A New Regularized Transform-Domain NLMS Adaptive Filtering Algorithm

S. C. Chan, Y. J. Chu and Z. G. Zhang
Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong, P. R. China
{scchan; yjchu; zgzhang}@eee.hku.hk

***Abstract*—The transform domain normalized LMS (TD-NLMS)-adaptive filtering algorithm is an efficient adaptive filter with fast convergence speed and reasonably low arithmetic complexity. However, it is sensitive to the level of the excitation signal, which may vary significantly over time in speech and audio signals. This paper proposes a new regularized transform domain NLMS (R-TDNLMS) algorithm and studies its mean and mean square convergence performance. The proposed algorithm extends the conventional TDNLMS algorithm by imposing a regularization term on the coefficients to reduce the variance of the estimator. The mean and mean square convergence behaviors of the proposed algorithm are studied to characterize its convergence condition and steady-state excess mean squares error (MSE). It shows that regularization can help to reduce the MSE for coloured inputs by trading slight bias for variance. Moreover, the immunity to varying input signal level is significantly reduced. Computer simulations are conducted to examine the effectiveness of the proposed algorithm and they are in good agreement with the theoretical analysis.**

***Keywords*—Regularization, performance analysis, NLMS, transform domain.**

I. INTRODUCTION

Adaptive filters are frequently used in applications such as system identification and related problems, in which the statistics of the underlying signals are either unknown *a priori*, or slowly varying. One of the most commonly used algorithms is the well known least mean square (LMS) algorithm [1] and its variants [2-5], due to their numerical stability and computational simplicity. An important class of the LMS algorithm is called the transform domain normalized LMS (TDNLMS) algorithm [2-5], which exploits the decorrelation property of transformations such as the discrete Fourier transform (DFT), discrete cosine transform (DCT), and wavelet transform (WT), to approximately whiten the input signal. This helps to reduce the eigenvalue spread of the input autocorrelation matrix and hence significantly improve the convergence speed.

One problem of the TDNLMS algorithm is its sensitivity to the level of the excitation signal, which may vary significantly over time as in speech and audio signals. In such situations, the estimated power of each transformed coefficient may become very small and the mean square errors may increase significantly. To address this issue, a commonly used technique is to introduce some kind of regularization into these algorithms. The regularization technique has attracted much interest recently as a useful tool for reducing the estimation variance, especially when only a small number of data samples are available [6]. It has also been successfully applied to a wide variety of areas such as audio signal processing [7]. For identification of systems with sparse impulse response coefficients, regularization technique was shown to improve the convergence speed and decrease the misadjustment in the LMS algorithms [8].

In this paper, a new regularized TDNLMS (R-TDNLMS) algorithm is proposed. A weighted regularization term on the adaptive filter coefficients is incorporated in the MSE cost function in order to reduce the estimation variance. To quantify the performance of the proposed algorithm, its mean and mean square convergence analyses are performed. Difference equations describing the mean and mean square convergence behavior of the proposed algorithm are derived. Simulation results show that the R-TDNLMS algorithm has lower excess MSE (EMSE) than the conventional TDNLMS algorithm for colored input and better immunity to the variation in input signal power. The theoretical analysis is also found to agree well with the simulation results. The rest of the paper is organized as follows: in Section II, the R-TDNLMS algorithm is derived. In Section III, the mean and mean square convergence behaviors of the proposed algorithm are derived. Simulation results are presented in Section IV. Finally, a conclusion is drawn in Section V.

II. THE R-TDNLMS ALGORITHM

Consider the identification of a linear time-invariant (LTI) finite impulse response (FIR) system with an impulse response coefficient vector $\boldsymbol{w}^*$ having L taps by an adaptive filter of the same length. The unknown system and adaptive filter are both excited by an input $x(n)$ and the measured output of the system is $d(n)$, which is assumed to be corrupted by an additive noise $\eta(n)$. Hence

$$d(n) = \boldsymbol{w}^{*T}\,\boldsymbol{x}(n) + \eta(n), \qquad (1)$$

where $\boldsymbol{x}(n) = [x(n), x(n-1), \cdots, x(n-L-1)]^T$ is the input signal vector. $d(n)$ is also applied to the desired input of the adaptive filter and the weight vector is estimated recursively by minimizing the penalized MSE between $d(n)$ and the output of the adaptive filter $y(n) = \boldsymbol{w}^T(n)\boldsymbol{x}(n)$ as follows

$$\min_{\boldsymbol{w}} \varepsilon_{RMSE} = E[(d(n) - \boldsymbol{w}^T(n)\boldsymbol{x}(n))^2] + \xi \,\|\, \boldsymbol{D}\boldsymbol{w}(n) \,\|_2^2, \qquad (2)$$

where $\boldsymbol{w}(n) = [w_1(n), w_2(n), \cdots, w_L(n)]^T$ is the coefficient vector of the adaptive filter, $E[\cdot]$ is the expectation operator and the last term represents a regularization which helps to reduce the variance of the estimation especially when the covariance matrix of $\boldsymbol{x}(n)$ is close to singular due to lacking of excitation or when $\boldsymbol{w}^*$ is spare. ξ is the regularization parameter and $\boldsymbol{D}$ can be made adaptive so as to approximate different regularization methods.

If the instantaneous MSE is used to approximate the MSE, then a recursive algorithm for determining $\boldsymbol{w}^*$ can be obtained by minimizing the objective function in (2) as

$$e(n) = d(n) - \boldsymbol{w}^T(n)\boldsymbol{x}(n)\,, \tag{3}$$

$$\boldsymbol{w}(n+1) = \boldsymbol{w}(n) + \mu[\boldsymbol{x}(n)e(n) - \xi\boldsymbol{R}_w\boldsymbol{w}(n)]\,, \tag{4}$$

where $\boldsymbol{R}_w = \boldsymbol{D}^T\boldsymbol{D}$.

Following the transformation approach in [9-10], the input can be transformed to speed up the convergence of the LMS algorithms above and the following updating equation can be obtained

$$e(n) = d(n) - \boldsymbol{W}^T(n)\boldsymbol{X}_C(n)\,, \tag{5}$$

$$\boldsymbol{W}(n+1) = \boldsymbol{W}(n) + \mu\boldsymbol{D}_x(\boldsymbol{X}_C(n)e(n) - \xi\boldsymbol{R}_W\boldsymbol{W}(n))\,, \tag{6}$$

where $\boldsymbol{W}(n) = \boldsymbol{C}\boldsymbol{w}(n) = [W_{C,1}(n), W_{C,2}(n), \cdots, W_{C,L}(n)]^T$, $\boldsymbol{X}_C(n) = \boldsymbol{C}\boldsymbol{x}(n) = [X_{C,1}(n), X_{C,2}(n), \cdots, X_{C,L}(n)]^T$ are, respectively, the transformed adaptive weight vector and signal vector. $\boldsymbol{C}$ is an $L \times L$ transformation matrix such as the DFT or DCT. $\boldsymbol{R}_W = \boldsymbol{C}\boldsymbol{R}_w\boldsymbol{C}^T$ and $\boldsymbol{D}_x = \mathrm{diag}[\varepsilon_1^{-1}(n), \varepsilon_2^{-1}(n), \cdots, \varepsilon_L^{-1}(n)]$ is an element-wise normalization matrix with $\varepsilon_i(n)$ being the estimated power of the i-th signal component after transformation.

III. Performance Analysis

In this section, we analyze the convergence performance of the proposed R-TDNLMS algorithm. The following assumptions are made:

(A1) $\{\boldsymbol{x}(n)\}$ is an independent identically distributed (i.i.d) Gaussian random sequence with zero-mean and covariance matrix $\boldsymbol{R}_{xx}$;

(A2) $\boldsymbol{W}(n)$, $x(n)$ and $\eta(n)$ are statistically independent;

(A3) the estimated power $\varepsilon_i(n)$ averages within a short period of time and hence $\boldsymbol{D}_x$ is a constant matrix.

(A2) is the independence assumption, which is a good approximation for large value of L and for small to medium step-size to simplify the convergence analysis of adaptive filtering algorithms. Moreover, we denote $\boldsymbol{W}^* = \boldsymbol{R}_{X_cX_c}^{-1}\boldsymbol{P}_{dX_c}$, where $\boldsymbol{P}_{dX_c} = E[d(n)\boldsymbol{X}_C(n)]$ is the ensemble-averaged cross-correlation vector between $\boldsymbol{X}_C(n)$ and $d(n)$. $\boldsymbol{W}^*$ is related to the optimal Wiener solution $\boldsymbol{w}_{\mathrm{OPT}}$ as $\boldsymbol{w}_{\mathrm{OPT}} = \boldsymbol{R}_{XX}^{-1}\boldsymbol{P}_{dX} = \boldsymbol{C}\boldsymbol{W}^*$.

A. Mean Convergence Analysis

Taking expectation on both sides of (6), we have

$$\begin{aligned} E[\boldsymbol{W}(n+1)] &= E[\boldsymbol{W}(n)] + \mu\boldsymbol{D}_x \\ &\cdot \{\boldsymbol{P}_{dX_C} - (\boldsymbol{R}_{X_CX_C} + \xi\boldsymbol{R}_W)E[\boldsymbol{W}(n)]\}. \end{aligned} \tag{7}$$

First we assume that the step-size is appropriately chosen so that the algorithm converges. At the steady state, (7) reads:

$$(\boldsymbol{R}_{X_CX_C} + \xi\boldsymbol{R}_W)\boldsymbol{W}_R = \boldsymbol{P}_{dX_C}\,, \tag{8}$$

where $\boldsymbol{W}_R = E[\boldsymbol{W}(\infty)]$ is the desired regularized Wiener solution.

Next we examine the convergence rate by introducing the weight error vector $\boldsymbol{v}(n) = \boldsymbol{W}(n) - \boldsymbol{W}_R$ in (7):

$$E[\boldsymbol{v}(n+1)] = (\boldsymbol{I} - \mu\boldsymbol{D}_x(\boldsymbol{R}_{X_CX_C} + \xi\boldsymbol{R}_W))E[\boldsymbol{v}(n)]\,, \tag{9}$$

where $\boldsymbol{I}$ is the identity matrix. Let $\widetilde{\boldsymbol{U}}\widetilde{\boldsymbol{\Lambda}}\widetilde{\boldsymbol{U}}^T$ be the eigendecomposition of $\overline{\boldsymbol{R}}_{xx} = \boldsymbol{D}_x^{1/2}\widetilde{\boldsymbol{R}}_{X_CX_C}\boldsymbol{D}_x^{1/2}$ with $\widetilde{\boldsymbol{R}}_{X_CX_C} = \boldsymbol{R}_{X_CX_C} + \xi\boldsymbol{R}_W$. Using (9) and expressing $\boldsymbol{v}(n)$ in the coordinate, $\boldsymbol{V}(n) = \widetilde{\boldsymbol{U}}^T\boldsymbol{D}_x^{-1/2}\boldsymbol{v}(n)$, we get the difference equation for the i-th element of $E[\boldsymbol{V}(n)]$ as follows

$$E[\boldsymbol{V}(n+1)]_i = (1 - \mu\widetilde{\lambda}_i)E[\boldsymbol{V}(n)]_i\,, \tag{10}$$

where $\widetilde{\lambda}_i$ is the i-th eigenvalue of $\overline{\boldsymbol{R}}_{xx}$. Thus, the mean weight vector of the adaptive filter will converge if

$$0 < \mu < 2/\widetilde{\lambda}_i\,. \tag{11}$$

Therefore, the maximum possible step-size is $\mu_{\max} = 2/\widetilde{\lambda}_{\max}$, where $\widetilde{\lambda}_{\max}$ is the maximum eigenvalue of $\overline{\boldsymbol{R}}_{xx}$.

It can be seen that if the input is so distributed that $\boldsymbol{R}_{X_CX_C}$ has zero eigenvalues, then these eigenmodes can never converge and the solution may be significantly biased. With sufficient regularization, the eigenvalues of $\overline{\boldsymbol{R}}_{xx}$ can be made nonzero and hence the solution will be given by (8) with a controllable bias.

B. Mean Square Convergence Analysis

To evaluate the mean square behavior, multiplying $\boldsymbol{v}(n)$ by its transpose and taking expectation, one gets a difference equation of the weight error covariance matrix $\boldsymbol{\Xi}(n) = E[\boldsymbol{v}(n)\boldsymbol{v}^T(n)]$ as follows

$$\begin{aligned} \boldsymbol{\Xi}(n+1) &= \boldsymbol{\Xi}(n) + \mu\boldsymbol{D}_x\{\boldsymbol{P}_{dX_C}E[\boldsymbol{v}(n)]^T - \widetilde{\boldsymbol{R}}_{X_CX_C}\boldsymbol{\Xi}_{wv}(n)\} \\ &+ \mu\{E[\boldsymbol{v}(n)]\boldsymbol{P}_{dX_C}^T - \boldsymbol{\Xi}_{vw}(n)\widetilde{\boldsymbol{R}}_{X_CX_C}\}\boldsymbol{D}_x^T \\ &+ \mu^2\boldsymbol{D}_xE[(\boldsymbol{X}_C(n)e(n) - \xi\boldsymbol{R}_W\boldsymbol{W}(n)) \\ &\cdot(\boldsymbol{X}_C^T(n)e(n) - \xi\boldsymbol{W}^T(n)\boldsymbol{R}_W)]\boldsymbol{D}_x^T, \end{aligned} \tag{12}$$

where $\boldsymbol{\Xi}_{vw}(n) = \boldsymbol{\Xi}_{wv}^T(n) = E[\boldsymbol{v}(n)\boldsymbol{W}^T(n)] = \boldsymbol{\Xi}_{vv}(n) + E[\boldsymbol{v}(n)]\boldsymbol{W}_R^T$ by noting that $e(n) = [\boldsymbol{W}^* - \boldsymbol{W}(n)]^T\boldsymbol{X}_C(n) + \eta(n) = \hat{\boldsymbol{v}}^T\boldsymbol{X}_C(n) + \eta(n)$, where $\hat{\boldsymbol{v}} = \boldsymbol{W}^* - \boldsymbol{W}_R - \boldsymbol{v}(n) = \Delta\boldsymbol{W} - \boldsymbol{v}(n)$.

It can be shown that [11] the R-TDNLMS algorithm will converge in the mean squares sense if the step-size satisfies the following convergence condition

$$\mu \le \frac{1}{2Tr(\widetilde{\boldsymbol{R}}_{X_CX_C}(n)\boldsymbol{D}_x) + Tr(\boldsymbol{R}_{X_CX_C}(n)\boldsymbol{D}_x)}, \tag{13}$$

which is very close to that of the conventional TDNLMS algorithm $\mu \le \dfrac{1}{3Tr(\boldsymbol{R}_{X_C X_C}(n)\boldsymbol{D}_x)}$ without regularization.

If the algorithm converges, $E[\boldsymbol{v}(\infty)] = \boldsymbol{0}$, and the last term in (12) will reduce to

$$S(\infty) = \mu^2 \boldsymbol{D}_x[\boldsymbol{A}_0(\infty) - \boldsymbol{A}_1(\infty) - \boldsymbol{A}_2(\infty) + \boldsymbol{A}_3(\infty) + \boldsymbol{R}_{X_C X_C}\sigma_\eta^2]\boldsymbol{D}_x^T,$$

$$\boldsymbol{A}_0(\infty) = 2\boldsymbol{R}_{X_C X_C}\boldsymbol{\Xi}_{vv}(\infty)\boldsymbol{R}_{X_C X_C} + 2\xi^2 \boldsymbol{R}_W \boldsymbol{W}_R \boldsymbol{W}_R^T \boldsymbol{R}_W^T + Tr(\{\Delta\boldsymbol{W}\Delta\boldsymbol{W}^T + \boldsymbol{\Xi}_{vv}(\infty)\}\boldsymbol{R}_{X_C X_C})\boldsymbol{R}_{X_C X_C},$$

$$\boldsymbol{A}_1(\infty) = \boldsymbol{A}_2^T(\infty) = \xi(\xi\boldsymbol{R}_W \boldsymbol{W}_R \boldsymbol{W}_R^T - \boldsymbol{R}_{X_C X_C}\boldsymbol{\Xi}_{vv}(\infty))\boldsymbol{R}_W,$$

$$\boldsymbol{A}_3(\infty) = \xi^2 \boldsymbol{R}_W(\boldsymbol{W}_R\boldsymbol{W}_R^T + \boldsymbol{\Xi}_{vv}(\infty))\boldsymbol{R}_W.$$

Therefore,

$$\begin{aligned}
&\boldsymbol{A}_0(\infty) - \boldsymbol{A}_1(\infty) - \boldsymbol{A}_2(\infty) + \boldsymbol{A}_3(\infty)\\
&\quad = 2(\boldsymbol{R}_{X_C X_C} + \xi\boldsymbol{R}_W)\boldsymbol{\Xi}_{vv}(\infty)(\boldsymbol{R}_{X_C X_C} + \xi\boldsymbol{R}_W)\\
&\quad - \xi(\boldsymbol{R}_{X_C X_C}\boldsymbol{\Xi}_{vv}(\infty)\boldsymbol{R}_W + \boldsymbol{R}_W^T\boldsymbol{\Xi}_{vv}(\infty)\boldsymbol{R}_{X_C X_C})\\
&\quad + Tr(\Delta\boldsymbol{W}\Delta\boldsymbol{W}^T\boldsymbol{R}_{X_C X_C} + \boldsymbol{\Xi}_{vv}(\infty)\boldsymbol{R}_{X_C X_C})\boldsymbol{R}_{X_C X_C}\\
&\quad + \xi^2\boldsymbol{R}_W\boldsymbol{W}_R\boldsymbol{W}_R^T\boldsymbol{R}_W - \xi^2\boldsymbol{R}_W\boldsymbol{\Xi}_{vv}(\infty)\boldsymbol{R}_W.
\end{aligned}$$

Assuming μ^2 and ξ^2 are small, we can drop the last term to get

$$\begin{aligned}
S(\infty) \approx \mu^2\boldsymbol{D}_x[&2(\boldsymbol{R}_{X_C X_C} + \xi\boldsymbol{R}_W)\boldsymbol{\Xi}_{vv}(\infty)(\boldsymbol{R}_{X_C X_C} + \xi\boldsymbol{R}_W)\\
&- \xi(\boldsymbol{R}_{X_C X_C}\boldsymbol{\Xi}_{vv}(\infty)\boldsymbol{R}_W + \boldsymbol{R}_W^T\boldsymbol{\Xi}_{vv}(\infty)\boldsymbol{R}_{X_C X_C})\\
&+ (J_* + \sigma_{\min}^2)\boldsymbol{R}_{X_C X_C} + \xi^2\boldsymbol{R}_W\boldsymbol{W}_R\boldsymbol{W}_R^T\boldsymbol{R}_W]\boldsymbol{D}_x^T,
\end{aligned} \tag{14}$$

where $\sigma_{\min}^2 = Tr(\Delta\boldsymbol{W}\Delta\boldsymbol{W}^T\boldsymbol{R}_{X_C X_C}) + \sigma_\eta^2$. This also gives an upper bound of the MSE since we dropped only the negative terms. Substituting (14) into (12) gives

$$\begin{aligned}
\boldsymbol{\Xi}_{vv}(\infty) \approx\ &\boldsymbol{\Xi}_{vv}(\infty) - \mu(\boldsymbol{D}_x\widetilde{\boldsymbol{R}}_{X_C X_C}\boldsymbol{\Xi}_{vv}(\infty) + \boldsymbol{\Xi}_{vv}(\infty)\widetilde{\boldsymbol{R}}_{X_C X_C}\boldsymbol{D}_x^T)\\
&+ 2\mu^2\boldsymbol{D}_x\widetilde{\boldsymbol{R}}_{X_C X_C}\boldsymbol{\Xi}_{vv}(\infty)\widetilde{\boldsymbol{R}}_{X_C X_C}\boldsymbol{D}_x + \mu^2\boldsymbol{D}_x[(J_* + \sigma_{\min}^2)\boldsymbol{R}_{X_C X_C}\\
&+ \xi^2\boldsymbol{R}'_w\boldsymbol{W}_R\boldsymbol{W}_R^T\boldsymbol{R}'_w]\boldsymbol{D}_x^T.
\end{aligned} \tag{15}$$

By expressing $\boldsymbol{v}(n)$ in the transformed coordinate: $\boldsymbol{V}(n) = \widetilde{\boldsymbol{U}}^T\boldsymbol{D}_x^{-1/2}\boldsymbol{v}(n)$, (15) can be further simplified to

$$\begin{aligned}
&\boldsymbol{\Xi}_{VV}(\infty) \approx \boldsymbol{\Xi}_{VV}(\infty) - \mu\widetilde{\boldsymbol{U}}^T\boldsymbol{D}_x^{1/2}\widetilde{\boldsymbol{R}}_{X_C X_C}\boldsymbol{D}_x^{1/2}\widetilde{\boldsymbol{U}}\boldsymbol{\Xi}_{VV}(n)\\
&- \mu\boldsymbol{\Xi}_{VV}(n)\widetilde{\boldsymbol{U}}\boldsymbol{D}_x^{1/2}\widetilde{\boldsymbol{R}}_{X_C X_C}\boldsymbol{D}_x^{1/2}\widetilde{\boldsymbol{U}}^T + \mu^2\Gamma(\infty)\\
&+ 2\mu^2\widetilde{\boldsymbol{U}}^T\boldsymbol{D}_x^{1/2}\widetilde{\boldsymbol{R}}_{X_C X_C}\boldsymbol{D}_x^{1/2}\widetilde{\boldsymbol{U}}\ \boldsymbol{\Xi}_{VV}(\infty)\widetilde{\boldsymbol{U}}^T\boldsymbol{D}_x^{1/2}\widetilde{\boldsymbol{R}}_{X_C X_C}\boldsymbol{D}_x^{1/2}\widetilde{\boldsymbol{U}}^T,
\end{aligned} \tag{16}$$

where $\Gamma(\infty) = (J_* + \sigma_{\min}^2)\Gamma_1 + \Gamma_2$, $\Gamma_1 = \widetilde{\boldsymbol{U}}\boldsymbol{D}_x^{1/2}\boldsymbol{R}_{X_C X_C}\boldsymbol{D}_x^{1/2}\widetilde{\boldsymbol{U}}^T$ and $\Gamma_2 = \xi^2\widetilde{\boldsymbol{U}}\boldsymbol{D}_x^{1/2}\boldsymbol{R}_W\boldsymbol{W}_R\boldsymbol{W}_R^T\boldsymbol{R}_W\boldsymbol{D}_x^{1/2}\widetilde{\boldsymbol{U}}^T$.

The diagonal values of (16) reads:

$$\begin{aligned}
&[\boldsymbol{\Xi}_{VV}(\infty)]_{i,i} \approx [\boldsymbol{\Xi}_{VV}(\infty)]_{i,i} - 2\mu\widetilde{\lambda}_i[\boldsymbol{\Xi}_{VV}(\infty)]_{i,i} + \mu^2\\
&\cdot(2\widetilde{\lambda}_i^2[\boldsymbol{\Xi}_{VV}(\infty)]_{i,i} + (J_* + \sigma_{\min}^2)[\Gamma_1(\infty)]_{i,i} + [\Gamma_2(\infty)]_{i,i}),
\end{aligned} \tag{17}$$

where $[\Gamma_1(\infty)]_{i,i} = \Gamma_{1,i}$, and $[\Gamma_2(\infty)]_{i,i} = \Gamma_{2,i}$.

On the other hand, the EMSE is given by

$$J_* = tr(\boldsymbol{\Xi}_{vv}(\infty)\boldsymbol{R}_{X_C X_C}) = tr(\boldsymbol{\Xi}_{VV}(\infty)\widetilde{\boldsymbol{U}}^T\boldsymbol{D}_x^{1/2}\boldsymbol{R}_{X_C X_C}\boldsymbol{D}_x^{1/2}\widetilde{\boldsymbol{U}}). \tag{18}$$

If the transform can approximately diagonalize $\boldsymbol{R}_{xx}$, then $\overline{\boldsymbol{R}}_{xx} = \boldsymbol{D}_x^{1/2}\widetilde{\boldsymbol{R}}_{X_C X_C}\boldsymbol{D}_x^{1/2} \approx \boldsymbol{I}$, and $\widetilde{\boldsymbol{U}}$ are approximately equal to the identity matrix. Hence,

$\widetilde{\boldsymbol{U}}^T\boldsymbol{D}_x^{1/2}\boldsymbol{R}_{X_C X_C}\boldsymbol{D}_x^{1/2}\widetilde{\boldsymbol{U}} = \widetilde{\boldsymbol{\Lambda}} - \xi\widetilde{\boldsymbol{U}}^T\boldsymbol{D}_x^{1/2}\boldsymbol{R}_W\boldsymbol{D}_x^{1/2}\widetilde{\boldsymbol{U}} \approx \widetilde{\boldsymbol{\Lambda}} - \xi\boldsymbol{D}_x^{1/2}\boldsymbol{R}_W\boldsymbol{D}_x^{1/2}$.

Moreover, for diagonal dominance $\boldsymbol{R}_W$, we further have

$$\begin{aligned}
J_* &= tr(\boldsymbol{\Xi}_{VV}(\infty)(\widetilde{\Lambda} - \xi\boldsymbol{D}_x^{1/2}\boldsymbol{R}_W\boldsymbol{D}_x^{1/2}))\\
&\approx \sum_{i=1}^{L}[\boldsymbol{\Xi}_{VV}(\infty)]_{i,i}(\widetilde{\lambda}_i - \xi\varepsilon_i^{-1}\boldsymbol{R}_{W_i,i}),
\end{aligned} \tag{19}$$

where $\boldsymbol{R}_{W\_i,i}$ is the *i-th* diagonal value of $\boldsymbol{R}_W$.

Solving for $[\boldsymbol{\Xi}_{VV}(\infty)]_{i,i}$ from (17), one gets

$$[\boldsymbol{\Xi}_{VV}(\infty)]_{i,i} = \frac{\mu}{2\widetilde{\lambda}_i(1 - \widetilde{\lambda}_i\mu)}((J_* + \sigma_{\min}^2)\Gamma_{1,i} + \Gamma_{2,i}). \tag{20}$$

Consequently, J_* is found to be

$$J_* = \frac{\frac{1}{2}\mu(\sigma_{\min}^2\phi_{\text{RNLMS}} + \phi_\xi)}{(1 - \frac{1}{2}\mu\phi_{\text{RNLMS}})}, \tag{21}$$

where $\phi_{\text{RNLMS}} = \sum_{i=1}^{L}\dfrac{\Gamma_{1,i}(\widetilde{\lambda}_i - \xi\varepsilon_i^{-1}R_{W\_i,i})}{\widetilde{\lambda}_i(1 - \widetilde{\lambda}_i r_\mu)}$, $\phi_\xi = \sum_{i=1}^{L}\dfrac{\Gamma_{2,i}(\widetilde{\lambda}_i - \xi\varepsilon_i^{-1}R_{W\_i,i})}{\widetilde{\lambda}_i(1 - \widetilde{\lambda}_i r_\mu)}$.

IV. Simulation Results

In this section, computer simulations are conducted to evaluate the convergence behavior of the proposed algorithm and verify the analytical results obtained in section III. As a comparison, we also consider the conventional TDNLMS algorithm. The DCT transformation is employed due to its wide usage and efficiency in practice. The power of the input element is estimated recursively by using a forgetting factor $\varepsilon_i(n) = (1 - \alpha_\varepsilon)\varepsilon_i(n-1) + \alpha_\varepsilon X_{C,i}^2(n)$ with α_ε =0.01. All simulations are performed using the system identification model and the results are averaged over 200 independent runs.

A. Experiment 1: Colored Gaussian Input

In this experiment, the following first order autoregressive (AR) process is employed as the input: $x(n) = 0.95x(n-1) + g(n)$, where $g(n)$ is a zero-mean and white Gaussian noise. The input power has been normalized. The unknown system to be estimated is an L-order (L=15) FIR filter. Different signal-to-noise ratios (SNRs) at the system output (SNR=0, 10 and 20 dB) are used to examine the effect of the regularization on the proposed algorithm. The step-size for the TDNLMS and R-TDNLMS algorithms is chosen as 0.007. The regularization parameter for R-TDNLMS is chosen as $\xi = 0.1$, 0.02 and 0.004 for different SNRs. The learning curves of EMSE for both algorithms are shown in Figs. 1(a), (b), and (c). It can be seen that the R-TDNLMS algorithm generally converges faster and to a lower steady-state EMSE compared to the conventional TDNLMS algorithm. The advantages are more significant when the SNR is low. The estimated steady-state EMSE agree well with the simulation results.

B. Experiment 2: Power Varying Input

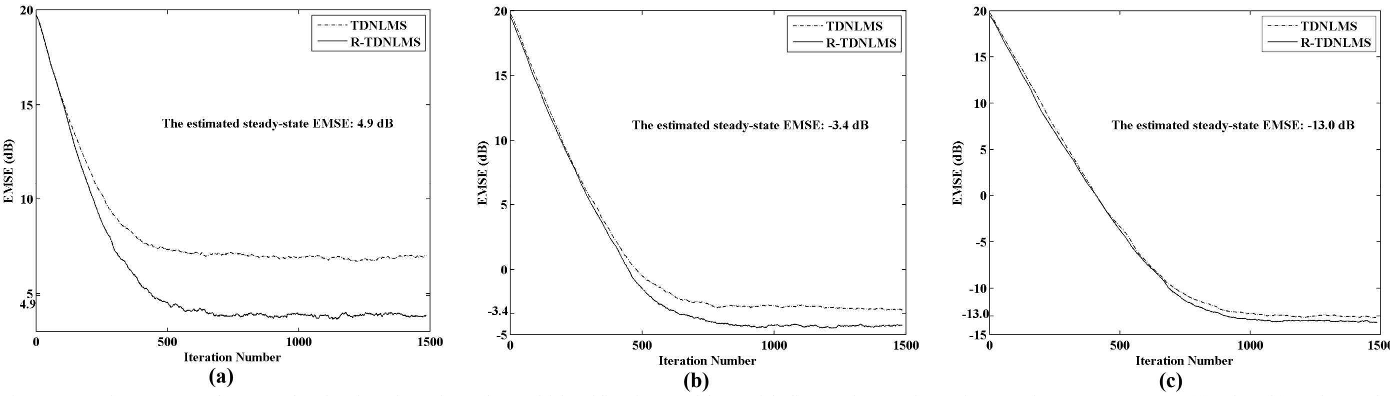

Fig. 1 Learning curves of EMSE for the time-invariant channel identification problem with first-order AR input in experiment 1 at SNR= (a) 0dB (b) 10dB and (c) 20dB.

In this experiment, the input signal is a segment of music. The SNR is set to be 15 dB. The system order is L=8. The step-size for both algorithms are 0.007. The regularization parameter for the R-TDNLMS algorithm is chosen as $\xi = 0.02$. The learning curves of EMSE are shown in Fig. 2. It can be seen that the TDNLMS algorithm is very sensitive to the input signal because the input power is varying considerably. The R-TDNLMS algorithm, on the other hand, has a high immunity to variation in input power and it achieves a lower steady-state EMSE values compared to the conventional TDNLMS algorithm.

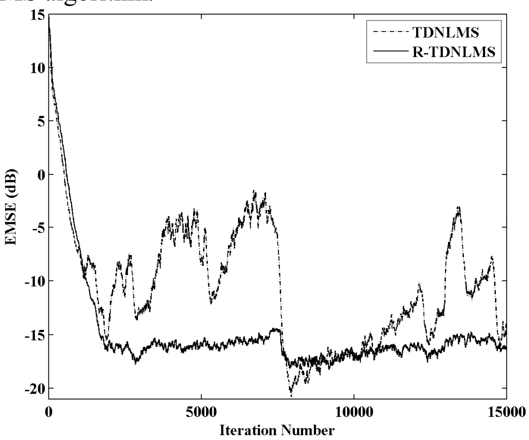

Fig. 2 Learning curves of EMSE for the time-invariant channel identification problem with music input in experiment 2 at SNR=15 dB.

V. CONCLUSION

A new regularized TD-NLMS adaptive filtering algorithm has been proposed. The algorithm is obtained by adding a penalizing term on the weighted 2-norm of the coefficient vector to the conventional MSE cost function. New difference equations describing the mean and mean square convergence behaviors of the proposed algorithm are developed. New expressions for step-size bound and EMSE are also derived. The R-TDNLMS algorithm is found to have a lower EMSE for coloured input and better immunity to variation of input signal power than the conventional TDNLMS. Theoretical analysis is also found to agree well with computer simulation under different conditions tested.

REFERENCES

[1] B. Widrow, J. M. McCool, M. G. Larimore, and C. R. Johnson, Jr., “Stationary and nonstationary learning characteristics of the LMS adaptive filter,” in *Proc. IEEE*, vol. 64, pp. 1151-1162, Aug. 1976.

[2] J. I. Nagumo and A. Noda, “A learning method for system identification,” *IEEE Trans. Automat. Contr*., vol. AC-12, pp. 282-287, June 1967.

[3] S. Narayan, A. M. Peterson and M. J. Narasimha, “Transform domain LMS algorithm,” *IEEE Trans. Acoust. Speech, Signal Process*., vol. ASSP-31, pp. 609-615, June 1983.

[4] J. C. Lee and C. K. Un, “Performance of transform-domain LMS adaptive digital filters,” *IEEE Trans. Acoustics, Speech, and Signal Process*., vol. 34, No. 3, pp. 499-510, 1986.

[5] S. C. Chan and Y. Zhou, “On the convergence analysis of the transform domain normalized LMS and related M-estimate algorithms,” in *Proc. IEEE APCCAS 2008*, Macao, Nov 30 - Dec 27, 2008.

[6] J. Fan and R. Li, “Variable selection via nonconcave penalized likelihood and its oracle properties,”. *J. Am. Stat. Assoc.*, vol. 96, no. 456, pp. 1348-1360, Dec. 2001.

[7] T. Van Waterschoot, G. Rombouts, and M. Moonen, “Optimally regularized adaptive filtering algorithms for room acoustic signal enhancement,” *Signal Process.*, vol. 88, pp. 594-611, 2008.

[8] Y. Chen, Y. Gu, and A. O. Hero, “Sparse LMS for system identification,” in *Proc. ICASSP2009*, Taipei, Taiwan, Apri 19-24, 2009, pp. 3125-3128.

[9] S. C. Chan, Y. J. Chu, Z. G. Zhang, and Y. Zhou, “On the convergence behavior of the noise-constrained NLMS algorithm,” in *Proc. IEEE ISCAS 2009*, Taipei, May 24-27, 2009.

[10] S. C. Chan, Z. G. Zhang Y. Zhou, and Y. Hu, “A new noise-constrained normalized least mean squares adaptive filtering algorithm,” in *Proc. IEEE APCCAS 2008*, Macao, China, Dec. 2008.

[11] S. C. Chan, “A new regularized transform domain NLMS adaptive filtering algorithm – performance analysis and applications,” Internal report, Department of EEE, The University of Hong Kong, 2010.