BLIND TENSOR-BASED IDENTIFICATION OF MEMORYLESS MULTIUSER VOLTERRA CHANNELS USING SOS AND MODULATION CODES

Carlos Alexandre R. FERNANDES, Gérard FAVIER and João Cesar M. MOTA*

I3S Laboratory, CNRS - University of Nice Sophia - Antipolis
2000 route des Lucioles, BP 121, 06903, Sophia-Antipolis Cedex, France
phone: + (33) 492942701, fax: + (33) 492942898, email: {acarlos,favier}@i3s.unice.fr

GTEL Laboratory - DETI, Federal University of Ceará
Campus do Pici, 60.755-640, 6007, Fortaleza, Brazil
phone: + (5585) 33669467, fax: + (5585) 33669468, email: {carlosalexandre,mota}@deti.ufc.br

ABSTRACT

In this paper, a channel identification technique using Second Order Statistics (SOS) is proposed for memoryless multiuser Volterra communication channels. The Parallel Factor (PARAFAC) decomposition of a third order tensor formed from spatio-temporal covariance matrices of the received signals is considered by using the Alternating Least Squares (ALS) algorithm. Modulation codes (constrained codes) are used to ensure some orthogonality constraints of the transmitted signals. That constitutes a new application of modulation codes, aiming to introduce temporal redundancy and ensure some statistical properties. Identifiability conditions for the problem under consideration are addressed and simulation results illustrate the performance of the proposed estimation method.

1. INTRODUCTION

In this paper, a channel estimation technique for memoryless multiuser Volterra communication systems is proposed. The channel is modelled as a complex-valued linear-cubic Multiple-Input-Multiple-Output (MIMO) Volterra filter, consisting of a generic representation of instantaneous linear-cubic polynomial mixtures. The nonlinear polynomial mixtures models have important applications in the field of telecommunications to model wireless communication links with nonlinear power amplifiers [1] and uplink channels in Radio Over Fiber (ROF) multiuser communication systems [2]. The ROF links have found a new important application with their introduction in microcellular wireless networks [3, 4]. In such systems, the uplink transmission is done from a mobile station towards a Radio Access Point, the transmitted signals being converted in optical frequencies by a laser diode and then retransmitted through optical fibers. Important nonlinear distortions are introduced by the electrical-optical (E/O) conversion [3, 4]. When the length of the optical fiber is short (few kilometers) and the radio frequency has an order of GHz, the dispersion of the fiber is negligible [5]. In this case, the nonlinear distortion arising from the E/O conversion process becomes preponderant [3, 4, 5]. Up to several Mbps, the ROF channel can be considered as a memoryless link [2, 3]. Thus, the channel is composed of a wireless link, which can be modelled as a linear instantaneous mixture, followed by electrical-optical (E/O) conversions, which can be modelled as memoryless nonlinearities [2]. Moreover, in this application, the channel nonlinearity is modelled as a third order polynomial [2, 3]. So, the overall channel corresponds to a third order memoryless multiuser Volterra system.

There are few works dealing with the problem of blind source separation or/and identification of nonlinear systems in the context of multiuser communication channels. Among them, we cite [6] that proposes a blind Zero Forcing technique for Code Division Multiple Access (CDMA) systems and [7], where semi-blind and blind source separation algorithms are developed for memoryless Volterra channels in ultra-wide-band systems.

The technique proposed in this paper exploits the use of Second Order Statistics (SOS) of the received signals. Modulation codes (constrained codes) [8] are used to ensure the orthogonality of nonlinear combinations of the transmitted signals for several time delays, allowing the application of the Parallel Factor (PARAFAC) decomposition [9] of a third order tensor formed from estimated spatio-temporal covariance matrices. A two-step Alternating Least Squares (ALS) algorithm [9, 10] is used to estimate the channel. The proposed modulation codes introduce redundancy by expanding the signal constellation, generating multilevel modulations. Modulation expanding is often used in bandwidth-constrained channels, where a performance gain can be achieved without expanding the channel bandwidth or the transmission power [8]. Modulation codes have applications in magnetic record, optical recording and digital communications over cable systems, with the goal of achieving spectral shaping and minimizing the DC content in the baseband signal [8]. This kind of coding was also used in [11] to reduce intrachannel nonlinear effects in high-speed optical transmissions. In this paper, the modulation codes are explored with a different purpose: the nonlinear channel identification. The redundancy provided by the codes introduces temporal correlation in a controlled way, in order that the transmitted signals verify some statistical constraints associated with the channel nonlinearities.

Blind channel identification and source separation using PARAFAC has been addressed earlier by some authors in the case of linear channels. In [12], a time-varying user power loading is proposed to enable the application of the PARAFAC analysis, in order to perform blind estimation of spatial signatures. PARAFAC decompositions have also been considered in the context of Code Division Multiple Access (CDMA) systems, with parameter estimation and/or source separation purposes [10]. In the case of nonlinear channels, a deterministic blind PARAFAC-based receiver was presented for Single-Input-Multiple-Output (SIMO) channels in [13].

2. SYSTEM MODEL

The discrete-time baseband equivalent model of the communication channel under consideration is assumed to be expressed as complex-valued linear-cubic polynomials of the form:

$$x^{(i)}(n) = \sum_{m_1=1}^{M} h_1^{(i)}(m_1) s_{m_1}(n) + \sum_{m_1=1}^{M} \sum_{m_2=m_1}^{M} \sum_{\substack{m_3=1 \\ m_3\neq m_1 \\ m_3\neq m_2}}^{M} h_3^{(i)}(m_1,m_2,m_3) s_{m_1}(n) s_{m_2}(n) s_{m_3}^{*}(n) + \upsilon^{(i)}(n), \quad (1)$$

where $x^{(i)}(n)$ is the received signal by the antenna i $(i = 1,2,...,I)$ at the time instant n, I is the number of antennae, M is the number

*C. A. R. FERNANDES is scholarship supported by CAPES/Brazil agency.

of users, $h^{(i)}_{2k+1}(m_1,\ldots,m_{2k+1})$, for $k=0,1$, are the channel coefficients, $s_m(n)$, for $1 \le m \le M$, are the unknown stationary and statistically independent transmitted signals and $\upsilon^{(i)}(n)$ is the Additive White Gaussian Noise (AWGN). The noise components $\upsilon^{(i)}(n)$ are assumed to be zero mean, independent from each other and from the input signals $s_m(n)$.

The cubic terms corresponding to $m_3=m_1$ and $m_3=m_2$ are absent in (1) due to the fact that, for constant modulus signals, like Phase-Shift Keying (PSK) modulated signals, they have the form: $s_{m_j}(n)\ |s_{m_3}(n)|^2$, $j\in\{1,2\}$, where $|s_{m_3}(n)|^2$ is a multiplicative constant absorbed by the channel coefficients. As a consequence, some cubic terms degenerate in linear terms. In addition, the quadratic terms are absent in (1) due to the fact that distortions generated by even-power terms produce spectral components lying outside the channel bandwidth, which can be eliminated by band-pass filters at the receiver.

The channel model (1) represents a complex-valued truncated triangular MIMO Volterra filter, the inputs of which are user indexed signals, instead of time indexed inputs as in traditional Single-Input-Single-Output (SISO) Volterra filters. It represents a general representation of instantaneous linear-cubic polynomial mixtures.

The signals received on the I antennae, at the time instant n, can also be expressed in a compact way:

$$\mathbf{x}(n) \quad = \quad \mathbf{H}\mathbf{s}(n)+\mathbf{v}(n), \tag{2}$$

where $\mathbf{x}(n) = [x^{(1)}(n)\ldots x^{(I)}(n)]^T \in \mathbb{C}^{I\times 1}$, $\mathbf{v}(n) = [\upsilon^{(1)}(n)\ldots\upsilon^{(I)}(n)]^T \in \mathbb{C}^{I\times 1}$ and $\mathbf{H}=[\mathbf{h}^{(1)}\ldots\mathbf{h}^{(I)}]^T \in \mathbb{C}^{I\times M_V}$, with the vector $\mathbf{h}^{(i)}$ ($1\le i\le I$) containing the parameters $h^{(i)}_{2k+1}(m_1,\ldots,m_{2k+1})$ and M_V being the number of coefficients of each subchannel in (1). Moreover, $\mathbf{s}(n)\in\mathbb{C}^{M_V\times 1}$ is the input vector containing the linear $\{s_{m_1}(n)\}$ $(1\le m_1\le M)$ and cubic terms $\{s_{m_1}(n)s_{m_2}(n)\ s^*_{m_3}(n)\}$ $(1\le m_1,m_2,m_3\le M,\ m_1\ne m_3,\ m_2\ne m_3,\ m_2\ge m_1)$. Note that $M_V=\frac{M}{2}(M^2-M+2)$.

3. SOS TENSOR

The proposed channel identification method relies on the PARAFAC decomposition of a tensor composed of spatio-temporal covariances of the received signals, given by:

$$\mathbf{R}(\tau)=\mathbb{E}\left[\mathbf{x}(n+\tau)\mathbf{x}^H(n)\right]=\mathbf{H}\mathbf{C}(\tau)\mathbf{H}^H+\sigma^2\mathbf{I}_I\delta(\tau), \tag{3}$$

with

$$\mathbf{C}(\tau)=\mathbb{E}\left[\mathbf{s}(n+\tau)\mathbf{s}^H(n)\right], \tag{4}$$

where $\tau\in\Upsilon=\{\tau_1,\tau_2,...,\tau_T\}$, T is the number of covariance matrices taken into account, the superscript H denotes the complex conjugate transpose of a matrix, $\delta(\tau)$ is the Kronecker symbol, σ^2 is the AWGN variance and $\mathbf{I}_I$ is the $I\times I$ identity matrix. The noise variance σ^2 can be estimated as the mean of the $(I-M_V)$ smallest eigenvalues of $\mathbf{R}(0)$, allowing the subtraction of the noise term in (3). Thus, we may ignore, from now on, this noise term.

A three-way tensor $\underline{\mathbf{R}}\in\mathbb{C}^{T\times N\times N}$ can be defined from the matrices $\mathbf{R}(\tau)$, for $\tau\in\Upsilon$, in such a way that the *first-mode slices* of $\underline{\mathbf{R}}$, denoted by $\mathbf{R}_{k\cdot\cdot}$, have the form:

$$\mathbf{R}_{k\cdot\cdot}=\mathbf{R}(\tau_k), \quad k=1,...,T, \tag{5}$$

where a *first-mode slice* of $\underline{\mathbf{R}}$ is obtained by fixing the first dimension index of $\underline{\mathbf{R}}$ and varying the indices of the two other modes. In order to enable the application of the PARAFAC decomposition of the tensor $\underline{\mathbf{R}}$, the matrices $\mathbf{C}(\tau)$ must be diagonal for $\tau\in\Upsilon$, leading to the following scalar notation:

$$r_{k,p,l}=\sum_{q=1}^{M_V} h_{p,q}h^*_{l,q}c_{q,q}(\tau_k)\ , \tag{6}$$

where $r_{k,p,l}=[\underline{\mathbf{R}}]_{k,p,l}=\mathbb{E}[x_p(n+\tau_k)x^*_l(n)]$, $h_{p,q}=[\mathbf{H}]_{p,q}$ and $c_{q,q}(\tau_k)=[\mathbf{C}(\tau_k)]_{q,q}$. To satisfy this property, the components of the nonlinear regression vector $\mathbf{s}(n)$ must be uncorrelated for $\tau\in\Upsilon$. The following theorem states sufficient conditions to ensure this constraint.

Theorem 1: Let us assume that all the signals transmitted by the users are mutually independent and have constant moduli. The following conditions are sufficient to ensure the diagonality of the covariance matrices $\mathbf{C}(\tau)$, $\tau\in\Upsilon$:

(i). $\mathbb{E}[s_m(n)]=0$, for all the users;
(ii). $\mathbb{E}\left[s^2_m(n)\right]=0$, for $(M-1)$ users;
(iii). $\mathbb{E}\left[s^2_m(n+\tau)s_m(n)\right]=0$ and $\mathbb{E}\left[s^2_m(n)s_m(n+\tau)\right]=0$, for $(M-1)$ users, $\forall\,\tau\in\Upsilon$;
(iv). $\mathbb{E}[s_m(n+\tau)s_m(n)]=0$, for $(M-1)$ users, $\forall\,\tau\in\Upsilon$.

The proof is omitted due to a lack of space. It can be derived by writing each non-diagonal term of the matrix $\mathbf{C}(\tau)$ as a product of the individual contributions of each user.

The following theorem proves that some conditions of Theorem 1 are verified if all the users transmit uniformly distributed PSK signals with more than 2 symbols in the constellation.

Theorem 2: Let us assume that all the users transmit uniformly distributed PSK signals with $R_m>2, \forall\, m\in\{1,2,...,M\}$, where R_m is the number of constellation symbols of the m^{th} user. Then, conditions (i) and (ii) of Theorem 1, and conditions (iii) and (iv), for $\tau=0$, are verified.

Proof: If $s_m(n)$, $m=1,...,M$, takes an equiprobable value from the set $\left\{A_m.e^{j2\pi(r-1)/R_m}; r=1,2,...,\ R_m; R_m>2\right\}$, then we have

$$\mathbb{E}\left[s^p_m(n)\right]=\frac{A^p_m}{R_m}\sum_{r=1}^{R_m}e^{j2\pi(r-1)p/R_m}=\frac{A^p_m\left(e^{j2\pi p}-1\right)}{R_m\left(e^{j2\pi p/R_m}-1\right)}, \tag{7}$$

which is equal to zero for $p=1,2,3$ and $R_m>2$. □

4. DESIGN OF CODING SCHEMES

In this section, some modulation codes are designed to ensure that the transmitted signals satisfy the constraints listed in Theorem 1. In these modulation code schemes, the modulation makes part of the encoding process and it introduces redundancy by expanding the signal constellation. This means that a modulation memory is introduced in a controlled way with the purpose of keeping the orthogonality between nonlinear combinations of the transmitted signals.

This constitutes a new application of modulation codes, since they are used to ensure some statistical properties associated with the channel nonlinearities. Moreover, the code redundancy could also be explored in the symbol recovery process to provide Bit Error Rate (BER) improvements, by exploiting the fact that introduced redundancy imposes some constraints on the symbol transitions. This subject will be investigated in future works.

The modulated signals are characterized by Discrete Time Markov Chains (DTMC) with R_m states, given by the PSK symbols $a_r=\{A_m\cdot e^{j2\pi(r-1)/R_m}\}$, for $r=1,2,...,R_m$, where A_m is the amplitude of the signal of the m^{th} user. The state transitions are defined by a block of k_m bits, denoted by $B_n=\{b^{(1)}_n,b^{(2)}_n,...,b^{(k_m)}_n\}$, where $b^{(k)}_n$, for k=1,...,k_m, is uniformly distributed over the set $\{0,1\}$ and $2^{k_m}<R_m$. In addition, it is assumed that $b^{(k)}_n$ (k=1,...,k_m) are mutually independent. For each of the R_m states, the block of bits B_n defines 2^{k_m} equiprobable possible transitions. Therefore, the coding imposes some restrictions on the symbol transitions. For each state, there is $\left(R_m-2^{k_m}\right)$ not assigned transitions. The code rate of the m^{th} user is then given by (k_m/l_m), where $l_m=\log_2 R_m$.

Let us denote by $\mathbf{T}=\{T_{r_1,r_2}\}$, with $r_1,r_2\in\{1,2,...,R_m\}$ the Transition Probability Matrix, T_{r_1,r_2} being the probability of a transition from the state r_1 to the state r_2. Note that $\sum_{r_2=1}^{R_m}T_{r_1,r_2}=1$

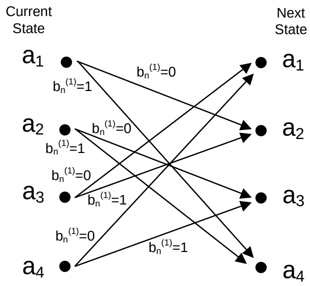

Fig. 1. Miller Code State diagram.

and $T_{r_1,r_2} \in \{0, 1/2^{k_m}\}$. So, the matrix $\mathbf{T}$ defines which are the possible state transitions for each state. An example of mapping from the bits B_n to the corresponding PSK symbols is illustrated in Fig. 1 for a 4-PSK signal, where $\{a_1, a_2, a_3, a_4\}$ are the constellation symbols (states) and $k_m = 1$. This state diagram corresponds to the run-length-limited code known as Miller Code, associated with the transition probability matrix $\mathbf{T}_{2,B}$ given in (20). The Miller Code is widely used in digital magnetic recording and in Binary-PSK carrier modulation systems [8]. Similar state diagrams can be obtained for the other transition probability matrices given in the Appendix.

According to Theorem 2, if all the users transmit uniformly distributed PSK signals, then conditions of Theorem 1 are verified for $\tau = 0$. So, the following theorem proposes some constraints in the transition probability matrix $\mathbf{T}$ in such a way that the corresponding user transmits an uniformly distributed PSK signal.

Theorem 3: Let us assume that the DTMC associated with the coding is irreducible and aperiodic. If $\sum_{r_1=1}^{R_m} T_{r_1,r_2} = 1$, for $1 \le r_2 \le R_m$, then, for a large number of time steps, the average fraction of time steps in which the DTMC is in the state a_{r_1} converges to $1/R_m$, for $1 \le r_1 \le R_m$.

Proof: The aperiodicity and irreducibility properties assure that [14]: (i) all the limiting probabilities of a DTMC exist and are positive, (ii) the stationary distribution exists and is unique, and (iii) the limiting probabilities distribution is equal to the stationary distribution. So, the limiting probabilities $\mathbf{P} = [p_1\ p_2\ \dots\ p_{R_m}]$ can be obtained by solving the following system of equations:

$$\begin{cases} \mathbf{P}\mathbf{T} = \mathbf{P}, \\ \sum_{r=1}^{R_m} p_r = 1. \end{cases} \tag{8}$$

It can be easily verified that if $\sum_{r_1=1}^{R_m} T_{r_1,r_2} = 1$, then $\mathbf{P} = [1/R_m\ \dots\ 1/R_m]$ is a solution of the system (8). And finally, it can be proved (the proof is omitted due to a lack of space) that if the limiting probability of a state a_{r_1} exists, then it is equal to the long-run time average spent in the state a_{r_1}, i.e. for a large number of time steps, the average fraction of time steps that the DTMC spends in the state a_{r_1} converges to the limiting probability of the state a_{r_1}. □

Now we develop some restrictions to the transition probability matrix so that the conditions of Theorem 1 be verified for $\tau \neq 0$. Let $T^n_{r_1,r_2}$ be the $(r_1, r_2)^{th}$ entry of $\mathbf{T}^n$. By definition, $T^n_{r_1,r_2}$ represents the probability of being in the state a_{r_2} after n transitions, supposing that the current state is a_{r_1}. So, we may write:

$$\mathbb{E}\left[s^k_m(n+\tau)s^l_m(n)\right] = \frac{1}{R_m}\mathbf{a}_l^T\mathbf{T}^\tau\mathbf{a}_k, \tag{9}$$

where $\mathbf{a} = [a_1, a_2, \dots a_{R_m}]^T$ and $\mathbf{a}_k = \left[a^k_1, a^k_2, \dots a^k_{R_m}\right]^T$. Thus, the conditions (iii) and (iv) of Theorem 1 can be rewritten as:

$$\mathbf{a}^T\mathbf{T}^\tau\mathbf{a}_2 = 0, \quad \mathbf{a}_2^T\mathbf{T}^\tau\mathbf{a} = 0 \quad \text{and} \quad \mathbf{a}^T\mathbf{T}^\tau\mathbf{a} = 0. \tag{10}$$

The results of this section may be summarized in the following corollary.

Corollary 1: If the following conditions hold for all the users:

(i). the transition probability matrix corresponds to an irreducible and aperiodic DTMC;

(ii). $\sum_{r_1=1}^{R_m} T_{r_1,r_2} = 1, \forall r_2,\ 1 \le r_2 \le R_m$;

and, in addition, equations (10) are verified for $(M-1)$ users $\forall \tau \in \Upsilon$, then all the conditions of Theorem 1 are satisfied and, therefore, the covariance matrix $\mathbf{C}(\tau)$ is diagonal $\forall \tau \in \Upsilon$.

It should be highlighted that equations (10) only depend on the matrix $\mathbf{T}$ and the constellation order. That means that the transition probability matrices can be a priori designed to verify these equations. In the Appendix, some examples of matrices verifying these constraints are listed, with the corresponding admissible delays.

5. CHANNEL ESTIMATION ALGORITHM

Let us denote respectively by $\mathbf{R_1} \in \mathbb{C}^{NT\times N}$ and $\mathbf{R_3} \in \mathbb{C}^{NT\times N}$ the first and third-mode unfolded matrices of the tensor $\underline{\mathbf{R}}$, defined as:

$$\mathbf{R_1} = \begin{bmatrix} \mathbf{R}_{1\cdot\cdot} \\ \vdots \\ \mathbf{R}_{T\cdot\cdot} \end{bmatrix}, \qquad \mathbf{R_3} = \begin{bmatrix} \mathbf{R}_{\cdot\cdot 1} \\ \vdots \\ \mathbf{R}_{\cdot\cdot N} \end{bmatrix}, \tag{11}$$

where the matrices $\mathbf{R}_{i\cdot\cdot}$, and $\mathbf{R}_{\cdot\cdot k}$ are respectively the first and third-mode slices of $\underline{\mathbf{R}}$. Provided that the conditions of Corollary 1 are satisfied, these slice matrices can be expressed respectively by:

$$\mathbf{R}_{i\cdot\cdot} = \mathbf{H}\mathrm{diag}_i[\mathbf{C}_d]\mathbf{H}^H \quad \text{and} \quad \mathbf{R}_{\cdot\cdot k} = \mathbf{C}_d\mathrm{diag}_k[\mathbf{H}^*]\mathbf{H}^T, \tag{12}$$

where the rows of the matrix $\mathbf{C}_d \in \mathbb{C}^{T\times M_V}$ contain the diagonal components of $\mathbf{C}(\tau)$ and $\mathrm{diag}_i[\mathbf{A}]$ is the diagonal matrix formed from the i^{th} row of $\mathbf{A}$. So, the unfolded matrices are given by:

$$\mathbf{R_1} = (\mathbf{C}_d \diamond \mathbf{H})\mathbf{H}^H \quad \text{and} \quad \mathbf{R_3} = (\mathbf{H}^* \diamond \mathbf{C}_d)\mathbf{H}^T, \tag{13}$$

where $\diamond$ denotes the Khatri-Rao (column-wise Kronecker) product. The following theorem imposes a sufficient condition for the uniqueness of the considered blind estimation problem.

Theorem 4: Let the two pairs of matrices $(\hat{\mathbf{H}}_a, \hat{\mathbf{H}}_b)$ and $(\hat{\mathbf{H}}'_a, \hat{\mathbf{H}}'_b)$ be solutions to (13) for a given matrix $\mathbf{C}_d$. If

$$2k_{\mathbf{H}} + k_{\mathbf{C}_d} \ge 2M_V + 2, \tag{14}$$

then $\hat{\mathbf{H}}_a = \hat{\mathbf{H}}'_a\Lambda$ and $\hat{\mathbf{H}}_b = \hat{\mathbf{H}}'_b\Lambda^{-1}$, where Λ is a $M_V \times M_V$ diagonal matrix and $k_{\mathbf{A}}$ is the k-rank of matrix $\mathbf{A}$, i.e. the greatest integer k such that every set of k columns of $\mathbf{A}$ is linearly independent.

Proof: The proof of the above theorem is a direct result of the Kruskal Theorem [15]. Denoting by $(\hat{\mathbf{H}}_a, \hat{\mathbf{H}}_b, \hat{\mathbf{C}_d})$ and$(\hat{\mathbf{H}}'_a, \hat{\mathbf{H}}'_b, \hat{\mathbf{C}_d}')$ two solutions of (13) for $\mathbf{C}_d$ unknown, the Kruskal Theorem says that if $k_{\mathbf{H}} + k_{\mathbf{H}^*} + k_{\mathbf{D}} \ge 2M_V + 2$, then $\hat{\mathbf{H}}_a = \hat{\mathbf{H}}'_a\Pi\Lambda_a$, $\hat{\mathbf{H}}_b = \hat{\mathbf{H}}'_b\Pi\Lambda_b$ and $\hat{\mathbf{C}_d} = \hat{\mathbf{C}_d}'\Pi\Lambda_c$, where Λ_a, Λ_b and Λ_c are diagonal matrices such that $\Lambda_a\Lambda_b\Lambda_c = \mathbf{I}_{M_V}$ and Π is a permutation matrix. If we assume that $\mathbf{C}_d$ is known, then $\hat{\mathbf{C}_d} = \hat{\mathbf{C}_d}' = \mathbf{C}_d$ and, therefore, $\Pi = \mathbf{I}_{M_V}$, $\Lambda_c = \mathbf{I}_{M_V}$ and $\Lambda_b = \Lambda_a^{-1} = \Lambda^{-1}$.

The scaling ambiguity in the channel estimation introduced by the matrix Λ does not represent an effective problem, as it can be removed by a gain control at the receiver. In the Appendix, some examples of configurations of transition probability matrices for 2 users are shown, verifying $k_{\mathbf{C}_d} = M_V$. In this case, (14) becomes $\min(I, M_V) \ge M_V/2 + 1$.

The channel estimation is obtained by using the ALS algorithm [9, 10], the principle of which is to estimate, in the least square sense, a subset of the parameters by using a previous estimation of other subsets of parameters. This process continues until the convergence of the parameters is achieved. The ALS algorithm is monotonically convergent but it may require a large number of iterations to converge [16]. For the proposed technique, each ALS iteration corresponds to two updating steps. It computes two estimates, denoted by $\hat{\mathbf{H}}_a$ and $\hat{\mathbf{H}}_b$, for the matrices $\mathbf{H}$ and $\mathbf{H}^*$, respectively, while the matrix $\mathbf{C}_d$ is assumed to be known, due to the fact that the matrices $\mathbf{C}(\tau)$ only depend on the modulation codes and the transmission power of the users. The algorithm does not take the fact that $\hat{\mathbf{H}}_a$ is the complex conjugate of $\hat{\mathbf{H}}_b$ into account. The channel estimation problem is solved by minimizing the following least squares cost function:

$$J = \left\|\hat{\mathbf{R}}_\mathbf{1} - \left(\mathbf{C}_d \diamond \hat{\mathbf{H}}_a\right)\hat{\mathbf{H}}_b^T\right\|_F^2 = \left\|\hat{\mathbf{R}}_\mathbf{3} - \left(\hat{\mathbf{H}}_b \diamond \mathbf{C}_d\right)\hat{\mathbf{H}}_a^T\right\|_F^2, \qquad (15)$$

where $\hat{\mathbf{R}}_\mathbf{1}$ and $\hat{\mathbf{R}}_\mathbf{3}$ are the sample estimated unfolded matrices and $\|\cdot\|_F$ denotes the *Frobenius norm*. Thus, the it^{th} iteration of the ALS algorithm can be described by the following steps:

$$\hat{\mathbf{H}}_b^{(it)} = \left[\left(\mathbf{C}_d \diamond \hat{\mathbf{H}}_a^{(it-1)}\right)^\dagger \hat{\mathbf{R}}_\mathbf{1}\right]^T, \qquad (16)$$

$$\hat{\mathbf{H}}_a^{(it)} = \left[\left(\hat{\mathbf{H}}_b^{(it)} \diamond \mathbf{C}_d\right)^\dagger \hat{\mathbf{R}}_\mathbf{3}\right]^T, \qquad (17)$$

where $\hat{\mathbf{H}}_a^{(0)}$ and $\hat{\mathbf{H}}_b^{(0)}$ are $I \times M_V$ Gaussian random matrices and $(\cdot)^\dagger$ denotes the pseudo-inverse. This process continues until the convergence of the estimated parameters is achieved. Each iteration of this ALS algorithm corresponds to about $(M_V + 8M_V^2)2IT + 2I^2TM_V + \frac{22}{3}M_V^3 + 2M_V^2$ multiplications. Three channel estimates can be then obtained: $\hat{\mathbf{H}}_a^{(it)}$, $(\hat{\mathbf{H}}_b^{(it)})^*$ and $0.5 \cdot [\hat{\mathbf{H}}_a^{(it)} + (\hat{\mathbf{H}}_b^{(it)})^*]$. The final channel estimate is chosen as the one providing the small value of the cost function (15).

6. SIMULATION RESULTS

In this section, the proposed channel estimation method is evaluated by means of computational simulations. The considered channel is a memoryless MIMO Wiener filter corresponding to the model of an uplink channel of a Radio Over Fiber (ROF) multiuser communication system [4]. The wireless interface is a memoryless $I \times M$ MIMO linear channel, consisting in an uniform spaced array of I antennae. The antennae are half-wavelength spaced and the transmitted signals are narrowband with respect to the array aperture. Moreover, the propagation scenario is characterized by two users, the angles of arrival of which are 30° and 70°. The E/O conversion in each antenna is modelled by the linear-cubic polynomial [4]: $-0.291x + 1.078|x|^2x$. The used modulation is 4-PSK and all the results were obtained via Monte Carlo simulations using $N_R = 100$ independent data realizations.

The channel estimation method is evaluated by means of the (Normalized Mean Squared Error) NMSE of the estimated channel parameters, defined as:

$$e_H = \frac{1}{N_R}\sum_{l=1}^{N_R}\frac{\|\mathbf{H} - \hat{\mathbf{H}}_l\|_F^2}{\|\mathbf{H}\|_F^2} \qquad (18)$$

where $\|\cdot\|_F$ denotes the *Frobenius norm* and $\hat{\mathbf{H}}_l \in \mathbb{C}^{N \times M_V}$ represents the channel matrix estimated at the j^{th} Monte Carlo simulation. Fig. 2 shows the NMSE versus SNR for the configurations of transition probability matrices given in Table 1 of the Appendix, for $M = 2$, $I = 4$, $T = 5$ and $N_s = 3000$, N_s being the length of the data block used in the estimation of the covariance matrices. We remark that all the tested schemes provide roughly similar NMSE performances. In this case, the ALS algorithm takes approximately 15 iterations to achieve the convergence .

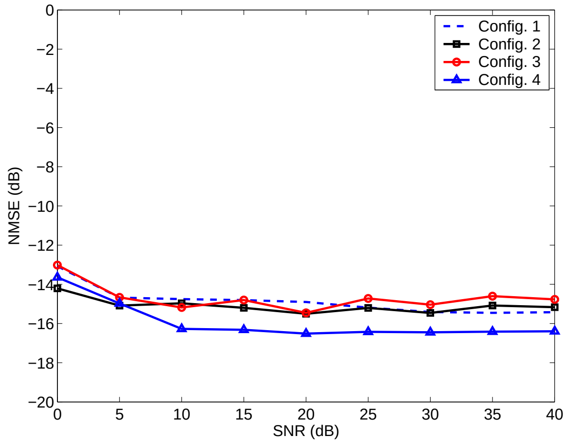

Fig. 2. NMSE versus SNR using the configurations of transition probability matrices of Table 1, for M=2, I=4, T=5 and $N_s = 3000$.

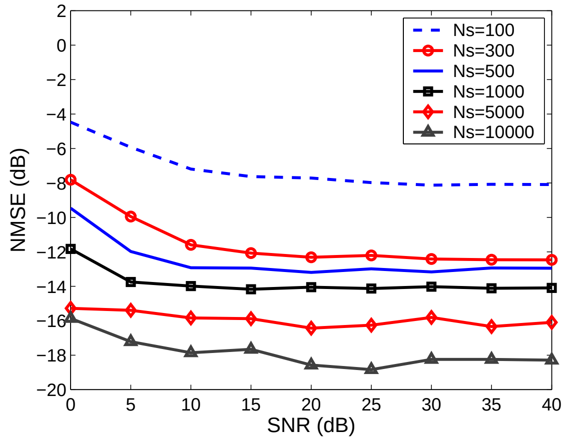

Fig. 3. NMSE versus SNR for various values of N_s using the Config. 2, $M = 2$, $I = 4$ and $T = 5$.

The influence of the data block length N_s used in the estimation of the covariance matrices, is illustrated in Fig. 3. It shows the NMSE versus SNR using the Config. 2 of Table 1, with $M = 2$, $I = 4$ and $T = 5$. The quality of the channel estimation can be considerably improved by increasing the value of N_s, which indicates that the errors in the estimation of the covariance matrices constitute one of the main sources of the degradation in this channel estimation technique. In fact, if the theoretical values of the covariance matrices $\mathbf{R}(\tau)$ are used, the estimation algorithm can attain very low NMSE values, limited by the machine precision. Moreover, we have also found that the number of ALS iterations needed to achieve the convergence decreases when Ns increases.

It should be highlighted from these results that the proposed channel estimator has a good robustness to AWGN, having no great performance degradation for low SNR's.

Fig. 4 shows the Symbol Error Rate (SER) versus SNR provided by the Minimum Mean Square Error (MMSE) receiver, given by $\hat{\mathbf{W}}_{MMSE} = \mathbf{C}(0)\hat{\mathbf{H}}^H\left[\hat{\mathbf{H}}\mathbf{C}(0)\hat{\mathbf{H}}^H + \sigma^2\mathbf{I}_\mathbf{I}\right]^{-1} \in \mathbb{C}^{M_V \times N}$. We have used Config. 2, $M = 2$, $I = 4$, $T = 5$ and $N_s = 3000$. In order to have a performance reference for our estimation technique, we have also simulated the MMSE receiver assuming perfect channel knowledge. Note that the SER performance using the estimated channel is very close to the one using the perfect channel.

Table 1. Examples of Configurations of transition probability matrices for 2 users.

Config.	User 1	User 2
1	$\mathbf{T}_{1,A}$	$\mathbf{T}_{2,B}$
2	$\mathbf{T}_{1,A}$	$\mathbf{T}_{2,A}$
3	$\mathbf{T}_{1,B}$	$\mathbf{T}_{2,B}$
4	$\mathbf{T}_{1,B}$	$\mathbf{T}_{2,A}$

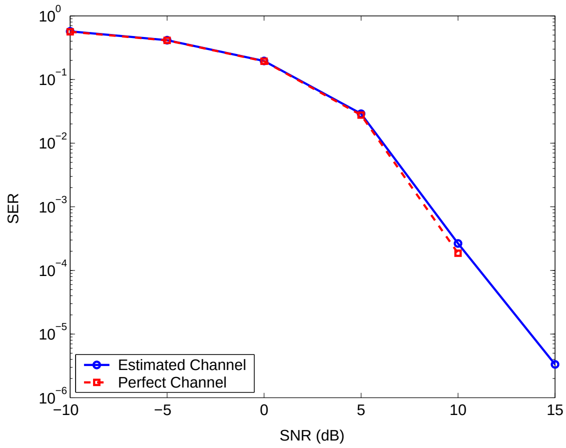

Fig. 4. SER versus SNR using the Config. 2, $M = 2$, $I = 4$, $T = 5$ and $N_s = 3000$.

7. CONCLUSION

The problem of blind estimation of memoryless multiuser Volterra channels has been studied in this paper. The proposed method is based on the PARAFAC decomposition of a tensor formed of spatio-temporal covariance matrices. Some constraints for the transmitted signals are developed to ensure the application of the PARAFAC analysis, using a two-step version of the ALS algorithm. Modulation codes are used to achieve these constraints, constituting a new application of this kind of coding. The proposed technique was applied to the identification of an uplink channel in a ROF multiuser communication system, providing good and consistent performances. The proposed blind identification method is robust to AWGN and the estimation errors of the covariance matrices are the main source of the performance degradation. In future works, other estimation algorithms will be tested and the impact of the modulation codes in the bit recovery process will be investigated.

A. APPENDIX - EXAMPLES OF CONFIGURATIONS OF TRANSITION PROBABILITY MATRICES

As pointed out, the transition probability matrices can be a priori designed to verify the conditions listed in Corollary 1. In the following, we present some examples of such matrices corresponding to $1/2$-code rate for 4-PSK signals. It can be proved by mathematical induction that the following matrices:

$$\mathbf{T}_{1,A} = 0.5 \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 \end{pmatrix}, \mathbf{T}_{1,B} = 0.5 \begin{pmatrix} 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \end{pmatrix}, \quad (19)$$

verify all the conditions of Corollary 1 $\forall \tau \in \mathbb{N}$. In this case $\mathbf{a} = [1 \; j \; -1 \; -j]^T$. In addition,

$$\mathbf{T}_{2,A} = 0.5 \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix}, \mathbf{T}_{2,B} = 0.5 \begin{pmatrix} 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \end{pmatrix} \quad (20)$$

verify conditions (i) and (ii) of Corollary 1.

The identifiability test in Theorem 4 depends on the covariance matrices $\mathbf{C}(\tau)$, for $\tau \in \Upsilon$, which can be calculated from the transition probability matrices by using (10) and:

$$\mathbb{E}\left[s_m^k(n+\tau) s_m^{l*}(n)\right] = \frac{1}{R_m} \mathbf{a}_l^H \mathbf{T}^\tau \mathbf{a}_k. \quad (21)$$

Thus, it can be verified that the configurations of transition probability matrices for 2 users given in Table 1 verify $k_{\mathbf{C}_d} = M_V$. The corresponding admissible delays are $\Upsilon = \{0, 1, \ldots, T-1\}$, with $T \geq 4$.

REFERENCES

[1] A. Ziehe, M. Kawanabe, S. Harmeling, and K.-R. Muller, "Blind separation of post-nonlinear mixtures using linearizing transformations and temporal decorrelation," *J. of Machine Learning Research*, vol. 4, no. 7-8, pp. 1319–1338, 2003.

[2] S. Z. Pinter and X. N. Fernando, "Estimation of radio-over-fiber uplink in a multiuser CDMA environment using PN spreading codes," in *Canadian Conf. on Elect. and Comp. Eng.*, May 1-4, 2005, pp. 1–4.

[3] X. N. Fernando and A. B. Sesay, "Higher order adaptive filter based predistortion for nonlinear distortion compensation of radio over fiber links," in *Intern. Conf. on Comm. (ICC)*, New-Orleans, LA, USA, June 2000, vol. 1/3, pp. 367–371.

[4] X. N. Fernando and A. B. Sesay, "A Hammerstein-type equalizer for concatenated fiber-wireless uplink," *IEEE Trans. Vehicular Tech.*, vol. 54, no. 6, pp. 1980–1991, 2005.

[5] W.I. Way, "Optical fiber based microcellular systems. An overview," *IEICE Trans. Commun.*, vol. E76-B, no. 9, pp. 1091–1102, Sept. 1993.

[6] A. J. Redfern and G. T. Zhou, "Blind zero forcing equalization of multichannel nonlinear CDMA systems," *IEEE Trans. Sig. Proc.*, vol. 49, no. 10, pp. 2363–2371, Oct. 2001.

[7] N. Petrochilos and K. Witrisal, "Semi-blind source separation for memoryless Volterra channels in UWB and its uniqueness," in *IEEE Workshop on Sensor Array and Multichannel Proc.*, 12-14 Jul. 2006, pp. 566–570.

[8] J. G. Proakis, *Digital Communications*, McGraw-Hill, 4*rd* edition, 2001.

[9] R. A. Harshman, *Foundations of the PARAFAC procedure: Models and conditions for an "explanatory" multimodal factor analysis*, UCLA Working Papers in Phonetics, 16 edition, Dec. 1970.

[10] N. D. Sidiropoulos, G. B. Giannakis, and R. Bro, "Blind PARAFAC receivers for DS-CDMA systems," *IEEE Trans. Sig. Proc.*, vol. 48, no. 3, pp. 810–822, March 2000.

[11] I. B. Djordjevic, B. Vasic, and V. S. Rao, "Rate 2/3 modulation code for suppression of intrachannel nonlinear effects in high-speed optical transmission," *IEE Proc.-Optoelectron.*, vol. 153, no. 2, pp. 87–92, Apr 2006.

[12] Y. Rong, S. A. Vorobyov, A. B. Gershman, and N. D. Sidiropoulos, "Blind spatial signature estimation via time-varying user power loading and parallel factor analysis," *IEEE Trans. Sig. Proc.*, vol. 53, no. 5, pp. 1697–1709, May 2005.

[13] A. Y. Kibangou, G. Favier, and M. M. Hassani, "Blind receiver based on the PARAFAC decomposition for nonlinear communication channels," in *Proc. Colloque GRETSI*, Louvain-la-neuve, Belgium, Sept. 2005, pp. 177–180.

[14] O. Haggstrom, *Finite Markov Chains and Algorithmic Applications*, Cambridge University Press, 2002.

[15] J. Kruskal, "Three way arrays: Rank and uniqueness of trilinear decomposition with applications to arithmetic complexity and statistics," *Linear Algebra Appl.*, vol. 18, pp. 95–138, 1977.

[16] R. Bro, *Multi-way analysis in the food industry: Models, algorithms and applications*, Ph.D. thesis, University of Amsterdam, Amsterdam, 1998.