Cramér-Rao-Induced Bounds for CANDECOMP/PARAFAC tensor decomposition

Petr Tichavský[1], Anh Huy Phan[2], and Zbyněk Koldovský[1,3]

***Abstract*—This paper presents a Cramér-Rao lower bound (CRLB) on the variance of unbiased estimates of factor matrices in Canonical Polyadic (CP) or CANDECOMP/PARAFAC (CP) decompositions of a tensor from noisy observations, (i.e., the tensor plus a random Gaussian i.i.d. tensor). A novel expression is derived for a bound on the mean square angular error of factors along a selected dimension of a tensor of an arbitrary dimension. The expression needs less operations for computing the bound, $O(NR^6)$, than the best existing state-of-the art algorithm, $O(N^3R^6)$ operations, where N and R are the tensor order and the tensor rank. Insightful expressions are derived for tensors of rank 1 and rank 2 of arbitrary dimension and for tensors of arbitrary dimension and rank, where two factor matrices have orthogonal columns.**

The results can be used as a gauge of performance of different approximate CP decomposition algorithms, prediction of their accuracy, and for checking stability of a given decomposition of a tensor (condition whether the CRLB is finite or not). A novel expression is derived for a Hessian matrix needed in popular damped Gauss-Newton method for solving the CP decomposition of tensors with missing elements. Beside computing the CRLB for these tensors the expression may serve for design of damped Gauss-Newton algorithm for the decomposition.

Index Terms

Multilinear models; canonical polyadic decomposition; Cramér-Rao lower bound; stability; uniqueness

I. Introduction

Order-3 and higher-order data arrays need to be analyzed in diverse research areas such as chemistry, astronomy, and psychology [1]–[3]. The analyses can be done through finding multi-linear dependencies among elements within the arrays. The most popular model is Parallel factor analysis (PARAFAC), also called Canonical decomposition (CANDECOMP) or CP, which is an extension of a low rank decomposition of matrices to higher-way arrays, usually called tensors. In signal processing, the tensor decompositions have become popular for their usefulness in blind source separation [4].

Note that a best-fitting CP decomposition may not exist for some tensors. In that case, trying to find a best-fitting CP decomposition results in diverging factors [5], [6]. This paper is focussed on studying CP decompositions of a noisy observations of tensors, which admit an exact CP decomposition. The decomposition of the noiseless tensor is taken as a *ground truth* for computing errors.

An important issue is the essential uniqueness of CP decomposition as it entails identifiability of the model (the factor matrices) from the tensor. The adjective "essential" means that the model is unique up to a scale and permutation ambiguity, which is inherent to the problem. Initial works in the field can be traced back in 70's in works of Harshman [7], [8]. A popular sufficient condition for the uniqueness was derived by Kruskal in [9]. Recently, the problem has been addressed again, namely by Stegeman, Ten Berge, De Lathauwer, Jiang, Sidiropoulos et al.; see [10]-[23].

This paper is focussed on stability of the CP decomposition rather than on the uniqueness. By stability we mean existence of a finite Cramér-Rao bound in a stochastic set-up, where tensor elements are corrupted by additive Gaussian-distributed noise. Relation of this kind of stability to a deterministic stability and to the uniqueness was studied in [24]. It is not true, in general, that stability of a solution of a nonlinear problem implies uniqueness of the solution. For example, there might always be a permutation or sign ambiguity. It is yet an open theoretical question if stability of the CP tensor decomposition problem implies its *essential* uniqueness. Regardless of the missing link to identifiability, the stability is an interesting concept which is worth to be studied, because different kind of noise is very common.

In general, in order to evaluate performance of a tensor decomposition, the approximation error between the data tensor and its approximate is sometimes used. Unfortunately, such measure does not imply quality of the estimated components. In practice, in some difficult scenarios such as decomposition of tensor with linear dependency among components of factor matrices, or large difference in magnitude between components [25], [26], most CP algorithms explained the data tensor at almost identical fit, but only few algorithms can accurately retrieve the hidden components from the tensor [27], [25]. In order to verify theoretically the quality of the estimated components and evaluate robustness of an algorithm, an appropriate measure is an essential prerequisite. The squared angular error between the estimated component and its original one is such a measure [28], [29]. Working with angular errors is practical, because the scaling ambiguity does not play a role. Only the permutation ambiguity has to be solved in practical examples, because order of the factor can be quite arbitrary.

Cramér-Rao lower bound for CP decomposition was first studied in [30], and later, a more compact asymptotic ex-

This work was supported by the Grant Agency of the Czech Republic through projects 102/09/1278 and P103/11/1947.
[1]Institute of Information Theory and Automation, Pod vodárenskou věží 4, P.O.Box 18,182 08 Prague 8, Czech Republic. E-mail: tichavsk@utia.cas.cz.
[2]Brain Science Institute, RIKEN, Wakoshi, Japan. E-mail: phan@brain.riken.jp.
[3]Faculty of Mechatronic and Interdisciplinary Studies, Technical University of Liberec, Studentská 2, 461 17 Liberec, Czech Republic. E-mail: zbynek.koldovsky@tul.cz.

pression was derived in [31] for tensors of order 3 appearing in wireless communications. A non-asymptotic (exact) CRLB-induced bound (CRIB) on squared angular deviation of columns of the factor matrices with respect to their nominal values has been studied in [28]. Similar results for symmetric tensors are derived in [32]. Nevertheless, the study is limited to the case of three-way tensors. In the general case, CRIB can be, indeed, calculated through the approximate Hessian which is often huge, and is impractical to directly invert. Note that such task normally costs $O(R^3T^3)$ where $T = \sum_n I_n$. Seeking a cheaper method for CRIB is a challenge to made it applicable.

This paper presents new CRIB expressions for tensors of arbitrary dimension and rank, and specialized expressions for rank 1 and rank 2 tensors. The results rely on compact expressions for Hessian of the problem derived in [27]. Alternative expressions for the Hessian exist in [38]. Note, however, that unlike [27], this paper presents different expressions for inverse of the Hessian, which have lower computational complexity. In particular, complexity of inversion of the Hessian is reduced from $O(N^3R^6)$ operations to $O(NR^6)$, where N and R are the tensor order and the tensor rank, respectively.

On basis of new discovered properties of the CRIB, we established connection between theoretical and practical results in CP decomposition (CPD):

- Stability of CPD for rank-1 and rank-2 tensors of arbitrary dimension.
- The work may serve as theoretical support for a novel CP decomposition algorithm through tensor reshaping [33], which was designed to decompose high-dimensional and high-order tensors. In particular, it appears that higher-order orthogonally constrained CPD [34], [35], [36], [37] can be decomposed efficiently through tensor unfolding.
- Stability when factor matrices occur linear dependence problem and especially the rank-overlap problem [1], [22], [35]. The problem is related to a variant of CPD for linear dependent loadings which was investigated in chemometric data and in flow injection analysis [1], [35]. A partial uniqueness condition of the related model is discussed in [22].
- CP decomposition of tensors with missing entries, which is quite frequent in practice, is addressed. An approximate Hessian for this case is derived, which is the core for the damped Gauss-Newton algorithm for the decomposition.
- A maximum tensor rank, given dimension of the tensor, which admits a stable decomposition is discussed.

The paper is organized as follows. Section II presents the main result, the Cramér-Rao induced bound on angular error of one factor vector in full generality. In Section III, this result is specialized for tensors of rank 1 and rank 2, and for the case when two factor matrices have mutually orthogonal columns. Section IV is devoted to a possible application of the bound: investigation of loss of accuracy of the tensor decomposition when the tensor is reshaped to a lower-dimensional form. Section V deals with the bound for tensors with missing entries, Section VI contains examples – CRIB computed for CP decomposition of a fluorescence tensor, stability of the tensor investigated by Brie *et al*, and a discussion of a maximum stable rank given the tensor dimension. Section VII concludes the paper.

II. Presentation of the CRIB

A. Cramér-Rao bound for CP decomposition

Let $\mathcal{Y}$ be an $N-$ way tensor of dimension $I_1 \times I_2 \times \ldots \times I_N$. The tensor is said to be of rank R, if R is the smallest number of rank-one tensors which admit the decomposition of $\mathcal{Y}$ of the form

$$\mathcal{Y} = \sum_{r=1}^{R} \mathbf{a}_r^{(1)} \circ \mathbf{a}_r^{(2)} \circ \ldots \circ \mathbf{a}_r^{(N)} \tag{1}$$

where $\circ$ denotes the outer vector product, $\mathbf{a}_r^{(n)}$, $r = 1, \ldots, R$, $n = 1, \ldots, N$ are vectors of the length I_n called factors. The tensor in (1) can be characterized by N factor matrices $\mathbf{A}_n = [\mathbf{a}_1^{(n)}, \mathbf{a}_2^{(n)}, \ldots, \mathbf{a}_R^{(n)}]$ of the size $I_n \times R$ for $n = 1, \ldots, N$. Sometimes (1) is referred to as a Kruskal form of a tensor [44].

In practice, CP decomposition of a given rank (R) is used as an approximation of a given tensor, which can be a noisy observation $\hat{\mathcal{Y}}$ of the tensor $\mathcal{Y}$ in (1). Owing to the symmetry of (1), we can focus on estimating the first factor matrix $\mathbf{A}_1$, without any loss of generality, and we can assume that all other factor matrices have columns of unit norm. Then the "energy" of the parallel factors is determined by the squared Euclidean norm of columns of $\mathbf{A}_1$.

It is common to assume that the noise has a zero mean Gaussian distribution with variance σ^2, and is independently added to each element of the tensor in (1).

Let a vector parameter $\boldsymbol{\theta}$ containing all parameters of our model be arranged as

$$\boldsymbol{\theta} = [(\mathtt{vec}\,\mathbf{A}_1)^T, \ldots, (\mathtt{vec}\,\mathbf{A}_N)^T]^T \ . \tag{2}$$

The maximum likelihood solution for $\boldsymbol{\theta}$ consists in minimizing the least squares criterion

$$\mathcal{Q}(\boldsymbol{\theta}) = \|\hat{\mathcal{Y}} - \mathcal{Y}(\boldsymbol{\theta})\|_F^2 \tag{3}$$

where $\|\cdot\|_F$ stands for the Frobenius norm.

We wish to compute the Cramér-Rao lower bound for estimating $\boldsymbol{\theta}$. In general, for this estimation problem, the CRLB is given as the inverse of the Fisher information matrix, which is equal to [28]

$$\mathbf{F}(\boldsymbol{\theta}) = \frac{1}{\sigma^2}\, \mathbf{J}^T(\boldsymbol{\theta})\mathbf{J}(\boldsymbol{\theta}) \tag{4}$$

where $\mathbf{J}(\boldsymbol{\theta})$ is the Jacobi matrix (matrix of the first-order derivatives) of $\mathcal{Q}(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$. In other words, the Fisher information matrix is proportional to the approximate Hessian matrix of the criterion, $\mathbf{H}(\boldsymbol{\theta}) = \mathbf{J}^T(\boldsymbol{\theta})\mathbf{J}(\boldsymbol{\theta})$ [38].

Let $\boldsymbol{\Gamma}_{nm}$ denote the Hadamard (elementwise) product of matrices $\mathbf{C}_k = \mathbf{A}_k^T \mathbf{A}_k, k \in \{1, \ldots, N\} - \{n, m\}$,

$$\boldsymbol{\Gamma}_{nm} \;=\; \mathop{\circledast}_{k \neq n,m} \mathbf{C}_k\,, \qquad \mathbf{C}_k = \mathbf{A}_k^T \mathbf{A}_k \ . \tag{5}$$

Theorem 1 [27]: The Hessian $\mathbf{H}$ can be decomposed into low rank matrices under the form as

$$\mathbf{H} = \mathbf{G} + \mathbf{Z}\,\mathbf{K}\,\mathbf{Z}^T \tag{6}$$

where $\mathbf{K} = [\mathbf{K}_{nm}]_{n,m=1}^N$ contains submatrices $\mathbf{K}_{nm}$ given by

$$\mathbf{K}_{nm} = (1-\delta_{nm})\mathbf{P}_R \,\mathrm{dvec}\,(\mathbf{\Gamma}_{nm}) \tag{7}$$

$\mathbf{P}_R$ is the permutation matrix of dimension $R^2 \times R^2$ defined in [27] such that $\mathtt{vec}\,\mathbf{M} = \mathbf{P}_R \,\mathtt{vec}\,(\mathbf{M}^T)$ for any $R \times R$ matrix $\mathbf{M}$, and δ_{nm} is the Kronecker delta, and $\mathtt{dvec}(\mathbf{M})$ is a short-hand notation for $\mathtt{diag}(\mathtt{vec}(\mathbf{M}))$, i.e. a diagonal matrix containing all elements of a matrix $\mathbf{M}$ on its main diagonal. Next,

$$\mathbf{G} \;=\; \mathtt{bdiag}\,(\mathbf{\Gamma}_{nn} \otimes \mathbf{I}_{I_n})_{n=1}^N \tag{8}$$

and

$$\mathbf{Z} = \mathtt{bdiag}\,(\mathbf{I}_R \otimes \mathbf{A}_n)_{n=1}^N \tag{9}$$

where $\otimes$ denotes the Kronecker product, $\mathbf{I}_{I_n}$ is an identity matrix of the size $I_n \times I_n$, and $\mathtt{bdiag}(\cdot)$ is a block diagonal matrix with the given blocks on its diagonal. Note that the Hessian $\mathbf{H}$ in (6) is rank deficient because of the scale ambiguity of columns of factor matrices [26], [40]. It has dimension $(R\sum_n I_n) \times (R\sum_n I_n)$ but its rank is at most $R\sum_n I_n - (N-1)R$.

A regular (reduced) Hessian can be obtained from $\mathbf{H}$ by deleting $(N-1)R$ rows and corresponding columns in $\mathbf{H}$, because the estimation of one element in the vectors $\mathbf{a}_r^{(n)}$, $r = 1,\ldots,R$, $n = 2,\ldots,N$ can be skipped. The reduced Hessian may have the form

$$\mathbf{H}_E = \mathbf{E}\mathbf{H}\mathbf{E}^T \tag{10}$$

where

$$\mathbf{E} = \mathtt{bdiag}\,(\mathbf{I}_{RI_1}, \mathbf{I}_R \otimes \mathbf{E}_2, \ldots, \mathbf{I}_R \otimes \mathbf{E}_N) \tag{11}$$

and $\mathbf{E}_n$ is an $(I_n-1)\times I_n$ matrix of rank $I_n - 1$. For example, one can put $\mathbf{E}_n = [\mathbf{0}_{(I_n-1)\times 1}\,\mathbf{I}_{I_n-1}]$ for $n = 2,\ldots,N$. With this definition of $\mathbf{E}_n$, $\mathbf{H}_E$ is a Hessian for estimating the first factor matrix $\mathbf{A}_1$ and all other vectors $\mathbf{a}_r^{(n)}$, $r = 1,\ldots,R$, $n = 2,\ldots,N$ without their first elements. In the sequel, however, we use a different definition of $\mathbf{E}_n$. Note that each $\mathbf{E}_n$ can be quite arbitrary, together facilitate a regular transformation of nuisance parameters, which does not influence CRLB of the parameter of interest.

The CRLB for the first column of $\mathbf{A}_1$, denoted simply as $\mathbf{a}_1$, is defined as σ^2 times the left-upper submatrix of $\mathbf{H}_E^{-1}$ of the size $I_1 \times I_1$,

$$\mathrm{CRLB}(\mathbf{a}_1) = \sigma^2\,[\mathbf{H}_E^{-1}]_{1:I_1,1:I_1}\;. \tag{12}$$

Substituting (6) in (10) gives

$$\mathbf{H}_E = \mathbf{G}_E + \mathbf{Z}_E\mathbf{K}\mathbf{Z}_E^T \tag{13}$$

where $\mathbf{G}_E = \mathbf{E}\mathbf{G}\mathbf{E}^T$ and $\mathbf{Z}_E = \mathbf{E}\mathbf{Z}$. Inverse of $\mathbf{H}_E$ can be written using a Woodbury matrix identity [39] as

$$\mathbf{H}_E^{-1} = \mathbf{G}_E^{-1} - \mathbf{G}_E^{-1}\mathbf{Z}_E\mathbf{K}(\mathbf{I}_{NR^2} + \mathbf{Z}_E^T\mathbf{G}_E^{-1}\mathbf{Z}_E\mathbf{K})^{-1}\mathbf{Z}_E^T\mathbf{G}_E^{-1} \tag{14}$$

provided that the involved inverses exist.

Next,

$$\begin{aligned}\mathbf{G}_E \;&=\; \mathtt{bdiag}\left(\mathbf{\Gamma}_{11}\otimes\mathbf{I}_1, \mathbf{\Gamma}_{22}\otimes(\mathbf{E}_2\mathbf{E}_2^T),\right.\\ &\qquad \left.\ldots,\mathbf{\Gamma}_{NN}\otimes(\mathbf{E}_N\mathbf{E}_N^T)\right) \qquad (15)\\ \mathbf{G}_E^{-1} \;&=\; \mathtt{bdiag}\left((\mathbf{\Gamma}_{11})^{-1}\otimes\mathbf{I}_1, \mathbf{\Gamma}_{22}^{-1}\otimes(\mathbf{E}_2\mathbf{E}_2^T)^{-1},\right.\\ &\qquad \left.\ldots,\mathbf{\Gamma}_{NN}^{-1}\otimes(\mathbf{E}_N\mathbf{E}_N^T)^{-1}\right)\;. \qquad (16)\end{aligned}$$

Put

$$\begin{aligned}\mathbf{\Psi} \;&=\; \mathbf{Z}_E^T\mathbf{G}_E^{-1}\mathbf{Z}_E \qquad (17)\\ \mathbf{B} \;&=\; \mathbf{K}(\mathbf{I}_{NR^2} + \mathbf{\Psi}\mathbf{K})^{-1} \qquad (18)\end{aligned}$$

and let $\mathbf{B}_0$ be the upper–left $R^2 \times R^2$ submatrix of $\mathbf{B}$, symbolically $\mathbf{B}_0 = \mathbf{B}_{1:R^2,1:R^2}$. Finally, let g_{11} and $\mathbf{g}_{1,:}$ be the upper–left element and the first row of $\mathbf{\Gamma}_{11}^{-1}$, respectively. Then

$$\begin{aligned}&[\mathbf{H}_E^{-1}]_{1:I_1,1:I_1}\\ &= [\mathbf{G}_E^{-1}]_{1:I_1,1:I_1} + [\mathbf{G}_E^{-1}\mathbf{Z}_E]_{1:I_1,1:R^2}\mathbf{B}_0[\mathbf{G}_E^{-1}\mathbf{Z}_E]_{1:I_1,1:R^2}^T\\ &= g_{11}\mathbf{I}_{I_1} + (\mathbf{g}_{1,:}\otimes\mathbf{A}_1)\,\mathbf{B}_0\,(\mathbf{g}_{1,:}\otimes\mathbf{A}_1)^T\;. \qquad (19)\end{aligned}$$

The CRLB represents a lower bound on the error covariance matrix $\mathrm{E}[(\hat{\mathbf{a}}_1 - \mathbf{a}_1)(\hat{\mathbf{a}}_1 - \mathbf{a}_1)^T]$ for any unbiased estimator of $\mathbf{a}_1$. The bound is asymptotically tight in the case of Gaussian noise and least squares estimator, which is equivalent to maximum likelihood estimator, under the assumptions that the permutation ambiguity has been solved out (order of the estimated factors was selected to match the original factors) and scaling of the estimator is in accord with the selection of the matrix $\mathbf{E}$.

B. Cramér-Rao-induced bound for angular error

CRLB($\mathbf{a}_1$) considered in the previous subsection is a matrix. In applications it is practical to characterize the error of the factor $\mathbf{a}_1$ in the decomposition by a scalar quantity. In [29] it was proposed to characterize the error by an angle between the true and the estimated vector, and compute a Cramér-Rao-induced bound (CRIB) for the squared angle. The CRIB may serve a gauge of achievable accuracy of estimation/CP decomposition. Again, it is an asymptotically (in the sense of variance of the noise going to zero) tight bound on the angular error between an estimated and true factor.

The angle α_1 between the true factor $\mathbf{a}_1$ and its estimate $\hat{\mathbf{a}}_1$ obtained through the CP decomposition is defined through its cosine

$$\cos\alpha_1 = \frac{\mathbf{a}_1^T\hat{\mathbf{a}}_1}{\|\mathbf{a}_1\|\,\|\hat{\mathbf{a}}_1\|}\;. \tag{20}$$

The Cramér-Rao induced bound for the squared angular error α_1^2 [radians^2] will be denoted CRIB($\mathbf{a}_1$) in the sequel. CRIB($\mathbf{a}_1$) in decibels (dB) is then defined as $-10\log_{10}[\mathrm{CRIB}(\mathbf{a}_1)]$ [dB].

Before computing CRIB($\mathbf{a}_1$) we present another interpretation of this quantity. Let the estimate $\hat{\mathbf{a}}_1$ be decomposed into a sum of a scalar multiple of $\mathbf{a}_1$ and a reminder, which is orthogonal to $\mathbf{a}_1$,

$$\hat{\mathbf{a}}_1 = \beta\mathbf{a}_1 + \mathbf{r}_1 \tag{21}$$

where $\beta = \mathbf{a}_1^T\hat{\mathbf{a}}_1/\|\mathbf{a}_1\|^2$ and $\mathbf{r}_1 = \hat{\mathbf{a}}_1 - \beta\mathbf{a}_1$. Then, the Distortion-to-Signal Ratio (DSR) of the estimate $\hat{\mathbf{a}}_1$ can be defined as

$$\mathrm{DSR}(\hat{\mathbf{a}}_1) = \frac{\|\mathbf{r}_1\|^2}{\beta^2\|\mathbf{a}_1\|^2}\ . \tag{22}$$

A straightforward computation gives

$$\mathrm{DSR}(\hat{\mathbf{a}}_1) = \frac{1-\cos^2\alpha_1}{\cos^2\alpha_1} \approx \alpha_1^2\ . \tag{23}$$

The approximation in (23) is valid for small α_1^2. We can see that CRIB($\mathbf{a}_1$) serves not only as a bound on the mean squared angular estimation error, but also as a bound on the achievable Distortion-to-Signal Ratio.

Theorem 2 [29]: Let CRLB($\mathbf{a}_1$) be the Cramér-Rao bound on covariance matrix of unbiased estimators of $\mathbf{a}_1$. Then the Cramér–Rao–induced bound on the squared angular error between the true and estimated vector is

$$\mathrm{CRIB}(\mathbf{a}_1) = \frac{\mathrm{tr}[\Pi^{\perp}_{\mathbf{a}_1}\mathrm{CRLB}(\mathbf{a}_1)]}{\|\mathbf{a}_1\|^2} \tag{24}$$

where

$$\Pi^{\perp}_{\mathbf{a}_1} = \mathbf{I}_{I_1} - \mathbf{a}_1\mathbf{a}_1^T/\|\mathbf{a}_1\|^2 \tag{25}$$

is the projection operator to the orthogonal complement of $\mathbf{a}_1$ and tr[.] denotes trace of a matrix.

Proof: A sketch of a proof can be found in [29]. It is based on analysis of a mean square angular error of a maximum likelihood estimator, which is known to be asymptotically tight (achieving the Cramér-Rao bound). Note that a conceptually more straightforward but longer proof would be obtained through the formula for CRLB on a transformed parameter, see e.g., Theorem 3.4 in [43]. In particular,

$$\mathrm{CRIB}(\mathbf{a}_1) = \mathbf{G}_a(\mathbf{a}_1)\mathrm{CRLB}(\mathbf{a}_1)\mathbf{G}_a^T(\mathbf{a}_1) \tag{26}$$

where $\mathbf{G}_a(\hat{\mathbf{a}}_1)$ is the Jacobi matrix of the mapping representing the angular error as a function of the estimate $\hat{\mathbf{a}}_1$.

Theorem 3: The CRIB($\mathbf{a}_1$) can be written in the form

$$\mathrm{CRIB}(\mathbf{a}_1) = \frac{\sigma^2}{\|\mathbf{a}_1\|^2}\left\{(I_1-1)g_{11} - \mathrm{tr}\left[\mathbf{B}_0\left((\mathbf{g}_{1,:}^T\mathbf{g}_{1,:})\otimes\mathbf{X}_1\right)\right]\right\} \tag{27}$$

where $\mathbf{B}_0$ is the submatrix of $\mathbf{B}$ in (18), $\mathbf{B}_0 = \mathbf{B}_{1:R^2,1:R^2}$,

$$\mathbf{X}_n = \mathbf{C}_n - \frac{1}{\mathbf{C}_{11}^{(n)}}\mathbf{C}_{:,1}^{(n)}\mathbf{C}_{:,1}^{(n)\,T} \tag{28}$$

for $n = 1,\ldots,N$, $\mathbf{C}_{11}^{(n)}$ and $\mathbf{C}_{:,1}^{(n)}$ denote the upper–right element and the first column of $\mathbf{C}_n$, respectively, and $\mathbf{\Psi}$ in the definition of $\mathbf{B}$ (18) takes, for a special choice of matrices $\mathbf{E}_n$, the form

$$\mathbf{\Psi} = \mathtt{bdiag}\left(\mathbf{\Gamma}_{11}^{-1}\otimes\mathbf{C}_1, \mathbf{\Gamma}_{22}^{-1}\otimes\mathbf{X}_2,\ldots,\mathbf{\Gamma}_{NN}^{-1}\otimes\mathbf{X}_N\right)\ . \tag{29}$$

Proof: Substituting (12) and (19) into (24) gives, after some simplifications,

$$\begin{aligned}&\mathrm{CRIB}(\mathbf{a}_1)\\ &= \frac{\sigma^2}{\|\mathbf{a}_1\|^2}\mathrm{tr}\left[\Pi^{\perp}_{\mathbf{a}_1}\left(g_{11}\mathbf{I}_{I_1} - (\mathbf{g}_{1,:}\otimes\mathbf{A}_1)\mathbf{B}_0(\mathbf{g}_{1,:}\otimes\mathbf{A}_1)^T\right)\right]\\ &= \frac{\sigma^2}{\|\mathbf{a}_1\|^2}\left\{(I_1-1)g_{11} - \mathrm{tr}\left[\Pi^{\perp}_{\mathbf{a}_1}(\mathbf{g}_{1,:}\otimes\mathbf{A}_1)\mathbf{B}_0(\mathbf{g}_{1,:}\otimes\mathbf{A}_1)^T\right]\right\}\\ &= \frac{\sigma^2}{\|\mathbf{a}_1\|^2}\left\{(I_1-1)g_{11} - \mathrm{tr}\left[\mathbf{B}_0\left((\mathbf{g}_{1,:}^T\mathbf{g}_{1,:})\otimes(\mathbf{A}_1^T\Pi^{\perp}_{\mathbf{a}_1}\mathbf{A}_1)\right)\right]\right\}\end{aligned} \tag{30}$$

This is (27), because

$$\mathbf{A}_1^T\Pi^{\perp}_{\mathbf{a}_1}\mathbf{A}_1 = \mathbf{C}_1 - \frac{1}{\mathbf{C}_{11}^{(1)}}\mathbf{C}_{:,1}^{(1)}\mathbf{C}_{:,1}^{(1)\,T} = \mathbf{X}_1\ . \tag{31}$$

Next, assume that $\mathbf{E}$ is defined as in (11), but $\mathbf{E}_n$ are arbitrary full rank matrices of the dimension $(I_n-1)\times I_n$. Then, combining (17), (9), (11) and (16) gives

$$\begin{aligned}\mathbf{\Psi} &= \mathbf{Z}_E^T\mathbf{G}_E^{-1}\mathbf{Z}_E\\ &= \mathtt{bdiag}\left(\mathbf{\Gamma}_{11}^{-1}\otimes\mathbf{C}_1, \mathbf{\Gamma}_{22}^{-1}\otimes\widetilde{\mathbf{X}}_2,\ldots,\mathbf{\Gamma}_{NN}^{-1}\otimes\widetilde{\mathbf{X}}_N\right)\end{aligned} \tag{32}$$

where

$$\widetilde{\mathbf{X}}_n = \mathbf{A}_n^T\mathbf{E}_n^T(\mathbf{E}_n\mathbf{E}_n^T)^{-1}\mathbf{E}_n\mathbf{A}_n \tag{33}$$

for $n = 2,\ldots,N$. Note that the expression $\mathbf{E}_n^T(\mathbf{E}_n\mathbf{E}_n^T)^{-1}\mathbf{E}_n$ is an orthogonal projection operator to the columnspace of $\mathbf{E}_n^T$. If $\mathbf{E}_n$ is chosen as the first (I_n-1) rows of

$$\Pi^{\perp}_{\mathbf{a}_1^{(n)}} = \mathbf{I}_{I_n} - \mathbf{a}_1^{(n)}\mathbf{a}_1^{(n)\,T}/\|\mathbf{a}_1^{(n)}\|^2 \tag{34}$$

then $\mathbf{E}_n^T(\mathbf{E}_n\mathbf{E}_n^T)^{-1}\mathbf{E}_n = \Pi^{\perp}_{\mathbf{a}_1^{(n)}}$ and consequently $\widetilde{\mathbf{X}}_n = \mathbf{A}_n^T\Pi^{\perp}_{\mathbf{a}_1^{(n)}}\mathbf{A}_n = \mathbf{X}_n$. ■

Note that the first row and the first column of $\mathbf{X}_n$ are zero.

Theorem 4: Assume that all elements of the matrices $\mathbf{C}_n$ in (5) are nonzero. Then, the matrix $\mathbf{B}_0$ in Theorem 3 can be written in the form

$$\mathbf{B}_0 = [-\mathbf{I}_{R^2} + \mathbf{V}(\mathbf{I}_{R^2}+\mathbf{V})^{-1}]\mathbf{Y} \tag{35}$$

where

$$\mathbf{V} = \mathbf{W} - \mathbf{Y}(\mathbf{\Gamma}_{11}^{-1}\otimes\mathbf{C}_1) \tag{36}$$

$$\mathbf{W} = \mathbf{P}_R\sum_{n=2}^{N}\mathtt{dvec}(\mathbf{\Gamma}_{1n})\mathbf{S}_n^{-1}(\mathbf{\Gamma}_{nn}^{-1}\otimes\mathbf{X}_n)\mathtt{dvec}(\mathbf{C}_1\oslash\mathbf{C}_n) \tag{37}$$

$$\mathbf{Y} = \mathbf{P}_R\sum_{n=2}^{N}\mathtt{dvec}(\mathbf{\Gamma}_{1n})\mathbf{S}_n^{-1}(\mathbf{\Gamma}_{nn}^{-1}\otimes\mathbf{X}_n)\mathbf{P}_R\mathtt{dvec}(\mathbf{\Gamma}_{1n}) \tag{38}$$

$$\mathbf{S}_n = \mathbf{I}_{R^2} - (\mathbf{\Gamma}_{nn}^{-1}\otimes\mathbf{X}_n)\mathtt{dvec}(\mathbf{\Gamma}_{nn}\oslash\mathbf{C}_n)\mathbf{P}_R \tag{39}$$

for $n = 2,\ldots,N$. In (37) and (39), "$\oslash$" stands for the element-wise division.

Proof: See Appendix B.

Note that in place of inverting the matrix $\mathbf{B}$ of the size $NR^2\times NR^2$, Theorem 4 reduces the complexity of the CRIB computation to N inversions of the matrices of the size $R^2\times R^2$. The Theorem can be extended to computing the inverse of the whole Hessian in $O(NR^6)$ operations, see [47].

Finally, note that the assumption that elements of $\mathbf{C}_n$ must not be zero is not too restrictive. Basically, it means that no pair of columns in the factor matrices must be orthogonal. The

Cramér-Rao bound does not exhibit any singularity in these cases, and is continuous function of elements of $\mathbf{C}_n$. If some element of $\mathbf{C}_n$ is closer to zero than say 10^{-5}, it is possible to increase its distance from zero to that value, and the resultant CRIB will differ from the true one only slightly.

Theorem 5 (Properties of the CRIB)

1) The CRIB in Theorems 3 and 4 depends on the factor matrices $\mathbf{A}_n$ only through the products $\mathbf{C}_n = \mathbf{A}_n^T \mathbf{A}_n$.
2) The CRIB is inversely proportional to the signal-to-noise ratio (SNR) of the factor of the interest (i.e. $\|\mathbf{a}_1\|^2/(\sigma^2 I_1)$) and independent of the SNR of the other factors, $\|\mathbf{a}_r\|^2/(\sigma^2 I_r)$, $r = 2, \ldots, R$.

Proof: Property 1 follows directly from Theorem 3. Property 2 is proven in Appendix C.

III. Special cases

A. Rank 1 tensors

In this case, the matrix $\mathbf{X}_1$ is zero, and

$$\mathrm{CRIB}(\mathbf{a}_1) = \frac{\sigma^2}{\|\mathbf{a}_1\|^2}(I_1 - 1)g_{11} = \frac{\sigma^2}{\|\mathbf{a}_1\|^2}(I_1 - 1) \,. \quad (40)$$

In (40), $g_{11} = 1$ due to the convention that the factor matrices $\mathbf{A}_n$, $n \geq 2$, have columns of unit norm. The result (40) is in accord with Harshman's early results on uniqueness of rank-1 tensor decomposition [8].

B. Rank 2 tensors

Consider the scaling convention that all factor vectors except the first factor have unit norm. Let c_n, $|c_n| \leq 1$, be defined as

$$c_n = \begin{cases} (\mathbf{a}_1^{(n)})^T \mathbf{a}_2^{(n)} & \text{for} \quad n = 2, \ldots, N \\ (\mathbf{a}_1^{(1)})^T \mathbf{a}_2^{(1)} / (\|\mathbf{a}_1^{(1)}\| \, \|\mathbf{a}_2^{(1)}\|) & \text{for} \quad n = 1 \,. \end{cases} \quad (41)$$

It follows from Theorem 5 that the CRIB on $\mathbf{a}_1$ is a function of $c_1, \ldots, c_N$ multiplied by $\sigma^2/\|\mathbf{a}_1\|^2$. It is symmetric function in $c_2, \ldots, c_N$ and possibly nonsymmetric in c_1. A closed form expression for the CRIB in the special case is subject of the following theorem.

Theorem 6 It holds for rank 2 tensors

$$\mathrm{CRIB}(\mathbf{a}_1) = \frac{\sigma^2}{\|\mathbf{a}_1\|^2}\frac{1}{1-h_1^2} \cdot \left[I_1 - 1 + \frac{(1-c_1^2)h_1^2[y^2 + z - h_1^2 z(z+1)]}{(1 - c_1 y - h_1^2(z+1))^2 - h_1^2(y + c_1 z)^2} \right] \quad (42)$$

where

$$h_n = \prod_{2 \leq k \neq n}^{N} c_n \qquad \text{for} \quad n = 1, \ldots, N \quad (43)$$

$$y = -c_1 \sum_{n=2}^{N} \frac{h_n^2(1-c_n^2)}{c_n^2 - h_n^2 c_1^2} \quad (44)$$

$$z = \sum_{n=2}^{N} \frac{1 - c_n^2}{c_n^2 - h_n^2 c_1^2} \,. \quad (45)$$

Proof: See Appendix D.

Note that the expressions (44)-(45) contain, in their denominators, terms $c_n - h_n c_1$. If any of these terms goes to zero, then quantities y and z go to infinity. In despite of this, the whole CRIB remain finite, because y and z appear both in the numerator and denominator in (42).

For example, for order-3 tensors ($N = 3$) we get (using e.g., Symbolic Matlab or Mathematica)

$$\mathrm{CRIB}_{N=3}(\mathbf{a}_1) = \frac{\sigma^2}{\|\mathbf{a}_1\|^2}\frac{1}{1-h_1^2}\left[I_1 - 1 + \frac{c_2^2}{1-c_2^2} + \frac{c_3^2}{1-c_3^2} \right] . \quad (46)$$

The above result coincides with the one derived in [28]. As far as the stability is concerned, the CRIB is finite unless either the second or third factor have co-linear columns. Note that the fact that the CRIB for $\mathbf{a}_1$ does not depend on c_1 can be linked to the uni-mode uniqueness conditions presented in [22].

For $N = 4$, the similar result is hardly tractable. Unlike the case $N = 3$, the result depends on c_1. A closer inspection of the result shows that the CRIB, as a function of c_1, achieves its maximum at $c_1 = 0$, and minimum at $c_1 = \pm 1$. Therefore we shall treat these two limit cases separately. We get

$$\mathrm{CRIB}_{N=4,c_1=0}(\mathbf{a}_1) = \quad (47)$$
$$\frac{\sigma^2}{\|\mathbf{a}_1\|^2}\frac{1}{1-h_1^2}\left[I_1 - 1 + \frac{c_2^2c_3^2 + c_2^2c_4^2 + c_3^2c_4^2 - 3c_2^2c_3^2c_4^2}{2c_2^2c_3^2c_4^2 - c_2^2c_3^2 - c_2^2c_4^2 - c_3^2c_4^2 + 1} \right]$$

$$\mathrm{CRIB}_{N=4,c_1=\pm 1}(\mathbf{a}_1) = \quad (48)$$
$$\begin{cases} \frac{\sigma^2}{\|\mathbf{a}_1\|^2}\frac{I_1-1}{1-h_1^2} & \text{for} \quad (|c_2|<1) \& (|c_3|<1) \& (|c_4|<1) \\ \frac{\sigma^2}{\|\mathbf{a}_1\|^2}\frac{1}{1-h_1^2}\left[I_1 - 1 + \frac{c_2^2+c_3^2-2c_2^2c_3^2}{(1-c_2^2)(1-c_3^2)} \right] & \text{for} \quad |c_4| = 1 \\ \frac{\sigma^2}{\|\mathbf{a}_1\|^2}\frac{1}{1-h_1^2}\left[I_1 - 1 + \frac{c_2^2+c_4^2-2c_2^2c_4^2}{(1-c_2^2)(1-c_4^2)} \right] & \text{for} \quad |c_3| = 1 \\ \frac{\sigma^2}{\|\mathbf{a}_1\|^2}\frac{1}{1-h_1^2}\left[I_1 - 1 + \frac{c_3^2+c_4^2-2c_3^2c_4^2}{(1-c_3^2)(1-c_4^2)} \right] & \text{for} \quad |c_2| = 1 \,. \end{cases}$$

As far as the stability is concerned, we can see that the CRIB is always finite unless two of the factor matrices have co-linear columns.

Similarly, for a general N, we have for $c_1 = 0$

$$\mathrm{CRIB}_{c_1=0}(\mathbf{a}_1) = \frac{\sigma^2}{\|\mathbf{a}_1\|^2}\frac{1}{1-h_1^2}\left[I_1 - 1 + \frac{h_1^2 z}{1 - h_1^2(z+1)} \right] \quad (49)$$

C. A case with two factor matrices having orthogonal columns

This subsection presents a closed-form CRIB for a tensor of a general order and rank, provided that two of its factor matrices have mutually orthogonal columns. The result cannot be derived from Theorem 5, because assumptions of the theorem are not fulfilled.

Theorem 7 When the factor matrices $\mathbf{A}_1$ and $\mathbf{A}_2$ both have mutually orthogonal columns, it holds

$$\mathrm{CRIB}(\mathbf{a}_1) = \frac{\sigma^2}{\|\mathbf{a}_1\|^2}\left[I_1 - 1 + \sum_{r=2}^{R} \frac{\gamma_r^2}{1 - \gamma_r^2} \right] \quad (50)$$

where $\gamma_r = \prod_{n=3}^{N} (\mathbf{a}_1^{(n)})^T \mathbf{a}_r^{(n)}$ for $r = 2, \ldots, R$.

Proof: See Appendix E.

Theorem 7 represents an important example when a tensor reshaping (see Section V.A. and [33] for more details) enables very efficient (fast) CP decomposition without compromising accuracy. It has close connection with orthogonally constrained CPD [35], [36], [37].

IV. CRIB FOR TENSORS WITH MISSING OBSERVATIONS

It happens in some applications, that tensors to be decomposed via CP have missing entries (some observations are simply missing). In this case, it is possible to treat stability of the decomposition through the CRIB as well. The only problem is that it is not possible to use expressions in Theorems 3-8 in such cases.

Assume that the tensor to be studied is given by its factor matrices $\mathbf{A}_1, \ldots, \mathbf{A}_N$ and a 0-1 "indicator" tensor $\mathcal{W}$ of the same dimension as $\mathcal{Y}$, which determines which tensor elements are available (observed). The task is to compute CRIB for columns of the factor matrices, like in the previous sections. The CRIB is computed through the Hessian matrix $\mathbf{H}$ as in (12) and (20), but its fast inversion is no longer possible. The Hessian itself can be computed as in its earlier definition

$$\mathbf{H} = \mathbf{J}_W^T(\boldsymbol{\theta})\mathbf{J}_W(\boldsymbol{\theta}), \qquad \mathbf{J}_W(\boldsymbol{\theta}) = \frac{\partial \mathtt{vec}(\mathcal{Y} \circledast \mathcal{W})}{\partial \boldsymbol{\theta}} \tag{51}$$

where $\boldsymbol{\theta}$ is the parameter of the model (2). More specific expressions for the Hessian can be derived in a straightforward manner.

Theorem 8: Consider the Hessian for tensor with missing data as an $N \times N$ partitioned matrix $\mathbf{H} = [\mathbf{H}^{(n,m)}]_{n=1,m=1}^{N,N}$ where $\mathbf{H}^{(n,m)} = [\mathbf{H}_{r,s}^{(n,m)}]_{r=1,s=1}^{R,R} \in \mathbb{R}^{RI_n \times RI_m}$. Then

$$\mathbf{H}_{r,s}^{(n,m)} = \tag{52}$$

$$\begin{cases} \mathtt{diag}\left(\mathcal{W}\,\bar{\times}_{-n}\left\{\mathbf{a}_r^{(1)} \circledast \mathbf{a}_s^{(1)}, \cdots, \mathbf{a}_r^{(N)} \circledast \mathbf{a}_s^{(N)}\right\}\right) & \text{for} \quad n = m \\ (\mathbf{a}_r^{(n)}\,\mathbf{a}_s^{(m)T}) \circledast \left(\mathcal{W}\,\bar{\times}_{-\{n,m\}}\left\{\mathbf{a}_r^{(1)} \circledast \mathbf{a}_s^{(1)}, \cdots, \mathbf{a}_r^{(N)} \circledast \mathbf{a}_s^{(N)}\right\}\right), & \text{for} \quad n \neq m \end{cases}$$

$\mathcal{Y}\bar{\times}_n\mathbf{u}_n$ denotes the mode-n tensor-vector product between $\mathcal{Y}$ and $\mathbf{u}_n$ [4], and

$$\mathcal{Y}\bar{\times}_{-n}\{\mathbf{u}\} = \mathcal{Y}\,\bar{\times}_1\mathbf{u}_1 \cdots \bar{\times}_{n-1}\mathbf{u}_{n-1}\bar{\times}_{n+1}\mathbf{u}_{n+1} \cdots \bar{\times}_N\mathbf{u}_N. \tag{53}$$

Proof: See Appendix F.

Theorem 8 can be used either to compute the CRIB for tensors with missing elements, or for implementing damped Gauss-Newton method for finding the decomposition in difficult cases, where ALS converges poorly.

V. APPLICATION AND EXAMPLES

A. *Tensor decomposition through reshape*

Assume that the tensor to-be decomposed is of dimension $N \geq 4$. The tensor can be reshaped to a lower dimensional tensor, which is computationally easier to decompose, so that the first factor matrix remains unchanged. The topic will be better elaborated in our next paper [33], in this paper we present only the main idea on two examples, to demonstrate usefulness of the CRIB.

In the first example, consider $N = 4$. The tensor in (1) can be reshaped to an order-3 tensor

$$\mathcal{Y}_{res} = \sum_{r=1}^{R} \mathbf{a}_r^{(1)} \circ \mathbf{a}_r^{(2)} \circ (\mathbf{a}_r^{(4)} \otimes \mathbf{a}_r^{(3)}) \;. \tag{54}$$

Both the original and the re-shaped tensors have the same number of elements ($I_1 I_2 I_3 I_4$) and the same noise added to them.

The question is, what is the accuracy of the factor matrix of the reshaped tensor compared to the original one. The former accuracy should be worse, because a decomposition of the reshaped tensor ignores structure of the third factor matrix. The question is, by how much worse. If the difference were negligible, then it is advised to decompose the simpler tensor (of lower dimension).

If the tensor has rank one, accuracy of both decompositions is the same. It is obvious from (40).

Let us examine tensors of rank 2. If the original tensor has scalar products of columns of the factor matrices c_1, c_2, c_3 and c_4, the reshaped tensor has scalar products c_1, c_2, and $c_3 c_4$, respectively. CRIB$(\mathbf{a}_1)$ of the reshaped tensor is independent of c_1, while CRIB of the original tensor is dependent on c_1, so there is a difference, in general. The difference will be smallest for $c_1 = 0$ (orthogonal factors) and largest for c_1 close to ± 1 (nearly or completely co-linear factors along the first dimension).

The smallest difference between CRIB$(\mathbf{a}_1)$ for the reshaped tensor and for the original one is

$$\frac{\sigma^2}{\|\mathbf{a}_1\|^2}\left[\frac{c_2^2 + c_3^2 c_4^2 - 2c_2^2 c_3^2 c_4^2}{(1-c_2^2)(1-c_3^2 c_4^2)} - \frac{c_2^2 c_3^2 + c_2^2 c_4^2 + c_3^2 c_4^2 - 3c_2^2 c_3^2 c_4^2}{(1-c_2^2 c_3^2 c_4^2)(2c_2^2 c_3^2 c_4^2 - c_2^2 c_3^2 - c_2^2 c_4^2 - c_3^2 c_4^2 + 1)}\right]$$

and the largest difference is

$$\frac{\sigma^2}{\|\mathbf{a}_1\|^2}\left[\frac{c_2^2 + c_3^2 c_4^2 - 2c_2^2 c_3^2 c_4^2}{(1-c_2^2)(1-c_3^2 c_4^2)}\right] = \frac{\sigma^2}{\|\mathbf{a}_1\|^2}\left[\frac{c_2^2}{1-c_2^2} + \frac{c_3^2 c_4^2}{1-c_3^2 c_4^2}\right] \;.$$

We can see that the difference may be large if the second or third factor matrix of the reshaped tensor has nearly co-linear columns ($c_2^2 \approx 1$ or $c_3^2 c_4^2 \approx 1$) . For example, for a tensor with $I_1 = 5$, $c_1 = 0$, $c_2 = 0.99$, $c_3 = c_4 = 0.1$ the loss of accuracy in decomposing reshaped tensor in place of the original one is 11.22 dB. If c_1 is changed to 1, the loss is only slightly higher, 11.23 dB. If $c_1 = c_2 = 0$, the loss is 0 dB for any c_3, c_4 (compare Theorem 7). If $c_1 = 1$, $c_2 = 0$ and $c_3 = c_4 = 0.99$, the loss is 8.5 dB.

Another example is a tensor of an arbitrary order and rank considered in Theorem 7. Let this tensor be reshaped to the order-3 tensor of the size $I_1 \times I_2 \times (I_3 \ldots I_N)$. Comparing the CRIB$(\mathbf{a}_1)$ of the original tensor and of the reshaped tensor shows that these two coincide. It follows that the decomposition based on reshaping is lossless in terms of accuracy.

B. *Amino Acids Tensor*

A data set consisting of five simple laboratory-made samples of fluorescence excitation-emission (5 samples $\times$ 201 emission wavelengths $\times$ 61 excitation wavelengths) is considered. Each sample contains different amounts of tryptophan, tyrosine, and phenylalanine dissolved in phosphate buffered water. The samples were measured by fluorescence on a spectrofluorometer [42]. Hence, a CP model with $R = 3$ is appropriate to the fluorescence data.

The tensor was factorized for several possible ranks R using the fLM algorithm [27]. CRIBs on the extracted components were then computed with the noise levels deduced from the error tensor $\mathcal{E} = \mathcal{Y} - \hat{\mathcal{Y}}$

$$\sigma^2 = \frac{\|\mathcal{Y} - \hat{\mathcal{Y}}\|_F^2}{\prod_n I_n}. \tag{55}$$

The resultant CRIB's are computed for all columns of all factor matrices and are summarized in Table 1.

Note that due to the "$-10\log_{10}$" definition, high CRIB in dB means high accuracy, and vice versa. A CRIB of 50 dB means that the standard angular deviation (square root of mean square angular error) of the factor is cca 0.18^o; a CRIB of 20 dB corresponds to the standard deviation 5.7^o.

The second mode to the decomposition, which represents intensity of the data versus the emission wavelength, for $R = 2, 3, 4$ and 8 is shown in Figure 1. We can see that the CRIB allows to distinguish between strong/significant modes of the decomposition and possibly artificial modes due to over-fitting the model. The criterion is different in general than the plain "energy" of the factor; if a factor has a low energy, it will probably have high CRIB, but it might not hold true vice versa. A high energy component might have a high CRIB.

In the next experiment, we have studied how much the accuracy of the decomposition is affected in case that some data are missing (not available). The decomposition with the correct rank $R = 3$ and σ^2 estimated as in (55) was taken as a ground truth; the 0-1 indicator tensor $\mathcal{W}$ of the same size was randomly generated with a given percentage of missing values. The CRIB of the second mode factors was plotted in Figure 2 as a function of this missing value rate. The figure also contains mean square angular error of the components obtained in simulations. Here an artificial Gaussian noise with zero mean and variance σ^2 was added to the "ground truth" tensor. The decomposition was obtained by a Levenberg-Marquardt algorithm [27] modified for tensors with missing entries.

A few observations can be made here.

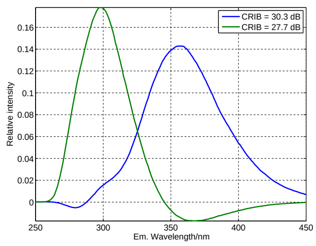
- CRIB coincides with MSAE for the percentage of the missing entries smaller than 70%. If the percentage exceeds the threshold, CRIB becomes overly optimistic.
- In general, accuracy of the decomposition declines slowly with the number of missing entries. If the number of missing entries is about 20%, loss of accuracy of the decomposition is only about 1-2 dB.

C. Stability of the decomposition of Brie's tensor

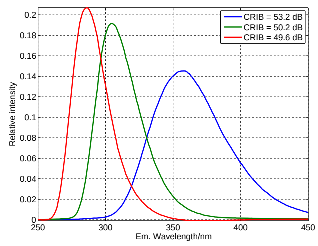
Brie *et al* [20] presented an example of a four-way tensor of rank 3, which arises while studying the response of bacterial bio-sensors to different environmental agents. The tensor has co-linear columns in three of four modes and the main message of the paper is that its CP decomposition is still unique. In this subsection we verify stability of the decomposition. The factor matrices of the tensor have the form

$$\mathbf{A}_1 = [\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3], \quad \mathbf{A}_2 = [\mathbf{a}_4, \mathbf{a}_4, \mathbf{a}_5],$$
$$\mathbf{A}_3 = [\mathbf{a}_6, \mathbf{a}_7, \mathbf{a}_6], \quad \mathbf{A}_4 = [\mathbf{a}_8, \mathbf{a}_9, \mathbf{a}_9] \ .$$

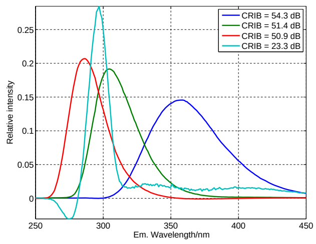
Assume for simplicity that all factors have unit norm, $\|\mathbf{a}_n\| = 1$, $n = 1, \ldots, 9$. Due to Theorem 5 it holds that CRIB on



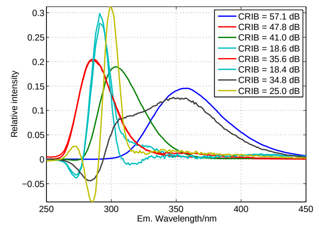
(a) Estimated components as $R = 2$.



(b) Estimated components as $R = 3$.



(c) Estimated components as $R = 4$.



(d) Estimated components as $R = 8$.

Fig. 1. Illustration for emission components from best-fit decompositions over 100 Monte Carlo runs for example VI-A.

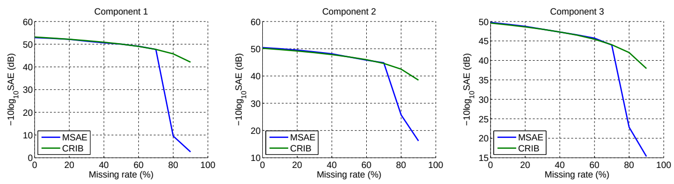


Fig. 2. CRIB for the second-mode components of CP decomposition of tensor in section VI.A with missing elements and mean square angular error obtained in simulations versus percentage of the missing elements.

$\mathbf{a}_1$ is a function of scalars $c_{11} = \mathbf{a}_1^T\mathbf{a}_2$, $c_{12} = \mathbf{a}_1^T\mathbf{a}_3$, $c_{13} = \mathbf{a}_2^T\mathbf{a}_3$, $c_2 = \mathbf{a}_4^T\mathbf{a}_5$, $c_3 = \mathbf{a}_6^T\mathbf{a}_7$, $c_4 = \mathbf{a}_8^T\mathbf{a}_9$ and I_1, which is the dimension of $\mathbf{a}_1$. Then, the matrices $\mathbf{C}_n = \mathbf{A}_n^T\mathbf{A}_n$, $n = 2, 3, 4$, have the form

$$\mathbf{C}_2 = \begin{bmatrix} 1 & 1 & c_2 \\ 1 & 1 & c_2 \\ c_2 & c_2 & 1 \end{bmatrix}, \mathbf{C}_3 = \begin{bmatrix} 1 & c_3 & 1 \\ c_3 & 1 & c_3 \\ 1 & c_3 & 1 \end{bmatrix}$$
$$\mathbf{C}_4 = \begin{bmatrix} 1 & c_4 & c_4 \\ c_4 & 1 & 1 \\ c_4 & 1 & 1 \end{bmatrix} .$$

A straightforward usage of Theorem 4 is not possible, because some of the involved matrices become singular. The CRIB itself, however, is finite and can be computed using an artificial parameter ε as a limit. The limit CRIB is computed for modified matrices at $\varepsilon \to 0$,

$$\mathbf{C}_{2\varepsilon} = \begin{bmatrix} 1 & 1-\varepsilon & c_2 \\ 1-\varepsilon & 1 & c_2 \\ c_2 & c_2 & 1 \end{bmatrix}$$
$$\mathbf{C}_{3\varepsilon} = \begin{bmatrix} 1 & c_3 & 1-\varepsilon \\ c_3 & 1 & c_3 \\ 1-\varepsilon & c_3 & 1 \end{bmatrix}$$
$$\mathbf{C}_{4\varepsilon} = \begin{bmatrix} 1 & c_4 & c_4 \\ c_4 & 1 & 1-\varepsilon \\ c_4 & 1-\varepsilon & 1 \end{bmatrix} .$$

If any of the correlations c_2, c_3, c_4 is zero, it is also augmented by ε.

The limit CRIB can be shown to be independent of off-diagonal elements of $\mathbf{C}_1$, unless $\mathbf{C}_1$ is singular. Assume that $\mathbf{C}_1$ is regular. The result, obtained by Symbolic Matlab, is

$$\begin{aligned}&\text{CRIB}_{\varepsilon=0}(\mathbf{a}_1) = \\ &= \frac{\sigma^2}{\|\mathbf{a}_1\|^2}\frac{1}{2c_2^2c_3^2c_4^2 - c_2^2c_3^2 - c_2^2c_4^2 - c_3^2c_4^2 + 1} \\ &\cdot\left[(I_1-1)(1-c_2^2c_3^2) - \frac{c_3^4(c_2^2+1)-3c_3^2+1}{1-c_3^2}\right. \\ &\left. -\frac{c_2^4(c_3^2+1)-3c_2^2+1}{1-c_2^2} + \frac{2-c_2^2-c_3^2}{1-c_4^2}\right] .\end{aligned} \tag{56}$$

It follows that the decomposition is stable, unless all three factors in some mode are collinear.

D. Maximum Stable Rank

A theoretically interesting question is, what is the maximum rank of a tensor of a given dimension which has a stable CP decomposition (with finite CRIB). For easy reference, we shall call it *maximum stable rank* and denote it $R_{smax}(I_1,\ldots,I_N)$.

An upper bound for the maximum stable rank can be deduced from the requirement that the number of free parameters in the model, which is $R(\sum_{n=1}^N I_n - N + 1)$ in CP decomposition, cannot exceed dimension of the available data, which is $\prod_{n=1}^N I_n$. It follows that

$$R_{smax}(I_1,\ldots,I_N) \le \left\lfloor \frac{\prod_{n=1}^N I_n}{\sum_{n=1}^N I_n - N + 1}\right\rfloor \tag{57}$$

where $\lfloor x \rfloor$ denotes the lower integer part of x. It can be verified numerically that for many (and maybe all[1]) tensor dimensions, an equality in (57) holds. In other words, it means that the CRIB computed, e.g., via Theorem 4 for a CP decomposition with rank $R = R_{smax}$ and some (e.g. random) factor matrices is finite. For example, the maximum stable rank is $R_{smax} = 2$ for $2\times 2\times 2$ tensors, and $R_{smax} = 3$ for $3\times 3\times 3$ tensors. For order-8 tensors of dimension $2\times\ldots\times 2$, $(8\times)$, it holds $R_{smax} = 28$.

It might be interesting to compare the maximum stable rank with the maximum rank and the maximum typical rank (to be explained below) for given tensor dimension, if they are known [45]. If the elements of a tensor are chosen randomly according to a continuous probability distribution, there is not a rank which occurs with probability 1 in general. Such rank, if exists, is called generic. Ranks which occur with strictly positive probabilities are called typical ranks. For example it was computed in [10] that probability for a real random Gaussian tensor of the size $2\times 2\times 2$ to be 2 and 3 is $\pi/4$, and $1-\pi/4$, respectively. We can see that no tensor of the rank 3 and the dimension has a stable decomposition. For tensors of the dimension $3\times 3\times 3$ the typical rank is 5 [10], it is a generic rank - but no decomposition of these rank-5 tensors is stable, as $R_{smax} = 3$.

[1]We do not have yet a formal proof that the equality in (57) holds for *all* tensor dimensions and orders.

Next, it might be interesting to compare the maximum stable rank with the maximum rank for unique tensor decomposition, or prove that these two coincide. Liu and Sidiropoulos [11], [30] derived a necessary condition for uniqueness of the CP decomposition, which, according to a formulation in [44] reads

$$\min_{n=1,\ldots,N} \text{rank}\,(\mathbf{A}_1\odot\ldots\odot\mathbf{A}_{n-1}\odot\mathbf{A}_{n+1}\odot\ldots\odot\mathbf{A}_N) = R \tag{58}$$

where $\odot$ means the Khatri-Rao product. The condition (58) is equivalent to the condition that the matrices $\mathbf{\Xi}_n = \mathbf{A}_1\odot\ldots\odot\mathbf{A}_{n-1}\odot\mathbf{A}_{n+1}\odot\ldots\odot\mathbf{A}_N$ have all full column rank, $n = 1,\ldots,N$, which is further equivalent to the condition that the product $\mathbf{\Xi}_n^T\mathbf{\Xi}_n$ are regular for $n = 1,\ldots,N$. Finally note that

$$\mathbf{\Xi}_n^T\mathbf{\Xi}_n = \mathbf{\Gamma}_{nn}, \qquad n = 1,\ldots,N .$$

where $\mathbf{\Gamma}_{nn}$ was defined in (5) and appears in computation of the CRIB.

Unfortunately, it appears that the condition (58) is only necessary, but not sufficient for uniqueness. It is often fulfilled for R higher than R_{smax}. Thus a relation between the stability and uniqueness of the CP decomposition remains open question for now.

VI. Conclusions

Cramér-Rao bounds for CP tensor decomposition represent an important tool for studying accuracy and stability of the decomposition. The bounds derived in this manuscript serve as a theoretical support for a method of the decomposition through tensor reshaping [33]. As a side result, a novel method of inverting Hessian matrix, which is more computationally efficient, is derived for the problem. It enables a further improvement of speed of the fast Gauss-Newton for the problem [27], [47]. A novel expression for Hessian for CP decomposition of tensor with missing entries has been derived. It can serve for assessing accuracy of CP decomposition of these tensors without need of long Monte Carlo simulations, and for implementing a damped Gauss-Newton algorithm for CP decomposition of these tensors.

A direct link between stability and essential uniqueness remains to be an open theoretical question. In particular, it is not known yet for sure if stability implies the essential uniqueness.

CRB expressions similar to the ones derived in this paper can be also derived for other important special tensor decomposition models such as INDSCAL (where two or more factor matrices coincide) [16], [38], or for the PARALIND model, where the factor matrices have certain structure [22], and for block factorization methods.

Appendix A

Matrix Inversion Lemma (Woodbury identity)
Let $\mathbf{A}$, $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{R}$ are matrices of compatible dimensions such that the following products and inverses exist. Then

$$(\mathbf{A}+\mathbf{XRY})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{X}(\mathbf{R}^{-1}+\mathbf{YA}^{-1}\mathbf{X})^{-1}\mathbf{YA}^{-1} . \tag{59}$$

APPENDIX B

Proof of Theorem 4

Let the matrices $\mathbf{K}$ and $\mathbf{\Psi}$ in (18) be partitioned as

$$\mathbf{K} = \begin{bmatrix} \mathbf{0} & \mathbf{K}_1 \\ \mathbf{K}_1^T & \mathbf{K}_2 \end{bmatrix}, \qquad \mathbf{\Psi} = \begin{bmatrix} \mathbf{\Psi}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{\Psi}_2 \end{bmatrix} \tag{60}$$

where the left-upper blocks have the size $R^2 \times R^2$. Then, using a formula for inverse of partitioned matrices, the left-upper block of $\mathbf{B}$ in (18) can be written as

$$\begin{aligned} \mathbf{B}_0 &= \mathbf{K}_1(\mathbf{I}_{(N-1)R^2} + \mathbf{\Psi}_2\mathbf{K}_2 - \mathbf{\Psi}_2\mathbf{K}_1^T\mathbf{\Psi}_1\mathbf{K}_1)^{-1}\mathbf{\Psi}_2\mathbf{K}_1^T \\ &\stackrel{\triangle}{=} \mathbf{K}_1\mathbf{K}_3^{-1}\mathbf{\Psi}_2\mathbf{K}_1^T \ . \end{aligned} \tag{61}$$

A key observation which enables a fast inversion of the term $\mathbf{K}_3$ is that

$$\mathbf{K} = \mathbf{K}_0 + \mathbf{D}\mathbf{F}\mathbf{D}^T \tag{62}$$

where

$$\mathbf{K}_0 = -\mathtt{bdiag}\left(\mathbf{P}_R\mathbf{F}(\mathtt{dvec}(1 \oslash \mathbf{C}_n))^2\right)_{n=1}^N \tag{63}$$

$$\mathbf{F} = \mathbf{P}_R \prod_{n=1}^N \mathtt{dvec}(\mathbf{C}_n) = \mathbf{P}_R\,\mathtt{dvec}(\mathbf{\Gamma}_{11} \circledast \mathbf{C}_1) \tag{64}$$

$$\mathbf{D} = [\mathtt{dvec}(1 \oslash \mathbf{C}_1), \ldots, \mathtt{dvec}(1 \oslash \mathbf{C}_N)]^T \ . \tag{65}$$

Similarly,

$$\mathbf{K}_2 = \mathbf{K}_{02} + \mathbf{D}_2\mathbf{F}\mathbf{D}_2^T \tag{66}$$

where

$$\mathbf{K}_{02} = -\mathtt{bdiag}\left(\mathbf{P}_R\mathbf{F}(\mathtt{dvec}(1 \oslash \mathbf{C}_n))^2\right)_{n=2}^N \tag{67}$$

$$\mathbf{D}_2 = [\mathtt{dvec}(1 \oslash \mathbf{C}_2), \ldots, \mathtt{dvec}(1 \oslash \mathbf{C}_N)]^T \ . \tag{68}$$

Then the matrix $\mathbf{K}_3$ in (61) can be written as

$$\begin{aligned} \mathbf{K}_3 &= \mathbf{I}_{(N-1)R^2} + \mathbf{\Psi}_2\mathbf{K}_2 - \mathbf{\Psi}_2\mathbf{K}_1^T\mathbf{\Psi}_1\mathbf{K}_1 \\ &= \mathbf{I}_{(N-1)R^2} + \mathbf{\Psi}_2(\mathbf{K}_{02} - \mathbf{K}_1^T\mathbf{\Psi}_1\mathbf{K}_1) + \mathbf{\Psi}_2\mathbf{D}_2\mathbf{F}\mathbf{D}_2^T \\ &= \mathbf{Q} + \mathbf{\Psi}_2\mathbf{D}_2\mathbf{F}\mathbf{D}_2^T \end{aligned} \tag{69}$$

where

$$\mathbf{Q} = \mathtt{bdiag}(\mathbf{Q}_n)_{n=2}^N \tag{70}$$

$$\begin{aligned} \mathbf{Q}_n = \mathbf{I}_{R^2} - (\mathbf{\Gamma}_{nn}^{-1} \otimes \mathbf{X}_n)\mathbf{P}_R\,&(\mathbf{F}(\mathtt{dvec}(1 \oslash \mathbf{C}_n))^2 \\ &+\mathtt{dvec}(\mathbf{\Gamma}_{1n})(\mathbf{\Gamma}_{11}^{-1} \otimes \mathbf{C}_1)\mathtt{dvec}(\mathbf{\Gamma}_{1n})\mathbf{P}_R) \ . \end{aligned} \tag{71}$$

Now, $\mathbf{K}_3$ can be easily inverted using the matrix inversion lemma (59),

$$\mathbf{K}_3^{-1} = \mathbf{Q}^{-1} - \mathbf{Q}^{-1}\mathbf{D}_2^T(\mathbf{I}_{R^2} + \mathbf{D}_2^T\mathbf{Q}^{-1}\mathbf{\Psi}_2\mathbf{D}_2\mathbf{F})^{-1}\mathbf{\Psi}_2\mathbf{D}_2\mathbf{F}\mathbf{Q}^{-1} \ . \tag{72}$$

Inserting (72) in (61) gives, after some simplifications, the result (35). ∎

APPENDIX C

Proof of Theorem 5

Consider the change of scale of columns of factor matrices up to their first columns. As in Section II assume that the scale change is realized in $\mathbf{A}_1$, while the other factor matrices have columns of unit norm. The theorem claims that the substitution $\mathbf{A}_1 \leftarrow \mathbf{A}_1\mathbf{D}$ into (27) where $\mathbf{D} = \mathtt{diag}(1, \lambda_2, \ldots, \lambda_R)$, $\lambda_r \neq 0$, has no influence on CRIB$(\mathbf{a}_1)$.

The substitution $\mathbf{A}_1 \leftarrow \mathbf{A}_1\mathbf{D}$ leads to $\mathbf{C}_1 \leftarrow \mathbf{D}\mathbf{C}_1\mathbf{D}$ and $\mathbf{X}_1 \leftarrow \mathbf{D}\mathbf{X}_1\mathbf{D}$ while $\mathbf{C}_n$ and $\mathbf{X}_n$, $n = 2, \ldots, N$, remain the same. Consequently, $\mathbf{\Gamma}_{1n}$, $n = 1, \ldots, N$, remain unchanged while $\mathbf{\Gamma}_{nn} \leftarrow \mathbf{D}\mathbf{\Gamma}_{nn}\mathbf{D}$ for $n = 2, \ldots, N$. Now, we can substitute into (35) assuming that the condition of Theorem 4 is satisfied.

Let $\widetilde{\mathbf{S}}_n$ denote the matrix $\mathbf{S}_n$ in (39) after the substitution $\mathbf{A}_1 \leftarrow \mathbf{A}_1\mathbf{D}$. It can be shown that $(\mathbf{D} \otimes \mathbf{I}_R)\widetilde{\mathbf{S}}_n = \mathbf{S}_n(\mathbf{D} \otimes \mathbf{I}_R)$ using the rules

$$(\mathbf{D}\mathbf{\Gamma}_{nn}\mathbf{D})^{-1} \otimes \mathbf{X}_n = (\mathbf{D}^{-1} \otimes \mathbf{I}_R)(\mathbf{\Gamma}_{nn}^{-1} \otimes \mathbf{X}_n)(\mathbf{D}^{-1} \otimes \mathbf{I}_R) \tag{73}$$

$$\mathtt{dvec}(\mathbf{D}\mathbf{\Gamma}_{nn}\mathbf{D} \oslash \mathbf{C}_n) = (\mathbf{D} \otimes \mathbf{D})\mathtt{dvec}(\mathbf{\Gamma}_{nn} \oslash \mathbf{C}_n) \tag{74}$$

$$(\mathbf{I}_R \otimes \mathbf{D})\mathbf{P}_R = \mathbf{P}_R(\mathbf{D} \otimes \mathbf{I}_R) \tag{75}$$

and the fact that diagonal matrices commute. Using the same rules in further substitutions, after some computations, the independence of CRIB$(\mathbf{a}_1)$ on $\mathbf{D}$ follows.

APPENDIX D

Proof of Theorem 6

Again, assume for simplicity that all factors have unit norms. It holds

$$\mathbf{\Gamma}_{11} = \begin{bmatrix} 1 & h_1 \\ h_1 & 1 \end{bmatrix}, \quad \mathbf{X}_n = \begin{bmatrix} 0 & 0 \\ 0 & 1 - c_n^2 \end{bmatrix}, \quad n = 1, \ldots, N \ .$$

and

$$g_{11} = [\mathbf{\Gamma}_{11}^{-1}]_{11} = \frac{1}{1 - h_1^2} \tag{76}$$

$$\mathbf{g}_{1,:} = g_{11}\,[1, \quad -h_1] \ . \tag{77}$$

The matrix $\mathbf{\Psi}$ in (32) can be decomposed as $\mathbf{\Psi} = \mathbf{J}\mathbf{\Phi}$ where

$$\mathbf{J} = \mathtt{bdiag}\left(\mathbf{I}_4, \mathbf{I}_2 \otimes [0,\ 1]^T, \ldots, \mathbf{I}_2 \otimes [0,\ 1]^T\right) \tag{78}$$

$$\begin{aligned} \mathbf{\Phi} = \mathtt{bdiag}\big(&\mathbf{\Gamma}_{11}^{-1} \otimes \mathbf{C}_1, (1 - c_2^2)\mathbf{\Gamma}_{22}^{-1} \otimes [0,\ 1], \\ &\ldots, (1 - c_N^2)\mathbf{\Gamma}_{NN}^{-1} \otimes [0,\ 1]\big) \ . \end{aligned} \tag{79}$$

Then the matrix $\mathbf{B}$ in (18) can be rewritten using the Woodbury identity (59) as

$$\begin{aligned} \mathbf{B} &= \mathbf{K}(\mathbf{I}_{4N} + \mathbf{J}\mathbf{\Phi}\mathbf{K})^{-1} \\ &= \mathbf{K} - \mathbf{K}\mathbf{J}(\mathbf{I}_{2N+2} + \mathbf{\Phi}\mathbf{K}\mathbf{J})^{-1}\mathbf{\Phi}\mathbf{K} \ . \end{aligned} \tag{80}$$

Now, put $\mathbf{B}_4 = \mathbf{I}_{2N+2} + \mathbf{\Phi}\mathbf{K}\mathbf{J}$ and write it in the block form as

$$\mathbf{B}_4 = \mathbf{I}_{2N+2} + \mathbf{\Phi}\mathbf{K}\mathbf{J} = \begin{bmatrix} \mathbf{B}_{41} & \mathbf{B}_{42} \\ \mathbf{B}_{43} & \mathbf{B}_{44} \end{bmatrix} \tag{81}$$

where $\mathbf{B}_{41}$ has the size 4×4. The bottom-right block $\mathbf{B}_{44}$ of dimension $(2N-2) \times (2N-2)$ is easy to be inverted using the Woodbury identity again, because it can be written as

$$\mathbf{B}_{44} = \mathbf{B}_5 + \mathbf{s}\mathbf{f}^T \tag{82}$$

where

$$\mathbf{B}_5 = \mathtt{bdiag}(\mathbf{B}_{52}, \ldots, \mathbf{B}_{5N}) \tag{83}$$

$$\mathbf{B}_{5n} = \begin{bmatrix} 1 & -\frac{h_n c_1 (1-c_n^2)}{1-h_n^2 c_1^2} \\ 0 & \frac{c_n^2 - h_n^2 c_1^2}{1-h_n^2 c_1^2} \end{bmatrix}, \qquad n = 2, \ldots, N \tag{84}$$

$$\mathbf{s} = \Big[-\frac{h_2 c_1(1-c_2^2)}{1-h_2^2 c_1^2}, \frac{(1-c_2^2)}{1-h_2^2c_1^2}, \ldots, -\frac{h_N c_1(1-c_N^2)}{1-h_N^2c_1^2}, \frac{(1-c_N^2)}{1-h_N^2 c_1^2}\Big]^T \tag{85}$$

$$\mathbf{f} = [0, 1, 0, 1, \ldots, 1]^T . \tag{86}$$

After some computations, we receive the result (42). ∎

Appendix E

Proof of Theorem 7

Under the assumption of the Theorem, it holds that the matrix $\mathbf{C}_1 = \mathbf{A}_1^T \mathbf{A}_1$ is diagonal and $\mathbf{C}_2 = \mathbf{I}_R$ (identity matrix). Thanks to Theorem 5 we can assume, without any loss of generality, that $\mathbf{C}_1 = \mathbf{I}_R$ as well. It can be shown for $\mathbf{\Gamma}_{mn}$ in (5) that $\mathbf{\Gamma}_{mn} = \mathbf{I}_R$ for all pairs (m,n), $(m,n) \neq (1,2),(2,1)$. Only $\mathbf{\Gamma}_{12}$ and $\mathbf{\Gamma}_{21} = \mathbf{\Gamma}_{12}$ are possibly different. Note that the first row of $\mathbf{\Gamma}_{12}$ is $(1, \gamma_2, \ldots, \gamma_N)$.

It follows from these observations that all non-diagonal $R^2 \times R^2$ blocks $\mathbf{K}_{mn}$ of $\mathbf{K}$ in (6) with $(m,n) \neq (1,2),(2,1)$ are identical, diagonal, having 1 at positions (p,p), $p = 1, R+1, 2R+2, \ldots, R^2$ and 0 elsewhere. In other words, these $\mathbf{K}_{mn}$ can be written as $\mathbf{K}_{mn} = \mathbf{Q}\mathbf{Q}^T$, where $\mathbf{Q}$ is a 0-1 matrix of the size $R^2 \times R$, the $p-$th column of $\mathbf{Q}$ has the value 1 at position $(p-1)(R+1)+1$ and 0 elsewhere.

Computation of the CRIB can proceed from equation (61) by inserting the special form of the blocks of $\mathbf{K}_1$ and $\mathbf{K}_2$ and using the Woodbury identity (59). ∎

Appendix F

Proof of Theorem 8

The following identities are used in this proof

$$\mathrm{vec}(\mathbf{A} \circledast \mathbf{B}) = \mathrm{dvec}(\mathbf{B})\, \mathrm{vec}(\mathbf{A}), \tag{87}$$

$$\mathbf{a}^T \,\mathrm{diag}(\mathbf{b})\, \mathbf{c} = (\mathbf{a} \circledast \mathbf{c})^T \mathbf{b}, \tag{88}$$

$$(\mathbf{a} \otimes \mathbf{b}) \circledast (\mathbf{c} \otimes \mathbf{d}) = (\mathbf{a} \circledast \mathbf{c}) \otimes (\mathbf{b} \circledast \mathbf{d}) . \tag{89}$$

Here, dimensions of $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ and $\mathbf{d}$ are assumed to match accordingly.

The approximate Hessian in (51) is given by

$$\mathbf{H} = \mathbf{J}_W^T(\boldsymbol{\theta})\mathbf{J}_W(\boldsymbol{\theta}) = \mathbf{J}(\boldsymbol{\theta})^T \,\mathrm{dvec}(\boldsymbol{\mathcal{W}})\mathbf{J}(\boldsymbol{\theta}), \tag{90}$$

where $\mathbf{J}(\boldsymbol{\theta})$ is the Jacobian for the complete data.

We have

$$\frac{\partial\, \mathrm{vec}(\boldsymbol{\mathcal{Y}})}{\partial \mathbf{a}_{ir}^{(n)}} = \left(\bigotimes_{k=n+1}^{N} \mathbf{a}_r^{(k)}\right) \otimes \mathbf{e}_i^{(n)} \otimes \left(\bigotimes_{k=1}^{n-1} \mathbf{a}_r^{(k)}\right) \tag{91}$$

where unit vector $\mathbf{e}_i^{(n)}$ for $i = 1, 2, \ldots, I_n$ is the i-th column of the identity matrix of size $I_n \times I_n$.

An (i,j) entry of a sub matrix $\mathbf{H}_{r,s}^{(n,n)}$ for $i = 1, 2, \ldots, I_n$, and $j = 1, 2, \ldots, I_n$ is given by

$$\mathbf{H}_{r,s}^{(n,n)}(i,j) \tag{92}$$

$$= \left(\frac{\partial\, \mathrm{vec}(\boldsymbol{\mathcal{Y}})}{\partial \mathbf{a}_{ir}^{(n)}}\right)^T \mathrm{dvec}\,(\boldsymbol{\mathcal{W}}) \left(\frac{\partial\, \mathrm{vec}(\boldsymbol{\mathcal{Y}})}{\partial \mathbf{a}_{js}^{(n)}}\right)$$

$$= \left(\left(\bigotimes_{k=n+1}^{N} \mathbf{a}_r^{(k)} \circledast \mathbf{a}_s^{(k)}\right) \otimes \left(\mathbf{e}_i^{(n)} \circledast \mathbf{e}_j^{(n)}\right) \otimes \otimes \left(\bigotimes_{k=1}^{n-1} \mathbf{a}_r^{(k)} \circledast \mathbf{a}_s^{(k)}\right)\right)^T \mathrm{vec}(\boldsymbol{\mathcal{W}})$$

$$= \boldsymbol{\mathcal{W}} \,\bar{\times}_{-n} \{\mathbf{b}^{(k)}\} \,\bar{\times}_n \,\delta_{ij}\, \mathbf{e}_i^{(n)} , \tag{93}$$

where δ_{ij} is the Kronecker delta, $\mathbf{b}^{(n)} = \mathbf{a}_r^{(n)} \circledast \mathbf{a}_s^{(n)}$, for $n = 1, \ldots, N$. This leads to that a diagonal sub-matrix $\mathbf{H}_{r,s}^{(n,n)}$ is a diagonal matrix as in Theorem IV.

For off-diagonal sub matrices $\mathbf{H}_{r,s}^{(n,m)}$ of size $I_n \times I_m$ $(1 \leq n < m \leq N)$, we have

$$\mathbf{H}_{r,s}^{(n,m)}(i,j) \tag{94}$$

$$= \left(\frac{\partial\, \mathrm{vec}(\boldsymbol{\mathcal{Y}})}{\partial \mathbf{a}_{ir}^{(n)}}\right)^T \mathrm{dvec}\,(\boldsymbol{\mathcal{W}}) \left(\frac{\partial\, \mathrm{vec}(\boldsymbol{\mathcal{Y}})}{\partial \mathbf{a}_{js}^{(m)}}\right)$$

$$= \left(\left(\bigotimes_{k=m+1}^{N} \mathbf{a}_r^{(k)} \circledast \mathbf{a}_s^{(k)}\right) \otimes \left(\mathbf{a}_r^{(m)} \circledast \mathbf{e}_j^{(m)}\right) \otimes \left(\bigotimes_{k=n+1}^{m-1} \mathbf{a}_r^{(k)} \circledast \mathbf{a}_s^{(k)}\right) \otimes \left(\mathbf{e}_i^{(n)} \circledast \mathbf{a}_s^{(n)}\right) \otimes \left(\bigotimes_{k=1}^{n-1} \mathbf{a}_r^{(k)} \circledast \mathbf{a}_s^{(k)}\right)\right)^T \mathrm{vec}(\boldsymbol{\mathcal{W}})$$

$$= a_{jr}^{(m)}\, a_{is}^{(n)} \left(\boldsymbol{\mathcal{W}} \bar{\times}_{-\{n,m\}} \{\mathbf{b}^{(k)}\} \,\bar{\times}_n \mathbf{e}_i^{(n)} \,\bar{\times}_m \mathbf{e}_j^{(m)}\right). \tag{95}$$

This leads to the compact form in Theorem 8. ∎

References

[1] R. Bro, *Multi-way Analysis in the Food Industry Models, Algorithms, and Applications*, University of Amsterdam, http://www/models.life.ku.dk/research/theses/, 1998.

[2] P.M. Kroonenberg, *Applied Multiway Data Analysis*, Wiley, 2008.

[3] A. Smilde, R. Bro, P. Geladi, *Multi-way Analysis: Applications in the Chemical Sciences*, Wiley, 2004.

[4] A. Cichocki, R. Zdunek, A. H. Phan and S. I. Amari, *Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-way Data Analysis and Blind Source Separation*, Wiley, 2009.

[5] V. De Silva, L.-H. Lim, "Tensor rank and the ill-posedness of the best low-rank approximation problem," *SIAM Journal on Matrix Analysis and Applications*, vol. 30, pp. 1084–1127, 2008.

[6] W.P. Krijnen, T.K. Dijkstra, and A. Stegeman, "On the non-existence of optimal solutions and the occurrence of "degeneracy" in the Candecomp/Parafac model, *Psychometrika*, vol. 73, pp. 431–439, 2008.

[7] R.A. Harshman, "Foundations of the PARAFAC procedure: model and conditions for an "explanatory" multimode factor analysis", *UCLA Working Papers Phonet.* vol. 16, pp. 1-84, 1970.

[8] R.A. Harshman, "Determination and proof of minimum uniqueness conditions for PARAFAC", *UCLA Working Papers Phonet.* vol. 22, pp. 111-117, 1972.

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

[10] J B Kruskal, "Rank, decomposition, and uniqueness for 3-way and N-way arrays" in *Multiway data analysis*, pp. 718, North-Holland (Amsterdam), 1989.

[11] N. D. Sidiropoulos and R. Bro, "On the uniqueness of multilinear decomposition of N-way arrays," *J. Chemometrics*, vol. 14, pp. 229239, May 2000.

[12] J. M. F. Ten Berge, and N.D. Sidiropoulos, "On Uniqueness in CANDECOMP / PARAFAC," *Psychometrika*, Vol. 67, No. 3, pp.399–409, Sep. 2002.

[13] T. Jiang, N.D. Sidiropoulos, "Kruskal's permutation lemma and the identification of Candecomp/Parafac and bilinear models with constant modulus constraints. *IEEE Transactions on Signal Processing*, vol. 52, pp. 2625–2636. (2004)

[14] L. De Lathauwer, "A link between the canonical decomposition in multilinear algebra and simultaneous matrix diagonalization", *SIAM Journal on Matrix Analysis and Applications*, vol. 28, pp. 642–666, 2006.

[15] A. Stegeman, N.D. Sidiropoulos, "On Kruskal's uniqueness condition for the Candecomp/Parafac decomposition", *Linear Algebra and its Applications*, vol. 420, pp. 540–552, 2007.

[16] A. Stegeman, "On uniqueness conditions for Candecomp/Parafac and Indscal with full column rank in one mode," *Linear Algebra and its Applications*, Vol. 431, No. 1–2, pp. 211–227, 2009.

[17] A. Stegeman and A.L.F. de Almeida, "Uniqueness conditions for constrained three-way factor decompositions with linearly dependent loadings," *SIAM Journal on Matrix Analysis and Applications*, Vol. 31, No. 3, pp. 1469–1490, Aug. 2009.

[18] A. Stegeman, "On uniqueness of the n-th order tensor decomposition into rank-1 terms with linear independence in one mode," *SIAM Journal on Matrix Analysis and Applications*, Vol. 31, No. 5, pp. 2498–2516, 2010.

[19] A. Stegeman, "On uniqueness of the canonical tensor decomposition with some form of symmetry," *SIAM Journal on Matrix Analysis and Applications*, Vol. 32, No. 3, pp. 561–583, 2011.

[20] D. Brie, S. Miron, F. Caland and C. Mustin, "An uniqueness condition for the 4-way CANDECOMP/PARAFAC model with collinear loadings in three modes", *Proc. ICASSP 2011*, pp. 4112–4115, 2011.

[21] I. Domanov and L. De Lathauwer, "On the uniqueness of the canonical polyadic decomposition of third-order tensors — Part I: Basic results and uniqueness of one factor matrix", eprint arXiv:1301.4602.
"Part II: Uniqueness of the overall decomposition", eprint arXiv:1301.4603.

[22] X. Guo, S. Miron, D. Brie, and A. Stegeman, "Uni-mode and partial uniqueness conditions for CANDECOMP/PARAFAC of three-way arrays with linearly dependent loadings, " *SIAM Journal on Matrix Analysis and Applications*, Vol. 33, No. 1, pp. 111–129, 2012.

[23] A. Stegeman, "Candecomp/Parafac: from diverging components to a decomposition in block terms," *SIAM Journal on Matrix Analysis and Applications*, 33, 291-316, 2012.

[24] S. Basu and Y. Bresler, "The stability of nonlinear least squares problems and the Cramér-Rao bound," *IEEE Trans. Signal Processing*, vol. 48, pp. 3426-3436, Dec. 2000.

[25] 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, April 2006.

[26] P. Paatero, "A weighted non-negative least squares algorithm for three-way 'PARAFAC' factor analysis", *Chemometrics and Intelligent Laboratory Systems*, vol. 38, pp. 223–242, 1997.

[27] A.-H. Phan, P. Tichavský, and A. Cichocki, "Low Complexity Damped Gauss-Newton Algorithms for CANDECOMP/PARAFAC", *SIAM, SIMAX* (accepted for publication), available at http://arxiv.org/abs/1205.2584.

[28] P. Tichavský and Z. Koldovský, "Stability of CANDECOMP-PARAFAC tensor decomposition", *Proc. ICASSP 2011*, Prague, Czech Republic, pp. 4164-4167, 2011.

[29] P. Tichavský and Z. Koldovský, "Weight Adjusted Tensor Method for Blind Separation of Underdetermined Mixtures of Nonstationary Sources, " *IEEE Trans. on Signal Processing*, Vol. 59, No. 3, pp. 1037–1047, March 2011.

[30] X. Liu, and N.D. Sidiropoulos, "Cramér-Rao Lower Bounds for Low-rank Decomposition of Multidimensional Arrays," *IEEE Trans. on Signal Processing*, vol. 49, no. 9, pp. 2074–2086, Sep. 2001.

[31] T. Jiang, and N.D. Sidiropoulos, "Blind Identification of Out of Cell Users in DS-CDMA", *EURASIP Journal on Applied Signal Processing (JASP)*, special issue on Advances in Smart Antennas, 2004, no. 9, pp. 1212–1224, Aug. 2004.

[32] Z. Koldovský, P. Tichavský, and Anh Huy Phan, "Stability Analysis and Fast Damped-Gauss-Newton Algorithm for INDSCAL Tensor Decomposition," *Proc. of IEEE Workshop on Statistical Signal Processing*, pp. 581–584, Nice, France, June 2011.

[33] A.-H. Phan, P. Tichavský, and A. Cichocki, "CANDECOMP/PARAFAC Decomposition of High-order Tensors Through Tensor Reshaping", submitted.

[34] R. Bro, R. A. Harshman, N. D. Sidiropoulos, and M. E. Lundy, "Modeling multi-way data with linearly dependent loadings," *Journal of Chemometrics*, vol. 23, no. 7–8, pp. 324-340, 2009.

[35] M. B. Dosse, J. M.F. Berge, and J. N. Tendeiro, "Some new results on orthogonally constrained candecomp," Journal of Classification, vol. 28, pp. 144-155, 2011.

[36] M. Sorensen, L. De Lathauwer, P. Comon, S. Icart, and L. Deneire, " Canonical Polyadic Decomposition with a Columnwise Orthonormal Factor Matrix," *SIAM Journal on Matrix Analysis and Applications*, Vol. 33, No. 4, pp. 1190–1213, 2012.

[37] J. Chen and Y. Saad, "On the Tensor SVD and the Optimal Low Rank Orthogonal Approximation of Tensors", Vol. 30, pp. 1709–1734, 2009.

[38] J. Tendeiro, M. Bennani Dosse, and J.M.F. Ten Berge, "First and second-order derivatives for CP and INDSCAL". *Chemometrics and Intelligent Laboratory Systems*, vol. 106, pp.27–36, 2011.

[39] A. Householder, *The Theory of Matrices in Numerical Analysis*. New York: Blaisdell Publishing Co., 1964.

[40] G. Tomasi and R. Bro, "PARAFAC and missing values", *Chemometrics and Intelligent Laboratory Systems*, vol. 75, pp. 163–180, 2005.

[41] R. Bro, "Exploratory study of sugar production using fluorescence spectroscopy and multi-way analysis", *Chemom. Intell. Lab. Syst.*, vol 46, pp. 133–147, 1999.

[42] R. Bro, *Multi-way Analysis in the Food Industry - Models, Algorithms, and Applications*, PhD thesis, University of Amsterdam, Holland, 1998.

[43] B. Porat, *Digital Processing of Random Signals*, Prentice Hall, 1994.

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

[45] P. Comon, J.M.F. ten Berge, L. De Lathauwer and J. Castaing, "Generic and Typical Ranks of Multi-Way Arrays", *Linear Algebra and its Applications*, vol. 430, no. 11 (2009), pp. 2997-3007, 2009.

[46] T.G. Kolda, "Orthogonal tensor decompositions", *SIAM Journal of Matrix Analysis and Applications*, vol. 23, pp. 243–255, 2001.

[47] P. Tichavský, A.H. Phan, and A. Cichocki, "A further improvement of a fast damped Gauss–Newton algorithm for CP tensor decomposition", submitted to *ICASSP 2013*.

TABLE I
ESTIMATED CRIBS [dB] ON BEST FIT CP COMPONENTS OF FLUORESCENCE TENSOR COMPUTED FOR ASSUMED RANK $R = 1, 2, 3, 4$

Factor	$R=1$	$R=2$		$R=3$			$R=4$			
n	1	1	2	1	2	3	1	2	3	4
1	44.43	44.44	41.87	**64.76**	**61.34**	**64.98**	65.78	60.96	65.77	38.17
2	27.44	30.28	27.71	**53.15**	**50.17**	**49.60**	54.33	51.39	50.87	23.29
3	32.67	36.23	33.66	**58.96**	**55.75**	**54.87**	60.25	56.28	54.27	25.74

Petr Tichavský (M'98, SM'04) received the M.S. degree in mathematics in 1987 from the Czech Technical University, Prague, Czechoslovakia and the Ph.D. degree in theoretical cybernetics from the Czechoslovak Academy of Sciences in 1992. Since that time he is with the Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic in Prague. In 1994 he received the *Fulbright grant* for a 10 month fellowship at Yale University, Department of Electrical Engineering, in New Haven, U.S.A. In 2002 he received the *Otto Wichterle Award* from Academy of Sciences of the Czech Republic.

He is author and co-author of research papers in the area of sinusoidal frequency/frequency-rate estimation, adaptive filtering and tracking of time varying signal parameters, algorithm-independent bounds on achievable performance, sensor array processing, independent component analysis and blind source separation.

Petr Tichavský served as associate editor of the *IEEE Signal Processing Letters* from 2002 to 2004, and as associate editor of the *IEEE Transactions on Signal Processing* from 2005 to 2009 and from 2011 to now. Since 2009 he is a member of the IEEE Signal Processing Society's Signal Processing Theory and Methods (SPTM) Technical Committee. Petr Tichavský has also served as a general co-chair of the 36th IEEE Int. Conference on Acoustics, Speech and Signal Processing ICASSP 2011 in Prague, Czech Republic.

Anh-Huy Phan received the master degree from Hochiminh University of Technology, Vietnam in 2005, and the Ph.D degree from the Kyushu Institute of Technology, Japan in 2011. He worked as Deputy head of Research and Development Department, Broadcast Research and Application Center, Vietnam Television, and is currently Research Scientist at the Laboratory for Advanced Brain Signal Processing, and a visiting research scientist in TOYOTA Collaboration Center, Brain Science Institute, RIKEN. His research interests include multilinear algebra, tensor computation, blind source separation and brain computer interface.

Zbyněk Koldovský (S'03-M'04) was born in Jablonec nad Nisou, Czech Republic, in 1979. He received the M.S. degree and Ph.D. degree in mathematical modeling from Faculty of Nuclear Sciences and Physical Engineering at the Czech Technical University in Prague in 2002 and 2006, respectively.

He is currently an associate professor at the Institute of Information Technology and Electronics, Technical University of Liberec. He has also been with the Institute of Information Theory and Automation of the Academy of Sciences of the Czech Republic since 2002. His main research interests are focused on audio signal processing, blind source separation, statistical signal processing and multilinear algebra.