The HKU Scholars Hub The University of Hong Kong 香港大學學術庫



Title	A New Variable Regularized QR Decomposition-Based Recursive Least M-Estimate Algorithm-Performance Analysis and Acoustic Applications
Author(s)	Chan, SC; Chu, Y; Zhang, Z; Tsui, KM
Citation	IEEE Transactions on Audio, Speech and Language Processing, 2013, v. 21, p. 907-922
Issued Date	2013
URL	http://hdl.handle.net/10722/189061
Rights	Creative Commons: Attribution 3.0 Hong Kong License

A New Variable Regularized QR Decomposition-Based Recursive Least M-Estimate Algorithm—Performance Analysis and Acoustic Applications

S. C. Chan, Member, IEEE, Y. J. Chu, Z. G. Zhang, Member, IEEE, and K. M. Tsui

Abstract—This paper proposes a new variable regularized QR decomposition (QRD)-based recursive least M-estimate (VR-QRRLM) adaptive filter and studies its convergence performance and acoustic applications. Firstly, variable L_2 regularization is introduced to an efficient QRD-based implementation of the conventional RLM algorithm to reduce its variance and improve the numerical stability. Difference equations describing the convergence behavior of this algorithm in Gaussian inputs and additive contaminated Gaussian noises are derived, from which new expressions for the steady-state excess mean square error (EMSE) are obtained. They suggest that regularization can help to reduce the variance, especially when the input covariance matrix is ill-conditioned due to lacking of excitation, with slightly increased bias. Moreover, the advantage of the M-estimation algorithm over its least squares counterpart is analytically quantified. For white Gaussian inputs, a new formula for selecting the regularization parameter is derived from the MSE analysis, which leads to the proposed VR-QRRLM algorithm. Its application to acoustic path identification and active noise control (ANC) problems is then studied where a new filtered-x (FX) VR-QRRLM ANC algorithm is derived. Moreover, the performance of this new ANC algorithm under impulsive noises and regularization can be characterized by the proposed theoretical analysis. Simulation results show that the VR-QRRLM-based algorithms considerably outperform the traditional algorithms when the input signal level is low or in the presence of impulsive noises and the theoretical predictions are in good agreement with simulation results.

Index Terms—Adaptive filters, ANC, performance analysis, recursive M-estimation, variable regularization.

I. INTRODUCTION

S YSTEM identification is frequently encountered in many applications such as acoustics [1], communications, power system measurement in industrial and biomedical engineering/informatics, etc. Recursive least-squares (RLS) algorithm is an effective adaptive filtering algorithm which has been widely used in these applications. Traditional RLS algorithms estimate

The authors are with the Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong (e-mail: scchan@eee.hku.hk; yjchu@eee.hku.hk; zgzhang@eee.hku.hk; kmtsui@eee.hku.hk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TASL.2012.2236315

the 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, the performance of conventional LS-based adaptive filters will be considerably degraded in impulsive noise environment [2], [3]. To address this problem, the robust least mean square M-estimate (LMM) [4], [5] or the RLS M-estimate (RLM) [6], [7] algorithms were proposed, so that the adverse effect of impulsive noises can be effectively mitigated. To improve the numerical stability of the RLM algorithms in finite wordlength implementation, a QRD-based RLM (QRRLM) algorithm was developed [8].

Another possible problem with the RLS-based algorithms is that the covariance matrix of input signals may become poorly conditioned or even singular. This is often encountered in acoustic applications such as adaptive echo cancellation (AEC) where the level of the excitation signal may vary significantly over times. In such situation, most RLS-like algorithms will show dramatically increased estimation variance or even suffer from instability. To address this ill-conditioned problem, a commonly used technique is to introduce some kind of regularization into these algorithms. In fact, regularization techniques have attracted much interest recently as a useful tool for reducing the estimation variance [9], especially when only a small number of data samples are available [10]. They have also been successfully applied to a wide variety of areas such as audio signal processing, etc, to improve the convergence speed and decrease the steady-state misadjustment in conventional RLS [11]-[15] and LMS-based [16]-[19] adaptive filtering algorithms. Most of the above regularized RLS algorithms are very sensitive to the round off error. It is therefore desirable to devise a regularized RLS algorithm using the QRD, which has much better numerical properties and can be effectively implemented in hardware without multiplications using the coordinate rotation digital computer (CORDIC) algorithm.

In this paper, a new VR-QRRLM algorithm is proposed. It employs M-estimation in combating impulsive noises and a variable weighted L_2 regularization to reduce the estimation variance. An efficient QRD implementation that may improve the numerical stability is also considered. In addition, to quantify the performance of the R-QRRLM algorithm, the mean and mean square convergence analyses with a fixed regularization parameter for Gaussian inputs and additive noises are first performed. They rely on the Price's theorem [20] and the generalized Abelian integral functions recently

Manuscript received November 23, 2011; revised March 22, 2012 and August 30, 2012; accepted August 30, 2012. Date of publication January 23, 2013; date of current version February 01, 2013. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Lauri Savioja.

introduced in [4], [21] to evaluate various expectations. While convergence behaviors of the RLS and the RLM algorithms have been considered previously in [22] and [7] respectively, the performance of the regularized case is considerably more complicated. Using an extension of the Price's theorem for contaminated Gaussian (CG) noise [23], we are able to analyze the performance of the R-QRRLM algorithm. A similar approach has been successfully applied to the analysis of the NLMS and M-estimate NLMS (NLMM) [21] algorithms for Gaussian inputs and CG noise. Difference equations describing the mean and mean square performance of the algorithm for Gaussian inputs and Gaussian or CG additive noises are obtained. Moreover, the convergence of these equations is thoroughly analyzed to obtain analytical expressions for the steady-state excess mean square error (EMSE). By analyzing the effect of L_2 regularization parameter on the steady-state MSE of the weight vector for white Gaussian inputs, a new formula for selecting the regularization parameter is obtained. This gives rise to the proposed VR-QRRLM algorithm.

In the preliminary version of the R-QRRLM algorithm [19], the regularization parameter is determined using the cross validation as is commonly used in the statistical communities [10]. This is later extended by transforming the input and using smoothly-clipped-absolute-deviation (SCAD) [25] to further promote sparsity. Similar to this performance analysis, we further exploit in this work the simplicity offered by white Gaussian inputs so as to derive an explicit formula for the regularization parameter of the VR-QRRLM algorithm.

To illustrate the usefulness of the R-QRRLM and VR-QRRLM algorithms, their applications to system identification and ANC problems are studied. In particular, a new FX-VR-QRRLM ANC algorithm is proposed with its convergence performance characterized. The new FX-VR-QRRLM employs the error signal, which is the only signal available from the error microphone, directly for updating the weight vector. Thus, improved robustness in CG noises and better convergence performance over conventional ANC realizations using the approximated desired signal can be obtained. The proposed theoretical analysis helps to address challenging problems in its performance analysis and related algorithms under impulsive noises and regularization. It is also found to be in good agreement with computer simulation.

It should be noted that we considerably extend our previous work in [13] in the following aspects. Different from the brief analytic results in [13], the detailed derivations of the difference equations are provided and the effect of the regularization term on the tradeoff between variance and bias is thoroughly investigated. Also, the analysis is further extended to the case of CG noises and instead of choosing the regularization parameter using an empirical formula as in [13], an analytic formula is derived to determine this parameter by considering the steady-state MSE deviation of the regularized optimal solution from the Wiener solution. Apart from these contributions, extensive simulations have been carried out to demonstrate the effectiveness of the proposed algorithm as well as the performance analyses. In particular, a practical application of the proposed algorithm to an ANC system is considered and a new variant of the proposed algorithm, called FX-VR-QRRLM algorithm, is developed.

The rest of the paper is organized as follows. In Section II, the R-QRRLM algorithm is proposed. Section III is devoted to the performance analysis of the R-QRRLM algorithm and the selection of the regularization parameter for the VR-QRRLM algorithm is addressed in Section IV. Computer simulations for verifying the theoretical analyses and comparison with other RLS-like algorithms are presented in Section V. Its applications to acoustic system identification and ANC will also be presented. Finally, conclusions are drawn in Section VI.

II. REGULARIZED QRRLM ALGORITHM

A. QRRLM Algorithm

Consider the adaptive system identification problem where an input signal x(n) is applied simultaneously to an *L*-order adaptive transversal filter with weight vector $\boldsymbol{W}(n) = [w_1(n), w_2(n), \dots, w_L(n)]^T$ and an unknown system to be identified with an impulse response $\boldsymbol{H}_0 = [h_1, h_2, \dots, h_L]^T$. Let $\boldsymbol{X}(n) = [x(n), x(n-1), \dots x(n-L+1)]^T$ be the input vector. Then the output of the adaptive filter is $y(n) = \boldsymbol{X}^T(n)\boldsymbol{W}(n)$. The measured output of the system is used as the desired signal d(n) of the adaptive filter

$$d(n) = \boldsymbol{X}^{T}(n)\boldsymbol{H}_{0} + \eta(n), \qquad (1)$$

where $\eta(n)$ denotes the additive noise or possible modeling error. The adaptive filter aims to minimize error measurement of the estimation error e(n) = d(n) - y(n).

In the RLM algorithm [6], [7], the following M-estimate cost function is minimized

$$J_{\rho}(n) = \sum_{i=0}^{n} \lambda_{n-i}(n) \rho\left(e(i)\right), \qquad (2)$$

where $\lambda_{n-i}(n)$ serves the purpose of an exponential window which puts less weight to errors at distant past. For example, it can be chosen as λ^{n-i} , where λ is a constant forgetting factor (FF) or can be updated adaptively as in variable FF algorithms [22]. $\rho(\cdot)$ is an M-estimate function such as the modified Huber (MH) function in (3) or other appropriate functions, say the Hampel's three part redescending function [26] or the bisquare functions [27],

$$\rho(e) = \begin{cases} \frac{e^2}{2} & 0 \le |e| < \xi, \\ |e| - \xi + \xi^2/2 & |e| \ge \xi. \end{cases}$$
(3)

 ξ is a threshold used to control the suppression of outliers and adaptation speed. It can be seen that large estimation error, which may be due to outliers, is significantly suppressed by the M-estimate function. In the adaptive threshold selection (ATS) method proposed in [6], [7], e(n) is assumed to be Gaussian distributed except being corrupted occasionally by additive impulsive noises. By estimating the variance of "impulse-free" estimation error $\hat{\sigma}^2(n)$, it is possible to detect and reject the impulses in e(n). Specifically, the probability of |e(n)| greater than a given threshold T is $\operatorname{erfc}(T\hat{\sigma}(n)/\sqrt{2})$, where erfc is the complementary error function, and $\hat{\sigma}(n)$ is the standard deviation of $\hat{\sigma}^2(n)$. By using the robust variance estimate [6]:

$$\hat{\sigma}^2(n) = \lambda_{\sigma} \hat{\sigma}^2(n-1) + c_1(1-\lambda_{\sigma}) \text{median} \left[A_e(n)\right], \quad (4)$$

the following adaptive threshold can be obtained:

$$\xi = k_{\xi} \hat{\sigma}^2(n), \tag{5}$$

where $\operatorname{median}[A_e(n)]$ is the median value of the vector $A_e(n) = \{e^2(n), \ldots, e^2(n - N_w + 1)\}, N_w$ is the length of the estimation window, $c_1 = 2.13$ is a finite sample correction factor, λ_{σ} is a positive FF close to but smaller than one and k_{ξ} is a constant used to control the suppression of impulsive interference and the corresponding reduction in convergence rate. A reasonable value of k_{ξ} is 2.576 and the window length N_w is usually chosen between 5 and 9 [5]. It should be noted that, the error nonlinearity of the M-estimation function (M-nonlinearity) helps to suppress the impulsive noise in exchange for a slightly slower adaptation rate than that in the Gaussian noise, which has been studied in [24]. Therefore, the performance of the M-estimation based algorithms, such as convergence speed, is generally very slightly degraded in the Gaussian noise environment, as compared to its LS counterparts.

By setting the first partial derivative of $J_{\rho}(n)$ with respect to (w.r.t.) $\boldsymbol{W}(n)$ to zero, it was shown in [6] that the optimal weight vector satisfies the M-estimate normal equation:

$$\mathbf{R}_{X\rho}(n)\boldsymbol{W}(n) = \mathbf{P}_{X\rho}(n),\tag{6}$$

where $\mathbf{R}_{X\rho}(n) = \sum_{i=0}^{n} \lambda_{n-i}(n)q(e(i))\boldsymbol{X}(i)\boldsymbol{X}^{T}(i)$ and $\mathbf{P}_{X\rho}(n) = \sum_{i=0}^{n} \lambda_{n-i}(n)q(e(i))d(i)\boldsymbol{X}(i)$ are the M-estimate covariance matrix of $\boldsymbol{X}(n)$ and the M-estimate cross-correlation vector of d(n) and $\boldsymbol{X}(n)$, respectively, and $q(e) = \rho'(e)/e = [d\rho(e)/de]/e$. 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 encountered, $\lambda_{n-i}(n)$ can be updated as follows [7]

$$\lambda_{n-i}(n) = \lambda_e(n)\lambda_{n-i-1}(n-1), \tag{7}$$

where $\lambda_e(n) = \lambda$ if $|e(n)| < \xi$, or $\lambda_e(n) = 1$ otherwise.

Applying the iterative reweighted LS method to (6), the following RLM algorithm can be obtained [8]

$$\mathbf{P}(n) = \lambda_e^{-1}(n) \left(\boldsymbol{I} - \boldsymbol{K}(n) \boldsymbol{X}^T(n) \right) \mathbf{P}(n-1),$$

$$\boldsymbol{K}(n) = \frac{q(e(n)) \mathbf{P}(n-1) \boldsymbol{X}(n)}{\lambda_e(n) + q(e(n)) \boldsymbol{X}^T(n) \mathbf{P}(n-1) \boldsymbol{X}(n)},$$

$$\boldsymbol{W}(n) = \boldsymbol{W}(n-1) + \left[d(n) - \boldsymbol{W}^T(n-1) \boldsymbol{X}(n) \right] \boldsymbol{K}(n), \quad (8)$$

where $\mathbf{P}(n)$ is the recursive update of $R_{X\rho}^{-1}(n)$. For the LS cost function, q(e(n)) = 1, whereas for the MH cost function, q(e(n)) = 1 when |e(n)| is smaller than the threshold ξ , and zero otherwise. For the latter, the factor q(e(n)) = 1 in the denominator of $\mathbf{K}(n)$ can be dropped without affecting the solution. Eqn. (8) can also be efficiently implemented using a QR-based implementation [8] as summarized in Tables I (the first QRD only). This QRRLM algorithm is mathematically equivalent to but has higher numerical stability than the RLM algorithm. Also, it is known that the arithmetic complexity of the QRRLM algorithm is of the order $O(L^2)$ [8].

B. R-QRRLM and VR-QRRLM

In some acoustic and related applications, the input to the adaptive filter is speech or other acoustic signals. Hence, the adaptive filter may not be persistently excited when the input

TABLE I THE R-QRRLM/VR-QRRLM ALGORITHM

Initialization:
$\mathbf{R}(0) = \sqrt{\delta} \mathbf{I}$, δ is a small positive constant;
$\mathbf{U}(0) = 0 ; W(0) = 0;$
Recursion:
Given $\mathbf{R}(n-1)$, $\mathbf{U}(n-1)$, $W(n-1)$, $X(n-1)$ and $d(n)$,
compute at time <i>n</i> : (i) The first undete:
$\begin{bmatrix} \mathbf{R}^{(1)}(n) & \mathbf{U}^{(1)}(n) \\ 0 & c^{(1)}(n) \end{bmatrix} = \mathbf{Q}^{(1)}(n) \begin{vmatrix} \sqrt{\lambda_e} \mathbf{R}(n-1) & \sqrt{\lambda_e} \mathbf{U}(n-1) \\ \sqrt{q(e(n))} \mathbf{X}^{T}(n) & \sqrt{q(e(n))} d(n) \end{vmatrix}$
The second update:
$\begin{bmatrix} \mathbf{R}(n) & \mathbf{U}(n) \\ 0 & c(n) \end{bmatrix} = \mathbf{Q}(n) \begin{bmatrix} \mathbf{R}^{(1)}(n) & \mathbf{U}^{(1)}(n) \\ \sqrt{q(e(n))\mu(n)}d_1 & 0 \end{bmatrix},$
where $\mathbf{Q}^{(1)}(n)$ and $\mathbf{Q}(n)$ are calculated by Givens rotation.
(ii). $W(n) = \mathbf{R}^{-1}(n)\mathbf{U}(n)$ (back-substitution).

signal level is very low. Consequently, the covariance matrix $\mathbf{R}_{X\rho}(n)$ may be ill-conditioned and a large estimation variance will result. To address this potential problem, we propose to include a regularization term on the adaptive filter coefficients in the M-estimate objective function to limit the variation in the coefficient vector $\boldsymbol{W}(n)$. This gives the following regularized M-estimate objective function:

$$J_{R_{-\rho}}(n) = J_{\rho}(n) + \sum_{i=1}^{L} p_{\kappa}(w_i(n)), \qquad (9)$$

where $p_{\kappa}(.)$ is the regularization function with κ being the regularization parameter. The commonly-used regularization functions include the L_2 regularization function: $p_{\kappa}(w_i(n)) = \kappa w_i^2(n)$; L_1 regularization function: $p_{\kappa}(w_i(n)) = \kappa |w_i(n)|$, and SCAD function [10]. To solve (9), the iterative reweighted LS method is frequently used, where the regularization function $p_{\kappa}(.)$ is approximated locally by a quadratic function of $w_i(n)$ and the resultant objective function becomes

$$\hat{J}_{R_{-\rho}}(n) = J_{\rho}(n) + \kappa \|\mathbf{D}(n)\mathbf{W}(n)\|_{2}^{2},$$
(10)

where $\mathbf{D}(n)$ is a weighting matrix used to approximate different regularization methods. For instance, $\mathbf{D}(n)$ is an identity matrix for L_2 regularization or the generalized inverse of the matrix $diag\{\sqrt{|w_1(n-1)|}, \dots, \sqrt{|w_L(n-1)|}\}$ for L_1 regularization. Unlike the cost function in (2), the regularized problem in (9) or (10) cannot be solved simply by the recursive QRD algorithm due to the additional regularization term. A solution is to apply the regularization successively using QRD. More precisely, at each time instant, the QRD is executed once for the data vector $\sqrt{q(e(n))}[\mathbf{X}^T(n), d(n)]$ and once for the regularization vector $\sqrt{q(e(n))}[\sqrt{\mu}\mathbf{d}_l, 0]$, where \mathbf{d}_l is the *l*-th row of the regularization matrix $\mathbf{D}(n)$ [19] and $\mu = \kappa L$. If the vector is applied sequentially, then $l = (n \mod L) + 1$. Therefore the complexity of the R-QRRLM algorithm is twice that of the QRRLM algorithm.

If the regularization parameter $\kappa(n)$ is made variable at each iteration, the R-QRRLM algorithm yields the VR-QRRLM algorithm. The selection of variable regularization parameter will be presented later in Section IV, since it is developed from the mean square convergence analysis introduced in Section III. As

will be seen later in Section IV, since only a few extra multiplications are needed for calculating $\kappa(n)$, the complexity of the VR-QRRLM algorithm is comparable to that of the R-QRRLM algorithm.

III. PERFORMANCE ANALYSIS OF R-QRRLM

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

$$\boldsymbol{W}(n+1) = \boldsymbol{W}(n) + \mathbf{R}_{X\rho}^{-1}(n) \left[\frac{\boldsymbol{X}(n)\psi(\boldsymbol{e}(n))}{\lambda + \boldsymbol{X}^{T}(n)\mathbf{R}_{E,\rho}^{-1}(n)\boldsymbol{X}(n)} + \frac{\boldsymbol{r}(n)\psi(\boldsymbol{e}_{r}(n))}{\lambda + \boldsymbol{r}^{T}(n)\mathbf{R}_{E,\rho}^{-1}(n)\boldsymbol{r}(n)} \right], \quad (11)$$

where $e_r(n) = -\sqrt{\mu} \boldsymbol{W}^T(n) \boldsymbol{d}_k^T$, $\boldsymbol{r}(n) = \sqrt{\mu} \boldsymbol{d}_k^T$. For notational simplicity, we have also used $\psi(e) = \rho'(e)$, $\mathbf{R}_{X\rho}^{-1}(n) = \mathbf{P}(n)$ and $\mathbf{R}_{E,\rho}^{-1}(n) = q(e(n))\mathbf{P}(n)$. If the MH cost function or non-linearity is used, $\mathbf{R}_{E,\rho}^{-1}(n)$ can be chosen as $\mathbf{R}_{X\rho}^{-1}(n)$ without affecting the update. The last two terms on the right hand side correspond, respectively, to the data and regularization updates.

We shall study the case for Gaussian inputs and additive noises first and then extend this to CG noises later in Section III-C. For simplicity, we shall consider the case with fixed FF and use the following assumptions:

(A1) $\{X(n)\}\$ is zero-mean Gaussian distributed with covariance matrix \mathbf{R}_{XX} ;

(A2) $\{\eta(n)\}$ is white Gaussian-distributed with zero-mean and variance σ_n^2 and it is uncorrelated with $\{\boldsymbol{X}(n)\}$;

(A3) the weight error vector $\{\boldsymbol{W}(n)\}$ is independent of

 $\{\boldsymbol{X}(n)\} \text{ and } \{\boldsymbol{\eta}(n)\};$ (A4) $\mathbf{P}(n) = \mathbf{R}_{X\rho}^{-1}(n) \approx \mathbf{R}_{E,\mathbf{X}}^{-1}(n) \approx \mathbf{R}_{E,\mathbf{X}}^{-1}(n),$ where $\mathbf{R}_{E,X}(n) = \lambda \mathbf{R}_{E,X}(n-1) + [\boldsymbol{X}(n)\boldsymbol{X}^{T}(n) + N]$ $\boldsymbol{r}(n)\boldsymbol{r}^{T}(n)];$

(A5) The MH or LS cost function is used.

(A3) is the independent assumption which is commonly introduced to simplify the analysis. On the other hand, (A4) assumes that the robust covariance estimate $\mathbf{R}_{X\rho}^{-1}(n)$ is close to the conventional covariance estimate $\mathbf{R}_{E,X}^{-1}(n)$, which in turn can be approximated by its expected value. (A1) has been verified in [7] while (A2) is commonly used in the analysis of the RLS [22] and RLM algorithms. From (A4), we can evaluate $\mathbf{R}_{X\rho}^{-1}(n)$ by using the following identity: $\mathbf{I} = E[\mathbf{R}_{E,X}(n)\mathbf{R}_{E,X}^{-1}(n)]$, where $\mathbf{R}_{E,X}(n) = \sum_{l=1}^{n} \lambda^{n-l} (\boldsymbol{X}(l)\boldsymbol{X}^{T}(l) + \boldsymbol{r}(l)\boldsymbol{r}^{T}(l))$. According to the averaging principle, $\mathbf{R}_{E,X}^{-1}(n)$ is assumed to be independent of $\boldsymbol{X}(l)\boldsymbol{X}^{T}(l)$ and $\boldsymbol{r}(l)\boldsymbol{r}^{T}(l)$. Hence, we have

$$\frac{1-\lambda^n}{1-\lambda} \left(\mathbf{R}_{XX} + \frac{\mu}{L} \mathbf{D}^2 \right) E \left[\mathbf{R}_{E,X}^{-1}(n) \right] = \mathbf{I}.$$
 (12)

Here, $\mathbf{D}(n)$ is assumed to be a constant matrix \mathbf{D} . Therefore, $E[\mathbf{R}_{E,X}^{-1}(n)] = [(1-\lambda)/(1-\lambda^{n})](\mathbf{R}_{XX} + (\mu/L)\mathbf{D}^{2})^{-1} \text{ and} \\ \lim_{k \to \infty} E[\mathbf{R}_{E,X}^{-1}(n)] = \mathbf{R}_{E,X}^{-1} = (1-\lambda)(\mathbf{R}_{XX} + (\mu/L)\mathbf{D}^{2})^{-1}$ for large n.

A. Mean Convergence Analysis

We now present the mean convergence behavior of the proposed algorithm with the LS or MH cost function for Gaussian inputs and additive noises. First, we assume that the algorithm is convergent. The condition of convergence will be shown later in this section. Then, we can subtract the optimal solution of the R-QRRLM algorithm, $W(\infty)$, from (11), and take the expectation on its both sides over $\{v, X, \eta\}$ to obtain:

$$E\left[\boldsymbol{v}(n+1)\right] = E\left[\boldsymbol{v}(n)\right] - \mathbf{R}_{E,\boldsymbol{X}}^{-1}(n)\left(\boldsymbol{L}_{1} + \boldsymbol{L}_{1}'\right), \quad (13)$$

where $\boldsymbol{L}_1 = E[\boldsymbol{X}(n)\psi(e(n))/(\lambda + \boldsymbol{X}^T(n)\mathbf{R}_{E,X}^{-1}(n)\boldsymbol{X}(n))],$ $\boldsymbol{L}_1' = E[\boldsymbol{r}(n)\psi(e_r(n))/(\lambda + \boldsymbol{r}^T(n)\mathbf{R}_{E,X}^{-1}(n)\boldsymbol{r}(n))],$ and $\boldsymbol{v}(n) = \boldsymbol{W}(\infty) - \boldsymbol{W}(n)$ is the weight error vector. For notational simplicity, the expectation over $\{\boldsymbol{v}(n), \boldsymbol{X}(n), \eta(n)\},\$ $E_{\{v,X,\eta\}}[.]$, is denoted as E[.]. Note, we have used the approximation (A4) in obtaining (13) and the MH and LS cost functions assumed in (A5) allow us to drop the factor q(e(n)) in the denominator to simplify the evaluation. For other M-nonlinearity [21], it also serves as a good approximation. By dropping the time index of X, e, and η , and using (A3), L_1 and L'_1 can be simplified to

$$\boldsymbol{L}_{1} = E_{\{\boldsymbol{v}\}} \left\{ E_{\{\boldsymbol{X},\eta\}} \left[\frac{\boldsymbol{X}\psi(e)}{\lambda + \boldsymbol{X}^{T} \mathbf{R}_{E,X}^{-1}(n) \boldsymbol{X}} \middle| \boldsymbol{v} \right] \right\} \equiv E_{\{\boldsymbol{v}\}}[\boldsymbol{A}], \quad (14)$$

$$\boldsymbol{L}_{1}^{\prime} = \frac{\sqrt{\mu}}{L} \sum_{k=1}^{\infty} E_{\{\boldsymbol{r}\}} \left\{ E_{\{\boldsymbol{r},\eta\}} \left[\frac{\boldsymbol{d}_{k}^{*} \psi(\boldsymbol{e}_{r})}{\lambda + \mu \boldsymbol{d}_{k} \mathbf{R}_{E,X}^{-1}(n) \boldsymbol{d}_{k}^{T}} \middle| \boldsymbol{v} \right] \right\}.$$
 (15)

Moreover, it is shown in Appendix A that

$$\boldsymbol{A} \approx A_{\psi} \left(\sigma_e^2(n) \right) \hat{\mathbf{R}}_{xx}(n) \boldsymbol{v}'(n).$$
(16)

Here $A_{\psi}(\sigma_e^2(n)) = E_{\{\boldsymbol{X},\eta\}}[\psi'(e)|\boldsymbol{v}], \ \sigma_e^2(n) = E[\boldsymbol{v}'^T(n)\mathbf{R}_{XX}\boldsymbol{v}'(n)] + \sigma_{\eta}^2, \ \boldsymbol{v}'(n) = \Delta \boldsymbol{W} + \boldsymbol{v}(n),$ $\Delta \boldsymbol{W} = \boldsymbol{H}_{0} - \boldsymbol{W}(\infty); \ \hat{\mathbf{R}}_{xx}(n) = \mathbf{Q}\mathbf{I}(\tilde{\Lambda})\mathbf{Q}^{T}, \ \mathbf{Q} = \mathbf{U}\Lambda^{1/2}\tilde{\mathbf{U}},$ with $\mathbf{R}_{XX} = \mathbf{U}\Lambda\mathbf{U}^{T}$ and $\mathbf{L}_{X}^{-1}\mathbf{R}_{E,X}(n)\mathbf{L}_{X}^{-T} = \tilde{\mathbf{U}}\tilde{\Lambda}\tilde{\mathbf{U}}^{T}$ being respectively the eigen-decomposition of \mathbf{R}_{XX} and $\mathbf{L}_X^{-1} \mathbf{R}_{E,X}(n) \mathbf{L}_X^{-T}$, $\mathbf{L}_X = \mathbf{U} \Lambda^{(1/2)}$. For notational convenience, we have dropped the time argument n of $\mathbf{U}(n)$, $\Lambda(n)$ and $\mathbf{Q}(n)$. $\mathbf{I}(\Lambda)$ is a diagonal matrix with its (i,i)-th entry given by the Abelian integral, $I_i(\Lambda)$ = $\int_0^\infty [\exp(-\beta\lambda)/(2\beta\widetilde{\lambda}_i^{-1} + 1)] (\prod_{j=1}^L [2\beta\widetilde{\lambda}_j^{-1} + 1]^{-1/2}) d\beta,$ where λ_i is the *i*-th eigenvalue of Λ . Similarly,

$$\boldsymbol{L}_{1}^{\prime} = \frac{\sqrt{\mu}}{L} \mathbf{D}^{\prime} \boldsymbol{\overline{\psi}}(n), \qquad (17)$$

where $\overline{\boldsymbol{\psi}}(n) = [\overline{\psi}(-\sqrt{\mu}d_{1,1}w_1(n)), \dots, \overline{\psi}(-\sqrt{\mu}d_{L,L}w_L(n))]^T$,
$$\begin{split} & \overline{\psi}(-\sqrt{\mu}d_{k,k}w_k(n)) = E_{\{v\}}[\psi(-\sqrt{\mu}d_{k,k}(w_{\infty,k}-v_k(n))] \text{ with } \\ & d_{k,k} \text{ and } w_{\infty,k} \text{ being the } k\text{-th element of } \boldsymbol{d}_k \text{ and } W(\infty), \mathbf{D}' = \\ & \text{diag}\{d_{1,1}/(\lambda + \mu d_{1,1}^2\sigma_{R_{E,X}^{-1},1}^{-1}), \dots, d_{L,L}/(\lambda + \mu d_{L,L}^2\sigma_{R_{E,X}^{-1},L}^{-1})\} \\ & \text{ is a diagonal matrix, and } \sigma_{R_{E,X}^{-1},k}^{2} \text{ is the } (k,k)\text{-th element of } \end{split}$$
 $\mathbf{R}_{E,X}^{-1}(n).$

Substituting (14)-(17) into (13), the following difference equation is obtained

$$E\left[\boldsymbol{v}(n+1)\right] \approx \left(\mathbf{I} - A_{\psi}\left(\sigma_{e}^{2}\right)\mathbf{R}_{E,X}^{-1}(n)\hat{\mathbf{R}}_{xx}(n)\right)E\left[\boldsymbol{v}(n)\right] \\ -\mathbf{R}_{E,X}^{-1}(n)\left[A_{\psi}\left(\sigma_{e}^{2}\right)\hat{\mathbf{R}}_{xx}(n)\Delta\boldsymbol{W} + \frac{\sqrt{\mu}}{L}\mathbf{D}'\overline{\boldsymbol{\psi}}(n)\right], \quad (18)$$

where for notational simplicity, $A(\sigma_e^2(n))$ is simply written as $A(\sigma_e^2)$. As mentioned, we first assume that the algorithm converges so as to determine $W(\infty)$. Then, we shall show that the algorithm is convergent for λ sufficiently close to 1 with persistent excitation. If the algorithm converges, $\lim_{n \to \infty} E[v(n)] = 0$ and the second term in (18) becomes zero. Consequently,

$$A_{\psi}\left(\sigma_{e}^{2}\right)\hat{\mathbf{R}}_{xx}\Delta\boldsymbol{W}=-\frac{\sqrt{\mu}}{L}\mathbf{D}'\boldsymbol{\overline{\psi}}(\infty),$$
(19)

where $\hat{\mathbf{R}}_{xx} = \hat{\mathbf{R}}_{xx}(\infty)$. At the steady state, $\overline{\boldsymbol{\psi}}(\infty) = -\sqrt{\mu}\mathbf{D}\boldsymbol{W}(\infty)$ since the M-estimate should clip error signal with extremely large amplitude, which is possibly corrupted by impulses. Then the optimal solution to the R-QRRLM algorithm can be derived

$$\boldsymbol{r}_{xd} = \mathbf{Q}\mathbf{I}^{-1}(\widetilde{\Lambda})\mathbf{Q}^{-1}\left(\hat{\mathbf{R}}_{xx} + \frac{\mu}{A_{\psi}\left(\sigma_e^2(\infty)\right)L}\mathbf{D}''\right)\boldsymbol{W}(\infty),$$
(20)

where $\mathbf{D}'' = \mathbf{D}'\mathbf{D}$ and $\mathbf{H}_0 = \mathbf{R}_{XX}^{-1}\boldsymbol{r}_{xd}$ has been used.

Moreover, as $\mathbf{R}_{E,X}(n) \approx (1/(1 - \lambda))(\mathbf{R}_{XX} + (\mu/L)\mathbf{D}^2) = \widetilde{\mathbf{R}}_{XX}$ for large *n*, it follows that $\widetilde{\mathbf{U}}\Lambda\widetilde{\mathbf{U}}^T = \mathbf{L}_X^{-1}\widetilde{\mathbf{R}}_{XX}\mathbf{L}_X^{-T} \approx (1/(1 - \lambda))(\mathbf{I} + (\mu/L)\mathbf{L}_X^{-1}\mathbf{D}^2\mathbf{L}_X^{-T})$. If $(\mu/L)\|\mathbf{D}^2\|_2$ is sufficiently small compared with the identity matrix, then $\widetilde{\mathbf{U}} \approx \mathbf{I}$ and $\mathbf{I}(\Lambda) \approx (1/(1 - \lambda))Erf(\lambda/1 - \lambda)\mathbf{I} \equiv E(\lambda)\mathbf{I}$, where $Erf(\lambda) = \int_0^\infty (\exp(-\beta\lambda)/(2\beta + 1)^{(L/2)+1})d\beta$ is the exponential integral function. Thus, $\widehat{\mathbf{R}}_{xx} \approx E(\lambda)\mathbf{R}_{XX}$ and (20) is reduced to

$$\left(\mathbf{R}_{XX} + \frac{\mu}{A_{\psi}\left(\sigma_{e}^{2}(\infty)\right)E(\lambda)L}\mathbf{D}^{\prime\prime}\right)\boldsymbol{W}(\infty) \approx \boldsymbol{r}_{xd}.$$
 (21)

To achieve a given regularization or diagonal loading, say $\mu\gamma_k$, at the *k*-th diagonal, we can set the regularization matrix **D** as $d_{k,k}^2 = LA_{\psi}(\sigma_e^2(\infty))E(\lambda)(\lambda + \mu d_k^2 \sigma_{R_{E,X},k}^{2-1})\gamma_k$. Note, the solution $\boldsymbol{W}(\infty)$ is biased. Let $\mathbf{G}(\mu) = \mathbf{R}_{XX} + [\mu/(A_{\psi}(\sigma_e^2(\infty))E(\lambda)L)]\mathbf{D}''$. The one gets from (21) that $\mathbf{G}(\mu)(\boldsymbol{H}_0 - \Delta \boldsymbol{W}) \approx \mathbf{R}_{XX}\boldsymbol{H}_0$ and after some manipulation $\Delta \boldsymbol{W} = [\mu/(A_{\psi}(\sigma_e^2(\infty))E(\lambda)L)]\mathbf{G}^{-1}(\mu)\mathbf{D}''\boldsymbol{H}_0$.

To study the convergence rate of the proposed algorithm, we shall focus on the term in the curved bracket in (18), which can be further simplified to

$$E\left[\mathbf{V}(n+1)\right] = \left(\mathbf{I} - A_{\psi}\left(\sigma_{e}^{2}\right)\widetilde{\Lambda}^{-1}\mathbf{I}(\widetilde{\Lambda})\right)E\left[\mathbf{V}(n)\right]$$
$$-\mathbf{Q}^{T}\mathbf{R}_{E,X}^{-1}(n)\left[A_{\psi}\left(\sigma_{e}^{2}\right)\widehat{\mathbf{R}}_{xx}(n)\Delta\mathbf{W} + \frac{\sqrt{\mu}}{L}\mathbf{D}'\overline{\boldsymbol{\psi}}(n)\right],\quad(22)$$

by expressing $\boldsymbol{v}(n)$ in the canonical coordinate $\boldsymbol{V}(n) = \mathbf{Q}^T \boldsymbol{v}(n)$. Therefore, the mean weight error vector will converge if $|1 - A_{\psi}(\sigma_e^2) \widetilde{\lambda}_i^{-1} I_i(\widetilde{\Lambda})| < 1$. It can be seen from (I-4) in Appendix A that, for small μ ,

It can be seen from (I-4) in Appendix A that, for small μ , $\lambda_i \ge 1$ and λ_i increases with μ . For $\mu = 0$ and LS objective function, we obtain the conventional RLS algorithm and the Abelian integral reduces to the exponential integral. Moreover,

$$\widetilde{\lambda}_{i}^{-1}I_{i}(\widetilde{\Lambda}) < \int_{0}^{\infty} \frac{\exp(-\beta\lambda)}{(2\beta+\widetilde{\lambda}_{i})} d\beta = \frac{1}{2} \int_{\lambda_{i}}^{\infty} \frac{\exp\left(-\frac{1}{2}(t-\widetilde{\lambda}_{i})\lambda\right)}{t} dt$$
$$\leq \frac{1}{2} \exp\left(\frac{1}{2}\widetilde{\lambda}_{i}\lambda\right) \exp\left(-\frac{1}{2}\widetilde{\lambda}_{i}\lambda\right) \ln\left(1+\frac{2}{\widetilde{\lambda}_{i}\lambda}\right), \quad (23)$$

where we have used the identity $(1/2) \exp(x) \ln(1 + (2/x)) < E_1(x) < \exp(x) \ln(1 + (1/x))$. From (23), it can be seen that for λ sufficiently close to 1, one gets

$$\widetilde{\lambda}_i^{-1} I_i(\widetilde{\Lambda}) = \frac{1}{2} \ln \left(1 + \frac{2}{\widetilde{\lambda}_i \lambda} \right) < \frac{1}{2} \ln(1+2) = 0.549.$$
 (24)

Furthermore, since $|A_{\psi}(\sigma_e^2)| < 1$ for most M-nonlinearity [21], $|A_{\psi}(\sigma_e^2)\tilde{\lambda}_i^{-1}I_i(\tilde{\Lambda})| < 1$ and hence the mean error weight vector of the algorithm is convergent if $\mathbf{R}_{E,X}(n)$ is nonsingular or the system is persistently exciting. If $\mathbf{R}_{E,X}(n)$ is singular due to the lacking of input excitation, the regularization term will help to reduce the variance of the estimator as revealed by the mean square convergence analysis below. However, it can be seen from (21) that a bias which is proportional to μ will be introduced. Hence, regularization should be applied only when $\mathbf{R}_{E,X}(n)$ is ill-conditioned. The selection of the regularization parameter will be investigated in Section IV.

B. Mean Square Convergence Analysis

Assuming $\overline{\boldsymbol{\psi}}(n)$ depends weakly on $E[\boldsymbol{v}^T(n)]$, then post-multiplying $\boldsymbol{v}(n)$ by its transpose and taking expectation gives

$$\begin{aligned} \boldsymbol{\Xi}_{vv}(n+1) &= \boldsymbol{\Xi}_{vv}(n) \\ &- \mathbf{R}_{E,X}^{-1}(n) \left\{ \mathbf{S}_1 + \frac{\sqrt{\mu}}{L} \mathbf{D}' \overline{\boldsymbol{\psi}}(n) E\left[\boldsymbol{v}^T(n)\right] \right\} \\ &- \left\{ \mathbf{S}_2 + \frac{\sqrt{\mu}}{L} E\left[v(n)\right] \overline{\boldsymbol{\psi}}^T(n) \mathbf{D}' \right\} \mathbf{R}_{E,X}^{-1}(n) \\ &+ \mathbf{R}_{E,\mathbf{X}}^{-1}(n) \mathbf{S}_3 \mathbf{R}_{E,\mathbf{X}}^{-1}(n), \end{aligned}$$
(25)

where

$$\begin{split} \mathbf{S}_{1} &= E \Biggl[\frac{\psi(e) \boldsymbol{X}(n) \boldsymbol{v}^{T}}{\left(\boldsymbol{\lambda} + \boldsymbol{X}^{T}(n) \mathbf{R}_{E,\mathbf{X}}^{-1}(n) \boldsymbol{X}(n) \right)} \Biggr] = \mathbf{S}_{2}^{T} \\ &\approx A_{\psi} \left(\sigma_{e}^{2} \right) \hat{\mathbf{R}}_{xx}(n) \mathbf{\Xi}_{v'v}(n), \\ \mathbf{S}_{3} &= \mathbf{S}_{3}' + \mathbf{S}_{4} + \mathbf{S}_{5} + \mathbf{S}_{6}, \\ \mathbf{S}_{3}' &= E \Biggl[\frac{\psi^{2}(e) \boldsymbol{X}(n) \boldsymbol{X}^{T}(n)}{\left(\boldsymbol{\lambda} + \boldsymbol{X}^{T}(n) \mathbf{R}_{E,X}^{-1}(n) \boldsymbol{X}(n) \right)^{2}} \Biggr] \\ &= E_{\{\mathbf{v}\}}[\mathbf{s}_{3}], \\ \mathbf{S}_{4} &= E \Biggl[\frac{\psi^{2}(e_{r}(n)) \boldsymbol{r}(n) \boldsymbol{r}^{T}(n)}{\left(\boldsymbol{\lambda} + \boldsymbol{r}^{T}(n) \mathbf{R}_{E,X}^{-1}(n) \boldsymbol{r}(n) \right)^{2}} \Biggr], \\ \mathbf{S}_{5} &= E \Biggl[\frac{\psi(e(n)) \psi(e_{r}(n)) \boldsymbol{X}(n) \boldsymbol{r}^{T}(n)}{\left(\boldsymbol{\lambda} + \boldsymbol{X}^{T}(n) \mathbf{R}_{E,X}^{-1}(n) \boldsymbol{X}(n) \right) \left(\boldsymbol{\lambda} + \boldsymbol{r}^{T}(n) \mathbf{R}_{E,X}^{-1}(n) \boldsymbol{r}(n) \right)} \Biggr] \\ &= \mathbf{S}_{6}^{T}, \end{split}$$

 $\boldsymbol{\Xi}_{vv'}(n) = E[\boldsymbol{v}(n)(\boldsymbol{v}^T(n) + \Delta \boldsymbol{W}^T)] = \boldsymbol{\Xi}_{vv}(n) + E[\boldsymbol{v}(n)\Delta \boldsymbol{W}^T]$ and $\boldsymbol{\Xi}_{vv}(n) = E[\boldsymbol{v}(n)\boldsymbol{v}^T(n)]$. s₃ is evaluated in Appendix B to be

$$\mathbf{s}_{3} \approx C_{\psi} \left(\sigma_{e}^{2} \right) \mathbf{I}_{1} + B_{\psi} \left(\sigma_{e}^{2} \right) \mathbf{I}_{2}, \tag{26}$$

where $\mathbf{I}_1 = 2\mathbf{Q}(\mathbf{R}_{v'v'} \circ \mathbf{I}''(\Lambda))\mathbf{Q}^T$, the symbol "o" denotes element-wise matrix product and $\mathbf{R}_{v'v'} = \mathbf{Q}^T v' v'^T \mathbf{Q}$, $\mathbf{I}_2 = \mathbf{Q}\mathbf{I}'(\tilde{\Lambda})\mathbf{Q}^T$, $C_{\psi}(\sigma_e^2) = (d/d\sigma_e^2)E[\psi^2(e)]$, $B_{\psi}(\sigma_e^2) = E_{\{\mathbf{X},\eta\}}[\psi^2(e)|v]$, $\mathbf{I}''(\tilde{\Lambda})$ is a matrix with its (i, j)-th entry given by the integral and $\mathbf{I}'_{i,j}(\tilde{\Lambda}) = \int_0^\infty [\beta \exp(-\beta \lambda)/((2\beta \tilde{\lambda}_i^{-1} + 1))]\prod_{k=1}^L (2\beta \tilde{\lambda}_k^{-1} + 1)^{-1/2} d\beta$ and $\mathbf{I}'(\tilde{\Lambda})$ is a diagonal matrix with the *i*-th element given by $I'_i(\tilde{\Lambda}) = \int_0^\infty (\beta \exp(-\beta \lambda)/(2\beta \tilde{\lambda}_i^{-1} + 1)) \prod_{k=1}^L (2\beta \tilde{\lambda}_k^{-1} + 1) d\beta$. Similarly, it was found after some manipulation that

$$\mathbf{S}_4 = \frac{\mu}{L} \mathbf{D}' \mathbf{D}_{\psi}(n) \mathbf{D}', \qquad (27)$$

where $\mathbf{D}_{\psi}(n) = (\mu/L) \sum_{k=1}^{L} E[w_k^2(n)] \mathbf{d}_k^T \mathbf{d}_k$ is a diagonal matrix with its *i*-th element $(\mu/L) d_{k,k}^2 (E[v_k^2(n)] + w_k^2(\infty) - 2w_k(\infty) E[v_k(n)])$.

$$\mathbf{S}_{5} = \frac{\sqrt{\mu}}{L} \sum_{k=1}^{L} \frac{\psi\left(e_{d_{k}}(n)\right) \boldsymbol{d}_{k}^{T}}{\left(\lambda + \mu \boldsymbol{d}_{k}(n) \mathbf{R}_{E,X}^{-1}(n) \boldsymbol{d}_{k}^{T}(n)\right)} \\ \times E\left[\frac{\psi\left(e(n)\right) \boldsymbol{X}(n)}{\left(\lambda + \boldsymbol{X}^{T}(n) \mathbf{R}_{E,X}^{-1}(n) \boldsymbol{X}(n)\right)}\right] \\ = \frac{\sqrt{\mu}}{L} \left[A_{\psi}\left(\sigma_{e}^{2}\right) \hat{\mathbf{R}}_{xx}(n) (E\left[\boldsymbol{v}(n)\right] + \Delta \boldsymbol{W}) \overline{\boldsymbol{\psi}}^{T}(n) \mathbf{D}'\right]. \quad (28)$$

Substituting (26)–(28) into (25) yields

$$\begin{aligned} \mathbf{\Xi}_{vv}(n+1) \\ \approx \mathbf{\Xi}_{vv}(n) - \mathbf{R}_{E,X}^{-1}(n) \\ \times \left\{ A_{\psi} \left(\sigma_{e}^{2} \right) \hat{\mathbf{R}}_{xx}(n) \mathbf{\Xi}_{v'v}(n) + \frac{\sqrt{\mu}}{L} \mathbf{D}' \overline{\boldsymbol{\psi}}(n) E \left[\boldsymbol{v}^{T}(n) \right] \right\} \\ - \left\{ A_{\psi} \left(\sigma_{e}^{2} \right) \mathbf{\Xi}_{vv'}(n) \hat{\mathbf{R}}_{xx}(n) + \frac{\sqrt{\mu}}{L} E \left[\boldsymbol{v}(n) \right] \overline{\boldsymbol{\psi}}^{T}(n) \mathbf{D}' \right\} \\ \times \mathbf{R}_{E,X}^{-1}(n) + \frac{\mu}{L} \mathbf{R}_{E,X}^{-1}(n) \mathbf{D}' \mathbf{D}_{\psi}(n) \mathbf{D}' \mathbf{R}_{E,X}^{-1}(n) \\ + 2C_{\psi} \left(\sigma_{e}^{2} \right) \mathbf{Q}^{-T} \widetilde{\Lambda}^{-1} \left(\mathbf{Q}^{T} \mathbf{\Xi}_{v'v'}(n) \mathbf{Q} \circ \mathbf{I}''(\widetilde{\Lambda}) \right) \widetilde{\Lambda}^{-1} \mathbf{Q}^{-1} \\ + B_{\psi} \left(\sigma_{e}^{2} \right) \mathbf{Q}^{-T} \widetilde{\Lambda}^{-1} \mathbf{I}'(\widetilde{\Lambda}) \widetilde{\Lambda}^{-1} \mathbf{Q}^{-1} + \frac{\sqrt{\mu}}{L} \mathbf{R}_{E,X}^{-1}(n) \\ \times \left[A_{\psi} \left(\sigma_{e}^{2} \right) \hat{\mathbf{R}}_{xx}(n) \left(E \left[\boldsymbol{v}(n) \right] + \Delta \boldsymbol{W} \right) \overline{\boldsymbol{\psi}}^{T}(n) \mathbf{D}' \right] \\ \times \mathbf{R}_{E,X}^{-1}(n) + \frac{\sqrt{\mu}}{L} \mathbf{R}_{E,X}^{-1}(n) \\ \times \left[A_{\psi} \left(\sigma_{e}^{2} \right) \mathbf{D}' \overline{\boldsymbol{\psi}}(n) \left(E \left[\boldsymbol{v}^{T}(n) \right] + \Delta \boldsymbol{W}^{T} \right) \hat{\mathbf{R}}_{xx}(n) \right] \\ \times \mathbf{R}_{E,X}^{-1}(n) \end{aligned}$$

$$(29)$$

We now analyze the steady-state EMSE of the algorithm. It is shown in Appendix C that the algorithm is also convergent in the mean square sense. Hence, at the steady state, we have

$$\mathbf{\Xi}_{VV}(\infty) = \mathbf{\Xi}_{VV}(\infty) - A_{\psi} \left(\sigma_e^2(\infty) \right) \widetilde{\Lambda}^{-1} \mathbf{I}(\widetilde{\Lambda}) \mathbf{\Xi}_{VV}(\infty)$$

$$-A_{\psi}\left(\sigma_{e}^{2}(\infty)\right) \mathbf{\Xi}_{VV}(\infty) \mathbf{I}(\widetilde{\Lambda}) \widetilde{\Lambda}^{-1} + 2C_{\psi}\left(\sigma_{e}^{2}(\infty)\right) \widetilde{\Lambda}^{-1}\left(\mathbf{\Xi}_{VV}(\infty) \circ \mathbf{I}''(\widetilde{\Lambda})\right) \widetilde{\Lambda}^{-1} + B_{\psi}\left(\sigma_{e}^{2}(\infty)\right) \widetilde{\Lambda}^{-1} \mathbf{\Gamma}_{1} \widetilde{\Lambda}^{-1} + \mathbf{\Gamma}_{2}.$$
(30)

Here $\Gamma_1 = \mathbf{I}'(\widetilde{\Lambda}), \ \Delta \mathbf{W}_L = \mathbf{Q}^T \Delta \mathbf{W} \text{ and } \Gamma_2 = \mathbf{Q}^T \Gamma(\infty) \mathbf{Q} + 2C_{\psi}(\sigma_e^2(\infty)) \widetilde{\Lambda}^{-1} (\Delta \mathbf{W}_L \Delta \mathbf{W}_L^T \circ \mathbf{I}''(\widetilde{\Lambda})) \widetilde{\Lambda}^{-1}, \ \Gamma(\infty) = (\mu/L) \mathbf{R}_{E,X}^{-1} \mathbf{D}' \cdot \mathbf{D}_{\psi}(\infty) \mathbf{D}' \mathbf{R}_{E,X}^{-1} + (\sqrt{\mu}/L) \mathbf{R}_{E,X}^{-1} [A_{\psi}(\sigma_e^2(\infty)) \widehat{\mathbf{R}}_{xx} \Delta \mathbf{W} \overline{\psi}^T(\infty) \mathbf{D}'] \mathbf{R}_{E,X}^{-1} + (\sqrt{\mu}/L) \mathbf{R}_{E,X}^{-1} [A_{\psi}(\sigma_e^2(\infty)) \mathbf{D}' \overline{\psi}(n) \Delta \mathbf{W}^T \widehat{\mathbf{R}}_{xx}] \mathbf{R}_{E,X}^{-1}.$ Therefore, the *i*-th diagonal element of $\Xi_{VV}(\infty), \ [\Xi_{VV}(\infty)]_{i,i},$ satisfies:

$$\begin{aligned} \left[\mathbf{\Xi}_{VV}(\infty) \right]_{i,i} &\approx \left(1 - 2A_{\psi} \left(\sigma_e^2(\infty) \right) \widetilde{\lambda}_i^{-1} I_i(\widetilde{\Lambda}) \right. \\ &\left. + 2C_{\psi} \left(\sigma_e^2(\infty) \right) \widetilde{\lambda}_i^{-2} I_{i,i}^{\prime\prime}(\widetilde{\Lambda}) \right) \left[\mathbf{\Xi}_{VV}(\infty) \right]_{i,i} \\ &\left. + B_{\psi} \left(\sigma_e^2(\infty) \right) \widetilde{\lambda}_i^{-2} \left[\mathbf{\Gamma}_1 \right]_{i,i} + \left[\mathbf{\Gamma}_2 \right]_{i,i}, \end{aligned}$$
(31)

where $[\Gamma_k]_{i,i}$, (k = 1, 2) denotes the (i, i)-th element of Γ_k . Solving (31), one gets

$$[\mathbf{\Xi}_{VV}(\infty)]_{i,i} \approx \frac{B_{\psi}\left(\sigma_{e}^{2}(\infty)\right)\widetilde{\lambda}_{i}^{-2}[\mathbf{\Gamma}_{1}]_{i,i} + [\mathbf{\Gamma}_{2}]_{i,i}}{2A_{\psi}\left(\sigma_{e}^{2}(\infty)\right)\widetilde{\lambda}_{i}^{-1}I_{i}(\widetilde{\Lambda}) - 2C_{\psi}\left(\sigma_{e}^{2}(\infty)\right)\widetilde{\lambda}_{i}^{-2}I_{i,i}^{\prime\prime}(\widetilde{\Lambda})}, \quad (32)$$

from which the steady-state EMSE can be determined as

$$J_{\psi_G} = tr\left(\mathbf{\Xi}_{vv}(\infty)\mathbf{R}_{XX}\right) = tr\left(\mathbf{\Xi}_{VV}(\infty)\mathbf{Q}^{-1}\mathbf{R}_{XX}\mathbf{Q}^{-T}\right).$$

Since $\mathbf{Q}^{-1}\mathbf{R}_{XX}\mathbf{Q}^{-T} = \widetilde{\mathbf{U}}\widetilde{\mathbf{U}}^T = \mathbf{I}$, we have

$$J_{\psi_G} \approx \sum_{i=1}^{L} \frac{B_{\psi}\left(\sigma_{e}^{2}(\infty)\right)\widetilde{\lambda}_{i}^{-2}[\Gamma_{1}]_{i,i} + [\Gamma_{2}]_{i,i}}{2A_{\psi}\left(\sigma_{e}^{2}(\infty)\right)\widetilde{\lambda}_{i}^{-1}I_{i}(\widetilde{\Lambda}) - 2C_{\psi}\left(\sigma_{e}^{2}(\infty)\right)\widetilde{\lambda}_{i}^{-2}I_{i,i}^{''}(\widetilde{\Lambda})}.$$
 (33)

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 [21]. Hence

$$J_{\psi_G} \approx \sum_{i=1}^{L} \frac{S_{\psi} \sigma_e^2(\infty) \lambda_i^{-2} [\mathbf{\Gamma}_1]_{i,i} + [\mathbf{\Gamma}_2]_{i,i}}{2A_{\psi} \tilde{\lambda}_i^{-1} I_i(\tilde{\Lambda}) - 2C_{\psi} \tilde{\lambda}_i^{-2} I_{i,i}''(\tilde{\Lambda})}.$$
 (34)

Since $\sigma_e^2(\infty) = \sigma_\eta^2 + E[(\boldsymbol{v}(\infty) + \Delta \boldsymbol{W})^T \mathbf{R}_{XX}(\boldsymbol{v}(\infty) + \Delta \boldsymbol{W})] = \sigma_\eta^2 + \Delta \boldsymbol{W}^T \mathbf{R}_{XX} \Delta \boldsymbol{W} + tr[\boldsymbol{\Xi}_{vv}(\infty) \mathbf{R}_{XX}] = \sigma_\eta^2 + \Delta \boldsymbol{W}^T \mathbf{R}_{XX} \Delta \boldsymbol{W} + J_{\psi_{-}G}$, one gets from (34) after some manipulation that:

$$J_{\psi_G} \approx \frac{\frac{1}{2} \left(\sigma_{\min}^2 \phi_{\psi} + \phi_{RW_\psi} \right)}{1 - \frac{1}{2} \phi_{\psi}}, \qquad (35)$$

where $\phi_{\psi} = S_{\psi} \sum_{i=1}^{L} \widetilde{\lambda}_{i}^{-2} [\Gamma_{1}]_{i,i} / [A_{\psi} \widetilde{\lambda}_{i}^{-1} I_{i}(\widetilde{\Lambda}) - C_{\psi} \widetilde{\lambda}_{i}^{-2} I_{i,i}''(\widetilde{\Lambda})],$ $\phi_{\mathrm{RW},\psi} \approx \sum_{i=1}^{L} [\Gamma_{2}]_{i,i} / [A_{\psi} \widetilde{\lambda}_{i}^{-1} I_{i}(\widetilde{\Lambda}) - C_{\psi} \widetilde{\lambda}_{i}^{-2} I_{i,i}''(\widetilde{\Lambda})] \text{ and }$ $\sigma_{\min}^{2} = \sigma_{\eta}^{2} + \Delta \boldsymbol{W}^{T} \mathbf{R}_{XX} \Delta \boldsymbol{W}.$

For the RLS algorithm, $A_{\psi} = S_{\psi} = C_{\psi} = 1$. For R-QRRLM with MH-nonlinearity and ATS, $\widetilde{A}_{MH} \approx A_c = (\text{erf}(k_{\xi}/\sqrt{2}) - (2k_{\xi}/\sqrt{2\pi}) \cdot \exp(-(k_{\xi}^2/2)))$, which is a constant close to one, $\widetilde{S}_{MH} \approx A_c$ and $\widetilde{C}_{MH} \approx \widetilde{A}_{MH} = A_c$. Hence, $\phi_{MH} = \phi_{RLS}$ and $\phi_{\rm RW_MH}$ is slightly larger than $\phi_{\rm RW_RLS}$. Thus, their performances in Gaussian noise are very similar. In the presence of outliers, however, the performance of the regularized RLS algorithm, like the conventional RLS algorithm, will be significantly degraded compared to the regularized RLM algorithm. This is also observed from simulation results to be presented in Section V.

C. Mean and Mean Square Behaviors in CG Noise

We now analyze the mean and mean square behaviors of the R-QRRLM algorithm in CG noise environment. It will shed light on how the LS-based algorithm is affected by the impulsive noise and why the M-estimate algorithms are more robust theoretically. From the previous analysis, we note that the assumption of Gaussian input and additive noise allows us to use the Price's theorem to decouple with the complicated effect of the nonlinearity. For most M-estimate functions which suppress outliers with large amplitude, the convergence rate will only be slightly impaired after employing ATS. We shall show in the following that the impulsive noise can be effectively suppressed and the EMSE is similar to the case where only Gaussian noise is present. On the other hand, the EMSE of the LS-based algorithm will be substantially affected by the impulsive CG noise. Although the Price's theorem is originally proposed for Gaussian variants, it was shown in [23] to be applicable to independent mixtures, and hence Gaussian mixtures. This extension of the Price's theorem [20] was employed in the analysis of the LMS and NLMS algorithms with MH nonlinearity and CG noise in [21] and a related algorithm in [4]. Similar techniques were also employed in analyzing the RLM [7] and the N-RLS algorithms [24].

1) Mean Behavior: For the analysis in this section, η is a CG noise defined as

$$\eta(n) = \eta_g(n) + \eta_{im}(n) = \eta_g(n) + b(n)\eta_w(n),$$
(36)

where $\eta_a(n)$ and $\eta_w(n)$ are both independent and identically distributed (i.i.d.) and zero mean Gaussian sequences with variance σ_a^2 and σ_w^2 , respectively. b(n) is an i.i.d. Bernoulli random sequence whose value at any time instant is either 0 or 1 with occurrence probability $P(b(n) = 1) = p_r$. Therefore, $\eta(n)$ is a Gaussian mixture consisting of two components $\eta_g(n)(b(n) =$ 0) and $\eta_{\Sigma}(n) = \eta_g(n) + \eta_w(n)(b(n) = 1)$, each with zero mean and variance σ_q^2 and σ_{Σ}^2 , respectively. Accordingly,

$$E_{\{\boldsymbol{v},\boldsymbol{X},\eta\}} [f(\boldsymbol{X}(n), e(n))] = (1 - p_r) E_{\{\boldsymbol{v},\boldsymbol{X},\eta_g\}} [f(\boldsymbol{X}(n), e(n))] + p_r E_{\{\boldsymbol{v},\boldsymbol{X},\eta_\Sigma\}} [f(\boldsymbol{X}(n), e(n))], \quad (37)$$

where $f(\boldsymbol{X}(n), e(n))$ is an arbitrary quantity whose statistical average is to be evaluated. Since X(n), $\eta_g(n)$ and $\eta_{\Sigma}(n)$ are Gaussian distributed, each of the expectation on the right hand side can be evaluated using the Price's theorem. Consequently, the results in Section III-A can be carried forward to the CG noise case by firstly changing the noise power respectively to σ_g^2 and σ_{Σ}^2 , and then combining the two results using (37). Recall the relation of the mean weight-error vector in (13):

$$E\left[\boldsymbol{v}(n+1)\right] = E\left[\boldsymbol{v}(n)\right] - \mathbf{R}_{E,X}^{-1}(n)\left(\widetilde{\boldsymbol{L}}_{1} + \widetilde{\boldsymbol{L}}_{1}'\right), \quad (38)$$

where $\widetilde{\boldsymbol{L}}_1 = E[\boldsymbol{X}(n)\psi(e(n))/(\lambda + \boldsymbol{X}^T(n)\mathbf{R}_{E,X}^{-1}(n)\boldsymbol{X}(n))] = p_r \boldsymbol{L}_{\Sigma} + (1 - p_r)\boldsymbol{L}_g,$

 $\widetilde{\boldsymbol{L}}_1' = E[\boldsymbol{r}(n)\psi(e_r(n))/(\lambda + \boldsymbol{r}^T(n)\mathbf{R}_{E,X}^{-1}(n)\boldsymbol{r}(n))] = p_r\boldsymbol{L}_{\Sigma}' +$ $(1 - p_r) \boldsymbol{L}_g'$. \boldsymbol{L}_{Σ} and \boldsymbol{L}_g are, respectively, the expectation for L_1 w.r.t. $\{v, X, \eta_{\Sigma}\}$ and $\{v, X, \eta_g\}$. Similarly L'_{Σ} and L'_q are, respectively, the expectation for L'_1 w.r.t. $\{\boldsymbol{v}, \boldsymbol{X}, \eta_{\Sigma}\}$ and $\{\boldsymbol{v}, \boldsymbol{X}, \eta_g\}$. From (14), $\tilde{\boldsymbol{L}}_i = A_{\psi}(\sigma_{e_i}^2(n))\hat{\mathbf{R}}_{xx}(n)\boldsymbol{v}'(n)$, for $i = g, \Sigma$, where $\sigma_{e_g}^2(n) = E[\boldsymbol{v}'^T(n)\mathbf{R}_{XX}\boldsymbol{v}'(n)] + \sigma_g^2$ and $\sigma_{e_{\Sigma}}^2(n) = E[\boldsymbol{v}'^T(n)\mathbf{R}_{XX}\boldsymbol{v}'(n)] + \sigma_{\Sigma}^2$. Hence, $\tilde{\boldsymbol{L}}_1 \approx \tilde{A}_{\psi}(n)\hat{\mathbf{R}}_{xx}(n)\boldsymbol{v}'(n)$, where $\tilde{A}_{\psi}(n) = (1 - p_r)A_{\psi}(\sigma_{e_g}^2(n)) + p_rA_{\psi}(\sigma_{e_{\Sigma}}^2(n))$. On the other hand, it was found that $\mathbf{L}'_i = (\sqrt{\mu}/L) \mathbf{D}' \overline{\boldsymbol{\psi}}(n)$. Thus $\widetilde{L}'_1 = (\sqrt{\mu}/L) \mathbf{D}' \overline{\psi}(n)$. Consequently, the difference (38) becomes

$$E\left[\boldsymbol{v}(n+1)\right] \approx \left(\mathbf{I} - \widetilde{A}_{\psi}(n)\mathbf{R}_{E,X}^{-1}(n)\hat{\mathbf{R}}_{xx}(n)\right) E\left[\boldsymbol{v}(n)\right] \\ -\mathbf{R}_{E,X}^{-1}(n)\left[\widetilde{A}_{\psi}(n)\hat{\mathbf{R}}_{xx}(n)\Delta\boldsymbol{W} + \frac{\sqrt{\mu}}{L}\mathbf{D}'\overline{\boldsymbol{\psi}}(n)\right].$$
(39)

At the steady state, we have from (39) the optimal solution to the R-QRRLM algorithm after some manipulations as:

$$\boldsymbol{r}_{xd} = \mathbf{Q}\mathbf{I}^{-1}(\widetilde{\Lambda})\mathbf{Q}^{-1}\left(\hat{\mathbf{R}}_{xx} + \frac{\mu}{\widetilde{A}_{\psi}(\infty)L}\mathbf{D}''\right)\boldsymbol{W}(\infty).$$
(40)

Again, if $(\mu/L) \|\mathbf{D}^2\|_2$ is sufficiently small, then $\mathbf{U} \approx \mathbf{I}$ and $\mathbf{I}(\widetilde{\Lambda}) \approx E(\lambda)\mathbf{I}$. Consequently, $\mathbf{QI}(\widetilde{\Lambda})\mathbf{Q}^T \cong E(\lambda)\mathbf{QQ}^T =$ $E(\lambda)\mathbf{R}_{XX}$ and (40) is reduced to

$$\left(\mathbf{R}_{XX} + \frac{\mu}{\widetilde{A}_{\psi}(\infty)E(\lambda)L}\mathbf{D}''\right)\boldsymbol{W}(\infty) = \boldsymbol{r}_{xd}.$$
 (41)

For notational convenience, the approximate symbol has been replaced by the equality symbol. This yields the same form as (21), except for $A_{\psi}(n)$. Similar argument regarding the mean convergence in Section III-A also applies to (41).

2) Mean Square Behavior: Using a similar approach as in Section III-B, it can be shown that

$$\begin{aligned} \boldsymbol{\Xi}_{vv}(n+1) \\ &= \boldsymbol{\Xi}_{vv}(n) - \boldsymbol{R}_{E,X}^{-1}(n) \\ &\times \left\{ \widetilde{A}_{\psi}(n) \hat{\boldsymbol{R}}_{xx}(n) \boldsymbol{\Xi}_{v'v}(n) + \frac{\sqrt{\mu}}{L} \boldsymbol{D}' \overline{\boldsymbol{\psi}}(n) E\left[\boldsymbol{v}^{T}(n)\right] \right\} \\ &- \left\{ \widetilde{A}_{\psi}(n) \boldsymbol{\Xi}_{vv'}(n) \hat{\boldsymbol{R}}_{xx}(n) + \frac{\sqrt{\mu}}{L} E\left[\boldsymbol{v}(n)\right] \overline{\boldsymbol{\psi}}^{T}(n) \boldsymbol{D}' \right\} \\ &\times \boldsymbol{R}_{E,X}^{-1}(n) + \frac{\mu}{L} \boldsymbol{R}_{E,X}^{-1}(n) \boldsymbol{D}' \boldsymbol{D}_{\psi}(n) \boldsymbol{D}' \boldsymbol{R}_{E,X}^{-1}(n) \\ &+ 2 \widetilde{C}_{\psi}(n) \boldsymbol{Q}^{-T} \widetilde{\Lambda}^{-1} \left(\boldsymbol{Q}^{T} \boldsymbol{\Xi}_{v'v'}(n) \boldsymbol{Q} \circ \mathbf{I}''(\widetilde{\Lambda}) \right) \widetilde{\Lambda}^{-1} \boldsymbol{Q}^{-1} \\ &+ \widetilde{B}_{\psi}(n) \boldsymbol{Q}^{-T} \widetilde{\Lambda}^{-1} \mathbf{I}'(\widetilde{\Lambda}) \widetilde{\Lambda}^{-1} \boldsymbol{Q}^{-1} + \frac{\sqrt{\mu}}{L} \boldsymbol{R}_{E,X}^{-1}(n) \\ &\times \left[\widetilde{A}_{\psi}(n) \hat{\boldsymbol{R}}_{xx}(n) \left(E\left[\boldsymbol{v}(n)\right] + \Delta \boldsymbol{W} \right) \overline{\boldsymbol{\psi}}^{T}(n) \boldsymbol{D}' \right] \\ &\times \boldsymbol{R}_{E,X}^{-1}(n) + \frac{\sqrt{\mu}}{L} \boldsymbol{R}_{E,X}^{-1}(n) \\ &\times \left[\widetilde{A}_{\psi}(n) \boldsymbol{D}' \overline{\boldsymbol{\psi}}(n) \left(E\left[\boldsymbol{v}^{T}(n)\right] + \Delta \boldsymbol{W}^{T} \right) \hat{\boldsymbol{R}}_{xx}(n) \right] \\ &\times \boldsymbol{R}_{E,X}^{-1}(n), \end{aligned}$$

where $\widetilde{C}_{\psi}(n) = (1 - p_r)C_{\psi}(\sigma_{e_g}^2(n)) + p_rC_{\psi}(\sigma_{e_{\Sigma}}^2(n))$ and $\widetilde{B}_{\psi}(n) = (1 - p_r)B_{\psi}(\sigma_{e_g}^2(n)) + p_rB_{\psi}(\sigma_{e_{\Sigma}}^2(n))$. Hence, the steady-state EMSE is

$$J_{\psi_CG} \approx \sum_{i=1}^{L} \frac{\widetilde{B}_{\psi}(\infty)\widetilde{\lambda}_{i}^{-2}[\Gamma_{1}]_{i,i} + [\Gamma_{2}]_{i,i}}{2\widetilde{A}_{\psi}(\infty)\widetilde{\lambda}_{i}^{-1}I_{i}(\widetilde{\Lambda}) - 2\widetilde{C}_{\psi}(\infty)\widetilde{\lambda}_{i}^{-2}I_{i,i}^{''}(\widetilde{\Lambda})}.$$
(43)

For the regularized RLS algorithm, $\tilde{A}_{\psi} = \tilde{C}_{\psi} = 1$, $\tilde{B}_{\psi} = (1 - p_r)\sigma_{e_g}^2(\infty) + p_r\sigma_{e_{\Sigma}}^2(\infty)$, then

$$\approx \frac{\frac{1}{2} \left((1-p_r) \sigma_{e_g}^2(\infty) + p_r \sigma_{e_{\Sigma}}^2(\infty) \right) \phi_{\text{RLS}} + \frac{1}{2} \phi_{\text{RW}_\text{RLS}}}{1 - \frac{1}{2} \phi_{\text{RLS}}}.$$
 (44)

For the R-QRRLM algorithm with MH-nonlinearity and ATS, if $\sigma_{e_g}^2(\infty) \ll \sigma_{e_{\Sigma}}^2(\infty)$, then $\widetilde{A}_{\rm MH} \approx (1 - p_r)A_c$ is a constant close to one given p_r is not too large. Similarly, $\widetilde{B}_{\rm MH}(n) \approx (1 - p_r)\sigma_{e_g}^2(\infty)A_c$ as $n \to \infty$ and $\widetilde{C}_{\rm MH}(n) \approx \widetilde{A}_{\rm MH}(n) \approx (1 - p_r)A_c$. Consequently, the steady-state EMSE for the R-QRRLM algorithm with MH nonlinearity and ATS is

$$J_{MH_CG} \approx \frac{\frac{1}{2}\sigma_{e_g}^2(\infty)\phi_{\rm MH} + \frac{1}{2}\phi_{\rm RW_MH}}{1 - \frac{1}{2}\phi_{\rm MH}}.$$
 (45)

Since $\sigma_{e_g}^2(\infty) = \sigma_{\min}^2$, the EMSE is still similar to its conventional LMS-based counterparts in Gaussian noise environment. This illustrates the robustness of the regularized RLM algorithms. The increase in EMSE of the RLS algorithm over the RLM algorithm is

$$\Delta EMSE_{RLS-RLM_CG}(\infty) \approx \frac{\frac{1}{2}p_r \left(\sigma_{e_{\Sigma}}^2(\infty) - \sigma_{e_g}^2(\infty)\right) \phi_{\text{RLS}}}{1 - \frac{1}{2}\phi_{\text{RLS}}}.$$
 (46)

We note from (46) that the improvement of the M-estimation algorithms over the LS-based algorithm is proportional to the power of the impulsive components, which is reasonable and expected.

IV. VARIABLE REGULARIZATION PARAMETER SELECTION

The regularization parameter plays an important role in the performance of the R-QRRLM algorithm in aspects such as steady-state EMSE, convergence rate and tracking capability. Using the performance analysis obtained previously, we now derive the regularization parameter μ from the MSE deviation of $W(\infty)$ around the Wiener solution.

To proceed further, we assume that the input is white with variance σ_x^2 and $\mathbf{D} = \mathbf{I}$. Hence $\phi_{\text{RLS}_W} \approx [(1 - \lambda)/(I(\tilde{\Lambda})(1 + \mu/(L\sigma_x^2)))]tr(\mathbf{\Gamma}_1) = ((1 - \lambda)/(I(\tilde{\Lambda})(1 + \mu/(L\sigma_x^2)))]tr(\mathbf{I}'(\Lambda)) = (1 - \lambda)LI'(\tilde{\Lambda})/(I(\tilde{\Lambda})(1 + \mu/(L\sigma_x^2)))$. In addition, for mild regularization, $\Delta \mathbf{W}_L$ is small and from (30), we have $\mathbf{\Gamma}_2 \approx (\mu/L)\Lambda^{1/2}[\mathbf{R}_{E,X}^{-1}\mathbf{D}'\mathbf{D}_{\psi}(\infty)\mathbf{D'R}_{E,X}^{-1}]\Lambda^{1/2}$. Using a similar approach, we can obtain that
$$\begin{split} \phi_{\text{RW}_\text{RLS}} &\approx \sum_{i=1}^{L} [\mathbf{\Gamma}_2]_{i,i} / (\widetilde{\lambda}_i^{-1} I_i(\widetilde{\Lambda}) - \widetilde{\lambda}_i^{-2} I_{i,i}''(\widetilde{\Lambda})) \approx [(1 + \mu/(L\sigma_x^2))/((1 - \lambda)I(\widetilde{\Lambda}))] tr(\mathbf{\Gamma}_2) = [\mu(1 + \mu/(L\sigma_x^2))/((1 - \lambda)I(\widetilde{\Lambda})L)] tr(\Lambda^{1/2}[\mathbf{R}_{E,X}^{-1} \mathbf{D}' \mathbf{D}_{\psi}(\infty) \mathbf{D}' \mathbf{R}_{E,X}^{-1}] \Lambda^{1/2}), \text{ where we} \\ \text{have used the fact } \widetilde{\lambda}_i^{-1}(\infty) I_i(\widetilde{\Lambda}) \gg \widetilde{\lambda}_i^{-2}(\infty) I_{i,i}''(\widetilde{\Lambda}) \text{ in deriving} \\ \text{the approximated expression. Since } \mu d_{i,i}^2 \sigma_{R_{E,X}^{-1},i}^{2-1} \text{ is small compared to } \lambda, \mathbf{D}' \approx \mathbf{I}/\lambda \text{ and } \mathbf{R}_{E,X}^{-1} \approx (1 - \lambda)(\Lambda + (\mu/L)\mathbf{I})^{-1}, \\ \phi_{\text{RW_RLS}} &= [\mu^2(1 + \mu/(L\sigma_x^2))(1 - \lambda)/(\lambda^2 I(\widetilde{\Lambda})L)] tr((\Lambda + (\mu/L)\mathbf{I})^{-2}\Lambda diag(\mathbf{W}(\infty)\mathbf{W}^T(\infty))). \end{split}$$

As this expression is somewhat difficult to be simplified further, we shall employ the following upper bound of $\phi_{\text{RW}_\text{RLS}}$ by using the identity $tr(\mathbf{AB}) \leq tr(\mathbf{A})tr(\mathbf{B})$ when \mathbf{A} and \mathbf{B} are positive definite matrices:

$$\phi_{\text{RW}_\text{RLS}} \leq \frac{\mu^2 \left(1 + \mu / \left(L\sigma_x^2\right)\right) (1 - \lambda)}{\lambda^2 I(\widetilde{\Lambda}) L} tr$$
$$\times \left(\left(\Lambda + \frac{\mu}{L} \mathbf{I} \right)^{-2} \Lambda \right) \| \boldsymbol{W}(\infty) \|_2^2$$

Combining these results, we obtain an approximation of (35):

$$J_{*} \approx \frac{1}{2} \left(\sigma_{\min}^{2} \phi_{\text{RLS}_w} + \phi_{\text{RW}_\text{RL}} \right)$$

$$\leq \frac{1}{2} \sigma_{\min}^{2} \frac{(1-\lambda)L \cdot I'(\tilde{\Lambda})}{I(\tilde{\Lambda})(1+\mu/L\sigma_{x}^{2})}$$

$$+ \frac{\mu^{2} \left(1+\mu/\left(L\sigma_{x}^{2}\right)\right)(1-\lambda)}{2\lambda^{2} I(\tilde{\Lambda})L} \phi_{\text{R}} \left\| \boldsymbol{W}(\infty) \right\|_{2}^{2}, \quad (47)$$

where $\phi_R = L\sigma_x^2(\sigma_x^2 + (\mu/L))^{-2}$. We now can determine the regularization parameter from the steady-state MSE of $\boldsymbol{W}(\infty)$ around \boldsymbol{H}_0 as follows $J_w = E[\|\boldsymbol{v}(\infty)\|_2^2] + \|\Delta \boldsymbol{W}\|_2^2$.

First of all, we note that for white input $E[\|\boldsymbol{v}\|_2^2] = (1/\sigma_x^2)J_*$. On the other hand, from Section III-A, we can obtain $\|\Delta \boldsymbol{W}\|_2^2 = (\mu^2/A_\psi^2 E^2(\lambda)L^2) \|\mathbf{G}^{-1}(\mu)\mathbf{D}''\boldsymbol{H}_0\|_2^2 \approx (\mu^2/A_\psi^2\lambda^2 E^2(\lambda)L^2\sigma_x^4) \|\boldsymbol{H}_0\|_2^2$, where we have assumed that the regularization is mild and hence $\mathbf{G}^{-1}(\mu) \cong \mathbf{R}_{XX}^{-1}$ and used the fact that $\mathbf{D}'' = \mathbf{I}$ for white input. Moreover, $\sigma_{\min}^2 = \sigma_\eta^2 + \sigma_x^2 \|\Delta \boldsymbol{W}\|_2^2$. Therefore,

$$\begin{split} J_w &= \frac{1}{2} \sigma_\eta^2 \frac{(1-\lambda)LI'(\tilde{\Lambda})}{I(\tilde{\Lambda}) \left(\sigma_x^2 + \frac{\mu}{L}\right)} + \left[\frac{\mu^2(1-\lambda)}{2\sigma_x^2 \lambda^2 I(\tilde{\Lambda}) \left(\sigma_x^2 + \frac{\mu}{L}\right)} \left\| \boldsymbol{W}(\infty) \right\|_2^2 \right. \\ &+ \frac{\mu^2}{A_\psi^2 \lambda^2 E^2(\lambda) L^2 \sigma_x^4} \left(1 + \frac{(1-\lambda)LI'(\tilde{\Lambda})\sigma_x^2}{2I(\tilde{\Lambda}) \left(\sigma_x^2 + \frac{\mu}{L}\right)} \right) \left\| \boldsymbol{H}_0 \right\|_2^2 \right]. \end{split}$$

It can be seen that the first and second (inside the square bracket) terms on the right hand side correspond, respectively, to the bias and variance of the MSE. In order to obtain a balanced performance in practical applications, we propose to choose μ so that the two terms are equal to each other. Consequently, the desired regularization parameter, μ_{opt} satisfies

$$\begin{split} & \mu_{opt}^{2} \\ = & \frac{\frac{1}{2}\sigma_{\eta}^{2}\frac{(1-\lambda)LI'(\widetilde{\Lambda})}{I(\widetilde{\Lambda})\left(\sigma_{x}^{2}+\frac{\mu}{L}\right)}}{\left[\frac{1}{A_{\psi}^{2}\lambda^{2}E^{2}(\lambda)L^{2}\sigma_{x}^{4}}\left(l\!\!+\!\frac{(1-\lambda)LI'(\widetilde{\Lambda})\sigma_{x}^{2}}{2I(\widetilde{\Lambda})\left(\sigma_{x}^{2}\!+\!\frac{\mu}{L}\right)}\right)\!\!+\!\frac{(1-\lambda)}{2\sigma_{x}^{2}\lambda^{2}I(\widetilde{\Lambda})\left(\sigma_{x}^{2}\!+\!\frac{\mu}{L}\right)}\right]\!\|\boldsymbol{H}_{0}\|_{2}^{2}, \end{split}$$



Fig. 1. Convergence curves for EMSE in CG noise with (a) white Gaussian input and (b) first-order AR input. L = 5, $\lambda = 0.98$, SNR = 0 dB.

where it is assumed that $\|\boldsymbol{W}(\infty)\|_{2}^{2} \approx \|H_{0}\|_{2}^{2}$. Since $I(\tilde{\Lambda}) \approx E(\lambda) = 1/\lambda$, $\lambda I'(\tilde{\Lambda}) \approx 1/\lambda$ and $\tilde{\lambda}^{-1}(\infty) = (1-\lambda)/(1+\mu/(L\lambda_{i}))$ are very small when λ is close to 1, we further have:

$$\begin{split} \mu_{opt} &\approx L \sigma_x^2 \lambda \sqrt{1-\lambda} \\ &\times \sqrt{\frac{\frac{\sigma_\eta^2}{\|\mathbf{H}_0\|_2^2}}{\frac{1}{A_\psi^2 \lambda L} \left(2\left(\sigma_x^2 + \frac{\mu}{L}\right) + \frac{(1-\lambda)L \sigma_x^2}{\lambda}\right) + \frac{(1-\lambda)L \sigma_x^2}{\lambda^2}}}. \end{split}$$

Since $\mu = \kappa L$, the desired κ is

$$\kappa_{opt} = \sigma_x^2 \lambda^{-1} \sqrt{1 - \lambda} \sqrt{\frac{\gamma \left(\frac{\sigma_n^2}{\sigma_x^2}\right)}{\left\|\boldsymbol{H}_0\right\|_2^2}},\tag{48}$$

where $\gamma = 1/((1/A_{\psi}^2 \lambda L)(2 + ((1-\lambda)L/\lambda)) + ((1-\lambda)L/\lambda^2)).$

V. EXPERIMENTAL RESULTS AND ACOUSTIC APPLICATIONS

We now evaluate the performance of the proposed algorithm and compare its performance with other conventional RLS algorithms. First, a system identification problem is used to examine the convergence speed, tracking capability, robustness in CG noises and steady-state EMSE of various algorithms. The mean and mean square convergence analyses developed in Section III will also be evaluated. Then the VR-QRRLM algorithm is applied to the ANC system, resulting in a new FX-VR-QRRLM algorithm. Unless specified otherwise, all the simulation results have been averaged over 300 Monte Carlo runs. The Abelian integrals as in Appendices are calculated by using the steady-state value of $\mathbf{R}_{E,X}^{-1}(n)$, i.e. $\mathbf{R}_{E,X}^{-1}$ for simplicity since it has been found that the theoretical prediction is in good agreement with simulations.

A. Performance in CG Noises

Experiment 1: This experiment is carried out in the system identification. The convergence performance and robustness to impulsive noises of the VR-QRRLM algorithm are evaluated and compared with the QRRLS and QRRLM algorithms using white and colored inputs. The system order is L = 5. The white

Gaussian input is of zero mean and unit variance whereas the colored input is simulated by a first-order auto-regressive (AR) process: x(n)x(n) = 0.9x(n-1) + g(n), where g(n) is a Gaussian process with zero mean and variance 0.1. For the CG noise, $\sigma_w^2 = 100$ is used to generate impulsive noises, and the impulse occurrence probability is p = 0.001. The signal-to-noise ratio (SNR) is defined by $10 \log(\sigma_{d_0}^2/\sigma_g^2)$, where $\sigma_{d_0}^2$ is the power of the system output $d_0(n) = \mathbf{X}^T(n)\mathbf{H}_0$. σ_g^2 is chosen to achieve an SNR = 0 dB. For illustration purpose, the locations of the impulses are fixed at time index 400, 800, and 1300, in each independent run. The FF for all the algorithms is set to be $\lambda = 0.98$. The parameters for estimating the noise variance in (4) are $\lambda_{\sigma} = 0.9$ and $N_w = 9$. The performance of the algorithms is evaluated using EMSE with respect to the optimal Wiener solution. Figs. 1(a) and (b) depict the performance of algorithms with white and colored inputs, respectively. First, it can be seen that the QRRLM and VR-QRRLM algorithms are robust to impulsive noises. In contrast, the performance of the QRRLS algorithm is deteriorated by these impulses severely. Secondly, all algorithms in test seem to have similar initial convergence performance. After the initial stage, however, the VR-QRRLM algorithm converges at a faster speed and to a lower steady-state EMSE. The improvement for the colored input is significant as shown in Fig. 1(b) because the small eigenvalues of the input covariance matrix have been greatly compensated by regularization and hence the steady-state EMSE is reduced.

Experiment 2: This experiment examines the tracking capability of the proposed algorithm for sudden system changes under different SNRs and compares it with the traditional QRRLM [8] algorithm. The impulsive noise here is the same as that in Experiment 1. The system order is L = 25 and the channel changes suddenly after the 800-th iteration. The FF λ is chosen as 0.99. The learning curves for EMSE are plotted in Figs. 2(a) and (b) for white Gaussian and first-order AR inputs, respectively. As can be seen, in the first 800 iterations, the VR-QRRLM algorithm generally converges faster and to a lower steady-state EMSE compared to the QRRLM algorithm. The advantages of the VR-QRRLM algorithm over QRRLM are more significant when the input is colored and the SNR is low. After the system changes at the 800-th sample, the



Fig. 2. Learning curves for EMSE in CG noise for sudden-change channels with (a) white Gaussian input and (b) first-order AR input. $L = 25, \lambda = 0.99$.

 TABLE II

 COMPARISON OF THE EXPERIMENTAL AND PREDICTED STEADY-STATE EMSES FOR WHITE GAUSSIAN INPUT

L					9	SNR = 0	dB			SNR = 10 dB									
	λ	0.98			0.99			0.995			0.98			0.99			0.995		
5	κ	0.35	0.07	0.007	0.30	0.06	0.006	0.25	0.05	0.005	0.12	0.024	0.002	0.1	0.02	0.002	0.075	0.015	0.002
	Simu	-4.7	-8.4	-8.2	-8.6	-11.3	-11.1	-10.0	-14.1	-14.1	-13.6	-17.8	-17.8	-15.1	-20.6	-20.6	-20.2	-23.4	-23.4
	Theo	-6.9	-8.4	-8.2	-9.8	-11.1	-11.0	-12.5	-14.2	-14.1	-15.2	-18.3	-17.6	-17.2	-21.2	-20.6	-21.0	-23.6	-23.3
	κ	0.5	0.1	0.01	0.5	0.1	0.01	0.5	0.1	0.01	0.15	0.03	0.003	0.15	0.03	0.003	0.15	0.03	0.003
25	Simu	6.0	5.1	5.8	4.3	2.2	2.6	3.3	-0.3	-0.1	-2.2	-4.1	-4	-4.2	-7.3	-7.3	-5.5	-9.9	-9.9
	Theo	8.5	5.3	5.3	6.9	2.4	2.5	4.4	0	0	-1.3	-4.2	-4.1	-2.6	-7.3	-7.3	-3.3	-9.9	-9.9



Fig. 3. Convergence curves for $||\mathbf{v}||_2$ with white Gaussian input for L = 5 and $\lambda = 0.98$ at SNR = 10 dB.

VR-QRRLM algorithm maintains a better convergence performance and lower steady-state EMSE.

B. Mean Convergence Behavior

We now evaluate the mean convergence analysis of the R-QRRLM algorithm based on system identification. The system $\boldsymbol{H}_0 = [0.3886, 0.1632, -0.8294, -0.9247, -0.6148]$ is employed. The norm of the mean weight-error vector is used as the performance measure: $\|\boldsymbol{v}(n)\|_2 = 10 \log(\sqrt{\sum_{i=1}^{L} [(1/K) \sum_{j=1}^{K} \boldsymbol{v}_i^{(j)}(n)]^2})$, where $v_i^{(j)}(n)$ is the *i*-th component of the weight-error vector at time *n* in the *j*-th independent run w.r.t. 1) the Wiener solution:

 $\mathbf{v}^*(n) = \mathbf{W}(n) - \mathbf{H}_0$, or 2) the optimal solution $\mathbf{W}(\infty)$ in (21): $\mathbf{v}_{\infty}(n) = \mathbf{W}(n) - \mathbf{W}(\infty)$. K is the total number of independent runs, which is set to be 5000 in this experiment. Fig. 3 plots the learning curves of $\|\mathbf{v}^*(n)\|_2$ and $\|\mathbf{v}_{\infty}(n)\|_2$ for the R-QRRLM algorithm. The settings are as follows: the FF is $\lambda = 0.98$, the regularization parameter is $\kappa = 0.008$ or 0.04 and SNR = 10 dB. It can be seen from the $\|\mathbf{v}^*(n)\|_2$ curves that the bias w.r.t. the Wiener solution increases significantly with κ . Since $\|\mathbf{v}_{\infty}(n)\|_2$ converges to a lower value than $\|\mathbf{v}^*(n)\|_2$, it suggests that the algorithm converges to $\mathbf{W}(\infty)$ rather than to \mathbf{H}_0 and the bias is introduced because of the regularization. Simulations with different FFs show similar results and are not presented due to page limitation.

C. Mean Square Convergence Behavior

The mean square convergence behavior is also examined by system identification. Simulation results for the steady-state EMSE of the R-QRRLM algorithm are compared with theoretical predictions as shown in Tables II and III for both white and colored inputs. In the simulations, we set L = 5, 25; $\lambda = 0.995, 0.99, 0.98$; $\kappa = 0.1\kappa_{opt}, \kappa_{opt}, 5\kappa_{opt}$ with κ_{opt} being the selected regularization parameter as suggested in (48); and SNR = 0 dB, 10 dB. It can be seen that the simulation and theoretical results agree well with each other, given the approximation used. In general, the theoretical results are closer to simulations when the regularization parameter κ is smaller. It can also be seen that the EMSE with $\kappa = \kappa_{opt}$ is the smallest compared to that with $\kappa = 0.1\kappa_{opt}$ and $5\kappa_{opt}$ in each case, which suggests the effectiveness of (48).

To further verify the performance analysis in Section III, the theoretical predictions of the convergence curves for EMSE are

TABLE III Comparison of the Experimental and Predicted Steady-State EMSEs for First Order AR Input

L						SNR = 0) dB			SNR = 10 dB									
	λ	0.98			0.99			0.995			0.98			0.99			0.995		
5	κ	0.85	0.17	0.017	0.75	0.15	0.015	0.7	0.14	0.014	0.3	0.06	0.006	0.25	0.05	0.005	0.18	0.035	0.003
	Simu	-3.2	-6.7	-4.8	-4.2	-8.8	-7.8	-5.2	-10.6	-10.6	-9.4	-14.2	-13.9	-9.8	-16.9	-16.9	-11.2	-18.9	-19.2
	Theo	1.0	-6.5	-4.9	0	-9.1	-7.6	-0.3	-11.0	-10.4	-6.7	-14.9	-14.2	-7.3	-17.4	-17.2	-9.8	-19.6	-19.4
	κ	1.5	0.3	0.03	1.5	0.3	0.03	1.5	0.3	0.03	0.5	0.1	0.01	0.5	0.1	0.01	0.5	0.1	0.01
25	Simu	9.7	9.4	13.7	7	6.5	10.6	5.8	3.7	7.7	2	1.8	4.9	-0.1	-1.1	1.9	-1.3	-4.0	-1.3
	Theo	11.7	10.0	13.8	10.5	8.7	10.7	9.3	5.8	8.1	2.9	2.0	4.3	1.3	0.1	1.6	0.8	-3.0	-1.1

Simu: simulation results; Theo: theoretical results.



Fig. 4. Comparison of simulation and theoretical learning curves for EMSE with (a) white Gaussian input and (b) first-order AR input. L = 5, $\lambda = 0.98$, and $\kappa = \kappa_{opt}$.

compared to the simulation results in Figs. 4(a) and (b). The settings are L = 5, $\lambda = 0.98$, $\kappa = \kappa_{opt}$ with SNR = 0 dB and 10 dB. It can be seen that the theoretical learning curves of EMSE agree well with simulations for both white and colored inputs. Here the integrals as in the Appendices are calculated by using $\mathbf{R}_{E,X}^{-1}(n) \approx \mathbf{R}_{E,X}^{-1}(\infty)$ for calculation simplicity.

D. Application to Acoustic System Identification

For some online acoustic system identification problems, the input signal usually has varying power. In this simulation, the performance of several algorithms is compared when the input is a segment of music. The impulse response has 100 taps. It is assumed that there is a double talk from the 3000 to 3100-th sample. The SNR is 10 dB. The FF for all the algorithms tested is 0.997. The regularization parameter for the R-QRRLM algorithm is $\kappa = 0.1$ and 0.001 while $\lambda_{\sigma} = 0.95$ and $N_w = 200$. The performance of various algorithms is shown in Fig. 5. It can be seen that although the QRRLS algorithm has a faster initial convergence speed, it is very sensitive to the input signal level and double talk which resembles a long series of impulsive noises. For the R-QRRLM algorithm, if a small regularization parameter, say $\kappa = 0.001$, is used, it is still sensitive to input power. On the other hand, if κ is increased to a larger value, i.e. $\kappa = 0.1$, the algorithm becomes much less sensitive to the input signal power variation but it converges to a higher EMSE value. The VR-QRRLM algorithm, however, adaptively selects the regularization parameter and offers both high immunity to variation in input signal power and impulsive noises.



Fig. 5. Performance comparison of various algorithms for EMSE at SNR = 10 dB. $\lambda = 0.997$.

E. Application to ANC

In this experiment, an ANC system using the FX algorithm as shown in Fig. 6 is considered [28]. { $p_k(n), k = 0, 1, \ldots, L_p - 1$ } denotes the primary path while { $s_k(n), k = 0, 1, \ldots, L_s - 1$ } is the secondary path. The impulse responses for both paths are assumed to be of finite duration for simplicity. An error microphone is used to pick up the residual signal e(n) to be minimized. Thus, the ANC controller $\boldsymbol{W}(n) = [w_0(n), \ldots, w_{Lw-1}(n)]^T$ approximates { $-p_k(n)$ } after cascading with { $s_k(n)$ } so that the undesirable contribution from the noise source {x(n)} is minimized. Since

 $\{s_k(n)\}\$ is unknown, it is replaced by its estimate $\{\hat{s}_k(n)\}\$ and the input to the adaptive filter is $\hat{x}_s(n) = x(n)^* \hat{s}_k(n)$, where "*" stands for the discrete time convolution. Let $\eta_{BG}(n)$ be the background noise and other acoustic signals at the error microphone, then the error signal picked up by the microphone is

$$e_a(n) = d_a(n) - y_a(n),$$
 (49)

where $d_a(n) = p_k(n)^* x(n) + \eta_{BG}(n)$, and $y_a(n)$ $-x(n)^*w_k(n)^*s_k(n)$. The mean square error of $e_a(n)$ can be minimized by the well-known FX-LMS algorithms. Since they are very sensitive to impulsive interferences, robust M-estimate algorithms have been proposed to enhance the performance of the FX-LMS algorithm in the presence of impulses [29]. On the other hand, since the secondary path is not known exactly, regularization techniques can improve the performance of the ANC controller. Since a fast response and reliable performance are required in ANC systems in order to avoid undesirable artifacts generated by the speaker, the proposed VR-QRRLM algorithm is a good alternative to the existing FX-LMM algorithm [29] because of its fast response and robustness to errors in the secondary path modeling. However, in ANC system, only $e_a(n)$ is observed, rather than $d_a(n)$ as required by the VR-QRRLM algorithm. In [37], the problem is solved by approximating $d_a(n)$ as $d'_a(n) = e_a(n) + y_a(n)$. Since $\{s_k(n)\}\$ is usually unknown, $d'_a(n)$ will lead to an increased EMSE. To solve this problem, we make use of (8) and compute the Kalman gain vector $\boldsymbol{K}(n)$ using the QRD in the VR-QRRLM algorithm. In fact, $\mathbf{R}^{(1)}(n)$ obtained from the QRD in Table I is the Cholesky factor of the regularized input correlation matrix $\mathbf{R}_{XX}(n) + (\mu/A_{\psi}(\sigma_e^2(\infty))E(\lambda)L)\mathbf{D}''$, i.e. $\mathbf{G}(\mu) = \mathbf{R}^T(n)\mathbf{R}(n)$, which is also equal to $\mathbf{P}^{-1}(n)$. $\mathbf{K}(n) =$ $[(q(e(n))/(\lambda_e(n)+q(e(n))\boldsymbol{X}^T(n)\mathbf{P}(n - 1)\boldsymbol{X}(n))\mathbf{P}(n - 1)\boldsymbol{X}(n)] \mathbf{P}(n - 1)\mathbf{X}(n)] \mathbf{P}(n - 1)\mathbf{X}(n) \mathbf{P}(n - 1)\mathbf{X}(n) \mathbf{P}(n - 1)\mathbf{X}(n)]$ 1) $\mathbf{X}(n) = \mathbf{P}(n)\mathbf{X}(n)$ can be readily computed if $\mathbf{u}(n) =$ $\mathbf{P}(n)\mathbf{X}(n)$ is available. Using the Cholesky factorization, $\boldsymbol{u}(n) = \mathbf{R}^{-1}(n)[\mathbf{R}^{-T}(n)\boldsymbol{X}(n)]$. The term in the square bracket, $l(n) = [\mathbf{R}^{-T}(n)\mathbf{X}(n)]$, can be computed by back-substitution, since $\mathbf{R}^{-T}(n)$ is a lower triangular matrix. The same applies to $\boldsymbol{u}(n) = \mathbf{R}^{-1}(n)\boldsymbol{l}(n)$, since $\mathbf{R}^{-1}(n)$ is an upper triangular matrix. The input to the controller is $\hat{\boldsymbol{x}}_{s}(n)$, which is obtained by filtering the input $\{x(n)\}$ through the estimate of the secondary path $\{\hat{s}_k(n)\}\$ with an impulse response vector $\hat{\boldsymbol{s}}(n) = [\hat{s}_0(n), \hat{s}_1(n), \dots, \hat{s}_{L_s} - 1(n)]^T$ of L_s taps. Hence $\boldsymbol{X}(n)$ above should be replaced by $\hat{\boldsymbol{x}}_s(n)$ in ANC. Alternatively, a simpler implementation with slightly lower arithmetic complexity is to compute the change in W(n), $\Delta \boldsymbol{W}(n) = e_a(n)\boldsymbol{K}(n)$, directly from the QRD. More details can be found in [38].

The convergence performance of ANC systems using the conventional FX-LMS algorithm has been analyzed in [30]–[34]. The effect of online secondary-path modeling was recently addressed in [34]. Though LMM-based algorithms for robust ANC in impulsive noises have been proposed in [29], [35], the performance analysis of ANC under impulsive noise and regularization remains a challenging problem [36]. Here, we shall make use of the general analytical approach presented in Section III to study the performance of the proposed FX-VR-QRRLM algorithm for ANC with regularization and impulsive noise.



Fig. 6. Block diagram of an ANC system.

First we note that the input to the FX-VR-QRRLM is $\hat{x}_{s}(n), d_{a}(n) = p_{k}(n)^{*}x(n) + \eta_{BG}(n), \text{ and } y_{a}(n) =$ $-x(n)^*w_k(n)^*s_k(n)$. In our model in Section III, the desired signal is assumed to be $d_a(n) = x_s(n)^* w_{OPT_k}(n) + \eta(n)$, where $\{w_{OPT_k}, k = 0..., L_w - 1\}$ is the length- L_w optimal Wiener solution, which is given by $\boldsymbol{W}_{OPT} = \mathbf{R}_{\hat{x}_s \hat{x}_s}^{-1} \boldsymbol{r}_{\hat{x}_s d}$ $-\mathbf{R}_{\hat{x}_s\hat{x}_s}^{-1} \boldsymbol{r}_{\hat{x}_s x} \boldsymbol{p}, \text{ where } \boldsymbol{W}_{OPT} = [w_{OPT_0}, \dots, w_{OPT_L_w-1}]^T,$ $\mathbf{R}_{\hat{x}_{s}\hat{x}_{s}}(n) = E[\hat{\boldsymbol{x}}_{s}(n)\hat{\boldsymbol{x}}_{s}^{T}(n)], \quad \boldsymbol{r}_{\hat{x}_{s}d}(n) = E[\hat{\boldsymbol{x}}_{s}(n)d(n)], \\ \boldsymbol{r}_{\hat{x}_{s}x}(n) = E[\hat{\boldsymbol{x}}_{s}(n)\boldsymbol{x}_{L_{p}}^{T}(n)], \quad \boldsymbol{x}_{L_{p}}(n) = [x(n), \dots, x(n-1)],$ $L_p + 1)]^T$, $\hat{\boldsymbol{x}}_s(n) = [\hat{x}_s(n), \dots, \hat{x}_s(n - L_w + 1)]$ and $\boldsymbol{p} = [p_0, \dots, p_{L_p-1}]^T$ is the impulse response vector of the primary path and it is assumed to be stationary. The additive noise is now given by $\eta(n) = \eta_{BG}(n) + \eta_M(n)$, where $\eta_{BG}(n) = \eta_q(n) + \eta_{im}(n)$ is the disturbance recorded which is assumed to be CG distributed and $\eta_M(n)$ is the additional modeling error which is equal to $\eta_M(n) = p_k(n)^* x(n) + \boldsymbol{x}_s^T(n) \boldsymbol{W}_{OPT}$. Since x(n) is assumed to be zero mean, $\eta_M(n)$ is zero mean. For simplicity, it is approximated to be white Gaussian distributed with variance $\sigma_{\eta_M}^2 = E[\eta_M^2(n)]$. This will allow our performance analysis in Section III to be applied to the FX-VR-QRRLM algorithm. Though it is possible to extend the approach in Section III and evaluate all the expectations involving $\eta_M(n)$ in a more accurate analysis, the approach taken here simplifies considerably the derivation and gives us reasonably accurate information about the EMSE and other quantities. For instance, from (41), we immediately have the steady state solution of the ANC controller $W(\infty)$.

$$\left(\mathbf{R}_{\hat{x}_s\hat{x}_s} + \frac{\mu}{\widetilde{A}_{\psi}(\infty)E(\lambda)L}\mathbf{D}''\right)\boldsymbol{W}(\infty) = -\boldsymbol{r}_{\hat{x}_sd}.$$
 (50)

It can be seen that a bias over that Wiener solution is again introduced. The corresponding EMSE is given by (45). To verify the performance analysis above, an experiment is carried out. The primary path has 100 taps and the secondary path has 50 taps, which is estimated offline with a modeling error of -14 dB after being normalized by the true secondary path. The noise $\{\eta_g\}$ is a white Gaussian sequence and the SNR is 10 dB. A series of impulsive interference occurring from the 1500 to 1600-th sample is used to simulate the situation of double talks or other interferences. The FF of VR-QRRLM is chosen as 0.998. Parameters for estimating the noise variance are $\lambda_{\sigma} = 0.95$ and $N_w = 100$. The simulated and estimated MSEs of the residue error $\{e_a(n)\}$ are plotted in Fig. 7. It can be seen that the theoretical results agree well with simulation



Fig. 7. Comparison of simulation and theoretical learning curves for MSE with SNR = 10 dB. Modeling error: -14 dB. $\lambda = 0.998$.



Fig. 8. Performance comparison of various algorithms for the ANC system: the learning curves of MSE with SNR = 10 dB. Modeling error: -14 dB. Step-size for LMS and LMM:0.001. FF for QRRLM and VR-QRRLM: 0.998.

results. The deviation between the two curves in transient is mainly attributed to the assumptions used such as $\eta_M(n)$ is white.

In the second simulation, the FX-LMS, FX-LMM [29], FX-QRRLM and the proposed FX-VR-QRRLM algorithms are evaluated. As a comparison, the FX-VR-QRRLM using the estimated desired signal $d'_a(n)$ as proposed in [37] is also shown. The settings are identical to those in the previous example except that a series of impulsive interference occurs from 4000 to 4100-th samples. For the FX-LMS and FX-LMM algorithms, the step-size is 0.001; for the FX-QRRLM and FX-VR-QRRLM algorithms, the FF is chosen as 0.998 so that all algorithms achieve comparable MSE. Note, in calculating the regularization parameter, the modeling errors in both primary and secondary path have been absorbed to the background noise variance. The MSE of the residue error $\{e_a(n)\}\$ is plotted in Fig. 8. It can be seen that 1) all the robust algorithms are insensitive to impulsive interferences, 2) the QRRLM based algorithms converge faster than the LMS-based algorithms, 3) the FX-VR-QRRLM using the approximated desired signal $d'_a(n)$ has higher variance (and occasionally stability problems), and 4) the FX-VR-QRRLM is more stable compared to FX-QRRLM when the secondary path modeling error exists and the regularization technique further reduces the steady-state MSE over the FX-QRRLM algorithm.

VI. CONCLUSION

A new variable regularized QRD-based RLM adaptive filtering algorithm and its mean and mean square convergence performance have been presented. It extends the conventional RLM algorithm by imposing a variable L_2 regularization term on the coefficients to reduce the variance of the estimator. An efficient recursive QRD-based implementation is developed to improve its numerical stability. Difference equations describing the mean and mean square convergence behaviors of the algorithm in Gaussian inputs and additive CG noises are derived. The bias over the classical Wiener solution introduced by the regularization is quantified. New expressions for the steadystate EMSE are derived and it suggests that the variance will decrease while the bias will increase with the regularization parameter. Therefore, the algorithm is especially useful when the input covariance matrix is ill-conditioned or singular due to lacking of excitation. The advantage of the M-estimation algorithm over its least square counterparts is also analytically quantified. For white Gaussian inputs, the regularization parameter can be determined based on the analysis, leading to the proposed VR-QRRLM algorithm. The theoretical results are in good agreement with simulations. Moreover, simulation results of several acoustic applications, namely system identification and ANC, show that the VR-QRRLM based algorithms outperform the traditional ones considerably when the input signal level is low and in the presence of considerable modeling error or impulsive noises. Using the proposed theoretical analysis, the challenging problem of performance analysis of the proposed FX-VR-ORRLM ANC algorithm under impulsive noise and regularization is also characterized.

APPENDIX A EVALUATION OF A

The expectation $\boldsymbol{A} = E_{\{X,\eta\}}[\boldsymbol{X}\psi(e)/(\lambda + \boldsymbol{X}^T\mathbf{R}_{E,X}^{-1}(n)\boldsymbol{X})|\boldsymbol{v}]$ is evaluated in this Appendix. For simplicity, all time indices have been omitted here. As $\eta(n)$ and x(n) are assumed to be statistically independent, and \boldsymbol{X} are zero-mean jointly Gaussian with covariance matrix \mathbf{R}_{XX} , one gets

$$\boldsymbol{A} = C_R \iint_{L+1fold} \frac{\boldsymbol{X}\psi(e) \exp\left(-\frac{1}{2}\boldsymbol{X}^T \mathbf{R}_{XX}^{-1} \boldsymbol{X}\right)}{\lambda + \boldsymbol{X}^T \mathbf{R}_{E,X}^{-1}(n) \boldsymbol{X}} f_{\eta}(\eta) d\eta d\boldsymbol{X},$$
(I-1)

where $C_R = (2\pi)^{-L/2} |\mathbf{R}_{XX}|^{-1/2}$ and $f_{\eta}(\eta)$ is the probability density function (PDF) of the Gaussian noise. Similar to [21], we consider the integral

$$\boldsymbol{F}(\beta) = C_R \iint_{L+1fold} \frac{\boldsymbol{X}\psi(e) \exp\left[-\beta \left(\lambda + \boldsymbol{X}^T \mathbf{R}_{E,X}^{-1}(n) \boldsymbol{X}\right)\right]}{\lambda + \boldsymbol{X}^T \mathbf{R}_{E,X}^{-1}(n) \boldsymbol{X}} \times \exp\left(-\frac{1}{2} \boldsymbol{X}^T \mathbf{R}_{XX}^{-1}(n) \boldsymbol{X}\right) f_{\eta}(\eta) d\eta d\boldsymbol{X}.$$
 (I-2)

Then $\mathbf{A} = \mathbf{F}(0)$. Differentiating (I-2) w.r.t. β , one gets

$$\frac{d\boldsymbol{F}(\beta)}{d\beta} = -\exp(-\beta\lambda)C_R$$
$$\iint_{L+1fold} \boldsymbol{X}\psi(e)\exp\left(-\frac{1}{2}\boldsymbol{X}^T\mathbf{B}^{-1}\boldsymbol{X}\right)f_{\eta}(\eta)d\eta d\boldsymbol{X}, \quad (I-3)$$

where $\mathbf{B} = (2\beta \mathbf{R}_{E,X}^{-1}(n) + \mathbf{R}_{XX}^{-1})^{-1}$. For notational convenience, we drop the time argument n of the quantities derived from $\mathbf{R}_{E,X}(n)$. Assume that The quantities derived from $\mathbf{L}_{E,X}(n)$. Assume that $\mathbf{R}_{E,X}(n) \approx (1 - \lambda^n/1 - \lambda)(\mathbf{R}_{XX} + (\mu/L)\mathbf{D}^2) = \mathbf{\widetilde{R}}_{XX}$, then $\mathbf{B} \approx (2\beta \mathbf{\widetilde{R}}_{XX}^{-1} + \mathbf{R}_{XX}^{-1})^{-1}$. Let the eigende-composition of \mathbf{R}_{XX} be $\mathbf{R}_{XX} = \mathbf{U}\Lambda\mathbf{U}^T$ and $\mathbf{L}_X = \mathbf{U}\Lambda^{1/2}$. Then $\mathbf{R}_{XX} = \mathbf{L}_X\mathbf{L}_X^T$, and $\mathbf{B} \approx (2\beta \mathbf{\widetilde{R}}_{XX}^{-1} + \mathbf{L}_X^{-T}\mathbf{L}_X^{-1})^{-1} = \mathbf{L}_X(2\beta \mathbf{L}_X^T \mathbf{\widetilde{R}}_{XX}^{-1}\mathbf{L}_X + \mathbf{I})^{-1}\mathbf{L}_X^T$. Since $\mathbf{L}_X^T \mathbf{\widetilde{R}}_{XX}^{-1}\mathbf{L}_X$ is symmetric, it admits the eigen-decom-position $\mathbf{I}^T \mathbf{\widetilde{R}}_{XX}^{-1} \mathbf{L}_X = \mathbf{\widetilde{U}}\lambda^{-1}\mathbf{\widetilde{U}}^T$. Thus,

position $\mathbf{L}_{X}^{T} \widetilde{\mathbf{R}}_{XX}^{-1} \mathbf{L}_{X} = \widetilde{\mathbf{U}} \widetilde{\Lambda}^{-1} \widetilde{\mathbf{U}}^{T}$. Thus,

$$\mathbf{L}_{X}^{-1}\widetilde{\mathbf{R}}_{XX}\mathbf{L}_{X}^{-T} = \frac{1-\lambda^{n}}{1-\lambda} \left(\boldsymbol{I} + \frac{\mu}{L} \Lambda^{-1/2} \mathbf{U}^{T} \mathbf{D}^{2} \mathbf{U} \Lambda^{-1/2} \right)$$
$$= \widetilde{\mathbf{U}} \widetilde{\Lambda} \widetilde{\mathbf{U}}^{T}. \quad (I-4)$$

Hence $\mathbf{B}^{-1} \approx \mathbf{L}_X^{-T}(2\beta \mathbf{L}_X^T \widetilde{\mathbf{R}}_{XX}^{-1} \mathbf{L}_X + \mathbf{I})\mathbf{L}_X^{-1} = \mathbf{L}_X^{-T}(2\beta \widetilde{\mathbf{U}}\widetilde{\Lambda}^{-1}\widetilde{\mathbf{U}}^T + \mathbf{I})\mathbf{L}_X^{-1} = \mathbf{L}_X^{-T}\widetilde{\mathbf{U}}(2\beta \widetilde{\Lambda}^{-1} + \mathbf{I})\widetilde{\mathbf{U}}^T \mathbf{L}_X^{-1} = \mathbf{L}_X^{-T}\widetilde{\mathbf{U}}\mathbf{D}_B^{-1}\widetilde{\mathbf{U}}^T \mathbf{L}_X^{-1}$, where \mathbf{D}_B is a diagonal matrix with the *i*-th diagonal entry $\tilde{d}_{ii} = (2\beta \tilde{\lambda}_i^{-1} + 1)^{-1}$. Noting that the determinants of U and \mathbf{D}_B are respectively 1 and $|\mathbf{D}_B|^{-1/2} = \prod_{i=1}^{L} (2\beta \tilde{\lambda}_i^{-1} + 1)^{-1/2}$, one can rewrite (I-3) as

$$\frac{d\boldsymbol{F}(\beta)}{d\beta} = -\gamma(\beta)C_B \iint_{L+1fold} \boldsymbol{X}\psi(e) \exp\left(-\frac{1}{2}\boldsymbol{X}^T \mathbf{B}^{-1}\boldsymbol{X}\right) \\ \times f_{\eta}(\eta)d\eta d\boldsymbol{X} \\ = -\gamma(\beta)E_{\{\boldsymbol{X},\eta\}} \left[\boldsymbol{X}\psi(e)|\boldsymbol{v}\right]\Big|_{E[\boldsymbol{X}\boldsymbol{X}^T]=\boldsymbol{B}} \\ = -\gamma(\beta)\boldsymbol{l}_1, \tag{I-5}$$

where $C_B = (2\pi)^{-L/2} |\mathbf{B}|^{-1/2}$, $\gamma(\beta) = \exp(-\beta\lambda) |\mathbf{B}|^{1/2} / |\mathbf{R}|^{1/2} = \exp(-\beta\lambda) |\mathbf{D}_B|^{-1/2} = \exp(-\beta\lambda) \prod_{i=1}^{L} (2\beta\tilde{\lambda}_i^{-1} + 1)^{-1/2}$, and l_1 is the expectation of $X_q(e)e$ conditioned on v when $x_i, x_j \in X$ are jointly Gaussian with covariance matrix **B**. Since X and e are assumed as jointly Gaussian in Assumption (A1) and (A2), the Price theorem [20] can be invoked to obtain the *i*-th element of $\boldsymbol{l}_1, \boldsymbol{l}_{1,i}$, as follows:

$$\boldsymbol{l}_{1,i} = A_{\psi} \left(\sigma_e^2 \right) \mathbf{B}_i \boldsymbol{v}', \tag{I-6}$$

where $A_{\psi}(\sigma_e^2) = E_{\{\mathbf{X},\eta\}}[\psi'(e)|\mathbf{v}]|_{\mathbf{B}}, \mathbf{B}_i$ is the *i*-th row of **B**, $v' = \Delta W + v$ and $\Delta W = H_0 - W(\infty)$.

Substituting (I-6) into (I-5) and integrating w.r.t. β yields

$$\boldsymbol{F}(\beta) = \left(-\int^{\beta} A_{\psi}\left(\sigma_{e}^{2}\right)\gamma(\beta)\mathbf{B}d\beta\right)\boldsymbol{v}'(n)$$
$$\approx A_{\psi}\left(\sigma_{e}^{2}\right)\left(-\int^{\beta}\gamma(\beta)\mathbf{B}d\beta\right)\boldsymbol{v}'(n),\qquad(\text{I-7})$$

where the constant of integration is equal to zero because of the boundary condition $F(\infty) = 0$. Here, we assumed that the term $A_{\psi}(\sigma_e^2)$ inside the bracket depends weakly on β so that we can move it outside the integral of β . This is a good approximation for M-nonlinearity [21].

Next, we evaluate the term $I(\beta) = -\int^{\beta} \gamma(\beta) B d\beta$. Noting that $\mathbf{B} = \mathbf{Q}\mathbf{D}_B\mathbf{Q}^T$ with $\mathbf{Q} = \mathbf{U}\Lambda^{1/2}\widetilde{\mathbf{U}}$, one gets

$$\mathbf{I}(\beta) = \mathbf{L}_X \left(-\int_{-}^{\beta} \gamma(\beta) \mathbf{D}_B d\beta \right) \mathbf{L}_X^T = \mathbf{L}_X \widetilde{\mathbf{U}} I(\widetilde{\Lambda}) \widetilde{\mathbf{U}}^T \mathbf{L}_X^T,$$
(I-8)

where the integral of the *i*-th diagonal entry of $I(\Lambda)$ given $\beta = 0$ is $I_i(\tilde{\Lambda}) \approx \int_0^\infty [(\exp(-\beta\lambda)/(2\beta\tilde{\lambda}_i^{-1} +$ 1)] $(\prod_{j=1}^{L} [2\beta \tilde{\lambda}_{j}^{-1} + 1]^{-1/2})d\beta.$

Finally, from (I-6)–(I-8), we have

$$\mathbf{A} = \boldsymbol{F}(0) \approx A_{\psi} \left(\sigma_e^2 \right) \mathbf{Q} \mathbf{I}(\widetilde{\Lambda}) \mathbf{Q}^T \boldsymbol{v}'(n).$$
(I-9)

APPENDIX B EVALUATION OF s_3

We evaluate $\mathbf{s}_3 = E_{\{\mathbf{X},\eta\}}[\psi^2(e)\mathbf{X}\mathbf{X}^T/(\lambda +$ $\boldsymbol{X}^T \mathbf{R}_{E,X}^{-1} \boldsymbol{X})^2 | \boldsymbol{v}]$. Similar to the derivation of \boldsymbol{A} , \mathbf{s}_3 is given by

$$\mathbf{s}_{3} = C_{R} \iint_{L+1fold} \frac{\psi^{2}(e) \boldsymbol{X} \boldsymbol{X}^{T} \exp\left(-\frac{1}{2} \boldsymbol{X}^{T} \mathbf{R}_{XX}^{-1} \boldsymbol{X}\right)}{\left(\lambda + \boldsymbol{X}^{T} \mathbf{R}_{E,X}^{-1}(n) \boldsymbol{X}\right)^{2}} \times f_{\eta}(\eta) d\eta d\boldsymbol{X}. \quad (\text{II-1})$$

Similar to [21], we define

$$\overline{\mathbf{F}}(\beta) = C_R \iint_{L+1fold} \frac{\boldsymbol{X} \boldsymbol{X}^T \psi^2(e) \exp\left[-\beta \left(\lambda + \boldsymbol{X}^T \mathbf{R}_{E,X}^{-1}(n) \boldsymbol{X}\right)\right]}{\left(\lambda + \boldsymbol{X}^T \mathbf{R}_{E,X}^{-1}(n) \boldsymbol{X}\right)^2} \times \exp\left(-\frac{1}{2} \boldsymbol{X}^T \mathbf{R}_{XX}^{-1} \boldsymbol{X}\right) f_{\eta}(\eta) d\eta d\boldsymbol{X}. \quad \text{(II-2)}$$

Comparing (II-2) with (II-1), it can be seen that $\mathbf{s}_3 = \overline{\mathbf{F}}(0)$. To evaluate $\overline{\mathbf{F}}(\beta)$, differentiating (II-2) twice w.r.t. β , one gets $d^2 \overline{\mathbf{F}}(\beta)/d\beta^2 = \gamma(\beta) \mathbf{L}_3$, where the $\gamma(\beta)$ and **B** have been defined in Appendix A. $\mathbf{L}_3 = E_{\{\mathbf{X},\eta\}}[\psi^2(e)\mathbf{X}\mathbf{X}^T|\mathbf{v}]|_{\mathbf{B}}$ is the expectation of $\psi^2(e) \mathbf{X} \mathbf{X}^T$ taken over $\{\mathbf{X}, \eta\}$ and conditioned on \boldsymbol{v} . First, consider the (i, j)-th element of \boldsymbol{l}_3 . We get from the Price theorem $\partial l_{3,i,j}/\partial r_{x_ix_j} = E_{\{\mathbf{X},\eta\}}[\psi^2(e)|\mathbf{v}]|_{\mathbf{B}} = \overline{\psi^2}(\sigma_e^2) = (1/\sqrt{2\pi}\sigma_e)\int_{-\infty}^{\infty}\psi^2(e)\exp(-(e^2/2\sigma_e^2))de$, where $x_i, x_j \in \mathbf{X}$ are jointly Gaussian variables with covariance matrix B. For notational simplicity, we write $\overline{\psi^2}(\sigma_e^2)$ as $B_{\psi}(\sigma_e^2)$. Integrating $\partial l_{3,i,j}/\partial r_{x_ix_j}$ w.r.t. $r_{x_ix_j}$ gives: $l_{3,i,j} = B_{\psi}(\sigma_e^2)r_{x_ix_j} + c_{i,j}$, with $c_{i,j} = E[\psi^2(e)x_ix_j]|_{r_{x_ix_j=0}}$ being the integration constant. Using Price's theorem, we have $\partial c_{i,j}/\partial r_{x_ie}$ $E[(d\psi^{2}(e)/de)x_{j}]|_{r_{x_{i}x_{j}=0}} = E[(d^{2}\psi^{2}(e)/de^{2})]r_{x_{j}e}$ = $2(dE[\psi^{2}(e)]/d\sigma_{e}^{2})r_{x_{j}e} = 2C_{\psi}(\sigma_{e}^{2})r_{x_{j}e}.$ Integrating again, one gets $c_{i,j} = 2C_{\psi}(\sigma_{e}^{2})r_{x_{j}e}r_{x_{i}e}.$ Combining above, we have $\boldsymbol{l}_{3,i,j} = B_{\psi}(\sigma_{e}^{2})r_{x_{i}x_{j}} + 2C_{\psi}(\sigma_{e}^{2})r_{x_{j}e}r_{x_{i}e}$ and hence $\mathbf{L}_{3} = B_{\psi}(\sigma_{e}^{2})\mathbf{B} + 2C_{\psi}(\sigma_{e}^{2})\mathbf{B}\boldsymbol{v}'\boldsymbol{v}'^{T}\mathbf{B}.$ Substituting \mathbf{L}_{3} into $d^2 \overline{\mathbf{F}}(\beta) / d\beta^2$ gives

$$\frac{d^2 \overline{\mathbf{F}}(\beta)}{d\beta^2} = 2\gamma(\beta) C_{\psi} \left(\sigma_e^2\right) \mathbf{B} \boldsymbol{v}' \boldsymbol{v}'^T \mathbf{B} + \gamma(\beta) B_{\psi} \left(\sigma_e^2\right) \mathbf{B}.$$
(II-3)

Integrating (II-3) w.r.t. β and letting $\beta = 0$ yields

$$\mathbf{s}_{3} = \int_{0}^{\infty} \int_{\beta_{1}}^{\infty} 2C_{\psi} \left(\sigma_{e}^{2}\right) \gamma(\beta_{2}) \mathbf{B} \boldsymbol{v}' \boldsymbol{v}'^{T} \mathbf{B} d\beta_{2} d\beta_{1} + \int_{0}^{\infty} \int_{\beta_{1}}^{\infty} B_{\psi} \left(\sigma_{e}^{2}\right) \cdot \gamma(\beta_{2}) \mathbf{B} d\beta_{2} d\beta_{1} \approx C_{\psi} \left(\sigma_{e}^{2}\right) \boldsymbol{I}_{1} + B_{\psi} \left(\sigma_{e}^{2}\right) \mathbf{I}_{2}.$$
(II-4)

The last approximation is obtained by observing that $B_{\psi}(\sigma_e^2)$ and $C_{\psi}(\sigma_e^2)$ depend weakly on β for M-nonlinearity [21]. I₁ can be evaluated as in [21], I₁ = $\int_0^{\infty} \int_{\beta_1}^{\infty} 2\gamma(\beta_2) \mathbf{B} \mathbf{v}' \mathbf{v}'^T \mathbf{B} d\beta_2 d\beta_1 = 2\mathbf{Q}(\int_0^{\infty} \int_{\beta_1}^{\infty} \gamma(\beta_2) \mathbf{D}_B \mathbf{R}_{\mathbf{v}'\mathbf{v}'} \mathbf{D}_B d\beta_2 d\beta_1) \mathbf{Q}^T = 2\mathbf{Q}(\mathbf{R}_{\mathbf{v}'\mathbf{v}'} \circ \mathbf{I}''(\tilde{\Lambda})) \mathbf{Q}^T$, where the (i, j)-th element of the integral $\mathbf{I}''(\tilde{\Lambda})$ is $I''_{i,j}(\tilde{\Lambda}) = \int_0^{\infty} \int_{\beta_1}^{\infty} \gamma(\beta_2) d\tilde{i}_{ij} d\beta_2 d\beta_1 = \int_0^{\infty} [\beta \exp(-\beta \lambda)/((2\beta \tilde{\lambda}_i^{-1} + 1))] \prod_{k=1}^{L} (2\beta \tilde{\lambda}_k^{-1} + 1)^{-1/2} d\beta$, and $\mathbf{R}_{\mathbf{v}'\mathbf{v}'} = \mathbf{Q}^T \mathbf{v}' \mathbf{v}'^T \mathbf{Q}$. Similarly, $\mathbf{I}_2 = \int_0^{\infty} \int_{\beta_1}^{\infty} \gamma(\beta_2) \mathbf{B} d\beta_2 d\beta_1 = \int_0^{\infty} \int_{\beta_1}^{\infty} \gamma(\beta_2) \mathbf{Q} \mathbf{D}_B \mathbf{Q}^T d\beta_2 d\beta_1 = \mathbf{Q} \int_0^{\infty} \int_{\beta_1}^{\infty} \gamma(\beta_2) \mathbf{D}_B d\beta_2 d\beta_1 \mathbf{Q}^T = \mathbf{Q} \cdot \mathbf{I}'(\tilde{\Lambda}) \mathbf{Q}^T$, where $I'_i(\tilde{\Lambda}) = \int_0^{\infty} (\beta \exp(-\beta \lambda)/(2\beta \tilde{\lambda}_i^{-1} + 1)) \prod_{k=1}^{L} (2\beta \tilde{\lambda}_k^{-1} + 1)^{-(1/2)} d\beta$ is the *i*-th diagonal entry of $\mathbf{I}'(\tilde{\Lambda})$.

APPENDIX C MEAN SQUARE CONVERGENCE OF R-RLM

From (35), we can see that the EMSE is bounded as long as

$$\phi_{\text{RLS}} < 2 \text{ and } A_{\psi} \widetilde{\lambda}_i^{-1} I_i(\widetilde{\Lambda}) > C_{\psi} \widetilde{\lambda}_i^{-2} I_i''(\widetilde{\Lambda})$$

where $\phi_{\text{RLS}} = \sum_{i=1}^{L} \phi_i$ and $\phi_i = S_{\psi} \tilde{\lambda}_i^{-2} [\Gamma_1]_{i,i} / (A_{\psi} \tilde{\lambda}_i^{-1} I_i(\tilde{\Lambda})) - C_{\psi} \tilde{\lambda}_i^{-2} I_{i,i}''(\tilde{\Lambda}))$. Since $A_{\psi} \approx C_{\psi}$ for MH nonlinearity and $\tilde{\lambda}_i^{-1} I_i(\tilde{\Lambda}) > \tilde{\lambda}_i^{-2} I_i''(\tilde{\Lambda})$, the first inequality is more important than the other. From the definition of $\tilde{\Lambda}$, it follows that $\tilde{\lambda}_i = (1/(1-\lambda))(1+\mu^2\lambda_i') \ge 1$, where λ_i' is the eigenvalue of $\Lambda^{-1/2} \mathbf{U}^T \mathbf{D}^2 \mathbf{U} \Lambda^{-1/2}$. Hence $\tilde{\lambda}_i$ increases with μ^2 since $\lambda_i' > 0$. For $\mu = 0$, we have the RLS algorithm and the Abelian integrals will reduces to the exponential integral. If \mathbf{R}_{XX} is positive definite, it can be shown that ϕ_{RLS} is less than 2 and the algorithm is convergent in the mean square sense. For the regularized algorithm, $\tilde{\lambda}_i$ is greater than 1 and we need to study the upper bound of $\phi_{\text{RLS}} \approx \Sigma_{i=1}^{L} \phi_i$, where

$$\phi_i \approx \frac{S_{\psi} \widetilde{\lambda}_i^{-2} I_{i,i}''(\widetilde{\Lambda})}{A_{\psi} \widetilde{\lambda}_i^{-1} I_i(\widetilde{\Lambda}) - C_{\psi} \widetilde{\lambda}_i^{-2} I_{i,i}''(\widetilde{\Lambda})} \\ = \frac{\int_0^\infty \frac{S_{\psi} \beta}{(2\beta + \widetilde{\lambda}_i)^2} k(\widetilde{\Lambda}) d\beta}{\int_0^\infty \frac{(2\beta + \widetilde{\lambda}_i) A_{\psi} - C_{\psi} \beta}{(2\beta + \widetilde{\lambda}_i)^2} k(\widetilde{\Lambda}) d\beta},$$

with
$$k(\widetilde{\Lambda}) = exp(-\beta\lambda)\Pi_{j=1}^{L}(2\beta\widetilde{\lambda}_{j}^{-1}+1)^{-1/2}$$
.
Suppose that $\widetilde{\lambda}_{k}$ is increased to $\widetilde{\lambda}_{k} = (\widetilde{\lambda}_{k}^{-1}-\Delta\lambda_{k})^{-1}$. One get

$$\begin{split} \phi_i' &\approx \frac{\int_0^\infty \frac{S_{\psi}\beta}{(2\beta+\widetilde{\lambda_i})^2} \frac{k(\widetilde{\Lambda})}{\left(1-\frac{2\beta\Delta\lambda_k}{2\beta\widetilde{\lambda_k}^{-1}+1}\right)^{\frac{1}{2}} d\beta}}{\int_0^\infty \frac{(2A_{\psi}-C_{\psi})\beta+A_{\psi}\widetilde{\lambda_i}}{(2\beta+\widetilde{\lambda_i})^2} \frac{k(\widetilde{\Lambda})}{\left(1-\frac{2\beta\Delta\lambda_k}{2\beta\widetilde{\lambda_k}^{-1}+1}\right)^{\frac{1}{2}} d\beta} \\ &= \frac{\int_0^\infty \frac{S_{\psi}\beta}{(2\beta+\widetilde{\lambda_i})^2} k(\widetilde{\Lambda}) d\beta + \int_0^\infty \frac{S_{\psi}\beta^2\Delta\lambda_k\widetilde{\lambda_k}}{(2\beta+\widetilde{\lambda_k})(2\beta+\widetilde{\lambda_i})^2} k(\widetilde{\Lambda}) d\beta - \dots}{\int_0^\infty \frac{(2A_{\psi}-C_{\psi})\beta+A_{\psi}\widetilde{\lambda_i}}{(2\beta+\widetilde{\lambda_i})^2} k(\widetilde{\Lambda}) d\beta + \int_0^\infty \frac{\beta\left((2A_{\psi}-C_{\psi})\beta+A_{\psi}\widetilde{\lambda_i}}{(2\beta+\widetilde{\lambda_i})(2\beta+\widetilde{\lambda_i})^2} k(\widetilde{\Lambda}) d\beta - \dots}, \end{split}$$

where only two terms of the expansion have been written.

If
$$(2A_{\psi} - C_{\psi}) \ge S_{\psi}$$
, then

$$\phi' < \frac{\int_{0}^{\infty} \frac{S_{\psi}\beta}{(2\beta + \widetilde{\lambda}_{i})^{2}} k(\widetilde{\Lambda}) d\beta}{\int_{0}^{\infty} \frac{(2A_{\psi} - C_{\psi})\beta + A_{\psi}\widetilde{\lambda}_{i}}{(2\beta + \widetilde{\lambda}_{i})^{2}} k(\widetilde{\Lambda}) d\beta} = \phi_{\phi}$$

For most M-nonlinearity, the condition $(2A_{\psi} - C_{\psi}) \ge S_{\psi}$ is satisfied. Therefore, when μ increases, λ'_i increases and both ϕ_i and ϕ_{RLS} decrease. Since the RLS algorithm with $\mu = 0$ is convergent when \mathbf{R}_{XX} is positive definite, the regularized algorithm is also convergent.

Moreover, as μ increases, the variance part of the EMSE, i.e. $(1/2)\sigma_{\min}^2\phi_{\rm RLS}/(1 - (1/2)\phi_{\rm RLS})$, will decrease whereas the bias term $\phi_{\rm RW}$ will increase. This is in accordance with the behavior of the ridge regression and related regularization.

References

- [1] S. Haykin, *Adaptive Filter Theory*, 4th ed. Upper Saddle River, NJ: Prentice-Hall, 2001.
- [2] A. Rosado-Munoz, M. Bataller-Mompean, E. Soria-Olivas, C. Scarante, and J. Guerrero-Martinez, "FPGA implementation of an adaptive filter robust to impulsive noise: Two approaches," *IEEE Trans. Ind. Electron.*, vol. 58, no. 3, pp. 860–870, Mar. 2011.
 [3] H. Dai and N. K. Sinha, "A robust off-line output error method for
- [3] H. Dai and N. K. Sinha, "A robust off-line output error method for system identification," *IEEE Trans. Ind. Electron.*, vol. 39, no. 4, pp. 285–292, Aug. 1992.
 [4] Y. Zhou, S. C. Chan, and K. L. Ho, "New sequential partial update
- [4] Y. Zhou, S. C. Chan, and K. L. Ho, "New sequential partial update least mean M-estimate algorithms for robust adaptive system identification in impulsive noise," *IEEE Trans. Ind. Electron.*, vol. 58, no. 9, pp. 4455–4470, Sep. 2011.
- [5] 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, pp. 1564–1569, Dec. 2000.
- [6] Y. Zou, S. C. Chan, and T. S. Ng, "A recursive least M-estimate (RLM) adaptive filter for robust filtering in impulse noise," *IEEE Signal Process. Lett.*, vol. 7, no. 11, pp. 324–326, Nov. 2000.
 [7] S. C. Chan and Y. Zou, "A recursive least M-estimate algorithm for
- [7] 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, Apr. 2004.
- [8] 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 *Proc. EUSIPCO '06*, Florence, Italy, Sep. 4–8, 2006.
 [9] Y. Chen, Y. Gu, and A. O. Hero, "Sparse LMS for system identifi-
- [9] Y. Chen, Y. Gu, and A. O. Hero, "Sparse LMS for system identification," in *Proc. ICASSP '09*, Taipei, Taiwan, Apr. 19–24, 2009, pp. 3125–3128.
- [10] J. Fan and R. Li, "Variable selection via nonconcave penalized likelihood and its oracle properties," *J. Amer. Stat. Assoc.*, vol. 96, no. 456, pp. 1348–1360, Dec. 2001.
- [11] T. V. Waterschoot, G. Rombouts, and M. Moonen, "Optimally regularized adaptive filtering algorithms for room acoustic signal enhancement," *Signal Process.*, vol. 88, pp. 594–611, 2008.

- [12] J. W. Stokes and J. C. Platt, "Robust RLS with round robin regularization including application to stereo acoustic echo cancellation," in *Proc. ICASSP '06*, Toulouse, France, May 15–19, 2006, pp. III-376–III-379.
- [13] S. C. Chan, Y. J. Chu, and Z. G. Zhang, "A new regularized QRD recursive least M-estimate algorithm: Performance analysis and applications," in *Proc. ICGCS '10*, Shanghai, China, Jun. 21–23, 2010.
- [14] A. Houacine, "Regularized fast recursive least squares algorithms for adaptive filtering," *IEEE Trans. Signal Process.*, vol. 39, no. 4, pp. 860–871, Apr. 1991.
- [15] 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.
- [16] J. Ni and F. Li, "A variable regularization matrix normalized subband adaptive filter," *IEEE Signal Process. Lett.*, vol. 16, no. 2, pp. 105–108, Feb. 2009.
- [17] H. Rey, L. R. Vega, S. Tressens, and J. Benesty, "Variable explicit regularization in affine projection algorithm: Robustness issues and optimal choice," *IEEE Trans. Signal Process.*, vol. 55, no. 5, pp. 2096–2109, May 2007.
- [18] Y. Ś. Choi, H. C. Shin, and W. J. Song, "Adaptive regularization matrix for affine projection algorithm," *IEEE Trans. Circuits Syst. II*, vol. 54, no. 12, pp. 1087–1091, Dec. 2000.
- [19] 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. ISCAS '09*, Taipei, Taiwan, May 24–27, 2009, vol. 5, pp. 1193–1196.
- [20] R. Price, "A useful theorem for nonlinear devices having Gaussian inputs," *IEEE Trans. Inf. Theory*, vol. IT-4, no. 2, pp. 69–72, Jun. 1958.
- [21] S. C. Chan and Y. Zhou, "On the performance analysis of the least mean M-estimate and normalized least mean M-estimate algorithms with Gaussian inputs and additive Gaussian and contaminated Gaussian noises," J. Signal Process. Syst., vol. 60, no. 1, pp. 81–103, Oct. 2009.
- [22] S. H. Leung and C. F. So, "Gradient-based variable forgetting factor RLS algorithm in time-varying environments," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 3141–3150, Aug. 2005.
- Process., vol. 53, no. 8, pp. 3141–3150, Aug. 2005.
 [23] R. Price, "Comment on: "A useful theorem for nonlinear devices having Gaussian inputs"," *IEEE Trans. Inf. Theory*, vol. IT-10, no. 2, p. 171, Apr. 1964.
- [24] J. F. Weng and S. H. Leung, "Adaptive nonlinear RLS algorithm for robust filtering in impulse noise," in *Proc. ISCAS* '97, Jun. 9–12, 1997, vol. 4, pp. 2337–2340.
- [25] S. C. Chan, Z. G. Zhang, and Y. J. Chu, "A new transform-domain regularized recursive least M-estimate algorithm for robust linear estimation," *IEEE Trans. Circuits Syst. II*, vol. 58, no. 2, pp. 120–124, Feb. 2011.
- [26] R. E. Frank and M. Hampel, Robust Statistics: The Approach Based on Influence Functions. New York: Wiley, 1986.
- [27] J. W. Tukey, A Survey of Sampling From Contaminated Distribution. Stanford, CA: Stanford Univ. Press, 1960.
- [28] S. M. Kuo and D. R. Morgan, Active Noise Control Systems—Algorithms and DSP Implementations. New York: Wiley, 1996.
- [29] P. Thanigainayagam, S. M. Kuo, and R. Yenduri, "Nonlinear active noise control for infant incubators in Neo-Natal intensive care units," in *Proc. ICASSP* '07, Apr. 2007, vol. I, pp. 109–112.
- [30] Y. Xiao, L. Ma, and K. Hasegawa, "Properties of FXLMS-based narrowband active noise control with online secondary-path modeling," *IEEE Trans. Signal Process.*, vol. 57, no. 8, pp. 2931–2949, Aug 2009.
- IEEE Trans. Signal Process., vol. 57, no. 8, pp. 2931–2949, Aug. 2009.
 J. Liu, Y. Xiao, J. Sun, and L. Xu, "Analysis of online secondary-path modeling with auxiliary noise scaled by residual noise signal," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 18, no. 8, pp. 1978–1993, Nov. 2010.
- [32] M. Zhang, H. Lan, and W. Ser, "On comparison of online secondary path modeling methods with auxiliary noise," *IEEE Trans. Speech Audio Process.*, vol. 13, no. 4, pp. 618–628, Jul. 2005.
 [33] M. Zhang, H. Lan, and W. Ser, "Cross-updated active noise control
- [33] M. Zhang, H. Lan, and W. Ser, "Cross-updated active noise control system with online secondary-path modeling," *IEEE Trans. Speech Audio Process.*, vol. 9, no. 5, pp. 598–602, Jul. 2001.
 [34] S. C. Chan and Y. J. Chu, "Performance analysis and design of FxLMS
- [34] S. C. Chan and Y. J. Chu, "Performance analysis and design of FxLMS algorithm in broadband ANC system with online secondary-path modeling," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 20, no. 3, pp. 982–993, Mar. 2012.
- [35] X. Sun, S. M. Kuo, and G. Meng, "Adaptive impulsive noise," J. Sound Vibr., vol. 291, no. 1–2, pp. 516–522, Mar. 2006.
- [36] S. M. Kuo, K. Kuo, and W. S. Gan, "Active noise control: Open problems and challenges," in *Proc. ICGCS '10*, Shanghai, China, Jun. 21–23, 2010.

- [37] M. Shoaib, S. Werner, J. A. Aprolinário, Jr., and S. A. Alshebieli, "Multi- input multi-output fast QR decomposition algorithm for active noise control," in *Proc. EUSIPCO '09*, Glasgow, U.K., Aug. 24–28, 2009.
- [38] [Online]. Available: http://www.eee.hku.hk/~dsp/download/supplementary.pdf



S. C. Chan (S'87–M'92) received the B.Sc. (Eng) and Ph.D. degrees from The University of Hong Kong, Pokfulam, in 1986 and 1992, respectively. Since 1994, he has been with the Department of Electrical and Electronic Engineering, the University of Hong Kong, where he is currently a professor. His research interests include fast transform algorithms, filter design and realization, multirate and biomedical signal processing, communications and array signal processing, high-speed A/D converter architecture, bioinformatics, smart grid image-based

rendering. He is currently a member of the Digital Signal Processing Technical Committee of the IEEE Circuits and Systems Society, and associate editors of *Journal of Signal Processing Systems* and IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS II. He was the chair of the IEEE Hong Kong Chapter of Signal Processing in 2000–2002, an organizing committee member of the 2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, the 2010 International Conference on Image Processing, and an associate editor of IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS I from 2008 to 2009.



Y. J. Chu received the B. S. and M. S. degree in acoustics from the Nanjing University, Nanjing, China, in 2005 and 2008, respectively, and the Ph.D. degree from the electrical and electronic engineering department of The University of Hong Kong, Hong Kong. In 2006, she served as a research assistant in the department of building service engineering of the Polytechnic University of Hong Kong, Hong Kong. Her current research interests are in analysis, design and applications of robust adaptive filtering algorithms and recursive detection methods.



Z. G. Zhang (S'05–M'07) received the B.Sc. degree in electrical and electronic engineering from Tianjin University, Tianjin, China, in 2000, the M.Eng. degree in electrical and electronic engineering from the University of Science and Technology of China, Hefei, China, in 2003, and the Ph.D. degree from the Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong, in 2008. He is currently a Research Assistant Professor with The University of Hong Kong. His research interests include biomedical signal

processing, neural engineering and computational neuroscience.



K. M. Tsui received the B.Eng., M.Phil. and Ph.D. degrees in electrical and electronic engineering from The University of Hong Kong, in 2001, 2004 and 2008, respectively. He is currently working as Post-doctoral Fellow in the department of Electrical and Electronic Engineering at the same university. His main research interests are in smart grid, high-speed A/D converter architecture, biomedical signal processing, array signal processing, digital signal processing, multirate filter bank and wavelet design, and digital filter design, realization and application.