

A New Regularized QRD Recursive Least M-estimate Algorithm: Performance Analysis and Applications

S. C. Chan, Y. J. Chu, and Z. G. Zhang

Department of Electrical and Electronic Engineering
The University of Hong Kong, Pokfulam Road, Hong Kong
{schan; yjchu; zgzhang}@eee.hku.hk

Abstract—This paper proposes a new regularized QR decomposition based recursive least M-estimate (R-QRRLM) adaptive filtering algorithm and studies its mean and mean square convergence performance and application to acoustic echo cancellation (AEC). The proposed algorithm extends the conventional RLM algorithm by imposing a weighted L_2 regularization term on the coefficients to reduce the variance of the estimator. Moreover, a QRD-based algorithm is employed for efficient recursive implementation and improved numerical property. The mean convergence analysis shows that a bias solution to the classical Wiener solution will be introduced due to the regularization. The steady-state excess mean square error (EMSE) is derived and it suggests that the variance will decrease while the bias will increase with the regularization parameter. Therefore, regularization can help to trade bias for variance. In this study, the regularization parameter can be adaptively selected and the resultant variable regularization parameter QRRLM (VR-QRRLM) algorithm can obtain both high immunity to input variation and low steady-state EMSE values. The theoretical results are in good agreement with simulation results. Computer simulation results on AEC show that the R-QRRLM and VR-QRRLM algorithms considerably outperform the traditional RLS algorithm when the input signal level is low or during double talk.

I. INTRODUCTION

The recursive least-squares (RLS) algorithm is an effective adaptive filtering algorithm which has been widely applied in system identification, interference cancellation and many other applications [1]. The traditional RLS method estimates the regression coefficients of a linear model in a least-squares (LS) sense by minimizing the sum of squared residual errors. Since the LS estimation implicitly assumes that the additive noise is Gaussian distributed, its performance will be considerably degraded in impulsive noise environment. To address this problem, a recursive least M-estimate (RLM) algorithm was proposed in [2,11] and it employs an M-estimation function instead of the conventional quadratic LS function so that the adverse effect of the impulsive noise or outliers can be effectively suppressed. Furthermore, to lower the computational complexity and to improve the numerical stability of the RLM algorithm in finite wordlength implementation, a QR decomposition (QRD)-based RLM (QRRLM) algorithm was developed in [2].

One possible problem with the RLS-like algorithms is that the covariance matrix may become poorly conditioned or even singular. This may occur when the excitation is not persistence, say when input signal level is weak or during a long period of

silence. The mean squares error (MSE) of the RLS algorithm may increase dramatically or even suffering from instability. For example, in the adaptive echo cancellation (AEC) problem, the level of the excitation signal, which is usually a speech or audio signal, may vary significantly over time and may not be long enough to reliably estimate the coefficients of an acoustic channel with long impulse response. As a result, most RLS-type algorithms will suffer from large variance in the channel estimate. To address this ill-conditional problem, a commonly used technique is to introduce some kind of regularization into the algorithm [15].

Regularization techniques have attracted much interest recently as a tool for reducing the variance for estimation with small number of data samples and automatic model selection. It has been applied to a wide variety of areas such as audio signal processing [3], model selection [4], basis pursuit denoising [5], sparse signal recovery and compressed sensing [6-8], etc. In [14], a RLS-based algorithm for adaptive beamforming with a quadratic norm constraint was introduced. It employs an approximate expansion of the diagonal loaded covariance matrix to update the beamformer and determine the loading factor by solving a quadratic equation. Since the QRD-based RLS algorithm is more robust in terms of finite wordlength effect and hardware implementation using the CORDIC algorithm, it is highly desirable to develop a QRD-based regularized RLS algorithm. In a recent contribution [8], the authors have introduced a recursive QRD-based RLM algorithm for reweighted L_1 regularization.

In this paper, the performance analysis and application of a new time-recursive regularized QRD-based RLM algorithm are proposed. The regularized RLM employs M-estimation in combating impulsive outliers and a weighted L_2 regularization term on its coefficients in order to reduce the estimation variance and improves the numerical stability over the conventional RLS algorithm. In addition, an efficient QRD implementation of the regularized RLM, which improves the numerical stability and lead to efficient hardware implementation using the CORDIC algorithm, is proposed. The resulting algorithm is called the regularized QRRLM (R-QRRLM) algorithm. Furthermore, the regularization parameters of the R-QRRLM algorithm can be adaptively determined, resulting in a variable regularization parameter QRRLM (VR-QRRLM) algorithm. The mean and mean square convergence analysis of the R-QRRLM algorithm for Gaussian inputs and additive noise is then carried out by using the Price theorem and the generalized Abelian integral

This study was partially supported by a grant from the Research Grants Council of the Hong Kong SAR, China.

functions [9]. Extension to the contaminated Gaussian noise case can be carried out using the approach recently introduced in [9]. The R-QRRLM algorithm is applied to an adaptive echo cancellation problem to demonstrate its effectiveness in contaminated-Gaussian noise environments and improved performance under low signal input level.

The paper is organized as follows. In Section II, the R-QRRLM and VR-QRRLM algorithms are proposed. In Section III, the performance analysis is derived. Experimental results for verifying the theoretical analysis and comparing the algorithm's performance with RLS algorithm in AEC are presented in Section IV. Finally, conclusions are drawn in Section V.

II. REGULARIZED QRRLM ALGORITHM

A. QRRLM Algorithm

The conventional RLS algorithm is based on the LS criterion and hence its performance will deteriorate considerably when the desired or the input signal is corrupted by impulsive noise. Robust statistics based on M-estimation [10] is an effective method to reduce the hostile effects of impulsive noise on the estimates. Specifically, for the system identification problem, where $\mathbf{W}^* = [w_1, \dots, w_L]^T$ is the length- L impulse response to be identified and $\mathbf{X}(n) = [x(n), \dots, x(n-L+1)]^T$ is the input signal, the RLM algorithm aims to minimize the following cost function:

$$\begin{aligned} J_\rho(n) &= \sum_{i=0}^n \lambda_{n-i}(n) \rho(e(i)) \\ &= \sum_{i=0}^n \lambda_{n-i}(n) \rho(d(i) - \mathbf{W}^T(n) \mathbf{X}(i)) \\ &= \sum_{i=0}^n \lambda_{n-i}(n) \rho(\mathbf{W}^{*T} \mathbf{X}(i) + \eta(i) - \mathbf{W}^T(n) \mathbf{X}(i)), \end{aligned} \quad (1)$$

where $\mathbf{W}(n) = [w_1(n), \dots, w_L(n)]^T$ is the adaptive filter with length L , $d(i) = \mathbf{W}^{*T} \mathbf{X}(i) + \eta(i)$ is the desired signal and $\eta(i)$ is the additive zero mean white Gaussian noise; $\rho(\cdot)$ is an M-estimate function such as the following modified Huber (MH) function or other appropriate functions

$$\rho(e) = \begin{cases} e^2/2 & 0 \leq |e| < \xi, \\ |e| - \xi + \xi^2/2 & |e| \geq \xi, \end{cases} \quad (2)$$

where ξ is a threshold parameter used to control the suppression of outliers and adaptation speed. It can be recursively updated using the following adaptive threshold selection (ATS) method introduced in [11,12].

$$\begin{aligned} \hat{\sigma}^2(n) &= \lambda_\sigma \hat{\sigma}^2(n-1) + c_1(1 - \lambda_\sigma) \text{median}[A_e(n)], \\ \xi &= 2.576 \hat{\sigma}^2(n), \end{aligned} \quad (3)$$

where the forgetting factor λ_σ is a positive real number close to but smaller than one, $A_e(n) = \{e^2(n), \dots, e^2(n - N_w + 1)\}$, $c_1 = 1.483(1 + 5/(N_w - 1))$ is a finite sample correction factor, and N_w is the length of the data set. $\lambda_{n-i}(n)$ in (1) serves the purpose of an exponential window which pays less emphasis to errors at distant past. For example, it can be chosen as λ^{n-i} , where λ is a constant forgetting factor or adaptively in variable forgetting (VFF) algorithms. By setting the first partial derivatives of $J_\rho(n)$, with respect to $\mathbf{W}(n)$, to zero, it

was shown in [10] that the optimal weight vector satisfy the M -estimate normal equation:

$$\mathbf{R}_{X\rho}(n) \mathbf{W}(n) = \mathbf{P}_{X\rho}(n), \quad (4)$$

where $\mathbf{R}_{X\rho}(n) = \sum_{i=0}^n \lambda_{n-i}(n) q(e(i)) \mathbf{X}(i) \mathbf{X}^T(i)$ and $\mathbf{P}_{X\rho}(n) = \sum_{i=0}^n \lambda_{n-i}(n) q(e(i)) d(i) \mathbf{X}(i)$ are the M -estimate autocorrelation matrix of $\mathbf{X}(n)$ and the M -estimate cross-correlation vector of $d(n)$ and $\mathbf{X}(n)$, respectively, and $\rho'(e) = \partial \rho(e) / \partial e = q(e)e$. In order to prevent the values of $\mathbf{R}_{X\rho}(n)$ and $\mathbf{P}_{X\rho}(n)$ from continuously decreasing when a series of impulses is present in the underlying signals, $\lambda_{n-i}(n)$ can be updated as follows [11]

$$\lambda_{n-i}(n) = \lambda_e(n) \lambda_{n-i-1}(n-1), \quad \lambda_e(n) = \begin{cases} \lambda & |e(n)| < \xi \\ 1 & |e(n)| \geq \xi \end{cases}.$$

Applying the iterative reweighted LS approach, the following RLM algorithm can be obtained:

$$\begin{aligned} \mathbf{P}(n) &= \lambda_e^{-1}(n) (\mathbf{I} - \mathbf{K}(n) \mathbf{X}^T(n)) \mathbf{P}(n-1), \\ \mathbf{K}(n) &= \frac{q(e(n)) \mathbf{P}(n-1) \mathbf{X}(n)}{\lambda_e(n) + q(e(n)) \mathbf{X}^T(n) \mathbf{P}(n-1) \mathbf{X}(n)}, \end{aligned} \quad (5)$$

$$\mathbf{W}(n) = \mathbf{W}(n-1) + [d(n) - \mathbf{W}^T(n-1) \mathbf{X}(n)] \mathbf{K}(n).$$

Eq. (5) can be efficiently implemented using a QR-based implementation [2] and it is summarized in Table I. It is mathematically equivalent to but has higher numerical stability than the RLM algorithm above.

TABLE I. THE QRRLM ALGORITHM

<p>Initialization: $\mathbf{R}(0) = \sqrt{\delta} \mathbf{I}$, δ is a small positive constant; $\mathbf{U}(0) = \mathbf{0}$; $\mathbf{w}(0) = \mathbf{0}$;</p> <p>Recursion: Given $\mathbf{R}(n-1)$, $\mathbf{U}(n-1)$, $\mathbf{w}(n-1)$, $\mathbf{X}(n)$ and $d(n)$, compute at time n:</p> <p>(i). $\begin{bmatrix} \mathbf{R}(n) & \mathbf{U}(n) \\ \mathbf{0} & c(n) \end{bmatrix} = \mathbf{Q}(n) \begin{bmatrix} \sqrt{\lambda_e} \mathbf{R}(n-1) & \sqrt{\lambda_e} \mathbf{U}(n-1) \\ \sqrt{q(e(n))} \mathbf{X}^T(n) & \sqrt{q(e(n))} d(n) \end{bmatrix}$, where $\mathbf{Q}(n)$ is calculated by Givens rotation and λ is the forgetting factor.</p> <p>(ii). $\mathbf{w}(n) = \mathbf{R}^{-1}(n) \mathbf{U}(n)$ (back-substitution).</p>

B. R-QRRLM and VR-QRRLM

In some applications, the adaptive filter is not persistently excited say when the input signal level is very low. Then, the autocorrelation matrix $\mathbf{R}_{X\rho}(n)$ may be ill-conditioned. As a result, a large estimation variance will result. To address the problem, a regularization term on the adaptive filter coefficients can be imposed on the objective function at time n to form:

$$\mathbf{W}(n) = \arg \min_{\mathbf{W}} [J_\rho(n) + \mu \|\mathbf{D} \mathbf{W}(n)\|_p^2] \quad (6)$$

where $\|\cdot\|_p$ represents L_p -norm and \mathbf{D} is a weighting matrix which can be used to approximate different regularization methods such as L_1 norm or the smoothly clipped absolute deviation (SCAD) as shown in [8]. The regularization parameter μ can be determined adaptively as

$$\mu(n) = (1 - \lambda)L \sqrt{\frac{\sigma_\eta^2 / \text{Tr}(\mathbf{R}_{X\rho}(n))}{\mathbf{W}^* \mathbf{W}^*}}, \text{ where } \text{Tr}(\cdot) \text{ is the trace}$$

operator. If the variable regularization parameter $\mu(n)$ is used, the algorithm is modified to the variable regularization parameter QRRLM (VR-QRRLM) algorithm, which has improved performance than R-QRRLM with a fixed regularization parameter [13]. Due to page limitation, the derivation of $\mu(n)$ is omitted here and it will be reported elsewhere.

Unlike the cost function in (1), the regularized problem-estimation function in (6) cannot be solved simply by the recursive QRD in Table I due to the regularization term. A possible method to overcome this difficulty is to append the vector $\sqrt{\mu L} E_l$ randomly or sequentially to the previous QRD, where E_l is the l -th row of the regularization matrix \mathbf{D} [8]. If the vector is applied sequentially, then $l = (n \bmod L) + 1$. \mathbf{D} is the identity matrix for L_2 regularization or the generalized inverse of $\text{diag}\{|w_1(n-1)|, \dots, |w_L(n-1)|\}$ for L_1 regularization. More precisely, at each time instant, the algorithm (i) in Table I is executed once for the vector $\sqrt{q(e(n))}[X(n), d(n)]$ and again for the vector $\sqrt{q(e(n))}[\sqrt{\mu L} E_l, 0]$. That is, $\mathbf{R}(n)$ and $\mathbf{U}(n)$ are updated twice by appending $\sqrt{q(e(n))}[X(n), d(n)]$ and $\sqrt{q(e(n))}[\sqrt{\mu L} E_l, 0]$ respectively in the QRD operation. No forgetting, i.e. the forgetting factor is equal to one, is imposed in the second regularization update. We now study the convergence behaviors of the R-QRRLM algorithm for Gaussian input and additive Gaussian noise. The results for contaminated Gaussian noise can be similarly derived by means of the approach introduced in [9].

III. PERFORMANCE ANALYSIS OF R-QRRLM

The above QR implementation of the regularized RLM algorithm can be written as the following equivalent update:

$$\begin{aligned} \mathbf{W}(n+1) = & \mathbf{W}(n) - \mathbf{R}_{E,X}^{-1} \left[\frac{\mathbf{X}(n)\psi(e(n))}{\lambda + \mathbf{X}^T(n)\mathbf{R}_{E,X}^{-1}\mathbf{X}(n)} \right. \\ & \left. + \frac{\mathbf{r}(n)\psi(e_r(n))}{\lambda + \mathbf{r}^T(n)\mathbf{R}_{E,X}^{-1}\mathbf{r}(n)} \right]. \end{aligned} \quad (7)$$

For simplicity, we shall assume the case with fixed forgetting factor. The following assumptions are made:

- (A1) $\{\mathbf{X}(n)\}$ is zero-mean Gaussian distributed with covariance matrix \mathbf{R}_{XX} ;
- (A2) $\{\eta(n)\}$ is white Gaussian-distributed with zero-mean and is uncorrelated with $\{\mathbf{X}(n)\}$;
- (A3) the weight error vector $\{\mathbf{w}(n)\}$ is independent of $\{\mathbf{X}(n)\}$ and $\{\eta(n)\}$;
- (A4) $\mathbf{P}(n) = \mathbf{R}_{X\rho}^{-1}(n) \approx \mathbf{R}_{E,X}^{-1}(n) \approx E[\mathbf{R}_{E,X}^{-1}(n)] = \mathbf{R}_{E,X}^{-1}$, where $\mathbf{R}_{E,X}(n) = \lambda \mathbf{R}_{E,X}(n-1) + [\mathbf{X}(n)\mathbf{X}^T(n) + \mathbf{r}(n)\mathbf{r}^T(n)]$ and $\mathbf{r}(n) = \sqrt{\mu} \mathbf{d}_k$, \mathbf{d}_k is the k -th column of $\sqrt{\mathbf{A}(n-1)}$. Since:

$$E[\mathbf{R}_{E,X}(n)\mathbf{R}_{E,X}^{-1}(n)] = E\left[\sum_{l=1}^n \lambda^{n-l} \{\mathbf{X}(l)\mathbf{X}^T(l) + \mathbf{r}(l)\mathbf{r}^T(l)\} \mathbf{R}_{E,X}^{-1}(n)\right] = \mathbf{I}$$

and by the averaging principle, $\mathbf{R}_{E,X}^{-1}(n)$ is assumed to be independent of $\mathbf{X}(l)\mathbf{X}^T(l)$ and $\mathbf{r}(l)\mathbf{r}^T(l)$, we have

$$\begin{aligned} E\left[\sum_{l=1}^n \lambda^{n-l} (\mathbf{X}(l)\mathbf{X}^T(l) + \mathbf{r}(l)\mathbf{r}^T(l))\right] \cdot E[\mathbf{R}_{E,X}^{-1}(n)] \\ = \frac{(1-\lambda^n)}{(1-\lambda)} (\mathbf{R}_{XX} + \frac{\mu}{L} \mathbf{D}^2) \cdot E[\mathbf{R}_{E,X}^{-1}(n)] = \mathbf{I}. \end{aligned}$$

Therefore, $E[\mathbf{R}_{E,X}^{-1}(n)] = \frac{(1-\lambda)}{(1-\lambda^n)} (\mathbf{R}_{XX} + \frac{\mu}{L} \mathbf{D}^2)^{-1}$. For large n ,

$$\lim_{n \rightarrow \infty} E[\mathbf{R}_{E,X}^{-1}(n)] = \mathbf{R}_{E,X}^{-1} = (1-\lambda)(\mathbf{R}_{XX} + \frac{\mu}{L} \mathbf{D}^2)^{-1}.$$

A. Mean Convergence Analysis

First of all, we subtract the optimal solution of the R-QRRLM algorithm $\mathbf{W}(\infty)$, which will be derived in the process of the analysis, from (7) and take the expectation over $\{\mathbf{v}, \mathbf{X}, \eta_o\}$ on its both sides. This gives

$$E[\mathbf{v}(n+1)] = E[\mathbf{v}(n)] - \mathbf{R}_{E,X}^{-1} (\mathbf{L}_1 + \mathbf{L}'_1), \quad (8)$$

where $\mathbf{v}(n) = \mathbf{W}(\infty) - \mathbf{W}(n)$, $E[\cdot]$ denotes the expectation over $\{\mathbf{v}(n-1), \mathbf{X}(n), \eta_o(n)\}$, which is more clearly written here as $E_{\{\mathbf{v}, \mathbf{X}, \eta_o\}}[\cdot]$; by dropping the time index of \mathbf{X} , e , and η_o , \mathbf{L}_1 and \mathbf{L}'_1 are, respectively,

$$\begin{aligned} \mathbf{L}_1 &= E_{\{\mathbf{v}, \mathbf{X}, \eta_o\}} \left[\frac{\mathbf{X}\psi(e)}{\lambda + \mathbf{X}^T \mathbf{R}_{E,X}^{-1} \mathbf{X}} \right] \\ &= E_{\{\mathbf{v}\}} \left\{ E_{\{\mathbf{X}, \eta_o\}} \left[\frac{\mathbf{X}\psi(e)}{\lambda + \mathbf{X}^T \mathbf{R}_{E,X}^{-1} \mathbf{X}} \middle| \mathbf{v} \right] \right\} = E_{\{\mathbf{v}\}}[\mathbf{A}] \\ \mathbf{L}'_1 &= E_{\{\mathbf{v}, \mathbf{r}, \eta_o\}} \left[\frac{\mathbf{r}\psi(e_r)}{\lambda + \mathbf{r}^T \mathbf{R}_{E,X}^{-1} \mathbf{r}} \right] \\ &= \frac{\sqrt{\mu}}{L} \sum_{k=1}^L E_{\{\mathbf{v}\}} \left\{ E_{\{\mathbf{r}, \eta_o\}} \left[\frac{\mathbf{d}_k \psi(e_r)}{\lambda + \mu \mathbf{d}_k^T \mathbf{R}_{E,X}^{-1} \mathbf{d}_k} \middle| \mathbf{v} \right] \right\}, \end{aligned} \quad (9)$$

where $e_r = -\sqrt{\mu} \mathbf{w}^T(n) \mathbf{d}_k = -\sqrt{\mu} \mathbf{d}_{k,k} \mathbf{w}_k(n)$. Eq. (9) is obtained from the independence assumption of $\eta(n)$, $\mathbf{w}(n)$ and $\mathbf{X}(n)$ in (A3). The details for evaluating

$\mathbf{A} = E_{\{\mathbf{X}, \eta_o\}}[\mathbf{X}\psi(e)/(\lambda + \mathbf{X}^T \mathbf{R}_{E,X}^{-1} \mathbf{X}) | \mathbf{v}]$ is given as [13]

$$\mathbf{A} \approx A_\psi(\sigma_e^2) \cdot \mathbf{U} \Lambda^{1/2} \tilde{\mathbf{U}} \mathbf{I}(\tilde{\Lambda}) \tilde{\mathbf{U}}^T \Lambda^{1/2} \mathbf{U}^T \mathbf{v}(n), \quad (11)$$

where $\mathbf{R}_{XX} = \mathbf{U} \Lambda \mathbf{U}^T$ and $\mathbf{L}_X^{-1} \mathbf{R}_{EX} \mathbf{L}_X^{-T} = \tilde{\mathbf{U}} \tilde{\Lambda} \tilde{\mathbf{U}}^T$ are respectively the eigen-decomposition of \mathbf{R}_{XX} and $\mathbf{L}_X^{-1} \mathbf{R}_{EX} \mathbf{L}_X^{-T}$ with $\mathbf{L}_X = \mathbf{U} \Lambda^{\frac{1}{2}}$, $\mathbf{v}' = \Delta \mathbf{W} + \mathbf{v}$, $\Delta \mathbf{W} = \mathbf{W}^* - \mathbf{W}(\infty)$, $E_{\{\mathbf{X}, \eta_o\}}[\psi'(e) | \mathbf{v}] = A_\psi(\sigma_e^2)$; $\mathbf{I}(\tilde{\Lambda})$ is a diagonal matrix with its (i,i) -th entry given by:

$$I_i(\tilde{\Lambda}) = \int_0^\infty \frac{\exp(-\beta \lambda)}{(2\beta \tilde{\lambda}_i^{-1} + 1) \prod_{j=1}^L (2\beta \tilde{\lambda}_j^{-1} + 1)^{1/2}} d\beta, \quad (12)$$

where $\tilde{\lambda}_i$ is the i -th eigenvalue of $\tilde{\Lambda}$ and $\lambda(n)$ is the time-variable forgetting factor. Here $\lambda(n)$ is a constant. Similarly,

$$L'_1 = \frac{\sqrt{\mu}}{L} \mathbf{D}' \bar{\boldsymbol{\psi}}(n), \quad (13)$$

where $\bar{\boldsymbol{\psi}}(n) = [\bar{\boldsymbol{\psi}}(-\sqrt{\mu}d_{1,1}w_1(n)), \dots, \bar{\boldsymbol{\psi}}(\sqrt{\mu}d_{L,L}w_L(n))]^T$, $\bar{\boldsymbol{\psi}}(-\sqrt{\mu}d_{k,k}w_k(n)) = E_v[\boldsymbol{\psi}(-\sqrt{\mu}d_{k,k}(w_{opt,k} - v_k(n)))]$ with $w_{opt,k}$ being the k -th element of $\mathbf{W}(\infty)$, and $\mathbf{D}' = \text{diag}\{\frac{d_{1,1}}{\lambda + \mu d_{1,1}^2 \sigma_{R_{E,X,1}}^2}, \dots, \frac{d_{L,L}}{\lambda + \mu d_{L,L}^2 \sigma_{R_{E,X,L}}^2}\}$.

Substituting (9) to (13) into (8), the following relation between $E[\mathbf{v}(n+1)]$ and $E[\mathbf{v}(n)]$ is obtained

$$E[\mathbf{v}(n+1)] \approx (\mathbf{I} - A_{\boldsymbol{\psi}}(n) \mathbf{R}_{E,X}^{-1} \mathbf{U} \Lambda^{1/2} \tilde{\mathbf{U}} \mathbf{I}(\tilde{\Lambda}) \tilde{\mathbf{U}}^T \Lambda^{1/2} \mathbf{U}^T) \cdot E[\mathbf{v}(n)] - \mathbf{R}_{E,X}^{-1} [A_{\boldsymbol{\psi}}(n) \mathbf{U} \Lambda^{1/2} \tilde{\mathbf{U}} \mathbf{I}(\tilde{\Lambda}) \tilde{\mathbf{U}}^T \Lambda^{1/2} \mathbf{U}^T \Delta \mathbf{W} + \frac{\mu}{L} \mathbf{D}' \bar{\boldsymbol{\psi}}(n)]. \quad (14)$$

We first assume that the algorithm converge to determine the optimal solution. Then, we shall show that the algorithm is convergent for λ sufficiently close to 1. If the algorithm converges, then we have from (14) the following:

$$A_{\boldsymbol{\psi}}(\infty) \mathbf{U} \Lambda^{1/2} \tilde{\mathbf{U}} \mathbf{I}(\tilde{\Lambda}) \tilde{\mathbf{U}}^T \Lambda^{1/2} \mathbf{U}^T \Delta \mathbf{W} = -\frac{\mu}{L} \mathbf{D}' \bar{\boldsymbol{\psi}}(n), \quad (15)$$

Using $\mathbf{W}^* = \mathbf{R}_{XX}^{-1} \mathbf{r}_{xd}$, the optimal solution to R-QRRLM algorithm can be derived from (15) after some m:

$$\mathbf{r}_{xd} = \mathbf{U} \Lambda^{1/2} \tilde{\mathbf{U}} \mathbf{I}(\tilde{\Lambda}) \tilde{\mathbf{U}}^T \Lambda^{1/2} \mathbf{U}^T \cdot (\mathbf{U} \Lambda^{1/2} \tilde{\mathbf{U}} \mathbf{I}(\tilde{\Lambda}) \tilde{\mathbf{U}}^T \Lambda^{1/2} \mathbf{U}^T + \frac{\mu}{A_{\boldsymbol{\psi}}(\infty)L} \mathbf{D}'^2) \mathbf{W}(\infty). \quad (16)$$

Since $\mathbf{R}_{EX} \approx \mathbf{R}_{XX} + (\mu/L) \mathbf{D}^2 \approx \tilde{\mathbf{R}}_{XX}$ and $\mathbf{L}_X^T \tilde{\mathbf{R}}_{XX} \mathbf{L}_X^{-T} = \mathbf{I} + (\mu/L) \mathbf{L}_X^T \mathbf{D}^2 \mathbf{L}_X^{-T}$. If $(\mu/L) \|\mathbf{D}^2\|_2$ is sufficiently small, then $\tilde{\mathbf{U}} \approx \mathbf{I}$ and $\mathbf{I}(\tilde{\Lambda}) \approx \text{Erf}(\lambda) \cdot \mathbf{I}$ where $\text{Erf}(\mathbf{I}) = \int_0^{\infty} \frac{\exp(-\beta \lambda)}{(2\beta + 1)^{(L/2)+1}} d\beta$. Consequently, (16) is reduced to

$$(\mathbf{R}_{XX} + \frac{\mu}{A_{\boldsymbol{\psi}}(\infty) \text{Erf}(\lambda)L} \mathbf{D}') \mathbf{W}(\infty) \approx \mathbf{r}_{xd}. \quad (17)$$

To achieve a given regularization or diagonal loading γ_k at the k -th diagonal, we can set $d_k = A_{\boldsymbol{\psi}}(\infty) \text{Erf}(\lambda) (\lambda + \mu d_k^2 \sigma_{R_{E,X,k}}^2) \gamma_k L / \mu$.

To study the convergence rate, we shall focus on the terms in the curved bracket in (14). Let $\mathbf{V}(n) = \tilde{\mathbf{U}}^T \Lambda^{1/2} \mathbf{U}^T \mathbf{v}(n)$:

$$E[\mathbf{V}(n+1)] = (\mathbf{I} - A_{\boldsymbol{\psi}}(n) \tilde{\Lambda}^{-1} \mathbf{I}(\tilde{\Lambda})) E[\mathbf{V}(n)] - \tilde{\mathbf{U}}^T \Lambda^{1/2} \mathbf{U}^T \mathbf{R}_{E,X}^{-1} \cdot [A_{\boldsymbol{\psi}}(n) \mathbf{U} \Lambda^{1/2} \tilde{\mathbf{U}} \mathbf{I}(\tilde{\Lambda}) \tilde{\mathbf{U}}^T \Lambda^{1/2} \mathbf{U}^T \Delta \mathbf{W} + \frac{\mu}{L} \mathbf{D}' \bar{\boldsymbol{\psi}}(n)]. \quad (18)$$

Therefore, the mean weight error vector will converge if $|1 - A_{\boldsymbol{\psi}}(n) \tilde{\lambda}_i^{-1} I_i(\tilde{\Lambda})| < 1$. It can be shown that for λ sufficiently close to 1, we have

$$\tilde{\lambda}_i^{-1} I_i(\tilde{\Lambda}) = \frac{1}{2} \ln(1 + 2/(\tilde{\lambda}_i \lambda)) < \frac{1}{2} \ln(1 + 2) = 0.549. \quad (19)$$

Since $|A_{\boldsymbol{\psi}}(n)| < 1$, $|A_{\boldsymbol{\psi}}(n) \tilde{\lambda}_i^{-1} I_i(\tilde{\Lambda})| < 1$ and hence the mean error weight vector of the algorithm is convergent.

B. Mean Square Convergence Analysis

Post-multiplying $\mathbf{v}(n)$ by its transpose and taking expectation, one gets

$$\begin{aligned} & \boldsymbol{\Xi}(n+1) = \boldsymbol{\Xi}(n) \\ & - \mathbf{R}_{E,X}^{-1} \{A_{\boldsymbol{\psi}}(n) \mathbf{U} \Lambda^{1/2} \tilde{\mathbf{U}} \mathbf{I}(\tilde{\Lambda}) \tilde{\mathbf{U}}^T \Lambda^{1/2} \mathbf{U}^T \boldsymbol{\Xi}_{vv}(n) + \frac{\mu}{L} \mathbf{D}' \bar{\boldsymbol{\psi}}(n) E[\mathbf{v}^T(n)]\} \\ & - \{A_{\boldsymbol{\psi}}(n) \boldsymbol{\Xi}_{vv}(n) \mathbf{U} \Lambda^{1/2} \tilde{\mathbf{U}} \mathbf{I}(\tilde{\Lambda}) \tilde{\mathbf{U}}^T \Lambda^{1/2} \mathbf{U}^T + \frac{\mu}{L} E[\mathbf{v}(n)] \bar{\boldsymbol{\psi}}^T(n) \mathbf{D}'\} \mathbf{R}_{E,X}^{-1} \\ & + 2C(\sigma_e^2) \mathbf{U} \Lambda^{-1/2} \tilde{\mathbf{U}} \tilde{\Lambda}^{-1} \tilde{\mathbf{U}}^T (\mathbf{L}_X^T \boldsymbol{\Xi}_{vv}(n-1) \mathbf{L}_X \circ \mathbf{I}''(\tilde{\Lambda})) \tilde{\mathbf{U}} \tilde{\Lambda}^{-1} \tilde{\mathbf{U}}^T \Lambda^{-1/2} \mathbf{U}^T \\ & + B_{\boldsymbol{\psi}}(\sigma_e^2) \mathbf{U} \Lambda^{-1/2} \tilde{\mathbf{U}} \tilde{\Lambda}^{-1} \tilde{\mathbf{U}}^T \mathbf{I}'(\tilde{\Lambda}) \tilde{\mathbf{U}} \tilde{\Lambda}^{-1} \tilde{\mathbf{U}}^T \Lambda^{-1/2} \mathbf{U}^T \\ & + \frac{\sqrt{\mu}}{L} \mathbf{R}_{E,X}^{-1} [\alpha(n) \mathbf{U} \Lambda^{1/2} \tilde{\mathbf{U}} \mathbf{I}(\tilde{\Lambda}) \tilde{\mathbf{U}}^T \Lambda^{1/2} \mathbf{U}^T (E[\mathbf{v}(n)] + \Delta \mathbf{W}) \bar{\boldsymbol{\psi}}^T(n) \mathbf{D}'] \mathbf{R}_{E,X}^{-1} \\ & + \frac{\sqrt{\mu}}{L} \mathbf{R}_{E,X}^{-1} [\alpha(n) \mathbf{D}' \bar{\boldsymbol{\psi}}(n) (E[\mathbf{v}^T(n)] + \Delta \mathbf{W}^T) \mathbf{U} \Lambda^{1/2} \tilde{\mathbf{U}} \mathbf{I}(\tilde{\Lambda}) \tilde{\mathbf{U}}^T \Lambda^{1/2} \mathbf{U}^T] \mathbf{R}_{E,X}^{-1} \\ & + \frac{\mu}{L} \mathbf{R}_{E,X}^{-1} \mathbf{D}' \bar{\boldsymbol{\psi}}(n) \bar{\boldsymbol{\psi}}^T(n) \mathbf{D}' \mathbf{R}_{E,X}^{-1}, \end{aligned} \quad (20)$$

where $\sigma_e^2 = \sigma_{\eta_0}^2 + \mathbf{v}^T \mathbf{R}_{XX} \mathbf{v}$, $\frac{d}{d\sigma_e^2} E[\boldsymbol{\psi}^2(e)] = C_{\boldsymbol{\psi}}(\sigma_e^2)$, $\overline{\boldsymbol{\psi}^2}(\sigma_e^2) = B_{\boldsymbol{\psi}}(\sigma_e^2)$; $\boldsymbol{\Xi}_{vv}(n) = E[\mathbf{v}(n)(\mathbf{v}^T(n) + \Delta \mathbf{W}^T)] = \boldsymbol{\Xi}_{vv}(n) + E[\mathbf{v}(n) \Delta \mathbf{W}^T]$ with $\Delta \mathbf{W} = \mathbf{W}^* - \mathbf{W}(\infty)$ and $\boldsymbol{\Xi}_{vv}(n) = E[\mathbf{v}(n) \mathbf{v}^T(n)]$; $\mathbf{I}''(\tilde{\Lambda})$ is a matrix with its (i,j) -th entry given by:

$$I_{i,j} = \int_0^{\infty} \beta \exp(-\beta \lambda) (2\beta \tilde{\lambda}_i^{-1} + 1)^{-1} (2\beta \tilde{\lambda}_j^{-1} + 1)^{-1} \prod_{k=1}^L (2\beta \tilde{\lambda}_k^{-1} + 1)^{-1/2} d\beta;$$

and $\mathbf{I}'(\tilde{\Lambda})$ is a diagonal matrix with its (i,i) -th entry given by:

$$[I'(\tilde{\Lambda})]_{i,i} = \int_0^{\infty} \beta \exp(-\beta \lambda) (2\beta \tilde{\lambda}_i^{-1} + 1)^{-1} \prod_{k=1}^L (2\beta \tilde{\lambda}_k^{-1} + 1)^{-1/2} d\beta. \quad \text{Due}$$

to page limitation, the evaluation of the expectation, which is based on the Price's theorem, is omitted and interested readers are referred to [13]. Also due to page limitation, we first analyze the steady state EMSE of the algorithm below. It can be shown in [13] that the algorithm is also convergent in the mean squares sense. Therefore, at the steady state and using the transformation $\mathbf{V}(n) = \tilde{\mathbf{U}}^T \Lambda^{1/2} \mathbf{U}^T \mathbf{v}(n)$, we have

$$[\boldsymbol{\Xi}_{VV}(\infty)]_{i,i} \approx (1 - 2\alpha(\infty) \tilde{\lambda}_i^{-1} I_i(\tilde{\Lambda}) + 2C(\sigma_e^2) \tilde{\lambda}_i^{-2} I_i''(\tilde{\Lambda})) [\boldsymbol{\Xi}_{VV}(\infty)]_{i,i} + 2C(\sigma_e^2) \tilde{\lambda}_i^{-2} I_i''(\tilde{\Lambda}) \Delta \mathbf{W}_{L,i}^2 + B_{\boldsymbol{\psi}}(\sigma_e^2) \tilde{\lambda}_i^{-2} [\Gamma_1]_{i,i} + [\Gamma_2]_{i,i}, \quad (21)$$

where $\mathbf{L}_X^T \Delta \mathbf{W} = \Delta \mathbf{W}_L$, $\Gamma_1 = \tilde{\mathbf{U}}^T \mathbf{I}'(\tilde{\Lambda}) \tilde{\mathbf{U}}$ and $\Gamma_2 = \tilde{\mathbf{U}}^T \Lambda^{1/2} \mathbf{U}^T \mathbf{I}(\infty) \mathbf{U} \Lambda^{1/2} \tilde{\mathbf{U}} + 2C(\sigma_e^2) \tilde{\Lambda}^{-1} \tilde{\mathbf{U}}^T \Lambda^{-1} \Delta \mathbf{W}_L \Delta \mathbf{W}_L^T \circ \mathbf{I}''(\tilde{\Lambda}) \tilde{\Lambda}^{-1} \tilde{\mathbf{U}} \tilde{\Lambda}^{-1}$;

$$\begin{aligned} & \text{and } \mathbf{I}(\infty) = \frac{\mu}{L} \mathbf{R}_{E,X}^{-1} \mathbf{D}' \bar{\boldsymbol{\psi}}(\infty) \bar{\boldsymbol{\psi}}^T(\infty) \mathbf{D}' \mathbf{R}_{E,X}^{-1} \\ & + \frac{\sqrt{\mu}}{L} \mathbf{R}_{E,X}^{-1} [\alpha(\infty) \mathbf{U} \Lambda^{1/2} \tilde{\mathbf{U}} \mathbf{I}(\tilde{\Lambda}) \tilde{\mathbf{U}}^T \Lambda^{1/2} \mathbf{U}^T \Delta \mathbf{W} \bar{\boldsymbol{\psi}}^T(\infty) \mathbf{D}'] \mathbf{R}_{E,X}^{-1} \\ & + \frac{\sqrt{\mu}}{L} \mathbf{R}_{E,X}^{-1} [\alpha(\infty) \mathbf{D}' \bar{\boldsymbol{\psi}}(\infty) \Delta \mathbf{W}^T \mathbf{U} \Lambda^{1/2} \tilde{\mathbf{U}} \mathbf{I}(\tilde{\Lambda}) \tilde{\mathbf{U}}^T \Lambda^{1/2} \mathbf{U}^T] \mathbf{R}_{E,X}^{-1}. \end{aligned}$$

Assume the regularization is mild, then we have $\tilde{\mathbf{U}} \approx \mathbf{I}$ and the 4th term on the right hand side of (21) can be approximated as $2C(\sigma_e^2) \tilde{\Lambda}^{-1} (\boldsymbol{\Xi}_{vv}(n) \circ \mathbf{I}''(\tilde{\Lambda})) \tilde{\Lambda}^{-1}$. Consequently, we have

$$[\boldsymbol{\Xi}_{VV}(\infty)]_{i,i} \approx \frac{B_{\boldsymbol{\psi}}(\sigma_e^2) \tilde{\lambda}_i^{-2} [\Gamma_1]_{i,i} + [\Gamma_2]_{i,i}}{2\alpha(\infty) \tilde{\lambda}_i^{-1} I_i(\tilde{\Lambda}) - 2C(\sigma_e^2) \tilde{\lambda}_i^{-2} I_i''(\tilde{\Lambda})}. \quad (22)$$

The steady-state excess mean squares error (EMSE) is $J_* = \text{tr}(\boldsymbol{\Xi}_{vv}(\infty) \mathbf{R}_{xx}) = \text{tr}(\boldsymbol{\Xi}_{VV}(\infty) \tilde{\mathbf{U}}^T \Lambda^{-1/2} \mathbf{U}^T \mathbf{R}_{xx} \mathbf{U} \Lambda^{-1/2} \tilde{\mathbf{U}})$

$$= \text{tr}(\mathbf{E}_{VV}(n)) \approx \sum_{i=1}^L \frac{B_\psi(\sigma_e^2) \tilde{\lambda}_i^{-2} [\mathbf{I}_1]_{i,i} + [\mathbf{I}_2]_{i,i}}{2\alpha(\sigma_e^2) \tilde{\lambda}_i^{-1} I_i(\tilde{\Lambda}) - 2C_\psi(\sigma_e^2) \tilde{\lambda}_i^{-2} I_i''(\tilde{\Lambda})}. \quad (23)$$

If ATS is used, it can be shown that $S_\psi(\sigma_e^2) = B_\psi(\sigma_e^2) / \sigma_e^2$, $A_\psi(\sigma_e^2)$ and $C_\psi(\sigma_e^2)$ are nearly constant [9] and hence we have

$$J_* \approx \sum_{i=1}^L \frac{S_\psi \sigma_e^2(\infty) \tilde{\lambda}_i^{-2} [\mathbf{I}_1]_{i,i} + [\mathbf{I}_2]_{i,i}}{2A_\psi \tilde{\lambda}_i^{-1} I_i(\tilde{\Lambda}) - 2C_\psi \tilde{\lambda}_i^{-2} I_i''(\tilde{\Lambda})}. \quad (24)$$

Since $\sigma_e^2(\infty) = \sigma_{\eta_0}^2 + E[(\mathbf{v}(\infty) + \Delta\mathbf{W})^T \mathbf{R}_{XX} (\mathbf{v}(\infty) + \Delta\mathbf{W})] = \sigma_{\eta_0}^2 + \Delta\mathbf{W}^T \mathbf{R}_{XX} \Delta\mathbf{W} + E[\mathbf{E}_{vv}(\infty) \mathbf{R}_{XX}] = \sigma_{\eta_0}^2 + \Delta\mathbf{W}^T \mathbf{R}_{XX} \Delta\mathbf{W} + J_*$, one gets from (24) after some manipulation:

$$J_* \approx \frac{\frac{1}{2} \sigma_{\min}^2 \phi_{\text{RLS}} + \phi_{\text{RW}}}{(1 - \frac{1}{2} \phi_{\text{RLS}})}, \quad (25)$$

where $\phi_{\text{RLS}} = S_\psi \sum_{i=1}^L \frac{\tilde{\lambda}_i^{-2} [\mathbf{I}_1]_{i,i}}{A_\psi \tilde{\lambda}_i^{-1} I_i(\tilde{\Lambda}) - C_\psi \tilde{\lambda}_i^{-2} I_i''(\tilde{\Lambda})}$, and

$\phi_{\text{RW}} \approx \sum_{i=1}^L \frac{[\mathbf{I}_2]_{i,i}}{A_\psi \tilde{\lambda}_i^{-1} I_i(\tilde{\Lambda}) - C_\psi \tilde{\lambda}_i^{-2} I_i''(\tilde{\Lambda})}$. The results for the

contaminated Gaussian noise follows from the approach introduced in [9] and the details can be found in [13].

IV. EXPERIMENTAL RESULTS

Computer simulation of a system identification problem is used to evaluate the mean and mean square convergence analysis in section III. In addition, we shall consider the application of the R-QRRLM algorithm in an acoustic echo canceller to illustrate the usefulness of the M-estimate algorithm and regularization, respectively in suppressing the double talk and reducing the variance of the estimate at low signal level. All the simulation results were averaged over 200 runs.

A. Performance of Mean and Mean Square Convergence Analysis

In evaluating the convergence performance analysis, the randomly generated system impulse response is $\mathbf{W}^* = [0.3886, 0.1632, -0.8294, -0.9247, -0.6148]$. For the mean convergence, the norm of the mean weight-error vector is used as the performance measure: $\|\mathbf{v}(n)\|_2 = 10 \log(\sqrt{\sum_{i=1}^L [\frac{1}{K} \sum_{j=1}^K v_i^{(j)}(n)]^2})$, where $v_i^{(j)}(n)$ is the i -th component of the weight-error vector with respect to (wrt) the Wiener solution, $\mathbf{v}^*(n) = \mathbf{w}(n) - \mathbf{w}^*$, or the optimal solution $\mathbf{W}(\infty)$ in (17), $\mathbf{v}(n) = \mathbf{w}(n) - \mathbf{W}(\infty)$, at time n in the j -th independent run. K is the total number of independent runs, which is set to 5000 in this experiment. Fig. 1 plots the learning curves of $\|\mathbf{v}(n)\|_2$ for the L_2 R-QRRLS algorithm. For page limitation, the results of L_1 regularization are not presented. $\|\mathbf{v}^*\|_2$ denotes the error norm wrt to the Wiener solution while $\|\mathbf{v}_\infty\|_2$ denotes that to the optimal solution $\mathbf{W}(\infty)$. The settings are as follows: forgetting factor $\lambda = 0.98$, regularization parameter $\mu = 0.005, 0.05$ and SNR=10 dB. It can be seen from the two $\|\mathbf{v}^*\|_2$ curves that the bias wrt the Wiener solution increases with μ . Since $\|\mathbf{v}_\infty\|_2$ converge to

a lower value than the $\|\mathbf{v}^*\|_2$, it suggests that the algorithm converges to the optimal solution $\mathbf{W}(\infty)$ rather than to the Wiener solution. Simulations with different forgetting factors give similar results and are not presented here for page limitation. The steady-state EMSE of the simulation and the theoretical analysis are compared in Tables II and III for both white Gaussian input and first order autoregressive input $x(n) = 0.9x(n-1) + g(n)$, where $g(n)$ is a zero-mean and white Gaussian noise. In the simulation, we set $\lambda = 0.995, 0.99, 0.98$, $\mu/L = 0.0001, 0.001$ and SNR=0 dB, 10 dB. The results show that the analysis slightly underestimates the steady-state EMSE because of the independent assumption used.

B. Performance of R-QRRLM in AEC

In this experiment, the performance of the proposed R-QRRLM and VR-QRRLM algorithms are compared with the RLS algorithm in the adaptive echo cancellation problem. The AEC system has a similar structure with that of the system identification except that the additive noise is replaced by the near-end source signal which is to be retained. The impulse response is shown in Fig. 2(a) and its length is 100. The input signal is a segment of music as shown in Fig. 2(b). It is assumed that there is a double talk between 3000 to 3100-th samples where the algorithm has almost converged in order to visualize the adverse effect of the double talk. The SNR is 10 dB. The forgetting factor λ for the RLS algorithm is chosen to be 0.997, resulting in an exponential data window as long as 3 times of the impulse response. For page limitation, the impulse response is assumed to be time-invariant and only L_2 regularization is tested. In cases of time-varying impulse response, L_1 regularization can be easily imposed to improve the performance [8]. The forgetting factor and regularization parameter for the R-QRRLM algorithm are: $\lambda = 0.997$, $\mu/L = 0.1/0.001$, and the parameters for estimating the noise variance in (3) are $\lambda_\sigma = 0.9$ and $N_w = 200$. The performances of various algorithms are shown in Fig. 2(c). It can be seen that the RLS algorithm is very sensitive to the level of the input signal and the double talk which resembles a long series of impulsive noise. For the R-QRRLM algorithm, if a small regularization parameter, say $\mu = 0.001$, is used, it converges to a lower EMSE value; on the other hand, if μ is increased to a comparatively large value, i.e. $\mu = 0.1$, the algorithm becomes much less sensitive to the input signal power variation but converges to a higher EMSE value. The VR-QRRLM algorithm, however, adaptively selects the regularization parameters and obtains both high immunity to input variation and low steady-state EMSE values. In addition, algorithms using M-estimation show the robustness in impulsive noise environment if parameters are appropriately selected.

V. CONCLUSION

A QR decomposition based regularized RLM algorithm and its mean and mean square convergence performance analysis are presented. The mean convergence analysis suggests that the algorithm solves the desired regularized solution of the Wiener solution with an additional bias. The

mean square convergence analysis gives the steady-state EMSE and suggests that the variance of estimation will decrease while the bias will increase with the regularization parameter. The theoretical results are in good agreement with those by computer simulation. Results for the AEC show that the R-QRRLM and VR-QRRLM algorithms outperform the traditional RLS algorithm at low input signal level or during double talk.

References

[1] S. Haykin, *Adaptive Filter Theory*, 4th edition, Prentice Hall, 2001.
 [2] Y. Zhou, S. C. Chan, and K. L. Ho, "A new variable forgetting factor QR-based recursive least M-estimate algorithm for robust adaptive filtering in impulsive noise environment," in *EUSIPCO 2006*, Florence, Italy, Sep. 4-8, 2006.
 [3] C. Yilun, G. Yuantao, and A. O. Hero, "Sparse LMS for system identification," in *Proc. ICASSP2009*, Taipei, Taiwan, Apr 19-24, 2009, pp. 3125-3128.
 [4] P. Zhao and B. Yu, "On model selection consistency of lasso," *J. Mach. Learning Res.*, vol. 7, pp. 2541-2563, 2006.
 [5] S. Chen, D. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit," *SIAM Rev.*, vol. 43, no. 1, pp. 129-159, 2001.
 [6] E. Candès, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Commun. Pure Appl. Math.*, vol. 59, no. 8, pp. 1207-1223, 2005.

[7] D. Donoho, M. Elad, and V. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *IEEE Trans. Inform. Theory*, vol. 52, no. 1, pp. 6-18, Jan. 2006.
 [8] Z. G. Zhang, S. C. Chan, Y. Zhou, and Y. Hu, "Robust linear estimation using M-estimation and weighted L1 regularization: Model selection and recursive implementation," in *Proc. ISCAS2009*, Taipei, Taiwan, May 24-27, 2009, vol. 5, pp. 1193-1196.
 [9] S. C. Chan and Y. Zhou, "On the convergence analysis of the normalized LMS and the normalized least mean M-estimate algorithms," *J. Signal Process. Syst.*, Oct. 2009.
 [10] Y. Zou, S. C. Chan and T. S. Ng, "A recursive least Mestimate (RLM) adaptive filter for robust filtering in impulse noise," *IEEE Signal Process. Letters*, vol. 7, no. 11, 2000.
 [11] S. C. Chan and Y. Zou, "A recursive least M-estimate algorithm for robust adaptive filtering in impulsive noise: fast algorithm and convergence performance analysis," *IEEE Trans. Signal Process.*, vol. 52, no. 4, pp. 975-991, April 2004.
 [12] Y. Zou, S. C. Chan, and T. S. Ng, "Least mean M-estimate algorithms for robust adaptive filtering in impulse noise," *IEEE Trans. Circuits Syst. II*, vol. 47, no. 12, Dec. 2000, pp. 1564-1569.
 [13] S. C. Chan, "Regularized QRD-based Recursive Least M-estimate (RLM) Algorithm and its Performance Analysis," internal report. Department of EEE, The University of Hong Kong.
 [14] Z. Tian, K. L. Bell, and H. L. Van Trees, "A recursive least squares implementation for LCMP beamforming under quadratic constraint," *IEEE Trans. Signal Process.*, vol. 49, no. 6, pp. 1138-1145, Jun. 2001.
 [15] T. Van Waterschoot, G. Rombouts, and M. Moonen, "Optimally regularized adaptive filtering algorithms for room acoustic signal enhancement," *Signal Process.*, vol. 88, pp. 594-611, 2008.

TABLE II. STEADY-STATE EMSE FOR WHITE GAUSSIAN INPUT

L	μ/L	SNR = 0 dB						SNR = 10 dB					
		0.0001			0.001			0.0001			0.001		
		λ	0.995	0.99	0.98	0.995	0.99	0.98	0.995	0.99	0.98	0.995	0.99
5	Simulation	-15.91	-12.85	-9.69	-15.91	-12.79	-9.60	-25.80	-22.71	-19.71	-25.75	-22.66	-19.66
	Analysis	-15.84	-12.89	-9.94	-15.84	-12.86	-9.91	-25.75	-22.79	-19.84	-25.69	-22.64	-19.62
25	Simulation	-9.10	-6.11	-2.90	-9.00	-6.10	-2.80	-18.99	-15.90	-12.76	-18.98	-15.90	-12.72
	Analysis	-9.20	-6.42	-3.74	-9.24	-6.42	-3.82	-19.21	-16.42	-13.82	-19.21	-16.41	-13.78
50	Simulation	-6.01	-2.98	0.46	-6.00	-2.79	0.47	-16.02	-12.89	-9.90	-16.00	-12.85	-9.57
	Analysis	-6.45	-3.92	-1.62	-6.46	-3.88	-1.61	-16.49	-13.89	-11.50	-16.48	-13.88	-11.50

TABLE III. STEADY-STATE EMSE FOR FIRST-ORDER AR INPUT

L	μ/L	SNR = 0 dB						SNR = 10 dB					
		0.0001			0.001			0.0001			0.001		
		λ	0.995	0.99	0.98	0.995	0.99	0.98	0.995	0.99	0.98	0.995	0.99
5	Simulation	-14.10	-10.90	-8.02	-13.80	-10.50	-7.60	-23.80	-20.80	-17.90	-23.50	-20.50	-17.50
	Analysis	-13.91	-11.06	-7.90	-13.90	-10.70	-7.81	-23.80	-20.73	-17.81	-23.60	-20.70	-17.80
25	Simulation	-6.73	-3.62	-0.40	-6.69	-3.60	-0.32	-16.65	-13.58	-10.40	-16.59	-13.52	-10.37
	Analysis	-6.82	-4.05	-1.41	-6.80	-4.04	-1.40	-16.83	-14.05	-11.41	-16.79	-14.00	-11.33
50	Simulation	-7.88	-4.64	-1.37	-7.85	-4.60	-1.33	-18.00	-14.74	-11.38	-17.80	-14.69	-11.01
	Analysis	-8.10	-5.50	-3.20	-8.10	-5.50	-3.17	-18.12	-15.54	-13.20	-18.11	-15.49	-12.99

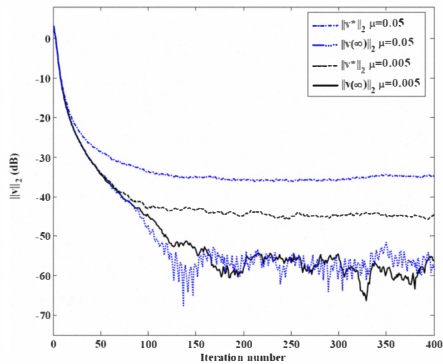


Figure 1. Convergence curves for $\|v\|_2$ with different regularization parameters.

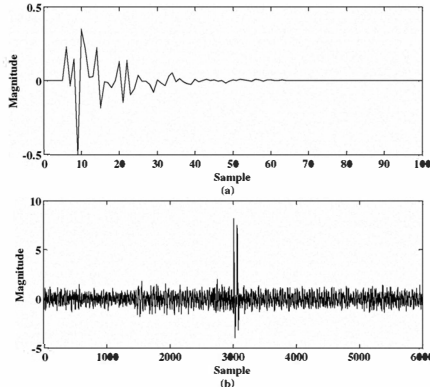


Figure 2. Performance comparison of various algorithms for the AEC problem in experiment 3. (a) impulse response of length 100. (b) the input music signal. (c) the learning curves of EMSE in 10 dB for various algorithms.