Wiener–Hammerstein Systems Modeling Using Diagonal Volterra Kernels Coefficients

Alain Y. Kibangou and Gérard Favier

Abstract—**In this letter, we first present explicit relations between block-oriented nonlinear representations and Volterra models. For an identification purpose, we show that the estimation of the diagonal coefficients of the Volterra kernels associated with the considered block-oriented nonlinear structures is sufficient to recover the overall model. An alternating least squares-type algorithm is provided to carry out this model identification.**

Index Terms—**Alternating least squares (ALS), bilinear decomposition, parameter estimation, Toeplitz matrix, Volterra models, Wiener–Hammerstein models.**

I. Introduction

MANY nonlinear systems can be represented by one of the three following cascade models: a linear filter followed by a memoryless nonlinearity (Wiener model), a memoryless nonlinearity followed by a linear filter (Hammerstein model), or a memoryless nonlinearity sandwiched between two linear filters (Wiener–Hammerstein model). Block-oriented nonlinear structures combining memoryless nonlinearity with linearly dispersive elements [1], [2] have been successfully employed in various areas, including digital communications and physiological systems [3], [4]. For example, in wireless and satellite digital communication systems, it is usual to operate the RF amplifier at or near saturation to improve its power efficiency [5]. The overall transmission system can then be modeled as a memoryless analytic nonlinearity representing the near saturation amplifier characteristic as well as a bandpass time invariant linear system with finite impulse response, modeling the effects of the bandpass transmit filter and the propagation channel [6].

This letter is concerned with the study of Wiener–Hammerstein-type nonlinear structures admitting Wiener and Hammerstein models as particular cases. We denote by $u(n)$, $y(n)$, $v_1(n)$, and $v_2(n)$ the input signal, the output signal, and intermediate variables, 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. The Wiener–Hammerstein model, as depicted in Fig. 1, is constituted by a polynomial $C(.)$, sandwiched between two linear filters with

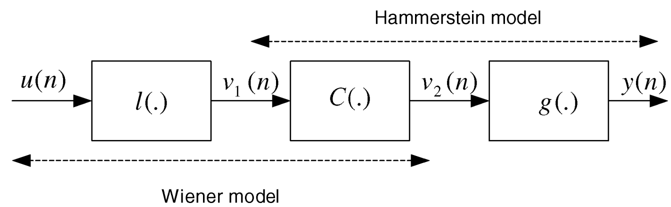

Fig. 1. Wiener–Hammerstein model.

impulse response $l(.)$ and $g(.)$ as well as memory M_l and M_g, respectively, i.e.,

$$v_1(n) = \sum_{i=0}^{M_l} l(i)u(n-i), \quad v_2(n) = \sum_{p=1}^{P} c_p v_1^p(n)$$
$$y(n) = \sum_{i=0}^{M_g} g(i)v_2(n-i).$$

The estimation of the parameters that characterize each block of the Wiener–Hammerstein model is not simple since this model is not linear with respect to its parameters. Instead of estimating the parameters of each block, it is more suitable to estimate the equivalent Volterra model, which is linear in its parameters. However, the number of parameters drastically increases with the nonlinearity order and the linear filters memories. In this letter, we show that the Volterra model associated with a Wiener–Hammerstein nonlinear system is completely characterized by the diagonal coefficients of its kernels. As a consequence, the effort dedicated to the estimation process of a Volterra model associated with a Wiener–Hammerstein structure can be limited to the diagonal coefficients estimation.

This letter is organized as follows. Section II gives the links between block-oriented nonlinear representations and Volterra models and states relationships between diagonal and nondiagonal coefficients of Volterra kernels associated with such block-oriented nonlinear representations when the nonlinearity is assumed to be a polynomial. In Section III, we present an alternating least squares (ALS)-type algorithm for estimating the diagonal kernels coefficients of the associated Volterra model. Some simulation results are provided to illustrate the performances of this algorithm. Conclusion and future work are given in Section IV.

Notations

T	Matrix transpose.
$\dagger$	Matrix pseudo-inverse.
δ	Kronecker symbol.
$\mathbf{A}_{.i}$	ith column of $\mathbf{A}$.
$vec(\mathbf{A})$	Stacking of the columns of $\mathbf{A}$ into a vector.

Manuscript received September 21, 2005; revised December 10, 2005. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Yue (Joseph) Wang.

The authors are with the Laboratoire I3S, CNRS-UNSA, Les Algorithmes-Bât. Euclide B, B.P. 121-06903 Sophia Antipolis Cedex, France (e-mail: favier@i3s.unice.fr; kibangou@i3s.unice.fr).

Digital Object Identifier 10.1109/LSP.2006.871705

$\mathbf{I}_N$	$(N \times N)$ identity matrix.
$\|\cdot\|_F$	Frobenius norm of a matrix.
$\|\cdot\|_2$	Euclidean norm of a vector.

II. Volterra Series Modeling of Block-Oriented Nonlinear Systems

The output of a discrete time Volterra model can be written as

$$y(n) = \sum_{p=1}^{P} \sum_{i_1=0}^{M} \cdots \sum_{i_p=0}^{M} h_p(i_1, \cdots, i_p) \prod_{j=1}^{p} u(n - i_j) \quad (1)$$

where $h_p(., \cdots, .)$ denotes the pth-order Volterra kernel. It is well known that the Volterra kernels $l_p(., \cdots, .)$ and $g_p(., \cdots, .)$, respectively associated with Wiener and Hammerstein models, are separable and given by

$$l_p(i_1, \cdots, i_p) = c_p \prod_{k=1}^{p} l(i_k), i_k = 0, \cdots, M_l, k = 1, \cdots, p \quad (2)$$

$$g_p(i, \cdots, i) = c_p g(i), \quad i = 0, \cdots, M_g \quad (3)$$

with $p = 1, \cdots, P$. The Wiener–Hammerstein model can be viewed as a concatenation of a Wiener model, represented by the Volterra kernels $l_p(., \cdots, .)$, followed by the linear filter $g(.)$.

By adopting the equivalent multichannel representation of the Volterra model associated with the Wiener–Hammerstein structure, the three representations in Fig. 2 are equivalent. Thus, the kernel $h_p(., \cdots, .)$ of this associated Volterra model results from the discrete convolution of the kernels $l_p(., \cdots, .)$ and $g(.)$, i.e., $h_p(i_1, \cdots, i_p) = \sum_{i=-\infty}^{\infty} g(i) l_p(i_1 - i, \cdots, i_p - i)$. By taking the finite memory and the causality properties of the filters $g(.)$ and $l(.)$ into account, we get

$$h_p(i_1, \cdots, i_p) = \sum_{i=0}^{M_g} g(i) l_p(i_1 - i, \cdots, i_p - i), \quad i_k = 0, \cdots, M$$

$$M = M_l + M_g, \quad k = 1, \cdots, p, \quad p = 1, \cdots, P. \quad (4)$$

Then, by using (2), the expression (4) of the Volterra kernels associated with a Wiener–Hammerstein model becomes

$$h_p(i_1, \cdots, i_p) = c_p \sum_{i=0}^{M_g} g(i) \prod_{k=1}^{p} l(i_k - i). \quad (5)$$

By setting respectively, $M_g = 0$ and $g(i) = \delta_{0,i}$, and $M_l = 0$ and $l(i) = \delta_{0,i}$, in (5), we get formulae (2) and (3), giving the Volterra kernels associated with Wiener and Hammerstein models. One can note that a Hammerstein model corresponds to a diagonal Volterra model, while the nondiagonal kernels coefficients associated with Wiener and Wiener–Hammerstein models are nonzero. In the sequel, we show that these nondiagonal kernels coefficients can be expressed in terms of the diagonal ones.

A. Wiener Model

Let us consider the Wiener model constituted by the serial concatenation of the filter $l(.)$ and the polynomial $C(.)$. We prove the following theorem.

Theorem 1: The nondiagonal coefficients of the pth-order Volterra kernel $l_p(., \cdots, .)$, $p = 1, \cdots, P$, associated with a Wiener model are linked with the diagonal coefficients $l_p(i) = l_p(i, \cdots, i)$, $i = 0, \cdots, M_l$, of the same kernel by the following formula:

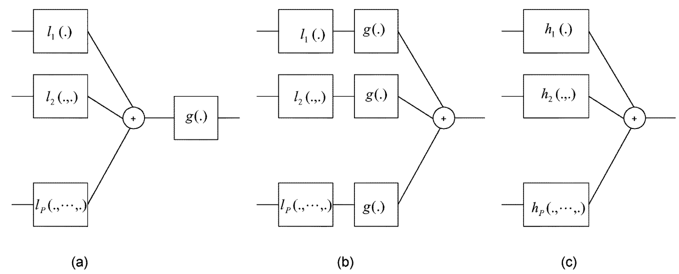

Fig. 2. Multichannel representations of a Wiener–Hammerstein structure.

$$l_p(i_1, \cdots, i_p) = \frac{1}{M_l + 1} \sum_{i=0}^{M_l} \frac{l_p(i)}{(l_1(i))^p} \prod_{j=1}^{p} l_1(i_j). \quad (6)$$

Proof: From (2), we have $l_p(i_1, \cdots, i_p) = c_p \prod_{j=1}^{p} l(i_j)$, $p = 1, \cdots, P$. Then, the diagonal coefficients of the kernel l_p are such as $l_p(i) = c_p(l(i))^p$, $i = 0, \cdots, M_l$, which gives $c_p = (l_p(i)/(l(i))^p)$. By summing this relation for all the possible values of $i = 0, \cdots, M_l$, we get $c_p = (1/(M_l + 1)) \sum_{i=0}^{M_l} (l_p(i)/(l(i))^p)$. Knowing that $l_1(i) = c_1 l(i)$, we deduce the following relation: $(c_p/c_1^p) = (1/(M_l + 1)) \sum_{i=0}^{M_l} (l_p(i)/(l_1(i))^p)$. So, the expression (2) of the pth-order kernel coefficients becomes

$$l_p(i_1, \cdots, i_p) = \frac{c_p}{c_1^p} \prod_{j=1}^{p} l_1(i_j) = \frac{1}{M_l + 1} \sum_{i=0}^{M_l} \frac{l_p(i)}{(l_1(i))^p} \prod_{j=1}^{p} l_1(i_j).$$

■

The Wiener model is therefore completely characterized by the diagonal coefficients of its associated Volterra kernels.

B. Wiener–Hammerstein Model

From (4), we can conclude that the diagonal coefficients of the Volterra kernels associated with a Wiener–Hammerstein model are given by

$$h_p(j) = h_p(j, \cdots, j) = \sum_{i=0}^{M_g} g(i) l_p(j - i), \ j = 0, \cdots, M. \quad (7)$$

Then, the generalization of (6) to Wiener–Hammerstein models is stated by the following theorem.

Theorem 2: The nondiagonal coefficients of the pth-order Volterra kernel $h_p(., \cdots, .)$, $p = 1, \cdots, P$, associated with a Wiener–Hammerstein model are given by

$$h_p(i_1, \cdots, i_p) = \frac{1}{M_l + 1} \sum_{k=0}^{M_g} \sum_{i=0}^{M_l} g(k) \frac{l_p(i)}{(l_1(i))^p} \prod_{j=1}^{p} l_1(i_j - k) \quad (8)$$

where $l_p(i)$, $i = 0, \cdots, M_l$, are the diagonal coefficients of the pth-order Volterra kernel associated with the Wiener part of the model, i.e., the cascade of the first linear system and the static nonlinearity.

Proof: Formula (8) directly results from (4) in replacing the coefficients $l_p(., \cdots, .)$ by their expression (6). ■

III. Estimation of the Nondiagonal Coefficients of the Volterra Model Associated With a Wiener–Hammerstein Model

In practice, for computing the nondiagonal coefficients of the equivalent Volterra model, Theorem 2 implies the need to estimate the coefficients $g(i)$, $i = 0, \cdots, M_g$ and $l_p(i)$, $i = 0, \cdots, M_l$, $p = 1, \cdots, P$. For this purpose, (7) is rewritten in the following matrix form:

$$\mathbf{H} = \mathbf{G}\mathbf{L} \tag{9}$$

with

$$\mathbf{G} = \begin{pmatrix} g(0) & 0 & \cdots & 0 \\ g(1) & g(0) & & \vdots \\ & g(1) & \ddots & 0 \\ \vdots & & \ddots & g(0) \\ & \vdots & & g(1) \\ g(M_g) & & \ddots & \\ 0 & g(M_g) & & \vdots \\ \vdots & \vdots & \ddots & \\ 0 & 0 & \cdots & g(M_g) \end{pmatrix} \in \Re^{(M+1)\times(M_l+1)} \tag{10}$$

$$\mathbf{L} = \begin{pmatrix} l_1(0) & l_2(0) & \cdots & l_P(0) \\ l_1(1) & l_2(1) & \cdots & l_P(1) \\ \vdots & \vdots & \ddots & \vdots \\ l_1(M_l) & l_2(M_l) & \cdots & l_P(M_l) \end{pmatrix} \in \Re^{(M_l+1)\times P}$$

$$\mathbf{H} = \begin{pmatrix} h_1(0) & h_2(0) & \cdots & h_P(0) \\ h_1(1) & h_2(1) & \cdots & h_P(1) \\ \vdots & \vdots & \ddots & \vdots \\ h_1(M) & h_2(M) & \cdots & h_P(M) \end{pmatrix} \in \Re^{(M+1)\times P}.$$

$\mathbf{G}$ and $\mathbf{L}$ are the factors of a bilinear decomposition of $\mathbf{H}$. Generally, this kind of decomposition is not unique. However, due to the Toeplitz structure of $\mathbf{G}$, uniqueness conditions are established in the lemma below. Then, assuming that $\mathbf{H}$ is known, an ALS-type algorithm is proposed to estimate the factors $\mathbf{G}$ and $\mathbf{L}$, which are used to determine the nondiagonal kernels coefficients of the associated Volterra model.

Lemma 1: Given $\mathbf{H}$, its factorization as $\mathbf{GL}$, with $\mathbf{G}$ in Toeplitz form, is unique up to scalar factors α and α^{-1} for $\mathbf{G}$ and $\mathbf{L}$, respectively, with α being a nonzero constant.

Proof: For any nonsingular matrix $\mathbf{T}$ of order (M_l+1), the set of matrices $(\tilde{\mathbf{G}}, \tilde{\mathbf{L}}) = (\mathbf{GT}, \mathbf{T}^{-1}\mathbf{L})$ is such as $\mathbf{H} = \tilde{\mathbf{G}}\tilde{\mathbf{L}} = \mathbf{GL}$. Since $\tilde{\mathbf{G}} = \mathbf{GT}$ must have the same Toeplitz structure as in (10), we get $\tilde{\mathbf{G}}_{.i} = \mathbf{GT}_{.i}$, $i = 1, \cdots, M_l + 1$. For any $i = 1, \cdots, M_l + 1$, $\mathbf{G}$ can be partitioned as follows:

$$\mathbf{G} = \begin{pmatrix} \mathbf{G}_1^{(i)} & 0_{(i-1)\times 1} & 0_{(i-1)\times(M_l-i+1)} \\ \mathbf{G}_2^{(i)} & \mathbf{g} & \mathbf{G}_3^{(i)} \\ 0_{(M_l-i+1)\times(i-1)} & 0_{(M_l-i+1)\times 1} & \mathbf{G}_4^{(i)} \end{pmatrix}$$

where $\mathbf{g} = (g(0)\ g(1)\ \cdots\ g(M_g))^T$, and

$$\mathbf{G}_1^{(i)} = \begin{pmatrix} g(0) & 0 & \cdots & 0 \\ g(1) & g(0) & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ g(i-2) & \cdots & g(1) & g(0) \end{pmatrix} \in \Re^{(i-1)\times(i-1)}$$

$$\mathbf{G}_4^{(i)} = \begin{pmatrix} g(M_g) & g(M_g-1) & \cdots & g(M_g-M_l+i) \\ 0 & g(M_g) & \ddots & \vdots \\ \vdots & \vdots & \ddots & g(M_g-1) \\ 0 & 0 & 0 & g(M_g) \end{pmatrix}$$

$\in \Re^{(M_l-i+1)\times(M_l-i+1)}$. By considering the following partitioning: $\mathbf{T}_{.i} = (\mathbf{t}_1^{(i)T}\ t_{i,i}\ \mathbf{t}_2^{(i)T})^T$, with $\mathbf{t}_1^{(i)}$ and $\mathbf{t}_2^{(i)}$ being, respectively, $(i-1)\times 1$ and $(M_l-i+1)\times 1$ vectors, we get

$$\tilde{\mathbf{G}}_{.i} = \begin{pmatrix} 0_{(i-1)\times 1} \\ \tilde{\mathbf{g}} \\ 0_{(M_l-i+1)\times 1} \end{pmatrix} = \begin{pmatrix} \mathbf{G}_1^{(i)}\mathbf{t}_1^{(i)} \\ \mathbf{G}_2^{(i)}\mathbf{t}_1^{(i)} + t_{i,i}\mathbf{g} + \mathbf{G}_3^{(i)}\mathbf{t}_2^{(i)} \\ \mathbf{G}_4^{(i)}\mathbf{t}_2^{(i)} \end{pmatrix}$$

where $\tilde{\mathbf{g}} = (\tilde{g}(0)\ \tilde{g}(1)\ \cdots\ \tilde{g}(M_g))^T$. Assuming that $g(0) \neq 0$ and $g(M_g) \neq 0$, the triangular matrices $\mathbf{G}_1^{(i)}$ and $\mathbf{G}_4^{(i)}$ are full rank. Then the equations $\mathbf{G}_1^{(i)}\mathbf{t}_1^{(i)} = 0_{(i-1)\times 1}$ and $\mathbf{G}_4^{(i)}\mathbf{t}_2^{(i)} = 0_{(M_l-i+1)\times 1}$ imply, respectively, $\mathbf{t}_1^{(i)} = 0$ and $\mathbf{t}_2^{(i)} = 0$; that means for any $i = 1, \cdots, M_l + 1$, all the components of $\mathbf{T}_{.i}$ are equal to zero, except $t_{i,i}$. So we get $\tilde{\mathbf{g}} = t_{i,i}\mathbf{g}$, $i = 1, \cdots, M_l+1$, i.e., $t_{i,i} = \alpha, \forall i = 1, \cdots, M_l + 1$. Consequently, $\mathbf{T} = \alpha \mathbf{I}_{M_l+1}$, showing that $\mathbf{G}$ and $\mathbf{L}$ are unique up to scalar factors α and α^{-1}, respectively. ■

Thanks to Lemma 1, using the bilinear decomposition (9) of $\mathbf{H}$ allows to get the coefficients $g(i)$, $i = 0, \cdots, M_g$, and $l_p(i)$, $i = 0, \cdots, M_l$, $p = 1, \cdots, P$, up to scaling factors α and $1/\alpha$, respectively. Then, the nondiagonal coefficients of the Volterra kernels associated with the Wiener–Hammerstein model can be obtained by means of formula (8). Note that the scaling factors resulting from the factorization (9) are compensated in (8), which allows to get all the coefficients without ambiguity.

In the sequel, by assuming that $\mathbf{H}$ is known, we propose an ALS procedure to determine the factors $\mathbf{G}$ and $\mathbf{L}$ in (9). The basic idea behind ALS is simple. At each iteration, there are two steps. In the first one, matrix $\mathbf{G}$ is updated using the least squares (LS) algorithm to minimize the cost function $\|\mathbf{H} - \mathbf{GL}\|_F^2$ conditioned on the previously obtained estimate of the matrix $\mathbf{L}$; in the second step, the same procedure is applied to update $\mathbf{L}$. These two steps are repeated until convergence of the algorithm. With $\mathbf{G}$ being a Toeplitz matrix, we have the following equality: $\mathbf{H}_{.p} = \mathbf{GL}_{.p} = \mathbb{T}(\mathbf{L}_{.p})\mathbf{g}$, where $\mathbb{T}(\mathbf{L}_{.p})$ is the $(M+1)\times(M_g+1)$ Toeplitz matrix constructed from the pth column of $\mathbf{L}$

$$\mathbb{T}(\mathbf{L}_{.p}) = \begin{pmatrix} l_p(0) & 0 & \cdots & 0 \\ l_p(1) & l_p(0) & & \vdots \\ \vdots & l_p(1) & \ddots & 0 \\ l_p(M_l) & \vdots & \ddots & l_p(0) \\ 0 & l_p(M_l) & & l_p(1) \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & l_p(M_l) \end{pmatrix}. \tag{11}$$

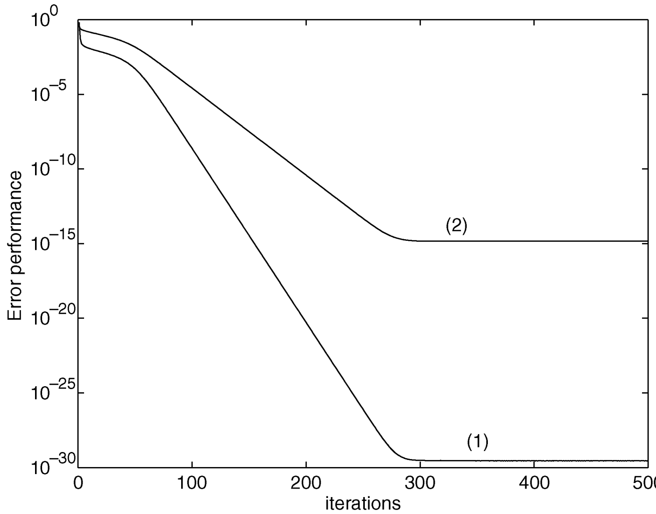

Fig. 3. (1) Square estimation error of $\mathbf{H}$. (2) NMP on $\hat{\mathbf{g}}$.

Thus, by defining the matrix

$$\boldsymbol{\Lambda} = \left(\mathbb{T}(\mathbf{L}_{.1})^T \cdots \mathbb{T}(\mathbf{L}_{.P})^T\right)^T \tag{12}$$

we have $\|\mathbf{H}-\mathbf{G}\mathbf{L}\|_F^2 = \|vec(\mathbf{H})-\boldsymbol{\Lambda}\mathbf{g})\|_2^2$. Then, the optimization problem is reformulated as follows:

$$\hat{\mathbf{g}} = \arg\min_{\mathbf{g}} = \|vec(\mathbf{H}) - \boldsymbol{\Lambda}\mathbf{g}\|_2^2 \tag{13}$$

which leads to the classical LS solution

$$\hat{\mathbf{g}} = \boldsymbol{\Lambda}^{\dagger} vec(\mathbf{H}). \tag{14}$$

Once $\hat{\mathbf{g}}$ has been determined, minimization of the cost function $\|\mathbf{H} - \hat{\mathbf{G}}\mathbf{L}\|_F^2$ with respect to $\mathbf{L}$ leads to the LS solution

$$\hat{\mathbf{L}} = \hat{\mathbf{G}}^{\dagger}\mathbf{H}. \tag{15}$$

The estimation procedure can then be summarized as follows.

Algorithm

1) Given the matrix $\mathbf{H}$, estimate the $\mathbf{G}$ and $\mathbf{L}$ factors by means of an ALS algorithm.
   a) Initialize $\hat{\mathbf{L}}_0$, and set $k = 0$.
   b) $k = k + 1$.
      - Construct $\boldsymbol{\Lambda}_k$ from $\hat{\mathbf{L}}_k$, using (11) and (12).
      - Compute $\hat{\mathbf{g}}_k = \boldsymbol{\Lambda}_k^{\dagger} vec(\mathbf{H})$, and then construct $\hat{\mathbf{G}}_k$, as in (10).
      - Compute $\hat{\mathbf{L}}_k = \hat{\mathbf{G}}_k^{\dagger}\mathbf{H}$.
   c) Repeat 1b) until convergence of the algorithm, i.e., $\|\mathbf{H} - \hat{\mathbf{G}}_k\hat{\mathbf{L}}_k\|_2^F \leq \varepsilon$, where ε is a very small positive constant.
2) Estimate of the nondiagonal coefficients of the Volterra kernels associated with the Wiener–Hammerstein model by means of formula (8).

Note that, under some restrictive conditions, instead of an ALS procedure, a subspace method can also be used to achieve the factorization (9) (see [7]).

Example: We consider a Wiener–Hammerstein system such that the linear filters and the polynomial coefficients are, respectively, given by $\mathbf{g} = (1,\ 0.5,\ 0.2)^T$, $\mathbf{l} = (1\ -0.3)^T$, and $\mathbf{c} = (2,\ 0.8,\ 0.5)^T$. We evaluate the performance of the above-described factorization algorithm by means of two measures: the square error $\|\mathbf{H} - \hat{\mathbf{G}}\hat{\mathbf{L}}\|_F^2$, i.e., the minimized criterion, and the normalized projection misalignment (NPM) on the vector $\hat{\mathbf{g}}$, defined as NPM $= \|\boldsymbol{\varepsilon}\|/\|\mathbf{g}\|$, where $\boldsymbol{\varepsilon} = \mathbf{g} - (\mathbf{g}^T\hat{\mathbf{g}}/\hat{\mathbf{g}}^T\hat{\mathbf{g}})\hat{\mathbf{g}}$. By projecting $\mathbf{g}$ onto $\hat{\mathbf{g}}$ and defining the projection error $\boldsymbol{\varepsilon}$, we only take the undesirable misalignment of the estimate of $\mathbf{g}$ into account, disregarding an arbitrary scaling factor inherently associated with it [8]. The plots of Fig. 3 were obtained by averaging the results of 500 independent runs with random initialization. One can note that the algorithm converges for both considered criteria. As the algorithm is randomly initialized, some misconvergence cases can occur. In such cases, the algorithm must be rerun with another initialization.

IV. Conclusion

In this letter, explicit relationships between block-oriented nonlinear systems and Volterra models have been derived. We have shown that Volterra kernels, associated with Wiener and Wiener–Hammerstein structures, are completely characterized by their diagonal coefficients. This result is particularly useful. It highlights that, for Volterra models associated with block-oriented nonlinear systems, the parameter estimation problem can be reduced to the estimation of the diagonal coefficients. In order to restrict the estimation process to the diagonal Volterra kernels, a special input can be designed, as it will be shown in a companion paper [9].

References

[1] N. Wiener, *Nonlinear Problems in Random Theory*. New York: Wiley, 1958.
[2] S. Billings and S. Fakhouri, "Identification of systems containing linear dynamics and static nonlinear elements," *Automatica*, vol. 18, pp. 15–26, 1982.
[3] N. Bershad, P. Celka, and S. Mc Laughlin, "Analysis of stochastic gradient identification of Wiener-Hammerstein systems for nonlinearities with Hermite polynomial expansions," *IEEE Trans. Signal Process.*, vol. 49, no. 5, pp. 1060–1072, May 2001.
[4] V. Z. Marmarelis, "Identification of nonlinear biological systems using Laguerre expansions of kernels," *Ann. Biomed. Eng.*, vol. 21, pp. 573–589, 1993.
[5] K. Feher, *Digital Communications-Satellite/Earth Station Engineering*. Englewood Cliffs, NJ: Prentice-Hall, 1993.
[6] R. Lopez-Valcarce and Dasgupta, "Second-order statistical properties of nonlinearly distorted phase-shift keyed (PSK) signals," *IEEE Commun. Lett.*, vol. 7, no. 7, pp. 323–325, Jul. 2003.
[7] L. Tong and S. Perreau, "Multichannel blind identification: From subspace to maximum likelihood methods," *Proc. IEEE*, vol. 86, no. 10, pp. 1951–1968, Oct. 1998.
[8] D. Morgan, J. Benesty, and M. Sondhi, "On the evaluation of estimated impulse responses," *IEEE Signal Process. Lett.*, vol. 5, no. 7, pp. 174–176, Jul. 1998.
[9] A. Kibangou and G. Favier, "Input design for estimating the diagonal kernels of Volterra models," in preparation.