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. Insightful expressions are derived for tensors of rank 1 and rank 2.

The existence of the bound reveals necessary conditions for essential uniqueness of the CP decomposition and, moreover, for identifiability of each column of each factor matrix separately. The results can be used for checking stability of a given decomposition of a tensor, and for evaluating performance of certain approximate CP decomposition methods based on reshaping the tensor.

Index Terms

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

I. Introduction

Three-way and higher-way data arrays need to be analyzed in diverse research areas such as chemistry, astronomy, and psychology. 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

[0]This work was supported by the Grant Agency of the Czech Republic through the project 102/09/1278.

[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.

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 [1].

An important issue is the essential uniqueness of CP decomposition as it entails identifiability of the model (the factor matrices) from the tensor. A sufficient condition was derived by Kruskal in [3]. Recently, the problem has been addressed again, namely by Stegeman, Ten Berge, De Lathauwer, Sidiropoulos et al.; see [4]-[11]. Uniqueness of the CP decomposition for four-way collinear tensors that have identical components was studied in [12].

Necessary conditions for the uniqueness of CP were derived in [13]. The stability of CP was investigated in [13] and more recently in [14], through the Cramér-Rao lower bound (CRLB) on an unbiased estimation of the factor matrices given a noisy observation of a tensor. The former paper presents the bound in the form of expressions that offer almost no insight into the problem, because the bound is derived for normalized factor matrices and for three and four-way tensors. In [14], the stability is studied in terms of a CRLB-induced bound (CRIB) on squared angular deviation of columns of the factor matrices with respect to their nominal values, which is much more practical, but the study is limited to the case of three-way tensors. Similar results for symmetric tensors are derived in [15].

This paper presents new CRIB expressions for tensors of arbitrary dimension, and specialized expressions for rank 1 and rank 2 tensors. The results can be used for checking stability of given CP decomposition of a tensor, and for evaluating performance of certain approximate CP decomposition methods based on reshaping the tensor.

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. Section IV is devoted to a possible application of the bound: investigation of loss in 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 for a fluorescence tensor, and stability of the tensor of Brie *et al*. 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 it can be written as a sum of R rank-one tensors (factors)

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

where $\circ$ denotes the outer tensor product, $\mathbf{a}_r^{(n)}$, $r = 1, \ldots, R$, $n = 1, \ldots, N$ are vectors of the length I_n.

Such tensor 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 $n = 1, \ldots, N$.

In practice, a CP decomposition of given rank (R) is used as an approximation of a given tensor, which can be a noisy observation $\hat{\boldsymbol{\mathcal{Y}}}$ of the tensor $\boldsymbol{\mathcal{Y}}$ in (1). Owing to the symmetry of (1), we can focus on estimating the first factor matrix $\mathbf{A}_1$, without any loss in 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 square criterion

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

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 [14]

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

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, $\mathbf{H}(\boldsymbol{\theta}) = \mathbf{J}^T(\boldsymbol{\theta})\mathbf{J}(\boldsymbol{\theta})$.

Let $\boldsymbol{\Gamma}_{nm}$ be defined as a 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 [18]: 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

$$\mathbf{K}_{nm} = (1 - \delta_{nm})\mathbf{P}_R\,\mathtt{diag}\,(\mathtt{vec}\,\boldsymbol{\Gamma}_{nm}) \tag{7}$$

$\mathbf{P}_R$ is the permutation matrix of dimension $R^2 \times R^2$ defined in [18] such that $\mathtt{vec}\,\mathbf{X} = \mathbf{P}_R\,\mathtt{vec}\,(\mathbf{X}^T)$ for any $R \times R$ matrix $\mathbf{X}$, and δ_{nm} is the Kronecker delta. Next,

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

and

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

where $\otimes$ denotes the Kronecker product, $\mathbf{I}_{I_n}$ is an identity matrix of the size $I_n \times I_n$, and $\mathtt{blkdiag}(\cdot)$ 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 [21], [19]. 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

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

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 parameter of the 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 [16] 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,

$$\mathbf{G}_E = \mathtt{blkdiag}\left(\boldsymbol{\Gamma}_{11} \otimes \mathbf{I}_1, \boldsymbol{\Gamma}_{22} \otimes (\mathbf{E}_2\mathbf{E}_2^T), \ldots, \boldsymbol{\Gamma}_{NN} \otimes (\mathbf{E}_N\mathbf{E}_N^T)\right) \tag{15}$$

$$\mathbf{G}_E^{-1} = \mathtt{blkdiag}\left((\boldsymbol{\Gamma}_{11})^{-1} \otimes \mathbf{I}_1, \boldsymbol{\Gamma}_{22}^{-1} \otimes (\mathbf{E}_2\mathbf{E}_2^T)^{-1}, \ldots, \boldsymbol{\Gamma}_{NN}^{-1} \otimes (\mathbf{E}_N\mathbf{E}_N^T)^{-1}\right) \ . \tag{16}$$

Put

$$\boldsymbol{\Psi} = \mathbf{Z}_E^T\mathbf{G}_E^{-1}\mathbf{Z}_E \tag{17}$$

$$\mathbf{B} = \mathbf{K}(\mathbf{I}_{NR^2} + \boldsymbol{\Psi}\mathbf{K})^{-1} \tag{18}$$

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 $\boldsymbol{\Gamma}_{11}^{-1}$, respectively. Then

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

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 [17] 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.

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 will be denoted CRIB($\mathbf{a}_1$) in the sequel.

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

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

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 [17]: 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.
Theorem 3: The CRIB($\mathbf{a}_1$) can be written in the form

$$\mathrm{CRIB}(\mathbf{a}_1) = \frac{\sigma^2}{\|\mathbf{a}_1\|^2}\left\{(I_1-1)g_{11} - \mathrm{tr}\left[\mathbf{B}_0\left((\mathbf{g}_{1,:}^T\mathbf{g}_{1,:})\otimes\mathbf{X}_1\right)\right]\right\} \tag{26}$$

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

$$\mathbf{X}_n = \mathbf{C}_n - \frac{1}{\mathbf{C}^{(n)}_{11}}\mathbf{C}^{(n)}_{:,1}\mathbf{C}^{(n)\,T}_{:,1} \tag{27}$$

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

$$\mathbf{\Psi} = \mathtt{blkdiag}\left(\mathbf{\Gamma}_{11}^{-1}\otimes\mathbf{C}_1, \mathbf{\Gamma}_{22}^{-1}\otimes\mathbf{X}_2, \ldots, \mathbf{\Gamma}_{NN}^{-1}\otimes\mathbf{X}_N\right) \ . \tag{28}$$

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

$$\begin{aligned}
\mathrm{CRIB}(\mathbf{a}_1) &= \frac{\sigma^2}{\|\mathbf{a}_1\|^2}\mathrm{tr}\left[\Pi^{\perp}_{\mathbf{a}_1}\left(g_{11}\mathbf{I}_{I_1} - (\mathbf{g}_{1,:}\otimes\mathbf{A}_1)\,\mathbf{B}_0\,(\mathbf{g}_{1,:}\otimes\mathbf{A}_1)^T\right)\right] \\
&= \frac{\sigma^2}{\|\mathbf{a}_1\|^2}\left\{(I_1-1)g_{11} - \mathrm{tr}\left[\Pi^{\perp}_{\mathbf{a}_1}(\mathbf{g}_{1,:}\otimes\mathbf{A}_1)\,\mathbf{B}_0\,(\mathbf{g}_{1,:}\otimes\mathbf{A}_1)^T\right]\right\} \\
&= \frac{\sigma^2}{\|\mathbf{a}_1\|^2}\left\{(I_1-1)g_{11} - \mathrm{tr}\left[\mathbf{B}_0\left((\mathbf{g}_{1,:}^T\mathbf{g}_{1,:})\otimes\left(\mathbf{A}_1^T\Pi^{\perp}_{\mathbf{a}_1}\mathbf{A}_1\right)\right)\right]\right\} \ .
\end{aligned} \tag{29}$$

This is (26), because

$$\mathbf{A}_1^T\Pi^{\perp}_{\mathbf{a}_1}\mathbf{A}_1 = \mathbf{C}_1 - \frac{1}{\mathbf{C}^{(1)}_{11}}\mathbf{C}^{(1)}_{:,1}\mathbf{C}^{(1)\,T}_{:,1} = \mathbf{X}_1 \ . \tag{30}$$

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

$$\mathbf{\Psi} = \mathbf{Z}_E^T\mathbf{G}_E^{-1}\mathbf{Z}_E = \mathtt{blkdiag}\left(\mathbf{\Gamma}_{11}^{-1}\otimes\mathbf{C}_1, \mathbf{\Gamma}_{22}^{-1}\otimes\widetilde{\mathbf{X}}_2, \ldots, \mathbf{\Gamma}_{NN}^{-1}\otimes\widetilde{\mathbf{X}}_N\right) \tag{31}$$

where

$$\widetilde{\mathbf{X}}_n = \mathbf{A}_n^T\mathbf{E}_n^T(\mathbf{E}_n\mathbf{E}_n^T)^{-1}\mathbf{E}_n\mathbf{A}_n \tag{32}$$

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

$$\Pi^{\perp}_{\mathbf{a}_1^{(n)}} = \mathbf{I}_{I_n} - \mathbf{a}_1^{(n)}\mathbf{a}_1^{(n)\,T}/\|\mathbf{a}_1^{(n)}\|^2 \tag{33}$$

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

$$\mathbf{B}_0 = [-\mathbf{I}_{R^2} + \mathbf{V}(\mathbf{I}_{R^2} + \mathbf{V})^{-1}]\mathbf{Y} \tag{34}$$

where

$$\mathbf{V} = \mathbf{W} - \mathbf{Y}(\mathbf{\Gamma}_{11}^{-1} \otimes \mathbf{C}_1) \tag{35}$$

$$\mathbf{W} = \mathbf{P}_R \sum_{n=2}^{N} \mathtt{dvec}(\mathbf{\Gamma}_{1n})\mathbf{S}_n^{-1}(\mathbf{\Gamma}_{nn}^{-1} \otimes \mathbf{X}_n)\mathtt{dvec}(\mathbf{C}_1 \oslash \mathbf{C}_n) \tag{36}$$

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

$$\mathbf{S}_n = \mathbf{I}_{R^2} - (\mathbf{\Gamma}_{nn}^{-1} \otimes \mathbf{X}_n)\mathtt{dvec}(\mathbf{\Gamma}_{nn} \oslash \mathbf{C}_n)\mathbf{P}_R, \qquad n = 2, \ldots, N\ . \tag{38}$$

Here, $\mathtt{dvec}(\mathbf{M})$ 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, and "$\oslash$" 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 $R^2 \times R^2$.

It can be shown that the first R rows and the first R columns in $\mathbf{B}_0$ are zeros. This fact indicates possibility of further simplifications and computational savings.

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

Because of Corollary 2 in the previous section, for simplicity we shall assume in this section that $\|\mathbf{a}_1\|^2/\sigma^2 = 1$.

A. Rank 1 tensors

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

$$\mathrm{CRIB}(\mathbf{a}_1) = (I_1 - 1)g_{11} = I_1 - 1 \ . \tag{39}$$

In (39), $g_{11} = 1$ due to the convention that the factor matrices $\mathbf{A}_n$, $n \geq 2$, have columns of unit norm.

B. Rank 2 tensors

Due to Corollary 2, we can assume, without any loss in generality, that all factor vectors have unit norm. Let c_n be the correlation of the factor vectors in the $n-$th mode. It means that the matrices $\mathbf{C}_n$ have the form

$$\mathbf{C}_n = \begin{bmatrix} 1 & c_n \\ c_n & 1 \end{bmatrix}, \quad n = 1, \ldots, N \ .$$

Due to Theorem 5 we note that the CRIB on $\mathbf{a}_1$ is only a function of the correlations $c_1, \ldots, c_N$. It is symmetric function in $c_2, \ldots, c_N$ and possibly nonsymmetric in c_1. In Appendix D, we derive the following result that does not need the assumption $c_n \neq 0$, $n = 2, \ldots, N$, unlike Theorem 4.

Theorem 6 It holds for rank 2 tensors

$$\mathrm{CRIB}(\mathbf{a}_1) = \frac{I_1 - 1}{1 - h_1^2} + \frac{(1 - c_1^2)h_1^2}{1 - h_1^2} \, \frac{y^2 + z - h_1^2 z(z+1)}{(1 - c_1 y - h_1^2(z+1))^2 - h_1^2(y + c_1 z)^2} \tag{40}$$

where

$$h_n = \prod_{2 \leq k \neq n}^{N} c_n \qquad \text{for} \quad n = 1, \ldots, N \tag{41}$$

$$y = -c_1 \sum_{n=2}^{N} \frac{h_n^2(1 - c_n^2)}{c_n^2 - h_n^2 c_1^2} \tag{42}$$

$$z = \sum_{n=2}^{N} \frac{1 - c_n^2}{c_n^2 - h_n^2 c_1^2} \ . \tag{43}$$

Proof: See Appendix C.

Note that the expressions (42)-(43) 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 (40).

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

$$\mathrm{CRIB}_{N=3}(\mathbf{a}_1) = \frac{1}{1-c_2^2c_3^2}\left[I_1 - 1 + \frac{c_2^2}{1-c_2^2} + \frac{c_3^2}{1-c_3^2}\right] \ . \tag{44}$$

The above result coincides with the one derived in [14]. As far as the stability is concerned, the CRIB is finite unless either the second or third factor have co-linear columns.

For $N = 4$, the similar result is hardly tractable. Unlike the case $N = 3$, the result depends on the correlation of the columns in the first factor matrix, which is 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

$$\mathrm{CRIB}_{N=4,c_1=0}(\mathbf{a}_1) = \frac{I_1-1}{1-c_2^2c_3^2c_4^2} + \frac{c_2^2c_3^2 + c_2^2c_4^2 + c_3^2c_4^2 - 3c_2^2c_3^2c_4^2}{(1-c_2^2c_3^2c_4^2)(2c_2^2c_3^2c_4^2 - c_2^2c_3^2 - c_2^2c_4^2 - c_3^2c_4^2 + 1)} \tag{45}$$

$$\mathrm{CRIB}_{N=4,c_1=\pm1}(\mathbf{a}_1) = \begin{cases} \frac{I_1-1}{1-c_2^2c_3^2c_4^2} & \text{for} \quad (|c_2|<1)\&(|c_3|<1)\&(|c_4|<1) \\ \frac{I_1-1}{1-c_2^2c_3^2} + \frac{c_2^2+c_3^2-2c_2^2c_3^2}{(1-c_2^2)(1-c_3^2)(1-c_2^2c_3^2)} & \text{for} \quad |c_4| = 1 \\ \frac{I_1-1}{1-c_2^2c_4^2} + \frac{c_2^2+c_4^2-2c_2^2c_4^2}{(1-c_2^2)(1-c_4^2)(1-c_2^2c_4^2)} & \text{for} \quad |c_3| = 1 \\ \frac{I_1-1}{1-c_3^2c_4^2} + \frac{c_3^2+c_4^2-2c_3^2c_4^2}{(1-c_3^2)(1-c_4^2)(1-c_3^2c_4^2)} & \text{for} \quad |c_2| = 1 \end{cases} \tag{46}$$

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 $c_1 = 0$

$$\mathrm{CRIB}_{c_1=0}(\mathbf{a}_1) = \frac{I_1-1}{1-h_1^2} + \frac{h_1^2 z}{(1-h_1^2)(1-h_1^2(z+1))} \ . \tag{47}$$

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 3-5 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 $\boldsymbol{\mathcal{W}}$ of the same dimension as $\boldsymbol{\mathcal{Y}}$, 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

$$\mathbf{H} = \mathbf{J}^T(\boldsymbol{\theta})\mathbf{J}(\boldsymbol{\theta}), \qquad \mathbf{J}(\boldsymbol{\theta}) = \frac{\partial \mathtt{vec}(\boldsymbol{\mathcal{Y}} \circledast \boldsymbol{\mathcal{W}})}{\partial \boldsymbol{\theta}} \tag{48}$$

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

Theorem 7: 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

$$\mathbf{H}_{r,s}^{(n,m)} = \begin{cases} \mathtt{diag}\left(\boldsymbol{\mathcal{W}} \,\bar{\times}_{-n} \left\{\mathbf{a}_r^{(1)} \circledast \mathbf{a}_s^{(1)}, \cdots, \mathbf{a}_r^{(N)} \circledast \mathbf{a}_s^{(N)}\right\}\right) \quad , & n = m, \\ (\mathbf{a}_r^{(n)}\, \mathbf{a}_s^{(m)T}) \circledast \left(\boldsymbol{\mathcal{W}} \,\bar{\times}_{-\{n,m\}} \left\{\mathbf{a}_r^{(1)} \circledast \mathbf{a}_s^{(1)}, \cdots, \mathbf{a}_r^{(N)} \circledast \mathbf{a}_s^{(N)}\right\}\right), & n \neq m \end{cases} \tag{49}$$

$\boldsymbol{\mathcal{Y}} \bar{\times}_n \mathbf{u}_n$ denotes the mode-n tensor-vector product between $\boldsymbol{\mathcal{Y}}$ and $\mathbf{u}_n$ [1], and

$$\boldsymbol{\mathcal{Y}} \bar{\times}_{-n} \{\mathbf{u}\} = \boldsymbol{\mathcal{Y}} \,\bar{\times}_1 \mathbf{u}_1 \cdots \bar{\times}_{n-1} \mathbf{u}_{n-1} \bar{\times}_{n+1} \mathbf{u}_{n+1} \cdots \bar{\times}_N \mathbf{u}_N. \tag{50}$$

V. Application

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 easier to decompose, so that the first factor matrix remains unchanged. For example, consider $N = 4$.

The tensor in (1) can be reshaped to a 3-way tensor

$$\boldsymbol{\mathcal{Y}}_{res} = \sum_{r=1}^{R} \mathbf{a}_r^{(1)} \circ \mathbf{a}_r^{(2)} \circ (\mathbf{a}_r^{(4)} \otimes \mathbf{a}_r^{(3)}) \ . \tag{51}$$

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 latter 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 (29).

Let us examine tensors of rank 2. If the original tensor has correlations between columns of the factor matrices c_1, c_2, c_3 and c_4, the reshaped tensor has correlations 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).

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

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

$$\frac{c_2^2 + c_3^2 c_4^2 - 2c_2^2 c_3^2 c_4^2}{(1-c_2^2)(1-c_3^2 c_4^2)} - \frac{c_2^2 c_3^2 + c_2^2 c_4^2 + c_3^2 c_4^2 - 3c_2^2 c_3^2 c_4^2}{(1-c_2^2 c_3^2 c_4^2)(2c_2^2 c_3^2 c_4^2 - c_2^2 c_3^2 - c_2^2 c_4^2 - c_3^2 c_4^2 + 1)}$$

and the largest difference is

$$\frac{c_2^2 + c_3^2 c_4^2 - 2c_2^2 c_3^2 c_4^2}{(1-c_2^2)(1-c_3^2 c_4^2)} = \frac{c_2^2}{1-c_2^2} + \frac{c_3^2 c_4^2}{1-c_3^2 c_4^2} \ .$$

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 $c_3^2 c_4^2 \approx 1$) .

VI. EXAMPLES

A. 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 [23]. 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 [18]. CRIBs on the extracted components were then computed with the noise levels deduced from the error tensor $\mathcal{E} = \mathcal{Y} - \hat{\mathcal{Y}}$

$$\sigma^2 = \frac{\|\mathcal{Y} - \hat{\mathcal{Y}}\|_F^2}{\prod_n I_n}. \tag{52}$$

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

Note that a CRIB of 50 dB means that the standard angular deviation (square root of mean square angular error) of the factor is cca $0.18^{\rm o}$; a CRIB of 20 dB corresponds to the standard deviation $5.7^{\rm o}$.

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.

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 (52) 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 means 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 [18] 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 in accuracy of the decomposition is only about 1-2 dB.

B. *Uniqueness of the decomposition of Brie's tensor*

Brie *et al* [12] 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 this property by computing the CRIB.

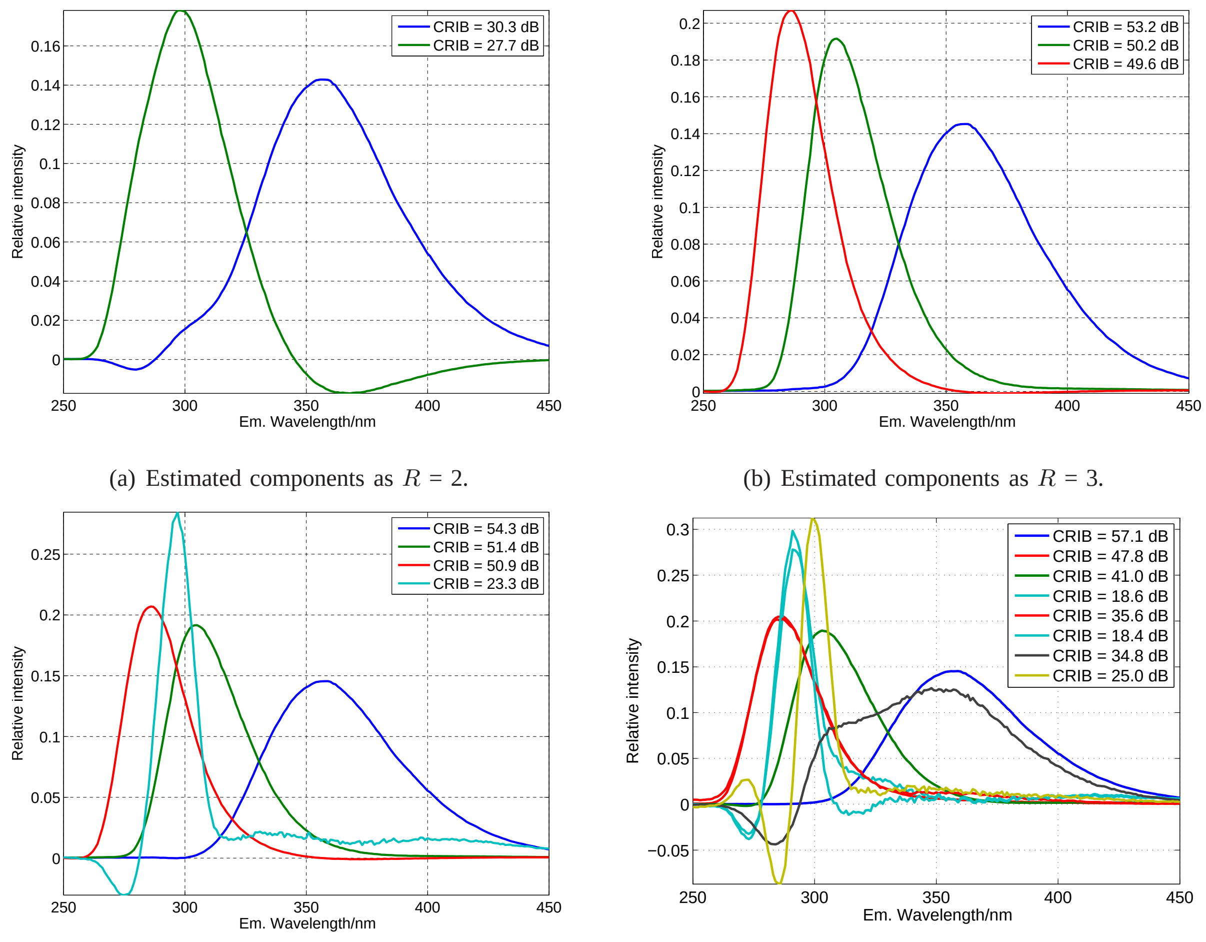
The factor matrices of the tensor have the form

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

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 $\mathbf{a}_1$ 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 correlation matrices $\mathbf{C}_n$, $n = 1, \ldots, 4$, have the form

$$\mathbf{C}_1 = \begin{bmatrix} 1 & c_{11} & c_{12} \\ c_{11} & 1 & c_{13} \\ c_{12} & c_{13} & 1 \end{bmatrix}, \mathbf{C}_2 = \begin{bmatrix} 1 & 1 & c_2 \\ 1 & 1 & c_2 \\ c_2 & c_2 & 1 \end{bmatrix}, \mathbf{C}_3 = \begin{bmatrix} 1 & c_3 & 1 \\ c_3 & 1 & c_3 \\ 1 & c_3 & 1 \end{bmatrix}, \mathbf{C}_4 = \begin{bmatrix} 1 & c_4 & c_4 \\ c_4 & 1 & 1 \\ c_4 & 1 & 1 \end{bmatrix} .$$

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) 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.

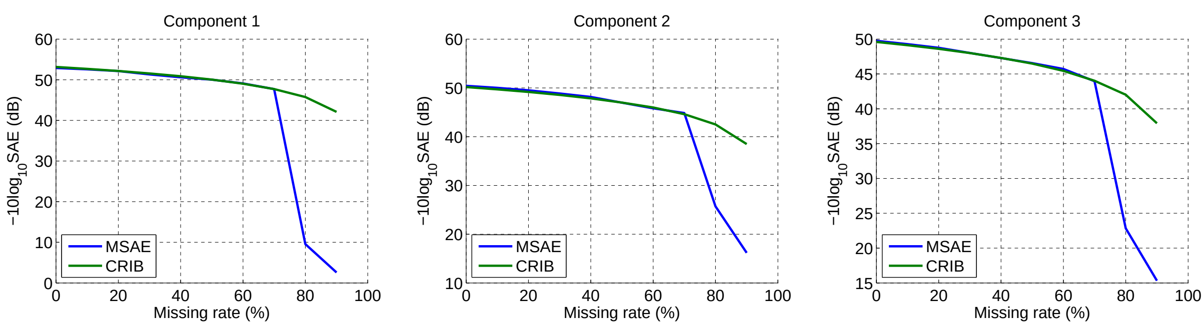


Fig. 2. CRIB for the second-mode components of CP decomposition of tensor in section VI.A with missing elements and mean square angular error obtained in simulations versus percentage of the missing elements.

a limit. The limit CRIB is computed for modified matrices at $\varepsilon \to 0$,

$$\mathbf{C}_{2\varepsilon} = \left[\begin{array}{ccc} 1 & 1-\varepsilon & c_2 \\ 1-\varepsilon & 1 & c_2 \\ c_2 & c_2 & 1 \end{array}\right], \mathbf{C}_{3\varepsilon} = \left[\begin{array}{ccc} 1 & c_3 & 1-\varepsilon \\ c_3 & 1 & c_3 \\ 1-\varepsilon & c_3 & 1 \end{array}\right], \mathbf{C}_{4\varepsilon} = \left[\begin{array}{ccc} 1 & c_4 & c_4 \\ c_4 & 1 & 1-\varepsilon \\ c_4 & 1-\varepsilon & 1 \end{array}\right].$$

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 c_{11}, c_{12}, c_{13}, unless any of these three correlations equals ± 1 (see a similar situation in (46)). Assume that they are all different. The result, obtained by Symbolic Matlab, is

$$\begin{aligned} \mathrm{CRIB}_{\varepsilon=0}(\mathbf{a}_1) \quad = \quad & \frac{1}{2c_2^2c_3^2c_4^2 - c_2^2c_3^2 - c_2^2c_4^2 - c_3^2c_4^2 + 1}\Bigg[(I_1-1)(1-c_2^2c_3^2) \\ & -\frac{c_3^4(c_2^2+1) - 3c_3^2+1}{1-c_3^2} - \frac{c_2^4(c_3^2+1)-3c_2^2+1}{1-c_2^2} + \frac{2-c_2^2-c_3^2}{1-c_4^2}\Bigg] . \end{aligned} \tag{53}$$

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

VII. Conclusions

In this paper, Cramér-Rao bounds for CP tensor decomposition are studied. Several novel expressions of the bound are derived, namely expression for a general dimension and general rank of the tensor, and specialized expressions for rank 2 tensors.

CRIB might be useful for assessing accuracy of the tensor decomposition in the presence of additive noise. It is a good indicator of stability and uniqueness of the decomposition. The tensor may have missing entries, and still have a stable CP decomposition.

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

$$(\mathbf{A} + \mathbf{XRY})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{X}(\mathbf{R}^{-1} + \mathbf{Y}\mathbf{A}^{-1}\mathbf{X})^{-1}\mathbf{Y}\mathbf{A}^{-1} . \tag{54}$$

Appendix B

Proof of Theorem 4

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

$$\mathbf{K} = \left[\begin{array}{cc} \mathbf{0} & \mathbf{K}_1 \\ \mathbf{K}_1^T & \mathbf{K}_2 \end{array}\right], \qquad \mathbf{\Psi} = \left[\begin{array}{cc} \mathbf{\Psi}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{\Psi}_2 \end{array}\right] \tag{55}$$

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

$$\mathbf{B}_0 = \mathbf{K}_1(\mathbf{I}_{(N-1)R^2} + \mathbf{\Psi}_2\mathbf{K}_2 - \mathbf{\Psi}_2\mathbf{K}_1^T\mathbf{\Psi}_1\mathbf{K}_1)^{-1}\mathbf{\Psi}_2\mathbf{K}_1^T \stackrel{\triangle}{=} \mathbf{K}_1\mathbf{K}_3^{-1}\mathbf{\Psi}_2\mathbf{K}_1^T \ . \tag{56}$$

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

$$\mathbf{K} \quad = \quad \mathbf{K}_0 + \mathbf{D}\mathbf{F}\mathbf{D}^T \tag{57}$$

where

$$\begin{aligned}
\mathbf{K}_0 \quad &= \quad -\mathtt{blkdiag}\left(\mathbf{P}_R\mathbf{F}(\mathtt{dvec}(1 \oslash \mathbf{C}_n))^2\right)_{n=1}^N && (58)\\
\mathbf{F} \quad &= \quad \mathbf{P}_R \prod_{n=1}^N \mathtt{dvec}(\mathbf{C}_n) = \mathbf{P}_R \, \mathtt{dvec}(\mathbf{\Gamma}_{11} * \mathbf{C}_1) && (59)\\
\mathbf{D} \quad &= \quad [\mathtt{dvec}(1 \oslash \mathbf{C}_1), \ldots, \mathtt{dvec}(1 \oslash \mathbf{C}_N)]^T \ . && (60)
\end{aligned}$$

Similarly,

$$\mathbf{K}_2 \quad = \quad \mathbf{K}_{02} + \mathbf{D}_2\mathbf{F}\mathbf{D}_2^T \tag{61}$$

where

$$\begin{aligned}
\mathbf{K}_{02} \quad &= \quad -\mathtt{blkdiag}\left(\mathbf{P}_R\mathbf{F}(\mathtt{dvec}(1 \oslash \mathbf{C}_n))^2\right)_{n=2}^N && (62)\\
\mathbf{D}_2 \quad &= \quad [\mathtt{dvec}(1 \oslash \mathbf{C}_2), \ldots, \mathtt{dvec}(1 \oslash \mathbf{C}_N)]^T \ . && (63)
\end{aligned}$$

Then the matrix $\mathbf{K}_3$ in (56) can be written as

$$\begin{aligned}
\mathbf{K}_3 \quad &= \quad \mathbf{I}_{(N-1)R^2} + \mathbf{\Psi}_2\mathbf{K}_2 - \mathbf{\Psi}_2\mathbf{K}_1^T\mathbf{\Psi}_1\mathbf{K}_1 \\
&= \quad \mathbf{I}_{(N-1)R^2} + \mathbf{\Psi}_2(\mathbf{K}_{02} - \mathbf{K}_1^T\mathbf{\Psi}_1\mathbf{K}_1) + \mathbf{\Psi}_2\mathbf{D}_2\mathbf{F}\mathbf{D}_2^T \\
&= \quad \mathbf{Q} + \mathbf{\Psi}_2\mathbf{D}_2\mathbf{F}\mathbf{D}_2^T && (64)
\end{aligned}$$

where

$$\begin{aligned}
\mathbf{Q} \quad &= \quad \mathtt{blkdiag}(\mathbf{Q}_n)_{n=2}^N && (65)\\
\mathbf{Q}_n \quad &= \quad \mathbf{I}_{R^2} - (\mathbf{\Gamma}_{nn}^{-1} \otimes \mathbf{X}_n)\mathbf{P}_R\left(\mathbf{F}(\mathtt{dvec}(1 \oslash \mathbf{C}_n))^2 + \mathtt{dvec}(\mathbf{\Gamma}_{1n})(\mathbf{\Gamma}_{11}^{-1} \otimes \mathbf{C}_1)\mathtt{dvec}(\mathbf{\Gamma}_{1n})\mathbf{P}_R\right) && (66)
\end{aligned}$$

Now, $\mathbf{K}_3$ can be easily inverted using the matrix inversion lemma (54),

$$\mathbf{K}_3^{-1} \quad = \quad \mathbf{Q}^{-1} - \mathbf{Q}^{-1}\mathbf{D}_2^T(\mathbf{I}_{R^2} + \mathbf{D}_2^T\mathbf{Q}^{-1}\mathbf{\Psi}_2\mathbf{D}_2\mathbf{F})^{-1}\mathbf{\Psi}_2\mathbf{D}_2\mathbf{F}\mathbf{Q}^{-1} \ . \tag{67}$$

Inserting (67) in (56) gives, after some simplifications, the result (34). ■

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 (26) 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 (34) assuming that the condition of Theorem 4 is satisfied.

Let $\widetilde{\mathbf{S}}_n$ denote the matrix $\mathbf{S}_n$ in (38) 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

$$(\mathbf{D}\mathbf{\Gamma}_{nn}\mathbf{D})^{-1} \otimes \mathbf{X}_n = (\mathbf{D}^{-1} \otimes \mathbf{I}_R)(\mathbf{\Gamma}_{nn}^{-1} \otimes \mathbf{X}_n)(\mathbf{D}^{-1} \otimes \mathbf{I}_R) \tag{68}$$

$$\mathtt{dvec}(\mathbf{D}\mathbf{\Gamma}_{nn}\mathbf{D} \oslash \mathbf{C}_n) = (\mathbf{D} \otimes \mathbf{D})\mathtt{dvec}(\mathbf{\Gamma}_{nn} \oslash \mathbf{C}_n) \tag{69}$$

$$(\mathbf{I}_R \otimes \mathbf{D})\mathbf{P}_R = \mathbf{P}_R(\mathbf{D} \otimes \mathbf{I}_R) \tag{70}$$

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

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

$$g_{11} \;=\; [\mathbf{\Gamma}_{11}^{-1}]_{11} = \frac{1}{1 - h_1^2} \tag{71}$$

$$\mathbf{g}_{1,:} \;=\; g_{11}\,[1, \quad -h_1] \ . \tag{72}$$

The matrix $\mathbf{\Psi}$ in (31) can be decomposed as $\mathbf{\Psi} = \mathbf{J}\mathbf{\Phi}$ where

$$\mathbf{J} \;=\; \mathtt{blkdiag}\left(\mathbf{I}_4, \mathbf{I}_2 \otimes [0,\ 1]^T, \ldots, \mathbf{I}_2 \otimes [0,\ 1]^T\right) \tag{73}$$

$$\mathbf{\Phi} \;=\; \mathtt{blkdiag}\left(\mathbf{\Gamma}_{11}^{-1} \otimes \mathbf{C}_1, (1 - c_2^2)\mathbf{\Gamma}_{22}^{-1} \otimes [0,\ 1], \ldots, (1 - c_N^2)\mathbf{\Gamma}_{NN}^{-1} \otimes [0,\ 1]\right) \ . \tag{74}$$

Then the matrix $\mathbf{B}$ in (18) can be rewritten using the Woodbury identity (54) as

$$\mathbf{B} \;=\; \mathbf{K}(\mathbf{I}_{4N} + \mathbf{J}\mathbf{\Phi}\mathbf{K})^{-1} = \mathbf{K} - \mathbf{K}\mathbf{J}(\mathbf{I}_{2N+2} + \mathbf{\Phi}\mathbf{K}\mathbf{J})^{-1}\mathbf{\Phi}\mathbf{K} \ . \tag{75}$$

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

$$\mathbf{B}_4 \quad = \quad \mathbf{I}_{2N+2} + \mathbf{\Phi}\mathbf{K}\mathbf{J} = \begin{bmatrix} \mathbf{B}_{41} & \mathbf{B}_{42} \\ \mathbf{B}_{43} & \mathbf{B}_{44} \end{bmatrix} \tag{76}$$

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

$$\mathbf{B}_{44} \quad = \quad \mathbf{B}_5 + \mathbf{s}\mathbf{f}^T \tag{77}$$

where

$$\mathbf{B}_5 \quad = \quad \texttt{blkdiag}\,(\mathbf{B}_{52}, \ldots, \mathbf{B}_{5N}) \tag{78}$$

$$\mathbf{B}_{5n} \quad = \quad \begin{bmatrix} 1 & -\frac{h_n c_1(1-c_n^2)}{1-h_n^2 c_1^2} \\ 0 & \frac{c_n^2 - h_n^2 c_1^2}{1-h_n^2 c_1^2} \end{bmatrix}, \qquad n = 2, \ldots, N \tag{79}$$

$$\mathbf{s} \quad = \quad \left[-\frac{h_2 c_1(1-c_2^2)}{1-h_2^2 c_1^2}, \frac{(1-c_2^2)}{1-h_2^2 c_1^2}, \ldots, -\frac{h_N c_1(1-c_N^2)}{1-h_N^2 c_1^2}, \frac{(1-c_N^2)}{1-h_N^2 c_1^2}\right]^T \tag{80}$$

$$\mathbf{f} \quad = \quad [0,\, 1,\, 0,\, 1, \ldots, 1]^T\ . \tag{81}$$

After some computations, we receive the result (40). ■

References

[1] 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.

[2] L. Albera, A. Ferreol, P. Comon, and P. Chevalier, "Blind Identification of Overcomplete Mixtures of Sources (BIOME)," *Lin. Algebra Appl.*, vol. 391, pp. 1–30, Nov. 2004.

[3] 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.

[4] A. Stegeman, J.M.F. Ten Berge, L. De Lathauwer, "Sufficient Conditions for Uniqueness in Candecomp/Parafac and Indscal with Random Component Matrices", *Psychometrika*, Vol. 71, No. 2, pp. 219-229, June 2006.

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

[6] 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.

[7] 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.

[8] 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.

[9] 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.

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

[11] 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.

[12] 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.

[13] 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.

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

[15] 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.

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

[17] 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.

[18] A.-H. Phan, P. Tichavský, and A. Cichocki, “Low complexity damped Gauss-Newton algorithms for parallel factor analysis”, *SIAM, SIMAX* (submitted), available at http://arxiv.org/abs/1205.2584.

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

[20] 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.

[21] 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.

[22] R. Bro (1999), Exploratory study of sugar production using fluorescence spectroscopy and multi-way analysis. Chemom. Intell. Lab. Syst, 46, 133–147.

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