BLIND JOINT IDENTIFICATION AND EQUALIZATION OF WIENER-HAMMERSTEIN COMMUNICATION CHANNELS USING PARATUCK-2 TENSOR DECOMPOSITION

Alain Kibangou† *and Gérard Favier*‡

†LAAS-CNRS, University of Toulouse, 7 avenue Colonel Roche, 31077 Toulouse, France
‡ I3S, University of Nice Sophia Antipolis, CNRS,
2000 route des Lucioles, Les Algorithmes, Euclide B, BP 121, 06903 Sophia Antipolis Cedex
phone: +(33) 5 61 33 69 26, +(33) 4 92 94 27 36, email: akibango@laas.fr, favier@i3s.unice.fr

ABSTRACT

In this paper, we consider the blind joint identification and equalization of Wiener-Hammerstein nonlinear communication channels. By considering a special design of the input signal, we show that output data can be organized into a third-order tensor. We show that the obtained tensor has a PARATUCK-2 representation. We derive new results on uniqueness of the PARATUCK-2 model by considering structural constraints such as Toeplitz and Vandermonde forms for some of its matrix factors. We also constrain the input signal to belong to a finite alphabet. Then an Alternating Least Squares (ALS) algorithm is proposed for estimating the factors of the PARATUCK-2 model and therefore the parameters of the Wiener-Hammerstein channel and the unknown input signal. The performances of the proposed joint identification and equalization method are illustrated by means of simulation results.

1. INTRODUCTION

Linear approximation is often inappropriate or inadequate for modelling practical systems. Therefore, it is necessary to consider nonlinear models. Among them the Volterra model has been the subject of several works [1]. However, Volterra model suffers generally of a huge parametric complexity. By taking the structural property of the considered nonlinear plant into account, it is sometimes possible to use block-structured nonlinear models such as Wiener-Hammerstein models. These models that combine a memoryless nonlinearity with linearly dispersive elements have been successfully employed in various areas including digital communications systems [2, 3, 4].

There are relatively few works in the area of blind identification of Wiener-Hammerstein or block-structured nonlinear systems. Existing identification methods mainly concern the stochastic case, the input signal being generally assumed to be Gaussian (see [5] for example), and make use of higher order statistics, or exploit the cyclostationarity of the input signal [6]. In this paper, we consider a deterministic approach based on a third-order tensor decomposition.

Tensor analysis is used in many areas of science and engineering. Since few years, the use of tensor concepts in signal processing has gained more attention, motivated by the field of higher order statistics. When dealing with tensors, two families of models are generally encountered: PARAFAC [7, 8] and TUCKER models [9]. Different models sharing some properties of both PARAFAC and TUCKER models have been proposed in the dedicated literature, as for instance the PARATUCK-2 model [10]. In this paper, we show that the output data of a Wiener-Hammerstein communication channel can be organized into a third-order PARATUCK-2 tensor with structural constraints for its matrix factors in Toeplitz and Vandermonde forms. After deriving uniqueness properties of this constrained PARATUCK-2 decomposition, a blind joint identification and equalization algorithm is proposed for such communication channels.

The paper is organized as follows. In Section 2, we briefly recall some tensor models. The link between Wiener-Hammerstein model and the PARATUCK-2 decomposition is stated in Section 3. Then in Section 4, the PARATUCK-2 uniqueness properties are considered before describing the estimation procedure in Section 5. Some simulation results are provided to illustrate the performances of the proposed algorithm in Section 6. Conclusions and future work are given in Section 7.

Notations:

$\dagger$	Matrix pseudo − inverse	T	Matrix transpose
$\odot$	Khatri − Rao product	$\|.\|_F$	Frobenius norm
$\mathbf{I}_N$	$(N\times N)$ Identity matrix		

$\mathbf{A}_{.p}$ (resp. $\mathbf{A}_{p.}$): pth column (resp. row) of the matrix $\mathbf{A}$.
$D_p(\mathbf{A})$: Operator that forms a diagonal matrix from the pth row of the matrix $\mathbf{A}$.
$D(\mathbf{a})$: Operator that forms a diagonal matrix from the elements of the vector $\mathbf{a}$. So, we have $D_p(\mathbf{A}) = D(\mathbf{A}_{p.})$.
$\mathbf{x}^p$ denotes a vector with the entries of $\mathbf{x} = (x_1\ x_2 \cdots x_M)^T$ raised to power p, i.e. $\mathbf{x}^p = (x_1^p\ x_2^p \cdots x_M^p)^T$.

2. TENSOR DECOMPOSITIONS

Let $\mathbb{X}$ be a three-way array, also called a $I_1\times I_2\times I_3$ third-order tensor, the elements of which are x_{i_1,i_2,i_3}, $i_j = 1,2,\cdots,I_j$, $j = 1,2,3$. In analysis of $\mathbb{X}$ the difference between independent and interacting factors is illustrated by the two following decompositions:

$$x_{i_1,i_2,i_3} = \sum_{n_1=1}^{N_1} a_{i_1,n_1} b_{i_2,n_1} c_{i_3,n_1},$$

$$x_{i_1,i_2,i_3} = \sum_{n_1=1}^{N_1}\sum_{n_2=1}^{N_2}\sum_{n_3=1}^{N_3} a_{i_1,n_1} b_{i_2,n_2} c_{i_3,n_3} h_{n_1,n_2,n_3}$$

The first representation is called PARAFAC(PARallel FACtors analysis)[7] or CANDECOMP (CANonical DECOMPosition)[8]. It exhibits independence between factor contributions to each data point. On the contrary, in the second representation, called TUCKER-3 [9], the factors interact in their contributions to the data. The TUCKER-3 representation is then more general than PARAFAC. In general, TUCKER models are not unique whereas uniqueness conditions exist for PARAFAC [11]. A family of models sharing features of the two above mentioned models has been suggested in the literature: the so-called PARATUCK-2 model [10] given by

$$x_{i_1,i_2,i_3} = \sum_{n_1=1}^{N_1}\sum_{n_2=1}^{N_2} a_{i_1,n_1} c^A_{i_3,n_1} h_{n_1,n_2} c^B_{i_3,n_2} b_{i_2,n_2}. \qquad (1)$$

By slicing the $I_1\times I_2\times I_3$ tensor $\mathbb{X}$ along its third mode, we get the following matrix representation of the PARATUCK-2 model:

$$\mathbf{X}_{..i_3} = \mathbf{A}D_{i_3}(\mathbf{C}^A)\mathbf{H}D_{i_3}(\mathbf{C}^B)\mathbf{B}^T \in \Re^{I_1\times I_2},\ i_3 = 1,2,\cdots,I_3, \qquad (2)$$

where $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}^A$, and $\mathbf{C}^B$ are matrices with the respective dimensions $I_1 \times N_1$, $I_2 \times N_2$, $I_3 \times N_1$, and $I_3 \times N_2$. $\mathbf{H}$ is a $N_1 \times N_2$ matrix called kernel-matrix. Uniqueness of the PARATUCK-2 decomposition has been proved when $N_1 = N_2$ provided $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{H}$ are full column rank matrices and $\mathbf{H}$ has no zero elements [10].

To the best of the authors knowledge, few applications and works are devoted to the PARATUCK-2 decomposition. In the sequel, we show how this decomposition can be used for jointly identifying and equalizing Wiener-Hammerstein channels, and we give uniqueness conditions when $N_1 \neq N_2$.

3. PARATUCK-2 MODELLING OF WIENER-HAMMERSTEIN COMMUNICATION CHANNELS

The Wiener-Hammerstein model, depicted in Fig. 1, is one of commonly used block-oriented nonlinear structures. We denote by $u(n)$, $y(n)$, $v_1(n)$, $v_2(n)$, and $e(n)$, the input signal, the output signal, intermediate variables, and the additive noise respectively. Assuming that the nonlinearity is continuous within the considered dynamic range, then, from the Weierstrass theorem, it can be approximated to an arbitrary degree of accuracy by a polynomial $C(.)$ of finite degree P, the coefficients of which are c_p. So, the Wiener-Hammerstein model is constituted by

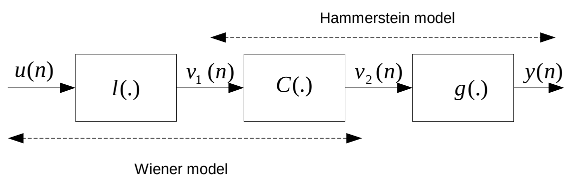

Figure 1: Wiener-Hammerstein model

a polynomial $C(.)$, sandwiched between two linear filters with impulse response $l(.)$ and $g(.)$ and memory M_l and M_g respectively, i.e., $v_1(n) = \sum_{i=0}^{M_l-1} l(i)u(n-i)$, $v_2(n) = \sum_{p=1}^{P} c_p v_1^p(n)$, and $y(n) = \sum_{i=0}^{M_g-1} g(i)v_2(n-i) + e(n)$. For characterizing the parametric representation of each block of the Wiener-Hammerstein system, we define the following vectors: $\mathbf{l} = (l(0)\ l(1) \cdots l(M_l-1))^T$, $\mathbf{g} = \big(g(0)\ g(1) \cdots g(M_g-1)\big)^T$, and $\mathbf{c} = (c_1\ c_2 \cdots c_P)^T$. One can note that if the linear subsystems $l(.)$ and $g(.)$ are respectively scaled by α_1 and α_2 then the polynomial defined by the coefficients $\tilde{c}_p = \frac{c_p}{\alpha_1^p \alpha_2}$ leads to the same input-output equation. In other words, the linear subsystems can be defined up to a scaling factor. In order to remove such an indeterminacy we enforce the linear subsystems to be monic, i.e. $l(0) = g(0) = 1$. This constraint is not restrictive. Indeed, if the actual linear subsystems are not monic then they can be determined up to the first coefficient of their respective impulse responses. One can also note that this constraint ensures the unicity of the polynomial subsystem. With these specifications, the equivalent Volterra representation is given by:

$$y(n) = \sum_{p=1}^{P} \sum_{j_1,\cdots,j_p=0}^{M-1} h_p(j_1,\cdots,j_p) \prod_{k=1}^{p} u(n-j_k) + e(n), \quad (3)$$

where $M = M_l + M_g$, and

$$h_p(j_1-1,\cdots,j_p-1) = c_p \sum_{j=1}^{M_g} g(j-1) \prod_{k=1}^{p} l(j_k-j), \quad (4)$$

with $j_k = 1,\cdots,M$, and $k = 1,\cdots,p$.

To remove the effect of the non-diagonal Volterra kernel coefficients on the output signal, assuming that the memory M and the nonlinearity degree are well-known, the input signal is designed such that:

$$u((nR+r-1)M+m-1) = \begin{cases} \phi_r s(n) & if \quad m=1 \\ 0 & if \quad m=2,\cdots,M \end{cases}$$

where ϕ_r, $r = 1,2,\cdots,R$, are distinct non-zero real valued known coefficients whereas $s(n)$ are the informative unknown symbols to be transmitted, belonging to a finite alphabet set $\Lambda = \{\pm\lambda_q\}_{q=1}^{Q/2}$. This choice of the input signal corresponds to a data transmission by block. Each block r associated with a symbol $s(n)$, is of dimension M and contains this symbol coded by a constant ϕ_r and $(M-1)$ zeroes that allow to remove the non-diagonal Volterra kernel coefficients effect on the output signal. Then, the output signal is given by:

$$y((nR+r-1)M+m-1) = \sum_{p=1}^{P} h_p(m-1,m-1,\cdots,m-1)\phi_r^p s^p(n) + e((nR+r-1)M+m-1).$$

By using (4) and denoting $y_{m,n,r} = y((nR+r-1)M+m-1)$ and $e_{m,n,r} = e((nR+r-1)M+m-1)$, we get:

$$y_{m,n,r} = \sum_{p=1}^{P} \sum_{j=1}^{M_g} c_p g(j-1) l^p(m-j) \phi_r^p s^p(n) + e_{m,n,r}, \quad (5)$$

or equivalently

$$y_{m,n,r} = \sum_{j=1}^{M_l} \sum_{p=1}^{P} c_p g(m-j) l^p(j-1) \phi_r^p s^p(n) + e_{m,n,r}. \quad (6)$$

The indices m, r, and n represent respectively the position m in the block r associated with the symbol $s(n)$. By defining the following correspondences:

$$(i_1,i_2,i_3,n_1,n_2,I_1,I_2,I_3,N_1,N_2) \Leftrightarrow (m,n,r,j,p,M,N,R,M_l,P),$$

and $a_{i_1,n_1} = g(m-j)$, $c^A_{i_3,n_1} = \phi_r l(j-1)$, $h_{n_1,n_2} = l^{p-1}(j-1)$, $c^B_{i_3,n_2} = c_p \phi_r^{p-1}$, and $b_{i_2,n_2} = s^p(n)$, we can conclude that (6) is a PARATUCK-2 model of the form (1).

We first reduce our discussion to the noiseless case. We define the $M \times N$ matrices $\mathbf{Y}_{..r}$ as:

$$\mathbf{Y}_{..r} = \begin{pmatrix} y_{1,1,r} & y_{1,2,r} & \cdots & y_{1,N,r} \\ y_{2,1,r} & y_{2,2,r} & \cdots & y_{2,N,r} \\ \vdots & \vdots & \ddots & \vdots \\ y_{M,1,r} & y_{M,2,r} & \cdots & y_{M,N,r} \end{pmatrix}.$$

Using the above stated correspondences and their equivalent matrix formulations, i.e.$(\mathbf{A},\mathbf{C}^A,\mathbf{H},\mathbf{C}^B,\mathbf{B}) \Leftrightarrow (\mathbf{G},\mathbf{F},\mathbf{L},\mathbf{C},\mathbf{S})$, equation (2) becomes :

$$\mathbf{Y}_{..r} = \mathbf{G} D_r(\mathbf{F}) \mathbf{L} D_r(\mathbf{C}) \mathbf{S}^T, \quad (7)$$

where

- $\mathbf{G}$ is a $M \times M_l$ Toeplitz matrix the first row and column of which are respectively given by $\mathbf{G}_{1.} = \begin{pmatrix} g(0)\ 0 \cdots 0 \end{pmatrix}$ and $\mathbf{G}_{.1} = \begin{pmatrix} \mathbf{g}^T\ 0 \cdots 0 \end{pmatrix}^T$,we also note $\mathbf{G} = \mathscr{T}(\mathbf{g})$,
- $\mathbf{S}$ is a $N \times P$ column-wise Vandermonde matrix: $\mathbf{S} = \begin{pmatrix} \mathbf{s}\ \mathbf{s}^2 \cdots \mathbf{s}^P \end{pmatrix}$, $\mathbf{s} = \begin{pmatrix} s(1)\ s(2) \cdots s(N) \end{pmatrix}^T$,
- $\mathbf{L}$ is a $M_l \times P$ column-wise Vandermonde matrix: $\mathbf{L} = \begin{pmatrix} \mathbf{l}^0\ \mathbf{l}^1 \cdots \mathbf{l}^{P-1} \end{pmatrix}$,
- $\mathbf{F}$ is a $R \times M_l$ rank one matrix: $\mathbf{F} = \boldsymbol{\phi}\mathbf{l}^T$, $\boldsymbol{\phi} = \begin{pmatrix} \phi_1\ \phi_2 \cdots \phi_R \end{pmatrix}^T$
- $\mathbf{C}$ is a $R \times P$ matrix: $\mathbf{C} = \boldsymbol{\Phi} D(\mathbf{c})$, $\boldsymbol{\Phi} = \begin{pmatrix} \boldsymbol{\phi}^0\ \boldsymbol{\phi}^1 \cdots \boldsymbol{\phi}^{P-1} \end{pmatrix}$.

The $M \times N \times R$ tensor $\mathbb{Y}$, defined by the slices $\mathbf{Y}_{..r}$, $r = 1, 2, \cdots, R$, is a PARATUCK-2 model characterized by structural constraints concerning the matrix factors $\mathbf{G}$, $\mathbf{L}$, and $\mathbf{S}$ that are in Toeplitz and Vandermonde forms respectively. One can note that uniqueness results given in [10] cannot be applied for such a PARATUCK-2 model since $M_l \neq P$ in general. In the sequel we investigate uniqueness properties of the PARATUCK-2 model (7) by exploiting structural constraints on the involved matrices.

4. UNIQUENESS ISSUE OF THE PARATUCK-2 MODEL WITH STRUCTURAL CONSTRAINTS

Before considering the uniqueness issue, we recall the useful property of *admissibility in the Vandermonde sense* of a given alphabet and a lemma on a related identifiability result concerning bilinear decompositions [12].

Definition: Let $\mathbf{V}_P$ be the Pth-order Vandermonde matrix associated with the finite alphabet $\Lambda = \{\pm\lambda_q\}_{q=1}^{Q/2}$ defined by

$$\mathbf{V}_P = \begin{pmatrix} \lambda_1 & -\lambda_1 & \cdots & \lambda_{Q/2} & -\lambda_{Q/2} \\ \lambda_1^2 & \lambda_1^2 & \cdots & \lambda_{Q/2}^2 & \lambda_{Q/2}^2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \lambda_1^P & (-1)^P\lambda_1^P & \cdots & \lambda_{Q/2}^P & (-1)^P\lambda_{Q/2}^P \end{pmatrix}.$$

The alphabet Λ is said to be admissible of order P in the Vandermonde sense if $Q \geq P$ and if the only admissible permutation matrices $\mathbf{\Pi}$ of order Q satisfying $\mathbf{V}_P\mathbf{\Pi}(\mathbf{V}_P^\dagger\mathbf{V}_P - \mathbf{I}_Q) = 0$ are $\mathbf{\Pi} = \mathbf{I}_Q$ and $\mathbf{\Pi} = \mathbf{J}_Q$, $\mathbf{J}_Q$ being a $Q \times Q$ block diagonal matrix constituted with $Q/2$ blocks defined as $\mathbf{J}_2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$.

Lemma 1 *Let $\mathbf{Y} = \mathbf{H}\mathbf{U}^T$, where $\mathbf{H}$ is an arbitrary full rank matrix, $\mathbf{U}$ is a column-wise Vandermonde matrix, constructed from a finite alphabet set $\Lambda = \{\pm\lambda_q\}_{q=1}^{Q/2}$. If the alphabet Λ is admissible of order P in the Vandermonde sense and if each symbol of Λ appears at least once in the first column of $\mathbf{U}$ then $\mathbf{U}$ and $\mathbf{H}$ can be uniquely identified up to a diagonal matrix $\mathbf{T} = diag(\beta, \quad \beta^2, \quad \cdots \quad, \beta^P)$, $\beta = \pm 1$.*

In the sequel, we consider the following assumptions:

A1: $g(0) = 1$ and $g(M_g - 1)$ is nonzero. This implies that any bilinear decomposition as $\mathbf{GK}$, $\mathbf{G}$ being in Toeplitz form and $\mathbf{K}$ with no particular structure, is unique [13].

A2: Λ is an admissible alphabet of order P in the Vandermonde sense and each symbol of Λ appears at least once in the sequence $\{s(n)\}_{n=1,2,\cdots,N}$.

A3: The weights ϕ_r, $r = 1, 2, \cdots, R$, and the coefficients of the polynomial c_p, $p = 1, 2, \cdots, P$ are nonzero.

A4: At least P coefficients of the impulse response $l(.)$ are distinct and non zero, and $(M_l \geq P)$.

The uniqueness property of the PARATUCK-2 model is stated in the theorem below.

Theorem 1 *Consider a $M \times N \times R$ third-order tensor $\mathbb{Y}$ with the PARATUCK-2 structure given in (7). Suppose there is an alternative representation of $\mathbb{Y}$ with matrices of the same size and structural form*

$$\mathbf{Y}_{..r} = \bar{\mathbf{G}}D_r(\bar{\mathbf{F}})\bar{\mathbf{L}}D_r(\bar{\mathbf{C}})\bar{\mathbf{S}}^T, r = 1, 2, \cdots, R.$$

Then, according to assumptions A1-A4, the two representations are necessarily linked by the following relations:

$$\bar{\mathbf{G}} = \mathbf{G},\ \bar{\mathbf{F}} = \mathbf{F},\ \bar{\mathbf{L}} = \mathbf{L},\ \bar{\mathbf{S}} = \mathbf{S}\mathbf{\Delta},\ \bar{\mathbf{C}} = \mathbf{C}\mathbf{\Delta}^{-1}, \tag{8}$$

with $\mathbf{\Delta} = diag(\beta, \quad \beta^2, \quad \cdots \quad, \beta^P)$, $\beta = \pm 1$.

Proof: see the Appendix.

Thanks to the uniqueness property of the PARATUCK-2 decomposition stated by the Theorem 1, we can estimate the vectors $\mathbf{g}$, $\mathbf{l}$, and $\mathbf{c}$ describing the different blocks of the Wiener-Hammerstein system. Obviously, both $\mathbf{g}$ and $\mathbf{l}$ are uniquely determined from $\mathbf{G}$ and $\mathbf{L}$ respectively. From $\mathbf{C} = \mathbf{\Phi}D(\mathbf{c})$ we can deduce $\mathbf{c}$. $\mathbf{\Phi}$ being known, the coefficients of the polynomial subsystems are then blindly estimated up to a sign.

5. ESTIMATION OF THE PARATUCK-2 FACTORS

As for PARAFAC and TUCKER decompositions, the estimation of the PARATUCK-2 decomposition factors can be carried out using the Alternating Least Squares (ALS) algorithm, as it was first proposed in [14] where the different factors were obtained by minimizing the cost function $f(\mathbf{G}, \mathbf{F}, \mathbf{L}, \mathbf{C}, \mathbf{S}) = \sum_{r=1}^{R} \left\|\tilde{\mathbf{Y}}_{..r} - \mathbf{G}D_r(\mathbf{F})\mathbf{L}D_r(\mathbf{C})\mathbf{S}^T\right\|_F^2$, $\tilde{\mathbf{Y}}_{..r}$ being the noisy version of $\mathbf{Y}_{..r}$. In this section, we propose a new method for estimating the matrix factors of the PARATUCK-2 model. We define the matrices $\mathbf{L}_2 = D(\mathbf{l})\mathbf{L}$ and $\mathbf{C}_2 = D(\boldsymbol{\phi})\mathbf{C}$. Then the slices $\mathbf{Y}_{..r}$ and the unfolded matrix $\mathbf{Y}_R$ defined in (14) can be rewritten as:

$$\mathbf{Y}_{..r} = \mathbf{G}\mathbf{L}_2 D_r(\mathbf{C}_2)\mathbf{S}^T, \quad \mathbf{Y}_R = (\mathbf{C}_2 \odot \mathbf{G}\mathbf{L}_2)\mathbf{S}^T.$$

We also consider the unfolded matrices associated with the first and second modes. For the first mode, the slices are given by $\mathbf{Y}_{m..} = \mathbf{S}D_m(\mathbf{G}\mathbf{L}_2)\mathbf{C}_2^T$, $m = 1, \cdots, M$ and we have $\mathbf{Y}_M = \left(\mathbf{Y}_{1..}^T \cdots \mathbf{Y}_{M..}^T\right)^T = (\mathbf{G}\mathbf{L}_2 \odot \mathbf{S})\mathbf{C}_2^T$. For the second mode, we have $\mathbf{Y}_{.n.} = \mathbf{C}_2 D_n(\mathbf{S})\mathbf{L}_2^T\mathbf{G}^T$, $n = 1, \cdots, N$, and $\mathbf{Y}_N = \left(\mathbf{Y}_{.1.}^T \cdots \mathbf{Y}_{.N.}^T\right)^T = (\mathbf{S} \odot \mathbf{C}_2)\mathbf{L}_2^T\mathbf{G}^T$. For estimating the factors $\mathbf{S}$, $\mathbf{G}$, $\mathbf{L}_2$, and $\mathbf{C}_2$, we alternatively minimize the cost functions $J_R = \left\|\tilde{\mathbf{Y}}_R - (\mathbf{C}_2 \odot \mathbf{G}\mathbf{L}_2)\mathbf{S}^T\right\|_F^2$, $J_N = \left\|\tilde{\mathbf{Y}}_N - (\mathbf{S} \odot \mathbf{C}_2)\mathbf{L}_2^T\mathbf{G}^T\right\|_F^2$, and $J_M = \left\|\tilde{\mathbf{Y}}_M - (\mathbf{G}\mathbf{L}_2 \odot \mathbf{S})\mathbf{C}_2^T\right\|_F^2$, where $\tilde{\mathbf{Y}}_R$, $\tilde{\mathbf{Y}}_N$, and $\tilde{\mathbf{Y}}_M$ are respective noisy versions of $\mathbf{Y}_R$, $\mathbf{Y}_N$, and $\mathbf{Y}_M$. Then, we can deduce $\mathbf{C}$ and $\mathbf{F}$. The proposed parameter estimation algorithm is now described.

5.1 Estimation of S

By minimizing J_R with respect to $\mathbf{S}$, we get:

$$\hat{\mathbf{S}}_{LS} = \left((\mathbf{C}_2 \odot \mathbf{G}\mathbf{L}_2)^\dagger\tilde{\mathbf{Y}}_R\right)^T.$$

In order to enforce the Vandermonde structure, $\hat{\mathbf{S}}$ is constructed from the first column of $\hat{\mathbf{S}}_{LS}$ after its projection onto the alphabet Λ:

$$\hat{\mathbf{S}} = (\hat{\mathbf{S}}_{LS,.1}\ \hat{\mathbf{S}}_{LS,.1}^2 \cdots \hat{\mathbf{S}}_{LS,.1}^P). \tag{9}$$

5.2 Estimation of G

One can note that $\mathbf{Y}_N^T = \mathbf{G}\mathbf{L}_2(\mathbf{S} \odot \mathbf{C}_2)^T = \mathbf{G}\mathbf{X}$, with $\mathbf{X} = \mathbf{L}_2(\mathbf{S} \odot \mathbf{C}_2)^T$. We define $\mathbf{\Theta} = \left(\mathscr{T}(\mathbf{X}_{.1})^T \cdots \mathscr{T}(\mathbf{X}_{.RN})^T\right)^T$, where $\mathscr{T}(\mathbf{X}_{.k})$ denotes the $M \times M_g$ Toeplitz matrix constructed from $\mathbf{X}_{.k}$. $\mathbf{G}$ being a Toeplitz matrix, we get: $vec(\mathbf{Y}_N^T) = \mathbf{\Theta}\mathbf{g}$. Then J_N can equivalently be written as: $J_N = \left\|vec(\tilde{\mathbf{Y}}_N^T) - \mathbf{\Theta}\mathbf{g}\right\|_2^2$. Its minimization with respect to $\mathbf{g}$ yields:

$$\hat{\mathbf{g}} = \mathbf{\Theta}^\dagger vec(\tilde{\mathbf{Y}}_N^T). \tag{10}$$

The entries of $\hat{\mathbf{g}}$ are first divided by the first entry and then $\hat{\mathbf{G}} = \mathscr{T}(\hat{\mathbf{g}})$.

5.3 Estimation of C

Since $\mathbf{C}_2 = D(\boldsymbol{\phi})\mathbf{C} = D(\boldsymbol{\phi})\mathbf{\Phi}D(\mathbf{c})$, for estimating $\mathbf{C}$, it is necessary to get $\mathbf{c}$. One can note that $\mathbf{Y}_M = (\mathbf{G}\mathbf{L}_2 \odot \mathbf{S})\mathbf{C}_2^T = (\mathbf{G}\mathbf{L}_2 \odot \mathbf{S})D(\mathbf{c})\bar{\mathbf{C}}^T$, with $\bar{\mathbf{C}} = D(\boldsymbol{\phi})\mathbf{\Phi}$. Then

J_M can be rewritten as $J_M = \left\| vec(\tilde{\mathbf{Y}}_M) - \left(\bar{\mathbf{C}} \odot (\mathbf{G}\mathbf{L}_2 \odot \mathbf{S})\right) \mathbf{c} \right\|_2^2$. As a consequence:

$$\hat{\mathbf{c}} = \left(\bar{\mathbf{C}} \odot (\mathbf{G}\mathbf{L}_2 \odot \mathbf{S})\right)^{\dagger} vec(\tilde{\mathbf{Y}}_M), \tag{11}$$

and $\hat{\mathbf{C}} = \mathbf{\Phi} D(\hat{\mathbf{c}})$.

5.4 Estimation of L and F

By minimizing J_N with respect to $\mathbf{L}_2$, we get:

$$\mathbf{L}_{2,LS} = \mathbf{G}^{\dagger}\tilde{\mathbf{Y}}_N^T \left((\mathbf{S} \odot \mathbf{C}_2)^T\right)^{\dagger}$$

Noticing that the estimated vector $\hat{\mathbf{l}}$ is given by the first column of $\mathbf{L}_{2,LS}$:

$$\hat{\mathbf{l}} = \left(\mathbf{G}\dagger\tilde{\mathbf{Y}}_N^T \left((\mathbf{S} \odot \mathbf{C}_2)^T\right)^{\dagger}\right)_{.1}, \tag{12}$$

all the entries of this vector are divided by its first element before constructing the following estimated factors : $\hat{\mathbf{F}} = \boldsymbol{\phi}\hat{\mathbf{l}}^T$ and $\hat{\mathbf{L}} = \left(\ \hat{\mathbf{l}}^0\ \hat{\mathbf{l}}^1\ \cdots\ \hat{\mathbf{l}}^{P-1}\ \right)$.

5.5 Wiener-Hammerstein identification algorithm

- Step 0: Initialize $\mathbf{g}$,$\mathbf{c}$, $\mathbf{l}$, with random values and deduce the initial values for $\mathbf{G}$, $\mathbf{C}_2$, and $\mathbf{L}_2$.
- Step 1: Estimate $\mathbf{S}$ using (9). If a threshold value σ_1 is reached by the cost function J_R, project the elements of the first column of $\mathbf{S}$ into the finite alphabet Λ.
- Step 2: Estimate $\mathbf{g}$ using (10) and then construct $\hat{\mathbf{G}} = \mathscr{T}(\hat{\mathbf{g}})$.
- Step 3: Estimate $\mathbf{c}$ using (11) and then deduce $\mathbf{C}_2$.
- Step 4: Estimate $\mathbf{l}$ using (12) and then construct $\mathbf{L}_2$.
- Step 5: If a convergence criterion is reached, the algorithm is stopped. Else, return to Step 1.

The algorithm is stopped when the estimates of both $\mathbf{g}$ and $\mathbf{l}$ do not significantly vary during consecutive iterations.

6. SIMULATIONS

We consider a Wiener-Hammerstein system such that the linear filters and the polynomial coefficients are, respectively, given by $\mathbf{l} = (1,\ -0.3\ 0.1)$, $\mathbf{g} = (1,\ 0.5\ 0.2)$, $\mathbf{c} = (2,\ 0.8\ 0.5)$. The input signal is a 6-PAM one, which is admissible of order $P = 3$ in the Vandermonde sense. We evaluate the performance of the proposed algorithm by means of two measures: the output normalized mean square error, i.e. $\left\|\tilde{\mathbf{Y}}_R - (\hat{\mathbf{C}}_2 \odot \hat{\mathbf{G}}\hat{\mathbf{L}}_2)\hat{\mathbf{S}}^T\right\|_F^2 / \|\mathbf{Y}_R\|_F^2$, and the parameters normalized mean square error. We consider the transmission of $N = 60$ unknown symbols with a repetition factor $R = 3$. A white Gaussian noise was added to the channel output.

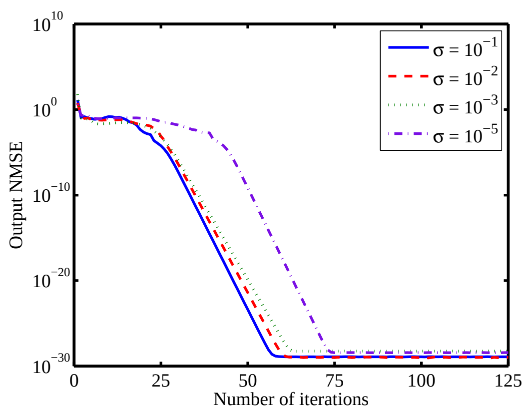

Figure 2: Output NMSE in the noiseless case.

In Fig. 2, corresponding to a noiseless case, the output NMSE is plotted with respect to the number of iterations. These results are averaged over 100 random initializations. One can note the convergence behavior of the algorithm. The convergence speed of the proposed algorithm increases with the convergence threshold value σ_1. In the sequel we set $\sigma_1 = 0.1$. In presence of noise, the choice of the estimated parameters initial values becomes more crucial. We consider 10 random initializations and for each one 100 noise sequences are generated. The results given in the following plots are averaged values over all of the initializations and noise sequences.

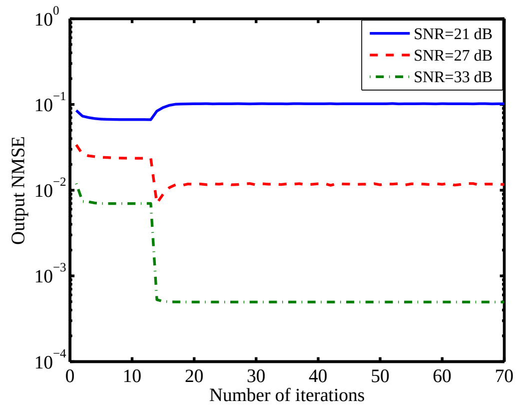

Figure 3: Output NMSE in noisy cases.

In Fig. 3, one can note the effect of the projection on the alphabet, which leads to get an NMSE corresponding to the additive noise level. However, this projection can deteriorate the algorithm performances especially for lower SNRs.

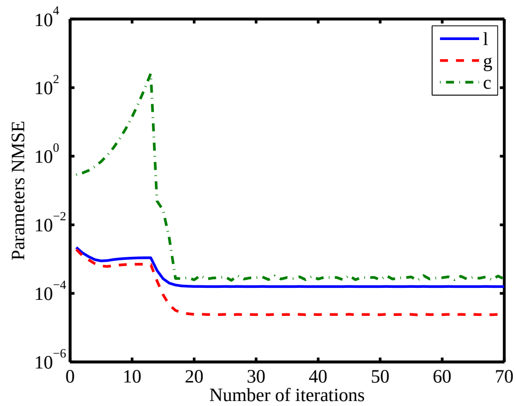

Figure 4: Parameters NMSE in a noisy case (SNR=30 dB).

In Figure 4, we can see convergence on the estimation of the subsystems parameters. The estimation of the polynomial subsystem is less precise than that of the linear ones. In Figure 5, we evaluate the equalization performance in terms of symbol error rate. We compare the proposed scheme with a non-blind one suggested in the literature [15]. In such approach, the channel parameters are assumed known. One can note that the performance significantly lowers when the SNR decreases. For high SNR the proposed algorithm gives performances close to those obtained with the non-blind scheme.

7. CONCLUSION

In this paper, we have proposed a blind joint identification and equalization algorithm for a Wiener-Hammerstein type channel.This algorithm uses a particular input design or precoding that yields a tensorial representation of output data. We have shown

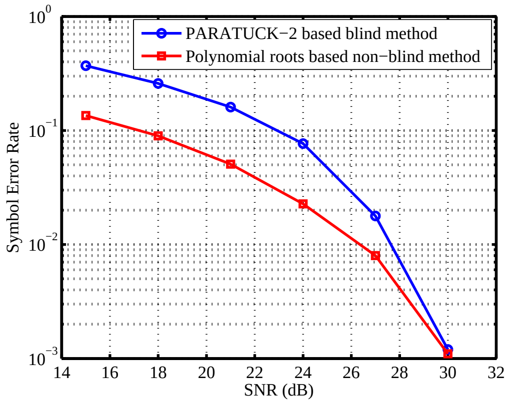

Figure 5: Error probability on symbols detection.

that the obtained tensor is a constrained PARATUCK-2 one. Its factors exhibit the Toeplitz and Vandermonde structures. We have established uniqueness conditions for such a PARATUCK-2 tensor model. Then an iterative LS type algorithm has been proposed for estimating the factors of the tensor model. From these factors we deduce the channel parameters and the input signal. The initialization of the proposed algorithm remains a crucial question. We intend to study this issue and that of the robustness to noise of the proposed scheme.

8. APPENDIX

For any slice $\mathbf{Y}_{..r}$, we can write: $\mathbf{Y}_{..r} = \mathbf{GZ}_r$, with $\mathbf{Z_r} = D_r(\mathbf{F})\mathbf{L}D_r(\mathbf{C})\mathbf{S}^T$. By concatenating all the R slices from left to right, we get:

$$\mathbf{Y} = (\mathbf{Y}_{..1} \quad \cdots \quad \mathbf{Y}_{..R}) = \mathbf{GZ}, \quad \mathbf{Z} = (\mathbf{Z}_1 \quad \cdots \quad \mathbf{Z}_R). \qquad (13)$$

Then according to assumption A1 and using the result in [13], we know that this bilinear decomposition is unique, i.e. $\bar{\mathbf{G}} = \mathbf{G}$. In the same way, by concatenating the slices from up to down, we get

$$\mathbf{Y}_R = \left(\mathbf{Y}^T_{..1} \quad \cdots \quad \mathbf{Y}^T_{..R}\right)^T = \mathbf{W}\mathbf{S}^T, \qquad (14)$$

where

$$\mathbf{W} = \begin{pmatrix} \mathbf{G}D_1(\mathbf{F})\mathbf{L}D_1(\mathbf{C}) \\ \vdots \\ \mathbf{G}D_R(\mathbf{F})\mathbf{L}D_R(\mathbf{C}) \end{pmatrix} \qquad (15)$$

Thanks to assumptions A3 and A4, by construction $\mathbf{W}$ is a full column rank matrix. Therefore, using assumption A2 and Lemma 1, the above bilinear decomposition is unique up to a diagonal matrix $\mathbf{\Delta}_1 = diag(\beta, \beta^2, \cdots, \beta^P)$, $\beta = \pm 1$. Thus we can write:

$$\mathbf{Y}_{..r} = \bar{\mathbf{G}} D_r(\mathbf{F})\mathbf{L}D_r(\mathbf{C})\mathbf{\Delta_1}^{-1}\bar{\mathbf{S}}^T, r = 1, 2, \cdots, R.$$

Let $\mathbf{T}_r$ and $\mathbf{X}_r$ be non-singular matrices. An alternative representation of the tensor $\mathbb{Y}$ can be obtained by defining

$$\begin{aligned} D_r(\bar{\mathbf{F}}) &= D_r(\mathbf{F})\mathbf{T}_r^{-1} && (16)\\ \bar{\mathbf{L}} &= \mathbf{T}_r\mathbf{L}\mathbf{X}_r && (17)\\ D_r(\bar{\mathbf{C}}) &= \mathbf{X}_r^{-1}D_r(\mathbf{C})\mathbf{\Delta}_1^{-1} && (18)\end{aligned}$$

In order to keep the matrices in the left side of (16) and (18) diagonal, it is obvious that $\mathbf{T}_r$ and $\mathbf{X}_r$ must be diagonal. Recall that $\mathbf{L}$ is a column-wise Vandermonde matrix with 1's in its first row and column. Since $\bar{\mathbf{L}}$ must have the same Vandermonde structure, then

$$\bar{\mathbf{L}}_{.1} = \begin{pmatrix} t_{1,1;r}x_{1,1;r} \\ \vdots \\ t_{M_l,M_l;r}x_{1,1;r} \end{pmatrix} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}, \quad \text{and} \quad \bar{\mathbf{L}}^T_{1.} = \begin{pmatrix} t_{1,1;r}x_{1,1;r} \\ \vdots \\ t_{1,1;r}x_{P,P;r} \end{pmatrix} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix},$$

where $t_{i,i;r}$, and $x_{i,i;r}$ denote respectively the diagonal entries of $\mathbf{T}_r$ and $\mathbf{X}_r$. As a consequence, $t_{1,1;r} = t_{2,2;r} = \cdots = t_{M_l,M_l;r} = \frac{1}{x_{1,1;r}}$ and $x_{1,1;r} = x_{2,2;r} = \cdots = x_{P,P;r} = \frac{1}{t_{1,1;r}}$, i.e. $t_{1,1;r}x_{1,1;r} = 1$. Then equations (16)-(18) can be written as: $D_r(\bar{\mathbf{F}}) = \frac{1}{t_{1,1;r}}D_r(\mathbf{F})$, $\bar{\mathbf{L}} = x_{1,1;r}t_{1,1;r}\mathbf{L} = \mathbf{L}$, and $D_r(\bar{\mathbf{C}}) = \frac{1}{x_{1,1;r}}D_r(\mathbf{C})\mathbf{\Delta}_1^{-1}$. By considering all values of r, we get: $\bar{\mathbf{F}} = \mathbf{\Delta}_2\mathbf{F}$ and $\bar{\mathbf{C}} = \mathbf{\Delta}_2^{-1}\mathbf{C}\mathbf{\Delta}_1^{-1}$. We also have $\bar{\mathbf{F}} = \mathbf{\Delta}_2\boldsymbol{\phi}\mathbf{l}^T = \bar{\boldsymbol{\phi}}\bar{\mathbf{l}}^T$. Since from $\mathbf{L}$, $\mathbf{l}$ can be uniquely determined, necessarily $\bar{\boldsymbol{\phi}} = \boldsymbol{\phi}$ and then $\mathbf{\Delta}_2 = \mathbf{I}_R$. As a consequence: $\bar{\mathbf{C}} = \mathbf{C}\mathbf{\Delta}_1^{-1}$.■

REFERENCES

[1] M. Schetzen, *The Volterra and Wiener theories of nonlinear systems*. John Wiley and Sons, Inc., New-York, 1980.

[2] N. Wiener, *Nonlinear problems in random theory*. New-York: Technology press of MIT and John Wiley and Sons Inc., 1958.

[3] E. Biglieri, A. Gersho, R. Gitlin, and T. Lim, “Adaptive cancellation of nonlinear intersymbol interference for voiceband data transmission,” *IEEE J. Select. Areas Commun.*, vol. 2, pp. 765–777, 1984.

[4] X. N. Fernando and A. Sesay, “Adaptive asymmetric linearization of radio over fiber links for wireless access,” *IEEE Trans. on Vehicular Technology*, vol. 51, no. 6, pp. 1576–1586, November 2002.

[5] S. Prakriya and D. Hatzinakos, “Blind identification of LTI-ZMNL-LTI nonlinear channel models,” *IEEE Trans. on Signal Processing*, vol. 43, no. 12, pp. 3007–3013, December 1995.

[6] ——, “Blind identification of linear subsystems of LTI-ZMNL-LTI models with cyclostationary inputs,” *IEEE Trans. on Signal Processing*, vol. 45, no. 8, pp. 2023–2036, August 1997.

[7] R. Harshman, “Foundation of the PARAFAC procedure: models and conditions for an "explanatory" multimodal factor analysis,” *UCLA working papers in phonetics*, vol. 16, pp. 1–84, 1970.

[8] J. Caroll and J. Chang, “Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition,” *Psychometrika*, vol. 35, pp. 283–319, 1970.

[9] L. Tucker, “Some mathematical notes of three-mode factor analysis,” *Psychometrika*, vol. 31, pp. 279–311, 1966.

[10] R. Harshman and M. Lundy, “Uniqueness proof for a family of models sharing features of Tucker's three-mode factor analysis and PARAFAC/CANDECOMP,” *Psychometrika*, vol. 61, no. 1, pp. 133–154, 1996.

[11] J. Kruskal, “Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics,” *Linear Algebra Applicat.*, vol. 18, pp. 95–138, 1977.

[12] A. Kibangou, G. Favier, and A. de Almeida, “Blind identification of series-cascade nonlinear channels,” in *ASILOMAR Conference on Signals, Systems, and Computers*, Pacific Grove, CA, USA, October 30-November 2, 2005, pp. 422–426.

[13] A. Kibangou and G. Favier, “Wiener-Hammerstein systems modelling using diagonal Volterra kernels coefficients,” *IEEE Signal Proc. Letters*, vol. 13, no. 6, pp. 381–384, June 2006.

[14] R. Bro, “Multi-way analysis in the food industry. models, algorithms and applications.” Ph.D. dissertation, University of Amsterdam, Netherlands, 1998.

[15] A. Redfern and G. Tong Zhou, “A root method for Volterra system equalization,” *IEEE Signal Processing Letters*, vol. 5, no. 5, pp. 285–288, 1998.