STABILITY OF CANDECOMP-PARAFAC TENSOR DECOMPOSITION

Petr Tichavský¹ and Zbyněk Koldovský¹,²

¹Institute of Information Theory and Automation, P.O.Box 18, 182 08 Prague 8, Czech Republic
²Faculty of Mechatronics, Informatics and Interdisciplinary Studies Technical University of Liberec, Studentská 2, 461 17 Liberec, Czech Republic

ABSTRACT

In this paper, stability of the CANDECOMP-PARAFAC (CP) tensor decomposition is addressed. It is done by deriving the Cramér-Rao lower bound (CRLB) on variance of an unbiased estimate of the tensor parameters, i.e., elements of its factor matrices, from its noisy observation (the tensor plus a random Gaussian i.i.d. tensor). The existence of the bound reveals necessary conditions for essential uniqueness of the CP decomposition, moreover, for identifiability of each column of each factor matrix separately. Analytical closed-form expressions of the bound are derived for 3 way tensors of rank 1 and 2. As a byproduct, a novel computationally efficient expression for the inverse of the approximate Hessian matrix is derived.

Index Terms—Multilinear models; canonical polyadic decomposition; Cramér-Rao lower bound

1. INTRODUCTION

Three-way and higher-way data arrays need to be analyzed in many research areas such as chemistry, astronomy, or even psychology. Parallel factor analysis (PARAFAC), or Canonical decomposition (CANDECOMP), or CP, is an extension of a low rank decomposition of matrices to higher way arrays, usually called tensors.

An important issue is the essential uniqueness of CP as it entails identifiability of the factor matrices from the tensor. A sufficient condition was derived by Kruskal in [1]. Recently, the problem has been addressed again, for instance, Stegeman et al. derived a condition that is closer to the necessity; see [2] and references therein.

In this paper, we study this issue by analyzing the local stability of the CP tensor decomposition. A tensor, determined by its factor matrices, is modified by adding a Gaussian distributed random noise independently to each of its elements. Stability of its CP decomposition means roughly saying that a small change of the tensor elements does not cause a large change in the CP decomposition. The stability is studied using the Cramér-Rao bound on the estimation of the factor matrices as parameters of the distribution of tensor elements. Finiteness of the bound points to necessary conditions for the identifiability of individual columns of factor matrices. The CRB for the CP decomposition has been studied already in [3]. However, in that paper, no closed-form CRB expressions are available.

In this paper, analytical closed-form expressions for the bound on mean square angle deviations of columns of the factor matrices are derived for 3 way tensors of rank 1 and 2. These expressions imply conditions on stability of the CP decomposition of 3 way tensors. As a byproduct, a novel computationally efficient expression for the inverse of the approximate Hessian matrix is derived.

2. PROBLEM FORMULATION

For simplicity, we restrict our presentation to three-way tensors, although an extension to higher way tensors is straightforward.

Assume that a three way tensor \( \mathbf{X} \) of the dimension \( I \times J \times K \) has elements

\[
X_{ijk} = \sum_{f=1}^{r} A_{ij} B_{jf} C_{kf} \tag{1}
\]

where \( A_{ij}, B_{jf}, \) and \( C_{kf} \), are elements of factor matrices \( \mathbf{A}, \mathbf{B}, \) and \( \mathbf{C} \), respectively, that have dimensions \( I \times r, J \times r \) and \( K \times r \), and their \( k \)th columns will be denoted by \( a_k, b_k \) and \( c_k \), respectively; \( r \) is the rank of \( \mathbf{X} \).

Assume that a noisy observation of the tensor \( \mathbf{X} \) is given,

\[
\mathbf{Y} = \mathbf{X} + \mathbf{E} \tag{2}
\]

where \( \mathbf{E} \) is a tensor of the same dimensions as \( \mathbf{X} \). Assume that elements of \( \mathbf{E} \) are independent Gaussian distributed random variables with zero mean and variance \( \sigma^2 \). The estimation problem is to find the factor matrices \( \mathbf{A}, \mathbf{B}, \) and \( \mathbf{C} \) from the noisy observation \( \mathbf{Y} \).

There is an inherent permutation and scale uncertainty in the problem. For the permutation ambiguity, we assume that
the order of columns of the estimated \( \hat{\mathbf{A}} \), \( \hat{\mathbf{B}} \), and \( \hat{\mathbf{C}} \), matches that of \( \mathbf{A} \), \( \mathbf{B} \), and \( \mathbf{C} \). To cope with the scale ambiguity of the factors, we shall only study the angular differences between the columns of these matrices. For example, the angle between the \( k \)th column of \( \hat{\mathbf{A}} \) and \( \mathbf{A} \) is defined through its cosine as

\[
\cos \alpha_k = \frac{|\hat{a}_k^T a_k|}{\|a_k\| \|\hat{a}_k\|}
\]

\( k = 1, \ldots, r \). Similarly, the angular deviations of columns of \( \hat{\mathbf{B}} \) and \( \hat{\mathbf{C}} \) can be defined.

Let a parameter vector \( \theta \) contain all parameters of our model. Let it be arranged as

\[
\theta = [\theta_1^T, \ldots, \theta_r^T]^T
\]

where \( \theta_k = [a_k^T, b_k^T, c_k^T]^T \). The maximum likelihood estimate of \( \theta \) consists in minimizing the least square criterion

\[
Q(\theta) = \left\| \text{vec} \left[ \mathbf{Y} - \mathbf{X}(\theta) \right] \right\|^2_2
\]

so it can be obtained by any algorithm that minimizes \( Q(\theta) \).

We wish to compute the Cramér-Rao lower bound for estimating \( \theta \). In general, for this estimation problem, the CRLB is given as the inverse of the Fisher information matrix (FIM), which is equal to \[7\]

\[
\mathbf{F}(\theta) = \frac{1}{\sigma^2} \mathbf{J}^T(\theta) \mathbf{J}(\theta)
\]

where \( \mathbf{J}(\theta) \) is the Jacobian matrix (matrix of the first-order derivatives) of \( \text{vec} \left[ \mathbf{X}(\theta) \right] \) with respect to \( \theta \). The FIM is proportional to \( \mathbf{H}(\theta) = \mathbf{J}^T(\theta) \mathbf{J}(\theta) \), which is an approximate Hessian matrix of \( Q(\theta) \) that occurs in Gauss-Newton or Levenberg-Marquardt optimization algorithms; see e.g. \[4\], or more recently \[5, 6\].

The CRLB of one column of one factor matrix can be found as an appropriate diagonal submatrix of the inverse of FIM. We derive the bound for \( a_k \) only, since then the bounds for other columns of all factor matrices follow thanks to the problem symmetry.

## 3. Cramér-Rao induced bound

It should be noted that the FIM (and the Hessian) is singular in our case because of the scale ambiguity problem. It is possible to fix scales of columns in two of three factor matrices, which reduces the number of free parameters to \( r(I + J + K - 2) \). We use another approach here: \( \mathbf{H}(\theta) \) is regularized by adding \( \mu \mathbf{I} \) to it, and a Cramér-Rao induced lower bound (CRLB) of the mean square angular deviation is derived for \( \mu \to 0 \), as follows.

Let \( \text{CRLB}(a_k) \) be the submatrix of \( \mathbf{F}^{-1} \) which bounds the mean square error in estimating \( a_k \). The angle \( \alpha_k \) between \( a_k \) and \( \hat{a}_k \) is defined through its cosine as

\[
\cos \alpha_k = \frac{|\hat{a}_k^T a_k|}{\|a_k\| \|\hat{a}_k\|} = \frac{x + \varepsilon}{\sqrt{x(x + 2\varepsilon + \nu)}},
\]

\( x = a_k^T a_k \), \( \varepsilon = a_k^T \Delta a_k \), \( \nu = \Delta a_k^T \Delta a_k \) and \( \Delta a_k = \hat{a}_k - a_k \). Taking the second-order Taylor series expansion on both sides of \[7\] and neglecting all higher-order terms of \( \omega, \varepsilon \) and \( \nu \) we get

\[
1 - \frac{1}{2} \alpha_k^2 = 1 + \frac{1}{x^2} \left( x - \frac{1}{2} x^2 - \frac{1}{2} \nu \right).
\]

Therefore

\[
\alpha_k^2 = \frac{x \nu - \varepsilon^2}{x^2} = \frac{1}{x^2} [x \Delta a_k^T \Delta a_k - a_k^T \Delta a_k \Delta a_k^T a_k].
\]

and consequently

\[
\begin{align*}
\mathbf{E}[\alpha_k^2] &= \frac{1}{x^2} \left\{ x \mathbf{E}[\Delta a_k^T \Delta a_k] - a_k^T \mathbf{E}[\Delta a_k \Delta a_k^T] a_k \right\} \\
&= \frac{1}{x^2} \{ x \mathbf{E}[\text{tr}(\Delta a_k \Delta a_k^T)] - a_k^T \mathbf{E}[\Delta a_k \Delta a_k^T] a_k \}.
\end{align*}
\]

If \( \hat{a}_k \) is the maximum likelihood estimate of \( a_k \), it holds asymptotically that \( \mathbf{E}[\Delta a_k \Delta a_k^T] = \text{CRLB}(a_k) \). Combining this fact with \[10\] it follows that the CRLB on the mean square angle deviation of \( \hat{a}_k \) can be defined as

\[
\begin{align*}
\text{CRLB}(\alpha_k^2) &= \frac{\text{tr}[\text{CRLB}(a_k)]}{\|a_k\|^2} - \frac{a_k^T \text{CRLB}(a_k) a_k}{\|a_k\|^4} \\
&= \frac{\text{tr}[\Pi_{a_k}^+ \text{CRLB}(a_k)]}{\|a_k\|^2}.
\end{align*}
\]

where

\[
\Pi_{a_k}^+ = \mathbf{I} - a_k a_k^T / \|a_k\|^2
\]

is the projection operator to the orthogonal complement of \( a_k \). It easily follows that the CRLB is always non-negative.

## 4. Analytical inversion of Hessian matrix

The Jacobian matrix and the Hessian matrix of the criterion for the 3-way tensor was derived in \[5\]. Similar expression for a general \( n \)-way tensor can be found in \[6\]. For the 3-way tensor it was shown that \( \mathbf{H}(\theta) \) can be partitioned into \( r \times r \) blocks of the size \( I + J + K \),

\[
\mathbf{H}(\theta) = \begin{bmatrix} \mathbf{H}_{11} & \ldots & \mathbf{H}_{1r} \\ \vdots & \ddots & \vdots \\ \mathbf{H}_{r1} & \ldots & \mathbf{H}_{rr} \end{bmatrix}
\]

where the \((j, i)\)th block can be written as

\[
\mathbf{H}_{ji} = \begin{bmatrix} \beta_{ij} \gamma_{ij} \mathbf{I}_j & \gamma_{ij} a_i b_j^T & \beta_{ij} a_i c_j^T \\ \gamma_{ij} b_i a_j^T & \alpha_{ij} \gamma_{ij} \mathbf{I}_j & \alpha_{ij} b_i c_j^T \\ \beta_{ij} c_i a_j^T & \alpha_{ij} c_i b_j^T & \alpha_{ij} \beta_{ij} \mathbf{I}_K \end{bmatrix}
\]

for \( i, j = 1, \ldots, F \). Next, \( \mathbf{I}_J, \mathbf{I}_J, \mathbf{I}_K \) stand for identity matrices of the dimension \( I, J \) and \( K \), respectively.
will be skipped in the sequel) and \(\alpha_{ij}, \beta_{ij}\) and \(\gamma_{ij}\) is the \((ij)\)th element of \(A^T A, B^T B,\) and \(C^T C,\) respectively.

In order to find a closer-form expression for the inverse of \(H(\theta) + \mu I,\) note that the blocks of the Hessian matrix can be written in the generic form

\[
H_{ji} = \begin{bmatrix}
  z_{ij} + M_{ij}^{AA} A^T & A M_{ij}^{AB} B^T & A M_{ij}^{AC} C^T \\
  B M_{ij}^{BA} A^T & z_{ij} + B M_{ij}^{BB} B^T & B M_{ij}^{BC} C^T \\
  C M_{ij}^{CA} A^T & C M_{ij}^{CB} B^T & z_{ij} + C M_{ij}^{CC} C^T 
\end{bmatrix}
\]  

(15)

where, by comparing (14) with (15), we get \(M_{ij}^{AA} = M_{ij}^{BB} = M_{ij}^{CC} = 0,\) \(M_{ij}^{AB} = (M_{ij}^{BA})^T = \gamma_{ij} e_i e_j^T,\) \(M_{ij}^{AC} = (M_{ij}^{CA})^T = \beta_{ij} e_i e_j^T,\) \(y_{ij} = \alpha_{ij} \gamma_{ij} + \mu \delta_{ij},\) \(z_{ij} = \alpha_{ij} \beta_{ij} + \mu \delta_{ij}\) for \(i,j = 1, \ldots, r,\) where \(e_k\) is the \(k\)th column of the \(r \times r\) identity matrix, \(k = 1, \ldots, r,\) and \(\delta_{ij}\) is the Kronecker’s delta.

Now, under the assumption that \(r \leq \min\{I, J, K\},\) the inverse of \(H(\theta) + \mu I\) can be sought in the same generic form. Let the inverse be partitioned into blocks \(\overline{H}_{ji}\) with the same structure as \(H_{ji},\) with constants \(\tau_{ij}, \gamma_{ij}\) and \(\tau_{ij}\) and matrices \(\overline{M}_{ij}^{PQ}, P, Q \in \{A, B, C\}.\) The expressions for these constants and matrices are derived in Appendix.

5. CRIB IN CLOSED FORMS

The computation of the CRIB on \(a_1\) is now straightforward. It can be found as the limit

\[
\text{CRIB}(\alpha_{12}^2) = \sigma^2 \lim_{\mu \to 0} \frac{1}{\|A_1\|^2} \text{tr}[\Pi_{A_1} \overline{H}_{\mu}]
\]

(16)

where \(\overline{H}_{\mu}\) is the left-upper diagonal block of \((H(\theta) + \mu I)^{-1}\) of the size \(I \times I,\) which is equal to \(\overline{H}_{\mu} = \tau_{11} I + A \overline{M}_{11}^{AA} A^T\) (the dependence of the right-hand side on \(\mu\) is not explicitly shown). Then,

\[
\text{tr}[\Pi_{A_1} \overline{H}_{\mu}] = \text{tr}[\Pi_{A_1} (\tau_{11} I + A \overline{M}_{11}^{AA} A^T)]
\]

\[
= \tau_{11} \text{tr}[\Pi_{A_1}] + \text{tr} \left[(A^T \Pi_{A_1} A) \overline{M}_{11}^{AA} \right]
\]

\[
= (I - 1) \tau_{11} + \text{tr} \left[(A^T \Pi_{A_1} A) \overline{M}_{11}^{AA} \right].
\]

(17)

Note that

\[
A \Pi_{A_1} A^T = \text{diag} \left(0, \alpha_{22} - \frac{\alpha_{12}^2}{\alpha_{11}}, \ldots, \alpha_{rr} - \frac{\alpha_{1r}^2}{\alpha_{11}} \right).
\]

(18)

5.1. Rank 1 tensors

In this case, \(\tau_{11} = (\beta_{11} \gamma_{11} + \mu)^{-1}\) and therefore

\[
\text{CRIB}(\alpha_{12}^2) = (I - 1) \frac{\sigma^2}{\alpha_{11} \beta_{11} \gamma_{11}}.
\]

(19)

5.2. Rank 2 tensors

In this case,

\[
\tau_{11} = \left(\frac{[B^T B] \odot (C^T C) + \mu I^{-1}}{11} \right) = \frac{\beta_{22} \gamma_{22}}{d_{BC}} + O(\mu)
\]

where \(d_{BC} = \text{det}[B^T B] \odot (C^T C)\) and \(\odot\) denotes the element-wise product. The matrix \(\overline{M}_{11}^{AA}\) can be obtained by solving the 12 \(\times\) 12 linear system (22). The \((2,2)\)th element of this matrix reads

\[
(\overline{M}_{11}^{AA})_{22} = \frac{\alpha_{11} \beta_{22} \gamma_{22}}{d_A d_B d_C d_{BC}} + O(\mu)
\]

where \(d_A = \text{det}[A^T A], d_B = \text{det}[B^T B],\) and \(d_C = \text{det}[C^T C].\) Therefore

\[
\text{CRIB}(\alpha_{12}^2) = \frac{\sigma^2}{\alpha_{11}} \left[ (I - 1) \tau_{11} + \frac{d_A}{\alpha_{11}} (\overline{M}_{11}^{AA})_{22} \right]
\]

\[
= \frac{\sigma^2 \beta_{22} \gamma_{22}}{\alpha_{11} d_{BC}} \left[ (I - 1) + \frac{\beta_{12}^2}{d_B} + \frac{\gamma_{12}^2}{d_C} \right].
\]

(21)

Since \(d_{BC} \geq d_B d_C\) [8], we can see that the tensor decomposition is unstable with respect to estimating the factor matrix \(A,\) that means that the CRIB is infinite, if \(d_B = 0\) or \(d_C = 0\) (i.e., if any of the factor matrices \(B\) and \(C\) has colinear columns).

5.2.1. Example

We generated random orthogonal factor matrices \(A, B\) and \(C\) of dimensions \(4 \times 2, 5 \times 2\) and \(6 \times 2,\) respectively. The first column of \(A\) was modified as \(a_1 \leftarrow \lambda a_1 + (1 - \lambda) a_2\) with \(\lambda \in [0, 1].\) In each trial, a noisy observation of the rank 2 tensor was generated according to (2) with \(\sigma = 0.01,\) and its CP decomposition was computed using the LM2 method from [5]. Note that for \(\lambda = 0,\) the Kruskal’s sufficient condition for essential uniqueness of the CP decomposition is not fulfilled.
The results shown in Fig. 1 demonstrate good agreement between the performance of the maximum likelihood estimates via LM2 and the CRIB. The example also shows that the rows of A are well identified even if \( \lambda = 0 \), which is in accordance with the necessary condition provided by the CRIB.

6. CONCLUSIONS

We have derived explicit forms of inversion of the Hessian matrix of the multilinear mapping that describes the CP factorization. These expressions can be used for determining whether a CP factorization of a tensor is stable or not. We have shown that the inverse of the Hessian matrix can be performed in \( O(r^6) \) operations, where \( r \) is the rank of the tensor, regardless of the size of the factor matrices, only the products \( A^T A, B^T B \) and \( C^T C \) are needed.

The analysis of the CRIB suggests that for stable estimation of one factor matrix in the decomposition, all other factor matrices must not have co-linear columns. The factor matrix of the interest may or may not contain colinear columns and still can be estimable in a stable way.

The CRIB for tensors of rank higher than two and more-than-three-way tensors can be treated numerically by analyzing the corresponding Hessian matrix.

Appendix

Inverse of the Hessian matrix can be found via the equation

\[
\sum_{k=1}^{r} \left[ \frac{1}{\delta_{ij}} \right]_{ik} \sum_{k=1}^{r} \left[ \begin{array}{ccc}
\tilde{x}_{kj} \delta_{ij} + \tilde{M}_{ik}^A A^T B M_{ij}^{BA} + \tilde{M}_{ik}^A A^T C M_{ij}^{CA} \\
B M_{ij}^{BA} C M_{ij}^{CA} \\
C M_{ij}^{CA}
\end{array} \right] = \delta_{ij}
\]

Theorem 3. If \( \sum_{k=1}^{r} x_{kj} \tilde{x}_{ik} = \delta_{ij} \) for \( i, j = 1, \ldots, r \), and hence the \( r \times r \) matrix \( \tilde{X} = (\tilde{x}_{ij})_{i,j=1}^{r} \) equals to the inverse of \( X = (x_{ij})_{i,j=1}^{r} \). Moreover, the matrices \( \tilde{M}_{ik}^A, \tilde{M}_{ik}^{BA} \) and \( \tilde{M}_{ik}^{CA} \) for \( i, k = 1, \ldots, r \) are solutions of the \( 3r^2 \times 3r^2 \) linear systems

\[
\sum_{k=1}^{r} (x_{kj} I + \tilde{M}_{ik}^A A^T A) \tilde{M}_{ik}^A + \sum_{k=1}^{r} (M_{kj}^{BA} B^T B) \tilde{M}_{ik}^{BA} + \sum_{k=1}^{r} (M_{kj}^{AC} C^T C) \tilde{M}_{ik}^{CA} = -\sum_{k=1}^{r} \tilde{x}_{ik} M_{kj}^{A^T}
\]
\[
\sum_{k=1}^{r} (M_{kj}^{BA} A^T A) \tilde{M}_{ik}^{BA} + \sum_{k=1}^{r} (y_{kj} I + M_{kj}^{BA} B^T B) \tilde{M}_{ik}^{BA} + \sum_{k=1}^{r} (M_{kj}^{BC} C^T C) \tilde{M}_{ik}^{CA} = -\sum_{k=1}^{r} \tilde{x}_{ik} M_{kj}^{BA}
\]
\[
\sum_{k=1}^{r} (M_{kj}^{CA} A^T A) \tilde{M}_{ik}^{CA} + \sum_{k=1}^{r} (M_{kj}^{CB} B^T B) \tilde{M}_{ik}^{CA} + \sum_{k=1}^{r} (z_{kj} I + M_{kj}^{CC} C^T C) \tilde{M}_{ik}^{CA} = -\sum_{k=1}^{r} \tilde{x}_{ik} M_{kj}^{CA}
\]

for \( j = 1, \ldots, r \). Solution of each of the linear systems requires \( O(r^6) \) operations, so that the whole inverse of the Hessian matrix can be done in the same number of operations plus \( O(r^2 (I + J + K)) \) operations needed to compute the products \( A^T A, B^T B \) and \( C^T C \).

7. REFERENCES