STABILITY OF CANDECOMP-PARAFAC TENSOR DECOMPOSITION

Petr Tichavský[1] *and Zbyněk Koldovský*[1,2]

[1]Institute of Information Theory and Automation,
P.O.Box 18, 182 08 Prague 8, Czech Republic
[2]Faculty 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 element. 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 ranks 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 $\underline{\mathbf{X}}$ of the dimension $I \times J \times K$ has elements

$$X_{ijk} = \sum_{f=1}^{r} A_{if} B_{jf} C_{kf} \tag{1}$$

where A_{if}, 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 kth columns will be denoted by $\mathbf{a}_k$, $\mathbf{b}_k$ and $\mathbf{c}_k$, respectively; r is the rank of $\underline{\mathbf{X}}$.

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

$$\underline{\mathbf{Y}} = \underline{\mathbf{X}} + \underline{\mathbf{E}} \tag{2}$$

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

There is an inherent permutation and scale uncertainty in the problem. For the permutation ambiguity, we assume that

[0]This work was supported by Ministry of Education, Youth and Sports of the Czech Republic through the project 1M0572 and by Grant Agency of the Czech Republic through the project 102/09/1278.

the order of columns of the estimated $\widehat{\mathbf{A}}$, $\widehat{\mathbf{B}}$, and $\widehat{\mathbf{C}}$, 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 kth column of $\widehat{\mathbf{A}}$ and $\mathbf{A}$ is defined through its cosine as

$$\cos\alpha_k = \frac{|\widehat{\mathbf{a}_k}^T\mathbf{a}_k|}{\|\widehat{\mathbf{a}_k}\|\,\|\mathbf{a}_k\|} \tag{3}$$

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

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

$$\boldsymbol{\theta} = [\boldsymbol{\theta}_1^T,\ldots,\boldsymbol{\theta}_r^T]^T \tag{4}$$

where $\boldsymbol{\theta}_k = [\mathbf{a}_k^T, \mathbf{b}_k^T, \mathbf{c}_k^T]^T$. The maximum likelihood estimate of $\boldsymbol{\theta}$ consists in minimizing the least square criterion

$$\mathcal{Q}(\boldsymbol{\theta}) = \left\|\mathrm{vec}[\underline{\mathbf{Y}} - \underline{\mathbf{X}}(\boldsymbol{\theta})]\right\|_2^2, \tag{5}$$

so it can be obtained by any algorithm that minimizes $\mathcal{Q}(\boldsymbol{\theta})$.

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 (FIM), which is equal to [7]

$$\mathbf{F}(\boldsymbol{\theta}) = \frac{1}{\sigma^2}\,\mathbf{J}^T(\boldsymbol{\theta})\mathbf{J}(\boldsymbol{\theta}) \tag{6}$$

where $\mathbf{J}(\boldsymbol{\theta})$ is the Jacobi matrix (matrix of the first-order derivatives) of $\mathrm{vec}[\underline{\mathbf{X}}(\boldsymbol{\theta})]$ with respect to $\boldsymbol{\theta}$. The FIM is proportional to $\mathbf{H}(\boldsymbol{\theta}) = \mathbf{J}^T(\boldsymbol{\theta})\mathbf{J}(\boldsymbol{\theta})$, which is an approximate Hessian matrix of $\mathcal{Q}(\boldsymbol{\theta})$ 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 $\mathbf{a}_1$ 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: $\mathbf{H}(\boldsymbol{\theta})$ is regularized by adding $\mu\mathbf{I}$ to it, and a Cramér-Rao induced lower bound (CRIB) of the mean square angular deviation is derived for $\mu \to 0$, as follows.

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

$$\cos\,\alpha_k = \frac{|\widehat{\mathbf{a}_k}^T\mathbf{a}_k|}{\|\widehat{\mathbf{a}_k}\|\,\|\mathbf{a}_k\|} = \frac{x+\varepsilon}{\sqrt{x(x+2\varepsilon+\nu)}} \tag{7}$$

where $x = \mathbf{a}_k^T\mathbf{a}_k$, $\varepsilon = \mathbf{a}_k^T\Delta\mathbf{a}_k$, $\nu = \Delta\mathbf{a}_k^T\Delta\mathbf{a}_k$, and $\Delta\mathbf{a}_k = \widehat{\mathbf{a}_k} - \mathbf{a}_k$. Taking the second-order Taylor series expansion on both sides of (7) and neglecting all higher-order terms of ω, ε and ν we get

$$1 - \frac{1}{2}\alpha_k^2 = 1 + \frac{1}{2}\frac{\varepsilon^2}{x^2} - \frac{1}{2}\frac{\nu}{x}\,. \tag{8}$$

Therefore

$$\alpha_k^2 = \frac{x\nu - \varepsilon^2}{x^2} = \frac{1}{x^2}[x\Delta\mathbf{a}_k^T\Delta\mathbf{a}_k - \mathbf{a}_k^T\Delta\mathbf{a}_k\Delta\mathbf{a}_k^T\mathbf{a}_k] \tag{9}$$

and consequently

$$\begin{aligned}\mathrm{E}[\alpha_k^2] &= \frac{1}{x^2}\{x\mathrm{E}[\Delta\mathbf{a}_k^T\Delta\mathbf{a}_k] - \mathbf{a}_k^T\mathrm{E}[\Delta\mathbf{a}_k\Delta\mathbf{a}_k^T]\mathbf{a}_k\}\\ &= \frac{1}{x^2}\{x\mathrm{E}[\mathrm{tr}(\Delta\mathbf{a}_k\Delta\mathbf{a}_k^T)] - \mathbf{a}_k^T\mathrm{E}[\Delta\mathbf{a}_k\Delta\mathbf{a}_k^T]\mathbf{a}_k\}\,. \end{aligned} \tag{10}$$

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

$$\begin{aligned}\mathrm{CRIB}(\alpha_k^2) &= \frac{\mathrm{tr}[\mathrm{CRLB}(\mathbf{a}_k)]}{\|\mathbf{a}_k\|^2} - \frac{\mathbf{a}_k^T\mathrm{CRLB}(\mathbf{a}_k)\mathbf{a}_k}{\|\mathbf{a}_k\|^4}\,.\\ &= \frac{\mathrm{tr}[\Pi_{\mathbf{a}_k}^{\perp}\mathrm{CRLB}(\mathbf{a}_k)]}{\|\mathbf{a}_k\|^2}\end{aligned} \tag{11}$$

where

$$\Pi_{\mathbf{a}_k}^{\perp} = \mathbf{I} - \mathbf{a}_k\mathbf{a}_k^T/\|\mathbf{a}_k\|^2 \tag{12}$$

is the projection operator to the orthogonal complement of $\mathbf{a}_k$. It easily follows that the CRIB is always non-negative.

4. ANALYTICAL INVERSION OF HESSIAN MATRIX

The Jacobi 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 $\mathbf{H}(\boldsymbol{\theta})$ can be partitioned into $r\times r$ blocks of the size $I+J+K$,

$$\mathbf{H}(\boldsymbol{\theta}) = \begin{bmatrix}\mathbf{H}_{11} & \ldots & \mathbf{H}_{1r}\\ \vdots & & \vdots\\ \mathbf{H}_{r1} & \ldots & \mathbf{H}_{rr}\end{bmatrix} \tag{13}$$

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

$$\mathbf{H}_{ji} = \begin{bmatrix}\beta_{ij}\gamma_{ij}\mathbf{I}_I & \gamma_{ij}\mathbf{a}_i\mathbf{b}_j^T & \beta_{ij}\mathbf{a}_i\mathbf{c}_j^T\\ \gamma_{ij}\mathbf{b}_i\mathbf{a}_j^T & \alpha_{ij}\gamma_{ij}\mathbf{I}_J & \alpha_{ij}\mathbf{b}_i\mathbf{c}_j^T\\ \beta_{ij}\mathbf{c}_i\mathbf{a}_j^T & \alpha_{ij}\mathbf{c}_i\mathbf{b}_j^T & \alpha_{ij}\beta_{ij}\mathbf{I}_K\end{bmatrix} \tag{14}$$

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

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

$$\mathbf{H}_{ji} = \begin{bmatrix} x_{ij}\mathbf{I} + \mathbf{A}\mathbf{M}_{ij}^{AA}\mathbf{A}^T & \mathbf{A}\mathbf{M}_{ij}^{AB}\mathbf{B}^T & \mathbf{A}\mathbf{M}_{ij}^{AC}\mathbf{C}^T \\ \mathbf{B}\mathbf{M}_{ij}^{BA}\mathbf{A}^T & y_{ij}\mathbf{I} + \mathbf{B}\mathbf{M}_{ij}^{BB}\mathbf{B}^T & \mathbf{B}\mathbf{M}_{ij}^{BC}\mathbf{C}^T \\ \mathbf{C}\mathbf{M}_{ij}^{CA}\mathbf{A}^T & \mathbf{C}\mathbf{M}_{ij}^{CB}\mathbf{B}^T & z_{ij}\mathbf{I} + \mathbf{C}\mathbf{M}_{ij}^{CC}\mathbf{C}^T \end{bmatrix} \quad (15)$$

where, by comparing (14) with (15), we get $\mathbf{M}_{ij}^{AA} = \mathbf{M}_{ij}^{BB} = \mathbf{M}_{ij}^{CC} = \mathbf{0}$, $\mathbf{M}_{ij}^{AB} = (\mathbf{M}_{ji}^{BA})^T = \gamma_{ij}\mathbf{e}_i\mathbf{e}_j^T$, $\mathbf{M}_{ij}^{AC} = (\mathbf{M}_{ji}^{CA})^T = \beta_{ij}\mathbf{e}_i\mathbf{e}_j^T$, $\mathbf{M}_{ij}^{BC} = (\mathbf{M}_{ji}^{CB})^T = \alpha_{ij}\mathbf{e}_i\mathbf{e}_j^T$, $x_{ij} = \beta_{ij}\gamma_{ij} + \mu\delta_{ij}$, $y_{ij} = \alpha_{ij}\gamma_{ij} + \mu\delta_{ij}$, $z_{ij} = \alpha_{ij}\beta_{ij} + \mu\delta_{ij}$ for $i, j = 1, \ldots, r$, where $\mathbf{e}_k$ is the kth column of the $r \times r$ identity matrix, $k = 1, \ldots, r$, and δ_{ij} is the Kronecker's delta.

Now, under the assumption that $r \leq \min\{I, J, K\}$[1], the inverse of $\mathbf{H}(\boldsymbol{\theta}) + \mu\mathbf{I}$ can be sought in the same generic form. Let the inverse be partitioned into blocks $\overline{\mathbf{H}}_{ji}$ with the same structure as $\mathbf{H}_{ji}$, with constants $\overline{x}_{ij}, \overline{y}_{ij}$ and $\overline{z}_{ij}$ and matrices $\overline{\mathbf{M}}_{ij}^{PQ}$, $P, Q \in \{A, B, C\}$. The expressions for these constants and matrices are derived in Appendix.

5. CRIB IN CLOSED FORMS

The computation of the CRIB on $\mathbf{a}_1$ is now straightforward. It can be found as the limit

$$\text{CRIB}(\alpha_1^2) = \sigma^2 \lim_{\mu \to 0} \left[\frac{1}{\|\mathbf{a}_1\|^2} \text{tr}[\Pi_{\mathbf{a}_1}^{\perp} \overline{\mathbf{H}}_{\mu}] \right] \quad (16)$$

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

$$\begin{aligned} & \text{tr}[\Pi_{\mathbf{a}_1}^{\perp} \overline{\mathbf{H}}_{\mu}] = \text{tr}[\Pi_{\mathbf{a}_1}^{\perp} (\overline{x}_{11}\mathbf{I} + \mathbf{A}\overline{\mathbf{M}}_{11}^{AA}\mathbf{A}^T)] \\ = \; & \overline{x}_{11}\text{tr}[\Pi_{\mathbf{a}_1}^{\perp}] + \text{tr}\left[(\mathbf{A}^T \Pi_{\mathbf{a}_1}^{\perp} \mathbf{A})\overline{\mathbf{M}}_{11}^{AA}\right] \\ = \; & (I-1)\overline{x}_{11} + \text{tr}\left[(\mathbf{A}^T \Pi_{\mathbf{a}_1}^{\perp} \mathbf{A})\overline{\mathbf{M}}_{11}^{AA}\right] . \end{aligned} \quad (17)$$

Note that

$$\mathbf{A}\Pi_{\mathbf{a}_1}^{\perp}\mathbf{A}^T = \text{diag}\left(0, \alpha_{22} - \frac{\alpha_{12}^2}{\alpha_{11}}, \ldots, \alpha_{rr} - \frac{\alpha_{1r}^2}{\alpha_{11}}\right) . \quad (18)$$

5.1. Rank 1 tensors

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

$$\text{CRIB}(\alpha_1^2) = (I-1)\frac{\sigma^2}{\alpha_{11}\beta_{11}\gamma_{11}} . \quad (19)$$

[1]If $r > \min\{I, J, K\}$, the generic form for the inverse of $\mathbf{H}(\boldsymbol{\theta}) + \mu\mathbf{I}$ may be still valid for suitable $\overline{\mathbf{M}}_{ij}^{PQ}$, $P, Q \in \{A, B, C\}$, but it is overparameterized. The CRIB can be still computed via (16) and can be finite.

5.2. Rank 2 tensors

In this case,

$$\overline{x}_{11} = \left([(\mathbf{B}^T\mathbf{B}) \odot (\mathbf{C}^T\mathbf{C}) + \mu\mathbf{I}]^{-1}\right)_{11} = \frac{\beta_{22}\gamma_{22}}{d_{BC}} + O(\mu) \quad (20)$$

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

$$(\overline{\mathbf{M}}_{11}^{AA})_{22} = \frac{\alpha_{11}\beta_{22}\gamma_{22}[\beta_{12}^2 d_C + \gamma_{12}^2 d_B]}{d_A d_B d_C d_{BC}} + O(\mu)$$

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

$$\begin{aligned} \text{CRIB}(\alpha_1^2) &= \frac{\sigma^2}{\alpha_{11}}\left[(I-1)\overline{x}_{11} + \frac{d_A}{\alpha_{11}}(\overline{\mathbf{M}}_{11}^{AA})_{22}\right] \\ &= \frac{\sigma^2\beta_{22}\gamma_{22}}{\alpha_{11}d_{BC}}\left[(I-1) + \frac{\beta_{12}^2}{d_B} + \frac{\gamma_{12}^2}{d_C}\right] . \end{aligned} \quad (21)$$

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 $\mathbf{A}$, that means that the CRIB is infinite, if $d_B = 0$ or $d_C = 0$ (i.e., if any of the factor matrices $\mathbf{B}$ and $\mathbf{C}$ has colinear columns).

5.2.1. *Example*

We generated random orthogonal factor matrices $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ of dimensions 4×2, 5×2 and 6×2, respectively. The first column of $\mathbf{A}$ was modified as $\mathbf{a}_1 \leftarrow \lambda\mathbf{a}_1 + (1 - \lambda)\mathbf{a}_2$ 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.

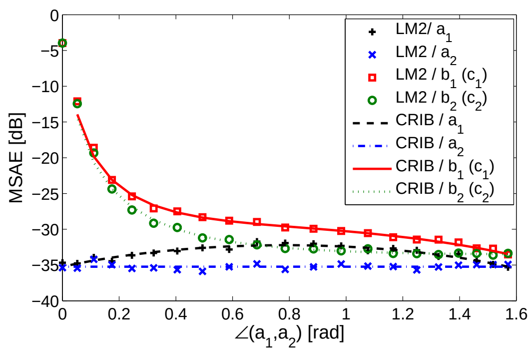

Fig. 1. Mean square angular deviation (MSAE) and the corresponding CRIB of estimated columns of factor matrices averaged over 100 independent trials.

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 $\mathbf{A}$ 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 $\mathbf{A}^T\mathbf{A}$, $\mathbf{B}^T\mathbf{B}$ and $\mathbf{C}^T\mathbf{C}$ 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.

7. REFERENCES

[1] J. B. Kruskal, "Three-way arrays: Rank and uniqueness of trilinear decompositions", *Linear Algebra and Applications*, vol. 18, pp. 95-138, 1977.

[2] 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.

[3] 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.

[4] 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.

[5] P. Tichavský, Z. Koldovský, "Simultaneous search for all modes in multilinear models", *Proc. ICASSP 2010*, Dallas, TX, March 14-19, 2010, pp. 4114-4117.

[6] A.H. Phan, P. Tichavský and A. Cichocki, "Low Complexity Damped Gauss-Newton Algorithms for Parallel Factor Analysis", submitted for publication.

[7] B. Porat, *Digital Processing of Random Signals, Theory and Methods*, Prentice Hall, Dec. 1993.

[8] G.P.H. Styan, "Hadamard Products and Multivariate Statistical Analysis", *Linear Algebra and its Applications*, vol. 6, pp. 217-240, 1973.

Appendix

Inverse of the Hessian matrix can be found via the equation $\sum_{k=1}^r \mathbf{H}_{jk}\overline{\mathbf{H}}_{ki} = \delta_{ij}\mathbf{I}$ for $\quad i,j = 1,\ldots,r$,

$$\sum_{k=1}^{r}\begin{bmatrix} x_{kj}\mathbf{I}+\mathbf{A}\mathbf{M}_{kj}^{AA}\mathbf{A}^T & \mathbf{A}\mathbf{M}_{kj}^{AB}\mathbf{B}^T & \mathbf{A}\mathbf{M}_{kj}^{AC}\mathbf{C}^T \\ \mathbf{B}\mathbf{M}_{kj}^{BA}\mathbf{A}^T & y_{kj}\mathbf{I}+\mathbf{B}\mathbf{M}_{kj}^{BB}\mathbf{B}^T & \mathbf{B}\mathbf{M}_{kj}^{BC}\mathbf{C}^T \\ \mathbf{C}\mathbf{M}_{kj}^{CA}\mathbf{A}^T & \mathbf{C}\mathbf{M}_{kj}^{CB}\mathbf{B}^T & z_{kj}\mathbf{I}+\mathbf{C}\mathbf{M}_{kj}^{CC}\mathbf{C}^T \end{bmatrix}\begin{bmatrix} \overline{x}_{ik}\mathbf{I}+\mathbf{A}\overline{\mathbf{M}}_{ik}^{AA}\mathbf{A}^T & \mathbf{A}\overline{\mathbf{M}}_{ik}^{AB}\mathbf{B}^T & \mathbf{A}\overline{\mathbf{M}}_{ik}^{AC}\mathbf{C}^T \\ \mathbf{B}\overline{\mathbf{M}}_{ik}^{BA}\mathbf{A}^T & \overline{y}_{ik}\mathbf{I}+\mathbf{B}\overline{\mathbf{M}}_{ik}^{BB}\mathbf{B}^T & \mathbf{B}\overline{\mathbf{M}}_{ik}^{BC}\mathbf{C}^T \\ \mathbf{C}\overline{\mathbf{M}}_{ik}^{CA}\mathbf{A}^T & \mathbf{C}\overline{\mathbf{M}}_{ik}^{CB}\mathbf{B}^T & \overline{z}_{ik}\mathbf{I}+\mathbf{C}\overline{\mathbf{M}}_{ik}^{CC}\mathbf{C}^T \end{bmatrix} = \delta_{ij}\mathbf{I}$$

$$\begin{aligned}
\sum_{k=1}^{r} x_{kj}\overline{x}_{ik}\mathbf{I} + \mathbf{A}\left[\sum_{k=1}^{r}\left(\overline{x}_{ik}\mathbf{M}_{kj}^{AA} + x_{kj}\overline{\mathbf{M}}_{ik}^{AA} + \mathbf{M}_{kj}^{AA}\mathbf{A}^T\mathbf{A}\overline{\mathbf{M}}_{ik}^{AA} + \mathbf{M}_{kj}^{AB}\mathbf{B}^T\mathbf{B}\overline{\mathbf{M}}_{ik}^{BA} + \mathbf{M}_{kj}^{AC}\mathbf{C}^T\mathbf{C}\overline{\mathbf{M}}_{ik}^{CA}\right)\right]\mathbf{A}^T &= \delta_{ij}\mathbf{I} \\
\mathbf{B}\left[\sum_{k=1}^{r}\left(\overline{x}_{ik}\mathbf{M}_{kj}^{BA} + y_{kj}\overline{\mathbf{M}}_{ik}^{BA} + \mathbf{M}_{kj}^{BA}\mathbf{A}^T\mathbf{A}\overline{\mathbf{M}}_{ik}^{AA} + \mathbf{M}_{kj}^{BB}\mathbf{B}^T\mathbf{B}\overline{\mathbf{M}}_{ik}^{BA} + \mathbf{M}_{kj}^{BC}\mathbf{C}^T\mathbf{C}\overline{\mathbf{M}}_{ik}^{CA}\right)\right]\mathbf{A}^T &= \mathbf{0} \\
\mathbf{C}\left[\sum_{k=1}^{r}\left(\overline{x}_{ik}\mathbf{M}_{kj}^{CA} + z_{kj}\overline{\mathbf{M}}_{ik}^{CA} + \mathbf{M}_{kj}^{CA}\mathbf{A}^T\mathbf{A}\overline{\mathbf{M}}_{ik}^{AA} + \mathbf{M}_{kj}^{CB}\mathbf{B}^T\mathbf{B}\overline{\mathbf{M}}_{ik}^{BA} + \mathbf{M}_{kj}^{CC}\mathbf{C}^T\mathbf{C}\overline{\mathbf{M}}_{ik}^{CA}\right)\right]\mathbf{A}^T &= \mathbf{0}
\end{aligned}$$

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

$$\begin{aligned}
\sum_{k=1}^{r}(x_{kj}\mathbf{I} + \mathbf{M}_{kj}^{AA}\mathbf{A}^T\mathbf{A})\,\overline{\mathbf{M}}_{ik}^{AA} + \sum_{k=1}^{r}(\mathbf{M}_{kj}^{AB}\mathbf{B}^T\mathbf{B})\,\overline{\mathbf{M}}_{ik}^{BA} + \sum_{k=1}^{r}(\mathbf{M}_{kj}^{AC}\mathbf{C}^T\mathbf{C})\,\overline{\mathbf{M}}_{ik}^{CA} &= -\sum_{k=1}^{r}\overline{x}_{ik}\mathbf{M}_{kj}^{AA} \\
\sum_{k=1}^{r}(\mathbf{M}_{kj}^{BA}\mathbf{A}^T\mathbf{A})\,\overline{\mathbf{M}}_{ik}^{AA} + \sum_{k=1}^{r}(y_{kj}\mathbf{I} + \mathbf{M}_{kj}^{BB}\mathbf{B}^T\mathbf{B})\,\overline{\mathbf{M}}_{ik}^{BA} + \sum_{k=1}^{r}(\mathbf{M}_{kj}^{BC}\mathbf{C}^T\mathbf{C})\,\overline{\mathbf{M}}_{ik}^{CA} &= -\sum_{k=1}^{r}\overline{x}_{ik}\mathbf{M}_{kj}^{BA} \\
\sum_{k=1}^{r}(\mathbf{M}_{kj}^{CA}\mathbf{A}^T\mathbf{A})\,\overline{\mathbf{M}}_{ik}^{AA} + \sum_{k=1}^{r}(\mathbf{M}_{kj}^{CB}\mathbf{B}^T\mathbf{B})\,\overline{\mathbf{M}}_{ik}^{BA} + \sum_{k=1}^{r}(z_{kj}\mathbf{I} + \mathbf{M}_{kj}^{CC}\mathbf{C}^T\mathbf{C})\,\overline{\mathbf{M}}_{ik}^{CA} &= -\sum_{k=1}^{r}\overline{x}_{ik}\mathbf{M}_{kj}^{CA}
\end{aligned} \tag{22}$$

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 $\mathbf{A}^T\mathbf{A}$, $\mathbf{B}^T\mathbf{B}$ and $\mathbf{C}^T\mathbf{C}$.