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

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):

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:**

\[ X = AS \quad \text{where} \quad X \in \mathbb{R}^{N \times T}, \ A \in \mathbb{R}^{N \times M}, \ S \in \mathbb{R}^{M \times T} \]

- \( N \) - 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.

Blind Source Separation – linear static problem

\( \mathbf{X} = \mathbf{A} \mathbf{S} \) and \( \mathbf{X} = \mathbf{A} \mathbf{T}^{-1} \mathbf{S} \) 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 \( \mathbf{X} \): \( \mathbf{L}(\mathbf{X}) = \mathbf{A} \mathbf{L}(\mathbf{S}) \). ICA is applied to \( \mathbf{L}(\mathbf{X}) \) to estimate \( \mathbf{A} \) and \( \mathbf{L}(\mathbf{S}) \). \( \mathbf{S} \) 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/](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](http://www.bsp.brain.riken.jp/ICALAB/nmflab.html)

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

\[
x_t = f(s_t) \quad t = 1, \ldots, T \quad ; \quad x_t \in \mathbb{R}_{0+}^{N \times 1}
\]

stands for nonnegative vector comprised of measurements acquired at \(T\) independent variables (pixel positions, \(m/z\) ratios, genes, etc.).

\[
s_t \in \mathbb{R}_{0+}^{M \times 1}
\]

stands for unknown vector of \(M\) sources. \(M > N \rightarrow \text{uBSS problem}\)

\[
f : \mathbb{R}_{0+}^{M} \mapsto \mathbb{R}_{0+}^{N}
\]

is an unknown multivariate mapping such that:

\[
f(s_t) = \begin{bmatrix} f_1(s_t) & \ldots & f_N(s_t) \end{bmatrix}^T \quad \text{and} \quad \{f_n : \mathbb{R}_{0+}^{M} \rightarrow \mathbb{R}_{0+}^{N} \}_{n=1}^N.
\]

Linear problem: \(f(s_t) = As_t\).


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 \( s_t, t=1,\ldots,T \) or simultaneously solving underdetermined linear systems of equations \( x_t = As_t \). Provided that \( s_t \) is sparse enough, solution is obtained at the minimum of \( L_p \)-norm, \( \| s_t \|_p, 0 \leq p \leq 1 \).

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

- NMF-based solution yields \( A \) and \( S \) simultaneously through sparseness and nonnegativity constrained factorization of \( X \).
When uBSS problems can(not) be solved?

Let us focus on underdetermined linear system:

\[ x = As, \quad x \in \mathbb{R}^N, \quad s \in \mathbb{R}^M, \quad M > N \]

Let \( s \) be \( K \)-sparse i.e. \( K = \|s\|_0 \).

Provided that \( A \) is \textit{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]:

\[ N \approx K \log(M/K) \]

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, [8]. In chemometrics $\mathbf{A}$ contains concentration profiles of pure components present in the mixtures, [9].

One result for deterministic $\mathbf{A}$ is given in [10]. 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 $\mathbf{A}$. $K$ 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.**

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

\[
\mu \{ A \} = \max_{1 \leq i, j \leq M \text{ and } i \neq j} \frac{|a_i^T a_j|}{\| a_i \| \cdot \| a_j \|}
\]

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 \( s \) relates sparseness requirement on \( s \) and coherence of \( A \), [11,12]:

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

When uBSS problems can(not) be solved?

When the mutual coherence \( \mu(A) \) 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 \( \mu(A) \approx 0.9999 \).

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].

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 $\mu(A) \approx 1$ it was proposed in [13,5,6] to transform problem:

\[ X = AS \text{ or } X = f(S) = GS \]

into:

\[ \Psi(X) \approx BS \]

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


Nonlinear uBSS [5]

\[ x_t = f(s_t) \quad t=1,\ldots,T \]

\[ x_t \in \mathbb{R}^{N \times 1}, s_t \in \mathbb{R}^{M \times 1}, M > N. \]

\[ f : \mathbb{R}^{M}_{0+} \mapsto \mathbb{R}^{N}_{0+} \]

\[ f(s_t) = \begin{bmatrix} f_1(s_t) & \cdots & f_N(s_t) \end{bmatrix}^T \quad \{ f_n : \mathbb{R}^{M}_{0+} \mapsto \mathbb{R}^{N}_{0+} \}_{n=1}^{N} \]

Nonlinear uBSS problem can be expanded into Taylor series around reference point \( s_0 \). Without loss of generality let us assume \( s_0 = 0_{M \times 1} \). Let us also assume \( f(s_0) = 0_{N \times 1} \).

Nonlinear uBSS

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

\[
  x = f(s) = Gs
\]

where \( G \) is a matrix

\[
  G = \sum_{j=1}^{J} (1/j!) G_{(1)}^{j}
\]

such that:

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

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 $X = AS$ nonlinearily into new one:

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

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

The nonlinear mapping has the following algebraic structure:

$$\phi(x(t)) = \left[ \{c_{q_{1}...q_{N}} x_{1}^{q_{1}}(t)...x_{N}^{q_{N}}(t)\}_{q_{1},...,q_{N}=0}^{\bar{N}} \right]^{T} \text{ such that } \sum_{n=1}^{N} q_{n} \leq \bar{N}, \ \forall t = 1,...,T.$$ 

Nonlinear mapping of linear uBSS problem?

The mapped problem becomes:

\[
\phi(x(t)) = c_0 e_1 + B \begin{bmatrix} 0 \\ s(t) \end{bmatrix} + B_{HOT} \begin{bmatrix} 0 \\ 0_{M \times 1} \\ s(t)_{HOT} \end{bmatrix} \quad \forall t = 1, ..., T
\]

where \( s(t)_{HOT} \) is \( \bar{N} - M - 1 \) column vector comprised of: \( \{s_1^{q_1}(t) \times s_2^{q_2}(t) \times \cdots \times s_M^{q_M}(t)\}_{q_1 \cdots q_M = 2}^{\bar{N}} \)

such that: \( \sum_{m=1}^{M} q_m \leq \bar{N} \).
Nonlinear mapping of linear uBSS problem?

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

\[
\phi(X) \approx \left[ c_0 e_1 \ldots c_0 e_1 \right] \times T \text{times} + B \left[ 0 \right] \left[ S \right] \left[ \left\{ s_{m_1} s_{m_2} \right\}_{m_1, m_2 = 1}^M \right]
\]

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 $\{x_i \in \mathbb{R}^N\}_{i=1}^T$ and real numbers $\{\alpha_i\}_{i=1}^T : \sum_{i,j=1}^T \alpha_i \alpha_j \kappa(x_i, x_j) \geq 0$.

Theorem 1. The Morre-Aronszjan theorem [18]. Given any nonnegative definite function $\kappa(x, y)$ there exists a uniquely determined RKHS $H_\kappa$ consisting of real valued functions on set $X \subset \mathbb{R}^N$ such that: (i) $\forall x \in X, \kappa(o, x) \in H_\kappa$; (ii) $\forall x \in X, \forall f \in H_\kappa, f(x) = \langle f, \kappa(o, x) \rangle_{H_\kappa}$. Here, $\langle o, o \rangle$ denotes inner product associated with $H_\kappa$.

Reproducible kernel Hilbert spaces

**Definition 2.** Replacing $f(x)$ in (ii) in Theorem 1 by $\kappa(\cdot, x)$ it follows
\[ \kappa(x_t, x) = \langle \kappa(\cdot, x_t), \kappa(\cdot, x) \rangle_{H_x}. \]
By selecting the nonlinear map as $\phi(x) = \kappa(\cdot, x)$ it follows
\[ \kappa(x_t, x) = \langle \phi(x_t), \phi(x) \rangle_{H_x}. \]
That is known as kernel trick. The nonlinear mapping $\phi(x)$ is known as explicit feature map (EFM) associated with kernel $\kappa(\cdot, x)$.

**Definition 3.** Empirical kernel map (EKM), [19]. For a given set of patterns
\[ \{v_d \in \mathbb{R}^N\}_{d=1}^D \subset X, \quad D \in \mathbb{N}, \]
we call $\psi : \mathbb{R}^N \rightarrow \mathbb{R}^D :$
\[ \{x_t \mapsto \kappa(\cdot, x_t)\}_{d=1}^D = \left[ \kappa(v_1, x_t), \ldots, \kappa(v_D, x_t) \right]_t^T \]
the EKM with respect to $\{v_d\}_{d=1}^D$.

Nonlinear mapping of linear uBSS problem?

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

\[
\phi(X) \approx \begin{bmatrix} c_0 e_1 \ldots c_0 e_1 \end{bmatrix}_{\times T \text{times}} + \bar{B}_0 S \approx X e e B S_{/dncurlybracketleft /horizcurlybracketext /dncurlybracketmid/horizcurlybracketext /dncurlybracketright}
\]

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

\[
\text{span} \{ v_d \}_{d=1}^D \approx \text{span} \{ x_i \}_{t=1}^T
\]

Then:

\[
\text{span} \{ \phi(v_d) \}_{d=1}^D \approx \text{span} \{ \phi(x_t) \}_{t=1}^T
\]
Nonlinear mapping of linear uBSS problem?

Projection yields:

$$\phi(V)^T \phi(x_i) = \psi(x_i)_V = \left[ \langle \phi(v_1), \phi(x_i) \rangle \ldots \langle \phi(v_D), \phi(x_i) \rangle \right]^T$$

When $\phi(x) = k(\circ, x)$ it follows: $\langle \phi(v), \phi(x) \rangle = k(v, x)$. It is shown in [13] that for non-overlapped binary sources non-overlapping binary sources:

$$s_i s_j = \delta(i-j), \quad i, j = 1, \ldots, M$$

that is, the mapping $\Psi(X)_V$ is $S$-invariant. Hence, by parameters of the mapping (kernel function $\kappa(v_d, x_i)$ and basis $V$) it is possible to tune $\mu(B)$. 
Nonlinear mapping of linear uBSS problem?

Basis \( V = \{ v_d \in \mathbb{R}^{N \times 1} \}_{d=1}^D \) needs to fulfill:

\[
\text{span}\{ v_d \}_{d=1}^D \approx \text{span}\{ x_t \}_{t=1}^T
\]

Thus, \( V \) can be found by clustering \( \{ x_t \}_{t=1}^T \) into \( D \leq T \) clusters. That, for example, can be accomplished by \textit{kmeans} algorithm.

When in addition to \textit{sparseness} constraint \textit{nonnegativity} constraint applies as well (that is the case in applications in imaging) \textit{sparseness constrained NMF} algorithms can be applied to \( \psi(X)_\tau \) 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):

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

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: $A^{(0)}$, $S^{(0)}$,

For $k=1,2,…, until convergence do

Step 1: $S^{(k+1)} = \arg \min_{s_m \geq 0} D_s \left( X \| A^{(k)} S \right)_{S^{(k)}}$

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

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:

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

- 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:

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

- 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]:

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

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

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

\[ \frac{\partial D(A,S)}{\partial s_{mt}} = \left[ -A^TX + A^TAS \right]_{mt} + \alpha_S \frac{\partial J_S(S)}{\partial s_{mt}} \]

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:

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

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

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:

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

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

$$
A \leftarrow A \otimes \left[ X S^T - \alpha_A 1_{NM} \right]_+ \quad S \leftarrow S \otimes \left[ A^T X - \alpha_S 1_{MT} \right]_+
$$
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 [21] by estimating rank-1 one factors $a_m s_m$ one at a time. Each time $a_m s_m$ is estimated it is removed from $X \rightarrow X - a_m s_m$. To prevent subtraction from being negative the under-approximation constraint is imposed on $a_m s_m$: $a_m s_m \leq X$.

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

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

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 \textit{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.spasc.tugraz.at/people/robert-peharz.

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

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].

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: $\mu(A) = 0.9995$.

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

**Unstained** specimen of human liver with metastasis from colon cancer $\mu(A) > 0.9997$. 
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: $\mu(A)=0.9995$.

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


Segmentation of synthetic image

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

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

Color coded ground truth.

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

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. $\mu(A)=0.9995$. $\mu_{\text{average}}(A)=0.9956$.

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

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

<table>
<thead>
<tr>
<th>SNR [dB]</th>
<th>$\sigma^2$</th>
<th>$\text{SNR} \geq 29$</th>
<th>$18 \leq \text{SNR} \leq 28$</th>
<th>$17 \leq \text{SNR} \leq 14$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma^2$</td>
<td>0.001</td>
<td>0.01</td>
<td>0.1</td>
<td></td>
</tr>
</tbody>
</table>
Segmentation of image of unstained specimen of human liver with HCC

Ustained specimen of human liver with HCC: $\mu(A)>0.9999$.

Color-coded EKM-NMF_L0 ($D=50$, $\sigma^2=0.1$). $\mu(B)=0.9760$. 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.
\( \mu(A) > 0.9999. \ \mu_{\text{average}}(A) = 0.9985. \)

Spectral responses of 3 tissue components in non-physical color space.
\( \mu(B) = 0.9760. \ \mu_{\text{average}}(B) = 0.8937. \)
Segmentation of image of unstained specimen of human liver with metastasis from colon cancer

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

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


Staining with H&E (the same slide).


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

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

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


Staining with H&E (the same slide).
Segmentation of image of unstained specimen of mouse fatty liver

Ustained 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

Ustained 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 SUDAN III (different slide). Orange: fat storage granules.
THANK YOU !!!!!!!!