Tensor CP Decomposition with Structured Factor Matrices: Algorithms and Performance

José Henrique de Morais Goulart, Mélanie Boizard, Remy Boyer, Gérard Favier, Pierre Comon

Tensor CP Decomposition with Structured Factor Matrices: Algorithms and Performance

José Henrique de M. Goulart, Maxime Boizard, Rémy Boyer, Gérard Favier and Pierre Comon

***Abstract*—The canonical polyadic decomposition (CPD) of high-order tensors, also known as Candecomp/Parafac, is very useful for representing and analyzing multidimensional data. This paper considers a CPD model having structured matrix factors, as *e.g.* Toeplitz, Hankel or circulant matrices, and studies its associated estimation problem. This model arises in signal processing applications such as Wiener-Hammerstein system identification and cumulant-based wireless communication channel estimation. After introducing a general formulation of the considered structured CPD (SCPD), we derive closed-form expressions for the Cramér-Rao bound (CRB) of its parameters under the presence of additive white Gaussian noise. Formulas for special cases of interest, as when the CPD contains identical factors, are also provided. Aiming at a more relevant statistical evaluation from a practical standpoint, we discuss the application of our formulas in a Bayesian context, where prior distributions are assigned to the model parameters. Three existing algorithms for computing SCPDs are then described: a constrained alternating least squares (CALS) algorithm, a subspace-based solution and an algebraic solution for SCPDs with circulant factors. Subsequently, we present three numerical simulation scenarios, in which several specialized estimators based on these algorithms are proposed for concrete examples of SCPD involving circulant factors. In particular, the third scenario concerns the identification of a Wiener-Hammerstein system via the SCPD of an associated Volterra kernel. The statistical performance of the proposed estimators is assessed via Monte Carlo simulations, by comparing their Bayesian mean-square error with the expected CRB.**

***Index Terms*—Tensor, CP decomposition, Structured matrices, Bayesian Cramér-Rao bound**

I. Introduction

Nowadays, high-order tensor models, also called multiway decompositions, are very useful in numerous applications for representing and analyzing multidimensional data as encountered in signal/image processing, computer vision, data mining, chemometrics, among many other application areas. For a review of tensor decompositions and applications, see the tutorial papers [3]–[5] and the books [6,7]. The most common tensor model is the canonical polyadic decomposition (CPD), independently introduced under the names of canonical decomposition (CANDECOMP) in [8] and parallel factor model (PARAFAC) in [9].

The CPD consists in decomposing a tensor of order N, with dimensions $I_1 \times \cdots \times I_N$ and rank R, into a sum of R rank-one tensors, *i.e.*, outer products of N vectors $\mathbf{a}_r^{(n)}$. Each $\mathbf{a}_r^{(n)}$ is the rth column of a factor matrix $\mathbf{A}^{(n)} \in \mathbb{R}^{I_n \times R}$, $n = 1, \ldots, N$. A well-known fundamental property of the CPD is its essential uniqueness under mild conditions, meaning that its factor matrices are unique up to permutation and scaling ambiguities on their columns.

The CPD can be viewed as one extension of the matrix singular value decomposition (SVD) to higher orders, with the difference that the factor matrices $\mathbf{A}^{(n)}$ are generally not orthogonal. In some signal processing applications, these factors have special form. For instance, Toeplitz, Hankel, and Vandermonde factors occur in problems like cumulant-based channel estimation [10], nonlinear system identification using Wiener-Hammerstein models [11], and multidimensional harmonic retrieval [12]. As a consequence, several specialized algorithms have been developed for estimating the parameters of CPD factors having those forms [13]–[16].

Given the practical relevance of estimating CPD factors with special structure, a systematic way of statistically evaluating estimators is desirable, as it allows choosing an appropriate approach in applied settings and provides important information also for the study and development of estimation algorithms. Although a wide variety of algorithms have been developed for estimating CPDs, very few theoretical results exist for assessing their statistical performance. The first contribution on this subject is the work in [17], which has derived the Cramér-Rao bound (CRB) for unstructured third- and fourth-order tensors, and has applied it to evaluate the performance of the standard alternating least squares (ALS) algorithm. The CRB, which is a classical subject in estimation theory [18], allows studying the statistical efficiency of unbiased estimators, since it is a lower bound for their mean squared errors (MSE). Apart from [17], both [19] and [20] have derived CRBs for the estimation of CPDs having Vandermonde factors, motivated, respectively, by the estimation of the directions of arrival of multiple source signals and by the estimation of the multidimensional harmonic model.

In the present paper, we first formulate the considered structured CPD (SCPD) with factor matrices belonging to proper subspaces spanned by given basis matrices. A nonlinear model constituted by a SCPD corrupted by white Gaussian noise is then presented, having reduced parametric

J. H. de M. Goulart is supported by CNPq-Brazil (individual grant 245358/2012-9). M. Boizard, R. Boyer and G. Favier wish to thank DIGITEO for its financial support on the project DETMOTS-2A. P. Comon is supported by ERC grant AdG-2013-320594 DECODA.

J. H. de M. Goulart and G. Favier are with the I3S Laboratory, CNRS, Univ. of Nice Sophia-Antipolis, Sophia-Antipolis, 06900, France (e-mails: {last name}@i3s.unice.fr).

R. Boyer and M. Boizard are with Université Paris-Sud, L2S and SATIE laboratories, France (e-mails: {first name}.{last name}@lss.supelec.fr).

P. Comon is with Gipsa-Lab, CNRS and Univ. Grenoble Alpes, F-38000, France (e-mail: p.comon@ieee.org).

The contents of this paper have been partially published in ICASSP 2015 and EUSIPCO 2015 [1,2].

complexity in comparison with an unstructured CPD. Closed-form expressions for the deterministic CRB are subsequently derived, including formulas for special cases of interest, as when the SCPD contains identical factors. Aiming at a statistical evaluation of greater practical relevance, we discuss the application of our formulas in a Bayesian context, where statistical priors are assigned to the model parameters and the Bayesian MSE (BMSE) of estimators is measured. It is shown that, in our setting, the expected CRB (ECRB) [21] is tighter than Van Trees' Bayesian CRB [22]. Three specialized algorithms for computing SCPDs are then described. The first one, called constrained ALS (CALS), is iterative and consists in adapting the popular ALS algorithm so as to take the structure constraint into account. The two others are non-iterative. They correspond to a subspace-based solution on one hand and to an algebraic solution for SCPDs with circulant factors on the other hand. Subsequently, we present three numerical simulation scenarios, in which specialized estimators are proposed for concrete examples of SCPDs involving circulant factors. The third scenario, in particular, concerns the identification of a Wiener-Hammerstein system [23] via the SCPD of an associated Volterra kernel [11]. The statistical performance of the proposed estimators is assessed via Monte Carlo simulations, by comparing their BMSE with the ECRB.

The rest of the paper is organized as follows. In Section II, we present a general formulation of the considered SCPD model and briefly discuss identifiability issues. Then, Section III is devoted to the derivation of closed-form expressions for the deterministic CRB and the study of their application in a Bayesian context. Three specialized algorithms for computing SCPDs are then described in Section IV, as well as the combination of algorithms for performance improvement. Monte Carlo simulation results are presented in Section V, where we assess the statistical performance of proposed estimators by comparing their BMSE with the ECRB, both in the case of SCPDs with circulant factors and in the context of Wiener-Hammerstein system identification via the SCPD of an associated Volterra kernel. The paper is concluded in Section VI, where some perspectives are drawn for future research.

Notation: We denote scalars, vectors, matrices and tensors by, respectively, lowercase letters (as θ_i), lowercase boldface letters (as $\boldsymbol{\theta}$), boldface capitals (*e.g.*, $\boldsymbol{\Theta}$), and calligraphic letters (as $\mathcal{X}$). The superscript ^{T} stands for transposition, ^{H} for Hermitian transposition and † for both left- and right-sided inverses. The symbols $\boxtimes$, $\odot$, $\times_n$ and $\otimes$ denote the Kronecker, Khatri-Rao, mode-n and outer (tensor) products, respectively. For our purposes, a Nth-order tensor $\mathcal{X}$ is assimilated to its array of coordinates, indexed by N indices. Elements of a vector, matrix or tensor are denoted either as in $[\mathbf{v}]_i$, $[\mathbf{A}]_{i_1,i_2}$ and $[\mathcal{X}]_{i_1,\ldots,i_N}$ or as in v_i, a_{i_1,i_2} and $x_{i_1,\ldots,i_N}$, respectively. $\mathbb{E}$ denotes mathematical expectation, and $\langle\mathbf{A}\rangle$ denotes the column space of $\mathbf{A}$. The ith canonical basis vector (whose dimension should be clear from the context) is denoted by $\mathbf{e}_i$. The operator $\mathrm{Diag}(\cdot)$ yields a diagonal matrix containing the coefficients of the vector argument in its diagonal, while $\mathrm{vec}(\cdot)$ stacks the columns of a matrix in a long vector. Along the text, we resort to the following properties: $\mathrm{vec}(\mathbf{ABC}) = \left(\mathbf{C}^T \boxtimes \mathbf{A}\right)\mathrm{vec}(\mathbf{B})$ and $\mathrm{vec}(\mathbf{A}\,\mathrm{Diag}(\mathbf{b})\mathbf{C}) = \left(\mathbf{C}^T \odot \mathbf{A}\right)\mathbf{b}$.

II. CPD MODEL WITH STRUCTURED FACTORS

In this section, we introduce the model which underlies the estimation problem considered in this paper. To keep its presentation simple, we shall restrict it to the case of third-order tensors, but the extension to higher orders is straightforward.

A. Tensors and the CP decomposition

Let $\mathcal{X}$ be a real $I_1 \times I_2 \times I_3$ tensor. A *polyadic decomposition* of $\mathcal{X}$ consists of a sum of rank-one terms which yields $\mathcal{X}$, *i.e.*,

$$\mathcal{X} = \sum_{r=1}^{R} \lambda_r \left(\mathbf{a}_r^{(1)} \otimes \mathbf{a}_r^{(2)} \otimes \mathbf{a}_r^{(3)}\right), \tag{1}$$

where λ_r is a scaling factor and $\mathbf{a}_r^{(n)}$ is the rth column of the nth *matrix factor* $\mathbf{A}^{(n)} \in \mathbb{R}^{I_n \times R}$. It should be noted that the columns $\mathbf{a}_r^{(n)}$ are usually normalized to eliminate the scaling indeterminacy of each term of (1). When R is the minimal integer such that (1) holds, then it is called the *rank* of $\mathcal{X}$ and (1) is called a *canonical polyadic decomposition* (CPD) of $\mathcal{X}$.

Decomposition (1) can be written in many useful alternative forms. First, we can write it in scalar form as

$$x_{i_1,i_2,i_3} = \sum_{r=1}^{R} \lambda_r a_{i_1,r}^{(1)} a_{i_2,r}^{(2)} a_{i_3,r}^{(3)}. \tag{2}$$

With respect to the mode-n unfolding of $\mathcal{X}$, we have

$$\mathbf{X}_n = \mathbf{A}^{(n)}\,\mathrm{Diag}(\boldsymbol{\lambda})\left(\mathbf{A}^{(n_1)} \odot \mathbf{A}^{(n_2)}\right)^T \in \mathbb{R}^{I_n \times I_{n_1} I_{n_2}}, \tag{3}$$

where $\boldsymbol{\lambda} = [\lambda_1 \ \ldots \ \lambda_R]^T$ and $n_q = (n - q - 1 \mod 3) + 1$ for $n \in \{1,2,3\}$ and $q \in \{1,2\}$. In terms of mode-n products, the CPD can be expressed as

$$\mathcal{X} = \mathcal{I} \times_1 \mathbf{A}^{(1)} \times_2 \mathbf{A}^{(2)} \times_3 \left(\mathbf{A}^{(3)}\,\mathrm{Diag}(\boldsymbol{\lambda})\right), \tag{4}$$

where $\mathcal{I} \in \mathbb{R}^{R \times R \times R}$ is the third-order diagonal tensor such that $[\mathcal{I}]_{r_1,r_2,r_3} = \delta_{r_1,r_2,r_3}$, with $\delta_{q,r,s}$ denoting the Kronecker delta symbol. It should be noted that the scaling matrix $\mathrm{Diag}(\boldsymbol{\lambda})$ can be absorbed by any of the three factors. Of even more utility in our development is the vectorized form

$$\mathbf{x} = \mathrm{vec}(\mathcal{X}) = \sum_{r=1}^{R} \lambda_r \left(\mathbf{a}_r^{(3)} \boxtimes \mathbf{a}_r^{(2)} \boxtimes \mathbf{a}_r^{(1)}\right), \tag{5}$$

where we adopt the convention $\mathrm{vec}(\mathcal{X}) = \mathrm{vec}(\mathbf{X}_1)$. Using (3), we can also write

$$\mathbf{x} = \mathrm{vec}(\mathbf{X}_1) = \left(\mathbf{A}^{(3)} \odot \mathbf{A}^{(2)} \odot \mathbf{A}^{(1)}\right)\boldsymbol{\lambda}. \tag{6}$$

B. Structured CPD model

We now formulate the structured CPD model with zero-mean white Gaussian noise, for which we shall derive analytical CRB formulae. To this end, let us introduce some definitions.

Definition 1 *A matrix $\mathbf{A}^{(n)} \in \mathbb{R}^{I_n \times R}$ is said to be* structured *if it can be written as [14]*

$$\mathbf{A}^{(n)} = \sum_{u=1}^{U_n} \theta_u^{(n)} \mathbf{E}_u^{(n)}, \tag{7}$$

where $U_n < I_n R$ and the matrices $\mathbf{E}_u^{(n)} \in \mathbb{R}^{I_n \times R}$ are linearly independent. In other words, $\mathbf{A}^{(n)}$ belongs to a proper subspace of dimension U_n spanned by the basis $B_n = \left\{\mathbf{E}_1^{(n)}, \ldots, \mathbf{E}_{U_n}^{(n)}\right\}$. In vector form, we have

$$\text{vec}\left(\mathbf{A}^{(n)}\right) = \mathbf{E}^{(n)} \boldsymbol{\theta}^{(n)}, \tag{8}$$

where

$$\mathbf{E}^{(n)} = \left[\text{vec}\left(\mathbf{E}_1^{(n)}\right), \ldots, \text{vec}\left(\mathbf{E}_{U_n}^{(n)}\right)\right] \in \mathbb{R}^{RI_n \times U_n} \tag{9}$$

and $\boldsymbol{\theta}^{(n)} = \left[\theta_1^{(n)}, \ldots, \theta_{U_n}^{(n)}\right]^T \in \mathbb{R}^{U_n}$.

As particular cases of Definition 1, we can mention banded, (block-)Hankel, (block-)Toeplitz and (block-)circulant matrices. It is therefore of broad applicability, characterizing structured matrices which are very important in many signal processing applications.

Because a structured matrix $\mathbf{A}^{(n)}$ belongs to a linear subspace, its columns can be written as

$$\mathbf{a}_r^{(n)} = \mathbf{S}_r^{(n)} \boldsymbol{\theta}^{(n)}, \tag{10}$$

where $\mathbf{S}_r^{(n)} \in \mathbb{R}^{I_n \times U_n}$ is defined such that its uth column equals the rth column of $\mathbf{E}_u^{(n)}$. Furthermore, since any (structured or non-structured) matrix can evidently be put in the form (7) with $U_n = I_n R$, by simply choosing B_n as the canonical basis, the columns of any matrix can be written as in (10). For convenience, we provide in Table I expressions of U_n, $\mathbf{S}_r^{(n)}$ and $\mathbf{E}_u^{(n)}$ for some classes of matrices. It should be noted that the expressions of $\mathbf{S}_r^{(n)}$ and $\mathbf{E}_u^{(n)}$ are not unique, since they depend on the particular ordering adopted for the vector of parameters, $\boldsymbol{\theta}^{(n)}$.

Definition 2 *A rank-R tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ is said to admit a* structured CPD *(SCPD) when it can be written as in* (4)*, with at least one structured matrix factor $\mathbf{A}^{(n)} \in \mathbb{R}^{I_n \times R}$, $n \in \{1, 2, 3\}$.*

We are now ready to introduce the structured CPD model,

$$\boldsymbol{\mathcal{Y}} = \boldsymbol{\mathcal{X}} + \boldsymbol{\mathcal{N}} \quad \in \mathbb{R}^{I_1 \times I_2 \times I_3}, \tag{11}$$

where $\boldsymbol{\mathcal{X}}$ admits a SCPD and $\boldsymbol{\mathcal{N}}$ is an additive noise tensor with i.i.d. zero-mean Gaussian entries of variance σ^2. By exploiting (10) together with (5), we can conveniently express the SCPD of $\boldsymbol{\mathcal{X}}$ under its vectorized form as

$$\mathbf{x} = \sum_{r=1}^{R} \lambda_r \left[\left(\mathbf{S}_r^{(3)} \boldsymbol{\theta}^{(3)}\right) \boxtimes \left(\mathbf{S}_r^{(2)} \boldsymbol{\theta}^{(2)}\right) \boxtimes \left(\mathbf{S}_r^{(1)} \boldsymbol{\theta}^{(1)}\right)\right].$$

Defining $\boldsymbol{\Phi}_r \triangleq \mathbf{S}_r^{(3)} \boxtimes \mathbf{S}_r^{(2)} \boxtimes \mathbf{S}_r^{(1)}$, we can rewrite $\mathbf{x}$ as a function of the parameter vector $\boldsymbol{\nu} = [\boldsymbol{\theta}^T,\ \boldsymbol{\lambda}^T]^T \in \mathbb{R}^K$, where $\boldsymbol{\theta} = \left[\boldsymbol{\theta}^{(1)^T},\ \boldsymbol{\theta}^{(2)^T},\ \boldsymbol{\theta}^{(3)^T}\right]^T$, having the form

$$\mathbf{x}(\boldsymbol{\nu}) = \underbrace{\left(\sum_{r=1}^{R} \lambda_r \boldsymbol{\Phi}_r\right)}_{\triangleq \boldsymbol{\Phi}(\boldsymbol{\lambda})} \underbrace{\left(\boldsymbol{\theta}^{(3)} \boxtimes \boldsymbol{\theta}^{(2)} \boxtimes \boldsymbol{\theta}^{(1)}\right)}_{\triangleq f(\boldsymbol{\theta})}. \tag{12}$$

Observe that $K = U_1 + U_2 + U_3 + R$. Finally, defining also $\mathbf{n} \triangleq \text{vec}(\boldsymbol{\mathcal{N}})$, we have the vectorized model

$$\mathbf{y}(\boldsymbol{\nu}) \triangleq \text{vec}(\boldsymbol{\mathcal{Y}}) = \boldsymbol{\Phi}(\boldsymbol{\lambda}) f(\boldsymbol{\theta}) + \mathbf{n}.$$

Note that the above development allows us to separate the contributions of $\boldsymbol{\theta}$ and $\boldsymbol{\lambda}$, which will considerably simplify the derivation of the CRB. In addition, (12) can be readily specialized for particular instances of interest, as when:

- The scaling factors are such that $\lambda_1 = \cdots = \lambda_R = \lambda$, which implies $\mathbf{x} = \lambda \boldsymbol{\Phi} f(\boldsymbol{\theta})$ with $\boldsymbol{\Phi} = \sum_r \boldsymbol{\Phi}_r$.
- Two or three factors are identical. For instance, if $\mathbf{A}^{(1)} = \mathbf{A}^{(2)} = \mathbf{A}^{(3)}$, then $\boldsymbol{\theta} = \boldsymbol{\theta}^{(1)} = \boldsymbol{\theta}^{(2)} = \boldsymbol{\theta}^{(3)}$ and so $f(\boldsymbol{\theta}) = \boldsymbol{\theta} \boxtimes \boldsymbol{\theta} \boxtimes \boldsymbol{\theta}$.

C. Identifiability

We say that the parameter vector $\boldsymbol{\nu}$ is locally identifiable when any point $\boldsymbol{\nu}_0$ has a neighborhood in which the mapping $\boldsymbol{\nu} \mapsto \mathbf{x}(\boldsymbol{\nu})$ is injective. This property is necessary for existence of a finite CRB, since otherwise the Fisher information matrix (FIM) is singular. In this respect, the CRB indicates how far we are from the local identifiability limit. Global identifiability, in its turn, means that $\boldsymbol{\nu} \mapsto \mathbf{x}(\boldsymbol{\nu})$ is injective over the entire $\mathbb{R}^K$, and is not required for existence of the CRB.

As is well-known, the CPD is inherently subject to permutation and scaling ambiguities. The former arises because the order in which the terms in (1) are summed does not matter. It is easy to see that this fact has no influence[1] over the injectivity of $\mathbf{x}(\boldsymbol{\nu})$. On the other hand, the scaling ambiguity stems from the possibility of jointly rescaling some parameters in such a way that $\mathbf{x}$ remains unaffected. This clearly implies the existence of infinitely many vectors $\boldsymbol{\nu}$ yielding the same $\mathbf{x}(\boldsymbol{\nu})$, which thus precludes (12) from being (locally or globally) injective.

For suppressing the scaling ambiguity, degrees of freedom "in excess" must be eliminated. In particular, in a SCPD this aspect depends on the structure of the factors. If, *e.g.*, a factor $\mathbf{A}^{(n)}$ is circulant, imposing $\theta_u^{(n)} = 1$ for some u is sufficient to fix its scaling, due to its structure (this excludes, though, the possibility of having $\theta_u^{(n)} = 0$). Another option would be to impose unit norm for each column, which also eliminates one degree of freedom (as one $\theta_u^{(n)}$ is then fixed as a function of the others). Note that alike measures must be taken for the other factors, so that their scaling is absorbed by $\boldsymbol{\lambda}$.

Henceforth, we shall consider that for each $\boldsymbol{\theta}^{(n)}$, a reduced version $\bar{\boldsymbol{\theta}}^{(n)}$ has been appropriately defined so as to include only the minimal necessary number of degrees of freedom. For

[1] Yet, it must be taken into account when assessing performances through computer simulations, for correctly measuring estimation errors.

TABLE I
CHARACTERISTICS OF $\mathbf{A}^{(n)}$ FOR SOME CLASSES OF MATRICES.

	U_n	$\mathbf{S}_r^{(n)}$, $r \in \{1,\dots,R\}$	$\mathbf{E}_u^{(n)}$, $u \in \{1,\dots,U_n\}$
Unstructured	RI_n	$\left[\mathbf{0}_{I_n \times I_n(r-1)} \;\; \mathbf{I}_{I_n} \;\; \mathbf{0}_{I_n \times I_n(R-r)}\right]$	$\left[\mathbf{E}_u^{(n)}\right]_{i,j} = \begin{cases} 1, & u = (j-1)I_n + i, \\ 0, & \text{otherwise} \end{cases}$
Hankel	$I_n + R - 1$	$\left[\mathbf{0}_{I_n \times (r-1)} \;\; \mathbf{I}_{I_n} \;\; \mathbf{0}_{I_n \times (R-r)}\right]$	$\left[\mathbf{E}_u^{(n)}\right]_{i,j} = \begin{cases} 1, & u = i + j - 1, \\ 0, & \text{otherwise} \end{cases}$
Toeplitz	$I_n + R - 1$	$\left[\mathbf{0}_{I_n \times (r-1)} \;\; \mathbf{e}_{I_n} \;\; \dots \;\; \mathbf{e}_1 \;\; \mathbf{0}_{I_n \times (R-r)}\right]$	$\left[\mathbf{E}_u^{(n)}\right]_{i,j} = \begin{cases} 1, & u = I_n + j - i, \\ 0, & \text{otherwise} \end{cases}$
(Toeplitz) Circulant	I_n	$\mathbf{\Pi}_n^{r-1} = \begin{bmatrix} \mathbf{0} & 1 \\ \mathbf{I}_{I_n - 1} & \mathbf{0} \end{bmatrix}^{r-1}$	$\left[\mathbf{E}_u^{(n)}\right]_{i,j} = \begin{cases} 1, & u = ((i-j) \mod I_n) + 1, \\ 0, & \text{otherwise} \end{cases}$
Banded circulant	$I_n - R + 1$	$\begin{bmatrix} \mathbf{0}_{(r-1)\times U_n} \\ \mathbf{I}_{U_n} \\ \mathbf{0}_{(R-r)\times U_n} \end{bmatrix}$	$\left[\mathbf{E}_u^{(n)}\right]_{i,j} = \begin{cases} 1, & u = i - j + 1, \\ 0, & \text{otherwise} \end{cases}$

notational simplicity, we assume without loss of generality that the V_n first elements of $\boldsymbol{\theta}^{(n)}$ are fixed, and we denote $\bar{U}_n = U_n - V_n$. Note that $\bar{\boldsymbol{\theta}}^{(n)} = \mathbf{B}_n \boldsymbol{\theta}^{(n)}$, where $\mathbf{B}_n \in \mathbb{R}^{\bar{U}_n \times U_n}$ is a selection matrix containing as rows the last $\bar{U}_n$ canonical vectors of $\mathbb{R}^{U_n}$. A vector $\bar{\boldsymbol{\theta}}$ is defined analogously to $\boldsymbol{\theta}$, so that we can introduce the reduced vector of parameters $\boldsymbol{\eta} = [\bar{\boldsymbol{\theta}}^T, \; \boldsymbol{\lambda}^T]^T \in \mathbb{R}^M$, with $M = \bar{U}_1 + \bar{U}_2 + \bar{U}_3 + R$. Its corresponding model will thus be denoted by

$$\mathbf{y}(\boldsymbol{\eta}) = \mathbf{x}(\boldsymbol{\eta}) + \mathbf{n} \quad \in \mathbb{R}^{I_1 I_2 I_3}. \tag{13}$$

III. DETERMINISTIC AND BAYESIAN CRB FOR STRUCTURED CP DECOMPOSITIONS

In this section, the material for the derivation of the deterministic and Bayesian CRB is provided. We first propose, as a preliminary result, several closed-form expressions of the deterministic CRB for structured CP decompositions, *i.e.*, when the parameters of interest are fixed in a deterministic way. In the second part of this section, the deterministic CRB is extended to the Bayesian context *i.e.*, when the parameters of interest are random with known statistical priors.

A. Deterministic CRB

For the derivation of the deterministic CRB, it is important to satisfy the following regularity conditions:

1) The conditional probability distribution function (pdf) of the observation, $p(\mathbf{y}|\boldsymbol{\eta})$, has to be a C^2 function, where C^2 is the space of continuously twice differentiable functions with respect to $\boldsymbol{\eta}$.
2) The Fisher Information Matrix (FIM) defined by

$$[\mathbf{F}(\boldsymbol{\eta})]_{ij} = \mathbb{E}_{\mathbf{y}|\boldsymbol{\eta}} \left\{ -\frac{\partial^2 \log p(\mathbf{y}|\boldsymbol{\eta})}{\partial [\boldsymbol{\eta}]_i \partial [\boldsymbol{\eta}]_j} \right\} \tag{14}$$

has to be positive definite.

Due to the structured CPD model given in (13), the first condition is satisfied by remarking

$$\mathbf{y}|\boldsymbol{\eta} \sim \mathcal{N}(\mathbf{x}(\boldsymbol{\eta}), \sigma^2 \mathbf{I}) \tag{15}$$

where the Normal distribution is a C^2 function. According to the discussion in Section II-C, a convenient choice of $\boldsymbol{\eta}$ allows guaranteeing $\mathbf{F}(\boldsymbol{\eta})$ is nonsingular, which in turn implies condition 2), as $\mathbf{F}(\boldsymbol{\eta})$ is positive semidefinite by definition.

1) Three different factors: This section is dedicated to the derivation of the CRB for the vectorized model $\mathbf{y}(\boldsymbol{\eta})$ in the case when the factors $\mathbf{A}^{(1)}$, $\mathbf{A}^{(2)}$ and $\mathbf{A}^{(3)}$ of $\boldsymbol{\mathcal{X}}$ are distinct.

a) Arbitrary $\boldsymbol{\lambda}$: First, we do not make further assumptions concerning $\boldsymbol{\lambda}$. We follow the ideas introduced in [24]. Let $\mathbf{C}(\boldsymbol{\eta}) \in \mathbb{R}^{M \times M}$ be the CRB matrix, *i.e.*, the inverse of the FIM, $\mathbf{F}(\boldsymbol{\eta})$ [18].

Then, the mean square error (MSE) of any (locally) unbiased estimator, $\hat{\boldsymbol{\eta}}(\mathbf{y})$, admits the following lower bound [18]:

$$\mathrm{MSE}(\hat{\boldsymbol{\eta}}) = \mathbb{E}_{\mathbf{y}}\{\|\boldsymbol{\eta} - \hat{\boldsymbol{\eta}}(\mathbf{y})\|^2\} \geq \; \mathrm{Tr}(\mathbf{C}(\boldsymbol{\eta})) \tag{16}$$

$$= \sum_{n=1}^{3} \sum_{u=1}^{\bar{U}_n} \mathrm{CRB}\left(\bar{\theta}_u^{(n)}\right) + \sum_{r=1}^{R} \mathrm{CRB}(\lambda_r), \tag{17}$$

where $\mathrm{CRB}\left(\bar{\theta}_u^{(n)}\right) = [\mathbf{C}(\boldsymbol{\eta})]_{v,v}$, with $v = \sum_{n'=1}^{n-1} \bar{U}_{n'} + u$ and $\mathrm{CRB}(\lambda_r) = [\mathbf{C}(\boldsymbol{\eta})]_{s,s}$, with $s = \bar{U}_1 + \bar{U}_2 + \bar{U}_3 + r$. Due to the form of (15), $\mathbf{C}(\boldsymbol{\eta})$ can be derived by applying the Slepian-Bangs formula (see [25, Eq. B.3.3]), which yields

$$\mathbf{C}(\boldsymbol{\eta}) = \sigma^2 \left(\mathbf{J}(\boldsymbol{\eta})^T \mathbf{J}(\boldsymbol{\eta})\right)^{-1}, \tag{18}$$

where $\mathbf{J}(\boldsymbol{\eta}) \in \mathbb{R}^{I_1 I_2 I_3 \times M}$ is the Jacobian of $\mathbf{x}(\boldsymbol{\eta})$ with respect to $\boldsymbol{\eta}$. This Jacobian matrix is thus given by

$$\mathbf{J}(\boldsymbol{\eta}) = \left[\underbrace{\mathbf{J}_{\bar{\boldsymbol{\theta}}^{(1)}}(\boldsymbol{\eta}) \, \mathbf{J}_{\bar{\boldsymbol{\theta}}^{(2)}}(\boldsymbol{\eta}) \, \mathbf{J}_{\bar{\boldsymbol{\theta}}^{(3)}}(\boldsymbol{\eta})}_{\mathbf{J}_{\bar{\boldsymbol{\theta}}}} \; \mathbf{J}_{\boldsymbol{\lambda}}(\boldsymbol{\eta}) \right], \tag{19}$$

where $\mathbf{J}_{\bar{\boldsymbol{\theta}}^{(n)}}(\boldsymbol{\eta})$ contains the derivatives of $\mathbf{x}(\boldsymbol{\eta})$ with respect to $\bar{\boldsymbol{\theta}}^{(n)}$, and $\mathbf{J}_{\boldsymbol{\lambda}}(\boldsymbol{\eta})$ is likewise. From (12),

$$\mathbf{J}_{\bar{\boldsymbol{\theta}}^{(1)}}(\boldsymbol{\eta}) = \boldsymbol{\Phi}(\boldsymbol{\lambda}) \left(\boldsymbol{\theta}^{(3)} \boxtimes \boldsymbol{\theta}^{(2)} \boxtimes \mathbf{B}_1^T\right),$$

$$\mathbf{J}_{\bar{\boldsymbol{\theta}}^{(2)}}(\boldsymbol{\eta}) = \boldsymbol{\Phi}(\boldsymbol{\lambda}) \left(\boldsymbol{\theta}^{(3)} \boxtimes \mathbf{B}_2^T \boxtimes \boldsymbol{\theta}^{(1)}\right),$$

$$\mathbf{J}_{\bar{\boldsymbol{\theta}}^{(3)}}(\boldsymbol{\eta}) = \boldsymbol{\Phi}(\boldsymbol{\lambda}) \left(\mathbf{B}_3^T \boxtimes \boldsymbol{\theta}^{(2)} \boxtimes \boldsymbol{\theta}^{(1)}\right),$$

$$\mathbf{J}_{\boldsymbol{\lambda}}(\boldsymbol{\eta}) = \left[\boldsymbol{\Phi}_1 f(\boldsymbol{\theta}), \dots, \boldsymbol{\Phi}_R f(\boldsymbol{\theta})\right].$$

For simplicity of notation, we shall omit the argument of $\mathbf{J}(\boldsymbol{\eta})$ whenever convenient.

In order to identify the contributions of $\bar{\boldsymbol{\theta}}$ and $\boldsymbol{\lambda}$ in $\mathrm{CRB}\left(\bar{\theta}_u^{(n)}\right)$ and $\mathrm{CRB}(\lambda_r)$, we propose to extend the results presented in [1] by using oblique projection. This is the

purpose of the following proposition. The oblique projection whose range is $\langle \mathbf{A} \rangle$ and whose null space contains $\langle \mathbf{B} \rangle$ is denoted by $\mathbf{E}_{\mathbf{AB}}$ (see [26] for details).

Proposition 3 *The closed-form expression of the lower bound on the MSE is given by* (17)*, where the CRB for the uth element of $\bar{\boldsymbol{\theta}}^{(n)}$ is given by*

$$\mathrm{CRB}\left(\bar{\theta}_u^{(n)}\right) = \frac{\sigma^2}{\left\| \left(\mathbf{I} - \mathbf{E}_{\mathbf{G}_{u,n}\mathbf{J}_{\boldsymbol{\lambda}}} - \mathbf{E}_{\mathbf{J}_{\boldsymbol{\lambda}}\mathbf{G}_{u,n}}\right) \mathbf{g}_{u,n} \right\|^2} \quad (20)$$

where $\mathbf{g}_{u,n}$ is the column of $\mathbf{J}_{\bar{\boldsymbol{\theta}}}$ containing the derivatives w.r.t. $\bar{\theta}_u^{(n)}$ and $\mathbf{G}_{u,n}$ contains all the other columns of $\mathbf{J}_{\bar{\boldsymbol{\theta}}}$ (in any order). Similarly, the CRB for the rth element of $\boldsymbol{\lambda}$ is

$$\mathrm{CRB}(\lambda_r) = \frac{\sigma^2}{\left\| \left(\mathbf{I} - \mathbf{E}_{\mathbf{D}_r\mathbf{J}_{\bar{\boldsymbol{\theta}}}} - \mathbf{E}_{\mathbf{J}_{\bar{\boldsymbol{\theta}}}\mathbf{D}_r}\right) \mathbf{d}_r \right\|^2}, \quad (21)$$

where $\mathbf{d}_r$ is the column of $\mathbf{J}_{\boldsymbol{\lambda}}$ associated with λ_r, while $\mathbf{D}_r$ holds all other columns (in any order).

Proof: We focus on the derivation of (20), since that of (21) is similar. First, define any permutation matrix $\boldsymbol{\Pi}_u^{(n)} \in \mathbb{R}^{M\times M}$ such that $\mathbf{J}(\boldsymbol{\eta})\boldsymbol{\Pi}_u^{(n)} = \begin{bmatrix}\mathbf{g}_{u,n} & \mathbf{G}_{u,n} & \mathbf{J}_{\boldsymbol{\lambda}}(\boldsymbol{\eta})\end{bmatrix}$, where $\mathbf{g}_{u,n}$ and $\mathbf{G}_{u,n}$ are as described above. Then, it is easily verified that the matrix $\overline{\mathbf{C}}(\boldsymbol{\eta}) = \left(\boldsymbol{\Pi}_u^{(n)}\right)^T \mathbf{C}(\boldsymbol{\eta})\boldsymbol{\Pi}_u^{(n)}$ is such that $[\mathbf{C}(\boldsymbol{\eta})]_{u,u} = [\overline{\mathbf{C}}(\boldsymbol{\eta})]_{1,1}$. In addition,

$$\left(\boldsymbol{\Pi}_u^{(n)}\right)^T \mathbf{C}(\boldsymbol{\eta})\boldsymbol{\Pi}_u^{(n)} = \sigma^2 \left(\left(\mathbf{J}\boldsymbol{\Pi}_u^{(n)}\right)^T \left(\mathbf{J}\boldsymbol{\Pi}_u^{(n)}\right)\right)^{-1} \quad (22)$$

$$= \sigma^2 \left(\mathbf{J}_{u,n}^T \mathbf{J}_{u,n}\right)^{-1}, \quad (23)$$

where we denote $\mathbf{J}_{u,n} = \mathbf{J}\boldsymbol{\Pi}_u^{(n)}$. Now, defining $\mathbf{H}_{u,n} \triangleq \begin{bmatrix}\mathbf{G}_{u,n} & \mathbf{J}_{\boldsymbol{\lambda}}(\boldsymbol{\eta})\end{bmatrix}$ we can write

$$\mathbf{J}_{u,n}^T \mathbf{J}_{u,n} = \begin{bmatrix} \|\mathbf{g}_{u,n}\|^2 & \mathbf{g}_{u,n}^T \mathbf{H}_{u,n} \\ \mathbf{H}_{u,n}^T \mathbf{g}_{u,n} & \mathbf{H}_{u,n}^T \mathbf{H}_{u,n} \end{bmatrix}. \quad (24)$$

Hence, $[\overline{\mathbf{C}}(\boldsymbol{\eta})]_{1,1} = [\mathbf{C}(\boldsymbol{\eta})]_{u,u}$ can be calculated with the block matrix inversion formula [27, Sec. 9.1.3], which yields

$$[\mathbf{C}(\boldsymbol{\eta})]_{u,u} = [\overline{\mathbf{C}}(\boldsymbol{\eta})]_{1,1} = \frac{\sigma^2}{\|\mathbf{P}^{\perp}_{\mathbf{H}_{u,n}} \mathbf{g}_{u,n}\|^2}, \quad (25)$$

where $\mathbf{P}^{\perp}_{\mathbf{H}_{u,n}} = \mathbf{I}_{I_1 I_2 I_3} - \mathbf{P}_{\mathbf{H}_{u,n}}$ in which $\mathbf{P}_{\mathbf{H}_{u,n}} = \mathbf{H}_{u,n}\mathbf{H}_{u,n}^{\dagger}$ is the orthogonal projector onto $\langle \mathbf{H}_{u,n} \rangle$. The result (20) is thus obtained by noting that $\mathbf{P}_{\mathbf{H}_{u,n}}$ can be rewritten as $\mathbf{P}_{\mathbf{H}_{u,n}} = \mathbf{E}_{\mathbf{G}_{u,n}\mathbf{J}_{\boldsymbol{\lambda}}} + \mathbf{E}_{\mathbf{J}_{\boldsymbol{\lambda}}\mathbf{G}_{u,n}}$ [26]. ■

b) $\boldsymbol{\lambda}$ with identical elements: We suppose now that $\lambda_1 = \cdots = \lambda_R = \lambda$. Actually, in this case λ does not need to be estimated, since we can assume $\lambda = 1$ without loss of generality,[2] due to the scaling ambiguity. The vector which contains the parameters of interest is then $\boldsymbol{\eta} = \bar{\boldsymbol{\theta}}$. Thanks to expression (15), the Slepian-Bangs formula then yields $\mathbf{C}(\boldsymbol{\eta}) = \sigma^2 \left(\mathbf{J}_{\bar{\boldsymbol{\theta}}}^T \mathbf{J}_{\bar{\boldsymbol{\theta}}}\right)^{-1}$. This leads to the following result.

[2]In this case, to preserve the generality of the model, only two factors should have the scaling of their columns fixed.

Proposition 4 *The closed-form expression of the lower bound on the MSE is given by*

$$\mathrm{MSE}(\hat{\boldsymbol{\eta}}) \geq \mathrm{Tr}(\mathbf{C}(\boldsymbol{\eta})) = \sum_{n=1}^{3} \sum_{u=1}^{\bar{U}_n} \mathrm{CRB}\left(\bar{\theta}_u^{(n)}\right) \quad (26)$$

where the CRB for the uth element of $\bar{\boldsymbol{\theta}}^{(n)}$ is given by

$$\mathrm{CRB}(\bar{\theta}_u^{(n)}) = \frac{\sigma^2}{\|\mathbf{P}^{\perp}_{\mathbf{G}_{u,n}} \mathbf{g}_{u,n}\|^2}, \quad (27)$$

where $\mathbf{g}_{u,n}$ is the column of $\mathbf{J}_{\bar{\boldsymbol{\theta}}}$ associated with $\bar{\theta}_u^{(n)}$, $\mathbf{G}_{u,n}$ contains the other columns of $\mathbf{J}_{\bar{\boldsymbol{\theta}}}$ (in any order), and $\mathbf{P}^{\perp}_{\mathbf{G}_{u,n}} = \mathbf{I}_{I_1 I_2 I_3} - \mathbf{G}_{u,n}\mathbf{G}_{u,n}^{\dagger}$.

Proof: The proof is similar to that of Proposition 3, but now taking into account the absence of the block $\mathbf{J}_{\boldsymbol{\lambda}}$ in $\mathbf{J}$. ■

2) Identical factors: We consider now the case of partial or full symmetry, *i.e.*, where two or three factors are identical. It turns out that the previous formulas are still valid. Only the Jacobian matrix needs to be derived accordingly, as follows.

First, consider the case of two identical factors. Without loss of generality, we assume $\boldsymbol{\theta}^{(2)} = \boldsymbol{\theta}^{(3)}$. The parameter vector is thus written as $\boldsymbol{\eta} = [\bar{\boldsymbol{\theta}}^{(1)}, \bar{\boldsymbol{\theta}}^{(2)}, \boldsymbol{\lambda}]^T$. The Jacobian is then given by $\mathbf{J}(\boldsymbol{\eta}) = \begin{bmatrix}\mathbf{J}_{\bar{\boldsymbol{\theta}}^{(1)}}(\boldsymbol{\eta}) & \mathbf{J}_{\bar{\boldsymbol{\theta}}^{(2)}}(\boldsymbol{\eta}) & \mathbf{J}_{\boldsymbol{\lambda}}(\boldsymbol{\eta})\end{bmatrix}$, with

$$\mathbf{J}_{\bar{\boldsymbol{\theta}}^{(1)}}(\boldsymbol{\eta}) = \boldsymbol{\Phi}(\boldsymbol{\lambda}) \left(\boldsymbol{\theta}^{(2)} \boxtimes \boldsymbol{\theta}^{(2)} \boxtimes \mathbf{B}_1^T\right), \quad (28)$$

$$\mathbf{J}_{\bar{\boldsymbol{\theta}}^{(2)}}(\boldsymbol{\eta}) = \boldsymbol{\Phi}(\boldsymbol{\lambda}) \left[\mathbf{w}_{V_2+1}, \ldots, \mathbf{w}_{U_2}\right], \quad (29)$$

$$\mathbf{J}_{\boldsymbol{\lambda}}(\boldsymbol{\eta}) = \left[\boldsymbol{\Phi}_1 f(\boldsymbol{\theta}), \ldots, \boldsymbol{\Phi}_R f(\boldsymbol{\theta})\right], \quad (30)$$

where $\mathbf{w}_u = \left(\mathbf{e}_u \boxtimes \boldsymbol{\theta}^{(2)} \boxtimes \boldsymbol{\theta}^{(1)}\right) + \left(\boldsymbol{\theta}^{(2)} \boxtimes \mathbf{e}_u \boxtimes \boldsymbol{\theta}^{(1)}\right)$, $f(\boldsymbol{\theta}) = \left(\boldsymbol{\theta}^{(2)} \boxtimes \boldsymbol{\theta}^{(2)} \boxtimes \boldsymbol{\theta}^{(1)}\right)$ and V_2 is as defined in Sec. II-C.

When $\boldsymbol{\theta}^{(1)} = \boldsymbol{\theta}^{(2)} = \boldsymbol{\theta}^{(3)} = \boldsymbol{\theta}$ (and thus $\bar{U}_n = \bar{U}$, $V_n = V$ and $U_n = U$ for all n), the Jacobian satisfies $\mathbf{J}(\boldsymbol{\eta}) = \begin{bmatrix}\mathbf{J}_{\bar{\boldsymbol{\theta}}}(\boldsymbol{\eta}) & \mathbf{J}_{\boldsymbol{\lambda}}(\boldsymbol{\eta})\end{bmatrix}$, with blocks given by

$$\mathbf{J}_{\bar{\boldsymbol{\theta}}}(\boldsymbol{\eta}) = \boldsymbol{\Phi}(\boldsymbol{\lambda}) \left[\mathbf{w}_{V+1}, \ldots, \mathbf{w}_U\right], \quad (31)$$

$$\mathbf{J}_{\boldsymbol{\lambda}}(\boldsymbol{\eta}) = \left[\boldsymbol{\Phi}_1 \left(\boldsymbol{\theta} \boxtimes \boldsymbol{\theta} \boxtimes \boldsymbol{\theta}\right), \ldots, \boldsymbol{\Phi}_R \left(\boldsymbol{\theta} \boxtimes \boldsymbol{\theta} \boxtimes \boldsymbol{\theta}\right)\right], \quad (32)$$

now with $\mathbf{w}_u = (\mathbf{e}_u \boxtimes \boldsymbol{\theta} \boxtimes \boldsymbol{\theta}) + (\boldsymbol{\theta} \boxtimes \mathbf{e}_u \boxtimes \boldsymbol{\theta}) + (\boldsymbol{\theta} \boxtimes \boldsymbol{\theta} \boxtimes \mathbf{e}_u)$.

3) Symmetric model: When the three factors $\mathbf{A}^{(n)}$ are identical, $\boldsymbol{\mathcal{X}}$ is a symmetric third-order tensor. In practice, it may happen that the noisy observed tensor, $\boldsymbol{\mathcal{Y}}$, is also symmetric, because elements which are identical due to symmetry are not repeatedly estimated or observed. This may apply when $\boldsymbol{\mathcal{X}}$ represents a symmetric quantity as, *e.g.*, a symmetric Volterra kernel. As a consequence, the noise tensor $\boldsymbol{\mathcal{N}}$ is also symmetric, and the preceding results are no longer valid.

Nevertheless, it is easy to adapt our model to such a situation. Indeed, it suffices to introduce a selection matrix $\boldsymbol{\Psi} \in \mathbb{R}^{L\times I^3}$ ($I_1 = I_2 = I_3 = I$) containing as rows every Kronecker product of canonical basis vectors $\mathbf{e}_{i_3}^T \boxtimes \mathbf{e}_{i_2}^T \boxtimes \mathbf{e}_{i_1}^T$ such that $(i_1, i_2, i_3) \in D = \{(i_1, i_2, i_3) : 1 \leq i_1 \leq i_2 \leq i_3 \leq I\}$, in such a way that the product $\boldsymbol{\Psi}\,\mathrm{vec}(\boldsymbol{\mathcal{Y}})$ no longer contains redundant components due to symmetry. Note that $L = |D| = \binom{I^3+3-1}{3}$. Then, we redefine the vector model as

$$\mathbf{y} \triangleq \boldsymbol{\Psi}\,\mathrm{vec}(\boldsymbol{\mathcal{Y}}) = \boldsymbol{\Psi}\boldsymbol{\Phi}(\boldsymbol{\lambda}) f(\boldsymbol{\theta}) + \mathbf{n}, \quad (33)$$

where now $\mathbf{n} = \boldsymbol{\Psi}\,\mathrm{vec}(\boldsymbol{\mathcal{N}})$. As $\mathbf{n}$ is still zero-mean Gaussian i.i.d., the Slepian-Bangs formula remains valid. The Jacobian,

however, must now be pre-multiplied by a factor $\mathbf{\Psi}$. It should be noted that the particular ordering of the rows of $\mathbf{\Psi}$ is irrelevant in the above reasoning.

B. Extension to Bayesian framework

There exist two ways to extend the deterministic CRB to the Bayesian framework, *i.e.*, for random parameters $\boldsymbol{\eta}$ with known prior $p(\boldsymbol{\eta})$, in the low noise variance regime. The first one is well-known under the name of Van Trees' Bayesian CRB (BCRB) and is derived in [22]. The second one is called the Expected CRB (ECRB) and is based on the Bayesian-deterministic connection [21,22]. Even if both approaches share the same mathematical principle, *i.e.*, exploit the Cauchy-Schwarz inequality, several important differences have to be listed. Below, after having recalled the expressions of these two bounds, we discuss their regularity conditions and tightness for SCPD estimation.

1) Definition of the two bounds:

a) Van Trees' Bayesian CRB: The BCRB is given by

$$\mathrm{BCRB} = \mathrm{Tr}\left[\mathbf{B}^{-1}(\boldsymbol{\eta})\right] \tag{34}$$

where the Bayesian Information Matrix (BIM) is given by

$$\mathbf{B}(\boldsymbol{\eta}) = \mathbb{E}_{\boldsymbol{\eta}}\left\{\mathbf{F}(\boldsymbol{\eta})\right\} + \mathbf{B}_{\mathrm{prior}}, \tag{35}$$

in which the FIM is given in (14) and the prior matrix is

$$[\mathbf{B}_{\mathrm{prior}}]_{ij} = \mathbb{E}_{\boldsymbol{\eta}}\left\{-\frac{\partial^2 \log p(\boldsymbol{\eta})}{\partial[\boldsymbol{\eta}]_i \partial[\boldsymbol{\eta}]_j}\right\}. \tag{36}$$

b) Expected CRB (ECRB): Recall that the BMSE is defined according to

$$\mathrm{BMSE} = \mathbb{E}_{\boldsymbol{\eta}}\left\{\mathrm{MSE}(\hat{\boldsymbol{\eta}};\boldsymbol{\eta})\right\}, \tag{37}$$

where the MSE conditioned on $\boldsymbol{\eta}$ is defined as

$$\mathrm{MSE}(\hat{\boldsymbol{\eta}};\boldsymbol{\eta}) = \mathbb{E}_{\mathbf{y}|\boldsymbol{\eta}}\left\{||\hat{\boldsymbol{\eta}}(\mathbf{y}) - \boldsymbol{\eta}||^2\right\}. \tag{38}$$

So, a (deterministic) lower bound verifies the inequality

$$\mathrm{MSE}(\hat{\boldsymbol{\eta}};\boldsymbol{\eta}) \geq \mathrm{Tr}\left[\mathbf{F}^{-1}(\boldsymbol{\eta})\right], \tag{39}$$

where $\mathbf{F}(\boldsymbol{\eta})$ is given by expression (14). Finally, the ECRB takes the expression

$$\mathrm{ECRB} = \mathrm{Tr}\left[\mathbb{E}_{\boldsymbol{\eta}}\left\{\mathbf{F}^{-1}(\boldsymbol{\eta})\right\}\right] \tag{40}$$

and bounds (37) as $\mathrm{BMSE} \geq \mathrm{ECRB}$.

2) Regularity conditions: The regularity condition for the BCRB (resp., ECRB) is that log-joint pdf $\log p(\mathbf{y},\boldsymbol{\eta}) \in C^2$ (resp., log-conditional pdf $\log p(\mathbf{y}|\boldsymbol{\eta}) \in C^2$). As $\log p(\mathbf{y},\boldsymbol{\eta}) = \log p(\mathbf{y}|\boldsymbol{\eta}) + \log p(\boldsymbol{\eta})$, the regularity conditions for the BCRB are $\log p(\mathbf{y}|\boldsymbol{\eta}) \in C^2$ *and* $\log p(\boldsymbol{\eta}) \in C^2$ (see expression (36)). This is clearly stricter than the regularity condition for the ECRB. Indeed, unlike the BCRB, the regularity condition for the ECRB does not involve the prior $p(\boldsymbol{\eta})$. As a consequence, the ECRB allows a wider degree of freedom in the choice of the prior $p(\boldsymbol{\eta})$ than the BCRB. For instance, a uniform prior is not admissible for the BCRB but can be used with the ECRB.

3) Tightness of bounds and final expressions: In the following result, the above discussed lower bounds are compared.

Result 5 *For any statistical priors $p(\boldsymbol{\eta})$, the ECRB, defined in* (40)*, is a tighter bound of the BMSE than the BCRB, defined in* (34)*.*

Proof: From (35) and Woodbury's identity [27], we have

$$\begin{aligned}\mathbf{B}^{-1}(\boldsymbol{\eta}) = &\left[\mathbb{E}_{\boldsymbol{\eta}}\left\{\mathbf{F}(\boldsymbol{\eta})\right\}\right]^{-1} \\ &- \left[\mathbb{E}_{\boldsymbol{\eta}}\left\{\mathbf{F}(\boldsymbol{\eta})\right\}\right]^{-1}\mathbf{D}^{-1}\left[\mathbb{E}_{\boldsymbol{\eta}}\left\{\mathbf{F}(\boldsymbol{\eta})\right\}\right]^{-1},\end{aligned}$$

where $\mathbf{D} = \left[\mathbb{E}_{\boldsymbol{\eta}}\left\{\mathbf{F}(\boldsymbol{\eta})\right\}\right]^{-1} + \mathbf{B}_{\mathrm{prior}}^{-1}$. As both $\mathbb{E}_{\boldsymbol{\eta}}\left\{\mathbf{F}(\boldsymbol{\eta})\right\}$ and $\mathbf{B}_{\mathrm{prior}}$ are positive definite (by hypothesis and by definition, respectively), $\left[\mathbb{E}_{\boldsymbol{\eta}}\left\{\mathbf{F}(\boldsymbol{\eta})\right\}\right]^{-1}\mathbf{D}^{-1}\left[\mathbb{E}_{\boldsymbol{\eta}}\left\{\mathbf{F}(\boldsymbol{\eta})\right\}\right]^{-1}$ is also positive definite. It thus follows that the BCRB verifies

$$\mathrm{BCRB} < \mathrm{Tr}\left(\left[\mathbb{E}_{\boldsymbol{\eta}}\left\{\mathbf{F}(\boldsymbol{\eta})\right\}\right]^{-1}\right). \tag{41}$$

Using Jensen's inequality, it is straightforward to show that

$$\mathrm{Tr}\left(\left[\mathbb{E}_{\boldsymbol{\eta}}\left\{\mathbf{F}(\boldsymbol{\eta})\right\}\right]^{-1}\right) < \mathrm{Tr}\left(\mathbb{E}_{\boldsymbol{\eta}}\left\{\mathbf{F}^{-1}(\boldsymbol{\eta})\right\}\right) = \mathrm{ECRB}. \tag{42}$$

Note that the above inequality is strict. This can be shown in the following manner. Recall that the equality holds if the FIM, $\mathbf{F}(\boldsymbol{\eta})$, is no longer a function of $\boldsymbol{\eta}$. But due to the non-linear structure (with respect to parameters $\boldsymbol{\eta}$) of the SCPD, this never holds for the considered model. In conclusion, using inequalities (41) and (42), we have $\mathrm{BCRB} < \mathrm{ECRB}$. Recall that the BCRB and the ECRB are two lower bounds of the BMSE in the low noise regime. Then there cannot exist an estimator that reaches the BCRB because its variance would be lower than the ECRB. This argument tells us that the latter bound is tighter than the former for any parameter priors. ■

So, based on the above result, we adopt the ECRB for evaluating the performance of the estimation algorithms introduced next. Specifically, the ECRB for independent parameters of interest are given below.

1) For arbitrary $\boldsymbol{\lambda}$, the ECRB for parameters $\bar{\theta}_u^{(n)}$ and λ_r are

$$\mathrm{ECRB}(\bar{\theta}_u^{(n)}) = \mathbb{E}_{\bar{\boldsymbol{\theta}}}\mathbb{E}_{\boldsymbol{\lambda}}\left\{\mathrm{CRB}(\bar{\theta}_u^{(n)})\right\}, \tag{43}$$

$$\mathrm{ECRB}(\lambda_r) = \mathbb{E}_{\bar{\boldsymbol{\theta}}}\mathbb{E}_{\boldsymbol{\lambda}}\left\{\mathrm{CRB}(\lambda_r)\right\} \tag{44}$$

where $\mathrm{CRB}(\bar{\theta}_u^{(n)})$ and $\mathrm{CRB}(\lambda_r)$ are given by (20) and (21), respectively.

2) For $\boldsymbol{\lambda}$ with identical elements, the ECRB for parameters $\bar{\theta}_u^{(n)}$ is given by

$$\mathrm{ECRB}(\bar{\theta}_u^{(n)}) = \mathbb{E}_{\bar{\boldsymbol{\theta}}}\left\{\mathrm{CRB}(\bar{\theta}_u^{(n)})\right\} \tag{45}$$

where $\mathrm{CRB}(\bar{\theta}_u^{(n)})$ is given by (27).

IV. Specialized estimation algorithms

CPD computation algorithms are often subject to facing issues as very slow convergence, which increases computational cost, and early termination, which degrades estimation precision [17,28]. Therefore, in the case of a SCPD, it is desirable to resort to specialized algorithms that exploit the structure of the factors, with the goal of overcoming these difficulties. In the following, we briefly review some existing

TABLE II
THE CALS ALGORITHM.

Inputs: $\mathcal{Y} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$, and initial parameter vectors $\hat{\boldsymbol{\lambda}}_0, \hat{\boldsymbol{\theta}}_0^{(1)}, \hat{\boldsymbol{\theta}}_0^{(2)}, \hat{\boldsymbol{\theta}}_0^{(3)}$.
Outputs: Estimated parameter vectors $\hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\theta}}^{(1)}, \hat{\boldsymbol{\theta}}^{(2)}, \hat{\boldsymbol{\theta}}^{(3)}$.
repeat for $i = 1, 2, \ldots$

$$\begin{cases} \hat{\boldsymbol{\theta}}_i^{(1)} = \left\{ \left[\left(\left(\hat{\mathbf{A}}_{i-1}^{(3)} \odot \hat{\mathbf{A}}_{i-1}^{(2)} \right) \hat{\boldsymbol{\Lambda}}_{i-1} \right) \boxtimes \mathbf{I}_{I_1} \right] \mathbf{E}^{(1)} \right\}^{\dagger} \mathrm{vec}\left(\mathbf{Y}_1\right). \\ \hat{\mathbf{A}}_i^{(1)} = \mathrm{Unvec}\left(\mathbf{E}^{(1)} \hat{\boldsymbol{\theta}}_i^{(1)} \right) \\ \hat{\boldsymbol{\theta}}_i^{(2)} = \left\{ \left[\left(\left(\hat{\mathbf{A}}_{i}^{(1)} \odot \hat{\mathbf{A}}_{i-1}^{(3)} \right) \hat{\boldsymbol{\Lambda}}_{i-1} \right) \boxtimes \mathbf{I}_{I_2} \right] \mathbf{E}^{(2)} \right\}^{\dagger} \mathrm{vec}\left(\mathbf{Y}_2\right). \\ \hat{\mathbf{A}}_i^{(2)} = \mathrm{Unvec}\left(\mathbf{E}^{(2)} \hat{\boldsymbol{\theta}}_i^{(2)} \right) \\ \hat{\boldsymbol{\theta}}_i^{(3)} = \left\{ \left[\left(\left(\hat{\mathbf{A}}_{i}^{(2)} \odot \hat{\mathbf{A}}_{i}^{(1)} \right) \hat{\boldsymbol{\Lambda}}_{i-1} \right) \boxtimes \mathbf{I}_{I_3} \right] \mathbf{E}^{(3)} \right\}^{\dagger} \mathrm{vec}\left(\mathbf{Y}_3\right). \\ \hat{\mathbf{A}}_i^{(3)} = \mathrm{Unvec}\left(\mathbf{E}^{(3)} \hat{\boldsymbol{\theta}}_i^{(3)} \right) \\ \hat{\boldsymbol{\lambda}}_i = \left(\hat{\mathbf{A}}_i^{(3)} \odot \hat{\mathbf{A}}_i^{(2)} \odot \hat{\mathbf{A}}_i^{(1)} \right)^{\dagger} \mathrm{vec}\left(\mathcal{Y}\right), \quad \hat{\boldsymbol{\Lambda}}_i = \mathrm{Diag}(\hat{\boldsymbol{\lambda}}_i) \end{cases}$$

until convergence

TABLE III
THE CALS ALGORITHM IN THE SYMMETRIC CASE.

Inputs: $\mathcal{Y} \in \mathbb{R}^{I \times I \times I}$, and initial parameter vectors $\hat{\boldsymbol{\lambda}}_0, \hat{\boldsymbol{\theta}}_0$.
Outputs: Estimated parameter vectors $\hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\theta}}$.
repeat for $i = 1, 2, \ldots$

$$\begin{cases} \hat{\boldsymbol{\theta}}_i = \left\{ \left[\left(\left(\hat{\mathbf{A}}_{i-1} \odot \hat{\mathbf{A}}_{i-1} \right) \hat{\boldsymbol{\Lambda}}_{i-1} \right) \boxtimes \mathbf{I}_{I} \right] \mathbf{E} \right\}^{\dagger} \mathrm{vec}\left(\mathbf{Y}_1\right). \\ \hat{\mathbf{A}}_i = \mathrm{Unvec}\left(\mathbf{E}\, \mathrm{normalize}(\hat{\boldsymbol{\theta}}_i) \right) \\ \hat{\boldsymbol{\lambda}}_i = \left(\hat{\mathbf{A}}_i \odot \hat{\mathbf{A}}_i \odot \hat{\mathbf{A}}_i \right)^{\dagger} \mathrm{vec}\left(\mathcal{Y}\right), \quad \hat{\boldsymbol{\Lambda}}_i = \mathrm{Diag}(\hat{\boldsymbol{\lambda}}_i) \end{cases}$$

until convergence

SCPD algorithms, keeping their presentation as general as possible with respect to the structure of $\mathcal{X}$. Along their description, we will denote the data tensor which we wish to decompose by $\mathcal{Y}$, and its mode-n unfolding by $\mathbf{Y}_n$.

A. Constrained ALS

Recall that the alternating least squares (ALS) algorithm [9,28] tackles the least-squares problem

$$\min_{\mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \mathbf{A}^{(3)}} \left\| \mathcal{Y} - \mathcal{I} \times_1 \mathbf{A}^{(1)} \times_2 \mathbf{A}^{(2)} \times_3 \mathbf{A}^{(3)} \right\|_F^2 \tag{46}$$

by applying alternating updates to each factor $\mathbf{A}^{(n)}$ separately, given current estimates of the other factors. Denoting such estimates by $\hat{\mathbf{A}}^{(n)}$, the alternating updates have the form

$$\hat{\mathbf{A}}^{(n)} = \mathbf{Y}_n \left(\mathbf{W}^{(n)} \right)^{\dagger}, \tag{47}$$

where $\mathbf{W}^{(n)} = \left(\hat{\mathbf{A}}^{(n_1)} \odot \hat{\mathbf{A}}^{(n_2)} \right)^T$, with $n_q = (n - q - 1 \mod 3) + 1$, $n \in \{1,2,3\}$ and $q \in \{1,2\}$, is assumed to have full row rank. Therefore, a straightforward computational procedure for computing a SCPD is to take (7) into account when estimating the nth factor with ALS, via estimation of $\boldsymbol{\theta}^{(n)}$ in the least-squares sense using (3) and (8), *i.e.*,

$$\hat{\boldsymbol{\theta}}^{(n)} = \left\{ \left[\left(\mathrm{Diag}(\boldsymbol{\lambda}) \mathbf{W}^{(n)} \right)^T \boxtimes \mathbf{I}_{I_n} \right] \mathbf{E}^{(n)} \right\}^{\dagger} \mathrm{vec}\left(\mathbf{Y}_n\right).$$

Similarly, the vector $\boldsymbol{\lambda}$ can also be estimated in the least-squares sense using (6), which yields

$$\hat{\boldsymbol{\lambda}} = \left(\hat{\mathbf{A}}^{(3)} \odot \hat{\mathbf{A}}^{(2)} \odot \hat{\mathbf{A}}^{(1)} \right)^{\dagger} \mathrm{vec}\left(\mathcal{Y}\right), \tag{48}$$

provided $\hat{\mathbf{A}}^{(3)} \odot \hat{\mathbf{A}}^{(2)} \odot \hat{\mathbf{A}}^{(1)}$ has full column rank.

This general scheme—which preserves the property of guaranteed non-increasing cost function values of ALS [28]—can be specialized for any conceivable structure in the sense of Definition 1. For concreteness, a pseudocode of such a constrained ALS (CALS) algorithm[3] is given in Table II. To suppress the scaling ambiguity, a normalization step can be performed after convergence. This can be done in several ways, as discussed in Sec. II-C. When $\lambda_1 = \cdots = \lambda_R$, the scaling can be *e.g.* absorbed by the third factor $\hat{\boldsymbol{\theta}}^{(3)}$, as $\boldsymbol{\lambda}$ is no longer estimated (cf. Sec. III-A1b).

It should be noted that it is also easy to take the assumption of (partial) symmetry (cf. Sec. III-A2) into account, by estimating all identical factors only once per iteration. This is illustrated by the algorithm given in Table III, where the three factors are assumed to be identical. However, the cost function is no longer guaranteed to decrease or stay the same along the iterations. Consequently, the symmetric CALS is more sensitive to its initialization and prone to suffering from convergence problems than non-symmetric CALS. Also, normalization at each iteration is advised for numerical stability.

When estimating a CPD with the ALS scheme, an often used convergence criterion is based on the variation of the (normalized) cost function between two consecutive iterates [17,28]. More precisely, given values of $\mathrm{NMSE}_i = \left\| \mathcal{Y} - \mathcal{I} \times_1 \hat{\mathbf{A}}_i^{(1)} \times_2 \hat{\mathbf{A}}_i^{(2)} \times_3 \hat{\mathbf{A}}_i^{(3)} \mathrm{Diag}(\hat{\boldsymbol{\lambda}}_i) \right\|_F^2 \|\mathcal{Y}\|_F^{-2}$ for two consecutive iterations, one verifies whether the condition $|\mathrm{NMSE}_i - \mathrm{NMSE}_{i-1}| < \epsilon$ holds for a given tolerance level ϵ.

B. Subspace-based closed form solutions

We briefly describe in this section a low-complexity algorithm, which allows to find a good approximate solution. It follows the same lines as in [14]. The approximation is based on the following: it is assumed that there is no noise (we apply properties of $\mathcal{X}$ to $\mathcal{Y}$) and that model (4) is exact, where one or two matrices are structured. These assumptions permit to obtain a solution within a finite number of operations, as now explained. Our description holds valid in the complex case.

1) One structured matrix: Suppose that matrix $\mathbf{A}^{(3)}$ is structured. Because of scaling indeterminacy, we may decide to impose $\theta_1^{(3)} = 1$ in (7) so that

$$\mathbf{A}^{(3)} = \mathbf{E}_1^{(3)} + \sum_{u=2}^{U_3} \theta_u^{(3)} \mathbf{E}_u^{(3)}.$$

From (3), the third matrix unfolding of tensor $\mathcal{X}$ satisfies $\mathbf{X}_3 = \mathbf{A}^{(3)} \mathrm{Diag}(\boldsymbol{\lambda})(\mathbf{A}^{(2)} \odot \mathbf{A}^{(1)})^T$. Let $\mathbf{X}_3 = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^H$ denote the SVD of $\mathbf{X}_3$. Then there exists an $R \times R$ invertible matrix $\mathbf{M}$ such that $\mathbf{U}\mathbf{M} = \mathbf{A}^{(3)}$, and $\mathbf{M}^{-1}\boldsymbol{\Sigma}\mathbf{V}^H = \mathrm{Diag}(\boldsymbol{\lambda})(\mathbf{A}^{(2)} \odot \mathbf{A}^{(1)})^T$. The idea is to solve the former for $\mathbf{M}$ and $\boldsymbol{\theta}^{(3)}$, which is possible if the number of unknowns, $U_3 - 1 + R^2$, is smaller

[3] In [2,16], the same acronym is used for "circulant-constrained ALS". Here, we employ it in a more general sense.

than the number of equations, I_3R. This can be done by just solving the linear system:

$$\left[-\mathbf{I}_R \boxtimes \mathbf{U}, \operatorname{vec}\left(\mathbf{E}_2^{(3)}\right), \ldots, \operatorname{vec}\left(\mathbf{E}_{U_3}^{(3)}\right)\right] \cdot \begin{pmatrix} \operatorname{vec}(\mathbf{M}) \\ \bar{\boldsymbol{\theta}}^{(3)} \end{pmatrix} = -\operatorname{vec}\left(\mathbf{E}_1^{(3)}\right).$$

Once $\mathbf{M}$ is obtained, the product $\mathbf{M}^{-1}\boldsymbol{\Sigma}\mathbf{V}^H$ yields $\operatorname{Diag}(\boldsymbol{\lambda})(\mathbf{A}^{(2)} \odot \mathbf{A}^{(1)})^T$. Hence we have matrix $\boldsymbol{\Omega} = \mathbf{A}^{(2)} \odot \mathbf{A}^{(1)}$ up to scaling. If $\boldsymbol{\omega}_r$ denotes the rth column of matrix $\boldsymbol{\Omega}$, the rth column of matrices $\mathbf{A}^{(1)}$ and $\mathbf{A}^{(2)}$ can eventually be obtained by computing the best rank-one approximation of the $I_1 \times I_2$ matrix $\operatorname{Unvec}(\boldsymbol{\omega}_r)$. The appropriate normalization of columns of $\mathbf{A}^{(1)}$ and $\mathbf{A}^{(2)}$ finally gives $\boldsymbol{\lambda}$.

2) Two structured matrices: Now let's turn to the case where two factor matrices are structured. Assume that $\mathbf{A}^{(1)}$ and $\mathbf{A}^{(2)}$ are structured and independently follow model (7) with parameter vectors $\boldsymbol{\theta}^{(1)}$ and $\boldsymbol{\theta}^{(2)}$, respectively. We do not assume any structure for matrix $\mathbf{A}^{(3)}$. As above, we still have $\mathbf{X}_3 = \mathbf{A}^{(3)} \operatorname{Diag}(\boldsymbol{\lambda}) \left(\mathbf{A}^{(2)} \odot \mathbf{A}^{(1)}\right)^T = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^H$. Then there exists a $R \times R$ invertible matrix $\mathbf{N}$ such that $\mathbf{N}\mathbf{V}^H = (\mathbf{A}^{(2)} \odot \mathbf{A}^{(1)})^T$ and $\mathbf{U}\boldsymbol{\Sigma}\mathbf{N}^{-1} = \mathbf{A}^{(3)} \operatorname{Diag}(\boldsymbol{\lambda})$. We then end up with the following linear system to solve:

$$\left[-\mathbf{I}_R \boxtimes \mathbf{V}^*, \operatorname{vec}\left(\mathbf{E}_1^{(2)} \odot \mathbf{E}_1^{(1)}\right), \ldots, \operatorname{vec}\left(\mathbf{E}_{U_2}^{(2)} \odot \mathbf{E}_{U_1}^{(1)}\right)\right] \cdot \begin{pmatrix} \operatorname{vec}(\mathbf{N}^T) \\ \boldsymbol{\zeta} \end{pmatrix} = \mathbf{0} \quad (49)$$

where $\boldsymbol{\zeta} = \boldsymbol{\theta}^{(2)} \boxtimes \boldsymbol{\theta}^{(1)}$. Ignoring the structure of unknown $\boldsymbol{\zeta}$, the system can be solved in the total LS sense by computing the right singular vector associated with the smallest singular value of the matrix between brackets. This yields $\operatorname{vec}(\mathbf{N}^T)$ and $\boldsymbol{\zeta}$ up to scaling.

Once $\mathbf{N}$ and $\boldsymbol{\zeta}$ are available, $\boldsymbol{\theta}^{(1)}$ and $\boldsymbol{\theta}^{(2)}$ can be obtained (up to scaling) by computing the best rank-1 approximate of $\operatorname{Unvec}(\boldsymbol{\zeta})$ since $\operatorname{Unvec}(\boldsymbol{\zeta}) \approx \boldsymbol{\theta}^{(1)}\boldsymbol{\theta}^{(2)\,T}$ in the absence of noise. This will yield matrices $\mathbf{A}^{(1)}$ and $\mathbf{A}^{(2)}$. Finally, matrix $\mathbf{A}^{(3)}$ may be computed as $\mathbf{U}\boldsymbol{\Sigma}\mathbf{N}^{-1}$, with $\boldsymbol{\lambda}$ being subsequently estimated using (6) after proper normalization of $\mathbf{A}^{(1)}$, $\mathbf{A}^{(2)}$ and $\mathbf{A}^{(3)}$.

C. Algebraic solution for the estimation of circulant factors

In [16], it was shown that if a hypercubical Nth-order tensor admits a SCPD containing only circulant factors (called circulant-constrained CPD (CCPD)), then they can be computed by solving a system of Nth-order homogeneous monomial equations. Central to this derivation is the fact that all square circulant matrices are diagonalized by the Fourier matrix. Below, we briefly recall this approach for $N = 3$.

Let $\mathcal{X} = \mathcal{I} \times_1 \mathbf{C}^{(1)} \times_2 \mathbf{C}^{(2)} \times_3 \mathbf{C}^{(3)}$ be such that all $\mathbf{C}^{(n)} \in \mathbb{R}^{I \times R}$ are circulant and $I \geq R$. Recall that any *square* circulant matrix $\mathbf{C} \in \mathbb{R}^{I \times I}$ can be written as $\mathbf{C} = \mathbf{F}\boldsymbol{\Lambda}\mathbf{F}^H$, where $\boldsymbol{\Lambda}$ is a diagonal matrix containing the eigenvalues of $\mathbf{C}$ and $\mathbf{F} \in \mathbb{C}^{I \times I}$ is the Fourier matrix defined by $[\mathbf{F}]_{i_1,i_2} = \frac{1}{\sqrt{I}} \exp((i_1 - 1)(i_2 - 1)\frac{j2\pi}{I})$, with $j = \sqrt{-1}$. Thus, we can write $\mathbf{C}^{(n)} = \mathbf{F}\boldsymbol{\Lambda}^{(n)}\mathbf{F}_R^H$, where $\mathbf{F}_R \in \mathbb{C}^{R \times I}$ contains only the first R rows of $\mathbf{F}$ and $\boldsymbol{\Lambda}^{(n)}$ contains the eigenvalues of $\breve{\mathbf{C}}^{(n)} = \mathbf{F}\boldsymbol{\Lambda}^{(n)}\mathbf{F}^H \in \mathbb{R}^{I \times I}$. Hence, the multidimensional discrete Fourier transform (MDFT) of $\mathcal{X}$ yields

$$\mathcal{W} = \mathcal{X} \times_1 \mathbf{F}^H \times_2 \mathbf{F}^H \times_3 \mathbf{F}^H \quad (50)$$

$$= \mathcal{I} \times_1 \boldsymbol{\Lambda}^{(1)}\mathbf{F}_R^H \times_2 \boldsymbol{\Lambda}^{(2)}\mathbf{F}_R^H \times_3 \boldsymbol{\Lambda}^{(3)}\mathbf{F}_R^H. \quad (51)$$

Denoting $[\boldsymbol{\Lambda}^{(n)}]_{i,i} = \lambda_i^{(n)}$, we have in scalar form

$$w_{i_1,i_2,i_3} = \frac{1}{I^{\frac{3}{2}}} \lambda_{i_1}^{(1)} \lambda_{i_2}^{(2)} \lambda_{i_3}^{(3)} \sum_{r=1}^{R} e^{\frac{-j2\pi}{I}(r-1)(i_1+i_2+i_3-3)}. \quad (52)$$

Hence, assuming R is known, one can compute the MDFT of $\mathcal{X}$ and then solve the resulting system of equations (52) for the eigenvalues $\lambda_i^{(n)}$, from which each $\mathbf{C}^{(n)}$ is then reconstructed.

Necessary and sufficient conditions for the non-nullity of the sum of exponentials in (52) were given in [16]. The remaining equations admit, in general, infinitely many solutions, each one associated with a CCPD of $\mathcal{X}$. This phenomenon is of course related with the permutation and scaling indeterminacies. For the interested reader, a characterization of this relation is also developed in [16].

The computation of accurate solutions for (52) is challenging in general, especially in the presence of noise. Nevertheless, a solution obtained with a straightforward *ad-hoc* procedure can be useful in practice for initializing more sophisticated iterative algorithms. Furthermore, it is computationally cheap, as $\mathcal{W}$ can be computed with a multidimensional FFT algorithm [29]. This approach is called *ad-hoc* algebraic solution (AAS) [16]. Consider for example the case with $I = 4$ and $R = 3$, in which all $4^3 = 64$ equations are generally exploitable. Suppose $\lambda_1^{(1)} \neq 0$ and, to suppress the scaling ambiguity, assume $\lambda_1^{(2)} = \lambda_1^{(3)} = 1$. Then, one can compute an estimate from

$$\begin{array}{llll} (i) & \lambda_1^{(1)} = \frac{8}{3} w_{1,1,1} & (vi) & \lambda_3^{(2)} = 8\frac{w_{1,3,1}}{\lambda_1^{(1)}} \\ (ii) & \lambda_2^{(1)} = 8j w_{2,1,1} & (vii) & \lambda_3^{(3)} = 8\frac{w_{1,1,3}}{\lambda_1^{(1)}} \\ (iii) & \lambda_2^{(2)} = 8j\frac{w_{1,2,1}}{\lambda_1^{(1)}} & (viii) & \lambda_4^{(1)} = -8j w_{4,1,1} \\ (iv) & \lambda_2^{(3)} = 8j\frac{w_{1,1,2}}{\lambda_1^{(1)}} & (ix) & \lambda_4^{(2)} = -8j\frac{w_{1,4,1}}{\lambda_1^{(1)}} \\ (v) & \lambda_3^{(1)} = 8 w_{3,1,1} & (x) & \lambda_4^{(3)} = -8j\frac{w_{1,1,4}}{\lambda_1^{(1)}}. \end{array} \quad (53)$$

Clearly, this scheme is neither numerically reliable nor robust to noise. This can be alleviated by noting that other alike procedures exploiting disjoint subsets of equations can be derived, which enables one to compute multiple candidate solutions and keep that which yields the lowest quadratic error w.r.t. $\mathcal{Y}$. Yet, for other combinations of R and I, this is not possible due to the pattern of vanishing equations—in particular, $\mathcal{W}$ is sparse when $I = R$.

The equations are simplified when the factors $\mathbf{C}^{(n)}$ are identical, since we can simply drop the superscripts of the eigenvalues in (52). For instance, if all factors are identical with $I = 4$ and $R = 3$ and $\lambda_1 \neq 0$, then one possible straightforward procedure for computing a solution is to calculate

$$\begin{array}{llll} (i) & \lambda_1 = 2\sqrt[3]{\frac{1}{3} w_{1,1,1}} & (iii) & \lambda_3 = 8\frac{w_{3,1,1}}{\lambda_1^2} \\ (ii) & \lambda_2 = 8j\frac{w_{2,1,1}}{\lambda_1^2} & (vi) & \lambda_4 = -8j\frac{w_{4,1,1}}{\lambda_1^2}. \end{array} \quad (54)$$

Note that, due to symmetry, a number of equations become redundant (unless $\mathcal{W}$ is noisy and the noise portion is not symmetric).

D. Combined algorithms

Although the non-iterative approaches described in Sections IV-B and IV-C bear a small computational cost in comparison with iterative methods, they often fall short of precision, especially in the presence of noise. Consequently, it is natural to consider a "hybrid" approach in a practical setting, by refining a solution produced by a non-iterative algorithm with the use of an iterative scheme.

Apart from the CALS algorithm presented in Sec. IV-A, one can employ a gradient, Newton or quasi-Newton descent to improve some initial approximation. As our model involves white zero-mean Gaussian noise, the maximum likelihood estimator is the solution to the least-squares problem

$$\min_{\boldsymbol{\eta}} \|\mathbf{y} - \mathbf{x}(\boldsymbol{\eta})\|_F^2 = \min_{\boldsymbol{\eta}} \|\mathbf{y} - \boldsymbol{\Phi}(\boldsymbol{\lambda}) f(\boldsymbol{\theta})\|_F^2 .$$

The gradient of such a cost function can then be written as $\nabla(\boldsymbol{\eta}) = -\mathbf{J}^T(\boldsymbol{\eta})\left(\mathbf{y} - \boldsymbol{\Phi}(\boldsymbol{\lambda}) f(\boldsymbol{\theta})\right)$ (recall that $\mathbf{J}$ denotes the Jacobian of (12)). When both $\mathcal{X}$ and $\mathcal{N}$ are symmetric, as discussed in Sec. III-A3, we have the problem

$$\min_{\boldsymbol{\eta}} \|\boldsymbol{\Psi}\left(\mathbf{y} - \boldsymbol{\Phi}(\boldsymbol{\lambda}) f(\boldsymbol{\theta})\right)\|_F^2 , \tag{55}$$

and thus the gradient reads

$$\nabla(\boldsymbol{\eta}) = -\mathbf{J}^T(\boldsymbol{\eta}) \boldsymbol{\Psi}^T \boldsymbol{\Psi}\left(\mathbf{y} - \boldsymbol{\Phi}(\boldsymbol{\lambda}) f(\boldsymbol{\theta})\right). \tag{56}$$

In the next section, we shall employ the above expressions to refine estimates produced by non-iterative schemes with the use of the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm [30], which is a quasi-Newton optimization method.

V. Simulation Results

We have conducted Monte Carlo simulations with the purposes of (i) illustrating the utility of the derived CRB expressions and (ii) comparing the statistical performance of estimators based on the algorithms of Sec. IV using the ECRB. As the structured factors we deal with in the experiments are Toeplitz (circulant), the algorithm of Sec. IV-B will be called CPTOEP (for "CP Toeplitz"), as in [14].

We note that all reported computing times were measured in Matlab R2013a running on a Intel Xeon ES-2630v2 2.60 GHz. Also, all ensemble averages are calculated by taking the 2% trimmed mean of the data, in order to attenuate the degrading effect of some realizations whose results are outstandingly poor. The tolerance used for BFGS is $\epsilon_{\text{BFGS}} = 10^{-8}$. The convergence criterion used for CALS is as described at the end of Sec. IV-A, and its tolerance was set as $\epsilon_{\text{CALS}} = 10^{-10}$, because ϵ_{BFGS} was seen to be to loose for this algorithm. Finally, we note that both these iterative algorithms were allowed to run only for a maximum number of 2000 iterations.

Three scenarios were considered, as described next. In each of them, the reported estimates of BMSE and ECRB were obtained via sample averaging (w.r.t. η) of, respectively, the MSE and the deterministic CRB given by the corresponding formula (according to the model structure) derived in Sec. III.

A. Three distinct circulant factors (CCPD)

In the first scenario, the SCPD model is constructed with $\mathcal{X} = \mathcal{I} \times_1 \mathbf{C}^{(1)} \times_2 \mathbf{C}^{(2)} \times_3 \mathbf{C}^{(3)}$, where each $\mathbf{C}^{(n)} \in \mathbb{R}^{I \times R}$ is circulant with $I = 4$ and $R = 3$. Hence, $\boldsymbol{\lambda}$ is not present in $\boldsymbol{\eta}$. We fix $\theta_1^{(1)} = \theta_1^{(2)} = 1$ to avoid identifiability issues, as discussed in Sec. II-C. Several joint realizations of $\bar{\boldsymbol{\theta}}^{(1)}$, $\bar{\boldsymbol{\theta}}^{(2)}$, $\bar{\boldsymbol{\theta}}^{(3)} = \boldsymbol{\theta}^{(3)}$ and $\bar{\mathcal{N}}$ are generated by drawing their elements from the standard Gaussian distribution. The noise tensor is then obtained as $\mathcal{N} = \sigma\bar{\mathcal{N}}$, with σ varying to simulate different signal-to-noise ratio (SNR) conditions.

Given one realization $\mathcal{Y}$, we apply the following estimators:

1) AAS: Factors are computed by solving (52), from which we obtain the estimate $\hat{\eta}$. To reduce degradation due to noise, we employ three different *ad-hoc* procedures as that shown in (53) and keep the solution which yields the lowest quadratic error w.r.t. $\mathcal{Y}$. As some complex residual is generally present in $\hat{\eta}$, we take its real part.
2) N_i-CALS: The algorithm given in Table II is employed. We use a multi-initialization scheme to improve performance, running the algorithm N_i times with different random initializations and keeping the solution yielding the lowest quadratic error w.r.t. $\mathcal{Y}$.
3) CPTOEP: The approach of Sec. IV-B is applied to jointly compute circulant factors $\mathbf{C}^{(1)}$ and $\mathbf{C}^{(2)}$, as well as an unstructured first estimate of $\mathbf{A}^{(3)}$. Then, $\boldsymbol{\theta}^{(3)}$ is estimated from $\mathbf{A}^{(3)}$ in the LS sense, by using (8).
4) AAS-CALS: The estimates given by AAS are used (after normalization) as initial points for CALS.
5) CPTOEP-CALS: After obtaining (normalized) initial estimates with CPTOEP, the CALS algorithm is used for refining them.
6) CPTOEP-BFGS: Instead of using CALS to refine the estimates given by CPTOEP, the quasi-Newton algorithm BFGS [30] is used.

Note that, as no normalization is imposed in AAS, CALS and CPTOEP, the parameter vectors need to be normalized *a posteriori*, by dividing $\boldsymbol{\theta}^{(n)}$ by $\theta_1^{(n)}$ for $n \in \{1, 2\}$ and absorbing these scaling factors in $\boldsymbol{\theta}^{(3)}$.

In Figure 1-(A), we show the BMSE of each estimator for $N_r = 500$ realizations of $\mathcal{Y}$, as well as the estimated ECRB, for multiple SNR levels (in dB) computed via

$$\text{SNR} = 10 \log_{10} \frac{\frac{1}{N_r} \sum_{n=1}^{N_r} \|\mathcal{X}_n\|_F^2}{\sigma^2 I^3}, \tag{57}$$

where $\mathcal{X}_n$ stands for the nth realization of $\mathcal{X}$. It can be seen that CALS performs quite poorly with a single random initialization, due to frequent early termination or inability to converge, while all other iterative estimators approximately reach the ECRB for SNR ≥ 15 dB. With regard to the non-iterative ones, CPTOEP performs better than AAS, thanks to its better numerical properties. By inspecting also the average computing times reported in Table IV, we conclude that AAS-CALS and CPTOEP-CALS provide the best compromise between precision and computational cost.

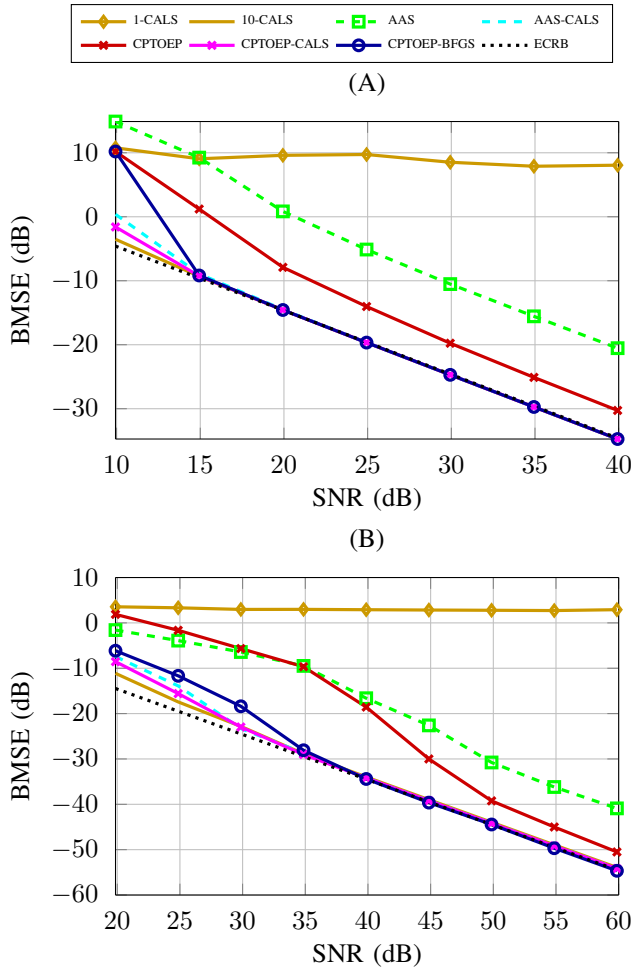

Fig. 1. Simulation results of scenarios A and B: BMSE of several estimators when applied to compute a SCPD having three distinct (A) and three identical (B) circulant factors, with $I_1 = I_2 = I_3 = 4$ and $R = 3$.

TABLE IV
AVERAGE COMPUTING TIME (IN SECONDS) MEASURED IN SCENARIO (A).

	N_i-CALS		AAS (*)		CPTOEP (‡)		
SNR	$N_i = 1$	$N_i = 10$	AAS	*-CALS	CPTOEP	‡-CALS	‡-BFGS
10	1.59e1	1.65e2	1.33e−3	1.73e−2	3.97e−3	2.13e−2	1.01e−1
20	1.36e1	1.43e2	1.16e−3	1.32e−2	3.56e−3	1.66e−2	6.77e−2
30	1.34e1	1.40e2	1.16e−3	1.17e−2	3.56e−3	1.48e−2	6.04e−2
40	1.35e1	1.42e2	1.15e−3	1.03e−2	3.57e−3	1.33e−2	5.48e−2

B. Three identical circulant factors (symmetric CCPD)

We consider now the setting where $\mathcal{X}$ admits a SCPD of the form $\mathcal{X} = \mathcal{I} \times_1 \mathbf{C} \times_2 \mathbf{C} \times_3 \mathbf{C}$, with $\mathbf{C} \in \mathbb{R}^{I \times R}$ circulant and $I = U = 4$, $R = 3$. The generation of model realizations is similar to that of Sec. V-A, but now there is a single $\bar{\boldsymbol{\theta}} = \boldsymbol{\theta} = \boldsymbol{\eta}$. Also, the noise tensor is symmetric (cf. the discussion of Sec. III-A3), with the elements n_{i_1,i_2,i_3} such that $i_1 \leq i_2 \leq i_3$ being generated as in the previous scenario and the other ones determined by symmetry.

We describe below how each algorithm was applied to estimate $\mathbf{C}$ from each realization of $\mathcal{Y}$.

1) AAS: Factors are computed by solving three disjoint subset of equations like (54) and keeping the best solution, in order to improve robustness against noise. As the parameter vector can only be estimated up to a complex scaling factor of the form $e^{j\frac{2\pi}{3}s}$, with $s \in \{0, 1, 2\}$, it is thus necessary to compensate it by taking into account the fact that $\boldsymbol{\theta}$ is real. This is done by computing

$$\hat{s} = \underset{s\in\{0,1,2\}}{\operatorname{argmin}} \sum_{u=1}^{U} \min \left\{ \left[\operatorname{Arg}(\theta_u) - \operatorname{Arg}\left(e^{j\frac{2\pi}{3}s}\right)\right]^2, \left[\operatorname{Arg}(\theta_u) - \operatorname{Arg}\left(e^{j(\frac{2\pi}{3}s-\pi)}\right)\right]^2 \right\}, \quad (58)$$

where $\operatorname{Arg} : \mathbb{C} \mapsto [-\pi, \pi[$ outputs the phase of its argument, and then estimating the generating vector as $\hat{\boldsymbol{\theta}} = \mathfrak{Re}\left\{e^{-j\frac{2\pi}{3}\hat{s}}\tilde{\boldsymbol{\theta}}\right\}$, with $\tilde{\boldsymbol{\theta}}$ denoting the output of AAS.
2) N_i-CALS: The CALS algorithm is applied this time taking symmetry into account, as shown in Table III. A multi-initialization scheme is again used with N_i random initial points.
3) CPTOEP: As this method does not take symmetry into account, this is done *a posteriori*, by averaging the three obtained factors, which are estimated as in Sec. V-A.

We also evaluate the combined approaches 4) AAS-CALS, 5) CPTOEP-CALS and 6) CPTOEP-BFGS.

The obtained results are shown in Fig. 1-(B). In comparison with the previous scenario, we can see that the algorithms perform in general worse. In the case of CALS, this is due to the imposition of symmetry. For both AAS and CPTOEP, an additional stage is employed (in AAS, for fixing the scaling factor; in CPTOEP, for computing a single factor estimate), which degrades performance. Nonetheless, all iterative algorithms (except for 1-CALS) get quite close to the ECRB for SNR $\geq$ 35 dB. Inspecting the average computing times reported in Table V, one can conclude that, once more, AAS-CALS and CPTOEP-CALS lead to the best compromise between statistical efficiency and computing cost.

TABLE V
AVERAGE COMPUTING TIME (IN SECONDS) MEASURED IN SCENARIO (B).

	N_i-CALS		AAS (*)		CPTOEP (‡)		
SNR	$N_i = 1$	$N_i = 10$	AAS	*-CALS	CPTOEP	‡-CALS	‡-BFGS
20	1.50e2	1.20e3	1.09e−3	2.98e−2	4.04e−3	4.11e−2	2.95e−2
30	6.44e1	8.45e2	1.09e−3	1.19e−2	4.03e−3	1.50e−2	2.53e−2
40	7.06e1	7.89e2	1.09e−3	9.94e−3	4.02e−3	1.29e−2	2.24e−2
50	6.26e1	7.14e2	1.09e−3	8.55e−3	4.02e−3	1.13e−2	2.05e−2
60	6.01e1	6.95e2	1.09e−3	7.23e−3	4.02e−3	9.80e−3	1.91e−2

C. Wiener-Hammerstein system identification

We now consider the Wiener-Hammerstein model identification problem, whose link with the computation of a SCPD was originally shown in [11]. In this context, an evaluation of several estimators in a deterministic setting has been recently conducted by the authors in [2]. Here, we extend it to a Bayesian context, considering the ECRB as the relevant bound. In the following, we first recall the problem formulation and then present the simulation procedure and its results.

1) Problem formulation: The Wiener-Hammerstein (WH) model is a well-known mathematical representation often used for modeling nonlinear dynamical systems [23]. Its time-invariant discrete-time version is illustrated in Fig. 2, where

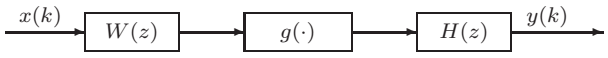

Fig. 2. Block-diagram of the Wiener-Hammerstein model.

$g(x)$ is a memoryless nonlinearity and $W(z)$, $H(z)$ are linear time-invariant systems. Because of its "modular" structure consisting of simple blocks, the WH model is said to belong to the class of block-oriented models [23].

Assume that the components of a given WH model have the form $g(x) = \sum_{n=1}^{N} g_n x^n$, $W(z) = \sum_{u=0}^{U-1} w_u z^{-u}$ and $H(z) = \sum_{r=0}^{R-1} h_r z^{-r}$. Then, the resulting input/output relation is

$$y(k) = \sum_{n=1}^{N} g_n \sum_{r=0}^{R-1} h_r \left[\sum_{m=r}^{U+r-1} w_{m-r} x(k-m) \right]^n, \quad (59)$$

where $x(k)$ is the input signal and $y(k)$ its corresponding output. It can be shown that this model admits an equivalent Volterra representation [31]

$$y(k) = \sum_{n=1}^{N} \sum_{i_1=0}^{I-1} \cdots \sum_{i_n=0}^{I-1} v^{(n)}(i_1, \ldots, i_n) \prod_{q=1}^{n} x(k - i_q),$$

having symmetric discrete-time finite-memory Volterra kernels

$$v^{(n)}(i_1, \ldots, i_n) = g_n \sum_{r=r_0}^{\tilde{R}} h_r \prod_{q=1}^{n} w_{i_q - r} \quad (60)$$

for $i_1, \ldots, i_n \in \{0, \ldots, I-1\}$, with $I = U + R - 1$, $r_0 = \max\{0, i_1 - U + 1, \ldots, i_n - U + 1\}$ and $\tilde{R} = \min\{R - 1, i_1, \ldots, i_n\}$.

Now, observe that (60) is of the same form as (2), provided we define $\lambda_r = h_{r-1}$ and $\mathbf{A}^{(n)} = \mathbf{A}$ such that

$$[\mathbf{A}]_{i_q, r} = \begin{cases} \theta_{i_q - r}, & r \le i_q \le U + r, \\ 0, & \text{otherwise}, \end{cases} \quad (61)$$

with $\theta_u = w_{u-1}$. In other words, $\mathbf{A}$ is banded circulant (as described by the last row of Table I). Hence, associating $v^{(n)}(i_1, \ldots, i_n)$ with a symmetric nth-order tensor $\boldsymbol{\mathcal{V}}^{(n)} \in \mathbb{R}^{I \times \cdots \times I}$ such that $[\boldsymbol{\mathcal{V}}^{(n)}]_{i_1, \ldots, i_n} = v^{(n)}(i_1 - 1, \ldots, i_n - 1)$, one can estimate the parameters of the linear subsystems (up to a factor g_n) by computing the SCPD of $\boldsymbol{\mathcal{V}}^{(n)}$.

Due to the scaling ambiguity of (60), one can assume, without loss of generality (as long as the kernel is not null), that $g_n = 1$. Also, assuming that $\theta_1 \neq 0$ (in other words, $W(z)$ does not comprise a pure delay), we can fix $\theta_1 = 1$, leaving the scaling in the vector $\boldsymbol{\lambda}$. So, the parameters of $W(z)$ and $H(z)$ can be estimated by computing the SCPD

$$\boldsymbol{\mathcal{V}}^{(n)} = \boldsymbol{\mathcal{I}} \times_1 \mathbf{A} \times_2 \cdots \times_{n-1} \mathbf{A} \times_n (\mathbf{A} \operatorname{Diag}(\boldsymbol{\lambda})). \quad (62)$$

This reasoning underlies the approach described in [11], where the parameters of $H(z)$ and $W(z)$ are obtained from the SCPD of a Volterra kernel estimated from input-output samples (by using any available method, such as, *e.g.*, [32]).

TABLE VI
AVERAGE COMPUTING TIME (IN SECONDS) MEASURED IN SCENARIO (C).

	N_i-CALS		CPTOEP (‡)		
SNR	$N_i = 1$	$N_i = 10$	CPTOEP	‡-CALS	‡-BFGS
-3.6	1.22e2	1.13e3	1.05e−2	3.30e−2	1.84e−1
6.4	1.97e1	2.98e2	7.61e−3	1.71e−2	7.16e−2
16.4	1.09e1	1.53e2	7.15e−3	1.42e−2	5.79e−2
26.4	1.06e1	1.46e2	7.34e−3	1.34e−2	5.36e−2

2) Evaluating estimators with the use of the ECRB: We turn now to the evaluation of different estimators when applied to compute (62), choosing $U = 5$ and $R = 3$; thus, $I = 7$. To perform this experiment, $N_r = 500$ realizations of the parameters $\bar{\boldsymbol{\theta}}$ and $\boldsymbol{\lambda}$ are generated by drawing each component $\bar{\theta}_u$ and λ_r uniformly over $[-1, 1]$. Note that we fix $\theta_1 = 1$ to ensure local identifiability.

We perform the described procedure for a third-order tensor $\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{7 \times 7 \times 7}$, with $\boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{V}}^{(3)}$ built from the exact Volterra kernel $v^{(3)}$ generated as (60). The kernel estimation error is modeled by the (symmetric) noise tensor $\boldsymbol{\mathcal{N}}$, which is generated exactly as in the previous scenario. The variance σ^2 of $\boldsymbol{\mathcal{N}}$ is again varied for simulating different SNR conditions.

The employed estimators are:

1) N_i-CALS: CALS is specialized to the particular structure of the Voltera kernel (62). This is done by estimating a single factor $\mathbf{A}$ per iteration. A multi-initialization scheme with N_i initializations is also used, for improving performance.
2) CPTOEP: The procedure is similar to that of scenario B, with the estimate of $\mathbf{A}$ being obtained by averaging the two structured factors estimated by the algorithm. After that, it is suitably normalized, and then $\boldsymbol{\lambda}$ is estimated by employing (48).

Again, we apply CALS and BFGS to refine the CPTOEP solution.

The BMSE estimated at several SNR levels is shown in Figure 3, with the corresponding time measurements reported in Table VI. It is seen that 1-CALS produces very poor results, due to the typical convergence problems encountered in practice. With 10 random initializations, this problem is overcome (for sufficiently high SNR), as it becomes more likely that at least one run will produce good estimates; however, the total computing cost is very high. In contrast, CPTOEP's BMSE lies only within moderate distance from the lower bound, but its computing cost is quite low. Taking advantage of this initial approximate solution, both CPTOEP-CALS and CPTOEP-BFGS are able to reach quite close to the ECRB for SNR values greater than approximately 11 dB, with a slight advantage for the latter. On the other hand, CPTOEP-CALS outperforms CPTOEP-BFGS from a computing cost perspective, thus offering the best compromise in this scenario.

VI. CONCLUSION

We have studied the structured CPD (SCPD) estimation problem, considering a tensor model in which the structured factor matrices belong to subspaces spanned by given basis matrices, and under the presence of additive white Gaussian

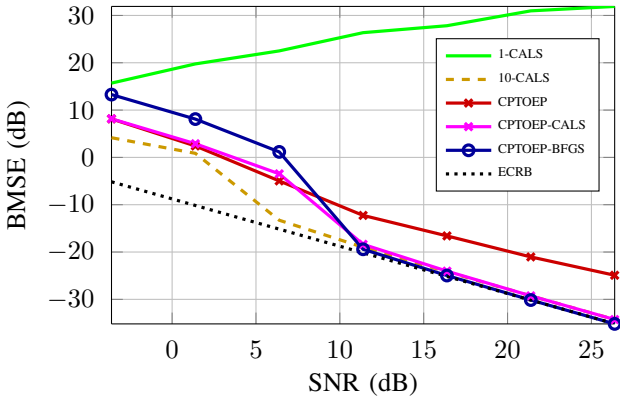

Fig. 3. WH model identification: BMSE of several estimators when applied to compute a symmetric SCPD with three identical circulant banded factors, one of them postmultiplied by a diagonal matrix.

noise. Closed-form expressions for the (deterministic) Cramér-Rao bound were derived for this model, taking into account the particularities of special cases, as that of a (possibly partially) symmetric SCPD. It was shown that the expected CRB (ECRB) provides a tighter lower bound than Van Trees' Bayesian CRB for the evaluation of estimators in a Bayesian context, where prior distributions are assigned for the parameters of interest. This idea was followed in three simulation scenarios, where several estimators based upon specialized SCPD computation algorithms were formulated and evaluated by comparing their Bayesian MSE with the ECRB. In particular, in one of these scenarios we considered the estimation of the linear subsystems of a Wiener-Hammerstein model as an example where the studied estimation problem arises.

Some of the formulated estimators were able to reach quite close to the bound for a wide range of SNR values. These statistically efficient estimators consist of two stages: first, a non-iterative method provides an initial solution, which is then refined by employing an iterative algorithm. It was also shown that this strategy can yield estimators which are also very efficient from a computational standpoint.

As perspectives for future work, we can mention the extension of the present study to other structured tensor model estimation problems, as the constrained CPD models described in [33], which find applications in psychometry, chemometrics, wireless communications and array signal processing.

REFERENCES

[1] M. Boizard, R. Boyer, G. Favier, and P. Comon, "Performance estimation for tensor CP decomposition with structured factors," in *Proc. Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Brisbane, 2015.

[2] J. H. de M. Goulart, M. Boizard, R. Boyer, G. Favier, and P. Comon, "Statistical efficiency of structured CPD estimation applied to Wiener-Hammerstein modeling," in *Proc. European Signal Process. Conf. (EUSIPCO)*, Nice, 2015, pp. 953–957.

[3] T. Kolda and B. Bader, "Tensor decompositions and applications," *SIAM Review*, vol. 51, no. 3, pp. 455–500, 2009.

[4] M. Mørup, "Applications of tensor (multiway array) factorizations and decompositions in data mining," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 1, no. 1, pp. 24–40, 2011.

[5] P. Comon, "Tensors : A brief introduction," *IEEE Signal Process. Mag.*, vol. 31, no. 3, pp. 44–53, 2014.

[6] A. Smilde, R. Bro, and P. Geladi, *Multi-way analysis: applications in the chemical sciences*, John Wiley & Sons, 2005.

[7] A. Cichocki, R. Zdunek, A. H. Phan, and S. Amari, *Nonnegative matrix and tensor factorizations: applications to exploratory multi-way data analysis and blind source separation*, John Wiley & Sons, 2009.

[8] J. D. Carroll and J.-J. Chang, "Analysis of individual differences in multidimensional scaling via an n-way generalization of "Eckart-Young" decomposition," *Psychometrika*, vol. 35, no. 3, pp. 283–319, 1970.

[9] R. A. Harshman, "Foundations of the PARAFAC procedure: models and conditions for an "explanatory" multi-modal factor analysis," *UCLA Working Papers in Phonetics*, vol. 16, pp. 1–84, 1970.

[10] C. E. R. Fernandes, G. Favier, and J. C. M. Mota, "Blind channel identification algorithms based on the PARAFAC decomposition of cumulant tensors: the single and multiuser cases," *Signal Process.*, vol. 88, no. 6, pp. 1382–1401, 2008.

[11] G. Favier and A. Y. Kibangou, "Tensor-based methods for system identification. Part 2: Three examples of tensor-based system identification methods," *Int. J. Sci. Tech. Automat. Control (IJ-STA)*, vol. 3, no. 1, 2009.

[12] D. Nion and N. D. Sidiropoulos, "Tensor algebra and multidimensional harmonic retrieval in signal processing for MIMO radar," *IEEE Trans. Signal Process.*, vol. 58, no. 11, pp. 5693–5705, 2010.

[13] A. Y. Kibangou and G Favier, "Non-iterative solution for PARAFAC with a Toeplitz matrix factor," in *EUSIPCO*, Glasgow, 2009, pp. 691–695.

[14] M. Sorensen and P. Comon, "Tensor decompositions with banded matrix factors," *Linear Algebra Applicat.*, vol. 438, no. 2, pp. 919–941, 2013.

[15] M. Sorensen and L. De Lathauwer, "Blind signal separation via tensor decomposition with Vandermonde factor: Canonical polyadic decomposition," *IEEE Trans. Signal Process.*, vol. 61, no. 22, pp. 5507–5519, 2013.

[16] J. H. de M. Goulart and G. Favier, "An algebraic solution for the Candecomp/PARAFAC decomposition with circulant factors," *SIAM J. Matrix Anal. Applicat.*, vol. 35, no. 4, pp. 1543–1562, 2014.

[17] X. Liu and N. D. Sidiropoulos, "Cramér-Rao lower bounds for low-rank decomposition of multidimensional arrays," *IEEE Trans. Signal Process.*, vol. 49, no. 9, pp. 2074–2086, 2001.

[18] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, Prentice-Hall, 1993.

[19] S. Sahnoun and P. Comon, "Tensor polyadic decomposition for antenna array processing," in *Proc. Int. Conf. Comput. Stat. (Compstat)*, Geneva, Switzerland, 2014, pp. 233–240.

[20] R. Boyer, "Deterministic asymptotic Cramér–Rao bound for the multidimensional harmonic model," *Signal Process.*, vol. 88, no. 12, pp. 2869–2877, 2008.

[21] Z. Ben-Haim and Y. C. Eldar, "A lower bound on the Bayesian MSE based on the optimal bias function," *IEEE Transactions on Information Theory*, vol. 55, no. 11, pp. 5179–5196, 2009.

[22] H. L. Van Trees and K. L. Bell, *Bayesian Bounds for Parameter Estimation and Nonlinear Filtering/Tracking*, Wiley, 2007.

[23] R. Haber and L. Keviczky, *Nonlinear system identification – Input-Output Modeling Approach*, vol. 1 of *Mathematical Modelling: Theory and Applications*, Kluwer Academic Publishers, 1999.

[24] L. L. Scharf and L. T. Mac Worther, "Geometry of the Cramér-Rao bound," *Signal Process.*, vol. 31, no. 3, pp. 301–311, Apr. 1993.

[25] P. Stoica and R. L. Moses, *Spectral analysis of signals*, Prentice Hall, 2005.

[26] R. T. Behrens and L. L. Scharf, "Signal processing applications of oblique projection operators," *IEEE Trans. Signal Process.*, vol. 42, no. 6, pp. 1413–1424, 1994.

[27] K. B. Petersen and M. S. Pedersen, "The matrix cookbook," 2012, Version 15/11/2012.

[28] G. Tomasi and R. Bro, "A comparison of algorithms for fitting the PARAFAC model," *Computational Statistics and Data Analysis*, vol. 50, no. 7, pp. 1700–1734, Apr. 2006.

[29] D. E. Dudgeon and R. M. Mersereau, *Multidimensional Digital Signal Processing*, Prentice Hall, NJ, USA, 1984.

[30] R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu, "A limited memory algorithm for bound constrained optimization," *SIAM J. Sci. Comput.*, vol. 16, no. 5, pp. 1190–1208, 1995.

[31] A. Y. Kibangou and G. Favier, "Wiener-Hammerstein systems modeling using diagonal Volterra kernels coefficients," *IEEE Signal Process. Lett.*, vol. 13, no. 6, pp. 381–384, 2006.

[32] C.-H. Tseng and E. J. Powers, "Identification of cubic systems using higher order moments of iid signals," *IEEE Trans. Signal Process.*, vol. 43, no. 7, pp. 1733–1735, 1995.

[33] G. Favier and A. L. F. de Almeida, "Overview of constrained PARAFAC models," *EURASIP Journal on Advances in Signal Processing*, vol. 2014, no. 1, pp. 1–25, 2014.

José Henrique de M. Goulart received the B.Sc. degree in Computer Science from Universidade Federal de Sergipe (UFS), Brazil, and the M.Sc. degree in Electronic Systems from Escola Politécnica da Universidade de São Paulo (EPUSP), Brazil, in 2006 and 2012, respectively. He is currently a Ph.D student at Université Nice Sophia Antipolis (UNS), France, working at the I3S Laboratory in Sophia Antipolis, France. His research interests include tensor models and decompositions, low-rank tensor recovery, tensor methods in signal processing and nonlinear system modeling and identification.

Maxime Boizard Maxime Boizard received the Ph.D degree from the Ecole Normale Supérieure de Cachan in 2013. From 2013 to 2015, he was a postdoctoral fellow at University Paris-Sud - Laboratory of Signals and Systems (L2S) in the context of a Digiteo-DigiCosme project.

Rémy Boyer received the B.Sc. and Ph.D degrees from the Ecole Nationale Supérieure des Telecommunications (ENST-Paris or Télécom ParisTech) in 1999 and 2002, respectively, in statistical signal processing. From 2002 to 2003, he was a postdoctoral fellow during six months at Sherbrooke University (Canada). From 2011 to 2012, Rémy Boyer was a visiting researcher at the SATIE Laboratory (Ecole Normale Supérieure de Cachan) and at the University of Aalborg (Danemark). From 2003, he is an associated professor at University Paris-Sud - Laboratory of Signals and Systems (L2S). Rémy Boyer received an "Habilitation à Diriger des Recherches (HDR)" from the University Paris-Sud in December 2012. His research interests include compressive sampling of non-bandlimited signals, array signal processing, Bayesian performance bounds for parameter estimation and detection, security in mobile networks as well as numerical linear and multi-linear algebra. Rémy Boyer is the author or co-author of over 100 publications and has been involved in more than 15 collaborative projects including 2 NoE (Networks of Excellence) projects. He was awarded the prize for excellence in scientific research ("prime d'excellence scientifique" in french). He is an elected member of the Commission of Experts of the University Paris-Sud, an elected member of the L2S laboratory board and a nominated member of the National Council of French Universities. Rémy Boyer is a member of the EURASIP Special Area Team SPMuS (Signal Processing for Multisensor Systems) and an affiliate member of the IEEE Sensor Array and Multichannel (SAM) Technical Committee.

Gérard Favier received the engineer diplomae from ENSCM (Ecole Nationale Supérieure de Chronométrie et de Micromécanique), Besançon, and ENSAE (Ecole Nationale Supérieure de l'Aéronautique et de l'Espace), Toulouse, the Dr.-Ing. (Ph.D.) and State doctorate degrees from the University of Nice-Sophia Antipolis, France, in 1973, 1974, 1977 and 1981, respectively. In 1976, he joined the CNRS (National Center for Scientific Research) and he is currently Emeritus Research Director at the CNRS and I3S Laboratory, in Sophia Antipolis, France. He was the Director of the I3S Laboratory from 1995 to 1999, and a member of the French National Committee of Scientific Research from 1995 to 2000.

His research interests include nonlinear system modeling and identification, tensor models/decompositions, tensor completion and low rank tensor recovery, and tensor approaches for cooperative communication systems. He is author or co-author of more than 200 papers published in international scientific conferences and journals, and 20 books or book chapters.

Pierre Comon (M'87 – SM'95 – F'07) received both the M.Sc. degree in 1982, and the Doctorate degree in 1985, from the University of Grenoble, France. He later received the Habilitation to Lead Researches in 1995, from the University of Nice, France. He has been for nearly 13 years in industry, first with Crouzet-Sextant, Valence, France, between 1982 and 1985, and then with Thomson Marconi, Sophia-Antipolis, France, between 1988 and 1997. He spent 1987 with the ISL laboratory, Stanford University, CA. He joined in 1997 the Eurecom Institute, Sophia-Antipolis, France. He is research director with CNRS since 1998, first at laboratory I3S, Sophia-Antipolis, France, until 2012, and then at Gipsa-Lab, Grenoble, France. His research interests include Blind techniques, Statistical Signal and Array Processing, Tensor decompositions, Multi-Way Factor Analysis and its applications to biomedical end environment.

Dr. Comon was Associate Editor of the *IEEE Transactions on Signal Processing* from 1995 to 1998, and a member of the French National Committee of Scientific Research from 1995 to 2000. He was the coordinator of the European Basic Research Working Group on High-Order Statistics, *ATHOS*, from 1992 to 1995. Between 1992 and 1998, he was a member of the Technical and Scientific Council of the Thomson/Thales multinational company. Between 2001 and 2004 he acted as launching Associate Editor with the *IEEE Transactions on Circuits and Systems I*, in the area of Blind Techniques. Ha has also been a member of the editorial board of the Elsevier journal *Signal Processing*, and a member of the IEEE SPTM, BSP and SAM Technical Commitees. He is presently in the editorial board of *SIAM Journal on Matrix Analysis and Applications*.