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 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 coefficients to reduce the variance of the estimator. 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 \( \mathbf{w}^* \) 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

\[
d(n) = \mathbf{w}^* x(n) + \eta(n),
\]

where \( x(n) = [x(n), x(n-1), \ldots, x(n-L-1)]^T \) 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 \( y(n) = \mathbf{w}^T(n) x(n) \) as follows

\[
\min_{\mathbf{w}} E_n[\varepsilon_n^2] = E[(d(n) - w^T(n) x(n))^2] + \frac{\lambda}{2} \| D w(n) \|_2^2,
\]

where \( w(n) = [w_1(n), w_2(n), \ldots, w_L(n)]^T \) 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 \( x(n) \) is close to singular due to lacking of excitation or when \( w^* \) is sparse. \( \xi \) is the regularization parameter and \( D \) 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 \( w^* \) can be obtained by minimizing the objective function in (2) as

\[
e(n) = d(n) - w^T(n)x(n),
\]

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

where \( \mathbf{R}_w = D^T D \).

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

\[
e(n) = d(n) - W^T(n)X_c(n),
\]

\[
W(n + 1) = W(n) + \mu D_\eta (X_v(n)e(n) - \xi \mathbf{R}_v W(n)),
\]

where \( W(n) = Cw(n) = [W_{c_1}(n), W_{c_2}(n), \ldots, W_{c_L}(n)]^T \), \( X_c(n) = Cx(n) = [X_{c_1}(n), X_{c_2}(n), \ldots, X_{c_L}(n)]^T \) are, respectively, the transformed adaptive weight vector and signal vector. \( C \) is an \( L \times L \) transformation matrix such as the DFT or DCT. \( \mathbf{R}_w = \mathbf{CR}_w \mathbf{C}^T \) and \( D_\eta = \text{diag}[\varepsilon_1^{-1}(n), \varepsilon_2^{-1}(n), \ldots, \varepsilon_L^{-1}(n)] \) 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) \( \{x(n)\} \) is an independent identically distributed (i.i.d) Gaussian random sequence with zero-mean and covariance matrix \( \mathbf{R}_x \);

(A2) \( W(n) \), \( x(n) \) and \( \eta(n) \) are statistically independent;

(A3) the estimated power \( \varepsilon_i(n) \) averages within a short period of time and hence \( D_\eta \) 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 \( W^* = R^T_{s_x} P_{s_x} \), where \( P_{s_x} = E[d(n)X_c(n)] \) is the ensemble-averaged cross-correlation vector between \( X_c(n) \) and \( d(n) \). \( W^* \) is related to the optimal Wiener solution \( w_{opt} \) as \( w_{opt} = R^{-1}_{s_x} P_{s_x} = CW^* \).

A. Mean Convergence Analysis

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

\[
E[W(n + 1)] = E[W(n)] + \mu D_\eta \left\{ P_{s_x} - (R_{s_x} + \xi \mathbf{R}_v)E[W(n)] \right\}.
\]

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

\[
(R_{s_x} + \xi \mathbf{R}_v)W_s = P_{s_x},
\]

where \( W_s = E[W(\infty)] \) is the desired regularized Wiener solution.

Next we examine the convergence rate by introducing the weight error vector \( v(n) = W(n) - W_s \) in (7):

\[
E[v(n + 1)] = (I - \mu D_\eta (R_{s_x} + \xi \mathbf{R}_v))E[v(n)],
\]

where \( I \) is the identity matrix. Let \( \mathbf{U} \mathbf{A} \mathbf{U}^T \) be the eigendecomposition of \( \mathbf{R}_x \mathbf{D}^{-1} \mathbf{R}_x^T \mathbf{D}^{-2} \) with \( \hat{R}_{s_x} = R_{s_x} + \xi \mathbf{R}_v \). Using (9) and expressing \( v(n) \) in the coordinate, \( V(n) = \mathbf{U}^T D_\eta^{-1} v(n) \), we get the difference equation for the \( i \)-th element of \( E[V(n)] \) as follows

\[
E[V(n + 1)] = (1 - \mu \lambda_i)E[V(n)],
\]

where \( \lambda_i \) is the \( i \)-th eigenvalue of \( \mathbf{R}_x \mathbf{D}^{-1} \). Thus, the mean weight vector of the adaptive filter will converge if

\[
0 < \mu < 2 / \lambda_{max},
\]

Therefore, the maximum possible step-size is \( \mu_{max} = 2 / \lambda_{max} \), where \( \lambda_{max} \) is the maximum eigenvalue of \( \mathbf{R}_x \mathbf{D}^{-1} \).

It can be seen that if the input is so distributed that \( R_{s_x} \) has zero eigenvalues, then these eigenmodes can never converge and the solution may be significantly biased. With sufficient regularization, the eigenvalues of \( \mathbf{R}_x \mathbf{D}^{-1} \) 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 \( v(n) \) by its transpose and taking expectation, one gets a difference equation for the weight error covariance matrix \( \mathbf{E}(n) = E[v(n)v^T(n)] \) as follows

\[
\mathbf{E}(n + 1) = \mathbf{E}(n) + \mu \mathbf{D}_\eta \left\{ \mathbf{P}_{s_x} - \mathbf{R}_{s_x} \mathbf{E}(n) \right\} + \mu^2 \mathbf{D}_\eta \left\{ \mathbf{E}(n) \mathbf{D}_\eta^{-1} \mathbf{D}_\eta^{-1} \right\} + \mu^2 \mathbf{E}(n) \mathbf{D}_\eta^{-1} \mathbf{D}_\eta^{-1} - \mathbf{R}_{s_x} \mathbf{E}(n) \mathbf{R}_{s_x}
\]

where \( \mathbf{E}(n) = \mathbf{E}(n) + \mu \mathbf{E}(n) \mathbf{D}_\eta^{-1} \mathbf{D}_\eta^{-1} - \mathbf{R}_{s_x} \mathbf{E}(n) \mathbf{R}_{s_x} \) by noting that \( e(n) = [W^* - W(n)]^T X_c(n) + \eta(n) = \tilde{v}^T X_c(n) + \eta(n) \), where \( \tilde{v} = W^* - W_s = \tilde{W} - v(n) \). 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

\[
\mu \leq \frac{1}{2 \text{Tr}(\mathbf{R}_{s_x} (n) \mathbf{D}_\eta) + \text{Tr}(\mathbf{R}_{s_x} (n) \mathbf{D}_\eta)}.
\]
which is very close to that of the conventional TDNLMS
good approximation.
If the algorithm converges, \( E[\nu(\infty)] = 0 \), and the last term in (12) will reduce to
\[
S(\infty) = \mu^2 \mathbf{D} \left[ (A(\infty) - A(\infty) + A(\infty) + R_{xx}) \right] \mathbf{D}^T.
\]
Therefore,
\[
A(\infty) = \frac{1}{2}(R_{xx} + \mathbf{D}^T) \mathbf{X}_n(\infty)(R_{xx} + \mathbf{D}) + (A(\infty) - A(\infty) + A(\infty))
\]
\[
- \xi(R_{xx} \mathbf{X}_n(\infty) R_{xx} + \mathbf{D}^T) + (J_1 + \sigma_{\text{min}}^2) R_{xx} \mathbf{D}^T + \xi^2 R_{xx} W_x^T W_x R_{xx}.
\]
Assuming \( \mu^2 \) and \( \xi^2 \) are small, we can drop the last term to get
\[
S(\infty) = \frac{1}{2}(R_{xx} + \mathbf{D}^T) \mathbf{X}_n(\infty)(R_{xx} + \mathbf{D}) + (A(\infty) - A(\infty) + A(\infty))
\]
\[
- \xi(R_{xx} \mathbf{X}_n(\infty) R_{xx} + \mathbf{D}^T) + (J_1 + \sigma_{\text{min}}^2) R_{xx} \mathbf{D}^T + \xi^2 R_{xx} W_x^T W_x R_{xx}.
\]
(14)

where \( \sigma_{\text{min}}^2 = \text{Tr}(\mathbf{D}^T \mathbf{R}_{xx} \mathbf{D}) + \xi^2 \mathbf{R}_{xx} W_x^T W_x \mathbf{R}_{xx} \).

By expressing \( \mathbf{v}(n) \) in the transformed coordinate: \( \mathbf{V}(n) = \mathbf{U}^T \mathbf{D}^{-1/2} \mathbf{v}(n) \), (15) can be further simplified to
\[
\mathbf{X}_n(\infty) = \mathbf{X}_n(\infty) - \mu \mathbf{U}^T \mathbf{D}^{1/2} \mathbf{R}_{xx} \mathbf{D}^{1/2} \mathbf{X}_n(\infty)
\]
\[
- \mu \mathbf{X}_n(\infty) \mathbf{U} \mathbf{D}^{1/2} \mathbf{R}_{xx} \mathbf{D}^{1/2} \mathbf{U}^T \mathbf{X}_n(\infty) + \mu \mathbf{U}^T \mathbf{G}(\infty)
\]
\[
+ 2 \mu \mathbf{U}^T \mathbf{D}^{1/2} \mathbf{R}_{xx} \mathbf{D}^{1/2} \mathbf{U} \mathbf{X}_n(\infty) \mathbf{U} \mathbf{D}^{1/2} \mathbf{D}^{1/2} \mathbf{U} + \mathbf{G}(\infty),
\]
(16)

where \( \mathbf{G}(\infty) = \mu \mathbf{G}(\infty) \mathbf{D}^{1/2} \mathbf{R}_{xx} \mathbf{D}^{1/2} \mathbf{U} + \mathbf{U} \mathbf{G}(\infty) \mathbf{D}^{1/2} \mathbf{R}_{xx} \mathbf{D}^{1/2} \mathbf{U} \).

The diagonal values of (16) reads:
\[
[\mathbf{X}_n(\infty)]_i = [\mathbf{X}_n(\infty)]_i - 2 \mu \xi [\mathbf{X}_n(\infty)]_i + \mu^2
\]
\[
- 2 \xi \mu [\mathbf{X}_n(\infty)]_i + (J_1 + \sigma_{\text{min}}^2) [\mathbf{G}(\infty)]_i + [\mathbf{G}(\infty)]_i,
\]
where \( [\mathbf{G}(\infty)]_i = [\mathbf{G}(\infty)]_i + [\mathbf{G}(\infty)]_i \).

On the other hand, the EMSE is given by
\[
J_1 = \text{tr}(\mathbf{X}_n(\infty) \mathbf{R}_{xx}) = \text{tr}(\mathbf{X}_n(\infty) \mathbf{U}^T \mathbf{D}^{1/2} \mathbf{R}_{xx} \mathbf{D}^{1/2} \mathbf{U}).
\]
(18)

If the transform can approximately diagonalize \( \mathbf{R}_{xx} \), then
\[
\mathbf{R}_{xx} = \mathbf{D}^{1/2} \mathbf{R}_{xx} \mathbf{D}^{1/2} = \mathbf{I},
\]
and \( \mathbf{U} \) are approximately equal to the identity matrix. Hence,
\[
\mathbf{U}^T \mathbf{D}^{1/2} \mathbf{R}_{xx} \mathbf{D}^{1/2} \mathbf{U} = \mathbf{A} - \xi \mathbf{U}^T \mathbf{D}^{1/2} \mathbf{R}_{xx} \mathbf{D}^{1/2} \mathbf{U}.
\]
Moreover, for diagonal dominance \( \mathbf{R}_{xx} \), we further have
\[
J_1 = \text{tr}(\mathbf{X}_n(\infty) \mathbf{A} - \xi \mathbf{U}^T \mathbf{D}^{1/2} \mathbf{R}_{xx} \mathbf{D}^{1/2} \mathbf{U})
\]
\[
= \sum_i [\mathbf{X}_n(\infty)]_i [\tilde{\chi}_i - \xi \mathbf{U}^T \mathbf{D}^{1/2} \mathbf{R}_{xx} \mathbf{D}^{1/2} \mathbf{U}]_i,
\]
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.

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