<table>
<thead>
<tr>
<th><strong>Title</strong></th>
<th>A new regularized transform-domain NLMS adaptive filtering algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Author(s)</strong></td>
<td>Chan, SC; Chu, YJ; Zhang, ZG</td>
</tr>
<tr>
<td><strong>Issued Date</strong></td>
<td>2010</td>
</tr>
<tr>
<td><strong>URL</strong></td>
<td><a href="http://hdl.handle.net/10722/137728">http://hdl.handle.net/10722/137728</a></td>
</tr>
<tr>
<td><strong>Rights</strong></td>
<td>IEEE Asia-Pacific Conference on Circuits and Systems Proceedings. Copyright © IEEE.; This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.; ©2010 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.</td>
</tr>
</tbody>
</table>
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
{sccchan; 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
signal. This helps to reduce the eigenvalue spread of the input
autocorrelation matrix and hence significantly improve the

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
count 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 \( w^* \) 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

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

(1)

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 \( y(n) = w^T(n)x(n) \) as follows

\[
\min_{w} \epsilon_{\text{new}} = E[(d(n) - w^T(n)x(n))^2] + \frac{\xi}{2} \|Dw(n)\|_2^2,
\]

(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

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

(3)

\[
w(n+1) = w(n) + \mu [x(n)e(n) - \xi R_x w(n)],
\]

(4)

where \( R_x = D^T D \).

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

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

(5)

\[
W(n+1) = W(n) + \mu D_t \{X_e(n)e(n) - \xi \bar{R}_x W(n)\},
\]

(6)

where \( W(n) = \text{Cov}(n) = [W_{c_1}(n), W_{c_2}(n), \ldots, W_{c_L}(n)]^T \), \( X_e(n) = \text{C}x(n) = [X_{c_1}(n), X_{c_2}(n), \ldots, X_{c_L}(n)]^T \) are, respectively, the transformed adaptive weight matrix and signal vector. \( C \) is an \( L \times L \) transformation matrix such as the DFT or DCT. \( R_w = CR_x C^T \) and \( D_t = \text{diag}[\varepsilon_1^{-1}(n), \varepsilon_2^{-1}(n), \ldots, \varepsilon_L^{-1}(n)] \) 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 \( R_x \);

(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 \( D_t \) 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 \( W^* = R_x^{1/2} P_{ax} \), where \( P_{ax} = E[d(n)X_e(n)] \) is the ensemble-averaged cross-correlation vector between \( X_e(n) \) and \( d(n) \). \( W^* \) is related to the optimal Wiener solution \( w_{opt} \) as \( w_{opt} = R_x^{1/2} P_{ax} = CW^* \).

A. Mean Convergence Analysis

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

\[
E[W(n+1)] = E[W(n)] + \mu D_t \{P_{ax} \{R_x^{1/2} + \xi \bar{R}_x\} E[W(n)]\}.
\]

(7)

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

\[
(R_x^{1/2} + \xi \bar{R}_x)W_s = P_{ax},
\]

(8)

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):

\[
E[v(n+1)] = (I - \mu D_t (R_x^{1/2} + \xi \bar{R}_x))E[v(n)],
\]

(9)

where \( I \) is the identity matrix. Let \( \tilde{U}A\tilde{U}^T \) be the eigendecomposition of \( \bar{R}_x = D_t^{-1/2} \bar{R}_x^{1/2} D_t^{1/2} \) with \( \bar{R}_x^{1/2} = R_x^{1/2} + \xi \bar{R}_x \). Using (9) and expressing \( v(n) \) in the coordinate, \( V(n) = \tilde{U} D_t^{-1/2} v(n) \), we get the difference equation for the \( i \)-th element of \( E[V(n)] \) as follows

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

(10)

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

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

(11)

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

It can be seen that if the input is so distributed that \( R_x^{1/2} \) has zero eigenvalues, then these eigenmodes can never converge and the solution may be significantly biased. With sufficient regularization, the eigenvalues of \( \bar{R}_x \) 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 of the weight error covariance matrix \( \Xi(n) = E[v(n)v^T(n)] \) as follows

\[
\Xi(n+1) = \Xi(n) + \mu P_{ax} \{P_{ax}^T \{E[v(n)]^2 - \bar{R}_x \Xi(n)\}\}
\]

\[
+ \mu \{E[v(n)]P_{ax}^T - \Xi(n)\bar{R}_x \bar{R}_x^T\} D_t^T
\]

\[
+ \mu^2 D_t \{[(X_e(n)e(n)) - \xi \bar{R}_x W(n)]^2
\]

\[
- \Xi(n) P_{ax} D_t^T\}
\]

(12)

where \( \Xi(n) = \Xi(n) \) 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 \( W^* = R_x^{1/2} P_{ax} \), where \( P_{ax} = E[d(n)X_e(n)] \) is the ensemble-averaged cross-correlation vector between \( X_e(n) \) and \( d(n) \). \( W^* \) is related to the optimal Wiener solution \( w_{opt} \) as \( w_{opt} = R_x^{1/2} P_{ax} = CW^* \).

A. Mean Convergence Analysis

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

\[
E[W(n+1)] = E[W(n)] + \mu D_t \{P_{ax} \{R_x^{1/2} + \xi \bar{R}_x\} E[W(n)]\}.
\]

(7)

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

\[
(R_x^{1/2} + \xi \bar{R}_x)W_s = P_{ax},
\]

(8)

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):

\[
E[v(n+1)] = (I - \mu D_t (R_x^{1/2} + \xi \bar{R}_x))E[v(n)],
\]

(9)

where \( I \) is the identity matrix. Let \( \tilde{U}A\tilde{U}^T \) be the eigendecomposition of \( \bar{R}_x = D_t^{-1/2} \bar{R}_x^{1/2} D_t^{1/2} \) with \( \bar{R}_x^{1/2} = R_x^{1/2} + \xi \bar{R}_x \). Using (9) and expressing \( v(n) \) in the coordinate, \( V(n) = \tilde{U} D_t^{-1/2} v(n) \), we get the difference equation for the \( i \)-th element of \( E[V(n)] \) as follows

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

(10)

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

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

(11)

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

It can be seen that if the input is so distributed that \( R_x^{1/2} \) has zero eigenvalues, then these eigenmodes can never converge and the solution may be significantly biased. With sufficient regularization, the eigenvalues of \( \bar{R}_x \) 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 of the weight error covariance matrix \( \Xi(n) = E[v(n)v^T(n)] \) as follows

\[
\Xi(n+1) = \Xi(n) + \mu P_{ax} \{P_{ax}^T \{E[v(n)]^2 - \bar{R}_x \Xi(n)\}\}
\]

\[
+ \mu \{E[v(n)]P_{ax}^T - \Xi(n)\bar{R}_x \bar{R}_x^T\} D_t^T
\]

\[
+ \mu^2 D_t \{[(X_e(n)e(n)) - \xi \bar{R}_x W(n)]^2
\]

\[
- \Xi(n) P_{ax} D_t^T\}
\]

(12)
which is very close to that of the conventional TDNLMS algorithm \( \mu \leq \frac{1}{3Tr(R_{x|x})} \) without regularization.

If the algorithm converges, \( E[v(\infty)] = 0 \), and the last term in (12) will reduce to \( S(\infty) = \mu D_s [A_1(\infty) - A_2(\infty) - A_3(\infty)] + R_{x|x}G^2 \),

\[ A_1(\infty) = 2R_{x|x}E_2(\infty)R_{x|x} + 2\xi R_{x|x}W_{x}^T W_{x}^r R_{x|x} \]

\[ + 2\xi R_{x|x}W_{x}^T W_{x}^r R_{x|x} \]

\[ + 2\xi R_{x|x}W_{x}^T W_{x}^r R_{x|x} \]

where \( \sigma^2 \) and \( \xi^2 \) are small, we can drop the last term to get

\[ S(\infty) = \mu D_s [2R_{x|x} + \xi R_{x|x}E_2(\infty) + \xi R_{x|x}] \]

\[ - \xi R_{x|x}E_2(\infty)R_{x|x} + R_{x|x}^r \xi E_2(\infty)R_{x|x} \]

\[ + (J_0 + \sigma^2)R_{x|x} + \xi R_{x|x}W_{x}^T W_{x}^r R_{x|x} \]

\[ + \xi R_{x|x}W_{x}^T W_{x}^r R_{x|x} \]

\[ + \xi R_{x|x}W_{x}^T W_{x}^r R_{x|x} \]

Therefore, \( A_1(\infty) = A_1(\infty) - A_1(\infty) + A_1(\infty) \)

\[ = 2R_{x|x}E_2(\infty)E_2(\infty)R_{x|x} + \xi R_{x|x} \]

\[ - \xi R_{x|x}E_2(\infty)R_{x|x} + R_{x|x}^r \xi E_2(\infty)R_{x|x} \]

\[ + (J_0 + \sigma^2)R_{x|x} + \xi R_{x|x}W_{x}^T W_{x}^r R_{x|x} \]

\[ + \xi R_{x|x}W_{x}^T W_{x}^r R_{x|x} \]

\[ + \xi R_{x|x}W_{x}^T W_{x}^r R_{x|x} \]

By expressing \( v(n) \) in the transformed coordinate: \( V(n) = \tilde{U}D_s^{1/2}v(n) \), (15) can be further simplified to

\[ \tilde{E}_s(\infty) = \tilde{E}_s(\infty) - \mu \tilde{U}D_s^{1/2} \tilde{R}_{x|x}D_s^{1/2} \tilde{U} \tilde{E}_s(\infty) \]

\[ - 2\mu \tilde{E}_s(\infty)\tilde{U}D_s^{1/2} \tilde{R}_{x|x}D_s^{1/2} \tilde{U} \tilde{E}_s(\infty) \]

\[ + 2\mu \tilde{U}D_s^{1/2} \tilde{R}_{x|x}D_s^{1/2} \tilde{U} \tilde{E}_s(\infty) \]

\[ + 2\mu \tilde{U}D_s^{1/2} \tilde{R}_{x|x}D_s^{1/2} \tilde{U} \]

\[ + 2\mu \tilde{U}D_s^{1/2} \tilde{R}_{x|x}D_s^{1/2} \tilde{U} \]

\[ + 2\mu \tilde{U}D_s^{1/2} \tilde{R}_{x|x}D_s^{1/2} \tilde{U} \]

\[ + 2\mu \tilde{U}D_s^{1/2} \tilde{R}_{x|x}D_s^{1/2} \tilde{U} \]

By the diagonal values of (16) reads:

\[ \{E_1(\infty)\}_{ij} = (\tilde{\xi}\tilde{E}_s(\infty))_{ij} - 2\mu \tilde{E}_s(\infty))_{ij} + \mu^2 \]

\[ \{E_2(\infty)\}_{ij} = (\tilde{\xi}\tilde{E}_s(\infty))_{ij} + (J_0 + \sigma^2)_{ij} + \mu^2 (\tilde{\xi}\tilde{E}_s(\infty))_{ij} \]

where \( \{E_1(\infty)\}_{ij} = \Gamma_{1,ij} \) and \( \{E_2(\infty)\}_{ij} = \Gamma_{2,ij} \).

On the other hand, the EMSE is given by

\[ J_0 = tr(E_s(\infty))_{x|x} = tr(\tilde{E}_s(\infty)\tilde{U}D_s^{1/2} \tilde{R}_{x|x}D_s^{1/2} \tilde{U}) \]

\[ \tilde{R}_{ss} = D_s^{1/2} \tilde{R}_{x|x} D_s^{1/2} = I \]

where \( \tilde{U} = D_s^{1/2} \tilde{U} \) and \( \tilde{U} = \tilde{D}_s^{1/2} \tilde{U} \).

Moreover, for diagonal dominance \( R_{w} \), we further have

\[ J_0 = tr(\tilde{E}_s(\infty)(\tilde{A} - \tilde{D}_s^{1/2} \tilde{R}_{w} D_s^{1/2})) \]

\[ = \sum_{i} (\tilde{E}_s(\infty))_{ii} (\tilde{A}_{ii} - \tilde{E}_s(\infty))_{ii} \]

where \( R_{w} \) is the i-th diagonal value of \( R_{w} \).

Solving for \( \tilde{E}_s(\infty) \), (17) gives

\[ \tilde{E}_s(\infty))_{ii} = \frac{\mu}{2\tilde{\xi}} ((J_0 + \sigma^2)_{ii} + \tilde{\xi}\tilde{E}_s(\infty))_{ii} \]

Consequently, \( J_0 \) is found to be

\[ J_0 = \frac{\mu}{2\tilde{\xi}} (\tilde{E}_s(\infty))_{ii} + \frac{\mu^2}{4\tilde{\xi}^2} \]

\[ \phi_{\text{RNLMS}} = \frac{\mu}{2\tilde{\xi}} \]

where \( \phi_{\text{RNLMS}} \) is the regularization parameter for R-TDNLMS. 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 \( \epsilon = (1 - \alpha_e) \epsilon (n-1) + \alpha_e \epsilon (n) \) with \( \alpha_e = 0.01 \). 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 \( \tilde{\xi} = 0.1, 0.02 \) 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

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 \( \epsilon = (1 - \alpha_e) \epsilon (n-1) + \alpha_e \epsilon (n) \) with \( \alpha_e = 0.01 \). All simulations are performed using the system identification model and the results are averaged over 200 independent runs.
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