Blind Deconvolution of DS-CDMA Signals by Means of Decomposition in Rank-$(1, L, L)$ Terms

Lieven De Lathauwer, *Senior Member, IEEE*, and Alexandre de Baynast, *Member, IEEE*

***Abstract*—In this paper, we present a powerful technique for the blind extraction of direct-sequence code-division multiple access (DS-CDMA) signals from convolutive mixtures received by an antenna array. The technique is based on a generalization of the canonical or parallel factor decomposition (CANDECOMP/PARAFAC) in multilinear algebra. We present a bound on the number of users under which blind separation and deconvolution is guaranteed. The solution is computed by means of an alternating least squares (ALS) algorithm. The excellent performance is illustrated by means of a number of simulations. We include an explicit expression of the Cramér–Rao bound (CRB) of the transmitted symbols.**

***Index Terms*—Blind deconvolution, block term decomposition, canonical decomposition, code division multiple access, parallel factor model.**

I. INTRODUCTION

IN this paper we present a new algebraic technique for the blind extraction of direct-sequence code-division multiple access (DS-CDMA) signals from convolutive mixtures received by an antenna array. The convolutive mixtures observed at each array element result from the superposition of all user's signals after propagation through transmission channels with memory. We tackle the problem by means of multilinear algebraic tools. Multilinear algebra is the algebra of higher-order tensors, which are quantities of which the elements are addressed by more than two indices; as such, higher-order tensors are the multi-way generalization of vectors (first order) and matrices (second order). Our approach more specifically fits in the framework of canonical decomposition (CANDECOMP) or parallel factor analysis (PARAFAC), which is a fundamental concept in multilinear algebra [5], [12]–[14], [16], [33]. We will use the abbreviation CP to denote CANDECOMP/PARAFAC.

Our technique is a generalization of the work of Sidiropoulos *et al.*, who were the first to adopt a multilinear algebraic point of view w.r.t. CDMA data [28]. They showed that, if there is no intersymbol interference (ISI), the data received by the antenna array can be stacked in a third-order tensor that can be decomposed in a sum of third-order rank-1 terms, where each term corresponds to the signal transmitted by one user. (A higher-order rank-1 term is defined as the outer product of a number of vectors. For a third-order tensor $\mathcal{A}$ and three vectors U, V, and W, this means that $a_{ijk} = u_i v_j w_k$ for all values of the indices, which will be written as $\mathcal{A} = U \circ V \circ W$.) This technique can be used in the case of small delay spread.

The case with large delay spread, in which there is ISI, was considered in [31]. The solution proposed in [31] consists of two stages: i) separate the users by exploiting partial uniqueness of the so-called parallel factor model with linear dependencies (PARALIND) [4], and ii) recover the sequence transmitted by each user by single-input multiple-output (SIMO) deconvolution of a finite impulse response (FIR) filter [19], [21], [42].

In this paper we treat the case with ISI by means of the concept of block term decompositions (BTDs), which we have recently introduced [8], [9], [10]. The decomposition that we will use in this paper is, contrary to CP, not a sum of outer products of three vectors, but a sum of outer products of a vector and a matrix, which itself results from the (inner) product of two matrices. Essential uniqueness of this decomposition can be demonstrated under conditions that are more relaxed than the ones that have so far been obtained for PARALIND. Our model has the peculiarity that one of the two matrices in each product has a Toeplitz structure. This structure will be exploited in the computations. We mention that in [31] the Toeplitz structure is only taken into account in the second stage of the algorithm.

Fig. 1 schematically represents the DS-CDMA system under consideration. Each symbol of a given user is multiplied by that user's spreading sequence. After transmission the signals are captured on an array of antennas.

We work under the same assumptions as in [31]. We assume that the spreading gain is known or has been estimated. Also the number of active users is assumed to be known. For simplicity we assume throughout the paper that the co-channel and adjacent-channel interferences are small and can be considered as additive Gaussian noise. (If needed, the procedure to be presented in this paper could be repeated for different user numbers and the most plausible value retained. This is the standard procedure for CP, of which our approach is a generalization. Several techniques have been developed for the estimation of the number of components in CP [2], [3]. These can probably be generalized to BTD. The generalization is outside the scope of this paper. If the number of users is bounded as in [31], it can be estimated as the number of different generalized eigenvalues

Manuscript received March 8, 2007; revised August 20, 2007. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Petr Tichavsky. Part of this research was carried out when L. De Lathauwer was with the Research Group ETIS, Cergy-Pontoise, France. Research supported by: 1) the Research Council K.U.Leuven: GOA-Ambiorics and CoE EF/05/006 Optimization in Engineering (OPTEC) and CIF1; 2) F.W.O.: a) project G.0321.06 and b) Research Communities ICCoS, ANMMM and MLDM; 3) the Belgian Federal Science Policy Office: IUAP P6/04 [(DYSCO), Dynamical Systems, Control and Optimization," 2007–2011); and 4) EU: ERNSI.

L. De Lathauwer is with the Research Group ESAT-SCD, Katholieke Universiteit Leuven (K.U.Leuven), B-3001 Leuven-Heverlee, Belgium (e-mail: delathau@esat.kuleuven.be), and also with the K.U.Leuven Campus Kortrijk, Subfaculty Sciences, 8500 Kortrijk, Belgium.

A. de Baynast is with the Wireless Networks Department of RWTH Aachen, D-52072, Aachen, Germany (e-mail: ade@mobnets.rwth-aachen.de).

Digital Object Identifier 10.1109/TSP.2007.910469

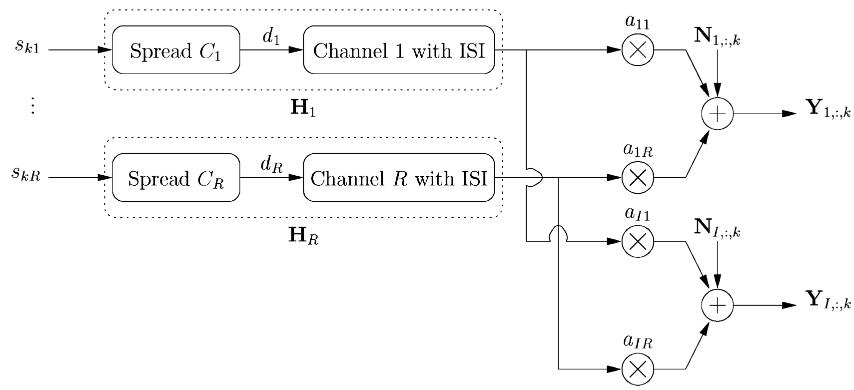

Fig. 1. Discrete-time baseband-equivalent scheme of CDMA transmission.

of a matrix pencil. See the discussion of the EVD-based solution in Section IV.) The signals are assumed to be synchronized at the symbol-level and the transmission channels are time-invariant over the measurement interval. Next, we assume that multipath reflections only take place in the far field of the receive antenna array. This implies that, for each user, the multipath/delay channels to the different antennas are the same up to a flat fading/antenna response factor [39]. Finally, we assume that an upper-bound of the maximum delay spread over all user's wireless channels is known. Our technique has the same conceptual advantages as the ISI-free technique [28].

- Because the (deterministic) algebraic structure of the data is exploited, the method works also for small sample sizes. Hence, the technique can be used in the case of block fading with a block size of the order of a few symbols. This is an important advantage over statistical techniques, where many more data are needed to obtain consistent estimates [38]. The same remark applies to adaptive algorithms, where the channel variations should be slow in comparison to the convergence speed (e.g., Fast-CMA requires at least 250 iterations to converge over a stationary channel [1]).
- The spreading codes need not be orthogonal and their knowledge is not required.
- No information is required regarding the multipath characteristics. The antennas do not have to be calibrated.
- The transmitted signals do not have to be constant modulus (CM) and the modulation does not have to be known.
- The transmitted signals need not be statistically independent nor uncorrelated (from a conceptual algebraic point of view). Of course, in practice, if signals are highly correlated, this may badly affect the conditioning of the problem and worsen the performance [slower convergence speed and higher bit error rate (BER)].

The paper is organized as follows. In the next section, we summarize the result obtained in [28]. In Section III we briefly state the central multilinear algebraic theorem on which our technique is based. For more details, the interested reader is referred to [8], [9], [10]. Section IV explains how this result can be applied to the problem at hand. Section V illustrates the technique by means of some simulations. Section VI is the conclusion. In the Appendix we determine the Cramér–Rao bound (CRB) for the blind deconvolution problem.

Notation: The Kronecker product is denoted by $\otimes$. The Moore–Penrose pseudoinverse is denoted by $(\cdot)^\dagger$. For $\mathbf{X} = [X_1 \dots X_N] \in \mathbb{C}^{M \times N}$, we define

$$\mathrm{vec}(\mathbf{X}) \stackrel{\mathrm{def}}{=} \begin{bmatrix} X_1 \\ \vdots \\ X_N \end{bmatrix}.$$

We sometimes use the MATLAB colon notation. $(\mathbf{A})_{i,:}$ and $(\mathbf{A})_{:,j}$ denote the ith row and the jth column of a matrix $\mathbf{A}$, respectively. The ith $(J \times K)$ slice of a tensor $\mathcal{A} \in \mathbb{C}^{I \times J \times K}$ is denoted by $(\mathbf{A})_{i,:,:}$.

II. CP Approach in the Absence of ISI

Let us start from the following noiseless/memoryless data model for multiuser DS-CDMA:

$$y_{ijk} = \sum_{r=1}^{R} a_{ir} c_{jr} s_{kr} \tag{1}$$

in which y_{ijk} is the output of the ith antenna for chip j and symbol k ($1 \le i \le I, 1 \le j \le J', 1 \le k \le K$, with I the number of antennas, J' the code length and K the number of transmitted symbols), a_{ir} is the fading/gain between user r and antenna element i, c_{jr} is the jth chip of the spreading sequence of user r and s_{kr} is the kth symbol transmitted by user r. Now let us assume that there is interchip interference (ICI) over at most L' chips. For any user, the length of the impulse response of the multipath channel is at most L' (at the chip rate). As proposed in [26] (see further), we add L' trailing zeros at the end of each spreading code. This makes that, at the receive antenna array, the signal related to a symbol s_{kr} has died out before the signal related to the following symbol $s_{k+1,r}$ arrives. In other words, due to the adding of a sufficient number of trailing zeros, there is no ISI. We have now the following data model:

$$y_{ijk} = \sum_{r=1}^{R} a_{ir} h_{jr} s_{kr}. \tag{2}$$

In this equation, h_{jr}, for varying j and fixed r, is the result of convolving the spreading sequence of user r with the impulse response of its propagation channel. Here we suppose that $1 \le$

$j \leq J = J' + L'$. Equation (2) can be written in a tensor format as

$$\mathcal{Y} = \sum_{r=1}^{R} A_r \circ H_r \circ S_r \tag{3}$$

with $\mathcal{Y} \in \mathbb{C}^{I\times J\times K}, A_r \in \mathbb{C}^I, H_r \in \mathbb{C}^J$ and $S_r \in \mathbb{C}^K$. The symbol $\circ$ denotes the tensor outer product. Equation (3) is a decomposition of $\mathcal{Y}$ in rank-1 terms. This is a CP model [5], [12]–[14], [16], [33]. This multilinear point of view w.r.t. CDMA data was adopted for the first time in [28].

Define $\mathbf{A} = [A_1 \ldots A_R] \in \mathbb{C}^{I\times R}, \mathbf{H} = [H_1 \ldots H_R] \in \mathbb{C}^{J\times R}, \mathbf{S} = [S_1 \ldots S_R] \in \mathbb{C}^{K\times R}$. Equation (3) has a number of inherent indeterminacies. First, the order of the rank-1 terms is arbitrary. Secondly, A_r, H_r, S_r may be rescaled $(1 \leq r \leq R)$ provided the scaling factors compensate each other.

Now let us introduce the following variant of the "rank" of a matrix [13]

Definition 1: The k-rank $k(\mathbf{A})$ of a matrix $\mathbf{A}$ is the maximal number such that *any* set of k columns of $\mathbf{a}$ is linearly independent.

It was shown in [15], [17], [28], [35] that decomposition (3) is unique, apart from the trivial indeterminacies mentioned in the previous paragraph, if

$$k(\mathbf{A}) + k(\mathbf{H}) + k(\mathbf{S}) \geq 2(R+1). \tag{4}$$

Because of user-independent fading and multipath, the CP matrices $\mathbf{A}$ and $\mathbf{H}$ are in practice full k-rank with probability 1. If the transmitted symbols belong to a finite alphabet (FA), then there is a chance that $\mathbf{S}$ is not full k-rank (the sequences of the different users might be equal, for instance). However, this becomes more and more unlikely for longer datasets. Hence, in practice, because of persistency of excitation, we assume that $\mathbf{S}$ is full k-rank as well. This means that the number of users that can simultaneously be processed, can be considered bounded as

$$\min(I,R) + \min(J,R) + \min(K,R) \geq 2(R+1). \tag{5}$$

The CP components can be estimated by means of an Alternating Least Squares (ALS) algorithm, in which the multi-linearity of the data is exploited [5], [28], [33]. (Other approaches are possible but will not be considered in this paper [6], [7], [18], [25], [36], [41].) In each step, the estimates of two of the matrices $\mathbf{A}, \mathbf{H}, \mathbf{S}$ are fixed and the third is conditionally updated. Computation of the update that is optimal in least-squares sense simply amounts to solving an overdetermined set of linear equations, because (3) is multi-linear in its unknowns. In ALS, one iterates over such conditional updates, thereby monotonically decreasing the cost function

$$f(\mathbf{A},\mathbf{H},\mathbf{S}) = \|\mathcal{Y} - \sum_{r=1}^{R} A_r \circ H_r \circ S_r\|^2 \stackrel{\text{def}}{=} \sum_{ijk} |y_{ijk} - \sum_{r=1}^{R} a_{ir}h_{jr}s_{kr}|^2. \tag{6}$$

In this approach it is essential that there is no ISI. To guarantee this in the case of convolutive transmission channels, one proposes to follow a "discard prefix" or "guard chips" strategy in [28]. This means that zeros are added at the beginning or the end of each code vector and that, at the receiver, the transient signal between consecutive symbols is discarded. To make sure that there is no ISI, the number of zeros should be at least equal to the maximal delay spread over all transmission channels. This is feasible when the maximal delay spread is small compared to the spreading gain. However, assume for instance that the propagation channel is of length $L' = 2J'$, which means that there is ISI over three symbols. In this case $2J'$ zeros have to be added per J' transmitted chips, and 67% of the received signal is discarded. In our approach we will not introduce superfluous zeros but exploit the algebraic structure of the convolved signal. The price that has to be paid is a moderate decrease of the maximum number of users that can be allowed. The new technique will be explained in Section IV. First we will briefly sketch the necessary multilinear algebraic background.

III. The Decomposition in Rank-$(1, L, L)$ Terms

Let us first introduce some basic definitions. Column and row vectors in matrix algebra are generalized to n-mode vectors in multilinear algebra (a column vector being a 1-mode vector and a row vector being a 2-mode vector). An n-mode vector of an $(I_1 \times I_2 \times I_3)$-tensor $\mathcal{A}$ is formally defined as an I_n-dimensional vector obtained from $\mathcal{A}$ by varying the index i_n and keeping the other indexes fixed. The n-rank of a higher-order tensor is the obvious generalization of the column (row) rank of matrices: it equals the dimension of the vector space spanned by the n-mode vectors. An important difference with the rank of matrices, is that the different n-ranks of a higher-order tensor are not necessarily the same. A tensor of which the 1-mode rank is equal to R_1, the 2-mode rank equal to R_2 and the 3-mode rank equal to R_3 is called a rank-(R_1, R_2, R_3) tensor. A rank-$(1,1,1)$ tensor is briefly called a rank-1 tensor; as mentioned before, it is equal to the outer product of three vectors.

Using these concepts and terminology, we have the following definition [9].

Definition 2: A decomposition of a tensor $\mathcal{T} \in \mathbb{C}^{I\times J\times K}$ in a sum of rank-$(1, L, L)$ terms is a decomposition of $\mathcal{T}$ of the form

$$\mathcal{T} = \sum_{r=1}^{R} A_r \circ \mathbf{E}_r \tag{7}$$

i.e.,

$$t_{ijk} = \sum_{r=1}^{R} (A_r)_i (\mathbf{E}_r)_{jk}, \quad \forall i,j,k$$

in which the $(J \times K)$ matrices $\mathbf{E}_r$ are rank-L.

If we factorize $\mathbf{E}_r$ as $\mathbf{B}_r \cdot \mathbf{C}_r^T$, in which the $(J \times L)$ matrix $\mathbf{B}_r$ and the $(K \times L)$ matrix $\mathbf{C}_r$ are rank-$L, r = 1, \ldots, R$, then we can write (7) as

$$\mathcal{T} = \sum_{r=1}^{R} A_r \circ \left(\mathbf{B}_r \cdot \mathbf{C}_r^T\right). \tag{8}$$

Note that the mode-1, mode-2 and mode-3 rank of each term are indeed equal to 1, L, and L, respectively: the mode-1 vectors are proportional to A_r, the vector space generated by the mode-2 vectors of the rth term is the column space of $\mathbf{B}_r$, and the vector

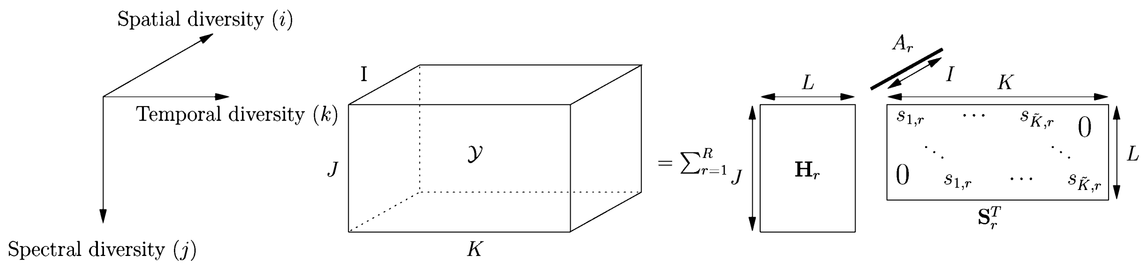

Fig. 2. Decomposition in rank-$(1, L, L)$ terms of the received data tensor.

space generated by the mode-3 vectors is the column space of $\mathbf{C}_r$. Decomposition (8) generalizes CP in the sense that, in CP, we have $L = 1$.

It is clear that in (8) one can arbitrarily permute the different rank-$(1, L, L)$ terms. Also, one can postmultiply $\mathbf{B}_r$ by any nonsingular $(L \times L)$ matrix $\mathbf{F}_r$, provided $\mathbf{C}_r^T$ is premultiplied by the inverse of $\mathbf{F}_r$. Moreover, the factors of a given rank-$(1, L, L)$ term may be arbitrarily scaled, as long as their product remains the same. We call the decomposition essentially unique when it is only subject to these trivial indeterminacies.

Define $\mathbf{A} = [A_1 \cdots A_R], \mathbf{B} = [\mathbf{B}_1 \cdots \mathbf{B}_R]$ and $\mathbf{C} = [\mathbf{C}_1 \cdots \mathbf{C}_R]$. Next, let us introduce the following generalization of the k-rank.

Definition 3: The k'-rank $k'(\mathbf{B})$ of a partitioned matrix $\mathbf{B} = [\mathbf{B}_1 \dots \mathbf{B}_R]$ is the maximal number such that the columns of any set of k' submatrices of $\mathbf{B}$ are linearly independent.

We now have the following uniqueness theorem.

Theorem 1 [9]: Consider the decomposition in rank-$(1, L, L)$ terms (8). This decomposition is unique, up to the trivial indeterminacies specified above, if

$$JK \geq R \quad \text{and} \quad k(\mathbf{A}) + k'(\mathbf{B}) + k'(\mathbf{C}) \geq 2(R+1). \quad (9)$$

This condition generalizes (4) to the decomposition in rank-$(1, L, L)$ terms. The proof in [8], [9] actually only holds under the condition that the entries of $\mathbf{A}, \mathbf{H}$ and $\mathbf{S}$ are drawn from continuous probability densities. Furthermore, we assumed that in an alternative decomposition, represented by $\bar{\mathbf{A}}, \bar{\mathbf{H}}$ and $\bar{\mathbf{S}}, k'_{\bar{\mathbf{H}}}$ and $k'_{\bar{\mathbf{S}}}$ are maximal under the given dimensionality constraints. In practice, these constraints are of no importance to the application studied in this paper.

IV. Generalized CP Approach in the Presence of ISI

A. Data Model

We consider the transmission of $\tilde{K}$ symbols. We assume that there is ICI over at most L' chips. Let $L = \lceil \frac{L'}{J} \rceil$ be the maximum channel length at the symbol rate, meaning that interference is occurring over maximally L symbols. The coefficients resulting from the convolution between the channel impulse response and the spreading sequence of the rth user are collected in a vector H_r of size JL. More specifically, $(H_r)_{j+(l-1)J}$ is the coefficient of the overall impulse response corresponding to the ith chip and the lth symbol. If the total number of coefficients is less than JL, the remaining entries are set equal to zero. We denote by $x_{jk}^{(r)}$ the jth chip of the kth symbol period of the signal of the rth user upon arrival at the antenna array. Denoting the kth symbol transmitted by the rth user by $s_{k,r}$, as before, we have

$$x_{jk}^{(r)} = \sum_{l=0}^{L-1} (H_r)_{j+(l-1)J} s_{k-l,r} \quad (10)$$

where $s_{k-l,r}$ is taken equal to zero if $k-l \leq 0$ or $k-l > \tilde{K}$. Let a_{ir} be the response of the ith antenna to the signal of the rth user, where we assume that the path loss is combined with the antenna gain. The jth chip of the kth symbol period of the overall signal received by the ith antenna array can now be written as

$$\begin{aligned} y_{ijk} &= \sum_{r=1}^{R} a_{ir} x_{jk}^{(r)} \\ &= \sum_{r=1}^{R} a_{ir} \sum_{l=0}^{L-1} (H_r)_{j+(l-1)J} s_{k-l,r}. \end{aligned} \quad (11)$$

Let $\mathbf{H}_r \in \mathbb{C}^{J \times L}$ be a matrix in which the coefficients of H_r are stacked column per column: $(\mathbf{H}_r)_{j,l} = (H_r)_{j+(l-1)J}, r = 1, \dots, R$. Then our data model is as follows:

$$\begin{aligned} y_{ijk} &= \sum_{r=1}^{R} a_{ir} \sum_{l=0}^{L-1} (\mathbf{H}_r)_{j,l} s_{k-l,r}, \\ & 1 \leq i \leq I, \quad 1 \leq j \leq J, \\ & 1 \leq k \leq K = \tilde{K} + L - 1. \end{aligned} \quad (12)$$

Equation (12) can be written in a tensor format as

$$\mathcal{Y} = \sum_{r=1}^{R} A_r \circ \left(\mathbf{H}_r \cdot \mathbf{S}_r^T\right) \quad (13)$$

in which $\mathbf{S}_r^T \in \mathbb{C}^{L \times K}$ is a Toeplitz matrix of which the first row consists of the $\tilde{K}$ subsequent symbols transmitted by user r, followed by $L-1$ zeros. Equation (13) is not an expansion in rank-1 terms, but a decomposition in rank-$(1, L, L)$ terms, i.e., each term consists of the outer product of a vector and a rank-L matrix. The decomposition is visualized in Fig. 2.

B. Uniqueness

Define $\mathbf{a} = [A_1 \dots A_R] \in \mathbb{C}^{I \times R}, \mathbf{H} = [\mathbf{H}_1 \dots \mathbf{H}_R] \in \mathbb{C}^{J \times LR}$ and $\mathbf{S} = [\mathbf{S}_1 \dots \mathbf{S}_R] \in \mathbb{C}^{K \times LR}$. According to Section III, decomposition (13) is essentially unique if

$$JK \geq R \quad \text{and} \quad k(\mathbf{A}) + k'(\mathbf{H}) + k'(\mathbf{S}) \geq 2(R+1). \quad (14)$$

The first inequality is always satisfied in practice (recall that K is lower-bounded by the number of transmitted symbols). Because of user-independent fading and multipath we may in practice also assume that $\mathbf{A}$ is full k-rank and $\mathbf{H}$ full k'-rank. By persistence of excitation of the transmitted signals we may further assume that $\mathbf{S}$ is also full k'-rank. (This becomes increasingly likely as K increases.) Hence, in practice (14) amounts to

$$\min(I,R)+\min\left(\left\lfloor\frac{J}{L}\right\rfloor,R\right)+\min\left(\left\lfloor\frac{K}{L}\right\rfloor,R\right)\geq 2(R+1). \quad (15)$$

This equation should be seen as a bound on the number of users that can simultaneously be processed. The condition is sufficient but not always necessary. (In [22] uniqueness is demonstrated for scenarios that do not satisfy condition (15).)

The structure of decomposition (13) allows for fewer indeterminacies than the general decomposition in rank-$(L,L,1)$ terms. According to the general theory of Section III, a term of the form $A_r \circ (\mathbf{H}_r\cdot\mathbf{S}_r^T)$ remains unchanged when i) A_r is multiplied with a scalar, provided $\mathbf{H}_r\cdot\mathbf{S}_r^T$ is multiplied with the inverse scalar, and ii) $\mathbf{H}_r$ is multiplied from the right with a nonsingular matrix $\mathbf{X}$, provided $\mathbf{S}_r^T$ is multiplied from the left with $\mathbf{X}^{-1}$. In our application however, the latter multiplication with $\mathbf{X}^{-1}$ would destroy the Toeplitz structure of $\mathbf{S}_r^T$.[1] Hence, we have that model (13) is unique up to i) the order of the terms and ii) a rescaling of the factors $A_r, \mathbf{H}_r$ and $\mathbf{S}_r$ in each term, provided the scaling factors compensate each other. These are the same indeterminacies as for the ordinary CP model. We conclude that the symbols transmitted by the different users may be found up to a scaling factor from the computation of decomposition (13).

C. Algorithm

For the computation of the components in decomposition (13) we follow an ALS approach. Note that the structure of (13) is such that, after fixing two of the sets $\{A_r\},\{\mathbf{H}_r\},\{\mathbf{S}_r\}$, a conditional update of the third set is a classical linear least-squares problem, like in the case of the ordinary CP model. In our algorithm we will take the block-Toeplitz structure of matrix $\mathbf{S}$ into account.

We will now derive explicit expressions for the conditional updates. Consider the noisy version of (13)

$$\mathcal{Y}=\sum_{r=1}^{R}A_r\circ\left(\mathbf{H}_r\cdot\mathbf{S}_r^T\right)+\mathcal{N} \quad (16)$$

in which $\mathcal{N}$ is a noise term.

First, let us consider the conditional update of $\mathbf{A}$, given $\mathbf{H}$ and $\mathbf{S}$. By "slicing" $\mathcal{Y}$ along the dimension corresponding to i (see Fig. 2), we obtain

$$\mathbf{Y}_{i,:,:}=\sum_{r=1}^{R}a_{ir}\mathbf{H}_r\cdot\mathbf{S}_r^T+\mathbf{N}_{i,:,:} \quad (17)$$

[1]The leftmost part of $\mathbf{X}^{-1}\cdot\mathbf{S}_r^T$ can only be upper triangular if $\mathbf{X}^{-1}$ is upper triangular. The rightmost part of $\mathbf{X}^{-1}\cdot\mathbf{S}_r^T$ can only be lower triangular if $\mathbf{X}^{-1}$ is lower triangular. Hence, $\mathbf{X}^{-1}$ is diagonal. Because the entries of $\mathbf{X}^{-1}\cdot\mathbf{S}_r^T$ have to be constant along the diagonals, $\mathbf{X}^{-1}$ is up to a scaling factor equal to the identity matrix.

in which $1\leq i\leq I$. Equation (17) is equivalent to

$$\mathrm{vec}(\mathbf{Y}_{i,:,:})=\left[\mathrm{vec}\left(\mathbf{H}_1\cdot\mathbf{S}_1^T\right)\ldots\mathrm{vec}\left(\mathbf{H}_R\cdot\mathbf{S}_R^T\right)\right]\times\mathrm{vec}(A_{i,:})+\mathrm{vec}(\mathbf{N}_{i,:,:}).$$

This will be written as

$$Y_{\underline{1},i}=\mathbf{M}(\mathbf{H},\mathbf{S})\cdot(\mathbf{A}^T)_{:,i}+N_{\underline{1},i},\quad 1\leq i\leq I. \quad (18)$$

This equation is used for a conditional update of $\mathbf{A}$.

Next, let us consider the conditional update of $\mathbf{H}$, given $\mathbf{A}$ and $\mathbf{S}$. By slicing $\mathcal{Y}$ along the dimension corresponding to j (see Fig. 2), we obtain

$$\begin{aligned}&[\,y_{i,j,1}\ \cdots\ y_{i,j,K}\,]\\&=\sum_{r=1}^{R}a_{ir}(\mathbf{H}_r)_{j,:}\cdot\mathbf{S}_r^T+[n_{i,j,1}\ \cdots\ n_{i,j,K}]\\&=\mathbf{H}_{j,:}\cdot\begin{bmatrix}a_{i1}\mathbf{S}_1\\ \vdots\\ a_{iR}\mathbf{S}_R\end{bmatrix}+[n_{i,j,1}\ \cdots\ n_{i,j,K}],\qquad 1\leq i\leq I.\end{aligned}$$

Stacking these equations for all values of i, we obtain

$$\begin{aligned}&[\,y_{1,j,1}\ \cdots\ y_{1,j,K}\ \cdots\ y_{I,j,1}\ \cdots\ y_{I,j,K}\,]^T\\&=\begin{bmatrix}a_{11}\mathbf{S}_1&\cdots&a_{1R}\mathbf{S}_R\\ \vdots&&\vdots\\ a_{I1}\mathbf{S}_1&\cdots&a_{IR}\mathbf{S}_R\end{bmatrix}(\mathbf{H}^T)_{:,j}\\&\quad+[\,n_{1,j,1}\ \cdots\ n_{1,j,K}\ \cdots\ n_{I,j,1}\ \cdots\ n_{I,j,K}\,]^T.\end{aligned}$$

This equation will be written as

$$Y_{\underline{2},j}=\mathbf{M}(\mathbf{A},\mathbf{S})\cdot(\mathbf{H}^T)_{:,j}+N_{\underline{2},j},\quad 1\leq j\leq J. \quad (19)$$

Finally, let us consider the conditional update of $\mathbf{S}$, given $\mathbf{A}$ and $\mathbf{H}$. Define the matrix $\tilde{\mathbf{S}}\in\mathbb{C}^{\tilde{K}\times R}$ that contains the symbols transmitted by the different users. Also define $\mathcal{T}(\mathbf{H}_r)$ as a $(JK\times\tilde{K})$ block Toeplitz matrix containing $(J\times 1)$ blocks. The first column of $\mathcal{T}(\mathbf{H}_r)$ is equal to $\mathrm{vec}(\mathbf{H}_r)$, followed by zeros and its first row consists of $(\mathbf{H}_r)_{11}$, followed by zeros. We have that $\mathrm{vec}(\mathbf{H}_r\cdot\mathbf{S}_r^T)=\mathcal{T}(\mathbf{H}_r)\cdot\mathrm{vec}(\tilde{S}_{:,r})$. Equation (17) can now be written as

$$\mathrm{vec}(\mathbf{Y}_{i,:,:})=\sum_{r=1}^{R}a_{ir}\mathcal{T}(\mathbf{H}_r)\mathrm{vec}(\tilde{S}_{:,r})+\mathrm{vec}(\mathbf{N}_{i,:,:}). \quad (20)$$

Stacking these equations for all values of i, we obtain

$$\begin{aligned}&[y_{1,1,1}\cdots y_{1,J,K}\cdots y_{I,1,1}\cdots y_{I,J,K}]^T\\&=\begin{bmatrix}a_{11}\mathcal{T}(\mathbf{H}_1)&\cdots&a_{1R}\mathcal{T}(\mathbf{H}_R)\\ \vdots&&\vdots\\ a_{I1}\mathcal{T}(\mathbf{H}_1)&\cdots&a_{IR}\mathcal{T}(\mathbf{H}_R)\end{bmatrix}\mathrm{vec}(\tilde{\mathbf{S}})\\&\quad+[n_{1,1,1}\cdots n_{1,J,K}\cdots n_{I,1,1}\cdots n_{I,J,K}]^T\end{aligned}$$

which is written as

$$Y=\mathbf{M}(\mathbf{A},\mathbf{H})\cdot S+N. \quad (21)$$

TABLE I
ALS ALGORITHM FOR THE COMPUTATION OF DECOMPOSITION (13)

Initialization: randomly initialize two of the three matrices, e.g., $\mathbf{A}$ and $\mathbf{H}$; if $R \leqslant \min(K, J)/L$, initial values may be obtained via an EVD.

l-th step:

1. Update the symbol estimates:
$$S^{(l)} = \mathbf{M}(\mathbf{A}^{(l-1)}, \mathbf{H}^{(l-1)})^{\dagger} \cdot Y. \quad (25)$$
Normalize the estimates of the sequences transmitted by the different users w.r.t. the scaling ambiguity:
$$(\tilde{\mathbf{S}})^{(l)}_{:,r} \leftarrow (\tilde{\mathbf{S}})^{(l)}_{:,r} / \|(\tilde{\mathbf{S}})^{(l)}_{:,r}\|, \qquad 1 \leqslant r \leqslant R.$$
2. Update the estimate of $\mathbf{H}$:
$$(\mathbf{H}^{(l)^T})_{:,j} = \mathbf{M}(\mathbf{A}^{(l-1)}, \mathbf{S}^{(l)})^{\dagger} \cdot Y_{\underline{2},j} \qquad 1 \leqslant j \leqslant J. \quad (26)$$
3. Update the estimate of $\mathbf{A}$:
$$(\mathbf{A}^{(l)^T})_{:,i} = \mathbf{M}(\mathbf{H}^{(l)}, \mathbf{S}^{(l)})^{\dagger} \cdot Y_{\underline{1},i}, \qquad 1 \leqslant i \leqslant I. \quad (27)$$
Normalize the columns of $\mathbf{A}$ w.r.t. the scaling ambiguity:
$$A^{(l)}_r \leftarrow A^{(l)}_r / \|A^{(l)}_r\|, \qquad 1 \leqslant r \leqslant R.$$

End: The iteration is terminated when $\|S^{(l)} - S^{(l-1)}\| < \epsilon$.

Since we directly update the different symbols in $\tilde{\mathbf{S}}$, the block-Toeplitz structure of $\mathbf{S}$ is preserved.

The ALS iteration is initialized with randomly chosen starting values. However, if the number of users is small, namely $R \leq \min(K, J)/L$, then the iteration can be initialized with the noise-free solution. Let $E \in \mathbb{R}^L$ be a vector of which all the entries are equal to 1. In the noise-free case, we have from (17)

$$\mathbf{Y}_{i,:,:} = \mathbf{H} \cdot \text{diag}(\text{vec}(A_{i,:}) \otimes E) \cdot \mathbf{S}^T \quad (22)$$
$$\mathbf{Y}_{i',:,:} = \mathbf{H} \cdot \text{diag}(\text{vec}(A_{i',:}) \otimes E) \cdot \mathbf{S}^T. \quad (23)$$

From these equations follows that the column spaces of $\{\mathbf{H}_r\}$ are invariant subspaces of $\mathbf{Y}_{i,:,:} \cdot \mathbf{Y}^{\dagger}_{i',:,:}$ and may hence be determined by means of an eigenvalue decomposition (EVD). The matrices $\{\mathbf{S}_r\}$ follow, up to right multiplication by nonsingular matrices $\mathbf{X}_r \in \mathbb{C}^{L \times L}$, from (22) or (23). These indeterminacies may be reduced to the inherent scaling ambiguities by imposing the Toeplitz structure of $\{\mathbf{S}_r\}$. This corresponds in fact to the identification of R FIR filters from SIMO measurements [21], [32], [42]. By substituting the results back in (22) or (23), the matrices $\{\mathbf{H}_r\}$ may be obtained up to a scaling factor as the solution of a set of linear equations. Finally, $\mathbf{A}$ may be calculated by solving (13) as a set of linear equations, given $\{\mathbf{H}_r\}$ and $\{\mathbf{S}_r\}$. This technique generalizes the procedure for the ordinary CP problem proposed in [18]. In [31] such a technique was for the first time proposed for W-CDMA with large delay spread. We refer to this paper for a detailed description of an EVD-based solution.

An outline of our algorithm is given in Table I. We will refer to this algorithm as Alg. 1. In substeps 1 and 3, respectively, the symbol sequences and the columns of $\mathbf{A}$ are normalized. This is to avoid arithmetic under- and overflow. Without the normalization, it could for instance happen that $\|\tilde{S}^{(l)}_r\|$ tends to infinity while $\|A^{(l)}_r\|$ tends to zero. The ALS algorithm monotonically decreases the cost function

$$f(\mathbf{A}, \mathbf{H}, \mathbf{S}) = \|\mathcal{Y} - \sum_{r=1}^{R} A_r \circ \left(\mathbf{H}_r \cdot \mathbf{S}^T_r\right) \|^2_F. \quad (24)$$

The algorithm converges to at least a local optimum (or, in odd cases, a saddle point) of the cost function. To increase the chance of finding the global optimum, one may run the algorithm a number of times, starting from different initial values.

Equations (25), (26), and (27) explicitly formulate the solution of overdetermined sets of linear equations. These can be solved by means of $O(IJK(\tilde{K}R)^2), O(IJK(LR)^2)$ and $O(IJKR^2)$ flops, respectively [11].

V. SIMULATIONS

In this section, we illustrate the performance of our algorithm by means of some Monte Carlo simulations. We compare against the probability of error based on the Cramér-Rao Bound CRB for the estimated symbols $s_{k,r}$. Whereas this probability of error is not, strictly speaking, a lower bound on the BER for blind detection, it provides a simple and useful benchmark. The CRB of the transmitted symbols $s_{k,r}$ is derived in Appendix. We also compare our algorithm to the nonblind least-squares (LS) receiver. In contrast to our algorithm, the LS receiver assumes perfect knowledge of channel fading coefficients, antenna gains and spreading codes. Its performance can usually not be reached, but it is often used as a benchmark for blind algorithms [28], [39]. The LS solution for the symbol estimates is

$$S_{\text{LS}} = \mathbf{M}(\mathbf{A}, \mathbf{H})^{\dagger} \cdot Y \quad (28)$$

in which perfect knowledge of $\mathbf{A}$ and $\mathbf{H}$ is assumed, as opposed to the ALS updating in (25).

For each Monte Carlo run, the channel fading coefficients, the antenna gains and the spreading sequences are redrawn from an i.i.d. complex Gaussian generator with zero mean and unit variance. The results are averaged over all users and all runs. The noise is zero-mean white (in all dimensions) Gaussian, with variance σ for all antennas. The observed tensor is given by $\mathcal{Y}_{\text{obs}} = \mathcal{Y} + \mathcal{N}$, where $\mathcal{Y}$ is the noise-free tensor that contains the data to be estimated and $\mathcal{N}$ represents the noise. The signal-to-noise ratio (SNR) at the input of the multiuser receiver is defined as

$$\text{SNR} = 10 \log_{10} \left(\frac{\|\mathcal{Y}\|^2}{\|\mathcal{N}\|^2} \right) \stackrel{\text{def}}{=} 10 \log_{10} \left(\frac{\sum_{ijk} |y_{ijk}|^2}{\sum_{ijk} |n_{ijk}|^2} \right) [\text{dB}].$$

In the first experiment, we compare our algorithm to the PARALIND-based algorithm of [31], in two scenarios where the latter can be used. The transmitted signals are of the BPSK-type, taking values in ± 1. There are two receive antennas $(I = 2)$.

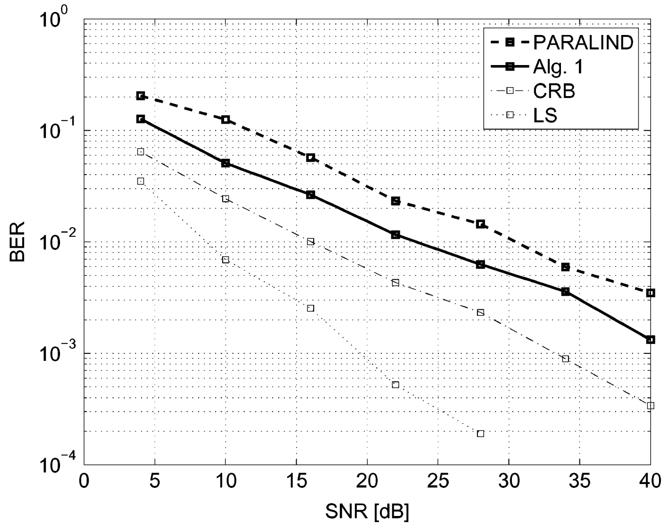

Fig. 3. BER versus SNR. The number of users $R = 4$, the spreading code length $J = 16$ and the number of antennas $I = 2$. The frame length is 50 symbols. We assume 2-taps channels for all users.

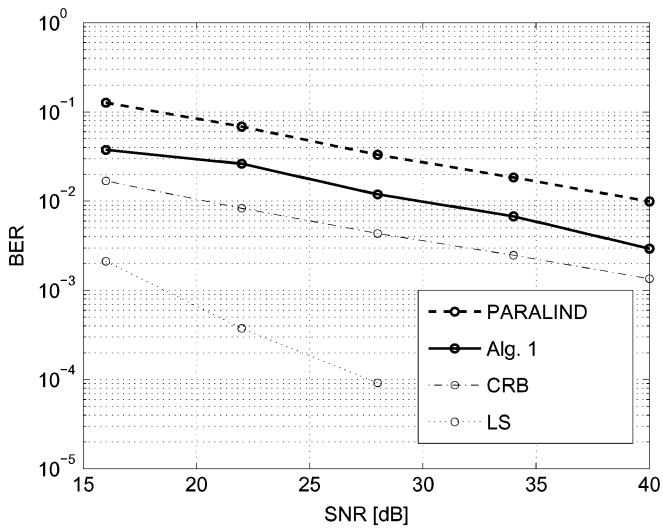

Fig. 4. BER versus SNR. The number of users $R = 6$, the spreading code length $J = 16$ and the number of antennas $I = 2$. The frame length is 50 symbols. We assume 2-taps channels for all users.

The spreading gain $J = 16$. All users transmit $\tilde{K} = 50$ symbols. The delay spread $L = 2$ symbols. The parameter ϵ in Alg. 1 was set equal to $1e-4$ and at most 1000 iterations were carried out. In each run, the algorithm started from a single random initialization. The obtained BERs are shown in Figs. 3 and 4, for $R = 4$ and $R = 6$ users, respectively. The number of Monte Carlo trials is equal to 1000 and 125 for Figs. 3 and 4, respectively.

We see that algorithm given in Table I was more accurate than the algorithm of [31]. The reason is that the Toeplitz structure of matrix $\mathbf{S}$ is exploited from the beginning of the iteration, and not just in the second stage of the algorithm as in [31]. On the other hand, the PARALIND-based algorithm is much cheaper than Alg. 1. It essentially involves the EVD of an $(RL \times RL)$ matrix. Recall that the computational complexity of Alg. 1 was discussed in Section IV-C. To save computations, Alg. 1 could be initialized with the result of the PARALIND-based algorithm.

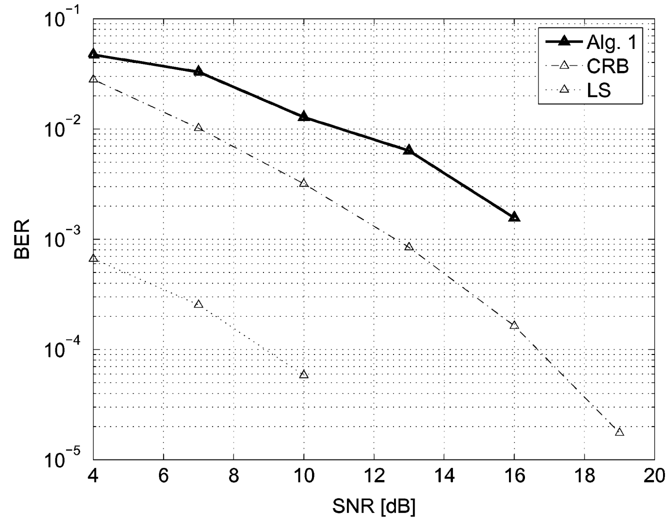

Fig. 5. BER versus SNR. The number of users R and the number of antennas I are equal to 4. The spreading gain $J = 9$. The frame length is 50 symbols and we assume 3-taps-channels for all users.

In the following experiment, we test Alg. 1 in a scenario where the conditions of [31] are not satisfied. The transmitted signals are of the QPSK-type, taking values in $\pm 1/\sqrt{2} \pm i/\sqrt{2}$. The number of users $R = 4$. The number of Monte Carlo trials is equal to 500. There are four receive antennas $(I = 4)$. The spreading gain $J = 9$. All users transmit $\tilde{K} = 50$ symbols. The delay spread $L = 3$ symbols. The parameter ϵ in Alg. 1 was set equal to $1e-6$ and at most 5000 iterations were carried out. In each run, the algorithm started from twenty random initializations. Assuming that the matrix $\mathbf{A}$ is full k-rank and that the matrices $\tilde{\mathbf{H}}$ and $\mathbf{S}$ are full k'-rank, the identifiability condition (14) is satisfied: $k(\mathbf{A}) + k'(\mathbf{H}) + k'(\mathbf{S}) = 4 + \lfloor 9/3 \rfloor + 4 = 11 \geq 2(R+1) = 10$. The obtained BER is shown in Fig. 5. For SNR ≥ 18 dB the estimation was perfect. Although this problem was difficult (the matrix $\mathbf{H}$ has more columns than rows), the BER curve is quite close to the CRB.

VI. Conclusion

We have derived a new algebraic algorithm for the blind separation-deconvolution of DS-CDMA signals received on an antenna array. The technique exploits the specific structure of the decomposition in rank-$(1, L, L)$ terms that underlies the data. For zero-mean white Gaussian noise the algorithm implements a maximum likelihood estimator. We have shown that the performance is quite close to the CRB over a broad SNR range. We have presented a bound on the number of users that guarantees unambiguous reconstruction of the CDMA sources. Current work includes relaxation of this bound.

The technique works well for small sample sizes. Neither DOA calibration information nor prior knowledge w.r.t. the multipath characteristics are required. The spreading codes need not be known and are allowed to be non-orthogonal. Besides the fact that they are of the CDMA-type, no information on the sources (such as FA, CM, statistical independence, whiteness, $\ldots$) is required. The same approach can be followed in other applications, such as the problems discussed in [27], [29], [30], and [39].

APPENDIX
CRAMÉR–RAO BOUND

In this section we calculate the CRB [34] of the transmitted symbols before detection. The derivation is similar to the ISI-free case [20]. The main difference resides in that we take into account the block-Toeplitz structure of the symbol matrix $\mathbf{S}$. We make the following standard assumptions: i) the symbols are independent and identically distributed (i.i.d.) and uncorrelated for different users. There is no correlation between the real and imaginary part and ii) the noise is zero-mean Gaussian with standard deviation equal to σ. The noise samples are i.i.d. in i, j and k.

A delicate point in the calculation of the CRB is the permutation and scale ambiguity in decomposition (13). We may get rid of the scale ambiguity by assuming that the entries on the first row of $\mathbf{A}$ and $\tilde{\mathbf{S}}$ are equal to one. Further we assume that the upper left entries of the channel matrices $\mathbf{H}_r, 1 \leq r \leq R$, are distinct and that $R-1$ of them are known. By the latter assumption, some information that is unknown to the algorithm is incorporated in the bound, which makes it somewhat harder to reach; however, this extra information allows us to resolve the permutation ambiguity. For notational convenience, we also assume that the noise variance is known, as in [20]. In this way, the number of unknown complex parameters is $R(I + JL + \tilde{K} - 3) + 1$.

Define a vector $\bar{H}_{1,:}$ of size $(R(L-1)+1)\times 1$ by stacking the unknown parts of the first rows of $\mathbf{H}_r, 1 \leq r \leq R$. Also define a vector $\bar{S}$ of size $R(\tilde{K}-1)\times 1$ resulting from the vectorization of matrix $\tilde{\mathbf{S}}$ after dropping the first row. Let $P = R(I + JL + \tilde{K} - 3) + 1$. Define the $(1 \times 2P)$ complex parameter vector

$$\Theta = (\bar{S}^T, A_{2,:}, \ldots, A_{I,:}, \bar{H}_{1,:}, H_{2,:}, \ldots, H_{J,:}, \bar{S}^H, A^*_{2,:}, \ldots, A^*_{I,:}, \bar{H}^*_{1,:}, H^*_{2,:}, \ldots, H^*_{J,:}).$$

Using (18), (19), and (21), the log-likelihood function $\mathcal{L}$ [40] can be written in three equivalent ways:

$$\mathcal{L}(\Theta) = c - \frac{1}{\sigma^2}\|Y - \mathbf{M}(\mathbf{A},\mathbf{H}) \cdot S\|^2 \tag{29}$$

$$= c - \frac{1}{\sigma^2}\sum_{i=1}^{I}\|Y_{\underline{1},i} - \mathbf{M}(\mathbf{H},\mathbf{S}) \cdot (\mathbf{A}^T)_{:,i}\|^2 \tag{30}$$

$$= c - \frac{1}{\sigma^2}\sum_{j=1}^{J}\|Y_{\underline{2},j} - \mathbf{M}(\mathbf{A},\mathbf{S}) \cdot (\mathbf{H}^T)_{:,j}\|^2 \tag{31}$$

with $c \stackrel{\text{def}}{=} -IJK\log(\pi\sigma^2)$. The Fisher information matrix $\mathbf{F}$ [34] is given by

$$\mathbf{F} = \mathbb{E}\{(\nabla_\Theta \mathcal{L}(\Theta))^H \cdot (\nabla_\Theta \mathcal{L}(\Theta))\}$$

in which $\mathbb{E}\{\cdot\}$ denotes the expectation. Because the noise is circular, $\mathbf{F}$ takes the form

$$\mathbf{F} = \begin{bmatrix} \Phi & \mathbf{0} \\ \mathbf{0} & \Phi^* \end{bmatrix}$$

in which the $(P \times P)$ Hermitian matrix Φ is given by (32), shown at the bottom of the page. Using (29)–(31), it can be shown that the elements of the upper triangular part of Φ can be computed as

$$\mathbb{E}\{\nabla_{\bar{S}^*}\mathcal{L} \cdot \nabla_{\bar{S}^T}\mathcal{L}\} = \frac{1}{\sigma^2}\mathbf{M}'(\mathbf{A},\mathbf{H})^* \cdot \mathbf{M}'(\mathbf{A},\mathbf{H})^T$$

$$\mathbb{E}\left\{\nabla_{(A^H)_{:,i}}\mathcal{L} \cdot \nabla_{A_{i',:}}\mathcal{L}\right\} = \frac{1}{\sigma^2}\mathbf{M}'(\mathbf{H},\mathbf{S})^* \cdot \mathbf{M}'(\mathbf{H},\mathbf{S})^T \cdot \delta_{ii'}, \quad 2 \leq i \leq i' \leq I$$

$$\mathbb{E}\left\{\nabla_{(\bar{H}^H)_{:,1}}\mathcal{L} \cdot \nabla_{\bar{H}_{1,:}}\mathcal{L}\right\} = \frac{1}{\sigma^2}\mathbf{M}'(\mathbf{A},\mathbf{S})^* \cdot \mathbf{M}'(\mathbf{A},\mathbf{S})^T$$

$$\mathbb{E}\left\{\nabla_{(\bar{H}^H)_{:,1}}\mathcal{L} \cdot \nabla_{H_{j,:}}\mathcal{L}\right\} = \mathbf{0}, \quad 2 \leq j \leq J$$

$$\mathbb{E}\left\{\nabla_{(H^H)_{:,j}}\mathcal{L} \cdot \nabla_{H_{j',:}}\mathcal{L}\right\} = \frac{1}{\sigma^2}\mathbf{M}(\mathbf{A},\mathbf{S})^* \cdot \mathbf{M}(\mathbf{A},\mathbf{S})^T \cdot \delta_{jj'}, \quad 2 \leq j \leq j' \leq J$$

$$\mathbb{E}\{\nabla_{\bar{S}^*}\mathcal{L} \cdot \nabla_{A_{i,:}}\mathcal{L}\} = \frac{1}{\sigma^4}\mathbf{M}'(\mathbf{A},\mathbf{H})^* \cdot \mathbb{E}\left\{N^* \cdot N^T_{\underline{1},i}\right\} \cdot \mathbf{M}'(\mathbf{H},\mathbf{S})^T, \quad 2 \leq i \leq I$$

$$\mathbb{E}\left\{\nabla_{\bar{S}^*}\mathcal{L} \cdot \nabla_{\bar{H}_{1,:}}\mathcal{L}\right\} = \frac{1}{\sigma^4}\mathbf{M}'(\mathbf{A},\mathbf{H})^* \cdot \mathbb{E}\left\{N^* \cdot N^T_{\underline{2},1}\right\} \cdot \mathbf{M}'(\mathbf{A},\mathbf{S})^T$$

$$\mathbb{E}\{\nabla_{\bar{S}^*}\mathcal{L} \cdot \nabla_{H_{j,:}}\mathcal{L}\} = \frac{1}{\sigma^4}\mathbf{M}'(\mathbf{A},\mathbf{H})^* \cdot \mathbb{E}\left\{N^* \cdot N^T_{\underline{2},j}\right\} \cdot \mathbf{M}(\mathbf{A},\mathbf{S})^T, \quad 2 \leq j \leq J$$

$$\mathbb{E}\left\{\nabla_{(A^H)_{:,i}}\mathcal{L} \cdot \nabla_{\bar{H}_{1,:}}\mathcal{L}\right\} = \frac{1}{\sigma^4}\mathbf{M}'(\mathbf{H},\mathbf{S})^* \cdot \mathbb{E}\left\{N^*_{\underline{1},i} \cdot N^T_{\underline{2},1}\right\} \cdot \mathbf{M}'(\mathbf{A},\mathbf{S})^T, \quad 2 \leq i \leq I$$

$$\Phi = \left[\begin{array}{c|ccc} \frac{\mathbf{E}\{\nabla_{\bar{S}^*}\mathcal{L}\cdot\nabla_{\bar{S}^T}\mathcal{L}\}}{\mathbf{E}\left\{\nabla_{(A^H)_{:,2}}\mathcal{L}\cdot\nabla_{\bar{S}^T}\mathcal{L}\right\}} & \frac{\mathbf{E}\{\nabla_{\bar{S}^*}\mathcal{L}\cdot\nabla_{A_{2,:}}\mathcal{L}\}}{\mathbf{E}\{\nabla_{(A^H)_{:,2}}\mathcal{L}\cdot\nabla_{A_{2,:}}\mathcal{L}\}} & \frac{\cdots}{\cdots} & \frac{\mathbf{E}\{\nabla_{\bar{S}^*}\mathcal{L}\cdot\nabla_{H_{J,:}}\mathcal{L}\}}{\mathbf{E}\{\nabla_{(A^H)_{:,2}}\mathcal{L}\cdot\nabla_{H_{J,:}}\mathcal{L}\}} \\ \vdots & \vdots & & \vdots \\ \mathbf{E}\{\nabla_{(H^H)_{:,J}}\mathcal{L}\cdot\nabla_{\bar{S}^T}\mathcal{L}\} & \mathbf{E}\{\nabla_{(H^H)_{:,J}}\mathcal{L}\cdot\nabla_{A_{2,:}}\mathcal{L}\} & \ldots & \mathbf{E}\{\nabla_{(H^H)_{:,J}}\mathcal{L}\cdot\nabla_{H_{J,:}}\mathcal{L}\} \end{array}\right]. \tag{32}$$

$$\mathbb{E}\left\{\nabla_{(A^H)_{:,i}}\mathcal{L}\cdot\nabla_{H_{j,:}}\mathcal{L}\right\} = \frac{1}{\sigma^4}\mathbf{M}'(\mathbf{H},\mathbf{S})^* \cdot\mathbb{E}\left\{N^*_{\underline{1},i}\cdot N^T_{\underline{2},j}\right\}\cdot\mathbf{M}(\mathbf{A},\mathbf{S})^T, \qquad 2\le i\le I, 2\le j\le J$$

where $\mathbf{M}'(\mathbf{A},\mathbf{H})$ results from $\mathbf{M}(\mathbf{A},\mathbf{H})$ by dropping the rows with row number $(r-1)\tilde{K}+1, 1\le r\le R$, and where $\mathbf{M}'(\mathbf{H},\mathbf{S})$ (respectively, $\mathbf{M}'(\mathbf{A},\mathbf{S})$) results from $\mathbf{M}(\mathbf{H},\mathbf{S})$ (respectively, $\mathbf{M}(\mathbf{A},\mathbf{S})$) by dropping the first row. The covariance matrix of N and $N_{\underline{1},i},\ 2\le i\le I$, is given by

$$\mathbb{E}\left[N^*\cdot N^T_{\underline{1},i}\right] = \sigma^2\begin{bmatrix}\mathbf{0}_{JK(i-1)\times JK}\\ \mathbf{I}_{JK\times JK}\\ \mathbf{0}_{JK(I-i)\times JK}\end{bmatrix}.$$

The covariance matrices $\mathbb{E}\left[N^*\cdot N^T_{\underline{2},j}\right], 1\le j\le J$, and $\mathbb{E}\left[N^*_{\underline{1},i}\cdot N^T_{\underline{2},j}\right], 2\le i\le I,\ 1\le j\le J$, can be obtained in a similar way.

Let the partitioning in (32) be represented by

$$\Phi = \begin{bmatrix}\Phi_{\bar{S}\bar{S}} & \Phi_2\\ \Phi_2^H & \Phi_1\end{bmatrix}.$$

The CRB on the variance of any unbiased estimator is proportional to the trace of the inverse of the Fisher Information Matrix [34]. In particular, denote the average CRB for the estimated symbols $s_{kr}, 2\le k\le K, 1\le r\le R$, over the whole frame as CRB_s. We have that the average variance of any unbiased estimator of the symbols is bounded below by

$$\mathrm{CRB}_s = \frac{1}{R(\tilde{K}-1)}\mathrm{trace}\left\{\left(\Phi_{\bar{S}\bar{S}} - \Phi_2\cdot\Phi_1^{-1}\cdot\Phi_2^H\right)^{-1}\right\}. \tag{33}$$

This result follows directly from applying to Φ the lemma of the inverse of a partitioned Hermitian matrix [43]

$$\begin{bmatrix}\Phi_{\bar{S}\bar{S}} & \Phi_2\\ \Phi_2^H & \Phi_1\end{bmatrix}^{-1} = \begin{bmatrix}(\Phi_{\bar{S}\bar{S}} - \Phi_2\Phi_1^{-1}\Phi_2^H)^{-1} & -\Phi_{\bar{S}\bar{S}}^{-1}\Phi_2\Phi_3\\ -\Phi_3\Phi_2^H\Phi_{\bar{S}\bar{S}}^{-1} & \Phi_3\end{bmatrix}$$

with $\Phi_3 = (\Phi_1 - \Phi_2^H\Phi_{\bar{S}\bar{S}}^{-1}\Phi_2)^{-1}$. We recall that the first symbol of each sequence is assumed to be known at the receiver in order to resolve the permutation ambiguity. Taking the average CRB over the other $R(\tilde{K}-1)$ symbols, i.e., the average over the upper $R(\tilde{K}-1)$ diagonal elements of matrix Φ^{-1}, CRB_s can be expressed as (33).

In practice, a lower bound on the BER is more useful than a lower bound on the variance of the estimated symbols. The derivation of the CRB of the symbols after detection is involved because of the nonlinearity of the detection operator. Instead, we propose a simple benchmark based on CRB (33). Assuming that the estimation errors of the symbols can be modeled as a zero-mean Gaussian random variable with variance greater than or equal to CRB_s, the corresponding probability of error after detection for binary and quaternary signaling can be expressed as ([24, p. 268, (5.2–57)] for binary signaling and [24, p. 269, (5.2–59)] and [24, p. 271, (5.2–62)] for quaternary signaling):

$$P_e^{\mathrm{CRB}} = \frac{1}{2}\mathrm{erfc}\left(\frac{1}{\sqrt{2\cdot\mathrm{CRB}_s}}\right) \text{ for binary signaling}$$
$$\approx \frac{1}{2}\mathrm{erfc}\left(\frac{1}{\sqrt{2\cdot\mathrm{CRB}_s}}\right) \times\left[1-\frac{1}{4}\mathrm{erfc}\left(\frac{1}{\sqrt{2\cdot\mathrm{CRB}_s}}\right)\right] \text{ for quaternary signaling}$$

where erfc is the complementary error function [24, p. 38, (2.1–94)]. Whereas P_e^{CRB} is, strictly speaking, not a lower bound on the probability of error for blind detection, the simulation results in Section V validate it as a useful benchmark.

References

[1] J. Benesty and P. Duhamel, "A fast constant modulus adaptive algorithm," in *Proc. Inst. Elect. Eng. F.*, Apr. 1991, vol. 138, pp. 379–387.

[2] R. Bro, "Multi-way analysis in the food industry. Models, algorithms and applications," Ph.D. dissertation, University of Amsterdam, Amsterdam, The Netherlands, 1998.

[3] R. Bro and H. A. L. Kiers, "A new efficient method for determining the number of components in PARAFAC models," *J. Chemometrics*, vol. 17, pp. 274–286, 2003.

[4] R. Bro, R. A. Harshman, and N. D. Sidiropoulos, "Modeling multi-way data with linearly dependent loadings," Royal Veterinary & Agricultural Univ., Frederiksberg, Denmark, Tech. Rep. 2005-176, 2005.

[5] J. Carroll and J. Chang, "Analysis of individual differences in multidimensional scaling via an N-way generalization of 'Eckart-Young' decomposition," *Psychometrika*, vol. 9, pp. 267–283, 1970.

[6] L. De Lathauwer, B. De Moor, and J. Vandewalle, "Computation of the canonical decomposition by means of simultaneous generalized Schur decomposition," *SIAM J. Matrix Anal. Appl.*, vol. 26, pp. 295–327, 2004.

[7] L. De Lathauwer, "A link between the canonical decomposition in multilinear algebra and simultaneous matrix diagonalization," *SIAM J. Matrix Anal. Appl.*, vol. 28, no. 3, pp. 642–666, 2006.

[8] L. De Lathauwer, "Decompositions of a higher-order tensor in block terms—Part I: Lemmas for partitioned matrices," *SIAM J. Matrix Anal. Appl.*, submitted for publication.

[9] L. De Lathauwer, "Decompositions of a higher-order tensor in block terms—Part II: Definitions and uniqueness," *SIAM J. Matrix Anal. Appl.*, submitted for publication.

[10] L. De Lathauwer and D. Nion, "Decompositions of a higher-order tensor in block terms—Part III: Alternating least squares algorithms," *SIAM J. Matrix Anal. Appl.*, submitted for publication.

[11] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed. Baltimore, MD: Johns Hopkins Univ. Press, 1996.

[12] R. A. Harshman, "Foundations of the PARAFAC procedure: Model and conditions for an "Explanatory" Multi-mode factor analysis," *UCLA Working Papers in Phonetics*, vol. 16, pp. 1–84, 1970.

[13] R. A. Harshman and M. E. Lundy, "The PARAFAC model for three-way factor analysis and multidimensional scaling," in *Research Methods for Multimode Data Analysis*, H. G. Law, C. W. Snyder, J. A. Hattie, and R. P. McDonald, Eds. New York: Praeger, 1984, pp. 122–215.

[14] F. L. Hitchcock, "The expression of a tensor or a polyadic as a sum of products," *J. Math. Phys.*, vol. 6, no. 1, pp. 164–189, 1927.

[15] F. L. Hitchcock, "Multiple invariants and generalized rank of a p-way matrix or tensor," *J. Math. Phys.*, vol. 7, no. 1, pp. 39–79, 1927.

[16] T. Jiang and N. D. Sidiropoulos, "Kruskal's permutation lemma and the identification of CANDECOMP/PARAFAC and bilinear models with constant modulus constraints," *IEEE Trans. Signal Process.*, vol. 52, no. 9, pp. 2625–2636, Sep. 2004.

[17] J. B. Kruskal, "Three-way arrays: Rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics," *Lin. Alg. Appl.*, vol. 18, pp. 95–138, 1977.

[18] S. E. Leurgans, R. T. Ross, and R. B. Abel, "A decomposition for three-way arrays," *SIAM J. Matrix Anal. Appl.*, vol. 14, pp. 1064–1083, 1993.

[19] H. Liu and G. Xu, "A deterministic approach to blind symbol estimation," *IEEE Signal Process. Lett.*, vol. 1, no. 12, pp. 205–207, Dec. 1994.

[20] X. Liu and N. Sidiropoulos, “Cramér-Rao lower bounds for low-rank decomposition of multidimensional arrays,” *IEEE Trans. Signal Process.*, vol. 49, no. 9, pp. 2074–2086, Sep. 2001.
[21] E. Moulines, P. Duhamel, J. F. Cardoso, and S. Mayrargue, “Subspace methods for the blind identification of multichannel FIR filters,” *IEEE Trans. Signal Process.*, vol. 43, no. 2, pp. 516–525, Feb. 1995.
[22] D. Nion and L. De Lathauwer, “A tensor-based blind DS-CDMA receiver using simultaneous matrix diagonalization,” in *Proc. IEEE Workshop Signal Processing Advances in Wireless Communications (SPAWC)*, Helsinki, Finland, Jun. 17–20, 2007.
[23] P. Paatero, “The multilinear engine—A table-driven, least squares program for solving multilinear problems, including the N-way parallel factor analysis model,” *J. Comput. Graphical Stat.*, vol. 8, pp. 854–888, 1999.
[24] J. G. Proakis, *Digital Communications*, 4th ed. New York: McGraw-Hill, 2001.
[25] E. Sanchez and B. R. Kowalski, “Tensorial resolution: A direct trilinear decomposition,” *J. Chemometrics*, vol. 4, pp. 29–45, 1990.
[26] A. Scaglione and G. Giannakis, “Design of user codes in QS-CDMA systems for MUI elimination in unknown multipath,” *IEEE Commun. Lett.*, vol. 3, pp. 25–27, Feb. 1999.
[27] A. Scaglione, G. Giannakis, and S. Barbarossa, “Redundant filterbank precoders and equalizers—Part I: Unification and optimal designs,” *IEEE Trans. Signal Processing*, vol. 47, no. 7, pp. 1988–2006, Jul. 1999.
[28] N. Sidiropoulos, G. Giannakis, and R. Bro, “Blind PARAFAC receivers for DS-CDMA systems,” *IEEE Trans. Signal Process.*, vol. 48, no. 3, pp. 810–823, Mar. 2000.
[29] N. Sidiropoulos, R. Bro, and G. Giannakis, “Parallel factor analysis in sensor array processing,” *IEEE Trans. Signal Process.*, vol. 48, no. 8, pp. 2377–2388, Aug. 2000.
[30] N. Sidiropoulos and X. Liu, “Identifiability results for blind beamforming in incoherent multipath with small delay spread,” *IEEE Trans. Signal Process.*, vol. 49, no. 1, pp. 228–236, Jan. 2001.
[31] N. D. Sidiropoulos and G. Z. Dimić, “Blind multiuser detection in W-CDMA systems with large delay spread,” *IEEE Signal Process. Letters*, vol. 8, no. 3, pp. 87–89, Mar. 2001.
[32] D. T. M. Slock and C. B. Papadias, “Further results on blind identification and equalization of multiple FIR channels,” in *Proc. Int. Conf. Acoustics, Speech, Signal Processing*, May 1995, vol. 3, pp. 1964–1967.
[33] A. Smilde, R. Bro, and P. Geladi, *Multi-way Analysis. Applications in the Chemical Sciences*. Chichester, U.K.: Wiley, 2004.
[34] H. W. Sorenson, *Parameter Estimation*. New York: Marcel Dekker, 1980.
[35] A. Stegeman and N. D. Sidiropoulos, “On Kruskal’s uniqueness condition for the CANDECOMP/PARAFAC decomposition,” *Lin. Alg. Appl.*, vol. 420, pp. 540–552, 2007.
[36] G. Tomasi and R. Bro, “A comparison of algorithms for fitting the PARAFAC model,” *Comp. Stat. Data Anal.*, vol. 50, pp. 1700–1734, 2006.
[37] A. Touzni, I. Fijalkow, M. G. Larimore, and J. R. Treichler, “A globally convergent approach for blind MIMO adaptive deconvolution,” *IEEE Trans. Signal Process.*, vol. 49, no. 6, pp. 1166–1178, Jun. 2001.
[38] J. K. Tugnait, “Approaches of FIR system identification with noisy data using higher order statistics,” *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 7, pp. 1307–1317, Jul. 1990.
[39] A.-J. van der Veen, “Algebraic methods for deterministic blind beamforming,” *Proc. IEEE*, vol. 86, no. 10, pp. 1987–2008, Oct. 1998.
[40] H. L. Van Trees, *Detection Estimation and Modulation Theory*. New York: Wiley, 1971.
[41] S. A. Vorobyov, Y. Rong, N. D. Sidiropoulos, and A. B. Gershman, “Robust iterative fitting of multilinear models,” *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 2678–2689, Aug. 2005.
[42] G. Xu, H. Liu, L. Tong, and T. Kailath, “A least-squares approch to blind channel identification,” *IEEE Trans. Signal Process.*, vol. 43, no. 12, pp. 2982–2993, Dec. 1995.
[43] F. Zhang, *Matrix Theory: Basic Results and Techniques*. New York: Springer, 1999.

Lieven De Lathauwer (M’04–SM’06) was born in Aalst, Belgium, on November 10, 1969. He received the Master’s degree in electromechanical engineering and the Ph.D. degree in applied sciences from the Katholieke Universiteit Leuven (K.U.Leuven), Leuven, Belgium, in 1992 and 1997, respectively. His Ph.D. dissertation concerned signal processing based on multilinear algebra.

From 2000 to 2007, he was with the Centre National de la Recherche Scientifique (C.N.R.S.), Cergy-Pontoise, France. He is currently with the K.U. Leuven. His research interests include linear and multilinear algebra, statistical signal and array processing, higher order statistics, independent component analysis, identification, blind identification, and equalization.

Dr. De Lathauwer is an Associate Editor of the *SIAM Journal on Matrix Analysis and Applications*.

Alexandre de Baynast (M’04) received the Diploma degree in electrical and computer engineering from ESME Sudria, Paris, France, in 1998, and the M.S. and Ph.D. degrees in electrical engineering from the University of Cergy-Pontoise, France, in 1999 and 2002, respectively.

From 2002 to 2006, he was a Postdoctoral Fellow with the Department of Electrical and Computer Engineering, Rice University, Houston, TX. In fall 2006, he joined the Department of Wireless Networks, RWTH Aachen University, Germany, as an Assistant Researcher. His research interests span the broad area of wireless communication and networking with special emphasis on signal processing for communication and architecture design for coding theory applications.