Convergence Analysis of The Recursive Least M-Estimate Adaptive Filtering Algorithm For Impulse Noise Suppression

Shing-Chow Chan [1] and Yue-Xian Zou [2]

scchan@eee.hku.hk, Tel: (852) 2859 8025 Fax: (852) 2559 8738

[1]Department of Electrical and Electronic Engineering, the University of Hong Kong, Hong Kong

[2]Department of Electrical and Electronic Engineering, Singapore Polytechnic, Singapore

Abstract

In this paper, we present the convergence analysis of the recursive least M-estimate (RLM) adaptive filter algorithm, which was recently proposed for robust adaptive filtering in impulse noise environment. The mean and mean squares behaviors of the RLM algorithm, based on the modified Huber M-estimate function (MHF), in contaminated Gaussian (CG) noise model are analyzed. Close-form expressions are derived. The simulation and theoretical results agree very well with each other and suggest that the RLM algorithm is more robust than the RLS algorithm under the CG noise model.

2. Introduction

Robust adaptive filtering in impulse noise environment has received considerable attention recently [2-6], due to its practical importance in communications and other applications. Since the distributions of the impulsive noise differ significantly from that of Gaussian noise, the performance of the conventional adaptive filtering algorithms degrade severely in impulse noise environment. To overcome this problem, the authors have proposed a RLS-like algorithm, called the recursive least M-estimate (RLM) algorithm, for impulse noise suppression [1]. It minimizes a new cost function based on robust M-estimators, instead of the conventional least squares (LS) estimator [1, 7, 8]. Simulation results show that the RLM algorithm is more robust than the conventional RLS [9], N-RLS [5], and OSFKF [6] algorithms when the input and desired signals are corrupted by CG noise or alpha-stable distributed noise [1, 7, 8]. In addition, a fast implementation of the RLM algorithm with a computational complexity of order $O(N)$ was recently reported in [16], where N is the length of the adaptive transversal filter.

In this paper, the convergence of the RLM algorithm under CG noise is analyzed. Both mean- and mean-square behaviors of the RLM algorithm under CG noise are examined. Simulations are also carried out to evaluate the analytical results. It is found that the theoretical and simulation results for the convergence behaviors of the RLM algorithm agree very well with each other.

Specifically, let's consider the system identification problem shown in Fig. 1 where $x(n)$ and $y(n)$ are respectively the input and output of the adaptive filter. The estimation error is $e(n)=d(n)-w^T(n)X(n)$, where $X(n)=[x(n),...,x(n-N+1)]^T$, $d(n)$ is the desired signal and $w(n)$ is the weight vector. Instead of the *LS* cost function, the *M*-estimate cost function $J_\rho(n) \triangleq \sum_{i=1}^{n} \rho(e(i))$ is adopted [1], where $\lambda>0$ is the forgetting factor, $e(i)$ is the estimation error and $\rho(\cdot)$ is an M-estimate function. In this paper, the modified Huber M-estimate function (MHF) is used due to its good performance and simplicity [1]:

$$\rho(e)=\begin{cases} e^2/2, & 0<|e|<\xi \\ \xi^2/2 & otherwise \end{cases}, \quad \left(q(e)=\begin{cases} 1 & 0<|e|<\xi \\ 0 & otherwise \end{cases} \right), \tag{1}$$

where ξ is the threshold parameter and $q(e)=(\partial\rho(e)/\partial e)/e$ is the weighting function of $\rho(e)$. It is found that the performance of the RLM algorithm using the MHF and the Hampel's three part redescending function are very close to each other under CG noise and alpha-stable distributed noise environment [8]. By minimizing $J_\rho(n)$, the following RLM algorithm can be obtained [1]:

$$V(n)=\lambda^{-1}\{I-K(n)X^T(n)\}V(n-1), \tag{2}$$

$$K(n)=\frac{q(e(n))V(n-1)X(n)}{\lambda+q(e(n))X^T(n)V(n-1)X(n)}, \tag{3}$$

$$w(n)=w(n-1)+(d(n)-X^T(n)w(n-1))K(n)=w(n-1)+e(n)K(n), \tag{4}$$

where $V(n)$ and $K(n)$ are the inverse matrix of the *M*-estimate autocorrelation matrix $R_{X_\rho}=\sum_{i=1}^{n}\lambda^{n-i}q(e(i))X(i)X^T(i)$ and gain vector, respectively. Interested readers are referred to [1,7,8] for details.

2. Mean and Mean Square convergence Analysis

Our analysis makes use of the following assumptions:

Ass 1): The input signal $x(n)$ is ergodic with zero mean.

Ass 2): The interference noise $\eta_o(n)$ is modeled as a CG noise, which is a frequently used model for analyzing impulsive interference [2, 3]. More specifically, a CG noise is given by

$$\eta_o(n)=\eta_g(n)+\eta_{im}(n)=\eta_g(n)+b(n)\eta_w(n), \tag{5}$$

where $\eta_g(n)$ and $\eta_w(n)$ are independent identically distributed (i.i.d.) and zero mean Gaussian noise with variance σ_g^2 and σ_w^2, respectively; $b(n)$ is a sequence of ones and zeros, which is modeled as i.i.d Bernoulli random process with probability $P_r(b(n)=1)=p_r$ and $P_r(b(n)=0)=1-p_r$. The ratio $r_{im}=p_r\sigma_w^2/\sigma_g^2$ is used to determine the impulsive characteristic of $\eta_o(n)$. The probability density function (pdf) of $\eta_o(n)$ in (5) is given by

$$f_{\eta_o}(\eta)=\frac{1-p_r}{(2\pi\sigma_g^2)^{1/2}}\exp(-\frac{\eta^2}{2\sigma_g^2})+\frac{p_r}{(2\pi\sigma_\Sigma^2)^{1/2}}\exp(-\frac{\eta^2}{2\sigma_\Sigma^2}), \tag{6}$$

where $\sigma_\Sigma^2=\sigma_g^2+\sigma_{im}^2=\sigma_g^2+p_r\sigma_w^2$ is the variance of $\eta_o(n)$.

Ass 3): $x(n)$ and $e(n)$ have a jointly Gaussian distribution. $w(n)$, $x(n)$ and $\eta_o(n)$ are statistically independent. Although this assumption might not be valid in general applications, it is commonly used to simplify the convergence analysis of a lot of adaptive filtering algorithms [2, 3].

Ass 4): $d_o(n)=(w^*)^T X(n)$, where w^* is the optimal Wiener solution given by $w^*=R_{E,X}^{-1}P_{E,Xd}$. $R_{E,X} \triangleq E[X(n)X^T(n)]$ is the ensemble averaged correlation matrix of $X(n)$, and $P_{E,dX} \triangleq E[d(n)X(n)]$ is the ensemble averaged cross-correlation vector between $X(n)$ and $d(n)$. The subscript E indicates the averaged is taken over the ensemble.

Ass 5): $R_{E,X}=E[X(n)X^T(n)]\approx R_{X_\rho}(n)/\Lambda(n)$ when n is large enough, where $\Lambda(n)=\sum_{i=1}^{n}\lambda^{n-i}$ is the normalization factor and $R_{X_\rho}(n)=\sum_{i=1}^{n}\lambda^{n-i}q(e(i))X(i)X^T(i)$. Note that when $X(i)$ is ergodic, the matrix $R_{E,X}$ can be approximated by $R_{E,X}(n)\approx R_X(n)/\Lambda(n)$ [5], where $R_X(n)=\sum_{i=1}^{n}\lambda^{n-i}X(i)X^T(i)$ is the time averaged value [9]. Whereas, when there are impulse noise, the estimated $R_{X_\rho}(n)$ is a small biased approximation to $\Lambda(n)R_{E,X}(n)$ due to the use of $\rho(\cdot)$ and $\lambda_i^{(n)}$. An experiment was carried out to evaluate the effect of using $R_{X_\rho}(n)$ when the desired signal is corrupted by a CG noise with $p_r=0.01$ to 0.1, $\xi=2.24\hat{\sigma}(n)$, and 2048 data length. It was found that the value $\|R_{X_\rho}-\Lambda(n)R_{E,X}\|_2/\|R_{E,X}\|_2$ averaged over 100 independent runs, is within 1.16% to 1.97%. This shows that the proposed estimator for $R_{X_\rho}(n)$ here is robust to different impulse noise density and **Ass 5)** is justified. Further theoretic results

0-7803-7503-3/02/$17.00 ©2002 IEEE

regarding the estimation error of $R_{X_p}(n)$ can be found in [11].

Let's define the weight–error vector as $v(n) \triangleq w^* - w(n)$. Substituting $V(n-1) = R_{X_p}^{-1}(n-1) = \Lambda^{-1}(n-1)R_{\xi X}^{-1}$, $v(n)$, $q(e(n))$, $\bar{\lambda} = \lambda\Lambda(n-1)$, and $w^* = R_{\xi X}^{-1}P_{\xi X d}$ into (4), we get

$$v(n) \approx v(n-1) - R_{\xi X}^{-1}X(n)q(e(n))e(n)/(\bar{\lambda} + X^T(n)R_{\xi X}^{-1}X(n)). \quad (7)$$

2.1 Mean behavior of the RLM Algorithm

Taking expectation over $\{v, X, \eta_o\}$ on (7), one gets

$$E[v(n)] = E[v(n-1)] - R_{\xi X}^{-1}L_1, \quad (8)$$

where $E[\cdot]$ on the right hand side of (8) denotes the expectation over $\{v(n-1), X(n), \eta_o(n)\}$, and is denoted as $E_{\{v,X,\eta_o\}}\{\cdot\}$. Dropping the time index from all variables and using the assumption of $\eta_o(n)$, $w(n)$ and $x(n)$ in **Ass 3**), one gets

$$L_1 = E[X(n)q(e(n))e(n)/(\bar{\lambda} + X^T(n)R_{\xi X}^{-1}X(n))] = E_{\{v\}}[A], \quad (9)$$

where $A \triangleq E_{\{X,\eta_o\}}[Xq(e)e/(\bar{\lambda} + X^T R_{\xi X}^{-1}X)\,|\,v]$. The detailed evaluation of A is given in Appendix A. From equation (33), we have

$$A = \{C_B \iint_{N \text{fold}} A_0(v, X)\exp(-\tfrac{1}{2}X^T B^{-1}X)dX/2\}I_1 R_{\xi X}v(n-1)/2, \quad (10)$$

where $B = R_{\xi X}/(1+2\beta)$, $C_B = (2\pi)^{-N/2}|B|^{-1/2}$, I_1 and $A_0(v, X)$ are given in (34) and (30), respectively. Note that $A_0(v, X)$ is bounded by $0 < A_0(v, X) \le 2$, but it is rather difficult to obtain a closed-form expression for the integral inside the $\{\cdot\}$ in (10). An approximation of $A_0(v, X)$ is therefore sought to simplify (10). First, let's consider the evaluation of $\xi = k_\xi\hat{\sigma}(n)$ in (32). In the RLM algorithm, the impulsive component $\eta_w(n)$ of $\eta_o(n)$ is likely to be rejected and the corresponding "impulse-free" estimation error can be approximated by $\hat{e}(n) \approx \eta_g(n) + v^T(n-1)X(n)$. Dropping the time index and using the condition that v, η_o and X are uncorrelated, one gets

$$\xi = k_\xi\{E[\eta_g^2] + E[(v^T X)^2]\}^{1/2} > k_\xi\{E[(v^T X)^2]\}^{1/2}. \quad (11)$$

If $k_\xi >> 1$, ξ will be much larger than $\{E[(v^T X)^2]\}^{1/2}$. Furthermore, if the probability density function of $v^T X$ decays sufficiently fast, it follows that $P_r(\xi >> |v^T X|) \approx 1$, which is a consequence of the condition $\xi >> \{E[(v^T X)^2]\}^{1/2}$ and the Chernoff Bound ([12] pp.54), though the latter bound is rather loose. This observation together with the property of the error function and the rapidly decaying exponential function $\exp(-\tfrac{1}{2}X^T B^{-1}X)$ in (10) allows us to use the following approximation of $A_0(v, X)$

$$A_0(v, X) \approx A_c = 2\{(1-p_r)erf(\xi/(\sqrt{2}\sigma_g)) + p_r erf(\xi/(\sqrt{2}\sigma_\Sigma))\} \quad (12)$$

(c.f.(30)) during the evaluation of the expectation. Since $\sigma_\Sigma > \sigma_g$, A_c in (12) is bounded by $0 < A_c \le 1$. From simulation results, it was found that A_c in (12) is a very good approximation to $A_0(v, X)$ under the specified simulation conditions. Inserting (12) into (10) yields

$$A \approx A_c I_1 R_{\xi X}v(n-1). \quad (13)$$

Substituting (13), (9), and (34) into (8), the following relation between $E[v(n)]$ and $E[v(n-1)]$ is obtained

$$E[v(n)] \approx \{1 - A_c\exp(\bar{\lambda}/2)E_{N/2+1}(\bar{\lambda}/2)\}E[v(n-1)]. \quad (14)$$

It can be verified that the term within the brace $\{\cdot\}$ in (14) is less than one because $0 < \exp(\bar{\lambda}/2)E_{N/2+1}(\bar{\lambda}/2) < 1$ and $0 < A_c \le 1$. The former inequality results from the properties of the function $E_i(x)$ and the fact that $\bar{\lambda}/2 > 0$ and $N/2+1 > 1$. Consequently, $\lim_{n\to\infty} E[v(n)] \approx 0$. This concludes that the RLM algorithm converges in the mean and $w(n)$ converges approximately to the system parameter w^* under the stated assumptions. It should be noted that **Ass 3**) cannot be applied to the conventional RLS algorithm when the impulse density, p_r, is high. The reason is that the impulses will change significantly the value of $w(n)$ and hence $v(n)$ in successive iterations. Therefore, $\eta_o(n)$ and $v(n)$ becomes correlated. Whereas in the proposed RLM algorithm, the impulses will most likely be rejected and $v(n)$ can be stabilized. Therefore, the independent assumption in **Ass 3**) remains valid. Actually, if ξ is taken as infinity as in the conventional RLS algorithm, the impulses cannot be detected and **Ass 3**) is violated. The choice of ξ and its sensitivities have been studied in [8]. It was found that, within a wide range of the threshold parameters, the RLM algorithm together with the robust parameter estimation method proposed in [1] is robust to impulse noise. This substantiates the validity of the proposed mean convergence analysis and theoretically explains the advantages of using the M-estimate cost function and the threshold estimation method in [1,8] over the conventional RLS algorithm.

2.2 Mean Square behavior of the RLM Algorithm

Post-multiplying (7) by its transpose and taking expectations on both sides over $\{v, X, \eta_o\}$ gives

$$\Gamma(n) = \Gamma(n-1) - R_{\xi X}^{-1}S_1 - S_2 R_{\xi X}^{-1} + R_{\xi X}^{-1}S_3 R_{\xi X}^{-1}, \quad (15)$$

where $\Gamma(n) \triangleq E[v(n)v^T(n)]$ is defined as the weight-error vector correlation matrix [9] and we have

$$S_1 = E_{\{v,X,\eta_o\}}\{q(e)eXv^T/(\bar{\lambda} + X^T R_{\xi X}^{-1}X)\} \approx A_c I_1 R_{\xi X}\Gamma(n-1)/2, \quad (16)$$

$$S_2 = E_{\{v,X,\eta_o\}}\{q(e)evX^T/(\bar{\lambda} + X^T R_{\xi X}^{-1}X)\} \approx A_c I_1 \Gamma(n-1)R_{\xi X}/2, \quad (17)$$

where A_c is given in (12). Note that the approximation introduced to derive A in (13) has been used to obtain (16) and (17). Moreover

$$S_3 = E_{\{v,X,\eta_o\}}\{q^2(e)e^2XX^T/(\bar{\lambda} + X^T R_{\xi X}^{-1}X)^2\} = E_{\{v\}}\{s_3\}$$
$$= \{I_1 - I_2\}A_c R_{\xi X}\Gamma(n-1)R_{\xi X}/2 + \{I_0 - I_1\}\mathcal{E}_c R_{\xi X}/4, \quad (18)$$

where I_0, I_1, I_2, s_3 are given in (34), (51), and (48) respectively. $s_3 = E_{\{X,\eta_o\}}[q^2(e)e^2XX^T/(\bar{\lambda} + X^T R_{\xi X}^{-1}X)^2\,|\,v]$ and other integrals are evaluated in Appendix B and S_c is given in (46). Substituting (16), (17) and (18) into (15) and then post-multiplying $R_{\xi X}$ and taking the trace operation on both sides of it, we get

$$\mathcal{E}(n) = \mathcal{E}(n-1)\Gamma_1 + \Gamma_2 S_c, \quad (19)$$

where $\mathcal{E}(n) \triangleq trace(\Gamma(n)R_{\xi X})$ is the misadjustment [9],

$$\Gamma_1 = 1 - A_c(I_1 + I_2)/2, \text{ and } \Gamma_2 = \{I_0 - I_1\}/4. \quad (20)$$

It can be seen from (19) that $\mathcal{E}(n)$ is prevented from going to zero by the last term. Because $0 < A_c < 1$, and $N >> 1$, the bound for Γ_1 from (20) can be deduced as $0 < \Gamma_1 < 1$. Similarly, with $\bar{\lambda} > 0$ and $N >> 1$, Γ_2 from (20) is found to be a small positive value with bound $0 < \Gamma_2 < 2/[(\bar{\lambda} + N - 2)(\bar{\lambda} + N + 2)]$. Therefore, from (19), it can be seen that the weight-error vector converges in the mean square sense under the stated assumptions and the steady-state error is mainly decided by the parameter S_c given in (46). Moreover, the convergence of $\mathcal{E}(n)$ is verified numerically by solving (19) under specified parameters and the results are presented in Section 3.2.

3. Simulation Results

In the following simulations, the adaptive filter is configured as the system identification problem as in Fig. 1.

3.1 Mean Convergence Performance

In this simulation, the mean convergence performance of the RLM algorithm will be evaluated. The unknown system is modeled as a FIR filter with $w^* = [0.2, -0.4, 0.6, -0.8, 1, -0.8, 0.6, -0.4, 0.2]^T$. The input signal $x(n)$ is a colored signal generated by passing a zero-mean white Gaussian noise through a linear-time invariant filter with coefficients [0.3887,1,0.3887] [9]. The parameters and initial values used in the simulation are $N=9$, $r_m=300$, $N_w=9$, $\lambda_0=\lambda=0.99$, $\hat{\sigma}^2(0)=d^2(0)$, $P(0)=0_{(N\times 1)}$, $w(0)=0_{(N\times 1)}$ and $V(0)=20I_{(N\times N)}$, where $0_{(N\times 1)}$ and $I_{(N\times N)}$ are the zero and identity matrices, respectively. The CG impulse noise is generated from (5) with $P_r(b(n)=1)=p_r=0.005$ in each independent run. The norm of the mean squared weight-error vector $\|v_A(n)\|_2$ is used as performance measure. The theoretical result is obtained from equation (14) with $\xi=2.24\cdot\hat{\sigma}(n)$, where $\hat{\sigma}(n)$ is obtained by averaging $\hat{\sigma}(n)$ over 200 independent runs. The resulting numerical and simulation results are illustrated in Fig. 2. It can be seen that there is a good match between the theoretical and simulation results, especially when n is sufficiently large. The small discrepancy at the beginning may result from the inaccurate estimation of $\hat{\sigma}(n)$, and the approximations used in the performance analysis.

3.2 Mean Square Convergence Performance

This experiment follows the same settings in Section 3.1 and it evaluates the mean square convergence performance of the RLM algorithm. The misadjustment $\triangleq(n)=trace(\Gamma(n)R_{E,X})$ is considered. The numerical result obtained from (19) and the simulation results averaged over 200 independent runs are plotted in Fig. 3. Again, there is a good match between the theoretical and simulation results.

4. Conclusion

This paper presents the mean and mean square convergence analysis of the RLM algorithm under the contaminated Gaussian (CG) impulse noise model. The simulation and theoretical convergence results agree very well with each other and suggest that the RLM algorithm is more robust than the RLS algorithm under the CG noise model.

5. Appendix A

The classical method proposed by Bershard in [13] is employed to evaluate $A\triangleq E_{\{X,\eta_g\}}[Xq(e)e/(\bar{\lambda}+X^T R_{E,X}^{-1}X)\,|\,v]$, where the subscript indicates that the expectation is taken over $\{X,\eta\}$. Using **Ass 2)** and X are jointly Gaussian with covariance matrix $R_{E,X}$, one gets

$$A=C_K\iint_{X,\eta_{fold}}\frac{Xq(e)e}{\bar{\lambda}+X^T R_{E,X}^{-1}X}\exp(-\frac{1}{2}X^T R_{E,X}^{-1}X)f_{\eta_o}(\eta)d\eta dX,\qquad(21)$$

where $C_K=(2\pi)^{-N/2}|R_{E,X}|^{-1/2}$. To evaluate (21), let's consider

$$F(\beta)=C_K\iint_{X,\eta}\frac{(Xq(e)e)e^{-\beta(\bar{\lambda}+X^T R_{E,X}^{-1}X)}}{\bar{\lambda}+X^T R_{E,X}^{-1}X}\exp(-\frac{X^T R_{E,X}^{-1}X}{2})f_{\eta_o}(\eta_o)d\eta_o dX\quad(22)$$

Notice that the desired result is obtained when $\beta=0$

$$A=F(0)\,.\qquad(23)$$

Differentiating (22) with respect to β, gives

$$dF(\beta)/d\beta=-\exp(-\beta\bar{\lambda})L_2/(2\beta+1)^{N/2}\,,\qquad(24)$$

where $L_2\triangleq E_{\{X,\eta_g\}}[Xq(e)e\,|v]|_{E[XX^T]=B}$ is the expectation of $Xq(e)e$ conditioned on v when $x_i,x_j\in X$ are jointly Gaussian with covariance matrix B and $B=R_{E,X}/(1+2\beta)$. In **Ass 3)**, X and e are assumed as jointly Gaussian, the Price theorem [14] for X and e can be invoked to obtain the following:

$$\partial L_2/\partial r_{E,Xe}=E_{\{X,\eta_g\}}[q'(e)e|v]\big|_B+E_{\{X,\eta_g\}}[q(e)\,|v]\big|_B=L_{21}+L_{22}\,,\qquad(25)$$

where $q'(e)=dq(e)/de$, $r_{E,Xe}\triangleq E_{\{X,\eta_g\}}[Xe|v]|_B$ is the covariance of X and e, $L_{21}\triangleq E_{\{X,\eta_g\}}[q'(e)e|v]|_B$ and $L_{22}\triangleq E_{\{X,\eta_g\}}[q(e)|v]|_B$. These expectations will be evaluated as follows. First of all, it is noted that

$$\begin{aligned}r_{E,Xe}&\triangleq E_{\{X,\eta_g\}}[Xe|v]|_B=E_{\{X,\eta_g\}}[X(\eta_g+X^T v)|v]|_B\\&=Bv=(1+2\beta)^{-1}R_{E,X}v\,.\end{aligned}\qquad(26)$$

Since X and η_g are statistically independent, we get

$$L_{21}=C_B\iint_{X,fold}\{\int(q'(e)e)f_{\eta_o}(\eta_o)d\eta_o\}\exp(-\frac{X^T B^{-1}X}{2})dX=0\,,\qquad(27)$$

In fact, $I=\int(q'(e)e)f_{\eta_o}(\eta_o)d\eta_o$ is an odd function and we have

$$I=\int[\delta(\eta_o-\Xi)-\delta(\eta_o+\Im)](\eta_o+v^T X)f_{\eta_o}(\eta_o)d\eta_o=\xi[f_{\eta_o}(\Xi)-f_{\eta_o}(-\Im)]\quad(28)$$

where, $\Xi=\xi-v^T X$, $\Im=\xi+v^T X$, $\xi=k_\xi\hat{\sigma}(n)$. We also used the identities $e=\eta+v^T X$ and $q'(e)=\delta(e+\xi)+\delta(e-\xi)$, where $\delta(\cdot)$ is the Dirac-delta function. Next, we consider L_{22}. The derivation process is similar to deriving L_{21}, we get

$$L_{22}=C_B\iint_{X,fold}A_0(v,X)\exp(-\frac{X^T B^{-1}X}{2})dX\,,\qquad(29)$$

where $A_0(v,X)\triangleq\int q(e)f_{\eta_o}(\eta)d\eta$. Substituting $f_{\eta_o}(\eta)$ given in (6) into (29), one gets

$$\begin{aligned}A_0(v,X)&=(1-p_r)\{erf(\Xi/(\sqrt{2}\sigma_g))+erf(\Im/(\sqrt{2}\sigma_g))\}\\&+p_r\{erf(\Xi/(\sqrt{2}\sigma_\Sigma))+erf(\Im/(\sqrt{2}\sigma_\Sigma))\}\,,\end{aligned}\qquad(30)$$

where $erf(x)=2\int_0^x\exp(-t^2)dt/\sqrt{\pi}$. Using the result in (27), (25) can be simplified to

$$\frac{\partial L_2}{\partial r_{E,Xe}}=L_{22}=C_B\iint_{X,fold}A_0(v,X)\exp(-\frac{1}{2}X^T B^{-1}X)dX\qquad(31)$$

Integrating (31) with respect to $r_{E,Xe}$, one gets

$$L_2=\{C_B\iint_{X,fold}A_0(v,X)\exp(-\frac{1}{2}X^T B^{-1}X)dX\}\frac{R_{E,X}v(n-1)}{(1+2\beta)}+C_{L_2}\,,\qquad(32)$$

where (26) is used and $C_{L_2}=E_{\{X,\eta_g\}}[Xq(e)e\,|v]|r_{E,Xe=0}$ is the constant of integration, which can be shown to be zero. Inserting (32) into (24), integrating with respect to β and setting $\beta=0$ yields

$$A=F(0)=\{C_B\iint_{X,fold}A_0(v,X)\exp(-\frac{X^T B^{-1}X}{2})dX\}R_{E,X}v(n-1)I_1/2\quad(33)$$

and

$$I_1=\int_0^\infty\frac{\exp(-\beta\bar{\lambda})}{(2\beta+1)^{(N/2)+2}}d\beta=\exp(\frac{\bar{\lambda}}{2})E_{\frac{N}{2}+1}(\frac{\bar{\lambda}}{2})\,,\qquad(34)$$

where $E_k(x)=\int_1^\infty\exp(-x\beta)/\beta^k d\beta$ is the Exponential Integral function. Note that the constant of integration for $F(\beta)$ is equal to zero because of the boundary condition $F(\infty)=0$.

6. Appendix B

The evaluation of $s_3=E_{\{X,\eta_g\}}[q^2(e)e^2XX^T/(\bar{\lambda}+X^T R_{E,X}^{-1}X)^2\,|v]$ follows the similar approach used in Appendix A. Specifically,

$$s_3=C_K\iint_{X,\eta_{fold}}\frac{q^2(e)e^2XX^T}{(\bar{\lambda}+X^T R_{E,X}^{-1}X)^2}\exp(-\frac{1}{2}X^T R_{E,X}^{-1}X)f_{\eta_o}(\eta)d\eta dX,\qquad(35)$$

where $f_{\eta_o}(\eta)$ is given in (6). Let's define

$$\bar{F}(\beta)=C_K\iint_{X,\eta}\frac{XX^T q^2(e)e^2 e^{-\beta(\bar{\lambda}+X^T R_{E,X}^{-1}X)}}{(\bar{\lambda}+X^T R_{E,X}^{-1}X)^2}\exp(-\frac{X^T R_{E,X}^{-1}X}{2})f_{\eta_o}(\eta_o)d\eta_o dX\,,\ (36)$$

Comparing (36) with (35), it can be seen that $s_3=\bar{F}(0)$.

Differentiating (36) twice with respect to β, one gets

$$d^2\bar{F}(\beta)/d\beta=\exp(-\beta\bar{\lambda})L_3/(2\beta+1)^{N/2}\,,\qquad(37)$$

where $L_3\triangleq E_{\{X,\eta_g\}}[q^2(e)e^2XX^T\,|v]|_B$. Using the approach introduced in [15], one gets from the Price theorem the following

$$\partial L_3/\partial r_{E,Xe}=2(L_{31}+L_{32})\,,\qquad(38)$$

where $L_{31}\triangleq E_{\{X,\eta_g\}}[q(e)q(e)e^2X^T\,|v]|_B$, $L_{32}\triangleq E_{\{X,\eta_g\}}[q^2(e)eX^T\,|v]|_B$. Following a similar approach in deriving L_{21} and L_{22}, one gets

$$L_{31}=C_B\iint_{X,fold}\left\{\int q'(e)q(e)e^2 f_{\eta_o}(\eta)d\eta\right\}X^T\exp(-\frac{1}{2}X^T B^{-1}X)dX\approx 0\,,\ (39)$$

$$L_{12} = C_B \iint_{Nfold} B_0(v,X) X^T \exp(-X^T B^{-1} X/2) dX \,, \quad (40)$$

where $B_0(v,X) \triangleq \int q^2(e) e f_{\eta_o}(\eta_o) d\eta_o$. Inserting $q(e)$ defined in (1) and following the same approach in approximating A in (10) to (13), we obtained the approximation for $B_0(v,X)$ as

$$B_0(v,X) = \int q^2(e) e f_{\eta_o}(\eta_o) d\eta_o = \int_{-\xi}^{\xi} e f_{\eta_o}(\eta_o) d\eta_o$$

$$\approx v^T X\{(1-p_r) erf(\frac{\xi}{\sqrt{2}\sigma_g}) + p_r erf(\frac{\xi}{\sqrt{2}\sigma_\Sigma})\} = A_e v^T X \,, \quad (41)$$

where A_e is given in (12). Inserting (41) into (40) and then substituting (39) into (38) gives

$$\partial L_3 / \partial r_{E,Xe} = 2 A_e v^T B \,. \quad (42)$$

Integrating (42) with respect to $r_{E,Xe}$ and using $r_{E,Xe} = (1+2\beta)^{-1} R_{E,X} v(n-1)$, one gets

$$L_3 = 2 A_e (1+2\beta)^{-2} R_{E,X} v v^T R_{E,X} + C_{L_3} \,, \quad (43)$$

where C_{L_3} is the integration constant, which can be determined from $E_{|X,\eta_o|}[q^2(e) e^2 X X^T | v]$ conditioned on $r_{E,Xe} = Bv = 0$. In this case, the variables e and X are uncorrelated. Therefore, we have

$$C_{L_3} = E_{|X,\eta_o|}[q^2(e) e^2 X X^T | v] \Big|_{r_{E,Xe}=0} = \bar{C}_{L_3} \Big|_{r_{E,Xe}=0} B \,, \quad (44)$$

where $\bar{C}_{L_3} \triangleq E_{|X,\eta_o|}[q^2(e) e^2 | v] \big|_B = C_B \iint_{Nfold} C_0(v,X) \exp(-\frac{X^T B^{-1} X}{2}) dX$ and $C_0(v,X) \triangleq \int q^2(e) e^2 f_{\eta_o}(\eta_o) d\eta_o$. Using the same argument in deriving (12), $C_0(v,X)$ can be further approximated as follows:

$$C_0(v,X) \triangleq \int q^2(e) e^2 f_{\eta_o}(\eta_o) d\eta_o \approx A_e v^T X X^T v + S_e \,, \quad (45)$$

where A_e is given in (12) and

$$S_e = -\frac{2(1-p_r)\sigma_g \xi}{\sqrt{2\pi}} \exp(-\frac{\xi}{2\sigma_g^2}) - \frac{2 p_r \sigma_\Sigma \xi}{\sqrt{2\pi}} \exp(-\frac{\xi}{2\sigma_\Sigma^2})$$

$$+ (1-p_r)\sigma_g^2 erf(\xi/(\sqrt{2}\sigma_g)) + p_r \sigma_\Sigma^2 erf(\xi/(\sqrt{2}\sigma_\Sigma)) \,. \quad (46)$$

Substituting the results in (44), (45) and (46), into (43), we get

$$L_3 \approx 2 A_e (1+2\beta)^{-2} R_{E,X} v v^T R_{E,X} + S_e (1+2\beta)^{-1} R_{E,X} \,. \quad (47)$$

Finally, substituting (47) into (37) and integrating (37) with respect to β, we have

$$s_v = \bar{F}(0) = \{I_1 - I_2\} A_e R_{E,X} v v^T R_{E,X} / 2 + (I_0 - I_1) S_e R_{E,X} / 4 \quad (48)$$

with the boundary conditions: $\bar{F}(\infty) = 0$ and $\partial \bar{F}(\beta) / \partial \beta |_{\beta=\infty} = 0$. Actually, the integrations in (48) are double integral in the first half of the first quadrant in the β_1, β_2 plane. Interchanging the order of integration [13], we have

$$\int_0^\infty \int_{\beta_1}^\infty \frac{\exp(-\beta_2 \bar{\lambda})}{(2\beta_2+1)^{(N+4)/2}} d\beta_2 d\beta_1 = \int_0^\infty \int_0^{\beta_2} \frac{\exp(-\beta_2 \bar{\lambda})}{(2\beta_2+1)^{(N+4)/2}} d\beta_1 d\beta_2$$

$$= \int_0^\infty \frac{\beta_2 \exp(-\beta_2 \bar{\lambda})}{(2\beta_2+1)^{(N+4)/2}} d\beta_2 = (I_1 - I_2)/4 \,, \quad (49)$$

$$\int_0^\infty \int_{\beta_1}^\infty \frac{\exp(-\beta_2 \bar{\lambda}) d\beta_2 d\beta_1}{(2\beta_2+1)^{(N+2)/2}} = (I_0 - I_1)/4 \,, \quad (50)$$

$$\text{where } I_0 = \exp(\frac{\bar{\lambda}}{2}) E_{\frac{N}{2}}(\frac{\bar{\lambda}}{2}) \text{ and } I_2 = \exp(\frac{\bar{\lambda}}{2}) E_{\frac{N}{2}+2}(\frac{\bar{\lambda}}{2}) \,. \quad (51)$$

References:

[1] Y. Zou, S. C. Chan, and T. S. Ng, "A Recursive Least M-Estimate (RLM) Adaptive Filter For Robust Filtering in Impulsive Noise," *IEEE Signal Processing Letters*, vol. 7, No. 11, pp. 324 -326, 2000.

[2] S. Koike, "Adaptive Threshold Nonlinear Algorithm for Adaptive Filters with Robustness against Impulsive Noise," *IEEE Transactions on Signal Processing*, vol. 45, No. 9, pp. 2391-2395, 1997.

[3] T. I. Haweel and P. M. Clarkson, "A Class of Order Statistic LMS Algorithms," *IEEE Transactions on Signal Processing*, vol. 40, No.1, pp. 44-53, 1992.

[4] K. L. Blackard, T. S. Rappaport, and C. W. Bostian, "Measurements and Models of Radio Frequency Impulsive Noise for Indoor Wireless Communications," *IEEE Journal on Selected Areas in Communications*, vol. SAC-11, No.7, pp. 991-1001, 1993.

[5] J. F. Weng and S. H. Leung, "Adaptive Nonlinear RLS Algorithm for Robust Filtering in Impulse Noise," *IEEE ISCAS'97*, vol. 4, pp. 2337-2340, 1997.

[6] R. Settineri, M. Najim, and D. Ottaviani, "Order Statistic Fast Kalman Filter," *IEEE ISCAS'96*, vol. 2, pp. 116-119, 1996.

[7] Y. Zou, S. C. Chan, and T. S. Ng, "Least mean M-estimate algorithms for robust adaptive filtering in impulse noise," *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, vol. 47, No. 12, pp. 1564 -1569, 2000.

[8] Y. Zou, "Robust Statistics Based Adaptive Filtering Algorithms For Impulsive Noise Suppression," *Ph. D. Dissertation, The University of Hong Kong, December*, 2000.

[9] S. S. Haykin, *Adaptive Filter Theory*, 3rd ed. Englewood Cliffs, N.J.: Prentice Hall, 1996.

[10] P. J. Rousseeuw and A. M. Leroy, *Robust Regression and Outlier Detection*: New York: John Wiley & Sons Inc., 1987.

[11] C. J. Masreliez and R. D. Martin, "Robust Bayesian Estimation for the Linear Model and Robustifying the Kalman Filter," *IEEE Trans. on Automatic Control*, vol. AC-22, No. 3, pp. 361-371, 1977.

[12] J. G. Proakis, *Digital Communications*, 3rd ed. New York: McGraw-Hill, 1995.

[13] N. J. Bershard, "Analysis of the Normalised LMS Algorithm with Guassian Inputs," *IEEE Transactions On Acoustic, Speech, and Signal Processing*, vol. ASSP-34, No. 4, pp. 793-806, 1986.

[14] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*: New York, McGraw-Hill, 3rd, 1991.

[15] V. J. Mathews, "Performance Analysis of Adaptive Filters Equipped with the Dual Sign Algorithm," *IEEE Transactions on Signal Processing*, vol. 39, No.1, pp. 85-91, 1991.

[16] Yuexian Zou and Shing-Chow Chan, "A Huber recursive least squares adaptive lattice filter for impulse noise suppression," *Proc. IEEE ICASSP2001*.

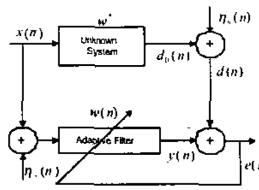

Fig. 1. System identification structure.

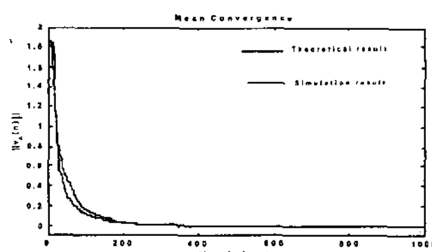

Fig.2. Mean convergence of the RLM algorithm.

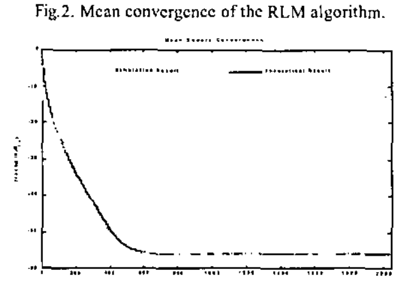

Fig. 3: Mean square convergence of the RLM algorithm.