A NORMALIZED MIXED–NORM ADAPTIVE FILTERING ALGORITHM ROBUST UNDER IMPULSIVE NOISE INTERFERENCE

Danilo P. Mandic, Eftychios V. Papoulis, and Christos G. Boukis

CSP Group, Department of Electrical and Electronic Engineering
Imperial College, London, UK
{d.mandic, eftychios.papoulis, christos.boukis}@ic.ac.uk

ABSTRACT

A Normalized Robust Mixed–Norm (NRMN) algorithm for system identification in the presence of impulsive noise is introduced. The standard Robust Mixed–Norm (RMN) algorithm, despite its ability to cope with impulsive noise by virtue of combining the first and second error norm in the cost function it minimizes, exhibits slow convergence, requires a stationary operating environment, and employs a constant step–size which needs to be determined a–priori. To overcome these limitations, the proposed NRMN algorithm introduces a time varying learning rate which is derived based upon the dynamics of the input signal, and thus no longer requires a stationary environment, a major drawback of the RMN algorithm. The normalized step–size is bounded from above and a parameter is introduced within its upper–bound, which provides a trade–off between the convergence rate and the steady–state coefficient error. The analysis and experimental results show that the proposed NRMN exhibits increased convergence rate and substantially reduces the steady–state coefficient error, as compared to the Least Absolute Deviation (LAD) and RMN algorithm.

1. INTRODUCTION

Interference noise in the form of sparsely distributed impulses arises frequently in a variety of practical situations, including speech, image, biomedical, and communications applications. Therefore, there is a need for adaptive filtering algorithms that are robust to impulsive interference. The Median Least Mean Square (MLMS) adaptive filter [1] [2] provides a solution to this problem by median filtering the most recent gradient terms and using this information for the weight adaptation. More recently, the Robust Mixed–Norm (RMN) algorithm [3] for adaptive Finite Impulse Response (FIR) filters in the system identification setting shown in Figure 1, has been proposed. It emerged as a modification of the mixed–norm adaptive algorithm proposed in [4], by replacing the fourth order error norm appearing therein with the first order one. The RMN algorithm minimizes a cost function defined as the following convex combination of the error norms

$$J(k) \triangleq \lambda(k)E\{e^2(k)\} + [1-\lambda(k)]E\{|e(k)|\}, \qquad (1)$$

where the mixing parameter $\lambda(k) \in [0,1]$. The instantaneous output error of the algorithm is $e(k) = d(k) - \mathbf{x}^T(k)\mathbf{w}(k)$ where $\mathbf{x}(k)$ denotes the length N tap–input–vector $\mathbf{x}(k) \triangleq [x(k),\ldots,x(k-N+1)]^T$, $\mathbf{w}(k)$ is the vector of adaptive weights $\mathbf{w}(k) \triangleq [w_0(k),\ldots,w_{N-1}(k)]^T$ at iteration k and $(\cdot)^T$ denotes the vector transpose. The desired signal $d(k)$ is comprised of the unknown system output $y(k) = \mathbf{x}^T(k)\mathbf{w}_{\text{opt}}$, of an impulsive noise component $n(k)$ and possibly of a zero–mean white noise component $u(k)$, which is assumed to be drawn from a normal distribution i.e., $u(k) \sim \mathcal{N}(0,\sigma_u^2)$. The impulsive noise component $n(k)$ can be modelled as $n(k) = \alpha(k)I(k)$ [5], with $\alpha(k)$ a binary process of independent and identically distributed (i.i.d.) random variables, described by the probability $p\{\alpha(k)=1\} = c$, $p\{\alpha(k)=0\} = 1-c$, where c represents the probability of the occurrence of impulsive interference. For $n(k)$ to be a realistic model of the impulsive noise, var$\{I(k)\} \gg$ var$\{y(k)\}$. Process $I(k)$ is assumed to be uncorrelated with $\alpha(k)$ and to have a symmetric amplitude distribution. This way, var$\{n(k)\} = c$var$\{I(k)\} = \sigma_n^2$. The weight update of RMN then takes the form [3]

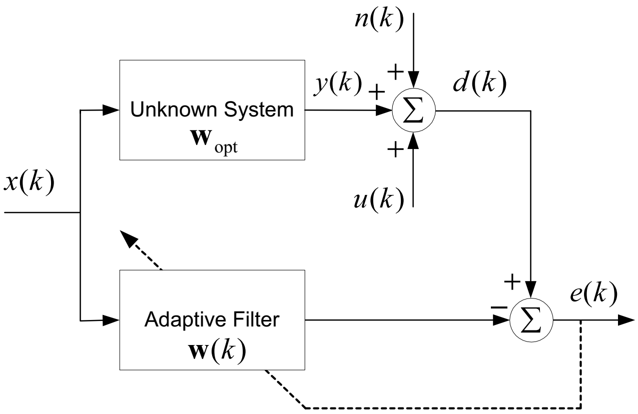

Fig. 1. The system identification setting. The signals $n(k)$ and $u(k)$ represent the impulsive and additive white Gaussian noise respectively and $\mathbf{w}_{\text{opt}}$ the vector of the unknown FIR system coefficients.

$$\begin{aligned}\mathbf{w}(k+1) = \mathbf{w}(k) + \mu\{\lambda(k)2e(k)\\ +[1-\lambda(k)]\text{sign}[e(k)]\}\mathbf{x}(k),\end{aligned} \qquad (2)$$

where μ is a constant step–size and $\lambda(k) \triangleq$ Prob$\{\mathbf{d} > |d(k)| \cup \mathbf{d} < -|d(k)|\}$, with the random variable $\mathbf{d}$ modelling the desired signal $d(k)$ in the absence of impulsive noise. For the case $\mathbf{d} \sim \mathcal{N}(0,\sigma_d^2)$[1], the mixing parameter $\lambda(k)$ is given by

[1]Due to the Central Limit Theorem [5], this assumption holds approximately in practice for large filter length N.

$$\lambda(k) = 2\mathrm{erfc}\Big\{\frac{|d(k)|}{\hat{\sigma}_d}\Big\}, \tag{3}$$

where $\hat{\sigma}_d$ is the estimate of the standard deviation of $d(k)$ in the absence of impulsive interference.

Although the RMN successfully solved the problem of divergence under impulsive interference due to the inherent sign–error LMS adaptation[2] and the automatic switching mechanism provided by (3), it suffers from slow and generally non–uniform convergence due to the use of the constant step–size μ in (2) and the possibly non–stationary behavior of the input. Furthermore, in a practical application, the value of μ cannot be known a–priori, thus rendering the RMN unsuitable for use in practice.

To this cause, an extension of the RMN, the Normalized Robust Mixed–Norm (NRMN) algorithm is proposed. It employs a time varying step–size $\mu(k)$ in (2) which takes into account the dynamics of the tap–input–vector, the current value of $\lambda(k)$ from (3), and the instantaneous output error value $e(k)$. For $\lambda(k) = 1$ the proposed NRMN degenerates into the Normalized Least Mean Square (NLMS) algorithm and thus inherits its fast convergence, during the periods where the impulsive noise is not present, whereas for $\lambda(k) \in (0,1)$ it essentially represents an NLMS followed by a normalized sign–error LMS update. A user–defined constant value parameter is incorporated in the upper–bound imposed on $\mu(k)$ so as to provide a trade–off between the steady–state coefficient error and the convergence rate. The theoretical analysis and experimental results demonstrate that the proposed NRMN exhibits superior performance as compared to that of RMN. For the same steady–state error it exhibits higher convergence rate, whereas for the the same convergence rate it offers substantially lower steady–state error. This improvement in the performance is achieved at practically no increase in computational complexity.

2. ANALYSIS OF THE NRMN ALGORITHM

The proposed Normalized Robust Mixed Norm (NRMN) algorithm is derived by replacing the constant step–size μ in (2) with a variable step–size $\mu(k)$ which takes into account the instantaneous tap–input–vector power $\parallel \mathbf{x}(k) \parallel_2^2$ and the current values of the mixing parameter $\lambda(k)$ and output error $e(k)$. To derive $\mu(k)$, the methodology applied in [6] for the NLMS algorithm is followed and the error signal $e(k, \mathbf{w}(k+1))$ is expressed in terms of $e(k, \mathbf{w}(k))$ using the Taylor series expansion

$$e(k, \mathbf{w}(k+1)) = e(k, \mathbf{w}(k)) + \sum_{j=0}^{N-1} \frac{\partial e(k, \mathbf{w}(k))}{\partial w_j(k)} \triangle w_j(k) + \sum_{j=0}^{N-1}\sum_{i=0}^{N-1} \frac{\partial^2 e(k, \mathbf{w}(k))}{\partial w_j(k) \partial w_i(k)} \triangle w_j(k) \triangle w_i(k) + \cdots, \tag{4}$$

where both $e(k, \mathbf{w}(k+1))$ and $e(k, \mathbf{w}(k))$ are conditioned on the tap–input–vector $\mathbf{x}(k)$ and the desired signal $d(k)$. From the error

$$e(k) = d(k) - \mathbf{x}^T(k)\mathbf{w}(k) \tag{5}$$

[2]The sign–error LMS applies the update term $\mu\mathrm{sign}[e(k)]\mathbf{x}(k)$ instead of $2\mu e(k)\mathbf{x}(k)$, which under impulsive interference has a lesser impact on the weight vector $\mathbf{w}(k)$.

and the update equation (2), all but the first order derivatives in (4) are zero, which yields

$$e(k, \mathbf{w}(k+1)) = e(k, \mathbf{w}(k))\Big\{1 - \mu(k) \parallel \mathbf{x}(k) \parallel_2^2 \big[2\lambda(k) + \frac{1-\lambda(k)}{|e(k)|}\big]\Big\}. \tag{6}$$

To minimize the a–posteriori error $e(k, \mathbf{w}(k+1))$, the term inside the large brackets in (6) is set to zero, hence yielding the normalization formula for $\mu(k)$, given by

$$\mu(k) = \frac{|e(k)|}{\big\{2\lambda(k)|e(k)| + [1-\lambda(k)]\big\} \parallel \mathbf{x}(k) \parallel_2^2}. \tag{7}$$

Denoting by $\mathbf{w}_{\text{opt}}$ the vector of the optimal weights, the dynamical behavior of the coefficient error vector

$$\mathbf{v}(k) \triangleq \mathbf{w}_{\text{opt}} - \mathbf{w}(k), \tag{8}$$

is governed by

$$\mathbf{v}(k+1) = \big\{\mathbf{I} - 2\mu(k)\lambda(k)\mathbf{x}(k)\mathbf{x}^T(k)\big\}\mathbf{v}(k) - \mu(k)\big\{[1-\lambda(k)]\mathrm{sign}[e(k)] + 2\lambda(k)n(k)\big\}\mathbf{x}(k). \tag{9}$$

According to the analysis in [7], for a constant step–size μ and mixing parameter λ, the approximation $E\{\mathrm{sign}[e(k)]\mathbf{x}(k)\} \approx \sqrt{2/\pi}[1/\sigma_e(k)]E\{e(k)\mathbf{x}(k)\}$ holds for small μ, where $\sigma_e(k)$ is the standard deviation of the error signal. From (9), the sufficient condition for convergence in the mean is given by

$$\mu(k) \leq \mu_{\text{UB}} \triangleq \frac{2A}{\Big[2\lambda(k) + [1-\lambda(k)]\sqrt{\frac{2}{\pi(\sigma_n^2+\sigma_u^2)}}\Big] N\sigma_x^2}. \tag{10}$$

In (10), $A = 1$, μ and λ are replaced by their time varying counterparts and the variance $\sigma_e^2(k)$ of the error signal by $(\sigma_n^2 + \sigma_u^2)$, thus yielding a lower value for the upper–bound. The significance of the constant value parameter A introduced in (10) will be explained in the sequel.

Assuming that the impulsive noise component occurs at time instant k_0, it yields a value $d(k_0)$ such that $|d(k_0)| \gg \hat{\sigma}_d$ as a result of $\mathrm{var}\{I(k)\} \gg \mathrm{var}\{y(k)\}$. For a large value of $\frac{|d(k_0)|}{\hat{\sigma}_d}$, the mixing parameter value $\lambda(k_0)$ of (3) used by NRMN will be very close to zero. If at the time instant k_0 a convergence state very close to $\mathbf{w}_{\text{opt}}$ had been attained, the value of the coefficient error vector $\mathbf{v}(k_0+1)$ after performing the weight update at time k_0 in the presence of impulsive noise is readily obtained. This can be found from (9) by setting $\mathbf{v}(k_0) \approx 0$ and $\lambda(k_0) \approx 0$, which yields $\mathbf{v}(k_0+1) = -\mu(k_0)\mathrm{sign}[e(k_0)]\mathbf{x}(k_0)$ and effectively

$$\parallel \mathbf{v}(k_0+1) \parallel_2^2 = \mu^2(k_0) \parallel \mathbf{x}(k_0) \parallel_2^2. \tag{11}$$

By combining (7), (10), (11), and taking into account that for $\lambda(k) = 0$, the upper–bound μ_{UB} in (10) is simply $\frac{A\sqrt{2\pi(\sigma_n^2+\sigma_u^2)}}{N\sigma_x^2}$, the value of the weight error vector norm right after the occurrence of impulsive interference is obtained as

$$\| \mathbf{v}(k_0+1) \|_2^2 = \min\left\{ \frac{e^2(k_0)}{\| \mathbf{x}(k_0) \|_2^2}, \frac{2\pi(\sigma_n^2+\sigma_u^2)A^2}{N\sigma_x^4} \frac{\| \mathbf{x}(k_0) \|_2^2}{N} \right\}. \quad (12)$$

If the length N of the adaptive filter is large, the second term in (12) can be approximated by $\frac{2\pi(\sigma_n^2+\sigma_u^2)A^2}{N\sigma_x^2}$. This term is independent of the instantaneous values of the signals, is inversely proportional to the SNR and the filter length N, and proportional to A^2. On the other hand, the first term in (12), namely $\frac{e^2(k_0)}{\|\mathbf{x}(k_0)\|_2^2}$, can take a large and unpredictable value under impulsive interference. Thus, according to the above analysis, the introduction of μ_{UB} guarantees that whenever the impulsive noise occurs after a state of convergence, the maximum value $\| \mathbf{v}(k_0+1) \|_2^2$ can take is known a–priori and equals $\frac{2\pi(\sigma_n^2+\sigma_u^2)A^2}{N\sigma_x^2}$. If a smaller value of $\| \mathbf{v}(k_0+1) \|_2^2$ is observed, this is $\frac{e^2(k_0)}{\|\mathbf{x}(k_0)\|_2^2}$. Therefore, the benefit of the constraint $\mu(k) \leq \mu_{\text{UB}}$ is that whenever impulsive noise occurs during a state of convergence, a well–defined, a–priori known and controllable by the constant value parameter A upper–bound is imposed on the coefficient error norm $\| \mathbf{v}(k_0+1) \|_2^2$. Allowing for values $A < 1$, reduces this upper–bound and provides a means of controlling $\| \mathbf{v}(k_0+1) \|_2^2$ and effectively the steady–state performance of the algorithm, since the steady–state coefficient error reduces with $\| \mathbf{v}(k_0+1) \|_2^2$. This is better illustrated in the next section. It should be noted that the upper–bound $\frac{2\pi(\sigma_n^2+\sigma_u^2)A^2}{N\sigma_x^2}$, which can be considered as the worst case scenario for the system, decreases with an increase in the value of the filter length N, a clear merit of the proposed NRMN algorithm, since in practical situations the length N is usually large.

The value of the mixing parameter $\lambda(k)$ is given by (3), where the standard deviation estimate $\hat{\sigma}_d$ is obtained from

$$\hat{\sigma}_d(k) = \sqrt{\frac{1}{N_w - K - 1}\, \mathbf{o}^T \mathbf{T} \mathbf{o}}, \quad (13)$$

where $\mathbf{T} \triangleq \text{Diag}[1,\ldots,1,0,\ldots,0]$ and the vector $\mathbf{o}(k) \triangleq \mathcal{O}([d(k),\ldots,d(k-N_w+1)]^T)$ contains the N_w most recent samples of $d(k)$, ordered from the smallest to the largest absolute value (the symbol $\mathcal{O}(\cdot)$ denotes the ordering operation). In (13), matrix $\mathbf{T}$ sets the last K elements of $\mathbf{o}$ to zero and forms an unbiased estimate $\hat{\sigma}_d(k)$ [5] using the remaining $(N_w - K)$ elements. The idea behind this is that if some of the N_w elements of $\mathbf{d}(k) \triangleq [d(k),\ldots,d(k-N_w+1)]^T$ are affected by impulsive noise, these will be the ones with the highest absolute values[3]. Thus, by setting the last K elements of $\mathcal{O}(k)$ to zero, an estimate $\hat{\sigma}_d(k)$ is obtained which is unaffected by the impulsive noise, provided that no more than K samples were affected by $n(k)$. In [3], the estimate $\hat{\sigma}_d(k)$ was proposed using $\mathbf{T} \triangleq \text{Diag}[0,1,\ldots,1,0]$ and ordering of the most recent N_w samples of $d(k)$ from the smallest algebraical value to the largest. A more robust estimate can be obtained by using $\mathbf{T} \triangleq \text{Diag}[1,\ldots,1,0,0]$ and ordering as in (13). This estimate can further be improved by allowing values of K grater than two, as we do in (13).

[3]This is a realistic assumption since the variance of $I(k)$ is substantially larger than that of $y(k)$.

3. EXPERIMENTAL RESULTS

Two sets of experiments were performed to test the proposed Normalized Robust Mixed–Norm (NRMN) algorithm. Both were conducted in the System Identification setting, depicted in Figure 1 and the performance of the algorithm was evaluated with respect to the normalized coefficient error quantified as $10\log_{10}(\| \mathbf{v}(k) \|_2^2 \,/\, \| \mathbf{w}_{\text{opt}} \|_2^2)$. For $n(k)$, the parameters $c = 10^{-2}$ and $var\{I(k)\} = \frac{10^4}{12}$ were used [3]. Expression (13) was used to provide the $\hat{\sigma}_d(k)$ estimates for RMN and NRMN for the values $K = 2$ and $N_w = N$. Zero–mean unit–variance white Gaussian noise was used as input i.e., $x(k) \sim \mathcal{N}(0,1)$. All the results were obtained by averaging 1000 outcomes of independent trials for the first and 100 outcomes of independent trials for the second experiment. For the Least Absolute Deviation[4] (LAD) and RMN, the parameters $\mu_{\text{LAD}} = 0.06$ and $\mu_{\text{RMN}} = 0.0324$ were used [3].

In the first experiment, the performance of NRMN was evaluated and compared to that of NLMS, LAD and RMN. The unknown system was a nine–tap filter described by $\mathbf{w}_{\text{opt}} = [1,2,3,4,5,4,3,2,1]^T$, which was also used for the simulations in [3], normalized so as to have unit power $\mathbf{w}_{\text{opt}}^T \mathbf{w}_{\text{opt}} = 1$. Figure 2 provides the results obtained for impulsive noise, while for Figure 3 white Gaussian noise for SNR $= 20$dB was also present. As seen from Figures 2 and 3, for the same steady–state coefficient error the proposed NRMN algorithm exhibits higher convergence rate, compared to RMN, whereas for the same convergence rate it exhibits a lower steady–state coefficient error. Its performance is also superior to that of LAD. The NLMS algorithm, which is not shown in the Figures, was also simulated. In both cases it failed to converge in the presence of impulsive interference, providing a steady–state coefficient error at a level of $+10$ dB. The Figures also demonstrate the effect of the variation of the parameter A in (10) on the performance of NRMN. When reducing the value of A, both the steady–state coefficient error and convergence rate decrease. On the other hand, by increasing the value of A, the steady–state coefficient error increases and so too does the convergence rate, which cannot however increase further than a certain value. This dependence of the steady–state error on the value of A is expected since the second term at the right hand side of (12) is proportional to A.

In the second experiment, the performance of NRMN was again compared to that of RMN, this time for an unknown system of length $N = 100$. For each trial, the coefficients of this system were drawn as samples of 100 i.i.d. random variables uniformly distributed between $[-\frac{1}{2}, \frac{1}{2}]$ and normalized so as to have unit power i.e., $\mathbf{w}_{\text{opt}}^T \mathbf{w}_{\text{opt}} = 1$. Additive white Gaussian noise of SNR $= 20$dB was also present. As seen from Figure 4, NRMN exhibits superior performance as compared to RMN. Similar conclusions as before for the effect of varying the parameter A on the performance of NRMN can also be drawn from Figure 4. Again, NLMS was tested and failed to converge.

4. CONCLUSIONS

A Normalized Robust Mixed–Norm (NRMN) algorithm for Finite Impulse Response (FIR) adaptive filters, which is robust to impulsive interference and exhibits fast convergence rate, has been presented. The cost function to minimize is a convex mixture

[4]The sign–error LMS algorithm.

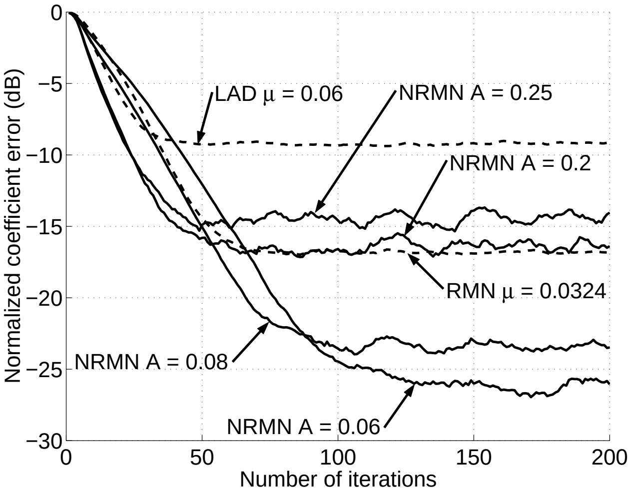

Fig. 2. Performance comparison of LAD, RMN and NRMN for varying A, under impulsive noise interference $n(k)$, for $u(k) = 0$ and a filter of length $N = 9$.

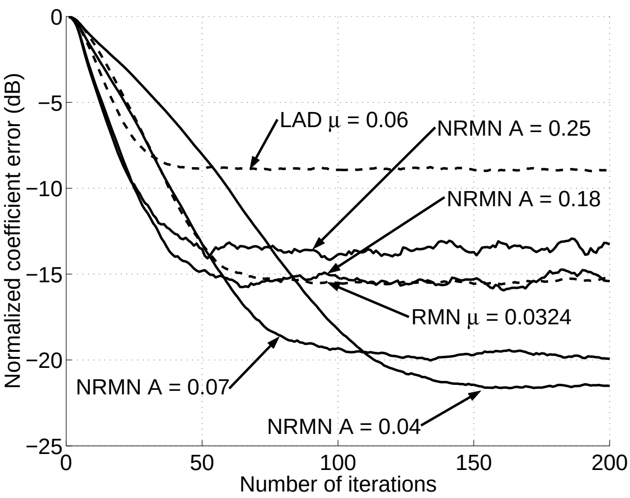

Fig. 3. Performance comparison of LAD, RMN and NRMN for varying A, under impulsive noise interference $n(k)$, for $u(k)$ of SNR $= 20$dB and a filter of length $N = 9$.

of the first and second error norm, controlled by a time varying mixing parameter. This parameter has been derived based upon an estimate of the variance of the desired signal –in the absence of impulsive interference and its current for each time instant value. The normalization scheme has been next introduced and an upper–bound has been imposed on the step–size employing a user–defined constant value parameter, which makes it possible to trade–off between the convergence rate and the steady–state coefficient error. The proposed NRMN has been shown to exhibit higher convergence rate than the Robust Mixed–Norm (RMN) algorithm for the same steady–state error, whereas by properly choosing the constant value parameter it employs within the upper–bound, it is possible to attain the same convergence rate for a substantially lower steady–state error. This significant improvement in the per-

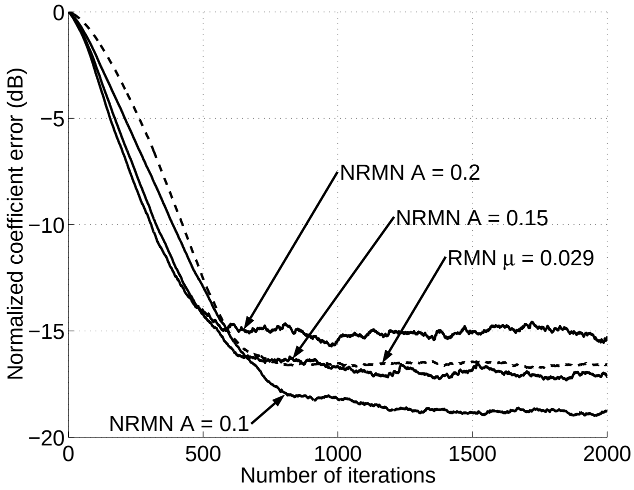

Fig. 4. Performance comparison of RMN and NRMN for varying A, under impulsive noise interference $n(k)$, for $u(k)$ of SNR $=$ 20dB and a filter of length $N = 100$ of random coefficients.

formance the proposed NRMN offers comes at practically no expense in computational complexity.

5. ACKNOWLEDGMENT

The second author would like to thank the reviewers for their comments and suggestions.

6. REFERENCES

[1] P. M. Clarkson and T. I. Haweel. Median LMS algorithm. *IEE Electronics Letters*, 25(8):520–522, Apr. 1989.

[2] G. A. Williamson, P. M. Clarkson, and W. A. Sethares. Performace Characteristics of the Median LMS Adaptive Filter. *IEEE Transactions on Signal Processing*, 41:667–680, 1993.

[3] J. Chambers and A. Avlonitis. A Robust Mixed-Norm Adaptive Filter Algorithm. *IEEE Signal Processing Letters*, 4(2):46–48, Feb. 1997.

[4] J. A. Chambers, O. Tanrikulu, and A. G. Constantinides. Least mean mixed-norm adaptive filtering. *IEE Electronics Letters*, 30(19):1574–1575, Sep. 1994.

[5] A. Papoulis. *Probability, Random Variables, and Stochastic Processes*. New York: McGraw-Hill, third edition, 1991.

[6] E. Soria-Olivas, J. Calpe-Maravilla, J. F. Guerrero-Martinez, M. Martinez-Sober, and J. Espi-Lopez. An Easy Demonstration of the Optimum Value of the Adaptation Constant in the LMS Algorithm. *IEEE Transactions on Education*, 41(1):81, Feb. 1998.

[7] V. J. Mathews and S. H. Cho. Improved Convergence Analysis of Stochastic Gradient Adaptive Filters Using the Sign Algorithm. *IEEE Transactions on Acoustics, Speech and Signal Processing*, ASSP–35(4):450–454, Apr. 1987.