EURASIP Journal on Applied Signal Processing 2004:17, 2715–2722

Performance Analysis of Adaptive Volterra Filters in the Finite-Alphabet Input Case

Hichem Besbes
Ecole Supérieure des Communications de Tunis (Sup'Com), Ariana 2083, Tunisia
Email: hichem.besbes@supcom.rnu.tn

Mériem Jaïdane
Ecole Nationale d'Ingénieurs de Tunis (ENIT), Le Belvedere 1002, Tunisia
Email: meriem.jaidane@enit.rnu.tn

Jelel Ezzine
Ecole Nationale d'Ingénieurs de Tunis (ENIT), Le Belvedere 1002, Tunisia
Email: jelel.ezzine@enit.rnu.tn

Received 15 September 2003; Revised 21 May 2004; Recommended for Publication by Fulvio Gini

This paper deals with the analysis of adaptive Volterra filters, driven by the LMS algorithm, in the finite-alphabet inputs case. A tailored approach for the input context is presented and used to analyze the behavior of this nonlinear adaptive filter. Complete and rigorous mean square analysis is provided without any constraining independence assumption. Exact transient and steady-state performances expressed in terms of critical step size, rate of transient decrease, optimal step size, excess mean square error in stationary mode, and tracking nonstationarities are deduced.

Keywords and phrases: adaptive Volterra filters, LMS algorithm, time-varying channels, finite-alphabet inputs, exact performance analysis.

1. INTRODUCTION

Adaptive systems have been extensively designed and implemented in the area of digital communications. In particular, nonlinear adaptive filters, such as adaptive Volterra filters, have been used to model nonlinear channels encountered in satellite communications applications [1, 2]. The nonlinearity is essentially due to the high-power amplifier used in the transmission [3]. When dealing with land-mobile satellite systems, the channels are time varying and can be modeled by a general Mth-order Markovian model to describe these variations [4]. Hence, to take into account the effect of the amplifier's nonlinearity and channel variations, one can model the equivalent baseband channel by a time-varying Volterra filter. In this paper, we analyze the behavior and parameters tracking capabilities of adaptive Volterra filters, driven by the generic LMS algorithm.

In the literature, convergence analysis of adaptive Volterra filters is generally carried out for small adaptation step size [5]. In addition, a Gaussian input assumption is used in order to take advantage of the Price theorem results. However, from a practical viewpoint, to maximize the rate of convergence or to determine the critical step size, one needs a theory that is valid for large adaptation step size range. To the best knowledge of the authors, no such exact theory exists for adaptive Volterra filters. It is important to note that the so-called independence assumption, well known of being a crude approximation for large step size range, is behind all available results [6].

The purpose of this paper is to provide an approach tailored for the finite-alphabet input case. This situation is frequently encountered in many digital transmission systems. In fact, we develop an exact convergence analysis of adaptive Volterra filters, governed by the LMS algorithm. The proposed analysis, pertaining to the large step size case, is derived without any independence assumption. Exact transient and steady-state performances, that is, critical step size, rate of transient decrease, optimal step size, excess mean square error (EMSE), and tracking capability, are provided.

The paper is organized as follows. In the second section, we provide the needed background for the analysis of adaptive Volterra filters. In the third section, we present the signal input model. In the fourth section, we develop the proposed approach to analyze the adaptive Volterra filter. Finally, the fifth section presents some simulation results to validate the proposed approach.

2. BACKGROUND

The FIR Volterra filter's output may be characterized by a truncated Volterra series consisting of q convolutional terms. The baseband model of the nonlinear time-varying channel is described as follows:

$$y_k = \sum_{m=1}^{q} \sum_{i_1=0}^{L-1} \sum_{i_2 \geq i_1}^{L-1} \cdots \sum_{i_m \geq i_{m-1}}^{L-1} f_k^m(i_1, \ldots, i_m) \times x_{k-i_1} \cdots x_{k-i_m} + n_k, \tag{1}$$

where x_k is the input signal, and n_k is the observation noise, assumed to be i.i.d and zero mean. In the above equation, q is the Volterra filter order, L is the memory length of the filter, and $f_m^k(i_1, \ldots, i_m)$ is a complex number, referred to as the mth-order Volterra kernel. This latter complex number may be a time-varying parameter.

The Volterra observation vector $\hat{X}_k$ is defined by

$$\hat{X}_k = [x_k, \ldots, x_{k-L+1}, x_k^2, x_k x_{k-1}, \ldots, x_k x_{k-L+1}, x_{k-1}^2, \ldots, x_{k-L+1}^q]^T, \tag{2}$$

where only one permutation of each product $x_{i_1} x_{i_2} \cdots x_{i_m}$ appears in $\hat{X}_k$. It is well known [7] that the dimension of the Volterra observation vector is $\beta = \sum_{m=1}^{q} \binom{L+m-1}{m}$.

The input/output recursion, corresponding to the above model, can then be rewritten in the following linear form:

$$y_k = \hat{X}_k^T F_k + n_k, \tag{3}$$

where $F_k = [f_k^1(0), \ldots, f_k^1(L-1), f_k^2(0,0), f_k^2(0,1), \ldots, f_k^q(L-1, \ldots, L-1)]^T$ is a vector containing all the Volterra kernels.

In this paper, we assume that the evolution of F_k is governed by an Mth-order Markovian model

$$F_{k+1} = \sum_{i=1}^{M} \Lambda_i F_{k-i+1} + \Omega_k, \tag{4}$$

where the Λ_i $(i = 1, \ldots, M)$ are matrices which characterize the behavior of the channel. $\Omega_k = [\omega_{1k}, \omega_{2k}, \ldots, \omega_{\beta k}]^T$ is an unknown zero-mean process, which characterizes the nonstationarity of the channel. It is to be noted that process $\{\Omega_k\}$ is independent of the input $\{\hat{X}_k\}$ as well as the observation noise $\{n_k\}$.

In this paper, we consider the identification problem of this time-varying nonlinear channel. To wit, an adaptive Volterra filter driven by the LMS algorithm is considered. This analysis is general, and therefore includes the stationary case, that is, $\Omega_k = 0$, as well as the linear case, that is, $q = 1$.

The coefficient update of the adaptive Volterra filter is given by

$$\begin{aligned} y_k^e &= \hat{X}_k^T G_k, \\ e_k &= y_k - y_k^e, \\ G_{k+1} &= G_k + \mu e_k \hat{X}_k^*, \end{aligned} \tag{5}$$

where y_k^e is the output estimate, G_k is the vector of (nonlinear) filter coefficients at time index k, μ is a positive step size, and $(\cdot)^*$ stands for the complex conjugate operator. Moreover, we assume that the channel and the Volterra filter have the same length.

By considering the deviation vector V_k, that is, the difference between the adaptive filter coefficients vector G_k and the optimum parameters vector F_k, that is, $V_k = G_k - F_k$, the behavior of the adaptive filter and the channel variations can be usefully described by an augmented vector Φ_k defined as

$$\Phi_k = [F_k^T, F_{k-1}^T, \ldots, F_{k-M+1}^T, V_k^T]^T. \tag{6}$$

From (3)–(6), it is readily seen that one can deduce that the dynamics of the augmented vector are described by the following linear time-varying recursion:

$$\Phi_{k+1} = C_k \Phi_k + B_k, \tag{7}$$

where

$$C_k = \begin{bmatrix} \Lambda_1 & \Lambda_2 & \cdots & \Lambda_{M-1} & \Lambda_M & 0 \\ I_{(\beta)} & 0 & & \cdots & 0 & 0 \\ 0 & I_{(\beta)} & 0 & \cdots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & I_{(\beta)} & 0 & 0 \\ I_{(\beta)} - \Lambda_1 & -\Lambda_2 & \cdots & -\Lambda_{M-1} & -\Lambda_M & I_{(\beta)} - \mu \hat{X}_k^* \hat{X}_k^T \end{bmatrix},$$

$$B_k = \begin{bmatrix} \Omega_k \\ 0 \\ 0 \\ \vdots \\ 0 \\ -\Omega_k + \mu n_k \hat{X}_k^* \end{bmatrix}, \tag{8}$$

and $I_{(\beta)}$ is the identity matrix with dimension β.

Note that V_k is deduced from Φ_k by the following simple relationship:

$$V_k = U \Phi_k, \quad U = \begin{bmatrix} 0_{(\beta, M\beta)} & I_{(\beta)} \end{bmatrix}, \tag{9}$$

where $0_{(l,m)}$ is a zero matrix with l rows and m columns.

The behavior of the adaptive filter can be described by the evolution of the mean square deviation (MSD) defined by

$$\text{MSD} = E(V_k^H V_k), \tag{10}$$

where $(\cdot)^H$ is the transpose of the complex conjugate of $(\cdot)$ and $E(\cdot)$ is the expectation operator. To evaluate the MSD, we must analyze the behavior of $E(\Phi_k \Phi_k^H)$. Since Ω_k and n_k are zero mean and independent of $\hat{X}_k$ and Φ_k, the nonhomogeneous recursion between $E(\Phi_{k+1} \Phi_{k+1}^H)$ and $E(\Phi_k \Phi_k^H)$ is given by

$$E(\Phi_{k+1} \Phi_{k+1}^H) = E(C_k \Phi_k \Phi_k^H C_k^H) + E(B_k B_k^H). \tag{11}$$

From the analysis of this recursion, all mean square performances in transient and in steady states of the adaptive Volterra filter can be deduced. However, (11) is hard to solve. In fact, since $\widehat{X}_k$ and $\widehat{X}_{k-1}$ are sharing $L-1$ components, they are dependent. Thus, C_k and C_{k-1} are dependent, which means that Φ_k and C_k are dependent as well. Hence, (11) becomes difficult to solve. It is important to note that even when using the independence assumption between C_k and Φ_k, equation (11) is still hard to solve due to its structure.

In order to overcome these difficulties, Kronecker products are required. Indeed, after transforming the matrix $\Phi_k\Phi_k^H$ to an augmented vector, by applying the $\mathrm{vec}(\cdot)$ linear operator, which transforms a matrix to an augmented vector, and by using some properties of tensorial algebra [8], that is, $\mathrm{vec}(ABC) = (C^T \otimes A)\,\mathrm{vec}(B)$, as well as the commutativity between the expectation and the $\mathrm{vec}(\cdot)$ operator, that is, $\mathrm{vec}(E(M)) = E(\mathrm{vec}(M))$, (11) becomes

$$E\left(\mathrm{vec}\left(\Phi_{k+1}\Phi_{k+1}^H\right)\right) = E\left(\left(C_k^* \otimes C_k\right)\mathrm{vec}\left(\Phi_k\Phi_k^H\right)\right) + E\left(\mathrm{vec}\left(B_kB_k^H\right)\right), \tag{12}$$

where $\otimes$ stands for the Kronecker product [8].

It is important to note that due to the difficulty of the analysis, few concrete results were obtained until now [9, 10]. When the input signal is correlated, and even in the linear case, the analysis is usually carried out for a first-order Markov model and a small step size [11, 12]. For a small step size, an independence assumption is made between C_k and Φ_k, which leads to a simplification of (12),

$$E\left(\mathrm{vec}\left(\Phi_{k+1}\Phi_{k+1}^H\right)\right) = E\left(\left(C_k^* \otimes C_k\right)\right)E\left(\mathrm{vec}\left(\Phi_k\Phi_k^H\right)\right) + E\left(\mathrm{vec}\left(B_kB_k^H\right)\right). \tag{13}$$

Equation (13) becomes a linear equation, and can be solved easily. However, the obtained results which are based on the independence assumption, are valid only for small step sizes.

The aim of this paper is to propose a valid approach to solve (12) for all step sizes, that is, from the range of small step sizes to the range of large step sizes, including the optimal and critical step sizes. To do so, we consider the case of baseband channel identification, where the input signal is a symbol sequence belonging to a finite-alphabet set.

3. ANALYSIS OF ADAPTIVE VOLTERRA FILTERS: THE FINITE-ALPHABET CASE

3.1. Input signal model

In digital transmission contexts, when dealing with baseband channel identification, the input signal x_k represents the transmitted symbols during a training phase. These symbols are known by the transmitter and by the receiver. The input signal belongs to a finite-alphabet set $S = \{a_1, a_2, \ldots, a_d\}$ with cardinality d, such as PAM, QAM, and so forth. For example, if we consider a BPSK modulation case, the transmitted sequence x_k belongs to $S = \{-1, +1\}$. Assuming that $\{x_k\}$ is an i.i.d. sequence, then x_k can be represented by an irreducible discrete-time Markov chain with finite states $\{1, 2\}$, and a probability transition matrix $P = \left[\begin{smallmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{smallmatrix}\right]$. This model for the transmitted signal is widely used, especially for the performance analysis of trellis-coded modulation techniques [13].

Consequently, the Volterra observation vector $\widehat{X}_k$ remains also in a finite-alphabet set

$$\mathcal{A} = \{\widehat{W}_1, \widehat{W}_2, \ldots, \widehat{W}_N\} \tag{14}$$

with cardinality $N = d^L$. Thus, the matrix C_k, defined in (8) and which governs the adaptive filter, belongs also to a finite-alphabet set

$$\mathcal{C} = \{\Psi_1, \ldots, \Psi_N\}, \tag{15}$$

where

$$\Psi_i = \begin{bmatrix} \Lambda_1 & \Lambda_2 & \cdots & \Lambda_{M-1} & \Lambda_M & 0 \\ I_{(\beta)} & 0 & & \cdots & 0 & 0 \\ 0 & I_{(\beta)} & 0 & \cdots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & I_{(\beta)} & 0 & 0 \\ I_{(\beta)} - \Lambda_1 & -\Lambda_2 & \cdots & -\Lambda_{M-1} & -\Lambda_M & I_{(\beta)} - \mu\widehat{W}_i^*\widehat{W}_i^T \end{bmatrix}. \tag{16}$$

As a result, the matrix C_k can be modeled as an irreducible discrete-time Markov chain $\{\theta(k)\}$ with finite state space $\{1, 2, \ldots, N\}$ and probability transition matrix $P = [p_{ij}]$, such that

$$C_k = \Psi_{\theta(k)}. \tag{17}$$

By using the proposed model of the input signal, we will analyze the convergence of the adaptive filter in the next subsection.

3.2. Exact performance evaluation

The main idea used to tackle (11), in the finite-alphabet input case, is very simple. Since there are N possibilities for $\Psi_{\theta(k)}$, we may analyze the behavior of $E(\Phi_k\Phi_k^H)$ through the following quantity, denoted by $Q_j(k)$, $j = 1, \ldots, N$, and defined by

$$Q_j(k) = E\left(\mathrm{vec}\left(\Phi_k\Phi_k^H 1_{(\theta(k)=j)}\right)\right), \tag{18}$$

where $1_{(\theta(k)=j)}$ stands for the indicator function, which is equal to 1 if $\theta(k) = j$ and is equal to 0 otherwise.

It is interesting to recall that at time k, $\Psi_{\theta(k)}$ can have only one value among the N possibilities, which means that

$$\sum_{j=1}^{N} 1_{(\theta(k)=j)} = 1. \tag{19}$$

From the last equation, it is easy to establish the relationship between $E(\Phi_k\Phi_k^H)$ and $Q_j(k)$. In fact, we have

$$\begin{aligned}\operatorname{vec}\left(E(\Phi_k\Phi_k^H)\right) &= \operatorname{vec}\left(E\left(\Phi_k\Phi_k^H\sum_{j=1}^{N}1_{(\theta(k)=j)}\right)\right)\\ &= \sum_{j=1}^{N}E\left(\operatorname{vec}\left(\Phi_k\Phi_k^H 1_{(\theta(k)=j)}\right)\right)\\ &= \sum_{j=1}^{N}Q_j(k).\end{aligned} \tag{20}$$

Therefore, we can conclude that the LMS algorithm converges if and only if all of the $Q_j(k)$ converge.

The recursive relationship between $Q_j(k+1)$ and all the $Q_i(k)$ can be established as follows:

$$\begin{aligned}Q_j(k+1) &= E\left(\operatorname{vec}\left(\Phi_{k+1}\Phi_{k+1}^H 1_{(\theta(k+1)=j)}\right)\right)\\ &= E\left(\left(C_k^*\otimes C_k\right)\operatorname{vec}\left(\Phi_k\Phi_k^H\right)1_{(\theta(k+1)=j)}\right)\\ &\quad + E\left(\operatorname{vec}(B_kB_k^H)1_{(\theta(k+1)=j)}\right)\\ &= \sum_{i=1}^{N}E\left(\left(C_k^*\otimes C_k\right)\operatorname{vec}\left(\Phi_k\Phi_k^H\right)1_{(\theta(k+1)=j)}1_{(\theta(k)=i)}\right)\\ &\quad + \sum_{i=1}^{N}E\left(\operatorname{vec}(B_kB_k^H)1_{(\theta(k+1)=j)}1_{(\theta(k)=i)}\right).\end{aligned} \tag{21}$$

In order to overcome the difficulty of the analysis found in the general context, we take into account the properties induced by the input characteristics, namely,

(1) C_k belongs to a finite-alphabet set

$$C_k 1_{(\theta(k)=i)} = \Psi_i 1_{(\theta(k)=i)}, \tag{22}$$

(2) Ψ_i are constant matrices independent of Φ_k.

Hence, the dependence difficulty found in (12) is avoided, and one can deduce that

$$\begin{aligned}Q_j(k+1) &= \sum_{i=1}^{N}(\Psi_i^*\otimes\Psi_i)E\left(\operatorname{vec}(\Phi_k\Phi_k^H)1_{(\theta(k+1)=j)}1_{(\theta(k)=i)}\right)\\ &\quad + \sum_{i=1}^{N}E\left(\operatorname{vec}(B_kB_k^H)1_{(\theta(k+1)=j)}1_{(\theta(k)=i)}\right)\\ &= \sum_{i=1}^{N}p_{ij}(\Psi_i^*\otimes\Psi_i)E\left(\operatorname{vec}\left(\Phi_k\Phi_k^H\right)1_{(\theta(k)=i)}\right)\\ &\quad + \sum_{i=1}^{N}p_{ij}E\left(\operatorname{vec}\left(B_kB_k^H\right)1_{(\theta(k)=i)}\right)\\ &= \sum_{i=1}^{N}p_{ij}(\Psi_i^*\otimes\Psi_i)Q_i(k)+\Gamma_j,\end{aligned} \tag{23}$$

where

$$\Gamma_j \triangleq \sum_{i=1}^{N}p_{ij}E\left(\operatorname{vec}\left(B_kB_k^H\right)1_{(\theta(k)=i)}\right). \tag{24}$$

From (18)–(24), along the same lines as in the linear case [10, 14], and by expressing the recursion between $Q_j(k+1)$ and the remaining $Q_i(k)$, we have proven, without any constraining independence assumption on the observation vector, that the terms $Q_j(k+1)$ satisfy the following exact and compact recursion:

$$\widetilde{Q}(k+1) = \Delta\widetilde{Q}(k)+\Gamma, \tag{25}$$

where $\widetilde{Q}(k) = [Q_1(k)^T,\ldots,Q_N(k)^T]^T$. The matrix Δ is defined by

$$\Delta \triangleq \left(P^T\otimes I_{((M+1)\beta^2)}\right)\operatorname{Diag}_\Psi, \tag{26}$$

where Diag_Ψ denotes a block diagonal matrix defined by

$$\operatorname{Diag}_\Psi = \begin{bmatrix}\Psi_1^*\otimes\Psi_1 & 0 & 0 & \cdots & 0\\ 0 & \Psi_2^*\otimes\Psi_2 & 0 & \cdots & 0\\ \vdots & \ddots & \ddots & \ddots & \vdots\\ 0 & \cdots & 0 & \Psi_{N-1}^*\otimes\Psi_{N-1} & 0\\ 0 & 0 & \cdots & 0 & \Psi_N^*\otimes\Psi_N\end{bmatrix}. \tag{27}$$

The vector Γ depends on the power of the observation noise and the input statistics and is defined by

$$\Gamma \triangleq [\Gamma_1^T,\ldots,\Gamma_N^T]^T \in \mathbb{C}^{N((M+1)\beta)^2}. \tag{28}$$

The compact linear and deterministic equation (25) will replace (11). From (25), we will deduce all adaptive Volterra filter performances.

3.3. Convergence conditions

Since the recursion (25) is linear, the convergence of the LMS is simply deduced from the analysis of the eigenvalues of Δ. We assume that the general Markov model (4) describing the channel behavior is stable, the algorithm stability can then be deduced from the stationary case, where $M = 1$, $\Omega_k = 0$, and $\Lambda_1 = I$. In this case, since F_k is constant, we choose $\Phi_k = V_k$ to analyze the behavior of the algorithm. Hence,

$$\Psi_i = I - \mu\widehat{W}_i^*\widehat{W}_i^T. \tag{29}$$

3.3.1. Excitation condition

Proposition 1. *The LMS algorithm converges only if the alphabet set $\mathcal{A} = \{\widehat{W}_1, \widehat{W}_2,\ldots,\widehat{W}_N\}$ spans the space $\mathbb{C}^\beta$.*

Physically, this condition means that, in order to converge to the optimal solution, we have to excite the algorithm in all directions which spans the space.

Proof. If the alphabet set does not span the space, we can find a nonzero vector, z, orthogonal to the alphabet set, and by constructing an augmented vector

$$Z = [z^H, \ldots, z^H, z^H, \ldots, z^H]^H, \tag{30}$$

it is easy to show that $\Delta Z = Z$, and so the matrix Δ has an eigenvalue equal to one. □

Proposition 2. *The set $\mathcal{A} = \{\widehat{W}_1, \widehat{W}_2, \ldots, \widehat{W}_N\}$ spans the space $\mathbb{C}^\beta$ only if the cardinality d of the alphabet $S = \{a_1, a_2, \ldots, a_d\}$ is greater than the order q of the Volterra filter nonlinearity.*

This can be explained by rearranging the rows of $\mathcal{W} = [\widehat{W}_1, \widehat{W}_2, \ldots, \widehat{W}_N]$ such that the first rows correspond to the memoryless case. We denote this matrix by

$$\widetilde{\mathcal{W}} = \begin{bmatrix} a_1 & a_2 & \cdots & a_d & \cdots & a_1 & \cdots & a_d \\ a_1^2 & a_2^2 & \cdots & a_d^2 & \cdots & a_1^2 & \cdots & a_d^2 \\ \vdots & \vdots & & \vdots & & \vdots & & \vdots \\ a_1^q & a_2^q & \cdots & a_d^q & \cdots & a_1^q & \cdots & a_d^q \end{bmatrix}. \tag{31}$$

This matrix is a Vandermonde matrix, and it is full rank if and only if $d > q$, which proves the excitation condition.

It is easy to note that this result is similar to the one obtained in [7]. As a consequence of this proposition, we can conclude that we cannot use a QPSK signal ($d = 4$) to identify a Volterra with order $q = 5$.

3.3.2. Convergence condition

We provide, under the persistent excitation condition, a very useful sufficient critical step size in the following proposition.

Proposition 3. *If the Markov chain $\{\theta(k)\}$ is ergodic, the alphabet set $\mathcal{A} = \{\widehat{W}_1, \widehat{W}_2, \ldots, \widehat{W}_N\}$ spans the space $\mathbb{C}^\beta$, and the noise n_k is zero mean, i.i.d., sequence independent of X_k, then there exists a critical step size μ_c such that*

$$\mu_c \geq \mu_{cNL}^{\min} = \frac{2}{\max_{i=1,\ldots,N} \widehat{W}_i^H \widehat{W}_i}, \tag{32}$$

and if $\mu \leq \mu_c$, then the amplitude of Δ's eigenvalues are less than one, and the LMS algorithm converges exponentially in the mean square sense.

Proof. Using the tensorial algebra property $(A \otimes B)(C \otimes D) = (AC) \otimes (BD)$, the matrix $\Delta\Delta^H$ is given by

$$\begin{aligned} \Delta\Delta^H = {} & (P^T \otimes I_{\beta^2}) \\ & \times \operatorname{diag}\left(\left(I - \mu \widehat{W}_i \widehat{W}_i^H\right)^2 \otimes \left(I - \mu \widehat{W}_i^* \widehat{W}_i^T\right)^2 \right) (P \otimes I_{\beta^2}). \end{aligned} \tag{33}$$

It is interesting to note that the matrix $\operatorname{diag}((I - \mu \widehat{W}_i \widehat{W}_i^H)^2 \otimes (I - \mu \widehat{W}_i^* \widehat{W}_i^T)^2)$ is a nonnegative symmetric matrix. By denoting $\{D_j,\ j = 1, \ldots, N-1\}$, the set of vectors orthogonal to the vector W_i, the eigenvalues of the matrix $((I - \mu \widehat{W}_i \widehat{W}_i^H)^2 \otimes (I - \mu \widehat{W}_i^* \widehat{W}_i^T)^2)$ are as follows:

(i) $(1 - \mu W_i^H W_i)^4$ associated with the eigenvectors $W_i \otimes W_i^*$,

(ii) $(1 - \mu W_i^H W_i)^2$ associated with the eigenvectors $W_i \otimes D_j^*$,

(iii) $(1 - \mu W_i^H W_i)^2$ associated with the eigenvectors $D_j \otimes W_i^*$,

(iv) 1 associated with the eigenvectors $D_j \otimes D_l^*$.

So, for $\mu \leq 2/\max_{i=1,\ldots,N} \widehat{W}_i^H \widehat{W}_i$, the eigenvalues λ_i of $\operatorname{diag}((I - \mu \widehat{W}_i \widehat{W}_i^H)^2 \otimes (I - \mu \widehat{W}_i^* \widehat{W}_i^T)^2)$ satisfy

$$0 \leq \lambda_i \leq 1. \tag{34}$$

Assuming that the Markov chain $\{\theta(k)\}$ is ergodic, the probability transition matrix P is acyclic [15], and it has 1 as the unique largest amplitude eigenvalue, corresponding to the vector $u = [1, \ldots, 1]^T$. This means that for a nonzero vector R in $\mathbb{C}^{N\beta^2}$, $R^H (P^T \otimes I_{\beta^2})(P \otimes I_{\beta^2}) R = R^H R$ if and only if R has the following structure:

$$R = u \otimes e, \tag{35}$$

where e is a nonzero vector in $\mathbb{C}^{\beta^2}$.

Now, for any nonzero vector R in $\mathbb{C}^{N\beta^2}$, there are two possibilities:

(1) there exists an e in $\mathbb{C}^{\beta^2}$ such that $R = u \otimes e$,

(2) R does not have the structure described by (35).

In the first case, we can express $R^H \Delta\Delta^H R$ as follows:

$$\begin{aligned} R^H \Delta\Delta^H R = {} & (u^T \otimes e^H)(P^T \otimes I_{\beta^2}) \\ & \times \operatorname{diag}\left(\left(I - \mu \widehat{W}_i \widehat{W}_i^H\right)^2 \otimes \left(I - \mu \widehat{W}_i^* \widehat{W}_i^T\right)^2 \right) \\ & \times (P \otimes I_{\beta^2})(u \otimes e) \\ = {} & (u^T \otimes e^H) \\ & \times \operatorname{diag}\left(\left(I - \mu \widehat{W}_i \widehat{W}_i^H\right)^2 \otimes \left(I - \mu \widehat{W}_i^* \widehat{W}_i^T\right)^2 \right) (u \otimes e) \\ = {} & \sum_{i=1}^{N} e^H \left(\left(I - \mu \widehat{W}_i \widehat{W}_i^H\right)^2 \otimes \left(I - \mu \widehat{W}_i^* \widehat{W}_i^T\right)^2 \right) e. \end{aligned} \tag{36}$$

Since $\mathcal{A} = \{\widehat{W}_1, \widehat{W}_2, \ldots, \widehat{W}_N\}$ spans the space $\mathbb{C}^\beta$, it is easy to show that

$$\begin{aligned} & \sum_{i=1}^{N} e^H \left(\left(I - \mu \widehat{W}_i \widehat{W}_i^H\right)^2 \otimes \left(I - \mu \widehat{W}_i^* \widehat{W}_i^T\right)^2 \right) e \\ & \qquad < N e^H e = R^H R, \end{aligned} \tag{37}$$

which means

$$R^H \Delta\Delta R < R^H R. \tag{38}$$

In the second case, it is easy to show that

$$R^H \Delta\Delta^H R \leq R^H (P^T \otimes I_{\beta^2})(P \otimes I_{\beta^2}) R. \tag{39}$$

This is due to the fact that Diag_Ψ is a symmetric nonnegative matrix, with largest eigenvalue equal to one.

Now, using the fact that R does not have the structure (35), this leads to

$$R^H \Delta \Delta^H R < R^H R. \tag{40}$$

If we resume the two cases, we conclude that for any nonnegative vector R in $\mathbb{C}^{N\beta^2}$,

$$\frac{R^H \Delta \Delta^H R}{R^H R} < 1, \tag{41}$$

which concludes the proof. □

It is interesting to note that when the input signal is a PSK signal, which has a constant modulus, all the quantities $2/\widehat{W}_i^H \widehat{W}_i$ are equal and thus they are also equal to the exact critical step size.

Moreover, in the general case, the exact critical step size μ_c and the optimum step size μ_{opt} for convergence are deduced by the analysis of the Δ eigenvalues as a function of μ. These important quantities depend on the transmitted alphabet and on the transition matrix P.

3.4. Steady-state performances

If the convergence conditions are satisfied, we determine the steady-state performances ($k \to \infty$) by

$$\widetilde{Q}_\infty = (I - \Delta)^{-1} \Gamma. \tag{42}$$

From $\lim_{k\to\infty} Q_i(k)$, and using the relationship (9) between V_k and Φ_k, we deduce that

$$\lim_{k\to\infty} E\big(\operatorname{vec}\left(V_k V_k^H 1_{(\theta(k)=i)}\right)\big) = (U \otimes U) \lim_{k\to\infty} Q_i(k), \tag{43}$$

and thus the exact value of MSD. In the same manner, we can compute the exact EMSE:

$$\begin{aligned} \text{EMSE} &= E(|y_k - y_k^e|^2) - E(|n_k|^2) \\ &= E(|\hat{X}_k^T V_k|^2) \\ &= E(\hat{X}_k^T V_k V_k^H \hat{X}_k^*) \\ &= E\big((\hat{X}_k^H \otimes \hat{X}_k^T) \operatorname{vec}(V_k V_k^H)\big). \end{aligned} \tag{44}$$

Using the relationship (9) between V_k and Φ_k, we can develop the EMSE as follows:

$$\begin{aligned} \text{EMSE} &= E\big((\hat{X}_k^H \otimes \hat{X}_k^T) \operatorname{vec}(U \Phi_k \Phi_k^H U^T)\big) \\ &= E\big((\hat{X}_k^H \otimes \hat{X}_k^T)(U \otimes U) \operatorname{vec}(\Phi_k \Phi_k^H)\big) \\ &= E\left((\hat{X}_k^H \otimes \hat{X}_k^T)(U \otimes U) \operatorname{vec}(\Phi_k \Phi_k^H) \sum_{i=1}^{N} 1_{\theta(k)=i}\right) \\ &= \sum_{i=1}^{N} E\big((\hat{X}_k^H \otimes \hat{X}_k^T)(U \otimes U) \operatorname{vec}(\Phi_k \Phi_k^H) 1_{\theta(k)=i}\big) \\ &= \sum_{i=1}^{N} E\big((\widehat{W}_i^H \otimes \widehat{W}_i^T)(U \otimes U) \operatorname{vec}(\Phi_k \Phi_k^H) 1_{\theta(k)=i}\big) \\ &= \sum_{i=1}^{N} (\widehat{W}_i^H \otimes \widehat{W}_i^T)(U \otimes U) E\big(\operatorname{vec}(\Phi_k \Phi_k^H) 1_{\theta(k)=i}\big). \end{aligned} \tag{45}$$

Under the convergence conditions, $E(\operatorname{vec}(\Phi_k \Phi_k^H) 1_{\theta(k)=i})$ converges to $\lim_{k\to\infty} Q_i(k)$, the mean square error (MSE) can be given by

$$\text{MSE} = \sum_{i=1}^{N} (\widehat{W}_i^H \otimes \widehat{W}_i^T)(U \otimes U) \lim_{k\to\infty} Q_i(k) + E\left(|n_k|^2\right). \tag{46}$$

In this section, we have proven that without using any unrealistic assumptions, we can compute the exact values of the MSD and the MSE.

It is interesting to note that the proposed approach remains valid even when the model order of the adaptive Volterra filter is overestimated, which means that the nonlinearity order and/or the memory length of the adaptive Volterra filter are greater than the real system to be identified. In fact, in this case the observation noise is still independent of the input signal, and the used assumptions remain valid. Indeed, this case is equivalent to identifying some coefficients which are set to zero. Of course, this will decrease the rate of convergence, and increase the MSE at the steady state.

In the next section, we will confirm our analysis through a study case.

4. SIMULATION RESULTS

The exact analysis of adaptive Volterra filters made for the finite-alphabet input case is illustrated in this section. We consider a case study, where we want to identify a nonlinear time-varying channel, modeled by a time-varying Volterra filter. The transmitted symbols are i.i.d. and belong to a QPSK constellation, that is, $x_k \in \{1+j, 1-j, -1+j, -1-j\}$ (where $j^2 = -1$). In this case, we have

$$\text{Prob}\,(x_{k+1}|x_k) = \text{Prob}\,(x_{k+1}) = \frac{1}{4}, \tag{47}$$

and x_k can be modeled by a discrete-time Markov chain with transition matrix equal to

$$P_x = \begin{bmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{bmatrix}. \tag{48}$$

In this example, we assume that the channel is modeled as follows:

$$\begin{aligned} y_k = {} & f_0(k) x_k + f_1(k) x_{k-1} + f_2(k) x_k^2 x_{k-1} \\ & + f_3(k) x_k x_{k-1}^2 + n_k. \end{aligned} \tag{49}$$

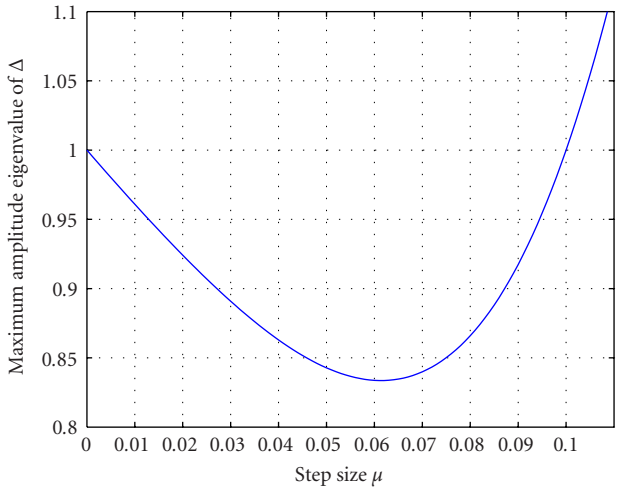

FIGURE 1: Evolution of the Δ's maximum eigenvalue versus the step size.

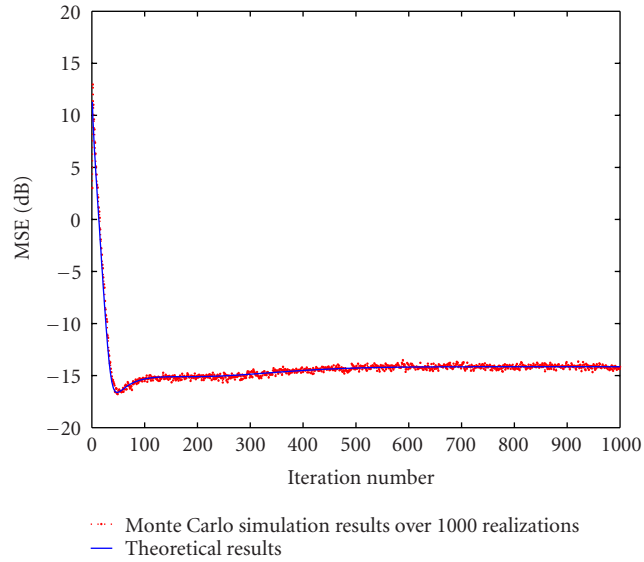

FIGURE 2: Transient behavior of the adaptive Volterra filter: the evolution of MSE.

The observation noise n_k is assumed to be i.i.d complex Gaussian with power $E(|n_k|^2) = 0.001$. The parameters vector $F_k = [f_0(k), f_1(k), f_2(k), f_3(k)]^T$ is assumed to be time varying, and its variations are described by a second-order Markovian model

$$F_{k+1} = 2\gamma\cos(\alpha)F_k - \gamma^2 F_{k-1} + \Omega_k, \tag{50}$$

where $\gamma = 0.995$, $\alpha = \pi/640$, and Ω_k is a complex Gaussian, zero mean, i.i.d., spatially independent, and with components power $E(|\omega_k|^2) = 10^{-6}$.

We assume that the adaptive Volterra filter has the same length as the channel model. In this case, the input observation vector is equal to $\widehat{X}_k = [x_k, x_{k-1}, x_k^2 x_{k-1}, x_k x_{k-1}^2]^T$, and it belongs to a finite-alphabet set with cardinality equal to 16, which is the number of all x_k and x_{k-1} combinations.

The sufficient critical step size computed using (32) is equal to $\mu_{cNL}^{\min} = 1/10$. To analyze the effect of the step size on the convergence rate of the algorithm, we report in Figure 1 the evolution of the largest absolute value of the eigenvalues of Δ, we deduce that

(i) the critical step size μ_c, deduced from the finite-alphabet case, corresponding to $\lambda_{\max}(\Delta) = 1$ is equal to $\mu_c = 0.100$, which has the same value as $\mu_{cNL}^{\min} = 1/10$. This result is expected since the amplitude of the input data x_k is constant;

(ii) the optimal step μ_{opt}, corresponding to the minimum value of $\lambda_{\max}(\Delta)$, is $\mu_{\text{opt}} = 0.062$. The optimal rate of convergence is found to be

$$\min_{\mu} \lambda_{\max}(\Delta) = 0.830. \tag{51}$$

In order to evaluate the evolution of the EMSE versus the iteration number, we compute the recursion (25), and we run a Monte Carlo simulation over 1000 realizations, for $\mu = 0.06$, for an initial deviation vector $V_0 = [1, 1, 1, 1]^T$,

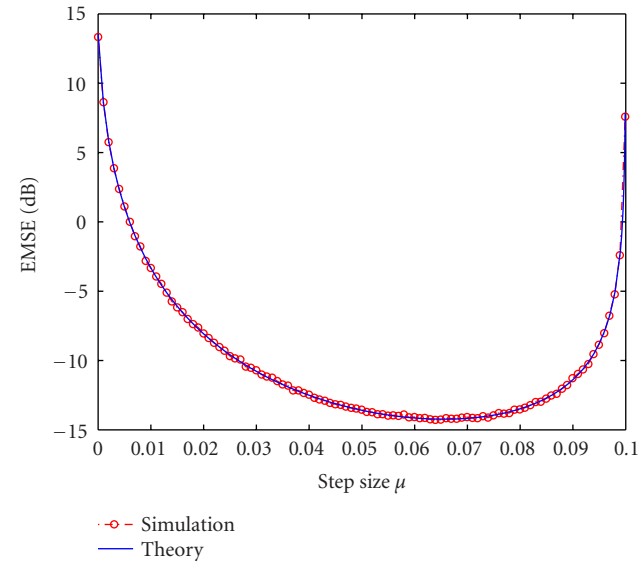

FIGURE 3: Variations of the EMSE versus μ in a nonstationary case.

and for an initial value of the channel parameters vector $F_0 = [0, 0, 0, 0]^T$. Figure 2 shows the superposition of the simulation results with the theoretical ones.

Figure 3 shows the variations of the EMSE at the convergence, versus the step size, which varies from 0.001 to 0.100. The simulation results are obtained by averaging over 100 realizations.

The simulations of transient and steady-state performances are in perfect agreement with the theoretical analysis. Note from Figure 3 the degradation of the tracking capabilities of the algorithm for small step size. The optimum step size is high, and it cannot be deduced from classical analysis.

5. CONCLUSION

In this paper, we have presented an exact and complete theoretical analysis of the generic LMS algorithm used for the identification of time-varying Volterra structures. The proposed approach is tailored for the finite-alphabet input case, and it was carried out without using any unrealistic independence assumptions. It reflects the exactness of the obtained performances in transient and in steady cases of the adaptive nonlinear filter. All simulations of transient and tracking capabilities are in perfect agreement with our theoretical analysis. Exact and practical bounds on the critical step size and optimal step size for tracking capabilities are provided, which can be helpful in a design context. The exactness and the elegance of the proof are due to the input characteristics, which is commonly used in the digital communications context.

REFERENCES

[1] V. J. Mathews, "Adaptive polynomial filters," *IEEE Signal Processing Magazine*, vol. 8, no. 3, pp. 10–26, 1991.

[2] S. Benedetto, E. Biglieri, and V. Castellani, *Digital Transmission Theory*, Prentice Hall, Englewood Cliffs, NJ, USA, 1987.

[3] H. Besbes, T. Le-Ngoc, and H. Lin, "A fast adaptive polynomial predistorter for power amplifiers," in *Proc. IEEE Global Telecommunications Conference (GLOBECOM '01)*, vol. 1, pp. 659–663, San Antonio, Tex, USA, November 2001.

[4] S. Ohmori, H. Wakana, and S. Kawase, *Mobile Satellite Communications*, Artech House Publishers, Boston, Mass, USA, 1998.

[5] T. Koh and J. E. Powers, "Second-order Volterra filtering and its application to nonlinear system identification," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 33, no. 6, pp. 1445–1455, 1985.

[6] M. V. Dokic and P. M. Clarkson, "On the performance of a second-order adaptive Volterra filter," *IEEE Trans. Signal Processing*, vol. 41, no. 5, pp. 1944–1947, 1993.

[7] R. D. Nowak and B. D. Van Veen, "Random and pseudorandom inputs for Volterra filter identification," *IEEE Trans. Signal Processing*, vol. 42, no. 8, pp. 2124–2135, 1994.

[8] J. W. Brewer, "Kronecker products and matrix calculus in system theory," *IEEE Trans. Circuits and Systems*, vol. 25, no. 9, pp. 772–781, 1978.

[9] H. Besbes, M. Jaidane, and J. Ezzine, "On exact performances of adaptive Volterra filters: the finite alphabet case," in *Proc. IEEE Int. Symp. Circuits and Systems (ISCAS '00)*, vol. 3, pp. 610–613, Geneva, Switzerland, May 2000.

[10] H. Besbes, M. Jaidane-Saidane, and J. Ezzine, "Exact analysis of the tracking capability of time-varying channels: the finite alphabet inputs case," in *Proc. IEEE International Conference on Electronics, Circuits and Systems (ICECS '98)*, vol. 1, pp. 449–452, Lisboa, Portugal, September 1998.

[11] M. Sayadi, F. Fnaiech, and M. Najim, "An LMS adaptive second-order Volterra filter with a zeroth-order term: steady-state performance analysis in a time-varying environment," *IEEE Trans. Signal Processing*, vol. 47, no. 3, pp. 872–876, 1999.

[12] E. Eweda, "Comparison of RLS, LMS, and sign algorithms for tracking randomly time-varying channels," *IEEE Trans. Signal Processing*, vol. 42, no. 11, pp. 2937–2944, 1994.

[13] E. Biglieri, D. Divsalar, P. J. McLane, and M.K. Simon, *Introduction to Trellis-Coded Modulation with Applications*, Macmillan Publishing Company, New York, NY, USA, 1991.

[14] H. Besbes, M. Jaidane-Saidane, and J. Ezzine, "On exact convergence results of adaptive filters: the finite alphabet case," *Signal Processing*, vol. 80, no. 7, pp. 1373–1384, 2000.

[15] F. R. Gantmacher, *The Theory of Matrices, Vol. 2*, Chelsea Publishing Company, New York, NY, USA, 1959.

Hichem Besbes was born in Monastir, Tunisia, in 1966. He received the B.S. (with honors), the M.S., and the Ph.D. degrees in electrical engineering from the Ecole Nationale d'Ingénieurs de Tunis (ENIT) in 1991, 1991, and 1999, respectively. He joined the Ecole Supérieure des Communications de Tunis (Sup'Com), where he was a Lecturer from 1991 to 1999, and then an Assistant Professor. From July 1999 to October 2000, he held a Postdoctoral position at Concordia University, Montréal, Canada. In July 2001, he joined Legerity Inc., Austin, Texas, USA, where he was a Senior System Engineer working on broadband modems. From March 2002 to July 2003, he was a member of the technical staff at Celite Systems Inc., Austin, Texas, where he contributed to definition, design, and development of Celite's high-speed data transmission systems over wireline networks, named Broadcast DSL. He is currently an Assistant Professor at Sup'Com. His interests include adaptive filtering, synchronisation, equalization, and multirates broadcasting systems.

Mériem Jaïdane received the M.S. degree in electrical engineering from the Ecole Nationale d'Ingénieurs de Tunis (ENIT), Tunisia, in 1980. From 1980 to 1987, she worked as a Research Engineer at the Laboratoire des Signaux et Systèmes, CNRS/Ecole Supérieure d'Electricité, France. She received the Doctorat d'Etat degree in 1987. Since 1987, she was with the ENIT, where she is currently a Full Professor at Communications and Information Technologies Department. She is a Member of the Unité Signaux et Systèmes, ENIT. Her teaching and research interests are in adaptive systems for digital communications and audio processing.

Jelel Ezzine received the B.S. degree in electromechanical engineering from the Ecole Nationale d'Ingénieurs de Tunis (ENIT), in 1982, the M.S.E.E. degree from the University of Alabama in Huntsville, in 1985, and the Ph.D. degree from the Georgia Institute of Technology, in 1989. From 1989 to 1995, he was an Assistant Professor at the Department of Systems Engineering, King Fahd University of Petroleum and Minerals, where he taught and carried out research in systems and control. Presently, he is an Associate Professor at the ENIT and an Elected Member of its scientific council. Moreover, he is the Director of Studies and the Vice Director of the ENIT. His research interests include control and stabilization of jump parameter systems, neuro-fuzzy systems, application of systems and control theory, system dynamics, and sustainability science. He has been a Visiting Research Professor at Dartmouth College from July 1998 to June 1999, the Automation and Robotics Research Institute, UTA, Texas, from March 1998 to June 1998. He was part of several national and international organizing committees as well as international program committees. He is an IEEE CEB Associate Editor and a Senior Member of IEEE, and is listed in Who's Who in the World and Who's Who in Science and Engineering.