Nonlinear sparse component analysis: low-contrast multichannel image decomposition

Ivica Kopriva

Ruđer Bošković Institute

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

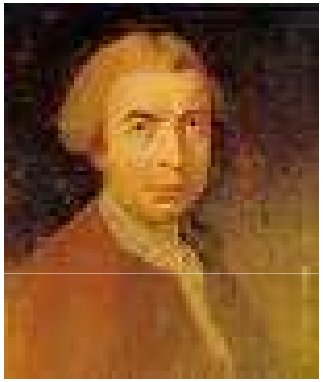
October 29, 2015

Acknowledgment:

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

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).
- Application: unsupervised decomposition of color (RGB) microscopic image of **unstained specimen** in histopathology.

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:

$\mathbf{X}=\mathbf{AS}$ $\mathbf{X}\in R^{NxT}$, $\mathbf{A}\in R^{NxM}$, $\mathbf{S}\in R^{MxT}$

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

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

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

Blind Source Separation – linear static problem

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

Independent component analysis (**ICA**) solves BSS problem provided that: source signals **S** are statistically independent and non-Gaussian; mixing matrix **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 **X**: $L(\mathbf{X})=\mathbf{A}L(\mathbf{S})$. ICA is applied to $L(\mathbf{X})$ to estimate **A** and $L(\mathbf{S})$. **S** is estimated from **S**≈**A**$^{-1}$**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 **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 **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

4. A. Cichocki, R. Zdunek, A. H. Phan, S. Amari, Nonnegative Matrix and Tensor Factorizations-Applications to Exploratory Multi-way Data Analysis and Blind Source Separation, John Wiley, 2009.

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

$\mathbf{x}_t = \mathbf{f}(\mathbf{s}_t) \quad t=1,...,T$; $\mathbf{x}_t \in \mathbb{R}_{0+}^{N\times 1}$ 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:

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

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 .

Linear Underdetermined BSS

- SCA-based solution of the linear uBSS problem is obtained in two stages:

  1) estimate basis or mixing matrix **A** using data clustering.

  2) estimating sources, with estimated **A**, one at a time $\mathbf{s}_t$, t=1,…,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, $\|\mathbf{s}_t\|_p$, $0 \le p \le 1$.

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

- NMF-based solution yields **A** and **S** simulatneously through sparseness and nonnegativity constrained factorization of **X**.

When uBSS problems can(not) be solved?

Let us focus on underdetermined linear system:

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

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

Provided that **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 **s,** [7]:

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

7. 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 **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, [8].In chemometrics **A** contains concentration profiles of pure components present in the mixtures, [9].

One result for deterministic **A** is given in [10]. For cyclic polynomial matrix **A** it applies $\boldsymbol{N=O(K^2)}$. That is significantly worse than $N \approx K\log(M/K)$ for random **A**. *K* correponds 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.

8. Kopriva I, Cichocki A. Blind decomposition of low-dimensional multi-spectral image by sparse component analysis. *J. Chemometrics* 2009; **23** (11): 590-597.
9. 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.
10. 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 **s** certain degree of incoherence of the mixing matrix **A** is required as well. Mutual coherence is defined as the largest absolute and normalized inner product between different columns in **A**, what reads as

$$\mu\{\mathbf{A}\} = \max_{1 \le i,j \le M \text{ and } i \ne j} \frac{\left|\mathbf{a}_i^T \mathbf{a}_j\right|}{\left\|\mathbf{a}_i\right\| \left\|\mathbf{a}_j\right\|}$$

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 **s** relates sparseness requirement on **s** and coherence of **A**, [11,12]:

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

11. R. Gribonval and M. Nielsen, "Sparse representations in unions of bases," *IEEE Transactions on Information Theory* **49**, 3320-3325 (2003).
12. J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Transactions on Information Theory* **50**, 2231-2242 (2004).

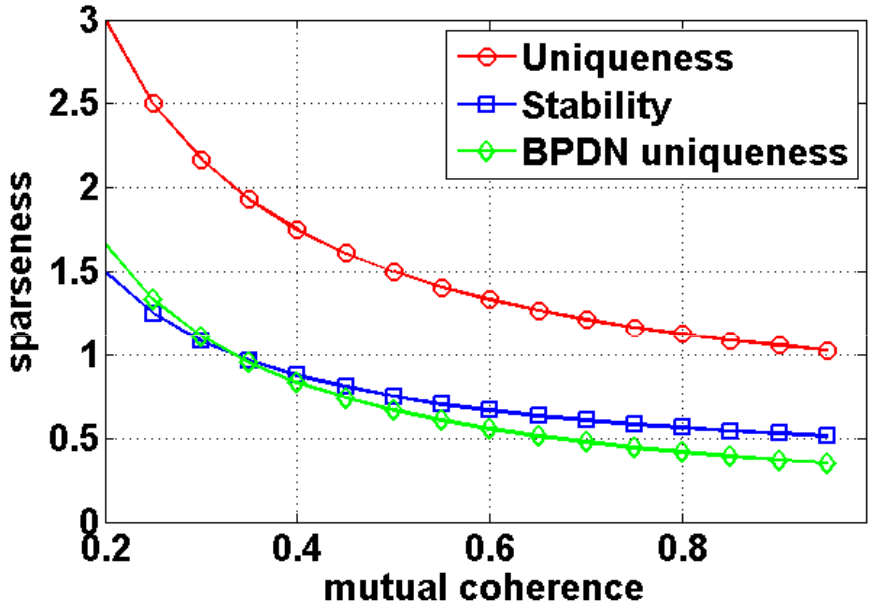
When uBSS problems can(not) be solved?

When the mutual coherence $\mu(\mathbf{A})$ is ***very close to*** 1 possibility to obtain meaningful solution of **x=As** is reduced drastically. Such scenario occurs when, as na example, **X** represent RGB microscopic image of **unstained specimen** in histopathology, [13]. 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 [14, 15], unstable [16, 17].

13. 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.
14. 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).
15. Tibshirani, R., "Regression shrinkage and selection via the Lasso," J. Roy. Statist. Soc. B 58 (1), 267-288 (1996).
16. Chen, S.S., Donoho, D.L., and Saunders, M.A., "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput*. 20, 33-61 (1998).
17. 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 $\|\mathbf{s}_p\|_0 = 1$. Thus, while uniqueness condition **in the absence of noise** is satisfied even when **μ(A)≈1** it is seen that **approximately μ(A)<0.33** is required to satisfy uinqueness condition in the presence of **modelling erros or noise**.

When uBSS problems can(not) be solved?

For scenarios when mutual coherence $\mu(\mathbf{A})\approx 1$ it was proposed in [13,5,6] to transform problem:

$$\mathbf{X}=\mathbf{AS} \text{ or } \mathbf{X}=\mathbf{f(S)}=\mathbf{GS}$$

into:

$$\Psi(\mathbf{X})\approx\mathbf{BS}$$

such that $\mu(\mathbf{B})<\mu(\mathbf{A})$, resp. $\mu(\mathbf{B})<\mu(\mathbf{G})$.

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 .
13. 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.

Nonlinear uBSS [5]

$$\mathbf{x}_t = \mathbf{f}(\mathbf{s}_t) \quad t=1,...,T$$

$$\mathbf{x}_t \in \mathbb{R}_{0+}^{N\times 1}, \mathbf{s}_t \in \mathbb{R}_{0+}^{M\times 1}, \; M > N.$$

$$\mathbf{f} : \mathbb{R}_{0+}^{M} \mapsto \mathbb{R}_{0+}^{N}$$

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

Nonlinear uBSS problem can be expanded into Taylor series around reference point $\mathbf{s}_0$. Without loss of generality let us assume $\mathbf{s}_0=\mathbf{0}_{M\times 1}$. Let us also assume $\mathbf{f}(\mathbf{s}_0)=\mathbf{0}_{N\times 1}$.

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.

Nonlinear uBSS

For non-overlapping binary sources: $s_i s_j = \delta(i-j)$, $i,j = 1,\ldots,M$ we obtain [5]:

$$\mathbf{x} = \mathbf{f}(\mathbf{s}) = \mathbf{G}\mathbf{s}$$

where **G** is a matrix $\mathbf{G} = \sum_{j=1}^{J} (1/j!)\mathbf{G}_{(1)}^{j}$ such that:

$$\left\{ \left[\underline{\mathbf{G}}^{j} \right]_{nm_1 \ldots m_j} = \frac{\partial^j f_n(\mathbf{s})}{\partial s_{m_1} \ldots \partial s_{m_j}} \delta\left(m_1 - \ldots - m_j\right) \right\}_{j=1}^{J}$$

Thus, nonlinear BSS problem **x=f(s)** becomes linear one **x=Gs**.

Nonlinear mapping of linear uBSS problem?

In [6] a new concept was proposed by mapping original uBSS problem **X=AS** nonlinearly into new one:

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

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

The nonlinear mapping has the following algebraic structure:

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

6. Kopriva I, Jerić I, Brkljačić, L. Nonlinear mixtures-wise expansion approach to underdetermined blind separation of nonnegative depedent sources. *J. Chemometrics* 2013; **27**: 189-197.

Nonlinear mapping of linear uBSS problem?

The mapped problem becomes:

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

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

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

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:

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

becomes computationally intractable.

Reproducible kernel Hilbert spaces

Definition 1. A real function $\kappa:\mathbb{R}^N\times\mathbb{R}^N\to\mathbb{R}$ 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 [18]. 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 .

18. Aronszajn, N., "The theory of reproducing kernels," *Trans. of the Amer. Math. Soc*. 68, 337-404 (1950).

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 as **explicit feature map** (EFM) associated with kernel $\kappa(\circ,\mathbf{x})$.

Definition 3. Empirical kernel map (EKM), [19]. 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 $\psi:\mathbb{R}^N\to\mathbb{R}^D$:

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

19. Schölkopf, B., and Smola, A., *Learning with kernels*, MIT Press, 2002, pp. 42-45.

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:

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

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:

Then:

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

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

Nonlinear mapping of linear uBSS problem?

Projection yields:

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

When $\phi(\mathbf{x})=k(\circ,\mathbf{x})$ it follows: $\langle\phi(\mathbf{v}), \phi(\mathbf{x})\rangle=k(\mathbf{v},\mathbf{x})$. It is shown in [13] that for non-overlapped binary sources non-overlapping binary sources: $s_i s_j=\delta(i-j)$, i,j = 1,…,M :

$$\psi(\mathbf{X})_{\mathbf{V}} = \mathbf{BS}$$

that is, the mapping $\Psi(\mathbf{X})_{\mathbf{V}}$ is **S**-invariant. Hence, by parameters of the mapping (kernel function $\kappa(\mathbf{v}_d, \mathbf{x}_t)$ and basis **V**) it is possible to tune $\mu(\mathbf{B})$.

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:

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

Thus, **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.

When in addition to **sparseness** constraint **nonnegativity** constraint applies as well (that is the case in applications in imaging) **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):

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

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 **A** and **S**, to narrow down number of possible decomposition of **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, [20], where it is proposed to estimate **A** and **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,…, until convergence* ***do***

Step 1: $$\mathbf{S}^{(k+1)} = \arg\min_{s_{mt}\geq 0} D_{\mathbf{S}}\left(\mathbf{X}\middle\|\mathbf{A}^{(k)}\mathbf{S}\right)_{\mathbf{S}^{(k)}}$$

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

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

20. D D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature* **401** (6755), 788-791 (1999).

Nonnegative matrix factorization

ALS-based NMF:

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

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

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

- In many instances non-negativity constraints imposed on **A** and **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 [20]:

$$D\left(\mathbf{X}\|\mathbf{AS}\right)=\frac{1}{2}\|\mathbf{X}-\mathbf{AS}\|_F^2+\alpha_{\mathbf{S}}J_{\mathbf{S}}(\mathbf{S})+\alpha_{\mathbf{A}}J_{\mathbf{A}}(\mathbf{A})$$

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

$$\frac{\partial D(\mathbf{A},\mathbf{S})}{\partial a_{nm}}=\left[-\mathbf{X}\mathbf{S}^{\mathrm{T}}+\mathbf{A}\mathbf{S}\mathbf{S}^{\mathrm{T}}\right]_{nm}+\alpha_{\mathbf{A}}\frac{\partial J_{\mathbf{A}}(\mathbf{A})}{\partial a_{nm}}$$

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

20. A. Cichocki, R. Zdunek, and S. Amari, "Csiszár's Divergences for Non-negative Matrix Factorization: Family of New Algorithms," *LNCS* **3889,** 32-39 (2006).

Nonnegative matrix factorization

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

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

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:

$$\mathbf{A} \leftarrow \mathbf{A} \otimes \frac{\left[\mathbf{X}\mathbf{S}^{\mathrm{T}} - \alpha_{\mathbf{A}} \dfrac{\partial J_{\mathbf{A}}(\mathbf{A})}{\partial \mathbf{A}}\right]_{+}}{\mathbf{A}\mathbf{S}\mathbf{S}^{\mathrm{T}} + \varepsilon \mathbf{1}_{NM}} \qquad \mathbf{S} \leftarrow \mathbf{S} \otimes \frac{\left[\mathbf{A}^{\mathrm{T}}\mathbf{X} - \alpha_{\mathbf{S}} \dfrac{\partial J_{\mathbf{S}}(\mathbf{S})}{\partial \mathbf{S}}\right]_{+}}{\mathbf{A}^{\mathrm{T}}\mathbf{A}\mathbf{S} + \varepsilon \mathbf{1}_{MT}}$$

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.:

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

Non-negative matrix under-approximation (NMU)

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

A sequential approach to NMF has been recently proposed in [21] 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:

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

21. N. Gillis, and F. Glineur, "Using underapproximations for sparse nonnegative matrix factorization," *Patt. Recog.*, vol. 43, pp. 1676-1687, 2010.

Non-negative matrix under-approximation (NMU)

Theorem 1 in [21] proves that number of nonzero entries in **A** and **S** is less than in **X**. Thus, the underapproximation constraint ensures sparse (parts based) factorization of **X**. This, however, does not imply that **A** and **S** obtained by enforcing underapproximation constrain yields the sparseset decomposition of **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, [22], imposes explicit L_0-constraint on entries of **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, [23], 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. For non-overlaped sources *K*=1.

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

22. R. Peharz, F. Pernkopf, "Sparse nonnegative matrix factorization with ℓ^0 constraints," *Neurocomputing*, vol. 80, pp. 38-46, 2012.
23. C. Lawson, R. Hanson, *Solving Least Squares Problems*, Prentice-Hall, 1974.

Decomposition (segmentation) of multichannel (RGB) images composed of spectrally (highly) similar objects [13]

13. 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.

Segmentation of low-contrast images

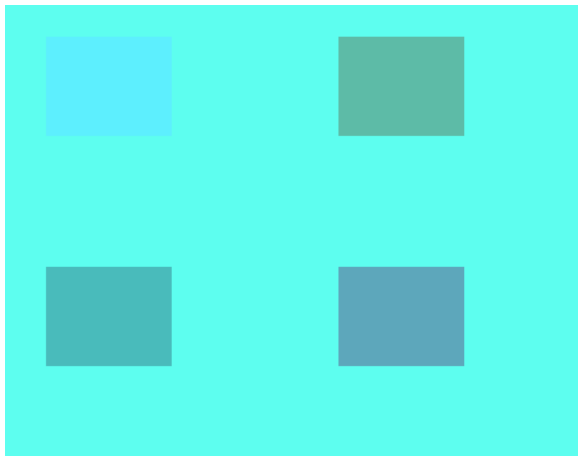
Image segmentation refers to the partitioning of an image into sets of pixels (segments) corresponding to distinct objects, [24]. Herein, distinct objects refer to spectrally distinct tissue components.

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, [25-29].

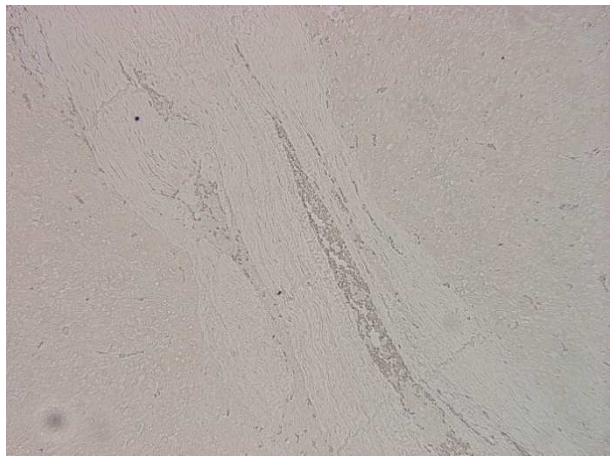
24. Jain, V., Seung, S.H. & Turaga, S.C. Machines that learn to segment images: a crucial technology for connectomics. *Curr. Opin. Neurobiol*. 20, 653-666 (2010).
25. Marr, D. & and Hildredth, E. Theory of edge detecion. *Proc. Royal Soc. London Series B Biol. Sci*. 207, 187-217 (1980).
26. Geman, S. & Geman, D. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Mach. Intell*. 6, 721-741 (1984).
27. Boykov, Y., Veksler, O. & Zabih, R. Fast Approximate Energy Minimization via Graph Cuts. *IEEE Trans. Pattern Anal. Mach. Intell*. 23, 1222-1239 (2001).
28. Kass, M., Witkin, A. & Terzopoulos, D. Snakes: Active contour models. *Int. J. Comput. Vis*. 1, 321-331 (1988).
29. 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: $\mu(\mathbf{A})=0.9995$.

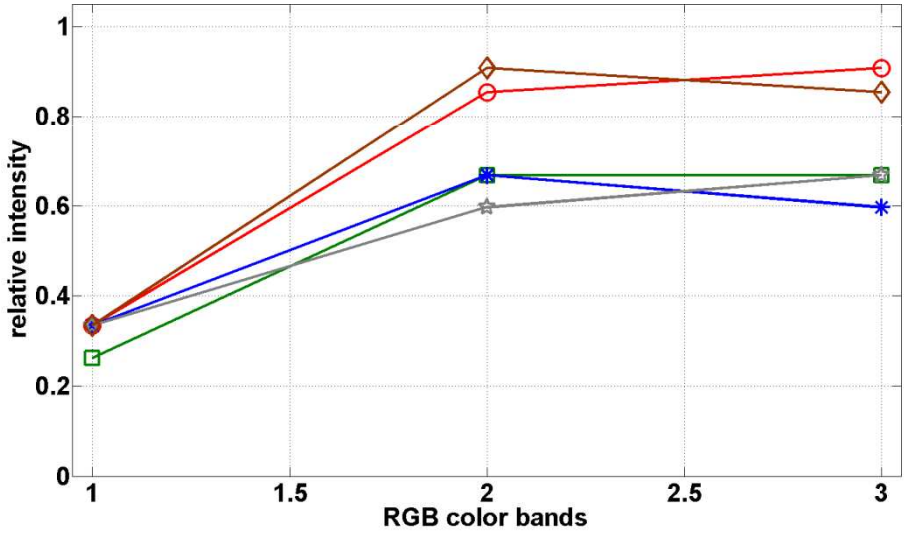


Unstained specimen of human liver with hepatocellular carcinoma: $\mu(\mathbf{A})>0.9999$.

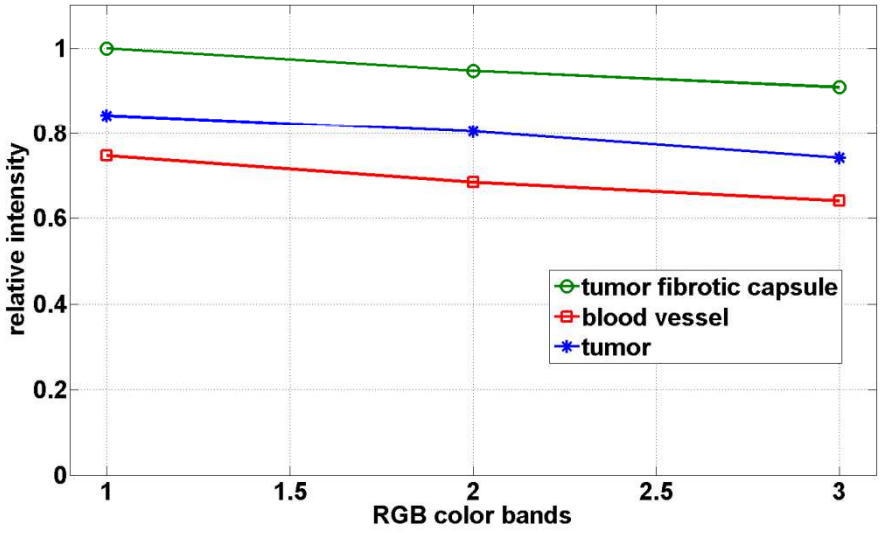
Unstained specimen of human liver with metastasis from colon cancer $\mu(\mathbf{A})>0.9997$.

Segmentation of low-contrast images

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



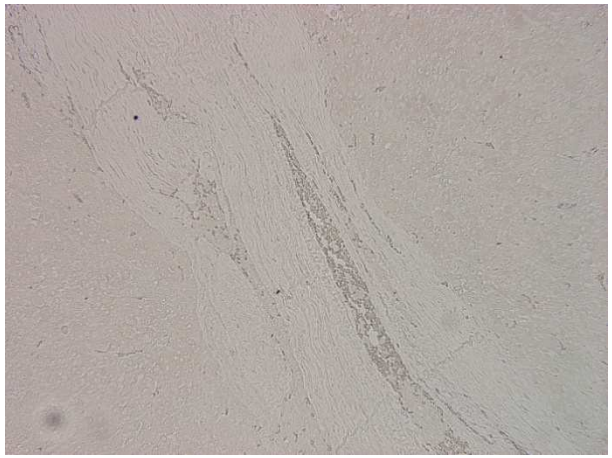
Synthetic image: $\mu(\mathbf{A})=0.9995$.



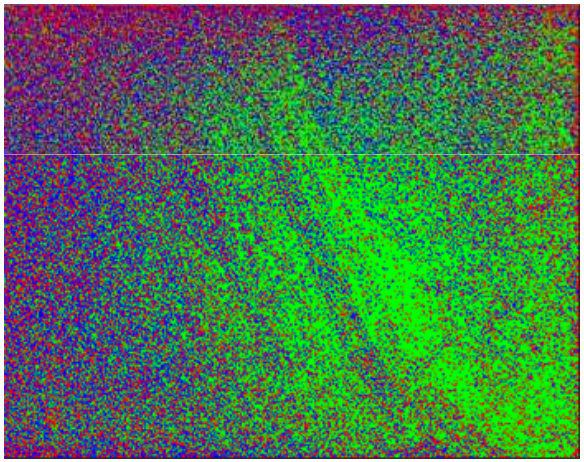
Unstained specimen of human liver with hepatocellular carcinoma: $\mu(\mathbf{A})>0.9999$.

Segmentation of low-contrast images

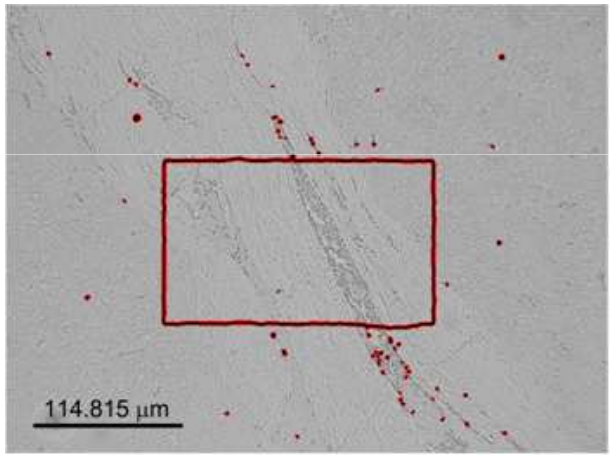
The intensity and/or texture-based segmentation methods, [28, 29], fail to segment tissue components correctly. Segmentation of the color image by means of clustering in the CIE $L^*a^*b^*$ color space, [30], also fails for the same reason.



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



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

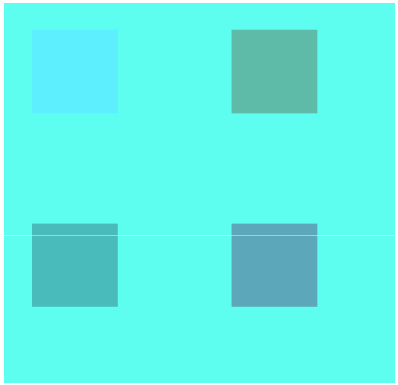


Geometric active contour method after 6000 iterations, [31].

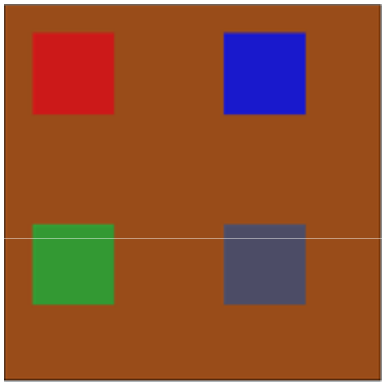
30.Chitade, A.Z. & Katiyar, S.K. Colour Based Image Segmentation Using K-Means Clustering. *Int. J. Eng. Sci. Tech*. 2, 5319-5325 (2010).
31. Sandhu, R., Georgiu, T., and Tannebaum, A., "A New Distribution Metric for Image Segmentation," *in Proc. SPIE 6914, Medical Imaging 2008: Image Processing*, 691404 (11 March 2008); doi: 10.1117/12.769010.

Segmentation of synthtetic image

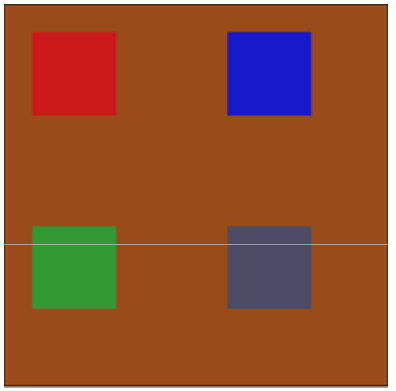
Mapping of the original image **X** by using EKM based on Gaussian kernel yields $\Psi(\mathbf{X})=\mathbf{BS}$. Applying NMU [21], resp. NMF_L0 [22], algorithms to $\Psi(\mathbf{X})$ executes image decomposition (segmentation). We name these methods EKM-NMU, resp. EKM-NMF_L0.



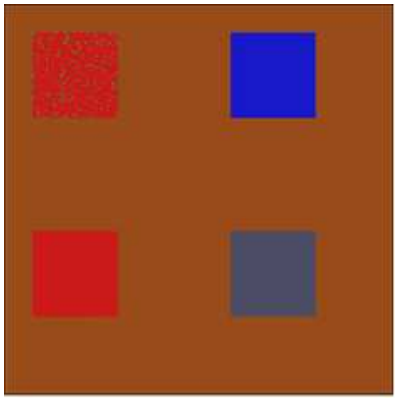
Synthetic image: $\mu(\mathbf{A})=0.9995$.
Per-channel SNR=70 dB.



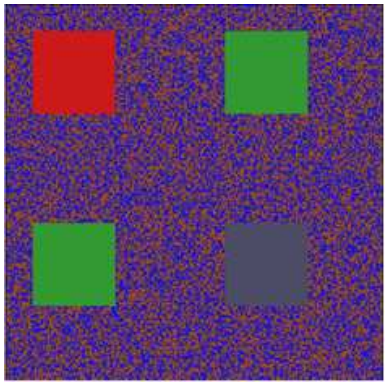
Color coded ground truth.



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

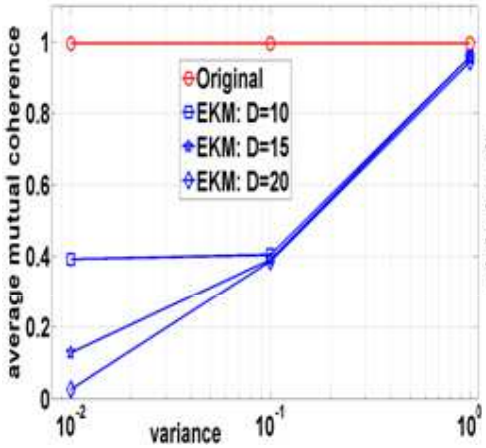


Color coded NMU.

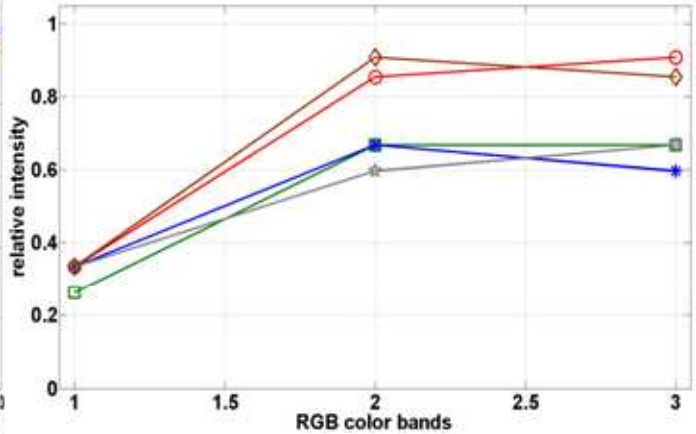


Color coded *K-means* in CIE $L^*a^*b^*$ color space.

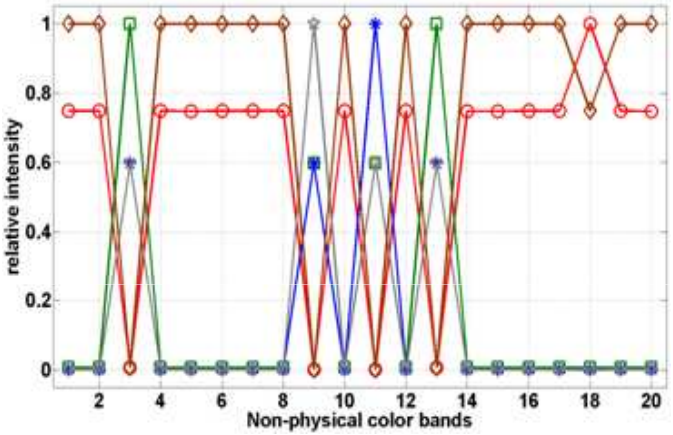
Segmentation of synthetic image



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



Spectral responses of 5 objects in RGB color space. μ(**A**)=0.9995. $\mu\_average$(**A**)=0.9956.

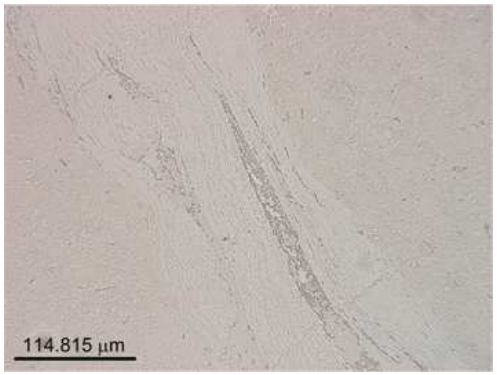


Spectral responses of 5 objects in non-physical color space. μ(**B**)=0.9807. $\mu\_average$(**B**)=0.3777.

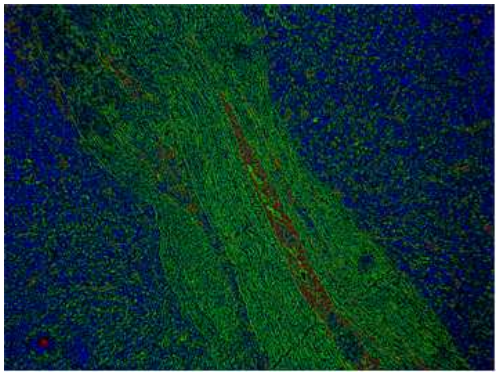
Variance σ^2 of the Gaussian kernel based EKM as a function of the per-spectral-channel SNR.

SNR [dB]	SNR≥ 29	18≤SNR≤28	17≤SNR≤14
σ^2	0.001	0.01	0.1

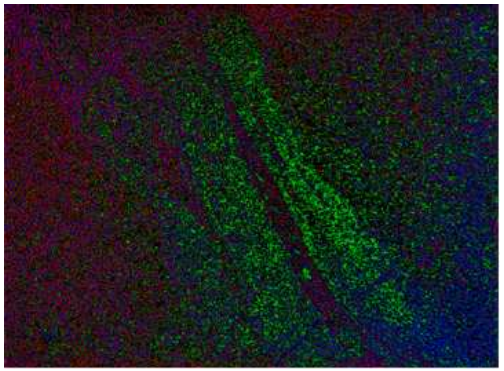
Segmentation of image of unstained specimen of human liver with HCC



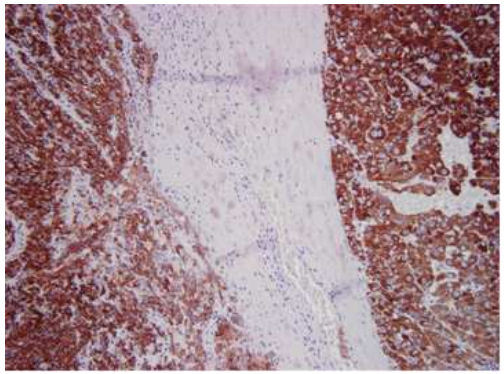
Ustained specimen of human liver with HCC: μ(**A**)>0.9999.



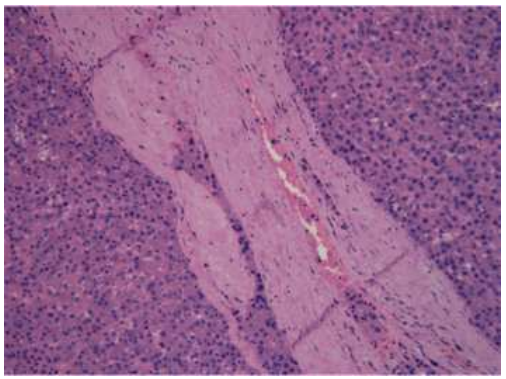
Color-coded EKM-NMF_L0 (D=50, σ^2=0.1). μ(**B**)=0.9760. Blue: HCC; green: tumor fibrotic capsule; red: blood vessel.



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

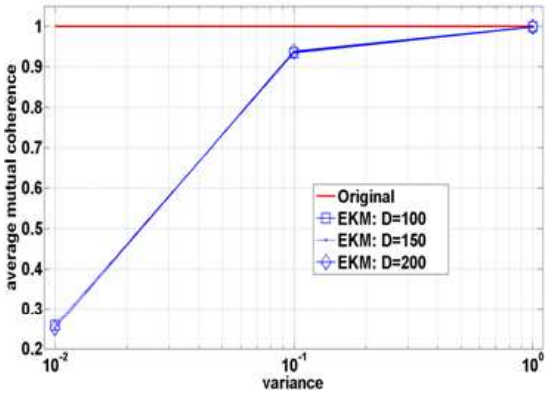


Staining with HepPar (different slide). Brown: hepatocytes, white: tumor fibrotic capsule; blue: enodthelium of blood vessel.

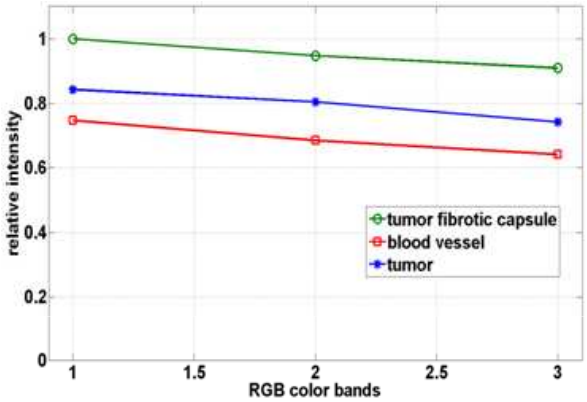


Staining with H&E (the same slide). Blue and dark pink: hepatocytes, light pink: tumor fibrotic capsule; white pink: blood vessel.

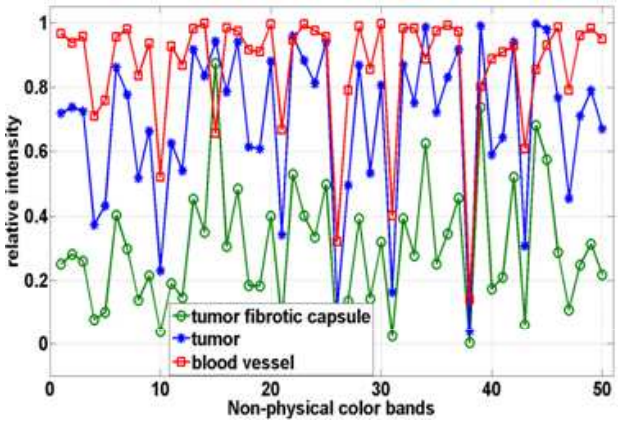
Segmentation of image of unstained specimen of human liver with HCC



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

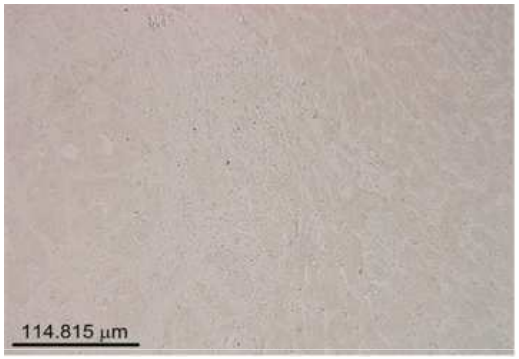


Spectral responses of 3 tissue components in RGB color space. μ(**A**)>0.9999. $\mu\_average$(**A**)=0.9985.

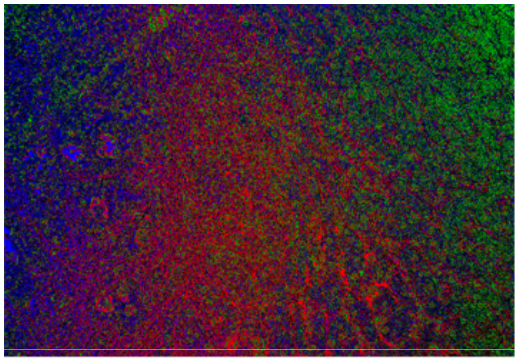


Spectral responses of 3 tissue components in non-physical color space. μ(**B**)=0.9760. $\mu\_average$(**B**)=0.8937.

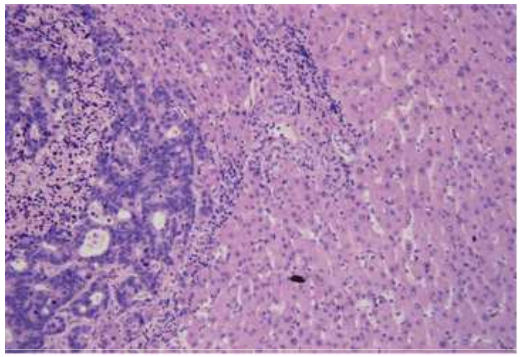
Segmentation of image of unstained specimen of human liver with metastasis from colon cancer



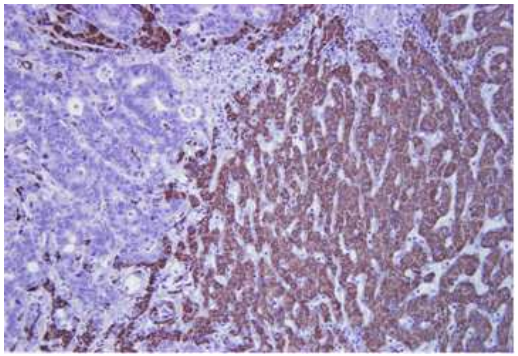
Ustained specimen: μ(**A**)=0.9997, $\mu\_average$(**A**)=0.9993.



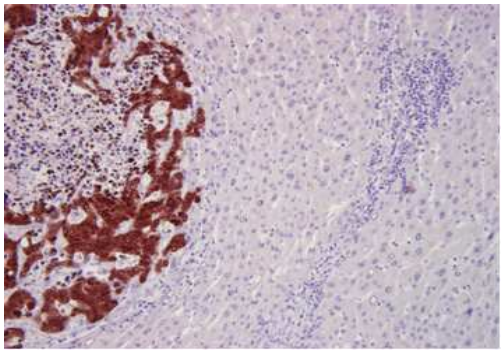
Color-coded EKM-NMF_L0 (D=50, σ^2=0.1). μ(**B**)=0.9998, $\mu\_average$(**B**)=0.9984. Blue: colon cancer; green: hepatocytes; red: border area between the tomor and liver tissue.



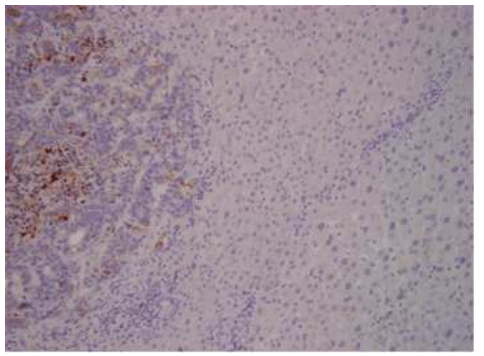
Staining with H&E (the same slide).



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

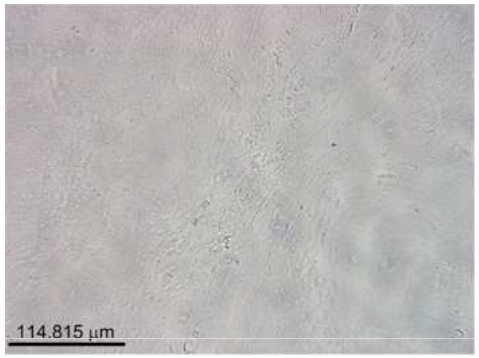


Staining with CDX2 (different slide). Brown: metastatic cells of colon cancer, blue: hepatocytes and inflamatory cells.

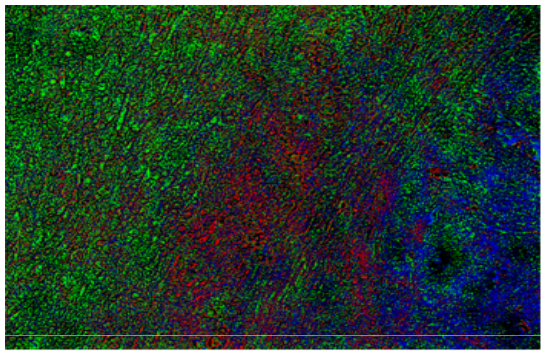


Staining with CK20 (different slide). Brown: metastatic cells of colon cancers, blue: hepatocytes and inflamatory cells.

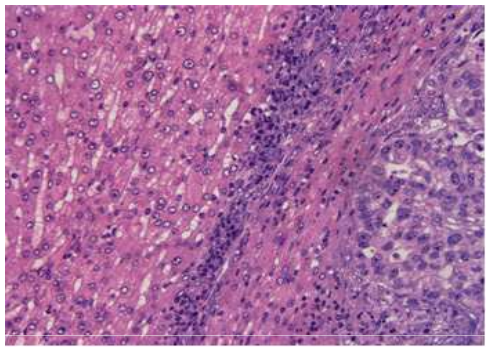
Segmentation of image of unstained specimen of human liver with metastasis from gastric cancer



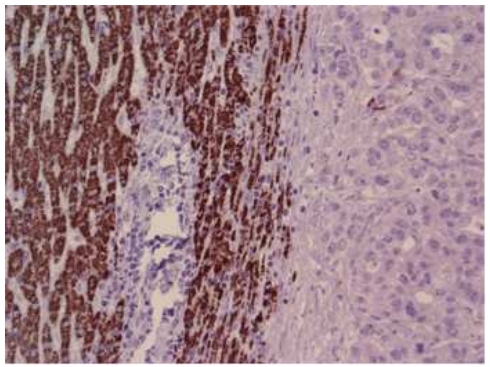
Ustained specimen: μ(**A**)=0.9999, $\mu\_average$(**A**)=0.9988.



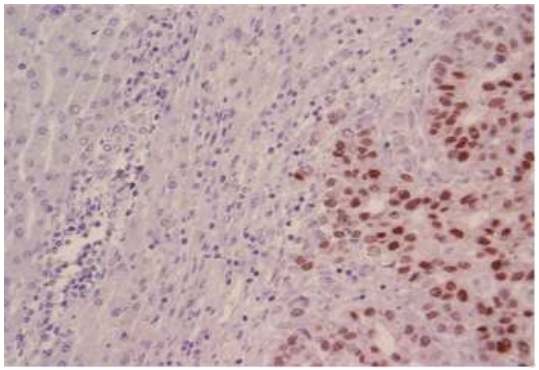
Color-coded EKM-NMF_L0 (D=50, σ^2=0.1). μ(**B**)=0.9994, $\mu\_average$(**B**)=0.9917. Blue: gastric cancer; green: hepatocytes; red: border area of inflamation.



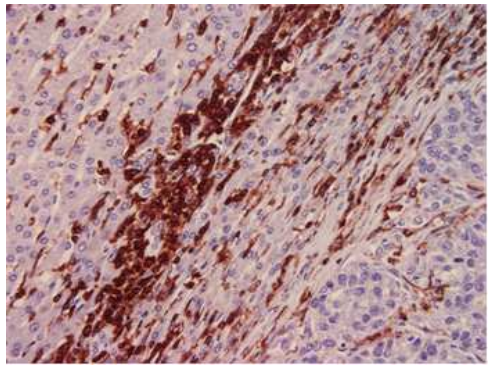
Staining with H&E (the same slide).



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

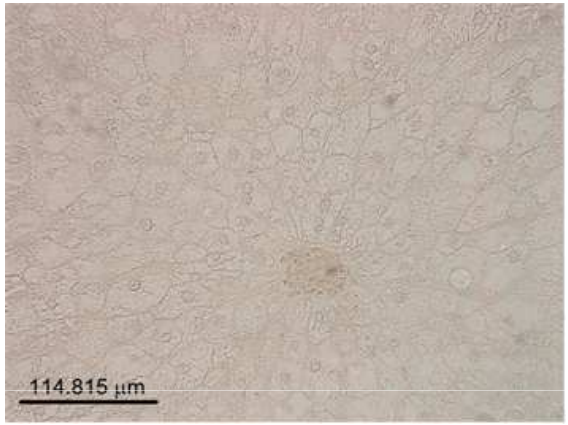


Staining with CDX2 (different slide). Brown: metastatic cells of gastric cancer, blue: hepatocytes and inflamatory cells.

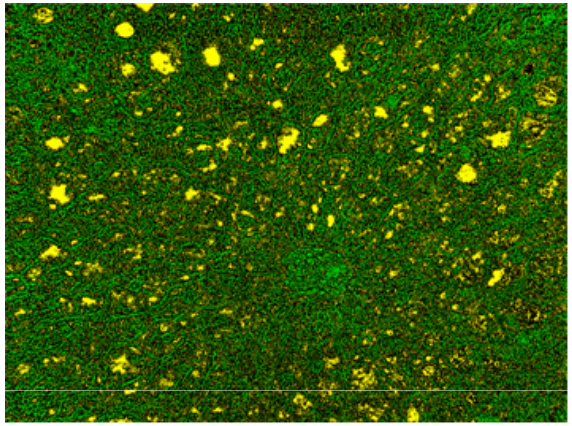


Staining with LCA (different slide). Brown: inflamatory cells; blue: hepatocytes and metastatic cells of gastric cancer.

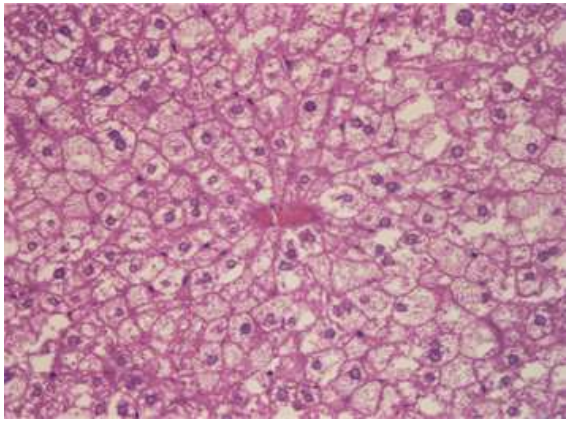
Segmentation of image of unstained specimen of mouse fatty liver



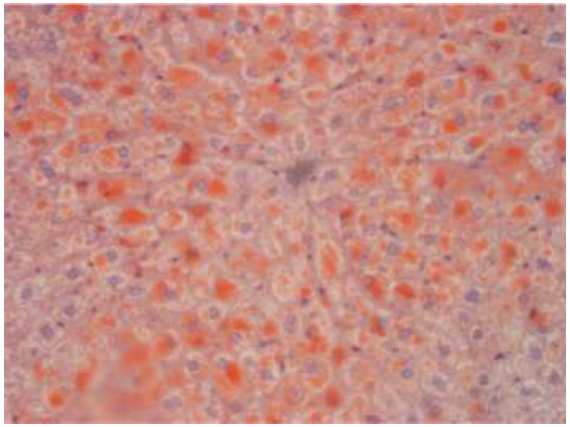
Ustained specimen.



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

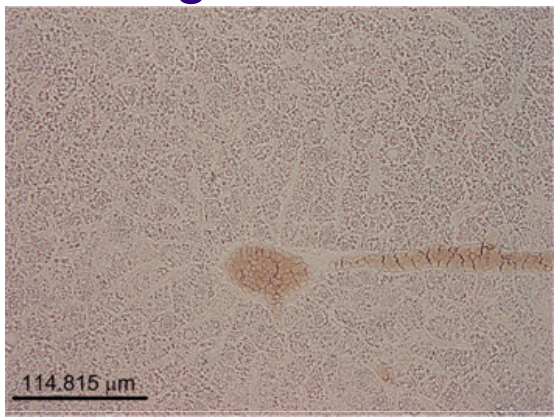


Staining with H&E (the same slide).

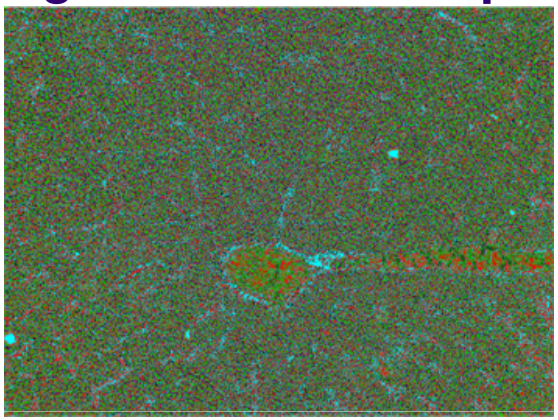


Staining with SUDAN III (different slide).
Orange: fat storage granules.

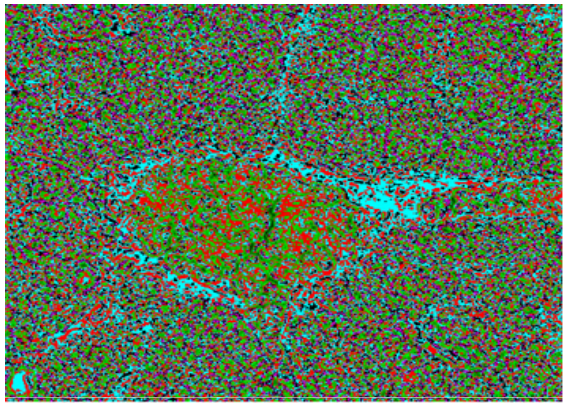
Segmentation of image of unstained specimen of mouse fatty liver



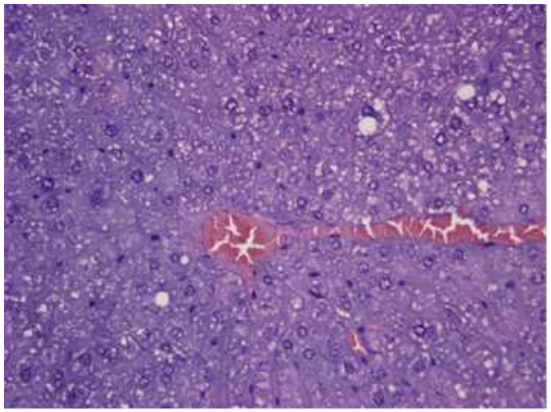
Ustained specimen.



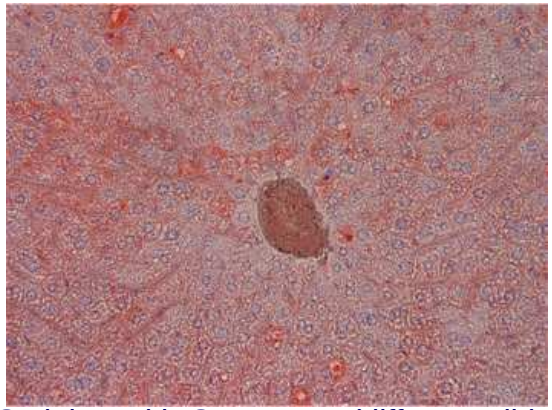
Color-coded EKM-NMF_L0 (D=50, σ^2=0.1). Red: blood vessel, sky blue: sinusoids, green: hepatocytes, magenta: reticular fiber.



Zoomed area



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



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

THANK YOU !!!!!!!