STABILITY ANALYSIS AND FAST DAMPED-GAUSS-NEWTON ALGORITHM FOR INDSCAL TENSOR DECOMPOSITION

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

[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
[3]Brain Science Institute, RIKEN, 2-1 Hirosawa, Wakoshi, Saitama 351-0198, Japan.

ABSTRACT

INDSCAL is a special case of the CANDECOMP-PARAFAC (CP) decomposition of three or more-way tensors, where two factor matrices are equal. This paper provides a stability analysis of INDSCAL that is done by deriving the Cramér-Rao lower bound (CRLB) on variance of an unbiased estimate of the tensor parameters from its noisy observation (the tensor plus an i.i.d. Gaussian random tensor). The existence of the bound reveals necessary conditions for the essential uniqueness of the INDSCAL decomposition. This is compared with previous results on CP. Next, analytical expressions for the inverse of the Hessian matrix, which is needed to compute the CRLB, are used in a damped Gaussian (Levenberg-Marquardt) algorithm, which gives a novel method for INDSCAL having a lower computational complexity.

Index Terms— INDSCAL; PARAFAC; CANDECOMP; tensor decomposition; Cramér-Rao lower bound; Levenberg-Marquardt algorithm

1. INTRODUCTION

Multilinear models of three-way and higher-way data arrays were applied in many research areas such as chemistry, astronomy, or even psychology. Recently, tensor decomposition techniques have become popular in signal processing for its usefulness, e.g., in blind source separation or feature extraction. A special attention is paid to the CP decomposition (also known as CANDECOMP or PARAFAC) that decomposes a given tensor into a sum of d rank-one tensors. For example, the decomposition of a three-way tensor $\underline{\mathbf{X}}$ is

$$\underline{\mathbf{X}} = \sum_{f=1}^{d} \mathbf{a}_f \circ \mathbf{b}_f \circ \mathbf{d}_f, \tag{1}$$

where $\mathbf{a}_f$, $\mathbf{b}_f$, $\mathbf{d}_f$, are columns of factor matrices $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{D}$, respectively, and $\circ$ denotes the outer vector product.

This paper addresses INDSCAL (INdividual Differences in multidimensional SCALing; see [1]) that is a special symmetric decomposition of three-way tensors where two of three factor matrices, say $\mathbf{A}$ and $\mathbf{B}$, are assumed to be the same. This model is useful, for example, in blind identification of underdetermined mixtures [2], where the factor matrices represent parameters of the mixing model

$$\mathbf{U} = \mathbf{AS} \tag{2}$$

where rows of $\mathbf{U}$ and $\mathbf{S}$ contain, respectively, samples of the observed signals and of unknown independent signals; $\mathbf{D}$ usually contains statistics of the original sources. An important issue here is the essential uniqueness[1] of the decomposition as it entails the estimableness of $\mathbf{A}$ and $\mathbf{D}$.

Necessary conditions for the uniqueness of CP were derived in [3], where it was shown that $\mathbf{A} * \mathbf{B}$, $\mathbf{B} * \mathbf{D}$ and $\mathbf{A} * \mathbf{D}$, where $*$ denotes the Khatri-Rao product, must have full column rank. The achievable accuracy of CP was studied in [3] through the Cramér-Rao lower bound (CRLB) on an unbiased estimation of the factor matrices, given a noisy observation of a tensor. More analytical results of the CRLB which have been recently derived in [5] provide a deeper insight into the stability of CP and into the identifiability of individual columns of factor matrices.

In this paper, we do an analogous stability analysis for INDSCAL, i.e., when $\mathbf{B} = \mathbf{A}$. As the number of parameters (factor matrices) is lower than in CP, the resulting CRLB is *different* and is shown to be *lower* than that of CP with $\mathbf{B} = \mathbf{A}$. The computation of CRLB requires inversion of a Hessian matrix also needed in Levenberg-Marquardt (LM) optimization algorithms [4, 6]. Using novel analytical simplifications, a faster algorithm for INDSCAL based on the LM optimization procedure is derived.

[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 projects 102/09/1278 and P103/11/1947.

[1]It means the uniqueness up to scaling and permutation of columns of factor matrices.

978-1-4577-0568-7/11/$26.00 ©2011 IEEE

2. PROBLEM FORMULATION

Let a three way tensor $\underline{\mathbf{X}}$ of the dimension $m \times m \times M$ has elements

$$X_{ijk} = \sum_{f=1}^{d} A_{if} A_{jf} D_{kf} \tag{3}$$

where A_{if} and D_{kf} are elements of factor matrices $\mathbf{A}$ and $\mathbf{D}$, respectively, whose dimensions are $m \times d$ and $M \times d$. Their kth columns will be denoted by $\mathbf{a}_k$ and $\mathbf{d}_k$, respectively; d is the rank of $\underline{\mathbf{X}}$, which in (2) corresponds to the number of original sources.

For the analysis, we consider a noisy observation of the tensor

$$\underline{\mathbf{Y}} = \underline{\mathbf{X}} + \underline{\mathbf{E}} \tag{4}$$

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

Let $\boldsymbol{\theta}$ be a $(m+M)d \times 1$ parameter vector arranged as

$$\boldsymbol{\theta} = [\boldsymbol{\theta}_1^T, \ldots, \boldsymbol{\theta}_d^T]^T \tag{5}$$

where $\boldsymbol{\theta}_k = [\mathbf{a}_k^T, \mathbf{d}_k^T]^T$. Since the distribution of elements of $\underline{\mathbf{E}}$ is Gaussian, the Fisher information matrix (FIM) of the vectorized version of the tensor $\underline{\mathbf{Y}}$, i.e. of vec$[\underline{\mathbf{Y}}]$, is

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

where $\mathbf{J}(\boldsymbol{\theta})$ is the Jacobian matrix (matrix of the first-order derivatives) of vec$[\underline{\mathbf{X}}]$ with respect to $\boldsymbol{\theta}$ [8]. The CRLB for unbiased estimation of $\boldsymbol{\theta}$ is given by the inverse of $\mathbf{F}(\boldsymbol{\theta})$ [7].

Thanks to the problem symmetry, we can derive the bound for $\mathbf{a}_1$ and $\mathbf{d}_1$ only. The bounds for the other columns of $\mathbf{A}$ and $\mathbf{D}$ follow.

2.1. Mean Square Angular Error and Cramér-Rao Induced Bound

To avoid the inherent permutation and scale uncertainty of the model, we assume that the order of columns of the estimated $\widehat{\mathbf{A}}$ and $\widehat{\mathbf{D}}$ matches the original order. As an evaluation criterion, we will consider the mean square angular error (MSAE) between the columns of the estimated and original matrices. The advantage of MSAE is that it is scale-invariant.

For example, the criterion between the kth column of $\widehat{\mathbf{A}}$ and $\mathbf{A}$ is defined as

$$\text{MSAE}(\mathbf{a}_k, \widehat{\mathbf{a}}_k) = \text{E}\left[\arccos^2 \alpha_k\right] \tag{7}$$

where $\alpha_k = |\widehat{\mathbf{a}}_k^T \mathbf{a}_k| / (\|\widehat{\mathbf{a}}_k\| \, \|\mathbf{a}_k\|)$, and E$[\cdot]$ stands for the expectation operator.

In [5], the leading term of an asymptotic approximation of the argument in (7) was shown to be

$$\arccos^2 \alpha_k \approx \frac{1}{x^2} [x \operatorname{tr}(\Delta\mathbf{a}_k \Delta\mathbf{a}_k^T) - \mathbf{a}_k^T \Delta\mathbf{a}_k \Delta\mathbf{a}_k^T \mathbf{a}_k] \tag{8}$$

where $x = \mathbf{a}_k^T \mathbf{a}_k$ and $\Delta\mathbf{a}_k = \widehat{\mathbf{a}}_k - \mathbf{a}_k$. Now, when $\widehat{\mathbf{a}}_k$ is an unbiased estimator of $\mathbf{a}_k$ achieving the CRLB, then $\text{E}[\Delta\mathbf{a}_k \Delta\mathbf{a}_k^T] = \text{CRLB}(\mathbf{a}_k)$ where CRLB$(\mathbf{a}_k)$ is the submatrix of $\mathbf{F}^{-1}$ which bounds the mean square error in estimating $\mathbf{a}_k$. Using this fact and the approximation (8) in (7), the Cramér-Rao induced lower bound (CRIB) on the MSAE of $\widehat{\mathbf{a}_k}$ can be defined as

$$\text{CRIB}(\mathbf{a}_k) = \frac{\text{tr}[\Pi_{\mathbf{a}_k}^{\perp} \text{CRLB}(\mathbf{a}_k)]}{\|\mathbf{a}_k\|^2} \tag{9}$$

where $\Pi_{\mathbf{a}_k}^{\perp} = \mathbf{I} - \mathbf{a}_k \mathbf{a}_k^T / \|\mathbf{a}_k\|^2$. See [5] for a more detailed derivation of (9).

Finally, we should note that $\mathbf{F}(\boldsymbol{\theta})$ is singular, because d parameters in $\boldsymbol{\theta}$ are free due to the scale ambiguity problem. We therefore use the same approach as in [5] and regularize $\mathbf{F}(\boldsymbol{\theta})$ by adding $\mu\mathbf{I}$ to it. The CRIB is then derived as a limit when $\mu \to 0$.

3. FAST INVERSION OF $\mathbf{F}(\boldsymbol{\theta})$

The partial derivatives of (3) by the elements of $\boldsymbol{\theta}$ are

$$\frac{\partial X_{ijk}}{A_{uv}} = D_{kv}(\delta_{iu} A_{jv} + \delta_{ju} A_{iv}) \qquad \frac{\partial X_{ijk}}{D_{uv}} = \delta_{ku} A_{iv} A_{jv}.$$

It can be shown that the structure of the Hessian matrix $\mathbf{H}(\boldsymbol{\theta}) \stackrel{\text{def.}}{=} \sigma^2 \mathbf{F}(\boldsymbol{\theta}) = \mathbf{J}^T(\boldsymbol{\theta}) \mathbf{J}(\boldsymbol{\theta})$ is such that it can be partitioned into $d \times d$ blocks of the size $(m+M) \times (m+M)$,

$$\mathbf{H}(\boldsymbol{\theta}) = \begin{bmatrix} \mathbf{H}_{11} & \ldots & \mathbf{H}_{1d} \\ \vdots & & \vdots \\ \mathbf{H}_{d1} & \ldots & \mathbf{H}_{dd} \end{bmatrix} \tag{10}$$

where the ijth block can be written as

$$\mathbf{H}_{ij} = \begin{bmatrix} 2\beta_{ij}(\alpha_{ij}\mathbf{I}_m + \mathbf{a}_j \mathbf{a}_i^T) & 2\alpha_{ij} \mathbf{a}_j \mathbf{d}_i^T \\ 2\alpha_{ij} \mathbf{d}_j \mathbf{a}_i^T & \alpha_{ij}^2 \mathbf{I}_M \end{bmatrix}, \tag{11}$$

where α_{ij} and β_{ij} are the elements of $\mathbf{A}^T\mathbf{A}$ and $\mathbf{D}^T\mathbf{D}$, respectively, and $\mathbf{I}_m$ denotes the identity matrix of the size $m \times m$. The blocks $\mathbf{H}_{ij}$ can be also written in a generic form

$$\mathbf{H}_{ij} = \begin{bmatrix} x_{ij}\mathbf{I}_m + \mathbf{A}\mathbf{M}_{ij}^{AA}\mathbf{A}^T & \mathbf{A}\mathbf{M}_{ij}^{AD}\mathbf{D}^T \\ \mathbf{D}\mathbf{M}_{ij}^{DA}\mathbf{A}^T & y_{ij}\mathbf{I}_M \end{bmatrix} \tag{12}$$

where, $\mathbf{M}_{ij}^{AA} = 2\beta_{ij}\mathbf{e}_j\mathbf{e}_i^T$, $\mathbf{M}_{ij}^{AD} = \mathbf{M}_{ij}^{DA} = 2\alpha_{ij}\mathbf{e}_j\mathbf{e}_i^T$, $x_{ij} = 2\alpha_{ij}\beta_{ij} + \mu$, and $y_{ij} = \alpha_{ij}^2 + \mu$, $i, j = 1, \ldots, d$, where $\mathbf{e}_k$ stands for the kth column of $\mathbf{I}_d$,

To find an analytical inverse of the matrix $\mathbf{H}(\boldsymbol{\theta}) + \mu\mathbf{I}$, we seek it in the same block-structure form as (10) with blocks whose structure is

$$\overline{\mathbf{H}}_{ij} = \begin{bmatrix} \overline{x}_{ij}\mathbf{I}_m + \mathbf{A}\overline{\mathbf{M}}_{ij}^{AA}\mathbf{A}^T & \mathbf{A}\overline{\mathbf{M}}_{ij}^{AD}\mathbf{D}^T \\ \mathbf{D}\overline{\mathbf{M}}_{ij}^{DA}\mathbf{A}^T & \overline{y}_{ij}\mathbf{I}_M + \mathbf{D}\overline{\mathbf{M}}_{ij}^{DD}\mathbf{D}^T \end{bmatrix} \tag{13}$$

where $\overline{x}_{ij}$ and $\overline{y}_{ij}$ are scalars and $\overline{\mathbf{M}}_{ij}^{AA}$, $\overline{\mathbf{M}}_{ij}^{DD}$ are matrices of the size, $d \times d$.

Based on this, we describe in Appendix how $\mathbf{H}^{-1}$ can be sought through the fact that $\sum_{k=1}^{d} \mathbf{H}_{ik}\overline{\mathbf{H}}_{kj} = \delta_{ij}\mathbf{I}$.

4. TOWARDS CRIB IN CLOSED FORMS

By (9), the CRIB on $\mathbf{a}_1$ can be found as the limit

$$\text{CRIB}(\alpha_1^2) = \sigma^2 \lim_{\mu \to 0} \left[\frac{1}{\|\mathbf{a}_1\|^2} \text{tr}[\Pi_{\mathbf{a}_1}^{\perp} \overline{\mathbf{H}}_{\mu}] \right] \tag{14}$$

where $\overline{\mathbf{H}}_{\mu}$ is the left-upper $m \times m$ diagonal block of $(\mathbf{H}(\boldsymbol{\theta}) + \mu\mathbf{I})^{-1}$, which is by (12) equal to $\overline{\mathbf{H}}_{\mu} = \overline{x}_{11}\mathbf{I} + \mathbf{A}\overline{\mathbf{M}}_{11}^{AA}\mathbf{A}^T$. Now, (14) can be simplified using the fact that

$$\text{tr}[\Pi_{\mathbf{a}_1}^{\perp} \overline{\mathbf{H}}_{\mu}] = (m-1)\overline{x}_{11} + \text{tr}\left[(\mathbf{A}^T \Pi_{\mathbf{a}_1}^{\perp} \mathbf{A}) \overline{\mathbf{M}}_{11}^{AA} \right],$$

and

$$\mathbf{A}\Pi_{\mathbf{a}_1}^{\perp}\mathbf{A}^T = \text{diag}\left(0, \alpha_{22} - \alpha_{12}^2/\alpha_{11}, \ldots, \alpha_{dd} - \alpha_{1d}^2/\alpha_{11}\right).$$

The analytical derivation of $\overline{\mathbf{M}}_{11}^{AA}$ poses a complex algebraical problem. We therefore resort our analytical computations to rank 1 and rank 2 tensors.

Rank-1 tensors

The simplest case when $d = 1$ gives $\overline{x}_{11} = (2\alpha_{11}\beta_{11} + \mu)^{-1}$ and therefore

$$\text{CRIB}(\mathbf{a}_1) = (m-1)\frac{\sigma^2}{2\alpha_{11}\beta_{11}}. \tag{15}$$

The bound is equal to one half of the corresponding bound for the CP decomposition [5].

Rank-2 tensors

For $d = 2$, we obtain

$$\overline{x}_{11} = \left([2(\mathbf{A}^T\mathbf{A}) \odot (\mathbf{D}^T\mathbf{D}) + \mu\mathbf{I}]^{-1}\right)_{11} = \frac{\alpha_{22}\beta_{22}}{2d_{AD}} + O(\mu)$$

where $d_{AD} = \det[(\mathbf{A}^T\mathbf{A}) \odot (\mathbf{D}^T\mathbf{D})]$ and $\odot$ denotes the Hadamard (element-wise) product. The matrix $\overline{\mathbf{M}}_{11}^{AA}$ can be obtained by solving the 8×8 linear system (18). The $(2,2)$th element of this matrix reads

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

where $d_A = \det[\mathbf{A}^T\mathbf{A}]$, $d_{AA} = \det[(\mathbf{A}^T\mathbf{A}) \odot (\mathbf{A}^T\mathbf{A})]$, $d_D = \det[\mathbf{D}^T\mathbf{D}]$, and $d_{AD} = \det[(\mathbf{A}^T\mathbf{A}) \odot (\mathbf{D}^T\mathbf{D})]$. Therefore

$$\begin{aligned} \text{CRIB}(\mathbf{a}_1) &= \frac{\sigma^2}{\alpha_{11}} \left[(m-1)\overline{x}_{11} + \frac{d_A}{\alpha_{11}} (\overline{\mathbf{M}}_{11}^{AA})_{22} \right] \\ &= \frac{\sigma^2 \alpha_{22}\beta_{22}}{\alpha_{11} d_{AD}} \left[\frac{m-1}{2} + \frac{\alpha_{12}^2}{d_A} + \frac{\beta_{12}^2}{d_D} \frac{d_{AA}}{2\alpha_{11} d_A} \right]. \end{aligned} \tag{16}$$

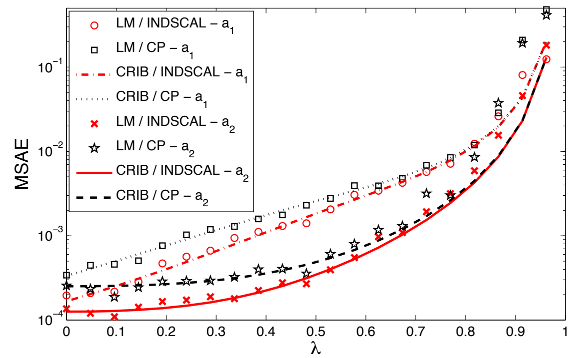

Fig. 1. Mean square angular error (MSAE) and the corresponding CRIB of estimated columns of $\mathbf{A}$ averaged over 100 independent trials.

This bound can be compared to that derived in [5] for CP. By taking $\mathbf{B} = \mathbf{A}$ it reads

$$\text{CRIB}^{(CP)}(\mathbf{a}_1) = \frac{\sigma^2\alpha_{22}\beta_{22}}{\alpha_{11}d_{AD}} \left[m - 1 + \frac{\alpha_{12}^2}{d_A} + \frac{\beta_{12}^2}{d_D} \right]. \tag{17}$$

It can be easily shown that $\text{CRIB}(\mathbf{a}_1) \le \text{CRIB}^{(CP)}(\mathbf{a}_1)$.

The CRIB for angular errors of columns of the matrix $\mathbf{D}$ can be derived in an analogous way. The resultant bound can be shown to be identical to that derived for the CP decomposition in [5], but we do not provide the derivations here due to lack of space.

5. EXPERIMENTS

5.1. Example 1

We generated a random rank-2 tensor of dimensions $3 \times 3 \times 4$ with $\mathbf{A}$ and $\mathbf{D}$ having orthogonal columns. Hundred noisy observations of the tensor were generated with $\sigma = 0.01$, and their CP and INDSCAL decompositions were computed using, respectively, the Levenberg-Marquardt (LM) method for CP from [4] and the proposed LM method for INDSCAL using the fast Hessian inversion through (18). Figure 1 compares the average performance of the algorithms with the theoretical value of CRIB when the first column of $\mathbf{A}$ was varied according to $\mathbf{a}_1 \leftarrow \lambda\mathbf{a}_1 + (1-\lambda)\mathbf{a}_2$ with $\lambda \in [0, 1]$.

The results show that the CRIB is in a good agreement with the performance of the algorithms[2] and confirm the better performance of INDSCAL compared to CP. For $\lambda = 1$, $\mathbf{A}$ is singular and conditions for essential uniqueness of INDSCAL (and CP) are not satisfied. This is revealed by the CRIB, which is going to infinity as $\lambda \to 1$.

[2]The algorithms achieve the CRIB because they, in fact, do the maximal likelihood estimation through minimizing the quadratic fit between the observed tensor and its model. In addition, the elements of $\underline{\mathbf{E}}$ in (4) are not correlated.

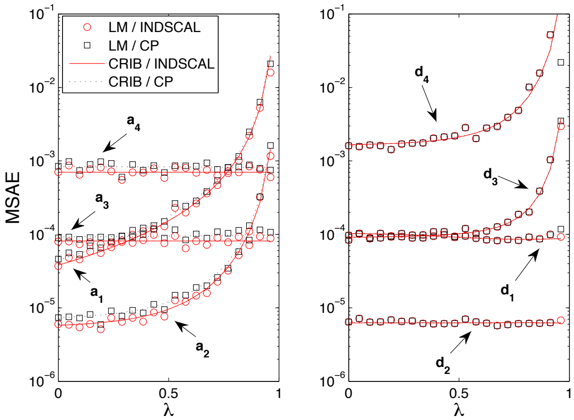

Fig. 2. Results of Example 2 in terms of MSAE for columns of factor matrices $\mathbf{A}$ and $\mathbf{D}$.

5.2. Example 2

An analogous experiment was done with a random tensor of rank 4 having dimensions $3 \times 3 \times 10$, but the first column of $\mathbf{D}$ was modified as $\mathbf{d}_1 \leftarrow \lambda \mathbf{d}_1 + (1-\lambda)\mathbf{d}_2$ while $\mathbf{A}$ was fixed. The results are shown in Figure 2.

While $\mathbf{d}_1$, $\mathbf{d}_2$, $\mathbf{a}_3$ and $\mathbf{a}_4$ are correctly estimated for all λ, the identification of $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{d}_3$ and $\mathbf{d}_4$ is not possible for $\lambda = 1$. For example, in the (underdetermined) blind separation problem (2) when there are four unknown sources mixed via $\mathbf{A}$ into three observed signals, it means that the directions of the first two sources and the characteristics of the other sources cannot be consistently retrieved, and vice versa.

6. CONCLUSIONS

Cramér-Rao-induced bounds for individual columns of factor matrices in the INDSCAL decomposition were derived. The bound can be used for determining whether the INDSCAL decomposition of a tensor is stable or not. Novel analytical expressions for the inversion of the Hessian matrix allow to compute it in $O(d^6)$ operations, which improves the computational burden of Levenberg-Marquardt-based optimization algorithms for INDSCAL.

7. REFERENCES

[1] J. D. Carroll and J. Chang, "Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition," *Psychometrika*, vol. 35, pp. 283-319, 1970.

[2] P. Comon and M. Rajih, "Blind identification of under-determined mixtures based on the characteristic function, " *Signal Processing*, vol. 86, no. 9, pp. 2271-2281, Sept. 2006.

[3] X. Liu and N. D. Sidiropoulos, "Cramér-Rao Lower Bounds for Low-Rank Decomposition of Multilinear Arrays," *IEEE Trans. on Sig. Proc.*, vol. 49, no. 9, pp. 2074-2086, Sept. 2001.

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

[5] P. Tichavský and Z. Koldovský, "Stability of CANDECOMP-PARAFAC tensor decomposition", *Proc. of ICASSP 2011*, pp. 4168-4171, Prague, Czech Republic, May 22-27, 2011.

[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] P. Comon, J. M. F. Ten Berge, L. De Lathauwer and J. Castaing, "Generic and Typical Ranks of Multi-Way Arrays", *Linear Algebra Appl.*, vol. 430, no. 11-12, pp. 2997-3007, Jun. 2009.

Appendix

From $\sum_{k=1}^d \mathbf{H}_{ik}\overline{\mathbf{H}}_{kj} = \delta_{ij}\mathbf{I}$ it follows that $\overline{\mathbf{M}}_{ij}^{AA}$, $\overline{\mathbf{M}}_{ij}^{DA}$ and the scalars $\overline{x}_{ij}$, $i = 1,\ldots,d$, $j = 1,\ldots,d$, are the solutions of

$$\sum_{k=1}^d x_{ik}\overline{x}_{kj}\mathbf{I} + \mathbf{A}\left[\sum_{k=1}^d \left(\overline{x}_{kj}\mathbf{M}_{ik}^{AA} + x_{ik}\overline{\mathbf{M}}_{kj}^{AA} + \mathbf{M}_{ik}^{AA}\mathbf{A}^T\mathbf{A}\overline{\mathbf{M}}_{kj}^{AA} + \mathbf{M}_{ik}^{AD}\mathbf{D}^T\mathbf{D}\overline{\mathbf{M}}_{kj}^{DA}\right)\right]\mathbf{A}^T = \delta_{ij}\mathbf{I},$$

$$\mathbf{D}\left[\sum_{k=1}^d \left(\overline{x}_{kj}\mathbf{M}_{ik}^{DA} + y_{ik}\overline{\mathbf{M}}_{kj}^{DA} + \mathbf{M}_{ik}^{DA}\mathbf{A}^T\mathbf{A}\overline{\mathbf{M}}_{kj}^{AA}\right)\right]\mathbf{A}^T = \mathbf{0}.$$

Hence $\sum_{k=1}^d x_{ik}\overline{x}_{kj} = \delta_{ij}$ for $i,j = 1,\ldots,d$, and therefore the $d \times d$ matrix $\overline{\mathbf{X}} = (\overline{x}_{ij})_{i,j=1}^d$ equals to the inverse of $\mathbf{X} = (x_{ij})_{i,j=1}^d$. Next, the columns of matrices $\overline{\mathbf{M}}_{ij}^{AA}$ and $\overline{\mathbf{M}}_{ij}^{DA}$ for $i,j = 1,\ldots,d$ can be sought as the solutions of $2d^2 \times 2d^2$ linear systems following from

$$\sum_{k=1}^d (x_{ik}\mathbf{I} + \mathbf{M}_{ik}^{AA}\mathbf{A}^T\mathbf{A})\,\overline{\mathbf{M}}_{kj}^{AA} + \sum_{k=1}^d (\mathbf{M}_{ik}^{AD}\mathbf{D}^T\mathbf{D})\,\overline{\mathbf{M}}_{kj}^{DA} = -\sum_{k=1}^d \overline{x}_{kj}\mathbf{M}_{ij}^{AA},$$

$$\sum_{k=1}^d (\mathbf{M}_{ik}^{DA}\mathbf{A}^T\mathbf{A})\,\overline{\mathbf{M}}_{kj}^{AA} + \sum_{k=1}^d y_{ik}\overline{\mathbf{M}}_{kj}^{DA} = -\sum_{k=1}^d \overline{x}_{kj}\mathbf{M}_{ik}^{DA}. \tag{18}$$

The solution of each of such linear systems requires $O(d^6)$ operations, so that the whole inverse of the Hessian matrix can be done in the same number of operations plus $O(d^2(m+M))$ operations needed to compute the products $\mathbf{A}^T\mathbf{A}$ and $\mathbf{D}^T\mathbf{D}$.