

This item was submitted to [Loughborough's Research Repository](#) by the author.
Items in Figshare are protected by copyright, with all rights reserved, unless otherwise indicated.

A robust mixed-norm adaptive filter algorithm

PLEASE CITE THE PUBLISHED VERSION

PUBLISHER

© IEEE

VERSION

VoR (Version of Record)

LICENCE

CC BY-NC-ND 4.0

REPOSITORY RECORD

Chambers, Jonathon, and Apostolos Avlonitis. 2019. "A Robust Mixed-norm Adaptive Filter Algorithm".
figshare. <https://hdl.handle.net/2134/5849>.

This item was submitted to Loughborough's Institutional Repository (<https://dspace.lboro.ac.uk/>) by the author and is made available under the following Creative Commons Licence conditions.



CC creative commons
COMMONS DEED

Attribution-NonCommercial-NoDerivs 2.5

You are free:

- to copy, distribute, display, and perform the work

Under the following conditions:

 **Attribution.** You must attribute the work in the manner specified by the author or licensor.

 **Noncommercial.** You may not use this work for commercial purposes.

 **No Derivative Works.** You may not alter, transform, or build upon this work.

- For any reuse or distribution, you must make clear to others the license terms of this work.
- Any of these conditions can be waived if you get permission from the copyright holder.

Your fair use and other rights are in no way affected by the above.

This is a human-readable summary of the [Legal Code \(the full license\)](#).

[Disclaimer](#) 

For the full text of this licence, please go to:
<http://creativecommons.org/licenses/by-nc-nd/2.5/>

A Robust Mixed-Norm Adaptive Filter Algorithm

Jonathon Chambers, *Member, IEEE*, and Apostolos Avlonitis

Abstract—We propose a new member of the family of mixed-norm stochastic gradient adaptive filter algorithms for system identification applications based upon a convex function of the error norms that underlie the least mean square (LMS) and least absolute difference (LAD) algorithms. A scalar parameter controls the mixture and relates, approximately, to the probability that the instantaneous desired response of the adaptive filter does not contain significant impulsive noise. The parameter is calculated with the complementary error function and a robust estimate of the standard deviation of the desired response. The performance of the proposed algorithm is demonstrated in a system identification simulation with impulsive and Gaussian measurement noise.

I. INTRODUCTION

THE FAMILY of mixed-norm adaptive filters has recently been introduced to combine the benefits of established stochastic gradient adaptive filter algorithms [1]. One such algorithm combines the least mean square (LMS) and the least mean fourth (LMF) algorithms. Analysis of this algorithm has confirmed the advantages found in simulations [2].

In this work, a mixed-norm adaptive filter algorithm is proposed that is suitable for system identification applications in which the desired response is corrupted by additive impulsive, or heavy tailed distribution, noise. The system identification structure is shown in Fig. 1. An alternative algorithm which is suitable for this problem is the median LMS algorithm, but this differs in that a block of past gradient terms must be stored [3]. The common input, $x(k)$, to the unknown system and to the adaptive finite impulse response (FIR) filter is assumed to be zero-mean Gaussian noise. The unknown system is also an FIR filter of known order q . The output distribution of the unknown system is therefore zero mean Gaussian, with a fixed, but unknown, standard deviation. Impulsive measurement noise, $n(k)$, is added to the output of the unknown system, $y(k)$, to form the desired response, $d(k)$, with a model of the form used in [3], that is, $n(k) = a(k)A(k)$, where $a(k)$ is a binary independent identically distributed occurrence process with $p[a(k) = 1] = c$, $p[a(k) = 0] = 1 - c$, and c is the arrival probability; whereas $A(k)$ is a process, with symmetric amplitude distribution, which is uncorrelated with $a(k)$. The variance of $A(k)$ is chosen to be substantially greater than that of $y(k)$ to represent impulsive noise. The mean value of $n(k)$

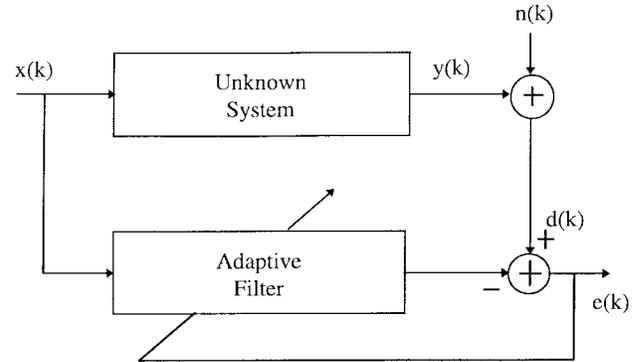


Fig. 1. System identification structure.

is zero and its variance is given by

$$\text{var}\{n(k)\} = c \text{var}\{A(k)\} = \sigma_n^2.$$

A robust estimate of the standard deviation of the Gaussian component of the desired response can be made through an outlier trimmed sliding window. In this window, the N_w most recent values of the desired response are ordered in terms of their amplitude, i.e., $\underline{d} = [d(k), d(k-1), \dots, d(k-N_w+1)]^t$ where $(\cdot)^t$ represents the vector transpose operation and \underline{d} = ord(\underline{d}) in which the elements of \underline{d} are algebraically ordered from smallest to largest, the minimum and maximum values of which are discounted by application of a diagonal trimming matrix, $\underline{T} = \text{Diag}[0, 1, 1, \dots, 1, 1, 0]$ that nullifies the first and last elements of \underline{d} ; and finally, the remainder are used to make a sample estimate of the standard deviation of the form

$$\hat{\sigma}_d(k) = \sqrt{\frac{1}{N_w - 3} \underline{d}^t \underline{T} \underline{d}}. \quad (1)$$

This estimate is next used to guide the mixture in the new robust mixed-norm (RMN) adaptation algorithm. This algorithm combines the conventional least mean square (LMS) and least absolute deviation (LAD), otherwise known as the sign-error LMS or pilot LMS [5], stochastic gradient algorithms. The motivation is that for identical initial convergence rates the LMS algorithm generally provides a more accurate final solution, less misadjustment, when there is no impulsive noise present, but is very sensitive to the presence of outliers; whereas, with fixed adaptation gain the LAD algorithm is less accurate, with higher misadjustment, but is more robust to the presence of outliers [4]. The algorithm is based on the minimization of the following convex combination of the error norms

$$J(k) = \lambda E\{e^2(k)\} + (1 - \lambda)E\{|e(k)|\} \quad (2)$$

Manuscript received April 22, 1996. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. J. J. Shynk.

The authors are with the Signal Processing and Digital Systems Section, Department of Electrical and Electronic Engineering, Imperial College of Science, Technology, and Medicine, London, SW7 2BT, U.K. (e-mail: j.chambers@is.ac.uk).

Publisher Item Identifier S 1070-9908(97)01698-2.

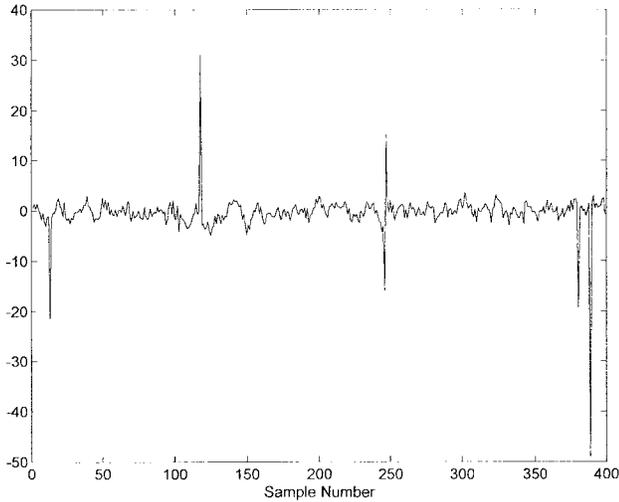


Fig. 2. Desired response signal with impulsive noise.

where $E\{\cdot\}$ is the statistical expectation operator and $\lambda \in [0, 1]$ is the mixing parameter. When $\lambda = 1$, (2) reverts to the error norm for the LMS algorithm, whereas for $\lambda = 0$, (2) becomes the error norm for the LAD algorithm. Careful choice of λ thereby provides an algorithm with intermediate performance between that of LMS and LAD, and a mechanism to mitigate the disturbance problem of outliers on the LMS algorithm. The error signal $e(k)$ is assumed to be related to the desired response signal $d(k)$, adaptive filter weight vector $\underline{w}(k)$, and input vector $\underline{x}(k)$ in the form $e(k) = d(k) - \underline{w}^t(k)\underline{x}(k)$ as is the convention in adaptive signal processing.

II. RMN ALGORITHM

The update equation for the robust mixed-norm (RMN) algorithm is derived from the following steepest descent type weight update equation

$$\underline{w}(k+1) = \underline{w}(k) - \mu \hat{\nabla}_{\underline{w}(k)} J(k) \quad (3)$$

where $\underline{w}(k)$ is the weight vector, μ is the fixed adaptation gain, and $\hat{\nabla}_{\underline{w}(k)} J(k)$ is the instantaneous estimate of the gradient of the error norm $J(k)$ evaluated at the current value of the weight vector $\underline{w}(k)$. Differentiation of (2) with respect to $\underline{w}(k)$ yields the RMN update equation

$$\underline{w}(k+1) = \underline{w}(k) + \mu \{ \lambda 2e(k) + (1 - \lambda) \text{sign}[e(k)] \} \underline{x}(k). \quad (4)$$

The convergence properties of this algorithm are controlled by the adaptation gain parameter μ and mixing parameter λ . An analysis based on [4], in which $E\{\text{sign}[e(k)]\underline{x}(k)\} = \sqrt{(2/\pi)} [1/\sigma_e(k)] E\{e(k)\underline{x}(k)\}$, for small μ , where $\sigma_e(k)$ is the standard deviation of the error sequence, yields the following sufficient conditions on μ for convergence of the mean

$$0 < \mu < \frac{2}{\left(2\lambda + [1 - \lambda] \sqrt{\frac{2}{\pi\sigma_n^2}}\right) N\sigma_x^2} \quad (5)$$

where σ_n^2 and σ_x^2 are, respectively, the measurement and input noise powers, and N is the length of the adaptive filter.

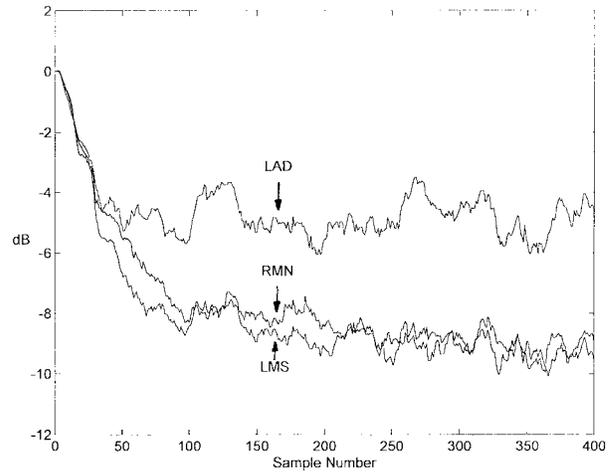


Fig. 3. Averaged log normalized weight error vector for LAD, LMS, and RMN algorithms.

The RMN algorithm, with fixed λ , requires at each iteration only two more additions than the LMS algorithm, and may be initialized with a null vector. To complete the definition of the RMN algorithm it is necessary to relate the outlier trimmed sliding window standard deviation estimate to the mixing parameter λ . The time-varying nature of the standard deviation estimate implies that the mixing parameter is a function of the sample index k . The selection of $\lambda(k)$ is based upon the following probability, where $d(k)$ has a symmetric distribution, and d_0 is positive:

$$\begin{aligned} \lambda(k) &\equiv \text{Prob} \{d(k) > d_0 \cup d(k) < -d_0\} \\ &= 2(1 - \text{Prob} \{d(k) \leq d_0\}) \\ &= 2 \left[1 - \int_{-\infty}^{d_0} f_D(x) dx \right] \\ &= 2 \left[\frac{1}{2} - \int_0^{d_0} f_D(x) dx \right] \end{aligned}$$

where $f_D(x)$ is the distribution of the desired response. Specializing this to the case that $d(k)$ has a zero-mean Gaussian distribution with standard deviation equal to the estimate in (1) yields

$$\begin{aligned} \lambda(k) &= 2 \left[\frac{1}{2} - \int_0^{|d(k)|/\hat{\sigma}_d} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \right] \\ &= 2 \text{erfc} \left[\frac{|d(k)|}{\hat{\sigma}_d} \right] \end{aligned} \quad (6)$$

where $\text{erfc}(\cdot)$ is the complementary error function. The probability that the instantaneous desired response contains significant impulsive noise is then approximated as $[1 - \lambda(k)]$. The underlying concept is that the LMS algorithm is progressively replaced by the LAD algorithm as the likelihood of an outlier increases. The value of this can be supported by considering the weight propagation equation for the RMN algorithm. With the definition that the weight error vector, $\underline{v}(k) \equiv \underline{w}(k) - \underline{w}_{opt}$,

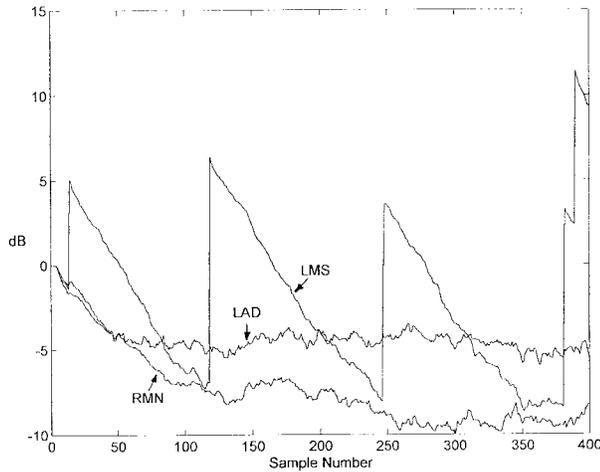


Fig. 4. Averaged log normalized weight error vector for LAD, LMS, and RMN algorithms in impulsive noise environment.

where \underline{w}_{opt} is the optimal Wiener solution, (4) becomes

$$\begin{aligned} \underline{v}(k+1) = & [I - 2\mu\lambda(k)\underline{x}(k)\underline{x}(k)^t]\underline{v}(k) \\ & + \{2\mu\lambda(k)n(k) \\ & + \mu[1 - \lambda(k)]\text{sgn}[e(k)]\}\underline{x}(k). \end{aligned} \quad (7)$$

If the adaptive filter is assumed to be close to the optimal solution at time sample k , i.e., $\underline{v}(k) \approx \underline{0}$, the arrival of a large impulsive noise component, $n(k)$, implies that the second term will dominate the right hand side of (7), so that

$$\underline{v}(k+1) \approx \{2\mu\lambda(k)n(k) + \mu[1 - \lambda(k)]\text{sgn}[n(k)]\}\underline{x}(k). \quad (8)$$

For the LMS case, i.e., $\lambda(k) = 1$, $E\{\underline{v}(k+1)^t\underline{v}(k+1)\} = 4\mu^2 N \sigma_n^2 \sigma_x^2$, whereas when the LAD algorithm is used, $\lambda(k) = 0$, $E\{\underline{v}(k+1)^t\underline{v}(k+1)\} = \mu^2 N \sigma_x^2$, and the weight error vector norm is independent of the impulsive noise statistics. Although the form of (6) is not well suited to real time implementation it could easily be replaced by a look-up table, or the calculation of $\lambda(k)$ could be restricted to those sample instances for which the ratio $|d(k)|/\hat{\sigma}_d$ is large.

III. SIMULATIONS

To demonstrate the performance of the RMN algorithm a system identification simulation similar to that reported in [1] is used. The desired response signal is formed by inputting white Gaussian distributed noise of unit power to a nine-tap FIR filter with coefficients [0.1, 0.2, 0.3, 0.4, 0.5, 0.4, 0.3, 0.2, 0.1] and, to test the algorithms severely, independent Gaussian distributed noise of fixed variance is added to its output so that the effective SNR is 0 dB prior to the addition of the impulsive noise. The impulsive noise is generated from the multiplicative model with $c = 10^{-2}$ and $\text{var}\{A(k)\} = 10^4/12$. The presence of the additional Gaussian measurement noise does not violate the assumptions in the derivation of the RMN algorithm, because it simply adds an additional independent Gaussian

component to the desired response of the adaptive filter, and is more representative of a real application. The desired response so produced is shown in Fig. 2. The length of the sliding window standard deviation estimator N_w is 10, so that the probability that more than one significant impulse lies within the window is very small. The adaptation gain for the LMS algorithm is set at 0.01, whereas for the RMN algorithm it is 0.018 and 0.04 for the LAD algorithm; chosen so that in simulation the initial convergence rates of the three algorithms were visually identical when no impulsive noise is present. This is demonstrated in Fig. 3, which shows the log normalized weight error vector norm as defined by

$$10^* \log_{10} \left[\frac{\|\underline{w}(k) - \underline{w}\|_2^2}{\|\underline{w}\|_2^2} \right]$$

averaged across ten independent trials. It is clear that the performance of the RMN algorithm is almost identical to that of LMS, both much improved on that of the LAD algorithm. Finally, in Fig. 4, the log normalized weight error vector norm is shown for simulations with the impulsive noise present and constant for each trial. It is evident how the behavior of the conventional LMS algorithm deteriorates, whereas that of the RMN algorithm is unaffected and remains much improved on than of the LAD algorithm.

IV. CONCLUSIONS

A new robust mixed-norm adaptive filter algorithm has been introduced. The combination of error norms yields an algorithm that is robust to the presence of significant impulsive noise in the desired response of the adaptive filter while maintaining good accuracy properties. This performance has been demonstrated in a system identification simulation. Current studies are addressing the extension of this work to adaptive filters applied to predictor structures in which both input and desired response are corrupted by impulsive noise.

ACKNOWLEDGMENT

The authors wish to thank the anonymous reviewers whose comments added much to the clarity of the presentation of this work.

REFERENCES

- [1] J. A. Chambers, O. Tanrikulu, and A. G. Constantinides, "Least mean mixed-norm adaptive filtering," *Electron. Lett.*, vol. 30, pp. 1574-1575, 1994.
- [2] O. Tanrikulu and J. A. Chambers, "Convergence and steady-state properties of the least-mean mixed norm (LMMN) adaptive algorithm," *IEE Proc. Vision, Image, Signal Processing*, vol. 143, pp. 137-142, June 1996.
- [3] G. A. Williamson, P. M. Clarkson, and W. A. Sethares, "Performance characteristics of the median LMS adaptive filter," *IEEE Trans. Signal Processing*, vol. 41, pp. 667-680, Feb. 1993.
- [4] V. J. Mathews and S. H. Cho, "Improved convergence analysis of stochastic gradient adaptive filters using the sign algorithm," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-35, no. 4, pp. 450-454, 1987.
- [5] O. Macchi, *Adaptive Processing: The Least Mean Squares Approach with Applications in Transmission*. New York: Wiley, 1995.