HIGH-DIMENSIONAL, LOW-RANK TENSOR APPROXIMATION: CRAMÉR-RAO LOWER BOUNDS AND APPLICATION TO MIMO CHANNELS

Clémence Prévost, Pierre Chainais

HIGH-DIMENSIONAL, LOW-RANK TENSOR APPROXIMATION: CRAMÉR-RAO LOWER BOUNDS AND APPLICATION TO MIMO CHANNELS*

C. PRÉVOST, P. CHAINAIS†

Abstract. Tensor factorization has been steadily used in the past decade to represent high-dimensional data. In particular, the canonical polyadic (CP) decomposition (CPD) is very appreciated for its modeling power and remarkable uniqueness properties. However, computing the CPD is challenging when the order of the tensor becomes high: numerical issues and high needs for storage and processing can lead the standard algorithms to diverge. To circumvent this limitation, the equivalence between the CPD and the Tensor Train Decomposition (TTD) is exploited. This approach is implemented in a new algorithm called Dimensionality Reduction, joint Estimation of the Ambiguity Matrices and the CP FACtors (DREAMFAC). A global coupled optimization scheme is proposed to break the curse of dimensionality and estimate the CP factors. DREAMFAC performs better than state-of-the-art methods. It avoids the usual propagation of the estimation error in the factors of the TTD. In particular, DREAMFAC reaches the Cramér-Rao lower bounds associated with the considered coupled CP-TT model, which is not the case for the state-of-the-art sequential procedure. Performances are illustrated on the problem of estimating the channels in a dual-polarized MIMO system. Numerical experiments show the competitive performance of the proposed method for recovery of the CP factors and estimation of the channel parameters, even with very low SNR.

Key words. Canonical polyadic decomposition, tensor train decomposition, coupled optimization, factor retrieval, MIMO systems

MSC codes. 15A23, 15A69

1. Introduction.

Tensors provide a faithful representation of higher-order observations by preserving their multidimensional structure [7, 18]. For instance, color images can be seen as data cubes with two spatial dimensions (the pixels) and a spectral dimension (the colours) encoding the red, green and blue channels. Tensor factorization has been steadily used in the past decade to model such data, due to its capabilities to capture the interactions between a set of latent factors. It has been successfully applied to various problems in, e.g., signal processing and machine learning [34], brain signal processing [1], video completion [41] or telecommunications [29]. The notion of low rank in tensors is not unified: various tensor decompositions are available, all carrying different properties and rank definitions. Perhaps the most natural generalization of the concept of matrix rank to tensors is the canonical polyadic (CP) decomposition (CPD) [10]. The CP rank of a tensor is defined as the minimal number K of rank-one tensors that, when linearly combined, lead to a perfect recovery of that tensor. A N-th order rank-one tensor is given by the outer product of N vectors. The CP decomposition has gained a lot of interest for data processing and analysis, due to its remarkable uniqueness properties under mild conditions [38]. However, the set of low-rank tensors is not closed, therefore computing the CPD is an ill-posed problem [20]. In practice, this task is usually carried out by suboptimal iterative algorithms. Furthermore, the storage and processing cost of tensors increase exponentially with their order N. This limitation is known as the curse of the dimensionality [28]. The Tensor Train (TT) decomposition (TTD) has been recently introduced [27]. This decomposition has two advantages. First, it exploits a stable numerical estimation procedure which avoids the iterative algorithms used to com-

*Submitted to the editors November 22, 2023.

Funding: This work was partly supported by the ANR project “Chaire IA Sherlock” ANR-20-CHIA-0031-01 hold by P. Chainais, as well as by the national support within the *programme d’investissements d’avenir* ANR-16-IDEX-0004 ULNE and Région HDF.

†Univ. Lille, CNRS, Centrale Lille, UMR 9189 CRIStAL, F-59000 Lille, France (clemence.prevost[at]univ-lille.fr, pierre.chainais[at]centralelille.fr).

pute the CPD. Second, it breaks the curse of dimensionality by operating on a set of matrices and third-order tensors, called TT-cores. In particular, its storage cost is linear in N. Numerous works were devoted to the task of dimensionality reduction in the CP model. In [43], the authors provided an equivalence between the CPD and the TTD through a set of non-singular change-of-basis matrices shared among the TT-cores. An algorithm, called Joint dImensionality Reduction And Factor Estimation (JIRAFE), was proposed. It recovered the factors of the CPD underlying a high-order tensor from its estimated TT-cores. A JIRAFE-like procedure was also recently extended to constrained CPD in [13]. These algorithms were based on a series of local and sequential optimization problems to obtain the CP factors at a low computational cost. The change-of-basis matrices appearing in the TTD were estimated once and propagated through the other optimization problems, without enforcing coherence between the TT-cores. For this reason, JIRAFE may lead to a sub-optimal estimation of the CP factors estimated in last position of the tensor train.

This work introduces a new algorithm for joint dimensionality reduction and estimation of the CP factors. A global optimization strategy is proposed. In contrast with JIRAFE, it ensures the coherence between TT-cores by considering coupled updates for the change-of-basis matrices. This approach is implemented in a new algorithm called Dimensionality Reduction, joint Estimation of the Ambiguity Matrices and the CP FACtors (DREAMFAC). Cramér-Rao Bound (CRB) [8, 12, 30] are used to evaluate and compare the relative performance of JIRAFE and DREAMFAC. Cramér-Rao bounds for tensor CP models have been extensively studied in the literature, including performance bounds for uncoupled CP models [4, 21, 32], Bayesian frameworks [6], and constrained models [31]. To take into account the coupling between the TT-cores, a coupled model that enforces constraints on the change-of-basis matrices is introduced. For such models the constrained Cramér-Rao bound (CCRB) can be used, whose versatility was shown by numerous works [22–24, 39, 40]. An application to the problem of harmonic retrieval in MIMO systems [2] permits to illustrate the behaviour of DREAMFAC on a realistic low-rank decomposition problem. Indeed, for MIMO channels modeling and estimation, it is important to accurately estimate channels parameters at the base station (angles of arrival, angles of departure, path gains, and polarization parameters) to perform beamforming and deal with multi-user interferences. In [29], a tensor-based approach for dual-polarized MIMO channel estimation was proposed. The MIMO channel was recast as a fourth-order tensor admitting a CPD. The authors proposed an Alternating Least Squares (ALS) to estimate CP factors. In [42], the authors adapted the JIRAFE procedure to treat the MIMO channel estimation problem. Therefore this problem can be used as a relevant benchmark.

The paper is organized as follows. Section 2 introduces important background on tensors and their low-rank decompositions. Section 3 describes the problem at hand, the JIRAFE procedure and its limitations. The proposed approach DREAMFAC is also detailed. Section 4 contains the derivation of the Cramér-Rao bounds for the proposed coupled model. It includes the detailed probabilistic framework, and closed-form expressions for the matrices to invert. Section 5 gathers an extensive set of numerical experiments to illustrate the performance of DREAMFAC with respect toe state of the art. They include simulations highlighting the robustness of DREAMFAC compared to JIRAFE, comparison of the algorithms' performance to the CRB, and results on the selected application to the estimation of MIMO channels.

Notation. The following notations [7, 18] are used: lower (a) or uppercase (A) plain font for scalars, boldface lowercase ($\boldsymbol{a}$) for vectors, boldface uppercase ($\mathbf{A}$) for matrices and calligraphic ($\boldsymbol{\mathcal{A}}$) for tensors. The elements of vectors, matrices, and tensors are denoted by a_d, A_{d_1,d_2} and $\mathcal{A}_{d_1,\ldots,D_N}$, respectively. The transpose of a matrix $\mathbf{A}$ is denoted by $\mathbf{A}^\mathsf{T}$. The matrix $\mathbf{I}_N$ is the $N \times N$ identity matrix and $\mathbf{0}_{L\times K}$ is the $L \times K$ matrix of zeros. The symbols $\boxtimes$, $\odot$ and $\otimes$ denote the Kronecker, Khatri-Rao and outer products. The operator vec$\{\cdot\}$ stands for the standard column-major vectorization of a matrix or a tensor. The operation Diag$\{\mathbf{A}, \mathbf{B}\}$ produces a block-diagonal matrix whose blocks are $\mathbf{A}$ and $\mathbf{B}$.

2. Background on low-rank tensor models.

2.1. Preliminaries on tensors. A tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{D_1\times\ldots\times D_N}$ is an N-dimensional array indexed by the elements $\mathcal{X}_{d_1,\ldots,d_N}$, for $d_n \in \{1,\ldots,D_n\}$ ($n \in \{1,\ldots,N\}$). Each dimension of a tensor is called a mode. A mode-p fiber of $\boldsymbol{\mathcal{X}}$ is a vector obtained by fixing all but the p-th dimension.

Definition 2.1 (Tensor unfoldings). *The mode-p unfolding of a tensor $\boldsymbol{\mathcal{X}}$, denoted by $\mathbf{X}^{(p)}$, is the matrix whose rows are the p-mode fibers of $\boldsymbol{\mathcal{X}}$, ordered according to the vectorization order. For a tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{D_1\times\ldots\times D_N}$, e.g., $\mathbf{X}^{(1)} \in \mathbb{R}^{D_N\ldots D_2\times D_1}$.*

Definition 2.2 (Matrix mode product). *The matrix p-mode product between a tensor $\boldsymbol{\mathcal{X}}$ and a matrix $\mathbf{M}$ is denoted by $\boldsymbol{\mathcal{X}} \bullet_p \mathbf{M}$ and is constructed such that each mode-p fiber of $\boldsymbol{\mathcal{X}}$ is multiplied by $\mathbf{M}$, e.g., the elements of the mode-1 product between $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{D_1\times\ldots\times D_N}$ and $\mathbf{M} \in \mathbb{R}^{L\times D_1}$ are accessed as*

$$\left(\boldsymbol{\mathcal{X}} \underset{1}{\bullet} \mathbf{M}\right)_{\ell,d_2,\ldots,d_N} = \sum_{i=1}^{D_1} \mathcal{X}_{i,d_2,\ldots,d_N} \mathbf{M}_{\ell,i}, \;\; \ell \in \{1,\ldots,L\}. \tag{2.1}$$

Moreover, it holds that $\boldsymbol{\mathcal{Y}} = \boldsymbol{\mathcal{X}} \bullet_p \mathbf{M} \Leftrightarrow \mathbf{Y}^{(p)} = \mathbf{X}^{(p)}\mathbf{M}^\mathsf{T}$.

Definition 2.3 (Tensor contraction product). *The contraction product on modes p, q between two tensors $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{D_1\times\ldots\times D_N}$ and $\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{J_1\times\ldots\times J_M}$ with $D_p = J_q$ is denoted by $\boldsymbol{\mathcal{X}} \bullet_p^q \boldsymbol{\mathcal{Y}}$. It produces a tensor of order $N + M - 2$ such that*

$$\left(\boldsymbol{\mathcal{X}} \overset{q}{\underset{p}{\bullet}} \boldsymbol{\mathcal{Y}}\right)_{d_1,\ldots,d_{p-1},d_{p+1},\ldots,d_N,j_1,\ldots,j_{q-1},j_{q+1},\ldots,j_M} = \tag{2.2}$$

$$\sum_{\ell=1}^{D_p=J_q} (\boldsymbol{\mathcal{X}})_{d_1,\ldots,d_{p-1},\ell,d_{p+1},\ldots,d_N} (\boldsymbol{\mathcal{Y}})_{j_1,\ldots,j_{q-1},\ell,j_{q+1},\ldots,j_M}. \tag{2.3}$$

Definition 2.4 (Outer product). *The outer product between N vectors $\boldsymbol{a}_n \in \mathbb{R}^{D_n}$ ($n \in \{1,\ldots,N\}$) is a rank-one tensor $\boldsymbol{\mathcal{X}} = \boldsymbol{a}_1 \otimes \ldots \otimes \boldsymbol{a}_N \in \mathbb{R}^{D_1\times\ldots\times D_N}$ whose elements are accessed as $\mathcal{X}_{d_1,\ldots,d_N} = (\boldsymbol{a}_1)_{d_1} \ldots (\boldsymbol{a}_N)_{d_N}$.*

2.2. The canonical polyadic decomposition. For all N-dimensional tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{D_1\times\ldots\times D_N}$, there exists an integer K such that it admits a canonical polyadic (CP) decomposition (CPD) as

$$\boldsymbol{\mathcal{X}} = [\![\mathbf{A}_1,\ldots,\mathbf{A}_N]\!], \tag{2.4}$$

where $\mathbf{A}_n \in \mathbb{R}^{D_n\times K}$ ($n \in \{1,\ldots,N\}$) are called the CP factors. When minimal, the integer K is the rank of the tensor $\boldsymbol{\mathcal{X}}$. Then each entry of $\boldsymbol{\mathcal{X}}$ can be expressed as

$$\mathcal{X}_{d_1,\ldots,d_N} = \sum_{k=1}^{K} (A_1)_{d_1,k} \ldots (A_N)_{d_N,k}. \tag{2.5}$$

The CP factors are essentially unique up to scaling and permutation ambiguities, if the rank K is not too large [7, 18]. The permutation ambiguity means that the columns of the latent CP factors can be reordered arbitrarily by any permutation matrix $\mathbf{\Pi} \in \mathbb{R}^{K \times K}$ as

$$\boldsymbol{\mathcal{X}} = [\![\mathbf{A}_1, \dots, \mathbf{A}_N]\!] = [\![\mathbf{A}_1\mathbf{\Pi}, \dots, \mathbf{A}_N\mathbf{\Pi}]\!]. \tag{2.6}$$

The scaling ambiguity means that the individual factors can be scaled as

$$\mathcal{X}_{d_1,\dots,d_N} = \sum_{k=1}^{K}\sum_{r=1}^{K} (\lambda_{1,k}(A_1)_{d_1,k}) \dots (\lambda_{N,k}(A_N)_{d_N,k}). \tag{2.7}$$

where $\lambda_{1,k} \dots \lambda_{N,k} = 1$, for all $k \in \{1, \dots, K\}$.

2.3. The tensor train decomposition (TTD). The TTD [27] factorizes a tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{D_1 \times \dots \times D_N}$ as a series of matrix-tensor products and contractions between third-order tensors as

$$\boldsymbol{\mathcal{X}} = \mathbf{G}_1 \overset{1}{\underset{2}{\bullet}} \boldsymbol{\mathcal{G}}_2 \dots \overset{1}{\underset{3}{\bullet}} \boldsymbol{\mathcal{G}}_{N-1} \overset{1}{\underset{3}{\bullet}} \mathbf{G}_N, \tag{2.8}$$

where $\mathbf{G}_1 \in \mathbb{R}^{D_1 \times K_1}$, $\mathbf{G}_N \in \mathbb{R}^{K_{N-1} \times D_N}$, and $\boldsymbol{\mathcal{G}}_n \in \mathbb{R}^{K_{n-1} \times D_n \times K_n}$, for $n \in \{2, \dots, N_1\})$, are referred to as TT-cores. The integers $K_1, \dots, K_{N_1}$ are called the TT-ranks. A fast and efficient way to estimate the TT-cores is to resort to the TT-SVD algorithm [27], a procedure that sequentially extracts dominant singular vectors from tensor unfoldings. As a result, the TTD in (2.8) is not unique. In fact, due to the use of the SVD, we can replace two successive TT-cores $\boldsymbol{\mathcal{G}}_n$ and $\boldsymbol{\mathcal{G}}_{n+1}$ by $\boldsymbol{\mathcal{G}}'_n$ and $\boldsymbol{\mathcal{G}}'_{n+1}$ such that

$$\boldsymbol{\mathcal{G}}'_n = \boldsymbol{\mathcal{G}}_n \overset{1}{\underset{3}{\bullet}} \mathbf{M}_n^{-1}, \quad \boldsymbol{\mathcal{G}}'_{n+1} = \mathbf{M}_n \overset{1}{\underset{2}{\bullet}} \boldsymbol{\mathcal{G}}_{n+1}, \tag{2.9}$$

where $\mathbf{M}_n \in \mathbb{R}^{K_n \times K_n}$ is a non-singular change-of-basis matrix. This means that the multiplicative ambiguities in the TTD correspond to post- and pre-multiplications by nonsingular matrices.

2.4. Equivalence between the CPD and the TTD. There exists an equivalence between the CP and TT decompositions. If the TT-ranks are such that $K_1 = \dots = K_{N-1} = K \leqslant \min(D_1, \dots, D_N)$, the TTD can be used to efficiently estimate the rank-K CP factors of a higher-order tensor [43]. In particular, using the TT-SVD, one can obtain TT factors such that:

$$\mathbf{G}_1 = \mathbf{A}_1\mathbf{M}_1^{-1},\ \boldsymbol{\mathcal{G}}_n = [\![\mathbf{M}_{n-1}, \mathbf{A}_n, \mathbf{M}_n^{-\mathsf{T}}]\!],\ \mathbf{G}_N = \mathbf{M}_{N-1}\mathbf{A}_N^{\mathsf{T}}, \tag{2.10}$$

where $\mathbf{M}_n \in \mathbb{R}^{K_n \times K_n}$ are non-singular change-of-basis matrices. These matrices can be related to the permutation and scaling ambiguity matrices of the CPD, see the proof of [43, Theorem 6] for more details.

3. Joint estimation and dimensionality reduction.

3.1. The model and the optimization problem. Let $\boldsymbol{\mathcal{X}}$ a tensor of rank K that admits a CPD of the form

$$\boldsymbol{\mathcal{X}} = [\![\mathbf{A}_1, \dots, \mathbf{A}_N]\!], \text{ where } \mathbf{A}_n \in \mathbb{R}^{D_n \times K}\ \forall n \in \{1, \dots, N\}. \tag{3.1}$$

Let $\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{D_1 \times \dots \times D_N}$ the noisy observation according to the following model

$$\boldsymbol{\mathcal{Y}} = \boldsymbol{\mathcal{X}} + \boldsymbol{\mathcal{E}}, \tag{3.2}$$

where the tensor $\boldsymbol{\mathcal{E}}$ represents isotropic white Gaussian noise. The model (3.2) can be rewritten

$$\boldsymbol{\mathcal{Y}} = [\![\mathbf{A}_1, \ldots, \mathbf{A}_N]\!] + \boldsymbol{\mathcal{E}}. \tag{3.3}$$

The problem of estimating the CP factors $\mathbf{A}_n$ based on model (3.3) has found numerous interests over the past decade. It has been applied, among others, to problems in signal processing and machine learning [34], brain signal processing [1], video completion [41] or telecommunications [29]. The popularity of the CPD lies in its high regularization power, mild uniqueness properties and versatility to model a variety of problems, from denoising to component analysis.

The most popular way to estimate the CP factors is an iterative Alternating Least-Squared (ALS) algorithm [9]. However, this procedure becomes exponentially expensive as the order N increases. One iteration of CP-ALS requires $\mathcal{O}\left(K^2 \prod_n D_N\right)$ flops. Furthermore, CP-ALS becomes less robust and is prone to convergence issues for high-order tensors. These limitations fall under the so-called curse of dimensionality [28]. Consequently, one may need to perform tensor dimensionality reduction to exploit the benefits of model (3.3).

Directly solving (3.3) using high-dimensional CP-ALS is costly. To circumvent this limitation, it is possible to perform low-rank denoising by considering the following two optimization problems. To ease the notation, let us denote $\mathbb{A} = \{\mathbf{A}_1, \ldots, \mathbf{A}_N\}$, $\mathbb{M} = \{\mathbf{M}_1, \ldots, \mathbf{M}_{N-1}\}$ and $\mathbb{G} = \{\mathbf{G}_1, \boldsymbol{\mathcal{G}}_2, \ldots, \boldsymbol{\mathcal{G}}_{N-1}, \mathbf{G}_N\}$.

First, a low rank-K factorized approximation is obtained by solving

$$\min_{\mathbb{G}} \|\boldsymbol{\mathcal{Y}} - \mathbf{G}_1 \overset{1}{\underset{2}{\bullet}} \boldsymbol{\mathcal{G}}_2 \ldots \overset{1}{\underset{N-1}{\bullet}} \boldsymbol{\mathcal{G}}_{N-1} \overset{1}{\underset{N}{\bullet}} \mathbf{G}_N\|_F^2. \tag{3.4}$$

This step can be interpreted as a denoising of $\boldsymbol{\mathcal{Y}}$ using the TT-SVD with rank K. This operation provides the best rank-K approximation of $\boldsymbol{\mathcal{Y}}$ in the least-squared sense, while being less costly than high-dimensional CP-ALS.

Then, the links between the TT-cores $\mathbb{G}$ and the CP factors $\mathbb{A}$ (2.10) are formulated through the optimization problem

$$\begin{aligned} \min_{\mathbb{A},\mathbb{M}} \; & \sum_{n=2}^{N-1} \left(\|\boldsymbol{\mathcal{G}}_n - [\![\mathbf{M}_{n-1}, \mathbf{A}_n, \mathbf{M}_n^{-\mathsf{T}}]\!]\|_F^2 \right) \\ & + \|\mathbf{G}_1 - \mathbf{A}_1 \mathbf{M}_1^{-1}\|_F^2 + \|\mathbf{G}_N - \mathbf{M}_{N-1} \mathbf{A}_N^{\mathsf{T}}\|_F^2, \end{aligned} \tag{3.5}$$

that accounts for the ambiguities of the TT-SVD through the $\mathbb{M}$ matrices. The n-th ambiguity matrix $\mathbf{M}_n$ appears twice in (3.5), once in the term related to $\boldsymbol{\mathcal{G}}_n$ and once in that related to $\boldsymbol{\mathcal{G}}_{n+1}$. As a result, (3.5) is a coupled optimization problem.

3.2. State-of-the-art and its limitations. In [43], the authors proposed to promote a fast local/sequential optimization method as a sub-optimal solution instead of minimizing (3.4)–(3.5). This procedure was called Joint dImensionality Reduction And Factor Estimation (JIRAFE) [43]. The first step of JIRAFE estimates the TT-cores using the TT-SVD algorithm with ranks $K_1 = \ldots = K_{N-1} = K$ by solving (3.4). Then, CP-ALS is performed on a single $\boldsymbol{\mathcal{G}}_n$ ($n \in \{2, \ldots, N-1\}$), e.g. on $\boldsymbol{\mathcal{G}}_2$, by minimizing the criterion

$$\min_{\mathbf{M}_1, \mathbf{A}_2, \mathbf{M}_2} \|\boldsymbol{\mathcal{G}}_2 - [\![\mathbf{M}_1, \mathbf{A}_2, \mathbf{M}_2^{-\mathsf{T}}]\!]\|_F^2, \tag{3.6}$$

Since the three CP factors $\mathbf{M}_1$, $\mathbf{A}_2$, $\mathbf{M}_2$ underlying $\boldsymbol{\mathcal{G}}_2$ are estimated thanks to an alternate least-squares algorithm, this step was referred to as Tri-ALS. In JIRAFE, the resulting estimate of $\mathbf{M}_2$ is then fixed and propagated to other $\boldsymbol{\mathcal{G}}_n$, $n \in \{3, \dots, N-1\}$. Then, for each $\boldsymbol{\mathcal{G}}_n$, the cost function to minimize is

$$\min_{\mathbf{A}_n, \mathbf{M}_n} \|\boldsymbol{\mathcal{G}}_n - [\![\mathbf{M}_{n-1}, \mathbf{A}_n, \mathbf{M}_n^{-\mathsf{T}}]\!]\|_F^2, \tag{3.7}$$

where $\mathbf{M}_{n-1}$ is supposed to be known. This operation was termed Bi-ALS. Finally, $\mathbf{A}_1$ and $\mathbf{A}_N$ were obtained using the links provided in (2.10). Algorithm 3.1 provides a complexity gain of approximately N`iter` flops with respect to CP-ALS. The JIRAFE procedure is summarized in Algorithm 3.1 below.

Algorithm 3.1 JIRAFE

input: Observation tensor $\boldsymbol{\mathcal{Y}}$, CP-rank K
output: CP factors $\mathbf{A}_1, \dots, \mathbf{A}_N$
Estimate $\mathbf{G}_1$, $\mathbf{G}_N$ and $\boldsymbol{\mathcal{G}}_n$ $\forall n \in \{2, \dots, N-1\}$ using TT-SVD on $\boldsymbol{\mathcal{Y}}$
repeat
  Estimate $\mathbf{M}_1$, $\mathbf{A}_2$, $\mathbf{M}_2^{-1}$ using (3.6) (Tri-ALS)
  for $n = 3, \dots, N-1$ **do**
    Estimate $\mathbf{A}_n$, $\mathbf{M}_n^{-1}$ using (3.7) with $\mathbf{M}_{n-1}$ known (Bi-ALS)
  end for
until convergence
Estimate $\mathbf{A}_1$ and $\mathbf{A}_N$ using (2.10)

However, JIRAFE suffers from a major limitation due to the sub-optimal estimation of the ambiguity matrices $\mathbf{M}_n$. For instance, $\mathbf{M}_2$ was estimated only from $\boldsymbol{\mathcal{G}}_2$ even though it was shared between the CPD of $\boldsymbol{\mathcal{G}}_2$ and $\boldsymbol{\mathcal{G}}_3$. More generally, each matrix $\mathbf{M}_n$ is shared between the CPD of $\boldsymbol{\mathcal{G}}_n$ and $\boldsymbol{\mathcal{G}}_{n+1}$. As such, some important information contained in (3.5) is ignored by JIRAFE so that the coherence between the TT-cores is lost. As a result, the performance of JIRAFE is highly dependent on the choice of the TT-core considered for the initial Tri-ALS step. By first estimating the CP factors underlying $\boldsymbol{\mathcal{G}}_2$, the estimation error on the $\mathbf{M}_n, \mathbf{A}_n$ for $n > 2$ is expected to increase with n. As a result, $\mathbf{M}_{N-1}$ and $\mathbf{A}_N$ shall inherit from the estimation errors of previous CP factors. Conversely, if Tri-ALS was performed first on $\boldsymbol{\mathcal{G}}_{N-1}$ before propagating estimations downto $n = 1$, $\mathbf{A}_1$ and $\mathbf{M}_1$ would not be estimated accurately. Therefore, the following subsection introduces a new estimation method to circumvent these limitations.

3.3. Proposed approach. This section proposes a new approach to solve problem (3.5). It fully takes into account the coupling induced by the $\mathbf{M}_n$ by carefully considering the global minimization of the cost function of (3.5). In place of a sequential estimation of factors that solves local optimization sub-problems, the global optimization is considered. Each iteration of the algorithm will update every factor, thus ensuring to reach the global minimum of (3.5). This is made possible thanks to a block-coordinate descent that alternates between all the parameters of interest. The resulting algorithm is called Dimensionality Reduction, joint Estimation of the Ambiguity Matrices and the CP FACtors (DREAMFAC).

First, DREAMFAC estimates the $\mathbf{G}_1$, $\mathbf{G}_N$ and $\boldsymbol{\mathcal{G}}_n$ $\forall n \in \{2, \dots, N-1\}$ thanks to a TT-SVD applied to $\boldsymbol{\mathcal{Y}}$. This step consists in solving for (3.4), hence this step is similar in JIRAFE and DREAMFAC. Then, the coupling constraints on the ambiguity matrices lead to the following coupled model:

$$
\begin{cases}
\mathbf{G}_1\mathbf{M}_1 &= \mathbf{A}_1\\
\mathbf{G}_2^{(1)\mathsf{T}} &= \mathbf{M}_1\left(\mathbf{M}_2^{-\mathsf{T}}\odot\mathbf{A}_2\right)^{\mathsf{T}}
\end{cases}
$$

$$
\begin{cases}
\mathbf{G}_n^{(3)}\mathbf{M}_n &= \left(\mathbf{A}_n\odot\mathbf{M}_{n-1}\right), & \forall\, 2\leqslant n\leqslant N-2,\\
\mathbf{G}_{n+1}^{(1)\,\mathsf{T}} &= \mathbf{M}_n\left(\mathbf{M}_{n+1}^{-\mathsf{T}}\odot\mathbf{A}_{n+1}\right)^{\mathsf{T}}, & \forall\, 2\leqslant n\leqslant N-2,
\end{cases}
\tag{3.8}
$$

$$
\begin{cases}
\mathbf{G}_{N-1}^{(3)}\mathbf{M}_{N-1} &= \left(\mathbf{A}_{N-1}\odot\mathbf{M}_{N-2}\right)\\
\mathbf{M}_{N-1}\mathbf{A}_N^{\mathsf{T}} &= \mathbf{G}_N
\end{cases}
$$

Considering that the ambiguity matrices $\mathbf{M}_n$ are fixed, the $\mathbf{A}_n$ can be updated from (3.8) by solving a simple least-squares problem, as in JIRAFE. Considering that the CP factors $\mathbf{A}_n$ are fixed, the coherence between TT-cores leads to new coupled updates for the $\mathbf{M}_n$ matrices:

$$
\mathbf{M}_1 = \arg\min_{\mathbf{M}_1} \|\mathbf{G}_1\mathbf{M}_1-\mathbf{A}_1\|_F^2 + \|\mathbf{G}_2^{(1)\mathsf{T}}-\mathbf{M}_1\left(\mathbf{M}_2^{-\mathsf{T}}\odot\mathbf{A}_2\right)^{\mathsf{T}}\|_F^2, \tag{3.9}
$$

$$
+\,\|\mathbf{G}_n^{(1)\mathsf{T}}-\mathbf{M}_{n-1}\left(\mathbf{M}_n^{-\mathsf{T}}\odot\mathbf{A}_n\right)^{\mathsf{T}}\|_F^2,
$$

$$
\begin{aligned}
\mathbf{M}_n = \arg\min_{\mathbf{M}_n} &\|\mathbf{G}_n^{(3)}\mathbf{M}_n-\left(\mathbf{A}_n\odot\mathbf{M}_{n-1}\right)\|_F^2\\
&+\|\mathbf{G}_{n+1}^{(1)\,\mathsf{T}}-\mathbf{M}_n\left(\mathbf{M}_{n+1}^{-\mathsf{T}}\odot\mathbf{A}_{n+1}\right)^{\mathsf{T}}\|_F^2,, \qquad \forall\, 2\leqslant n\leqslant N-2,
\end{aligned}
\tag{3.10}
$$

$$
\mathbf{M}_{N-1} = \arg\min_{\mathbf{M}_{N-1}} \|\mathbf{G}_{N-1}^{(3)}\mathbf{M}_{N-1}-\left(\mathbf{A}_{N-1}\odot\mathbf{M}_{N-2}\right)\|_F^2+\|\mathbf{G}_N-\mathbf{M}_{N-1}\mathbf{A}_N^{\mathsf{T}}\|_F^2. \tag{3.11}
$$

The solution to this mixture of least-squares problem is explicit. Appendix A gathers the resulting closed-form expressions for the ambiguity matrices. In the proposed approach, the $(\mathbf{M}_n)_{1\leqslant n\leqslant N-1}$ and $(\mathbf{A})_{2\leqslant n\leqslant N-1})$ matrices are all updated in the same loop, termed Multi-ALS. This new procedure, called Dimensionality Reduction, joint Estimation of the Ambiguity Matrices and the CP FACtors (DREAMFAC), is summarized in Algorithm 3.2 below.

Algorithm 3.2 DREAMFAC

input: Observation tensor $\boldsymbol{\mathcal{Y}}$, CP-rank K
output: CP factors $\mathbf{A}_1,\ldots,\mathbf{A}_N$
Estimate $\mathbf{G}_1$, $\mathbf{G}_N$ and $\boldsymbol{\mathcal{G}}_n$ $\forall n\in\{2,\ldots,N-1\}$ using TT-SVD on $\boldsymbol{\mathcal{Y}}$
repeat
  Estimate the $\mathbf{M}_n$ using (3.9)–(3.11), and the $\mathbf{A}_n$ using (3.7) (Multi-ALS)
  Estimate $\mathbf{A}_1$ and $\mathbf{A}_N$ using (2.10)
until convergence

Note that the coherence between the TT-cores is ensured thanks to the coupled estimation of the ambiguity matrices $\mathbf{M}_n$. Hence the information comprised in model (3.8) is fully exploited by this new approach. As a by-product, in contrast with JIRAFE, the performance of DREAMFAC is independent on the starting TT-core. Figure 1 shows a graphical representation of model (3.8) and summarizes the differences between JIRAFE and DREAMFAC.

Finally, turning to the algorithmic complexity, Algorithm 3.2 requires at most the same number of updates as JIRAFE. While updating the $\mathbf{A}_n$ matrices still requires

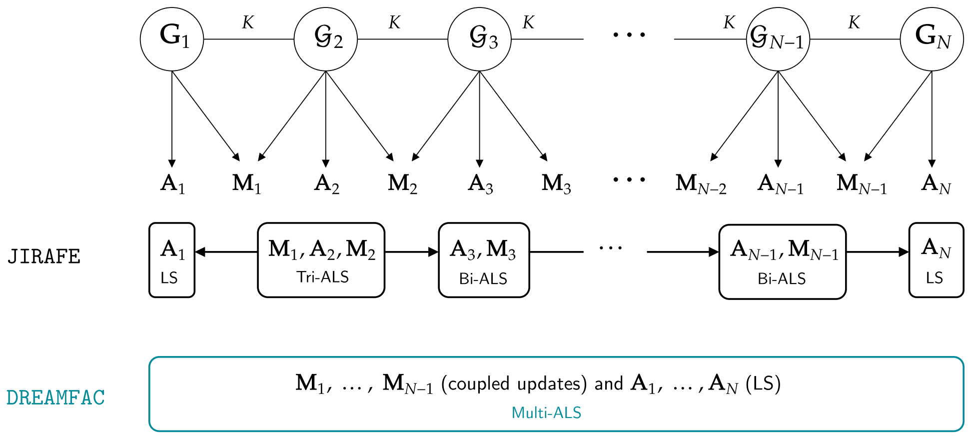

FIG. 1. *Graphical representation of JIRAFE (black) and DREAMFAC (blue).*

$\mathcal{O}\left(K^2 D_n\right)$ flops as in JIRAFE, the coupled updates of the $\mathbf{M}_n$'s can be made faster by using a fast solver for the Sylvester equations [3, 14, 35]. Such solvers require only $\mathcal{O}\left(2K^3\right)$ flops, which is smaller[1] than $\mathcal{O}\left(K^2 D_n\right)$ as soon as $K \ll D_n$, which is most likely.

4. Cramér-Rao bounds for the proposed approach.

4.1. General probabilistic framework. To derive appropriate performance bounds, it is necessary to embed the problem in an appropriate probabilistic framework. This requires to properly define the probabilistic model and the parameters of interest. A general coupled model has the form

$$\begin{cases} \boldsymbol{g}_n \sim \mathbf{f}_{\boldsymbol{\omega}}(\boldsymbol{g}_n) & \forall\, n \in \{1, \dots, N\}, \\ \mathbf{h}_n(\boldsymbol{\omega}) = 0, \end{cases} \tag{4.1}$$

where $\boldsymbol{g}_n$ denote the observations. The functions $\mathbf{f}_{\boldsymbol{\omega}}(\boldsymbol{g}_n)$ are the probability density functions (PDF) of the random real datasets $\boldsymbol{g}_n$, parameterized by the unknown deterministic real parameter vector $\boldsymbol{\omega}$. Several assumptions will be necessary: i) the PDFs $\mathbf{f}_{\boldsymbol{\omega}}(\boldsymbol{g}_n)$ are non-redundant functions differentiable w.r.t. $\boldsymbol{\omega}$, and their support as functions of $\boldsymbol{g}_n$ do not depend on $\boldsymbol{\omega}$; ii) the $\boldsymbol{g}_n$ are statistically independent. The constraints in the model are described by the functions $\mathbf{h}_n$ that are non-redundant deterministic vector functions, everywhere differentiable with respect to $\boldsymbol{\omega}$.

The model (3.8) can be rewritten equivalently under the form (4.1). First, the link between the TT-cores and the $\mathbf{A}_n$, $\mathbf{M}_n$ matrices is such that

$$\mathbf{G}_1 = \mathbf{A}_1 \mathbf{M}_{1,1}^{-1}, \ \boldsymbol{\mathcal{G}}_n = [\![\mathbf{M}_{n-1,n}, \mathbf{A}_n, \mathbf{M}_{n,n}^{-\mathsf{T}}]\!], \ \mathbf{G}_N = \mathbf{M}_{N-1,N} \mathbf{A}_N^{\mathsf{T}}. \tag{4.2}$$

This paper proposes an approach to solve the approximation problem (**??**) in the least-squares sense based on model (3.3). Therefore, it is assumed that the observations are Gaussian. To be more precise, let us note $\boldsymbol{g}_n = \mathrm{vec}\{\boldsymbol{\mathcal{G}}_n\}$ for $n \in \{1, N\}$ and $\boldsymbol{g}_n = \mathrm{vec}\{\boldsymbol{\mathcal{G}}_n\}$ for $n \in \{2, \dots, N-1\}$. The vectors $\boldsymbol{g}_n$ are random real Gaussian distributed datasets parameterized by their mean, *i.e.*,

$$\boldsymbol{g}_n \sim \boldsymbol{\mathcal{N}}(\boldsymbol{\mu}_n(\boldsymbol{\omega}), \sigma_n^2 \mathbf{I}) \ \text{ where } \ \boldsymbol{\mu}_n = \mathrm{vec}\{[\![\mathbf{M}_{n-1,n}, \mathbf{A}_n, \mathbf{M}_{n,n}^{-\mathsf{T}}]\!]\}, \tag{4.3}$$

[1]In many applications, e.g., harmonic retrieval in MIMO channels, the number of elements K is often very small.

and σ_n^2 denote the variances of the Gaussian noise on $\boldsymbol{g}_n$.

The unknown real deterministic parameter $\boldsymbol{\omega}$ is such that

$$\boldsymbol{\omega} = \left[\text{vec}\{\mathbf{A}_1\}^\mathsf{T} \ \text{vec}\{\mathbf{M}_{1,1}\}^\mathsf{T} \ \text{vec}\{\mathbf{M}_{1,2}\}^\mathsf{T} \ \text{vec}\{\mathbf{A}_2\}^\mathsf{T} \ \dots \ \text{vec}\{\mathbf{A}_N\}^\mathsf{T}\right]. \tag{4.4}$$

Second, the coherence between the TT-cores is ensured by the set of constraints

$$\mathbf{M}_{n,n} = \mathbf{M}_{n,n+1} \ \forall \ n \in \{1, \dots, N-1\}. \tag{4.5}$$

Reshaping (4.5) under the form of vector functions depending on $\boldsymbol{\omega}$ yields

$$\mathbf{h}_n(\boldsymbol{\omega}) = \text{vec}\{\mathbf{M}_{n,n}\} - \text{vec}\{\mathbf{M}_{n,n+1}\} \ \forall n \in \{1, \dots, N-1\}. \tag{4.6}$$

As a result, the model (4.3) under constraints (4.6) characterizes (3.8) under the form of a general coupled model.

4.2. Calculation of the Constrained Cramér-Rao bound. Evaluating the performance of the coupled model (4.3) under constraints (4.6) is necessary to compare the relative performance of JIRAFE and DREAMFAC for the estimation of $\boldsymbol{\omega}$. The standard tool for this task is the constrained Cramér-Rao bound (CCRB) [22, 40], defined[2] as

$$\mathbf{CCRB}(\boldsymbol{\omega}) = \mathbf{U}\left[\mathbf{U}^\mathsf{T}\mathbf{F}\mathbf{U}\right]^{-1}\mathbf{U}^\mathsf{T}, \tag{4.7}$$

where $\mathbf{F} \triangleq \mathbf{F}(\boldsymbol{\omega})$ is the Fisher Information Matrix (FIM) of the model, $\mathbf{J}_h$ is the Jacobian matrix related to the constraints $\mathbf{h}(\boldsymbol{\omega}) = 0$ and $\mathbf{U}$ is a basis of $\ker(\mathbf{J}_h)$.

Additionally to the model parameter $\boldsymbol{\omega}$, let us define $\boldsymbol{x} = \text{vec}\{\boldsymbol{\mathcal{X}}\} \in \mathbb{R}^\ell$ ($\ell = \Pi_n D_n$), that represents the vectorized low-rank approximation of $\boldsymbol{\mathcal{Y}}$. The parameter $\boldsymbol{x}$ can be linked to the model parameters through the relationship $\boldsymbol{x} - \mathbf{h}_\mathcal{X}(\boldsymbol{\omega}) = 0$, where $\mathbf{h}_\mathcal{X}$ is a non-redundant deterministic vector function, everywhere differentiable with respect to $\boldsymbol{\omega}$ such that

$$\mathbf{h}_\mathcal{X}(\boldsymbol{\omega}) = \text{vec}\{[\![\mathbf{A}_1, \dots, \mathbf{A}_N]\!]\}. \tag{4.8}$$

The above parameterization allows us to compute the CCRB on the reconstruction of $\boldsymbol{x}$ [19, p.125] as

$$\mathbf{CCRB}(\boldsymbol{x}) = \left[\frac{\mathbf{h}_\mathcal{X}(\boldsymbol{\omega})}{\partial \boldsymbol{\omega}^\mathsf{T}}\right]^\mathsf{T} \mathbf{CCRB}(\boldsymbol{\omega}) \left[\frac{\mathbf{h}_\mathcal{X}(\boldsymbol{\omega})}{\partial \boldsymbol{\omega}^\mathsf{T}}\right]. \tag{4.9}$$

4.2.1. Expression of the FIM. The FIM for $\boldsymbol{\omega}$ first needs to be computed in order to obtain the CCRB.

The FIM for $\boldsymbol{\omega}$ is a block-diagonal matrix of the form

$$\mathbf{F}(\boldsymbol{\omega}) = \begin{bmatrix} \mathbf{F}_1(\boldsymbol{\omega}) & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \ddots & & \vdots \\ \vdots & & \ddots & \mathbf{0} \\ \mathbf{0} & & \mathbf{0} & \mathbf{F}_N(\boldsymbol{\omega}) \end{bmatrix}, \tag{4.10}$$

where the blocks $\mathbf{F}_n(\boldsymbol{\omega})$ encode the contributions of each $\boldsymbol{g}_n$ to the estimation of $\boldsymbol{\omega}$. They are obtained by using the Slepian-Bangs formula [36]:

$$\mathbf{F}_n(\boldsymbol{\omega}) = \frac{1}{\sigma_n^2}\left[\frac{\partial \boldsymbol{\mu}_n(\boldsymbol{\omega})}{\partial \boldsymbol{\omega}^\mathsf{T}}\right]^\mathsf{T}\left[\frac{\partial \boldsymbol{\mu}_n(\boldsymbol{\omega})}{\partial \boldsymbol{\omega}^\mathsf{T}}\right]. \tag{4.11}$$

[2]If $\mathbf{F}$ is invertible, then (4.7) and the alternative expression for the CCRB provided in [15] are equivalent [40, Corollary 1].

The expression of $\left[\frac{\partial \boldsymbol{\mu}_n(\boldsymbol{\omega})}{\partial \boldsymbol{\omega}^\mathsf{T}}\right]$ is obtained simply for $n = 1$ and $n = N$:

$$\frac{\partial \boldsymbol{\mu}_1(\boldsymbol{\omega})}{\partial \operatorname{vec}\{\mathbf{A}_1\}^\mathsf{T}} = \mathbf{M}_{1,1}^{-\mathsf{T}} \boxtimes \mathbf{I}_{D_1}, \tag{4.12}$$

$$\frac{\partial \boldsymbol{\mu}_1(\boldsymbol{\omega})}{\partial \operatorname{vec}\{\mathbf{M}_{1,1}\}^\mathsf{T}} = \mathbf{I}_K \boxtimes \mathbf{A}_1, \tag{4.13}$$

$$\frac{\partial \boldsymbol{\mu}_1(\boldsymbol{\omega})}{\partial \operatorname{vec}\{\mathbf{M}_{N-1,N}\}^\mathsf{T}} = \mathbf{A}_N \boxtimes \mathbf{I}_K, \tag{4.14}$$

$$\frac{\partial \boldsymbol{\mu}_N(\boldsymbol{\omega})}{\partial \operatorname{vec}\{\mathbf{A}_N\}^\mathsf{T}} = \boldsymbol{\Pi}_N \left(\mathbf{M}_{N-1,N}^{-1} \boxtimes \mathbf{I}_{D_N}\right), \tag{4.15}$$

where $\boldsymbol{\Pi}_N$ is a permutation matrix linking the entries of $\operatorname{vec}\{\mathbf{G}_N^\mathsf{T}\}$ to those of $\operatorname{vec}\{\mathbf{G}_N\}$. For $n \in \{2, \dots, N-1\}$, $\left[\frac{\partial \boldsymbol{\mu}_n(\boldsymbol{\omega})}{\partial \boldsymbol{\omega}^\mathsf{T}}\right]$ can be computed using relationships between tensor unfoldings:

$$\begin{aligned}
\boldsymbol{g}_n &= \left[(\mathbf{M}_{n,n}^{-\mathsf{T}} \odot \mathbf{A}_n) \boxtimes \mathbf{I}_K\right] \operatorname{vec}\{\mathbf{M}_{n-1,n}\} && (4.16)\\
&= \boldsymbol{\Pi}_n^{(2,1)} \left[(\mathbf{M}_{n,n}^{-\mathsf{T}} \odot \mathbf{M}_{n-1,n}) \boxtimes \mathbf{I}_{D_n}\right] \operatorname{vec}\{\mathbf{A}_n\} && (4.17)\\
&= \boldsymbol{\Pi}_n^{(3,1)\mathsf{T}} \left[\mathbf{I}_K \boxtimes (\mathbf{A}_n \odot \mathbf{M}_{n-1,n})\right] \operatorname{vec}\{\mathbf{M}_{n,n}\}, && (4.18)
\end{aligned}$$

where $\boldsymbol{\Pi}_n^{(2,1)}$ and $\boldsymbol{\Pi}_n^{(3,1)}$ are permutation matrices that link the entries of $\operatorname{vec}\{\mathbf{G}_n^{(2)}\}$ (resp. $\operatorname{vec}\{\mathbf{G}_n^{(3)}\}$) to those of $\boldsymbol{g}_n = \operatorname{vec}\{\mathbf{G}_n^{(1)}\}$. To ease the notation, we define the matrices $\mathbf{S}_{1,n}, \mathbf{S}_{2,n}, \mathbf{S}_{3,n}$ ($n \in \{2, \dots, N-1\}$) as

$$\begin{aligned}
\mathbf{S}_{1,n} &= \left[(\mathbf{M}_{n,n}^{-\mathsf{T}} \odot \mathbf{A}_n) \boxtimes \mathbf{I}_K\right], && (4.19)\\
\mathbf{S}_{2,n} &= \boldsymbol{\Pi}_n^{(2,1)} \left[(\mathbf{M}_{n,n}^{-\mathsf{T}} \odot \mathbf{M}_{n-1,n}) \boxtimes \mathbf{I}_{D_n}\right], && (4.20)\\
\mathbf{S}_{3,n} &= \boldsymbol{\Pi}_n^{(3,1)\mathsf{T}} \left[\mathbf{I}_K \boxtimes (\mathbf{A}_n \odot \mathbf{M}_{n-1,n})\right]. && (4.21)
\end{aligned}$$

As a result, we have

$$\left[\frac{\partial \boldsymbol{\mu}_n(\boldsymbol{\omega})}{\partial \boldsymbol{\omega}^\mathsf{T}}\right] = \begin{cases} \left[\mathbf{M}_{1,1}^{-\mathsf{T}} \boxtimes \mathbf{I}_{D_1} \quad \mathbf{I}_K \boxtimes \mathbf{A}_1\right] & \text{for } n = 1,\\ \left[\mathbf{S}_{1,n} \quad \mathbf{S}_{2,n} \quad \mathbf{S}_{3,n}\right] & \forall\ n \in \{2, \dots, N-1\},\\ \left[\mathbf{A}_N \boxtimes \mathbf{I}_K \quad \boldsymbol{\Pi}_N \left(\mathbf{M}_{N-1,N}^{-1} \boxtimes \mathbf{I}_{D_N}\right)\right] & \text{for } n = N. \end{cases} \tag{4.22}$$

The closed-form expressions for the $\mathbf{F}_n(\boldsymbol{\omega})$ are provided in Appendix B.

4.2.2. Expression of the CCRB. The Jacobian matrix $\mathbf{J}_h$ is obtained by deriving the functions $\mathbf{h}_n(\boldsymbol{\omega})$ with respect to all the elements of $\boldsymbol{\omega}$ into a block-matrix. Given the constraints on $\boldsymbol{\omega}$ in (4.6), it holds that

$$\frac{\partial \mathbf{h}_n(\boldsymbol{\omega})}{\partial \operatorname{vec}\{\mathbf{A}_n\}^\mathsf{T}} = \mathbf{0}, \quad \forall n \in \{1, \dots, N-1\}, \tag{4.23}$$

$$\frac{\partial \mathbf{h}_n(\boldsymbol{\omega})}{\partial \operatorname{vec}\{\mathbf{M}_{m-1,m}\}^\mathsf{T}} = \begin{cases} \mathbf{I}_{K^2} \text{ if } m = n\\ \mathbf{0} \text{ if } m \neq n, \end{cases} \tag{4.24}$$

$$\frac{\partial \mathbf{h}_n(\boldsymbol{\omega})}{\partial \operatorname{vec}\{\mathbf{M}_{m,m}\}^\mathsf{T}} = \begin{cases} -\mathbf{I}_{K^2} \text{ if } m = n\\ \mathbf{0} \text{ if } m \neq n. \end{cases} \tag{4.25}$$

As a result, $\mathbf{J}_h$ is such that

$$\mathbf{J}_h = \begin{bmatrix} \frac{\partial \mathbf{h}_1(\boldsymbol{\omega})}{\partial \boldsymbol{\omega}^\mathsf{T}} \\ \vdots \\ \frac{\partial \mathbf{h}_{N-1}(\boldsymbol{\omega})}{\partial \boldsymbol{\omega}^\mathsf{T}} \end{bmatrix} = \begin{bmatrix} \mathbf{0} & \mathbf{I}_{K^2} & -\mathbf{I}_{K^2} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & & & \mathbf{0} \\ \vdots & \mathbf{0} & \mathbf{0} & \vdots & \mathbf{I}_{K^2} & -\mathbf{I}_{K^2} & \dots & & & \vdots \\ \vdots & \vdots & \vdots & \vdots & \mathbf{0} & \mathbf{0} & \dots & & & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \dots & & & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{I}_{K^2} & -\mathbf{I}_{K^2} & \mathbf{0} \end{bmatrix}. \tag{4.26}$$

The matrix $\mathbf{U}$ can then be obtained by simply solving $\mathbf{J}_h\mathbf{U} = \mathbf{0}$. Therefore $\mathbf{U}$ is the identity matrix of size $K(2N - 2 + \sum_n D_N)$. The CCRB submatrices for the $\mathbf{A}_n$ and the $\mathbf{M}_{n-1,n}$ are obtained by developing expression (4.7):

$$\mathbf{CCRB}(\mathbf{A}_n) = \mathbf{F}(\mathbf{A}_n)^{-1}, \quad \mathbf{CCRB}(\mathbf{M}_{n-1,n}) = \mathbf{F}(\mathbf{M}_{n-1,n})^{-1}. \tag{4.27}$$

4.3. Reparameterized CRB for reconstruction of the low-rank tensor.

The expression for $\mathbf{h}_{\mathcal{X}}(\boldsymbol{\omega})$ is necessary to compute the reparameterized CRB in (4.9). The relationships between tensor unfoldings give

$$\boldsymbol{x} = [(\mathbf{A}_N \odot \dots \odot \mathbf{A}_2) \boxtimes \mathbf{I}_{D_1}] \operatorname{vec}\{\mathbf{A}_1\} \tag{4.28}$$
$$= \dots$$
$$= \mathbf{\Pi}_{\mathcal{X}}^{(N,1)} [(\mathbf{A}_{N-1} \odot \dots \odot \mathbf{A}_1) \boxtimes \mathbf{I}_{D_N}] \operatorname{vec}\{\mathbf{A}_N\}, \tag{4.29}$$

where the $\mathbf{\Pi}_{\mathcal{X}}^{(n,1)}$ are permutation matrices that link the entries of the $\operatorname{vec}\{\mathbf{X}^{(n)}\}$ vectors to those of $\operatorname{vec}\{\mathbf{X}^{(1)}\} = \boldsymbol{x}$. The matrices $\mathbf{S}_{n,\mathcal{X}}$ are defined such that

$$\mathbf{S}_{n,\mathcal{X}} = \mathbf{\Pi}_{\mathcal{X}}^{(n,1)} [(\mathbf{A}_N \odot \dots \odot \mathbf{A}_{n+1} \odot \mathbf{A}_{n-1} \odot \dots \odot \mathbf{A}_1) \boxtimes \mathbf{I}_{D_n}], \tag{4.30}$$

finally yielding

$$\left[\frac{\mathbf{h}_{\mathcal{X}}(\boldsymbol{\omega})}{\partial \boldsymbol{\omega}^\mathsf{T}}\right] = [\mathbf{S}_{1,\mathcal{X}} \quad \mathbf{0} \quad \mathbf{0} \quad \mathbf{S}_{2,\mathcal{X}} \quad \mathbf{0} \quad \mathbf{0} \quad \dots \quad \mathbf{S}_{N,\mathcal{X}}]. \tag{4.31}$$

Therefore,

$$\mathbf{CCRB}(\boldsymbol{x}) = \operatorname{Diag}\{\mathbf{S}_{1,\mathcal{X}}^\mathsf{T}\mathbf{F}(\mathbf{A}_1)^{-1}\mathbf{S}_{1,\mathcal{X}}, \dots, \mathbf{S}_{N,\mathcal{X}}^\mathsf{T}\mathbf{F}(\mathbf{A}_N)^{-1}\mathbf{S}_{N,\mathcal{X}}\}. \tag{4.32}$$

Given the expressions of the constrained Cramér-Rao bounds in Equation (4.27) and Equation (4.32), the performance of the proposed approach can now be evaluated numerically.

5. Simulations.

5.1. Recovery of the CP factors. A 7-order ($N = 7$) tensor $\boldsymbol{\mathcal{X}}$ with $D_1 = \dots = D_N = 6$ and $K = 3$ is taken as a reference. This tensor admits a reference CPD as in (3.1) with i.i.d. entries generated from the normal distribution. The noisy tensor $\boldsymbol{\mathcal{Y}}$ was generated using isotropic white Gaussian noise to yield a SNR of 20dB.

The proposed approach is compared to two JIRAFE-like algorithms. In the first one, Tri-ALS is initially performed on $\boldsymbol{\mathcal{G}}_2$ and the factors is propagated towards the higher n. In the second one, Tri-ALS is performed on $\boldsymbol{\mathcal{G}}_{N-1}$ with propagation towards the lower n. The TT-cores are estimated using TT-SVD with $K = 3$. The CP factors were initialized as $\mathbf{A}_n = \mathbf{U}_n (\mathbf{\Sigma}_n)^{\frac{1}{2}}$ using the SVD of the n-th mode unfolding of $\boldsymbol{\mathcal{Y}}$

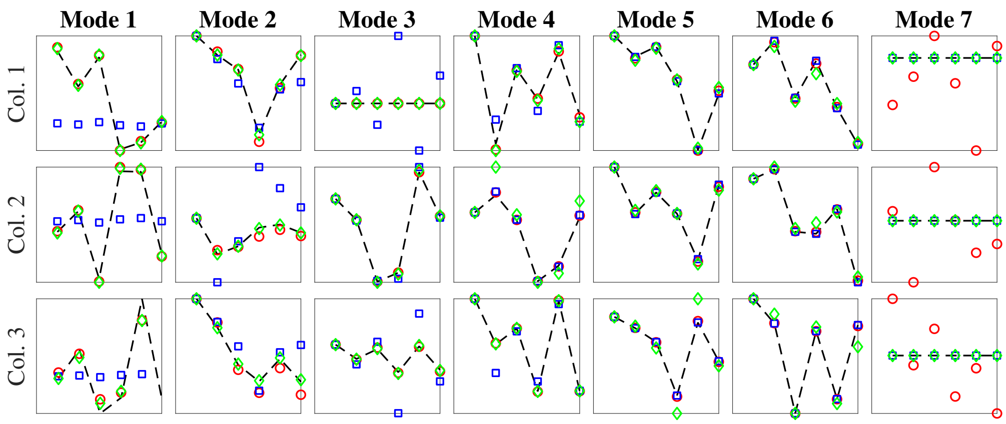

FIG. 2. *Columns of the reference (black dashed lines) and estimated factors as returned by JIRAFE with forward propagation (red circles), JIRAFE with backward propagation (blue squares), and DREAMFAC (green diamonds).*

with rank K, namely $\mathbf{Y}^{(n)} = \mathbf{U}_n \mathbf{\Sigma}_n \mathbf{V}_n$. The algorithms have a maximum of 1000 iterations, and the results are averaged over 100 noise realizations.

Figure 2 shows columns of the reference and estimated factors. The columns of the $\mathbf{A}_n$ factors for $n \in \{2, \dots, N-1\}$ are correctly estimated by all the algorithms, except for a few outliers in $\mathbf{A}_6$. Algorithm 3.2 usually provides a slightly better estimate than the JIRAFE procedures. While DREAMFAC also correctly estimates $\mathbf{A}_1$ and $\mathbf{A}_N$, the JIRAFE-like procedures provided an incorrect estimation of $\mathbf{A}_N$ (resp. $\mathbf{A}_1$).

To further investigate the propagation of the error through the TT-cores, the normalized mean square error between the estimated $\widehat{\mathbf{A}}_n$ and the reference $\mathbf{A}_n$ is considered:

$$\text{NMSE} = \frac{\|\widehat{\mathbf{A}}_n - \mathbf{A}_n\|_F^2}{\|\mathbf{A}_n\|_F^2}. \tag{5.1}$$

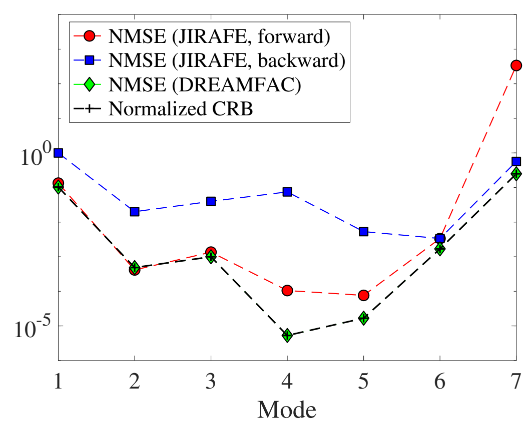

FIG. 3. *NMSE provided by JIRAFE with forward propagation (red dots), backward propagation (blue squares), and DREAMFAC (green diamonds), and normalized CRB.*

Figure 3 shows in semi-log scale the NMSE for each factor, as provided by the three algorithms. As a comparison, the value of the uniform CCRB, normalized by $\|\mathbf{A}_n\|_F^2$, is also displayed. For the two JIRAFE-like procedures, the highest NMSE corresponded to the last estimated factor, resp. $\mathbf{A}_N$ for the forward procedure and $\mathbf{A}_1$ for the backward procedure. The NMSE provided by DREAMFAC remained smaller for both $\mathbf{A}_1$ and $\mathbf{A}_N$. On average, it was also smaller than that provided by the JIRAFE algorithms, and reaches the optimal NMSE expected from the normalized CRB.

5.2. Relative efficiency of the algorithms. This subsection assesses the efficiency of JIRAFE and DREAMFAC by comparing their performance to the constrained Cramér-Rao bound obtained in Section 4.

The reference tensor $\boldsymbol{\mathcal{X}}$ is a 7th-order tensor with $D_1 = \ldots = D_N = 6$ and $K = 3$. The entries of the true CP factors $\mathbf{A}_n$ were generated once as i.i.d. real standard Gaussian variables. The model is simulated under additive Gaussian noise. We assume that the noise level was the same for all TT-cores: this assumption is reasonable since the $\boldsymbol{\mathcal{G}}_n$ are all estimated from $\boldsymbol{\mathcal{Y}}$ using TT-SVD. on the TT-cores. The SNR on the observed tensors in dB is defined as $SNR_i = 10\log_{10}\left(\|\boldsymbol{\mathcal{G}}_i\|_F^2/\|\boldsymbol{\mathcal{E}}_i\|_F^2\right)$, $(i = 1, \ldots, N)$.

The model parameters are retrieved using JIRAFE and DREAMFAC. The CP factors and ambiguity matrices are initialized randomly. The permutation ambiguities in the estimated factors are corrected using the Hungarian algorithm [25] to make comparisons aligned with the reference. The experiments show the uniform MSE and uniform CCRB obtained from the MSE and CCRB matrix traces, as widely considered in, e.g., [11,16,17]. The expressions for the bounds proposed in this paper permit the computation of the uniform CCRB by taking the trace of these matrices.

Figure 4 shows in semi-log scale the uniform bounds and MSEs for all the entries of the $\mathbf{A}_n$ and $\mathbf{M}_n$, $\boldsymbol{\omega}$ and $\boldsymbol{\mathcal{X}}$ as a function of the SNR. It is noticeable that the uniform MSE produced by DREAMFAC reaches the uniform CCRB. Therefore the proposed approach is optimal for estimation of the parameters and reconstruction of $\boldsymbol{\mathcal{X}}$. JIRAFE yields a higher MSE, which was expected since it is a suboptimal optimization method. The two algorithms depict the same kind of behavior with respect to the SNR. This is particularly visible for estimation of the ambiguity matrices. DREAMFAC permits to gain 5dB for the estimation of the $\mathbf{A}_n$, and 3dB for the reconstruction of $\boldsymbol{\mathcal{X}}$ compared to JIRAFE.

Figure 5 shows in semi-log scale the uniform CCRB for $\mathbf{A}_1$, $\mathbf{A}_3$, $\mathbf{A}_5$ and $\mathbf{A}_7$, as well as the uniform MSEs, as a function of the SNR. The results for $\mathbf{A}_2$, $\mathbf{A}_4$ and $\mathbf{A}_6$ are similar. The uniform MSE obtained with DREAMFAC reaches the CCRB for each factor, and its scale barely varies trough the $\mathbf{A}_n$. The MSE obtained with JIRAFE was higher, and progressively increased for high values of n. Therefore DREAMFAC is also efficient for the estimation of each CP factor.

5.3. Application to channel estimation in dual-polarized MIMO systems. The problem of channel estimation in dual-polarized massive MIMO aims at recovering the channel parameters at the base station (angles of arrival, angles of departure, path gains, and polarization parameters). In [29], a tensor-based approach for this task was proposed. The MIMO channel was recast as a fourth-order tensor admitting a CPD. In [42], the authors adapted the JIRAFE procedure to treat the channel estimation problem with receiver and transmitter rectangular arrays, hence the channel was viewed as a fifth-order tensor.

5.3.1. Model description. The steering vectors for the k-th path between a Uniform Rectangular Array (URA) transmitter of size $M_T^x \times M_T^y$ and a URA receiver of size $M_R^x \times M_R^y$ are such that

$$\boldsymbol{a}_T(k) = \boldsymbol{a}_T^x(k) \boxtimes \boldsymbol{a}_T^y(k), \;\; \boldsymbol{a}_R(k) = \boldsymbol{a}_R^x(k) \boxtimes \boldsymbol{a}_R^y(k), \tag{5.2}$$

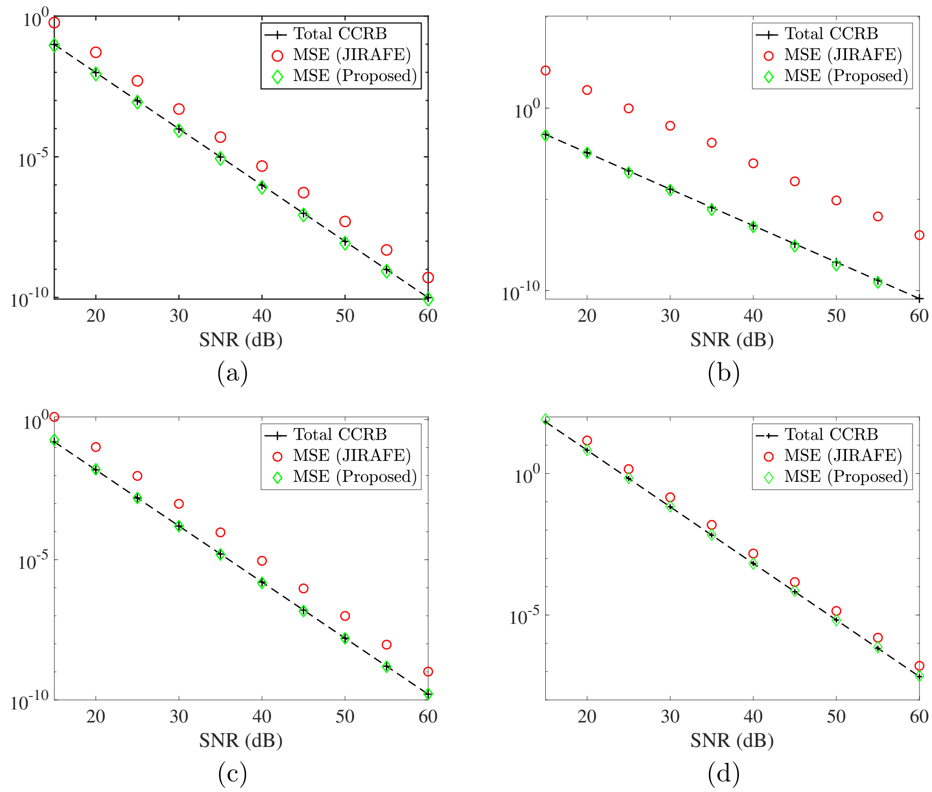

FIG. 4. *Uniform CCRB and MSE provided by JIRAFE and DREAMFAC, for (a) estimation of all the* $\mathbf{A}_n$*, (b) estimation of all the* $\mathbf{M}_n$*, (c) estimation of* $\boldsymbol{\omega}$ *and (d) reconstruction of* $\boldsymbol{\mathcal{X}}$*.*

with e.g., $\boldsymbol{a}_T^x(k) = \left[1, \exp\left(j\omega_T^x(k)\right), \ldots, \exp\left(j\omega_T^x(k)(M_T^x - 1)\right)\right]^\mathsf{T}$, and likewise for $\boldsymbol{a}_T^y(k), \boldsymbol{a}_R^x(k), \boldsymbol{a}_R^y(k)$. As a result, for K paths, the steering matrices in transmission and in reception are

$$\mathbf{A}_T = \mathbf{A}_T^x \odot \mathbf{A}_T^y, \ \mathbf{A}_R = \mathbf{A}_R^x \odot \mathbf{A}_R^y, \tag{5.3}$$

with e.g., $\mathbf{A}_T^x = [\boldsymbol{a}_T^x(1), \ldots, \boldsymbol{a}_T^x(K)]$.

The path-loss matrix $\mathbf{B} \in \mathbb{C}^{4 \times K}$ contains the path-loss parameters. Let the (p,q)-th subchannel correspond to $p \in \{V_R, H_R\}$ for the vertical (V) polarized and horizontal (H) polarized receive antennas, and $q \in \{V_T, H_T\}$ for the V-polarized and H-polarized transmit antennas. For the k-th path and for the (p,q)-th subchannel, $\beta_k^{(p,q)}$ is the path-loss parameter. Therefore, noting $\beta^{(p,q)} = \left[\beta_1^{(p,q)}, \ldots, \beta_K^{(p,q)}\right]$,

$$\mathbf{B} = \left[\beta^{(V_R,V_T)} \ \beta^{(V_R,H_T)} \ \beta^{(H_R,V_T)} \ \beta^{(H_R,H_T)}\right], \tag{5.4}$$

As a result, the channel tensor $\boldsymbol{\mathcal{H}} \in \mathbb{C}^{M_T^x \times M_T^y \times M_R^x \times M_R^y \times 4}$ can be written as

$$\boldsymbol{\mathcal{H}} = [\![\mathbf{A}_R^x, \mathbf{A}_R^y, (\mathbf{A}_T^x)^\star, (\mathbf{A}_T^y)^\star, \mathbf{B}]\!] + \boldsymbol{\mathcal{E}}, \tag{5.5}$$

which corresponds to a fifth-order CPD of rank K. The noise term $\boldsymbol{\mathcal{E}}$ encompasses the background noise and the estimation error due to the pre-estimation of the unstructured channel, and can be modeled as zero-mean circularly complex Gaussian random variables.

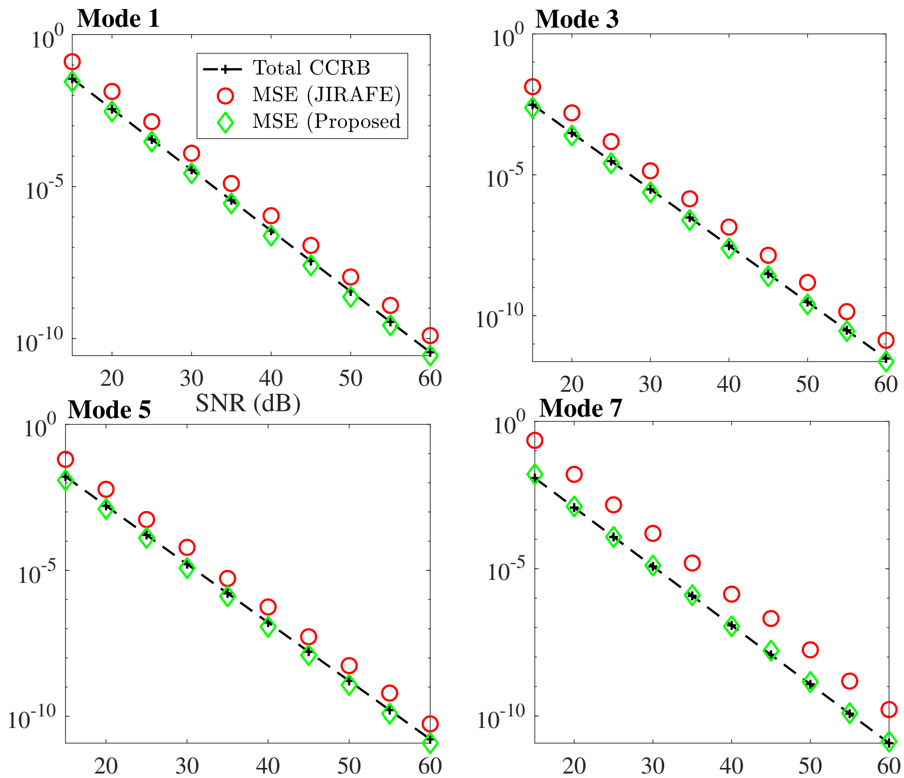

FIG. 5. *Uniform CCRB and uniform MSE provided by JIRAFE and DREAMFAC for the estimation of the* $\mathbf{A}_n$ *individually.*

The main assumption of the model (5.5) is that $K \leqslant \min(M_T^x, M_T^y, M_R^x, M_R^y)$, *i.e.*, the steering matrices are full column-rank. Furthermore, it is assumed that there are few dominant paths, *i.e.*, $K < 4$ and $\mathbf{B}$ is a full column-rank matrix. The assumption of a small number of paths is usually made in massive MIMO scenarios.

5.3.2. Results. The matrices $\mathbf{A}_T^x \, \mathbf{A}_T^y, \mathbf{A}_R^x, \mathbf{A}_R^y$ were generated once based on single random realizations of the angular frequencies $\omega_T^x(k), \omega_T^y(k), \omega_R^x(k), \omega_R^y(k)$ following a uniform distribution on $]0, \pi]$. The factor $\mathbf{B}$ was drawn from a complex Gaussian distribution with zero mean and unit variance. The following dimensions were considered: $M_T^x = M_R^x = 10$ and $M_T^y = M_R^y = 8$, and $K = 3$.

Figure 6 shows in semi-log scale true factors and the estimated factors provided by DREAMFAC and JIRAFE, with 30dB SNR. All factors are recovered correctly by DREAMFAC. The factors estimated by JIRAFE seem coherent with the reference factors, but with the wrong scale and angle.

To further assess the performance of the proposed approach, the MSE between

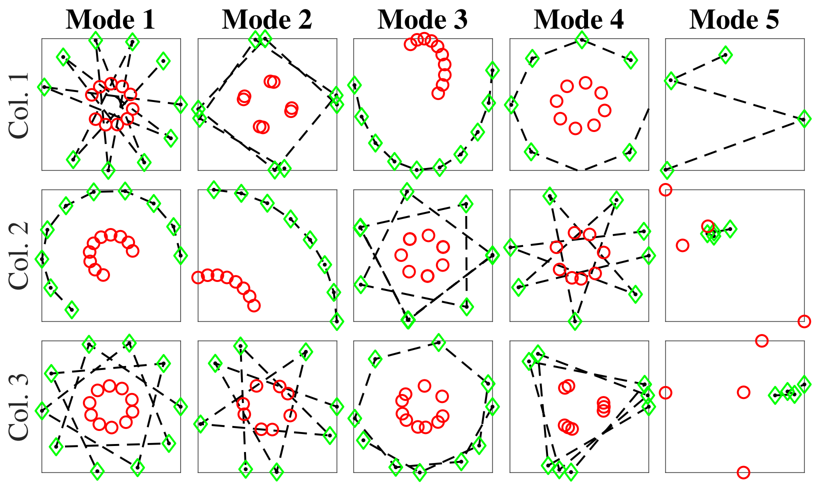

FIG. 6. *Columns of the reference factors (black dashed lines) and estimated factors as returned by JIRAFE (red circles) and DREAMFAC (green diamonds).*

the true and estimated factors, and between the angular frequencies, were considered:

(5.6)
$$\mathbf{MSE}_R^x = \sum_{k=1}^{K} \left(\mathbf{A}_R^x(:,k) - \widehat{\mathbf{A}}_R^x(:,k)\right)^2, \quad \mathbf{MSE}_R^y = \sum_{k=1}^{K} \left(\mathbf{A}_R^y(:,k) - \widehat{\mathbf{A}}_R^y(:,k)\right)^2,$$

(5.7)
$$\mathbf{MSE}_T^x = \sum_{k=1}^{K} \left(\mathbf{A}_T^x(:,k) - \widehat{\mathbf{A}}_T^x(:,k)\right)^2, \quad \mathbf{MSE}_T^y = \sum_{k=1}^{K} \left(\mathbf{A}_T^y(:,k) - \widehat{\mathbf{A}}_T^y(:,k)\right)^2,$$

(5.8)
$$\begin{aligned}\mathbf{MSE}_\omega = &\sum_{k=1}^{K} \left(\omega_T^x(k) - \widehat{\omega}_T^x(k)\right)^2 + \left(\omega_T^y(k) - \widehat{\omega}_T^y(k)\right)^2 \\ &+ \left(\omega_R^x(k) - \widehat{\omega}_R^x(k)\right)^2 + \left(\omega_R^y(k) - \widehat{\omega}_R^y(k)\right)^2.\end{aligned}$$

The MSE were evaluated over 10 values of the SNR in $[-20, 30]$ dB, and calculated by averaging the results over 500 independent noise realizations. Computing the complex CCRB is a delicate task that requires the calculation of Wirtinger derivatives [26]. For this reason, calculation of the CRB associated with model (5.5) is relegated to future works.

DREAMFAC was compared to four tensor-based methods. The first one was CP-ALS [9] followed by closed-form solutions to estimate the parameters from the factors. The second one was the so-called CP-VDM, for CPD with Vandermonde factor matrix, proposed in [37]. The third one was based on the generalized eigenvalue decomposition (GEVD) [33]. The fourth one was JIRAFE followed by a Vandermonde rectification strategy [5] (termed JIRAFE-VDM) to enforce the structure of the steering matrices. This step is crucial to estimate the factors with the correct scale and angle, as exemplified in Figure 6. All algorithms used at most 1000 iterations.

Figure 7 shows in semi-log scale the averaged MSE for recovery of the steering matrices. For a SNR superior or equal to 5dB, all approaches but GEVD yield the same MSE. Thanks to the Vandermonde rectification, JIRAFE could achieve good

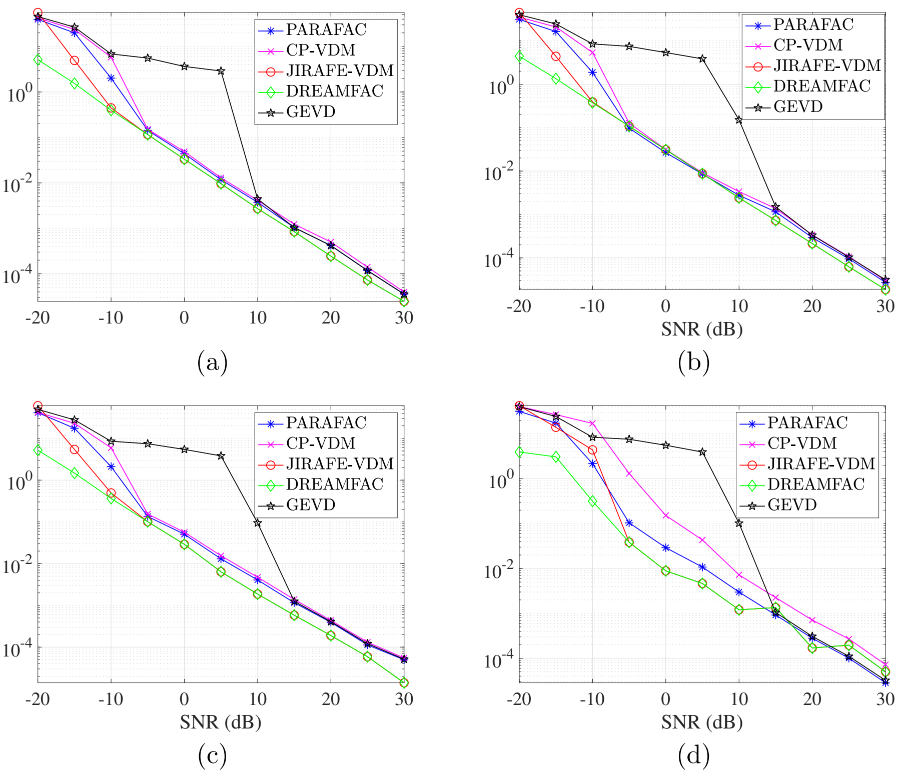

FIG. 7. *(a)* $\boldsymbol{MSE}_R^x$*, (b)* $\boldsymbol{MSE}_R^y$*, (c)* $\boldsymbol{MSE}_T^x$*, (d)* $\boldsymbol{MSE}_T^y$ *as a function of the SNR.*

performance for higher values of the SNR. For lower values of the SNR, DREAMFAC yields the lowest MSE. Its performance is better than that of JIRAFE, with a gain of approximately 3dB on the quality of the estimation.

Figure 8 shows in semi-log scale the averaged $\mathbf{MSE}_\omega$ provided by the algorithms, as a function of the SNR. For estimation of the angular frequencies, DREAMFAC yields the best MSE for a low SNR. Its gain with respect to JIRAFE for low SNR is of approximately 5dB. The proposed approach has similar performance to other approaches for a SNR superior or equal to 5dB.

6. Conclusion. This paper proposes a new approach called DREAMFAC for Dimensionality Reduction, joint Estimation of the Ambiguity Matrices and the CP FACtors. It relies on a global coupled optimization scheme, instead of a local and sequential strategy in previous approaches. DREAMFAC performs favorably with respect to the state-of-the-art of estimation of the CP factors. We have derived constrained Cramér-Rao bounds to evaluate the potential optimum performance under mild assumptions, as well as for comparison with state-of-the-art methods. Numerical experiments show that the proposed algorithm reaches the Cramér-Rao bound, as expected since it fully exploits the information of the full coupled model.

The performance of DREAMFAC is exemplified on the realistic problem of complex harmonics retrieval in dual-polarized MIMO channels. The proposed approach performs well on this task, providing better estimation of the base station parameter at low SNR than state-of-the-art methods. As a conclusion, this paper clarifies what are the best performance of a low-rank tensor CP factorization using tensor-train.

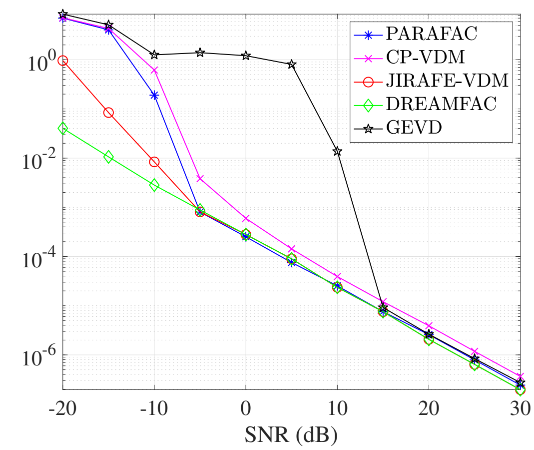

FIG. 8. *MSE on the estimation of the angular frequencies.*

The proposed approach, DREAMFAC, is an optimal solution to this problem.

Appendix A. Closed-form expression for the coupled $\mathbf{M}_n$ in the proposed approach.

Solutions the least-squared problems (3.9)–(3.11) can be obtained by solving the following Sylvester equations:

$$\mathbf{M}_1\left(\mathbf{M}_2^{-1}\mathbf{M}_2^{-\mathsf{T}} \boxdot \mathbf{A}_2^{\mathsf{T}}\mathbf{A}_2\right) + \left(\mathbf{G}_1^{\mathsf{T}}\mathbf{G}_1\right)\mathbf{M}_1 = \mathbf{G}_1^{\mathsf{T}}\mathbf{A}_1 + \mathbf{G}_2^{(1)\mathsf{T}}\left(\mathbf{A}_2 \odot \mathbf{M}_2\right), \tag{A.1}$$

$$\begin{aligned}\mathbf{M}_{n-1}\left(\mathbf{M}_n^{-1}\mathbf{M}_n^{-\mathsf{T}} \boxdot \mathbf{A}_n^{\mathsf{T}}\mathbf{A}_n\right) &+ \left(\mathbf{G}_{n-1}^{(3)}{}^{\mathsf{T}}\mathbf{G}_{n-1}^{(3)}\right)\mathbf{M}_{n-1} \\ &= \mathbf{G}_n^{(1)\mathsf{T}}\left(\mathbf{M}_n^{-\mathsf{T}} \odot \mathbf{A}_n\right) + \mathbf{G}_{n-1}^{(3)}{}^{\mathsf{T}}\left(\mathbf{A}_{n-1} \odot \mathbf{M}_{n-2}\right),\end{aligned} \tag{A.2}$$

$$\begin{aligned}\mathbf{M}_n\left(\mathbf{M}_{n+1}^{-1}\mathbf{M}_{n+1}^{-\mathsf{T}} \boxdot \mathbf{A}_{n+1}^{\mathsf{T}}\mathbf{A}_{n+1}\right) &+ \left(\mathbf{G}_n^{(3)\mathsf{T}}\mathbf{G}_n^{(3)}\right)\mathbf{M}_n \\ &= \mathbf{G}_{n+1}^{(1)}{}^{\mathsf{T}}\left(\mathbf{M}_{n+1}^{-\mathsf{T}} \odot \mathbf{A}_{n+1}\right) + \mathbf{G}_n^{(3)\mathsf{T}}\left(\mathbf{A}_n \odot \mathbf{M}_{n-1}\right),\end{aligned} \tag{A.3}$$

$$\begin{aligned}\mathbf{M}_{N-1}\left(\mathbf{A}_N^{\mathsf{T}}\mathbf{A}_N\right) &+ \left(\mathbf{G}_{N-1}^{(3)}{}^{\mathsf{T}}\mathbf{G}_{N-1}^{(3)}\right)\mathbf{M}_{N-1} \\ &= \mathbf{G}_N\mathbf{A}_N + \mathbf{G}_{N-1}^{(3)}{}^{\mathsf{T}}\left(\mathbf{A}_{N-1} \odot \mathbf{M}_{N-2}\right).\end{aligned} \tag{A.4}$$

Fast solvers (see [35]) can be used to solve (A.1)–(A.4).

Appendix B. Closed-form expressions for the uncoupled FIM.

The FIM matrices $\mathbf{F}_n(\boldsymbol{\omega})$ introduced in (4.11) are block-matrices such as

(B.1)
$$\mathbf{F}_1(\boldsymbol{\omega}) = \frac{1}{\sigma_1^2}\begin{bmatrix}\mathbf{M}_{1,1}^{-1}\mathbf{M}_{1,1}^{-\mathsf{T}} \boxtimes \mathbf{I}_{d_1} & \mathbf{M}_{1,1}^{-1} \boxtimes \mathbf{A}_1 \\ \mathbf{M}_{1,1}^{-\mathsf{T}} \boxtimes \mathbf{A}_1^{\mathsf{T}} & \mathbf{I}_R \boxtimes \mathbf{A}_1^{\mathsf{T}}\mathbf{A}_1\end{bmatrix},$$

(B.2)
$$\mathbf{F}_n(\boldsymbol{\omega}) = \frac{1}{\sigma_n^2}\begin{bmatrix}\mathbf{S}_{1,n}^{\mathsf{T}}\mathbf{S}_{1,n} & \mathbf{S}_{1,n}^{\mathsf{T}}\mathbf{S}_{2,n} & \mathbf{S}_{1,n}^{\mathsf{T}}\mathbf{S}_{3,n} \\ \mathbf{S}_{2,n}^{\mathsf{T}}\mathbf{S}_{1,n} & \mathbf{S}_{2,n}^{\mathsf{T}}\mathbf{S}_{2,n} & \mathbf{S}_{2,n}^{\mathsf{T}}\mathbf{S}_{3,n} \\ \mathbf{S}_{3,n}^{\mathsf{T}}\mathbf{S}_{1,n} & \mathbf{S}_{3,n}^{\mathsf{T}}\mathbf{S}_{2,n} & \mathbf{S}_{3,n}^{\mathsf{T}}\mathbf{S}_{3,n}\end{bmatrix} \text{ for } n \in \{2,\ldots,N-1\},$$

(B.3)
$$\mathbf{F}_N(\boldsymbol{\omega}) = \frac{1}{\sigma_N^2}\begin{bmatrix}\mathbf{A}_N^{\mathsf{T}}\mathbf{A}_N \boxtimes \mathbf{I}_R & \left(\mathbf{A}_N^{T} \boxtimes \mathbf{I}_R\right)\mathbf{\Pi}_N\left(\mathbf{M}_{N-1,N}^{-1} \boxtimes \mathbf{I}_{D_N}\right) \\ \left(\mathbf{M}_{N-1,N}^{-\mathsf{T}} \boxtimes \mathbf{I}_{D_N}\right)\mathbf{\Pi}_N^{\mathsf{T}}\left(\mathbf{A}_N \boxtimes \mathbf{I}_R\right) & \mathbf{M}_{N-1,N}^{-\mathsf{T}}\mathbf{M}_{N-1,N}^{-1} \boxtimes \mathbf{I}_{D_N}\end{bmatrix}.$$

It is possible to identify the blocks containing the contribution of the $\mathbf{A}_n$ and the $\mathbf{M}_{n-1,n}$ to the estimation of $\boldsymbol{\omega}$). These blocks, denoted to as $\mathbf{F}(\mathbf{A}_n)$ and $\mathbf{F}(\mathbf{M}_{n-1,n})$, are such that

$$\text{(B.4)} \qquad \mathbf{F}(\mathbf{A}_n) = \begin{cases}\mathbf{I}_R \boxtimes \mathbf{A}_1^{\mathsf{T}}\mathbf{A}_1 & \text{for } n = 1, \\ \mathbf{S}_{2,n}^{\mathsf{T}}\mathbf{S}_{2,n} & \text{for } n \in \{2,\ldots,N-1\}, \\ \mathbf{A}_N^{\mathsf{T}}\mathbf{A}_N \boxtimes \mathbf{I}_R & \text{for } n = N,\end{cases}$$

$$\text{(B.5)} \qquad \mathbf{F}(\mathbf{M}_{n-1,n}) = \begin{cases}\mathbf{S}_{1,n}^{\mathsf{T}}\mathbf{S}_{1,n} & \text{for } n \in \{2,\ldots,N-1\}, \\ \mathbf{M}_{N-1,N}^{-\mathsf{T}}\mathbf{M}_{N-1,N}^{-1} \boxtimes \mathbf{I}_{D_N} & \text{for } n = N.\end{cases}$$

REFERENCES

[1] E. Acar, C. Aykut-Bingol, H. Bingol, R. Bro, and B. Yener, *Multiway analysis of epilepsy tensors*, Bioinformatics, 23 (2007), pp. i10–i18.

[2] D. C. Araújo, T. Maksymyuk, A. L. F. de Almeida, T. Maciel, J. C. M. Mota, and M. Jo, *Massive mimo: survey and future research topics*, IET Communications, 10 (2016), pp. 1938–1946.

[3] R. Bartels and G. Stewart, *Solution of the matrix equation AX+XB=C*, Commun. ACM, 15 (1972), pp. 820–826.

[4] M. Boizard, R. Boyer, G. Favier, J. Cohen, and P. Comon, *Performance estimation for tensor CP decomposition with structured factors*, in Proc. ICASSP, 2015.

[5] R. Boyer and P. Comon, *Rectified als algorithm for multidimensional harmonic retrieval*, in 2016 IEEE SAM Signal Processing Workshop, IEEE, 2016, pp. 1–5.

[6] R. Cabral Farias, J. Cohen, and P. Comon, *Exploring multimodal data fusion through joint decompositions with flexible couplings*, IEEE Trans. Signal Process., 64 (2016), pp. 4830–4844.

[7] P. Comon, *Tensors: A brief introduction*, IEEE Signal Process. Mag., 31 (2014), pp. 44–53.

[8] H. Cramér, *Mathematical Methods of Statistics*, Univ. Press, Princeton, 1946.

[9] L. De Lathauwer and D. Nion, *Decompositions of a higher-order tensor in block terms–part III: Alternating least squares algorithms*, SIAM journal on Matrix Analysis and Applications, 30 (2008), pp. 1067–1083.

[10] L. Domanov, I.and De Lathauwer, *Canonical polyadic decomposition of third-order tensors: Reduction to generalized eigenvalue decomposition*, SIAM Journal on Matrix Analysis and Applications, 35 (2014), pp. 636–660.

[11] Y. C. Eldar, *Minimum variance in biased estimation: Bounds and asymptotically optimal estimators*, IEEE Transactions on Signal Processing, 52 (2004), pp. 1915–1930.

[12] M. Fréchet, *Sur l'extension de certaines évaluations statistiques au cas de petits échantillons*, Rev. Int. Stat, 11 (1943), pp. 182–205.

[13] M. Giraud, V. Itier, R. Boyer, Y. Zniyed, and A. L. F. de Almeida, *Tucker decomposition based on a tensor train of coupled and constrained cp cores*, IEEE Signal Process. Letters, (2023).

[14] G. Golub, S. Nash, and C. V. Loan, *A Hessenberg-Schur method for the problem AX+XB=C*, IEEE Trans. Autom. Control, 24 (1979), pp. 909–913.

[15] J. Gorman and A. Hero, *Lower bounds for parametric estimation with constraints*, IEEE Trans. Inf. Theory, 36 (1990), pp. 1285–1301.

[16] A. O. Hero, *A Cramér-Rao type lower bound for essentially unbiased parameter estimation*, tech. report, Massachussets Inst. of tech., Lexington Lincoln Lab, 1992.

[17] A. O. Hero, J. A. Fessler, and M. Usman, *Exploring estimator bias-variance tradeoffs using the uniform CR bound*, IEEE Transactions on Signal Processing, 44 (1996), pp. 2026–2041.

[18] T. G. Kolda and B. W. Bader, *Tensor Decompositions and Applications*, SIAM Review, 51 (2009), pp. 455–500.

[19] E. Lehmann and G. Casella, *Theory of Point Estimation (2nd ed.)*, Springer, 1998.

[20] N. Li, S. Kindermann, and C. Navasca, *Some convergence results on the regularized alternating least-squares method for tensor decomposition*, Linear Algebra and its Applications, 438 (2013), pp. 796–812.

[21] X. Liu and N. Sidiropoulos, *Cramér-Rao lower bounds for low-rank decomposition of multidimensionnal arrays*, IEEE Trans. Signal Process., 49 (2001), pp. 2074–2086.

[22] T. Menni, E. Chaumette, P. Larzabal, and J. P. Barbot, *New results on Deterministic Cramér-Rao bounds for real and complex parameters*, IEEE Trans. on SP, 60 (2012), pp. 1032–1049.

[23] T. Menni, J. Galy, E. Chaumette, and P. Larzabal, *Versatility of Constrained CRB for System Analysis and Design*, IEEE Trans. on AES, 50 (2014), pp. 1841–1863.

[24] T. J. Moore, B. M. Sadler, and R. J. Kozick, *Maximum-Likelihood Estimation, the Cramér-Rao Bound, and the Method of Scoring With Parameter Constraints*, IEEE Trans. on SP, 56 (2008), pp. 895–908.

[25] J. Munkres, *Algorithms for the assignment and transportation problems*, Journal of the society for industrial and applied mathematics, 5 (1957), pp. 32–38.

[26] E. Ollila, V. Koivunen, and J. Eriksson, *On the cramér-rao bound for the constrained and unconstrained complex parameters*, in 2008 5th IEEE Sensor Array and Multichannel Signal Processing Workshop, IEEE, 2008, pp. 414–418.

[27] I. V. Oseledets, *Tensor-train decomposition*, SIAM Journal on Scientific Computing, 33 (2011), pp. 2295–2317.

[28] I. V. Oseledets and E. E. Tyrtyshnikov, *Breaking the curse of dimensionality, or how to use svd in many dimensions*, SIAM Journal on Scientific Computing, 31 (2009), pp. 3744–3759.

[29] C. Qian, X. Fu, N. D. Sidiropoulos, and Y. Yang, *Tensor-based channel estimation for dual-polarized massive mimo systems*, IEEE Trans. Signal Process., 66 (2018), pp. 6390–6403.

[30] C. R. Rao, *Information and accuracy attainable in the estimation of statistical parameters*, Bull. Calcutta Math. Soc, 37 (1945), pp. 81–91.

[31] C. Ren, R. Cabral Farias, P.-O. Amblard, and P. Comon, *Performance bounds for coupled models*, in Proc. 2016 IEEE SAM, 2016. event-place: Rio de Janeiro, Brazil.

[32] S. Sahnoun and P. Comon, *Joint source estimation and localization*, IEEE Trans. Signal Process., 63 (2015), pp. 2485–2595.

[33] E. Sanchez and B. R. Kowalski, *Tensorial resolution: a direct trilinear decomposition*, Journal of Chemometrics, 4 (1990), pp. 29–45.

[34] N. D. Sidiropoulos, L. De Lathauwer, X. Fu, K. Huang, E. E. Papalexakis, and C. Faloutsos, *Tensor decomposition for signal processing and machine learning*, IEEE Transactions on Signal Processing, 65 (2017), pp. 3551–3582.

[35] V. Simoncini, *Computational methods for linear matrix equations*, SIAM Review, 58 (2016), pp. 377–441, https://doi.org/10.1137/130912839.

[36] D. Slepian, *Estimation of signal parameters in the presence of noise*, Trans. IRE Professional Group Inf. Theory, 3 (1954), pp. 68–69.

[37] M. Sørensen and L. De Lathauwer, *Blind signal separation via tensor decomposition with vandermonde factor: Canonical polyadic decomposition*, IEEE Transactions on Signal Processing, 61 (2013), pp. 5507–5519.

[38] A. Stegeman and N. D. Sidiropoulos, *On kruskal's uniqueness condition for the candecomp/parafac decomposition*, Linear Algebra and its applications, 420 (2007), pp. 540–552.

[39] P. Stoica and T. L. Marzetta, *Parameter estimation problems with singular information matrices*, IEEE Trans. on SP, 49 (2001), pp. 87–90.

[40] P. Stoica and B. C. Ng, *On the Cramér-Rao bound under parametric constraints*, IEEE SP Letters, 5 (1998), pp. 177–179.

[41] Q. Zhao, G. Zhou, L. Zhang, A. Cichocki, and S.-I. Amari, *Bayesian robust tensor factorization for incomplete multiway data*, IEEE Trans. Neural Networks and Learning Systems, 27 (2015), pp. 736–748.

[42] Y. Zniyed, R. Boyer, A. De Almeida, and G. Favier, *Tensor train representation of mimo channels using the jirafe method*, Signal Processing, 171 (2020), p. 107479.

[43] Y. Zniyed, R. Boyer, A. F. De Almeida, and G. Favier, *High-order tensor estimation via trains of coupled third-order cp and tucker decompositions*, Linear Algebra and its Applications, 588 (2020), pp. 304–337.