Lecture XI
Tensor Factorization: blind separation of multidimensional sources and feature extraction

Ivica Kopriva

Ruđer Bošković Institute

e-mail: ikopriva@irb.hr ikopriva@gmail.com
Web: http://www.lair.irb.hr/ikopriva/

Course outline

- Motivation with illustration of applications (lecture I)
- Mathematical preliminaries with principal component analysis (PCA)? (lecture II)
- Independent component analysis (ICA) for linear static problems: information-theoretic approaches (lecture III)
- ICA for linear static problems: algebraic approaches (lecture IV)
- ICA for linear static problems with noise (lecture V)
- Dependent component analysis (DCA) (lecture VI)

Course outline

- Underdetermined blind source separation (BSS) and sparse component analysis (SCA) (lecture VII/VIII)
- Nonnegative matrix factorization (NMF) for determined and underdetermined BSS problems (lecture VIII/IX)
- BSS from linear convolutive (dynamic) mixtures (lecture X/XI)
- Nonlinear BSS (lecture XI/XII)
- Tensor factorization (TF): BSS of multidimensional sources and feature extraction (lecture XIII/XIV)

Seminar problems

1. Blind separation of two uniformly distributed signals with maximum likelihood (ML) and AMUSE/SOBI independent component analysis (ICA) algorithm. Blind separation of two speech signals with ML and AMUSE/SOBI ICA algorithm. **Theory, MATLAB demonstration and comments of the results.**

2. Blind decomposition/segmentation of multispectral (RGB) image using ICA, dependent component analysis (DCA) and nonnegative matrix factorization (NMF) algorithms. **Theory, MATLAB demonstration and comments of the results.**

3. Blind separation of acoustic (speech) signals from convolutive dynamic mixture. **Theory, MATLAB demonstration and comments of the results.**

Seminar problems

4. Blind separation of images of human faces using ICA and DCA algorithms (innovation transform and ICA, wavelet packets and ICA) **Theory, MATLAB demonstration and comments of the results.**

5. Blind decomposition of multispectral (RGB) image using sparse component analysis (SCA): clustering + L_p norm ($0<p\leq 1$) minimization. **Theory, MATLAB demonstration and comments of the results.**

6. Blind separation of four sinusoidal signals from two static mixtures (a computer generated example) using sparse component analysis (SCA): clustering + L_p norm ($0<p\leq 1$) minimization in frequency (Fourier) domain. **Theory, MATLAB demonstration and comments of the results.**

Seminar problems

7. Blind separation of three acoustic signals from two static mixtures (a computer generated example) using sparse component analysis (SCA): clustering + L_p norm ($0<p\leq 1$) minimization in time-frequency (short-time Fourier) domain. **Theory, MATLAB demonstration and comments of the results.**

8. Blind extraction of five pure components from mass spectra of two static mixtures of chemical compounds using sparse component analysis (SCA): clustering a set of single component points + L_p norm ($0<p\leq 1$) minimization in m/z domain. **Theory, MATLAB demonstration and comments of the results.**

9. Feature extraction from protein (mass) spectra by ***tensor factorization*** of disease and control samples in joint bases. Prediction of prostate/ovarian cancer. **Theory, MATLAB demonstration and comments of the results.**

Blind Source Separation

A theory for multichannel blind signal recovery requiring minimum of *a priori* information.

Problem:

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

N-number of sensors;
M- *unknown* number of sources
T-number of samples

Goal: find **S, A** and number of sources *M* based on **X** only.

Meaningful solutions are characterized by scaling and permutation indeterminacies:

$$\mathbf{Y}\cong\mathbf{S}=\mathbf{WX} \rightarrow \mathbf{Y} \cong\mathbf{WAS}=\mathbf{P}\Lambda\mathbf{S}$$

A. Hyvarinen, J. Karhunen, E. Oja, “Independent Component Analysis,” John Wiley, 2001.
A. Cichocki, S. Amari, “Adaptive Blind Signal and Image Processing,” John Wiley, 2002.

Blind Source Separation

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

Constraints must be imposed on **A** and/or **S** in order to obtain solution of the BSS problem that is characterized with **T=P**Λ.

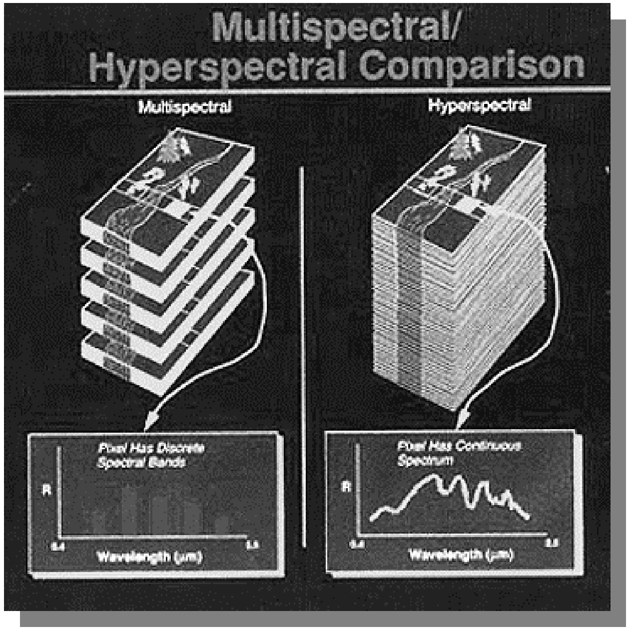
ICA solves BSS problem imposing statistical independence and non-Gaussianity constraints on source signals s_m, m=1,…,M.

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

NMF solves BSS problem imposing sparseness, smoothness or some other constraints on source signals s_m, m=1,…,M.

Tensor factorization[1,2,3]

A number of data sets is not naturally represented in 2D space but in 3D (ND, N>3 space). Few examples include: multispectral/hyperspectral image, video signal, EEG data, fluorescence spectroscopy data, magnetic resonance image, etc.



Multispectral-hyperspectral image cube (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

Magnetic resonance image cube (3D tensor)

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

1. A. Cichocki, R. Zdunek, A.H. Phan, S. Amari, Nonegative Matrix and Tensor Factorizations, John Wiley & Sons, 2009.
2. E. Acar, and B. Yener, "Unsupervised Multiway Data Analysis: A Literature Survey," IEEE Trans. Knowl. Data Eng. **21**, 6 (2009).
3. T.G. Kolda, and B.W. Bader, “Tensor Decompositions and Applications,” SIAM Review **51**, 453 (2009).

Tensor factorization[1]

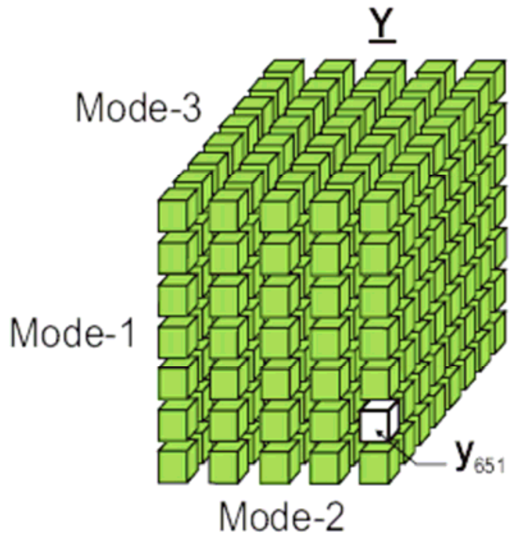


Fig. 1.14 Three-way array - third order tensor $\underline{\mathbf{Y}} \in \mathbb{R}^{7\times5\times8}$ with indicated individual elements y_{itq}.

Tensor factorization[1]

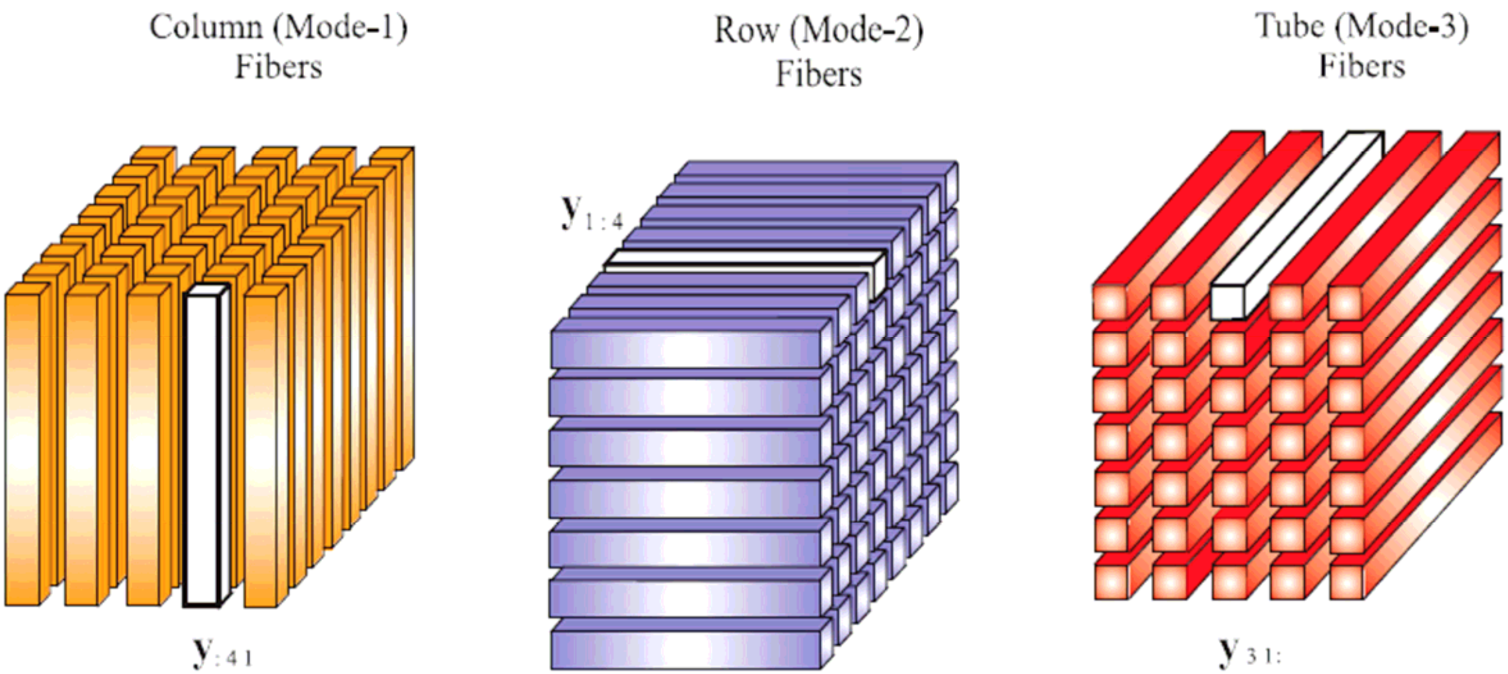


Fig. 1.16 Fibers: for a third order tensor $\underline{\mathbf{Y}} = [y_{itq}] \in \mathbb{R}^{I \times T \times Q}$ (All fibres are treated as column vectors).

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*.

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

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

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

<u>Problems:</u>

- local structure of 3D data is not used
- matrix factorization assumed by linear mixing model **X=AS** suffers from indeterminacies because $\mathbf{A}\mathbf{T}\mathbf{T}^{-1}\mathbf{S}=\mathbf{X}$ 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 factorization[1]

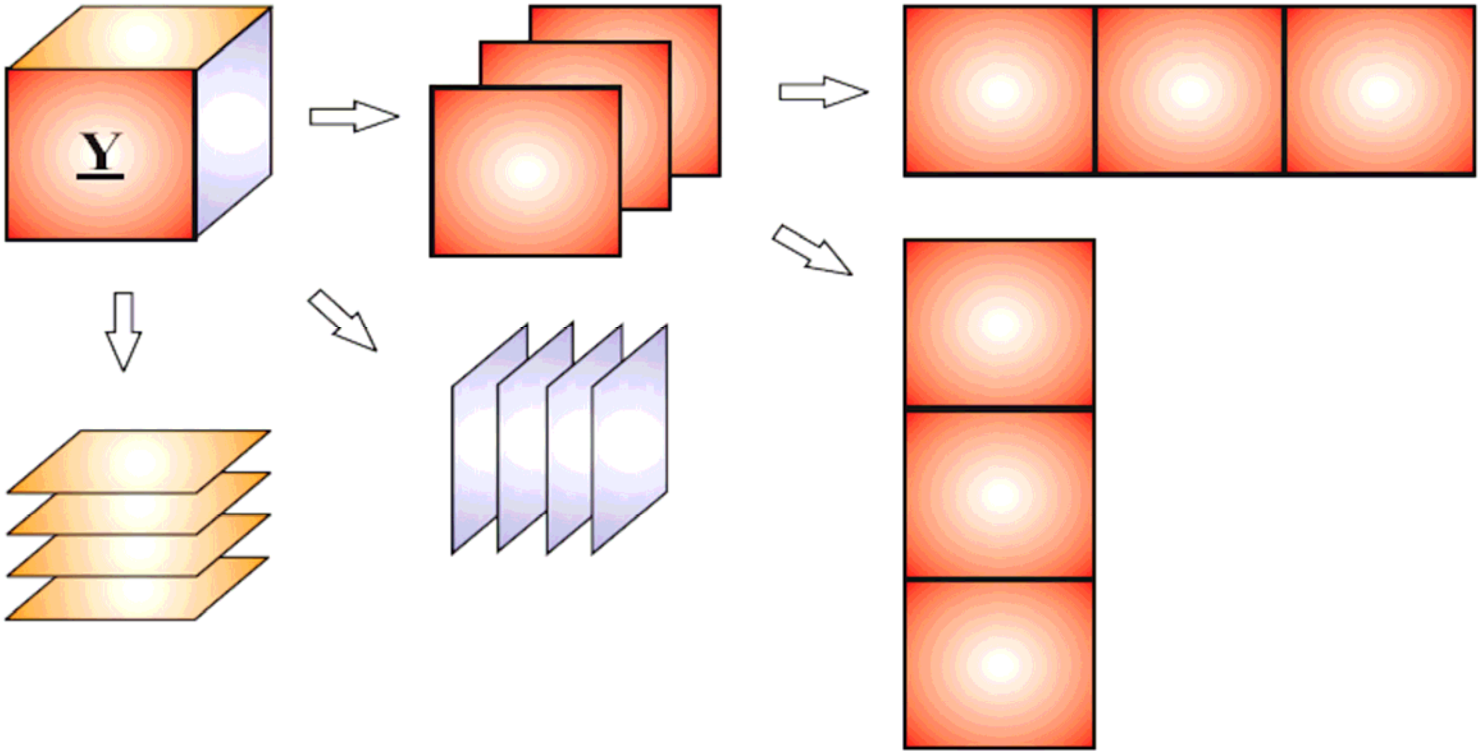


Fig. 1.19 Illustration of row-wise and column-wise unfolding (flattening, matricizing)of a third-order tensor.

Tensor factorization[1]

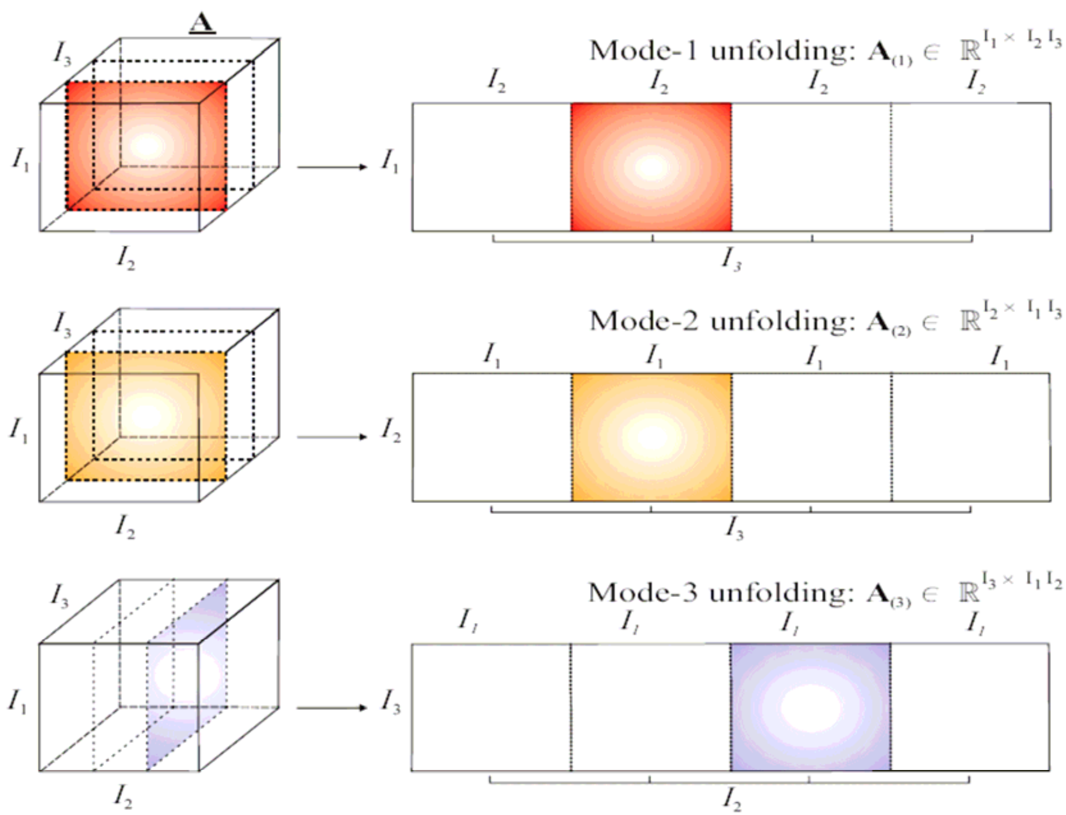


Fig. 1.20 Unfolding of a third-order tensor. The tensor can be unfolded in 3 ways to obtain matrices comprising its mode-1, mode-2 and mode-3 vectors.

Tensor products

Among several types of tensor products we are interested in *n*-mode product.

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.

For 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[a] and Canonic Polyadic Decomposition (CPD)/PARAlel FACtor (PARAFAC) analysis /CANonical DECOMPosition (CANDECOMP) model, [4-6]. The Tucker3 model for 3D tensor is defined as:

$$\underline{\mathbf{X}} \approx \underline{\mathbf{G}} \times_1 \mathbf{A}^{(1)} \times_2 \mathbf{A}^{(2)} \times_3 \mathbf{A}^{(3)} = \sum_{j_1=1}^{J_1} \sum_{j_2=1}^{J_2} \sum_{j_3=1}^{J_3} g_{j_1 j_2 j_3} \mathbf{a}_{j_1}^{(1)} \circ \mathbf{a}_{j_2}^{(2)} \circ \mathbf{a}_{j_3}^{(3)}$$

$$x_{pqr} \approx \sum_{j_1=1}^{J_1} \sum_{j_2=1}^{J_2} \sum_{j_3=1}^{J_3} g_{j_1 j_2 j_3} a_{pj_1}^{(1)} a_{qj_2}^{(2)} a_{rj_3}^{(3)}$$

where $\underline{\mathbf{G}} \in \mathbb{R}_{0+}^{J_1 \times J_2 \times J_3}$ is core tensor and $\left\{ \mathbf{A}^{(n)} \in \mathbb{R}_{0+}^{I_n \times J_n} \right\}_{n=1}^{3}$ are factors.

4. L. R. Tucker, "Some mathematical notes on three-mode factor analysis," Psychometrika 31, 279 (1966).
5. 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).
6. 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, uniqueness of the factorization up to permutation and scaling is not guaranteed. That is because:

$$
\begin{aligned}
\underline{\mathbf{X}} &\approx \underline{\mathbf{G}} \times_1 \mathbf{A}^{(1)} \times_2 \mathbf{A}^{(2)} \times_3 \mathbf{A}^{(3)} \\
&= \underline{\mathbf{G}} \times_1 \mathbf{T}^{(1)^{-1}} \times_2 \mathbf{T}^{(2)^{-1}} \times_3 \mathbf{T}^{(3)^{-1}} \times_1 \left(\mathbf{A}^{(1)}\mathbf{T}^{(1)}\right) \times_2 \left(\mathbf{A}^{(2)}\mathbf{T}^{(2)}\right) \times_3 \left(\mathbf{A}^{(3)}\mathbf{T}^{(3)}\right)
\end{aligned}
$$

where $\left\{\mathbf{T}^{(n)} \in \mathbb{R}^{J_n \times J_n}\right\}$. Hence, some constraints are necessary to be imposed on array factors and/or core tensor in order to ensure uniqueness of the factorization

The CPD tensor model

The CPD/PARAFAC model is a special case of the Tucker model when core tensor is superdiagonal i.e. $\underline{\mathbf{G}} = \underline{\mathbf{I}}$.

$$\underline{\mathbf{X}} \approx \underline{\mathbf{I}} \times_1 \mathbf{A}^{(1)} \times_2 \mathbf{A}^{(2)} \times_3 \mathbf{A}^{(3)} = \sum_{j_1=1}^{J_1} \sum_{j_2=1}^{J_2} \sum_{j_3=1}^{J_3} \mathbf{a}_{j_1}^{(1)} \circ \mathbf{a}_{j_2}^{(2)} \circ \mathbf{a}_{j_3}^{(3)}$$

$$x_{pqr} \approx \sum_{j_1=1}^{J_1} \sum_{j_2=1}^{J_2} \sum_{j_3=1}^{J_3} a_{pj_1}^{(1)} a_{qj_2}^{(2)} a_{rj_3}^{(3)}$$

Thus, factors in different mode can only interact factorwise. However, this restriction enables uniqueness of tensor factorization based the PARAFAC model within the permutation and scaling indeterminacies of the factors under very mild conditions, [4,5], without need to impose any special constraints on them such as sparseness or statistical independence.

The CPD tensor model

Assuming that *J*1=*J*2=*J*3=*J* uniqueness condition is reduced to:

$$k_{\mathbf{A}^{(1)}} + k_{\mathbf{A}^{(2)}} + k_{\mathbf{A}^{(3)}} \geq 2J + 2$$

where $k_{\mathbf{A}^{(n)}}$ is Kruskal’s rank of factor $\mathbf{A}^{(n)}$, [7]. The result is generalized in [8] for the CPD/PARAFAC model of the Nth order tensor:

$$\sum_{n=1}^{N} k_{\mathbf{A}^{(n)}} \geq 2J + (N-1)$$

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

The CPD tensor model

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

$k_{\mathbf{A}}$ (the Kruskal's rank of **A**)=r if *every* r columns are linearly independent, and this fails for at least one set of r+1 columns:

$$k_{\mathbf{A}} \leq r_{\mathbf{A}} \leq \min(I,J) \ \forall \ \mathbf{A}.$$

Condition that ensures uniqueness of the CPD/PARAFAC decomposition with probability one for 3-way tensor is, [9]:

$$J \leq I_3 \text{ and } J(J-1) \leq I_1(I_1-1)\, I_2(I_2-1)/2$$

9. L. De Lathauwer, “A link between the canonical decomposition in multilinear algebra and simultaneous matrix diagonalization,” SIAM Journal on Matrix Analysis and Applications 28, 642 (2006).

Tensor based vs. matrix based mixture models

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

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

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. *J* represents the unknown number of sources.

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

$$\underline{\mathbf{X}} \approx \underline{\mathbf{G}} \times_1 \mathbf{A}^{(1)} \times_2 \mathbf{A}^{(2)} \times_3 \mathbf{A}^{(3)}$$

$$\underline{\mathbf{X}} \in \mathbb{R}_{0+}^{I_1 \times I_2 \times I_3}, \underline{\mathbf{G}} \in \mathbb{R}_{0+}^{J_1 \times J_2 \times J_3}, \left\{ \mathbf{A}^{(n)} \in \mathbb{R}_{0+}^{I_n \times J_n} \right\}_{n=1}^{3}$$

Tensor based vs. matrix based mixture models

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

$$\mathbf{X}_{(3)} \approx \mathbf{A}^{(3)}\mathbf{G}_{(3)}\left[\mathbf{A}^{(2)} \otimes \mathbf{A}^{(1)}\right]^{\mathrm{T}}$$

Dimensionality analysis yields, [10,11]:

$$\mathbf{A} \approx \mathbf{A}^{(3)}$$

$$\underline{\mathbf{S}} \approx \underline{\mathbf{G}} \times_1 \mathbf{A}^{(1)} \times_2 \mathbf{A}^{(2)} = \underline{\mathbf{X}} \times_3 \left(\mathbf{A}^{(3)}\right)^{\dagger} \qquad \underline{\mathbf{S}} \in \mathbb{R}_{0+}^{I_1 \times I_2 \times J_3}$$

where '†' denotes Moore-Penrose pseudo-inverse and it is assumed $J_3 \leq I_3$.

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

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

The CPD tensor model

Despite limited accuracy in modeling complex relations among data the CPD is useful in many instances due to mild uniqueness conditions. Assuming *J*-component model (array factors have the same number of columns) Nth order tensor can also be written as linear combination of *J* rank-1 terms:

$$\underline{\mathbf{X}} \approx \sum_{j=1}^{J} \mathbf{a}_j^{(1)} \circ \mathbf{a}_j^{(2)} \circ \ldots \circ \mathbf{a}_j^{(N)}$$

Important property of CPD is that it can be written in a slice-wise fashion as:

$$\underline{\mathbf{X}}_{:,:,i_3,\ldots,i_N} \approx \mathbf{A}^{(1)} \cdot diag\left(a_{i_3 1}^{(3)} \ldots a_{i_N 1}^{(N)}, a_{i_3 2}^{(3)} \ldots a_{i_N 2}^{(N)}, \ldots, a_{i_3 J}^{(3)} \ldots a_{i_N J}^{(N)}\right) \mathbf{A}^{(2)^T}$$

that in a case of N=3 reduces to:

$$\underline{\mathbf{X}}_{:,:,i_3} \approx \mathbf{A}^{(1)} \cdot diag\left(a_{i_3 1}^{(3)}, a_{i_3 2}^{(3)}, \ldots, a_{i_3 J}^{(3)}\right) \mathbf{A}^{(2)^T}$$

Hence in a case of 3D tensor the CPD is nothing but simultaneous diagonalization of matrix slices.

The CPD tensor model

In the context of BSS the CPD allows, in principle, only for an approximation of the data.

Decomposition if fitted to the given data tensor $\underline{\mathbf{X}}$, very often in the least square sense. The “workhorse” approach is alternating least square (ALS) procedure:

1. To estimate the factor $\mathbf{A}^{(n)}$ data tensor $\underline{\mathbf{X}}$ is unfoled in mode n:

$$\mathbf{X}_{(n)} \approx \mathbf{A}^{(n)}\mathbf{Z}_{-n}$$

$$\mathbf{Z}_{-n} = \left[\mathbf{A}^{(N)} \otimes \ldots \otimes \mathbf{A}^{(n+1)} \otimes \mathbf{A}^{(n-1)} \otimes \ldots \otimes \mathbf{A}^{(1)}\right]^{\mathrm{T}}$$

2. the problem becomes least square NMF problem:

$$\left(\hat{\mathbf{A}}^{(n)}, \hat{\mathbf{Z}}_{(-n)}\right) = \underset{\mathbf{A}^{(n)}, \mathbf{Z}_{(-n)}}{\arg\min} \frac{1}{2}\left\|\mathbf{X} - \mathbf{A}^{(n)}\mathbf{Z}_{(-n)}\right\|_2^2 \quad s.t.\, \mathbf{A}^{(n)} \geq \mathbf{0}, \mathbf{Z}_{(-n)} \geq \mathbf{0}$$

The CPD tensor model

ALS method is easy and it often works well. However, it can be very slow, especially when the problem is ill conditioned. Moreover, convergence of the ALS method to the global minimum is not guaranteed. Also, for symmetric tensor the ALS procedure breaks symmetry. Convergence properties of the ALS method can be improved by the “enhanced line search” method, [12].

Although, CPD is often unique as such, it makes sense to impose orthogonality constraints on factor matrices if they are known to apply. That may improve convergence, increase the accuracy and even enable uniqueness for higher values of J (number of components).

The cost function associated with the J-rank approximation of the higher-order tensor does not always has a minimum but only infimum. Then, decrease of the cost function essentially makes some terms go to infinity.

12. M. Rajih, P. Comon, R. Harshman, “Enhanced line search: A novel method to accelerate PARAFAC,” ***SIAM J. Matrix Anal. Appl.,*** vol. 30, no. 3, pp. 1148-1171, Sept. 2008.

The CPD tensor model

These diverging terms almost completely cancel each other and the overall sum yields better and better approximation of the data tensor. However, the result (factors) is poor.

This problem indicates that CPD is not exact. In the BSS scenario, this indicates that the noise (error) level is too high, that underlying components can not be represented by rank-1 tensor, etc.

Sometimes, divergence can be prevented by imposing additional constraints (nonnegativity, orthogonality) on the factor matrices. However, if it is impossible to impose meaningful constraints then the CPD is not the right model to the data $\underline{\mathbf{X}}$.

The CPD tensor model

The CPD can be used to solve several algebraic ICA/BSS problems.

Let us have **X=AS**, $\mathbf{X} \in \mathbb{R}^{I_1 \times J}, \mathbf{A} \in \mathbb{R}^{I_1 \times J}, \mathbf{S} \in \mathbb{R}^{J \times I_2}$

Non-gaussian i.i.d. sources. The n-th order cumulant of **X** is defined as:

$$C_{\mathbf{X}}^{(n)} = \sum_{j=1}^{J} C_{s_j}^{(n)} \mathbf{a}_j \circ \mathbf{a}_j \circ \ldots \circ \mathbf{a}_j \quad n \geq 3$$

i.e. the CPD does apply. The cross-terms in the above expression vanished due to statistical independence of the sources. By re-writing above expression in a slice-wise fashion we obtain:

$$C_{\mathbf{X}}^{(n)} :,:,i_3,\ldots,i_n \approx \mathbf{A} \cdot diag\left(C_{s_1}^{n} a_{i_3 1} \ldots a_{i_n 1}, C_{s_2}^{n} a_{i_3 2} \ldots a_{i_n 2}, \ldots, C_{s_J}^{n} a_{i_3 J} \ldots a_{i_n J} \right) \mathbf{A}^{T}$$

i.e. mixing matrix **A** is obtained as CPD of symmetric cumulant tensor $C_{\mathbf{X}}^{(n)}$:

$$\hat{\mathbf{A}} = \mathbf{A}^{(1)}$$

The CPD tensor model

Individually correlated in time and mutually uncorrelated sources. The covariance matrices for different time lags τ_i are defined as:

$$C_{\mathbf{X}}(\tau_i) = E\left[\mathbf{X}(t)\mathbf{X}^T(t+\tau_i)\right] = \mathbf{A}C_{\mathbf{S}}(\tau_i)\mathbf{A}^T \; i = 1,\ldots,I_3$$

where possibly τ_1=0. $C_{\mathbf{S}}(\tau_i)$ are diagonal due to uncorrelatedness between the sources. Lets us stuck $C_{\mathbf{X}}(\tau_1)$, $C_{\mathbf{X}}(\tau_2)$,…, $C_{\mathbf{X}}(\tau_{I3})$ in a tensor $\underline{C}_{\mathbf{X}}(\boldsymbol{\tau}) \in \mathbb{R}^{I_1 \times I_1 \times I_3}$

Simultaneous matrix diagonalization of $C_{\mathbf{X}}(\tau_i)$ is equivalent to computation of CPD of:

$$\underline{C}_{\mathbf{X}}(\boldsymbol{\tau}) = \sum_{j=1}^{J} \mathbf{a}_j \circ \mathbf{a}_j \circ \tilde{\mathbf{a}}_j$$

in which $\tilde{a}_{i_3 j} = \left(C_{\mathbf{S}}\left(\tau_{i_3}\right)\right)_{jj} \quad 1 \le i_3 \le I_3,\, 1 \le j \le J$.

The CPD tensor model

By re-writing $\underline{C}_{\mathbf{X}}(\boldsymbol{\tau})$ in a slice-wise fashion we obtain:

$$\underline{C}_{\mathbf{X}}\left(\boldsymbol{\tau}\right):,:,i_3 \approx \mathbf{A}\cdot diag\left(\tilde{a}_{i_3 1},\tilde{a}_{i_3 2},\ldots,\tilde{a}_{i_3 J}\right)\cdot \mathbf{A}^T$$

i.e. mixing matrix **A** is obtained as CPD of tensor $\underline{C}_{\mathbf{X}}\left(\boldsymbol{\tau}\right)$: $\hat{\mathbf{A}} = \mathbf{A}^{(1)}$.

Individually non-stationary and mutually uncorrelated sources. The covariance matrix $\tilde{C}_{\mathbf{X}}\left(t\right) \in \mathbb{R}^{I_1\times I_1}$ at time *t* for zero mean mixture signals is defined by:

$$\left(\tilde{C}_{\mathbf{X}}\left(t\right)\right)_{i_1 i_2} = E\left[\tilde{x}_{i_1}(t)\tilde{x}_{i_2}(t)\right]\ \forall 1\le i_1, i_2 \le I_1$$

The CPD tensor model

We can write $\tilde{C}_{\mathbf{X}}(t)$ as $\tilde{C}_{\mathbf{X}}(t) = \mathbf{A}\tilde{C}_{\mathbf{S}}(t)\mathbf{A}^T$.

Since source signals are uncorrelated $\tilde{C}_{\mathbf{S}}(t)$ is diagonal. We can stack covariance matrices $\tilde{C}_{\mathbf{X}}(t)$ for $t=t_1, t_2,\ldots,t_{I3}$ in a tensor $\tilde{C}_{\mathbf{X}}(\mathbf{t}) \in \mathbb{R}^{I_1\times I_1\times I_3}$. This tensor admits CPD:

$$\tilde{C}_{\mathbf{X}}(\mathbf{t}) = \sum_{j=1}^{J} \mathbf{a}_j \circ \mathbf{a}_j \circ \tilde{\mathbf{a}}_j$$

in which $\tilde{a}_{i_3 j} = \left(\tilde{C}_{\mathbf{S}}(t_{i_3})\right)_{jj}$ $1 \le i_3 \le I_3,\ 1 \le j \le J$. Again, we can re-write $\tilde{C}_{\mathbf{X}}(\mathbf{t})$ in a slice-wise fashion:

$$\underline{C}_{\mathbf{X}}(\mathbf{t})_{:,:,i_3} \approx \mathbf{A} \cdot diag\left(\tilde{a}_{i_3 1}, \tilde{a}_{i_3 2}, \ldots, \tilde{a}_{i_3 J}\right) \cdot \mathbf{A}^T$$

Hence, the mixing matrix **A** is obtained as CPD of tensor $\tilde{C}_{\mathbf{X}}(\mathbf{t})$: $\hat{\mathbf{A}} = \mathbf{A}^{(1)}$.

The CPD tensor model

CPD to prewhitened data X. Let us assume that data **X** are pre-whitened by **Z**=**E**^{T}**X**, where **E** is matrix of eigenvectors obtained by eigenvalue decomposition of data covariance matrix *E*[**XX**^{T}]. Then, the n-th order cumulant of **Z** is: .

$$C_{\mathbf{Z}}^{(n)} = \sum_{j=1}^{J} \frac{C_{s_j}^{(n)}}{\sigma_{s_j}^{n}} \mathbf{e}_j \circ \mathbf{e}_j \circ \ldots \circ \mathbf{e}_j \quad n \geq 3$$

This is fully symmetric orthogonality-constrained CPD of a fully symmetric n-th order cumulant tensor.

The TuckerN tensor model

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

$$\begin{aligned}\underline{\mathbf{X}} &\approx \underline{\mathbf{G}} \times_1 \mathbf{A}^{(1)} \times_2 \mathbf{A}^{(2)} \times_3 \mathbf{A}^{(3)} \\ &= \underline{\mathbf{G}} \times_1 \mathbf{T}^{(1)^{-1}} \times_2 \mathbf{T}^{(2)^{-1}} \times_3 \mathbf{T}^{(3)^{-1}} \times_1 \left(\mathbf{A}^{(1)}\mathbf{T}^{(1)}\right) \times_2 \left(\mathbf{A}^{(2)}\mathbf{T}^{(2)}\right) \times_3 \left(\mathbf{A}^{(3)}\mathbf{T}^{(3)}\right)\end{aligned}$$

where $\left\{\mathbf{T}^{(n)} \in \mathbb{R}^{J_n \times J_n}\right\}$. Hence, some constraints are necessary to be imposed on array factors and/or core tensor in order to ensure uniqueness of the factorization

Multilinear SVD

The multilinear SVD, [13], implements orthogonality constrained factorization of the Tucker model of the Nth-order tensor. For this purpose tensor is unfolded in each mode n =1,…,N:

$$\mathbf{X}_{(n)} = \mathbf{A}^{(n)}\mathbf{G}_{(n)}\left[\mathbf{A}^{(N)} \otimes \ldots \otimes \mathbf{A}^{(n+1)} \otimes \mathbf{A}^{(n-1)} \otimes \ldots \otimes \mathbf{A}^{(1)}\right]^{\mathrm{T}}$$

SVD of $\mathbf{X}_{(n)}$ is performed and $\mathbf{A}^{(n)}$ is estimated from the left singular vectors of $\mathbf{X}_{(n)}$. The core tensor is estimated from:

$$\underline{\hat{\mathbf{G}}} = \underline{\mathbf{X}} \times_1 \hat{\mathbf{A}}^{(1)^T} \times_2 \hat{\mathbf{A}}^{(2)^T} \times_3 \ldots \times_N \hat{\mathbf{A}}^{(N)^T}$$

That is known as HOSVD algorithm, [13]. Orthogonality constraints ensure virtually unique tensor decomposition. However, meaningfulness of these constraints is application dependent.

13, L. De Lathauwer, B. De Moor, and J. Vandewalle, J., "A multilinear singular value decomposition,“ *SIAM J. Matrix Anal. and Appl., vol. 21, pp. 1253-1278,* 2000.

The Higher-order Orthogonal Iteration (HOOI) algorithm

Orthogonality-constrained tensor factorization of using TuckerN tensor model can be also obtained by means of HOOI algorithm, [14]. The ALS instead of SVD is used to estimate array factors. Due to orthogonality constraints the squared Euclidean distance based cost function becomes:

$$\left[\!\left[\underline{\hat{\mathbf{G}}}, \left\{ \hat{\mathbf{A}}^{(n)} \right\}_{n=1}^{N} \right]\!\right] = \min_{\underline{\mathbf{G}}, \left\{ \mathbf{A}^{(n)} \right\}_{n=1}^{N}} \frac{1}{2} \left\| \underline{\mathbf{X}} - \underline{\hat{\mathbf{X}}} \right\|_2^2 = \min_{\underline{\mathbf{G}}, \left\{ \mathbf{A}^{(n)} \right\}_{n=1}^{N}} : \left\| \underline{\mathbf{X}} \right\|_2^2 - \left\| \underline{\mathbf{G}} \right\|_2^2$$

Hence, it suffice to perform:

$$\left\{ \hat{\mathbf{A}}^{(n)} \right\}_{n=1}^{N} = \max_{\left\{ \mathbf{A}^{(n)} \right\}_{n=1}^{N}} \left\| \underline{\mathbf{G}} \right\|_2^2 = \max_{\left\{ \mathbf{A}^{(n)} \right\}_{n=1}^{N}} \left\| \underline{\mathbf{X}} \times_1 \mathbf{A}^{(1)\mathrm{T}} \times_2 \mathbf{A}^{(2)\mathrm{T}} \dots \times_{\mathrm{N}} \mathbf{A}^{(N)\mathrm{T}} \right\|_2^2$$

The core tensor is estimated as:

$$\underline{\hat{\mathbf{G}}} = \underline{\mathbf{X}}_1 \times_1 \hat{\mathbf{A}}^{(1)\mathrm{T}} \times_2 \hat{\mathbf{A}}^{(2)\mathrm{T}} \dots \times_{\mathrm{N}} \hat{\mathbf{A}}^{(N)\mathrm{T}}$$

14. L. De Lathauwer, B. De Moor, and J. Vandewalle, “On the best rank-1 and rank-(R_1,R_2,…,R_N) approximation of Higher-order tensors,” SIAM Journal on Matrix Analysis and Applications 21, 1324 (2000).

The Higher-order Orthogonal Iteration (HOOI) algorithm

Maximization is performed in alternating least square (ALS) manner. Matlab code is available (tucker_als function) at: **http://csmr.ca.sandia.gov/~tkolda/TensorToolbox**.

The ALS procedure (Euclidean distance based cost function) does not guarantee convergence toward global minimum. To improve convergence properties of the HOOI algorithm it is wise to use the HOSVD based estimation of the initial values of the array factors instead of purely random initialization.

Tensor factorization

Cost functions. Euclidean distance (leas square error) between tensor and its model is a *standard* choice of cost function, [14]:

$$D_F\left[\underline{\mathbf{X}}\middle\|\hat{\underline{\mathbf{X}}}\right]=\frac{1}{2}\left\|\underline{\mathbf{X}}-\hat{\underline{\mathbf{X}}}\right\|_2^2$$

α-divergence based cost function is adaptable to data and noise statistics, [15,16]:

$$D_\alpha\left[\underline{\mathbf{X}}\middle\|\hat{\underline{\mathbf{X}}}\right]=\frac{1}{\alpha\left(1-\alpha\right)}\sum_{i_1,i_2,i_3}\alpha x_{i_1,i_2,i_3}+(1-\alpha)\hat{x}_{i_1,i_2,i_3}-x^{\alpha}_{i_1,i_2,i_3}\hat{x}^{(1-\alpha)}_{i_1,i_2,i_3}$$

15. A. H. Phan, and A. Cichocki, "Fast and Efficient Algorithms for Nonnegative Tucker Decomposition," *Lect. Notes Comput. Sci.*, vol. 5264, pp.772-782, 2008.
16. A. Cichocki, A.H. Phan, "Fast Local Algorithms for Large Scale Nonnegative Matrix and Tensor Factorizations," IEICE Transaction on Fundamentals, E92-A(3), 708-721.

Tensor factorization

Update rules for α-NTF algorithm:

$$\underline{\mathbf{G}} \leftarrow \underline{\mathbf{G}} \otimes \left\{ \frac{\left(\underline{\mathbf{X}} / \underline{\hat{\mathbf{X}}}\right)^{.\alpha} \times_1 \mathbf{A}^{(1)^{\mathrm{T}}} \times_2 \mathbf{A}^{(2)^{\mathrm{T}}} \times_3 \mathbf{A}^{(3)^{\mathrm{T}}}}{\underline{\mathbf{E}} \times_1 \mathbf{A}^{(1)^{\mathrm{T}}} \times_2 \mathbf{A}^{(2)^{\mathrm{T}}} \times_3 \mathbf{A}^{(3)^{\mathrm{T}}}} \right\}^{.\frac{1}{\alpha}}$$

$$\mathbf{A}^{(n)} \leftarrow \mathbf{A}^{(n)} \otimes \left\{ \frac{\left[\left(\underline{\mathbf{X}} / \underline{\hat{\mathbf{X}}}\right)^{.\alpha}\right]_{(n)} \mathbf{G}_{\mathbf{A}}^{(n)^{\mathrm{T}}}}{\mathbf{1}\mathbf{1}^{\mathrm{T}} \mathbf{G}_{\mathbf{A}}^{(n)^{\mathrm{T}}}} \right\}^{.\frac{1}{\alpha}}$$

where $\underline{\mathbf{E}}$ is a tensor whose every element is one, $\otimes$ denotes element-wise multiplication and / denotes element-wise division. **1** denotes a vector whose every element is one.

Tensor factorization

$$\left[\left(\underline{\mathbf{X}}/\hat{\underline{\mathbf{X}}}\right)^{.\alpha}\right]_{(n)}\mathbf{G}_{\mathbf{A}}^{(n)^{\mathrm{T}}}=\left[\left(\underline{\mathbf{X}}/\hat{\underline{\mathbf{X}}}\right)^{.\alpha}\times_{m\neq n}\mathbf{A}^{(m)^{\mathrm{T}}}\right]_{n}\mathbf{G}_{(n)}^{\mathrm{T}}$$

where $\mathbf{G}_{(n)}$ represents *n*-mode flattened version of the core tensor.

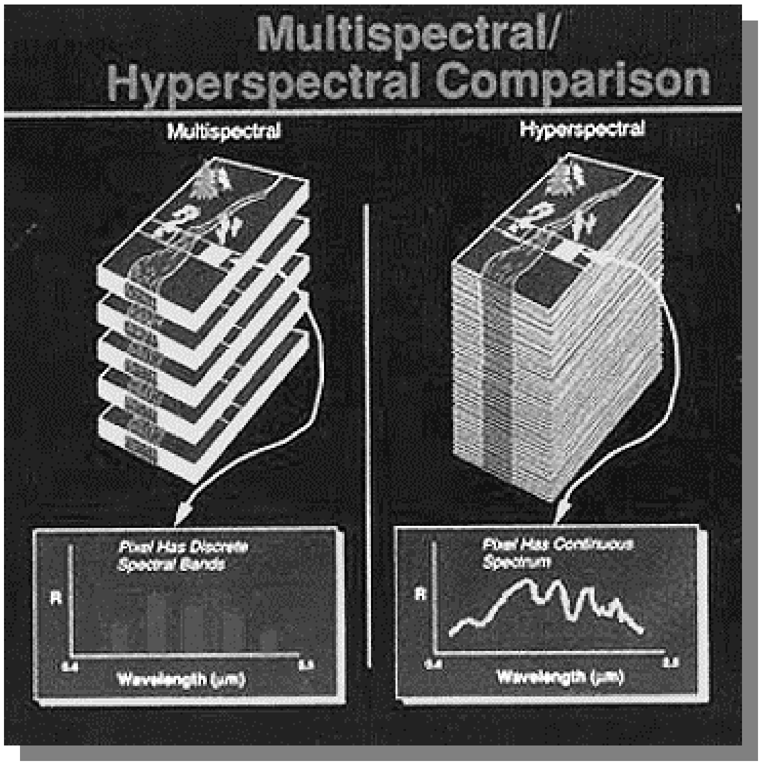
$$\mathbf{1}^{\mathrm{T}}\mathbf{G}_{\mathbf{A}}^{(n)^{\mathrm{T}}}=\left[\underline{\mathbf{G}}\times_{m\neq n}\mathbf{1}^{\mathrm{T}}\mathbf{A}^{(m)}\right]_{(n)}^{\mathrm{T}}$$

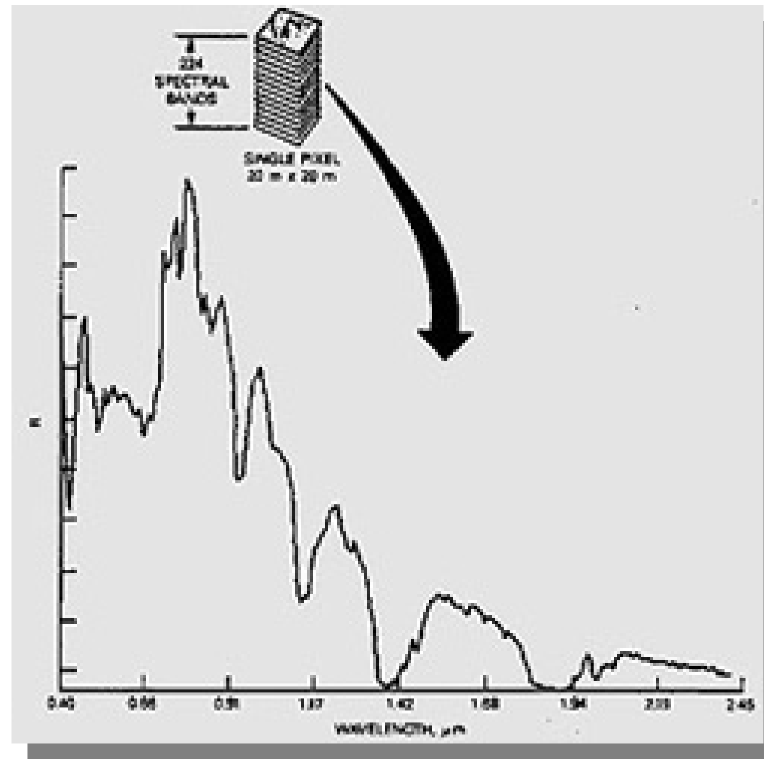
where $\underline{\mathbf{G}}\times_{m\neq n}\mathbf{1}^{\mathrm{T}}\mathbf{A}^{(m)}$ denotes *m*-mode products between core tensor and matrices $\mathbf{1}^{\mathrm{T}}\mathbf{A}^{(m)}$ for all m=1, ..., 3 and $m\neq n$.

Unsupervised Segmentation of Multispectral Images[17]

17. **I. Kopriva**, A. Cichocki, "Blind Multi-spectral Image Decomposition by 3D Nonnegative Tensor Factorization," *Optics Letters* vol. 34, No. 14, pp 2210-2212, 2009.

Unsupervised segmentation of multispectral images





- SPOT- 4 bands, LANDSAT -7 bands, AVIRIS-224 bands (0.38μ-2.4μ);

Unsupervised segmentation of multispectral images

Hyperspectral/multispectral image and static linear mixture model. For image consisting of I_3 bands and J materials linear data model is assumed:

$$\mathbf{X} = \mathbf{A}\mathbf{S} = \sum_{j=1}^{J} \mathbf{a}_j \underline{\mathbf{s}}_j$$

$$[\mathbf{a}_1 \quad \mathbf{a}_2 \quad \ldots \quad \mathbf{a}_J] \equiv \mathbf{A}$$

$$[\underline{\mathbf{s}}_1 \quad \underline{\mathbf{s}}_2 \quad \ldots \quad \underline{\mathbf{s}}_J]^T \equiv \mathbf{S}$$

X - measured data intensity matrix: $\mathbf{X} \in \mathbb{R}_{0+}^{I_3 \times I_1 I_2}$

S - unknown class matrix: $\mathbf{S} \in \mathbb{R}_{0+}^{J \times I_1 I_2}$

A – unknown spectral reflectance matrix: $\mathbf{A} \in \mathbb{R}_{0+}^{I_3 \times J}$

Unsupervised segmentation of multispectral images

True hyperspectral/multispectral image is a 3D tensor. Hence, blind image decomposition can be performed through 3D NTF, [17]:

$$\mathbf{A} \approx \mathbf{A}^{(3)}$$

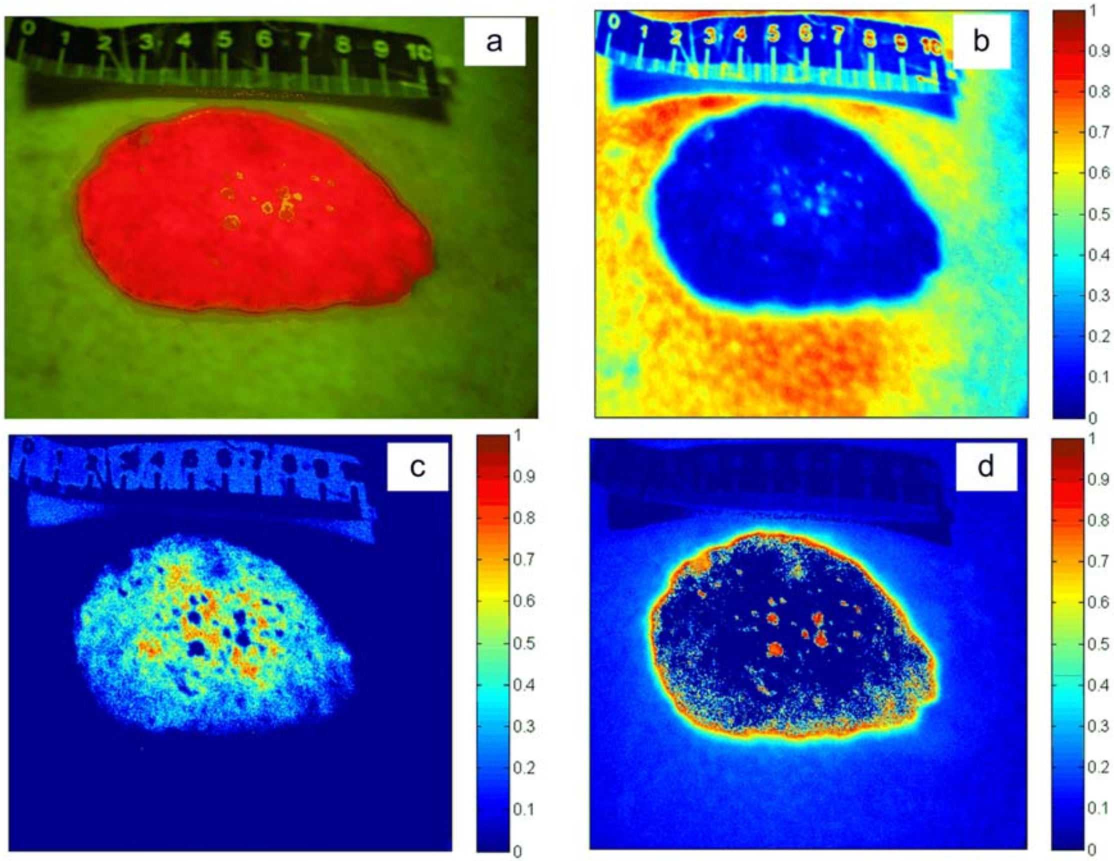
$$\underline{\mathbf{S}} \approx \underline{\mathbf{X}} \times_3 \left(\mathbf{A}^{(3)}\right)^{\dagger} \qquad \underline{\mathbf{S}} \in \mathbb{R}_{0+}^{I_1 \times I_2 \times J}$$

RGB fluorescent image (I_3=3) of the skin tumor is used to exemplify the concept. 3D α-NTF algorithm is compared against second order NMF algorithm, [18], and dependent component analysis algorithm, [19].

Note, that 3D α-NTF is based on Tucker3 model with non-negativity constraints only!!!.

18. R. Zdunek, and A. Cichocki, "Nonnegative matrix factorization with constrained second-order optimization," *Sig. Proc.* 87, 1904 (2007).
19. I. Kopriva, A. Peršin, "Unsupervised decomposition of low-intensity low-dimensional multi-spectral fluorescent images for tumour demarcation, *Medical Image Analysis* 13, 507-518, 2009.

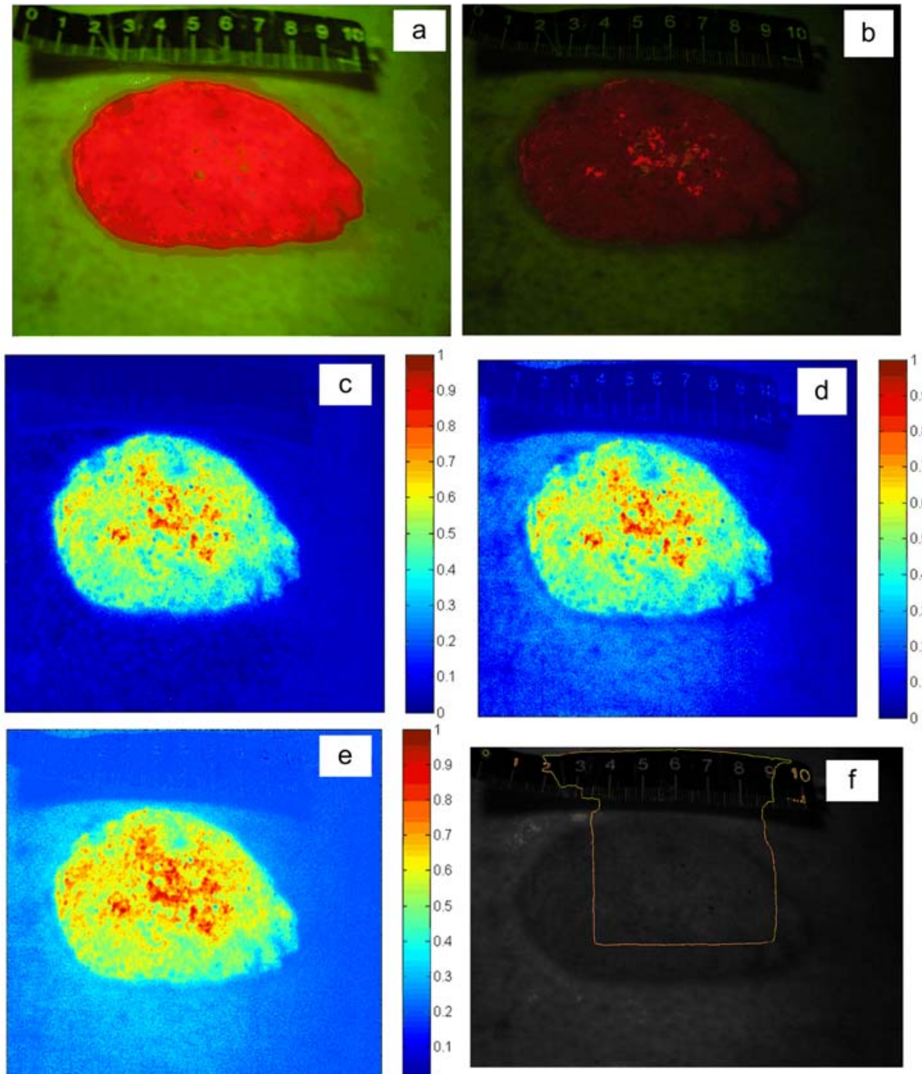
Unsupervised segmentation of multispectral images



(a) Experimental high-intensity fluorescent RGB image of the skin tumour (basal cell carcinoma). **(b)** to **(d):** Spatial maps of the objects extracted from RGB image shown in Figure 1a by means of α-NTF algorithm with α=0.1. Extracted maps of the objects were rescaled to the interval [0, 1] and shown it in pseudo colour scale, wherein dark blue colour represents 0, i.e. the absence of the object, and dark red colour represents 1, i.e. the presence of the object.

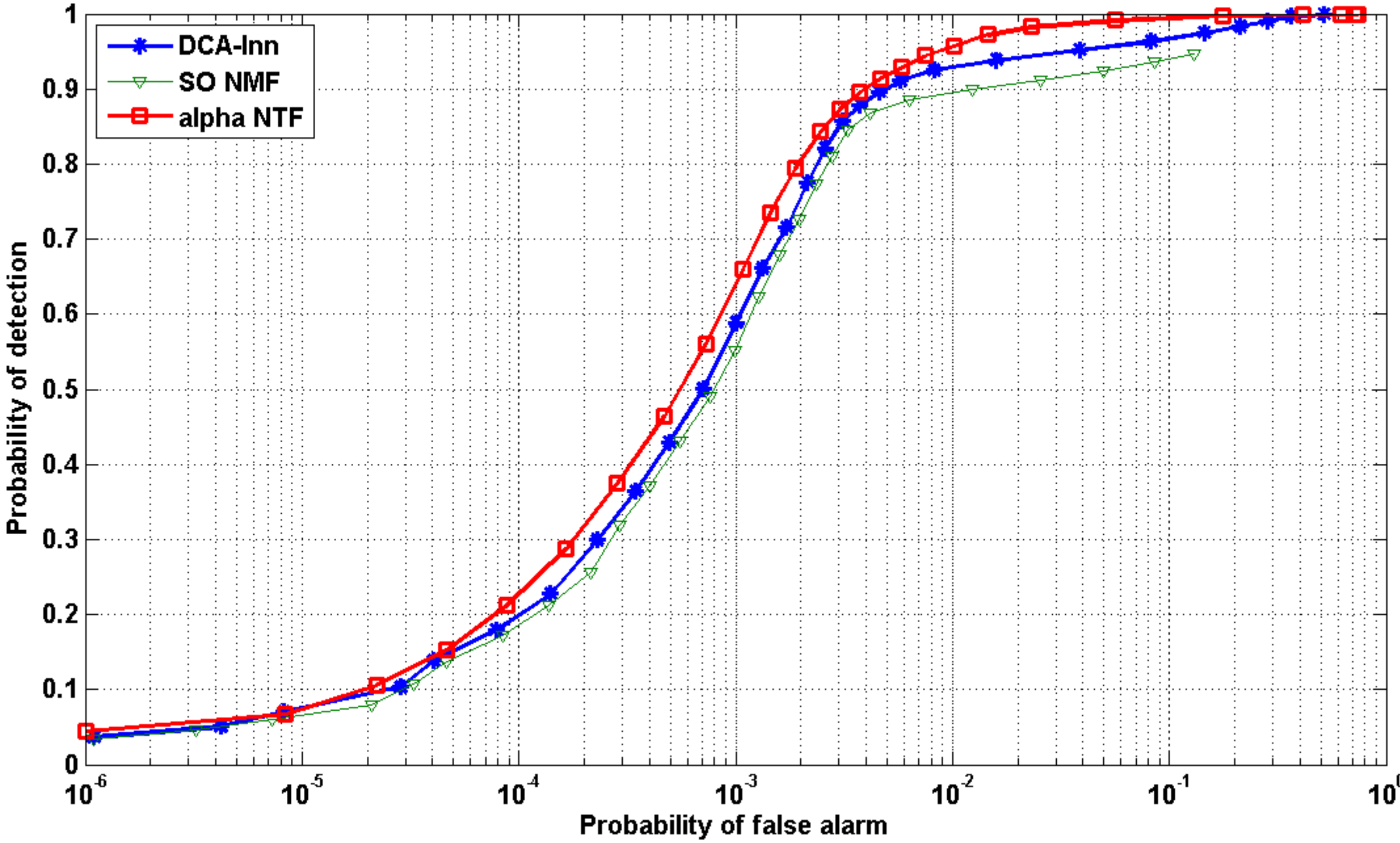
3D α-NTF yields result that is meaningful.

Unsupervised segmentation of multispectral images



Experimental fluorescent MSI RGB image of skin tumor: **a)** high-intensity version; **b)** <u>low-intensity version</u>. Spatial maps of the tumor extracted from Figure 1b by means of: **c)** α-NTF algorithm with α=0.1; **d)** SO NMF algorithm; **e)** DCA algorithm; **f)** evolution curve calculated by level set method on gray scale version of Figure 1b after 1000 iterations. Dark red color indicates that tumor is present with probability 1, while dark blue color indicates that tumor is present with probability 0.

Unsupervised segmentation of multispectral images



ROC curves calculated for spatial maps of the tumor shown in Figures 1c to 1e: red squares - α-NTF algorithm based on Tucker3 model with α=0.1; blue stars - DCA algorithm; green triangles - SO NMF algorithm.

Model-free Space-Variant Blind Image Deconvolution[20]

20. I. Kopriva, "Tensor factorization for model-free space-variant blind deconvolution of the single- and multi-frame multi-spectral image," *Optics Express*, Vol. 18, No.17, pp. 17819-17833, 2010.

• Blind image deconvolution (BID) relates to estimation of the original image $\mathbf{F} \in \mathbb{R}_{0+}^{I_1 \times I_2}$ from degraded observed image $\mathbf{G} \in \mathbb{R}_{0+}^{I_1 \times I_2}$ only assuming convolution observation model:

$$\mathbf{G}(i_1, i_2) = \sum_{s=-M}^{M} \sum_{t=-M}^{M} \mathbf{H}(s, t, i_1, i_2)\mathbf{F}(i_1 - s, i_2 - t)$$

i.e. it is assumed that convolution kernel **H**, which models degradation, is unknown.

• $\mathbf{H}(s,t,i_1,i_2)$ denotes space variant degradation which is better related to physical reality but it is used less often in BID algorithms (due to the mathematical difficulties) than space invariant degradation $\mathbf{H}(s,t)$.

An example

Original image

Blurred with circular kernel R=3, SNR=40dB;

Non-blind RL algorithm (`deconvlucy`)

PSF misspecified

True PSF

• Many BID algorithms assume that source of degradation is _**known**_ and **H** can be represented by appropriate parametric model.

•For an example 2D Gaussian with unknown variance and size of support is used to model turbulence blur. Circular kernel with unknown radius is used to model de-focus blur. A line with an unknown length is used to model blur cause by vibrations or relative motion between image and object planes.

•Then, the BID problem becomes parametric estimation problem that can be solved by expectation-maximization method. An example is blind Richardson-Lucy algorithm that is implemented in MATLAB function `deconvblind`.

•Herein, the **model-free** BID method for possibly **space-variant** degradation will be presented. This will be achieved by converting BID problem into BSS problem and solved by the higher order orthogonal iteration (HOOI) tensor factorization algorithm.

Model-free blind image deconvolution

• In recent work, [20-23], it has been demonstrated the possibility to formulate single frame blind image deconvolution as a multichannel BSS problem solved by NMF[23], DCA[22] and very recently by TF[20,21] algorithms.

•Benefit: no *a priori* knowledge about model (origin) of the degradation kernel is required.

$$\mathbf{G}(i_1, i_2) = \sum_{s=-M}^{M} \sum_{t=-M}^{M} \mathbf{H}(s,t)\mathbf{F}(i_1 - s, i_2 - t)$$

21. I. Kopriva, "3D Tensor Factorization Approach to Single-frame Model-free Blind Image Deconvolution," *Optics Letters*, Vol. 34, No.14, pp. 2210-2212, 2009.
22. I. Kopriva, "Approach to Blind Image Deconvolution by Multiscale Subband Decomposition and Independent Component Analysis," *Journal Optical Society of America A*, Vol. 24, No.4, pp. 973-983, 2007.
23. I. Kopriva, "Single Frame Multichannel Blind Deconvolution by Non-negative Matrix Factorization with Sparseness Constraint," *Optics Letters*, Vol. 30, No. 23, pp. 3135-3137, 2005.

Model-free blind image deconvolution

- Model-free BID is achieved by converting the BID problem into the BSS problem. This is achieved by using implicit Taylor series expansion of the $\mathbf{F}(i_1-\mathbf{s},i_2-t)$ around the origin (i_1,i_2) and (when necessary) using 2D Gabor filter bank to generate multi-channel representation.

- To make the presentation easier to follow this will be demonstrated on a case of space-invariant deconvolution of blurred single-frame gray scale image.

- The more complex scenarios that will be demonstrated herein include: single-frame gray scale image blurred by a space-variant blur, single-frame multi-spectral image blurred by space-variant and space-invariant blurs, and multi-frame multi-spectral image blurred by space-variant and space-invariant blurs.

Model-free blind image deconvolution

The single-frame model-free multi-channel deconvolution was first proposed by Umeyama, [24].The key insight is the Taylor series expansion of $\mathbf{F}(i_1\text{-}s, i_2\text{-}t)$ around $\mathbf{F}(i_1, i_2)$.

$$\mathbf{F}(i_1-s, i_2-t) = \mathbf{F}(i_1, i_2) + s\mathbf{F}_{i_1}(i_1, i_2) + t\mathbf{F}_{i_2}(i_1, i_2) + \ldots$$

Then the degraded image is obtained as

$$\mathbf{G}(i_1, i_2) = a_1\mathbf{F}(i_1, i_2) + a_2\mathbf{F}_{i_1}(i_1, i_2) + a_3\mathbf{F}_{i_2}(i_1, i_2) + \ldots$$

where

$$a_1 = \sum_{s=-K}^{K}\sum_{t=-K}^{K}\mathbf{H}(s,t)\; a_2 = \sum_{s=-K}^{K}\sum_{t=-K}^{K}s\mathbf{H}(s,t)\; a_3 = \sum_{s=-K}^{K}\sum_{t=-K}^{K}t\mathbf{H}(s,t)$$

The PSF coefficients are absorbed into mixing coefficients. <u>No *a priori* knowledge about the nature of the blurring process or size of the blurring kernel is required.</u>

24. S. Umeyama, Scripta Technica, Electron Comm Jpn, Pt 3, 84(12), 1-9 (2001).

Model-free blind image deconvolution

- A multi-channel representation is obtained by applying a bank of 2-D Gabor filters to degraded image **G**(i_1,i_2). A single-frame multi-channel image model **G**(i_1,i_2,i_3) is obtained as :

$$\underline{\mathbf{G}} = \begin{bmatrix} \mathbf{G}(:,:,1) \\ \mathbf{G}(:,:,2) \\ \cdots \\ \mathbf{G}(:,:,I_3) \end{bmatrix} \cong \begin{bmatrix} a_1 \;\; a_2 \;\; a_3 \cdots \\ a_{11}\, a_{12}\, a_{13} \cdots \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ a_{(I_3-1)1}\, a_{(I_3-1)2}\, a_{(I_3-1)3} \cdots \end{bmatrix} \begin{bmatrix} \mathbf{F}(:,:) \\ \mathbf{F}_{i_1}(:,:) \\ \mathbf{F}_{i_2}(:,:) \\ \cdots\cdot \end{bmatrix} = \underline{\mathbf{R}} \times_1 \mathbf{A}^{(1)} \times_2 \mathbf{A}^{(2)} \times_3 \mathbf{A}^{(3)} = \underline{\mathbf{F}} \times_3 \mathbf{A}^{(3)}$$

Above model suggests that tensor $\underline{\mathbf{F}}$ composed of original image and its spatial derivatives can be estimated from $\underline{\mathbf{F}} \cong \underline{\mathbf{G}} \times_3 \left(\hat{\mathbf{A}}^{(3)}\right)^{\dagger}$ where $\hat{\mathbf{A}}^{(3)}$ is obtained by the HOOI-based decomposition of $\underline{\mathbf{G}}$.

Sources represent original image **F** and its spatial derivatives. These sources are **<u>neiter statistically independent nor sparse</u>**. Hence ICA and NMF based decomposition of related multi-channel image are expected to be less accurate in performing BID.

Model-free blind image deconvolution

- Tensor representation is easily extendable to multi-spectral single-frame image. Multi-channel image model **G**(i_1,i_2,i_3 ,i_4) is then obtained as:

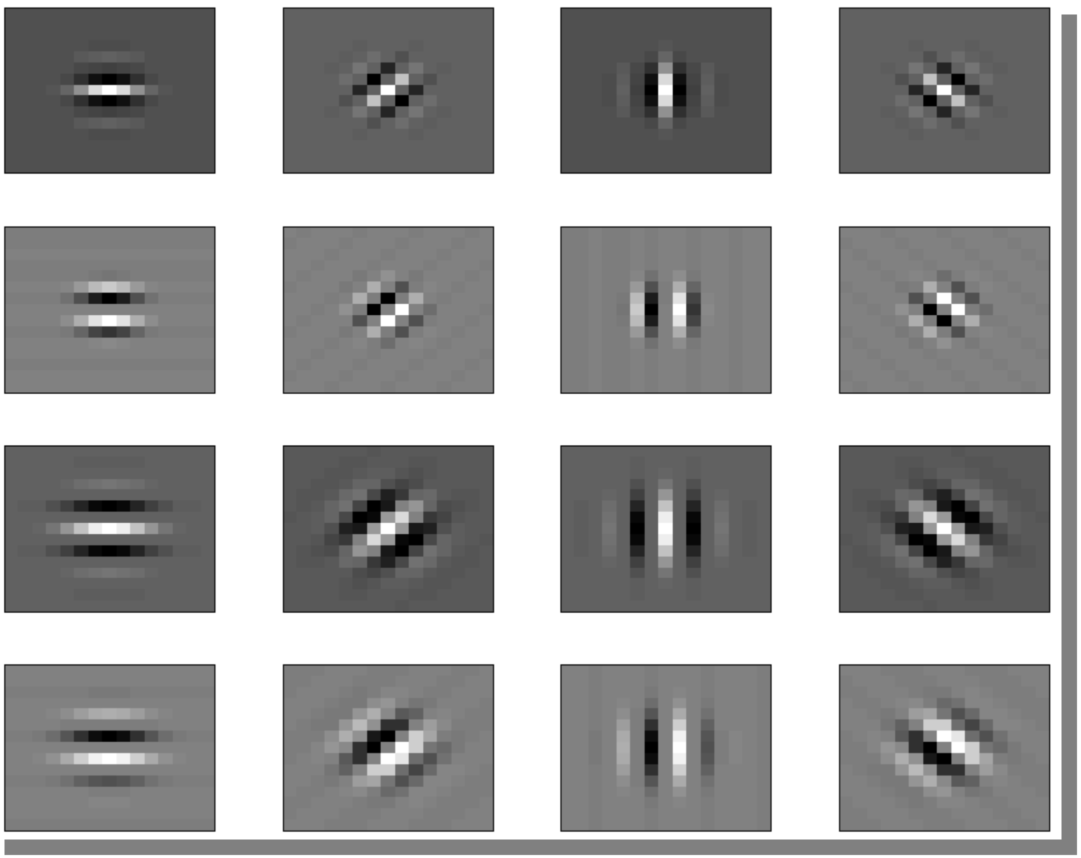
$$\underline{\mathbf{G}}(:,:,i_3,:)=\begin{bmatrix}\mathbf{G}(:,:,i_3,1)\\ \mathbf{G}(:,:,i_3,2)\\ \cdots\\ \mathbf{G}(:,:,i_3,I_4)\end{bmatrix}\cong\begin{bmatrix}a_{(i_3)1} & a_{(i_3)2} & a_{(i_3)3}\cdots\\ a_{(i_3)11} & a_{(i_3)12} & a_{(i_3)13}\cdots\\ \cdots\cdots\cdots\cdots\cdots\cdots & & \\ a_{(i_3)(I_4-1)1} & a_{(i_3)(I_4-1)2} & a_{(i_3)(I_4-1)3}\cdots\end{bmatrix}\begin{bmatrix}\mathbf{F}(:,:,:)\\ \mathbf{F}_{i_1}(:,:,:)\\ \mathbf{F}_{i_2}(:,:,:)\\ \cdots\end{bmatrix}$$

$$\underline{\mathbf{G}}=\underline{\mathbf{R}}\times_1\mathbf{A}^{(1)}\times_2\mathbf{A}^{(2)}\times_3\mathbf{A}^{(3)}\times_4\mathbf{A}^{(4)}=\underline{\mathbf{F}}\times_4\mathbf{A}^{(4)}$$

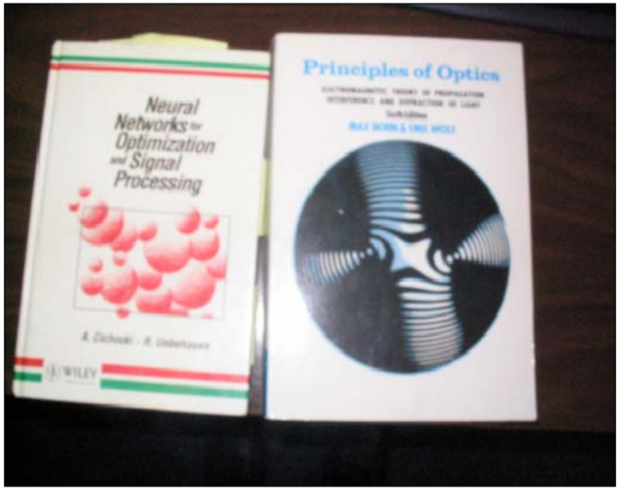
Tensor composed of original multi-spectral image and its spatial derivatives can be estimated from $\underline{\mathbf{F}}\cong\underline{\mathbf{G}}\times_4\left(\hat{\mathbf{A}}^{(4)}\right)^{\dagger}$ where $\hat{\mathbf{A}}^{(4)}$ is obtained by the HOOI-based decomposition of $\underline{\mathbf{G}}$.

Model-free blind image deconvolution

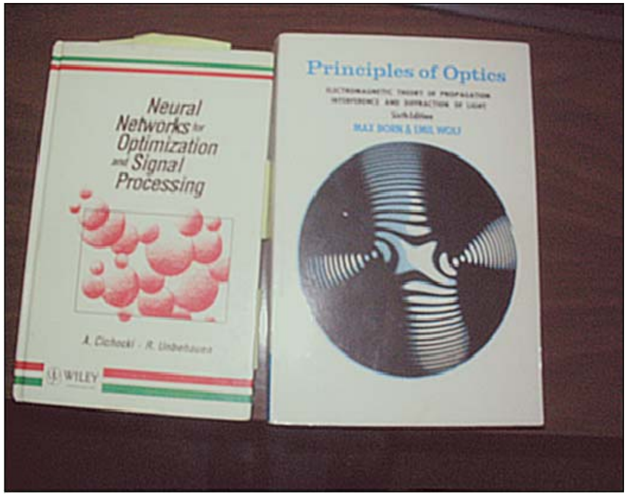
A Gabor filter bank of 7x7 filters with two spatial frequencies and four orientations.



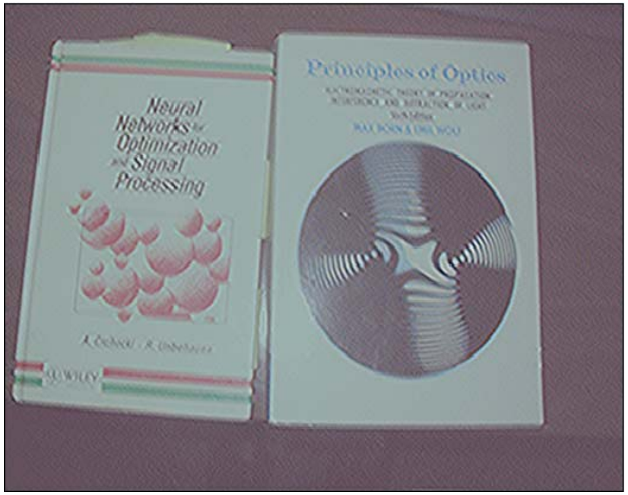
Single-frame multi-channel blind image deconvolution



Defocused image



Three 3D TF deconvolutions[21]

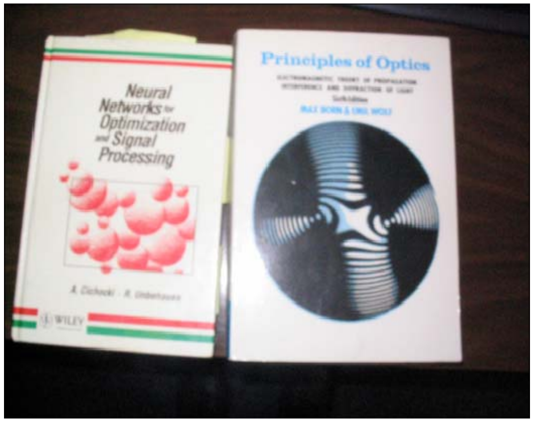


DCA deconvolution[22]

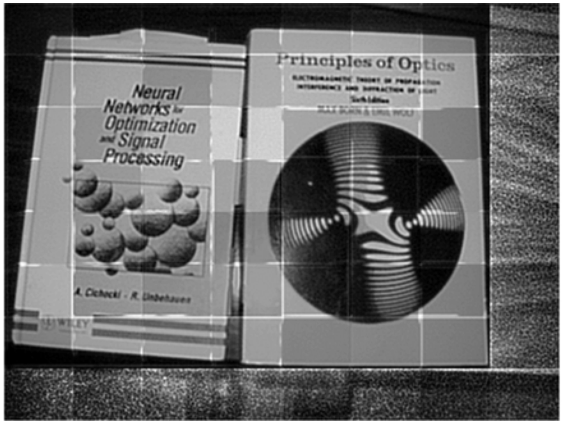
21. I. Kopriva, "3D Tensor Factorization Approach to Single-frame Model-free Blind Image Deconvolution," *Optics Letters*, Vol. 34, No.14, pp. 2210-2212, 2009.
22. I. Kopriva, "Approach to Blind Image Deconvolution by Multiscale Subband Decomposition and Independent Component Analysis," *Journal Optical Society of America A*, Vol. 24, No.4, pp. 973-983, 2007.

Space-variant blind deconvolution of de-focused single-frame gray scale image

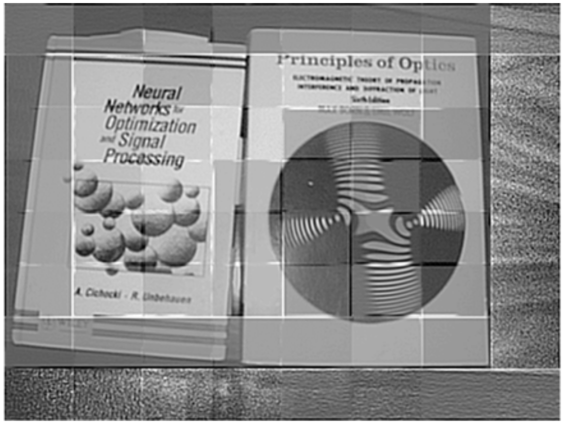
- Gray scale version of de-focused image is divided into 64 blocks of size 48x64 pixels. PSF is assumed to be constant within each block. Each block is filtered by a bank of 2D Gabor filters. 4D image tensor if characterized with: I_1=48, I_2=64, I_3=64 and I_4=17.



Defocused image



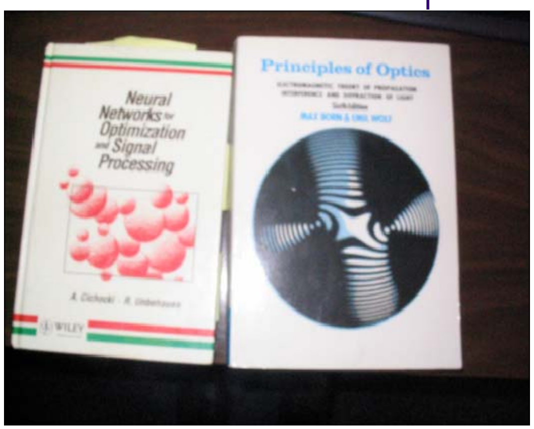
One 4D TF



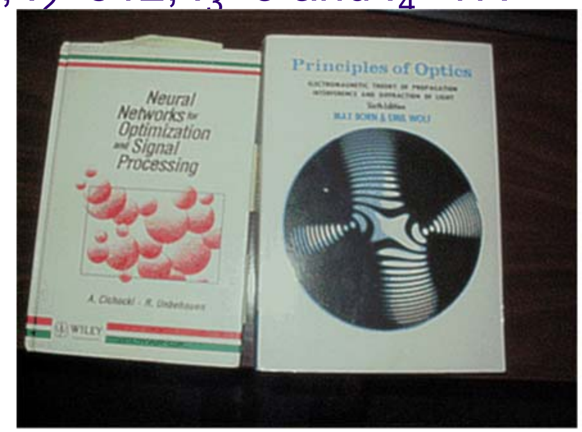
Sixty four 3D TF

Space-invariant blind deconvolution of de-focused single-frame multi-spectral image

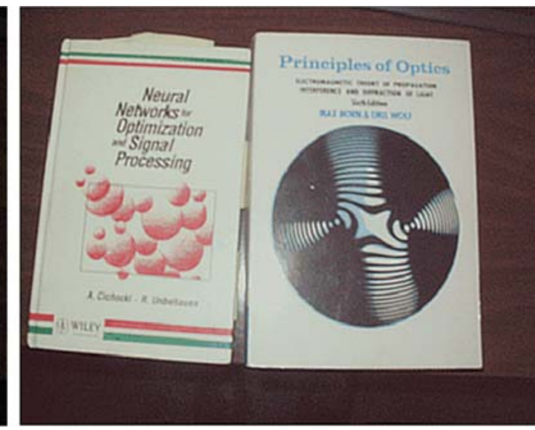
- RGB de-focused image is of the size 384x512 pixels. PSF is assumed to be constant within the image. Each spectral image is filtered by a bank of 2D Gabor filters. 4D image tensor if characterized with: I_1=384, I_2=512, I_3=3 and I_4=17.



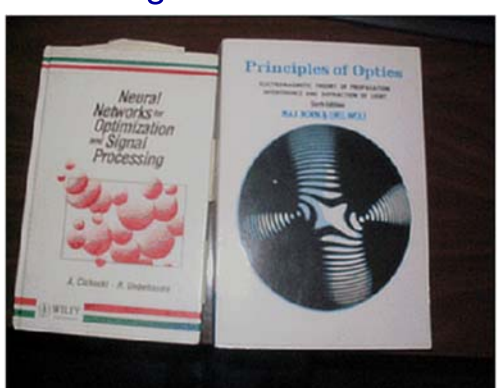
Defocused image

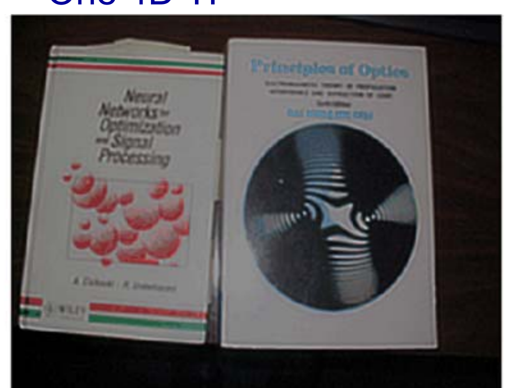


One 4D TF



Three 3D TF

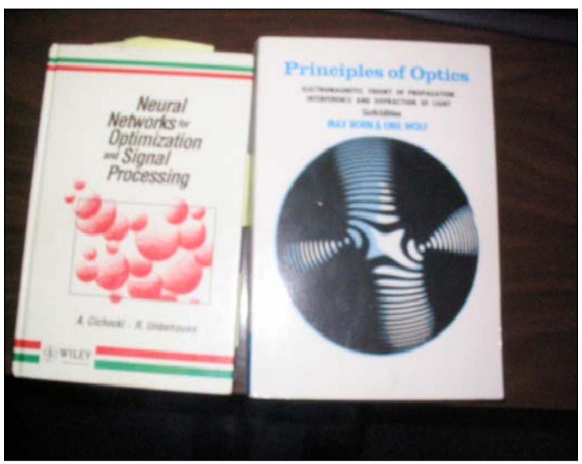




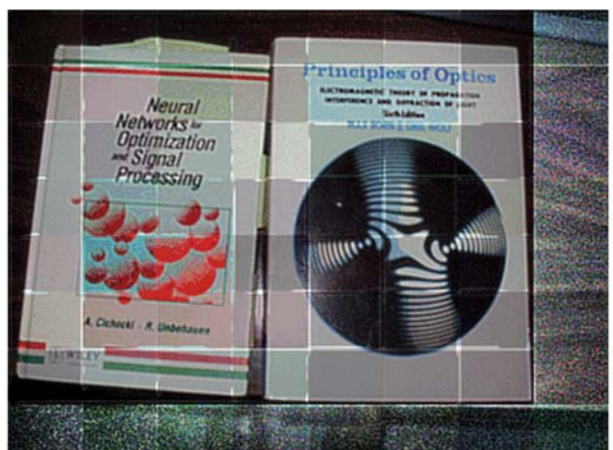
Blind R-L. R=2 pixels (left), 3 pixels (right).

Space-variant blind deconvolution of de-focused single-frame multi-spectral image

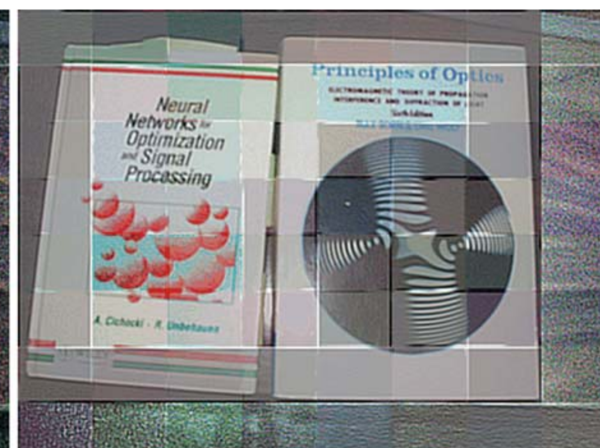
- RGB de-focused image is divided into 64 blocks of the size 48x64 pixels. PSF is assumed to be constant within the block. Each spectral image of each block is filtered by a bank of 2D Gabor filters. 5D image tensor is characterized with: I_1=48, I_2=64, I_3=64, I_4=3 and I_5=17.



Defocused image



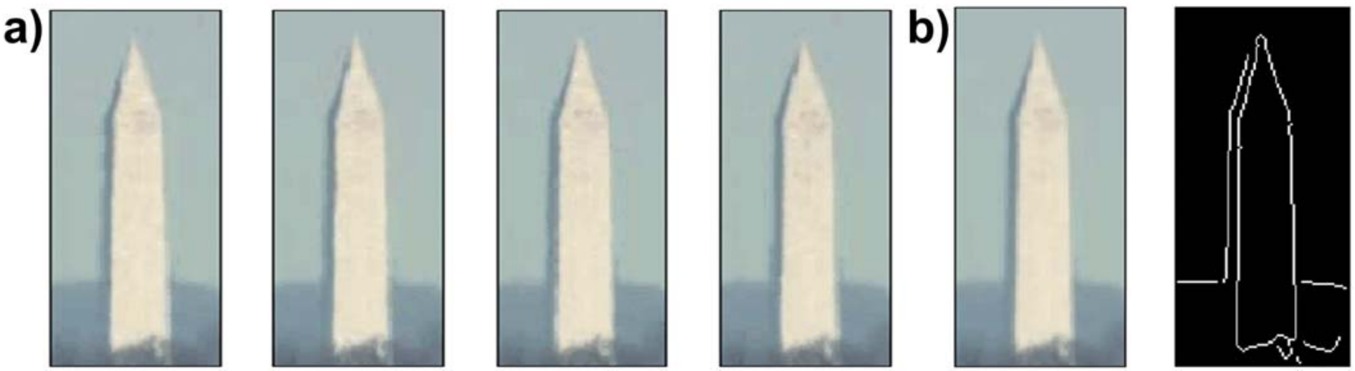
One 5D TF



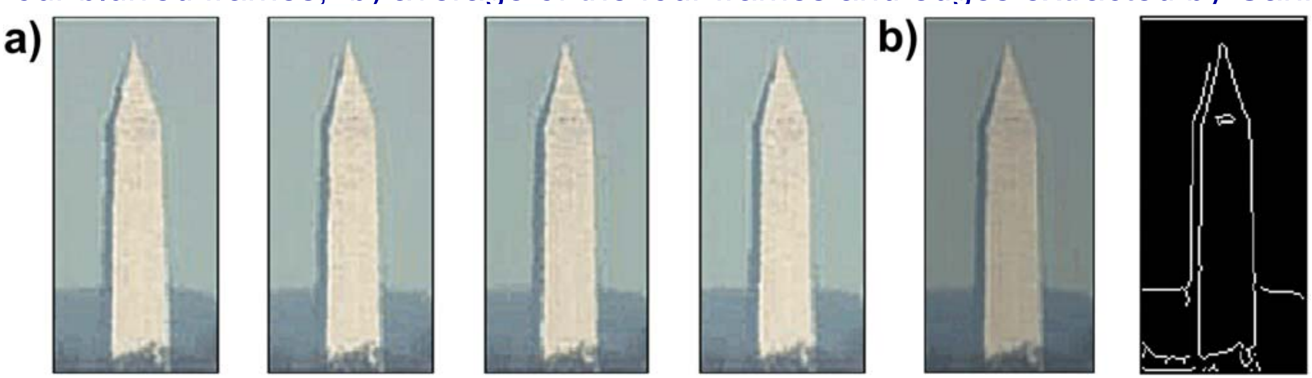
Sixty four 4D TF

Space-invariant blind deconvolution of a multi-frame multi-spectral image blurred by atmospheric turbulence

- Multi-frame image of the Washington monument has been used. Four frames were chosen randomly. PSF is assumed to be constant within the image. Each spectral image of each frame is filtered by a bank of 2D Gabor filters. 5D image tensor is characterized with: I_1=160, I_2=80, I_3=3, I_4=4 and I_5=17.



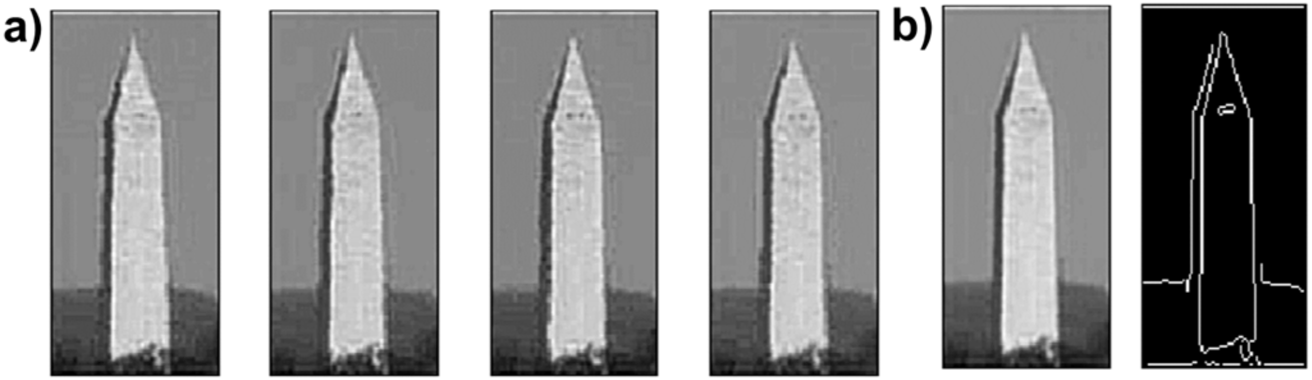
a) Four blurred frames; b) average of the four frames and edges extracted by Canny’s method



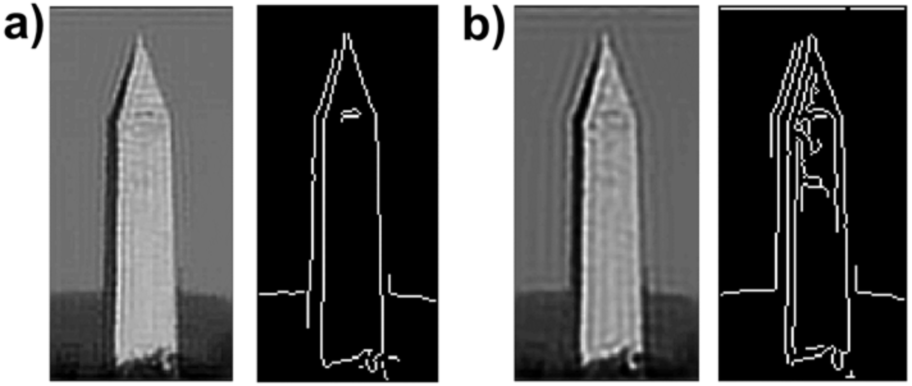
a) Time evolution of the source frame extracted by 5D TF; b) average of the four source frame

Space-invariant blind deconvolution of a multi-frame gray scale image blurred by atmospheric turbulence

- 4D image tensor is characterized with: I_1=160, I_2=80, I_3=4 and I_4=17.



a) Time evolution of the source frame extracted by 4D TF; b) average of the four source frame



Blind R-L with 2D Gaussian PSF model. Kernel width 18 pixels. a) σ^2=1.3 pixels; b) σ^2=1.9 pixels

Space-(in)variant blind deconvolution of a single-frame multi-spectral image blurred by a grating

- RGB image is blurred by a grating (photon sieve). The true image is composed of a palette of color pens, and a painting on the white board. Since there is no "real life" physical analogy to this grating-caused blur, it is virtually impossible to select a specialized method to perform blind deconvolution of this grating-blurred image. For space-variant deconvolution image is divided into 6 blocks of size 150x117 pixels. Each spectral image of each block is filtered by a bank of 2D Gabor filters. 5D image tensor is characterized with: I_1=150, I_2=117, I_3=6, I_4=3 and I_5=17. Space-invariant problem is reduced to 4D TF characterized with: I_1=301, I_2=351, I_3=3, and I_4=17.

From left to right: blurred image; 5D TF space-variant BID; 4D TF space-invariant BID; Blind R-L with circular PSF, R=2 pixels

Conclusion and open issues

•The HOOI-based tensor factorization approach has been proposed for the space-(in)variant model-free blind deconvolution of a single- and multi-frame multi-spectral image. This is achieved by converting blind deconvolution into blind source separation using implicit Taylor expansion of the original image in the convolution image-forming equation.

• Two major contributions of the proposed approach to blind image deconvolution are: (*i*) as opposed to matrix factorization methods the HOOI-based factorization of the tensor of the blurred image is virtually unique with no hard constraints imposed on source images; (*ii*) neither model nor size of the support of the point spread function is required to be *a priori* known or estimated.

Conclusion and open issues

•Use of the implicit Taylor expansion implies certain level of smoothness of the original image. This might limit the performance of the proposed approach to blind image deconvolution when the blurring process is strong or the original image contains sharp boundaries.

•Nevertheless, the proposed method is expected to be useful in scenarios when *a priori* information required by physically constrained iterative blind deconvolution methods are difficult or impossible to obtained.

•The two fundamental issues considered to be important exploring in the future work are: optimal selection of the size of the image blocks and neutralization of block-wise induced artifacts associated with space-variant deconvolution; looking for possible replacement of 2D Gabor filter bank approach to single-channel blind source separation.

FEATURE EXTRACTION FOR CANCER PREDICTION BY TENSOR DECOMPOSITION OF 1D PROTEIN EXPRESSION LEVELS[23]

23. I. Kopriva, A. Jukić, A. Cichocki (2011). Feature extraction for cancer prediction by tensor decomposition of 1D protein expression levels, accepted for 2nd IASTED Computational Bioscience Conference, G. Montana (ed.), pp. 277-283, Cambridge, UK, July 11-13.

Tensor factorization for feature extraction

•Feature extraction and selection are essential problems in the analysis of datasets with large number of variables. Typical areas where considered problems arise include text mining, combinatorial chemistry, proteomics, genomics, computational biology, etc.

•Extraction of suitable features is considered to have a major effect on classification/prediction performance. It is of paramount importance in problems characterized by small number of samples and large number of features (variables). Such situations are common in contemporary proteomics and genomics, where each sample (mass spectra in proteomics and gene expression levels in genomics) represents patient (healthy or disease) characterized by up to tens of thousands of variables (*m/z* ratios in proteomics and genes in genomics).

• The class prediction problem is too ill-posed and feature (dimensionality) reduction is necessary to prevent classifier to be tuned on training data and perform poorly on unseen (test) data (overfitting).

Tensor factorization for feature extraction

• Tensor factorization approach using the HOOI algorithm has been applied in [23] to reduce dimensionality (extract features) of the low resolution surface-enhanced laser desorption ionization time-of-flight (SELDI-TOF) mass spectra of a blood serum representing control group (healthy patients) and case group (patients with ovarian and prostate cancers: [24]).

• Used datasets are well known and were used previously by other researches to test accuracy of classifiers and feature extraction methods in cancer prediction.

• Prostate cancer dataset comprises 69 disease and 63 control samples. Ovarian cancer dataset comprises 100 disease and 100 control samples. In both datasets sample is represented by 15154 features representing intensity level for *m*/*z* ratios.

24. Center for Cancer Research, National Cancer Institute Program in Clinical Proteomics.
Available: http://home.ccr.cancer.gov/ncifdaproteomics/ppatterns.asp

Tensor factorization for feature extraction

• Each sample is a vector $\mathbf{x}^{(k)} \in \mathbb{R}^{I_1 I_2} \quad \forall k = 1,\ldots,K$. For feature extraction it is transformed in a matrix $\mathbf{X}^{(k)} \in \mathbb{R}^{I_1 \times I_2} \quad \forall k = 1,\ldots,K$

$\mathbf{x}^{(k)}$

1	2	3	.	.	.	.	.	34	35	36

$\mathbf{X}^{(k)}$

1	2	3	4	5	6
12	11	10	9	8	7
13	14	15	16	17	18
24	23	22	21	20	19
25	26	27	28	29	30
36	35	34	33	32	31

Example of transformation of vector to matrix

Tensor factorization for feature extraction

- Each sample matrix is represented by a trilinear Tucker2 model:

$$\mathbf{X}^{(k)} \approx \mathbf{A}^{(1)}\mathbf{F}^{(k)}\mathbf{A}^{(2)T} = \mathbf{F}^{(k)} \times_1 \mathbf{A}^{(1)} \times_2 \mathbf{A}^{(2)} \qquad \forall k = 1,\ldots,K$$

where the core tensor (matrix) is sample dependent and basis matrices (factors) are sample invariant (fixed). We can concatenate sample matrices in a 3D tensor $\underline{\mathbf{X}} \in \mathbb{R}^{I_1 \times I_2 \times K}$ that is described by the following Tucker2 model:

$$\underline{\mathbf{X}} \approx \underline{\mathbf{F}} \times_1 \mathbf{A}^{(1)} \times_2 \mathbf{A}^{(2)}$$

i.e. the core tensor $\underline{\mathbf{F}} \in \mathbb{R}^{J \times J \times K}$ is composed of *K* feature matrices, where the *k*th frontal slice matches $\mathbf{F}^{(k)}$. After HOOI decomposition of tensor $\underline{\mathbf{X}}$ to its Tucker2 model, [25], we obtain orthogonal factor matrices $\mathbf{A}^{(1)}$ and $\mathbf{A}^{(2)}$.

25. C. A. Andersson and R. Bro, The N-way Toolbox for MATLAB, Chemometrics & Intelligent Laboratory Systems, 52 (1), 2000, 1-4. http://www.models.life.ku.dk/source/nwaytoolbox/

Tensor factorization for feature extraction

Approximation of the core tensor is obtained from:

$$\underline{\hat{\mathbf{F}}} = \underline{\mathbf{X}} \times_1 \mathbf{A}^{(1)\mathrm{T}} \times_2 \mathbf{A}^{(2)\mathrm{T}}$$

Extracted features $\left\{ \hat{\mathbf{f}}^{(k)} \in \mathbb{R}^{J^2} \right\}_{k=1}^{K}$ for each of the *K* training data samples are obtained by vectorization of the frontal slices $\hat{\mathbf{F}}_k = \hat{\mathbf{F}}^{(k)}$. That yields set $S = \left\{ \left(\hat{\mathbf{f}}^{(1)}, c_1 \right), \ldots, \left(\hat{\mathbf{f}}^{(K)}, c_K \right) \right\}$ containing extracted features paired with class labels (c_k=1 for disease, -1 for healthy) for *K* training samples.

Features for unseen (test) samples are obtained as:

$$\underline{\hat{\mathbf{F}}}_{test} = \underline{\mathbf{X}}_{test} \times_1 \mathbf{A}^{(1)\mathrm{T}} \times_2 \mathbf{A}^{(2)\mathrm{T}}$$

	J^2	kNN	linSVM	polySVM	rbfSVM
10-fold	36	88.8±11.3 / 88.3±11.7 k=4	95.1±8.8 / 94.7±8.2	97.5±6.0 / 89.9±11.8 d=3	96.1±7.2 / 91.3±10.9 σ=3.2
	100	94.6±8.3 / 88.7±11.6 k=3	99.3±3.5 / 96.9±7.2	99.6±2.7 / 93.1±9.3 d=3	98.3±5.1 / 93.1±9.9 σ=6.8
	256	96.4±6.7 / 92.2±10.5 k=4	99.9±1.0 / 97.7±5.6	99.7±2.1 / 93.7±9.9 d=3	99.8±1.7 / 93.0±10.0 σ=13
	400	95.4±8.1 / 93.7±9.2 k=3	100±0 / 100±0	99.6±2.5 / 94.0±10.1 d=2	99.8±2.3 / 93.2±9.7 σ=15.6
	625	96.0±7.2 / 94.6±8.9 k=3	100±0 / 100±0	100±0 / 94.4±9.6 d=2	100±0 / 93.8±9.4 σ=19.8
2-fold	36	85.4±7.1 / 85.1±5.6 k=4	91.9±5.6 / 91.4±4.8	94.4±4.7 / 89.0±5.9 d=3	93.2±5.1 / 89.7±5.1 σ=3.2
	100	89.3±5.4 / 86.7±5.3 k=2	98.2±2.8 / 95.6±3.6	98.3±2.7 / 89.2±5.5 d=3	94.6±5.2 / 90.3±4.7 σ=6.6
	256	90.2±5.4 / 89.7±5.1 k=4	98.8±2.2 / 96.5±3.4	97.0±3.3 / 90.0±5.4. d=2	96.2±3.9 / 91.0±4.9 σ=12.6
	400	92.6±4.6 / 90.1±4.2 k=2	99.3±1.6 / 97.8±3.3	98.2±2.7 / 90.7±5.9 d=2	96.4±4.3 / 91.5±4.8 σ=13.2
	625	93.0±4.9 / 91.8±4.6 k=2	99.6±1.2 / 98.7±2.9	98.9±2.1 / 91.4±5.1 d=2	98.0±3.2 / 93.2±4.5 σ=19

Prostate cancer. Sensitivity and specificity in % (means +/- standard deviations). 200 random partitions.

Prostate cancer prediction results reported in the literature

Petricoin *et al.*[15]	sensitivity: 94.7%; specificity: 75.9%; 253 benign and 69 cancers. CV details not reported
Xu *et al.* [16]	sensitivity: 97.1%; specificity: 96.8%; 253 benign and 69 cancers. CV details not reported
Yang *et al.* [17]	Average error rate of 28.97 on four class problem with 3-fold CV.

	J^2	kNN	linSVM	polySVM	rbfSVM
10-fold	36	72.4±13.6 / 63.9±15.1 k=4	88.1±9.9 / 84.6±10.4	86.1±10.7 / 85.8±12.1 d=3	88.4±9.3 / 81.1±11.7 σ=4.2
	100	87.0±10.4 / 65.0±14.1 k=6	92.8±8.4 / 89.4±9.2	93.1±8.5 / 88.6±10.2 d=3	93.9±7.7 / 87.9±10.2 σ=6
	256	87.4±10.8 / 69.8±14.2 k=4	95.3±6.4 / 95.4±7.4	94.5±7.0 / 90.0±11.1 d=3	95.2±6.6 / 91.1±9.3 σ=12.2
	400	90.2±10.2 / 71.0±14.7 k=6	96.8±5.5 / 95.0±6.4	94.8±7.4 / 91.5±9.3 d=2	95.9±6.4 / 92.5±8.8 σ=14.4
	625	93.8±7.5 / 71.0±14.2 k=8	97.9±4.9 / 97.9±4.8	96.5±5.9 / 95.2±6.6 d=2	96.3±6.1 / 94.5±7.5 σ=19.4
2-fold	36	66.7±8.4 / 64.7±7.2 k=12	84.3±6.1 / 81.4±5.9	83.4±6.5 / 79.8±7.2 d=3	84.6±5.6 / 77.3±6.7 σ=4.4
	100	77.0±7.0 / 64.6±6.7 k=4	91.1±4.6 / 87.7±5.0	90.7±4.5 / 82.7±6.4 d=3	92.1±4.2 / 84.5±5.4 σ=8.2
	256	82.7±6.2 / 67.7±6.5 k=4	93.6±3.7 / 91.9±4.1	93.6±3.5 / 85.1±5.8 d=3	93.9±3.0 / 87.7±4.8 σ=14.2
	400	85.4±5.7 / 67.2±7.3 k=4	95.5±3.2 / 93.4±3.6	93.4±3.6 / 85.4±5.5 d=2	95.1±2.7 / 90.1±4.4 σ=17.4
	625	87.8±5.3 / 68.7±6.2 k=4	96.8±2.9 / 95.4±3.5	94.2±3.4 / 87.1±5.7 d=2	95.0±2.7 / 91.9±4.1 σ=18.6

Ovarian cancer. Sensitivity and specificity in % (means +/- standard deviations). 200 random partitions.

Petricoin *et al.* [18]	sensitivity: 100%; specificity: 95% (one partition only: 50/50 training; 66/50 test).
Assareh *et al.* [19]	accuracy averaged over ten 10-fold partitions: 98-99% (SD: 0.3-0.8)
Li *et al.* [20]	sensitivity: 98%; specificity: 95% (2-fold CV with 100 partitions)
Qiu *et al.* [21]	97.63% accuracy (other details not specified)
Yang *et al.* [17]	average error rate of 4.1 % with 3-fold CV.