"ERROR ANALYSIS OF LOW-RANK THREE-WAY TENSOR FACTORIZATION APPROACH TO BLIND SOURCE SEPARATION,"
2014 IEEE International Conference on Acoustics, Speech and Signal Processing, Florence, Italy, May 7, 2014.

ERROR ANALYSIS OF LOW-RANK THREE-WAY TENSOR FACTORIZATION APPROACH TO BLIND SOURCE SEPARATION

Ivica Kopriva[1*], Jean-Philipe Royer[2], Nadege Thirion-Moreau[2], Pierre Comon[4]

[1]Ruđer Bošković Institute, Zagreb, Croatia

[2] Aix-Marseille Université, CNR, ENSAM, LSIS, UMR 7296, 13397 Marseille and Université de Toulon, CNRS, LSIS, UMR 7296, 83957, La Garde, France

[3] GIPSA-Lab, CNRS UMR5216, Grenoble Campus, BP.46 F-38402 St Martin d'Heres Cedex, France

e-mail: ikopriva@irb.hr, jph.royer@gmail.com ,thirion@univ-tln.fr, pierre.comon@gipsa-lab.grenoble-inp.fr

[*] Work of I. Kopriva has been partially supported through the FP7-REGPOT-2012-2013-1,Grant Agreement Number 316289 – *InnoMol* and partially through the Grant 9.01/232 funded by Croatian Science Foundation.

Talk outline

- Blind separation of multidimensional sources
- Tensor notation, unfolding, norm, products, rank,…
- CPD and TuckerN tensor models:
  - uniqueness, properties, …
- Error analysis with demonstration:
  - multispectral image decomposition

Blind source separation

Recovery of signals from their multichannel linear superposition using minimum of *a priori* information i.e. multichannel measurements only [1-3].

Problem:

$\mathbf{X}=\mathbf{AS}$ $\quad \mathbf{X}\in \mathrm{R}^{N\mathrm{x}T}$, $\mathbf{A}\in \mathrm{R}^{N\mathrm{x}M}$, $\mathbf{S}\in \mathrm{R}^{M\mathrm{x}T}$

N - number of sensors/mixtures;
M - unknown number of sources
T - number of samples/observations

Goal: find **A** and **S** based on **X** only.

1. A. Hyvarinen, J. Karhunen, E. Oja, "Independent Component Analysis," John Wiley, 2001.
2. A. Cichocki, S. Amari, "Adaptive Blind Signal and Image Processing," John Wiley, 2002.
3. P. Comon, C. Jutten, editors, "Handbook of Blind Source Separation," Elsevier, 2010.

Blind Source Separation

$\mathbf{X}=\mathbf{AS}$ and $\mathbf{X}=(\mathbf{AT})(\mathbf{T}^{-1}\mathbf{S})$ are equivalent for any square invertible matrix $\mathbf{T}$. There are infinitely many pairs ($\mathbf{AT}$, $\mathbf{T}^{-1}\mathbf{S}$) satisfying linear mixture model $\mathbf{X}=\mathbf{AS}$.

Solutions unique up to permutation and scaling indeterminacies, $\mathbf{T}=\mathbf{P\Lambda}$, are meaningful. Constraints must be imposed on **A** and/or **S** in order to obtain solution of the BSS problem that is characterized with $\mathbf{T}=\mathbf{P\Lambda}$.

ICA solves BSS problem imposing statistical independence and non-Gaussianity constraints on source signals s_m, $m=1,\ldots,M$.

DCA improves accuracy of the ICA when sources s_m, $m=1,\ldots,M$, are not statistically independent.

SCA / NMF solves BSS problem imposing nonnegativity, sparseness, smoothness or some other constraints on source signals s_m, $m=1,\ldots,M$.

Multidimensional (tensorial) sources

N^{th} order tensor (also called N-way array) is N-dimensional array of, not necessary real, numbers:

$$\underline{\mathbf{X}} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$$

Each index is called way or mode and number of levels of a mode represents dimension of that mode, [4-7]. Eg., dimension of mode-1 is I_1.

Scalars, vectors and matrices are respectively tensors of order 0, 1 and 2. Sometimes tensors of order 3 are called hypermatrices.

4. H. A. L. Kiers, "Towards a standardized notation and terminology in multiway analysis," *J. Chemometrics*, **14**, no. 3, pp. 105-122, 2000.
5. A. Cichocki, R. Zdunek, A.H. Phan, S. Amari, Nonegative Matrix and Tensor Factorizations, John Wiley & Sons, 2009.
6. E. Acar, and B. Yener, "Unsupervised Multiway Data Analysis: A Literature Survey," *IEEE Trans. Knowl. Data Eng*. **21**, 6 (2009).
7. T.G. Kolda, and B.W. Bader, “Tensor Decompositions and Applications,” *SIAM Review* **51**, 453 (2009).

Multidimensional (tensorial) sources

A number of data sets is not naturally represented in 2D space but in N-D, $N \geq 3$, space. Few examples include: multispectral/hyperspectral image, video signal, EEG data, fluorescence spectroscopy data, magnetic resonance image, multi-phase CT image, etc.

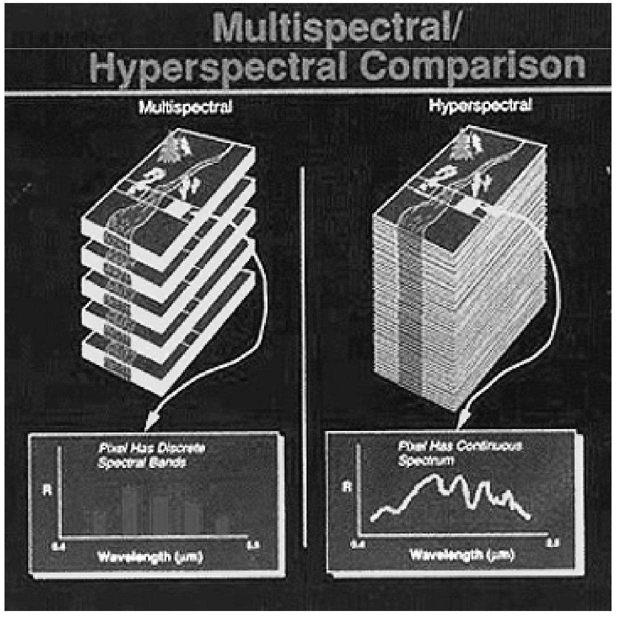
Multispectral-hyperspectral image (3D tensor)

$$\underline{\mathbf{X}} \in \mathbb{R}_{0+}^{I_1 \times I_2 \times I_3}$$

I_3 spectral images of the size $I_1 \times I_2$ pixels

RGB image contains I_3=3 spectral channels.

x_{i1i2i3} represents brightness intensity at spatial location indexed by (i_1, i_2) and spectral location indexed by i_3.



Multispectral magnetic resonance image (3D tensor)

I_3=3 (PD,T_1 and T_2) images of the size $I_1 \times I_2$ pixels

Tensor factorization

Very often for the purpose of exploratory data analysis, that includes the BSS methods such as ICA, DCA, SCA or NMF, 3D data are mapped to 2D data that is known as *matricization*, *unfolding* or *flattening*.

$$\underline{\mathbf{X}} \in \mathbb{R}_{0+}^{I_1 \times I_2 \times I_3} \xrightarrow{1} \mathbf{X}_{(1)} \in \mathbb{R}_{0+}^{I_1 \times I_2 I_3}$$

$$\underline{\mathbf{X}} \in \mathbb{R}_{0+}^{I_1 \times I_2 \times I_3} \xrightarrow{2} \mathbf{X}_{(2)} \in \mathbb{R}_{0+}^{I_2 \times I_1 I_3}$$

$$\underline{\mathbf{X}} \in \mathbb{R}_{0+}^{I_1 \times I_2 \times I_3} \xrightarrow{3} \mathbf{X}_{(3)} \in \mathbb{R}_{0+}^{I_3 \times I_1 I_2}$$

Problems:

- local structure of 3D data is lost (not exploited)
- matrix factorization assumed by linear mixing model **X**=**AS** suffers from indeterminacies because $\mathbf{ATT}^{-1}\mathbf{S}=\mathbf{X}$ for any invertible **T**, i.e. infinitely many (**A**,**S**) pairs can give rise to **X**.

- Meaningful solutions of the BSS problems are characterized by $\mathbf{T}=\mathbf{P}\Lambda$. To obtain them, matrix factorization methods such as ICA and/or NMF must respectively impose statistical independence and sparseness constraints on **S**.

Tensor products

The *n*-mode product of a tensor $\underline{\mathbf{X}}$ and a matrix **A** is written as:

$$\underline{\mathbf{X}} \times_n \mathbf{A}$$

Let $\underline{\mathbf{X}}$ be of size $I_1 \times I_2 \times I_3$ and let **A** be of size $J_1 \times J_2$.

The *n*-mode product multiplies vectors in mode *n* of $\underline{\mathbf{X}}$ with row vectors in **A**. Therefore, *n*-mode multiplication requires that $I_n = J_2$.

The result of the $\underline{\mathbf{X}} \times_n \mathbf{A}$ is a tensor with the same order (number of modes) as $\underline{\mathbf{X}}$ but with the size I_n replaced by J_1. After *n*-mode unfolding it follows:

$$\mathbf{Y}_{(n)} = \mathbf{A}\mathbf{X}_{(n)}$$

As an example, classical matrix product **AB** can be seen as a special case of *n*-mode product:

$$\mathbf{AB} = \mathbf{A} \times_2 \mathbf{B}^{\mathrm{T}} = \mathbf{B} \times_1 \mathbf{A}$$

Tensor models

Two most widely used tensor models are TuckerN model, [11], and Canonic Polyadic Decomposition (CPD)/PARAlel FACtor (PARAFAC) analysis /CANonical DECOMPosition (CANDECOMP) model, [12,13]. The Tucker3 model for 3D tensor is defined as:

$$\underline{\mathbf{X}} \approx \underline{\mathbf{G}} \times_1 \mathbf{A} \times_2 \mathbf{B} \times_3 \mathbf{C} = \sum_{j_1=1}^{R_1} \sum_{j_2=1}^{R_2} \sum_{j_3=1}^{R_3} g_{j_1 j_2 j_3} \mathbf{a}_{j_1} \circ \mathbf{b}_{j_2} \circ \mathbf{c}_{j_3}$$

$$x_{pqr} \approx \sum_{j_1=1}^{R_1} \sum_{j_2=1}^{R_2} \sum_{j_3=1}^{R_3} g_{j_1 j_2 j_3} a_{pj_1} b_{qj_2} a_{rj_3}$$

where $\underline{\mathbf{G}} \in \mathbb{R}^{R_1 \times R_2 \times R_3}$ is a core tensor and $\left\{\mathbf{A}, \mathbf{B}, \mathbf{C} \in \mathbb{R}^{I_n \times R_n}\right\}_{n=1}^{3}$ are factors.

11. L. R. Tucker, "Some mathematical notes on three-mode factor analysis," *Psychometrika* **31**, 279 (1966).
12. J. D. Carrol, and J. J. Chang, "Analysis of individual differences in multidimensional scaling via N-way generalization of Eckart-Young decomposition," *Psychometrika* **35**, 283 (1970).
13. R. A. Harshman, "Foundations of the PARAFAC procedure: models and conditions for an exploratory multi-mode factor analysis," *UCLA Working Papers in Phonetics* **16**, 1 (1970).

Tensor models

Tucker model has good generalization capability due to the fact that the core tensor allows interaction between a factor with any factor in other modes.

However, essential uniqueness of the factorization (up to permutation and scaling) is not guaranteed. That is because of:

$$\underline{\mathbf{X}} \approx \underline{\mathbf{G}} \times_1 \mathbf{A} \times_2 \mathbf{B} \times_3 \mathbf{C}$$
$$= \underline{\mathbf{G}} \times_1 \mathbf{T}_1^{-1} \times_2 \mathbf{T}_2^{-1} \times_3 \mathbf{T}_3^{-1} \times_1 \left(\mathbf{A}\mathbf{T}_1\right) \times_2 \left(\mathbf{B}\mathbf{T}_2\right) \times_3 \left(\mathbf{C}\mathbf{T}_3\right)$$

where $\left\{\mathbf{T}_n \in \mathbb{R}^{R_n \times R_n}\right\}_{n=1}^{3}$.

Some constraints have to be imposed on array factors and/or core tensor in order to ensure uniqueness of the factorization.

The CPD tensor model

The CPD model is a special case of the Tucker model when core tensor is diagonal i.e. $\underline{\mathbf{G}} = \underline{\mathbf{\Lambda}}$. CPD factorizes a tensor into a sum of rank-one tensors:

$$\underline{\mathbf{X}} \approx [\boldsymbol{\lambda}, \mathbf{A}, \mathbf{B}, \mathbf{C}] = \sum_{i=1}^{R} \lambda_i \mathbf{a}_i \circ \mathbf{b}_i \circ \mathbf{c}_i$$

where $\boldsymbol{\lambda} \in \mathbb{R}^{R}, \mathbf{A} \in \mathbb{R}^{I_1 \times R}, \mathbf{B} \in \mathbb{R}^{I_2 \times R}, \mathbf{C} \in \mathbb{R}^{I_3 \times R}$.

The mode-*3* matricized version of the tensor is given as:

$$\mathbf{X}_{(3)} = \mathbf{C}\mathbf{\Lambda}[\mathbf{B} \odot \mathbf{A}]^{T}$$

where $\mathbf{\Lambda} = diag(\boldsymbol{\lambda})$ and $\odot$ denotes the Khatri-Rao product.

The CPD tensor model

Assuming that $R \leq \min(I_1, I_2, I_3)$ uniqueness condition for CPD model is [14, 15]:

$$k(\mathbf{A}) + k(\mathbf{B}) + k(\mathbf{C}) \geq 2R + 3$$

where $k(\mathbf{A})$, $k(\mathbf{B})$ and $k(\mathbf{C})$ are Kruskal's ranks of factor **A**, **B** and **C**.

For a matrix $\mathbf{A} \in \mathbb{R}^{I \times J}$ standard rank $\mathbf{r}(\mathbf{A})$:=rank($\mathbf{A}$)=R if **A** contains collection of R linearly independent columns (rows), and this fails for R+1 columns (rows).

$k(\mathbf{A}) = R$ if *every* R columns are linearly independent, and this fails for at least one set of R+1 columns:

$$k(\mathbf{A}) \leq r(\mathbf{A}) \leq \min(I, J) \ \forall \ \mathbf{A}.$$

14. J. B. Kruskal, "Three-way arrays: Rank and uniqueness of trilinear decompositions," Linear Algebra Appl. **18, 95 (1977)**.
15. N. D. Sidiropoulos, and R. Bro, "On the uniqueness of multilinear decomposition of *N-way arrays," J. of Chemometrics* **14, 229 (2000)**.

Tensor factorization

Tensor based vs. matrix based mixture models.

•2D linear mixture model for 2D source signals:

$$\mathbf{X}_{(3)} = \mathbf{AS} \quad \mathbf{X}_{(3)} \in \mathbb{R}_{0+}^{I_3 \times I_1 I_2},\ \mathbf{A} \in \mathbb{R}_{0+}^{I_3 \times R}\ \mathbf{S} \in \mathbb{R}_{0+}^{R \times I_1 I_2}$$

In a case of MSI (or MRI) I_1 and I_2 represent image dimensions and I_3 represents number of spectral bands. In a case of video I_3 represents number of frames. R represents the unknown number of sources. Low-rank constraint implies: $R \leq I_3$.

•3D linear mixtures model with 2D sources signals:

$$\underline{\mathbf{X}} \approx \underline{\mathbf{G}} \times_1 \mathbf{A} \times_2 \mathbf{B} \times_3 \mathbf{C}$$

$$\underline{\mathbf{X}} \in \mathbb{R}_{0+}^{I_1 \times I_2 \times I_3}, \underline{\mathbf{G}} \in \mathbb{R}_{0+}^{R \times R \times R}, \left\{\mathbf{A}, \mathbf{B}, \mathbf{C} \in \mathbb{R}_{0+}^{I_n \times R}\right\}_{n=1}^{3}$$

Tensor factorization

3-mode unfolding of $\underline{\mathbf{X}}$ yields:

$$\mathbf{X}_{(3)} \approx \mathbf{C}\mathbf{G}_{(3)}\left[\mathbf{B}\otimes\mathbf{A}\right]^{\mathrm{T}}$$

Dimensionality analysis yields [16, 17]:

$$\mathbf{A} \approx \mathbf{C}$$

$$\underline{\mathbf{S}} \approx \underline{\mathbf{G}}\times_1 \mathbf{A}\times_2 \mathbf{B} \approx \underline{\mathbf{X}}\times_3 \left(\mathbf{C}\right)^{\dagger} \qquad \underline{\mathbf{S}} \in \mathbb{R}_{0+}^{I_1\times I_2\times R}$$

where '†' denotes Moore-Penrose pseudo-inverse and it is assumed $R \leq I_3$.

Thus, for MSI/MRI decomposition tensor factorization yields tensor of spatial distributions of materials/tissue substances present in the MSI/MRI.

16. **I. Kopriva,** A. Cichocki, "Blind Multi-spectral Image Decomposition by 3D Nonnegative Tensor Factorization," ***Optics Letters*** vol. 34, No. 14, pp 2210-2212, 2009.
17. **I. Kopriva**, "3D Tensor Factorization Approach to Single-frame Model-free Blind Image Deconvolution," ***Optics Letters***, Vol. 34, No.14, pp. 2210-2212, 2009.

Tensor factorization

Tensor factorization yields two formulas for calculating source signal tensor:

$$\underline{\mathbf{S}}^{dir} \approx \underline{\mathbf{G}} \times_1 \mathbf{A} \times_2 \mathbf{B}$$

$$\underline{\mathbf{S}}^{inv} \approx \underline{\mathbf{X}} \times_3 \left(\mathbf{C}\right)^{\dagger}$$

How to choose between these two formulas?

Which formula is more robust again perturbations of model factors $[\underline{\mathbf{G}}, \mathbf{A}, \mathbf{B}, \mathbf{C}]$ and/or data tensor $\underline{\mathbf{X}}$?

Perturbations of model factors

First order perturbation analysis of *direct* and *inverse* formulas for source tensor yields:

$$\delta\underline{\mathbf{S}}^{dir} \approx \delta\underline{\mathbf{G}} \times_1 \mathbf{A} \times_2 \mathbf{B} + \underline{\mathbf{G}} \times_1 \delta\mathbf{A} \times_2 \mathbf{B} + \underline{\mathbf{G}} \times_1 \mathbf{A} \times_2 \delta\mathbf{B}$$

$$\delta\underline{\mathbf{S}}^{inv} \approx -\underline{\mathbf{X}} \times_3 \begin{bmatrix} \mathbf{C}^{\dagger}\delta\mathbf{C}\mathbf{C}^{\dagger} + \mathbf{C}^{\dagger}\mathbf{C}^{\dagger\mathrm{T}}\delta\mathbf{C}^{\mathrm{T}}\left(\mathbf{I}_{I_3} - \mathbf{C}\mathbf{C}^{\dagger}\right) \\ + \left(\mathbf{I}_R - \mathbf{C}^{\dagger}\mathbf{C}\right)\delta\mathbf{C}^{\mathrm{T}}\mathbf{C}^{\dagger\mathrm{T}}\mathbf{C}^{\dagger} \end{bmatrix}$$

Where I_n denotes $n \times n$ identity matrix. For square invertible matrix **C** expression for *inverse* formula becomes [18]:

$$\delta\underline{\mathbf{S}}^{inv} \approx -\underline{\mathbf{X}} \times_3 \left[\mathbf{C}^{-1}\delta\mathbf{C}\mathbf{C}^{-1}\right]$$

18. A. Hjørungnes. and D. Gesbert, "Complex-Valued Matrix Differentiation: Techniques and Key Results," *IEEE Transactions on Signal Processing*, vol. 55, pp. 2740-2746, 2007.

Perturbations of model factors

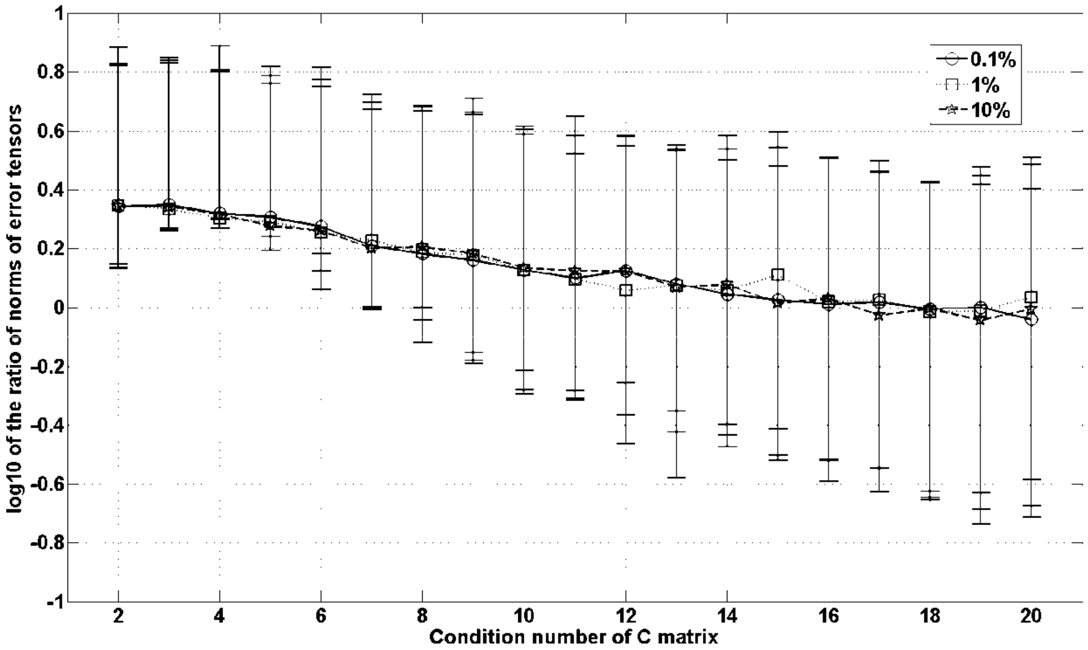
This type of perturbation relates to rounding errors and/or conversion to local minima.

Model factors $\underline{\mathbf{G}}, \mathbf{A}, \mathbf{B}, \mathbf{C}$ were perturbed by i.i.d. nonnegative uniformly distributed noise in the amounts of 0.1%, 1% and 10% of the Frobenius norms of the true values of loading factors.

Data tensor: 856×1144×3. Rank of factor matrices was *R*=3. The 3×3 **C** matrix has been generated with controlled condition number between 2 and 20 in steps of 1.

The crucial point is conditioning of **C** matrix. When **C** matrix is *sufficiently well* condtioned the *inverse* formula will be more accurate.

Perturbations of model factors



$\log_{10}$ of the ratio of the Frobenius norms of error tensors:

$$\log_{10}\left(\left\|\delta\underline{\mathbf{S}}^{dir}\right\| / \left\|\delta\underline{\mathbf{S}}^{inv}\right\|\right) \quad .$$

The *inverse* formula yields smaller error when condition number is less than 16.

Unsupervised Decomposition of Multispectral Image

True hyperspectral/multispectral image is a 3D tensor. Hence, blind image decomposition can be performed through 3D NTF, [16].

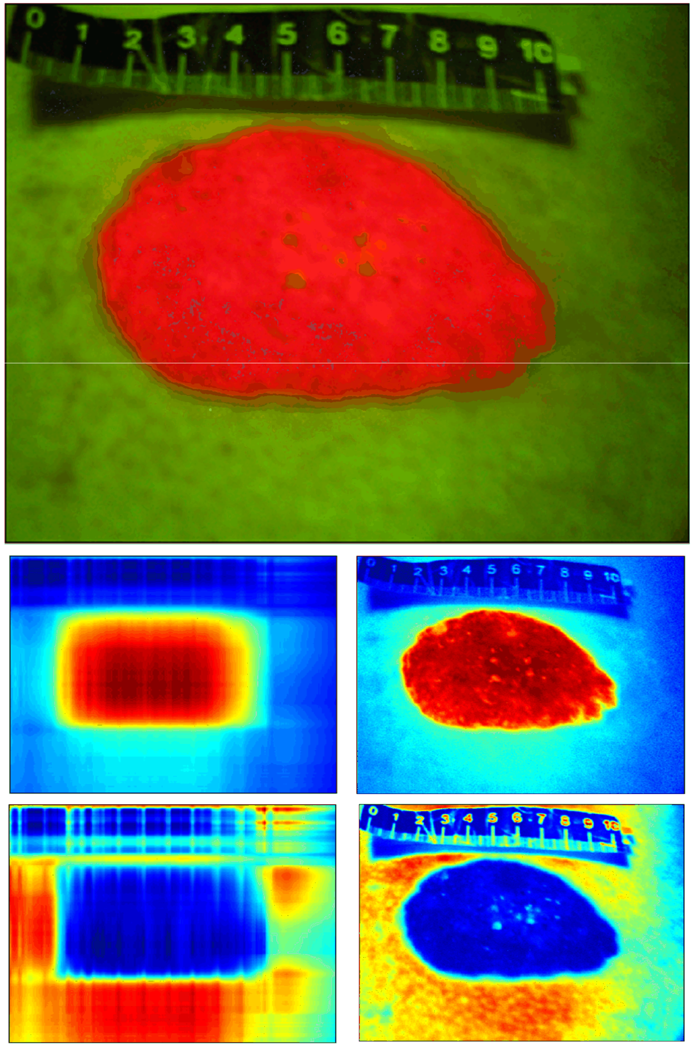
It is related to the unsupervised decomposition of fluorescent RGB image of a skin tumor (basal cell carcinoma) by means of the α-NTF algorithm (α=0.1 was chosen in this case), [19]. Here, α-divergence is just a choice and other cost functions could be used as well.

The ground truth is simple and visible on the RGB image itself. The image contains fluorescent tumor component in red color and background component (composed of surrounding healthy skin and the ruler) in green and black colors, that is the **C** matrix is 3×2 matrix.

19. Y. D. Kim, A. Cichocki, and S. Choi, "Nonnegative Tucker Decomposition with Alpha-Divergence," in Proceedings of the 2008 *IEEE International Conference on Acoustics, Speech and Signal Processing*, Las Vegas, NV, USA, March 30-April 4, 2008, pp. 1829-1832.

"ERROR ANALYSIS OF LOW-RANK THREE-WAY TENSOR FACTORIZATION APPROACH TO BLIND SOURCE SEPARATION,"
2014 IEEE International Conference on Acoustics, Speech and Signal Processing, Florence, Italy, May 7, 2014.

Unsupervised Decomposition of Multispectral Image



Top row: experimental RGB fluorescent image of a skin tumor that stands for measurement tensor of dimensions 856×1144×3. **Mid row**: intensity maps of tumor component. Left *direct* formula; right: *inverse* formula. **Bottom row**: intensity maps of background component. **Left**: *direct* formula; **right**: *inverse* formula. Intensity maps are scaled to [0,1] interval and shown in pseudo-color such that dark red indicates that component is present with probability 1, while dark blue indicates that component is present with probability 0.

Although result obtained by *inverse* formula is better, it is seen that *direct* formula also yields result that is meaningful. Thus, if the conditioning of the **C** matrix happens to be poor (it could be due to the existence of *spectrally similar* objects) *direct* formula can be useful.

Perturbations of measurement/data tensor

Data tensor $\underline{\mathbf{X}}$ has been perturbed. Corresponding perturbations of factor matrices **A, B** and **C** as well as diagonal core tensor $\underline{\mathbf{\Lambda}}$ were calculated.

This type of analysis makes sense only when decomposition is unique. Thus, CPD model is analyzed and Tucker3 model is discarded in this perturbation analysis.

It is necessary to obtain expressions for δ**A**, δ**B**, δ**C** and $\delta\underline{\mathbf{\Lambda}}$ as a function of $\delta\underline{\mathbf{X}}$. For this purpose we consider first order perturbation of the CPD model of $\underline{\mathbf{X}}$ [20]:

$$\delta\mathbf{X}_{(3)} \approx \delta\mathbf{C}\mathbf{\Lambda}[\mathbf{B}\odot\mathbf{A}]^{\mathrm{T}} + \mathbf{C}\delta\mathbf{\Lambda}[\mathbf{B}\odot\mathbf{A}]^{\mathrm{T}} + \mathbf{C}\mathbf{\Lambda}[\delta\mathbf{B}\odot\mathbf{A}]^{\mathrm{T}} + \mathbf{C}\mathbf{\Lambda}[\mathbf{B}\odot\delta\mathbf{A}]^{\mathrm{T}}$$

where $\odot$ denotes Khatri-Rao product and $\mathbf{\Lambda}$ is $R\times R$ diagonal matrix if R denotes rank of $\underline{\mathbf{X}}$.

20. C. R. Rao, Linear Statistical Inference and its Applications, ser. Probability and Statistics. Wiley, 1965.

Perturbations of measurement/data tensor

The expansion is based on the mode-*3* unfolding of $\underline{\mathbf{X}}$:

$$\mathbf{X}_{(3)} = \mathbf{C}\boldsymbol{\Lambda}\left[\mathbf{B} \odot \mathbf{A}\right]^{\mathrm{T}}$$

Perturbation is a linear system and can be written in more convenient form by defining the following vectors:

$$\delta\mathbf{z} \triangleq \begin{bmatrix} vec\{\delta\mathbf{A}\} \\ vec\{\delta\mathbf{B}\} \\ vec\{\delta\mathbf{C}\} \\ vecd\{\delta\boldsymbol{\Lambda}\} \end{bmatrix} \text{ and } \delta\mathbf{x} \triangleq vec\{\delta\mathbf{X}_{(3)}\}$$

Perturbations of measurement/data tensor

vec{.} and *vecd* {.} respectively mean :

$$vec\{\delta\mathbf{A}\} \triangleq \begin{bmatrix} \delta\mathbf{a}_1 \\ \ldots \\ \delta\mathbf{a}_R \end{bmatrix} \quad vec\{\delta\mathbf{B}\} \triangleq \begin{bmatrix} \delta\mathbf{b}_1 \\ \ldots \\ \delta\mathbf{b}_R \end{bmatrix} \quad vec\{\delta\mathbf{C}\} \triangleq \begin{bmatrix} \delta\mathbf{c}_1 \\ \ldots \\ \delta\mathbf{c}_R \end{bmatrix} \quad vecd\{\delta\mathbf{\Lambda}\} \triangleq \begin{bmatrix} \delta\lambda_{11} \\ \ldots \\ \delta\lambda_{RR} \end{bmatrix}$$

Thus, we have to solve for the linear system δ**x**= Mδ**z** . Thereby, **M** is the 4-block matrix:

$$\mathbf{M} \triangleq \left[\begin{array}{c|c} \mathbf{K}_{I_1I_2,R}\left(\mathbf{C\Lambda}\otimes\mathbf{I}_{I_1I_2}\right)D(\mathbf{B}) & \mathbf{K}_{I_1I_2,R}\left(\mathbf{C\Lambda}\otimes\mathbf{I}_{I_1I_2}\right)D(\mathbf{A}) \\ \left(\mathbf{B}\odot\mathbf{A}\right)\mathbf{\Lambda}\otimes\mathbf{I}_{I_3} & \left(\mathbf{B}\odot\mathbf{A}\right)\odot\mathbf{C} \end{array}\right]$$

Perturbations of measurement/data tensor

Here:

$$D(\mathbf{A}) = Diag\left\{\mathbf{I}_{I_2} \otimes \mathbf{a}_1, \ldots, \mathbf{I}_{I_2} \otimes \mathbf{a}_R\right\} \quad D(\mathbf{B}) = Diag\left\{\mathbf{b}_1 \otimes \mathbf{I}_{I_1}, \ldots, \mathbf{b}_R \otimes \mathbf{I}_{I_1}\right\}$$

are diagonal matrices of size $I_1I_2R\times I_2R$ and $I_1I_2R\times I_1R$.

$\mathbf{K}_{I_1I_2,R}$ is square permutation matrix of dimensions $I_1I_2R\times I_1I_2R$:

$$\mathbf{K}_{I_1I_2,R} \triangleq \sum_{i=1}^{I_1I_2}\sum_{r=1}^{R}\mathbf{E}_{ir}^{(I_1I_2\times R)}\mathbf{E}_{ri}^{(R\times I_1I_2)} \quad \mathbf{E}_{ir}^{(I_1I_2\times R)} \triangleq \mathbf{e}_i\mathbf{e}_r^T$$

$\mathbf{e}_i$ is unit vector in $\mathbb{R}^{I_1I_2}$ and $\mathbf{e}_r$ is unit vector in $\mathbb{R}^R$.

Perturbations of measurement/data tensor

The matrix **M** is of size $I_1 I_2 I_3 \times (I_1 + I_2 + I_3 + 1)R$.

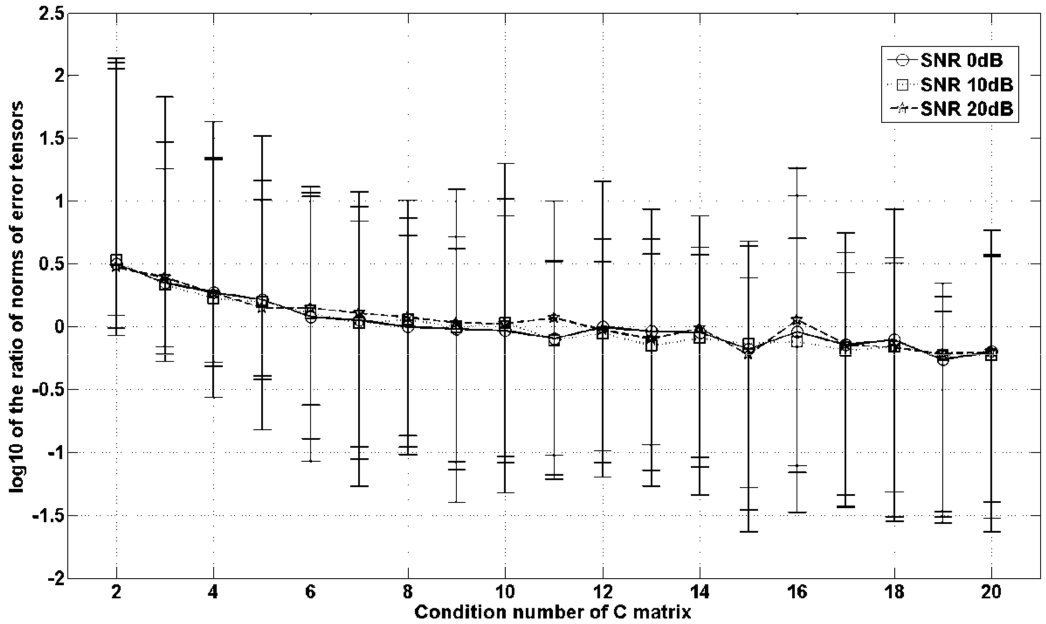
Due to low-rank constraint , $R \leq I_3$, and because I_3 is small (it stands for number of channels that in case of RGB or image is $I_3 = 3$) matrix **M** has less columns than rows. Thus, there are more equations than unknowns.

In computer simulation data tensor $\underline{\mathbf{X}}$ was of size 50×50×3. For each realization entries of loading matrices **A**, **B** and **C** were drawn from nonnegative uniform distribution with number of columns R=3. The core tensor $\underline{\mathbf{\Lambda}}$ was generated with nonnegative uniformly distributed values on diagonal.

The 3×3 **C** matrix has been generated with controlled conditioned number between 2 and 20 in steps of 1. Entries of perturbation tensor $\delta\underline{\mathbf{X}}$ were drawn independently according to nonnegative uniform distribution. Froebenius norm of $\delta\underline{\mathbf{X}}$ has been determined from predefined signal-to-noise-ratio:

$$SNR = 20\log_{10}\left(\|\underline{\mathbf{X}}\| / \|\delta\underline{\mathbf{X}}\|\right)$$

Perturbations of measurement/data tensor



Log_{10} of the ratio of the Frobenius norms of error tensors:

$$\log_{10}\left(\left\|\delta\underline{\mathbf{S}}^{dir}\right\| / \left\|\delta\underline{\mathbf{S}}^{inv}\right\|\right)$$

The *inverse* formula yields smaller error when condition number of **C** matrix is less than 8. The fact that *inverse* formula is more sensitive to measurement noise than noise in loading factors is expected since it amplifies noise via $\mathbf{C}^{\dagger}$.

Presented results supplement the one related to CRLB in [21]. While CRLB predicts error bounds on parameter of the CPD model under a white Gaussian noise assumption, the error analysis presented herein can be performed for arbitrary distribution of the additive noise.

21. X. Q. Liu, and N. D. Sidiropoulos, "Cramer-Rao lower bounds for low-rank decomposition of multidimensional arrays," *IEEE Transactions on Signal Processing,* vol. 49, pp. 2074-2086, 2001.

SUMMARY

In factorization of (three-way) tensors *direct* and *inverse* formulas for calculating source tensor emerge.

If errors are due to perturbations in loading matrices *inverse* formula is better when condition number of the mode-*3* loading matrix is smaller than or equal to 16.

In case of measurement noise, *inverse* formula is better when condition number of mode-*3* loading matrix is smaller than or equal to 8.

Topic for future analysis is related to probabilistic formulation that complies with some predefined (sparseness and/or nonnegativity) constraints on factors of the model. That can lead to interesting results regarding essential uniqueness of the Tucker3(N) tensor model?!