ANALYSIS OF LMS ALGORITHM BEHAVIOR WITH SUBSPACE INPUTS

N. J. Bershad, J. C. M. Bermudez† and J.-Y. Tourneret*⋆

* University of California, Irvine, 1621 Santiago Drive, Newport Beach, CA, 92660, USA
† Federal University of Santa Catarina, Florianopolis, SC Brazil
⋆ IRIT/ENSEEIHT/TéSA, 2 rue C. Camichel, BP 7122, 31071 Toulouse cedex 7, France

bershad@ece.uci.edu,jean-yves.tourneret@enseeiht.fr,j.bermudez@ieee.org

ABSTRACT

This paper studies the behavior of the LMS algorithm for a special system identification problem when partial wavelet transformations restrict the algorithm's input vector to a subspace of the unknown system's input vector space. It is shown that the independence theory is not applicable in this case. A new theoretical model for the weight mean and fluctuation behaviors is developed which incorporates the correlation between successive data vectors (as opposed to using the independence theory model). Comparison of the new model predictions with Monte Carlo simulations shows good-to-excellent agreement, certainly much better than predicted by the independence theory model.

Index Terms— Adaptive filters, adaptive signal processing, adaptive systems.

1. INTRODUCTION

The Least Mean Squares (LMS) is the most popular adaptive algorithm due to its simplicity and robustness [1, 2]. It has been studied for decades, and yet its exact behavior in certain practical situations is still to be determined.

A recent paper [3] presented a novel scheme for identifying sparse impulse responses. Two adaptive filters operate sequentially. The first adaptive filter adapts using a partial Haar transform of the input and yields an estimate of the location of the peak of the sparse response. The second adaptive filter is then centered about this estimate. Both filters are short in comparison to the delay uncertainty of the unknown channel. Hence, two short adaptive filters are used instead of one long filter, resulting in faster overall convergence and reduced computational complexity and storage.

The scheme was analyzed in [3] for two LMS adaptive filters. The analytic model used the so-called *independence assumption* (IA) [1]. For a stationary input vector $\boldsymbol{X}(n) = [x(n), \ldots, x(n-N+1)]^T$, IA assumes that $E[\boldsymbol{X}(n)\boldsymbol{X}^T(m)] = \delta(n-m)\boldsymbol{R}_x$, with $\boldsymbol{R}_x = E[\boldsymbol{X}(n)\boldsymbol{X}^T(n)]$. Monte Carlo simulations of the weight variance in [3] were shown to be in good agreement with the theoretical model for an independent signal model but in significant disagreement with a tapped delay-line (TDL) model.

Previous analyses (not relying on the IA) [4–8] do not consider an important property of the scheme in [3]. The partial Haar transform yields adaptive filter input vectors which lie in a subspace of the input vector space to the unknown system. Consider the adaptive filtering problem depicted in Fig. 1. Using vector notation [1], $\boldsymbol{W}_o$ is the impulse response to be identified. $\boldsymbol{W}(n)$ is the impulse response of the adaptive filter, and $\eta(n) = d(n) - \boldsymbol{W}_o{}^T\boldsymbol{X}(n)$ is the portion of $d(n)$ that cannot be estimated with the adaptive filter. Exact modelling implies that $\{\eta(n)\}$ and $\{x(n)\}$ are statistically independent random processes. Fig. 2 shows the sparse channel echo cancellation problem studied in [3], where $z(n)$ is the partial-Haar-transformed input signal. The portion of Fig. 2 corresponding to the partial Haar adaptive filter system is detailed in Fig. 3. The partial Haar transformation leads to an input vector $\boldsymbol{Z}_p(n) = [z_1(n), \ldots, z_q(n)]^T$ of lower dimension than $\boldsymbol{X}(n)$ $(q < N)$ [3]. Hence, the system in Fig. 3 does not satisfy the conditions for application of the IA. In Fig. 3, $d(n)$ can be modelled as $d(n) = s_1(n) + s_0(n) + \eta(n)$, where $s_1(n)$ can be obtained by linear filtering signals in the subspace spanned by $\boldsymbol{Z}_p(n)$, $s_0(n)$ cannot be cancelled by linearly filtering $\boldsymbol{Z}_p(n)$ (it is provided by the portion of the input subspace not captured in $\boldsymbol{Z}_p(n)$) but is correlated with $x(n)$, and $\eta(n)$ is statistically independent of $x(n)$.

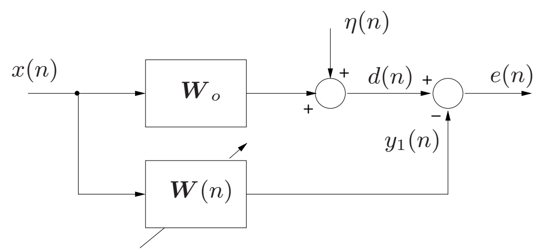

Fig. 1. Standard adaptive filtering.

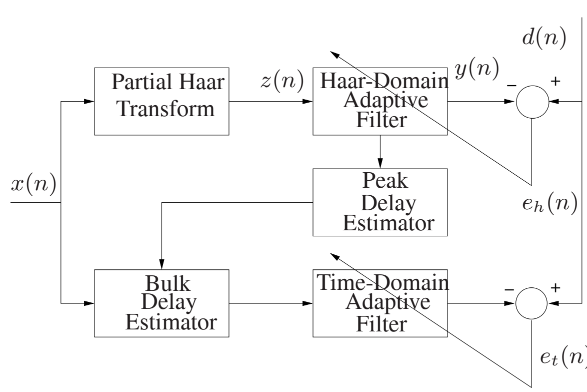

Fig. 2. Partial-Haar dual adaptive filter for sparse channels.

This work was partially supported by CNPq under grant No. 308095/2003-0.

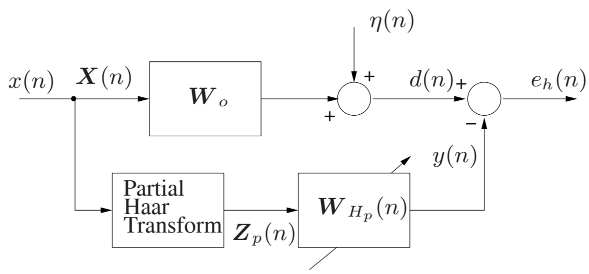

Fig. 3. Adaptive system analyzed

This paper derives a more accurate mathematical model for the LMS weight behavior than that derived in [3] using the IA for the system in Fig. 3 with a tapped delay line input. The residual estimation error is shown to be correlated with $\boldsymbol{Z}_p(n)$. Thus, the required analysis differs from previous analyses in two major ways. The residual estimation error is: 1) correlated in time and 2) statistically dependent on the input signal. The new model yields good-to-excellent agreement with the Monte Carlo simulations for TDL inputs.

2. PROBLEM DESCRIPTION

This paper studies the behavior of the system shown in Fig. 3. The input vector is $\boldsymbol{X}(n) = [x(n), \ldots, x(n-N+1)]^T$. The partial Haar transform is represented by a $q \times N$ matrix $\boldsymbol{H}_{Mp}$,[1] and $q < N$ is the number of adaptive weights in $\boldsymbol{W}_{H_p}(n) = [w_1(n), \ldots, w_q(n)]^T$. The dimension q is chosen according to design considerations discussed in [3]. The input to the adaptive filter is the Partial-Haar transformed vector $\boldsymbol{Z}_p(n) = [z_1(n), \ldots, z_q(n)]^T = \boldsymbol{H}_{Mp}\boldsymbol{X}(n)$. The signal $y(n)$ is an estimate of $d(n)$, which is related to $x(n)$ by

$$d(n) = \boldsymbol{W}_o{}^T \boldsymbol{X}(n) + \eta(n), \tag{1}$$

where $\boldsymbol{W}_o = E[\boldsymbol{X}(n)\boldsymbol{X}^T(n)]^{-1}E[d(n)\boldsymbol{X}(n)]$ is the Wiener solution for the linear estimation of $d(n)$ from the observations in $\boldsymbol{X}(n)$, and $\eta(n)$ is zero-mean, i.i.d. and statistically independent of $x(n)$. $e_h(n)$ is the error for the problem of linearly estimating $d(n)$ from the observations in $\boldsymbol{Z}_p(n)$.

2.1. Properties of the Estimation Error

The partial Haar transformation with $q < N$ leads to under-modelling, since $\boldsymbol{Z}_p(n)$ lies in a q-dimensional subspace spanned by the columns of $\boldsymbol{H}_{Mp}$. It has been shown in [3] that the optimum weight vector in the partial Haar domain is given by $\boldsymbol{W}_{H_po} = (\boldsymbol{H}_{Mp}\boldsymbol{R}_x\boldsymbol{H}_{Mp}^T)^{-1} \times \boldsymbol{H}_{Mp}\boldsymbol{R}_x\boldsymbol{W}_o$, which reduces to $\boldsymbol{W}_{H_po} = \boldsymbol{H}_{Mp}\boldsymbol{W}_o$ for white inputs.

An important consequence of the under-modelling is the nature of the optimum estimation error. Using Fig. 3 and the expression $\boldsymbol{Z}_p(n) = \boldsymbol{H}_{Mp}\boldsymbol{X}(n)$, the estimation error $e_h(n)$ is given by

$$e_h(n) = d(n) - \boldsymbol{W}_{H_p}^T(n)\boldsymbol{H}_{Mp}\boldsymbol{X}(n). \tag{2}$$

Using (1) in (2) and evaluating the optimum estimation error $e_o(n)$ (corresponding to $e_h(n)$ for $\boldsymbol{W}_{H_p} = \boldsymbol{W}_{H_po}$) yields

$$e_o(n) = \left(\boldsymbol{W}_o^T - \boldsymbol{W}_{H_po}^T\boldsymbol{H}_{Mp}\right)\boldsymbol{X}(n) + \eta(n). \tag{3}$$

[1] According to the notation used in [3], the subscript M relates to the dimension of the full Haar transform matrix of which $\boldsymbol{H}_{Mp}$ is part. The subscript p stands for *partial*.

From (3), the autocorrelation of $e_o(n)$ can be easily evaluated as

$$\begin{aligned} E\{e_o(n)e_o(m)\} &= \left(\boldsymbol{W}_o^T - \boldsymbol{W}_{H_po}^T\boldsymbol{H}_{Mp}\right) \\ &\times E\left\{\boldsymbol{X}(n)\boldsymbol{X}^T(m)\right\}\left(\boldsymbol{W}_o - \boldsymbol{H}_{Mp}^T\boldsymbol{W}_{H_po}\right). \end{aligned} \tag{4}$$

Since $E\{\boldsymbol{X}(n)\boldsymbol{X}^T(m)\} \neq \boldsymbol{0}$ even for white $x(n)$, the condition for $e_o(n)$ to be uncorrelated is

$$\boldsymbol{W}_o = \boldsymbol{H}_{Mp}^T\boldsymbol{W}_{H_po}. \tag{5}$$

Eq. (5) cannot be satisfied unless $\boldsymbol{W}_o$ is in the row space of $\boldsymbol{H}_{Mp}$ (a very special case). Thus, $e_o(n)$ is correlated in time. The use of the IA in (4) would lead to the opposite conclusion.

Straightforward calculation also shows that

$$\begin{aligned} E\{e_o(m)\boldsymbol{Z}_p(n)\} &= \boldsymbol{H}_{Mp}E\left\{\boldsymbol{X}(n)\boldsymbol{X}^T(m)\right\} \\ &\times \left(\boldsymbol{W}_o - \boldsymbol{H}_{Mp}^T\boldsymbol{W}_{H_po}\right). \end{aligned} \tag{6}$$

For $e_o(m)$ to be uncorrelated with $\boldsymbol{Z}_p(n)$, it is required that

$$\begin{aligned} \boldsymbol{W}_{H_po} &= \left(\boldsymbol{H}_{Mp}E\left\{\boldsymbol{X}(n)\boldsymbol{X}^T(m)\right\}\boldsymbol{H}_{Mp}^T\right)^{-1} \\ &\times \boldsymbol{H}_{Mp}E\left\{\boldsymbol{X}(n)\boldsymbol{X}^T(m)\right\}\boldsymbol{W}_o, \end{aligned} \tag{7}$$

which is true if (5) holds or if $m = n$, since $E\left\{\boldsymbol{X}(n)\boldsymbol{X}^T(n)\right\} = \boldsymbol{R}_x$ ($E\{e_o(n)\boldsymbol{Z}_p(n)\} = 0$ by the orthogonality principle). Neglecting the unlikely condition (5), this result shows that the residual estimation error is correlated with $z(n)$. The use of the IA would lead to erroneous results.

The results (4) and (6) show why the IA-based model derived in [3] leads to poor results for a tapped delay-line input model. These results also show that the analysis must consider the statistical correlation between the estimation error and $\boldsymbol{Z}_p(n)$. Such analysis is not available in the literature and requires a new approach.

3. FORMULATION OF THE ANALYSIS PROBLEM

The following analysis assumes that:

- A1: $x(n)$ is stationary, i.i.d., zero-mean and Gaussian. Thus, $\boldsymbol{R}_x = E[\boldsymbol{X}(n)\boldsymbol{X}^T(n)] = \sigma_x^2\boldsymbol{I}_N$, $\boldsymbol{I}_N$ the $N \times N$ identity matrix.
- A2: $x(n)$ and $d(m)$ for all n and m are zero-mean jointly stationary Gaussian sequences.

Vectors $\boldsymbol{X}(n)$ and $\boldsymbol{X}(m)$ are considered statistically dependent for $|n - m| < N$. Thus, the IA cannot be used. Moreover, the existing analysis techniques which avoid IA cannot be used because the residual estimation cannot be assumed i.i.d..

The LMS weight recursion for the Haar-domain adaptive filter in Fig. 3 is given by [3, Eq. (14)]

$$\begin{aligned} \boldsymbol{W}_{H_p}(n+1) &= \boldsymbol{W}_{H_p}(n) + \mu e_h(n)\boldsymbol{Z}_p(n) \\ &= \left[\boldsymbol{I}_q - \mu\boldsymbol{Z}_p(n)\boldsymbol{Z}_p^T(n)\right]\boldsymbol{W}_{H_p}(n) + \mu d(n)\boldsymbol{Z}_p(n) \end{aligned} \tag{8}$$

Subtracting $\boldsymbol{W}_{H_po} = \boldsymbol{H}_{Mp}\boldsymbol{W}_o$ from both sides of (8) and defining $\boldsymbol{V}_p(n) = \boldsymbol{W}_{H_p}(n) - \boldsymbol{W}_{H_po}$ yields a recursion for $\boldsymbol{V}_p(n)$,

$$\begin{aligned} \boldsymbol{V}_p(n+1) &= \left[\boldsymbol{I}_q - \mu\boldsymbol{Z}_p(n)\boldsymbol{Z}_p^T(n)\right]\boldsymbol{V}_p(n) \\ &+ \mu\left[d(n) - \boldsymbol{Z}_p^T(n)\boldsymbol{W}_{H_po}\right]\boldsymbol{Z}_p(n). \end{aligned} \tag{9}$$

The last term in brackets is the Wiener error $e_o(n) = d(n) - \boldsymbol{Z}_p^T(n)\boldsymbol{W}_{H_po}$. The expected value of the last term in (9) is zero from the orthogonality principle ($E[e_o(n)\boldsymbol{Z}_p(n)] = \boldsymbol{0}$).

The stochastic algorithm behavior is usually determined using (9) to derive recursions for the mean $E[\boldsymbol{V}_p(n)]$ and the covariance matrix $E[\boldsymbol{V}_p(n)\boldsymbol{V}_p^T(n)]$. However, recursion for $E[\boldsymbol{V}_p(n)\boldsymbol{V}_p^T(n)]$ occurs only when the assumption $E[\boldsymbol{X}(n)\boldsymbol{X}^T(m)] = \boldsymbol{0}$ for $n \neq m$ is used. Otherwise, the recursion involves $E[\boldsymbol{V}_p(n)\boldsymbol{V}_p^T(m)]$ for $n \neq m$ as well. A recursion for $E[\boldsymbol{V}_p(n)\boldsymbol{V}_p^T(m)]$ will include expectations involving $\boldsymbol{X}(m)$ and $\boldsymbol{V}_p(n)$ unless one invokes the IA again.

Our approach to this analysis requires an approximation to (9) that has a closed form solution so as to avoid the problems described above with the recursive solution. To this end, the term $\boldsymbol{Z}_p(n)\boldsymbol{Z}_p^T(n)$ can be written as a mean plus a fluctuating part,

$$\boldsymbol{Z}_p(n)\boldsymbol{Z}_p^T(n) = E\left[\boldsymbol{Z}_p(n)\boldsymbol{Z}_p^T(n)\right] + \boldsymbol{\Psi}(n) = \sigma_x^2\boldsymbol{I}_q + \boldsymbol{\Psi}(n). \tag{10}$$

Inserting (10) in (9) yields

$$\boldsymbol{V}_p(n+1) = \left(1-\mu\sigma_x^2\right)\boldsymbol{V}_p(n) + \mu e_o(n)\boldsymbol{Z}_p(n) - \mu\boldsymbol{\Psi}(n)\boldsymbol{V}_p(n). \tag{11}$$

Eq. (11) can be viewed as a deterministic recursion for $\boldsymbol{V}_p(n)$ driven by random inputs $e_o(n)\boldsymbol{Z}_p(n)$ and $\boldsymbol{\Psi}(n)\boldsymbol{V}_p(n)$. During transient, fluctuations of $\boldsymbol{V}_p(n)$ are small compared with $E[\boldsymbol{V}_p(n)]$, and $\boldsymbol{\Psi}(n)\boldsymbol{V}_p(n)$ can be approximated by $\boldsymbol{\Psi}(n)E[\boldsymbol{V}_p(n)]$. Close to convergence $E[\boldsymbol{V}_p(n)]$ tends to zero and the input to the recursion can be approximated by $e_o(n)\boldsymbol{Z}_p(n)$ if the fluctuations in $\boldsymbol{V}_p(n)$ are sufficiently small so that

$$e_o(n)\boldsymbol{Z}_p(n) \gg \boldsymbol{\Psi}(n)\left\{\boldsymbol{V}_p(n) - E[\boldsymbol{V}_p(n)]\right\}, \tag{12}$$

which is more valid for slower adaptation rates.

Assuming (12), (11) can be approximated by the recursion

$$\begin{aligned}\boldsymbol{V}_p(n+1) &\simeq \left(1-\mu\sigma_x^2\right)\boldsymbol{V}_p(n) \\ &\quad + \mu e_o(n)\boldsymbol{Z}_p(n) - \mu\boldsymbol{\Psi}(n)E[\boldsymbol{V}_p(n)].\end{aligned} \tag{13}$$

Eq. (13) can be used to determine the effects of $E\left[\boldsymbol{Z}_p(n)\boldsymbol{Z}_p^T(m)\right] \neq \boldsymbol{0}$ for $n \neq m$ on the behavior of the weight error vector. Viewing the last two terms on the r.h.s. of (13) as forcing terms, (13) has an explicit closed form solution

$$\begin{aligned}\boldsymbol{V}_p(n) &\simeq \left(1-\mu\sigma_x^2\right)^n\boldsymbol{V}_p(0) + \mu\sum_{m=0}^{n-1}\left(1-\mu\sigma_x^2\right)^{n-m-1} \\ &\times\left\{e_o(m)\boldsymbol{Z}_p(m) - \boldsymbol{\Psi}(m)E[\boldsymbol{V}_p(m)]\right\}.\end{aligned} \tag{14}$$

Eq. (14) represents a deterministic system with random inputs and can be used to determine the response to correlated inputs vectors $\boldsymbol{Z}_p(m), \ldots, \boldsymbol{Z}_p(m-k)$, $k = 2, \ldots, n-1$.

4. STOCHASTIC BEHAVIOR ANALYSIS

4.1. Mean Weight Behavior

Averaging (14) and using the orthogonality principle,

$$E\left[\boldsymbol{V}_p(n)\right] \simeq \left(1-\mu\sigma_x^2\right)^n\boldsymbol{V}_p(0), \tag{15}$$

since $E[\boldsymbol{\Psi}(n)] = \boldsymbol{0}$. This result coincides with that obtained from the IA model. Such coincidence is expected since the effects of input vector cross-correlation on the mean weight analysis are ignored in the approximation (13). Fortunately, this is not the case for the weight fluctuation behavior, as will become clear in the next section.

4.2. Weight Fluctuation Behavior

The covariance matrix $\boldsymbol{Q}(n)$ of $\boldsymbol{V}_p(n)$ is

$$\begin{aligned}\boldsymbol{Q}(n) &= E\left\{\left[\boldsymbol{V}_p(n) - E[\boldsymbol{V}_p(n)]\right]\left[\boldsymbol{V}_p(n) - E[\boldsymbol{V}_p(n)]\right]^T\right\} \\ &= \mu^2\sum_{m=0}^{n-1}\sum_{r=0}^{n-1}\left(1-\mu\sigma_x^2\right)^{n-m-1}\left(1-\mu\sigma_x^2\right)^{n-r-1} \\ &\quad\times\Bigg[E\Big\{\boldsymbol{Z}_p(m)e_o(m)e_o(r)\boldsymbol{Z}_p^T(r) \\ &\quad+\boldsymbol{\Psi}(m)E[\boldsymbol{V}_p(m)]E[\boldsymbol{V}_p^T(r)]\boldsymbol{\Psi}^T(r)\Big\} \\ &\quad-E\Big\{e_o(m)\boldsymbol{Z}_p(m)E[\boldsymbol{V}_p^T(r)]\boldsymbol{\Psi}^T(r) \\ &\quad+\boldsymbol{\Psi}(m)E[\boldsymbol{V}_p(m)]\boldsymbol{Z}_p^T(r)e_o(r)\Big\}\Bigg].\end{aligned} \tag{16}$$

The expected values in (16) is evaluated in [9], and leads to

$$\begin{aligned}\boldsymbol{Q}(n) &= \left(\frac{1-a}{1+a}\right)\boldsymbol{H}_{Mp}\Bigg\{\boldsymbol{G}_0(1-a^{2n}) \\ &+\sum_{\nu=1}^{n-1}\Big(\boldsymbol{G}_\nu + \boldsymbol{G}_{-\nu}\Big)\left(a^\nu - a^{2n}a^{-\nu}\right)\Bigg\}\boldsymbol{H}_{Mp}^T + (1-a)^2a^{2n-2} \\ &\times\boldsymbol{H}_{Mp}\left[n\boldsymbol{K}_0 + \sum_{\nu=1}^{n-1}(n-\nu)(\boldsymbol{K}_\nu + \boldsymbol{K}_{-\nu})\right]\boldsymbol{H}_{Mp}^T \\ &+(1-a)a^{n-1}\left\{\sum_{\nu=1}^{n-1}\boldsymbol{T}_\nu\left(1-a^{n-\nu}\right) + \sum_{\nu=1}^{n-1}\boldsymbol{T}_{-\nu}\left(a^\nu - a^n\right)\right\},\end{aligned} \tag{17}$$

where

$$a = (1-\mu\sigma_x^2) \tag{18a}$$

$$\boldsymbol{G}_\nu = \boldsymbol{W}_o^T\boldsymbol{Z}_2\boldsymbol{F}_\nu\boldsymbol{Z}_2\boldsymbol{W}_o\boldsymbol{F}_\nu + \boldsymbol{F}_\nu\boldsymbol{Z}_2\boldsymbol{W}_o\boldsymbol{W}_o^T\boldsymbol{Z}_2\boldsymbol{F}_\nu \tag{18b}$$

$$\begin{aligned}\boldsymbol{K}_\nu =& \boldsymbol{F}_\nu\boldsymbol{H}_{Mp}^T\boldsymbol{V}_p(0)\boldsymbol{V}_p^T(0)\boldsymbol{H}_{Mp}\boldsymbol{F}_\nu \\ &+\boldsymbol{V}_p^T(0)\boldsymbol{H}_{Mp}\boldsymbol{F}_{-\nu}\boldsymbol{H}_{Mp}^T\boldsymbol{V}_p(0)\boldsymbol{F}_\nu.\end{aligned} \tag{18c}$$

$$\begin{aligned}\boldsymbol{T}_\nu =& \boldsymbol{H}_{Mp}\boldsymbol{F}_{m-r}\Big[\boldsymbol{V}_p^T(0)\boldsymbol{H}_{Mp}\boldsymbol{F}_{r-m}\boldsymbol{Z}_2\boldsymbol{W}_o\boldsymbol{I}_N \\ &+\boldsymbol{H}_{Mp}^T\boldsymbol{V}_p(0)\boldsymbol{W}_o^T\boldsymbol{Z}_2\boldsymbol{F}_{m-r}\Big]\boldsymbol{H}_{Mp}^T\end{aligned} \tag{18d}$$

$$\begin{aligned}\boldsymbol{T}_{-\nu} =& \boldsymbol{H}_{Mp}\Big[\boldsymbol{I}_N\boldsymbol{W}_o^T\boldsymbol{Z}_2\boldsymbol{F}_{m-r}\boldsymbol{H}_{Mp}^T\boldsymbol{V}_p(0) \\ &+\boldsymbol{F}_{r-m}\boldsymbol{Z}_2\boldsymbol{W}_o\boldsymbol{V}_p^T(0)\boldsymbol{H}_{Mp}\Big]\boldsymbol{F}_{r-m}\boldsymbol{H}_{Mp}^T.\end{aligned} \tag{18e}$$

5. SIMULATION RESULTS

Consider the symmetric exponential channel impulse response $\boldsymbol{W}_o = [a^r, a^{r-1}, \ldots, a, 1, a, \ldots, a^{r-1}, a^r]^T$ for $r = 32$ and $a = 0.5$, located in a span of 1024 samples, leading to a sparse channel response. This response was used in [3], and is again used here for comparison. The optimum responses $\boldsymbol{W}_{H_po}$ for $N = 256$, 128 and 64 where obtained from the dot product of $\boldsymbol{W}_o$ and the rows of the associated $\boldsymbol{H}_{Mp}$ (inserting enough zeros so that the dot product is defined). The channel bulk delay was varied from zero to eight taps. The variable bulk delay represents the random delay of the channel with respect to the 1024-tap time span.

Figs. 4–6 show the tap weight variance over time, estimated from Monte Carlo simulations by computing tr$[\boldsymbol{Q}(n)]$ for $q = 256$, 128 and 64 and for different values of $\max\{\boldsymbol{W}_{H_p o}\}$, the maximum value of $\boldsymbol{W}_{H_p o}$. The step sizes used were given by $\mu = 0.1/(q+2)$ and $\sigma_x^2 = 1$ in all cases. $\boldsymbol{W}_{H_p}(n)$ was initialized at $\boldsymbol{W}_{H_p}(0) = \mathbf{0}$. The theoretical curves were obtained from (17). For comparison, the figures also present the plots corresponding to a sequence of statistically independent input vectors $\boldsymbol{X}(n)$. Curves identified as "TDL" correspond to theoretical results obtained using (17) and Monte Carlo simulations for a TDL input model. Curves identified as "Independent" correspond to the model in [3] and simulations for statistically independent input vectors. Note that there is good-to-excellent agreement between the theory and simulations, especially when one compares these results with those in [3] for the IA model.

6. CONCLUSIONS

This paper has developed a new theoretical model for the behavior of the LMS adaptive algorithm applied to an under-determined system identification problem with a TDL input signal model. The new theory is in good-to-excellent agreement with Monte Carlo simulations. This was not the case for the theoretical model based on the independence theory assumption. The new model can be used to better design the scheme proposed in [3] for estimating the location of the peak of an unknown sparse impulse response. The new approach can also be used for studying other systems in which the algorithm operates on a subspace of the input signal space, when the independence theory model is no longer valid.

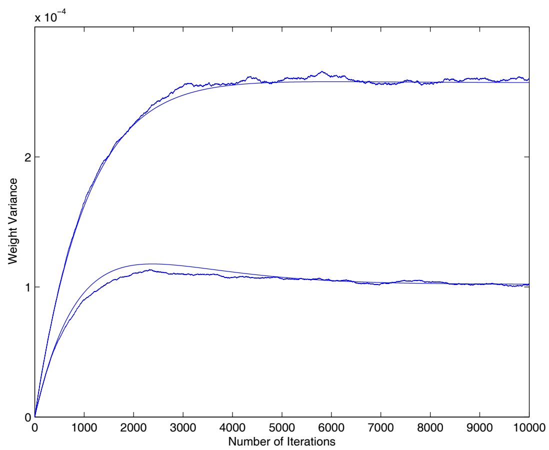

Fig. 4. tr$[\boldsymbol{Q}(n)]$ for $q = 256$, $\max\{w_{H_p o}\} = 0.5625$. Smooth plots: theory. Jagged plots: 100 MC Simulations. Upper plots: Independent (theory from [3]). Lower plots: TDL (theory from (17)).

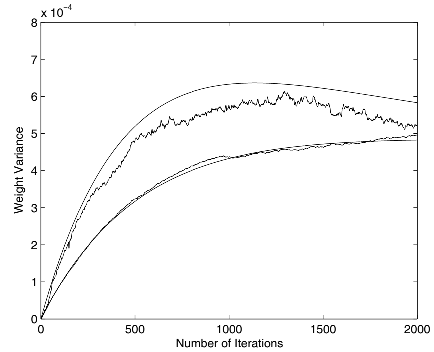

Fig. 5. tr$[\boldsymbol{Q}(n)]$ for $q = 128$, $\max\{w_{H_p o}\} = 0.6215$. Smooth plots: theory. Jagged plots: 100 MC Simulations. Upper plots: TDL (theory from (17)). Lower plots: Independent (theory from [3]).

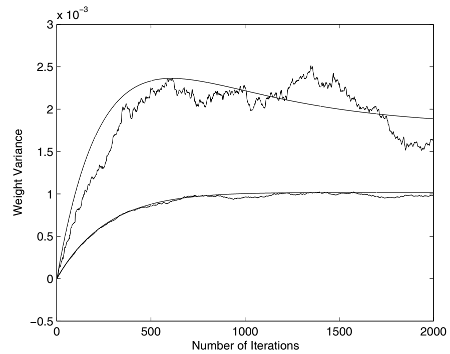

Fig. 6. tr$[\boldsymbol{Q}(n)]$ for $q = 64$, $\max\{w_{H_p o}\} = 0.6172$. Smooth plots: theory. Jagged plots: 100 MC Simulations. Upper plots: TDL (theory from (17)). Lower plots: Independent (theory from [3]).

7. REFERENCES

[1] S. Haykin, *Adaptive Filter Theory*, Prentice-Hall, New Jersey, 4th edition, 2002.

[2] A. H. Sayed, *Fundamentals of Adaptive Filtering*, Wiley-Interscience, New York, 2003.

[3] N. J. Bershad and A. Bist, "Fast coupled adaptation for sparse impulse responses using a partial haar transform," *IEEE Trans. on Signal Processing*, vol. 53, no. 3, pp. 966–976, March 2005.

[4] J. Chao, S. Kawabe, and S. Tsujii, "Convergence analysis of gradient adaptive algorithms for arbitrary inputs without the independence assumption," in *Proc. of the IEEE International Conference on Systems Engineering*, Kobe, Japan, Sept. 1992, pp. 556–559, IEEE.

[5] S. C. Douglas and T. H. Meng, "Exact expectation analysis of the LMS adaptive filter without the independence assumption," in *Proc. of ICASSP*, San Francisco, CA, Mar. 1992, pp. 61–64.

[6] S. C. Douglas and W. Pan, "Exact expectation analysis of the LMS adaptive filter," *IEEE Trans. on Signal Processing*, vol. 43, no. 12, pp. 2863–2871, December 1995.

[7] H. J. Butterweck, "A steady-state analysis of the LMS adaptive algorithm without use of the independence assumption," in *Proc. of ICASSP*, Detroit, MI, May 1995, vol. 2, pp. 1404–1407.

[8] M. Rupp and H. J. Butterweck, "Overcoming the independence assumption in LMS filtering," in *Proc. of the 37th Asilomar Conference*, Monterey, CA, Nov. 2003, pp. 607–611.

[9] N. J. Bershad, J. C. M. Bermudez, and J. -Y. Tourneret, "Stochastic analysis of the LMS algorithm for system identification with subspace inputs," *submitted to the IEEE Trans. on Signal Processing*, 2006.