

## A Robust $M$ -estimate Adaptive Filter For Impulse Noise Suppression

Yuxian Zou, S. C. Chan and T. S. Ng  
Department of Electrical and Electronic Engineering  
The University of Hong Kong, Pokfulam Road, Hong Kong  
{yxzou,scchan,tsng}@eee.hku.hk

**Abstract:** In this paper, a robust  $M$ -estimate adaptive filter for impulse noise suppression is proposed. The objective function used is based on a robust  $M$ -estimate. It has the ability to ignore or down weight large signal error when certain thresholds are exceeded. A systematic method for estimating such thresholds is also proposed. An advantage of the proposed method is that its solution is governed by a system of linear equation. Therefore, fast adaptation algorithms for traditional linear adaptive filters can be applied. In particular, a  $M$ -estimate recursive least square (M-RLS) adaptive algorithm is studied in detail. Simulation results show that it is more robust against individual and consecutive impulse noise than the MN-LMS and the N-RLS algorithms. It also has fast convergence speed and a low steady state error similar to its RLS counterpart.

### I. Introduction

Recently, there are considerable interests in studying adaptive filtering algorithms that are robust to impulse noise and interferences. Such impulse noise can be due to natural or man-made electromagnetic sources [1]. Under this adverse condition, the performance of the linear adaptive filters can deteriorate significantly. Nonlinear techniques are often employed to reduce these adverse effects of the impulse noise. For example, median filtering has been applied in the LMS algorithm to protect the filter weights from the effects of impulsive noise [11]. Another class of nonlinear technique is to smooth out the momentary fluctuation of the error signal in conventional adaptive filters by means of some nonlinear clipping functions. These include the nonlinear LMS algorithm (N-LMS) and the nonlinear recursive least squares algorithm (N-RLS) proposed in [6] and [7], respectively. A mixed-norm adaptive filter using a combination of  $L_1$  and  $L_2$  norm as the objective function has also been proposed recently [2,3,4]. Using the stochastic gradient method, an algorithm similar to a combination of the Least Mean Square (LMS) and the Least Absolute Difference (LAD) algorithms, called the mixed-norm LMS algorithm (MN-LMS), is obtained. The method is further extended to study the system identification problem in the presence of impulsive noise or noise with heavy tailed distribution. This algorithm is robust against the impulse noise occurred in the desired signal [2]. However, due to the LMS nature of the algorithm, it usually suffers from slow convergence speed when the input is coloured.

In this paper, a new adaptive filter using the robust  $M$ -estimate as the objective function is proposed.  $M$ -estimate, like the median, belongs to the general class of robust statistical

estimates, which are designed to perform robust estimation under model mismatch or the presence of outliers. A block-based adaptive filter based on the Huber  $M$ -estimate objective function has previously been proposed in [10]. The Huber measure uses the  $L_2$  norm when the signal error is smaller than a certain threshold and the  $L_1$  norm when the error signal is large. The adaptive filter down-weights the outliers and behaves like a least squares filter on impulse free condition. The  $M$ -estimate proposed in this paper is differed from the Huber estimate in that the error will be completely ignored if the signal error is larger than certain threshold. A systematic method for estimating such thresholds is also proposed. The solution is governed by a system of linear equation similar to the conventional normal equation. Using this  $M$ -estimate normal equation, it is possible to derive different fast adaptation algorithms as in the traditional linear adaptive filters. In particular, a  $M$ -estimate recursive least square (M-RLS) adaptive algorithm is proposed. Simulation results show that the proposed M-RLS algorithm in this paper is more robust against individual and consecutive impulse noise than the MN-LMS and the N-RLS algorithms. It also has a low steady state error similar to the RLS algorithm.

### II. Overview

Fig.1 shows the block diagram of a FIR adaptive filter being used in a system identification setting.  $x(n)$  and  $y(n)$  are, respectively, the input and output of the unknown system. The output of the adaptive filter is given by

$$\hat{y}(n) = \mathbf{w}^T(n)\mathbf{X}(n). \quad (1)$$

where,  $\mathbf{w}(n) = [w_0(n), \dots, w_{L-1}(n)]^T$ , and  $\mathbf{x}(n) = [x(n), \dots, x(n-L+1)]^T$  are the weight and signal vectors, respectively.  $d(n)$  is the desired signal and is assumed to be corrupted with additive Gaussian noise, and probably impulse noise. The problem is to identify the coefficients of the unknown system by minimizing certain objective function or distortion measure. In the conventional least squares adaptive filter, the exponentially weighted square error is used as the cost function,

$$J(n) = \sum_{i=1}^n \lambda^{n-i} e(i)e^*(i) \quad (2)$$

where,  $0 < \lambda \leq 1$  is the forgetting factor, and  $e(n) = d(n) - \hat{y}(n)$  is the instantaneous error. Differentiating (2) with respect to  $\mathbf{w}(n)$  and setting the derivative to zero, one gets the normal equation,

$$\mathbf{R}(n)\hat{\mathbf{w}}(n) = \mathbf{r}(n), \quad (3)$$

where,  $\mathbf{R}(n) = \sum_{i=1}^n \lambda^{n-i} \mathbf{X}(i) \mathbf{X}^T(i)$  is the autocorrelation matrix of  $\mathbf{X}(n)$  and  $\mathbf{r}(n) = \sum_{i=1}^n \lambda^{n-i} d(i) \mathbf{X}(i)$  is the cross-correlation vector.

Well-known techniques for solving (3) include the recursively least square (RLS) algorithm and its various extensions [5]. In practical situations,  $y(n)$  or  $x(n)$  may be corrupted by additive noise whose nature is impulsive. In this case,  $\mathbf{R}(n)$  and/or  $\mathbf{r}(n)$ ,  $e(n)$  and hence  $\hat{\mathbf{w}}(n)$ , will exhibit momentary fluctuation which might take many iterations to recover, affecting the convergence performance of the adaptive filter.

In the non-linear recursive least square algorithm (N-RLS) [7], a non-linear clipping function,  $f_c(\cdot)$ , is applied to  $e(n)$  to reduce its influence if  $e(n)$  is large. More specifically, the coefficient update equation is,

$$\hat{\mathbf{w}}(n) = \hat{\mathbf{w}}(n-1) + f_c(e(n)) \mathbf{K}(n), \quad (4)$$

where  $f_c(e(n)) = \begin{cases} e(n) & |e(n)| \leq h \\ h & |e(n)| > h \end{cases}$  and  $\mathbf{K}(n)$  is the Kalman gain vector. The parameter,  $h$ , is estimated from the variance as follows,

$$h = 2.24 \hat{\sigma}_e(n), \\ \hat{\sigma}_e^2(n) = \lambda_r \hat{\sigma}_e^2(n-1) + (1 - \lambda_r) e(n) e^*(n). \quad (5)$$

In the mixed-norm LMS (MN-LMS) algorithm, the weight vector is updated to minimize a combination of the  $L_1$  and  $L_2$  norm of the error [2],

$$\hat{\mathbf{w}}(n) = \hat{\mathbf{w}}(n-1) + \mu \{ \lambda(n) e(n) + (1 - \lambda(n)) \text{sign}(e(n)) \} \mathbf{X}(n) \quad (6)$$

where,  $\mu$  is the step-size and  $\lambda(n)$  is a mixing parameter that determines the relative importance of the two distortion measures. In the presence of the impulsive noise,  $e(n)$  will be very large and the weighting of the  $L_2$  norm will be decreased to minimize the adverse effects of the noise pulses. On the other hand, the weighting of the  $L_1$  norm will dominate in the noise free case to reduce the steady state error. Assuming that  $e(n)$  is Gaussian,  $\lambda(n)$  can be estimated as [2],

$$\lambda(n) = 2 \text{erfc}(|e(n)| / \hat{\sigma}_e(n)). \quad (7)$$

where  $\text{erfc}(\cdot)$  is the complementary error function. Here, the robust estimate of the error signal variance  $\hat{\sigma}_e(n)$  can be calculated as  $\hat{\sigma}_e(n) = \sqrt{\mathbf{O}^T(n) \mathbf{A} \mathbf{O}(n) / (N_o - 3)}$ , where  $\mathbf{A} = \text{diag}(0, 1, 1, \dots, 1, 0)$  is the diagonal trimming matrix,  $\mathbf{O}(n) = \text{sort}[e(n), \dots, e(n - N_o + 1)]$ , and  $\text{sort}[\dots]$  is the operator that rearranges the components in the square bracket in ascending order. The limitations of this algorithm are the slow convergence speed of the LMS-type algorithm and the increased steady-error, due to the use of the mixed-norm.

### III. Robust $M$ -estimate RLS Algorithm

The  $M$ -estimator proposed in this paper is given by,

$$\rho(e(i)) = \begin{cases} e(i)^2/2 & |e(i)| \leq \xi \\ \xi |e(i)| - \xi^2/2 & \xi < |e(i)| < \Delta \\ \Delta^2/2 & |e(i)| \geq \Delta \end{cases}, \quad (8)$$

The objective function is,

$$J(n) = \sum_{i=1}^n \lambda^{n-i} \rho(e(i)). \quad (9)$$

As shown in Fig. 2,  $\rho(\cdot)$  is an even real-valued function. It is quadratic when  $|e(i)|$  is smaller than  $\xi$ . For values of  $|e(i)|$  in the interval  $[\xi, \Delta]$ , the function is linear. For values of  $|e(i)|$  greater than  $\Delta$ , the function is equal to a constant,  $c = \Delta^2/2$ . It becomes apparent that the  $M$ -estimator is capable of suppressing outliers with large amplitude. Parameters  $\xi$  and  $\Delta$  control the degree of suppression of the outliers. They are usually chosen according to the applications or estimated continuously.  $J(n)$ , as defined as (9), helps to smooth out momentary fluctuation due to impulsive interferences and model mismatches. Using (8), (9) can also be rewritten as,

$$J(n) = \sum_{\xi < |e(i)| \leq \Delta} \lambda^{n-i} \{ \xi \text{sgn}[e(i)] e(i) - \xi^2/2 \} \\ + \frac{1}{2} \sum_{|e(i)| \leq \xi} \lambda^{n-i} e(i) e^*(i) + \sum_{|e(i)| > \Delta} \lambda^{n-i} c \quad (10)$$

Due to the nonlinear function, (10) will have multiple solutions but they are close to each other. Differentiating both sides of (10) and setting  $\partial J(n) / \partial w_k(n)$  to zero, one gets the following,

$$\sum_{k=1}^L \tilde{w}_k(n) \cdot \sum_{\substack{i=1 \\ |e(i)| \leq \xi}} \lambda^{n-i} x(n-k) x(n-i), \\ = \xi \sum_{\xi < |e(i)| \leq \Delta} \lambda^{n-i} \text{sgn}[e(i)] x(n-i) + \sum_{|e(i)| \geq \Delta} \lambda^{n-i} d(n) x(n-i) \quad (11)$$

or in matrix form,

$$\mathbf{R}_N(n) \tilde{\mathbf{w}}(n) = \mathbf{r}_N(n). \quad (12)$$

(12), which is referred to as the  $M$ -estimate normal equation here, can be viewed as the counterpart of the normal equation (3). Since (12) is linear, traditional adaptive filtering algorithms such as the recursive least square algorithm (RLS) can be used to solve (12) by properly updating  $\mathbf{R}_N(n)$  and  $\mathbf{r}_N(n)$ . Depending on the values of  $e(i)$ , there are three different cases for updating  $\mathbf{R}_N(n)$  and  $\mathbf{r}_N(n)$ , as suggested by (11).

**Case 1:**  $|e(n)| \leq \xi$ . In this case, the error function is quadratic and the update equation is similar to the conventional RLS algorithm.

$$\mathbf{R}_N(n) = \lambda \mathbf{R}_N(n-1) + \mathbf{X}(n) \mathbf{X}^T(n) \quad (13)$$

$$\mathbf{r}_N(n) = \lambda \mathbf{r}_N(n-1) + d(n) \mathbf{X}(n) \quad (14)$$

It is natural to use fast adaptive algorithms to update the filter weight recursively. Here, the conventional RLSA is adopted for its fast convergence speed.

**Case 2:**  $\xi < |e(n)| \leq \Delta$ . Here, the  $L_1$  norm of the error is used.  $\mathbf{R}_N(n)$  remains unchanged,  $\mathbf{R}_N(n) = \mathbf{R}_N(n-1)$ , and the cross correlation vector is updated as follows,

$$\mathbf{r}_N(n) = \mathbf{r}_N(n-1) + \xi \operatorname{sgn}(e(n))\mathbf{X}(n) \quad (15)$$

Since  $\mathbf{R}_N(n)$  is unchanged, the inverse in the previous iteration can be used to update the new weight vector,

$$\tilde{\mathbf{w}}(n) = \mathbf{R}_N^{-1}(n-1)\mathbf{r}_N(n). \quad (16)$$

**Case 3:**  $|e(n)| > \Delta$ . In this case, the error will be completely ignored and updating is not necessary.

$$\mathbf{R}_N(n) = \mathbf{R}_N(n-1), \mathbf{r}_N(n) = \mathbf{r}_N(n-1), \tilde{\mathbf{w}}(n) = \tilde{\mathbf{w}}(n-1). \quad (17)$$

The remaining problem is to estimate the parameters  $\xi$  and  $\Delta$ . We first compute the probability of  $e(n)$  greater than a given threshold  $T$ . Though the distribution of  $e(n)$  is unknown, it is assumed to be Gaussian distributed when there is no outlier, so that,

$$\theta_T(n) = \Pr\{|e(n)| > T\} = 1 - \operatorname{erf}\left(\frac{T}{\sqrt{2}\hat{\sigma}_e}\right), \quad (18)$$

where  $\operatorname{erf}(\cdot)$  is the error function, and  $\hat{\sigma}_e^2(n)$  is the estimated variance. The probability of  $e(n)$  greater than  $\xi$  and  $\Delta$  are therefore  $\theta_\xi$  and  $\theta_\Delta$ , respectively. By appropriate choice of  $\theta_\xi$  and  $\theta_\Delta$ , the values of  $\xi$  and  $\Delta$  can be determined. In this work,  $\theta_\xi$  and  $\theta_\Delta$  are chosen as 0.2 and 0.1, respectively, so that we have 80% confident to down weight the error in the interval  $[\xi, \Delta]$  and 90% confident to reject it completely when  $e(n) > \Delta$ . The traditional estimate of  $\hat{\sigma}_e^2$  can be computed by the second equation in (5), which is however not robust in impulse noise environment because the impulses can increase the variance of the error signal and hence the value of  $\xi$  and  $\Delta$ . Therefore the impulsive noise cannot be removed by the nonlinear function,  $\rho(\cdot)$ . Here,  $\hat{\sigma}_e^2$  is estimated by [11]

$$\hat{\sigma}_e^2 = T_{MAD}(n) = 1.483 * \operatorname{med}\left\{e^2(i) - \operatorname{med}_{i \neq j}(e^2(j))\right\}, \quad (19)$$

$$i, j = n, \dots, n - N_\sigma + 1,$$

where  $T_{MAD}(n)$  is the median absolute deviation from the median (MAD) of the block data in the data window with length  $N_\sigma$  and  $\operatorname{med}(\cdot)$  is the median operator.

## VI. Simulation Results

In order to evaluate the performance of the proposed M-RLS algorithm, simulation is performed on the system identification problem as shown in Fig. 1. The unknown system is modelled as a FIR filter with impulsive response [0.2, -0.4, 0.6, -0.8, 1, -0.8, 0.6, -0.4, 0.2]. The adaptive filter is assumed to have the same length as the unknown system, i.e.  $L = 9$ . The window length  $N_\sigma$  is chosen as  $L$  in our simulation. The input signal is a coloured signal generated by passing a zero-mean white Gaussian process through a linear time-invariant

filter with coefficients  $f_k = .5 * [1 + \cos(2\pi(k-2)/w_f)]$ ,  $k = 1, 2, 3$  [5].  $w_f$  is chosen to be 3.5 and the eigenvalues spread of the correlation matrix  $\mathbf{R}$  is approximately 46.8. The additive noise  $n(n)$  is assumed to be Gaussian with zero-mean and variance  $\sigma_n^2$ . The signal to noise ratio at the system output is given by  $SNR = 20 \log_{10}(\sigma_y^2 / \sigma_n^2)$ , where  $\sigma_y^2$  is the variance of the output of the unknown system. The impulses are generated from the same multiplication model proposed in [2]. The initial weights of the adaptive filter are set to zeros and the SNR is 50dB. The normalised square norm of the weight error vector, (NSWE), is given by,

$$NSWE = 10 \log_{10} \left( \frac{\sum_{i=1}^L |\hat{w}_i(n) - w_i^*|^2}{\sum_{i=1}^L |w_i^*|^2} \right), \quad (20)$$

which is used as a measure of the convergence performance. Here,  $w_i^*$  is the  $i$ th coefficient of the unknown system and  $\hat{w}_i(n)$  is its estimate at time instant  $n$ . In our experiments, the NSWE is obtained by averaging 20 independent runs.

### Example 1

This simulation is set up to compare the convergence performance of the M-RLS, RLS, N-RLS and MN-LMS algorithms. The impulses are generated with arrival probability  $par = 10^{-2}$  and variance  $A(n) = 200$ . The desired signal  $d(n)$  is shown in Fig.3a where impulses appear at  $n = 150, 250, 330, 420$ .  $\lambda$  and  $\lambda_e$  are chosen to be 0.99. For N-RLS, the scalar parameter,  $h$ , is estimated as in (5). For MN-LMS, the step-size,  $\mu$ , is set to 0.025 and the mixing parameter,  $\lambda(n)$ , is estimated as in (7). The NSWE results are plotted in Fig.4. It can be seen that for  $n = 1:149$ , the convergence of the M-RLS, N-RLS and RLS are identical with the same adaptation process. When impulsive noise is present in the desired signal, the convergence of the RLS is significantly affected, showing its sensitivity in impulse noise environment. The MN-LMS is able to suppress the impulses but its convergence speed is much slower than the other three algorithms. Both the M-RLS and the N-RLS algorithms are able to robustly identify the unknown system with comparable convergence performance.

### Example 2

This simulation is specifically performed to investigate the performance of these algorithms when successive impulses appear in the desired signal for the same system identification problem. The impulses are generated with arrival probability  $par = 2 * 10^{-2}$  and variance  $A(n) = 200$ . The desired signal  $d(n)$  is plotted in Fig.3b where impulses appear at  $n = 250, 330, 420$  and  $n = 150, 151, 153, 155, 156$ . All other parameters are identical to example 1. The NSWE results are plotted in Fig.5. Comparing Fig.5 and Fig.4, it can be concluded that

successive impulses have nearly no influence on the convergence performance of the M-RLS. On the other hand, the convergence performance of N-RLS is degraded even though it can suppress individual impulses occurred at  $n = 250, 330, 420$ . The performance of the MN-LMS is slightly impaired. It can be concluded that the proposed M-RLS algorithm can provide robust identification of the unknown system even in the presence of consecutive impulses. It also shares the fast convergence speed and the low steady-state error of RLS-type algorithms.

### V. Conclusion

In this paper, a robust  $M$ -estimate adaptive filter for impulse noise suppression is presented. The objective function used is based on a robust  $M$ -estimate. It has the ability to ignore or down weight large signal error when certain thresholds are exceeded. A systematic method for estimating such thresholds is also proposed. An advantage of the proposed method is that its solution is governed by a system of linear equation. Therefore, fast adaptation algorithms for traditional linear adaptive filters can be applied. In particular, a  $M$ -estimate recursive least square (M-RLS) adaptive algorithm is studied in detail. Simulation studies demonstrate that the proposed algorithm can provide robust identification of the unknown system and fast convergence speed in individual and consecutive impulse noise environment.

### References

- [1] K. L. Blackard, etc. "Measurements and models of radio frequency impulsive noise for indoor wireless communications", IEEE J. Select. Areas Com., Vol. SAC-11, No. 7 pp. 991-1001, Sept. 1993.
- [2] J. A. Chambers, A. Avlonitis, "A Robust Mixed-Norm (RMN) Adaptive Filter Algorithm", IEEE Signal Processing Letters, Vol. 4, No.2, pp. 46-48, Feb. 1997.
- [3] J. A. Chambers, etc. "Least mean mixed-norm adaptive filtering", Electronics Letters, Vol. 30, pp. 1574-1575, 1994.
- [4] O. Tanrikulu, J. A. Chambers, "Convergence and Steady-State Properties of the Least-Mean Mixed Norm (LMMN) Adaptive Algorithm", IEE Proc. Vision, Image and Signal Processing, Vol. 143(3), pp. 137-142, June 1996.
- [5] Simon Haykin, "Adaptive Filter Theory", Prentice-Hall, 1991.
- [6] S. Koile, "Adaptive threshold nonlinear algorithm for adaptive filters with robustness against impulsive noise" Proc. of ICASSP'96, Atlanta, pp. 1644-1647.
- [7] S. H. Leung, J. F. Weng, "Adaptive Nonlinear RLS Algorithm for Robust Filtering in Impulse Noise" IEEE ISCAS'97, pp. 2337-2340, June 9-12, 1997, Hong Kong.
- [8] P. J. Huber, "Robust Statistics", John Wiley, New York, 1981.
- [9] R. Frank, E. Hampel, M. Ronchetti and W.A. Stahel, Robust Statistics: the approach based on influence functions, John Wiley & Sons, Inc. 1986.
- [10] L. Mili, M.G. Cheniae, P.J. Rousseeuw, "Robust state estimation based on projection statistics", IEEE. Trans. On Power Systems, Vol. 11, No. 2, pp 1118-1126, May 1996
- [11] I.Pitas and A.N. Venetsanopoulos, "Nonlinear digital filters principles and applications", Kluwer Academic Publisher, London.

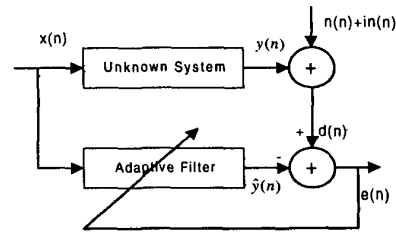


Fig.1 System Identification Structure

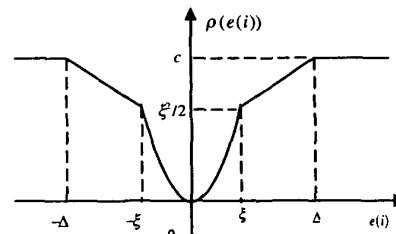


Fig.2 Modified M-estimator function

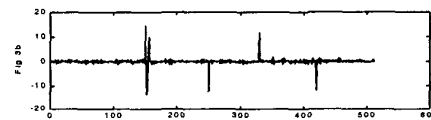
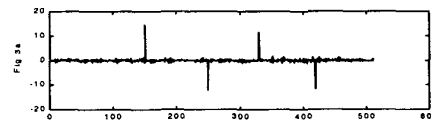


Fig.3 The input signal to unknown system

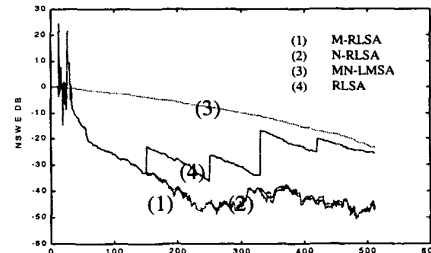


Fig.4 Normalized norm, squared weight error vector for experiment 1

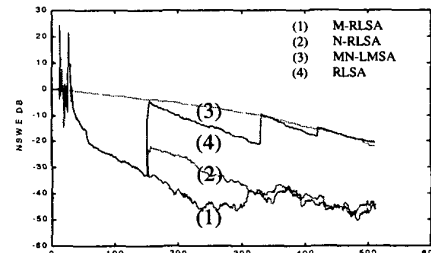


Fig.5 Normalized norm, squared weight error vector for experiment 2