Tensor Based Receivers for Nonlinear Radio Over Fiber Uplinks in Multiuser CDMA Systems

Carlos A. R. FERNANDES*
and Gérard FAVIER
I3S Laboratory
University of Nice-Sophia Antipolis/CNRS
Sophia-Antipolis, France
Email: {acarlos,favier}@i3s.unice.fr

João C. M. MOTA
Dept. of Teleinformatics Engineering
Federal University of Ceará
Fortaleza, Brazil
Email: mota@gtel.ufc.br

Abstract—Two semi-blind tensor based receivers for uplink Radio Over Fiber (ROF) nonlinear instantaneous channels in a multiuser Code Division Multiple Access (CDMA) environment are proposed in this paper. These receivers are based on the PARAFAC decomposition of a tensor (three-way array) composed of received signals with space, time and code diversities. A great advantage of the tensor approach is that it allows joint blind estimation of the channel and transmitted signals under weak uniqueness conditions compared with existing methods, which is particularly suitable when identifying nonlinear systems. Two estimation algorithms are proposed and illustrated by means of computer simulations.

I. INTRODUCTION

This paper proposes two tensor-based receivers for Radio Over Fiber (ROF) uplink channels in multiuser Code Division Multiple Access (CDMA) systems. The ROF links have found a new important application with their introduction in microcellular wireless networks. In such systems, the uplink transmission is done from a mobile station towards a Radio Access Point (RAP), 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 [1], [2], [3]. 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 [4]. In this case, the nonlinear distortion arising from the E/O conversion process becomes preponderant. Moreover, up to several Mbps, the ROF channel can be considered as a memoryless link [2]. Thus, the channel is composed of a wireless link, which can be modeled as a linear instantaneous mixture, followed by an electrical-optical (E/O) conversion, modeled as a memoryless nonlinearity. The overall channel can then be viewed as an instantaneous Multiple-Input-Multiple-Output (MIMO) Wiener filter, consisting in a particular case of MIMO Volterra filters (multiuser Volterra systems).

The presented techniques are based on the PARAFAC decomposition of a third-order tensor (three-way array) composed of received signals with space, time and code diversities. This tensor-based approach allows joint estimation of the channel and transmitted signals, using only one known pilot symbol. Another advantage of this approach is that it allows working with weak uniqueness conditions compared with existing methods [5], [6], [7], [8], [9], which require that the number of channel outputs be greater than the number of nonlinear terms. That provides a great flexibility on the number of antennas and spreading factor, which is particularly suitable when identifying Volterra systems with a high number of parameters.

Two semi-blind channel estimation algorithms are considered. First, the PARAFAC factors estimation is carried out by means of the Alternating Least Squares (ALS) algorithm. This technique can be viewed as an extension of [10] to nonlinear channels. However, as the ALS algorithm may present some convergence problems, a modified version of the ALS algorithm is considered by taking the structure of one of the factor matrices into account and by using a short training sequence.

PARAFAC-based channel estimation and source separation have also been addressed in the case of linear channels in the context of CDMA systems [10], [11], [12], [13]. In [14], 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. In the case of nonlinear channels, a deterministic blind PARAFAC-based receiver was presented for SIMO channels in [15] and a blind identification method based on the PARAFAC decomposition of an output data tensor was recently proposed for Wiener-Hammerstein type channels [16].

The rest of this paper is organized as follows. In Section II, the nonlinear ROF channel model used in this work is introduced. In Section III, the third-order tensor composed of received signals is characterized. In Section IV, the uniqueness issue is addressed. In Section V, two channel estimation algorithms are presented. In Section VI, the performance of these algorithms is illustrated by means of computer simulations and some conclusions are drawn in Section VII.

II. THE MULTIUSER NONLINEAR ROF CHANNEL MODEL

The discrete-time baseband equivalent model of the communication channel of a multiuser ROF CDMA system can
be expressed as a MIMO Wiener filter [1], [2], [3]. That means that the discrete-time baseband received signals after E/O conversion can be expressed as the output of a MIMO Volterra system:

\[
x_{r,n,p} = \sum_{r=0}^{R-1} \sum_{t=1}^{T} \sum_{i=1}^{K} \sum_{t_k+1}^{t_k+1} \sum_{q=1}^{T} \sum_{t_k+1}^{t_k+1} \sum_{r,n,p} \sum_{h_{2k+1}^{(r)}(t_1, \ldots, t_{2k+1})} u_{t_i,n,p} + v_{r,n,p},
\]

where \(x_{r,n,p} (1 \leq r \leq R, 1 \leq p \leq P)\) is the chip rate sampled signal received by antenna \(r\) and associated with chip \(p\) of time instant \(n\), i.e. received at the \((n-1)P + p\)th chip period, \(R\) is the number of receive antennas, \(P\) is the length of the spreading code (number of chips per symbol), \((2K+1)\) is the nonlinearity order of the model, \(h_{2k+1}^{(r)}(t_1, \ldots, t_{2k+1})\) are the kernel coefficients of the \(r\)-th sub-channel, \(u_{t_i,n,p}\) is the spread signal transmitted by user \(t\) and associated with chip \(p\) of time instant \(n\), \(T\) is the number of users and \(v_{r,n,p}\) is the Additive White Gaussian Noise (AWGN).

In the noiseless case, equation (1) can be rewritten as:

\[
x_{r,n,p} = \sum_{q=0}^{Q} h_{r,q} u_{n,p,q} = h_{r}^T \bar{u}_{n,p}, \tag{5}
\]

where the vector \(h_r = [h_{r,1}, h_{r,2}, \ldots, h_{r,Q}]^T \in \mathbb{C}^{Q \times 1}\) contains the Volterra kernel coefficients \(h_{2k+1}^{(r)}(t_1, \ldots, t_{2k+1})\) of the \(r\)-th sub-channel, the number of parameters of each sub-channel being given by \(Q = \sum_{k=0}^{K} C_{T,k} C_{T,k,1} + 1\), with \(C_{T,k} = \binom{T+k-1}{k}\). Moreover, \(\bar{u}_{n,p,q} = \[u_{n,p,1}, \ldots, u_{n,p,Q}]^T \in \mathbb{C}^{Q \times 1}\) is the nonlinear input vector containing the products of \(u_{t_i,n,p}\) and \(u_{t_i,n,p}\) in (1), given by:

\[
\bar{u}_{n,p,q} = \prod_{t=1}^{T} [u_{t_i,n,p}]^{\alpha_t} [u_{t_i,n,p}]^{\beta_i}, \tag{6}
\]

for some non negative integers \(\alpha_t\) and \(\beta_i\) verifying \(\sum_{t=1}^{T} \alpha_t = k + 1, \sum_{t=1}^{T} \beta_i = k\), with \(0 \leq \alpha_t \leq K + 1\) and \(0 \leq \beta_i \leq K\).

The information signal \(s_{n,t}\) of the \(t\)-th user is upsampled and multiplied by the spreading code before transmission, as:

\[
u_{t_i,n,p} = s_{n,t} c_{p,t}, \tag{7}
\]

where \(c_{p,t} (p = 1, \ldots, P)\) is the \(p\)-th element of the spreading code of the \(t\)-th user. Substituting (7) into (6), we get

\[
\bar{u}_{n,p,q} = \tilde{s}_{n,q} c_{p,q}, \tag{8}
\]

where

\[
\tilde{s}_{n,q} = \prod_{t=1}^{T} [s_{n,t}]^{\alpha_t} [s_{n,t}]^{\beta_i}. \tag{9}
\]

and

\[
\tilde{c}_{p,q} = \prod_{t=1}^{T} [c_{p,t}]^{\alpha_t} [c_{p,t}]^{\beta_i}. \tag{10}
\]

Substituting (8) into (5), we get

\[
x_{r,n,p} = \sum_{q=0}^{Q} h_{r,q} \tilde{s}_{n,q} \tilde{c}_{p,q}. \tag{11}
\]

If the information signals \(s_{n,t}\) are PSK modulated and the spreading codes \(c_{p,t}\) have an unitary modulus, then the transmitted signals \(u_{t_i,n,p}\) are constant modulus. In this case, the nonlinear terms corresponding to \(t_i = t_j\), for all \(i \in \{1, \ldots, k + 1\}\) and \(j \in \{k + 2, \ldots, 2k + 1\}\), are absent in (1) due to the fact that the term \(|u_{t_i,n,p}|^2\) reduces to a multiplicative constant that can be absorbed by the associated channel coefficient. As a consequence, some nonlinear terms degenerate in terms of smaller order, leading to a smaller value of \(Q\). Moreover, we have \(\alpha_t = 0\) or/and \(\beta_i = 0\), for all \(t = 1, \ldots, T\) in (6). For instance, for \(T = 2\) and \(K = 1\) (linear-cubic system), we have \(Q = 8\) when Quadrature amplitude modulation (QAM) is used, and \(Q = 4\) for PSK modulation.

### III. PARAFAC DECOMPOSITION

Let \(X \in \mathbb{C}^{T \times N \times P}\) be the third-order tensor composed of received signals \(x_{r,n,p}\) for \(1 \leq r \leq R, 1 \leq n \leq N\) and \(1 \leq p \leq P\), with \([X]_{r,n,p} \equiv x_{r,n,p}\), where \(N\) is the size of the data block. Equation (11) represents the scalar writing of the PARAFAC decomposition of the third-order tensor \(X\) of rank \(Q\), with matrix components \(H, \tilde{S}\) and \(\tilde{C}\), such that

\[
H = [h_1 \ldots h_R]^T \in \mathbb{C}^{R \times Q}, \tag{12}
\]

is the channel matrix,

\[
\tilde{S} = [s_1 \ldots s_N]^T \in \mathbb{C}^{N \times Q}, \tag{13}
\]

with \(\tilde{s}_n = [s_{n,1} \ldots s_{n,Q}]^T \in \mathbb{C}^{Q \times 1}\), is the nonlinear input matrix containing the products of information signals \(s_{n,t}\) given in (9), and

\[
\tilde{C} = [c_1 \ldots c_P]^T \in \mathbb{C}^{P \times Q}, \tag{14}
\]

with \(\tilde{c}_p = [c_{p,1} \ldots c_{p,Q}]^T \in \mathbb{C}^{Q \times 1}\), is the nonlinear code matrix, i.e. the matrix containing the products of codes given in (10).

For instance, for \(T = 2, K = 1\) and constant modulus transmitted signals, the matrices \(H, \tilde{S}\) and \(\tilde{C}\) are respectively given by:

\[
H = \begin{pmatrix}
h_1^{(1)}(1) & h_1^{(2)}(1) & h_3^{(1)}(1,1,2) & h_3^{(1)}(2,2,1) \\
... & ... & ... & ... \\
h_1^{(R)}(1) & h_1^{(R)}(2) & h_3^{(R)}(1,1,2) & h_3^{(R)}(2,2,1)
\end{pmatrix},
\]

\[
\tilde{S} = \begin{pmatrix}
s_{1,1} & s_{1,2} & s_{1,3}^2 & s_{1,4} \\
... & ... & ... & ... \\
s_{N,1} & s_{N,2} & s_{N,3}^2 & s_{N,4}^2
\end{pmatrix},
\]

\[
\tilde{C} = \begin{pmatrix}
c_1 & c_2 & c_3 & c_4 \\
... & ... & ... & ... \\
c_{P,1} & c_{P,2} & c_{P,3} & c_{P,4}
\end{pmatrix}.
\]
In particular, if we choose \( N \) and \( i \) then the matrix factors \( H \) if any other set of matrices \( k \) Let us denote by \( \text{diag}_i[] \) denotes the diagonal matrix formed from the \( i \)th row of the matrix argument.

All the elements of a tensor can be organized in unfolded matrices by stacking all the matrix slices of a given type. Let us denote respectively by \( X_1 \in \mathbb{C}^{N \times R} \), \( X_2 \in \mathbb{C}^{P \times N} \) and \( X_3 \in \mathbb{C}^{R \times P} \) the following unfolded matrices of the tensor \( X \):

\[
X_1 = \begin{bmatrix} X_{1,1} \\ \vdots \end{bmatrix}, \quad X_2 = \begin{bmatrix} X_{2,1} \\ \vdots \end{bmatrix}, \quad X_3 = \begin{bmatrix} X_{3,1} \\ \vdots \end{bmatrix},
\]

which gives

\[
X_1 = (\hat{H} \hat{S}) \tilde{C}, \quad X_2 = (\hat{S} \hat{C}) H^T, \quad X_3 = (\tilde{C} \hat{H}) \tilde{S}^T,
\]

where \( \circ \) denotes the Khatri-Rao (column-wise Kronecker) product.

IV. UNIQUENESS CONDITION

The main property of the PARAFAC decomposition is its essential uniqueness, assured by the Kruskal theorem [17]. Let us denote by \( k_A \) the k-rank of matrix \( A \), i.e. the greatest integer \( k_A \) such that every set of \( k_A \) columns of \( A \) is linearly independent.

The Kruskal theorem says that if:

\[
k_H + k_S + k_C \geq 2Q + 2.
\]

then the matrix factors \( H, \tilde{S} \) and \( \tilde{C} \) are essentially unique, i.e. if any other set of matrices \( H', \tilde{S}' \) and \( \tilde{C}' \) satisfies (11), then

\[
H' = \Pi_1 A_1, \quad \tilde{S}' = \Pi_2 A_2 \quad \text{and} \quad \tilde{C}' = \Pi_3 A_3 ,
\]

where \( A_1 \), \( A_2 \) and \( A_3 \) are diagonal matrices such that \( A_1 A_2 A_3 = I_Q \) and \( \Pi \) is a permutation matrix.

Assuming that the matrices \( H, \tilde{S} \) and \( \tilde{C} \) are full k-rank, the Kruskal condition becomes:

\[
\min(R,Q) + \min(N,Q) + \min(P,Q) \geq 2Q + 2. \tag{22}
\]

In particular, if we choose \( N \geq Q \), we get:

\[
\min(R,Q) + \min(P,Q) \geq Q + 2. \tag{23}
\]

A great advantage of using the PARAFAC decomposition is that it allows working in the underdetermined case, i.e. with \( R < Q \). That is particularly interesting as Volterra systems are generally characterized by a large number \( Q \) of parameters. Indeed, working in the underdetermined case may impose a strong constraint on the number of antennas to be used, as it is the case in many previous works [5], [6], [7], [18], [9].

The flexibility on the choice of \( R \) and \( P \) provided by the Kruskal condition is one of the main advantages of using a tensor-based approach.

V. ESTIMATION ALGORITHMS

A. Alternating Least Squares (ALS) algorithm

The first presented channel estimation algorithm is a two-steps version of the ALS algorithm [10], the principle of which is to estimate, in the least square sense, a subset of parameters by using previous estimates of the other parameters. Two estimates, denoted by \( \hat{H} \) and \( \hat{S} \), corresponding respectively to the matrices \( H \) and \( \tilde{S} \) are obtained. The matrix \( \tilde{C} \) is assumed to be known, as it can be calculated if the codes are known. So, if the Kruskal condition (21) is satisfied, we have \( C = \tilde{C} \) and, hence, \( \Pi = \Lambda_3 = I_Q \) and \( \Lambda_2 = \Lambda_1^{-1} \). Therefore, \( \hat{H} = \hat{H} A_1 \) and \( \hat{S} = \hat{S} A_1^{-1} \). This means that the permutation ambiguity is eliminated. Moreover, due to the structure of the matrix \( \tilde{S} \), the scaling ambiguity matrix \( \Lambda_1 \) can be identified by using one known pilot symbol for each user, i.e. by assuming that the first row of \( S \) is known, as:

\[
\Lambda_1 = \text{diag} \begin{bmatrix} \hat{s}_{1,1} & \cdots & \hat{s}_{1,Q} \end{bmatrix}^T,
\]

where \( \text{diag}(\cdot) \) denotes the diagonal matrix formed from the vector argument.

The channel estimation problem is solved by minimizing the two following cost functions in an alternate way:

\[
J_1 = \| X_2 - (\hat{S} \circ \tilde{C}) H^T \|_F^2, \quad J_2 = \| X_3 - (\tilde{C} \circ \hat{H}) \tilde{S}^T \|_F^2,
\]

where \( \| \cdot \|_F \) denotes the Frobenius norm. The ALS algorithm is summarized in Table I, where \( (\cdot)^T \) denotes the matrix pseudoinverse. The ALS algorithm is monotonically convergent but it may require a large number of iterations to converge.

The existence of the left inverse of the matrices \( (\hat{S}^{-1} \circ \tilde{C}) \) and \( (\tilde{C} \circ \hat{H}) \) is asymptotically assured if the Kruskal condition (21) is satisfied [19]. This technique can be viewed as a generalization of [10] to nonlinear channels, where the factor matrices \( H, S, C \) contain only the elements corresponding to the linear kernel.

B. ALS with Direct Decision and Block Initialization (ALS-DD-Bl) algorithm

The performance of the ALS algorithm can be improved by taking the structure of matrix \( \tilde{S} \) into account and assuming that the user symbols belong to a finite alphabet. The second
The proposed channel estimation method consists then in a modified version of the ALS algorithm where direct decisions are used to construct the matrix $\tilde{S}^{(it)}$. Moreover, some known pilot symbols are used to obtain an initial estimate for $H^{(it)}$.

Let us denote by $\tilde{S}_L^{(it)} \in \mathbb{C}^{N \times T}$ the matrix composed of the $T$ first columns of $\tilde{S}^{(it)}$, i.e. the matrix containing the linear part of $\tilde{S}^{(it)}$, and by $\tilde{S}_{NL}^{(it)} \in \mathbb{C}^{N \times (Q-T)}$ the matrix composed of the $(Q-T)$ last columns of $\tilde{S}^{(it)}$, i.e. the matrix containing the nonlinear part of $\tilde{S}^{(it)}$. That gives $\tilde{S}^{(it)} = [\tilde{S}_L^{(it)} \, \tilde{S}_{NL}^{(it)}]$. Moreover, let us define $\tilde{S}_{L,DD}^{(it)} \in \mathbb{C}^{N \times T}$ as the matrix composed of the elements of $\tilde{S}_{L}^{(it)}$ after a decision device. Finally, let $\tilde{S}_{NL,DD}^{(it)} \in \mathbb{C}^{N \times (Q-T)}$ be the nonlinear part of the input matrix reconstructed from $\tilde{S}_{L,DD}^{(it)}$. The ALS-DD-BI algorithm is summarized in Table II, where a short training sequence is used to obtain an initial estimate of the channel matrix $H^{(0)}$. Let us denote by $\hat{S}_0 \in \mathbb{C}^{N_t \times Q}$ the matrix composed of the $N_t$ first rows of $\hat{S}$ and by $X_{[2],0} \in \mathbb{C}^{P N_t \times R}$ the corresponding unfolded matrix of the tensor $X$, where $N_t$ is the length of the training sequence. An initial estimation of $H$ is obtained as

$$\hat{H}^{(0)} = \left[\left(\hat{S}_0 \circ \tilde{C}\right) \, \times_{[2],0} \right]^T.$$  \hspace{1cm} (27)

Note that a necessary identifiability condition for this initialization is $r(\hat{S}_0 \circ \tilde{C}) = Q$, which implies $N_t P \geq Q$.

### VI. Simulation Results

In this section, the two proposed channel estimation methods are evaluated by means of computational simulations. A MIMO Wiener model of an uplink channel of a ROF multi-user communication system [1], [5] is considered for the simulations. The flat fading wireless link is characterized by a memoryless $R \times 2$ linear mixer, with $T = 2$ users ($Q = 4$) and a half-wavelength spaced array of $R$ antennas. The E/O conversion for each channel is modeled by the following polynomial $c_1 x + c_3 x |^2 x$, with $c_1 = -0.35$ and $c_3 = 1$ [1], [20]. All the simulation results were obtained via Monte Carlo simulations using $N_R = 2000$ independent data realizations and the modulation of the transmitted signals is 4-PSK. The spreading codes are complex exponentials with an unitary modulus and a phase uniformly distributed over the set $[-\pi, \pi]$.

The performance of the proposed channel estimation methods is evaluated by means of the normalized mean square error (NMSE) of the estimated channel parameters, defined as:

$$\text{NMSE} = \frac{1}{N_t} \sum_{i=1}^{N_t} \frac{\|H_i - H^{(0)}_i\|^2}{\|H_i\|^2},$$

where $H_i$ represents the channel matrix estimated at the $i^{th}$ Monte Carlo simulation. Figure 1 shows the NMSE versus Signal to Noise Ratio (SNR) provided by the ALS and ALS-DD-BI techniques for $N = 50$, $N_t = 10$, $P = 3$, $R = 2$ and 4. From this figure, it can be concluded that the ALS-DD-BI algorithm provides better NMSE performances than the ALS algorithm. Moreover, it can be remarked that the performance of the proposed estimation methods is not deteriorated in the underdetermined case ($R = 2$), with respect to the overdetermined case ($R = 4$). In order to provide a performance reference to our method, we also show the NMSE obtained using (27) in a supervised scenario with $R = 4$, i.e. assuming that all the transmitted symbols are known ($N_t = N = 50$). Note that the performance of the ALS-DD-BI technique is quite similar to that of the supervised estimation.
improvement in terms of channel estimation and BER. The ALS algorithm described in this work can be viewed as a generalization of [10] to nonlinear channels. In future work, we will extend these methods to the case of nonlinear channels with memory and time-varying parameters.

VII. CONCLUSION

In this paper, we have proposed two new methods for jointly estimating the nonlinear ROF uplink channel and the transmitted symbols in a multiuser CDMA system. These methods are based on the PARAFAC decomposition of a third order tensor composed of received signals. The first method uses the classical ALS algorithm, while the second one, called the ALS-DD-BI algorithm, includes a block initialization and a decision device, which implies a significant performance improvement in terms of channel estimation and BER. The ALS algorithm described in this work can be viewed as a generalization of [10] to nonlinear channels. In future work, we will extend these methods to the case of nonlinear channels with memory and time-varying parameters.

REFERENCES


Figure 2 shows the number of iterations needed to achieve the convergence versus SNR for the ALS and ALS-DD-BI algorithms for $N = 50$, $N_f = 10$, $P = 3$, $R = 2$ and 4. It can be concluded that the ALS-DD-BI technique converges more quickly than the ALS algorithm in most of the cases. Note that the ALS-DD-BI algorithm converges after approximately 2 iterations when the SNR is higher than 5dB.

Figure 3 shows the Bit Error Rate (BER) versus SNR provided by the ALS and ALS-DD-BI techniques for $N = 50$, $N_f = 10$, $P = 3$, $R = 2$ and 4. It is also shown the BER provided by the zero forcing (ZF) receiver assuming the channel knowledge. Similar conclusions can be drawn from this figure, the ALS-DD-BI algorithm providing a BER better than that of the ALS algorithm and very close to that of the ZF receiver.

Fig. 2. Number of iterations needed to achieve the convergence versus SNR for $N = 50$, $N_f = 10$, $P = 3$, $R = 2$ and 4.

Fig. 3. BER versus SNR for $N = 50$, $N_f = 10$, $P = 3$, $R = 2$ and 4.