection sign algorithms for network echo cancellation**." *IEEE Transactions on Audio, Speech, and Language Processing* 19.8 (2011): 2273-2284.

[12] Wang, Shiyuan, Jiuchao Feng, and K. Tse Chi. "**Kernel affine projection sign algorithms for combating impulse interference**." *IEEE Transactions on Circuits and Systems II: Express Briefs* 60.11 (2013): 811-815.

[13] Bhotto, Md Zulfiquar Ali, M. Omair Ahmad, and M. N. S. Swamy. "**Robust shrinkage affine-projection sign adaptive-filtering algorithms for impulsive noise environments.**" *IEEE Transactions on Signal Processing* 62.13 (2014): 3349-3359.

[14] Vega, Leonardo Rey, et al. "**A new robust variable step-size NLMS algorithm**." *IEEE Transactions on Signal Processing* 56.5 (2008): 1878-1893.

[15] Song, Insun, PooGyeon Park, and Robert W. Newcomb. "**A normalized least mean squares algorithm with a step-size scaler against impulsive measurement noise**." *IEEE Transactions on Circuits and Systems II: Express Briefs* 60.7 (2013): 442-445.

[16] Zayyani, Hadi. "**Continuous mixed-norm adaptive algorithm for system identification**." *IEEE Signal Processing Letters* 21.9 (2014): 1108-1110.

[17] Weruaga, Luis, and Shihab Jimaa. "***Exact NLMS Algorithm with-Norm Constraint.***" *IEEE Signal Processing Letters* 22.3 (2015): 366-370.

[18] Homer, John. "**Detection guided NLMS estimation of sparsely parametrized channels**." *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing* 47.12 (2000): 1437-1442.

[19] Skolnik, Merril I. **Introduction to radar systems**. McGraw Hill, 3rd edition, 2003.