Maximum Likelihood based Identification for Nonlinear Multichannel Communications Systems

Ouahbi Rekik, Karim Abed-Meraim, Mohamed Nait Meziane, Anissa Mokraoui, Nguyen Linh-Trung

Maximum Likelihood based Identification for Nonlinear Multichannel Communications Systems

Ouahbi Rekik*, Karim Abed-Meraim#, Mohamed Nait-Meziane#, Anissa Mokraoui*, and Nguyen Linh Trung+
*L2TI, UR 3043, Université Sorbonne Paris Nord, France,
firstname.lastname@univ-paris13.fr
#PRISME, Université d'Orléans, France,
firstname.lastname@univ-orleans.fr
+AVITECH, University of Engineering and Technology, Vietnam National University, Hanoi, Vietnam,
linhtrung@vnu.edu.vn [1]

Abstract

Nonlinear distortions are important issues in many communications systems. Therefore, this paper deals with the blind and semi-blind identification of nonlinear SIMO/MIMO channels. Quadratic and cubic nonlinearities are considered for the system model as well as a discussion on how the developed work can be extended to more general nonlinear models. The proposed blind solution is initialized by using a subspace approach, which is followed by an appropriate ambiguity removal method, then refined by a Maximum Likelihood (ML) based processing using the Expectation-Maximization (EM) algorithm. The proposed semi-bind solution, involving both data and pilots, is fully based on the EM algorithm. These solutions are supported by some identifiability results and performance bounds analysis related to the considered models (blind and semi-blind). Finally, simulation results essentially show that the proposed algorithms exhibit very attractive channel estimation performance, with interesting convergence speed for the EM-based iterative processing.

1. Introduction

Nonlinear behaviors can be encountered in many practical situations, in which case appropriate (nonlinear) processing is needed, when such nonlinearities are too important to be disregarded [1, 2]. Indeed, because most of real-life systems are inherently nonlinear in nature, nonlinear problems have drawn important interest and extensive attention from engineers, physicists, mathematicians and many other scientists [2]. In communications systems, and due

[1]This work was supported by the National Foundation for Science and Technology Development of Vietnam under Grant No. 01/2019/TN.

to the presence of nonlinear devices such as optical equipments [3, 4] and power amplifiers, even in some MIMO-OFDM [5] and massive MIMO scenarios [6, 7] or millimeter wave based systems [8], communication channels are sometimes corrupted by nonlinear distortions such as nonlinear inter-symbol interference, nonlinear multiple access interference and nonlinear inter-carrier interference. These nonlinear distortions can significantly deteriorate the signal reception, leading to poor system performance. In order to overcome such an issue, nonlinear models are adopted to provide an accurate channel representation and to allow the development of efficient signal processing techniques capable of mitigating these nonlinear distortions. In the case of system identification, a widely used class of nonlinear models is the class of linear-in-the-parameters models. The input-output relation is essentially nonlinear but the estimation problem is linear with respect to the channel coefficients. Popular examples are polynomial filters, and more particularly Volterra filters [9]. They have been applied in many fields such as, electronic and electrical engineering, mechanical engineering, aeroelasticity problems and control engineering [10]. Indeed, the motivation for adopting these filters is that, they have the ability of modeling the behavior of nonlinear real-life phenomena, especially the ability to capture their "memory" effects [2]; and have mathematical relationship with other nonlinear system models namely the Wiener series, Hammerstein model, Wiener model, Wiener-Hammerstein model (block-oriented nonlinear systems), Taylor series or NARMAX model [10].

For nonlinear system identification, several approaches, most of them based on Volterra filters, have been proposed in the literature. Some works exploited training sequences and are essentially based on Least-Mean-Squares adaptive filters [11], Recursive Least-Squares algorithms [12, 13], and Affine Projection algorithms [14]. Other approaches are fully blind, thus, they seek to determine the system's kernel using the output data only. One could cite the higher order output cumulant-based approach [15], the subspace-based approach [16], the genetic programming-based method using Volterra filter [17], the tensor-based frameworks in [18], or the Reversible Jump Markov Chain Monte Carlo approach [19]. One can notice that, these methods have been adopted and adapted to nonlinear systems mainly due to their efficiency for the linear case. Consequently, and due to the attractive advantages of the maximum-likelihood (ML) approaches (which is used in the current work), namely the consistency, and the asymptotic efficiency of the estimates, some works proposed ML-based identification techniques of certain nonlinear systems [20, 21]. In these works, an approximation of the complex likelihood function is minimized via modified Gauss-Newton methods assuming the input data to be white Gaussian and a block-structured system model. However, a review of the current literature reveals that an ML solution for the case of nonlinear, finite alphabet, multi-channel communications systems does not exist.

Other works, like those in [22, 23], have considered a Hammerstein model with cascaded nonlinear and linear blocks, where the initialization and the system identification (channel estimation) are performed by firstly estimating the impulse response of the linear filter, which is then used to estimate the nonlin-

ear function parameters. By contrast, in the proposed work, both linear and nonlinear parameters are estimated simultaneously through solutions that fit into the framework of joint channel estimation and data detection.
Also, in [24] a blind nonlinear system identification is proposed based on the parallel factors (PARAFAC) tensor decomposition. However, it is shown that the input signals must satisfy some orthogonality constraints associated with the channel nonlinearities in order to allow the desired PARAFAC decomposition. Hence, a precoding scheme is introduced using temporal redundancy on the signals, which is carried out by imposing some constraints on the symbol transitions.

The aim of the current paper is to present ML-like blind and semi-blind channel estimators for Volterra-like nonlinear Single-Input Multiple-Output (SIMO) systems, that can be easily extended to MIMO scenarios. The proposed blind channel estimator combines a subspace-based estimation and an EM-based one. More precisely, firstly we exploit the received signal Second Order Statistics (SOS) using a subspace approach for channel estimation where the nonlinear SIMO system is treated as a linear Multiple-Input Multiple-Output (MIMO) system. A straightforward motivation is that, the use of SOS-based estimators avoids the need of high number of data symbols often required for High Order Statistics based methods, e.g. [15]. Then, unlike many blind-based works (e.g. [16]), we propose also an original method to remove the ambiguity inherent to such a blind approach. Finally, a second estimation is performed based on a maximum likelihood approach, where an iterative optimization is performed using the Expectation-Maximization (EM) algorithm. Indeed, due to its sensitivity to initialization, the EM-based estimator is initialized using the subspace-based one. Note that, efficient and practical initialization for blind EM-based techniques is often missing in the literature. Moreover, within the proposed EM-based framework, one could perform a joint channel estimation and data detection as will be highlighted later in the paper. The global scheme of the proposed blind EM-based estimator is given in Figure 1, in which the signal at the received antennas is the input for the subspace-based estimator, whereas the estimates of linear and nonlinear channel coefficients represent the output of the EM algorithm after convergence.

To make our method more flexible, and to consider the case where training sequences (pilots) are available, this work is extended to the semi-blind framework where data and pilot symbols are jointly exploited to improve the estimation accuracy and overcome certain limitations of the blind processing. In this case, the initialization of the EM algorithm is performed by exploiting the available pilots. The proposed blind and semi-blind approaches are supported by some identifiability results and performance bounds related to our context, that allow the reader getting more insights on the problem's identifiability and its inherent performance limits.

The rest of the paper is organized as follows. The adopted system model is presented in section 2. Section 3 describes the proposed blind channel estimation scheme. It starts by presenting the derivation of the subspace-based estimation, used for the initialization of our ML solution. Since this method suffers from

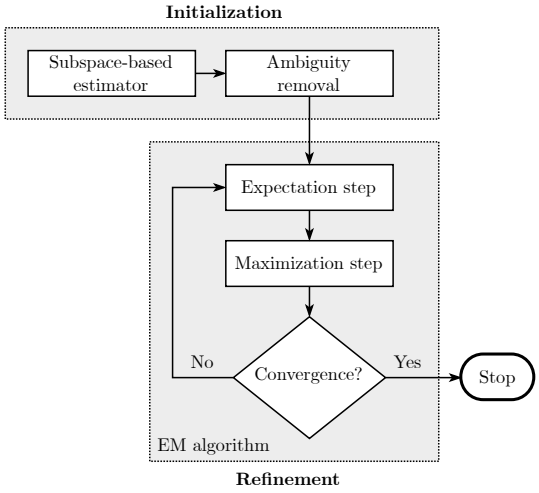


Figure 1: Proposed blind channel estimation scheme.

inherent matrix indeterminacy, it is followed by the introduction of an original ambiguity removal technique. Then, the ML-based method which exploits the EM algorithm is presented in details at the end of this section. Motivated by the widespread use of pilots in communications systems, we then derive, in section 4, an extension to the aforementioned ML method in a semi-blind context. In this section, we first detail the proposed semi-blind EM algorithm, then we provide a discussion on extending it to the MIMO case. In section 5, some identifiability results and performance bounds are given to corroborate the proposed solutions. Section 6 is dedicated to certain comments related to the computational complexity of our algorithms and a discussion on their potential extension to other nonlinear models. Section 7 provides comparative simulation result analysis of the proposed algorithms. Finally, the last section contains concluding remarks

Notations

In the sections below, the following notations have been adopted. Lower-case letters (e.g., x) denote scalars; lower-case boldface letters (e.g., $\mathbf{x}$) denote (column) vectors, and upper-case boldface letters (e.g., $\mathbf{X}$) denote matrices. Operators $(\cdot)^*, (\cdot)^T, (\cdot)^H, (\cdot)^\#$, and $\mathrm{tr}\,(\cdot)$ stand respectively for complex-conjugation, transposition, Hermitian transposition, matrix pseudo-inverse, and the trace of a matrix. Operator $\mathrm{diag}\,(\mathbf{x})$ denotes a diagonal matrix with entries of $\mathbf{x}$ as diagonal elements. $\mathbf{I}_M$ stands for an $M \times M$ identity matrix; $\mathbf{0}_{a\times b}$ is the all-zeros matrix of size $a \times b$; $\mathrm{vec}\,(\cdot)$ denotes the matrix (column) vectorization operator; and $\star$ denotes the convolution operator.

2. System model

This section details the data model adopted in this paper. A nonlinear SIMO system is considered as illustrated in Figure 2. It is composed of one

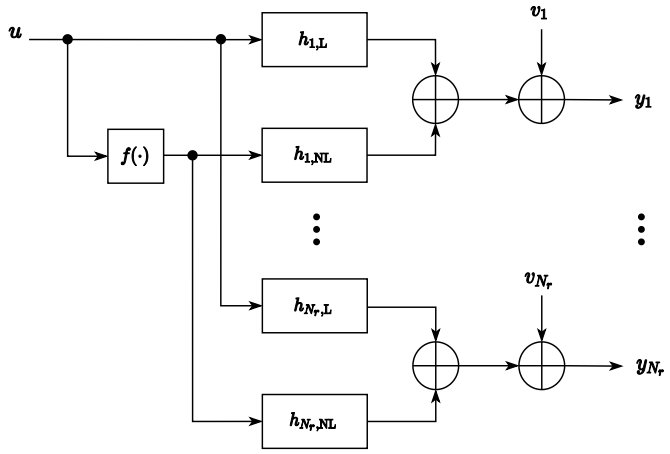


Figure 2: Architecture of the considered SIMO system, where $f(.)$ is a nonlinear function w.r.t. to the linear signal $\mathbf{u}$.

single-antenna transmitter and a receiver equipped with $N_r > 2$ antennas. The k-th received signal at the r-th receive antenna, denoted $y_r(k)$ with $1 \leq r \leq N_r$, is given by:

$$y_r(k) = \sum_{n=0}^{M_{r,\mathrm{L}}} h_{r,\mathrm{L}}(n)u(k-n) + \sum_{n=0}^{M_{r,\mathrm{NL}}} h_{r,\mathrm{NL}}(n)\tilde{u}(k-n) + v_r(k), \qquad (1)$$

where $h_{r,\mathrm{L}}(n)$ (resp. $h_{r,\mathrm{NL}}(n)$) refers to the elements of the linear (resp. nonlinear), r-th receiver, channel's finite impulse response coefficient vector of size $M_{r,\mathrm{L}} + 1$ (resp. $M_{r,\mathrm{NL}} + 1$), $u(k)$ is the transmitted (input) symbol sequence assumed to be independent and identically distributed (i.i.d.) complex random variables taking values, with equal probabilities, in a finite alphabet set $\mathcal{A} = \{a_1, a_2, \ldots, a_{2^B}\}$ where B is the number of bits per symbol. $\tilde{u}(k)$ stands for a nonlinear combination of the input signal so that $\tilde{u}(k) = f(u(k), u(k-1), \ldots)$ (f being an appropriate nonlinear function, chosen to accurately model the system's non-linearity) and $v_r(k)$ is a white circular Gaussian noise (uncorrelated from sensor to sensor) with variance σ_v^2.

In the sequel, two models for the nonlinear function $\tilde{u}(k)$ will be considered. The first one is a second-order (quadratic) model (e.g., [11, 25]), where $\tilde{u}(k) = u(k)^2$, whereas the second one is a third-order (cubic) model (e.g., [15, 26, 3])

where $\tilde{u}(k) = |u(k)|^2 u(k)$. These models have been used to model real-life nonlinearties, namely those related to power amplifiers and optical devices.

Since the system model is nonlinear with respect to the transmitted signal but linear in regards to the channel coefficients, we propose to treat this nonlinear SIMO model as a linear MIMO model with two inputs ($u(k)$ and $\tilde{u}(k)$) [27]. For this, let us define the following vectors: $\mathbf{h}_{r,\mathrm{L}} = [h_{r,\mathrm{L}}(0), \ldots, h_{r,\mathrm{L}}(M)]^T$, $\mathbf{h}_{r,\mathrm{NL}} = [h_{r,\mathrm{NL}}(0), \ldots, h_{r,\mathrm{NL}}(M)]^T$, $\mathbf{u}(k) = [u(k), u(k-1), \ldots, u(k-M)]^T$ and $\tilde{\mathbf{u}}(k) = [\tilde{u}(k), \tilde{u}(k-1), \ldots, \tilde{u}(k-M)]^T$ where $M = \max(M_{r,\mathrm{L}}, M_{r,\mathrm{NL}})$, $1 \leq r \leq N_r$. The received signal, given by equation (1), can be represented as:

$$y_r(k) = \mathbf{h}_r^T \bar{\mathbf{u}}(k) + v_r(k), \tag{2}$$

where $\mathbf{h}_r = [\mathbf{h}_{r,\mathrm{L}}^T, \mathbf{h}_{r,\mathrm{NL}}^T]^T$ and $\bar{\mathbf{u}}(k) = [\mathbf{u}(k)^T, \tilde{\mathbf{u}}(k)^T]^T$. Moreover, by considering the N_r receive antennas, one can write:

$$\mathbf{y}(k) = \mathbf{H}\bar{\mathbf{u}}(k) + \mathbf{v}(k), \tag{3}$$

where $\mathbf{y}(k) = [y_1(k), \ldots, y_{N_r}(k)]^T$, $\mathbf{H} = [\mathbf{h}_1, \ldots, \mathbf{h}_{N_r}]^T$ and $\mathbf{v}(k) = [v_1(k), \ldots, v_{N_r}(k)]^T$.

The system model, provided by equation (3), is considered as a Markov process where the state vector defined as $\mathbf{s}(k) = [u(k-1), \ldots, u(k-M)]^T$ contains M successive symbols. The set of $N = 2^{BM}$ states is denoted $\mathcal{Q}_{state} = \{\mathbf{q}_1, \ldots, \mathbf{q}_N\}$. The transition vector is defined as $\mathbf{x}_{nm} = [u(k), \ldots, u(k-M)]^T$, containing $(M+1)$ symbols associated with the transition between two connected successive states $\mathbf{q}_n$ and $\mathbf{q}_m$. The set of the $2^{B(M+1)}$ possible transitions is denoted $\mathcal{X}$. For convenience, sequences of observations $\mathbf{y}(n), \ldots, \mathbf{y}(m)$ and states $\mathbf{s}(n), \ldots, \mathbf{s}(m)$ are denoted $\mathbf{Y}_{[n:m]}$ and $\mathbf{S}_{[n:m]}$, respectively. Note that the number of possible transitions and states depends only on the number of linear terms since the nonlinear terms are directly obtained from the former.

In what follows, the parameters to be estimated are the channel coefficients and the noise variance, grouped in a single vector denoted $\boldsymbol{\theta} = [\mathrm{vec}\left(\mathbf{H}^T\right)^T, \sigma_v^2]^T$.

3. Blind EM-based estimation

This section details the proposed blind channel estimation approach. A subspace-based estimation, for the nonlinear SIMO system, is firstly considered. Then, solutions for ambiguity removal are proposed to get rid of the inherent ambiguity of the blind processing. After that, an EM-based channel estimation is detailed, which helps refining the subspace-based estimation already performed, as illustrated in Figure 1. Besides, a data estimation scheme, within the EM framework, is provided.

3.1. Subspace-based estimation

Blind subspace-based techniques have been used successfully for channel estimation based on Second Order Statistics (SOS) in the case of linear MIMO systems [28]. In what follows, the nonlinear SIMO system will be considered as a linear MIMO ($2 \times N_r$) system, where the term $\tilde{u}(k)$ will be treated as the second source signal. The following assumptions are supposed to hold:

- The polynomial matrix $\mathbf{H}(z) = [h_{\mathrm{L}}(z), h_{\mathrm{NL}}(z)] = \sum_{l=0}^{M} \mathbf{H}(l) z^{-l}$ of size $(N_r \times 2)$ is irreducible [28] and its highest order coefficient $\mathbf{H}(M)$ is full column-rank[2], with $\mathbf{H}(l) \in \mathbb{C}^{N_r \times 2}$ contains the l-th channel's tap matrix coefficients. The (r, i)-th input of $\mathbf{H}(z)$ is the transfer function given by $h_{r,i}(z) = \sum_{l=0}^{M} h_{r,\mathrm{L}}(l) z^{-l}$ for $i = 1$ and $h_{r,i}(z) = \sum_{l=0}^{M} h_{r,\mathrm{NL}}(l) z^{-l}$ for $i = 2$.

- The 2×2 covariance matrix of the input signal $[u(k), \tilde{u}(k)]^T$ is of full rank.

With the aforementioned assumptions, the subspace channel estimation technique introduced in [28] will be adopted. It allows us to identify the polynomial matrix $H(z)$ up to a constant matrix $\mathbf{Q}$ of size (2×2), which represents the inherent ambiguity of the blind subspace technique.

This technique exploits the received signal's SOS through the use of the covariance matrix of random vector $\mathbf{y}_w(k) = [\mathbf{y}(k)^T, \mathbf{y}(k-1)^T, \ldots, \mathbf{y}(k-N_w)^T]^T$, N_w being the window length[3] assumed sufficiently large ($N_w \geq 2M$). In this case the channel matrix has a Sylvester block-Toeplitz structure [29], denoted by $\mathcal{T}_{N_w}(\mathbf{h})$ ($\mathbf{h} = \mathrm{vec}\left(\mathbf{H}^T\right)$), of size $N_r(N_w + 1) \times 2(M + N_w + 1)$ where its first block row is given by $[\mathbf{H}(0), \ldots, \mathbf{H}(M), \mathbf{0}_{N_r \times 2}, \ldots, \mathbf{0}_{N_r \times 2}]$.

Under the previous assumptions, the noise variance σ_v^2 is the smallest eigenvalue of the covariance matrix $\mathcal{R}_w$ of $\mathbf{y}_w(k)$. The eigenspace associated with σ_v^2 is referred to as the noise subspace, which is the orthogonal complement of the signal subspace (i.e., range $(\mathcal{T}_{N_w}(\mathbf{h}))$ where range $(\mathbf{X})$ denotes the subspace generated by the column vectors of matrix $\mathbf{X}$) and is denoted $\mathbf{\Pi}_N = \mathrm{range}\,(\mathcal{T}_{N_w}(\mathbf{h}))^{\perp}$. The subspace identification method used is ultimately related to the fact that:

$$\mathbf{\Pi}_N^H \mathcal{T}_{N_w}(\mathbf{h}) = 0. \tag{4}$$

To take into account the noise effect, one solves equation (4) in the Least-Squares (LS) sense, which leads to the following quadratic expression (see [28] for details):

$$\hat{\mathbf{h}}_{ss} = \arg\min_{\mathbf{h}} \ \mathbf{h}^H \mathcal{Q} \mathbf{h}, \tag{5}$$

where the subscript "ss" stands for subspace and $\mathcal{Q}$ is a $2N_r(M+1) \times 2N_r(M+1)$ symmetric matrix corresponding to the considered LS cost function.
Note that to avoid degenerate solutions, different constraints on $\mathbf{h}$ can be considered including the unit-norm constraint for which the solution of equation (5) is given by the eigenvector associated to the smallest eigenvalue of $\mathcal{Q}$.

This method allows to estimate the polynomial matrix $H(z)$ up to a 2×2 constant matrix $\mathbf{Q}$, i.e., $\hat{H}(z) = H(z)\mathbf{Q}$. In order to use the estimated channels' vector $\hat{\mathbf{h}}_{ss}$ for further processing, one needs to remove the latter matrix

[2]The assumption that $\mathbf{H}(M)$ is full rank is not fundamental and is considered here for simplicity. Indeed, as shown in [28], one can get rid of it for subspace methods or simply use other SOS-based methods (like the linear prediction or the mutually referenced equalizers) which do not rely on this particular assumption

[3]This technique processes the signal using windows of length N_w.

ambiguity. In the sequel, solutions are proposed to get rid of $\mathbf{Q}$ by exploiting the nonlinear relation between $u(k)$ and $\tilde{u}(k)$.

3.2. Ambiguity removal for quadratic nonlinearity

In this section, it is assumed that $\tilde{u}(k) = u^2(k)$ (see equation (1)). At first, a channel equalization is performed based on the subspace channel estimate $\hat{\mathbf{h}}$. More precisely, a zero-forcing equalizer of delay M has been used in our simulations. It is given by the $(2M+1)$-th row[4] of the pseudo inverse matrix $\mathcal{T}^{\#}_{N_w}(\hat{\mathbf{h}})$. The obtained signal, in the noiseless case, is given by:

$$\mathbf{x}(k) = \mathbf{Q}^{-1}\mathbf{u}(k) \text{ with } \mathbf{u}(k) = [u(k), u^2(k)]^T, \tag{6}$$

The unknown matrix $\mathbf{Q} = (q_{i,j})_{1\leq i,j\leq 2}$ can be seen here as the separating matrix of the previous mixture, characterized (up to a diagonal) by the fact that vector $\mathbf{z}(k) = \mathbf{Q}\mathbf{x}(k)$ verifies $z_2(k) = z_1(k)^2$. Thus, in order to estimate the unknown matrix $\mathbf{Q}$, we propose to minimize the following LS criterion:

$$\sum_{k=1}^{N_s} |z_2(k) - z_1^2(k)|^2 = \|\overline{\mathbf{X}}\overline{\mathbf{q}}\|^2, \tag{7}$$

where $\overline{\mathbf{q}} = [q_{2,1}, q_{2,2}, q_{1,1}^2, q_{1,2}^2, q_{1,1}q_{1,2}]^T$; N_s is the size of the equalized signal and

$$\overline{\mathbf{X}} = \begin{bmatrix} -x_1(1) & -x_2(1) & x_1^2(1) & x_2^2(1) & 2x_1(1)x_2(1) \\ \vdots & & & & \vdots \\ -x_1(N_s) & -x_2(N_s) & x_1^2(N_s) & x_2^2(N_s) & 2x_1(N_s)x_2(N_s) \end{bmatrix}.$$

Vector $\overline{\mathbf{q}}$ is estimated up to a constant. It is proportional to the eigenvector $\mathbf{v} = [v_1, v_2, v_3, v_4, v_5]^T$ associated to the smallest eigenvalue of $\overline{\mathbf{X}}^H\overline{\mathbf{X}}$. The latter is shown to be uniquely identified under some mild assumptions on the input signal as it will be detailed in section 5. Consequently, it is possible to estimate $\mathbf{Q}$ as follows:

$$\overline{\mathbf{q}} = \begin{bmatrix} q_{2,1} \\ q_{2,2} \\ q_{1,1}^2 \\ q_{1,2}^2 \\ q_{1,1}q_{1,2} \end{bmatrix} = \alpha \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \end{bmatrix} \tag{8}$$

where α is an unknown scalar factor. Hence taking into account the structure of $\overline{\mathbf{q}}$, one can write

$$\mathbf{Q} = \begin{bmatrix} \sqrt{\alpha v_3} & \frac{\sqrt{\alpha}v_5}{\sqrt{v_3}} \\ \alpha v_1 & \alpha v_2 \end{bmatrix} = \sqrt{\alpha}\begin{bmatrix} 1 & 0 \\ 0 & \lambda \end{bmatrix}\tilde{\mathbf{Q}}. \tag{9}$$

[4]Note that the odd rows of $\mathcal{T}^{\#}_{N_w}(\hat{\mathbf{h}})$ extract symbols $u(n)$ at different delays while even rows of $\mathcal{T}^{\#}_{N_w}(\hat{\mathbf{h}})$ extract delayed samples of $\tilde{u}(k)$.

The common scalar factor $\sqrt{\alpha}$ can be disregarded due to the scale ambiguity of such blind processing. However, for the diagonal matrix term, scalar $\lambda = \sqrt{\alpha}$ is estimated by substituting the expression (9) in the criterion given by equation (7), leading to:

$$\lambda = \frac{\sum_{k=1}^{N_s} \tilde{z}_1^2(k)\tilde{z}_2^*(k)}{\sum_{k=1}^{N_s} |\tilde{z}_2(k)|^2}, \tag{10}$$

where $\tilde{\mathbf{z}}(k) = \tilde{\mathbf{Q}}\mathbf{x}(k)$.

In the case of blind processing, the source signal (and also the channel coefficients) is estimated up to a constant, which represents the inherent ambiguity of the considered problem. However, it is possible to further reduce this indetermination (unknown phase multiple of $\pi/2$ for Quadrature Amplitude Modulation (QAM) signals) by exploiting the independence of the real and imaginary parts of the transmitted symbols. In our simulations, we have used the phase rotation given in [30].

3.3. Ambiguity removal for cubic nonlinearity

In this section, it is assumed that $\tilde{u}(k) = |u(k)|^2 u(k)$. Similarly to the previous case, after channel equalization the obtained signal, in the noiseless case, is expressed as:

$$\mathbf{x}(k) = \mathbf{Q}^{-1}\mathbf{u}(k) \text{ with } \mathbf{u}(k) = [u(k), |u(k)|^2 u(k)]^T, \tag{11}$$

To get rid of the unknown mixing matrix $\mathbf{Q}$, one can minimize:

$$\sum_{k=1}^{N_s} |z_2(k) - |z_1(k)|^2 z_1(k)|^2, \tag{12}$$

with $\mathbf{z}(k) = \mathbf{Q}\mathbf{x}(k)$. This multivariate optimization problem can be reduced to the search of one complex parameter $b = \bar{q}_{1,2}$, by normalizing $\bar{q}_{1,1} = 1$ (this is possible thanks to the inherent scale ambiguity of the considered problem) and by solving equation (12) with respect to $[\bar{q}_{2,1}, \bar{q}_{2,2}]$ in terms of b. Indeed, for a fixed value of b, the criterion in (12) reduces to a LS optimization problem with respect to $[\bar{q}_{2,1}, \bar{q}_{2,2}]$, for which a closed-form solution exists:

$$[\bar{q}_{2,1}, \bar{q}_{2,2}] = [1, b](\mathbf{X}\mathbf{X}^{\#}), \tag{13}$$

with $\mathbf{X} = [\mathbf{x}(1), \ldots, \mathbf{x}(N_s)]$. Plugging equation (13) into equation (12) leads to a nonlinear cost function in terms of parameter b that can be solved using numerical optimization techniques. Note that, as it will be shown in section 5, criterion given by equation (12) has spurious solutions for small or moderate size constellations (e.g., QAM 4, 16 and 32) in which case an alphabet matching cost function (see [31] and references therein) should be used instead.

3.4. EM-based estimation

This section is devoted to the proposed EM-based, maximum likelihood channel estimator for the system model given by equation (1). Note that, due to its sensitivity to initialization, the EM-based estimator is initialized using the subspace-based estimates given in 3.1, after performing an ambiguity removal proposed in 3.2 or in 3.3.

The EM algorithm is an iterative method aiming at finding maximum likelihood or maximum *a posteriori* estimates of parameters in statistical models, where the model depends on unobserved latent variables. Thus, the sequence of N_s states $\mathbf{S}_{[1:N_s]}$, which are not observed, represent the missing data, whereas the N_s received symbols $\mathbf{Y}_{[1:N_s]}$ stand for the incomplete data (observations). Moreover, the complete-data is given by the sequence $(\mathbf{Y}_{[1:N_s]}, \mathbf{S}_{[1:N_s]})$. Each EM iteration alternates between two steps: E-step and M-step as described below.

3.4.1. E-step

The objective of this step is to find the auxiliary function, denoted by $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(m)})$, which is defined as the conditional expectation of the complete-data log-likelihood, with respect to the conditional distribution of the missing data $\mathbf{S}_{[1:N_s]}$, given the observations $\mathbf{Y}_{[1:N_s]}$ and the current estimated parameter value at the m-th iteration $\boldsymbol{\theta}^{(m)} = \left(\text{vec}\left(\mathbf{H}^{(m)}\right)^T, \sigma_v^{2(m)}\right)$. Thus, such an auxiliary function can be expressed as:

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(m)}) \quad = \quad E\left(\log f_\theta\left(\mathbf{Y}_{[1:N_s]}, \mathbf{S}_{[1:N_s]}\right) \middle| \mathbf{Y}_{[1:N_s]}, \boldsymbol{\theta}^{(m)}\right), \tag{14}$$

where $E(\cdot)$ refers to the expectation with respect to the distribution of the missing data.

After some straightforward derivations and by ignoring terms that are independent of $\boldsymbol{\theta}$, $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(m)})$ is shown to be proportional to (see [32]):

$$\sum_{\mathbf{x}_{ij} \in \mathcal{X}} \sum_{k=1}^{N_s} \left(-N_r \log(\sigma_v^2) - \|\mathbf{y}(k) - \mathbf{H}\bar{\mathbf{x}}_{ij}\|^2 / \sigma_v^2\right) \gamma_{\theta^{(m)}}(k; i, j), \tag{15}$$

where $\bar{\mathbf{x}}_{ij} = [\mathbf{x}_{ij}^T, \tilde{\mathbf{x}}_{ij}^T]^T$ ($\tilde{\mathbf{x}}_{ij}$ being the nonlinear term associated to $\mathbf{x}_{ij}$), and $\gamma_{\theta^{(m)}}(k; i, j) = f_{\theta^{(m)}}\left(\mathbf{s}(k) = \mathbf{q}_i, \mathbf{s}(k+1) = \mathbf{q}_j | \mathbf{Y}_{[1:N_s]}\right)$ represents the posterior probability of the trellis branch $(\mathbf{s}(k) = \mathbf{q}_i, \mathbf{s}(k+1) = \mathbf{q}_j)$ given the observations $\mathbf{Y}_{[1:N_s]}$ and the current estimate of the parameter $\boldsymbol{\theta}^{(m)}$. This probability can be efficiently computed using the forward-backward variables ([33, 34]) denoted by $\alpha_{\theta^{(m)}}(k; i)$ and $\beta_{\theta^{(m)}}(k; j)$. Consequently, by omitting multiplicative scaling factors independent of k, i, j, it can be shown that:

$$\gamma_{\theta^{(m)}}(k; i, j) \quad \propto \quad \alpha_{\theta^{(m)}}(k; i)\ \beta_{\theta^{(m)}}(k+1; j)\ b_{\theta^{(m)}}(k; i, j),$$

where

$$\alpha_{\theta^{(m)}}(k; i) \quad = \quad f_{\theta^{(m)}}(\mathbf{Y}_{[1:k-1]}, \mathbf{s}(k) = \mathbf{q}_i), \tag{16}$$

$$\beta_{\theta^{(m)}}(k; j) \quad = \quad f_{\theta^{(m)}}(\mathbf{Y}_{[k:N_s]} \mid \mathbf{s}(k) = \mathbf{q}_j), \tag{17}$$

$$b_{\theta^{(m)}}(k,i,j) \propto \left(\sigma_v^{2(m)}\right)^{-N_r} \exp\left(\frac{-\|\mathbf{y}(k)-\mathbf{H}^{(m)}\bar{\mathbf{x}}_{ij}\|^2}{\sigma_v^{2(m)}}\right). \tag{18}$$

The forward and backward variables can be evaluated recursively as:

$$\alpha_{\theta^{(m)}}(k+1;i) = \frac{1}{N}\sum_{l\in\mathcal{F}(i)} \alpha_{\theta^{(m)}}(k;l) b_{\theta^{(m)}}(k,l,i), \tag{19}$$

$$\beta_{\theta^{(m)}}(k;j) = \frac{1}{N}\sum_{l\in\mathcal{B}(j)} \beta_{\theta^{(m)}}(k+1;l) b_{\theta^{(m)}}(k,j,l), \tag{20}$$

where $\mathcal{F}(i)$ (resp. $\mathcal{B}(j)$) denotes the set of states connected to $\mathbf{q}_i$ (resp. $\mathbf{q}_j$) in forward (predecessors) (resp. backward (successors)) directions. Note that, the predecessors of the state $\mathbf{q}_i = \mathbf{s}(k) = [u(k-1),\ldots,u(k-M)]^T$ take the form $\mathbf{s}(k-1) = [u(k-2),\ldots,u(k-M-1)]^T$, whereas its successors are given by $\mathbf{s}(k+1) = [u(k),\ldots,u(k-M+1)]^T$, where the symbols $u(k)$ and $u(k-M-1)$ take values from the set $\mathcal{A}$ with equal probabilities. Hence, the total number of a state's predecessors and successors is 2^B.

3.4.2. M-step

The objective of this step is to find the parameter $\boldsymbol{\theta}^{(m+1)}$ that maximizes $Q\left(\boldsymbol{\theta};\boldsymbol{\theta}^{(m)}\right)$, i.e.,

$$\boldsymbol{\theta}^{(m+1)} = \arg\max_{\boldsymbol{\theta}} Q\left(\boldsymbol{\theta};\boldsymbol{\theta}^{(m)}\right). \tag{21}$$

This process is shown in [35] and [36] to increase the likelihood function, and consequently it leads to the algorithm's convergence to a global maximum point.

Since $Q\left(\boldsymbol{\theta};\boldsymbol{\theta}^{(m)}\right)$ is quadratic in its argument, the maximization step reduces to:

$$\mathbf{H}^{(m+1)} = \mathbf{R}_{yx}\mathbf{R}_{xx}^{-1}, \tag{22}$$

$$(\sigma_v^2)^{(m+1)} = \frac{1}{N_s N_r}\,\mathrm{tr}\left(\mathbf{R}_{yy} - \mathbf{H}^{(m+1)}\mathbf{R}_{xy}\right), \tag{23}$$

where $\mathbf{R}_{yy}$ is the autocorrelation matrix of observations, $\mathbf{R}_{xy}$ is the "weighted" cross-correlation matrix between the unobserved symbol transitions and the observations, and $\mathbf{R}_{xx}$ stands for the "weighted" auto-correlation matrix of the

unobserved symbol transitions. These matrices are given by:

$$\mathbf{R}_{yy} = \sum_{k=1}^{N_s} \mathbf{y}(k)\mathbf{y}(k)^H, \tag{24}$$

$$\mathbf{R}_{xy} = \mathbf{R}_{yx}^H = \sum_{k=1}^{N_s} \sum_{\mathbf{x}_{ij}\in\mathcal{X}} \bar{\mathbf{x}}_{ij}\mathbf{y}(k)^H \gamma_{\theta^{(m)}}(k;i,j)$$

$$= \sum_{k=1}^{N_s} E_{\theta^{(m)}}(\mathbf{x}(k)|\mathbf{Y}_{[1:N_s]})\mathbf{y}(k)^H \tag{25}$$

$$\mathbf{R}_{xx} = \sum_{k=1}^{N_s} \sum_{\mathbf{x}_{ij}\in\mathcal{X}} \bar{\mathbf{x}}_{ij}\bar{\mathbf{x}}_{ij}^H \gamma_{\theta^{(m)}}(k;i,j) = \sum_{k=1}^{N_s} E_{\theta^{(m)}}(\mathbf{x}(k)\mathbf{x}(k)^H|\mathbf{Y}_{[1:N_s]}). \tag{26}$$

The iterative procedure can be stopped as soon as $\frac{\left\|\boldsymbol{\theta}^{(m+1)}-\boldsymbol{\theta}^{(m)}\right\|}{\left\|\boldsymbol{\theta}^{(m)}\right\|} < \epsilon$, for a chosen positive threshold ϵ.

3.5. Data detection within EM framework

Several data detection methods can be designed given the value of the estimated parameter obtained at the end of the iterations, denoted by $\boldsymbol{\theta}^{(\infty)}$; the observation sequence $\mathbf{Y}_{[1:N_s]}$ and the trellis diagram of the channel. The optimal criterion retained in the current work is the minimum symbol-error probability [37], which can be easily implemented within an EM framework. Minimizing the symbol-error probability aims at choosing, at each instant k, the data symbol which maximizes the posterior probability of the symbol $u(k)$ given the observations $\mathbf{Y}_{[1:N_s]}$ and the channel parameters $\boldsymbol{\theta}^{(\infty)}$ as follows:

$$\hat{u}(k) = \arg\max_{a_{i_0}\in\mathcal{A}} f_{\boldsymbol{\theta}^{(\infty)}}\left(u(k) = a_{i_0}\middle|\mathbf{Y}_{[1:N_s]};\boldsymbol{\theta}^{(\infty)}\right). \tag{27}$$

This quantity may be simply expressed, as a function of the posterior probability $\gamma_{\boldsymbol{\theta}^{(\infty)}}(k;i,j)$ of the trellis branch $(\mathbf{s}(k) = \mathbf{q}_i, \mathbf{s}(k+1) = \mathbf{q}_j)$ given the observations $\mathbf{Y}_{[1:N_s]}$ and $\boldsymbol{\theta}^{(\infty)}$ as:

$$f_{\boldsymbol{\theta}^{(\infty)}}\left(u(k) = a_{i_0}\middle|\mathbf{Y}_{[1:N_s]};\boldsymbol{\theta}^{(\infty)}\right) = \sum_{i,j\in\mathcal{S}(i_0)} \gamma_{\boldsymbol{\theta}^{(\infty)}}(k;i,j), \tag{28}$$

where $\mathcal{S}(i_0)$ is the set of all trellis branch values so that $\mathbf{x}_{ij} = [u(k) = a_{i_0}, u(k-1), \ldots, u(k-M)]$. Consequently, the EM-based solutions presented in this work can be further considered as a joint channel estimation and data detection within a maximum likelihood framework.

4. Semi-blind EM-based estimation

In most communications systems, some training symbols (pilots) are usually sent periodically within the wireless network frames besides the unknown data.

Hence, a Semi-Blind (SB) approach, exploiting pilots, can be adopted in order to take advantage of this available data and reduce the different difficulties and issues related to the blind processing. To do so, and without loss of generality, the transmitted sequences and the received observations are assumed to be composed of N_p pilots and N_d data symbols so that $\mathbf{S}_{[1:N_s]} = [\mathbf{S}_{p_{[1:N_p]}}, \mathbf{S}_{d_{[N_p+1:N_s]}}]$ and $\mathbf{Y}_{[1:N_s]} = [\mathbf{Y}_{p_{[1:N_p]}}, \mathbf{Y}_{d_{[N_p+1:N_s]}}]$ where $N_p + N_d = N_s$ (indices p and d stand for pilot and data, respectively). Moreover, since the initialization can be performed using the available pilot symbols, the constraints on the number of receive antennas $N_r > 2$ as well as channel diversity conditions defined in section 2, are no longer required.

In what follows, we describe the E-step and the M-step for the case of EM-based semi-blind framework.

4.1. E-step

By considering pilots and data, the auxiliary function, given in equation (15) for the blind case, has now an additional term corresponding to the pilot sequence. The new function becomes:

$$\begin{aligned} Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(m)}) \propto & \sum_{k=1}^{N_p} \left(-N_r \log(\sigma_v^2) - \|\mathbf{y}(k) - \mathbf{H}\bar{\mathbf{u}}_p(k)\|^2/\sigma_v^2\right) + \\ & \sum_{\mathbf{x}_{ij} \in \mathcal{X}} \sum_{k=N_p+1}^{N_s} \left(-N_r \log(\sigma_v^2) - \|\mathbf{y}(k) - \mathbf{H}\bar{\mathbf{x}}_{ij}\|^2/\sigma_v^2\right) \gamma_{\theta^{(m)}}(k; i, j), \end{aligned} \tag{29}$$

where $\bar{\mathbf{u}}_p(k) = [u_p(k), \ldots, u_p(k-M), \tilde{u}_p(k), \ldots, \tilde{u}_p(k-M)]^T$, of size $2(M+1)$, is composed of linear and nonlinear terms of the pilot signal.

4.2. M-step

In the case of semi-blind processing, equations (25) and (26) become:

$$\mathbf{R}_{yx} = \mathbf{R}_{xy}^H = \sum_{k=1}^{N_p} \mathbf{y}(k)\bar{\mathbf{u}}_p(k)^H + \sum_{k=N_p+1}^{N_s} \sum_{\mathbf{x}_{ij} \in \mathcal{X}} \mathbf{y}(k)\bar{\mathbf{x}}_{ij}^H \gamma_{\theta^{(m)}}(k; i, j), \tag{30}$$

$$\mathbf{R}_{xx} = \sum_{k=1}^{N_p} \bar{\mathbf{u}}_p(k)\bar{\mathbf{u}}_p(k)^H + \sum_{k=N_p+1}^{N_s} \sum_{\mathbf{x}_{ij} \in \mathcal{X}} \bar{\mathbf{x}}_{ij}\bar{\mathbf{x}}_{ij}^H \gamma_{\theta^{(m)}}(k; i, j). \tag{31}$$

Remark: A normalization of the posterior probability of the data terms might be performed when considering such semi-blind context. To do so, one can write[5]:

$$\gamma_{\theta^{(m)}}(k; i, j) = \frac{\gamma_{\theta^{(m)}}(k; i, j)}{\sum_{\mathbf{x}_{ij} \in \mathcal{X}} \gamma_{\theta^{(m)}}(k; i, j)}. \tag{32}$$

[5]To avoid introducing a new notation, we kept the same expression for the normalized posterior probability $\gamma_{\theta^{(m)}}(k; i, j)$.

4.3. Extension to Nonlinear MIMO systems

The previous results can be easily extended to the multi-user case (i.e., MIMO system). To do so, the system model given in equation (1), is re-written as follows:

$$y_r(k) = \sum_{i=1}^{N_t} \sum_{n=0}^{M_{i,r,\mathrm{L}}} h_{i,r,\mathrm{L}}(n) u_i(k-n) + \sum_{i=1}^{N_t} \sum_{n=0}^{M_{i,r,\mathrm{NL}}} h_{i,r,\mathrm{NL}}(n) \tilde{u}_i(k-n) + v_r(k), v_r(k), \quad (33)$$

where $h_{i,r,\mathrm{L}}$ (resp. $h_{i,r,\mathrm{NL}}$) refers to the linear (resp. nonlinear) channel impulse response between the i-th user and the r-th receive antenna; while $u_i(k)$ represents the transmitted symbol of the i-th user.

Consequently, the system model given by equation (2) will be based on the following vectors: $\mathbf{h}_{r,\mathrm{L}} = [h_{1,r,\mathrm{L}}(0), \ldots, h_{1,r,\mathrm{L}}(M), \ldots, h_{N_t,r,\mathrm{L}}(M)]^T$, $\mathbf{h}_{r,\mathrm{NL}} = [h_{1,r,\mathrm{NL}}(0), \ldots, h_{1,r,\mathrm{NL}}(M), \ldots, h_{N_t,r,\mathrm{NL}}(M)]^T$, $\mathbf{u}(k) = [u_1(k), \ldots, u_1(k-M), \ldots, u_{N_t}(k-M)]^T$ and $\tilde{\mathbf{u}}(k) = [\tilde{u}_1(k), \ldots, \tilde{u}_1(k-M), \ldots, \tilde{u}_{N_t}(k-M)]^T$, where $M = \max_{i,r}\{M_{i,r,\mathrm{L}}, M_{i,r,\mathrm{NL}}\}$.

In such a case, the state vector is given by $\mathbf{s}(k) = [u_1(k-1), \ldots, u_1(k-M), \ldots, u_{N_t}(k-1), \ldots, u_{N_t}(k-M)]^T$ containing $N_t M$ symbols, with 2^{BN_tM} possible state values. Whereas, the transition vector is defined by $\mathbf{x}_{nm} = [u_1(k), \ldots, u_1(k-M), \ldots, u_{N_t}(k), \ldots, u_{N_t}(k-M)]^T$ of $N_t(M+1)$ symbols.

By using this new vectors, equations (30), (31), (22) and (23) are still valid, leading to an EM-based channel estimation for nonlinear MIMO communications systems.

5. Identifiability results and performance bounds

The aim of this section is to corroborate the proposed blind and semi-blind approaches by providing some results related to the channel identifiability conditions. Also, in order to get more insight on the proposed solutions performance limits, we perform the derivation of the deterministic Cramér-Rao Bound (CRB), corresponding to the adopted nonlinear system model.

5.1. Identifiability results

In the blind context, there exist certain inherent ambiguities with respect to the identification of the channel parameters. In particular, since we are using the subspace method for the initialization of our EM algorithm, we are interested in the SOS-based identifiability. Under the assumption of i.i.d. input symbols considered for the data model, the power spectral density (PSD) of the observed data is expressed as:

$$\mathbf{P}_y(e^{j2\pi f}) = [\mathbf{h}_{\mathrm{L}}(e^{j2\pi f}), \mathbf{h}_{\mathrm{NL}}(e^{j2\pi f})] \mathbf{R}_u [\mathbf{h}_{\mathrm{L}}(e^{j2\pi f}), \mathbf{h}_{\mathrm{NL}}(e^{j2\pi f})]^H + \sigma_v^2 \mathbf{I}, \quad (34)$$

where $\mathbf{h}_{\mathrm{L}}(e^{j2\pi f}) = [h_{1,\mathrm{L}}(e^{j2\pi f}), \ldots, h_{M,\mathrm{L}}(e^{j2\pi f})]^T$ (resp. $\mathbf{h}_{\mathrm{NL}}(e^{j2\pi f}) = [h_{1,\mathrm{NL}}(e^{j2\pi f}), \ldots, h_{M,\mathrm{NL}}(e^{j2\pi f})]^T$) is the frequency response of the linear (resp. the nonlinear) channels, while $\mathbf{R}_u$ is the 2×2 covariance matrix (assumed full rank) of $[u(k), \tilde{u}(k)]^T$. As shown in [28], by considering $u(k)$ and $\tilde{u}(k)$ as two different source signals, the SOS allow us to identify $H(z)$ up to a constant matrix $\mathbf{Q}$. Indeed, the PSD can be rewritten as

$$\mathbf{P}_y(e^{j2\pi f}) = \left([\mathbf{h}_{\mathrm{L}}(e^{j2\pi f}), \mathbf{h}_{\mathrm{NL}}(e^{j2\pi f})]\mathbf{Q}\right) \mathbf{R}_{u,Q} \left([\mathbf{h}_{\mathrm{L}}(e^{j2\pi f}), \mathbf{h}_{\mathrm{NL}}(e^{j2\pi f})]\mathbf{Q}\right)^H + \sigma_v^2 \mathbf{I} \quad (35)$$

where $\mathbf{R}_{u,Q} = \mathbf{Q}^{-1}\mathbf{R}_u\mathbf{Q}^{-H}$. Now, to get rid of this ambiguity, we need to use higher order information through the nonlinear cost functions in equations (7) and (12). The latter helps to reduce the ambiguity from a 2×2 matrix factor to a scalar factor under certain additional assumptions given in the following proposition.

Proposition 1. *For the quadratic nonlinearity case with $\tilde{u}(k) = u^2(k)$, the minimization of the criterion given by equation (7) in the large sample size and noiseless case leads to the desired input signal (up to a constant factor c), i.e., $z_1(k) = cu(k)$, if and only if the correlation matrix of vector $\overline{\mathbf{u}}(k) = [1, u(k), u^2(k), u^3(k)]^T$ is full rank, i.e., $E(\overline{\mathbf{u}}(k)\overline{\mathbf{u}}(k)^H) > 0$.*

For the cubic nonlinearity with $\tilde{u}(k) = |u(k)|^2 u(k)$, the minimization of the criterion given by equation (12) in the asymptotic and noiseless case leads to the desired input signal (up to a constant factor), if and only if the number of possible modulus values of the input signal, denoted d, satisfies $d > 4$.

Proof. Consider an instantaneous mixture of $u(k)$ and $\tilde{u}(k)$:

$$\begin{bmatrix} z_1(k) \\ z_2(k) \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix} \begin{bmatrix} u(k) \\ \tilde{u}(k) \end{bmatrix}.$$

For the quadratic nonlinearity case, we would like to prove that criterion given by equation (7) is minimum (null in the noiseless case) if and only if (iff):

$$\begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix} = \begin{bmatrix} c & 0 \\ 0 & c^2 \end{bmatrix}, \tag{36}$$

for a given constant c. We have

$$\begin{aligned} z_2(k) = z_1^2(k) \quad &\Longleftrightarrow \quad u(k)\mathbf{m}^T\overline{\mathbf{u}}(k) = 0 \\ &\Longleftrightarrow \quad \mathbf{m}^T\overline{\mathbf{u}}(k) = 0 \quad \text{since } u(k) \neq 0, \end{aligned}$$

where $\mathbf{m} = [-m_{21}, m_{11}^2 - m_{22}, 2m_{11}m_{12}, m_{12}^2]^T$. Hence, by taking the mean value, $E(|z_2(k) - z_1^2(k)|^2) = 0$ is equivalent to $\mathbf{m}^T E(\overline{\mathbf{u}}(k)\overline{\mathbf{u}}(k)^H) \times \mathbf{m}^* = 0$. This latter equality has a unique solution $\mathbf{m} = 0$ under the full-rank condition, i.e., $E(\overline{\mathbf{u}}(k)\overline{\mathbf{u}}(k)^H) > 0$. Finally, the vector $\mathbf{m}$ is null iff equation (36) holds.

For the cubic nonlinearity, we would like to prove that the criterion given by equation (12) is null (in the noiseless case) iff:

$$\begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix} = \begin{bmatrix} c & 0 \\ 0 & |c|^2 c \end{bmatrix}, \tag{37}$$

for a given constant c. We have:

$$\begin{aligned} z_2(k) = |z_1(k)|^2 z_1(k) \quad &\Longleftrightarrow \quad u(k)\overline{\mathbf{m}}^T\overline{\overline{\mathbf{u}}}(k) = 0 \\ &\Longleftrightarrow \quad \overline{\mathbf{m}}^T\overline{\overline{\mathbf{u}}}(k) = 0 \quad \text{since } u(k) \neq 0, \end{aligned}$$

where $\overline{\mathbf{m}} = [-m_{21}, (m_{11}|m_{11}|^2 - m_{22}), 2m_{11}\,\mathrm{Re}\,(m_{11}m_{12}^*) + m_{12}|m_{11}|^2, m_{11}|m_{12}|^2 + 2m_{12}\,\mathrm{Re}\,(m_{11}m_{12}^*), m_{12}|m_{12}|^2]^T$ and $\overline{\overline{\mathbf{u}}}(k) = [1,\ |u(k)|^2,\ |u(k)|^4,\ |u(k)|^6,\ |u(k)|^8]^T$. Now, the equation system $\overline{\mathbf{m}}^T\overline{\overline{\mathbf{u}}}(k) = 0$, for all k will have a unique solution $\overline{\mathbf{m}} = 0$, if the Vandermonde-like matrix formed by vectors $\overline{\overline{\mathbf{u}}}(k)$ has a full (row) rank equal to 5. This is the case iff symbols $u(k)$ have at least $d > 4$ different modulus values. Finally, vector $\overline{\mathbf{m}}$ is null iff (37) holds. □

For M-QAM modulations, one can easily check that in the case of quadratic nonlinearity, the lemma's condition is met for $M > 4$. However, for the cubic nonlinearity, the lemma's condition is quite restrictive and requires large modulation sizes with $M > 64$. In such a case, when the modulation size is small or moderate (i.e. $M \leq 64$), the ambiguity removal requires the use of another criterion such as the alphabet matching one [31].

Also, note that when the polynomial degrees of the nonlinear and linear channels are not the same, i.e., $\deg(h_{\mathrm{L}}(z)) \neq \deg(h_{\mathrm{NL}}(z))$, the subspace method would identify the channel matrix $H(z)$ up to a certain 2×2 polynomial matrix $Q(z)$ (see [28] for details). In such a case, the proposed ambiguity removal method does not apply and consequently the EM algorithm's initialization might be ineffective leading to potential ill-convergence of the considered blind algorithm. All these issues can be avoided in the semi-blind context where the knowledge of pilot signals can be exploited to initialize our EM algorithm but also to remove the previous blind processing indeterminations. Next, we derive the deterministic[6] Cramér Rao Bound (CRB) relative to the SB context, that will be used later for our algorithm's performance benchmarking.

5.2. Deterministic Cramér-Rao Bound (CRB)

Given a parametric statistical model, the CRB provides a lower bound of the error variance for all unbiased estimators of the system's parameter vector. In particular, the Gaussian CRB (G-CRB) represents a lower bound within the class of estimators using only the SOS of the observed data. It is also the least favorable CRB as shown in [38]. In the sequel, we derive the expression of the deterministic G-CRB for our system model.

The data model in equation (1) can be rewritten in a more compact way by considering all data samples N_s and all outputs N_r as:

$$\mathbf{y} = \mathcal{H}_{\mathrm{L}}\mathbf{u}_{\mathrm{L}} + \mathcal{H}_{\mathrm{NL}}\mathbf{u}_{\mathrm{NL}} + \mathbf{v} = \mathcal{U}_{\mathrm{L}}\mathbf{h}_{\mathrm{L}} + \mathcal{U}_{\mathrm{NL}}\mathbf{h}_{\mathrm{NL}} + \mathbf{v}, \tag{38}$$

where $\mathbf{y} = [y_1(0), \ldots, y_1(N-1), \ldots, y_{N_r}(0), \ldots, y_{N_r}(N-1)]^T$ ($N = N_p + N_s$ the total number of transmitted pilot and data symbols), $\mathbf{v} = [v_1(0), \ldots, v_1(N-1), \ldots, v_{N_r}(0), \ldots, v_{N_r}(N-1)]^T$, $\mathbf{u}_{\mathrm{L}} = [u(0), \ldots, u(N-1)]^T$, $\mathbf{u}_{\mathrm{NL}} = [\tilde{u}(0), \ldots, \tilde{u}(N-1)]^T$, $\mathbf{h}_{\mathrm{L}} = [h_{1,\mathrm{L}}(0), \ldots, h_{1,\mathrm{L}}(M), \ldots, h_{N_r,\mathrm{L}}(0), \ldots, h_{N_r,\mathrm{L}}(M)]^T$, $\mathbf{h}_{\mathrm{NL}} = [h_{1,\mathrm{NL}}(0), \ldots, h_{1,\mathrm{NL}}(M), \ldots, h_{N_r,\mathrm{NL}}(0), \ldots, h_{N_r,\mathrm{NL}}(M)]^T$, $\mathcal{H}_\ddagger = [\mathbf{H}_{1,\ddagger}^T, \ldots, \mathbf{H}_{N_r,\ddagger}^T]^T$, $\mathcal{U}_\ddagger = I \otimes \mathbf{U}_\ddagger$ ($\ddagger = \mathrm{L}$ or NL) where $\otimes$ denotes the Kronecker product, and $\mathbf{H}_{r,\ddagger}$ and $\mathbf{U}_\ddagger$ are defined as ($u_\ddagger(k)$ equals $u(k)$ if $\ddagger = \mathrm{L}$ or $\tilde{u}(k)$ if $\ddagger = \mathrm{NL}$):

$$\mathbf{H}_{r,\ddagger} = \begin{bmatrix} h_{r,\ddagger}(M) & \cdots & h_{r,\ddagger}(0) & & 0 \\ & \ddots & & \ddots & \\ 0 & & h_{r,\ddagger}(M) & \cdots & h_{r,\ddagger}(0) \end{bmatrix}_{N \times (N+M)}, \tag{39}$$

$$\mathbf{U}_\ddagger = \begin{bmatrix} u_\ddagger(0) & \cdots & u_\ddagger(-M) \\ & \ddots & \vdots \\ \vdots & & u_\ddagger(0) \\ & & \vdots \\ u_\ddagger(N_s-1) & \cdots & u_\ddagger(N_s-1-M) \end{bmatrix}_{N \times (M+1)}. \tag{40}$$

Since we assumed a semi-blind approach, we can write $\mathbf{u}_\ddagger = [\mathbf{u}_{\ddagger,p}^T, \mathbf{u}_{\ddagger,d}^T]^T$ where $\mathbf{u}_{\ddagger,p}$ contains pilot-data samples and $\mathbf{u}_{\ddagger,d}$ contains unknown-data samples. We consider the vector of unknown parameters to be $\boldsymbol{\theta} = [\mathbf{h}_{\mathrm{L}}^T, \mathbf{h}_{\mathrm{NL}}^T, \mathbf{u}_d^T]^T$ (of size $2(M+1)N_r + N_s$) where $\mathbf{u}_d = \mathbf{u}_{\mathrm{L},d}$.

[6] The input symbols are treated as deterministic unknown parameters.

The *unconstrained complex* Fisher Information Matrix (FIM) is defined as:

$$\mathbf{J} = \begin{bmatrix} \mathbf{J}_{\boldsymbol{\theta\theta}} & \mathbf{J}_{\boldsymbol{\theta\theta}^*} \\ \mathbf{J}_{\boldsymbol{\theta}^*\boldsymbol{\theta}} & \mathbf{J}_{\boldsymbol{\theta}^*\boldsymbol{\theta}^*} \end{bmatrix} = \begin{bmatrix} \mathbf{J}_{\boldsymbol{\theta\theta}} & \mathbf{J}_{\boldsymbol{\theta\theta}^*} \\ (\mathbf{J}_{\boldsymbol{\theta\theta}^*})^H & (\mathbf{J}_{\boldsymbol{\theta\theta}})^* \end{bmatrix}, \tag{41}$$

where $\mathbf{J}_{\boldsymbol{\theta\theta}} = \frac{1}{\sigma_v^2}\left(\frac{\partial\tilde{\boldsymbol{\mu}}}{\partial\boldsymbol{\theta}}\right)^H \frac{\partial\tilde{\boldsymbol{\mu}}}{\partial\boldsymbol{\theta}}$, $\tilde{\boldsymbol{\mu}} = [\boldsymbol{\mu}^T, \boldsymbol{\mu}^H]^T$, $\boldsymbol{\mu} = \mathcal{H}_{\mathrm{L},p}\mathbf{u}_p + \mathcal{H}_{\mathrm{L},d}\mathbf{u}_d + \mathcal{H}_{\mathrm{NL},p}\tilde{\mathbf{u}}_p + \mathcal{H}_{\mathrm{NL},d}\tilde{\mathbf{u}}_d$, and $\frac{\partial\mathbf{f}}{\partial\mathbf{x}}$ ($\mathbf{f} \in \mathbb{C}^{m\times 1}, \mathbf{x} \in \mathbb{C}^{n\times 1}$) denotes the differentiation operator defined as $\frac{\partial\mathbf{f}}{\partial\mathbf{x}} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}$, where the matrix elements represent Wirtinger's derivatives [39]. Note from equation (41) that we only need to find $\mathbf{J}_{\boldsymbol{\theta\theta}}$ and $\mathbf{J}_{\boldsymbol{\theta\theta}^*}$ in order to find $\mathbf{J}$. After derivation, we find that:

$$\mathbf{J}_{\boldsymbol{\theta\theta}} = \frac{1}{\sigma_v^2}\begin{bmatrix} \mathcal{U}_{\mathrm{L}}^H\mathcal{U}_{\mathrm{L}} & \mathcal{U}_{\mathrm{L}}^H\mathcal{U}_{\mathrm{NL}} & \mathcal{U}_{\mathrm{L}}^H\boldsymbol{\Lambda} \\ \mathcal{U}_{\mathrm{NL}}^H\mathcal{U}_{\mathrm{L}} & \mathcal{U}_{\mathrm{NL}}^H\mathcal{U}_{\mathrm{NL}} & \mathcal{U}_{\mathrm{NL}}^H\boldsymbol{\Lambda} \\ \boldsymbol{\Lambda}^H\mathcal{U}_{\mathrm{L}} & \boldsymbol{\Lambda}^H\mathcal{U}_{\mathrm{NL}} & \boldsymbol{\Lambda}^H\boldsymbol{\Lambda} + \boldsymbol{\Gamma}^T\boldsymbol{\Gamma}^* \end{bmatrix}, \tag{42}$$

$$\mathbf{J}_{\boldsymbol{\theta\theta}^*} = \frac{1}{\sigma_v^2}\begin{bmatrix} 0 & 0 & \mathcal{U}_{\mathrm{L}}^H\boldsymbol{\Gamma} \\ 0 & 0 & \mathcal{U}_{\mathrm{NL}}^H\boldsymbol{\Gamma} \\ \boldsymbol{\Gamma}^T\mathcal{U}_{\mathrm{L}}^* & \boldsymbol{\Gamma}^T\mathcal{U}_{\mathrm{NL}}^* & \boldsymbol{\Lambda}^H\boldsymbol{\Gamma} + (\boldsymbol{\Lambda}^H\boldsymbol{\Gamma})^T \end{bmatrix}, \tag{43}$$

where $\boldsymbol{\Lambda} = \mathcal{H}_{\mathrm{L},d} + 2\mathcal{H}_{\mathrm{NL},d}\,\mathrm{diag}\left(|\mathbf{u}_d|^2\right)$, and $\boldsymbol{\Gamma} = \mathcal{H}_{\mathrm{NL},d}\,\mathrm{diag}\left(\mathbf{u}_d^2\right)$ for the cubic case. For the quadratic case, we found $\mathbf{J}_{\boldsymbol{\theta\theta}^*} = \mathbf{0}, \boldsymbol{\Lambda} = \mathcal{H}_{\mathrm{L},d} + 2\mathcal{H}_{\mathrm{NL},d}\,\mathrm{diag}\left(\mathbf{u}_d\right)$, and $\boldsymbol{\Gamma} = \mathbf{0}$. The G-CRB is computed as the inverse[7] of equation (41). The G-CRB for the channel parameters is given by the top left submatrix of the global G-CRB.

6. Generalization and numerical complexity

This section is dedicated to the extension of the proposed methods to other nonlinear models. It provides also details about the computational complexity of the proposed algorithms.

6.1. Generalization

The proposed semi-blind estimator can be easily generalized to other, finite memory, nonlinear models, like the one considered in [41]. Indeed, since the initialization is performed by using the pilots, and the state vector, exploited in the EM procedure, depends only on the linear terms, the extension remains possible upon certain derivation adaptations. However, for the blind case, it is feasible only when using 'another' appropriate initialization, since the main difficulty comes from the ambiguity removal of the blind initialization (subspace-based in our case). To do that, one could rely on some algorithms given in the literature (e.g., [41]). In fact, by assuming a linear-in-the-parameters model, a known nonlinear function and a finite impulse response, the proposed techniques can be applied to the following general model (given also in [25]):

$$y_r(k) = \sum_{n=0}^{M_{r,\mathrm{L}}} h_{r,\mathrm{L}}(n)u(k-n) + v_r(k) + \sum_{n=0}^{M_{r,\mathrm{NL}}}\sum_{l=n}^{M_{r,\mathrm{NL}}} h_{r,\mathrm{NL}}^{(2)}(n,l)u(k-n)u(k-l)$$

[7]Note that, in the blind case, the FIM is singular (due to the problem's ambiguities), in which case one needs to rely on the constrained CRB [40].

$$+\sum_{n}\sum_{l}\sum_{m} h^{(3)}_{r,\mathrm{NL}}(n,l,m)u(k-n)u(k-l)u^{*}(k-m)+\cdots \tag{44}$$

Furthermore, by considering the input-output relation, a block-oriented nonlinear model (e.g., Hammerstein model) can be expressed (approximated) as in (44) and hence, treated by the proposed techniques. However in this case, the estimated coefficients would represent products of the block-oriented linear and nonlinear parameters.

6.2. Algorithms' complexity discussion

It can be seen that the parameter estimation given by equations (22) and (23) requires, at each iteration, the calculation of the different conditional expectations given in equations (25), (26), (30) and (31). Hence, the global complexity of the proposed techniques is of order $\mathcal{O}(N_{iter}N_s 2^{BN_t(M+1)}2N_t(M+1)(4N_t(M+1)+2N_r))$, where N_t is the number of transmitters and N_{iter} is the total number of iterations needed for convergence.

It can be noticed that the algorithms' computational complexity is of the same order as the linear case[8]. It is worth pointing out that the computational complexity of our EM-based algorithm is mainly affected by the number of transitions, given by $2^{BN_t(M+1)}$. However, as it will be seen through simulation results, a relatively small number of pilots, data symbols and iterations are needed for convergence.

On the other hand, to reduce the computational complexity, one can apply some approximations for the posterior probability (as in [32] and references therein), use independent symbols through OFDM coding to avoid the forward-backward variables calculation or exploit some approximation/simplification approaches as proposed in [42]. Such complexity reduction is left for future work.

7. Performance analysis and discussion

This section provides the performance analysis of the proposed blind and semi-blind channel estimators for the considered nonlinear systems. For benchmarking, we consider a 'full' training-based (fully-pilot) estimator, as done in many works (e.g., [19, 25]), where all transmitted symbols (pilots and data) are assumed known and used to estimate the channel parameters. The estimation performance is evaluated in terms of the Normalized Root-Mean-Square Error (NMSE) given by:

$$\mathrm{NMSE} = \frac{1}{N_{mc}}\sum_{mc=1}^{N_{mc}} \frac{\left\|\hat{\mathbf{h}}_{mc}-\mathbf{h}_{exact}\right\|^2}{\left\|\mathbf{h}_{exact}\right\|^2}, \tag{45}$$

where $N_{mc}=500$ represents the number of independent Monte-Carlo runs used, $\hat{\mathbf{h}}_{mc}$ is the vector of estimated channel parameters at the mc-th run, and $\mathbf{h}_{exact}$ contains the true (exact) values of the channel coefficients. For data detection, the performance is evaluated in terms of the Symbol Error Rate (SER), which is the ratio between the wrongly detected symbols and the total number of transmitted data symbols. The channel coefficients are generated as i.i.d., unit-power, zero-mean (complex) Gaussian random variables, whereas the pilots and data symbols are uniformly randomly drawn from different QAM modulations (specified later for each experiment).

[8]For our model, there is approximately a factor 2 between the costs of the NL and the L cases.

The Signal-to-Noise Ratio (SNR) was defined as SNR $= \|\mathbf{H}\overline{\mathbf{u}}\|_F^2/\|\mathbf{v}\|_F^2$, where $\|\cdot\|_F$ is the Frobenius norm, $\mathbf{H}$ is defined in equation (3), and $\overline{\mathbf{u}} = [\overline{\mathbf{u}}(1), \ldots, \overline{\mathbf{u}}(N_s)] \in \mathbb{C}^{2(M+1)\times N_s}$ and $\mathbf{v} = [\mathbf{v}(1), \ldots, \mathbf{v}(N_s)] \in \mathbb{C}^{N_r \times N_s}$ are formed by stacking all values of $\overline{\mathbf{u}}(k)$ and $\mathbf{v}(k)$ (see equation (3)), respectively.

In the following, different experiments highlighting different aspects of our estimators are presented. Simulation parameters are summarized in Table 1. They are used for all experiments, unless otherwise specified.

Table 1: Simulation parameters

Parameter	Specification
Number of data symbols	$N_s = 100$
Number of pilot symbols	$N_p = 10$
Number of receive antennas	$N_r = 4$
Window length for SS	$N_w = 9$
Order of linear and nonlinear channels	$M = 4$

Experiment 1: Effect of neglecting the nonlinear terms (Figures 3–4)
In order to illustrate the effect of ignoring the nonlinear term and considering only a linear model, the system model given by (1) is rewritten as follows:

$$y_r(k) = \sum_{n=0}^{M_{r,\mathrm{L}}} h_{r,\mathrm{L}}(n)u(k-n) + \alpha \sum_{n=0}^{M_{r,\mathrm{NL}}} h_{r,\mathrm{NL}}(n)\tilde{u}(k-n) + v_r(k), \tag{46}$$

where α determines the weight of the nonlinear term that exists in the assumed underlying model (e.g., $\alpha = 0$: the underlying model is linear, $\alpha = 1$: the weight of the nonlinear term is equal to the linear one).
Figure 3 illustrates the NMSE vs. SNR behave when there is a nonlinearity (assuming different weights α indicated in superscript) in the underlying model. One can notice that, depending on the weight of the nonlinear term, the performance of the linear EM-based blind and semi-blind estimators could be highly degraded.
Furthermore, Figure 4 illustrates the SER vs. the weight α, for SNR = 5 dB. We notice a clear gain (especially for high α values) obtained by considering a nonlinear model (NL-B, NL-SB) as compared to the case where only linear terms are taken into account (L-B, L-SB).
In the sequel, the nonlinear model will be considered by setting $\alpha = 1$.
Experiment 2: NMSE vs. SNR (Figures 5– 7)
Figure 5 investigates the performance, in terms of NMSE, of the subspace(SS)-based estimator ($\mathrm{h_{SS}}$), and the blind and semi-blind EM-based estimators ($\mathrm{h_{EM-B}}, \mathrm{h_{EM-SB}}$) with respect to SNR for the quadratic nonlinearity considering 4-QAM (Figure 5a) and 16-QAM (Figure 5b) modulations. These estimators are benchmarked against the semi-blind Gaussian CRB (G-$\mathrm{CRB_{SB}}$) and the fully-pilot-based estimator ($\mathrm{h_{PILOT}}$). The blind EM-based estimator is initialized by the subspace-based one, whereas the semi-blind estimator is initialized by some pilots. One can observe that $\mathrm{h_{SS}}$ presents sub-optimal performance compared to $\mathrm{h_{PILOT}}$ (to all other estimators as well), whereas a significant improvement is observed with $\mathrm{h_{EM-B}}$ and $\mathrm{h_{EM-SB}}$. Moreover, the exploitation of the available pilots enhances the performance as $\mathrm{h_{EM-SB}}$ outperforms $\mathrm{h_{EM-B}}$ and hugs very tightly $\mathrm{h_{PILOT}}$ for moderate and high SNRs. On the other

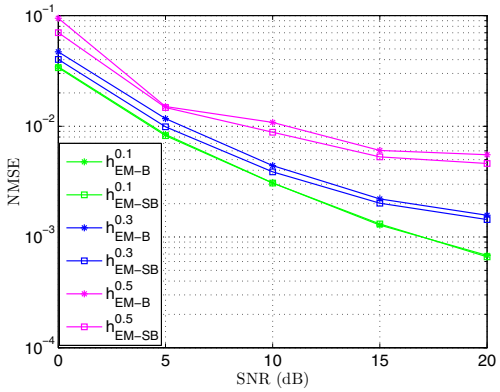


Figure 3: NMSE vs. SNR with different values of the weight α (see (46)), indicated in superscript, for the blind and semi-blind 'linear' EM-based estimators, with quadratic model and 4-QAM modulation

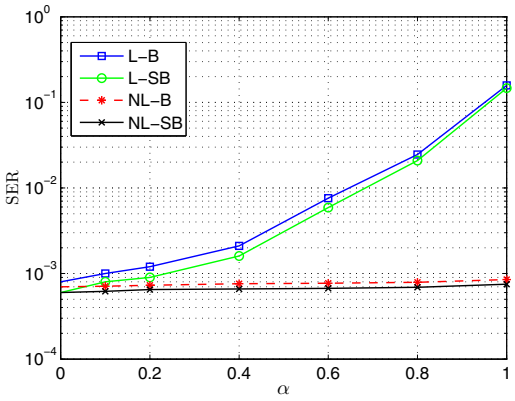


Figure 4: SER vs. α with SNR = 5 dB for linear blind and semi-blind (L-B, L-SB) EM-based estimators and for nonlinear (NL-B, NL-SB) ones, with quadratic model and 4-QAM modulation

hand, the EM-based estimators are found, interestingly, below the G-CRB$_{\mathrm{SB}}$, which reflects the fact that our channel identification solution outperforms all SOS-based identification methods. This result strongly supports the effectiveness of the aforementioned estimators for non-Gaussian QAM signals.

Also, a comparison has been performed with a cumulant-based technique [15], as illustrated in Figure 5a and Figure 5b, where, in the context of our scenarios, poor performance has been obtained since such a method requires higher number of symbols for convergence (at least 16000 data symbols as mentioned in [15]).

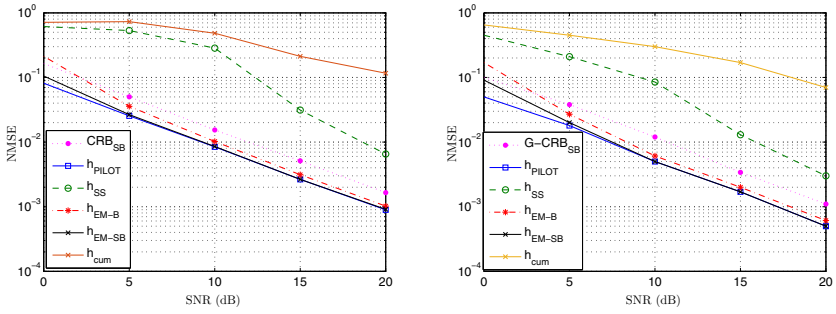
As mentioned in section 4.3, the proposed estimators can be easily applied to MIMO systems, where similar performance to the SIMO case is observed (see Figure 6).

From Figure 7a, one can observe that an important spatial diversity (higher number

of receive antennas, as is the case for massive MIMO systems) improves the performance for both $\mathrm{h}_{\mathrm{EM-B}}$ and $\mathrm{h}_{\mathrm{EM-SB}}$ at low SNRs.

In Figure 7b, a cubic nonlinear model has been used with a 16-QAM modulated input signal. One can note similar performance as provided in Figures 5 and 7a; EM-based estimators outperform the SS-based estimator and all SOS-based estimators (represented by the G-$\mathrm{CRB}_{\mathrm{SB}}$) for moderate to high SNRs, and $\mathrm{h}_{\mathrm{EM-SB}}$ outperforms $\mathrm{h}_{\mathrm{EM-B}}$ as it hugs more tightly the $\mathrm{h}_{\mathrm{PILOT}}$ curve.

Note that using a 4-QAM signal modulation with our cubic model will render the model linear, leading to a loss of identifiability. This can be easily seen by writing the 4-QAM sequence as $u(k) = \sqrt{2}e^{j(\frac{\pi}{4}+\eta(k)\frac{\pi}{2})}$ where $\eta(k) \sim \mathcal{U}\{0,1,2,3\}$, then, $\tilde{u}(k) = |u(k)|^2 u(k) = 2u(k)$, which is linear.



(a) (b)

Figure 5: NMSE vs. SNR for subspace-based estimator (h_{SS}), blind ($\mathrm{h}_{\mathrm{EM-B}}$) and semi-blind ($\mathrm{h}_{\mathrm{EM-SB}}$) EM-based estimators benchmarked against the semi-blind Gaussian CRB (G-$\mathrm{CRB}_{\mathrm{SB}}$), the cumulant-based technique [15] ($\mathrm{h}_{\mathrm{cum}}$) and the fully-pilot-based estimator ($\mathrm{h}_{\mathrm{PILOT}}$), in both cases: (a) quadratic model and 4-QAM modulation, and (b) quadratic model and 16-QAM modulation.

Experiment 3: NMSE vs. SNR with different linear and nonlinear channel orders (Figure 8)

In many practical situations, the channel order of the linear and nonlinear channels are different. Figure 8 illustrates the behavior of two different scenarios considered in *Experiment 1* with a channel order $M_{\mathrm{NL}} = 2$ for the nonlinear channel (the last two channel coefficients out of the previously used five coefficients have been considered null). On can note, particularly, that for high SNRs h_{SS} performs badly affecting the performance of $\mathrm{h}_{\mathrm{EM-B}}$.

Experiment 4: Speed of convergence (number of iterations vs. SNR, Figure 9)

Figure 9 shows the number of iterations needed for convergence for blind (B) and semi-blind (SB) EM-based estimators with respect to SNR. Considering 4-QAM and 16-QAM signal modulations and the considered nonlinear models (Quadratic and Cubic), we observe that for low SNRs ($0-5$ dB), the number of iterations varies according to the signal's model and modulation but is still relatively small (less than 30 at 5 dB). For moderate to high SNRs (> 10 dB), very few iterations (less than 8) are needed. In fact, for high SNRs (≥ 15 dB) only 2 or 3 iterations are needed independently of the signal's model and modulation. Moreover, using a higher number of antennas at the receiver (e.g., as in massive MIMO systems) leads to further reducing the number

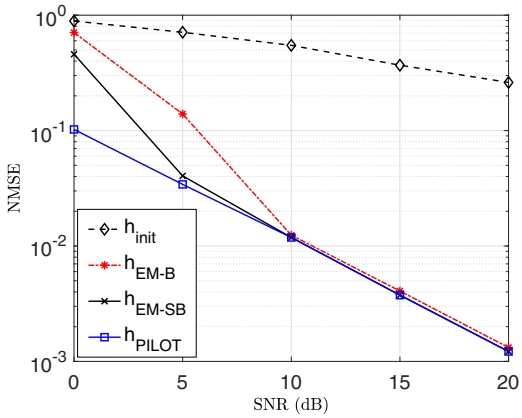


Figure 6: NMSE vs. SNR for blind ($\mathrm{h}_{\mathrm{EM-B}}$) and semi-blind ($\mathrm{h}_{\mathrm{EM-SB}}$) EM-based estimators benchmarked against the fully-pilot-based estimator ($\mathrm{h}_{\mathrm{PILOT}}$). We consider a quadratic model, 4-QAM modulation and a MIMO system with $N_t = 2, N_r = 4$, and $M = 2$. $\mathrm{h}_{\mathrm{init}}$ is included for reference and refers to a pilot-based estimator using available pilot symbols.

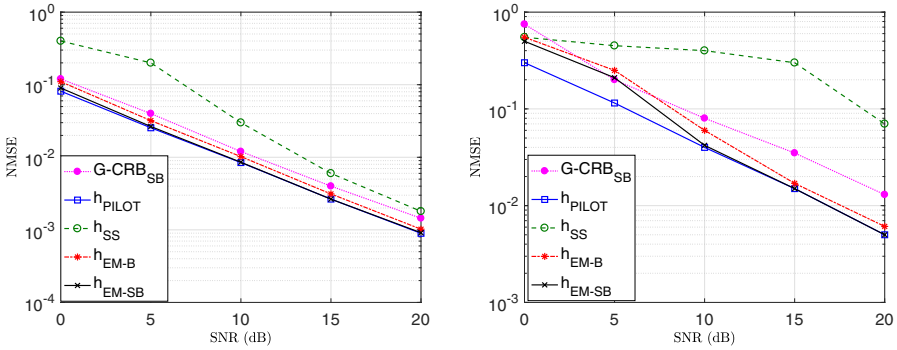


Figure 7: NMSE vs. SNR for subspace-based estimator (h_{SS}), blind ($\mathrm{h}_{\mathrm{EM-B}}$) and semi-blind ($\mathrm{h}_{\mathrm{EM-SB}}$) EM-based estimators benchmarked against the semi-blind Gaussian CRB ($\mathrm{G\text{-}CRB}_{\mathrm{SB}}$) and the fully-pilot-based estimator ($\mathrm{h}_{\mathrm{PILOT}}$) (a) in the case of a quadratic model, 4-QAM modulation and $N_r = 9$ receive antennas (b) in the case of a cubic model and 16-QAM modulation.

of iterations needed for convergence, especially at low SNRs, as can be seen from Figure 9b.

Experiment 5: NMSE vs. number of iterations at fixed SNR (Figure 10)

Figure 10 illustrates the variation of the NMSE with respect to the number of iterations needed for convergence of the subspace(SS)-based estimator (h_{SS}), and the blind and semi-blind EM-based estimators ($\mathrm{h}_{\mathrm{EM-B}}, \mathrm{h}_{\mathrm{EM-SB}}$) at SNR = 10 dB. Note that $\mathrm{h}_{\mathrm{EM-B}}$ is initialized by h_{SS}, whereas $\mathrm{h}_{\mathrm{EM-SB}}$ is initialized by some pilots such that $\mathrm{h}_{\mathrm{EM-B}}$

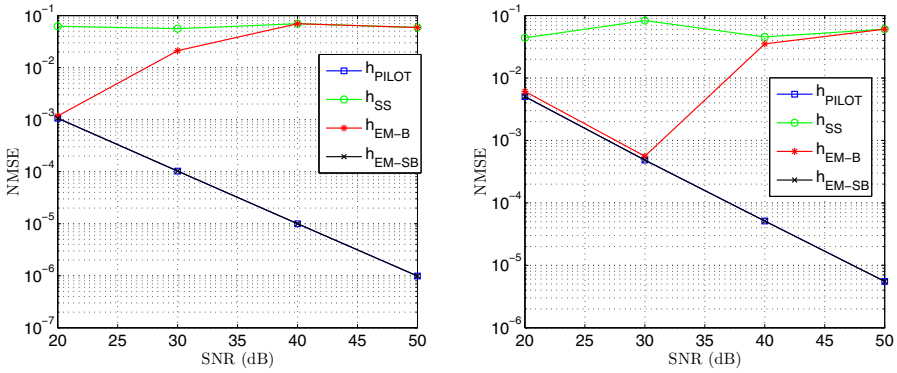


Figure 8: NMSE vs. SNR with $M_{\mathrm{L}} = 4$ and $M_{\mathrm{NL}} = 2$ for subspace-based estimator (h_{SS}), blind ($\mathrm{h}_{\mathrm{EM-B}}$) and semi-blind ($\mathrm{h}_{\mathrm{EM-SB}}$) EM-based estimators benchmarked against the fully-pilot-based estimator ($\mathrm{h}_{\mathrm{PILOT}}$). We consider: (a) quadratic model and 4-QAM modulation, and (b) cubic model and 16-QAM modulation.

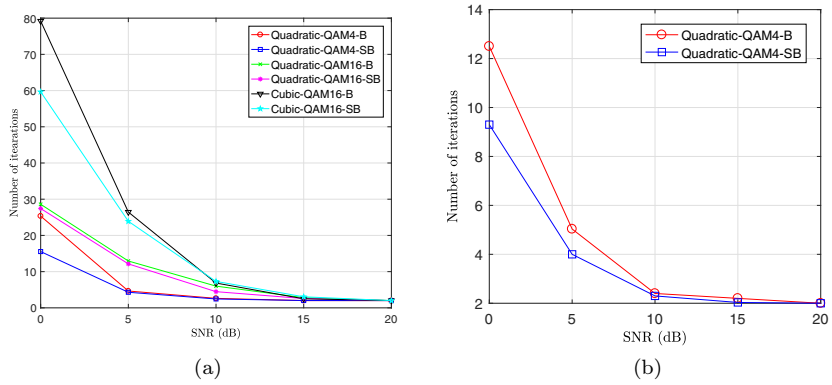


Figure 9: Number of iterations for convergence vs. SNR for blind (B) and semi-blind (SB) EM-based estimators (a) considering different nonlinearities and modulations (b) considering a quadratic model, a 4-QAM modulation, and $N_r = 9$ receive antennas.

and $\mathrm{h}_{\mathrm{EM-SB}}$ start from the same initial point (i.e., no iterations yet). We observe that after two iterations, $\mathrm{h}_{\mathrm{EM-SB}}$ converges to the solution given by the fully-pilot-based estimator ($\mathrm{h}_{\mathrm{PILOT}}$). We also observe that $\mathrm{h}_{\mathrm{EM-B}}$ comes very close to $\mathrm{h}_{\mathrm{PILOT}}$ after being initialized by h_{SS}. This result illustrates better the use of the word "refinement" in Figure 1, for the blind EM-based processing and is very interesting with regards to the computational complexity, especially for the nonlinear case where the number of channel coefficients is, in general, higher than the linear case.

Experiment 6: NMSE vs. number of pilots N_p (Figure 11)

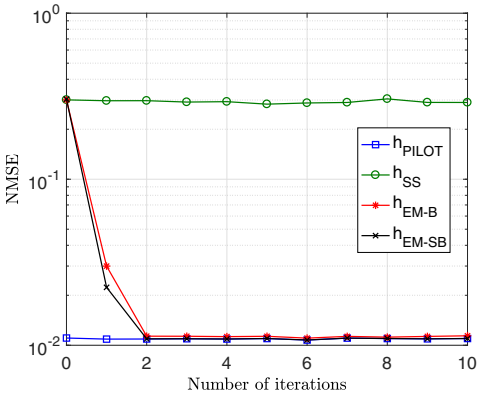
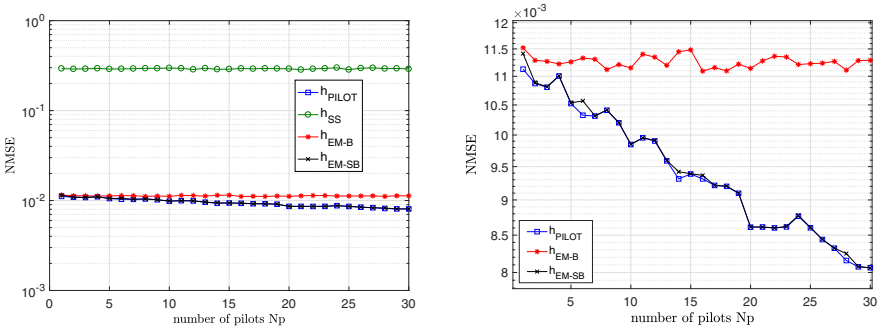


Figure 10: NMSE vs. number of iterations for subspace-based estimator ($\mathrm{h_{SS}}$), blind ($\mathrm{h_{EM-B}}$) and semi-blind ($\mathrm{h_{EM-SB}}$) EM-based estimators benchmarked against the fully-pilot-based estimator ($\mathrm{h_{PILOT}}$), in the case of a quadratic model and 4-QAM modulation at SNR = 10 dB.

We consider the quadratic model and 4-QAM modulation and we investigate the impact of the number of pilots N_p on the NMSE of the semi-blind EM-based estimator $\mathrm{h_{EM-SB}}$, at SNR = 10 dB (Figure 11; $\mathrm{h_{SS}}$ and $\mathrm{h_{EM-B}}$ are included for reference). We observe (Figure 11b) that only a slight decrease (around 3×10^{-3}) in the NMSE accompanies the increase in the number of pilots from 1 to 30, which indicates that only few pilots are needed to allow a quasi-optimal semi-blind channel estimation within the EM framework.



(a) NMSE vs. number of pilots.

(b) zoom for NMSE around 10^{-2}.

Figure 11: NMSE vs. number of pilots N_p for subspace-based estimator ($\mathrm{h_{SS}}$), blind ($\mathrm{h_{EM-B}}$) and semi-blind ($\mathrm{h_{EM-SB}}$) EM-based estimators benchmarked against the fully-pilot-based estimator ($\mathrm{h_{PILOT}}$), with a quadratic model and 4-QAM modulation at SNR = 10 dB.

Experiment 7: Symbol Error Rate (SER) vs. SNR (Figure 12)
Figure 12 investigates the performance of the proposed estimators in terms of Symbol Error Rate (SER) with respect to SNR. For blind and semi-blind EM-based estimators

(EM-B and EM-SB), a data detection is performed within an EM-based framework as described in section 3.5. For the subspace (SS) and the fully-pilots approaches, a zero-forcing is applied using the estimated channel coefficients. It can be noticed that performing data detection, using a ML-based approach leads to a significant performance gain and is part of a joint channel estimation and data detection.

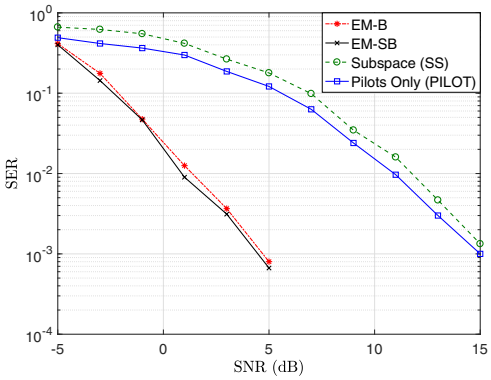


Figure 12: SER vs. SNR for subspace-based estimator (SS), blind (EM-B) and semi-blind (EM-SB) EM-based estimators benchmarked against the fully-pilot-based estimator (PILOT), with a quadratic model and 4-QAM modulation.

8. Conclusion

In this paper, blind and semi-blind Maximum Likelihood (ML) solutions are proposed for the identification of nonlinear multichannel communications systems. The ML criterion is maximized through the Expectation-Maximization (EM) algorithm. In the blind case, the EM algorithm is initialized by the subspace method followed by an original ambiguity removal technique. An identifiability study reveals that the success of the initialization step requires some stringent conditions that might not be verified for low order QAM modulations. Thus, an alternative solution is proposed, based on the semi-blind approach. Besides, theoretical performance bounds and identifiability results are provided for the considered problem and as well as the extension to a large class of nonlinear models. Moreover, it is shown that the proposed methods, not only outperforms the existing solutions in the literature but also, allows to achieve efficient joint channel estimation and data detection.

References

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

[2] J. Schoukens and L. Ljung, "Nonlinear system identification: A user-oriented road map," *IEEE Control Systems Magazine*, vol. 39, no. 6, pp. 28–99, 2019.

[3] M. R. Zefreh and J. A. Salehi, "Statistical modeling and performance characterization of ultrashort light pulse communication system using power-cubic optical nonlinear preprocessor," *IEEE Transactions on Communications*, vol. 63, no. 8, pp. 2948–2958, 2015.

[4] A. Amari, O. A. Dobre, R. Venkatesan, O. S. Kumar, P. Ciblat, and Y. Jaouën, "A survey on fiber nonlinearity compensation for 400 gb/s and beyond optical communication systems," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 4, pp. 3097–3113, 2017.

[5] P. Aggarwal and V. A. Bohara, "Analytical characterization of dual-band multi-user MIMO-OFDM system with nonlinear transmitter constraints," *IEEE Transactions on Communications*, vol. 66, no. 10, pp. 4536–4549, 2018.

[6] L. Chu, F. Wen, L. Li, and R. Qiu, "Efficient nonlinear precoding for massive MIMO downlink systems with 1-bit DACs," *IEEE Transactions on Wireless Communications*, vol. 18, no. 9, pp. 4213–4224, 2019.

[7] M. Yao, M. Sohul, R. Nealy, V. Marojevic, and J. Reed, "A digital predistortion scheme exploiting degrees-of-freedom for massive MIMO systems," in *IEEE International Conference on Communications (ICC)*, 2018, pp. 1–5.

[8] Y. Cai, Y. Xu, Q. Shi, B. Champagne, and L. Hanzo, "Robust joint hybrid transceiver design for millimeter wave full-duplex MIMO relay systems," *IEEE Transactions on Wireless Communications*, vol. 18, no. 2, pp. 1199–1215, 2019.

[9] T. Ogunfunmi, *Adaptive nonlinear system identification: the Volterra and Wiener model approaches*. Springer Science & Business Media, 2007.

[10] C. Cheng, Z. Peng, W. Zhang, and G. Meng, "Volterra-series-based nonlinear system modeling and its engineering applications: A state-of-the-art review," *Mechanical Systems and Signal Processing*, vol. 87, pp. 340–364, 2017.

[11] T. Ogunfunmi and S.-L. Chang, "Second-order adaptive Volterra system identification based on discrete nonlinear Wiener model," *IEE Proceedings-Vision, Image and Signal Processing*, vol. 148, no. 1, pp. 21–29, 2001.

[12] Y. V. Zakharov, G. P. White, and J. Liu, "Low-complexity RLS algorithms using dichotomous coordinate descent iterations," *IEEE Transactions on Signal Processing*, pp. 3150–3161, 2008.

[13] R. Claser, V. H. Nascimento, and Y. V. Zakharov, "A low-complexity RLS-DCD algorithm for Volterra system identification," in *24th European Signal Processing Conference (EUSIPCO)*, 2016, pp. 6–10.

[14] G. L. Sicuranza and A. Carini, "A multichannel hierarchical approach to adaptive Volterra filters employing filtered-X affine projection algorithms," *IEEE Transactions on Signal Processing*, vol. 53, no. 4, pp. 1463–1473, 2005.

[15] N. Kalouptsidis and P. Koukoulas, "Blind identification of Volterra-Hammerstein systems," *IEEE Transactions on Signal Processing*, vol. 53, no. 8, pp. 2777–2787, 2005.

[16] J. C. Gómez and E. Baeyens, "Subspace-based identification algorithms for Hammerstein and Wiener models," *European Journal of Control*, vol. 11, no. 2, pp. 127–136, 2005.

[17] L. Yao and C.-C. Lin, "Identification of nonlinear systems by the genetic programming-based Volterra filter," *IET Signal Processing*, vol. 3, no. 2, pp. 93–105, 2009.

[18] K. Batselier, Z. Chen, and N. Wong, "Tensor Network alternating linear scheme for MIMO Volterra system identification," *Automatica*, vol. 84, pp. 26–35, 2017.

[19] O. Karakuş, E. Kuruoğlu, and M. A. Altınkaya, "Bayesian Volterra system identification using reversible jump MCMC algorithm," *Signal Processing*, vol. 141, pp. 125–136, 2017.

[20] L. Vanbeylen, R. Pintelon, and J. Schoukens, "Blind Maximum-Likelihood Identification of Wiener Systems," *IEEE Transactions on Signal Processing*, vol. 57, no. 8, pp. 3017–3029, 2009.

[21] L. Vanbeylen and R. Pintelon, "Blind Maximum-Likelihood identification of Wiener and Hammerstein nonlinear block structures," in *Block-oriented nonlinear system identification*. Springer, 2010, pp. 273–291.

[22] E.-W. Bai and M. Fu, "A blind approach to Hammerstein model identification," *IEEE Transactions on Signal Processing*, vol. 50, no. 7, pp. 1610–1619, 2002.

[23] L. Vanbeylen, R. Pintelon, and J. Schoukens, "Blind maximum likelihood identification of Hammerstein systems," *Automatica*, vol. 44, no. 12, pp. 3139–3146, 2008.

[24] C. A. R. Fernandes, "Nonlinear MIMO communication systems: Channel estimation and information recovery using Volterra models," Ph.D. dissertation, Université de Nice Sophia Antipolis, 2009.

[25] L. Tao, H. Tan, C. Fang, and N. Chi, "Volterra series based blind equalization for nonlinear distortions in short reach optical CAP system," *Optics Communications*, vol. 381, pp. 240–243, 2016.

[26] C.-H. Tseng, "Estimation of cubic nonlinear bandpass channels in orthogonal frequency-division multiplexing systems," *IEEE Transactions on Communications*, vol. 58, no. 5, pp. 1415–1425, 2010.

[27] G. B. Giannakis and E. Serpedin, "Linear multichannel blind equalizers of nonlinear FIR Volterra channels," *IEEE Transactions on Signal Processing*, vol. 45, no. 1, pp. 67–81, 1997.

[28] K. Abed-Meraim, P. Loubaton, and E. Moulines, "A subspace algorithm for certain blind identification problems," *IEEE Transactions on Information Theory*, vol. 43, no. 2, pp. 499–511, 1997.

[29] M. Van Barel and A. Bultheel, "A lookahead algorithm for the solution of block Toeplitz systems," *Linear Algebra and its Applications*, vol. 266, pp. 291–335, 1997.

[30] E. Serpedin, P. Ciblat, G. B. Giannakis, and P. Loubaton, "Performance analysis of blind carrier phase estimators for general QAM constellations," *IEEE Transactions on Signal Processing*, vol. 49, no. 8, pp. 1816–1823, 2001.

[31] S. A. Wahab Shah, K. Abed-Meraim, and T. Al-Naffouri, "Blind source separation algorithms using hyperbolic and Givens rotations for high-order QAM constellations," *IEEE Transactions on Signal Processing*, vol. 66, no. 7, pp. 1802 – 1816, 2017.

[32] K. Abed-Meraim and E. Moulines, "A maximum likelihood solution to blind identification of multichannel FIR filters," in *7th European Signal Processing Conference (EUSIPCO)*, 1994, pp. 1011–1014.

[33] G. K. Kaleh and R. Vallet, "Joint parameter estimation and symbol detection for linear or nonlinear unknown channels," *IEEE Transactions on Communications*, vol. 42, no. 7, pp. 2406–2413, 1994.

[34] S.-Z. Yu and H. Kobayashi, "An efficient forward-backward algorithm for an explicit-duration hidden Markov model," *IEEE Signal Processing Letters*, vol. 10, no. 1, pp. 11–14, 2003.

[35] S. M. Kay, *Fundamentals of Statistical Signal Processing*. Prentice Hall PTR, 1993.

[36] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 39, no. 1, pp. 1–22, 1977.

[37] J. R. Barry, E. A. Lee, and D. G. Messerschmitt, *Digital communication*. Springer Science & Business Media, 2012.

[38] S. Park, E. Serpedin, and K. Qaraqe, "Gaussian assumption: the least favorable but the most useful," *IEEE Signal Processing Magazine*, vol. 30, no. 4, pp. 183 – 186, 2013.

[39] W. Wirtinger, "Zur formalen theorie der funktionen von mehr komplexen veränderlichen," *Mathematische Annalen*, vol. 97, no. 1, pp. 357–375, 1927.

[40] A. K. Jagannatham and B. D. Rao, "Cramer-Rao lower bound for constrained complex parameters," *IEEE Signal Processing Letters*, vol. 11, no. 11, pp. 875–878, 2004.

[41] R. Lopez-Valcarce and S. Dasgupta, "Blind equalization of nonlinear channels from second-order statistics," *IEEE Transactions on Signal Processing*, vol. 49, no. 12, pp. 3084–3097, 2001.

[42] A. Ladaycia, A. Belouchrani, K. Abed-Meraim, and A. Mokraoui, "Semi-blind MIMO-OFDM channel estimation using expectation maximisation like techniques," *IET Communications*, vol. 13, no. 20, pp. 3452–3462, 2019.

of Interest Statement

Manuscript title: Maximum Likelihood based Identification for Nonlinear Multichannel Communications Systems

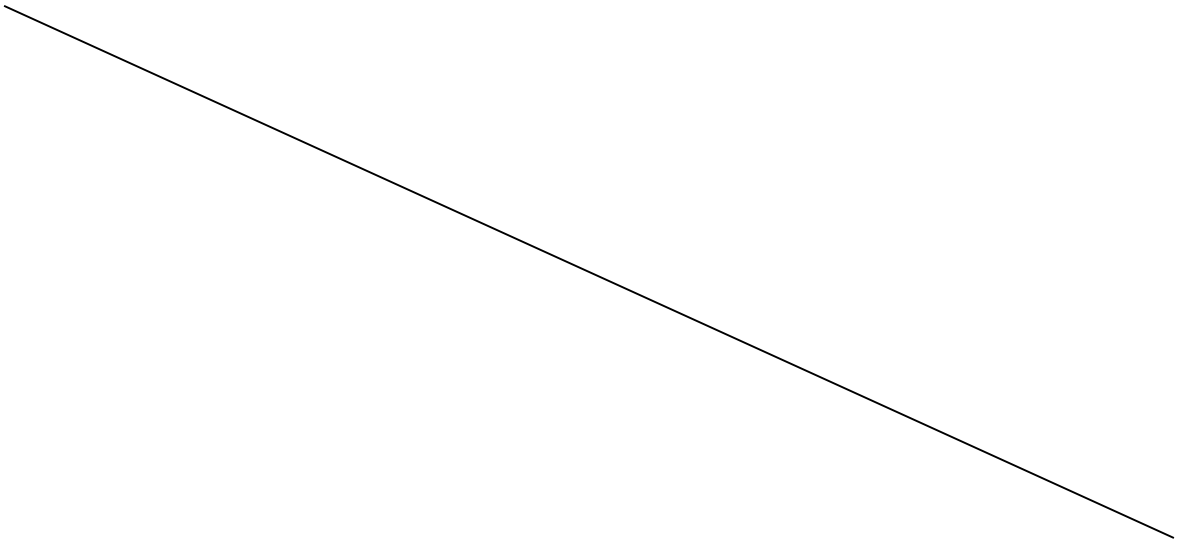
The authors whose names are listed immediately below certify that they have NO affiliations with or involvement in anyorganization or entity with any financial interest (such as honoraria; educational grants; participation in speakers' bureaus;membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements), or non-financial interest (such as personal or professional relationships, affiliations, knowledge or beliefs) in the subject matter or materials discussed in this manuscript.

Author names:

- Ouahbi Rekik
- Karim Abed-Meraim
- Mohamed Nait-Meziane
- Anissa Mokraoui
- Nguyen Linh Trung

The authors whose names are listed immediately below report the following details of affiliation or involvement in an organization or entity with a financial or non-financial interest in the subject matter or materials discussed in this manuscript. Please specify the nature of the conflict on a separate sheet of paper if the space below is inadequate.

Author names:



Ouahbi Rekik: Writing- Reviewing and Editing.

Karim Abed-meraim: Theoritical development, editing.

Mohamed Nait-Meziane: Writing- Reviewing and Editing.

Anissa Mokraoui : Reviewing and Editing.

Nguyen Linh Trung: Reviewing and Editing.

HIGHLIGHTS

June 2021

- Nonlinear distortions are important issues in many communications systems. This work proposes original blind and semi-blind solutions for the identification of nonlinear, finite alphabet, multichannel systems.
- The proposed solutions rely on the maximum likelihood approach and are based on the EM-algorithm. The blind one is initialized by a subspace method combined with an original ambiguity removal technique. They lead to a significant performance gain as compared to the state-of-the-art.
- We provide theoretical performance bounds and identifiability results for the considered problem and show how our solutions can be extended to a large class of nonlinear models.
- Our methods, not only outperform the existing solutions in the literature but also, allow to achieve channel estimation and data detection simultaneously.