



HAL
open science

Stochastic Analysis of the LMS Algorithm for System Identification with Subspace Inputs

Neil J. Bershad, José Carlos Moreira Bermudez, Jean-Yves Tournet

► **To cite this version:**

Neil J. Bershad, José Carlos Moreira Bermudez, Jean-Yves Tournet. Stochastic Analysis of the LMS Algorithm for System Identification with Subspace Inputs. *IEEE Transactions on Signal Processing*, 2008, 5 (3), pp.1018-1027. 10.1109/TSP.2007.908967 . hal-03579602

HAL Id: hal-03579602

<https://hal.science/hal-03579602>

Submitted on 18 Feb 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Stochastic Analysis of the LMS Algorithm for System Identification With Subspace Inputs

Neil J. Bershad, *Fellow, IEEE*, José Carlos M. Bermudez, *Senior Member, IEEE*, and Jean-Yves Tournet, *Member, IEEE*

Abstract—This paper studies the behavior of the low-rank least mean squares (LMS) adaptive algorithm for the general case in which the input transformation may not capture the exact input subspace. It is shown that the Independence Theory and the independent additive noise model are not applicable to 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 the Independence Theory model). The new theory is applied to a network echo cancellation scheme which uses partial-Haar input vector transformations. Comparison of the new model predictions with Monte Carlo simulations shows good-to-excellent agreement, certainly much better than predicted by the Independence Theory based model available in the literature.

Index Terms—Adaptive filters, least mean square methods, sparse impulse response, system identification.

I. INTRODUCTION

THE least mean squares (LMS) algorithm 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 the impulse response of a sparse channel. An important practical application for this scheme is in network echo cancellation (NEC). The advent of voice-over internet protocol (VoIP) [4] has revived interest in the NEC problem. Communications networks must incorporate hundreds of echo cancellers for VoIP. This usage creates almost a mandatory need for NEC solutions that use very limited processing power. Adaptive NEC systems must identify a relatively small number of active samples in a long impulse response [5]. The scheme proposed in [3] addressed the sparse response identification through a very fast initial estimation of the peak of the channel's impulse response. It consists of two adaptive filters operating 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 impulse 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. This way, two short adaptive filters are used instead of one long filter, resulting in faster overall convergence and reduced computational complexity and storage. The choice of the Haar transformation is crucial for this application. The Haar transformation combines a very simple wavelet implementation with an excellent time localization performance. The latter property is important for an accurate peak location estimation.

The scheme was analyzed in detail for a structure using the LMS algorithm in both adaptive filters. However, it can be implemented using any combination of adaptive algorithms. The analysis consisted of two major parts: 1) mean and fluctuation behavior of the weights for both LMS algorithms for independent and identically distributed (i.i.d.) Gaussian input data; and 2) an approximate analysis of the mean and variance of a peak delay estimator scheme. The analytic model in 1) used the so-called *independence assumption* (IA) [2], [6]. A fundamental assumption in IA is that the sequence of input vectors to the algorithm is i.i.d. [6]. This assumption is clearly incorrect. However, its use in several analyses has led to important and useful conclusions about the algorithm's behavior. 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 for a tapped delay-line (TDL) filtering structure. Such results indicate the need for improvement in the statistical analysis for the important case of TDL filtering, when lagged input vectors are not independent.

The limitations of the IA have been studied by several authors. Studies of the LMS properties without employing IA started about two decades ago with convergence analyses [6, and references therein], [7]. More recently, analytical models have been derived for the behavior of the LMS algorithm without relying on the IA. For instance, [8] derived conditions for convergence of first and second moments of the adaptive weights. A recursive algorithmic procedure was also proposed in [9]–[11] for the determination of the LMS algorithm behavior. An iterative solution has also been proposed in [12] for determining the steady-state algorithm behavior in which the first and second moments of the weight error vector are determined as a sum of partial functions [1]. This procedure has been later extended to the analysis of the transient adaptive behavior [13]. All these analyses assume that the residual estimation error is i.i.d.

The above and other related analyses do not take into account an important property of the scheme proposed in [3]. The par-

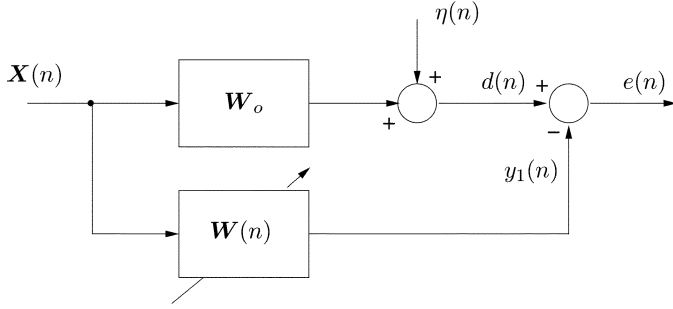


Fig. 1. Standard adaptive filtering. $\mathbf{X}(n) = [x(n), \dots, x(n - N + 1)]^T$.

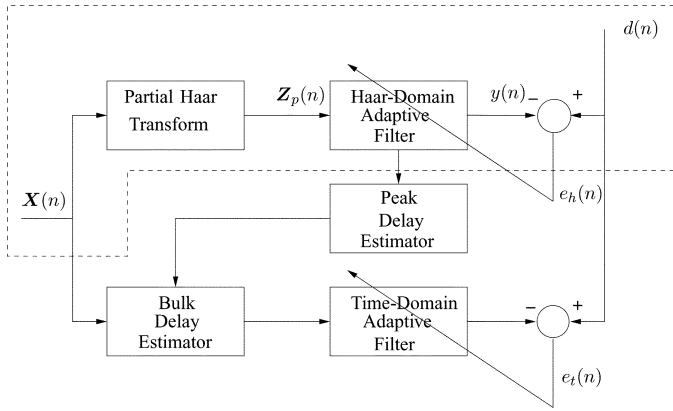


Fig. 2. Partial-Haar dual adaptive filter for sparse channels. $\mathbf{X}(n) = [x(n), \dots, x(n - N + 1)]^T$. $\mathbf{Z}_p(n) = [z_1(n), \dots, z_q(n)]^T$.

tial Haar transform yields an input to the adaptive filter which lies in a vector subspace of the unknown system input vector space.¹ Consider the model depicted in Fig. 1. \mathbf{W}_o and $\mathbf{W}(n)$ are vectors with the impulse responses of FIR filters. This model can be used as a standard representation of the adaptive filtering problem [6]. In this diagram, \mathbf{W}_o is the impulse response to be identified. The lower branch $\mathbf{W}(n)$ represents the impulse response of the adaptive filter. $\eta(n)$ is a disturbance term that is statistically independent of the input $x(n)$ [14]. Exact modeling implies that the optimum estimation error² $e_o(n)$ is $\eta(n)$, and, thus, statistically independent of $x(n)$ [6].

Fig. 2 shows the sparse channel echo cancellation problem studied in [3]. $\mathbf{Z}_p(n)$ is the Partial-Haar-transformed input signal vector. The portion of Fig. 2 within the dashed line is detailed in Fig. 3. The partial Haar transformation results in an input vector to the Haar domain adaptive filter, $\mathbf{Z}_p(n)$, that is of lower dimension than the input vector $\mathbf{X}(n)$ [3]. This leads to an undermodeled system identification problem [14], [15] where the residual estimation error is correlated with the input signal. Hence, the system in Fig. 3 does not satisfy either the condition for application of the IA or the statistical independence of the optimum estimation error and the input signal.

The problem described in Fig. 3 can be interpreted as a low rank adaptive filtering problem [16]–[22] since the adaptive

¹To simplify the notation, the words space and subspace will be used from now on with the meanings of vector space and vector subspace, respectively.

² $e_o(n)$ is the estimation error when $\mathbf{W}(n)$ equals the weight vector that minimizes $E\{e^2(n)\}$.

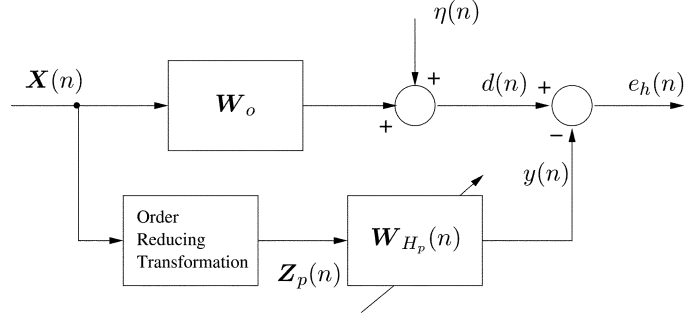


Fig. 3. Adaptive system analyzed, corresponding to the highlighted portion (within the dashed line) in Fig. 2. $\mathbf{W}_{H_p}(n)$ is the coefficient vector of the partial Haar domain adaptive filter.

filter operates on a subspace of the input vector space. Low rank adaptive filtering deals with estimation problems in which the desired signal $d(n)$ in Fig. 3 is composed of a signal derived from a subspace of the input vector space and an additive noise that is statistically independent of the input signal. Thus, the order of the adaptive filter can be reduced to the minimum necessary subspace dimension for efficiency. The reduced order basis is usually found from an analysis of the input data. If the optimal basis can be determined, the reduced order adaptive filter leads to the same optimal performance as the full order adaptive filter but with the least complexity. In general, a close-to-optimal basis is obtained and part of the residual estimation error power is correlated with the input signal [17], [18].

Much of the low rank adaptive filtering literature addresses the problem of searching the optimal subspace. These analyses focus on the least squares (LS) formulation. The LS formulation results in a simple structure based on the properties of orthogonal transformations and eigenvalue decompositions. No analysis is generally available for the transient behavior of the low rank LMS algorithm. [19] presents a transient analysis of low rank transform domain adaptive filtering, assuming the knowledge of the exact subspace dimension. In this case, most of the classical results available for the LMS algorithm are readily applicable [19].

The motivation for the system studied in this paper is different from most low rank adaptive filtering studies. Here, the transformation basis is chosen to enable a good estimate of the time location of the impulse response's peak with the least possible complexity. An accurate identification of the unknown response is not required. The accurate estimate will be performed later by a second time-domain adaptive filter centered about the peak's estimate. Of great concern also is convergence speed. Nearly all of the time, devoted to the unknown response estimation, is used by the second time domain adaptive filter. Thus, these are the reasons for choosing a partial (low order) Haar (time localization) transformation – few adaptive weights and moderate accuracy for the peak estimate. The problem studied here in a broader sense corresponds to a low rank finite impulse response (FIR) LMS adaptive filtering problem with a suboptimal input subspace representation. Thus, the analysis results presented here are valid for any order reducing transformation that is a projection of the input data onto an orthonormal basis.

A mathematical model is derived for the first and second order moments of the LMS weight vector. The resulting model is more accurate than the model in [3] for the tapped delay-line structure which used an i.i.d. input vector sequence assumption. The new model can also be used to study the transient behavior of low rank adaptive filters employing order reducing orthonormal input transformations and the LMS algorithm. The new analysis differs from previous analyses in two major ways: 1) The residual estimation error is correlated in time and 2) the residual estimation error is not statistically independent of the input signal. The new model yields good-to-excellent agreement with Monte Carlo simulations of the mean and fluctuation behaviors of $\mathbf{W}_{H_p}(n)$ in Fig. 3.

For the structure in Fig. 3 and a tapped delay line filter, this paper shows that:

- a) the optimum estimation error is correlated in time and correlated with the input regressor of the Haar-domain adaptive filter; and
- b) the adaptive filtering problem corresponds to a linear combiner with time-correlated orthogonal input signals.

A new formulation is introduced for the statistical analysis of the LMS weight vector behavior operating on a subspace of the input signal space. This problem cannot be studied using the i.i.d. input vector sequence assumption. The new formulation takes into account the correlation between the estimation error and the input vector.

This paper is organized as follows. Section II describes the problem studied and discusses important properties of both the transformed input vector and the estimation error. These properties justify the need for a new analysis approach. Section III formulates the analysis problem, presents the statistical assumptions used and introduces a mathematical model for the adaptive weight updating. Section IV computes the weight mean and fluctuation behavior using this solution and avoids the use of the independence assumption. Section V presents Monte Carlo simulation results which support the theoretical approximations made in the analysis.

II. PROBLEM DESCRIPTION

This paper studies the behavior of the system in Fig. 3. The input vector is $\mathbf{X}(n) = [x(n), x(n-1), \dots, x(n-N+1)]^T$. The order reducing transform is represented by a $q \times N$ matrix \mathbf{H}_{M_p} ,³ and $q < N$ is the number of adaptive weights in the transformed domain adaptive filter. The dimension q is chosen according to design considerations discussed in [3]. The adaptive weights are elements of the $q \times 1$ vector $\mathbf{W}_{H_p}(n) = [w_1(n), \dots, w_q(n)]^T$. The input to the transformed domain adaptive filter $\mathbf{W}_{H_p}(n)$ is the transformed $q \times 1$ input vector $\mathbf{Z}_p(n) = [z_1(n), \dots, z_q(n)]^T = \mathbf{H}_{M_p}\mathbf{X}(n)$. The transformed domain adaptive filter attempts to estimate the desired signal $d(n)$ which is assumed related to $x(n)$ by

$$d(n) = \mathbf{W}_o^T \mathbf{X}(n) + \eta(n) \quad (1)$$

³In keeping with the notation used in [3], the subscript M relates to the dimension of the full Haar transform matrix of which \mathbf{H}_{M_p} is part. The subscript p stands for *partial*.

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

Given the above conditions, two important properties of adaptive filtering in the transformed domain must be clearly understood. These properties will be derived here for the particular case of the partial-Haar transformation used in [3]. The conclusions are valid, however, for any order reducing transformation that projects the input vector onto an orthonormal basis.

A. The Nature of $\mathbf{Z}_p(n)$

A full Haar transform with $M = 3$ (3 scales in the wavelet transform) is implemented by a matrix \mathbf{H}_M given by [3]

$$\mathbf{H}_3 = 2^{-3/2} \times \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ \sqrt{2} & \sqrt{2} & -\sqrt{2} & -\sqrt{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sqrt{2} & \sqrt{2} & -\sqrt{2} & -\sqrt{2} \\ 2 & -2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & -2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & -2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 2 & -2 \end{bmatrix}. \quad (2)$$

If the second scale is used for the partial Haar transformation, the 2×8 partial matrix becomes:

$$\mathbf{H}_{3_p} = 2^{-3/2} \times \begin{bmatrix} \sqrt{2} & \sqrt{2} & -\sqrt{2} & -\sqrt{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sqrt{2} & \sqrt{2} & -\sqrt{2} & -\sqrt{2} \end{bmatrix}. \quad (3)$$

In general, for a given value on $N = 2^M$ and considering the choice of scale s ($s = 1, \dots, \log_2 N$) for the partial transformation, the number of adaptive coefficients is $q = N/2^s$. The number of nonzero elements in each row of \mathbf{H}_{M_p} will be 2^s . For instance, if $N = 1024$ ($M = 10$), using the second scale ($s = 2$) leads to $q = 256$ coefficients, and each row of \mathbf{H}_{M_p} will have only four nonzero elements.

Since $\mathbf{Z}_p(n) = \mathbf{H}_{M_p}\mathbf{X}(n)$, each sequence $\{z_i(n)\}$, $i = 1, \dots, q$, corresponding to the time evolution of the i th component of $\mathbf{Z}_p(n)$ will be a linearly filtered version of the input sequence $\{x(n)\}$. Thus, the $\{z_i(n)\}$ are random sequences correlated in time. Moreover, the rows of \mathbf{H}_{M_p} are orthogonal to each other (in the vector sense). Then, input sequences $\{z_i(n)\}$ and $\{z_j(n)\}$, $i \neq j$, are statistically orthogonal for an i.i.d. input process $\{x(n)\}$. These sequences are the inputs to adaptive weights $w_i(n)$ and $w_j(n)$, respectively. Thus, for $\mathbf{R}_x = \sigma_x^2 \mathbf{I}_N$, where \mathbf{I}_N is the N -dimensional identity matrix, it is easy to verify that $\mathbf{R}_z = E[\mathbf{Z}_p(n)\mathbf{Z}_p^T(n)] = \sigma_x^2 \mathbf{I}_q$. Also, $\mathbf{H}_{M_p}\mathbf{H}_{M_p}^T = \mathbf{I}_q$ [3].

B. Properties of the Estimation Error

The order reducing transformation \mathbf{H}_{M_p} leads to transformed input vectors $\mathbf{Z}_p(n)$ that lie in a q -dimensional subspace spanned by the columns of \mathbf{H}_{M_p} . \mathbf{H}_{M_p} is a

linear transformation on the input signal. Thus, the optimum weight vector in the transformed domain⁴ is given by $\mathbf{W}_{H_p o} = (\mathbf{H}_{M_p} \mathbf{R}_x \mathbf{H}_{M_p}^T)^{-1} \mathbf{H}_{M_p} \mathbf{R}_x \mathbf{W}_o$, which reduces to $\mathbf{W}_{H_p o} = \mathbf{H}_{M_p} \mathbf{W}_o$ for white inputs.

An important consequence of undermodeling (an order reducing transformation that does not capture the exact input subspace) is the nature of the optimum estimation error. From Fig. 3 and using the expression $\mathbf{Z}_p(n) = \mathbf{H}_{M_p} \mathbf{X}(n)$, the estimation error $e_h(n)$ is given by

$$\begin{aligned} e_h(n) &= d(n) - \mathbf{W}_{H_p}^T(n) \mathbf{Z}_p(n) \\ &= d(n) - \mathbf{W}_{H_p}^T(n) \mathbf{H}_{M_p} \mathbf{X}(n). \end{aligned} \quad (4)$$

Using (1) in (4) and evaluating the optimum estimation error $e_o(n)$, which corresponds to $e_h(n)$ for $\mathbf{W}_{H_p} = \mathbf{W}_{H_p o}$, yields

$$e_o(n) = \left(\mathbf{W}_o^T - \mathbf{W}_{H_p o}^T \mathbf{H}_{M_p} \right) \mathbf{X}(n) + \eta(n). \quad (5)$$

From (5), the lagged autocorrelation of $e_o(n)$ can be easily evaluated as

$$\begin{aligned} E\{e_o(n)e_o(m)\} &= \left(\mathbf{W}_o^T - \mathbf{W}_{H_p o}^T \mathbf{H}_{M_p} \right) \\ &\times E\left\{ \mathbf{X}(n) \mathbf{X}^T(m) \right\} \left(\mathbf{W}_o - \mathbf{H}_{M_p}^T \mathbf{W}_{H_p o} \right) + \sigma_\eta^2 \delta(n-m) \end{aligned} \quad (6)$$

where σ_η^2 is the power of $\eta(n)$ and $\delta(n-m) = 1$ for $m = n$ and equal to zero otherwise.

Since $E\left\{ \mathbf{X}(n) \mathbf{X}^T(m) \right\} \neq \mathbf{0}$ even for white $\{x(n)\}$,⁵ the condition for $e_o(n)$ to be uncorrelated is

$$\mathbf{W}_o = \mathbf{H}_{M_p}^T \mathbf{W}_{H_p o} \quad (7)$$

which cannot be satisfied unless \mathbf{W}_o is in the row space of \mathbf{H}_{M_p} . This special case would correspond to the optimal transformation, leading to a complete cancellation of the part of $d(n)$ which is correlated with $x(n)$. Thus, in general, $e_o(n)$ must be considered correlated in time, while the use of the IA in (6) would lead to the erroneous conclusion that $e_o(n)$ is uncorrelated in time.

Straightforward calculation also shows that

$$\begin{aligned} E\{e_o(m) \mathbf{Z}_p(n)\} \\ = \mathbf{H}_{M_p} E\left\{ \mathbf{X}(n) \mathbf{X}^T(m) \right\} \left(\mathbf{W}_o - \mathbf{H}_{M_p}^T \mathbf{W}_{H_p o} \right). \end{aligned} \quad (8)$$

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

$$\begin{aligned} \mathbf{W}_{H_p o} &= \left(\mathbf{H}_{M_p} E\left\{ \mathbf{X}(n) \mathbf{X}^T(m) \right\} \mathbf{H}_{M_p}^T \right)^{-1} \\ &\times \mathbf{H}_{M_p} E\left\{ \mathbf{X}(n) \mathbf{X}^T(m) \right\} \mathbf{W}_o \end{aligned} \quad (9)$$

which is true if (7) holds or if $m = n$, since $E\left\{ \mathbf{X}(n) \mathbf{X}^T(n) \right\} = \mathbf{R}_x$.⁶ Excluding the case of exact subspace modeling [(7) satisfied], this result shows that the residual estimation error is correlated with $z(n)$. Thus, the

⁴The weight vector that minimizes $E\{e_h^2(n)\}$ for the given transformation matrix \mathbf{H}_{M_p} .

⁵For $x(n)$ white, $E\left\{ \mathbf{X}(n) \mathbf{X}^T(m) \right\} \neq 0$ for $n - (N-1) \leq m \leq n + (N-1)$.

⁶Note that for $m = n$ (8) becomes $E\{e_o(n) \mathbf{Z}_p(n)\} = 0$, which must be satisfied by the orthogonality principle.

use of the IA would lead to erroneous results. Statistical tests of the residual estimation error autocorrelation and the residual estimation error and the input signal cross-correlation are frequently used in linear system identification for model validation [14], [15].

Equations (5) and (7) show that the optimum estimation error $e_o(n)$ (for a given transformation \mathbf{H}_{M_p}) consists of the additive white noise $\eta(n)$, statistically independent of $x(n)$, plus a second nonzero term (except when \mathbf{H}_{M_p} is the exact subspace modeling transformation). This second term is correlated in time [see (6)] and cannot be cancelled by the subspace adaptive filter. The second term is also correlated with $x(n)$ [see (8)]. The second term is the main reason why the IA-based model derived in [3] leads to poor results for a tapped delay line filter structure. Equations (5) and (7) also show that the analysis technique must consider the statistical correlation between the estimation error and $\mathbf{Z}_p(n)$. Such analysis is not available in the literature and requires a new approach.

III. FORMULATION OF THE ANALYSIS PROBLEM

This section presents the statistical analysis of the LMS adaptive algorithm under the conditions described in the previous section. This analysis corresponds to the following adaptive filtering setup.

- 1) The adaptive structure is characterized as a linear combiner with q statistically orthogonal inputs $z_i(n)$, $i = 1, \dots, q$;
- 2) The input sequences $\{z_i(n)\}$ are correlated in time;
- 3) The sequence of $q \times 1$ input vectors $\mathbf{Z}_p(n) = [z_1(n), \dots, z_q(n)]^T$ is obtained from a sequence of $N \times 1$ random vectors $\mathbf{X}(n) = [x(n), x(n-1), \dots, x(n-N+1)]^T$ with $N > q$ by the linear transformation

$$\mathbf{Z}_p(n) = \mathbf{H}_{M_p} \mathbf{X}(n) \quad (10)$$

where \mathbf{H}_{M_p} is a $p \times N$ matrix as defined in Section II-A.

A. Model Assumptions

The following analysis assumes that:

A1: The input signal $x(n)$ is stationary, i.i.d., zero-mean and Gaussian. Thus, the input vector $\mathbf{X}(n)$ has autocorrelation matrix $\mathbf{R}_x = E[\mathbf{X}(n) \mathbf{X}^T(n)] = \sigma_x^2 \mathbf{I}_N$, where \mathbf{I}_N is the $N \times N$ identity matrix.

A2 : $\{x(n)\}$ and $\{d(m)\}$ for all n and m are zero-mean jointly stationary Gaussian sequences.

A3: An appropriate delay has been introduced in the signal path to compensate for the extra delay introduced in the adaptive path by the transformation.

It has been shown that the minimum delay required for a wavelet decomposition with J levels is $\Delta = (N_{H_{\max}} + N_{H_{\min}})/2 - J - 2$, where $N_{H_{\max}}$ and $N_{H_{\min}}$ are, respectively, the lengths of the longest and the shortest analysis filters associated to the wavelet decomposition [23]. Implementations using linear transformations which are not wavelet decompositions will have specific delays which should be evaluated. This is a necessary step for proper system implementation. Assume the correct delay has been introduced. Thus, the analysis can be performed independent of Δ .

Input vectors $\mathbf{X}(n)$ and $\mathbf{X}(m)$ 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 error is not i.i.d..

The stationary Gaussian signal model is not necessarily representative of speech, since speech is highly nonstationary and non-Gaussian. However, the theoretical approach and results obtained with Gaussian signal models are suggestive of the fundamental analysis and design issues [2], [24]–[29]. The i.i.d. assumption for $x(n)$ also departs from the speech signal model. However, the problem studied corresponds to a linear combiner with input sequences correlated in time. Thus, the following analysis already considers a form of input signal correlation. The i.i.d. nature of the input signal $x(n)$ limits the correlation between different input sequences to the linear combiner. This assumption keeps the mathematical analysis problem tractable, while preserving the time correlation property of the input sequences $\{z_i(n)\}$. Furthermore, sparse systems are not restricted to speech applications [30] and the results derived in this paper may prove useful for other applications.

The analysis here assumes that \mathbf{W}_o (see Fig. 1) has a finite impulse response. The analysis is performed in the time domain. Thus, the analytical approach here applies to systems with any frequency response whose impulse response can be modeled by an FIR filter.

B. A Mathematical Model

The LMS weight recursion for the transformed domain adaptive filter Fig. 3 is given by [3, eq. (14)]

$$\begin{aligned} \mathbf{W}_{H_p}(n+1) &= \mathbf{W}_{H_p}(n) + \mu e_h(n) \mathbf{Z}_p(n) \\ &= [\mathbf{I}_q - \mu \mathbf{Z}_p(n) \mathbf{Z}_p^T(n)] \mathbf{W}_{H_p}(n) + \mu d(n) \mathbf{Z}_p(n). \end{aligned} \quad (11)$$

Subtracting the optimum weight vector $\mathbf{W}_{H_p o} = \mathbf{H}_{M_p} \mathbf{W}_o$ from both sides of (11) and defining $\mathbf{V}_p(n) = \mathbf{W}_{H_p}(n) - \mathbf{W}_{H_p o}$ yields a recursion for $\mathbf{V}_p(n)$,

$$\begin{aligned} \mathbf{V}_p(n+1) &= [\mathbf{I}_q - \mu \mathbf{Z}_p(n) \mathbf{Z}_p^T(n)] \mathbf{V}_p(n) \\ &\quad + \mu [d(n) - \mathbf{Z}_p^T(n) \mathbf{W}_{H_p o}] \mathbf{Z}_p(n). \end{aligned} \quad (12)$$

Equation (12) is analogous to [28, eq. (9.46)] but is for the real rather than complex LMS algorithm. The last term in brackets is the Wiener error (i.e., the error obtained using the optimum Wiener filter), defined as $e_o(n) = d(n) - \mathbf{Z}_p^T(n) \mathbf{W}_{H_p o}$. The expected value of the last term in (12) is $E[e_o(n) \mathbf{Z}_p(n)] = \mathbf{0}$ from the orthogonality principle.

To determine the stochastic algorithm behavior, the well-known approach [28] uses (12) to derive recursions for the mean $E[\mathbf{V}_p(n)]$ and the covariance matrix $E[\mathbf{V}_p(n) \mathbf{V}_p^T(n)]$. The problem with this approach is that the recursion for $E[\mathbf{V}_p(n) \mathbf{V}_p^T(n)]$ occurs only when the assumption $E[\mathbf{X}(n) \mathbf{X}^T(m)] = \mathbf{0}$ for $n \neq m$ is used. Otherwise, the

recursion involves $E[\mathbf{V}_p(n) \mathbf{V}_p^T(m)]$ for $n \neq m$ as well. If one attempts to derive a recursion for $E[\mathbf{V}_p(n) \mathbf{V}_p^T(m)]$, then expectations involving $\mathbf{X}(m)$ and $\mathbf{V}_p(n)$ appear unless one invokes the independence assumption again.

Our approach to this analysis requires an approximation to (12) that has a closed form solution so as to avoid the problems described above with the recursive solution. To this end, the term $\mathbf{Z}_p(n) \mathbf{Z}_p^T(n)$ can be written as a mean plus a fluctuating part

$$\mathbf{Z}_p(n) \mathbf{Z}_p^T(n) = E[\mathbf{Z}_p(n) \mathbf{Z}_p^T(n)] + \Psi(n) = \sigma_x^2 \mathbf{I}_q + \Psi(n). \quad (13)$$

Inserting (13) in (12) yields

$$\mathbf{V}_p(n+1) = (1 - \mu \sigma_x^2) \mathbf{V}_p(n) + \mu e_o(n) \mathbf{Z}_p(n) - \mu \Psi(n) \mathbf{V}_p(n). \quad (14)$$

Equation (14) can be viewed as a deterministic recursion for $\mathbf{V}_p(n)$ driven by two random inputs: $e_o(n) \mathbf{Z}_p(n)$ and $\Psi(n) \mathbf{V}_p(n)$. During the beginning of the transient phase of adaptation (small n), the fluctuations of $\mathbf{V}_p(n)$ are small compared with $E[\mathbf{V}_p(n)]$, and $\Psi(n) \mathbf{V}_p(n)$ can be approximated by $\Psi(n) E[\mathbf{V}_p(n)]$. This approximation becomes less accurate for moderate values of n and as the step size μ increases. Close to convergence $E[\mathbf{V}_p(n)]$ tends to zero and the input to the recursion can be approximated by $e_o(n) \mathbf{Z}_p(n)$ if the fluctuations in $\mathbf{V}_p(n)$ are sufficiently small so that⁷

$$\left| \left\{ e_o(n) \mathbf{Z}_p(n) \right\}_i \right| \gg \left| \left\{ \Psi(n) \{ \mathbf{V}_p(n) - E[\mathbf{V}_p(n)] \} \right\}_i \right| \quad (15)$$

for $i = 1, \dots, q$, where the subscript i stands for the i th component of the vector. Approximation (15) is more valid for smaller step sizes.

Assuming (15), (14) can be approximated by the recursion

$$\begin{aligned} \mathbf{V}_p(n+1) &\simeq (1 - \mu \sigma_x^2) \mathbf{V}_p(n) + \mu e_o(n) \mathbf{Z}_p(n) \\ &\quad - \mu \Psi(n) E[\mathbf{V}_p(n)]. \end{aligned} \quad (16)$$

Equation (16) can be used to explicitly determine the effects of $E[\mathbf{Z}_p(n) \mathbf{Z}_p^T(m)] \neq \mathbf{0}$ for $n \neq m$ on the behavior of the weight error vector. Equation (16) implicitly assumes that IA holds for the mean behavior of $\mathbf{V}_p(n)$.

Viewing the last two terms on the right-hand side (RHS) of (16) as forcing terms, (16) has an explicit closed form solution

$$\begin{aligned} \mathbf{V}_p(n) &\simeq (1 - \mu \sigma_x^2)^n \mathbf{V}_p(0) + \mu \sum_{m=0}^{n-1} (1 - \mu \sigma_x^2)^{n-m-1} \\ &\quad \left\{ e_o(m) \mathbf{Z}_p(m) - \Psi(m) E[\mathbf{V}_p(m)] \right\}. \end{aligned} \quad (17)$$

Equation (17) represents a deterministic system with random inputs and can be used to determine the response to correlated inputs vectors $\mathbf{Z}_p(m), \mathbf{Z}_p(m-1), \dots, \mathbf{Z}_p(m-k)$, $k = 2, \dots, n-1$.

⁷Condition (15) is a small fluctuation assumption, not uncommon in analysis of adaptive filters.

TABLE I
SIGNIFICANT COEFFICIENTS OF THE PARTIAL-HAAR TRANSFORMS OF THE SPARSE RESPONSE FOR DIFFERENT BULK DELAYS

256-tap Partial-Haar	$\mathbf{W}_{H_p o} = [0.0005, -0.0088, -0.1406, 0.3750, 0.00703, 0.0044, 0.0003]^T$
	$\mathbf{W}_{H_p o} = [-0.0001, -0.0022, -0.0352, -0.5625, 0.2812, 0.0176, 0.0011, 0.0001]^T$
128-tap Partial-Haar	$\mathbf{W}_{H_p o} = [-0.0012, -0.3107, 0.6215, 0.0024]^T$
	$\mathbf{W}_{H_p o} = [-0.0002, -0.0388, 0.3315, 0.0194, 0.0001]^T$

IV. STOCHASTIC BEHAVIOR ANALYSIS

A. Mean Weight Behavior

Averaging (17) and using the orthogonality principle

$$E[\mathbf{V}_p(n)] \simeq (1 - \mu\sigma_x^2)^n \mathbf{V}_p(0) \quad (18)$$

since $E[\Psi(n)] = \mathbf{0}$. 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 (16). Fortunately, this is not the case for the weight fluctuation behavior, as will become clear in the next section.

B. Weight Fluctuation Behavior

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

$$\begin{aligned} \mathbf{Q}(n) &= E \left\{ [\mathbf{V}_p(n) - E[\mathbf{V}_p(n)]] [\mathbf{V}_p(n) - E[\mathbf{V}_p(n)]]^T \right\} \\ &= \mu^2 \sum_{m=0}^{n-1} \sum_{r=0}^{n-1} (1 - \mu\sigma_x^2)^{n-m-1} (1 - \mu\sigma_x^2)^{n-r-1} \\ &\quad \times \left[E \left\{ \mathbf{Z}_p(m) e_o(m) e_o(r) \mathbf{Z}_p^T(r) \right. \right. \\ &\quad \left. \left. + \Psi(m) E[\mathbf{V}_p(m)] E[\mathbf{V}_p^T(r)] \Psi^T(r) \right\} \right. \\ &\quad \left. - E \left\{ e_o(m) \mathbf{Z}_p(m) E[\mathbf{V}_p^T(r)] \Psi^T(r) \right. \right. \\ &\quad \left. \left. + \Psi(m) E[\mathbf{V}_p(m)] \mathbf{Z}_p^T(r) e_o(r) \right\} \right]. \quad (19) \end{aligned}$$

The evaluation of the expected values in (19) is detailed in Appendix I, and leads to

$$\begin{aligned} \mathbf{Q}(n) &= \left(\frac{1-a}{1+a} \right) \mathbf{H}_{Mp} \left\{ \mathbf{G}_0 (1 - a^{2n}) \right. \\ &\quad \left. + \sum_{\nu=1}^{n-1} (\mathbf{G}_\nu + \mathbf{G}_{-\nu}) (a^\nu - a^{2n} a^{-\nu}) \right\} \mathbf{H}_{Mp}^T \\ &\quad + (1-a)^2 a^{2n-2} \mathbf{H}_{Mp} \\ &\quad \times \left[n \mathbf{K}_0 + \sum_{\nu=1}^{n-1} (n-\nu) (\mathbf{K}_\nu + \mathbf{K}_{-\nu}) \right] \mathbf{H}_{Mp}^T \\ &\quad + (1-a) a^{n-1} \left\{ \sum_{\nu=1}^{n-1} \mathbf{T}_\nu (1 - a^{n-\nu}) \right. \\ &\quad \left. + \sum_{\nu=1}^{n-1} \mathbf{T}_{-\nu} (a^\nu - a^n) \right\} \quad (20) \end{aligned}$$

where, defining $\nu = m - r$,

$$a = (1 - \mu\sigma_x^2) \quad (21a)$$

$$\begin{aligned} \mathbf{G}_\nu &= \left[\mathbf{W}_o^T \mathbf{Z}_2 \mathbf{F}_\nu \mathbf{Z}_2 \mathbf{W}_o + \frac{\sigma_\eta^2}{\sigma_x^4} \delta(\nu) \right] \mathbf{F}_\nu \\ &\quad + \mathbf{F}_\nu \mathbf{Z}_2 \mathbf{W}_o \mathbf{W}_o^T \mathbf{Z}_2 \mathbf{F}_\nu \quad (21b) \end{aligned}$$

$$\begin{aligned} \mathbf{K}_\nu &= \mathbf{F}_\nu \mathbf{H}_{Mp}^T \mathbf{V}_p(0) \mathbf{V}_p^T(0) \mathbf{H}_{Mp} \mathbf{F}_\nu \\ &\quad + \mathbf{V}_p^T(0) \mathbf{H}_{Mp} \mathbf{F}_{-\nu} \mathbf{H}_{Mp}^T \mathbf{V}_p(0) \mathbf{F}_\nu. \quad (21c) \end{aligned}$$

$$\begin{aligned} \mathbf{T}_\nu &= \mathbf{H}_{Mp} \mathbf{F}_\nu \left[\mathbf{V}_p^T(0) \mathbf{H}_{Mp} \mathbf{F}_{-\nu} \mathbf{Z}_2 \mathbf{W}_o \mathbf{I}_N \right. \\ &\quad \left. + \mathbf{H}_{Mp}^T \mathbf{V}_p(0) \mathbf{W}_o^T \mathbf{Z}_2 \mathbf{F}_\nu \right] \mathbf{H}_{Mp}^T \quad (21d) \end{aligned}$$

and

$$\begin{aligned} \mathbf{T}_{-\nu} &= \mathbf{H}_{Mp} \left[\mathbf{I}_N \mathbf{W}_o^T \mathbf{Z}_2 \mathbf{F}_\nu \mathbf{H}_{Mp}^T \mathbf{V}_p(0) \right. \\ &\quad \left. + \mathbf{F}_{-\nu} \mathbf{Z}_2 \mathbf{W}_o \mathbf{V}_p^T(0) \mathbf{H}_{Mp} \right] \mathbf{F}_{-\nu} \mathbf{H}_{Mp}^T \quad (21e) \end{aligned}$$

with \mathbf{Z}_2 and $\mathbf{F}_\nu = \mathbf{F}_{m-r}$ as defined in Appendix I [(25) and (26), respectively].

V. SIMULATION RESULTS

To verify the accuracy of the theoretical model, several Monte Carlo simulations were carried out for different sets of parameters. Consider the symmetric exponential channel impulse response

$$\mathbf{W}_o = [a^r, a^{r-1}, \dots, a, 1, a, \dots, a^{r-1}, a^r]^T \quad (22)$$

for $r = 32$ and $a = 0.5$, which is located in a span of $N = 1024$ samples, leading to a sparse channel response. This is the same response used in [3], and is again used here to facilitate comparison. The optimum Partial-Haar responses $\mathbf{W}_{H_p o}$ for $q = 256$, 128 and 64 were obtained from the dot product of \mathbf{W}_o and the rows of the associated \mathbf{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. The additive measurement noise $\eta(n)$ was made equal to zero to emphasize the effects of the rank reducing transformation on the mean square deviation.

The Partial-Haar transforms of the different channels are shown in Table I [3].⁸ Figs. 4–6 show the tap weight variance over time, estimated from Monte Carlo simulations by computing $\text{tr}[\mathbf{Q}(n)] = E \left\{ [\mathbf{V}_p(n) - E\{\mathbf{V}_p(n)\}]^T \times [\mathbf{V}_p(n) - E\{\mathbf{V}_p(n)\}] \right\}$ for $q = 256$, 128 and 64 and for different

⁸Different bulk channel delays yield different partial Haar transforms for the same impulse response. Table I presents some of these. All results for this impulse response can be found in [3, Table I].

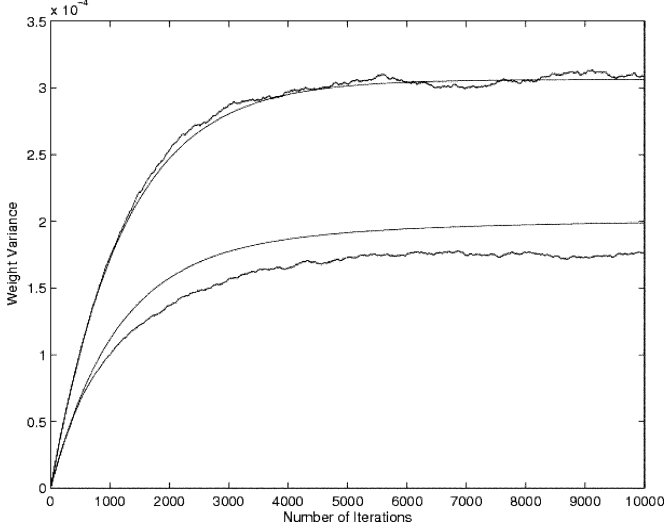


Fig. 4. Weight variance $\text{tr}[\mathbf{Q}(n)]$ for $q = 256$, $\max\{w_{H_p o}\} = 0.375$. Theory (smooth plots), 100 MC Simulations (jagged plots) – Upper plots -Independent (theory from [3]), Lower plots -TDL [theory from (20)].

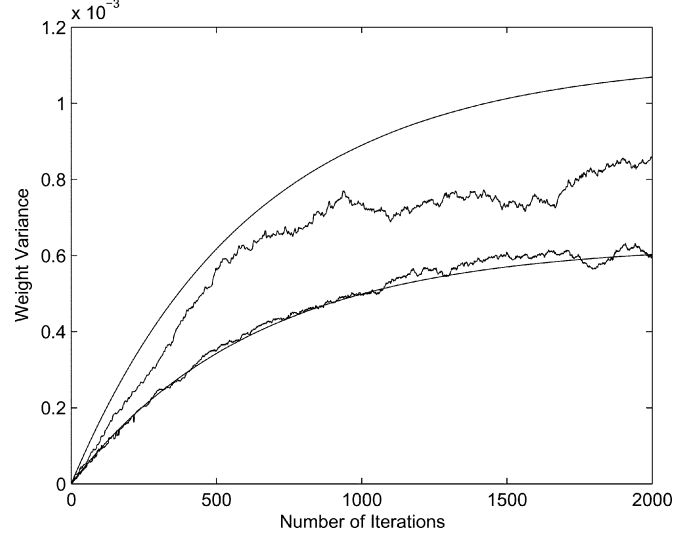


Fig. 6. Weight variance $\text{tr}[\mathbf{Q}(n)]$ for $q = 128$, $\max\{w_{H_p o}\} = 0.3315$. Theory (smooth plots), 10 MC Simulations (jagged plots) – Upper plots -TDL [theory from (20)], Lower plots – Independent (theory from [3]).

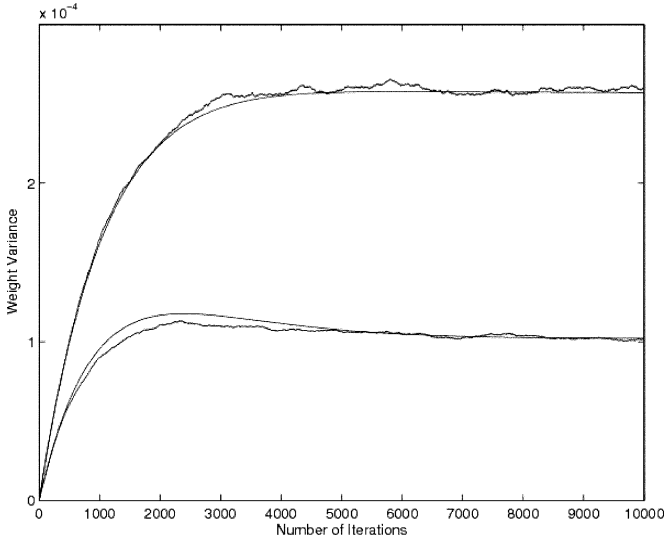


Fig. 5. Weight variance $\text{tr}[\mathbf{Q}(n)]$ for $q = 256$, $\max\{w_{H_p o}\} = 0.5625$, Theory (smooth plots), 100 MC Simulations (jagged plots) – Upper plots -Independent (theory from [3]), Lower plots -TDL [theory from (20)].

values of $\max\{\mathbf{W}_{H_p o}\}$, the maximum value of $\mathbf{W}_{H_p o}$. The step sizes used were given by $\mu = 0.1/(q + 2)$ (about 1/20 of the stability limit) and $\sigma_x^2 = 1$ in all cases. The adaptive filter coefficients were initialized at $\mathbf{W}_{H_p}(0) = \mathbf{0}$, or $\mathbf{V}_p(0) = -\mathbf{W}_{H_p o} = -\mathbf{H}_{M_p} \mathbf{W}_o$. The theoretical curves were obtained from (20). For comparison purposes, the simulations are shown for a representative sample of the cases studied in [3] and the figures also present the plots corresponding to a sequence of statistically independent input vectors $\mathbf{X}(n)$. The curves identified as “TDL” in each plot correspond to the theoretical results obtained using (20) and the Monte Carlo simulations for a tapped delay line filter structure. The curves identified as “Independent” correspond to the model in [3] and simulations for statistically independent input vectors satisfying $E\{\mathbf{X}(n)\mathbf{X}^T(m)\} = \mathbf{0}$ for $n \neq m$. 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. Fig. 6 ($\max\{w_{H_p o}\} = 0.3315$) illustrates the difficulty of the new model to accurately predict the algorithm behavior for large step sizes, small values of q (large scales) and small peak values of the partial-Haar transformed impulse response. Fig. 4 shows some mismatch in the steady-state behavior between the TDL theory and simulations. However, the independence theory model clearly does not provide good predictions, even in steady state, for this case. Fig. 5 shows that the accuracy of the new model is excellent throughout all phases of adaptation, especially in steady state. It is conjectured that the mismatch in Fig. 4 is due to the relatively small value of $\max\{\mathbf{W}_{H_p o}\}$ as compared to that of Fig. 5. This is because $\max\{\mathbf{W}_{H_p o}\}$ displays the attenuation of \mathbf{W}_o ($\max\{\mathbf{W}_o\} = 1$) caused by the partial Haar transformation. The spreading caused by the transformation can cause a significant loss of the peak value of the impulse response (see Table I) as seen through the partial Haar transform. This conjecture was also supported by simulation results obtained for the remaining cases in [3].

VI. CONCLUSION

This paper has developed a new theoretical model for predicting the behavior of the first and second moments of the LMS algorithm with a tapped delay line filtering structure for a low rank system identification problem. The analysis considered the general case of order reducing transformations which do not capture the exact input subspace. The new theory was applied to the partial-Haar transformed adaptive filtering scheme proposed in [3] for network echo cancellation. The theoretical predictions were found in good-to-excellent agreement with Monte Carlo simulations. This was not the case for the theoretical model based on the Independence Theory assumption derived in [3]. The new model can be used to better design the scheme proposed in [3] for estimating the location of the peak of an unknown impulse response for a sparse channel. It can also be

used to study the transient behavior of low rank adaptive systems driven by the LMS algorithm and using different input subspace estimation strategies. Finally, it is anticipated that the present analysis can be easily extended to the case of complex signals, coefficients and transformations with the proper considerations regarding specific mathematical evaluations such as the use of the Gaussian moment factoring theorem.

APPENDIX I

EVALUATION OF EXPECTED VALUES IN (19)

The expectations in the third line of (19) can be factored using the Gaussian moment factoring theorem for real variates [28, p. 318]. The first expectation can be written

$$\begin{aligned} & E [\mathbf{Z}_p(m)e_o(m)e_o(r)\mathbf{Z}_p^T(r)] \\ &= E [e_o(m)e_o(r)] E [\mathbf{Z}_p(m)\mathbf{Z}_p^T(r)] \\ & \quad + E [e_o(m)\mathbf{Z}_p(m)] E [e_o(r)\mathbf{Z}_p^T(r)] \\ & \quad + E [e_o(r)\mathbf{Z}_p(m)] E [e_o(m)\mathbf{Z}_p^T(r)]. \end{aligned} \quad (23)$$

The second term on the RHS is zero from the orthogonality principle and disappears from the summation. When the projection onto the correct input subspace is used (and, thus, IA holds), the third term on the RHS is zero for all r and m and the first term on the RHS is also zero for $r \neq m$. This is not the case in general, and (23) needs to be carefully evaluated. The various terms are evaluated in Appendix II yielding

$$\begin{aligned} & E [e_o(m)e_o(r)] E [\mathbf{Z}_p(m)\mathbf{Z}_p^T(r)] \\ & \quad + E [e_o(r)\mathbf{Z}_p(m)] E [e_o(m)\mathbf{Z}_p^T(r)] \\ &= \sigma_x^4 \left[\left(\mathbf{W}_o^T \mathbf{Z}_2 \mathbf{F}_{m-r} \mathbf{Z}_2 \mathbf{W}_o + \frac{\sigma_\eta^2}{\sigma_x^4} \delta(m-r) \right) \mathbf{H}_{Mp} \mathbf{F}_{m-r} \mathbf{H}_{Mp}^T \right. \\ & \quad \left. + \mathbf{H}_{Mp} \mathbf{F}_{m-r} \mathbf{Z}_2 \mathbf{W}_o \mathbf{W}_o^T \mathbf{Z}_2 \mathbf{F}_{m-r} \mathbf{H}_{Mp}^T \right] \end{aligned} \quad (24)$$

where

$$\mathbf{Z}_2 = \mathbf{I}_N - \mathbf{H}_{Mp}^T \mathbf{H}_{Mp} \quad (25)$$

and

$$\mathbf{F}_{m-r} = \frac{1}{\sigma_x^2} E [\mathbf{X}(m)\mathbf{X}^T(r)]. \quad (26)$$

The expectation of the second term in the brackets in (19) can be evaluated as follows:

$$\begin{aligned} & E \{ \Psi(m) E [\mathbf{V}_p(m)] E [\mathbf{V}_p^T(r)] \Psi^T(r) \} \\ &= E \{ [\mathbf{Z}_p(m)\mathbf{Z}_p^T(m) - \sigma_x^2 \mathbf{I}_q] \mathbf{Z}_1(m)\mathbf{Z}_1^T(r) \\ & \quad \times [\mathbf{Z}_p(r)\mathbf{Z}_p^T(r) - \sigma_x^2 \mathbf{I}_q] \} \end{aligned} \quad (27)$$

where $\mathbf{Z}_1(m) = (1 - \mu\sigma_x^2)^m \mathbf{V}_p(0)$ from (18). Expanding the term in brackets yields

$$\begin{aligned} & E \{ \Psi(m) E [\mathbf{V}_p(m)] E [\mathbf{V}_p^T(r)] \Psi^T(r) \} \\ &= E [\mathbf{Z}_p(m)\mathbf{Z}_p^T(m)\mathbf{Z}_1(m)\mathbf{Z}_1^T(r)\mathbf{Z}_p(r)\mathbf{Z}_p^T(r)] \\ & \quad - \sigma_x^4 \mathbf{Z}_1(m)\mathbf{Z}_1^T(r). \end{aligned} \quad (28)$$

Note that the product of the middle four factors in the expectation is equivalent to the product of two scalars so that Gaussian moment factorization can be immediately applied, yielding

$$\begin{aligned} & E [\mathbf{Z}_p(m)\mathbf{Z}_p^T(m)\mathbf{Z}_1(m)\mathbf{Z}_1^T(r)\mathbf{Z}_p(r)\mathbf{Z}_p^T(r)] \\ &= E [\mathbf{Z}_p(m)\mathbf{Z}_p^T(m)\mathbf{Z}_1(m)] E [\mathbf{Z}_1^T(r)\mathbf{Z}_p(r)\mathbf{Z}_p^T(r)] \\ & \quad + E [\mathbf{Z}_p(m)\mathbf{Z}_p^T(r)\mathbf{Z}_1(r)] E [\mathbf{Z}_1^T(m)\mathbf{Z}_p(m)\mathbf{Z}_p^T(r)] \\ & \quad + E [\mathbf{Z}_p(m)\mathbf{Z}_p^T(r)] E [\mathbf{Z}_1^T(m)\mathbf{Z}_p(m)\mathbf{Z}_p^T(r)\mathbf{Z}_1(r)] \\ &= \sigma_x^4 \mathbf{Z}_1(m)\mathbf{Z}_1^T(r) \\ & \quad + \mathbf{H}_{Mp} \mathbf{F}_{m-r} \mathbf{H}_{Mp}^T \mathbf{Z}_1(r)\mathbf{Z}_1^T(m)\mathbf{H}_{Mp} \mathbf{F}_{m-r} \mathbf{H}_{Mp}^T \\ & \quad + \mathbf{Z}_1^T(r)\mathbf{H}_{Mp} \mathbf{F}_{r-m} \mathbf{H}_{Mp}^T \mathbf{Z}_1(m)\mathbf{H}_{Mp} \mathbf{F}_{m-r} \mathbf{H}_{Mp}^T. \end{aligned} \quad (29)$$

Using (14), the third term within the brackets in (19) can be written as

$$\begin{aligned} & E \left\{ e_o(m)\mathbf{Z}_p(m) E [\mathbf{V}_p^T(r)] \Psi^T(r) \right\} \\ &= E \left\{ e_o(m)\mathbf{Z}_p(m) E [\mathbf{V}_p^T(r)] [\mathbf{Z}_p(r)\mathbf{Z}_p^T(r) - \sigma_x^2 \mathbf{I}_q] \right\}. \end{aligned} \quad (30)$$

Noting that $e_o(m)$ and $E[\mathbf{V}_p^T(r)]\mathbf{Z}_p(r)$ are scalars and using the moment factorization theorem yields

$$\begin{aligned} & E \left\{ e_o(m)\mathbf{Z}_p(m) E [\mathbf{V}_p^T(r)] \Psi^T(r) \right\} \\ &= E [e_o(m)\mathbf{Z}_p(m)] E [\mathbf{V}_p^T(r)] E [\mathbf{Z}_p(r)\mathbf{Z}_p^T(r) - \sigma_x^2 \mathbf{I}_q] \\ & \quad + E \left\{ e_o(m) E [\mathbf{V}_p^T(r)] \mathbf{Z}_p(r) \right\} E [\mathbf{Z}_p(m)\mathbf{Z}_p^T(r)] \\ & \quad + E \left\{ \mathbf{Z}_p(m) E [\mathbf{V}_p^T(r)] \mathbf{Z}_p(r) \right\} E [e_o(m)\mathbf{Z}_p^T(r)]. \end{aligned} \quad (31)$$

The first term in (31) is zero both because the first term is zero by the orthogonality principle and because $E[\Psi(r)] = 0$. Using the expressions for $e_o(m)$ and $d(n)$, the second term can be expanded as

$$\begin{aligned} & E \left\{ e_o(m) E [\mathbf{V}_p^T(r)] \mathbf{Z}_p(r) \right\} E [\mathbf{Z}_p(m)\mathbf{Z}_p^T(r)] \\ &= E [\mathbf{V}_p^T(r)] E \left\{ [\mathbf{W}_o^T \mathbf{X}(m) + \eta(m) - \mathbf{Z}_p^T(m)\mathbf{W}_{H_p o}] \mathbf{Z}_p(r) \right\} \\ & \quad \times E [\mathbf{H}_{Mp} \mathbf{X}(m)\mathbf{X}^T(r)\mathbf{H}_{Mp}^T] \\ &= E [\mathbf{V}_p^T(r)] E \left[\mathbf{Z}_p(r)\mathbf{X}^T(m)\mathbf{W}_o - \mathbf{Z}_p(r)\mathbf{Z}_p^T(m)\mathbf{W}_{H_p o} \right] \\ & \quad \times \sigma_x^2 \mathbf{H}_{Mp} \mathbf{F}_{m-r} \mathbf{H}_{Mp}^T \end{aligned} \quad (32)$$

where the definition of \mathbf{F}_{m-r} and the fact that $\eta(m)$ is independent and zero-mean have been used to obtain the last expression.

The second expectation in (32) is

$$\begin{aligned} & E [\mathbf{Z}_p(r)\mathbf{X}^T(m)\mathbf{W}_o - \mathbf{Z}_p(r)\mathbf{Z}_p^T(m)\mathbf{W}_{H_p o}] \\ &= \sigma_x^2 [\mathbf{H}_{Mp} \mathbf{F}_{r-m} \mathbf{W}_o - \mathbf{H}_{Mp} \mathbf{F}_{r-m} \mathbf{H}_{Mp}^T \mathbf{W}_{H_p o}] \\ &= \sigma_x^2 \mathbf{H}_{Mp} \mathbf{F}_{r-m} [\mathbf{I}_N - \mathbf{H}_{Mp}^T \mathbf{H}_{Mp}] \mathbf{W}_o \\ &= \sigma_x^2 \mathbf{H}_{Mp} \mathbf{F}_{r-m} \mathbf{Z}_2 \mathbf{W}_o \end{aligned} \quad (33)$$

with \mathbf{Z}_2 defined below (24). Inserting (33) in (32) yields

$$\begin{aligned} E \left\{ e_o(m) E[\mathbf{V}_p^T(r)] \mathbf{Z}_p(r) \right\} E[\mathbf{Z}_p(m) \mathbf{Z}_p^T(r)] \\ = \sigma_x^4 E[\mathbf{V}_p^T(r)] \mathbf{H}_{Mp} \mathbf{F}_{r-m} \mathbf{Z}_2 \mathbf{W}_o \mathbf{H}_{Mp} \mathbf{F}_{m-r} \mathbf{H}_{Mp}^T \end{aligned} \quad (34)$$

for the second term in (31).

For the third term in (31), the expectation $E[e_o(m) \mathbf{Z}_p^T(r)]$ is the transpose of (33). The remaining expectation is given by

$$\begin{aligned} E \left\{ \mathbf{Z}_p(m) E[\mathbf{V}_p^T(r)] \mathbf{Z}_p(r) \right\} &= E[\mathbf{Z}_p(m) \mathbf{Z}_p^T(r)] E[\mathbf{V}_p(r)] \\ &= \sigma_x^2 \mathbf{H}_{Mp} \mathbf{F}_{r-q} \mathbf{H}_{Mp}^T E[\mathbf{V}_p(r)]. \end{aligned} \quad (35)$$

Thus, the third term in (31) is

$$\begin{aligned} E \left\{ \mathbf{Z}_p(m) E[\mathbf{V}_p^T(r)] \mathbf{Z}_p(r) \right\} E[e_o(m) \mathbf{Z}_p^T(r)] \\ = \sigma_x^4 \mathbf{H}_{Mp} \mathbf{F}_{r-q} \mathbf{H}_{Mp}^T E[\mathbf{V}_p(r)] \mathbf{W}_o^T \mathbf{Z}_2 \mathbf{F}_{m-r} \mathbf{H}_{Mp}^T. \end{aligned} \quad (36)$$

Finally, entering (34) and (36) in (31), using (18) and factoring out common matrix terms from both sides yields

$$\begin{aligned} E \left\{ e_o(m) \mathbf{Z}_p(m) E[\mathbf{V}_p^T(r)] \mathbf{\Psi}^T(r) \right\} \\ = -\sigma_x^4 (1 - \mu \sigma_x^2)^r \mathbf{H}_{Mp} \mathbf{F}_{m-r} \left[\mathbf{V}_p^T(0) \mathbf{H}_{Mp} \mathbf{F}_{r-m} \mathbf{Z}_2 \mathbf{W}_o \mathbf{I}_N \right. \\ \left. + \mathbf{H}_{Mp}^T \mathbf{V}_p(0) \mathbf{W}_o^T \mathbf{Z}_2 \mathbf{F}_{m-r} \right] \mathbf{H}_{Mp}^T. \end{aligned} \quad (37)$$

The last expectation in (19) is the transpose of (37) with r and m exchanged. Thus

$$\begin{aligned} E \left\{ \mathbf{\Psi}(m) E[\mathbf{V}_p(m)] \mathbf{Z}_p^T(r) e_o(r) \right\} \\ = -\sigma_x^4 (1 - \mu \sigma_x^2)^r \mathbf{H}_{Mp} \left[\mathbf{I}_N \mathbf{W}_o^T \mathbf{Z}_2 \mathbf{F}_{m-r} \mathbf{H}_{Mp}^T \mathbf{V}_p(0) \right. \\ \left. + \mathbf{F}_{r-m} \mathbf{Z}_2 \mathbf{W}_o \mathbf{V}_p^T(0) \mathbf{H}_{Mp} \right] \mathbf{F}_{r-m} \mathbf{H}_{Mp}^T. \end{aligned} \quad (38)$$

Inserting (29) in (28), and then using (24), (28), (37) and (38) in (19) yields (39), shown at the bottom of the page.

The double sums in (39) can be simplified to single sums by a change of variable $m - r = \nu$ and an interchange of the order of the double sum yielding (20).

APPENDIX II EVALUATION OF (24)

Consider first

$$\begin{aligned} E[e_o(m) e_o(r)] &= \mathbf{W}_o^T \left[\mathbf{I}_N - \mathbf{H}_{Mp}^T \mathbf{H}_{Mp} \right] E[\mathbf{X}(m) \mathbf{X}^T(r)] \\ &\quad \times \left[\mathbf{I}_N - \mathbf{H}_{Mp}^T \mathbf{H}_{Mp} \right] \mathbf{W}_o + \sigma_\eta^2 \delta(m - r). \end{aligned} \quad (40)$$

The matrices $\mathbf{Z}_0 = \mathbf{H}_{Mp}^T \mathbf{H}_{Mp}$ and $\mathbf{Z}_2 = \mathbf{I}_N - \mathbf{H}_{Mp}^T \mathbf{H}_{Mp}$ appear in many places in the subsequent calculations. In the case of partial Haar transformations \mathbf{Z}_0 is an $N \times N$ sparse matrix with all zeros except on the main diagonal and the adjacent 3 or 7 or 15 off-diagonal terms for Haar 256, Haar 128, and Haar 64 transforms, respectively. \mathbf{Z}_2 is idempotent. The middle expectation term is a sparse matrix with σ_x^2 only on the $r - m$ off-diagonal and zeros elsewhere. If $r - m > 0$, then the upper diagonal is nonzero. If $r - m < 0$, then the lower diagonal is nonzero. When $|r - m| \geq N$, the matrix is null, displaying the independence of input vectors which are separated in time by more than the length of the delay line.

Thus, the first term in (23) is

$$\begin{aligned} E[e_o(m) e_o(r)] E[\mathbf{Z}_p(m) \mathbf{Z}_p^T(r)] \\ = E[e_o(m) e_o(r)] \mathbf{H}_{Mp} E[\mathbf{X}(m) \mathbf{X}^T(r)] \mathbf{H}_{Mp}^T \\ = \sigma_x^4 \left[\mathbf{W}_o^T \mathbf{Z}_2 \mathbf{F}_{m-r} \mathbf{Z}_2 \mathbf{W}_o + \frac{\sigma_\eta^2}{\sigma_x^2} \delta(m - r) \right] \mathbf{H}_{Mp} \mathbf{F}_{m-r} \mathbf{H}_{Mp}^T. \end{aligned} \quad (41)$$

The third line in (23) is $E[e_o(r) \mathbf{Z}_p(m)] E[e_o(m) \mathbf{Z}_p^T(r)]$. Now

$$\begin{aligned} E[e_o(m) \mathbf{Z}_p^T(r)] \\ = \mathbf{W}_o^T \left[\mathbf{I}_N - \mathbf{H}_{Mp}^T \mathbf{H}_{Mp} \right] E[\mathbf{X}(m) \mathbf{X}^T(r)] \mathbf{H}_{Mp}^T \\ = \sigma_x^2 \mathbf{W}_o^T \mathbf{Z}_2 \mathbf{F}_{m-r} \mathbf{H}_{Mp}^T. \end{aligned} \quad (42)$$

Hence

$$\begin{aligned} E[e_o(r) \mathbf{Z}_p(m)] E[e_o(m) \mathbf{Z}_p^T(r)] \\ = \sigma_x^4 \mathbf{H}_{Mp} \mathbf{F}_{m-r} \mathbf{Z}_2 \mathbf{W}_o \mathbf{W}_o^T \mathbf{Z}_2 \mathbf{F}_{m-r} \mathbf{H}_{Mp}^T. \end{aligned} \quad (43)$$

Inserting (41) and (43) in the left-hand side (LHS) of (24) leads to the RHS of (24).

$$\begin{aligned} \mathbf{Q}(n) &= \mu^2 \sigma_x^4 \sum_{m=0}^{n-1} \sum_{r=0}^{n-1} (1 - \mu \sigma_x^2)^{n-m-1} (1 - \mu \sigma_x^2)^{n-r-1} \\ &\quad \times \mathbf{H}_{Mp} \left[\begin{aligned} &\left[\mathbf{W}_o^T \mathbf{Z}_2 \mathbf{F}_{m-r} \mathbf{Z}_2 \mathbf{W}_o + \frac{\sigma_\eta^2}{\sigma_x^2} \delta(m - r) \right] \mathbf{F}_{m-r} + \mathbf{F}_{m-r} \mathbf{Z}_2 \mathbf{W}_o \mathbf{W}_o^T \mathbf{Z}_2 \mathbf{F}_{m-r} \\ &+ \mathbf{F}_{m-r} \mathbf{H}_{Mp}^T \mathbf{Z}_1(r) \mathbf{Z}_1^T(m) \mathbf{H}_{Mp} \mathbf{F}_{m-r} + \mathbf{Z}_1^T(r) \mathbf{H}_{Mp} \mathbf{F}_{r-m} \mathbf{H}_{Mp}^T \mathbf{Z}_1(m) \mathbf{F}_{m-r} \\ &+ (1 - \mu \sigma_x^2)^r \mathbf{F}_{m-r} \left[\mathbf{V}_p^T(0) \mathbf{H}_{Mp} \mathbf{F}_{r-m} \mathbf{Z}_2 \mathbf{W}_o \mathbf{I}_N + \mathbf{H}_{Mp}^T \mathbf{V}_p(0) \mathbf{W}_o^T \mathbf{Z}_2 \mathbf{F}_{m-r} \right] \\ &+ (1 - \mu \sigma_x^2)^r \left[\mathbf{I}_N \mathbf{W}_o^T \mathbf{Z}_2 \mathbf{F}_{m-r} \mathbf{H}_{Mp}^T \mathbf{V}_p(0) + \mathbf{F}_{r-m} \mathbf{Z}_2 \mathbf{W}_o \mathbf{V}_p^T(0) \mathbf{H}_{Mp} \right] \mathbf{F}_{r-m} \end{aligned} \right] \mathbf{H}_{Mp}^T. \end{aligned} \quad (39)$$

REFERENCES

- [1] S. Haykin, *Adaptive Filter Theory*, 4th ed. Englewood Cliffs, NJ: Prentice-Hall, 2002.
- [2] A. H. Sayed, *Fundamentals of Adaptive Filtering*. New York: Wiley-Interscience, 2003.
- [3] N. J. Bershad and A. Bist, "Fast coupled adaptation for sparse impulse responses using a partial Haar transform," *IEEE Trans. Signal Process.*, vol. 53, no. 3, pp. 966–976, Mar. 2005.
- [4] M. M. Sondhi, "The history of echo cancellation," *IEEE Signal Process. Mag.*, vol. 23, no. 5, pp. 95–102, Sept. 2006.
- [5] Digital Network Echo Cancellers ITU-T, 2004, ITU-T Recommendation G. 168.
- [6] O. Macchi, *Adaptive Processing: The Least Mean Squares Approach with Applications in Transmission*. West Sussex, U.K.: Wiley, 1995.
- [7] O. Macchi and E. Eweda, "Second order convergence analysis of stochastic adaptive linear filtering," *IEEE Trans. Autom. Control*, vol. AC-28, no. 1, pp. 76–85, Jan. 1983.
- [8] J. Chao, S. Kawabe, and S. Tsujii, "Convergence analysis of gradient adaptive algorithms for arbitrary inputs without the independence assumption," in *Proc. IEEE Int. Conf. Syst. Eng.*, Kobe, Japan, Sep. 1992, pp. 556–559.
- [9] S. C. Douglas and T. H. Y. Meng, "Exact expectation analysis of the LMS adaptive filter without the independence assumption," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, San Francisco, CA, Mar. 1992, vol. 4, pp. 61–64.
- [10] S. C. Douglas, "Exact expectation analysis of the LMS adaptive filter for correlated Gaussian input data," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Minneapolis, MN, Apr. 1993, vol. 3, pp. 519–522.
- [11] S. C. Douglas and W. Pan, "Exact expectation analysis of the LMS adaptive filter," *IEEE Trans. Signal Process.*, vol. 43, no. 12, pp. 2863–2871, Dec. 1995.
- [12] H. J. Butterweck, "A steady-state analysis of the LMS adaptive algorithm without use of the independence assumption," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Detroit, MI, May 1995, vol. 2, pp. 1404–1407.
- [13] M. Rupp and H. J. Butterweck, "Overcoming the independence assumption in LMS filtering," in *Proc. 37th Asilomar Conf. Signals, Syst. Comput.*, Monterey, CA, Nov. 2003, pp. 607–611.
- [14] L. Ljung, *System Identification: Theory for the User*. T. Kailath, Ed. Englewood Cliffs, NJ: Prentice-Hall, 1987, Inf. Syst. Sci.
- [15] T. Söderström and P. Stoica, *System Identification*, M. J. Grimble, Ed. Cambridge, U.K.: Prentice-Hall, 1989, Syst. Control Eng.
- [16] P. Strobach, "Low-rank adaptive filters," *IEEE Trans. Signal Process.*, vol. 44, no. 12, pp. 2932–2947, Dec. 1996.
- [17] J. S. Goldstein and I. S. Reed, "Reduced-rank adaptive filtering," *IEEE Trans. Signal Process.*, vol. 45, no. 2, pp. 492–496, Feb. 1997.
- [18] J. S. Goldstein, I. S. Reed, and L. L. Scharf, "A multistage representation of the wiener filter based on orthogonal projections," *IEEE Trans. Inf. Theory*, vol. 44, no. 7, pp. 2943–2959, Nov. 1998.
- [19] B. Raghoeaman, D. Linebarger, and D. Begusija, "Analysis of low rank transform domain adaptive filtering algorithm," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 1999, vol. 4, pp. 1869–1872.
- [20] M. Honig and M. K. Tsatsanis, "Adaptive techniques for multiuser CDMA receivers," *IEEE Signal Process. Mag.*, vol. 17, no. 3, pp. 49–61, May 2000.
- [21] T. Gustafsson and B. D. Rao, "Statistical analysis of subspace-based estimation of reduced-rank linear regressors," *IEEE Trans. Signal Process.*, vol. 50, no. 1, pp. 151–159, Jan. 2002.
- [22] K. Werner and M. Jansson, "Reduced rank linear regression and weighted low rank approximations," *IEEE Trans. Signal Process.*, vol. 54, no. 6, pt. 1, pp. 2063–2075, Jun. 2006.
- [23] M. R. Petraglia and J. C. B. Torres, "Performance analysis of adaptive filter structure employing wavelet and sparse subfilters," *Inst. Elect. Eng. Proc.-Vis. Image Signal Process.*, vol. 149, no. 2, pp. 115–119, Apr. 2002.
- [24] H. L. V. Trees, *Detection, Estimation, and Modulation Theory – Part I*. New York: Wiley Intersci., 2001.
- [25] H. L. V. Trees, *Detection, Estimation, and Modulation Theory – Part II: Nonlinear Modulation Theory*. Hoboken, NJ: Wiley Interscience, 2003.
- [26] H. L. V. Trees, *Detection, Estimation, and Modulation Theory – Part III: Radar-Sonar Signal Processing and Gaussian Signals in Noise*. New York: Wiley Interscience, 2001.
- [27] H. L. V. Trees, *Detection, Estimation, and Modulation Theory – Part IV: Optimum Array Processing*. New York: Wiley Interscience, 2002.
- [28] S. Haykin, *Adaptive Filter Theory*, 2nd ed. Englewood Cliffs, NJ: Prentice-Hall, 1991.
- [29] D. G. Manolakis, V. K. Ingle, and S. M. Kogon, *Statistical and Adaptive Signal Processing*. Boston, MA: McGraw-Hill, 1999.
- [30] D. Han, S.-P. Kim, and J. C. Principe, "Sparse channel estimation with regularization method using convolution inequality for entropy," in *Proc. 2005 IEEE Int. Joint Conf. Neural Networks*, Quebec, Canada, 2005, vol. 4, pp. 2359–2362.



Neil J. Bershad (S'60–M'62–SM'81–F'88) received the B.E.E. degree from Rensselaer Polytechnic Institute, Troy, NY, in 1958, the M.S. degree in electrical engineering from the University of Southern California, Los Angeles, CA in 1960, and the Ph.D. degree in electrical engineering from Rensselaer Polytechnic Institute in 1962.

He joined the Faculty of the Henry Samueli School of Engineering, University of California, Irvine, in 1966, where he is now an Emeritus Professor of Electrical Engineering and Computer Science. His

research interests have involved stochastic systems modeling and analysis. His recent interests are in the area of stochastic analysis of adaptive filters. He has published a significant number of papers on the analysis of the stochastic behavior of various configurations of the LMS adaptive filter. His present research interests include the statistical learning behavior of adaptive filter structures for echo cancellation, active acoustic noise cancellation, and variable gain (μ) adaptive algorithms.

Dr. Bershad has served as an Associate Editor of the IEEE TRANSACTIONS ON COMMUNICATIONS in the area of phase-locked loops and synchronization. More recently, he was an Associate Editor of the IEEE TRANSACTIONS ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING in the area of adaptive filtering



José Carlos M. Bermudez (S'78–M'85–SM'02) received the B.E.E. degree from the Federal University of Rio de Janeiro (UFRJ), Rio de Janeiro, Brazil, the M.Sc. degree in electrical engineering from COPPE/UFRJ, and the Ph.D. degree in electrical engineering from Concordia University, Montreal, QC, Canada, in 1978, 1981, and 1985, respectively.

He joined the Department of Electrical Engineering, Federal University of Santa Catarina (UFSC), Florianópolis, Brazil, in 1985, where he is currently a Professor of electrical engineering. In the

winter of 1992, he was a Visiting Researcher with the Department of Electrical Engineering, Concordia University. In 1994, he was a Visiting Researcher with the Department of Electrical Engineering and Computer Science, University of California, Irvine (UCI). His research interests have involved analog signal processing using continuous-time and sampled-data systems. His recent research interests are in digital signal processing, including linear and nonlinear adaptive filtering, active noise and vibration control, echo cancellation, image processing, and speech processing.

Prof. Bermudez served as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING in the area of adaptive filtering from 1994 to 1996 and from 1999 to 2001, and as the Signal Processing Associate Editor for the *Journal of the Brazilian Telecommunications Society* (2005–2006). He was a member of the Signal Processing Theory and Methods Technical Committee of the IEEE Signal Processing Society from 1998 to 2004. He is currently an Associate Editor for the *EURASIP Journal on Advances in Signal Processing*.



Jean-Yves Tournet (M'94) received the Ingénieur degree in electrical engineering from Ecole Nationale Supérieure d'Electronique, d'Electrotechnique, d'Informatique et d'Hydraulique in Toulouse (ENSEIHT), France, in 1989 and the Ph.D. degree from the National Polytechnic Institute, Toulouse, France, in 1992.

He is currently a Professor in the University of Toulouse (ENSEIHT), France, and a member of the IRIT laboratory (UMR 5505 of the CNRS). His research activities are centered around statistical

signal processing with a particular interest to Markov chain Monte Carlo methods.