

Nonlinear sparse component analysis: pure components extraction and multichannel image decomposition

Ivica Kopriva

Ruđer Bošković Institute

e-mail: ikopriva@irb.hr ikopriva@gmail.com

Web: <http://www.lair.irb.hr/ikopriva/>

September 18, 2015

Acknowledgment:



Croatian Science Foundation Grant 9.01/232 "Nonlinear component analysis with applications in chemometrics and pathology".

Rudjer Boskovich Institute, Zagreb, Croatia

<https://www.irb.hr/eng>

The Ruđer Bošković Institute is regarded as Croatia's leading scientific institute in the natural and biomedical sciences as well as marine and environmental research, owing to its size, scientific productivity, international reputation in research, and the quality of its scientific personnel and research facilities.

The Institute is the leading and internationally most competitive Croatian institute by virtue of its participation in international research projects, such as the IAEA and EC FP5-7 programs funded by the European Commission, NATO, NSF, SNSF, DAAD and other international scientific foundations.

Today, the Ruđer Bošković Institute has over 550 scientists and researchers in more than 80 laboratories pursuing research in theoretical and experimental physics, physics and materials chemistry, electronics, physical chemistry, organic chemistry and biochemistry, molecular biology and medicine, the sea and the environment, informational and computer sciences, laser and nuclear research and development.



Roger Joseph Boskovich

http://en.wikipedia.org/wiki/Roger_Joseph_Boscovich



Ruđer Bošković (18 May 1711 – 13 February 1787) was a physicist, astronomer, mathematician, philosopher, diplomat, poet, theologian, Jesuit priest, and a polymath from the city of Dubrovnik in the Republic of Ragusa (today Croatia), who studied and lived in Italy and France where he also published many of his works.

Among his many achievements he was ***the first*** to suggest least absolute deviation based regression (1757). That was studied by Laplace (1793) and predated the least square technique originally developed by Legendre (1805) and Gauss (1823):

P. Bloomfield and W. L. Steiger. *Least Absolute Deviations: Theory, Applications, and Algorithms*. Birkhauser, Boston, MA, 1983.

Talk outline

- ◆ Instantaneous blind source separation (BSS): problem definition and overview of main methods.
- ◆ Nonlinear underdetermined BSS (uBSS): motivation, conversion to linear uBSS.
- ◆ uBSS and sparse component analysis (SCA):
 - ◆ asymptotic results from compressed sensing theory,
 - ◆ SCA by sparseness constrained non-negative matrix factorization (NMF),
 - ◆ SCA/NMF in reproducible kernel Hilbert spaces (RKHS).
- ◆ Applications: (i) pure components extraction from mass spectra of nonlinear chemical reactions; (ii) unsupervised decomposition of color (RGB) microscopic image of unstained specimen in histopathology.



Blind separation of sources

Linear problems

Nonlinear problems

Dynamic problems

Static problems

Over- and
determined
problems

Underdetermined
problems

Principal component analysis
Independent component analysis
Dependent component analysis
Nonnegative matrix factorization
Nonnegative tensor factorization

Sparse component analysis:
* clustering + l_p ($0 < p \leq 1$) min
* Hierarchical nonnegative matrix factorization

Blind Source Separation – linear static problem

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

Problem:

$$\mathbf{X} = \mathbf{A}\mathbf{S} \quad \mathbf{X} \in \mathbb{R}^{N \times T}, \quad \mathbf{A} \in \mathbb{R}^{N \times M}, \quad \mathbf{S} \in \mathbb{R}^{M \times T}$$

N - number of sensors/mixtures;

M - unknown number of sources

T - number of samples/observations

Goal: find \mathbf{S} , \mathbf{A} and number of sources M based on \mathbf{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 – linear static problem

$\mathbf{X}=\mathbf{A}\mathbf{S}$ and $\mathbf{X}=\mathbf{A}\mathbf{T}\mathbf{T}^{-1}\mathbf{S}$ are equivalent for any square invertible matrix \mathbf{T} . There are infinitely many pairs $(\mathbf{A}\mathbf{T}, \mathbf{T}^{-1}\mathbf{S})$ satisfying linear mixture model $\mathbf{X}=\mathbf{A}\mathbf{S}$. Solutions unique up to permutation and scaling indeterminacies, $\mathbf{T}=\mathbf{P}\mathbf{\Lambda}$, are meaningful. For such solutions constraints must be imposed on \mathbf{A} and/or \mathbf{S} .

Independent component analysis (ICA) solves BSS problem provided that: source signals \mathbf{S} are statistically independent and non-Gaussian; mixing matrix \mathbf{A} is full column rank i.e. $M \leq N$.

Dependent component analysis (DCA) improves accuracy of ICA when sources are not statistically independent. Linear high-pass filtering type of preprocessing transform is applied row-wise to \mathbf{X} : $L(\mathbf{X})=\mathbf{A}L(\mathbf{S})$. ICA is applied to $L(\mathbf{X})$ to estimate \mathbf{A} and $L(\mathbf{S})$. \mathbf{S} is estimated from $\mathbf{S} \approx \mathbf{A}^{-1}\mathbf{X}$.

Matlab implementation of many ICA algorithms can be found in the ICALAB:
<http://www.bsp.brain.riken.go.jp/ICALAB/>



Blind Source Separation – linear static problem

Sparse component analysis (SCA) solves BSS problem imposing sparseness constraints on source signals \mathbf{S} . M can be less than, equal to or greater than N .

Thus, SCA can be used to solve underdetermined BSS problems where number of source signals is greater than number of mixtures.

Nonnegative matrix factorization (NMF) solves BSS problem imposing nonnegativity, sparseness, smoothness or constraints on source signals. NMF algorithms that enforce sparse decomposition of \mathbf{X} can be seen as SCA algorithms [4]

Matlab implementation of many NMF algorithms can be found in the NMFLAB:
<http://www.bsp.brain.riken.jp/ICALAB/nmflab.html>

Underdetermined BSS: (nonlinear) static problem [3,2,5,6]

$\mathbf{x}_t = \mathbf{f}(\mathbf{s}_t) \quad t=1, \dots, T$; $\mathbf{x}_t \in \mathbb{R}_{0+}^{N \times 1}$ stands for nonnegative vector comprised of measurements acquired at T independent variables (pixel positions, m/z ratios, genes, etc.).

$\mathbf{s}_t \in \mathbb{R}_{0+}^{M \times 1}$ stands for unknown vector of M sources. $M > N \rightarrow$ **uBSS problem**

$\mathbf{f} : \mathbb{R}_{0+}^M \mapsto \mathbb{R}_{0+}^N$ is an unknown multivariate mapping such that:

$$\mathbf{f}(\mathbf{s}_t) = \left[f_1(\mathbf{s}_t) \dots f_N(\mathbf{s}_t) \right]^T \quad \text{and} \quad \left\{ f_n : \mathbb{R}_{0+}^M \rightarrow \mathbb{R}_{0+} \right\}_{n=1}^N .$$

Linear problem: $\mathbf{f}(\mathbf{s}_t) = \mathbf{A}\mathbf{s}_t$.

5. I. Kopriva, I. Jerić, M. Filipović, L. Brkljačić (2014). Empirical Kernel Map Approach to Nonlinear Underdetermined Blind Separation of Sparse Nonnegative Dependent Sources: Pure Components Extraction from Nonlinear Mixtures Mass Spectra. *J. of Chemometrics* , vol. 28, pp. 704-715.
6. I. Kopriva, I. Jerić, L. Brkljačić, (2013). Nonlinear Mixture-wise Expansion Approach to Underdetermined Blind Separation of Nonnegative Dependent Sources. *J. of Chemometrics*, vol. 27, pp.189-197 .



Underdetermined Blind Source Separation: motivation

In biomarker identification studies *number of mixture spectra* of biological samples (urine, blood, tissue extract, saliva, etc.) *is rather small, while number of components/analytes* (some of them are candidates for biomarkers) *can be large*.

For example, 326 analytes were quantified in extracts of *Arabidopsis thaliana* leaf tissue [7], while the independent gas chromatography-mass spectrometry (GC-MS) study of *Arabidopsis thaliana* leaves detected 497 unique chemical components [8].

Analysis of human adult urinary metabolome by liquid chromatography-mass spectrometry (LC-MS) revealed presence of 1484 components, while 384 of them were characterized by matching their spectra with references stored in libraries [9].

7. Fiehn O, Kopka J, Dörmann P, Altmann T, Trethewey R N, Willmitzer L. Metabolite profiling for plant functional genomics. *Nature Biotechnology* 2000; 18: 1157-1161.

8. Jonsson P, Johansson A I, Gullberg J, Trygg J, Jiye A, Grung B, Marklund S, Sjöström M, Antti H, Moritz T. High-throughput data analysis for detecting and identifying differences between samples in GC/MS-based metabolomic analyses," *Analytical Chem.* 2005; 77: 5635-5642.

9. Roux A, Xu Y, Heilier J-F, Olivier M-F, Ezan E, Tabet J-C, Junot C. Annotation of the human adult urinary metabolome and metabolite identification using ultra high performance liquid chromatography coupled to a linear quadrupole ion trap-orbitrap mass spectrometer. *Anal. Chem.* 2012; 84: 6429–6437.



Nonlinear u-Blind Source Separation: motivation

While linear mixture model is adequate for many scenarios, nonlinear model offers more accurate description of processes and interactions occurring in biological systems.

Living organisms are best examples of complex nonlinear systems that function far from equilibrium. Internal and external stimuli (disease, drug treatment, environmental changes) cause perturbations in the system as a result of highly synchronized molecular interactions, [10].

Furthermore, interactions within genes in components that are parts of gene regulating networks are nonlinear, [11].

10. Walleczek J (ed). Self-organized biological dynamics and non-linear control. Cambridge University Press: Cambridge, UK. 2000.

11. Yuh, C. H., Bolouri, H., Davidson, E. H.: Genomic cis-regulatory logic: experimental and computational analysis of a sea urchin gene. *Science* 279, 1896-1902 (1998).

Linear Underdetermined BSS

- SCA-based solution of the linear uBSS problem is obtained in two stages:
 - 1) estimate basis or mixing matrix \mathbf{A} using data clustering.
 - 2) estimating sources, with estimated \mathbf{A} , one at a time \mathbf{s}_t , $t=1, \dots, T$ or simultaneously solving underdetermined linear systems of equations $\mathbf{x}_t = \mathbf{A}\mathbf{s}_t$. Provided that \mathbf{s}_t is sparse enough, solution is obtained at the minimum of L_p -norm, $\|\mathbf{s}_t\|_p$, $0 \leq p \leq 1$.

Here:
$$\|\mathbf{s}_t\|_p = \left(\sum_{m=1}^M |s_{mt}|^p \right)^{1/p}.$$

- NMF-based solution yields \mathbf{A} and \mathbf{S} simultaneously through sparseness and nonnegativity constrained factorization of \mathbf{X} .



When uBSS problems can(not) be solved?

Let us focus on underdetermined linear system:

$$\mathbf{x} = \mathbf{A}\mathbf{s}, \mathbf{x} \in \mathbb{R}^N, \mathbf{s} \in \mathbb{R}^M, M > N$$

Let \mathbf{s} be K -sparse i.e. $K = \|\mathbf{s}\|_0$.

Provided that \mathbf{A} is random, with entries from Gaussian or Bernoulli distributions, compressed sensing theory has established necessary and sufficient condition on N , M and K to obtain, with probability one, unique solution at the minimum of L_1 -norm of \mathbf{s} , [12]:

$$N \approx K \log(M/K)$$

12. Candès E, Tao T. Near optimal signal recovery from random projections: universal encoding strategy?. *IEEE Trans. Information Theory* 2006; **52**: 5406-5425.



When uBSS problems can(not) be solved?

However in BSS problems \mathbf{A} is not random matrix but deterministic matrix with a structure. For example, in multispectral imaging it contains spectral profiles of the objects/materials present in the image, [13]. In chemometrics \mathbf{A} contains concentration profiles of pure components present in the mixtures, [14].

One result for deterministic \mathbf{A} is given in [15]. For cyclic polynomial matrix \mathbf{A} it applies $N = O(K^2)$. That is significantly worse than $N \approx K \log(M/K)$ for random \mathbf{A} . K corresponds with number of sources that are active/present at the specific coordinate t (time, pixel, m/z variable, frequency, etc). Thus, K is application dependent.

13. Kopriva I, Cichocki A. Blind decomposition of low-dimensional multi-spectral image by sparse component analysis. *J. Chemometrics* 2009; **23** (11): 590-597.

14. Kopriva I, Jerić I. Blind separation of analytes in nuclear magnetic resonance spectroscopy and mass spectrometry: sparseness-based robust multicomponent analysis. *Anal. Chem.* 2010; **82**: 1911-1920.

15. DeVore R A. Deterministic constructions of compressed sensing matrices. *Journal of Complexity* 2007; **23**: 918-925.

When uBSS problems can(not) be solved?

In addition to sparseness requirement on \mathbf{s} certain degree of incoherence of the mixing matrix \mathbf{A} is required as well. Mutual coherence is defined as the largest absolute and normalized inner product between different columns in \mathbf{A} , what reads as

$$\mu\{\mathbf{A}\} = \max_{1 \leq i, j \leq M \text{ and } i \neq j} \frac{|\mathbf{a}_i^T \mathbf{a}_j|}{\|\mathbf{a}_i\| \|\mathbf{a}_j\|}$$

The mutual coherence provides a **worst case** measure of similarity between the basis vectors. It indicates how much two closely related vectors may confuse any pursuit algorithm (solver of the underdetermined linear system of equations). The worst-case perfect recovery condition for \mathbf{s} relates sparseness requirement on \mathbf{s} and coherence of \mathbf{A} , [16,17]:

$$\|\mathbf{s}\|_0 < \frac{1}{2} \left(1 + \frac{1}{\mu\{\mathbf{A}\}} \right)$$

16. R. Gribonval and M. Nielsen, "Sparse representations in unions of bases," *IEEE Transactions on Information Theory* **49**, 3320-3325 (2003).

17. J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Transactions on Information Theory* **50**, 2231-2242 (2004).

When uBSS problems can(not) be solved?

When the mutual coherence $\mu(\mathbf{A})$ is **very close to** 1 possibility to obtain meaningful solution of $\mathbf{x}=\mathbf{A}\mathbf{s}$ is reduced drastically. Such scenario occurs when, as an example, \mathbf{X} represent RGB microscopic image of **unstained specimen** in histopathology, [18]. In such scenario $\mu(\mathbf{A})\approx 0.9999$.

Even though uniqueness condition holds formally, **only small amount of noise or modelling error** will make the algorithms, such as basis pursuit denoising algorithm [19, 20], unstable [21, 22].

18. I. Kopriva, M. Popović Hadžija, M. Hadžija, G. Aralica, "Unsupervised segmentation of low-contrast multi-channel images: discrimination of tissue components in microscopic images of unstained specimens," *Scientific Reports* 5: 11576, DOI: 10.1038/srep11576.

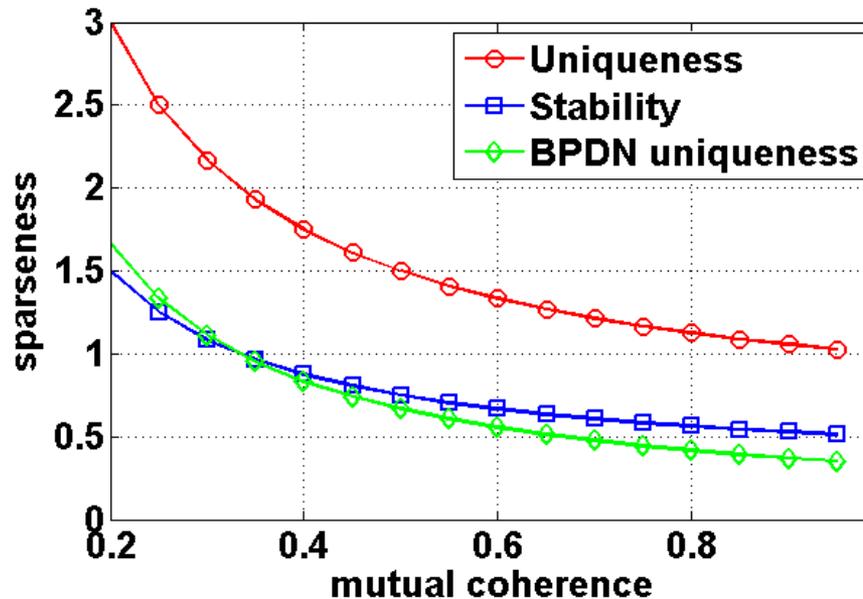
19. Bruckstein, A.M., Donoho, D.L., and Elad, M., "From Sparse Solutions of Systems of Equations to Sparse Modeling of Signals and Images," *SIAM Review* 51 (1), 34-81 (2009).

20. Tibshirani, R., "Regression shrinkage and selection via the Lasso," *J. Roy. Statist. Soc. B* 58 (1), 267-288 (1996).

21. Chen, S.S., Donoho, D.L., and Saunders, M.A., "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.* 20, 33-61 (1998).

22. Donoho, D.L., Elad, M., and Temlyakov, V., "Stable recovery of sparse overcomplete representations in the presence of noise," *Information Theory, IEEE Transactions on* 52, 6-18 (2006).

When uBSS problems can(not) be solved?



The amount of sparseness as a function of mutual coherence imposed by various uniqueness/stability conditions. Circles: uniqueness condition without noise. Squares: stability condition. Diamonds: uniqueness condition for basis pursuit denoising algorithm. For non-overlapping (orthogonal) histological structures sparseness equals $\|s_p\|_0 = 1$. Thus, while uniqueness condition in the absence of noise is satisfied even when $\mu(A) \approx 1$ it is seen that approximately $\mu(A) < 0.33$ is required to satisfy uniqueness condition in the presence of modelling errors or noise.

When uBSS problems can(not) be solved?

For scenarios when mutual coherence $\mu(\mathbf{A}) \approx 1$ it was proposed in [18,5,6] to transform problem:

$$\mathbf{X} = \mathbf{A}\mathbf{S} \text{ or } \mathbf{X} = \mathbf{f}(\mathbf{S}) = \mathbf{G}\bar{\mathbf{S}}$$

into:

$$\Psi(\mathbf{X}) = \mathbf{B}\bar{\mathbf{S}}$$

such that $\mu(\mathbf{B}) < \mu(\mathbf{A})$, resp. $\mu(\mathbf{B}) < \mu(\mathbf{G})$. For $\{0, 1\}$ binary non-overlapping sources it was shown in [18]: $\mathbf{X} = \mathbf{f}(\mathbf{S}) = \mathbf{G}\mathbf{S}$ and $\Psi(\mathbf{X}) = \mathbf{B}\mathbf{S}$.

5. I. Kopriva, I. Jerić, M. Filipović, L. Brkljačić (2014). Empirical Kernel Map Approach to Nonlinear Underdetermined Blind Separation of Sparse Nonnegative Dependent Sources: Pure Components Extraction from Nonlinear Mixtures Mass Spectra. *J. of Chemometrics*, vol. 28, pp. 704-715.
6. I. Kopriva, I. Jerić, L. Brkljačić, (2013). Nonlinear Mixture-wise Expansion Approach to Underdetermined Blind Separation of Nonnegative Dependent Sources. *J. of Chemometrics*, vol. 27, pp.189-197.
18. I. Kopriva, M. Popović Hadžija, M. Hadžija, G. Aralica, "Unsupervised segmentation of low-contrast multi-channel images: discrimination of tissue components in microscopic images of unstained specimens," *Scientific Reports* 5: 11576, DOI: 10.1038/srep11576.



Linear uBSS: summary

Linear uBSS problem is characterized with a triplet (N, M, K) . Under L_1 -norm constraints unique solution is possible if $N \approx K \log(M/K)$.

In biological experiments M corresponds with number of analytes (metabolites) present in mixture spectra and, thus, can be large. K represents maximal number of overlapping components. Depending on the resolution of the spectrometer it can be large as well. N stands for number of biological samples and is usually (very) small. Thus, requirement $N \approx K \log(M/K)$ can often failed to be fulfilled!!!

In [6] solution was proposed to transform original uBSS problem $\mathbf{x} = \mathbf{A}\mathbf{s}$ into new one $\Psi(\mathbf{x}) = \bar{\mathbf{A}}\bar{\mathbf{s}}$, $\bar{\mathbf{A}} \in \mathbb{R}_{0+}^{D \times P}$, $\bar{\mathbf{s}} \in \mathbb{R}_{0+}^{P \times 1}$, with maximal number of overlapping components equal to Q . Thus, uniqueness condition becomes: $D \approx Q \log(P/Q)$. That is fulfilled if: $(D/N) \gg (P/M)$ as well as $(D/N) \gg (Q/K)$.

Nonlinear mapping of linear uBSS problem?

In [6] a new concept was proposed by mapping original uBSS problem $\mathbf{X}=\mathbf{A}\mathbf{S}$ nonlinearly into new one:

$$\left\{ \mathbf{x}(t) \rightarrow \phi(\mathbf{x}(t)) \right\}_{t=1}^T \quad \text{s.t.} \quad \mathbf{x}(t) \in \mathbb{R}_{0+}^N, \quad \phi(\mathbf{x}(t)) \in \mathbb{R}_{0+}^{\bar{N}} \quad \text{and} \quad \bar{N} \gg N$$

since mapping $\phi(\mathbf{x}(t))$ is nonlinear new measurements are linearly independent.

The nonlinear mapping has the following algebraic structure:

$$\phi(\mathbf{x}(t)) = \left[\left\{ c_{q_1 \dots q_N} x_1^{q_1}(t) \dots x_N^{q_N}(t) \right\}_{q_1, \dots, q_N=0}^{\bar{N}} \right]^T \quad \text{such that} \quad \sum_{n=1}^N q_n \leq \bar{N}, \quad \forall t = 1, \dots, T.$$

Nonlinear mapping of linear uBSS problem?

The mapped problem becomes:

$$\phi(\mathbf{x}(t)) = c_0 \mathbf{e}_1 + \mathbf{B} \begin{bmatrix} 0 \\ \mathbf{s}(t) \end{bmatrix} + \mathbf{B}_{HOT} \begin{bmatrix} 0 \\ \mathbf{0}_{M \times 1} \\ \mathbf{s}(t)_{HOT} \end{bmatrix} \quad \forall t = 1, \dots, T$$

where $\mathbf{s}(t)_{HOT}$ is $\bar{N} - M - 1$ column vector comprised of: $\left\{ s_1^{q_1}(t) \times \dots \times s_M^{q_M}(t) \right\}_{q_1, \dots, q_M=2}^{\bar{N}}$

such that: $\sum_{m=1}^M q_m \leq \bar{N}$.

Sparse probabilistic model of sources

Let us assume sparse probabilistic model of the sources, i.e. each source signal is distributed according to p.d.f. based on mixed state random variable model [23, 24, 5]:

$$p(s_{mt}) = \rho \delta(s_{mt}) + (1 - \rho) \delta^*(s_{mt}) f(s_{mt}) \quad \forall m = 1, \dots, M \quad \forall t = 1, \dots, T$$

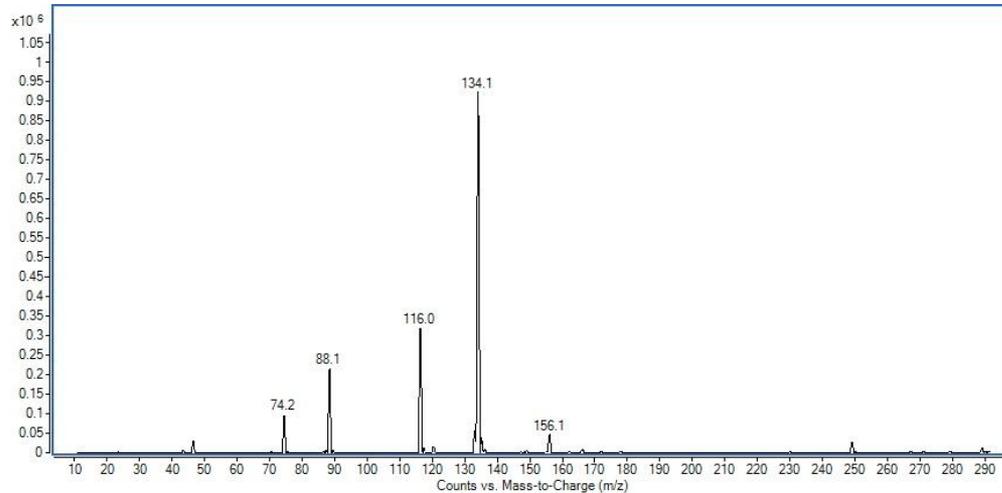
where $\delta(s_{mt})$ is an indicator function and $\delta^*(s_{mt}) = 1 - \delta(s_{mt})$ is its complementary function. $\rho = \left\{ P(s_{mt} = 0) \right\}_{t=1}^T$ Thus, $\left\{ P(s_{mt} > 0) = 1 - \rho \right\}_{t=1}^T$.

23. Bouthemy P, Piriou C H G, Yao J. Mixed-state auto-models and motion texture modeling. *J. Math Imaging Vision* 2006; 25: 387-402.

24. Caifa C, Cichocki A. Estimation of Sparse Nonnegative Sources from Noisy Overcomplete Mixtures Using MAP. *Neural Comput.* 2009; 21: 3487-3518.

5. I. Kopriva, I. Jerić, M. Filipović, L. Brkljačić (2014). Empirical Kernel Map Approach to Nonlinear Underdetermined Blind Separation of Sparse Nonnegative Dependent Sources: Pure Components Extraction from Nonlinear Mixtures Mass Spectra. *J. of Chemometrics*, vol. 28, pp. 704-715.

Sparse probabilistic model of sources



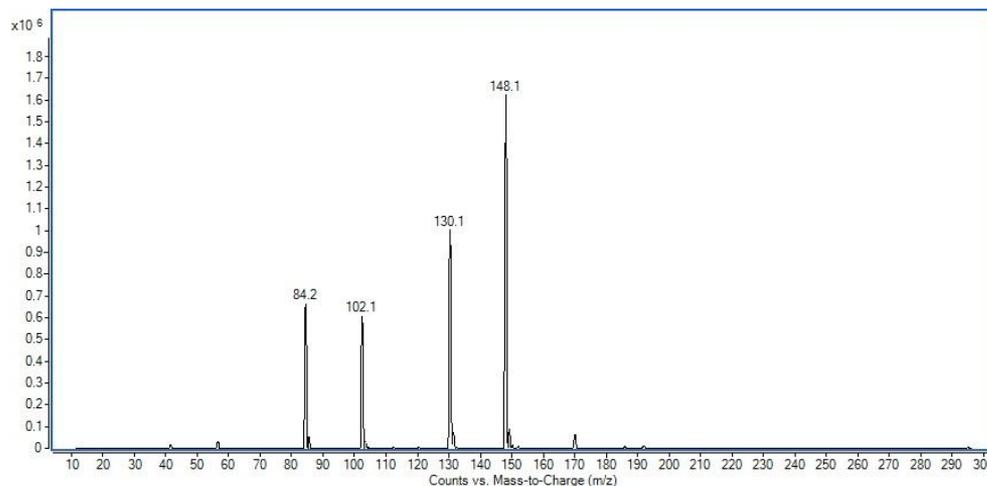
Examples of mass spectra of sources (pure components, analytes,...). They are sparse in support and amplitude. We can use exponential distribution for:

$$f(s_{mt}) = (1/\mu_m) \exp(-s_{mt}/\mu_m)$$

In [5] using mass spectra of 25 pure components it has been estimated:

$$\hat{\rho}_m \in [0.27, 0.74]$$

$$\hat{\mu}_m \in [0.0012, 0.0014]$$



Nonlinear mapping of linear uBSS problem?

Thus, with high probability at least one source will not be present at location t . Thus, many cross-products will vanish. Also, by assuming $0 \leq s_{mt} \leq 1$ it follows that $s_m^{q_m}(t) \rightarrow 0$ when q_m grows.

Thus, by hard or soft thresholding of $\phi(\mathbf{x}(t))$ higher-order terms can be suppressed. Under sparse probabilistic prior validated on experimental mass spectra mostly second order terms will survive. That yields:

$$\phi(\mathbf{X})_\tau \approx \underbrace{\begin{bmatrix} c_0 \mathbf{e}_1 & \dots & c_0 \mathbf{e}_1 \\ \times T \text{ times} \end{bmatrix}}_{\times T \text{ times}} + \bar{\mathbf{B}} \begin{bmatrix} 0 \\ \mathbf{S} \\ \left\{ \mathbf{s}_{m_1} \mathbf{s}_{m_2} \right\}_{m_1, m_2=1}^M \end{bmatrix}$$

where : $\phi(\mathbf{X})_\tau \in \mathbb{R}_{0+}^{\bar{N} \times T}$ $\bar{\mathbf{B}} \in \mathbb{R}_{0+}^{\bar{N} \times P+1}$ and $P \approx 2M + M(M-1)/2$.

Nonlinear mapping of linear uBSS problem?

Thus, linear uBSS problem characterized by (N, M, K) is converted into new one characterized by (\bar{N}, P, Q) , where Q denotes maximal number of overlapping sources in mapped domain. If sources do not overlap heavily and higher-order terms are suppressed we have:

$$(\bar{N} / N) \gg (P / M) \quad \text{and} \quad (\bar{N} / N) \gg (Q / K)$$

where $Q \approx 2K + K(K-1)/2$. $P \approx 2M + M(M-1)/2$ above condition becomes:

$$(\bar{N} / N) \gg (M / 2 - 3/2) \quad \text{and} \quad (\bar{N} / N) \gg (K / 2 - 3/2)$$

The same procedure can be applied to equivalent linear representation $\mathbf{x} = \mathbf{G}\bar{\mathbf{s}}$ of the nonlinear BSS problem $\mathbf{x} = \mathbf{f}(\mathbf{s})$.

Nonlinear mapping of linear uBSS problem?

The problem with using explicit feature maps $\phi(\mathbf{x}(t))$ is that \bar{N} can be very large or even infinite. Thus, factorization problem:

$$\phi(\mathbf{X})_{\tau} \approx \underbrace{\begin{bmatrix} c_0 \mathbf{e}_1 & \dots & c_0 \mathbf{e}_1 \end{bmatrix}}_{\times T \text{ times}} + \bar{\mathbf{B}} \begin{bmatrix} 0 \\ \mathbf{S} \\ \left\{ \mathbf{s}_{m_1} \mathbf{s}_{m_2} \right\}_{m_1, m_2=1}^M \end{bmatrix}$$

becomes computationally intractable.

Reproducible kernel Hilbert spaces

Definition 1. A real function $\kappa: \mathbb{R}^N \times \mathbb{R}^N \rightarrow \mathbb{R}$ is positive semi-definite if it is symmetric and satisfies for any finite set of points $\{\mathbf{x}_t \in \mathbb{R}^N\}_{t=1}^T$ and real numbers $\{\alpha_t\}_{t=1}^T$: $\sum_{i,j=1}^T \alpha_i \alpha_j \kappa(\mathbf{x}_i, \mathbf{x}_j) \geq 0$.

Theorem 1. The Morre-Aronszjan theorem [25]. Given any nonnegative definite function $\kappa(\mathbf{x}, \mathbf{y})$ there exists a uniquely determined RKHS H_κ consisting of real valued functions on set $\mathbf{X} \subset \mathbb{R}^N$ such that: (i) $\forall \mathbf{x} \in \mathbf{X}, \kappa(\circ, \mathbf{x}) \in H_\kappa$;
 (ii) $\forall \mathbf{x} \in \mathbf{X}, \forall f \in H_\kappa, f(\mathbf{x}) = \langle f, \kappa(\circ, \mathbf{x}) \rangle_{H_\kappa}$. Here, $\langle \circ, \circ \rangle$ denotes inner product associated with H_κ .

Reproducible kernel Hilbert spaces

Definition 2. Replacing $f(\mathbf{x})$ in (ii) in Theorem 1 by $\kappa(\circ, \mathbf{x})$ it follows

$\kappa(\mathbf{x}_t, \mathbf{x}) = \langle \kappa(\circ, \mathbf{x}_t), \kappa(\circ, \mathbf{x}) \rangle_{H_\kappa}$. By selecting the nonlinear map as $\phi(\mathbf{x}) = \kappa(\circ, \mathbf{x})$ it follows $\kappa(\mathbf{x}_t, \mathbf{x}) = \langle \phi(\mathbf{x}_t), \phi(\mathbf{x}) \rangle_{H_\kappa}$. That is known as **kernel trick**. The nonlinear mapping $\phi(\mathbf{x})$ is known as an **explicit feature map** (EFM) associated with kernel $\kappa(\circ, \mathbf{x})$.

Definition 3. Empirical kernel map (EKM), [26]. For a given set of patterns

$\{\mathbf{v}_d \in \mathbb{R}^N\}_{d=1}^D \subset \mathbf{X}$, $D \in \mathbb{N}$, we call $\psi: \mathbb{R}^N \rightarrow \mathbb{R}^D$:

$\left\{ \mathbf{x}_t \mapsto \kappa(\circ, \mathbf{x}_t) \Big|_{\{\mathbf{v}_d\}_{d=1}^D} = \left[\kappa(\mathbf{v}_1, \mathbf{x}_t), \dots, \kappa(\mathbf{v}_D, \mathbf{x}_t) \right] \right\}_{t=1}^T$ the EKM with respect to $\{\mathbf{v}_d\}_{d=1}^D$.

Nonlinear mapping of linear uBSS problem?

The problem with using explicit feature maps $\phi(\mathbf{x}(t))$ is that \bar{N} can be very large or even infinite. Thus, factorization problem:

$$\phi(\mathbf{X})_{\tau} \approx \underbrace{\begin{bmatrix} c_0 \mathbf{e}_1 & \dots & c_0 \mathbf{e}_1 \end{bmatrix}}_{\times T \text{ times}} + \bar{\mathbf{B}} \begin{bmatrix} 0 \\ \mathbf{S} \\ \left\{ \mathbf{s}_{m_1} \mathbf{s}_{m_2} \right\}_{m_1, m_2=1}^M \end{bmatrix}$$

becomes computationally intractable. That is fixed by projecting $\phi(\mathbf{x}(t))$ onto $\phi(\mathbf{V})$ where $\mathbf{V} = \left\{ \mathbf{v}_d \in \mathbb{R}^{N \times 1} \right\}_{d=1}^D$ stands for basis such that:

$$\text{span} \left\{ \mathbf{v}_d \right\}_{d=1}^D \approx \text{span} \left\{ \mathbf{x}_t \right\}_{t=1}^T$$

Then:

$$\text{span} \left\{ \phi(\mathbf{v}_d) \right\}_{d=1}^D \approx \text{span} \left\{ \phi(\mathbf{x}_t) \right\}_{t=1}^T$$

Nonlinear mapping of linear uBSS problem?

Projection yields:

$$\phi(\mathbf{V})^T \phi(\mathbf{x}_t) = \psi(\mathbf{x}_t)_{\mathbf{V}} = \left[\langle \phi(\mathbf{v}_1), \phi(\mathbf{x}_t) \rangle \dots \langle \phi(\mathbf{v}_D), \phi(\mathbf{x}_t) \rangle \right]^T$$

When $\phi(\mathbf{x}) = k(\cdot, \mathbf{x})$ it follows: $\langle \phi(\mathbf{v}), \phi(\mathbf{x}) \rangle = k(\mathbf{v}, \mathbf{x})$. It is shown in [6] that when sources comply with sparse probabilistic model it applies:

$$\psi(\mathbf{X})_{\tau} \approx \left[\underbrace{c_0 \mathbf{e}_1 \dots c_0 \mathbf{e}_1}_{\times T \text{ times}} \right] + \bar{\mathbf{B}} \begin{bmatrix} 0 \\ \mathbf{S} \\ \left\{ \mathbf{s}_{m_1} \mathbf{s}_{m_2} \right\}_{m_1, m_2=1}^M \end{bmatrix}$$

$$\psi(\mathbf{X})_{\tau} \in \mathbb{R}_{0+}^{D \times T} \quad \bar{\mathbf{B}} \in \mathbb{R}_{0+}^{D \times P+1} \quad \text{and } P \approx 2M + M(M-1)/2.$$

Subscript τ indicates that some type of thresholding was applied on $\psi(\mathbf{X})$ to suppress *HOT*. For $\{0, 1\}$ binary non-overlapping sources: $\psi(\mathbf{X}) = \mathbf{BS}$, [18].

Nonlinear mapping of linear uBSS problem?

Nonlinear uBSS problem (N, M, K) is substituted by the linear BSS problem $(D, 2M + M(M-1)/2, Q)$, $Q \approx 2K + K(K-1)/2$. Equivalent linear BSS problem is solvable when:

$$(D/N) \gg (P/M) \quad \text{and} \quad (D/N) \gg (Q/K)$$

Since $P \approx 2M + M(M-1)/2$ and $Q \approx 2K + K(K-1)/2$ above condition becomes:

$$(D/N) \gg (M/2 - 3/2) \quad \text{and} \quad (D/N) \gg (K/2 - 3/2)$$

That is possible to fulfill by finding basis $\mathbf{V} = \left\{ \mathbf{v}_d \in \mathbb{R}^{N \times 1} \right\}_{d=1}^D$ with sufficiently large dimension D .

Nonlinear mapping of linear uBSS problem?

Basis $\mathbf{V} = \left\{ \mathbf{v}_d \in \mathbb{R}^{N \times 1} \right\}_{d=1}^D$ needs to fulfill:

$$\text{span} \left\{ \mathbf{v}_d \right\}_{d=1}^D \approx \text{span} \left\{ \mathbf{x}_t \right\}_{t=1}^T$$

Thus, \mathbf{V} can be found by **clustering** $\left\{ \mathbf{x}_t \right\}_{t=1}^T$ into $D \leq T$ clusters. That, for example, can be accomplished by *kmeans* algorithm.

For $D=T$ each data sample is a basis vector and clustering is not necessary. But, computational costs in matrix factorization stage (that follows) is very large.

When in addition to sparseness constraint nonnegativity constraints apply as well (that is the case in applications in imaging and/or mass spectrometry) sparseness constrained NMF algorithms can be applied to $\psi(\mathbf{X})_\tau$ to estimate source components.



Nonnegative matrix factorization

Many BSS problems arising in imaging, chemo- and/or bioinformatics are described by superposition of non-negative latent variables (sources):

$$\mathbf{X} = \mathbf{AS} \quad \mathbf{X} \in \mathbb{R}_{0+}^{N \times T}, \quad \mathbf{A} \in \mathbb{R}_{0+}^{N \times M} \quad \text{and} \quad \mathbf{S} \in \mathbb{R}_{0+}^{M \times T}$$

where N represents number of sensors, M represents number of sources and T represents number of samples.

Thus, solution of related decomposition problem can be obtained by imposing non-negativity constraints on \mathbf{A} and \mathbf{S} , to narrow down number of possible decomposition of \mathbf{X} . This leads to NMF algorithms.

Due to non-negativity constraints some other constraints (statistical independence) can be relaxed/replaced in applications where they are not fulfilled.

Nonnegative matrix factorization

Modern approaches to NMF problems have been initiated by Lee-Seung' Nature paper, [27], where it is proposed to estimate \mathbf{A} and \mathbf{S} through alternative minimization procedure of the two possibly different cost functions:

Set Randomly initialize: $\mathbf{A}^{(0)}$, $\mathbf{S}^{(0)}$,

For $k=1,2,\dots$, until convergence do

$$\text{Step 1: } \mathbf{S}^{(k+1)} = \arg \min_{s_{mi} \geq 0} D_s \left(\mathbf{X} \parallel \mathbf{A}^{(k)} \mathbf{S} \right)_{\mathbf{S}^{(k)}}$$

$$\text{Step 2: } \mathbf{A}^{(k+1)} = \arg \min_{a_{nm} \geq 0} D_A \left(\mathbf{X} \parallel \mathbf{A} \mathbf{S}^{(k+1)} \right)_{\mathbf{A}^{(k)}}$$

If both cost functions represent squared Euclidean distance (Froebenius norm) we obtain alternating least square (ALS) approach to NMF.

Nonnegative matrix factorization

ALS-based NMF:

$$\left(\mathbf{A}^*, \mathbf{S}^*\right) = \arg \min_{\mathbf{A}, \mathbf{S}} D(\mathbf{X} \| \mathbf{AS}) = \frac{1}{2} \|\mathbf{X} - \mathbf{AS}\|_F^2 \quad s.t. \mathbf{A} \geq \mathbf{0}, \mathbf{S} \geq \mathbf{0}$$

- Minimization of the square of Euclidean norm of approximation error $\mathbf{E} = \mathbf{X} - \mathbf{AS}$ is, from the maximum likelihood viewpoint, justified only if error distribution is Gaussian:

$$p(\mathbf{X} | \mathbf{A}, \mathbf{S}) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{\|\mathbf{X} - \mathbf{AS}\|_2^2}{2\sigma^2}\right)$$

- In many instances non-negativity constraints imposed on \mathbf{A} and \mathbf{S} do not suffice to obtain solution that is unique up to standard BSS indeterminacies: permutation and scaling.

Nonnegative matrix factorization

In relation to original Lee-Seung NMF algorithm additional constraints are necessary to obtain factorization unique up to permutation and scaling. Generalization that involves constraints is given in [28]:

$$D(\mathbf{X} \parallel \mathbf{AS}) = \frac{1}{2} \|\mathbf{X} - \mathbf{AS}\|_F^2 + \alpha_S J_S(\mathbf{S}) + \alpha_A J_A(\mathbf{A})$$

where $J_S(\mathbf{S}) = \sum_{m,t} s_{mt}$ and $J_A(\mathbf{A}) = \sum_{n,m} a_{nm}$ are sparseness constraints that correspond with L_1 -norm of \mathbf{S} and \mathbf{A} respectively. α_S and α_A are regularization constants. Gradient components in matrix form are:

$$\frac{\partial D(\mathbf{A}, \mathbf{S})}{\partial a_{nm}} = \left[-\mathbf{XS}^T + \mathbf{ASS}^T \right]_{nm} + \alpha_A \frac{\partial J_A(\mathbf{A})}{\partial a_{nm}}$$

$$\frac{\partial D(\mathbf{A}, \mathbf{S})}{\partial s_{mt}} = \left[-\mathbf{A}^T \mathbf{X} + \mathbf{A}^T \mathbf{AS} \right]_{mt} + \alpha_S \frac{\partial J_S(\mathbf{S})}{\partial s_{mt}}$$

Maximum a posteriori probability BSS/NMF

Maximization of *a-posterior* probability (MAP) $P(\mathbf{A}, \mathbf{S} | \mathbf{X})$ yields:

$$\left(\mathbf{A}^*, \mathbf{S}^* \right) = \max_{\mathbf{A}, \mathbf{S}} P(\mathbf{A}, \mathbf{S} | \mathbf{X}) \propto \max_{\mathbf{A}, \mathbf{S}} P(\mathbf{X} | \mathbf{A}, \mathbf{S}) P(\mathbf{A}) P(\mathbf{S}) \quad s.t. \mathbf{A} \geq \mathbf{0}, \mathbf{S} \geq \mathbf{0}$$

Above formulation is equivalent to maximizing likelihood $P(\mathbf{X} | \mathbf{A}, \mathbf{S})$ and maximizing prior probabilities $P(\mathbf{A})$ and $P(\mathbf{S})$. Assuming normal distribution of approximation error $\mathbf{E} = \mathbf{X} - \mathbf{AS}$ this yields:

$$\left(\mathbf{A}^*, \mathbf{S}^* \right) = \arg \min_{(\mathbf{A}, \mathbf{S})} \frac{1}{2} \|\mathbf{X} - \mathbf{AS}\|_F^2 + \alpha_S J_S(\mathbf{S}) + \alpha_A J_A(\mathbf{A}) \quad s.t. \mathbf{A} \geq \mathbf{0}, \mathbf{S} \geq \mathbf{0}.$$

Maximum a posteriori probability BSS/NMF

Assuming non-informative prior on \mathbf{A} : $P(\mathbf{A}) = \text{const}$ and Laplacian (sparse) prior on \mathbf{S} : $P(\mathbf{S}) = \exp(-(|s_1| + \dots + |s_M|))$ yields:

$$(\mathbf{A}^*, \mathbf{S}^*) = \arg \min_{(\mathbf{A}, \mathbf{S})} \frac{1}{2} \|\mathbf{X} - \mathbf{AS}\|_F^2 + \alpha_S \|\mathbf{S}\|_1 \quad s.t. \mathbf{A} \geq \mathbf{0}, \mathbf{S} \geq \mathbf{0}.$$

It is possible to select for $P(\mathbf{S})$ prior other than Laplacian. That leads to general sparseness constrained factorization:

$$(\mathbf{A}^*, \mathbf{S}^*) = \arg \min_{(\mathbf{A}, \mathbf{S})} \frac{1}{2} \|\mathbf{X} - \mathbf{AS}\|_F^2 + \alpha_S \|\mathbf{S}\|_p \quad s.t. 0 < p \leq 1, \mathbf{A} \geq \mathbf{0}, \mathbf{S} \geq \mathbf{0}.$$

Nonnegative matrix factorization

Since NMF problem deals with non-negative variables the idea is to automatically ensure non-negativity of \mathbf{A} and \mathbf{S} through learning. That can be achieved by multiplicative learning equations:

$$\mathbf{A} \leftarrow \mathbf{A} \otimes \frac{\nabla_{\mathbf{A}}^{-} D(\mathbf{A}, \mathbf{S})}{\nabla_{\mathbf{A}}^{+} D(\mathbf{A}, \mathbf{S})} \quad \mathbf{S} \leftarrow \mathbf{S} \otimes \frac{\nabla_{\mathbf{S}}^{-} D(\mathbf{A}, \mathbf{S})}{\nabla_{\mathbf{S}}^{+} D(\mathbf{A}, \mathbf{S})}$$

where \otimes denotes entry-wise multiplication, $\nabla_{\mathbf{A}}^{-} D(\mathbf{A}, \mathbf{S})$ and $\nabla_{\mathbf{A}}^{+} D(\mathbf{A}, \mathbf{S})$ denote respectively negative and positive part of the gradient $\nabla_{\mathbf{A}} D(\mathbf{A}, \mathbf{S})$. Likewise, $\nabla_{\mathbf{S}}^{-} D(\mathbf{A}, \mathbf{S})$ and $\nabla_{\mathbf{S}}^{+} D(\mathbf{A}, \mathbf{S})$ are negative and positive part of the gradient $\nabla_{\mathbf{S}} D(\mathbf{A}, \mathbf{S})$.

When gradients converge to zero corrective terms converge to one. Since learning equations include multiplications and divisions of non-negative terms, non-negativity is ensured automatically.

Nonnegative matrix factorization

Multiplicative learning rules for NMF based on regularized squared L_2 -norm of the approximation are obtained as:

$$\mathbf{A} \leftarrow \mathbf{A} \otimes \frac{\left[\mathbf{X}\mathbf{S}^T - \alpha_A \frac{\partial J_A(\mathbf{A})}{\partial \mathbf{A}} \right]_+}{\mathbf{A}\mathbf{S}\mathbf{S}^T + \varepsilon \mathbf{1}_{NM}} \quad \mathbf{S} \leftarrow \mathbf{S} \otimes \frac{\left[\mathbf{A}^T \mathbf{X} - \alpha_S \frac{\partial J_S(\mathbf{S})}{\partial \mathbf{S}} \right]_+}{\mathbf{A}^T \mathbf{A}\mathbf{S} + \varepsilon \mathbf{1}_{MT}}$$

where $[x]_+ = \max\{\varepsilon, x\}$ with small ε . For L_1 -norm based regularization, derivatives of sparseness constraints in above expressions are equal to 1, i.e.:

$$\mathbf{A} \leftarrow \mathbf{A} \otimes \frac{\left[\mathbf{X}\mathbf{S}^T - \alpha_A \mathbf{1}_{NM} \right]_+}{\mathbf{A}\mathbf{S}\mathbf{S}^T + \varepsilon \mathbf{1}_{NM}} \quad \mathbf{S} \leftarrow \mathbf{S} \otimes \frac{\left[\mathbf{A}^T \mathbf{X} - \alpha_S \mathbf{1}_{MT} \right]_+}{\mathbf{A}^T \mathbf{A}\mathbf{S} + \varepsilon \mathbf{1}_{MT}}$$

Non-negative matrix under-approximation (NMU)

NMF algorithms outlined before require a priori knowledge of sparseness related regularization (trade off) constant.

A sequential approach to NMF has been recently proposed in [29] by estimating rank-1 one factors $\mathbf{a}_m \mathbf{s}_m$ one at a time. Each time $\mathbf{a}_m \mathbf{s}_m$ is estimated it is removed from $\mathbf{X} \rightarrow \mathbf{X} - \mathbf{a}_m \mathbf{s}_m$. To prevent subtraction from being negative the under-approximation constraint is imposed on $\mathbf{a}_m \mathbf{s}_m$: $\mathbf{a}_m \mathbf{s}_m \leq \mathbf{X}$.

Hence, the NMU algorithm is obtained as a solution of:

$$(\mathbf{A}^*, \mathbf{S}^*) = \arg \min_{(\mathbf{A}, \mathbf{S})} \frac{1}{2} \|\mathbf{X} - \mathbf{AS}\|_F^2 \quad s.t. \quad \mathbf{A} \geq \mathbf{0}, \mathbf{S} \geq \mathbf{0}, \mathbf{AS} \leq \mathbf{X}.$$



Non-negative matrix under-approximation (NMU)

Theorem 1 in [29] proves that number of nonzero entries in \mathbf{A} and \mathbf{S} is less than in \mathbf{X} . Thus, the underapproximation constraint ensures sparse (parts based) factorization of \mathbf{X} . This, however, does not imply that \mathbf{A} and \mathbf{S} obtained by enforcing underapproximation constrain yields the sparsest decomposition of \mathbf{X} .

However, since no explicit regularization is used there are no difficulties associated with selecting values of regularization constants.

MATLAB code for NMU algorithm is available at:
<https://sites.google.com/site/nicolasgillis/code>



Non-negative matrix factorization with L_0 -constraint (NMF_L0)

The NMF_L0 algorithm, [30], imposes explicit L_0 -constraint on entries of \mathbf{S} , i.e. number of nonzero entries is tried to be minimized explicitly by integrating nonnegativity constraint in the OMP algorithm. That is achieved through modifications of the nonnegative least square (NNLS) algorithm, [31], called sparse NNLS and recursive sparse NNLS. The mixing matrix is updated by some of standards dictionary update methods.

The „weak” side of the NMF_L0 algorithm is that, in addition to number of sources M , the maximal number of overlapped sources K has to be known *a priori*. Quite often that is hard to achieve in practice.

MATLAB code for NMF_L0 algorithm is available at:
<http://www3.spsc.tugraz.at/people/robert-peharz>.

30. R. Peharz, F. Pernkopf, "Sparse nonnegative matrix factorization with ℓ^0 constraints," *Neurocomputing*, vol. 80, pp. 38-46, 2012.

31. C. Lawson, R. Hanson, *Solving Least Squares Problems*, Prentice-Hall, 1974.

Nonlinear underdetermined blind source separation: numerical experiments and separation of pure components mass spectra from mixtures of nonlinear chemical reactions [5]

5. I. Kopriva, I. Jerić, M. Filipović, L. Brkljačić (2014). Empirical Kernel Map Approach to Nonlinear Underdetermined Blind Separation of Sparse Nonnegative Dependent Sources: Pure Components Extraction from Nonlinear Mixtures Mass Spectra. *J. of Chemometrics* , vol. 28, pp. 704-715.

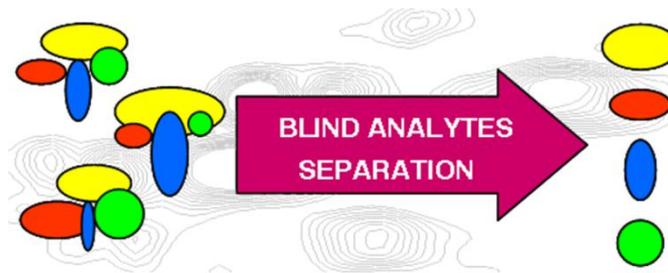
Linear mixing model

$$\mathbf{X} = \mathbf{A}\mathbf{S}$$

$$\mathbf{X} \in \mathbb{R}_{0+}^{N \times T}, \mathbf{S} \in \mathbb{R}_{0+}^{M \times T}, \mathbf{A} \in \mathbb{R}_{0+}^{N \times M}$$

In chemometrics (NMR spectroscopy or mass spectrometry) rows of \mathbf{X} represent spectra of mixture samples, columns of \mathbf{A} represent concentration profiles of analytes (a.k.a. pure components) present in mixture spectra \mathbf{X} , while rows of \mathbf{S} represent spectra of analytes present in mixture spectra \mathbf{X} .

The (u)BSS problem relates to extraction of analytes (and their concentrations) using mixture spectra \mathbf{X} only:



Pure components can represent compounds indicative for disease. Thus, they can be useful for ***biomarker analysis***. They can be isolated from spectra (NMR, mass) of biological samples (urine, blood, tissues).



Implementation details

Studies on numerical and experimental data reported below were executed on personal computer running under Windows 64-bit operating system with 64GB of RAM using Intel Core i7-3930K processor and operating with a clock speed of 3.2 GHz. MATLAB 2012b environment has been used for programming.

Electrospray ionization-mass spectrometry (ESI-MS) measurements operating in a positive ion mode were performed on a HPLC-MS triple quadrupole instrument equipped with an autosampler (Agilent Technologies, Palo Alto, CA, USA). The desolvation gas temperature was 300°C with flow rate of 8.0 L/min. The fragmentor voltage was 135 V and capillary voltage was 4.0 kV. Mass spectra were recorded in m/z segment of 10-2000. All data acquisition and processing was performed using Agilent MassHunter software. Acquired mass spectra are composed of intensities at T=9901 m/z coordinates.

Numerical experiment

Nonlinear uBSS problem characterized by $N=3, M=8, K=3$ and $T=1000$ is simulated:

$$f_1(\mathbf{s}) = s_1^3 + s_2^2 + \tan^{-1}(s_3) + s_4^2 + s_5^3 + s_6^3 + \tanh(s_7) + \sin(s_8)$$

$$f_2(\mathbf{s}) = \tanh(s_1) + s_2^3 + s_3^3 + \tan^{-1}(s_4) + \tanh(s_5) + \sin(s_6) + s_7^2 + s_8^2$$

$$f_3(\mathbf{s}) = \sin(s_1) + \tan^{-1}(s_2) + s_3^2 + s_4^3 + \tanh(s_5) + \sin(s_6) + s_7^3 + \tan^{-1}(s_8)$$

Each source signal is according to p.d.f. based on mixed state random variable model with exponential prior [5]:

$$p(s_{mt}) = \rho \delta(s_{mt}) + (1 - \rho) \delta^*(s_{mt}) f(s_{mt}) \quad \forall m = 1, \dots, M \quad \forall t = 1, \dots, T$$

$$f(s_{mt}) = (1/\mu_m) \exp(-s_{mt}/\mu_m)$$

where $\rho_m=0.8$ and $\mu_m=1.5 \times 10^{-3} \quad \forall m=1, \dots, M.$



Numerical experiment

Comparative performance analysis of NMU, NMF_L0, EKM-NMU, EKM-NMF_L0, PTs-EKM-NMU and PTs-EKM-NMF_L0 algorithms. Probability of zero state was $\rho_m=0.8$.

Four metrics used in comparative performance analysis were: number of associated components with normalized correlation coefficient greater than or equal to 0.6, mean value of correlation coefficient over all associated components, minimal value of correlation coefficient and number of pure components assigned incorrectly (that occurs due to poor separation).

All four metrics were calculated with respect to predefined labeling of the pure components stored in the library. Incorrect assignment implies that, based on maximal correlation criterion, two or more pure components are assigned to the same separated component.

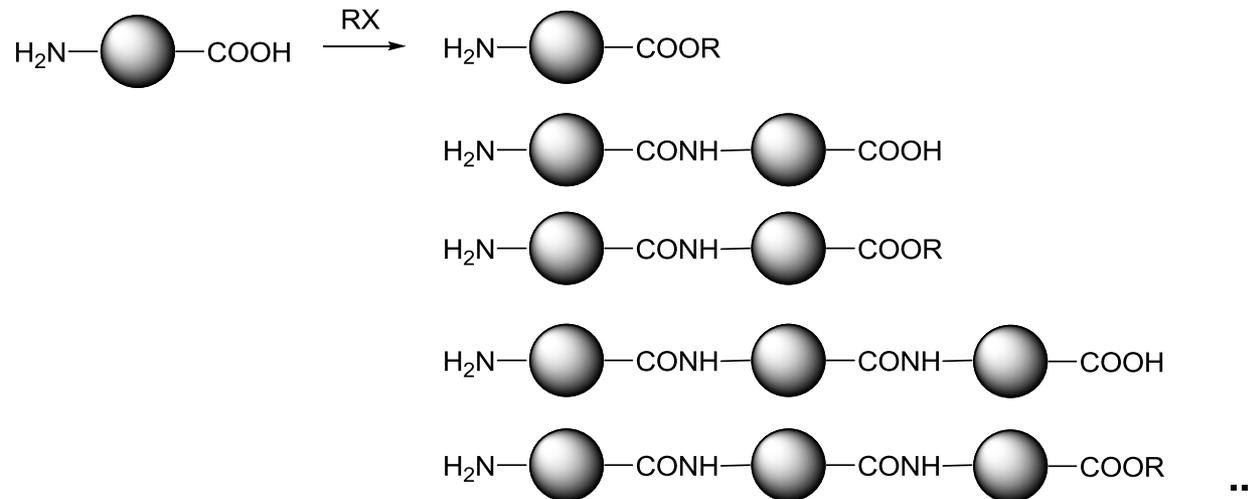
Mean values and variance are reported and estimated over 10 Monte Carlo runs. The best result in each metric is in bold. The first three metrics are calculated only for correctly assigned components. That is why NMU and NMF_L0 appear to have comparable performance.

Numerical experiment

	NMU	NMF_L0	EKM-NMU	EKM-NMF_L0	PTs_EKM-NMU	PTs-EKM-NMF_L0
correlation G.E. 0.6	2.8±0.92	2.3±1.34	3.7±0.48	3.2±0.63	3.8±0.42	3.7±0.48
mean correllation	0.70±0.03	0.61±0.11	0.69±0.02	0.64±0.03	0.70±0.03	0.69±0.04
minimal correlation	0.53±0.04	0.42±0.08	0.51±0.03	0.45±0.04	0.52±.04	0.49±0.06
inccorect assignments	3.4±0.70	3.1±0.57	2.4±0.97	2.2±0.63	2.0±0.88	1.5±1.43

Nonlinear chemical reaction

9 nonlinear mixtures mass spectra were recorded in nonlinear chemical reaction related to peptide bond synthesis.



25 pure components were present in the mixtures. They were separated chromatographically which enabled formation of pure components library and validation of the algorithms' performances.

	s_2	s_6	s_7	s_9	s_{10}	s_{12}
s_1	0.9839	0.1416	0.1218	0.1796	0.1072	0.3343
s_2	s_6	s_7	s_9	s_{10}	s_{12}	
	0.1418	0.1268	0.1797	0.1075	0.3305	
s_3	s_{16}	s_{17}	s_{18}			
	0.3575	0.3103	0.1716			
s_4	s_6	s_{19}	s_{21}			
	0.3077	0.3947	0.4005			
s_5	s_7					
	0.7824					
s_7	s_9					
	0.3297					
s_8	s_{13}					
	0.1293					
s_{11}	s_{12}	s_{22}				
	0.2666	0.1622				
s_{14}	s_{17}					
	0.1024					
s_{15}	s_{22}					
	0.1349					
s_{16}	s_{17}					
	0.9783					
s_{17}	s_{18}					
	0.1186					
s_{19}	s_{21}					
	0.9962					
s_{23}	s_{24}	s_{25}				
	0.4409	0.4339				
s_{24}	s_{25}					
	0.3008					

Pure components correlation matrix. 30 pairs of pure components have correlation greater than or equal to 0.1.

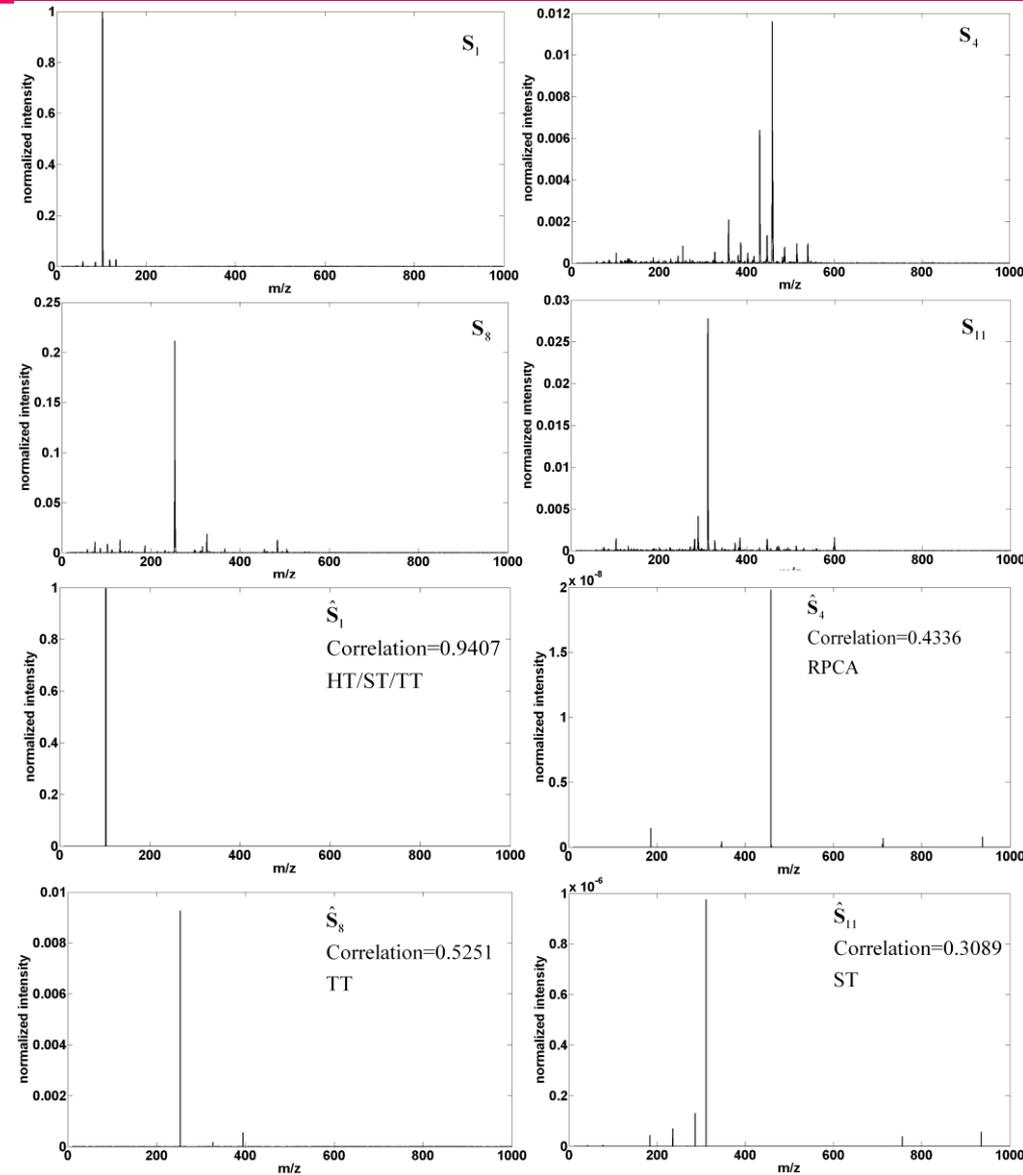
Nonlinear chemical reaction

	NMU	NMF_L0	EKM-NMU	PTs_EKM-NMU <i>D=T=9901</i>	PTs-EKM-NMU <i>D=4000</i>
correlation G.E. 0.6	8	14	16	18	18
mean correlation	0.342	0.518	0.673	0.702	0.708
minimal correlation	0.038	0.039	0.267	0.419	0.283
incorrect assignments	15	7	0	0	1
CPU time	1.3s	40 s	78.78h	4×78h*	4×13.7h*

* Sparseness constrained NMF had to be executed 4 times because 4 methods for suppression of *HOT* have been applied to $\psi(\mathbf{X})$: hard, soft and trimmed thresholding as well as robust PCA



Mass spectra of several true and estimated pure components.



Decomposition (segmentation) of multichannel (RGB) images composed of spectrally (highly) similar objects [19, 32]

19. I. Kopriva, M. Popović Hadžija, M. Hadžija, G. Aralica (2015). Unsupervised segmentation of low-contrast multi-channel images: discrimination of tissue components in microscopic images of unstained specimens," *Scientific Reports* 5: 11576, DOI: 10.1038/srep11576.
32. A. G. Savić, S. Živković, K. K. Jovanović, L. Duponchel, I. Kopriva (2015). Complete determination of plant tissues and advanced image analysis - study of needles and stamen. *Journal of Chemometrics*, doi: 10.1002/cem/2735

Segmentation of low-contrast images

Segmentation of nontrivial images is considered one of the most difficult tasks in image processing, [33].

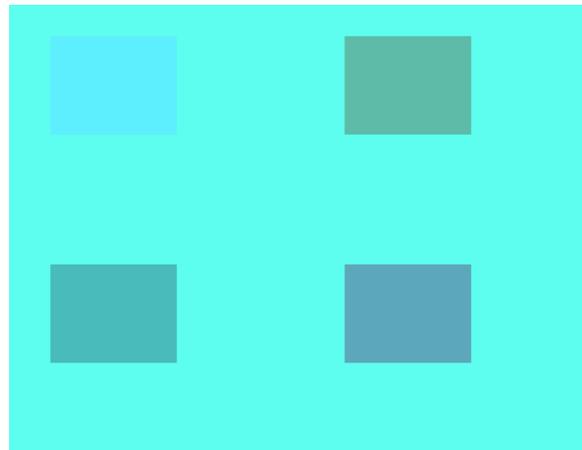
Image segmentation refers to the partitioning of an image into sets of pixels (segments) corresponding to distinct objects, [34]. Herein, distinct objects refer to spectrally distinct tissue components.

It is important to distinguish between single (grayscale)- and multi-channel images. In the former case, segmentation is performed by detection of changes of intensity or texture by thresholding some type of spatial derivative of an image, [35-39].

33. Gonzalez, R.C. & Woods, R.E. in Digital Image Processing (Prentice Hall, 2007).
34. Jain, V., Seung, S.H. & Turaga, S.C. Machines that learn to segment images: a crucial technology for connectomics. *Curr. Opin. Neurobiol.* 20, 653-666 (2010).
35. Marr, D. & Hildreth, E. Theory of edge detection. *Proc. Royal Soc. London Series B Biol. Sci.* 207, 187-217 (1980).
36. Geman, S. & Geman, D. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Mach. Intell.* 6, 721-741 (1984).
37. Boykov, Y., Veksler, O. & Zabih, R. Fast Approximate Energy Minimization via Graph Cuts. *IEEE Trans. Pattern Anal. Mach. Intell.* 23, 1222-1239 (2001).
38. Kass, M., Witkin, A. & Terzopoulos, D. Snakes: Active contour models. *Int. J. Comput. Vis.* 1, 321-331 (1988).
39. Osher, S. & Fedkiw, R.P. Level Set Methods: An Overview and Some Recent Results. *J. Comput. Phys.* 169, 463-502 (2001).

Segmentation of low-contrast images

Images that comprise components with very similar profiles (spectral, density, and/or concentration) have **very poor visual contrast**. For an example, if staining is not used, the spectral similarity between the tissue components present in the specimen is very high and the visual contrast is very poor, i.e., tissue components appear colorless and virtually texture-less when viewed under a light microscope.



Synthetic image: $\mu(\mathbf{A})=0.9995$.



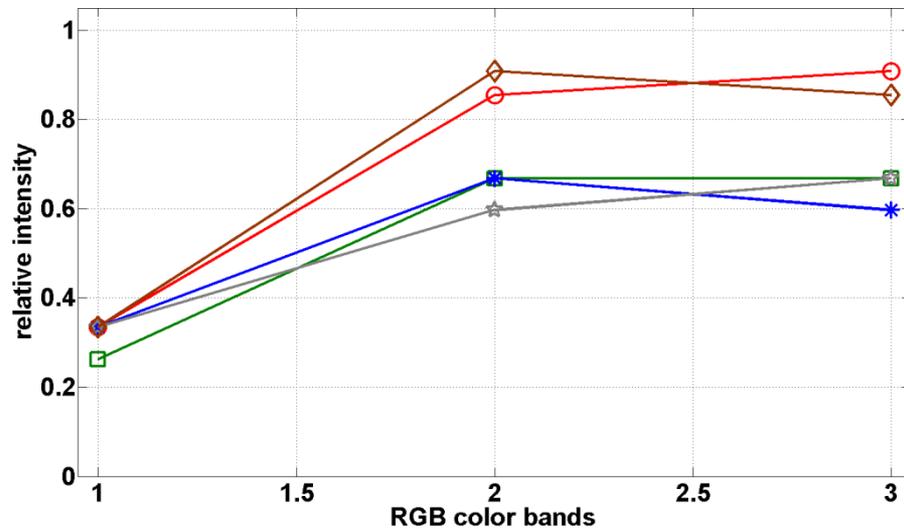
Unstained specimen of human liver with hepatocellular carcinoma: $\mu(\mathbf{A})>0.9999$.



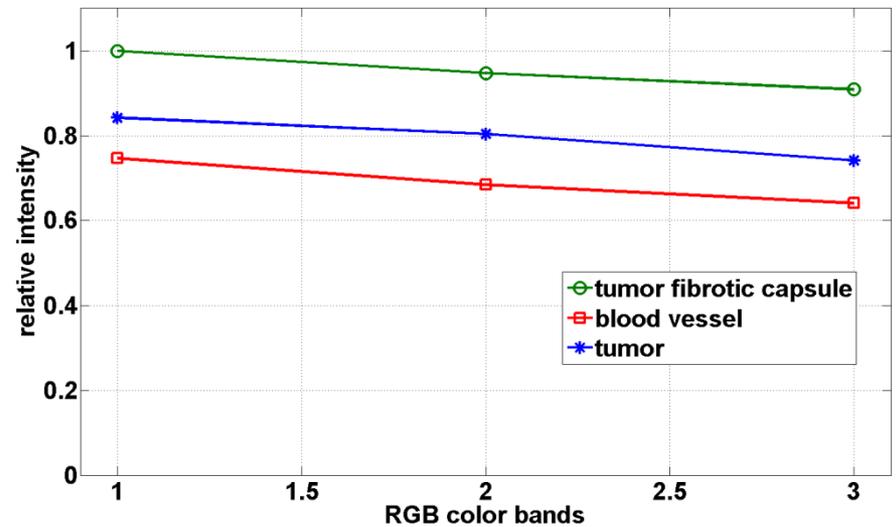
Unstained specimen of human liver with metastasis from colon cancer $\mu(\mathbf{A})>0.9997$.

Segmentation of low-contrast images

When spectral vectors are plotted vs. their indices (corresponding red, green and blue colors) they are virtually parallel.



Synthetic image: $\mu(\mathbf{A})=0.9995$.



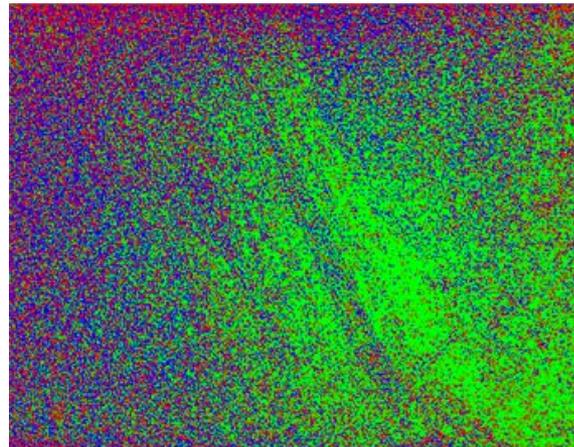
Ustained specimen of human liver with hepatocellular carcinoma: $\mu(\mathbf{A})>0.9999$.

Segmentation of low-contrast images

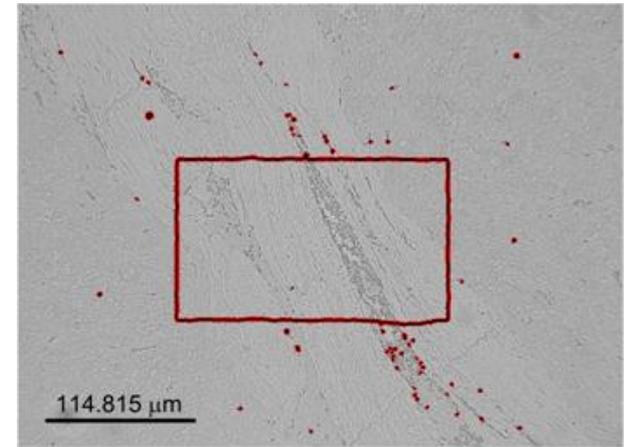
The intensity and/or texture-based segmentation methods, [35-39], fail to segment tissue components correctly. Segmentation of the color image by means of clustering in the CIE L*a*b* color space, [40], also fails for the same reason.



Unstained specimen of human liver with hepatocellular carcinoma: $\mu(\mathbf{A}) > 0.9999$.



K-means in CIE L*a*b* color space



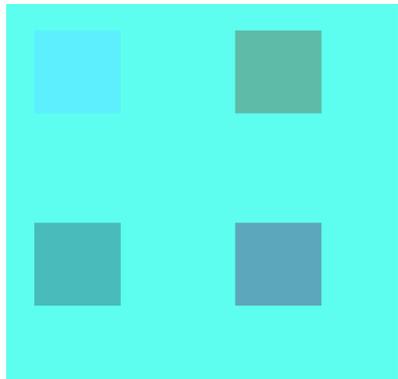
Geometric active contour method after 6000 iterations, [541].

40. Chitade, A.Z. & Katiyar, S.K. Colour Based Image Segmentation Using K-Means Clustering. *Int. J. Eng. Sci. Tech.* 2, 5319-5325 (2010).

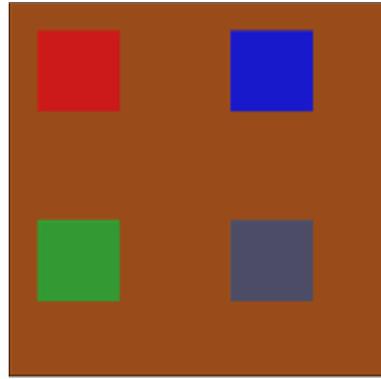
41. Sandhu, R., Georgiu, T., and Tannebaum, A., "A New Distribution Metric for Image Segmentation," in *Proc. SPIE 6914, Medical Imaging 2008: Image Processing*, 691404 (11 March 2008); doi: 10.1117/12.769010.

Segmentation of synthetic image

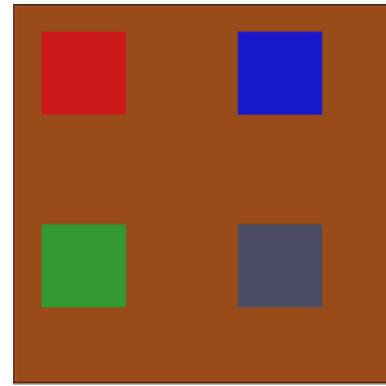
Mapping of the original image \mathbf{X} by using EKM based on Gaussian kernel yields $\Psi(\mathbf{X})=\mathbf{BS}$. Applying NMU [29], resp. NMF_L0 [30], algorithms to $\Psi(\mathbf{X})$ executes image decomposition (segmentation). We name these methods EKM-NMU, resp. EKM-NMF_L0.



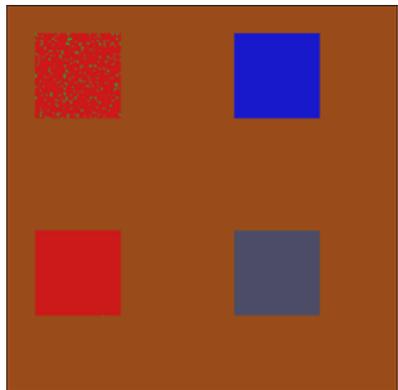
Synthetic image: $\mu(\mathbf{A})=0.9995$.
Per-channel SNR=70 dB.



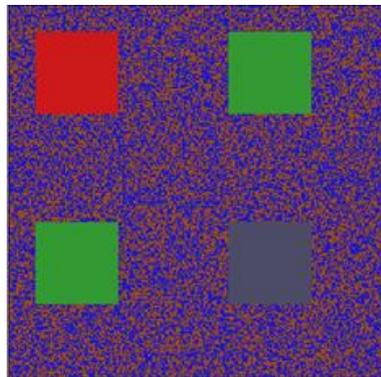
Color coded ground truth.



Color coded EKM-NMU ($D=20$, $\sigma^2=0.01$). $\mu(\mathbf{B})=0.9807$.

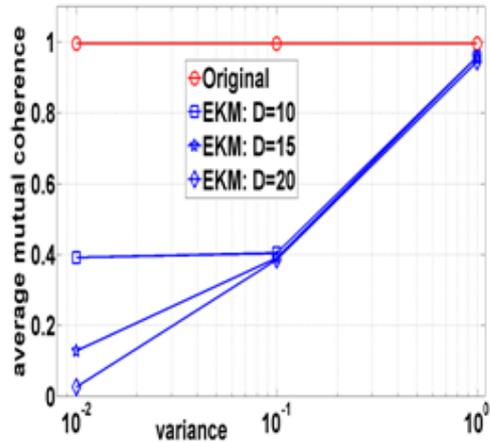


Color coded NMU.

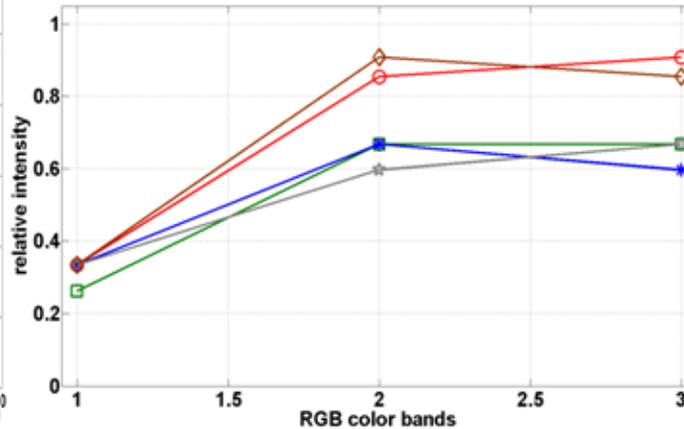


Color coded *K-means* in CIE $L^*a^*b^*$ color space.

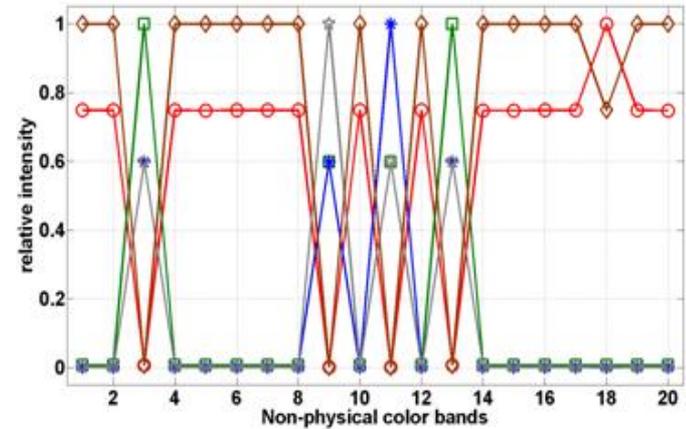
Segmentation of synthetic image



Average coherence for **A** (red) and **B** (blue).



Spectral responses of 5 objects in RGB color space. $\mu(\mathbf{A})=0.9995$.
 $\mu_{average}(\mathbf{A})=0.9956$.

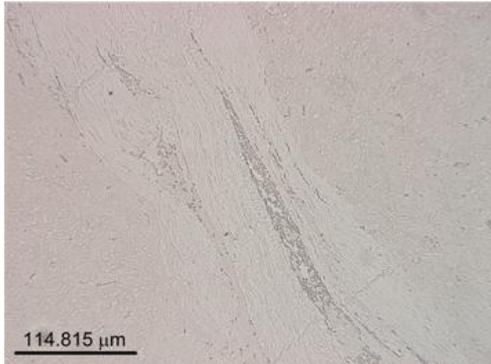


Spectral responses of 5 objects in non-physical color space. $\mu(\mathbf{B})=0.9807$.
 $\mu_{average}(\mathbf{B})=0.3777$.

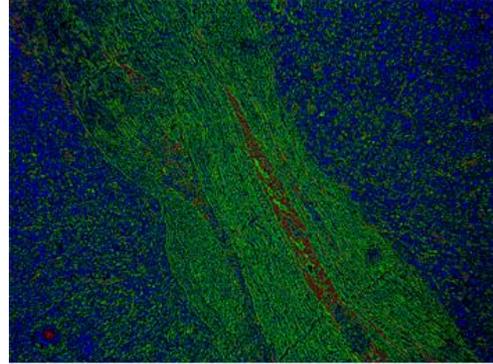
Variance σ^2 of the Gaussian kernel based EKM as a function of the per-spectral-channel SNR.

SNR [dB]	$\text{SNR} \geq 29$	$18 \leq \text{SNR} \leq 28$	$17 \leq \text{SNR} \leq 14$
σ^2	0.001	0.01	0.1

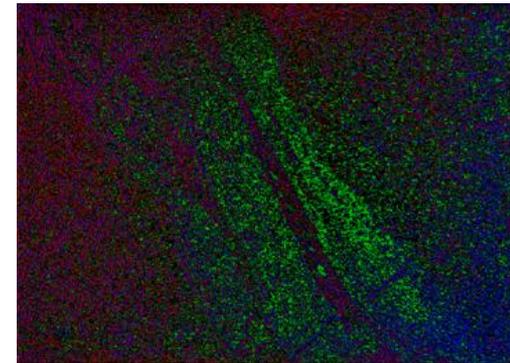
Segmentation of image of unstained specimen of human liver with HCC



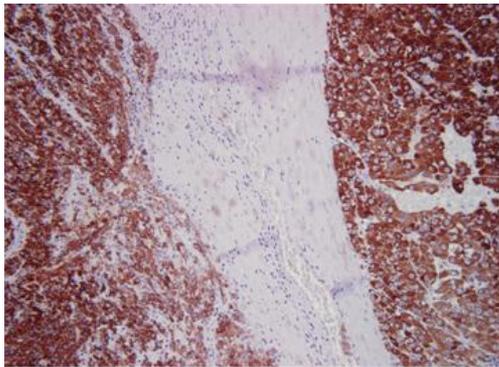
Unstained specimen of human liver with HCC: $\mu(\mathbf{A}) > 0.9999$.



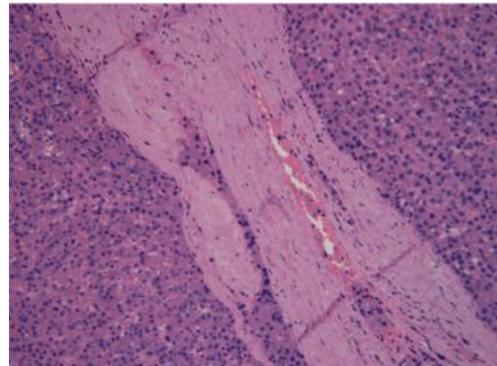
Color-coded EKM-NMF_L0 ($D=50$, $\sigma^2=0.1$). $\mu(\mathbf{B})=0.9760$. Blue: HCC; green: tumor fibrotic capsule; red: blood vessel.



Color-coded *K-means* in CIE L*a*b* color space.

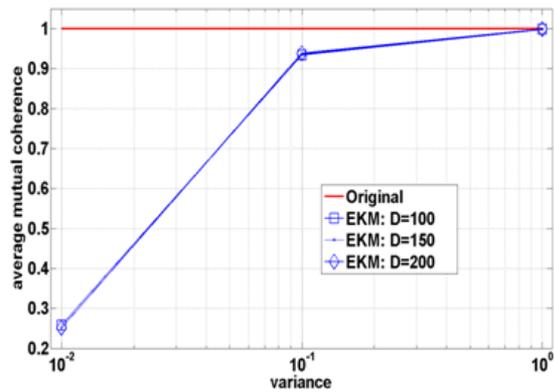


Staining with HepPar (different slide). Brown: hepatocytes, white: tumor fibrotic capsule; blue: endothelium of blood vessel.

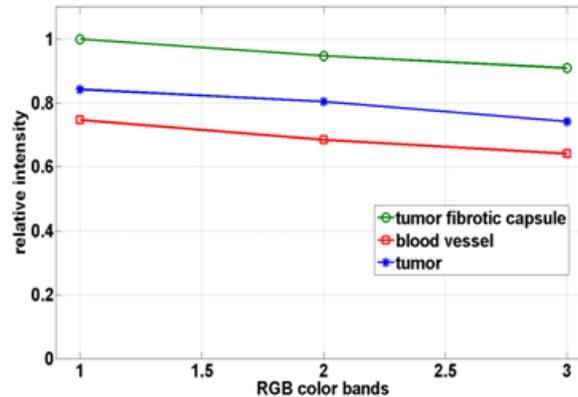


Staining with H&E (the same slide). Blue and dark pink: hepatocytes, light pink: tumor fibrotic capsule; white pink: blood vessel.

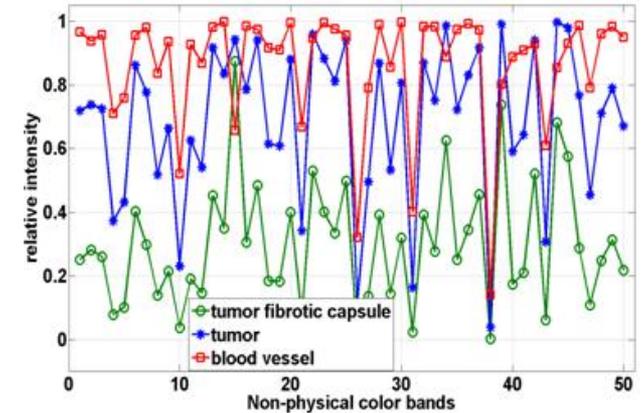
Segmentation of image of unstained specimen of human liver with HCC



Average coherence for **A** (red) and **B** (blue).

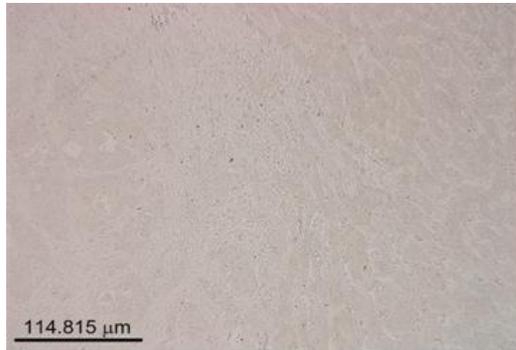


Spectral responses of 3 tissue components in RGB color space.
 $\mu(\mathbf{A}) > 0.9999$. $\mu_{average}(\mathbf{A}) = 0.9985$.

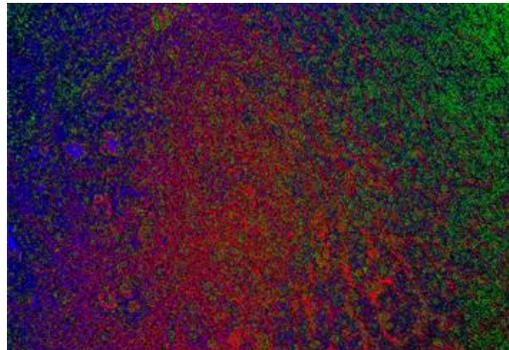


Spectral responses of 3 tissue components in non-physical color space.
 $\mu(\mathbf{B}) = 0.9760$. $\mu_{average}(\mathbf{B}) = 0.8937$.

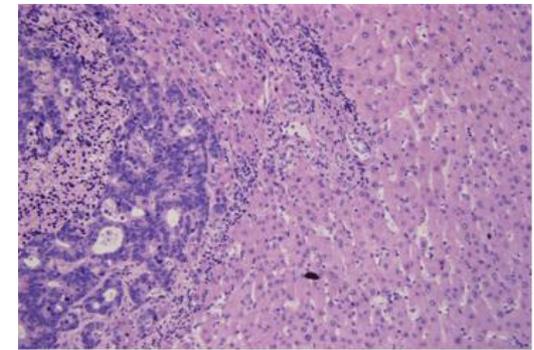
Segmentation of image of unstained specimen of human liver with metastasis from colon cancer



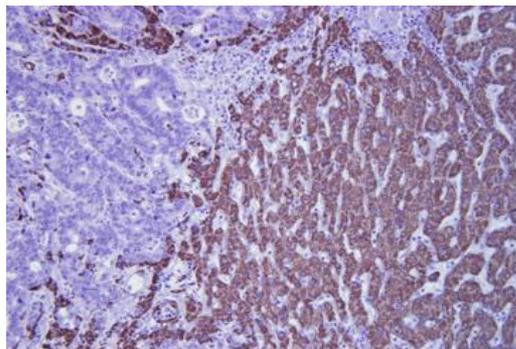
Unstained specimen:
 $\mu(\mathbf{A})=0.9997$,
 $\mu_{average}(\mathbf{A})=0.9993$.



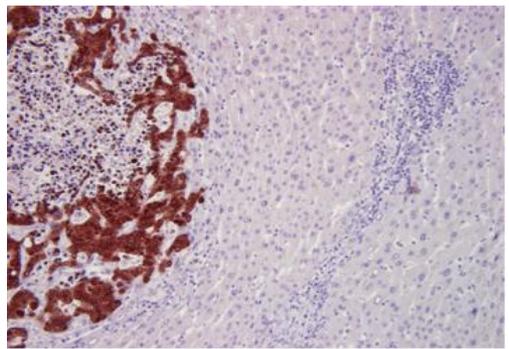
Color-coded EKM-NMF_L0 ($D=50$, $\sigma^2=0.1$).
 $\mu(\mathbf{B})=0.9998$, $\mu_{average}(\mathbf{B})=0.9984$. Blue:
 colon cancer; green: hepatocytes; red: border
 area between the tumor and liver tissue.



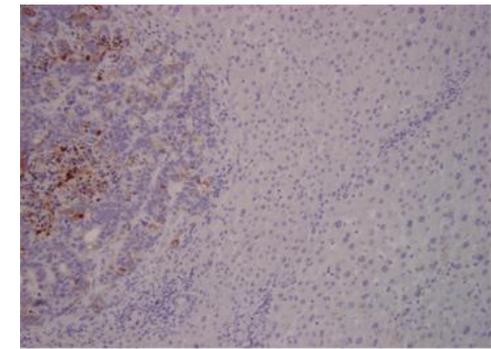
Staining with H&E (the same slide).



Staining with HepPar (different slide). Brown: hepatocytes, blue: metastatic cells of colon cancer and inflammatory cells.

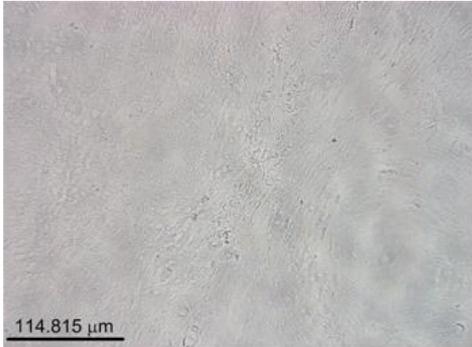


Staining with CDX2 (different slide). Brown: metastatic cells of colon cancer, blue: hepatocytes and inflammatory cells.

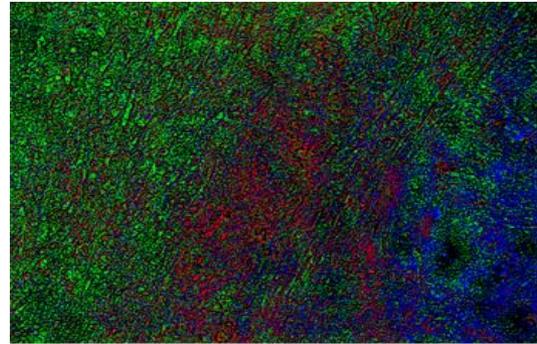


Staining with CK20 (different slide). Brown: metastatic cells of colon cancers, blue: hepatocytes and inflammatory cells.

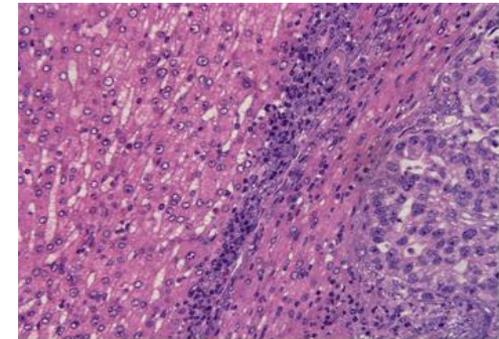
Segmentation of image of unstained specimen of human liver with metastasis from gastric cancer



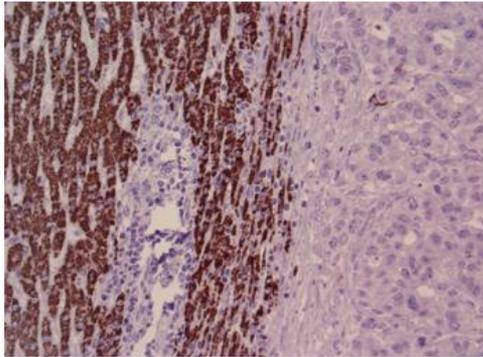
Unstained specimen:
 $\mu(\mathbf{A})=0.9999$,
 $\mu_{average}(\mathbf{A})=0.9988$.



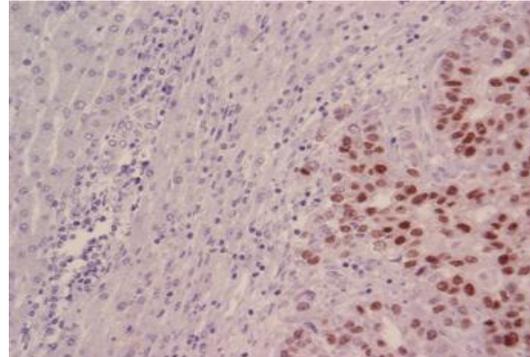
Color-coded EKM-NMF_L0 ($D=50$, $\sigma^2=0.1$).
 $\mu(\mathbf{B})=0.9994$, $\mu_{average}(\mathbf{B})=0.9917$. Blue:
 gastric cancer; green: hepatocytes; red:
 border area of inflammation.



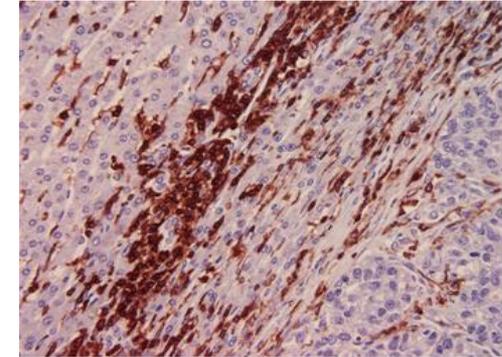
Staining with H&E (the same slide).



Staining with HepPar (different slide). Brown: hepatocytes, blue: metastatic cells of gastric cancer and inflammatory cells.

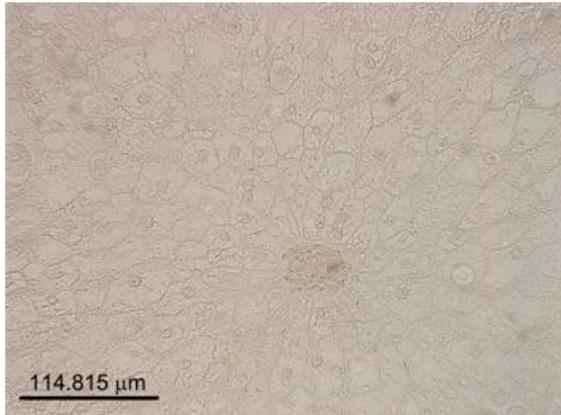


Staining with CDX2 (different slide). Brown: metastatic cells of gastric cancer, blue: hepatocytes and inflammatory cells.

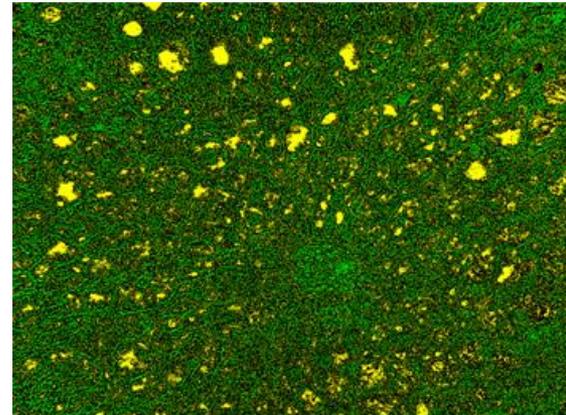


Staining with LCA (different slide). Brown: inflammatory cells; blue: hepatocytes and metastatic cells of gastric cancer.

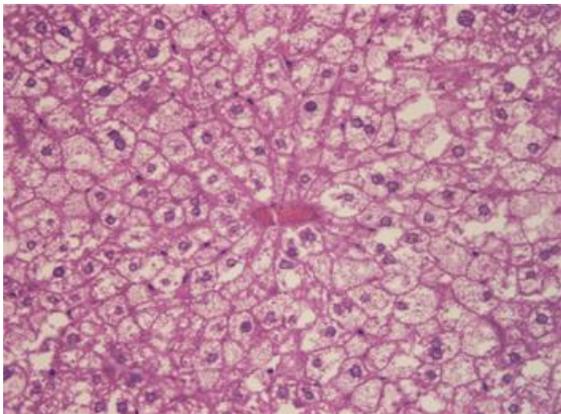
Segmentation of image of unstained specimen of mouse fatty liver



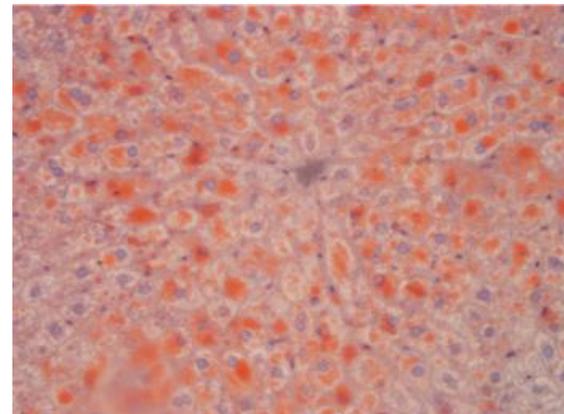
Unstained specimen.



Color-coded EKM-NMF_L0 ($D=50$, $\sigma^2=0.1$).
 Yellow: vacuoles, green: liver parenchyma.

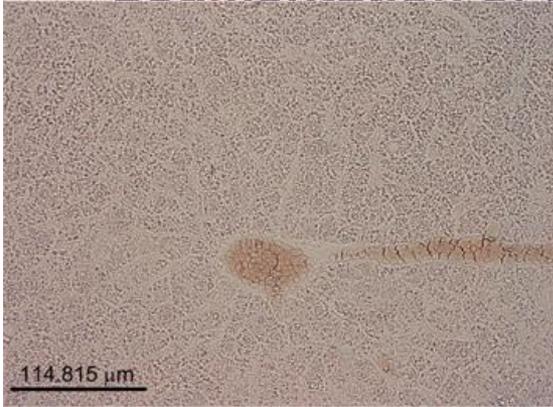


Staining with H&E (the same slide).

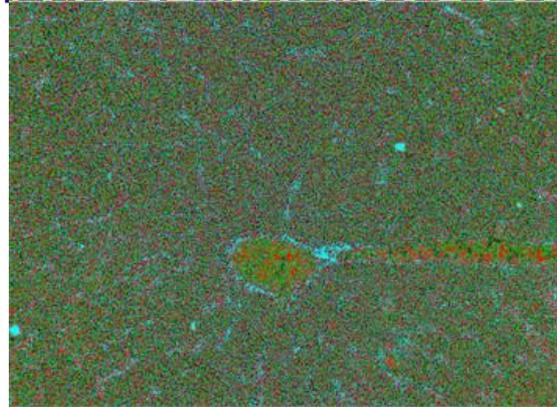


Staining with SUDAN III (different slide).
 Orange: fat storage granules.

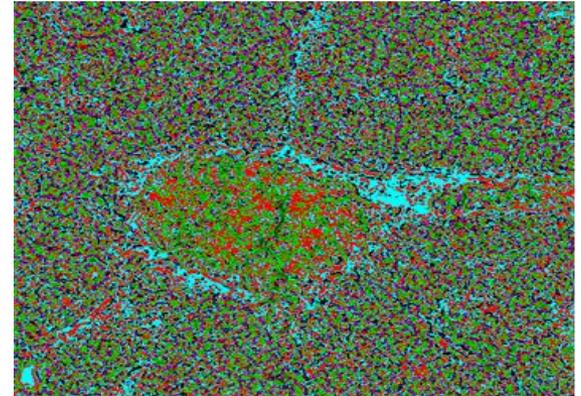
Segmentation of image of unstained specimen of mouse fatty liver



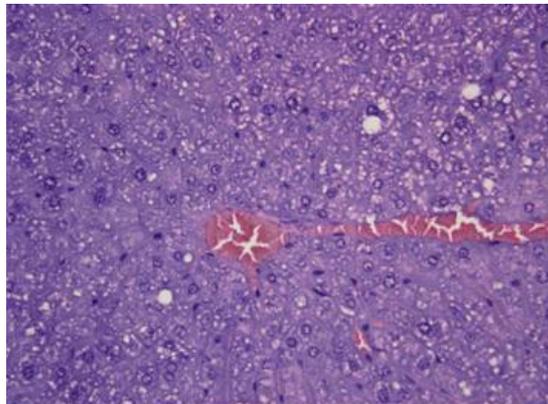
Unstained specimen.



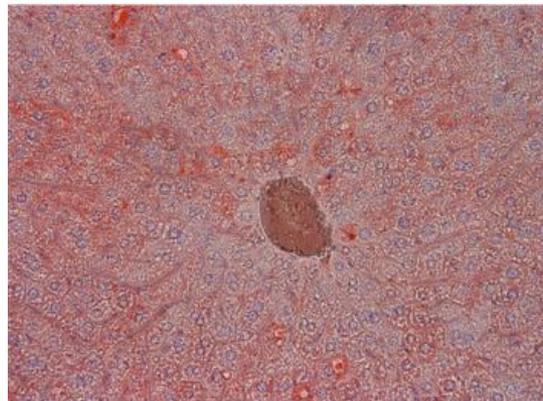
Color-coded EKM-NMF_L0 ($D=50$, $\sigma^2=0.1$).
 Red: blood vessel, sky blue: sinusoids, green: hepatocytes, magenta: reticular fiber.



Zoomed area



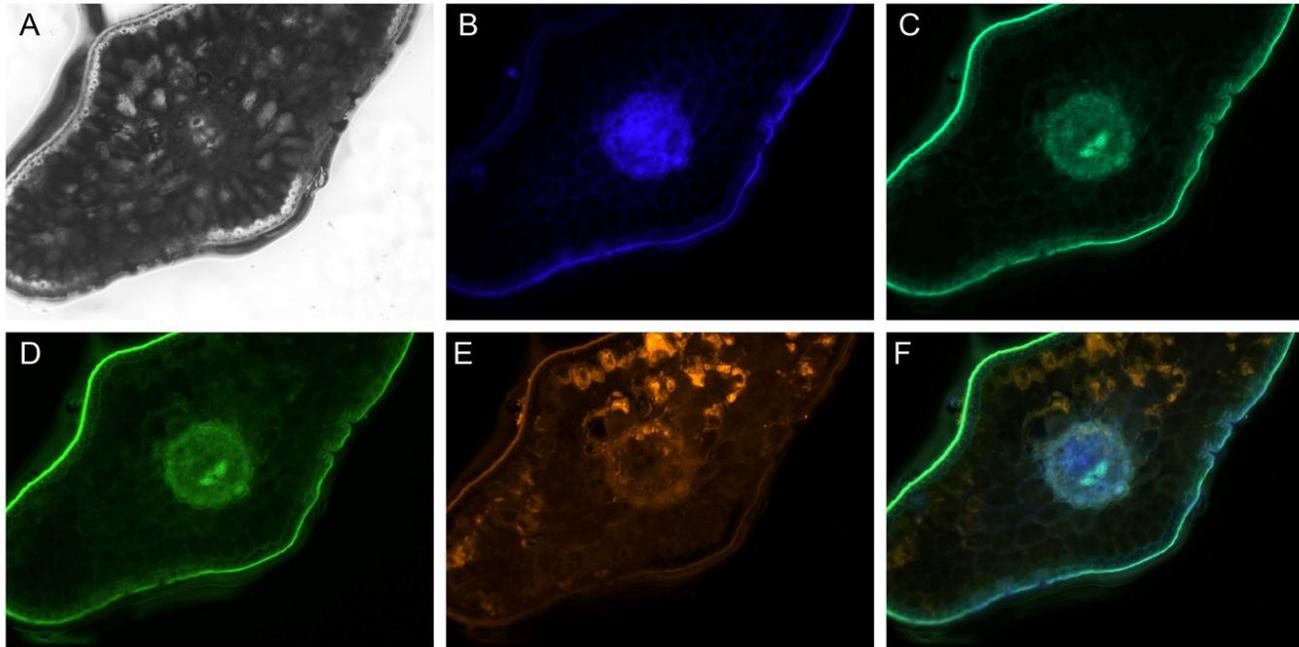
Staining with H&E (the same slide). Brown: hepatocytes, blue: metastatic cells of gastric cancer and inflammatory cells.



Staining with SUDAN III (different slide). Orange: fat storage granules.

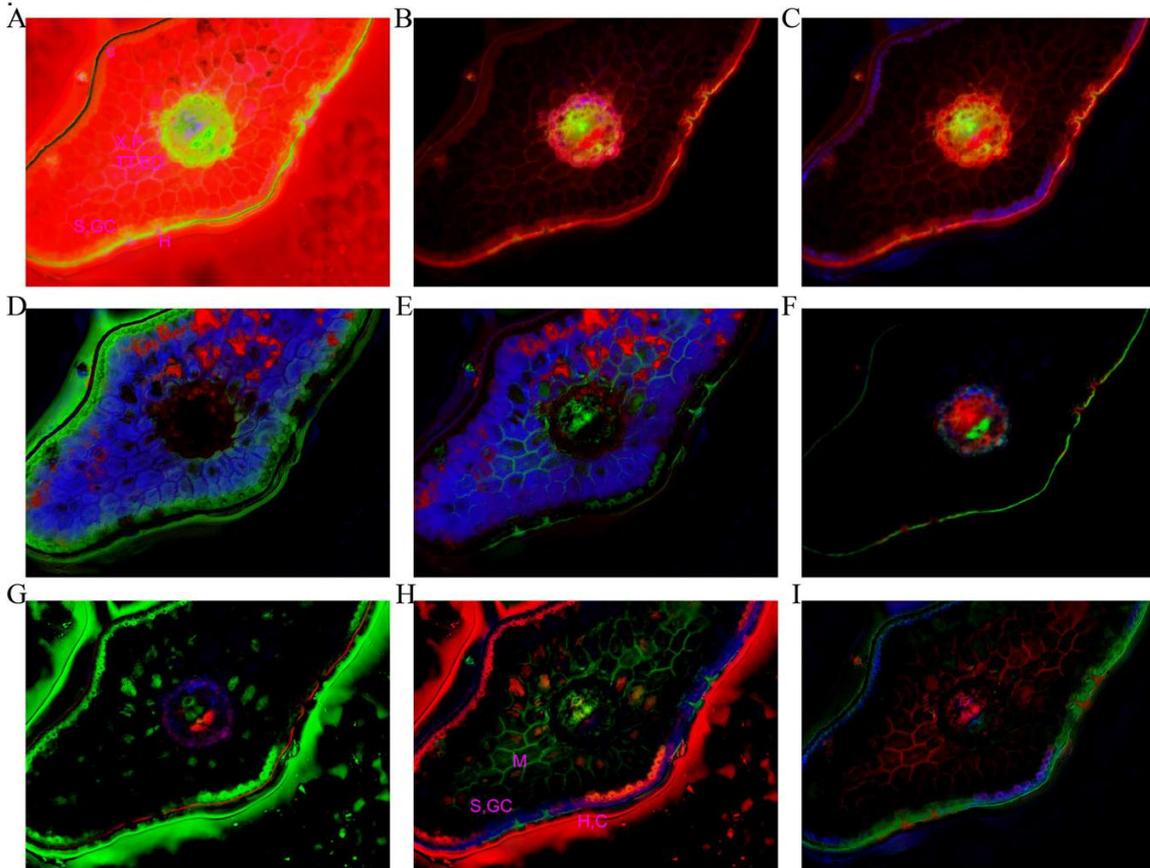
Segmentation of microscopic fluorescent image of plant specimen

Mapping of the original image \mathbf{X} by using EFM based on Gaussian kernel yields $\phi(\mathbf{X})=\mathbf{BS}$. Applying NMF_L0 [30], algorithm to $\phi(\mathbf{X})$ executes image decomposition (segmentation). We named this method EFM-NMF_L0. The method has been demonstrated in segmentation of a cross-section of a needle of *Picea Omorika*, [32].



A – transmission image, B-E – fluorescence images obtained by application of DAPI, FAM, GFP and DsRED filter respectively. F – composite image made using fluorescence images according to CIE convention.

Segmentation of microscopic fluorescent image of plant specimen



11 tissue types were segmented and identified from 5-channel image.

A – Composite RGB images obtained by combining some of the extracted components as channels. Each combination (A-I) uncovers cell types and describes it .

A = C+H+ED+TT+X+P+S+GC, B = C+E+H+M+ED+X+S,

C = C+E+H+M+ED+TT+X+P+S+GC, D = C+H+M+GC, E = C+M+X+S+GC,

F = C+ED+P+TT+S, G = C+H+ED+X+P, H = C+E+H+M+TT+X+S+GC, I = C+E+H+M+TT+X+P+S+GC.

THANK YOU !!!!!!!!