Lecture I

Motivation with illustration of applications

Ivica Kopriva

Ruđer Bošković Institute

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

Course outline

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

Course outline

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

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 + lp (0<p≤1) min
* Hierarchical nonnegative matrix factorization

Homework problems

1. Understanding natural (relative) gradient. Convergence analysis (comparison) of algorithms in adaptive minimum mean square error problem with matrix argument using Riemanian and Euclidean gradient. **Theory, MATLAB demonstration and comments of the results.**

2. Principal component analysis (PCA) based separation of two Gaussian signals.PCA based separation of two uniformly distributed signals. Scatter plots of true sources, mixtures and estimated sources. PCA based separation of two images (histograms). **Theory, MATLAB demonstration and comments of the results.**

3. Independent component analysis (ICA) based separation of two uniformly distributed signals (scatter plots of true sources, mixtures and estimated sources).ICA based separation of two images (histograms). **Theory, MATLAB demonstration and comments of the results.**

Seminar problems

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

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

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

Seminar problems

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

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

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

Seminar problems

7. Blind separation of three acoustic signals from two static mixtures (a computer generated example) using sparse component analysis (SCA): clustering + L_p norm ($0<p\leq1$) minimization in time-frequency (short-time Fourier) domain. **Theory, MATLAB demonstration and comments of the results.**

8. Blind extraction of five pure components from mass spectra of two static mixtures of chemical compounds using sparse component analysis (SCA): clustering a set of single component points + L_p norm ($0<p\leq1$) minimization in m/z domain. **Theory, MATLAB demonstration and comments of the results.**

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

Blind Source Separation – linear static problem

Signal recovery from multichannel linear superposition using minimum of *a priori* information i.e. multichannel measurements only.

Problem:

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

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

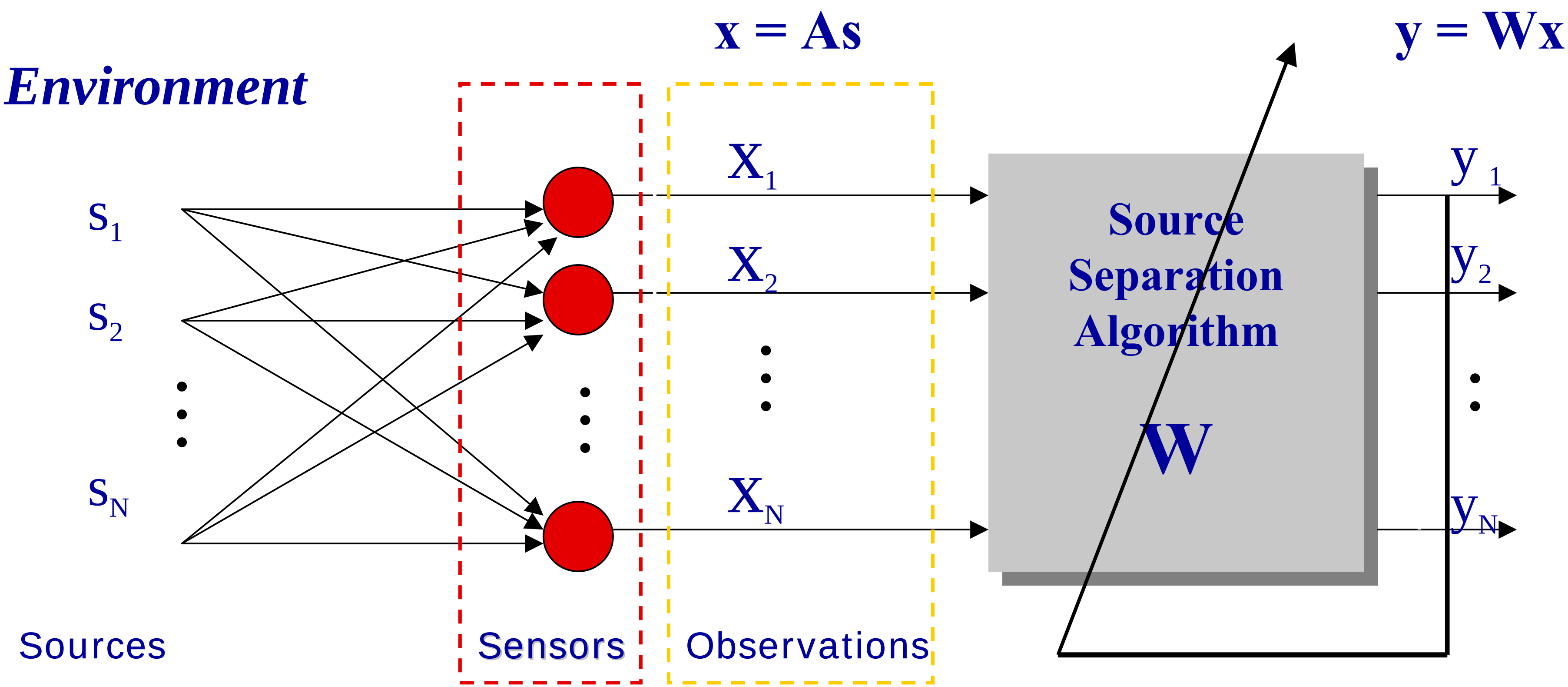
Goal: find **S, A** and number of sources *M* based on **X** only.

Meaningful solutions are characterized by scaling and permutation indeterminacies:

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

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

Blind Source Separation – linear static problem



Blind Source Separation – linear dynamic problem

In many situations related to acoustics and data communications we are confronted with multiple signals received from a multipath mixture.

Sometimes, this is known under a popular name of *cocktail-party* problem.

A multipath mixture can be described by a mixing matrix whose elements are the individual transfer functions between a source and a sensor.

When both mixing matrix and sources are unknown the problem is referred to as the multichannel blind deconvolution (MBD) problem.

A. Hyvarinen, J. Karhunen and E. Oja, *Chapter 19* in Independent Component Analysis, J. Wiley, 2001.

A. Cichocki, S. Amari, *Chapter 9* in Adaptive Blind Signal and Image Processing – Learning Algorithms and Applications, J. Wiley, 2002.

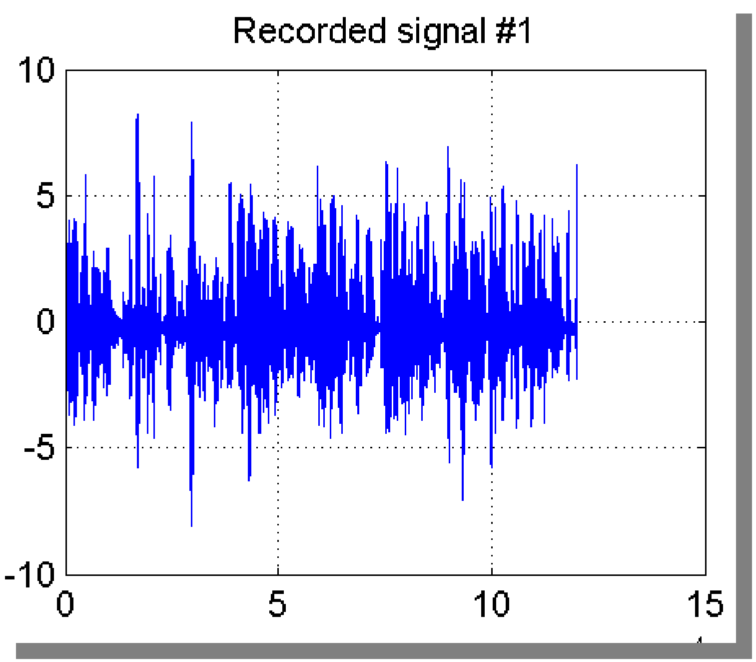
R. H. Lambert and C.L. Nikias, *Chapter 9* in Unsupervised Adaptive Filtering – Volume I Blind Source Separation, S.Haykin, ed., J. Wiley, 2000.

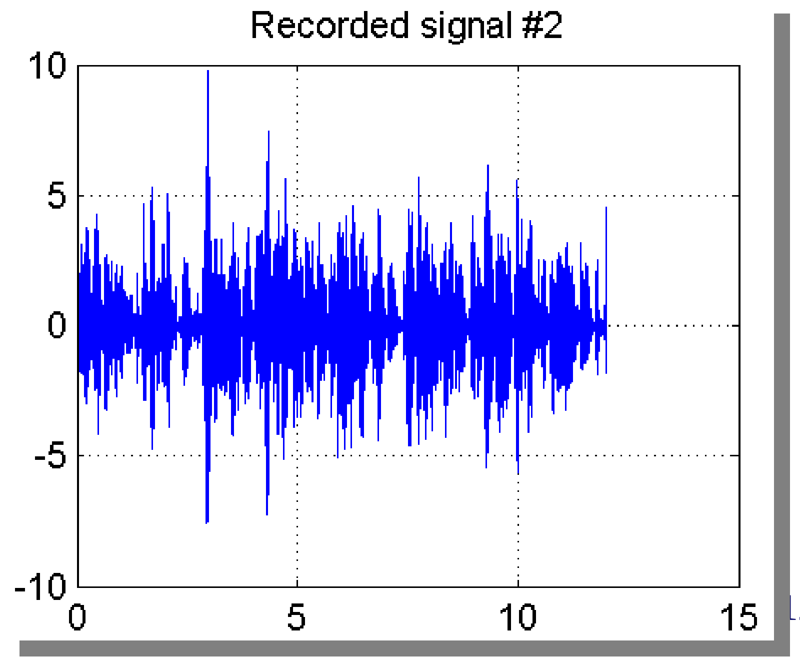
S.C. Douglas and S. Haykin, *Chapter 3* in Unsupervised Adaptive Filtering – Volume II Blind Deconvolution, S. Haykin, ed., J. Wiley, 2000.

Blind Source Separation – linear dynamic problem

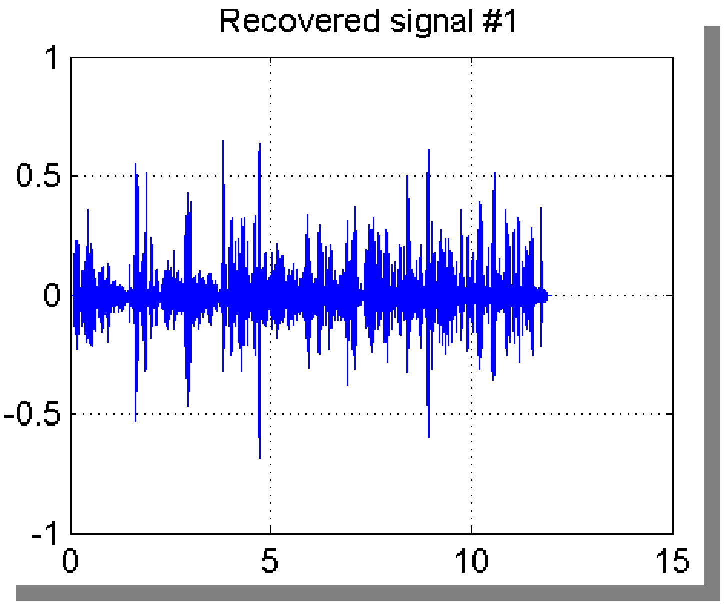
Speech separation in reverberant acoustic environment. Two recorded signals were downloaded from Russel Lamberts’ home page: http://home.socal.rr.com/russdsp/ .

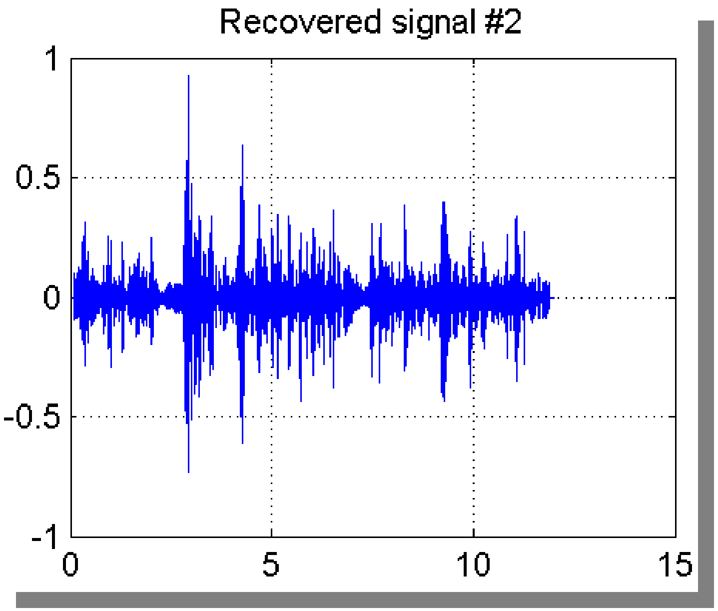
Signals were sampled with 8kHz and contain male and female speakers talking simultaneously for 12 seconds.





Blind Source Separation – linear dynamic problem



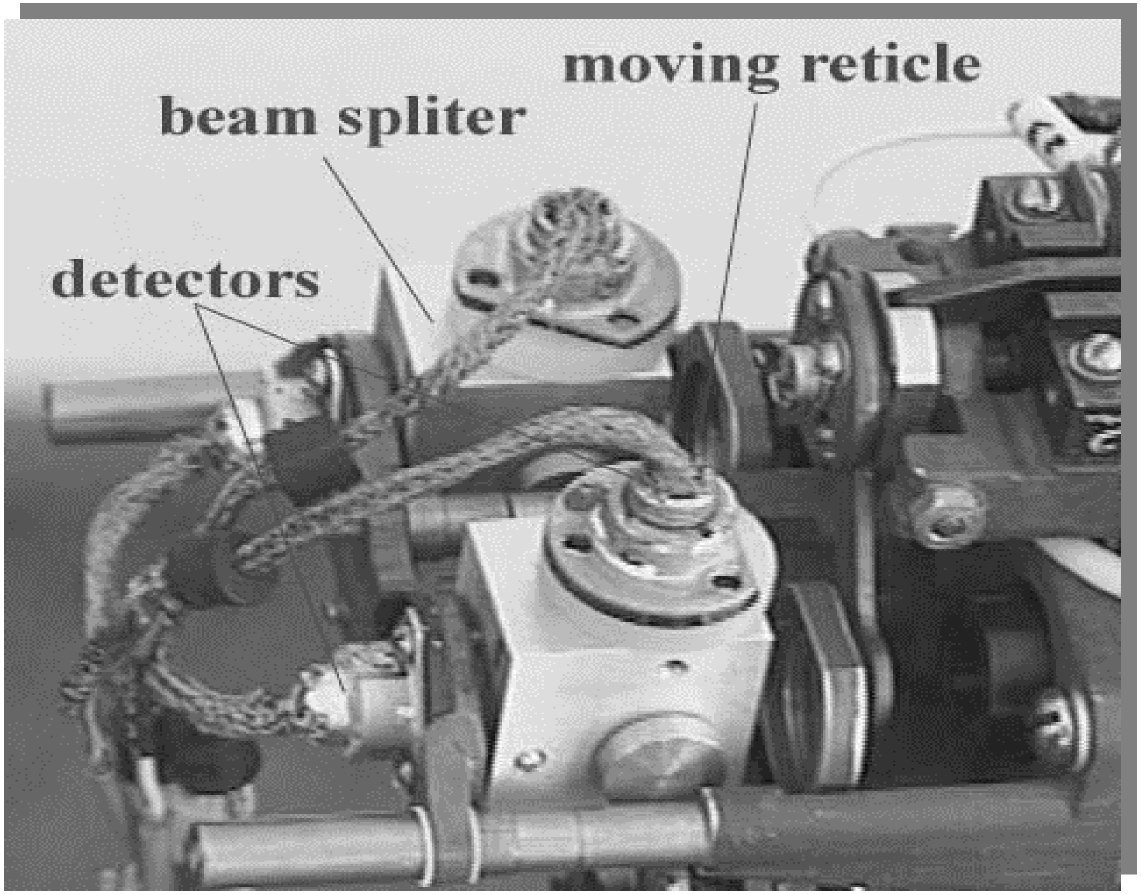


https://www.scientificamerican.com/article.cfm?id=solving-the-cocktail-party-problem

"Computers have great trouble deciphering voices that are speaking simultaneously. That may soon change.."

https://domino.research.ibm.com/comm/research_projects.nsf/pages/speechseparation.index.html

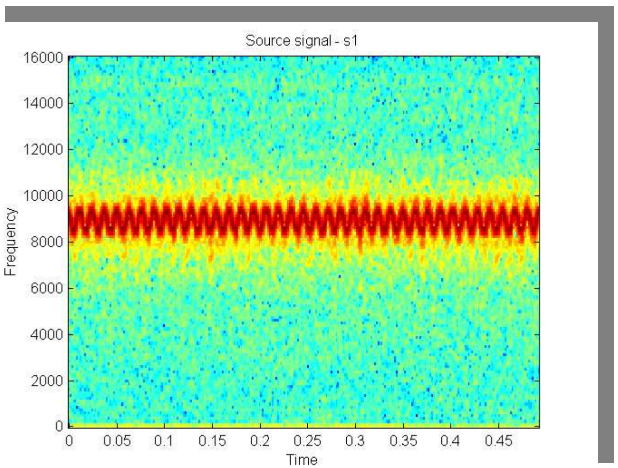
ICA and reticle based IR tracker

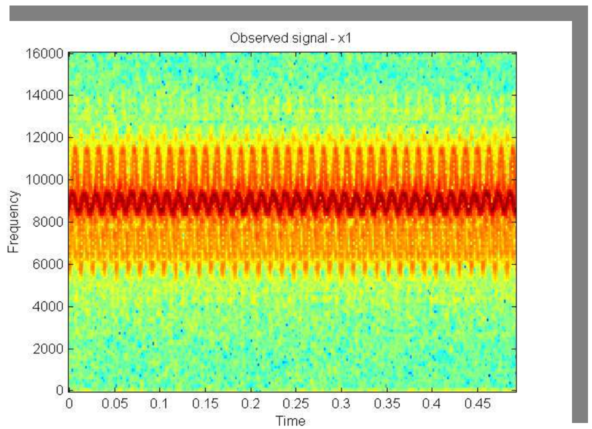


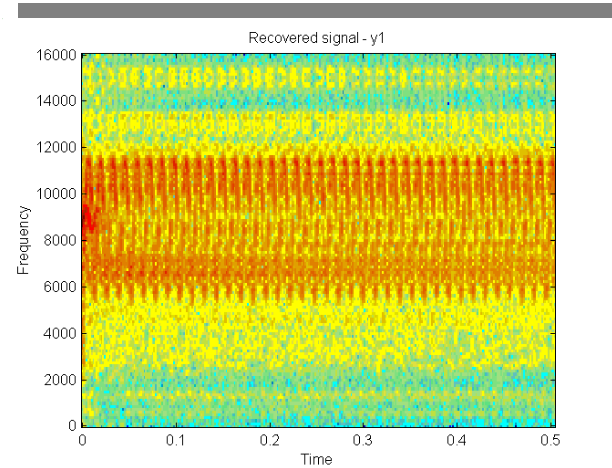
I. Kopriva, A. Peršin, Applied Optics, Vol. 38, No. 7, pp. 1115-1126, 1999.
I.Kopriva, H. Szu, A.Persin, Optics Communications, Vol. 203, Issue 3-6, pp. 197-211, 2002.

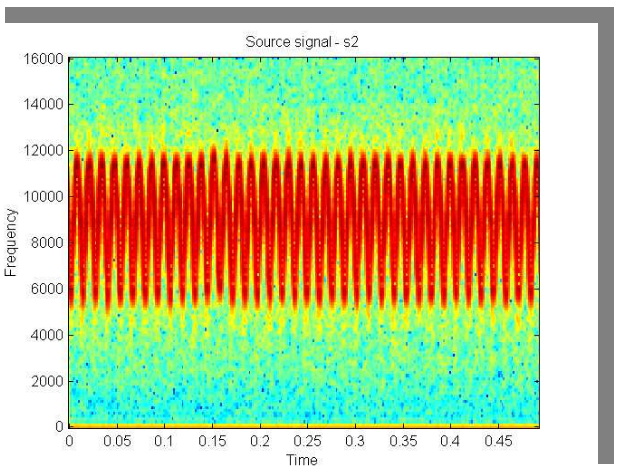
ICA and reticle based IR tracker

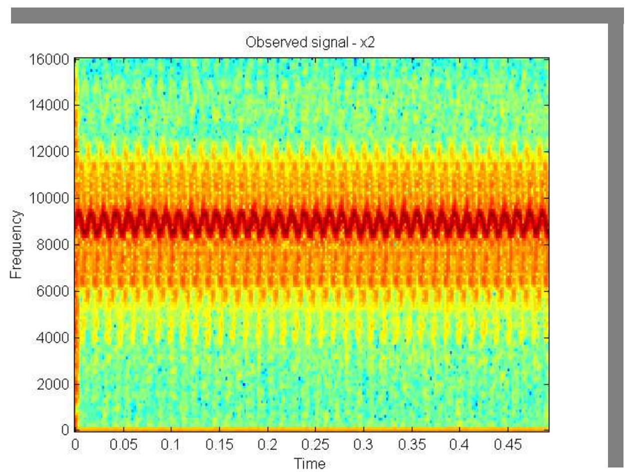
$\mathbf{s}$ $\qquad$ $\mathbf{x} = \mathbf{A} * \mathbf{s}$ $\qquad$ $\mathbf{y} = \mathbf{W} * \mathbf{x}$

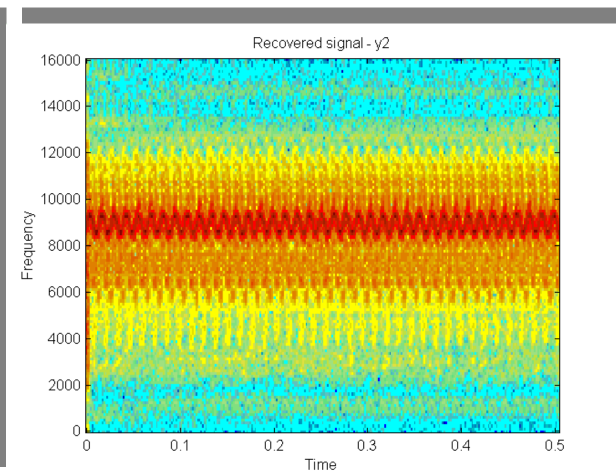












Blind Source Separation – nonlinear static problem

Problem:

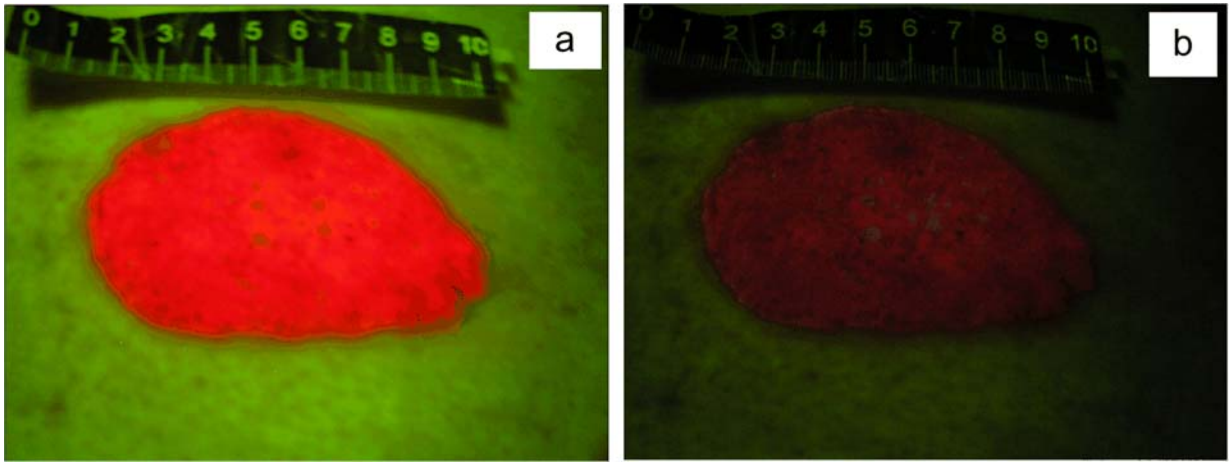
$\mathbf{X}=F(\mathbf{S})$ $\mathbf{X}\in R^{N\times T}$, $\mathbf{S}\in R^{M\times T}$

N-number of sensors;
M- *unknown* number of sources
T-number of samples/observations
F – ***unknown*** vector valued function with vector argument.

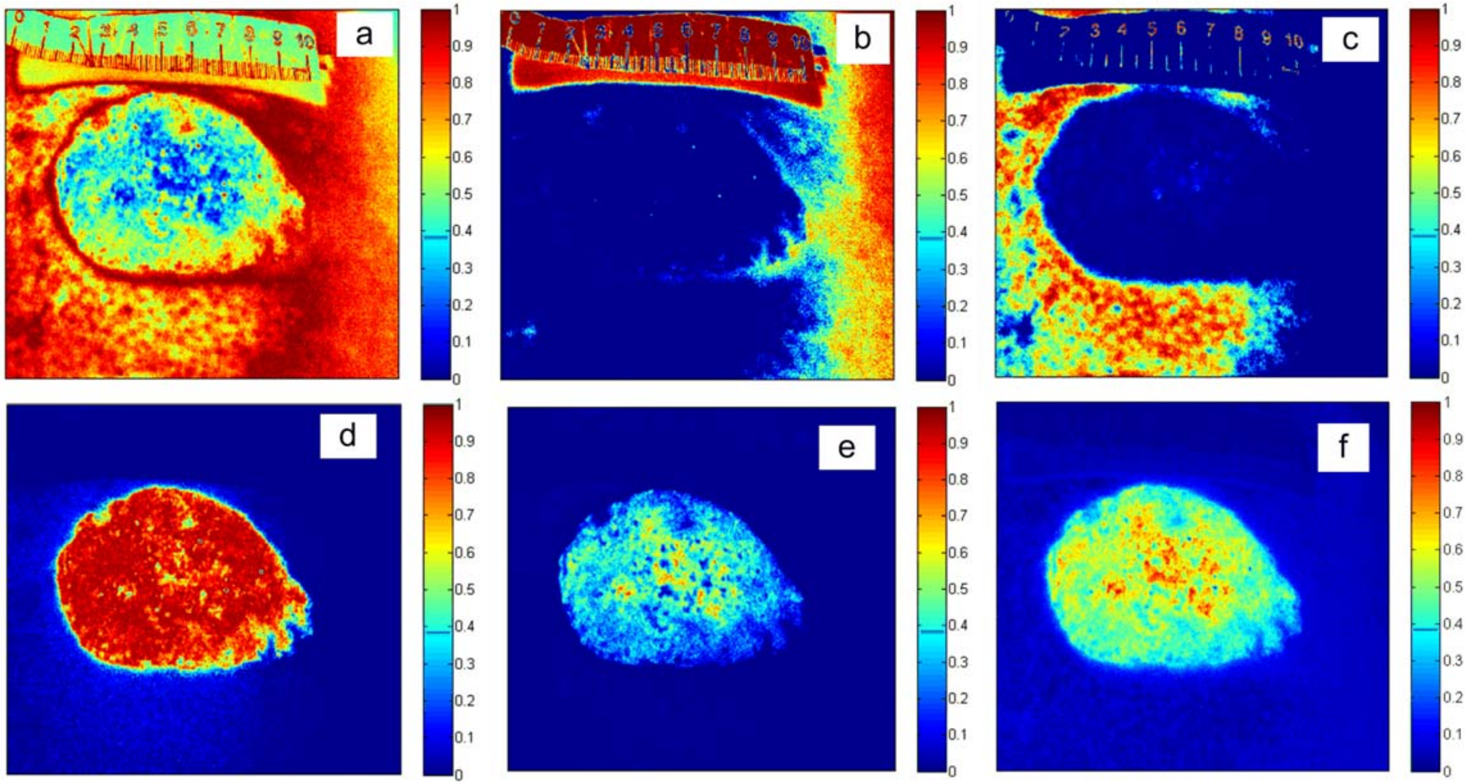
Goal: find **S** based on **X** only. Solution is possible without preconditions on the type of nonlinearity *F* by transforming original problem $\mathbf{X}=F(\mathbf{S})$ into reproducible kernel Hilbert space (RKHS) where mapped sources possibly become linearly separable: $\Phi(\mathbf{X})\approx\mathbf{A}\,\Phi(\mathbf{S})$. Constraints stronger than statistical independence must be imposed on **S**.

"Nonlinear Blind Source Separation," *Chapter 18* in:, "Handbook of Blind Source Separation," Academic Press, 2010, P. Comon, C. Jutten, editors.

Blind Source Separation – nonlinear static problem



Blind Source Separation – nonlinear static problem



a) – d) parts decomposed by nonlinear method.
e) and f) tumor component decomposed by nonlinear method.

I. Kopriva and A. Peršin (2009). Unsupervised decomposition of low-intensity low-dimensional multi-spectral fluorescent images for tumour demarcation, *Medical Image Analysis* 13, 507-518.

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 (**A**,**S**) satisfying linear mixture model **X**=**AS**.Constraints must be imposed on **A** and/or **S** in order to obtain solution of the BSS problem that is characterized with **T**=**P**Λ.

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

Dependent component analysis (DCA) improves accuracy of the ICA when sources are not statistically independent.

Sparse component analysis (SCA) solves BSS problem imposing sparseness constraints on source signals.

Nonnegative matrix factorization (NMF) solves BSS problem imposing nonnegativity, sparseness, smoothness or constraints on source signals.

Statistical independence - ICA

Imagine situation in which two microphones recording weighted sums of the two signals emitted by the speaker and background noise.

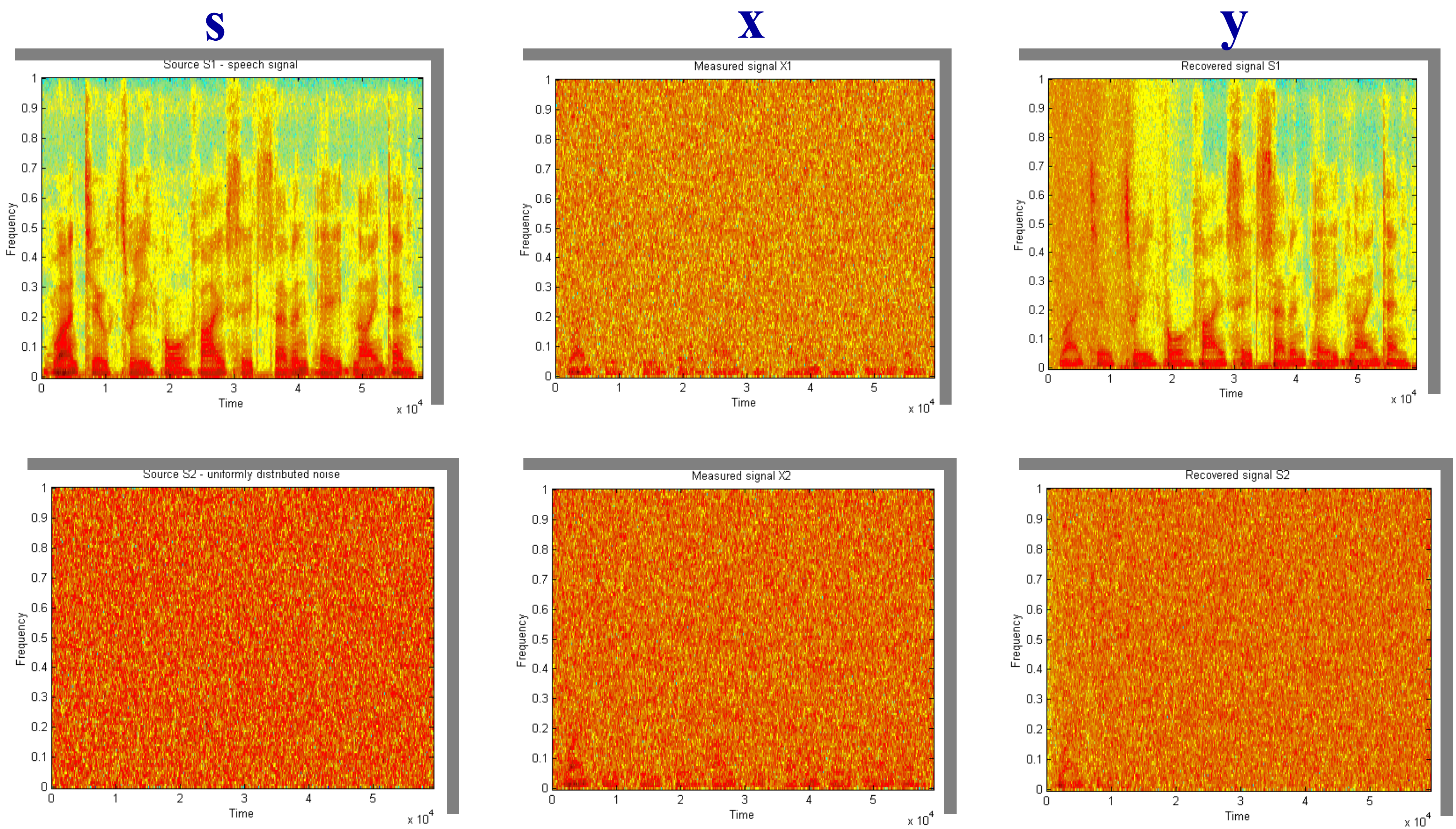
$$x_1 = a_{11}s_1 + a_{12}s_2$$
$$x_2 = a_{21}s_1 + a_{22}s_2$$

The problems is to estimated the speech signal (s_1) and noise signal (s_2) from observations x_1 and x_2.

If mixing coefficients a_{11}, a_{12}, a_{21} and a_{22} are known problem would be solvable by simple matrix inversion.

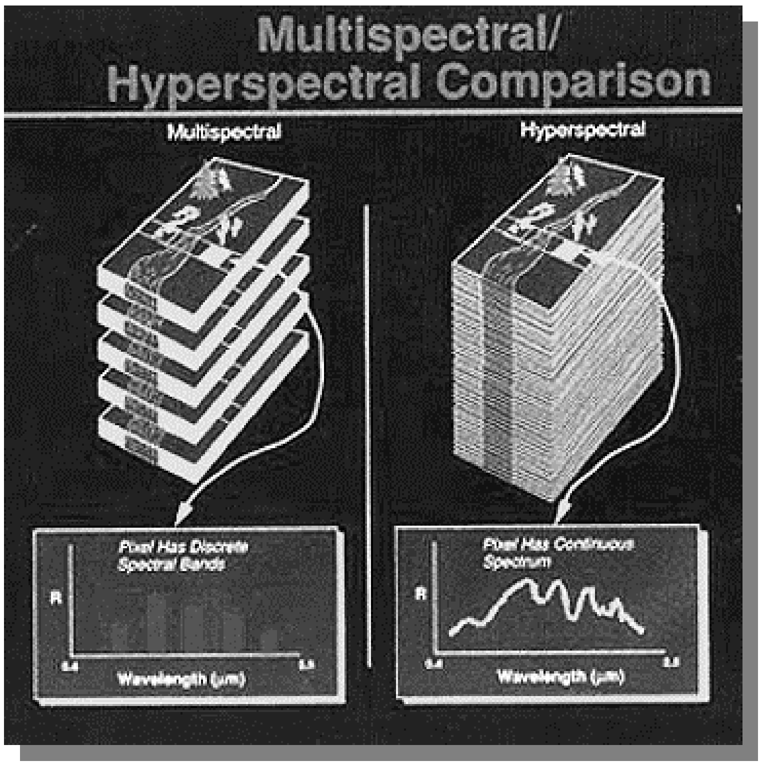
ICA enables to estimated speech signal (s_1) and noise signal (s_2) without knowing the mixing coefficients a_{11}, a_{12}, a_{21} and a_{22}. This is why the problem of recovering source signals s_1 and s_2 is called *blind source separation* problem.

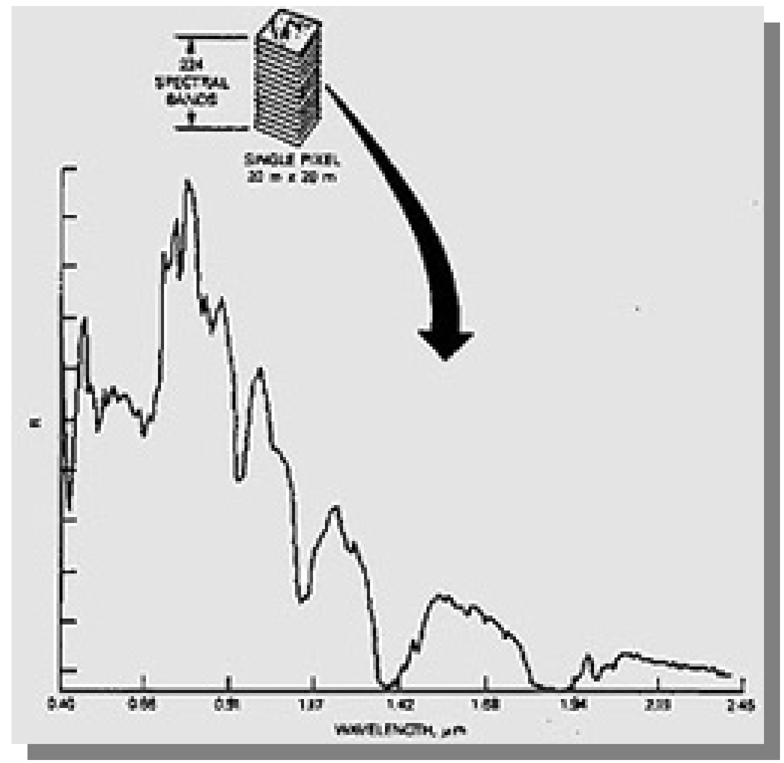
Speech from noise separation



ICA and multispectral remote sensing

Hyperspectral vs. Multispectral Remote sensing





- ❑SPOT- 4 bands, LANDSAT -7 bands, AVIRIS-224 bands (0.38μ-2.4μ);
- ❑ Objects with very similar reflectance spectra can be discriminated.

Hyperspectral/Multispectral Linear Mixing Data Model

For sensor consisting of N bands and M pixels linear data model is assumed:

$$\mathbf{x} = \mathbf{A}\mathbf{s} = \sum_{i=1}^{M} \mathbf{a}_i s_i \qquad [\mathbf{a}_1 \quad \mathbf{a}_2 \quad \ldots \quad \mathbf{a}_M] \equiv \mathbf{A}$$

x - measured data intensity vector, $\mathbf{x} \in \mathbf{R}^{N\times1}$

s - unknown class vector, $\mathbf{s} \in \mathbf{R}^{1\times M}$

A – unknown spectral reflectance matrix nonsingulairty condition implies $\mathbf{a}_i \neq \mathbf{a}_j$. $\mathbf{A} \in \mathbf{R}^{N\times M}$

Unknown endmembers s_i are can be recovered by ICA based de-mixing:

$$\hat{\mathbf{s}} = \mathbf{W}\mathbf{x}$$

Statistical independence assumption between sources (classes) fails when they become spectrally similar. Thus, ICA will be less accurate for low-dimensional multispectral image than for high-dimensional hyperspectral image.

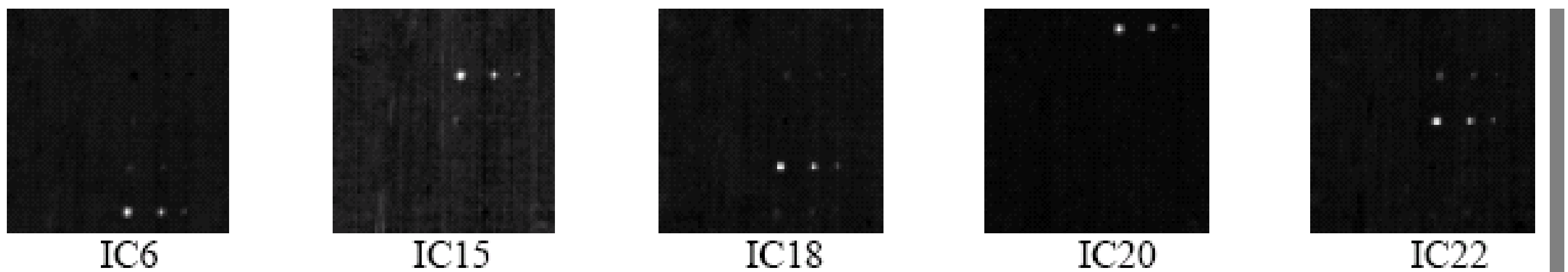
ICA and unsupervised classification of the hyperspectral image

- HYDICE Panel scene (a) that contains 15 panels in 5x3 matrix. Image is collected in Maryland in 1995 from the flight altitude of 10000 feet with approximately 1.5m spatial resolution.
- Original HYDICE image had 210 channels with spectral coverage 0.4-2.5μm. After removing atmospheric bands with low SNR number of bands was reduced to 169.
- In each row panels are made from the same material but differ in size that varies as 3x3m 2x2m and 1x1m.

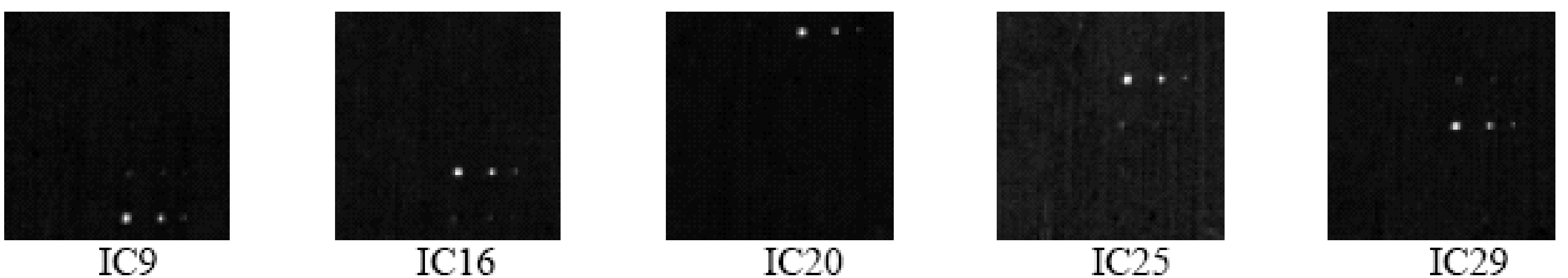
(a) (b)

Q. Du, I. Kopriva and H. Szu, “ Independent Component Analysis for Hyperspectral Remote Sensing Imagery Classification,” *Optical Engineering*, vol. 45, 017008, January 2006.
Q. Du, I. Kopriva, “Automated Target Detection and Discrimination Using Constrained Kurtosis Maximization,” *IEEE Geoscience Remote Sensing Letters*, vol. 5, No. 1, pp. 38-42, 2008.

❑ With noise adjusted PCA algorithm for dimensionality reduction and JADE ICA algorithm for image classification all five panel classes have been correctly classified with only 30 principal components in image representation.



5(b) From 30 PC components



5(c) From 40 PC components

ICA and fMRI signal processing

- Separating fMRI data into independent spatial components involves determining three-dimensional brain maps and their associated time courses of activation that together sum up to the observed fMRI data.

- The primary assumption is that the component maps, specified by fixed spatial distributions of values (one for each brain voxel), are spatially independent.

- This is equivalent to saying that voxel values in any one map do not convey any information about the voxel values in any of the other maps.

- With these assumptions, fMRI signals recorded from one or more sessions can be separated by the ICA algorithm into a number of independent component maps with unique associated time courses of activation.

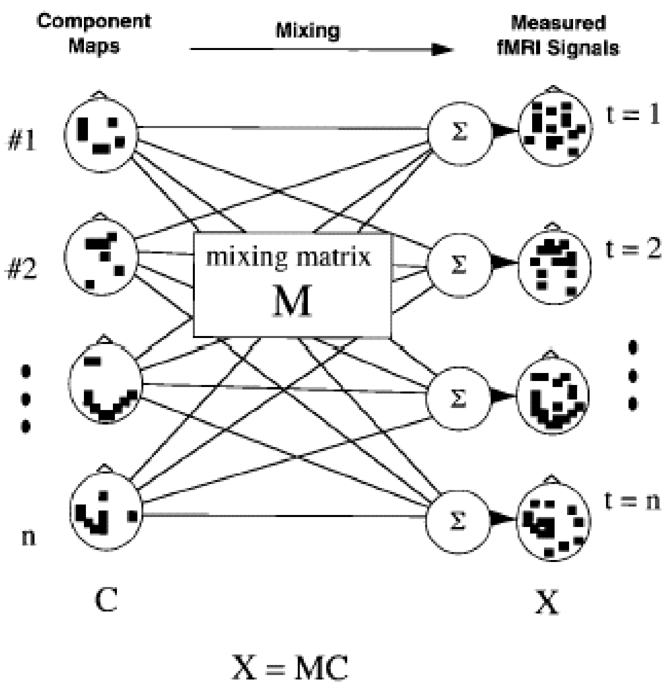


Figure 3.
fMRI data as a mixture of independent components. The mixing matrix M specifies the relative contribution of each component at each time point. ICA finds an *unmixing* matrix that separates the observed component mixtures into the independent component maps and time courses.

McKeown, et. al, “Analysis of fMRI Data by Blind Separation Into Independent Spatial Components,” Human Brain Mapping 6: 160-188 (1998).
M. J. McKewon, et. al, Spatially independent activity patterns in functional MRI data during the Stroop color-naming task,” Proc. Natl. Acad. Sci, USA, Vol. 95, pp.803-810, February 1998.

ICA and fMRI signal processing

❑ The matrix of component map values can be computed by multiplying the observed data by the ICA learned de-mixing matrix ***W***.

Where ***X*** is the *NxM* matrix of fMRI signal data (*N*, the number of time point in the trial, and *M*, the number of brain voxels and C_{ij} is the value of the *j* voxel of the *i*th component.

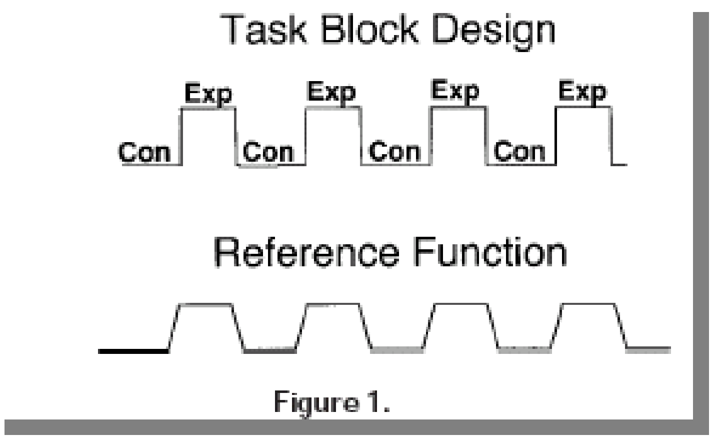
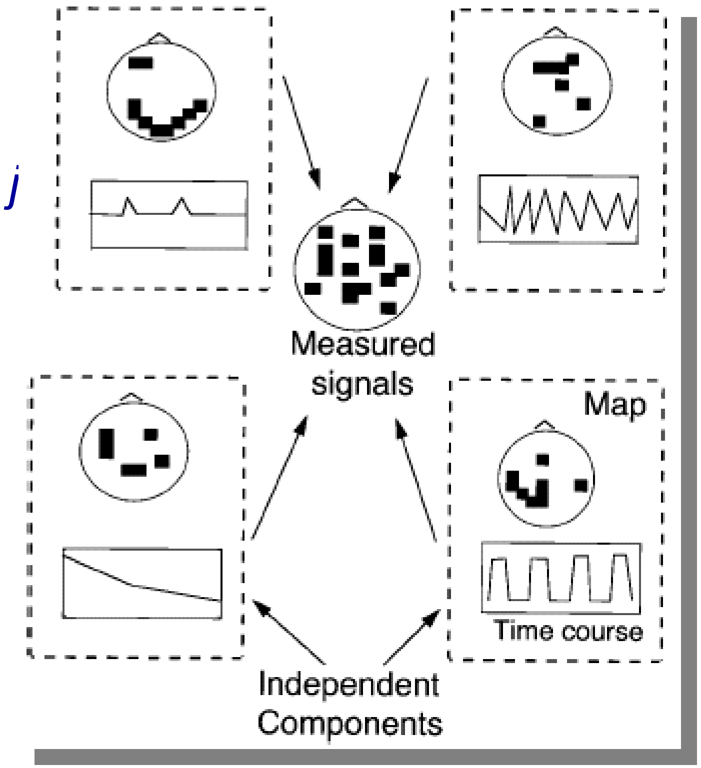


Figure 1.



ICA and fMRI signal processing

ICA has been successfully used to distinguish between task related and non-task related signal components.

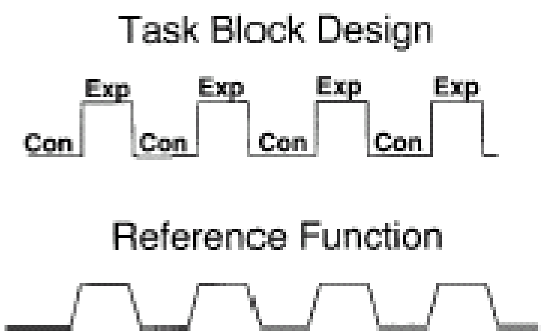


Figure 1.

BOLD signal complexity and task reference function. a: Time courses of 10 randomly selected voxels from a 6-min fMRI trial of the Stroop color-naming task illustrate the typical complexity of BOLD signals. b: Convolving an a priori estimate of the hemodynamic response function with the square-wave function representing the task block structure of the trial, alternating experimental (Exp) and control (Con) blocks (upper trace) produce the reference function for the trial (bottom trace).

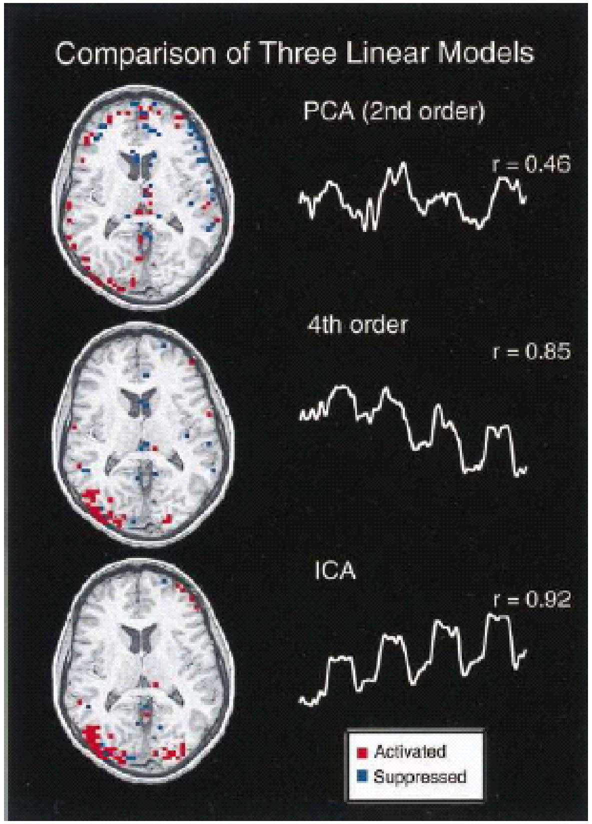
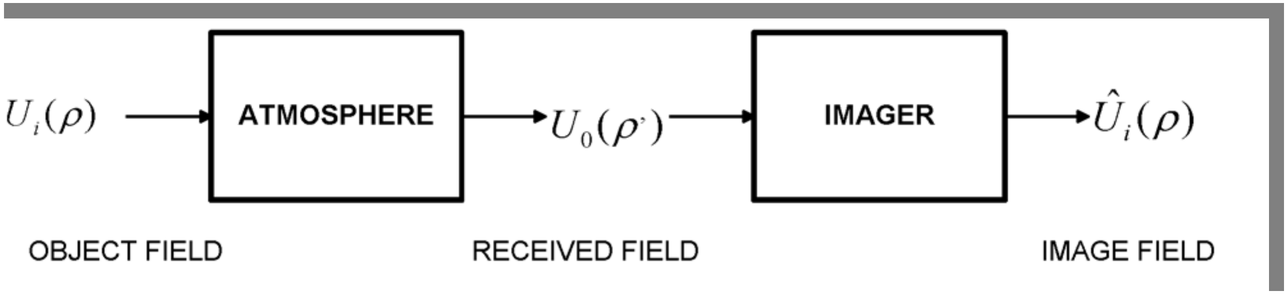


FIG. 2. Comparison of three linear models for analyzing fMRI data. PCA and two versions of ICA were used to linearly separate the data into partially spatially independent maps. The most consistently task-related component determined by each of the three methods from the first trial are shown, along with the correlation coefficient between the associated time courses and the reference function for the behavioral experiment. The ICA algorithm components resembled the task reference function much more strongly than the most highly correlated PCA components.

ICA and image sharpening in the atmospheric turbulence

- Random fluctuations of the refractive index in space and time along the atmospheric path will degrade performance of the imaging system much beyond the classical Rayleigh’s diffraction limit.



Intensity in the image plane at time point t_k can be approximated as linear superposition of the intensities of the original image and sources of turbulence placed at reference time t_0.

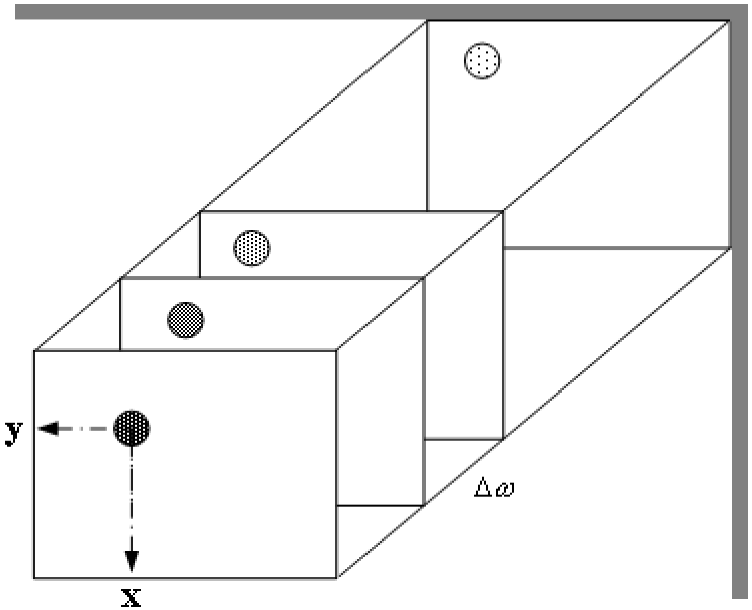
$$I_{ik}(t_k, x, y) = \sum_{n=1}^{N0} a_{kn}(\Delta t_{kn}) I_{0n}(t_0, x, y)$$

I.Kopriva, et al., Optics Communications, Vol. 233, Issue 1-3, pp.7-14, 2004.

ICA representation of the image sequence

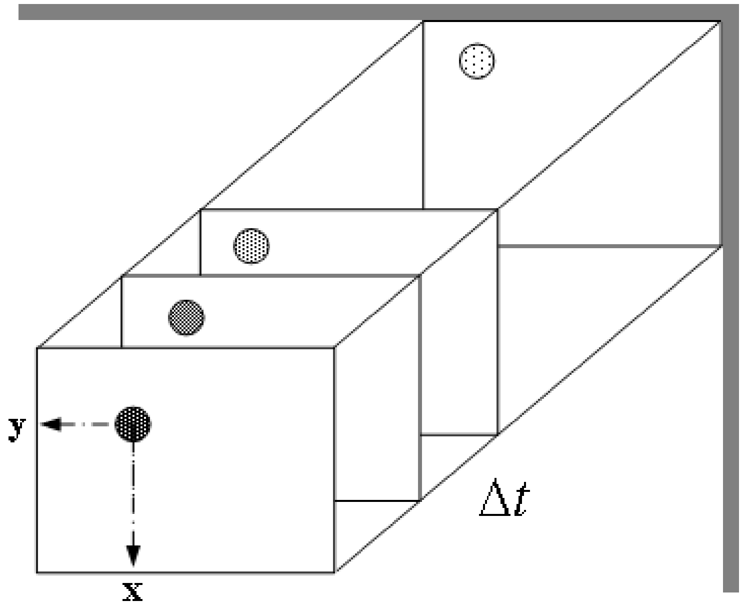
$$\mathbf{I}_i(x,y) = \mathbf{A}\mathbf{I}_\mathbf{o}(x,y) \quad + \quad \nu(x,y)$$

Image cube for multispectral imaging

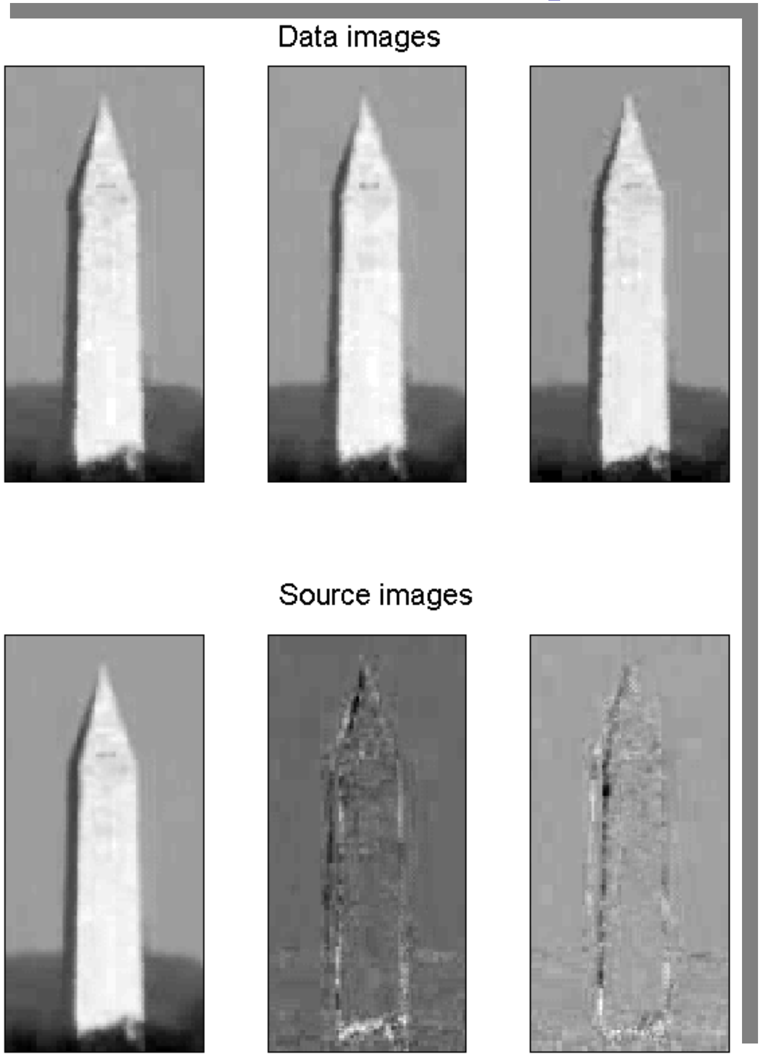


$$\omega \Leftrightarrow t$$

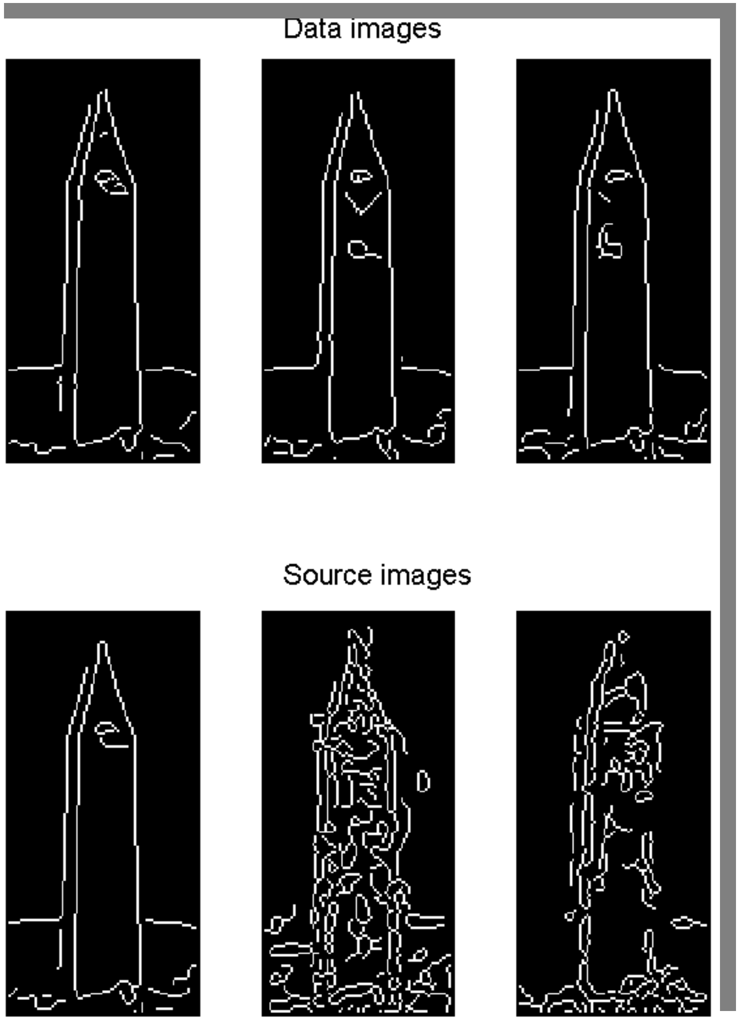
Image cube for video sequence



Experimental results



- Three randomly selected frames with nonzero mutual information



- Cany’s method of edge extraction gives the best result for the ICA recovered object image.
- Important to reduce the false alarm rate in automatic target recognition (ATR).

ICA and for dictionary learning: inpainting and denoising of high density impulsive noise

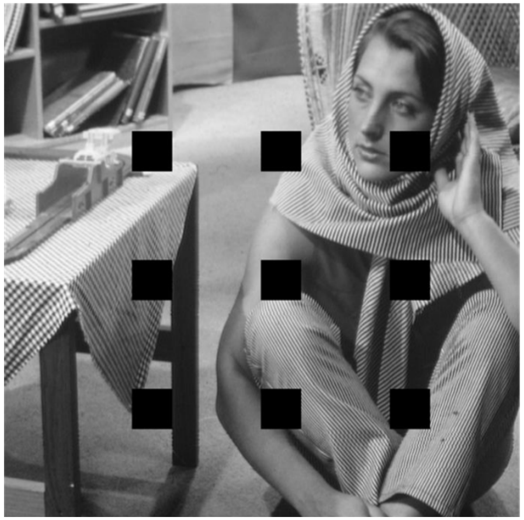
M. Filipovic, I.Kopriva, Inverse Problems and Imaging, Vol. 5, no. 4, pp.815-841, 2011.

Inpainting and denoising of salt and pepper noise

- recovery of lost or corrupted part of images/missing data reconstruction (also known as matrix/tensor completion)

Random pattern/pepper noise

Random pattern/salt noise



Pattern with structure (wholes)

Problem formulation

Column-wise vectorized image $\mathbf{x}\in R^{n}$ is to be reconstructed from the vector of known pixels $\mathbf{y}\in R^{l}$, $l < n$.

It is assumed that $\mathbf{x}$ can be represented by k<<n significant coefficients over the known m-dimensional dictionary $\mathbf{D}\in R^{n\times m}$: $\mathbf{x}\approx\mathbf{Dc}$, $\|\mathbf{c}\|_0=k$ and $m\geq n$.

Dictionary $\mathbf{D}$ is a basis when $m=n$ and is a frame when $m>n$.

Vector of known pixels $\mathbf{y}$ is related to vector of unknown pixels $\mathbf{x}$: $\mathbf{y}=\mathbf{Mx}$, $\mathbf{M}\in R^{l\times n}$. $\mathbf{M}$ contains 0s and 1s representing layout of the missing values.

Hence inpainting problem is described as:

$$\mathbf{y}=\mathbf{Mx}=\mathbf{MDc}=\Phi\mathbf{c}$$

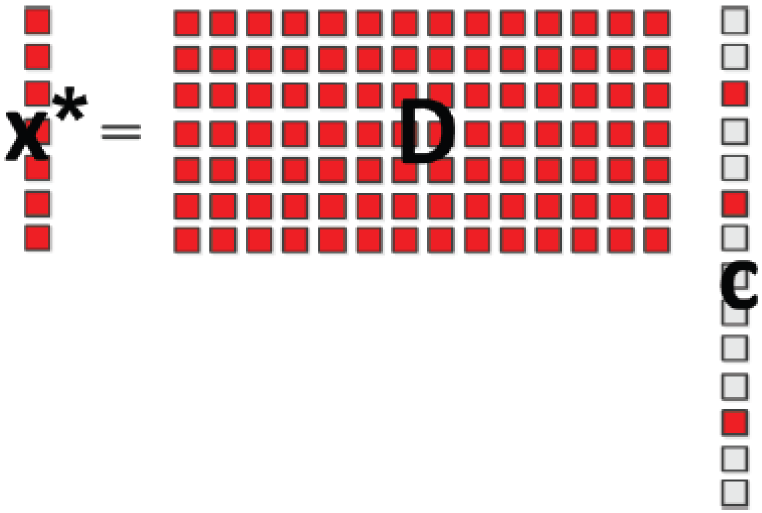
Problem formulation

Once the estimate of $\mathbf{c}$ denoted as $\hat{\mathbf{c}}$ is obtained the estimate of $\mathbf{x}$ is obtained as: $\hat{\mathbf{x}} = \mathbf{D}\hat{\mathbf{c}}$.

In the no noise scenario underdetermined system of equations $\mathbf{y} = \Phi\mathbf{c}$ has a unique solution if the number of **known pixels *l* satisfies: *l* ≥2*k***.*

Sparseness assumption $k \ll n$ is necessary because $\mathbf{y} = \Phi\mathbf{c}$ is ill-posed.

k depends on **D**, whereas **D** can be learned/ trained such that *k* is small !!!!



*J. Tropp and S. J. Wright, Computational methods for sparse solution of linear inverse problems, Proc. of the IEEE, 98 (2010), 948-958.

Problem formulation

The dictionary learning problem is organized in the *patch* space. Then, **x** denotes vectorized image *patch* $\mathbf{I} \in R^{\mathrm{sqrt}(n) \times \mathrm{sqrt}(n)}$.

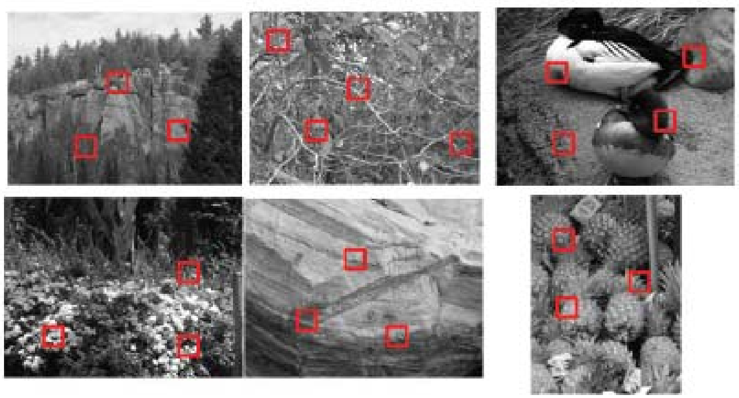
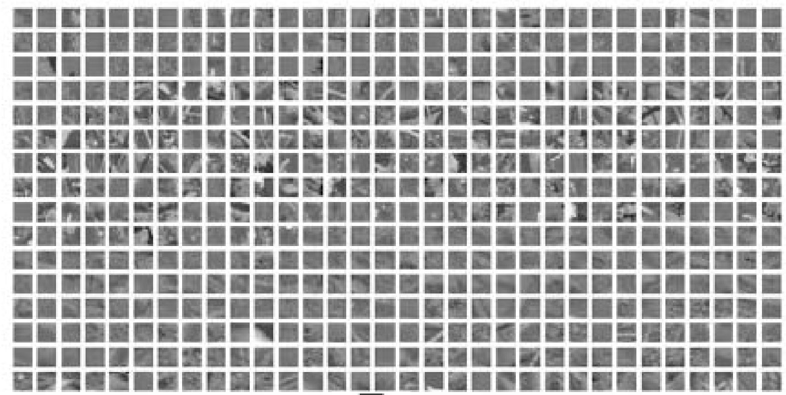


Image patches

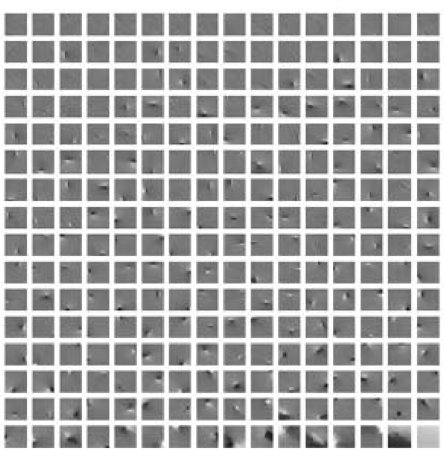


Vectorization preprocessing

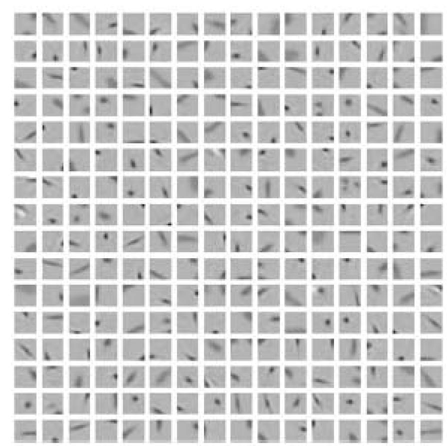
$$X = \left[x_1, \ldots, x_N \right]$$

ICA, K-SVD

$$X = DC$$



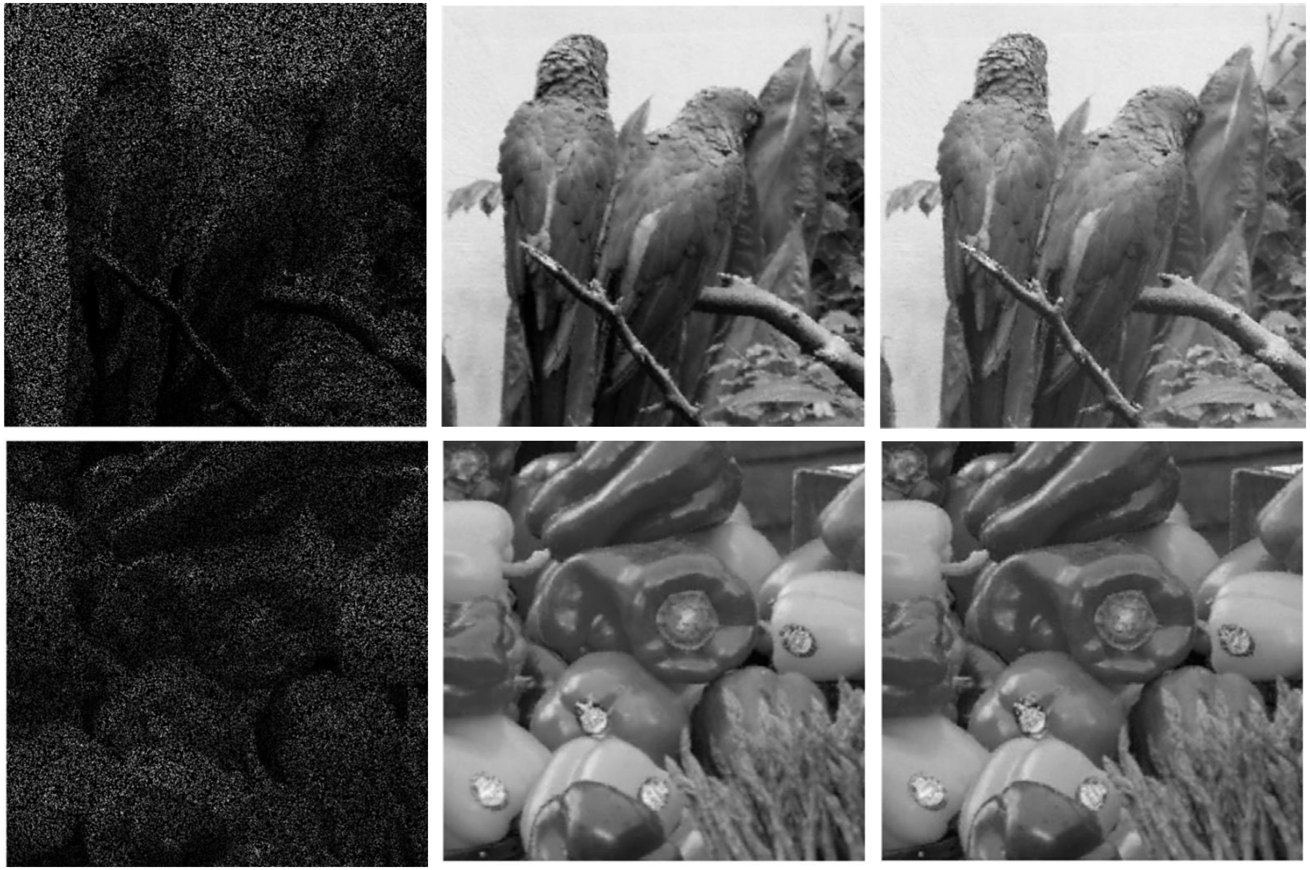
ICA basis



K-SVD basis

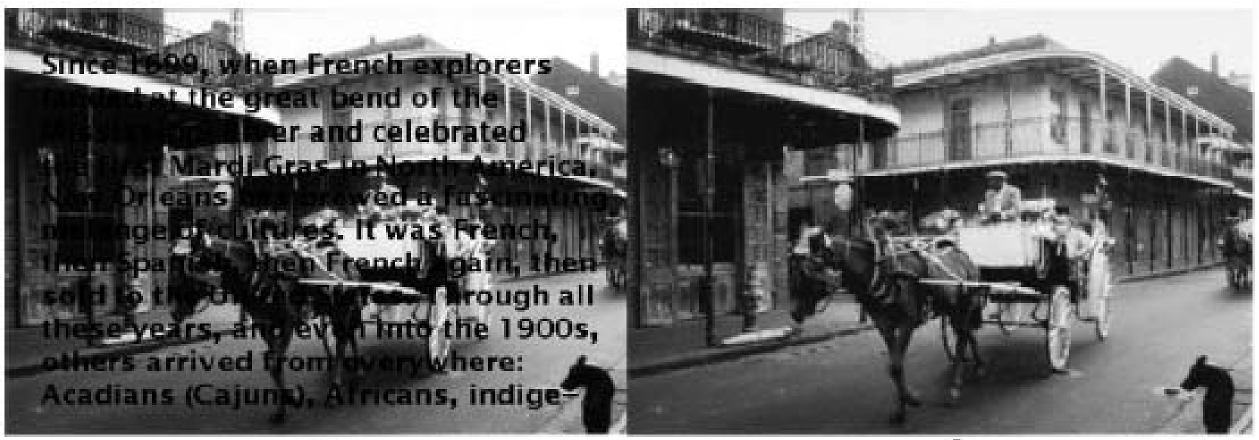
Experiments: denoising 80% dense pepper noise

Salt and pepper noise generates random pattern of missing values and that is the easiest inpainting problem to solve. **Complete basis: 256**x**256**.



Corrupted images ICA basis K-SVD basis

Experiments: inpainting missing pattern with a “text” structure



a b

FIGURE 12. Inpainting for text removal. a) Image with text. b) Inpainting using ICA learned basis.

For this example the images and corresponding mask of missing pixels were taken from: http://www.dtic.upf.edu/~mbertalmio/restoration0.html.

Dependent component analysis

Increasing statistical independence

- We want to find a linear operator T with the property that $T(s_m)$ and $T(s_n)$ are more independent than s_m and s_n $\forall m, n$.

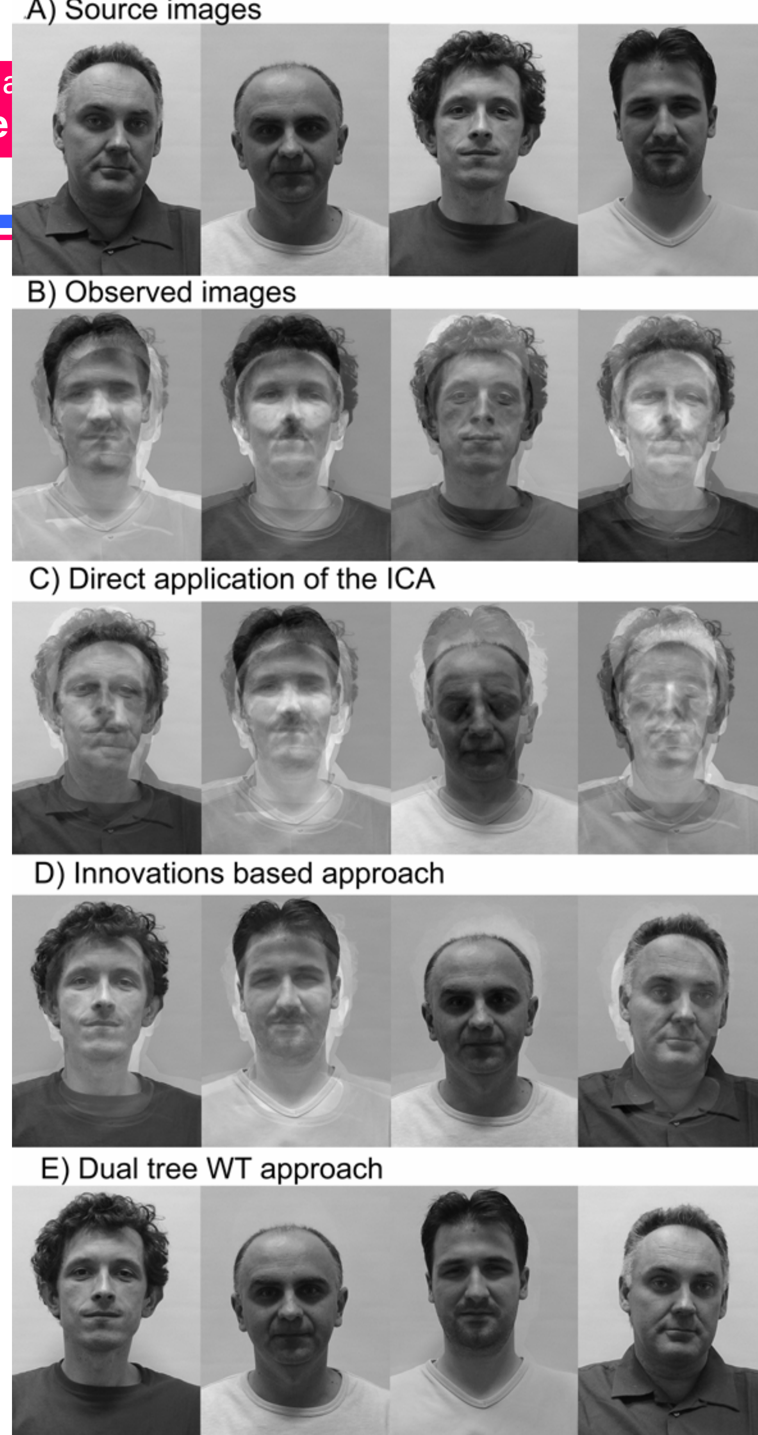
- Then, $\mathbf{W} \cong \mathbf{A}^{-1}$ is learnt by applying ICA on $T(\mathbf{x})=\mathbf{A}T(\mathbf{s})$.

- How to find linear operator T (problem for DCA)?

Separation of images of human faces

• Wavelet packets approach to blind separation of statistically dependent sources is tested on separation of the images of human faces. They are known to be highly dependent i.e. people are quite similar (statistically).

• Background Gaussian noise has been added as wide-band interferer to all source images with an average SNR $\cong$30dB.


A) Source images
B) Observed images
C) Direct application of the ICA
D) Innovations based approach
E) Dual tree WT approach

Robust demarcation of the basal cell carcinoma

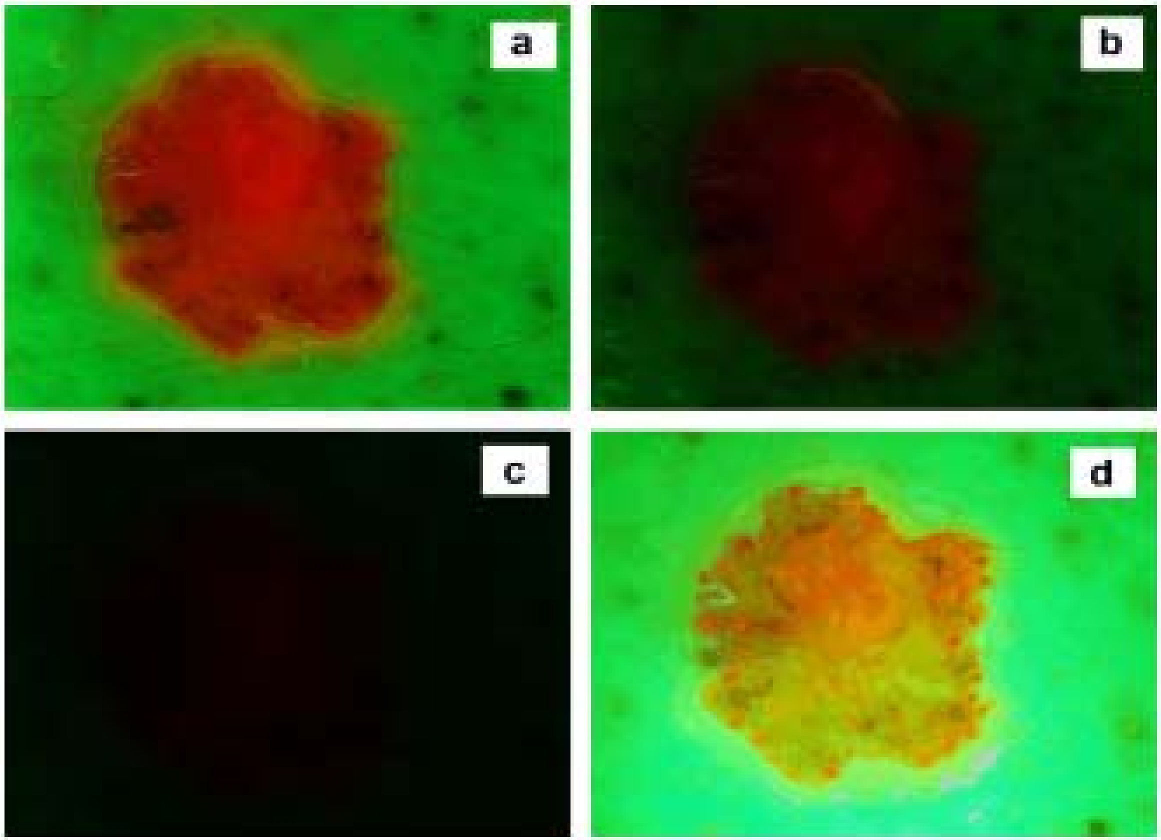
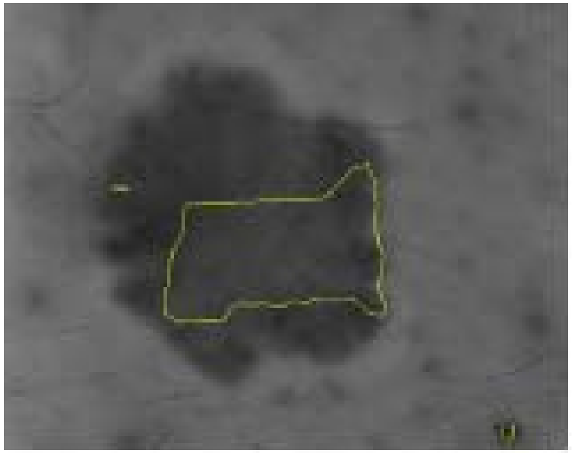


Fig. 1. RGB fluorescent images of the BCC from the first patient acquired under different intensities of illumination: (a) illumination with the maximal intensity I_0; (b) illumination with the intensity $I_0/9.15$; illumination with the intensity $I_0/73.47$; (d) RGB fluorescent image with demarcation line manually marked by the red dots.

I. Kopriva, A. Peršin, N. Puizina-Ivić, L. Mirić (2010). Robust demarcation of basal cell carcinoma by dependent component analysis-based segmentation of multi-spectral fluorescence image, *Journal Photochemistry and Photobiology B: Biology*, vol. 100, pp. 10-18

Robust demarcation of the basal cell carcinoma



Evolution curve after 700 iterations on gray scale image of the tumor.

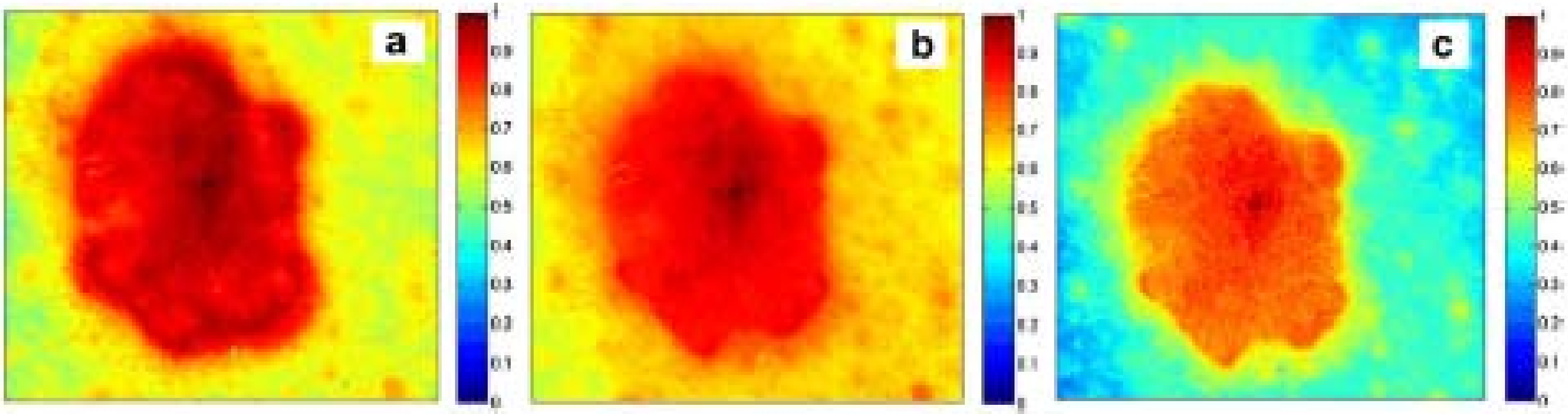


Fig. 6. BCC spatial maps ... extracted from fluorescent RGB images shown in Fig. 1a–c by means of EFICA algorithm [36]. Extracted maps are normalized on interval [0, 1] and shown in pseudo-color scale.

Robust demarcation of the basal cell carcinoma

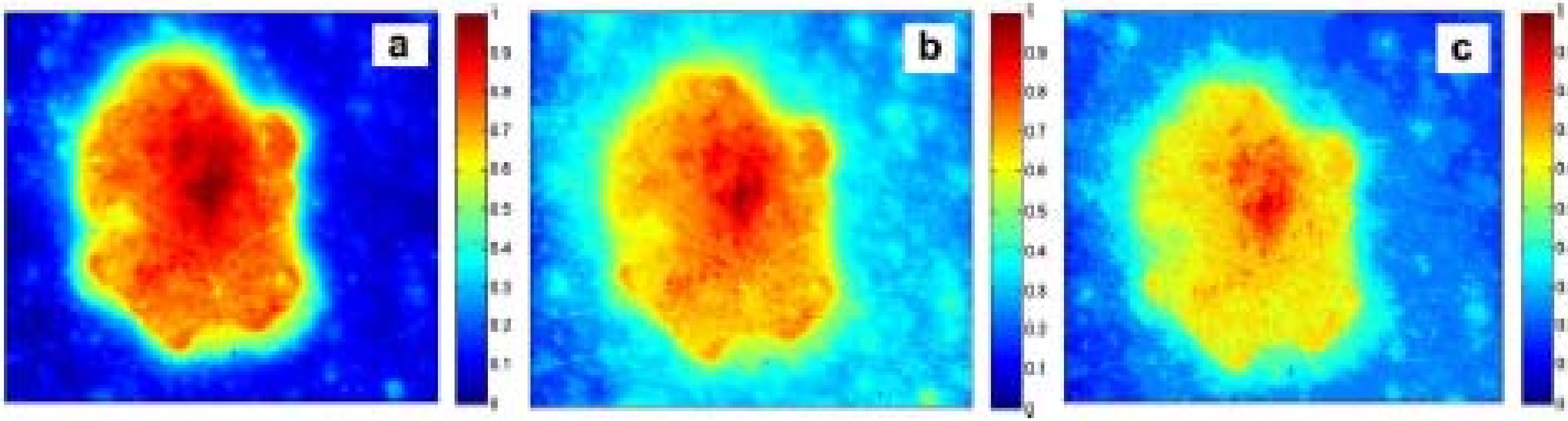


Fig. 7. BCC spatial maps in extracted from fluorescent RGB images shown in Fig. 1a–c by means of DCA-HPF algorithm. Extracted maps are normalized on interval [0, 1] and shown in pseudo-color scale.

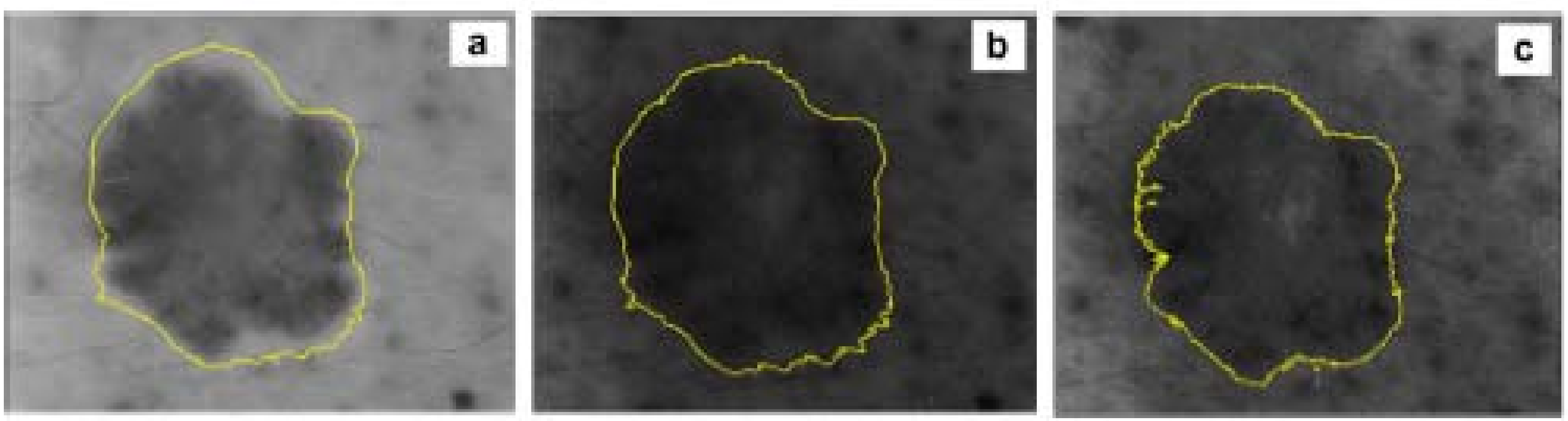


Fig. 8. BCC demarcation lines calculated by means of Canny's edge extraction method from spatial maps shown in Fig. 7a–c, with a fixed threshold set to 0.5. Demarcation lines were superimposed on the gray scale version of the fluorescent RGB images shown in Fig. 1a–c.

Underdetermined blind source separation:

sparse component analysis (SCA)

and

nonnegative matrix factorization (NMF)

Underdetermined BSS

•uBSS occurs when number of measurements *N* is less than number of sources *M*. Resulting system of linear equations

$$\mathbf{x}=\mathbf{A}\mathbf{s}$$

is underdetermined. Without constraints on **s** unique solution does not exist even if **A** is known:

$$\mathbf{s}=\mathbf{s}_p + \mathbf{s}_h = \mathbf{A}^{\dagger}\mathbf{x} + \mathbf{V}\mathbf{z} \quad \mathbf{A}\mathbf{V}\mathbf{z}=\mathbf{0}$$

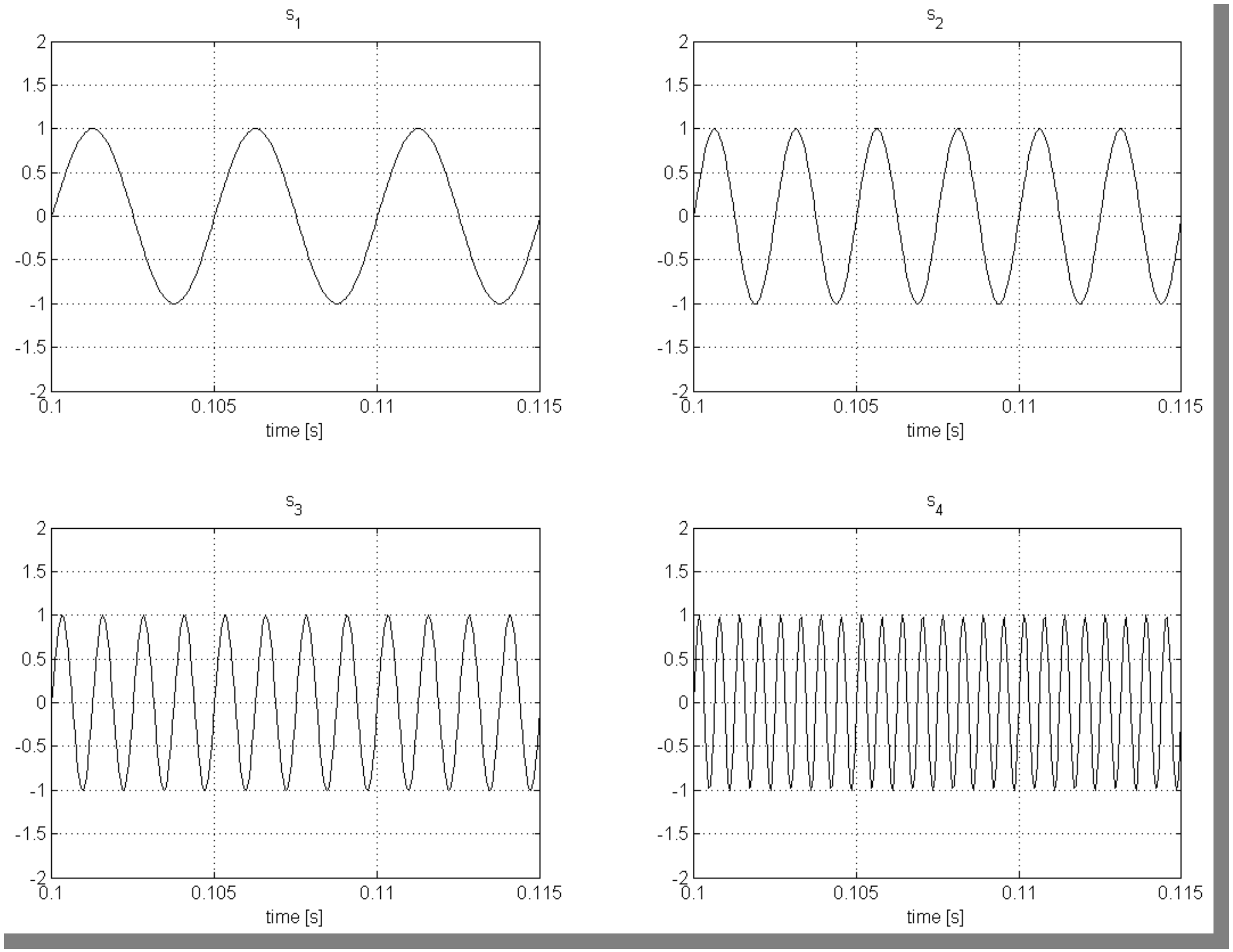
where **V** spans null-space of **A** that is *M*-*N* dimensional.

• However, if **s** is sparse enough **A** can be identified and unique solution for **s** can be obtained. This is known as sparse component analysis (SCA).

Blind separation of four sine signals from two mixtures

Four sinusoidal signals with frequencies 200Hz, 400Hz, 800Hz and 1600Hz.

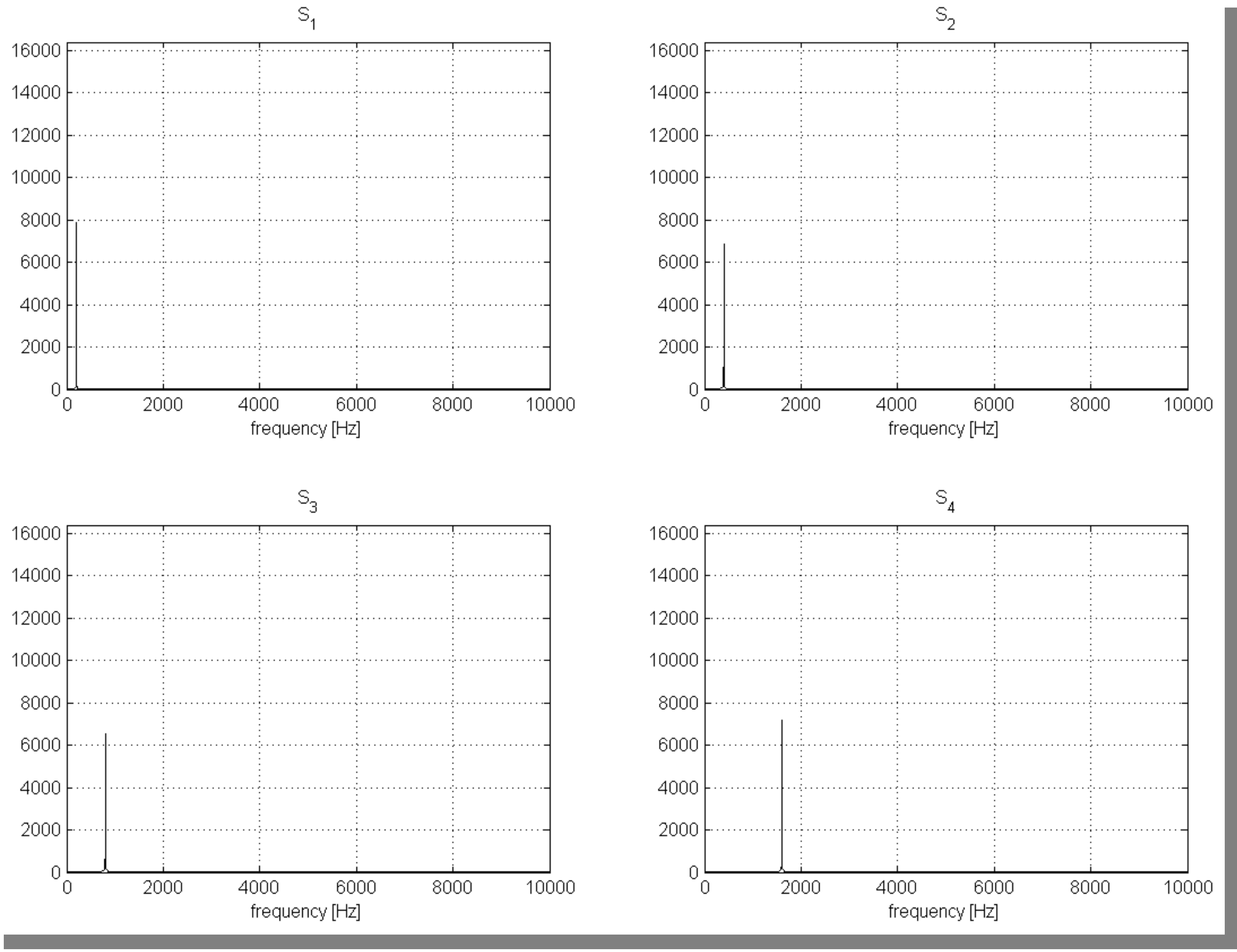
<u>TIME DOMAIN</u>



Blind separation of four sine signals from two mixtures

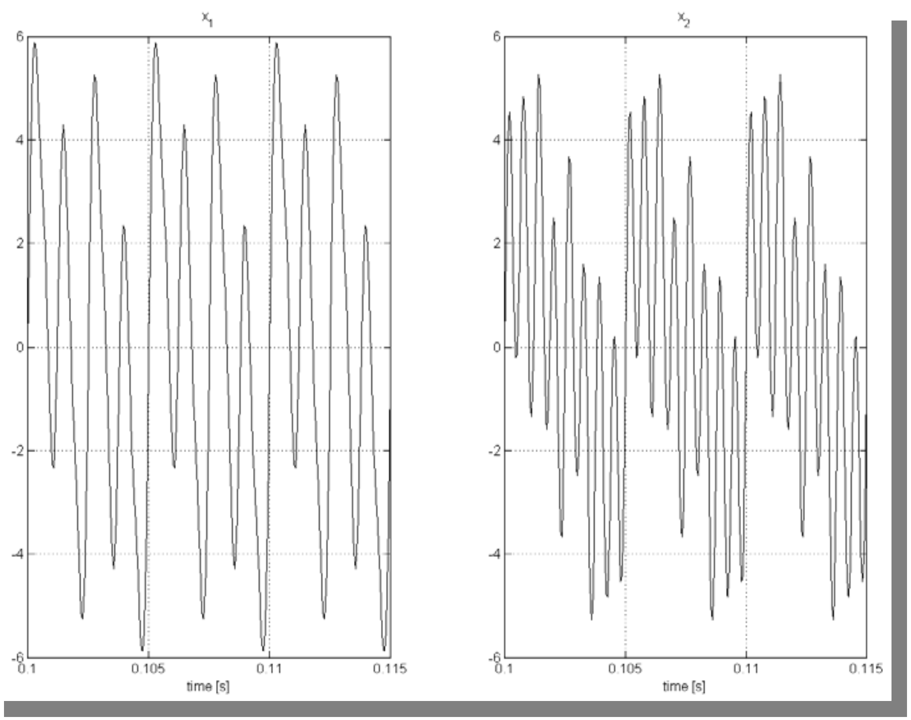
Four sinusoidal signals with frequencies 200Hz, 400Hz, 800Hz and 1600Hz.

FREQUENCY DOMAIN

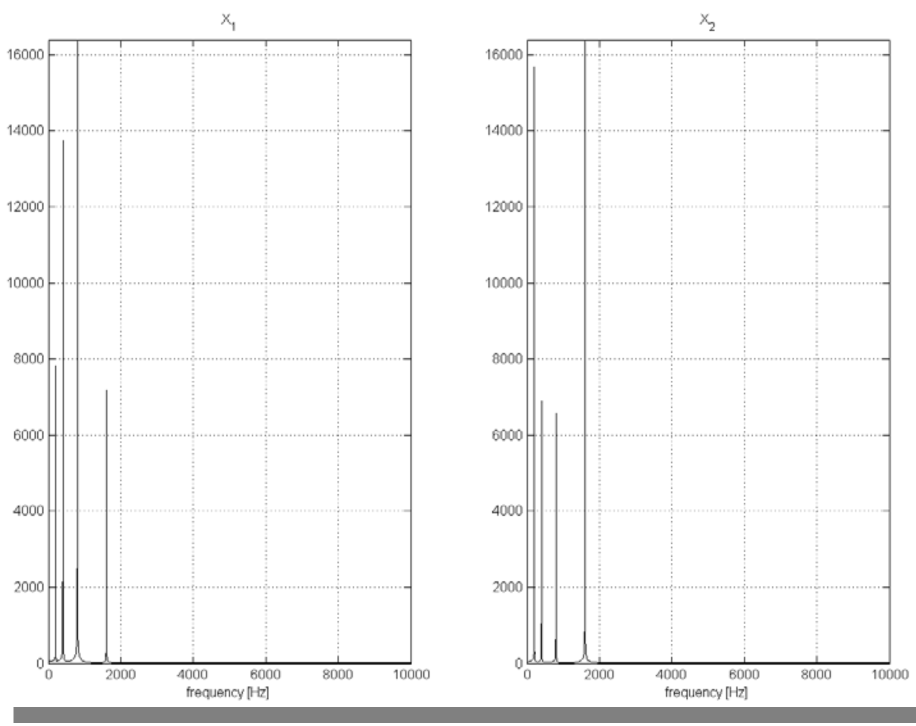


Blind separation of four sine signals from two mixtures

Two mixed signals



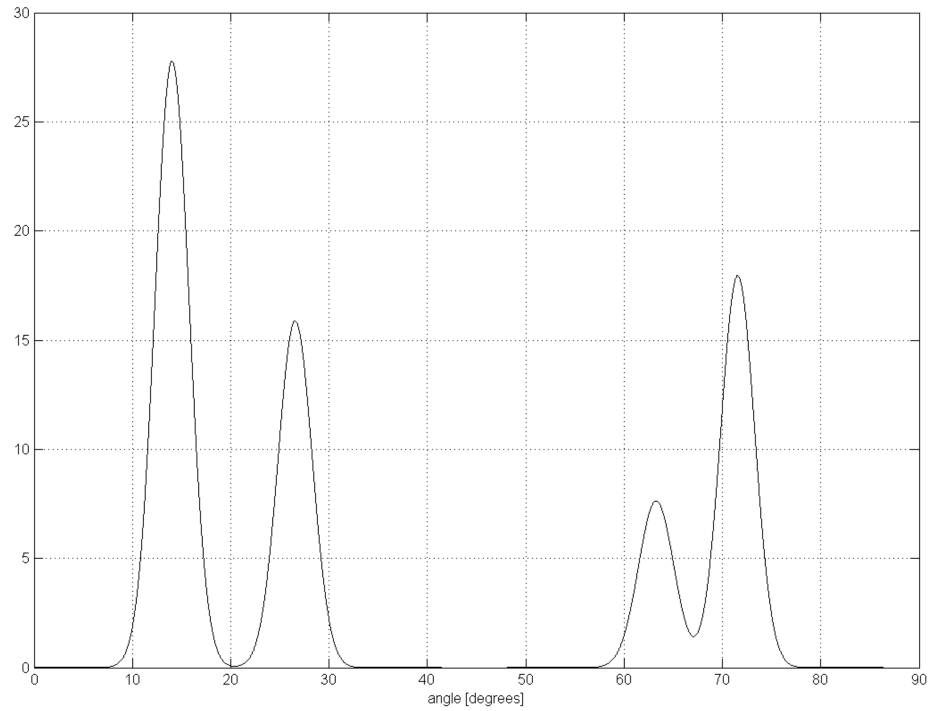
TIME DOMAIN



FREQUENCY DOMAIN

Blind separation of four sine signals from two mixtures

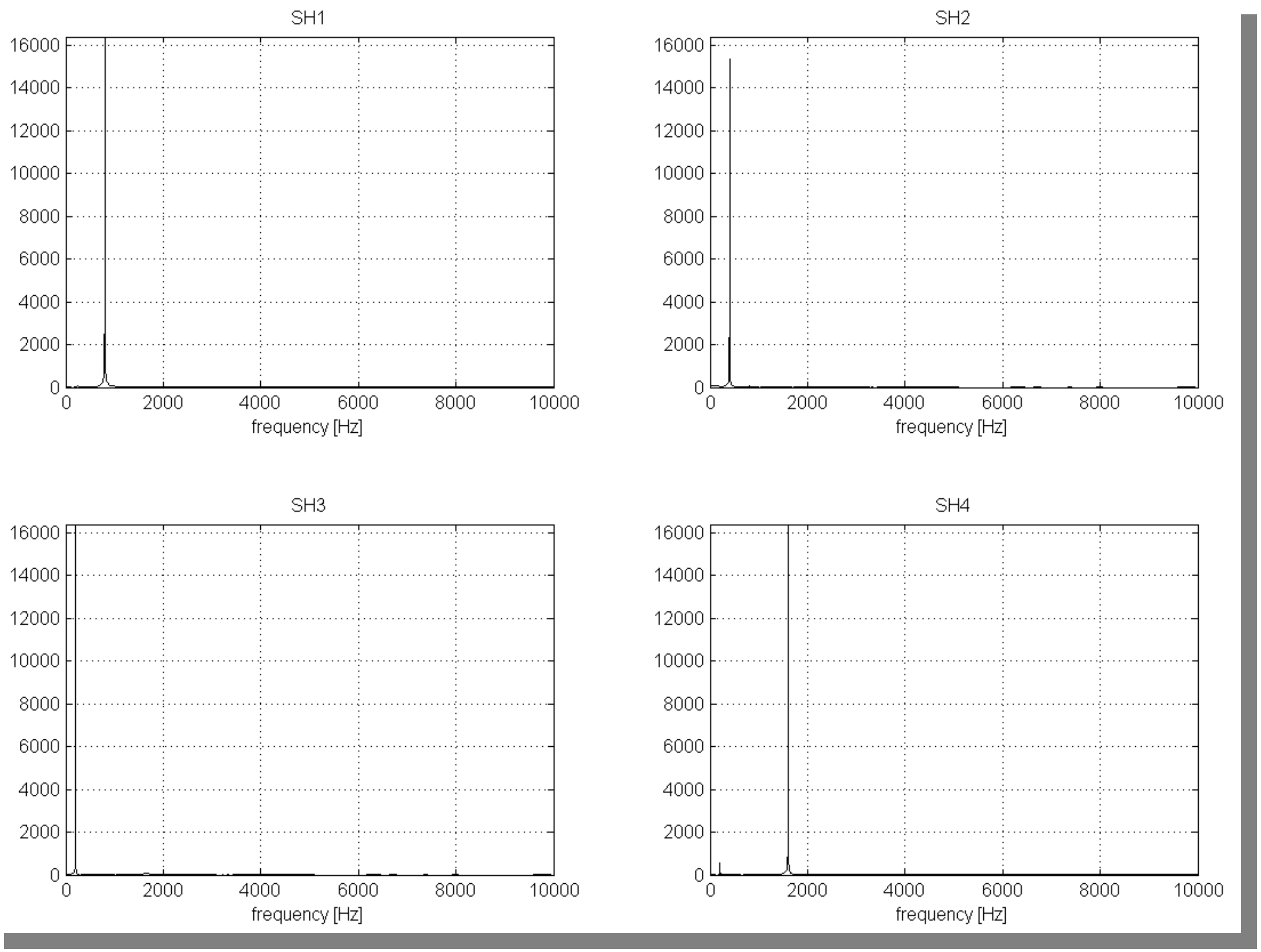
Clustering function



A=[63.44^0 26.57^0 14.04^0 71.57^0]

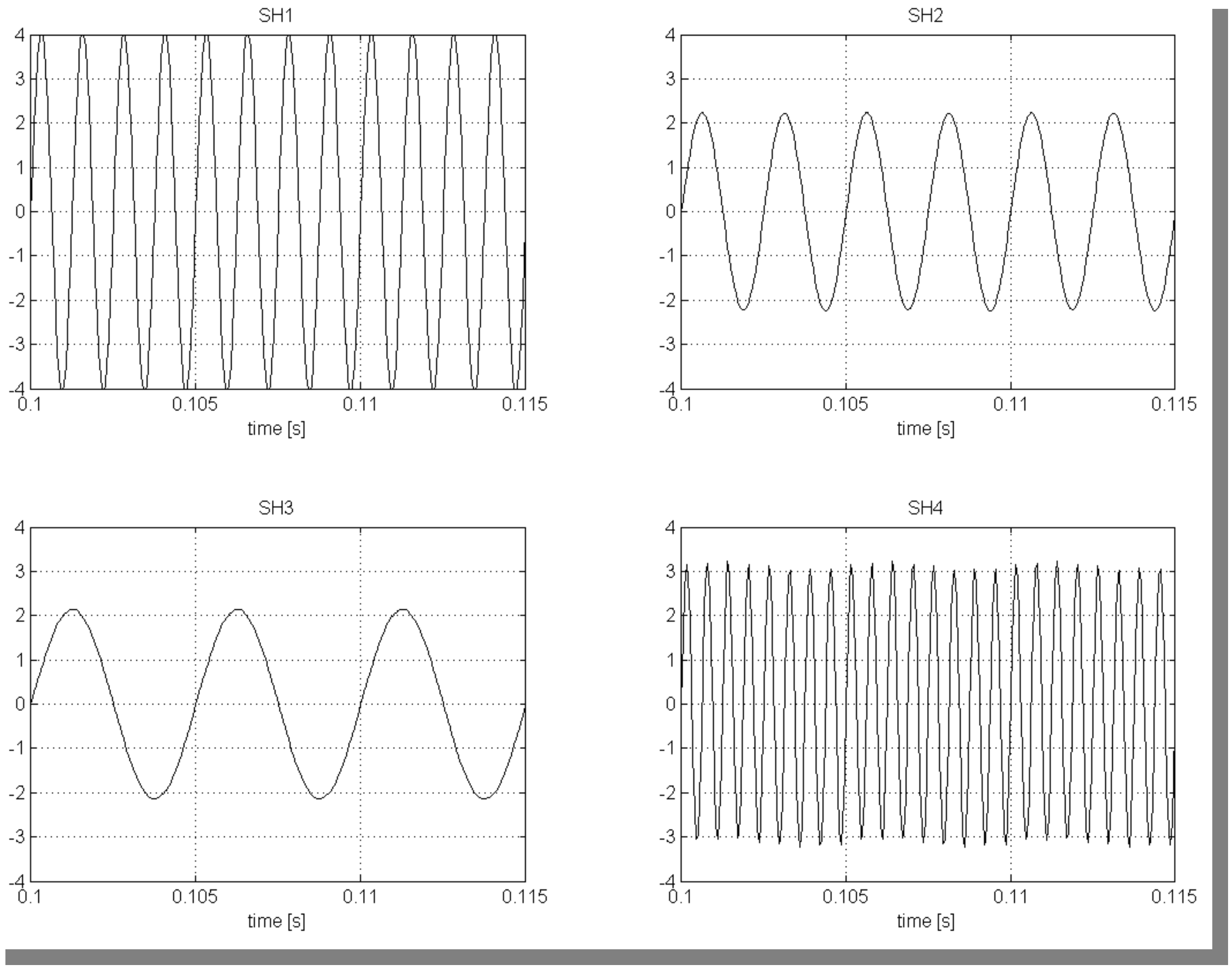
AH=[14.03^0 26.55^0 63.26^0 71.55^0]

Blind separation of four sine signals from two mixtures



Magnitudes of the estimated sources in FREQUENCY DOMAIN

Blind separation of four sine signals from two mixtures

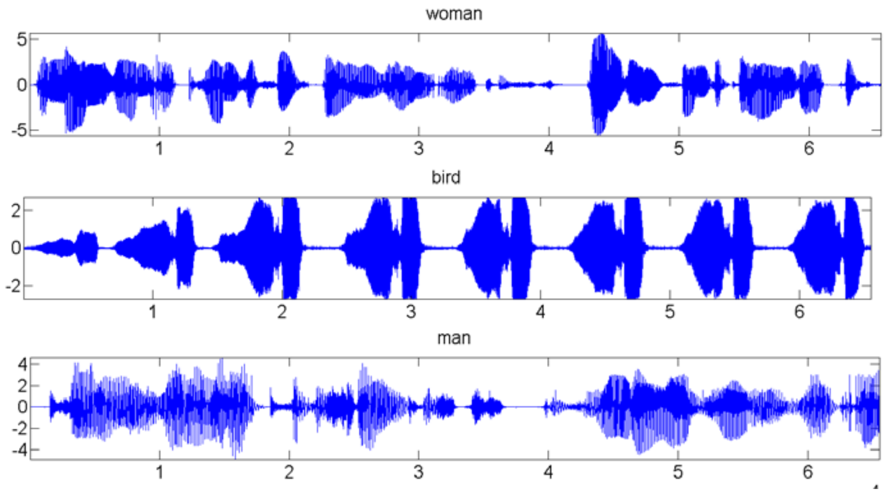


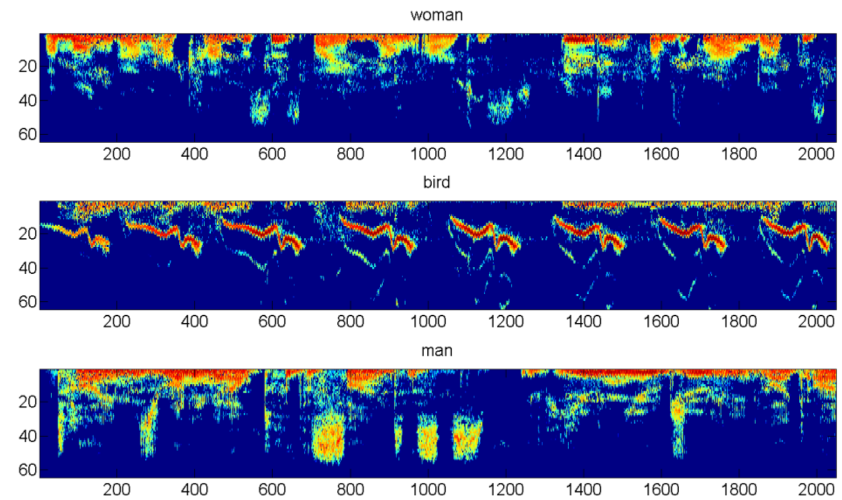
Estimated sources in TIME DOMAIN

Blind separation of three sounds from two mixtures

Blind separation of three sounds from two mixtures

Three source signals are female and male voice and bird’s sound:



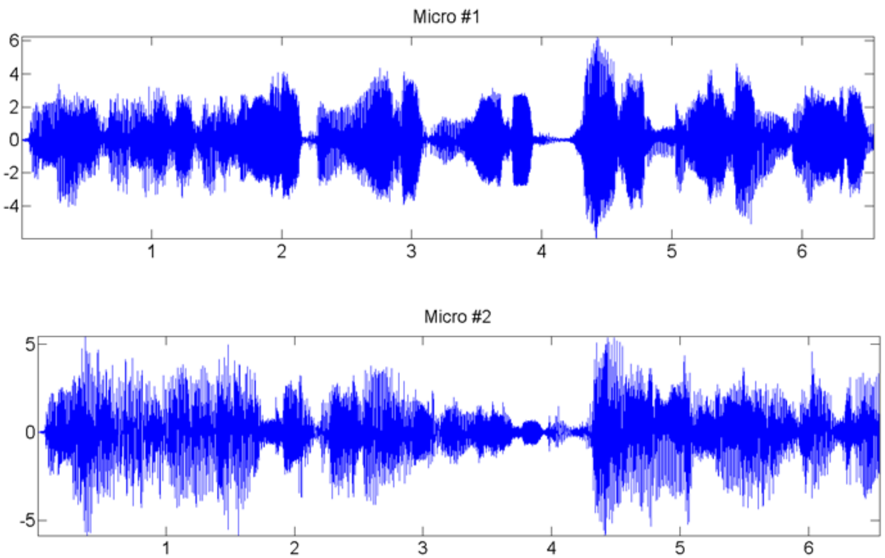


Time domain waveforms

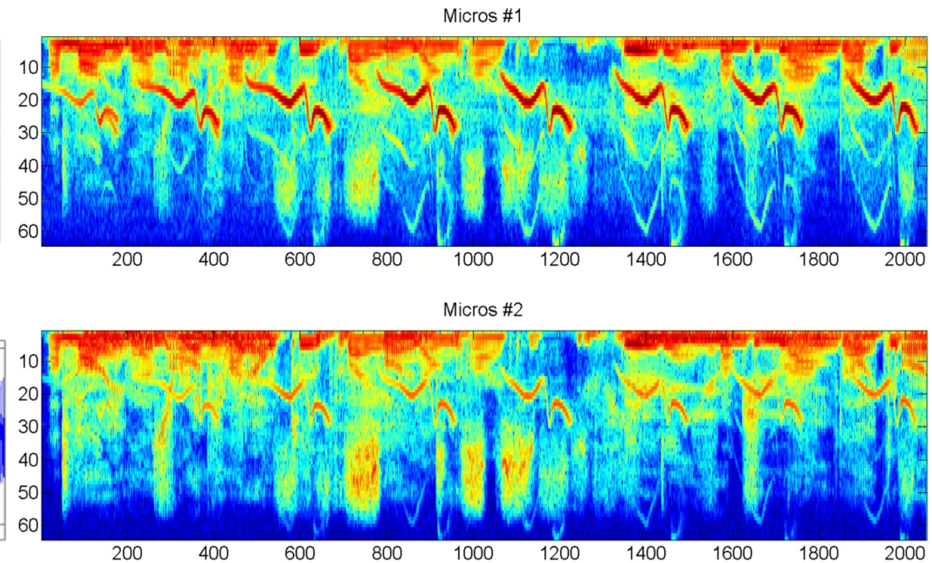
Time-frequency representations

Blind separation of three sounds from two mixtures

Two mixtures of sounds:



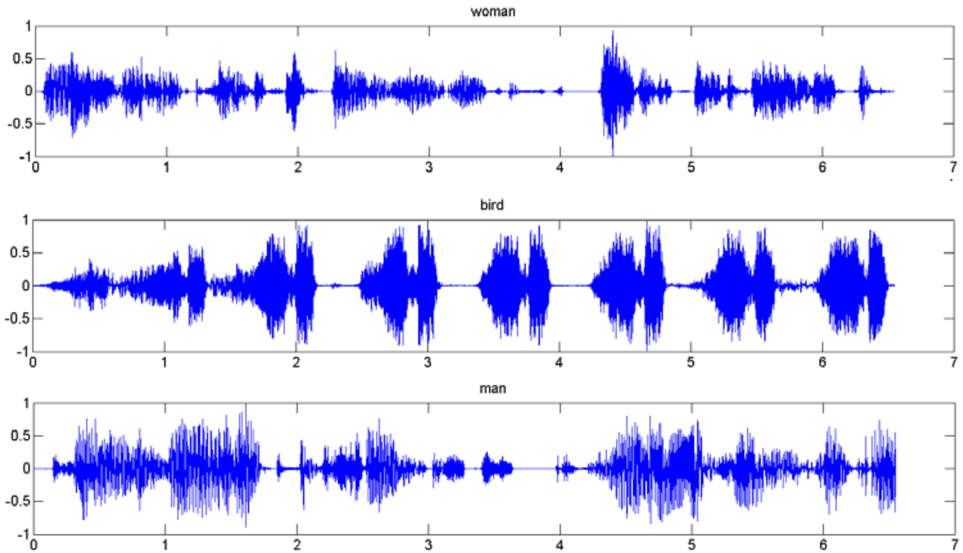
Time domain waveforms



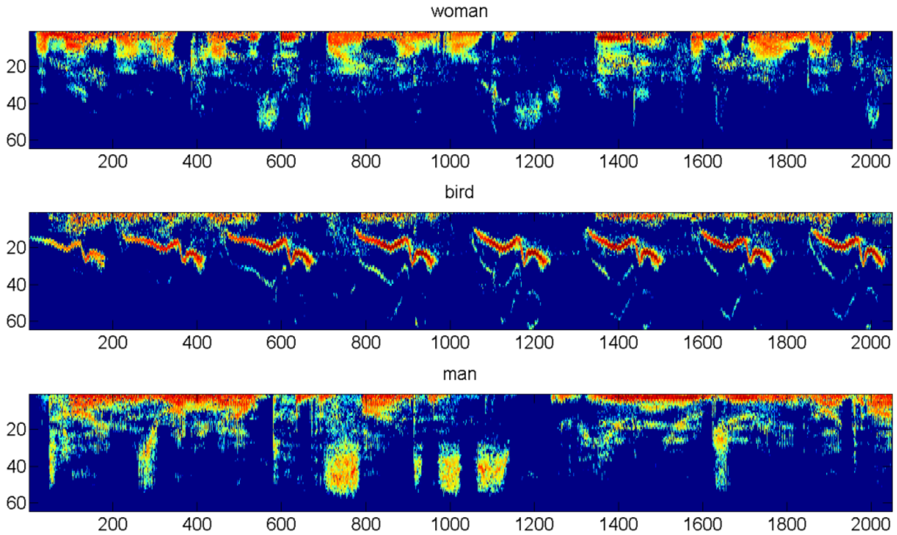
Time-frequency representations

Blind separation of three sounds from two mixtures

Three extracted sounds combining clustering on a set of single source points and linear programming in time-frequency domain:



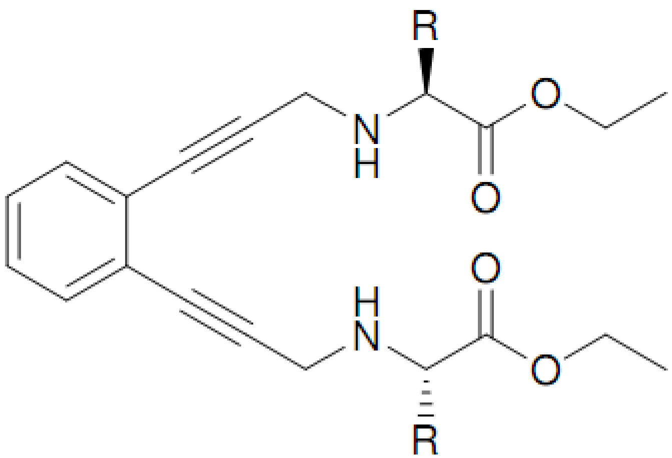
Time domain waveforms



Time-frequency representations

Blind extraction of analytes (pure components) from mixtures of chemical compounds

I. Kopriva, I. Jerić (2010). Blind separation of analytes in nuclear magnetic resonance spectroscopy and mass spectrometry: sparseness-based robust multicomponent analysis, **Analytical Chemistry 82**:1911-1920.
I. Kopriva, I. Jerić, V. Smrečki (2009). Extraction of multiple pure component ^{1}H and ^{13}C NMR spectra from two mixtures: novel solution obtained by sparse component analysis-based blind decomposition, **Analytica Chimica Acta**, vol. 653, pp. 143-153.
I. Kopriva, I. Jerić (2009). Multi-component Analysis: Blind Extraction of Pure Components Mass Spectra using Sparse Component Analysis, **Journal of Mass Spectrometry**, vol. 44, issue 9, pp. 1378-1388.
I. Kopriva, I. Jerić, A. Cichocki (2009). Blind Decomposition of Infrared Spectra Using Flexible Component Analysis," **Chemometrics and Intelligent Laboratory Systems** 97**.**



5 R=H

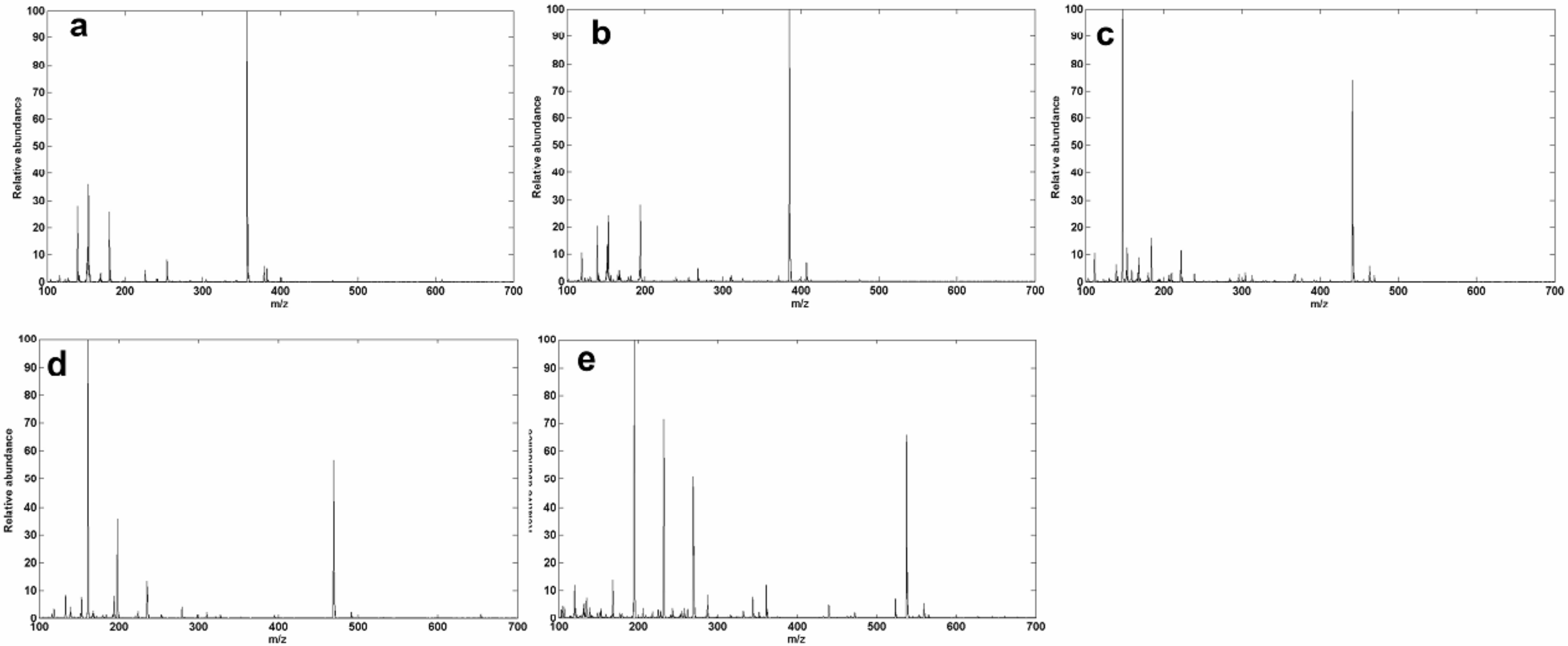
6 R=CH_3

7 R=$CH(CH_3)_2$

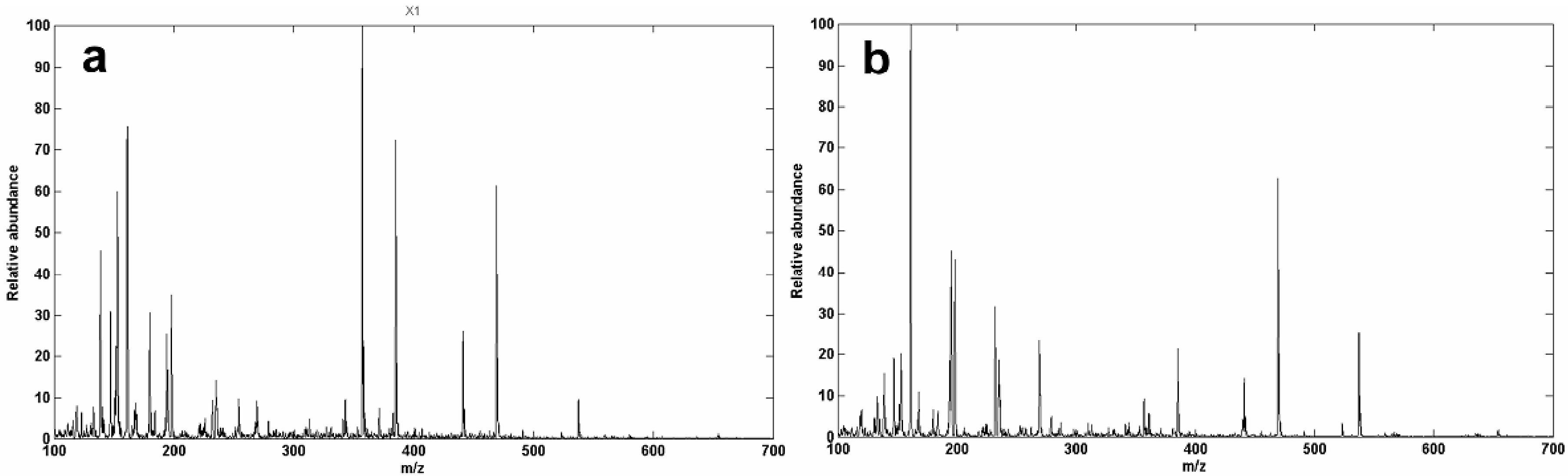
8 R=$CH_2CH(CH_3)_3$

9 R=$CH_2C_6H_5$

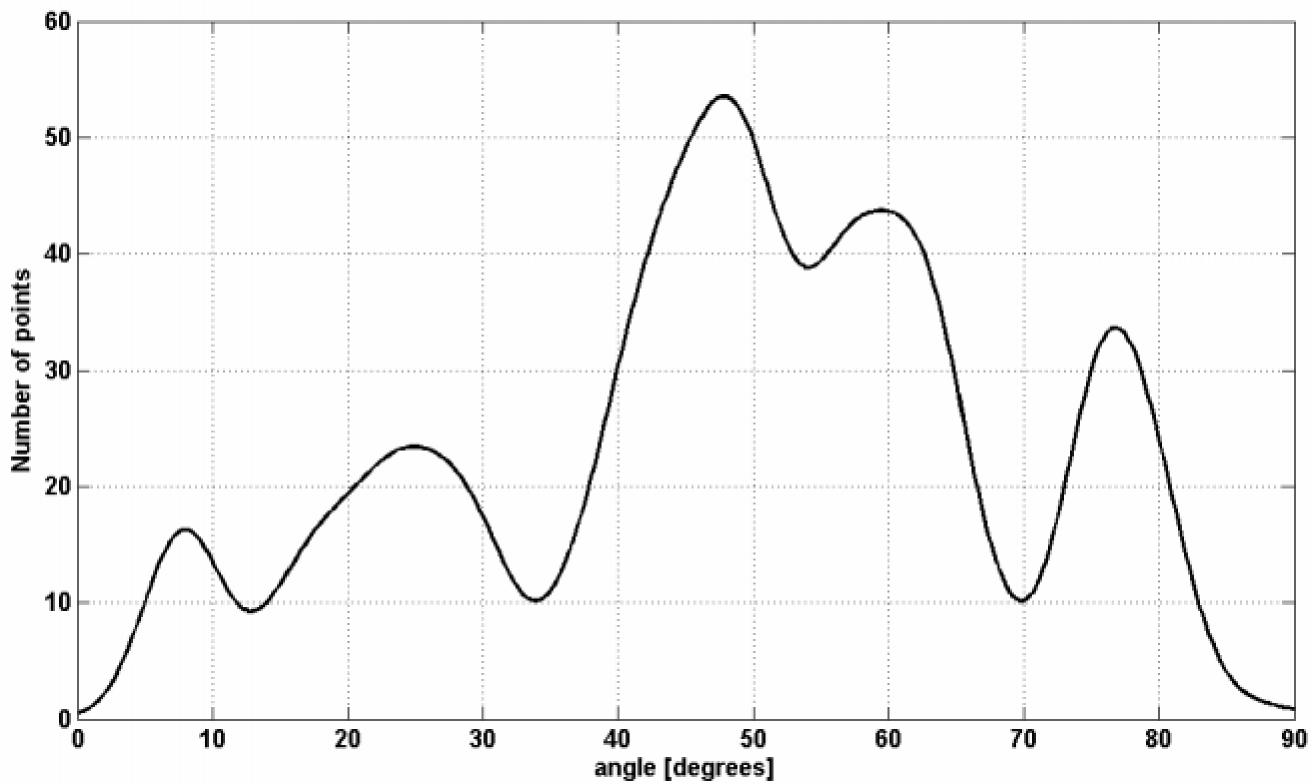
Figure S-1. Chemical structure of five pure components.



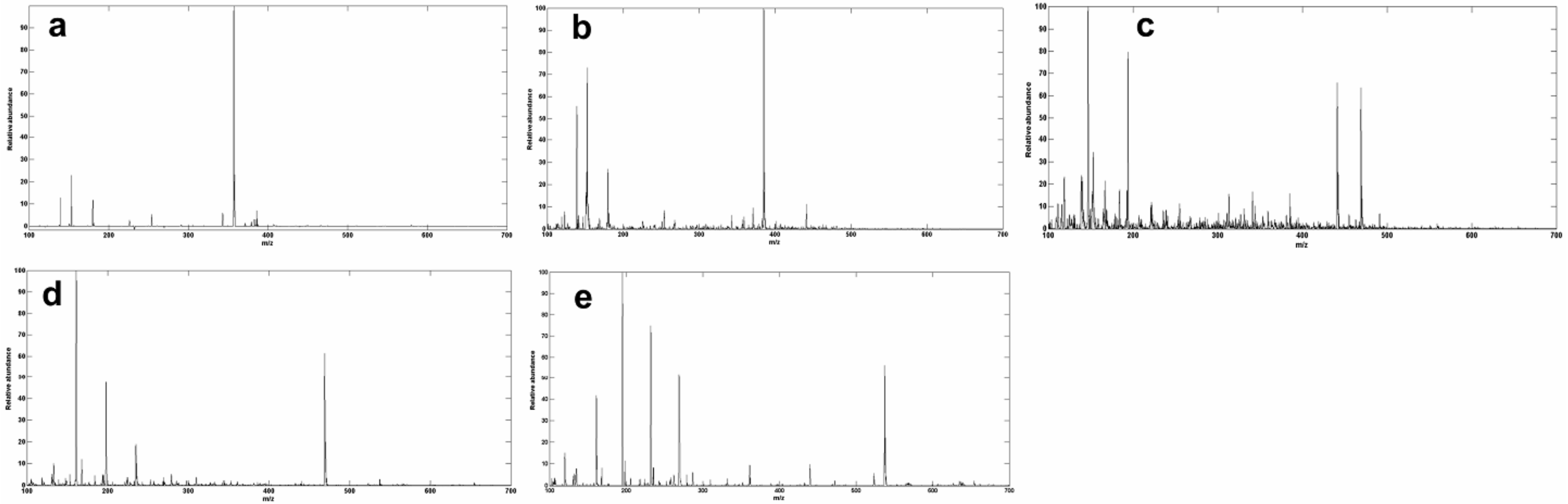
Mass spectra of five pure components.



Mass spectra of two mixtures



Dana clustering function in the mixing anagle domain. Five peaks indicate presence of five components in the mixtures spectra.



Estimated mass spectra of five pure components.

Table S-1. Normalized correlation coefficients for (a) pure analytes **5-9**; (b) analytes **5-9** estimated on 290 SAPs detected by using analytical representation (3) and *clusterdata* algorithm.*

entry		**An_5**	**An_6**	**An_7**	**An_8**	**An_9**
a	**An_5**	**1**	0.1268	0.0456	0.0266	0.0075
	An_6	0.1268	**1**	0.0321	0.0332	0.0379
	An_7	0.0456	0.0321	**1**	0.0134	0.0030
	An_8	0.0265	0.0332	0.0134	**1**	0.0029
	An_9	0.0075	0.0379	0.0030	0.0029	**1**
b	**$\hat{A}n_5$**	**0.9038**	0.0305	0.0044	0.0002	0.0120
	$\hat{A}n_6$	0.3162	**0.8294**	0.1198	0.0325	0.0043
	$\hat{A}n_7$	0.0959	0.2334	**0.7275**	0.2009	0.0038
	$\hat{A}n_8$	0.0043	0.0038	0.0124	**0.9736**	0.0293
	$\hat{A}n_9$	0.0121	0.0161	0.0073	0.2097	**0.9437**

*An_5-An_9 pure analytes **5-9**; $\hat{A}n_5$- $\hat{A}n_9$ estimated analytes **5-9**.

Nonnegative matrix factorization (NMF)

NMF algorithms solve blind decomposition problem

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

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

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

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

R. Zdunek, A. Cichocki, *Nonnegative matrix factorization with constrained second order optimization,* Signal Proc. **87** (2007) 1904-1916.

A. Cichocki, R. Zdunek, S.I. Amari, Hierarchical ALS Algorithms for Nonnegative Matrix Factorization and 3D Tensor Factorization, LNCS **4666** (2007) 169-176

A. Cichocki, A-H. Phan, R. Zdunek, and L.-Q. Zhang, "Flexible component analysis for sparse, smooth, nonnegative coding or representation," LNCS **4984**, 811-820 (2008).

A. Cichocki, R. Zdunek, S. Amari, Nonnegative Matrix and Tensor Factorization, IEEE Sig. Proc. Mag. **25** (2008) 142-145. A. Cichocki, and R. Zdunek, “Multilayer Nonnegative Matrix Factorization,” El. Letters **42**, 947-948 (2006).

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.

Nonnegative matrix factorization

Modern approaches to NMF problems have been initiated by Lee-Seung Nature paper, where it is proposed to estimate **A** and **S** through alternative minimization procedure of the possibly two different cost functions:

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

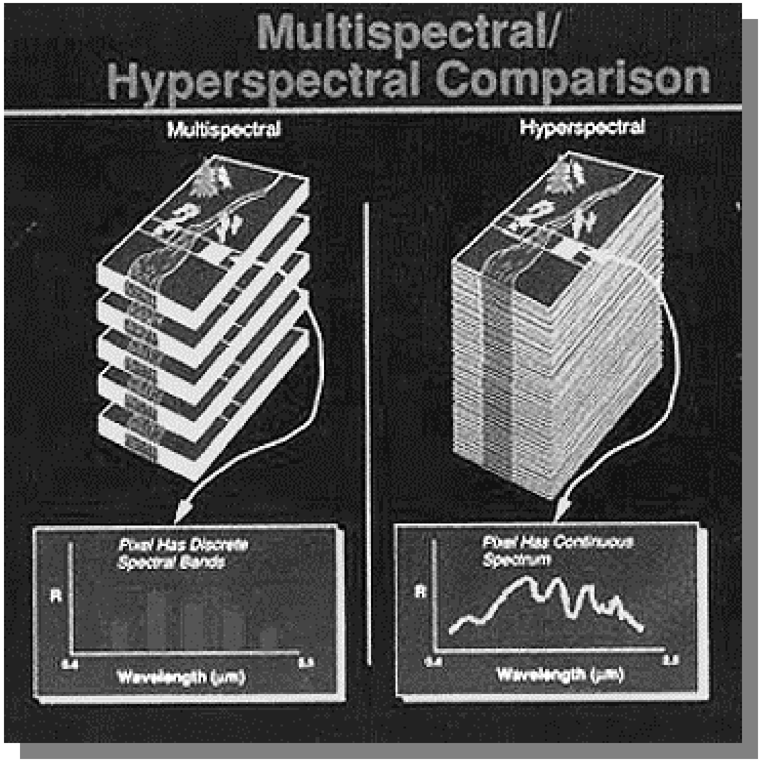
For ***k****=1,2,…, until convergence* **do**

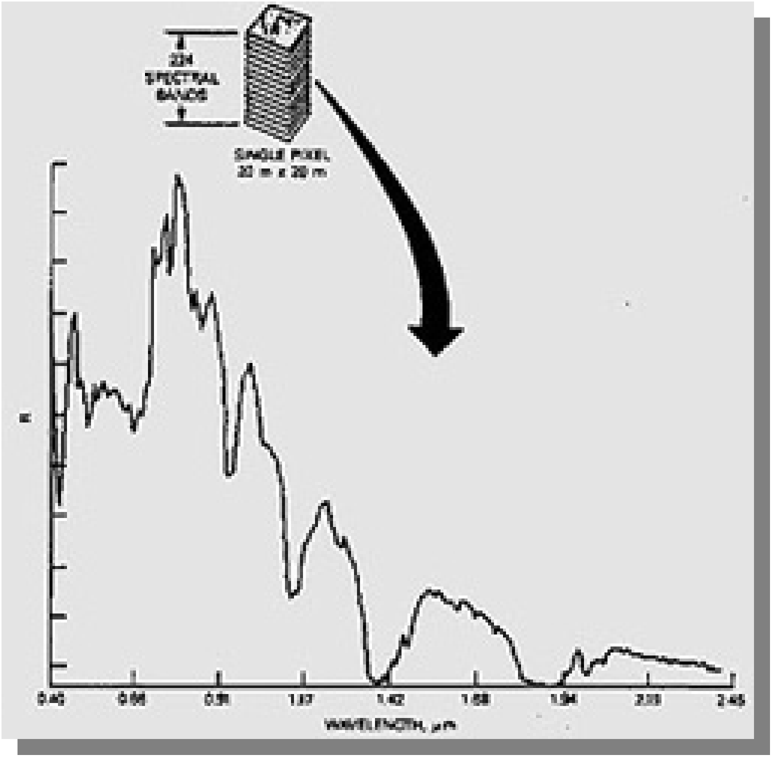
Step 1: $$\mathbf{S}^{(k+1)} = \underset{s_{mt}\geq 0}{\arg\min}\, D_{\mathbf{s}}\left(\mathbf{X} \middle\| \mathbf{A}^{(k)}\mathbf{S}\right)_{\mathbf{S}^{(k)}}$$

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

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

Unsupervised segmentation of multispectral images





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

❑ Objects with very similar reflectance spectra are *difficult to discriminate*.

Unsupervised segmentation of multispectral images

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

$$\mathbf{X} = \mathbf{A}\mathbf{S} = \sum_{m=1}^{M} \mathbf{a}_m \mathbf{s}_m$$

$$\begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \ldots & \mathbf{a}_M \end{bmatrix} \equiv \mathbf{A}$$

$$\begin{bmatrix} \mathbf{s}_1 & \mathbf{s}_2 & \ldots & \mathbf{s}_M \end{bmatrix}^T \equiv \mathbf{S}$$

X - measured data intensity matrix, $\mathbf{X} \in \mathbb{R}_{0+}^{N \times T}$

S - unknown class matrix, $\mathbf{S} \in \mathbb{R}_{0+}^{M \times T}$

A – unknown spectral reflectance matrix. $\mathbf{A} \in \mathbb{R}_{0+}^{N \times M}$

Unsupervised segmentation of multispectral images

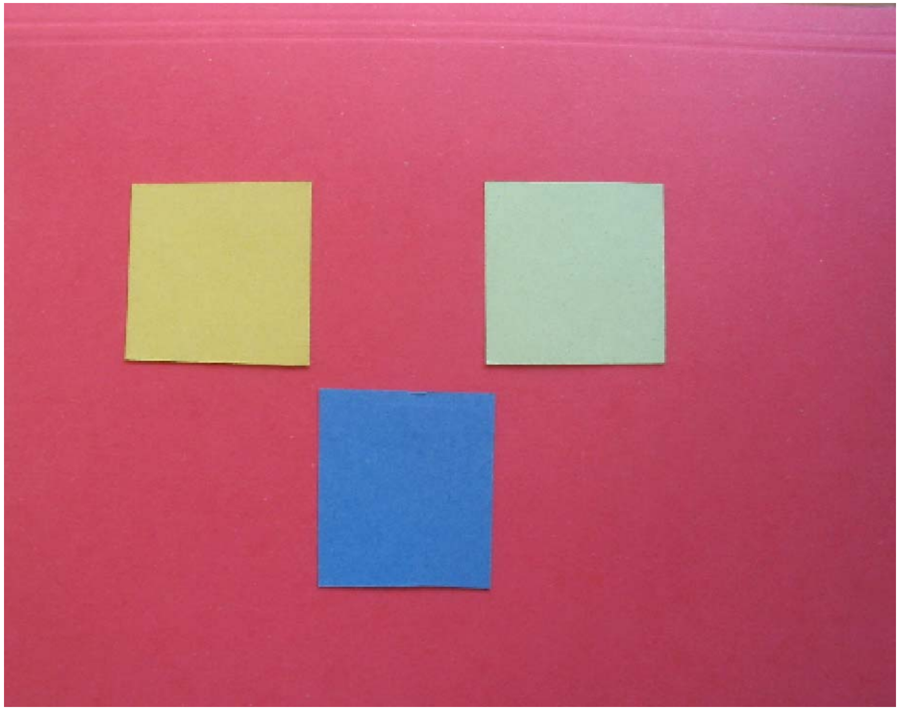
Spectral similarity between the sources s_m and s_n implies that corresponding column vectors are close to collinear i.e. $\mathbf{a}_m \cong c\mathbf{a}_n$.

Contribution at certain pixel location *t* is: $\mathbf{a}_m s_{mt} + \mathbf{a}_n s_{nt} \cong c\mathbf{a}_n s_{mt} + \mathbf{a}_n s_{nt}$. This implies that $\mathbf{s}_n$ and $c\mathbf{s}_m$ are indistinguishable i.e. they are statistically dependent.

Thus, spectral similarity between the sources causes ill-conditioning problems of the basis matrix as well as statistical dependence among the sources. **Both conditions imposed by ICA algorithm on SLMM are not satisfied.**

Unsupervised segmentation of RGB image with four materials

Consider blind decomposition of the RGB image (N=3) composed of four materials (M=4):



I. Kopriva and A . Cichocki, “Sparse component analysis-based non-probabilistic blind decomposition of low-dimensional multi-spectral images,” *Journal of Chemometrics*, vol. 23, Issue 11, pp. 590-597 (2009).

Unsupervised segmentation of multispectral images

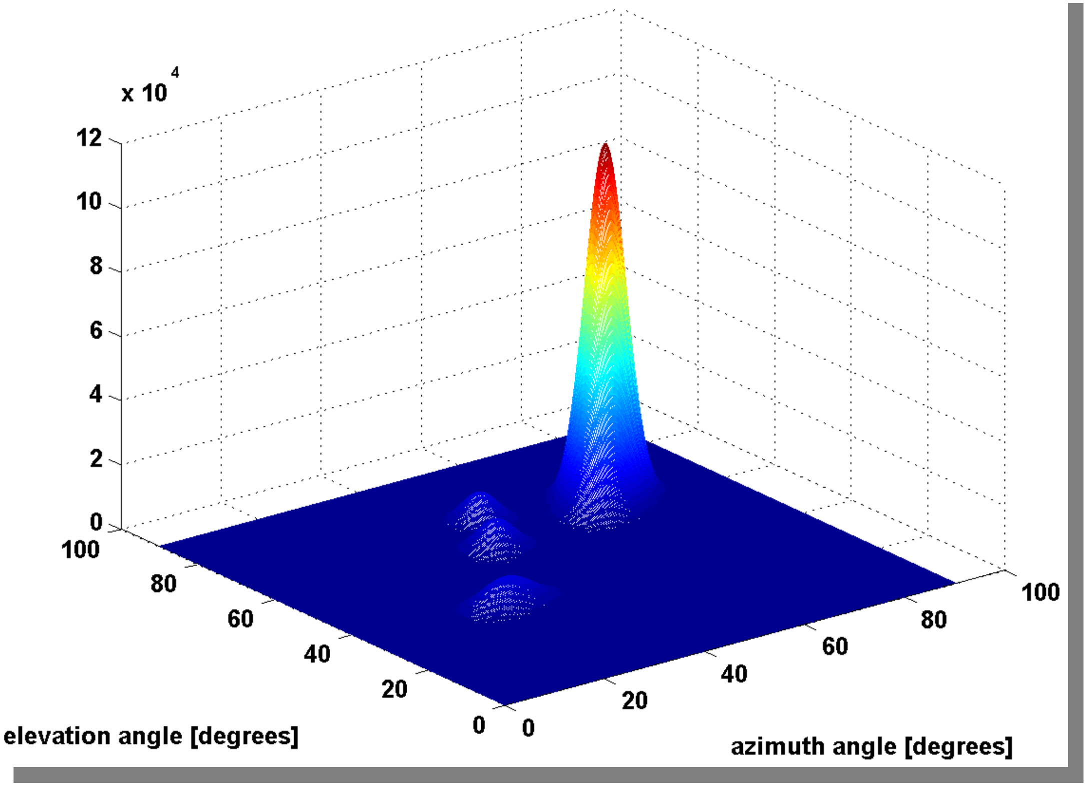
Evidently degree of overlap between materials in spatial domain is very small i.e. $s_{\mathrm{m}}(t)*s_n(t)\approx\delta_{nm.}$. Hence RGB image decomposition problem can be solved either with clustering and L_1-norm minimization or with HALS NMF algorithm with sparseness constraints.

For the L_1-norm minimization estimate of the mixing (spectral reflectance matrix) **A** and number of materials *M* is necessary. For HALS NMF only estimate of *M* is necessary. Both tasks can be accomplished by data clustering algorithm].

Since materials in do not overlap in spatial domain it applies $\|\mathbf{s}(t)\|_0\approx 1$.

Unsupervised segmentation of multispectral images

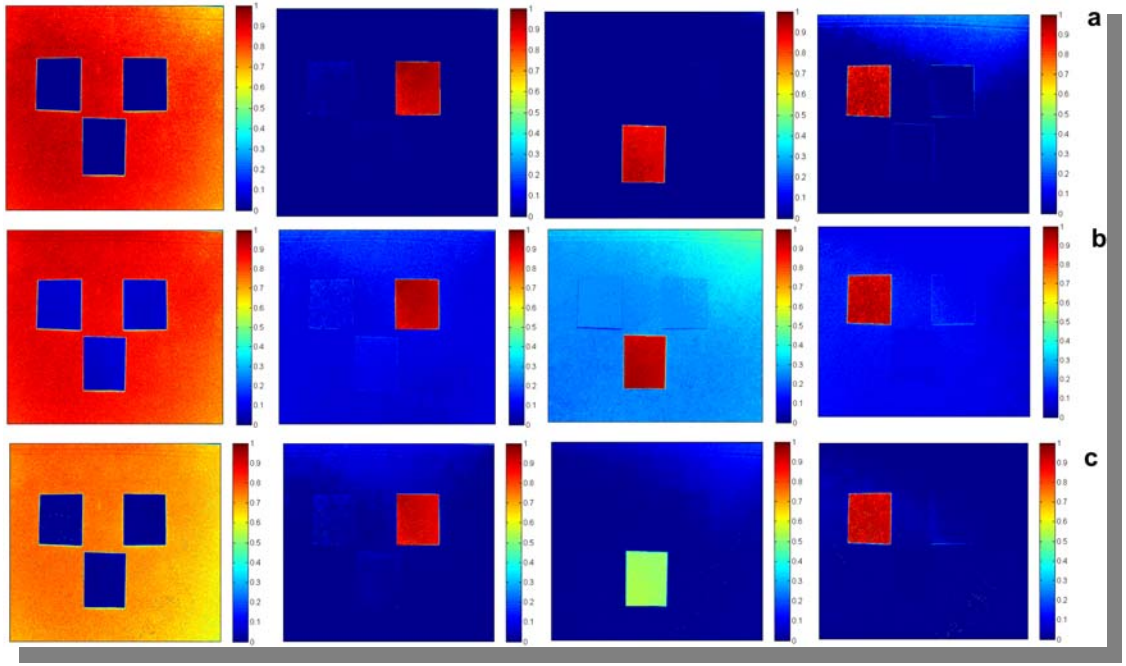
For shown experimental RGB image clustering function is obtained as:



Four peaks suggest existence of four materials in the RGB image **i.e. *M*=4**.

Unsupervised segmentation of multispectral images

Spatial maps of the materials extracted by HALS NMF with 25 layers, linear programming and interior point method are obtained as:

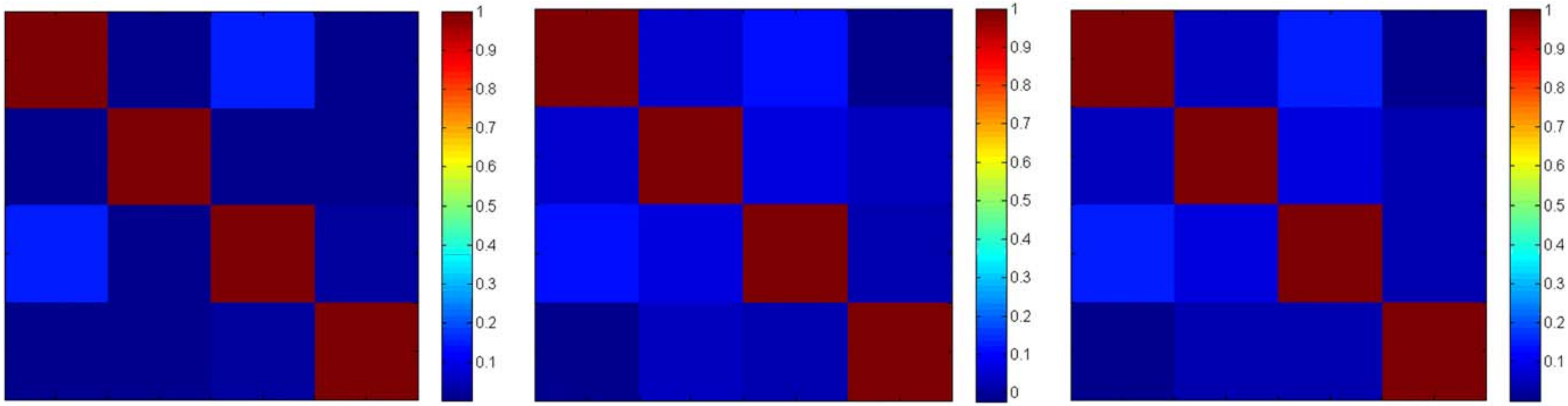


a) 25 layers HALS NMF; b) Interior point method; c) Linear programming.

S.J. Kim, K. Koh, M. Lustig, S. Boyd, D. Gorinevsky, "An Interior-Point Method for Large-Scale L_1 -Regularized Least Squares,"IEEE Journal of Selected Topics in Signal Processing **1**, 606-617 (2007).
http://www.stanford.edu/~boyd/l1_ls/.

Unsupervised segmentation of multispectral images

Correlation matrices



From left to right: 25 layers HALS NMF; Interior point method, [74,90]; c) Linear programming.

CR performance measure in dB

	Multilayer HALS NMF	**Interior-point method**	**Linear program**
CR [dB]	13.67	9.97	7.77
CPU time [s]*	3097	7751	3265

*MATLAB environment on 2.4 GHz Intel Core 2 Quad Processor Q6600 desktop computer with 4GB RAM.