Blind source separation: theory and applications

Ivica Kopriva

Ruđer Bošković Institute

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

Blind separation of sources

- **Linear problems**
  - Dynamic problems
  - **Static problems**
    - **Over- and determined problems**
      - **Principal component analysis**
      - **Independent component analysis**
      - **Dependent component analysis**
      - **Nonnegative matrix factorization**
      - Nonnegative tensor factorization
    - **Underdetermined problems**
      - **Sparse component analysis:**
        - *** clustering + lp (0<p≤1) min**
        - *** Hierarchical nonnegative matrix factorization**
- Nonlinear problems

Blind Source Separation – linear static problem

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

Problem:

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

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

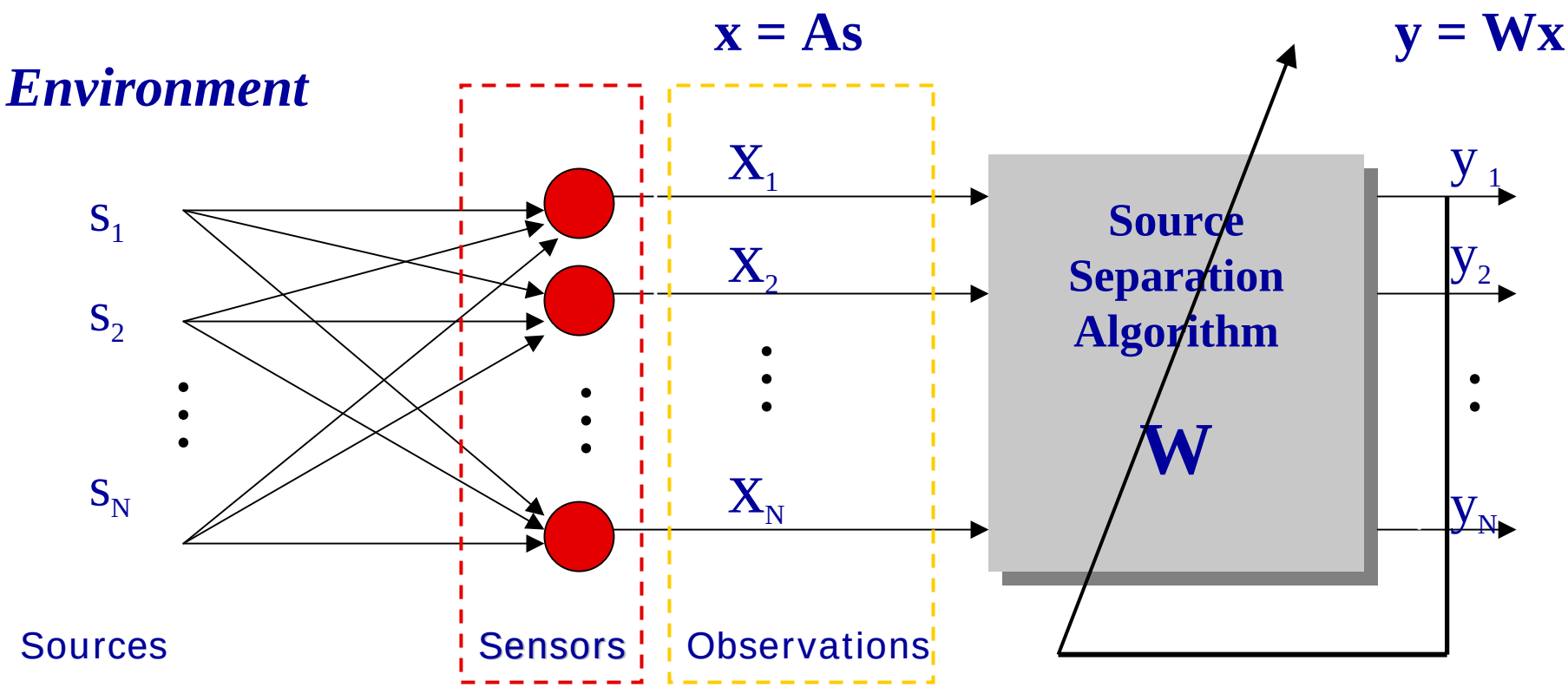
Goal: find **S, A** and number of sources *M* based on **X** only.

Meaningful solutions are characterized by scaling and permutation indeterminacies:

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

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.

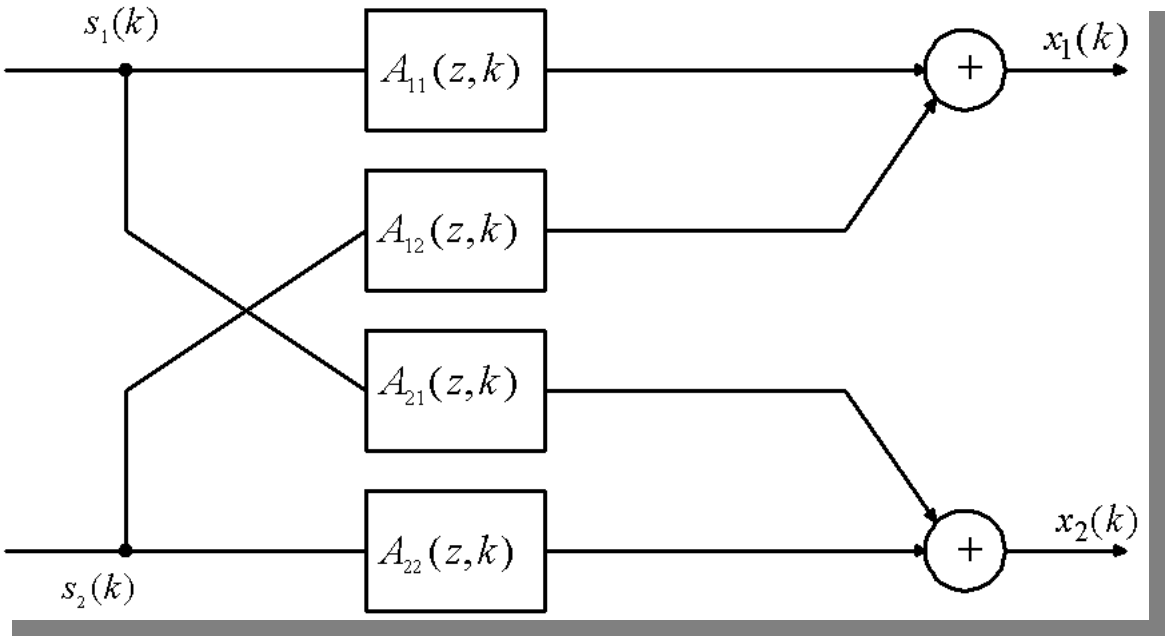
A. Cichocki, S. Amari, *Chapter 9* in Adaptive Blind Signal and Image Processing – Learning Algorithms and Applications, J. Wiley, 2002.

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

Dynamic (convolutive) model for 2x2 system.



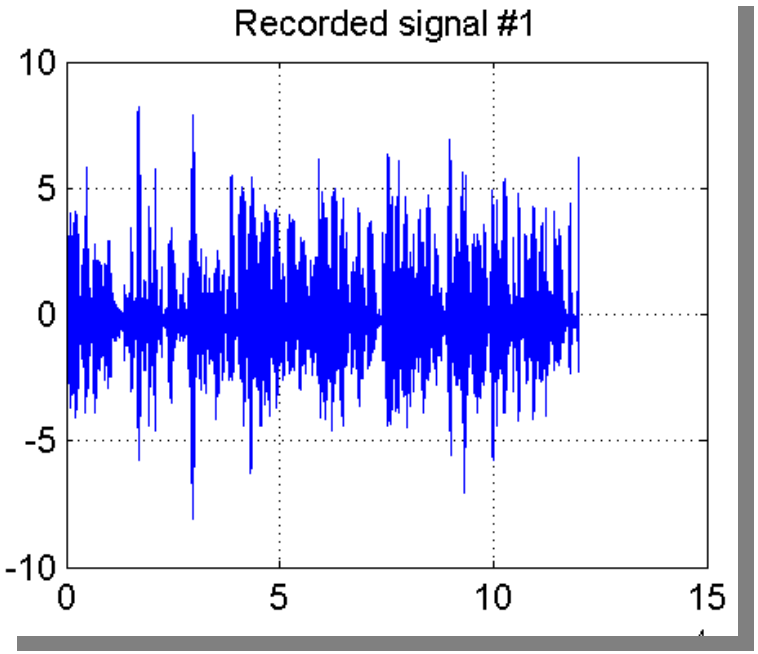
$$x_n(k) = \sum_{m=1}^{2} \sum_{l=0}^{L} a_{nm}(l) s_m(k-l) \quad n = 1, 2$$

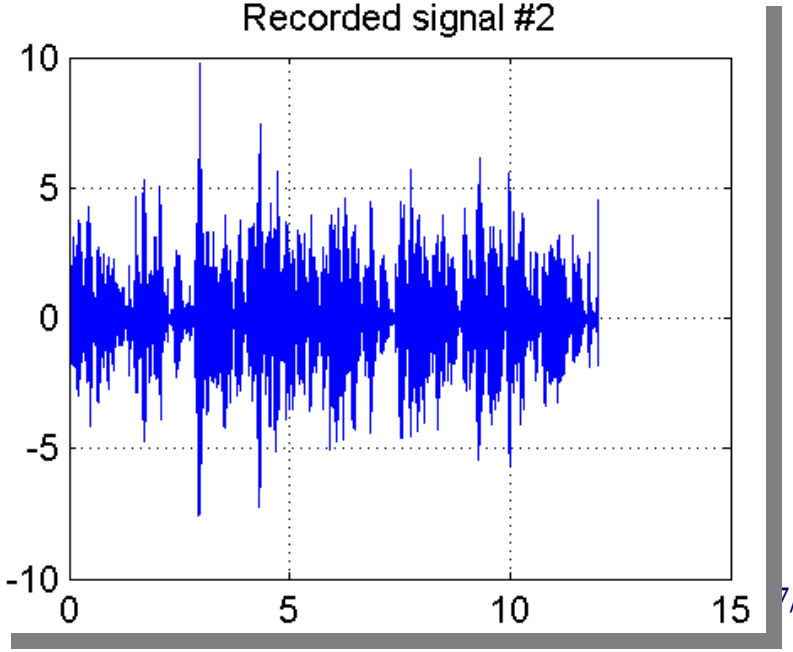
Blind Source Separation – linear dynamic problem

Speech separation in reverberant acoustic environment. Two recorded signals were downloaded from Russel Lamberts' home page:

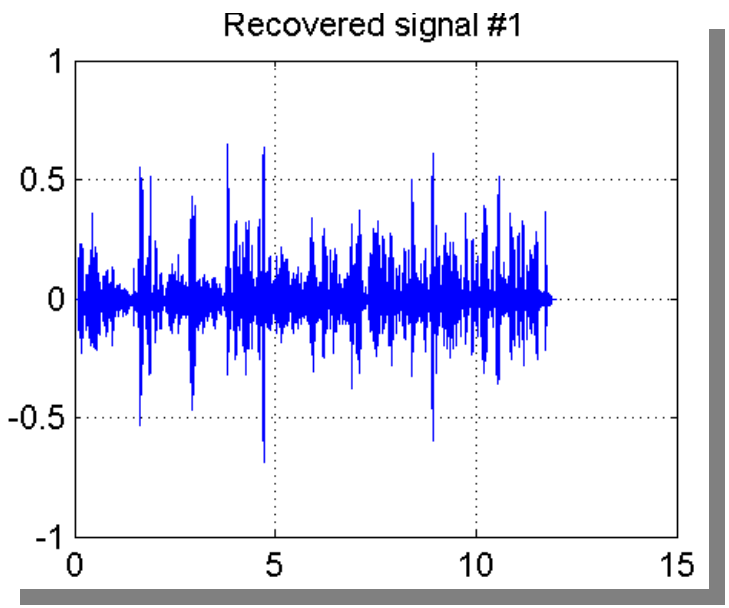
http://home.socal.rr.com/russdsp/ .

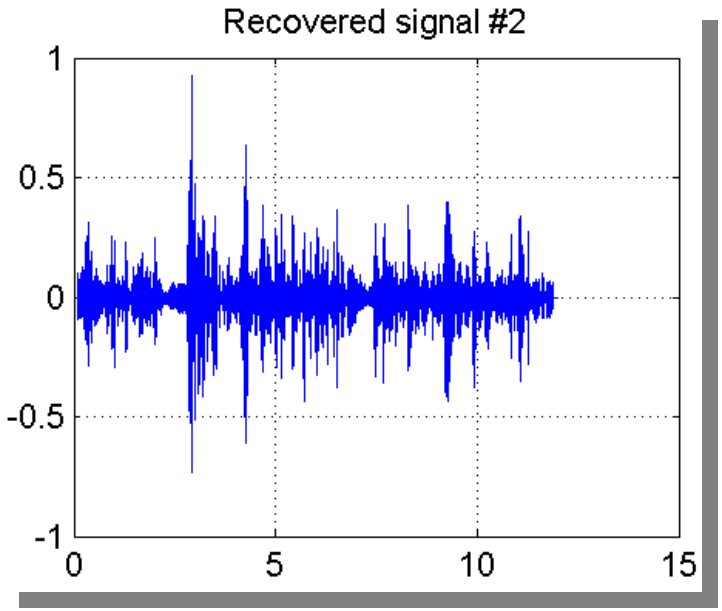
Signals were sampled with 8kHz and contain male and female speakers talking simultaneously for 12 seconds.





Blind Source Separation – linear dynamic problem



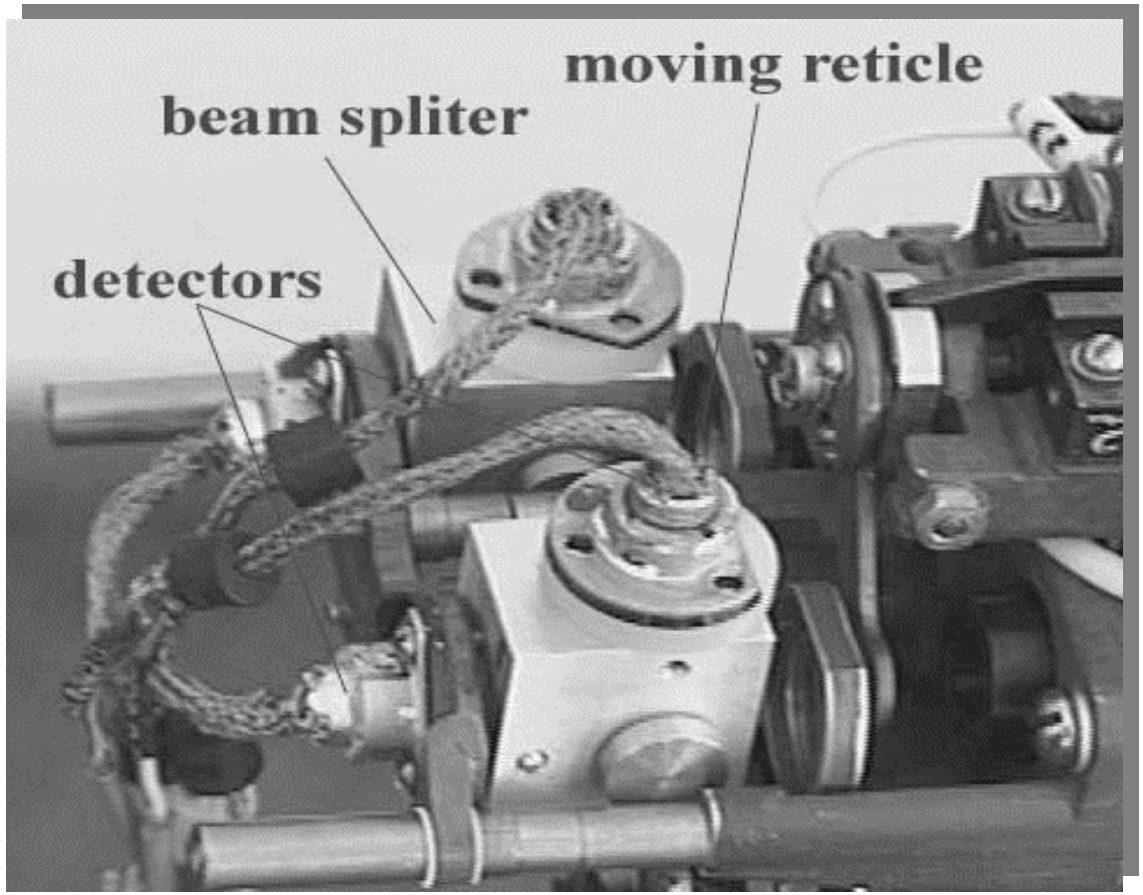


https://www.scientificamerican.com/article.cfm?id=solving-the-cocktail-party-problem

“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

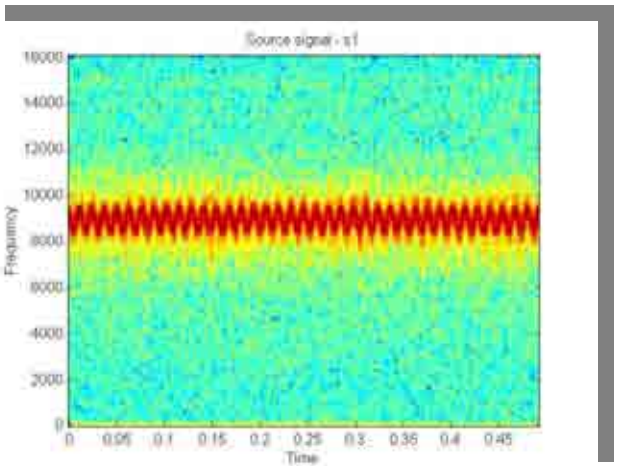
ICA and reticle based IR tracker

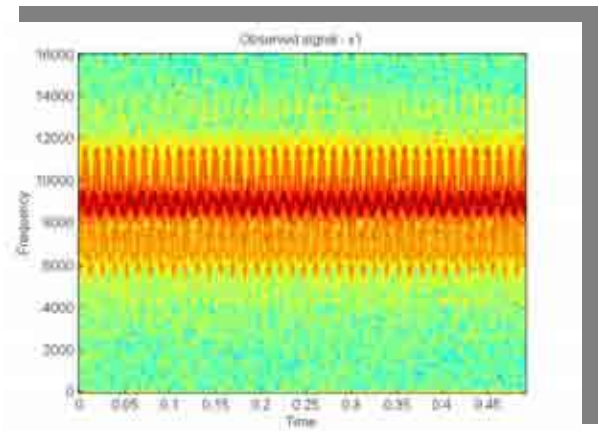


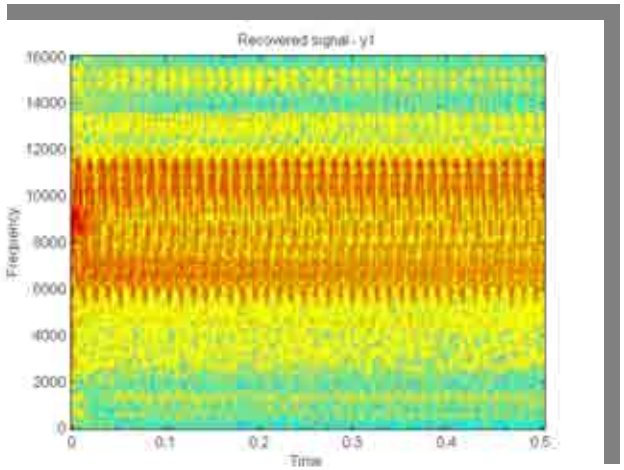
I. Kopriva, A. Peršin, Applied Optics, Vol. 38, No. 7, pp. 1115-1126, 1999.
I.Kopriva, H. Szu, A.Persin, Optics Communications, Vol. 203, Issue 3-6, pp. 197-211, 2002.

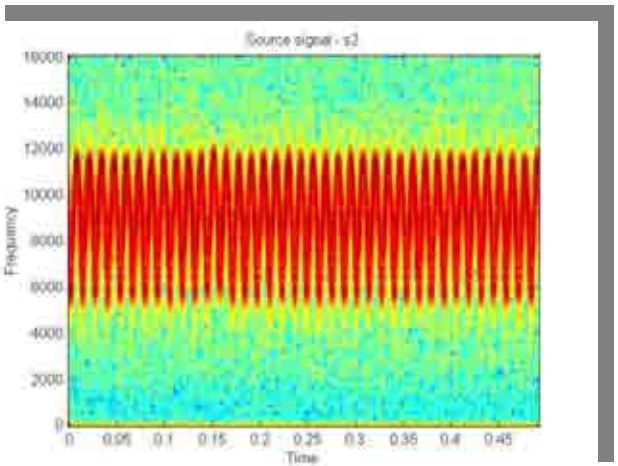
ICA and reticle based IR tracker

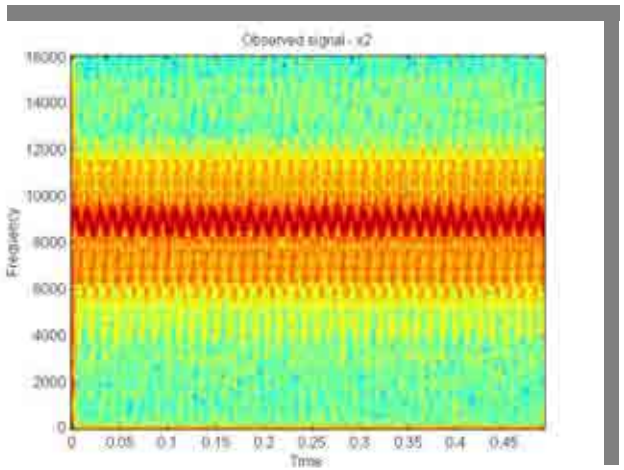
$\mathbf{s}$ $\mathbf{x = A * s}$ $\mathbf{y = W * x}$

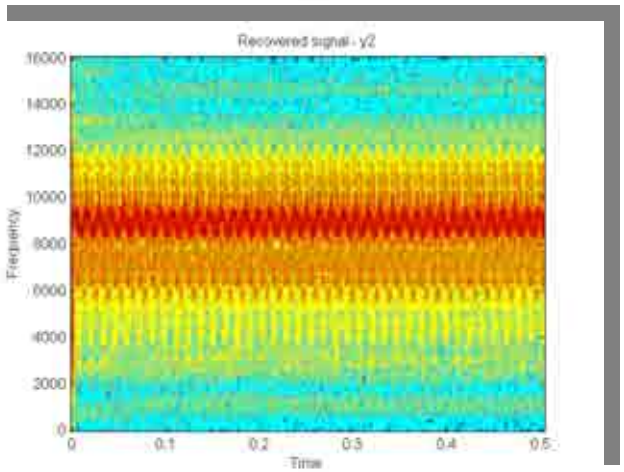












Blind Source Separation – nonlinear static problem

Problem:

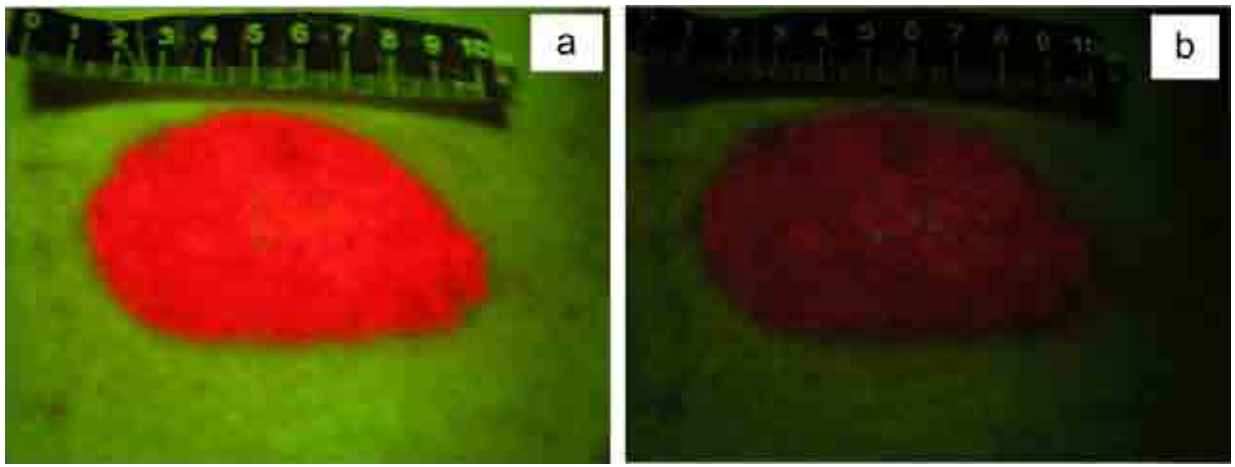
$\mathbf{X}=F(\mathbf{S})$ $\mathbf{X}\in R^{NxT}$, $\mathbf{S}\in R^{MxT}$

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

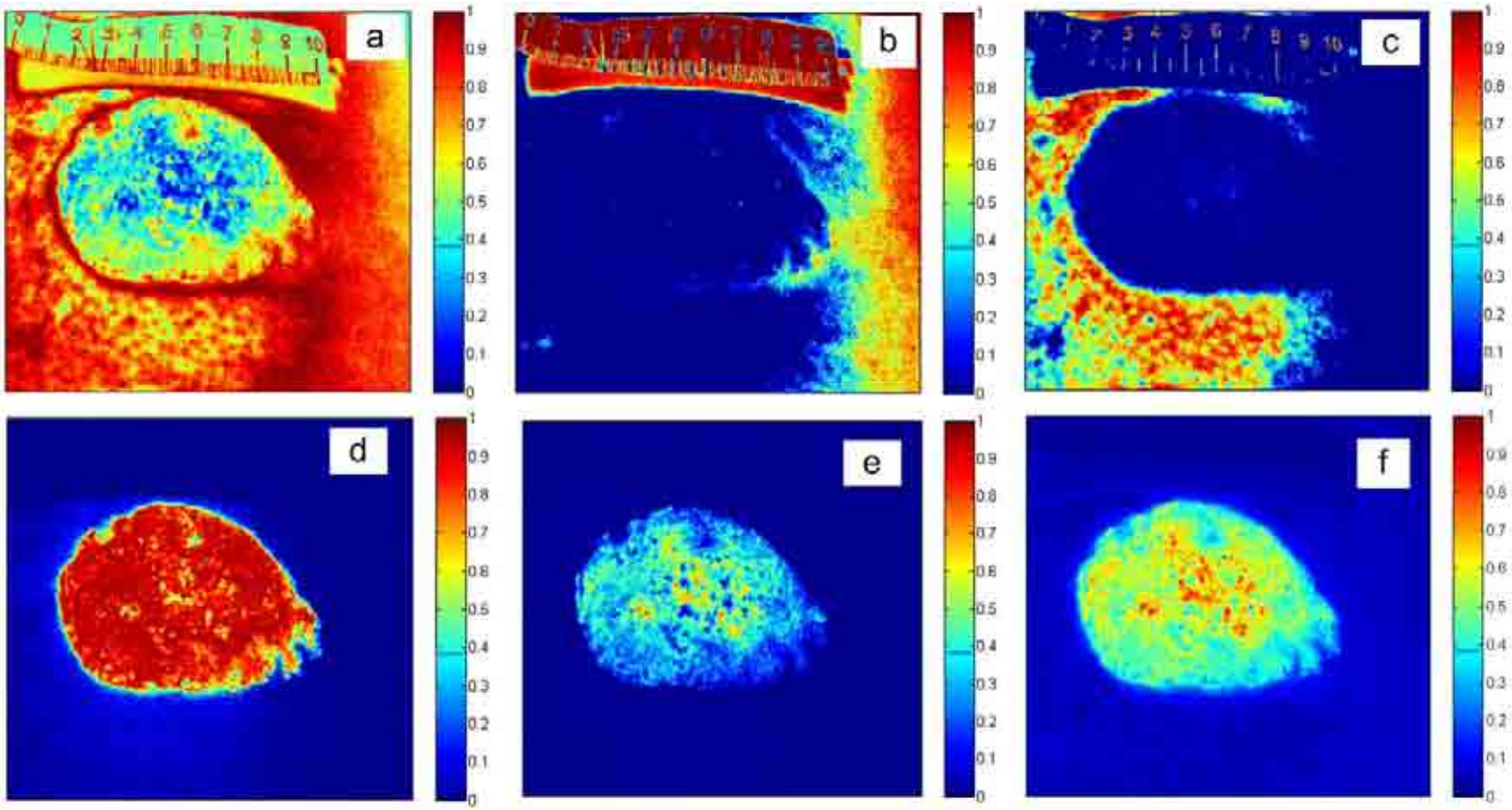
Goal: find **S** based on **X** only. Solution is possible without preconditions on the type of nonlinearity *F* by transforming original problem $\mathbf{X}=F(\mathbf{S})$ into reproducible kernel Hilbert space (RKHS) where mapped sources possibly become linearly separable: $\Phi(\mathbf{X})\approx\mathbf{A}\,\Phi(\mathbf{S})$.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



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

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

First stage: principal component analysis (PCA) and whitening (batch and online). PCA is decorellation transform used in multivariate data analysis. In connection with ICA it is very often a useful preprocessing step used in the whitening transformation after which multivariate data become uncorrelated with unit variance.

$$\mathbf{R}_{\mathbf{xx}} \approx (1/T)\sum_{t=1}^{T} \mathbf{x}(t)\mathbf{x}^{\mathrm{T}}(t)$$

It is assumed data **x** is zero mean. If not this is achieved by **x ← x- E{x}**.

Eigendecomposition of $\mathbf{R}_{\mathbf{xx}}$ is obtained as

$$\mathbf{R}_{\mathbf{xx}} = \mathbf{E}\Lambda\mathbf{E}^{\mathrm{T}}$$

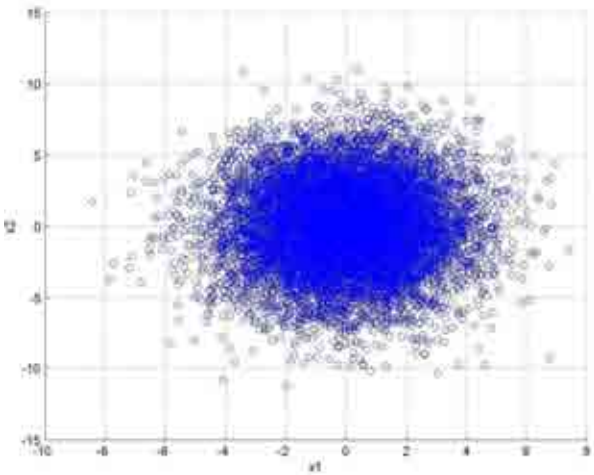
Where **E** is matrix of eigenvectors and Λ is diagonal matrix of eigenvalues of $\mathbf{R}_{\mathbf{xx}}$.

Batch form of PCA/whitening transform is obtained as

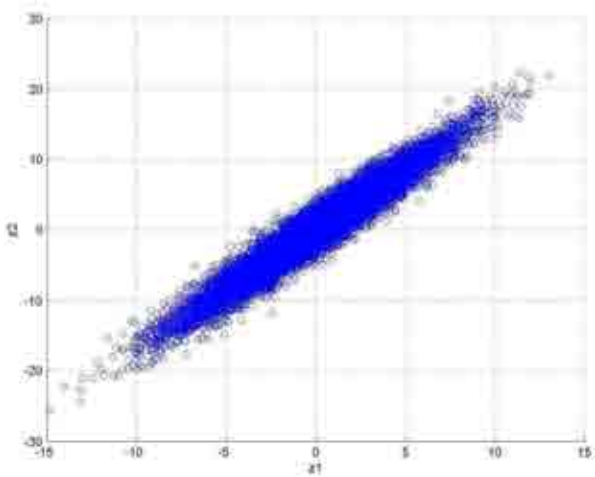
$$\mathbf{z} = \mathbf{V}\mathbf{x} = \Lambda^{-1/2}\mathbf{E}^{\mathrm{T}}\mathbf{x}$$

Statistical independence

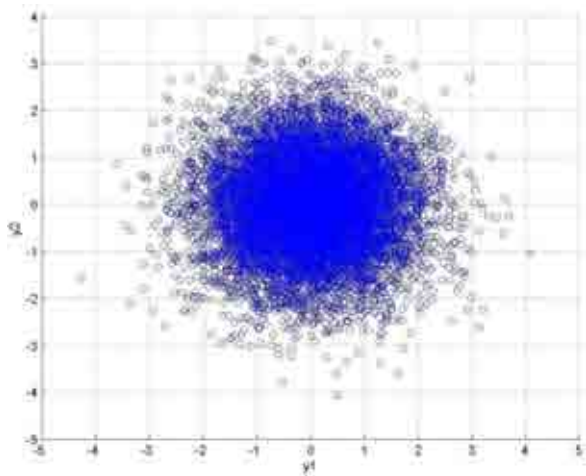
Scatter plots of two uncorrelated Gaussian signals (left); two correlated signals obtained as linear combinations of the uncorrelated Gaussian signals (center); two signals after PCA transform (right).



x_1=N(0,4); x_2 = N(0,9)



$z_1=x_1 + x_2$
$z_2=x_1 + 2x_2$



$\mathbf{y=\Lambda^{-1/2}E^{T}z}$
$\mathbf{z=[z_1;z_2]}$

Statistical independence

s_1

s_2

$x_1 = 2s_1 + s_2$
$x_2 = s_1 + s_2$

x_1

x_2

$y_1 \cong s_1$ (?)
$y_2 \cong s_2$ (?)

Statistical independence - ICA

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

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

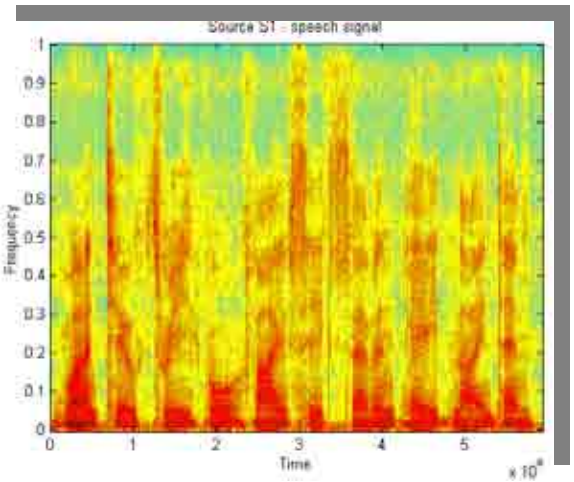
The problems is to estimated the speech signal (s_1) and noise signal (s_2) from observations x_1 and x_2.

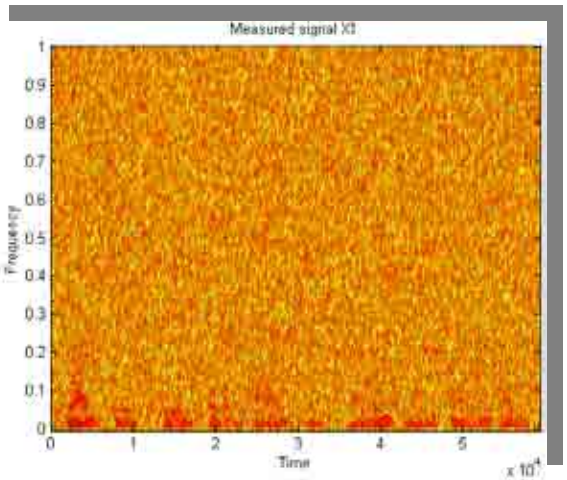
If mixing coefficients a_{11}, a_{12}, a_{21} and a_{22} are known problem would be solvable by simple matrix inversion.

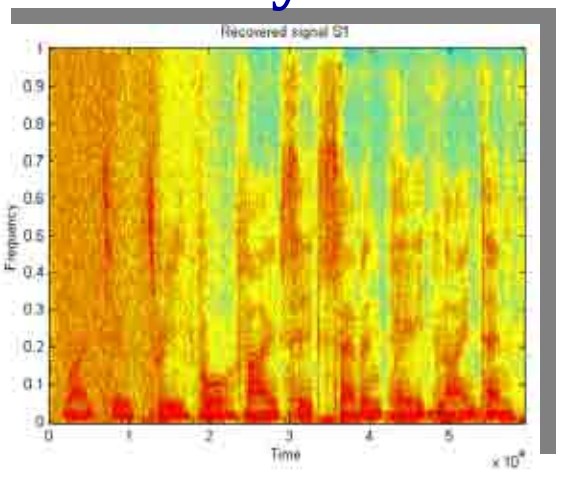
ICA enables to estimated speech signal (s_1) and noise signal (s_2) without knowing the mixing coefficients a_{11}, a_{12}, a_{21} and a_{22}. This is why the problem of recovering source signals s_1 and s_2 is called *blind source separation* problem.

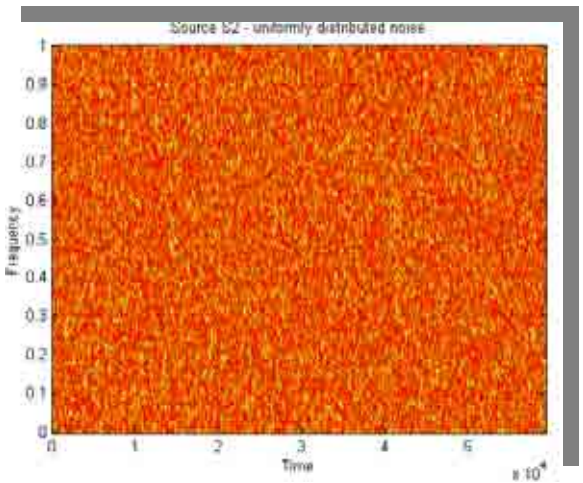
Speech from noise separation

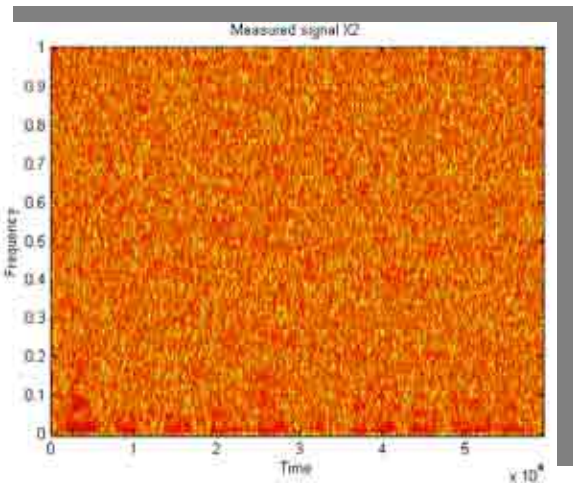
s **x** **y**

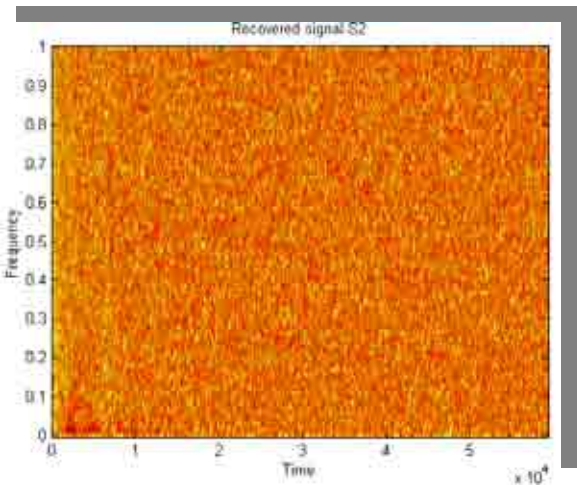












When does ICA work !?

- source signals $s_i(t)$ must be statistically independent.

$$\mathrm{p}(\mathbf{s}) = \prod_{i=}^{N} p_i(s_i)$$

- source signals $s_i(t)$, except one, must be non-Gaussian.

$$C_n(s_i) \neq 0 \quad n > 2$$

- mixing matrix ***A*** must be nonsingular and full column rank.

$$\mathbf{W} \cong \mathbf{A}^{-1}$$

When does ICA work !?

Ambiguities of ICA.

a) Variances (energies) of the independent components can not be determined. This is called *scaling indeterminacy*. The reason is that both **s** and **A** being unknown any scalar multiplier in one of the sources can always be canceled by dividing the corresponding column of **A** by the same multiplier:

$$\mathbf{x} = \sum_i \left(\frac{1}{\alpha_i} \mathbf{a}_i \right) \left(s_i \alpha_i \right)$$

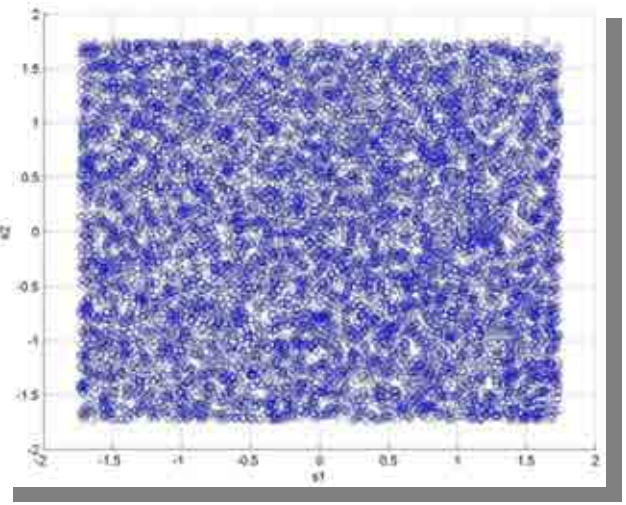
b) Order of the independent components can not be determined. This is called *permutation indeterminacy*. The reason is that components of the source vector **s** and columns of the mixing matrix **A** could be freely changed in such that

$$\mathbf{x} = \mathbf{A}\mathbf{P}^{-1}\mathbf{P}\mathbf{s}$$

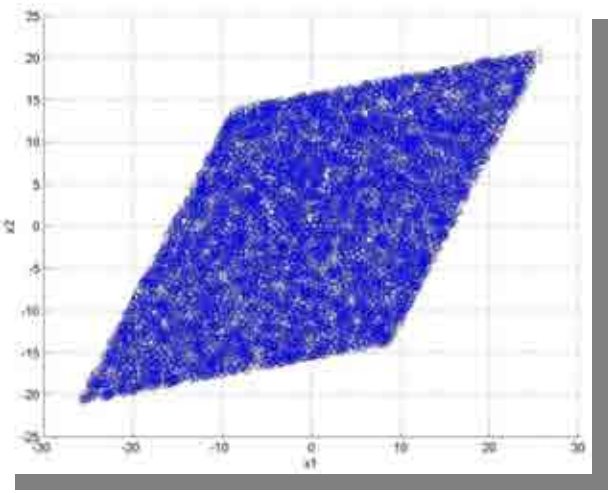
where **P** permutation matrix, **Ps** is new source vector with original components but in different order and $\mathbf{AP}^{-1}$ is a new unknown mixing matrix.

When does ICA work !?

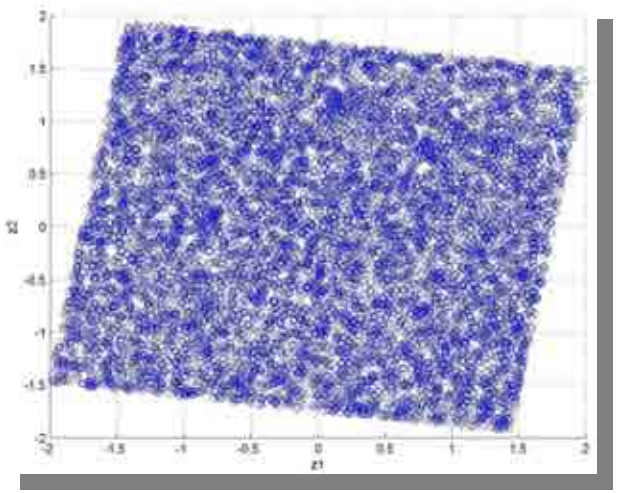
Whitening is only half of the ICA. Whitening transform decorrelates signals. If signals are non-Gaussian it does not make them statistically independent. Whitening transform is useful first processing step in ICA. A second rotation stage achieved by an unitary matrix can be obtained by ICA exploiting non-Gaussianity of the signals.



Source signals



Mixed signals



Whitened signals

When does ICA work !?

PCA applied to blind image separation:

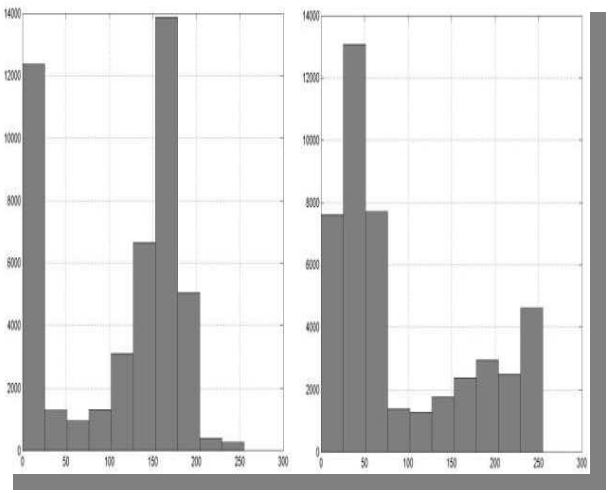
z_1

z_2

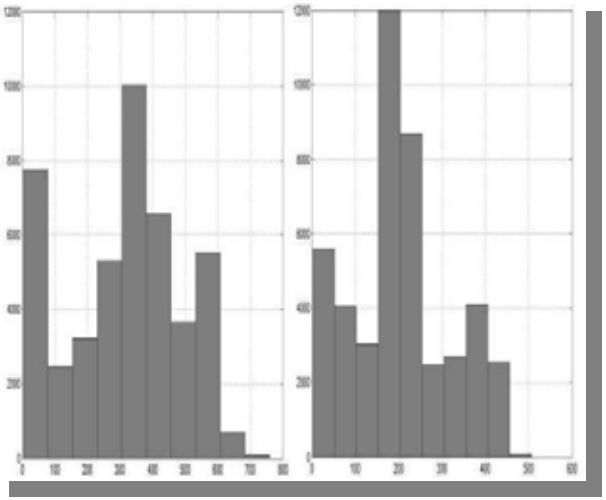
MATLAB code:

```
Rx=cov(X');                    % estimate of the data covariance matrix
[E,D] = eig(Rx);   % eigen-decomposition of the data covariance matrix
Z = E'*X;                      % PCA transform
z1=reshape(Z(1,:),P,Q);    % transforming vector into image
figure(1); imagesc(z1);      % show first PCA image
z2=reshape(Z(2,:),P,Q);    % transforming vector into image
figure(2); imagesc(z2);      % show second PCA image
```

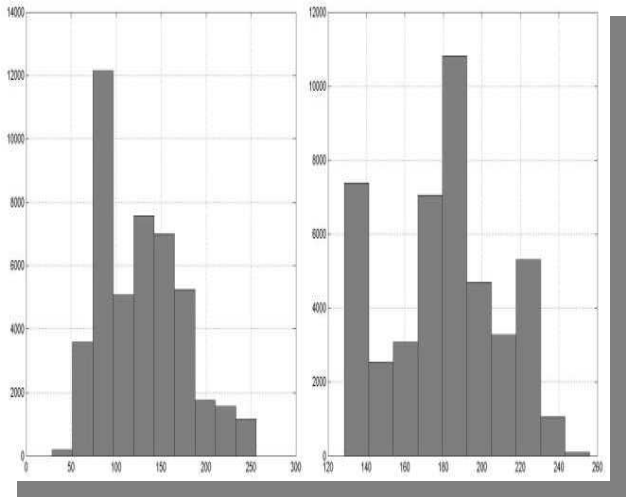
Histograms of source, mixed and PCA extracted images



Source image



Mixed images



PCA extracted images

ICA for linear instantaneous models

- Information theoretic ICA
- Tensorial methods (Fourth order cumulants) ICA
- ICA by time-delayed correlations
- Applications

Information theoretic ICA

ICA by maximum likelihood (ML). Likelihood of the noise free ICA model **x=As** is formulated as:

$$p_x(\mathbf{x}) = \left|\det \mathbf{W}\right| p_s(\mathbf{s}) = \left|\det \mathbf{W}\right| \prod_i p_i(s_i)$$

where $\mathbf{W}=[\mathbf{w}_1\ \mathbf{w}_2\ \ldots\ \mathbf{w}_N]^T=\mathbf{A}^{-1}$. ML means that we want to maximize probability that data **x** were observed under model **x=As**. Because $s_i=\mathbf{w}_i^T\mathbf{x}$, $p_x(\mathbf{x})$ can be written as:

$$p_x(\mathbf{x}) = \left|\det \mathbf{W}\right| \prod_i p_i(\mathbf{w}_i^T\mathbf{x})$$

If this is evaluated across T observations we obtain likelihood $L(\mathbf{W})$ as:

$$L(\mathbf{W}) = \prod_{t=1}^{T}\prod_{i=1}^{N} p_i(\mathbf{w}_i^T\mathbf{x}(t))\left|\det \mathbf{W}\right|$$

Normalized log-likelihood is obtained as:

$$\frac{1}{T}\log L(\mathbf{W}) = E\left\{\sum_{i=1}^{N}\log p_i(\mathbf{w}_i^T\mathbf{x}(t))\right\} + \log\left|\det \mathbf{W}\right|$$

D. T. Pham, "Blind separation of mixtures of independent sources through a quasimaximum likelihood approach," *IEEE Trans. Signal Processing* 45, pp. 1712-1725, 1997.

Information theoretic ICA

Gradient maximization of the log-likelihood function gives:

$$\Delta \mathbf{W} = \frac{1}{T}\frac{\partial \log L}{\partial \mathbf{W}} = \left[\mathbf{W}^{\mathrm{T}}\right]^{-1} - E\left\{\varphi(\mathbf{W}\mathbf{x})\mathbf{x}^{\mathrm{T}}\right\}$$

where nonlinearity $\varphi(y_i)$ is called score function and is given with

$$\varphi_i = -\frac{1}{p_i}\frac{dp_i}{dy_i}$$

Correcting Euclidean gradient with metric tensor $\mathbf{W}^{\mathrm{T}}\mathbf{W}$ we get ML batch ICA algorithm:

$$\mathbf{W}(k+1) = \mathbf{W}(k) + \eta\left[\mathbf{I} - E\left\{\varphi(\mathbf{y})\mathbf{y}^{\mathrm{T}}\right\}\right]\mathbf{W}(k)$$

ML adaptive ICA algorithm is obtained by dropping expectation:

$$\mathbf{W}(t+1) = \mathbf{W}(t) + \eta\left[\mathbf{I} - \varphi(\mathbf{y}(t))\mathbf{y}(t)^{\mathrm{T}}\right]\mathbf{W}(t)$$

S. Amari, "Natural gradient works efficiently in learning," *Neural Computation* **10**(2), pp. 251-276, 1998.
J. F. Cardoso, and B. Laheld, "Equivariant adaptive source separation," *IEEE Trans. Signal Processing* **44**(12), pp. 3017-3030, 1996.

Information theoretic ICA

The central problem is that optimal value of φ**(y)** requires knowledge of the probability density of the source signals:

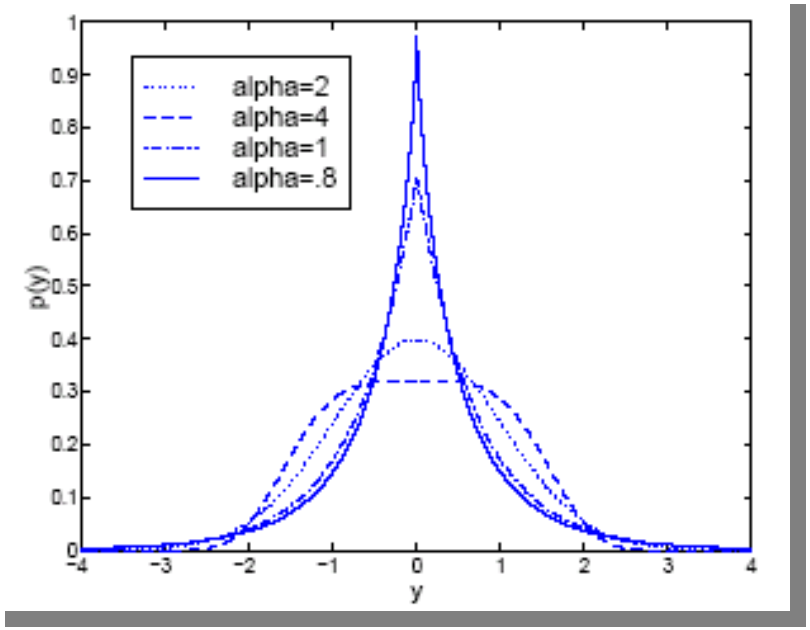
$$\varphi_i = -\frac{1}{p_i}\frac{dp_i}{dy_i}$$

which by definition is not known (the problem is **blind**).

Information theoretic ICA

Flexible nonlinearity concept is derived from the generalized Gaussian distribution model:

$$p_i(y_n) = \frac{\alpha_i}{2\sigma_i \Gamma(1/\alpha_i)} \exp\left(-\frac{1}{\alpha_i} \left| \frac{y_i}{\sigma_i} \right|^{\alpha_i} \right)$$



With the single parameter α_i (called Gaussian exponent) super-Gaussian distributions (α_i <2) and sub-Gaussian distributions (α_i >2) could be modeled.

S. Choi, A. Cihcocki, S. Amari, "Flexible Independent Component Analysis," Journal VLSI, KAP, 2000.
L. Zhang, A. Cichocki, S. Amari, "Self-adaptive Blind Source Separation Based on Activation Function adaptation", *IEEE Tran. On Neural Networks,* vol. 15, No. 2, pp. 233-244, March, 2004.

Information theoretic ICA

If generalized Gaussian probability density function is inserted in the optimal form for score function the expression for flexible nonlinearity is obtained:

$$\varphi_i(y_i) = sign(y_i)\left|y_i\right|^{\alpha_i - 1}$$

If *a priory* knowledge about statistical distributions of the source signals is available α_i can be fixed in advance. This is not always impossible. For example if source signals are speech or music signals α_i can be set to $\alpha_i=1$ because speech and music are super-Gaussian signals. If source signals are various communication signals α_i can be set to $\alpha_i=2.5$ or $\alpha_i=3$ because communication signals are sub-Gaussian signals.

Alternative way is to estimate α_i adaptively from data.

Information theoretic ICA

Score functions can be estimated from data based on estimation of the probability density function from using, as an example, Gaussian kernel estimator.

$$\hat{p}_i(y_i(t), \mathbf{y}_i) = \frac{1}{T}\sum_{tt=1}^{T} G\left(y_i(t) - y_i(tt), \sigma^2\mathbf{I}\right)$$

$$\frac{d\hat{p}_i(y_i)}{dy_i} = -\frac{1}{T}\sum_{tt=1}^{T} \frac{y_i(t) - y_i(tt)}{\sigma^2} G\left(y_i(t), \sigma^2\mathbf{I}\right)$$

$$G\left(y_i(t), \sigma^2\mathbf{I}\right) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{y_i^2(t)}{2\sigma^2}\right)$$

S J.C. Principe, D. Xu and J.W. Fisher, "Information-Theoretic Learning," *Chapter 7* in *Unsupervised Adaptive Filtering- Volume I Blind Source Separation*, ed. S. Haykin, J. Wiley, 2000.

Tensorial methods based ICA

Tensorial methods minimize only second and fourth order statistical dependence between components of **y**. Second order dependence is minimized by whitening transform **z=Vx**. Minimization of the fourth order statistical dependence is formulated as joint diagonalization problem:

$$\mathbf{W} = \arg\min \sum_i \sum_j \sum_k \sum_l \mathbf{off}\left(\mathbf{W}^T \hat{C}_4(y_i, y_j, y_k, y_l)\mathbf{W}\right)$$

$$\mathbf{off}(\mathbf{A}) = \sum_{1 \le i \ne j \le N} \left|a_{ij}\right|^2$$

Where **y=Wz** and $\hat{C}_4(y_i, y_j, y_k, y_l)$ represents sample estimate of the FO crosscumulant:

$$\hat{C}_4(y_i, y_j, y_k, y_l) = \langle y_i y_j y_k y_l \rangle - \langle y_i y_j \rangle \langle y_k y_l \rangle - \langle y_i y_k \rangle \langle y_j y_l \rangle - \langle y_i y_l \rangle \langle y_j y_k \rangle$$

Algorithm is know as JADE (Joint Approximate Diagonalization of Eigen-matrices) and can be downloaded from:http://www.tsi.enst.fr/~cardoso/Algo/Jade/jade.m

J. F. Cardoso and A. Souloumiac, "Blind beamforming for non-Gaussian signals," *IEE-Proc. – F,* vol. 140, pp. 1362-1370, 1993.

ICA by time-delayed correlations

When source signals have time structure i.e. their correlations and cross-correlations are nonzero for different time lags:

$$E\left[s_i(t)s_i(t-\tau)\right] \neq 0 \text{ for } \tau = 1,2,3,\ldots$$

it is possible to generate enough equations in order to solve the BSS problem without usage of the higher order statistics. If source signals have time structure (colored statistics) they are even allowed to be Gaussian. If data are already whitened with **z=Vx**, it is possible to formulate symmetric one-lag covariance matrix as:

$$\overline{\mathbf{C}}_\tau^{\mathbf{z}} = \frac{1}{2}\left[\mathbf{C}_\tau^{\mathbf{z}} + \left(\mathbf{C}_\tau^{\mathbf{z}}\right)^{\mathrm{T}}\right]$$

L. Molgedey and H. G. Schuster, "Separation of mixture of independent signals using time delayed correlations," *Physical Review Letters,* vol. 72, pp. 3634-3636, 1994.
L. Tong, R.W. Liu, V.C. Soon, and Y. F. Huang, "Indeterminacy and identifiability of blind identification," *IEEE Trans. on Circuits and Systems,* 38:499-509, 1991.

ICA by time-delayed correlations

Symmetric one-time lag covariance matrix has the following structure **(Wz=s; z=W$^{\mathrm{T}}$s):**

$$\bar{\mathbf{C}}_{\tau}^{\mathbf{z}} = \frac{1}{2}\mathbf{W}^{\mathrm{T}}\left[E\left\{\mathbf{s}(t)\mathbf{s}(t-\tau)^{\mathrm{T}}\right\} + E\left\{\mathbf{s}(t-\tau)\mathbf{s}(t)^{\mathrm{T}}\right\}\right]\mathbf{W} = \mathbf{W}^{\mathrm{T}}\bar{\mathbf{C}}_{\tau}^{\mathbf{s}}\mathbf{W}$$

Because source signals are independent by assumption one-time lag covariance matrix $\bar{\mathbf{C}}_{\tau}^{\mathbf{s}}$ is diagonal matrix:

$$\bar{\mathbf{C}}_{\tau}^{\mathbf{s}} = E\left\{\mathbf{s}(t)\mathbf{s}(t-\tau)^{\mathrm{T}}\right\} + E\left\{\mathbf{s}(t-\tau)\mathbf{s}(t)^{\mathrm{T}}\right\} = \boldsymbol{\Lambda}$$

data covariance matrix can be written as:

$$\bar{\mathbf{C}}_{\tau}^{\mathbf{z}} = \mathbf{W}^{\mathrm{T}}\boldsymbol{\Lambda}\mathbf{W}$$

which shows that rows of de-mixing matrix **W** are the eigen-vectors of the symmetrical one-lag data covariance matrix $\bar{\mathbf{C}}_{\tau}^{\mathbf{z}}$. This is how BSS problems is solved by the ***AMUSE*** algorithm.

ICA by time-delayed correlations

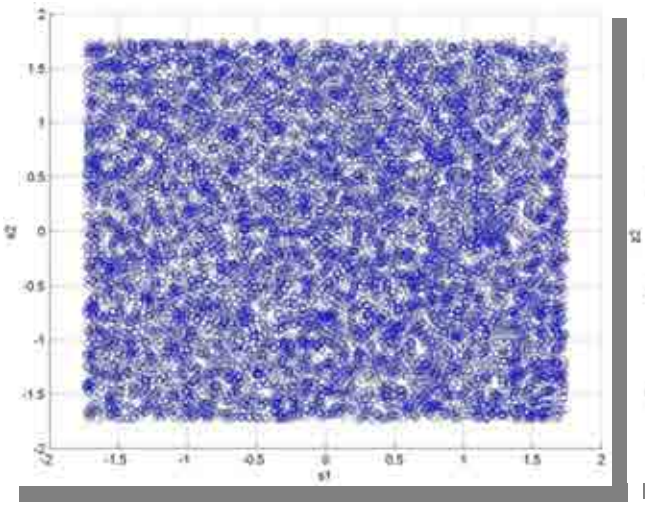
Approach could be extended by using multiple time lags. The ICA algorithm is formulated as joint diagonalization problem:

$$J(\mathbf{W}) = \sum_{\tau \in S} \text{off}\left(\mathbf{W}\bar{\mathbf{C}}_{\tau}^{\mathbf{z}}\mathbf{W}^{\mathrm{T}}\right)$$

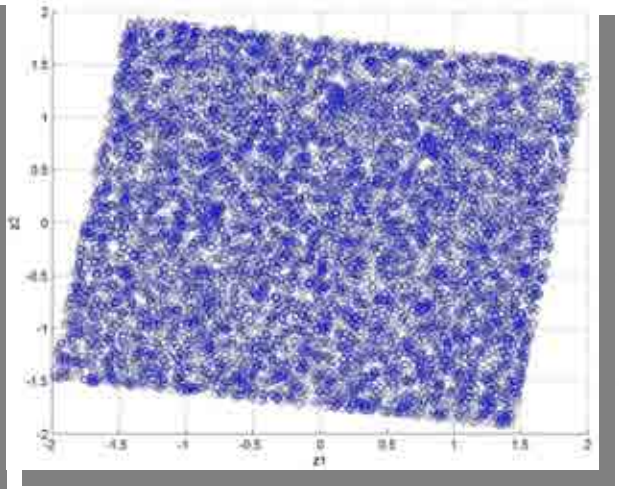
Representative algorithms are SOBI (second order blind identification) and TDSEP.

A. Belouchrami, K.A. Meraim, J.F. Cardoso, and E. Moulines, "A blind source separation technique based on second order statistics," *IEEE Trans. on Signal Processing,* 45(2), pp. 434-444, 1997.
A. Ziehe, K.R. Muller, G. Nolte, B. M. Mackert, and G. Curio, "TDSEP-an efficient algorithm for blind separation using time structure," *Proc. ICANN'98,* pp. 675-680, Skovde, Sweden, 1998.

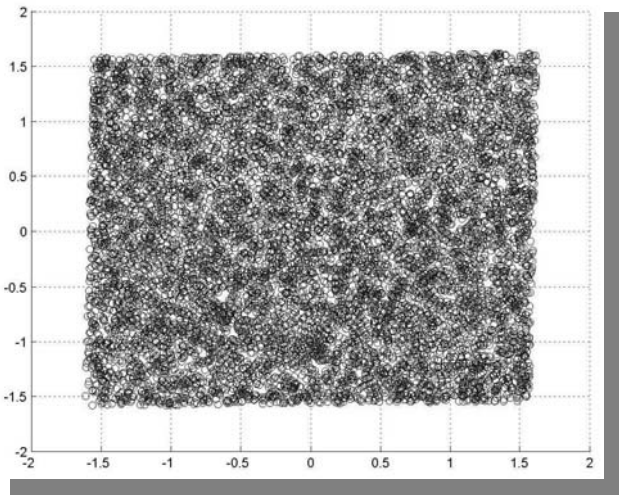
Scatter diagrams of PCA and ICA extracted signals



Source signals



PCA extracted signalsi



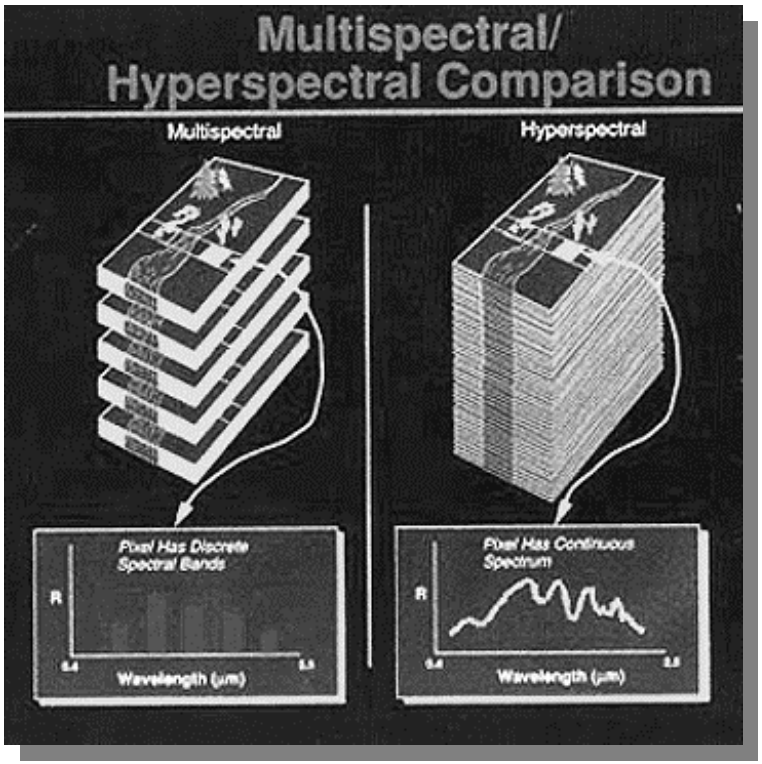
ICA extracted signals (min $MV(\mathbf{y})$).

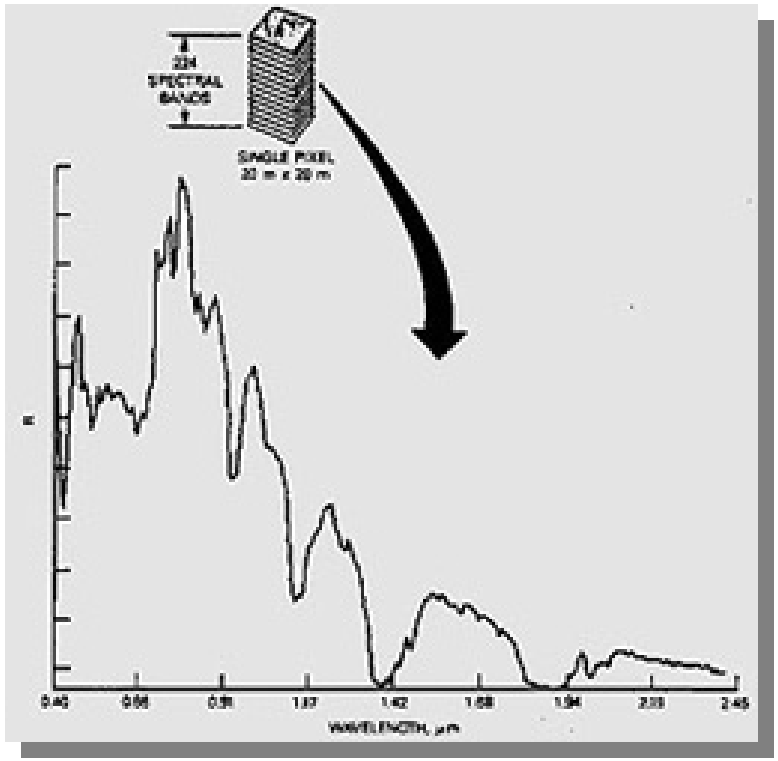
PCA

ICA (min $MI(\mathbf{y})$).

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:

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

x - measured data intensity vector, $\mathbf{x} \in \mathbf{R}^{Nx1}$

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

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

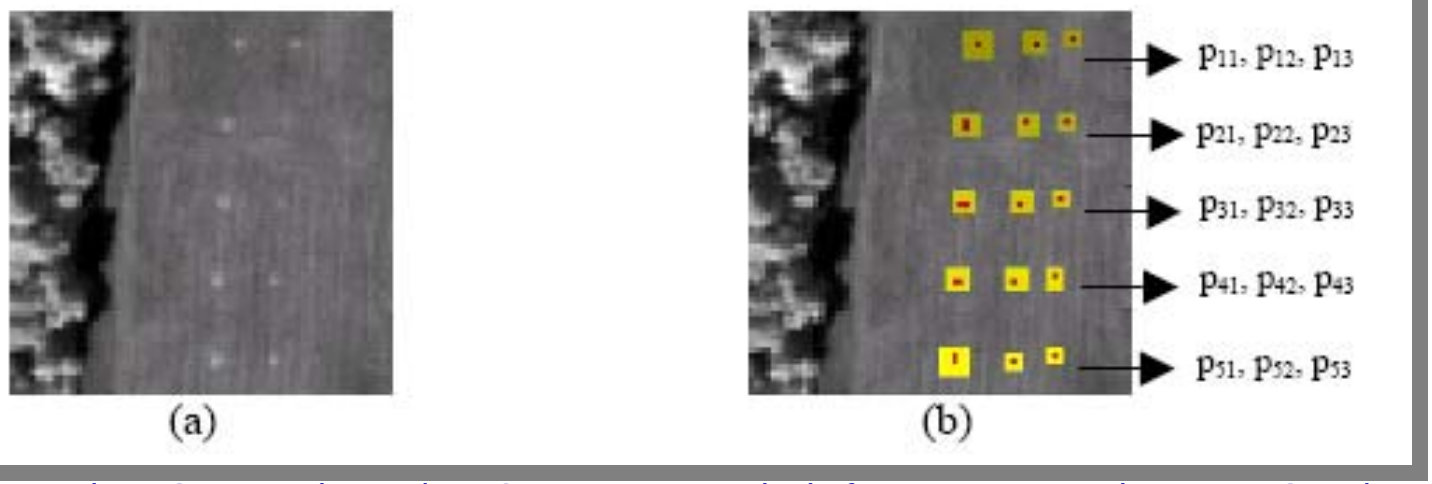
Unknown endmembers s_i are can be recovered by ICA based de-mixing:

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

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.

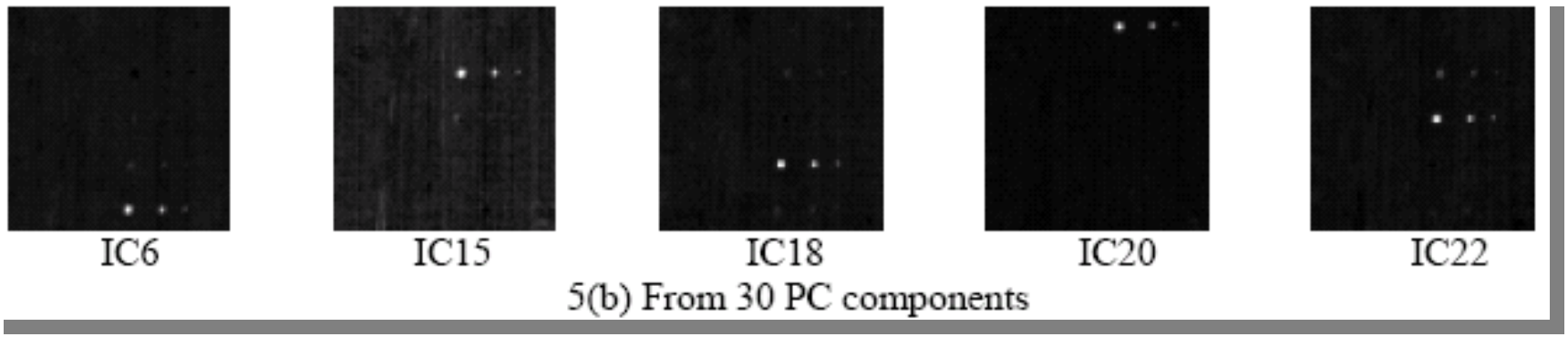
ICA and unsupervised classification of the hyperspectral image

- 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.

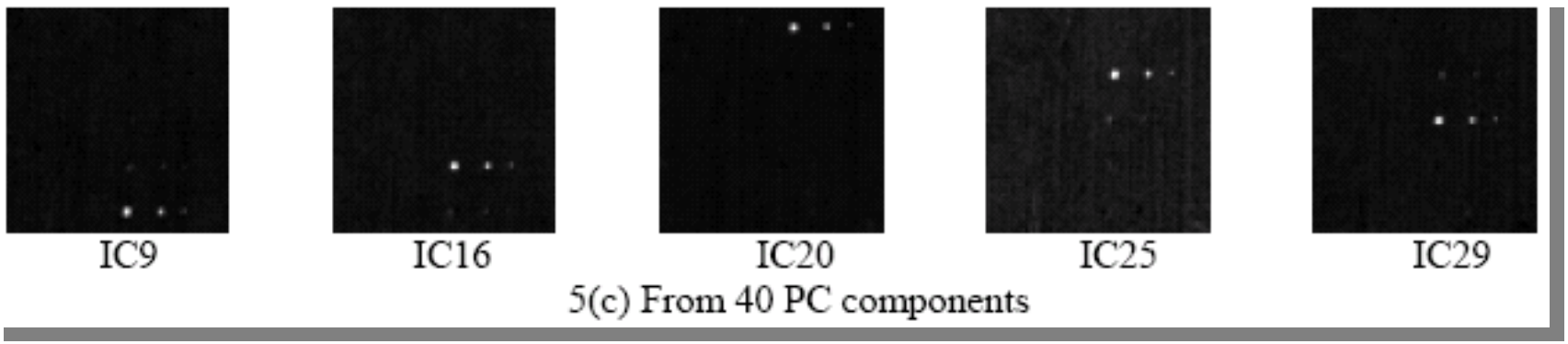


Q. Du, I. Kopriva and H. Szu, " Independent Component Analysis for Hyperspectral Remote Sensing Imagery Classification," *Optical Engineering*, vol. 45, 017008, January 2006.
Q. Du, I. Kopriva, "Automated Target Detection and Discrimination Using Constrained Kurtosis Maximization," *IEEE Geoscience Remote 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.

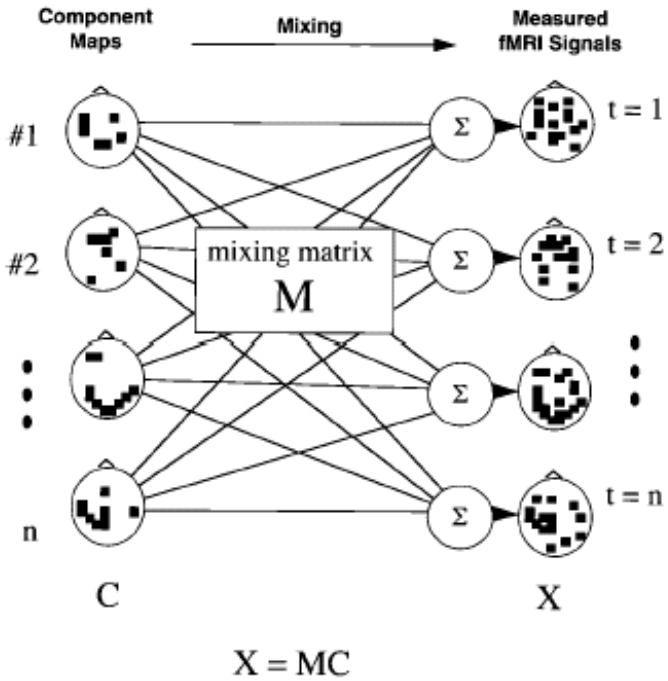


Figure 3.
fMRI data as a mixture of independent components. The mixing matrix M specifies the relative contribution of each component at each time point. ICA finds an *unmixing* matrix that separates the observed component mixtures into the independent component maps and time courses.

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.

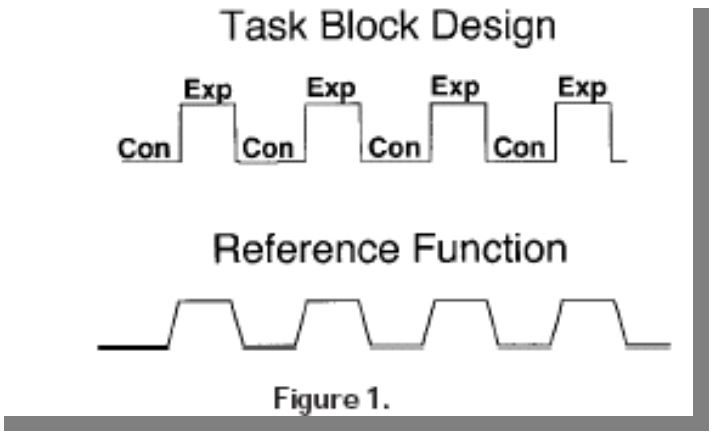
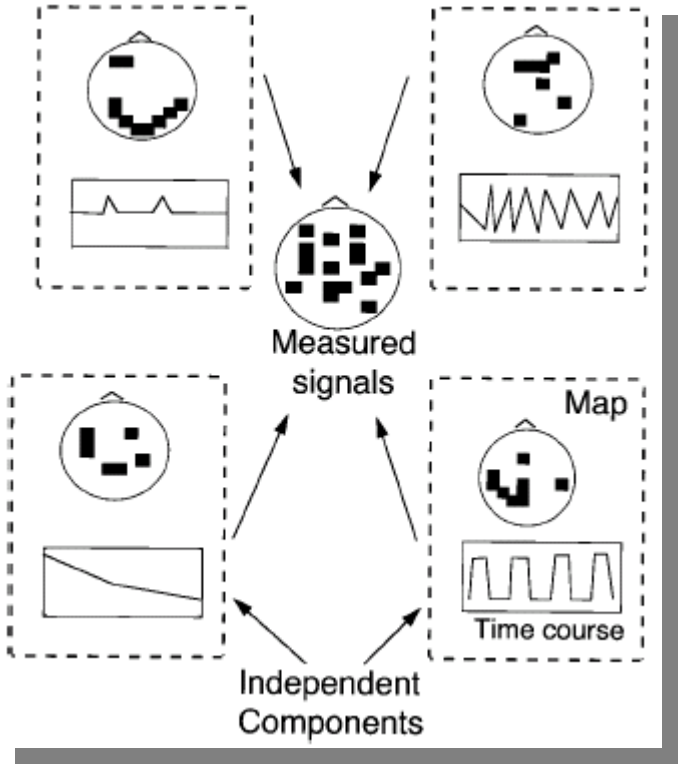


Figure 1.



ICA and fMRI signal processing

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

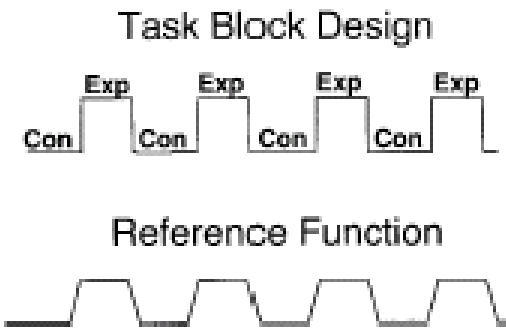


Figure 1.
BOLD signal complexity and task reference function. a: Time courses of 10 randomly selected voxels from a 6-min fMRI trial of the Stroop color-naming task illustrate the typical complexity of BOLD signals. b: Convolving an a priori estimate of the hemodynamic response function with the square-wave 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).

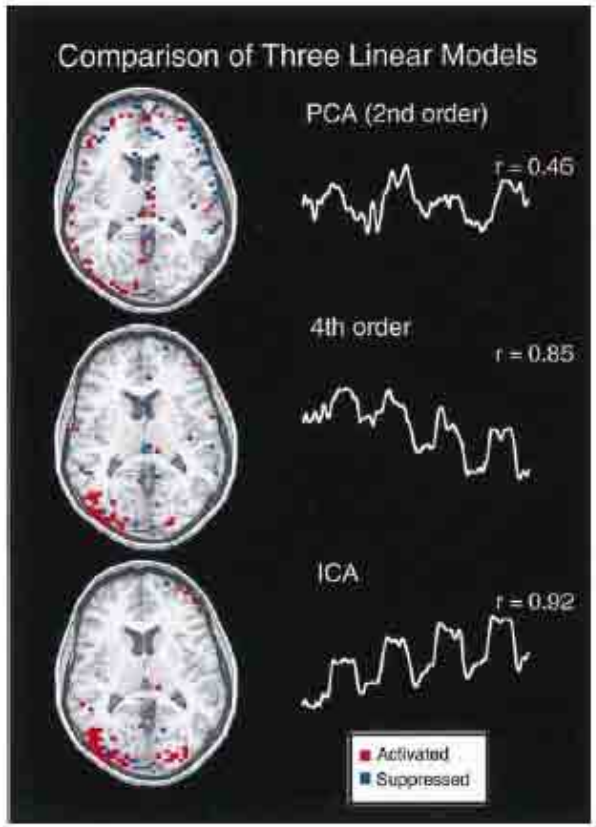
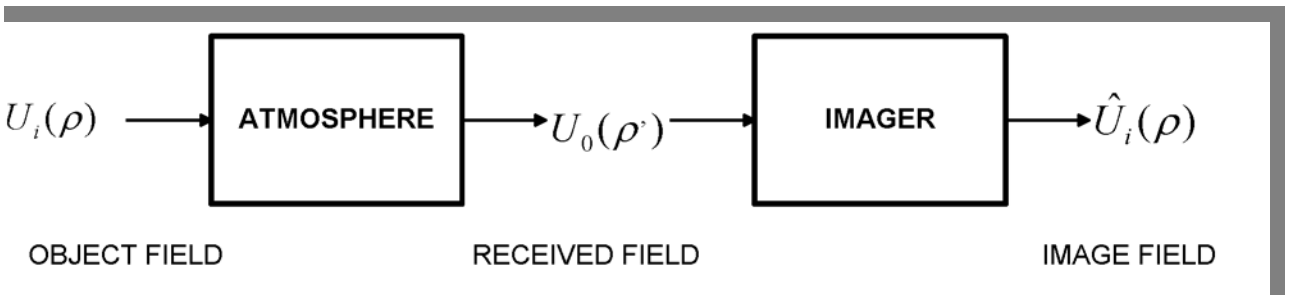


FIG. 2. Comparison of three linear models for analyzing fMRI data. PCA and two versions of ICA were used to linearly separate the data into partially spatially independent maps. The most consistently task-related component determined by each of the three methods from the first trial are shown, along 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.

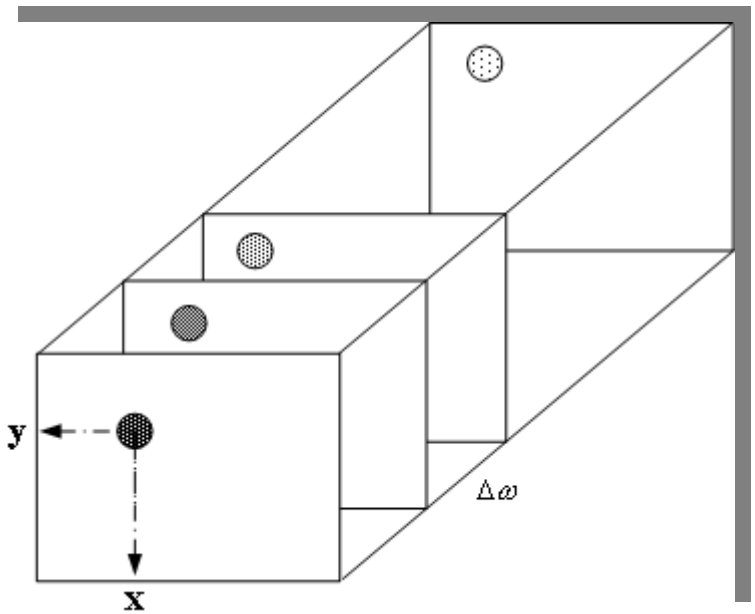
$$I_{ik}(t_k, x, y) = \sum_{n=1}^{N} a_{kn}(\Delta t_{kn}) I_{0n}(t_0, x, y)$$

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

ICA representation of the image sequence

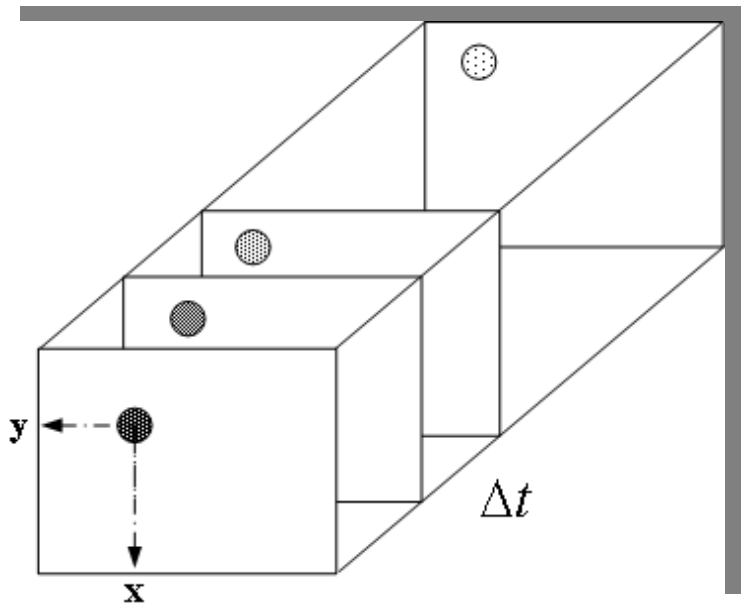
$$\mathbf{I}_i(x,y) = \mathbf{A}\mathbf{I}_\mathbf{o}(x,y) \quad + \quad \nu(x,y)$$

Image cube for multispectral imaging

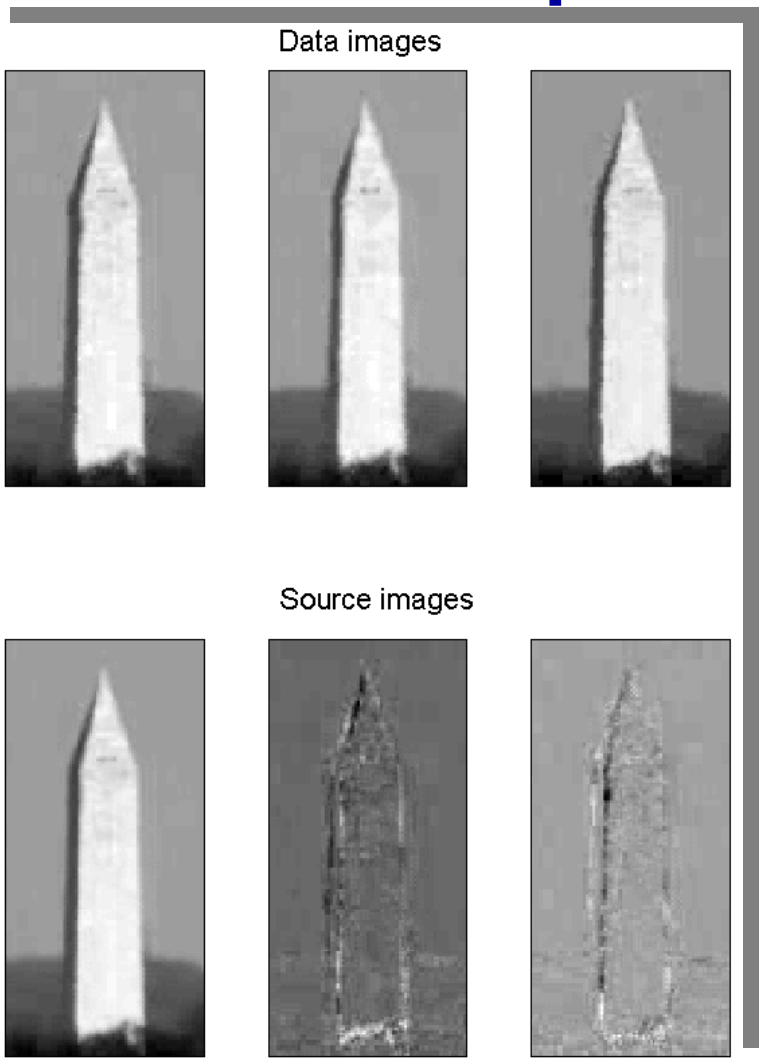


$\omega \Leftrightarrow t$

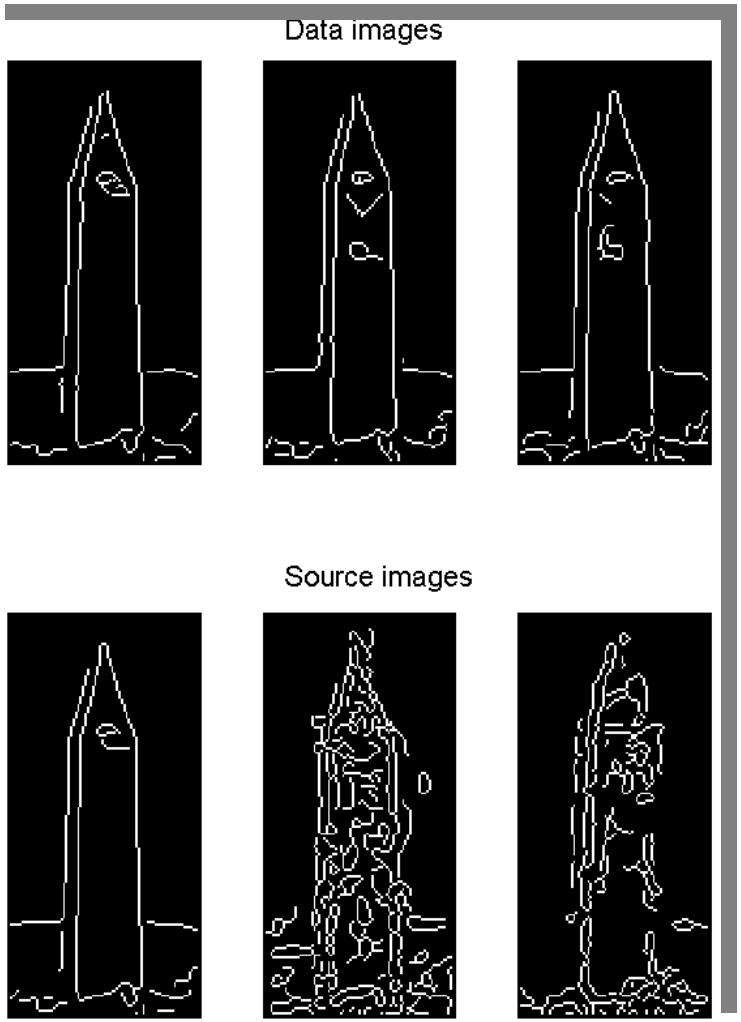
Image cube for video sequence



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).

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 ?

Increasing statistical independence

- Sub-band decomposition ICA (SDICA): wideband source signals are dependent, but there exist sub-bands where they are less dependent.
- Innovations-based approach.

A. Cichocki, P. Georgiev, Blind source separation algorithms with matrix xonstraints, IEICE Trans. Fund. Electron. Commun. Comput. Sci. E86-A (2003) 522-531.
T. Tanaka, A. Cichocki, Subband decomposition independent component analysis and new performance criteria, Proc. ICASSP, 2004.
I. Kopriva, D. Sersic, Wavelet packets approach to blind separation of statistically dependent sources, Neurocomputing **71**,1642-1655 (2008).
I. Kopriva, D. Sersic, Robust blind separation of statistically dependent sources using dual tree wavelets, ICIP 2007.
A. Hyvarinnen, Independent component analysis for time-dependent stochastic processes, ICANN’98, Skvode, Sweden, 1998.

Increasing statistical independence: innovations-based approach

•Argument for using innovations (prediction errors) is that they are *more independent* from each other and *more non-Gaussian* than original processes → *essentially important* for the success of the ICA algorithms.

•Innovations: $\tilde{s}_m(t) = s_m(t) - E\left[s_m(t) \middle| s_m(t-1), s_m(t-2), \ldots \right]$

$$\begin{aligned}
\tilde{\mathbf{x}}(t) &= \mathbf{x}(t) - E[\mathbf{x}(t) | \mathbf{x}(t-1), \mathbf{x}(t-2), \ldots] \\
&= \mathbf{A}\mathbf{s}(t) - E[\mathbf{A}\mathbf{s}(t) | \mathbf{A}\mathbf{s}(t-1), \mathbf{A}\mathbf{s}(t-2), \ldots] \\
&= \mathbf{A}\left[\mathbf{s}(t) - E[\mathbf{s}(t) | \mathbf{s}(t-1), \mathbf{s}(t-2), \ldots] \right] \\
&= \mathbf{A}\tilde{\mathbf{s}}(t)
\end{aligned}$$

Increasing statistical independence: innovations-based approach

- Innovation is realized through prediction error filtering:

$$\tilde{x}_n(t) = x_n(t) - \sum_{k=1}^{K} h_n(k) x_n(t-K)$$

$\mathbf{h}_n$ is learned for each x_n separately. Final prediction error filter is obtained as an average:

$$\mathbf{h} = \frac{1}{N} \sum_{n=1}^{N} \mathbf{h}_n$$

- Linear time invariant prediction error filter is efficiently estimated from data by means of Levinson algorithm (MATLAB command `lpc`). Thus, innovations actually are data adaptive *high-pass* filtering.

Increasing statistical independence: innovations-based approach

•Innovations are data adaptive *high-pass* filtering due to the fact that linear prediction error filter removes slow varying (predictable) part of the signal. Thus, through innovations a low frequency part of the spectrum is removed.

•In this regard even *fixed high pass* filters are efficient in enhancing statistical independence between the source signals.

•The first order high pass filter **h**=[1 –1] is very useful in various image processing problems.

Increasing statistical independence: SDICA approach

• In SDICA approach the operator *T* represents prefilter applied to all observed signals.

• The wideband source signals are dependent, but some of their subcomponents are independent.

$$\mathbf{s}(t) = \sum_{l=1}^{L} \mathbf{s}_l(t)$$

•The challenge is how to find a subband index $1 \le k \le L$, such that $\mathbf{s}_k$ contains least dependent subcomponents?

Increasing statistical independence: SDICA approach

• To locate sub-band with least dependent components small cumulant based approximation is used to measure the mutual information between the components of the measured signals in the corresponding nodes of the wavelet trees.

$$\hat{I}_k^j\left(x_{k1}^j, x_{k2}^j, ..., x_{kN}^j\right) \approx \frac{1}{4}\sum_{\substack{0\le n<l\le N\\ n\neq l}} cum^2(x_{kn}^j, x_{kl}^j) + \frac{1}{12}\sum_{\substack{0\le n<l\le N\\ n\neq l}} \left(cum^2(x_{kn}^j, x_{kn}^j, x_{kl}^j) + cum^2(x_{kn}^j, x_{kl}^j, x_{kl}^j)\right)$$

$$+\frac{1}{48}\sum_{\substack{0\le n<l\le N\\ n\neq l}} \left(cum^2(x_{kn}^j, x_{kn}^j, x_{kn}^j, x_{kl}^j) + cum^2(x_{kn}^j, x_{kn}^j, x_{kl}^j, x_{kl}^j) + cum^2(x_{kn}^j, x_{kl}^j, x_{kl}^j, x_{kl}^j)\right)$$

where *j* represents scale index and *k* represents sub-band index at the appropriate scale.

I. Kopriva, D. Sersic, Wavelet packets approach to blind separation of statistically dependent sources, Neurocomputing **71**, 1642-1655 (2008).

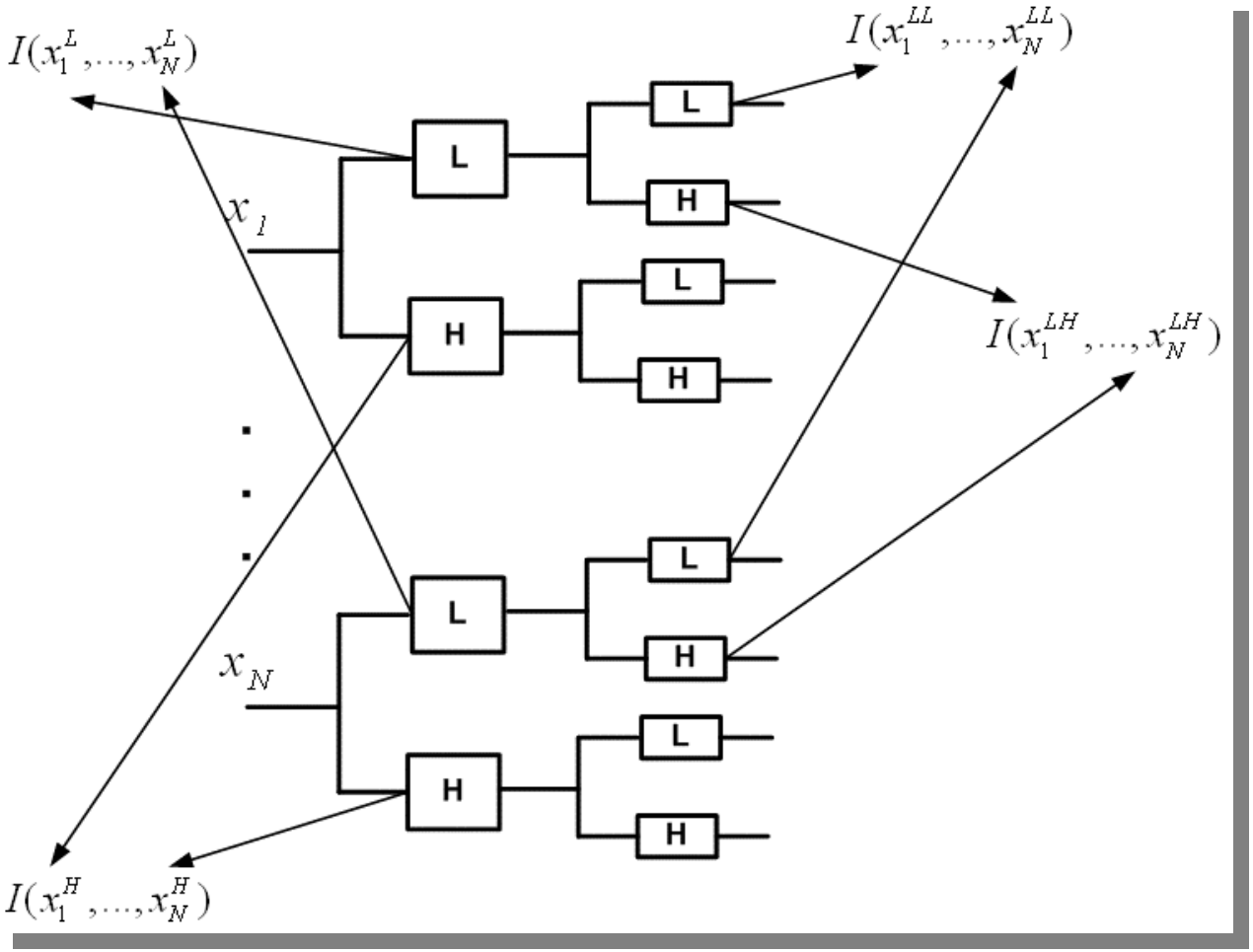
Increasing statistical independence: SDICA approach

Mutiscale analysis SDICA

$$k^* = \arg\min_k I\left(x_1^k, \ldots, x_N^k\right)$$

$$\mathbf{W} \cong ICA(\mathbf{x}_k)$$

$$\mathbf{s} \cong \mathbf{W}\mathbf{x}$$



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).

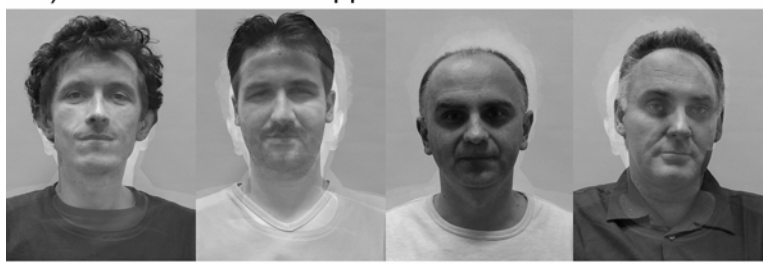
• Background Gaussian noise has been added as wide-band interferer to all source images with an average SNR ≅30dB.

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

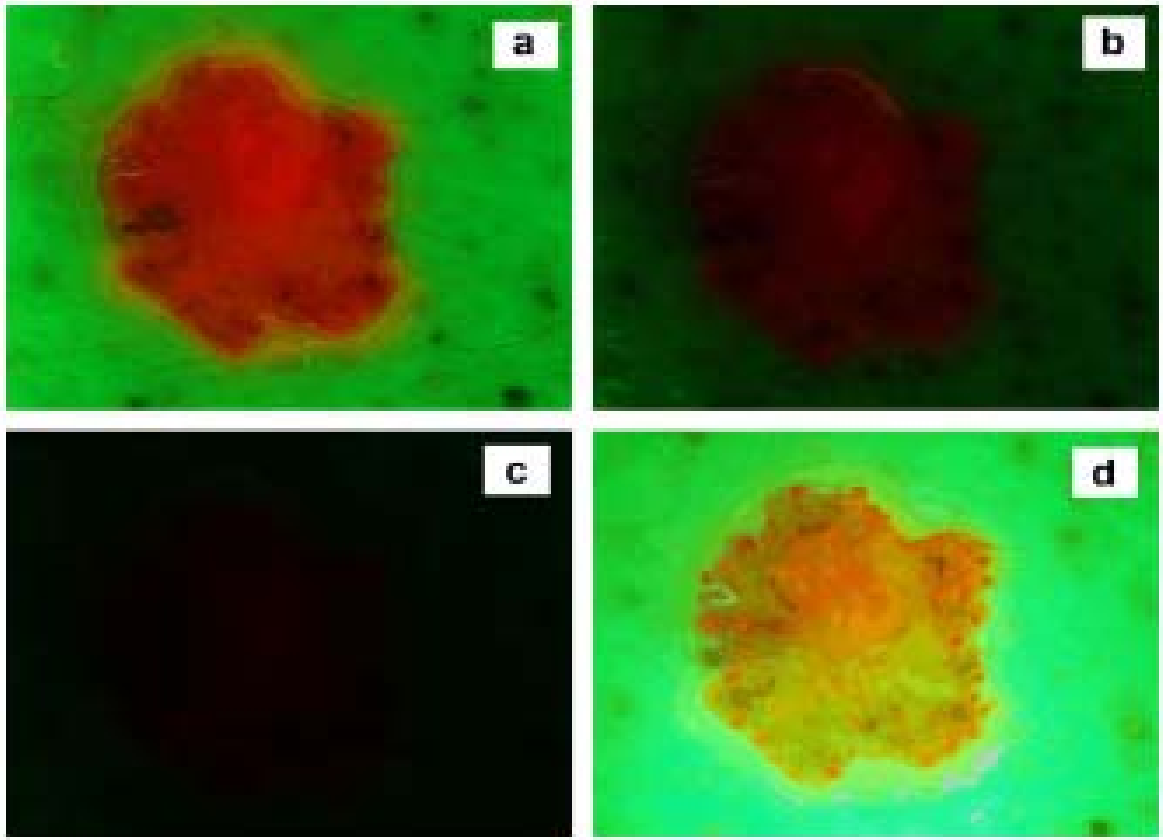
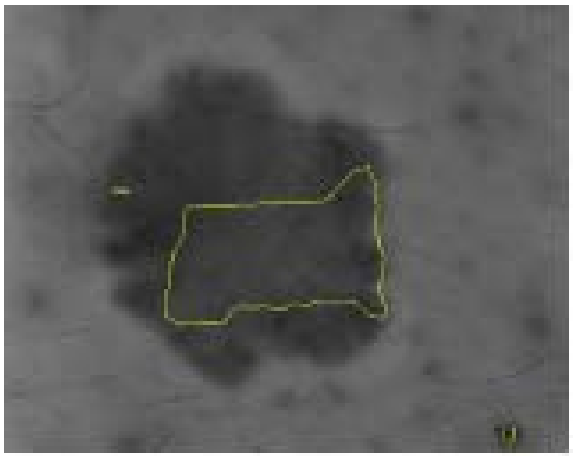


Fig. 1. RGB fluorescent images of the BCC from the first patient acquired under different intensities of illumination: (a) illumination with the maximal intensity I_0; (b) illumination with the intensity $I_0/9.15$; illumination with the intensity $I_0/73.47$; (d) RGB fluorescent image with demarcation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

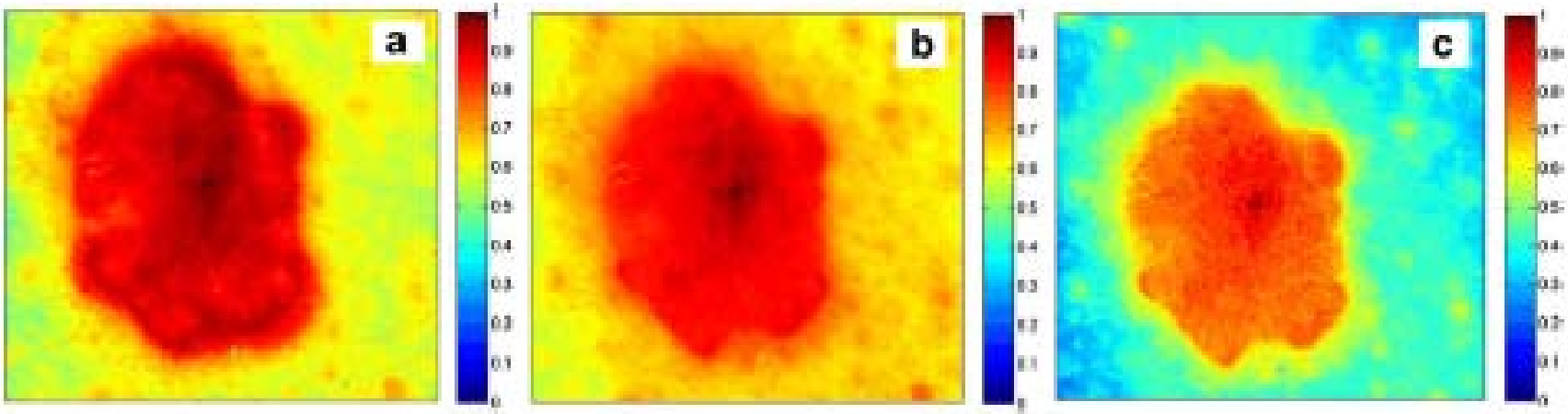


Fig. 6. BCC spatial maps $\mathbf{s}_i$ extracted from fluorescent RGB images shown in Fig. 1a–c by means of ERICA algorithm [36]. Extracted maps are normalized on interval $[0, 1]$ and shown in pseudo-color scale.

Robust demarcation of the basal cell carcinoma

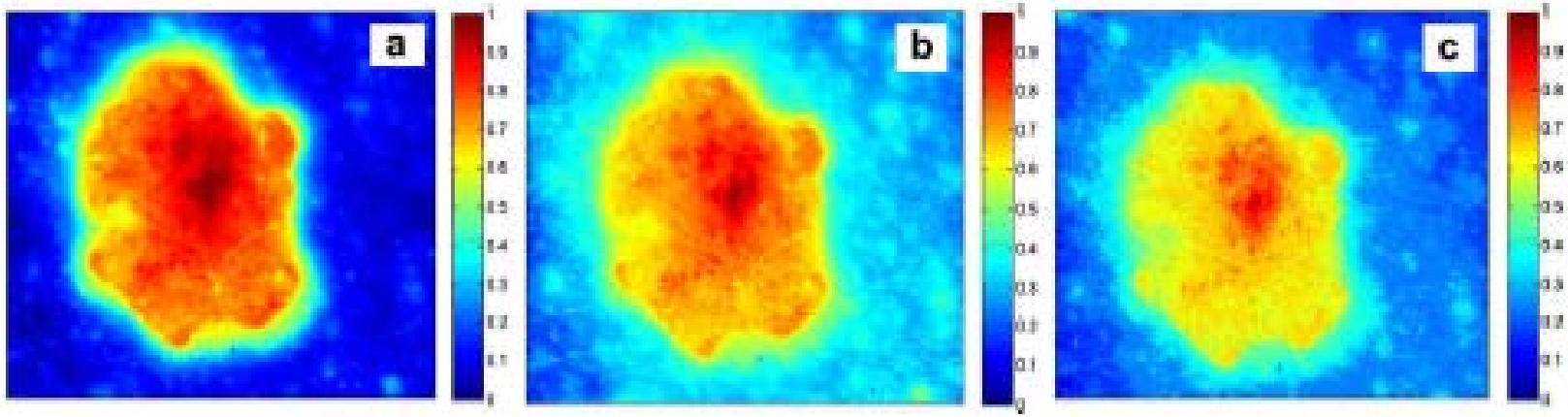


Fig. 7. BCC spatial maps is extracted from fluorescent RGB images shown in Fig. 1a–c by means of DCA-HPF algorithm. Extracted maps are normalized on interval [0, 1] and shown in pseudo-color scale.

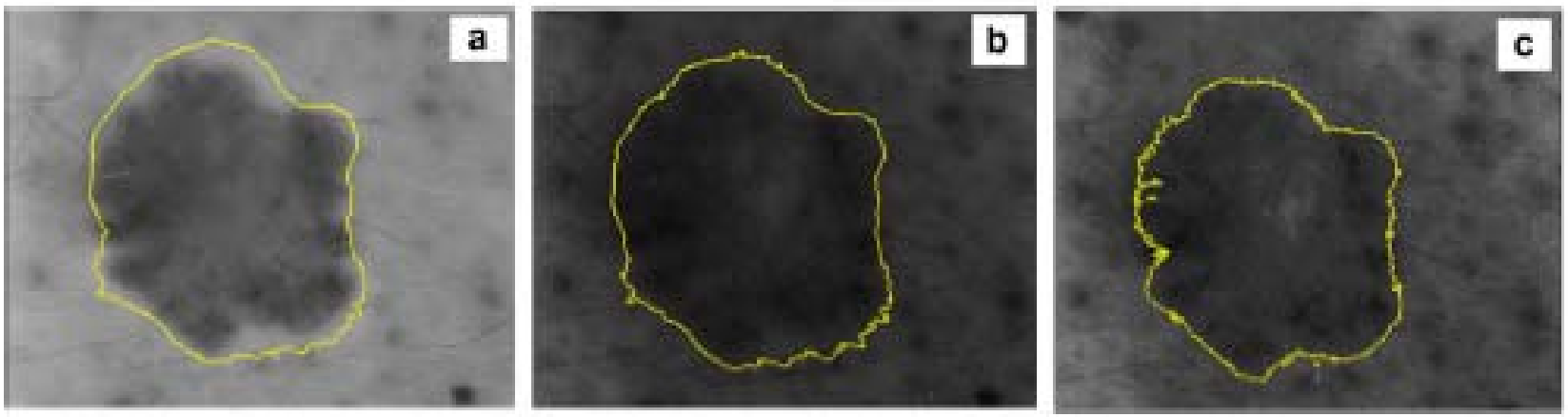


Fig. 8. BCC demarcation lines calculated by means of Canny's edge extraction method from spatial maps shown in Fig. 7a–c, with a fixed threshold set to 0.5. Demarcation lines were superimposed on the gray scale version of the fluorescent RGB images shown in Fig. 1a–c.

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).

Underdetermined BSS

Provided that prior on **s**(*t*) is Laplacian, maximum likelihood approach to maximization of posterior probability *P*(**s**|**x,A**) yields minimum L_1-norm as the solution:

$$\hat{\mathbf{s}}(t) = \max_{\hat{\mathbf{A}}\mathbf{s}(t)=\mathbf{x}(t)} P\left(\mathbf{s}(t)\middle|\mathbf{x}(t),\hat{\mathbf{A}}\right)$$

$$= \max_{\hat{\mathbf{A}}\mathbf{s}(t)=\mathbf{x}(t)} P\left(\mathbf{x}(t)\middle|\mathbf{s}(t),\hat{\mathbf{A}}\right)P\left(\mathbf{s}(t)\right)$$

$$= \max_{\hat{\mathbf{A}}\mathbf{s}(t)=\mathbf{x}(t)} P\left(\mathbf{s}(t)\right)$$

$$= \max_{\hat{\mathbf{A}}\mathbf{s}(t)=\mathbf{x}(t)} \exp-\left(\left|\mathbf{s}_1(t)\right|+\ldots+\left|\mathbf{s}_M(t)\right|\right)$$

$$= \min_{\hat{\mathbf{A}}\mathbf{s}(t)=\mathbf{x}(t)} \left|\mathbf{s}_1(t)\right|+\ldots+\left|\mathbf{s}_M(t)\right|$$

$$= \min_{\hat{\mathbf{A}}\mathbf{s}(t)=\mathbf{x}(t)} \left\|\mathbf{s}(t)\right\|_1$$

uBSS – L_1 norm minimization

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

- estimate basis or mixing matrix **A** using data clustering.

- estimate sources **s** solving underdetermined linear system of equations **x**=**As**. Provided that **s** is sparse enough, solution is obtained at the minimum of L_1-norm. Due to convexity L_1-norm is used as a replacement for L_0-quasi-norm.

- accuracy of the estimation of the mixing matrix **A** can be improved significantly when it is estimated on a set of single component points i.e. points where only one component/source is present.

uBSS – L_1 norm minimization

- at the points *t* of single source activity the following relation holds:

$$\mathbf{x}_t = \mathbf{a}_j s_{jt}$$

where *j* denotes the source index that is present at point *t*. At these points the mixing vector $\mathbf{a}_j$ is collinear with data vector $\mathbf{x}_t$.

- it is assumed that data vector and source components are complex. If not, Hilbert transform-based analytical expansion can be used to obtain complex representation.
- if single source points can not be found in original domain a linear transform such as wavelet transform, Fourier transform or Short-time Fourier transform can be used to obtain sparse representation:

$$T(\mathbf{x})_t = \boldsymbol{a}_j \, T(\mathrm{s_j})_t$$

I.Kopriva, I. Jerić, “Blind separation of analytes in nuclear magnetic resonance spectroscopy and mass spectrometry: sparseness-based robust multicomponent analysis,“ Analytical Chemistry **82**: 1911-1920 (2010).

uBSS – L_1 norm minimization

- since the mixing vector is real the real and imaginary part of data vector $\mathbf{x}_t$ must point in the same direction when real and imaginary part of s_{jt} have the same sign. Otherwise, they must point into opposite directions.

Thus, such points can be identified using:

$$\left| \frac{R\{\mathbf{x}_t\}^{\mathrm{T}} I\{\mathbf{x}_t\}}{\|R\{\mathbf{x}_t\}\| \|I\{\mathbf{x}_t\}\|} \right| \geq \cos(\Delta\theta)$$

where $R\{\mathbf{x}_t\}$ and $I\{\mathbf{x}_t\}$ denote real and imaginary part of $\mathbf{x}_t$, and $\Delta\theta$ denotes angular displacement from a direction of 0 or π radians.

V.G. Reju, S.N. Koh, I. Y. Soon, "An algorithm for mixing matrix estimation in instantaneous blind source separation," Signal Processing **89**, 1762-1773 (2009).

S.G. Kim, C.D. Yoo, "Underdetermined Blind Source Separation Based on Subspace Representation," IEEE Trans. Signal Processing **57**, 2604-2614 (2009).

uBSS – L_1 norm minimization

- several methods to solve underdetermined linear system of equations are linear programming:

$$\hat{\mathbf{s}}(t) = \arg\min_{\mathbf{s}(t)} \sum_{m=1}^{\hat{M}} s_m(t) \quad \text{s.t. } \hat{\mathbf{A}}\mathbf{s}(t) = \mathbf{x}(t) \quad \forall t = 1,\ldots,T$$

$$\text{s.t. } \mathbf{s}(t) \geq 0$$

L_1-regularized least square problem:

$$\hat{\mathbf{s}}(t) = \arg\min_{\mathbf{s}(t)} \frac{1}{2}\left\|\hat{\mathbf{A}}\mathbf{s}(t) - \mathbf{x}(t)\right\|_2^2 + \lambda \left\|\mathbf{s}(t)\right\|_1 \qquad \forall t = 1,\ldots,T$$

and L_2-regularized linear problem:

$$\hat{\mathbf{s}}(t) = \arg\min_{\mathbf{s}(t)} \left\|\mathbf{s}(t)\right\|_1 \quad \text{s.t. } \left\|\hat{\mathbf{A}}\mathbf{s}(t) - \mathbf{x}(t)\right\|_2^2 \leq \varepsilon \qquad \forall t = 1,\ldots,T$$

S.J. Kim, K. Koh, M. Lustig, S. Boyd, D. Gorinevsky, "An Interior-Point Method for Large-Scale -Regularized Least Squares,"IEEE Journal of Selected Topics in Signal Processing **1**, 606-617 (2007).
E. van den Berg, M.P. Friedlander, "Probing the Pareto Frontier for Basis Pursuit Solutions," SIAM J. Sci. Comput. 31, 890-912 (2008).
M.A.T. Figuiredo, R.D. Nowak, S.J. Wright, "Gradient Projection for Sparse Reconstruction: Application to Compressed Sensing and Other Inverse Problems," IEEE Journal on Selected Topics in Signal Processing **1**, 586-597 (2007).

uBSS – clustering

Assuming unit L_2-norm of $\boldsymbol{a}_m$ and N=2 we can parameterize column vectors in a plane by one angle

$$\mathbf{a}_m = [\cos(\varphi_m) \;\; \sin(\varphi_m)]^{\mathrm{T}}$$

Assuming that **s** is *1*-sparse in representation domain estimation of **A** and M is obtained by data clustering algorithms.

•We remove all data points close to the origin for which applies: $\left\{ \left| \mathbf{x}(t) \right|_2 \le \varepsilon \right\}_{t=1}^{T}$ where ε represents some predefined threshold.

•Normalize to unit L_2-norm remaining data points $\mathbf{x}(t)$, i.e., $\left\{ \mathbf{x}(t) \to \mathbf{x}(t) / \left| \mathbf{x}(t) \right|_2 \right\}_{t=1}^{\bar{T}}$

F.M. Naini, G.H. Mohimani, M. Babaie-Zadeh, Ch. Jutten, "Estimating the mixing matrix in Sparse Component Analysis (SCA) based on partial *k*-dimensional subspace clustering," Neurocomputing **71**, 2330-2343 (2008).

uBSS – clustering

• Calculate function $f(\mathbf{a})$:

$$f(\mathbf{a}) = \sum_{t=1}^{\bar{T}} \exp\left(-\frac{d^2(\mathbf{x}(t), \mathbf{a})}{2\sigma^2} \right)$$

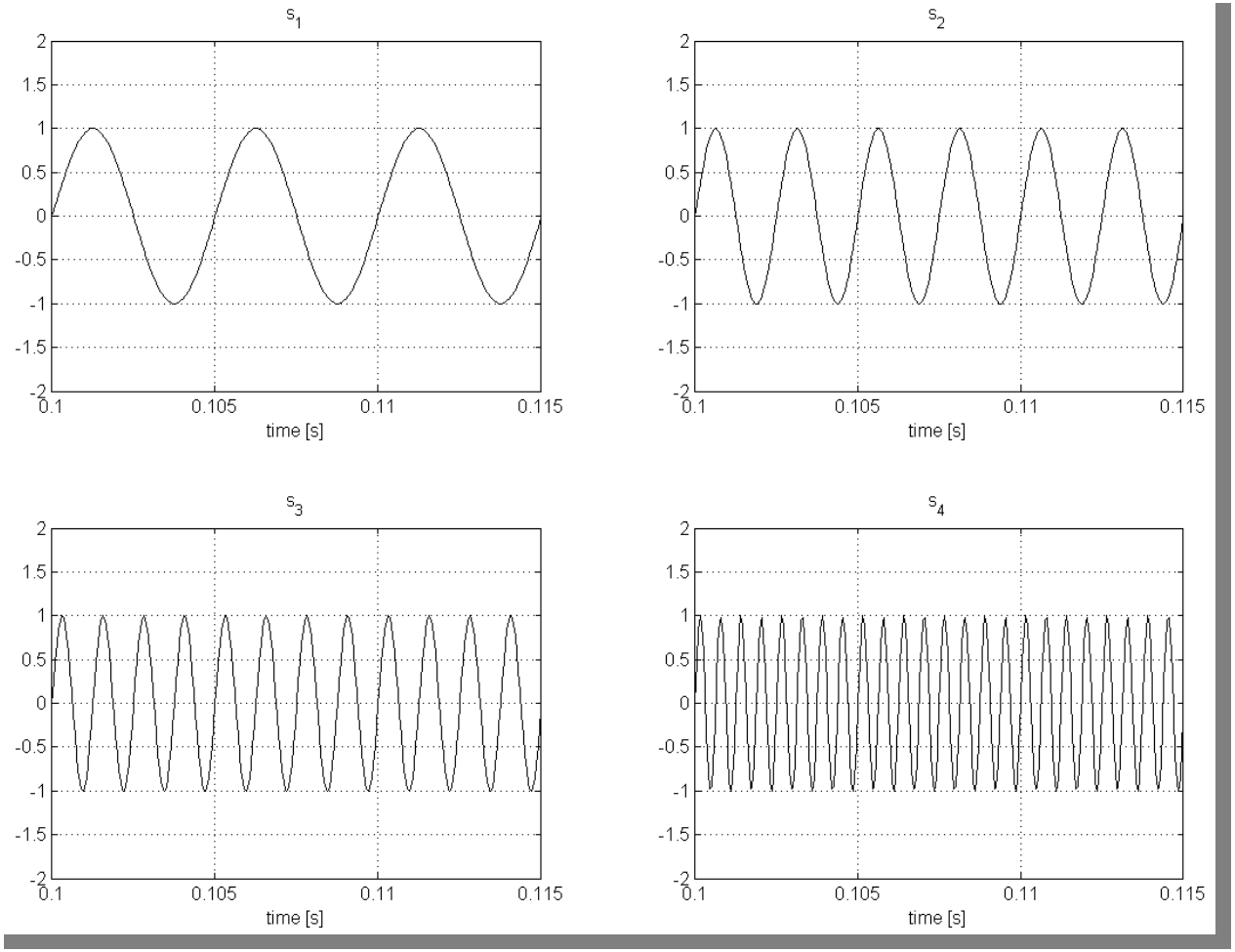
Where $d(\mathbf{x}(t), \mathbf{a}) = \sqrt{1 - (\mathbf{x}(t) \cdot \mathbf{a})^2}$ and $(\mathbf{x}(t) \cdot \mathbf{a})$ denotes inner product. Parameter σ is called dispersion. If set to sufficiently small value the value of the function $f(\mathbf{a})$ will approximately equal the number of data points close to **a**. Thus by varying mixing angle φ we effectively cluster data.

• Number of peaks of the function $f(\mathbf{a})$ corresponds with the estimated number of materials M. Locations of the peaks correspond with the estimates of the mixing angles $\{(\hat{\varphi}_m)\}_{m=1}^{\hat{M}}$, i.e., mixing vectors $\{\hat{\mathbf{a}}_m\}_{m=1}^{\hat{M}}$.

Blind separation of four sine signals from two mixtures

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

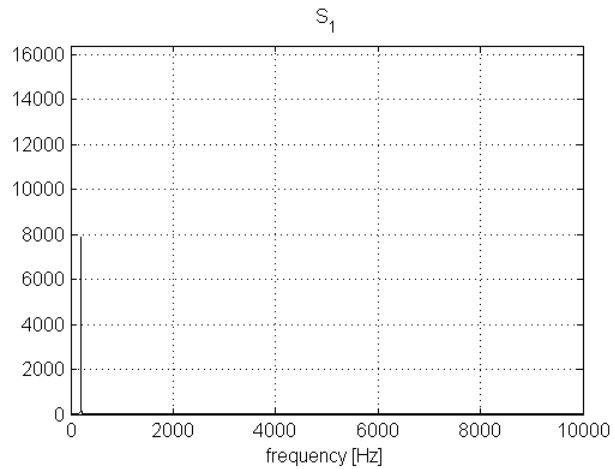
TIME DOMAIN

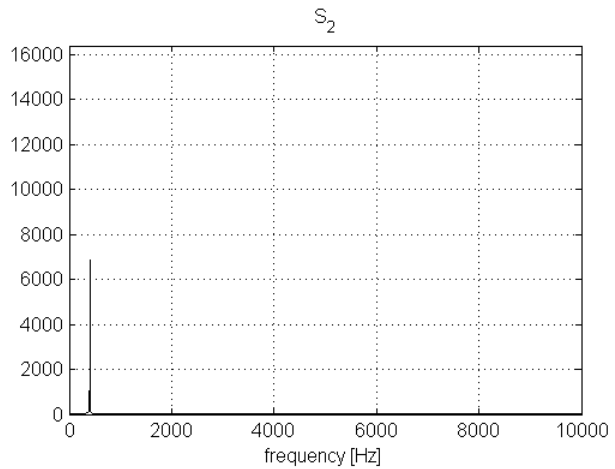


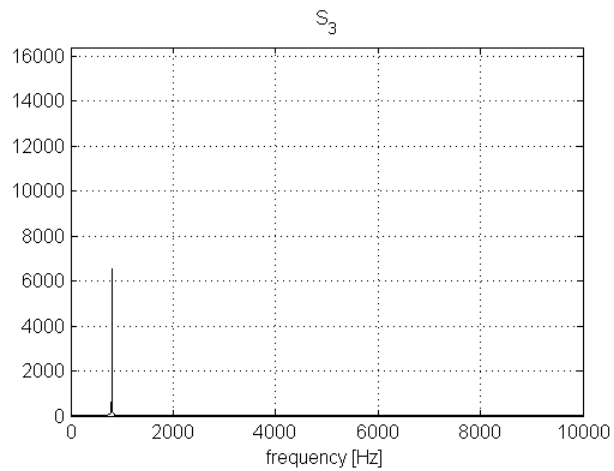
Blind separation of four sine signals from two mixtures

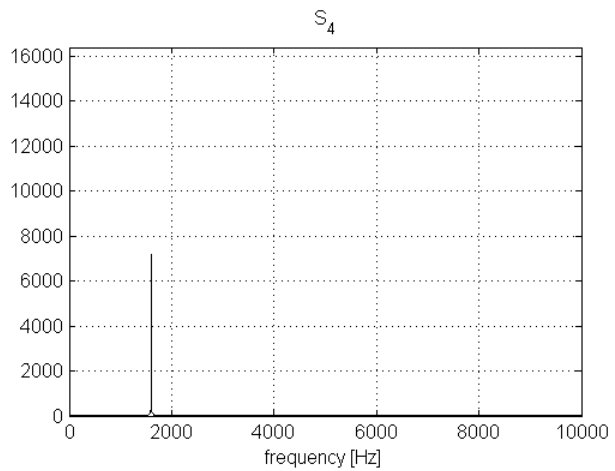
Four sinusoidal signals with frequencies 200Hz, 400Hz, 800Hz and 1600Hz.

FREQUENCY DOMAIN



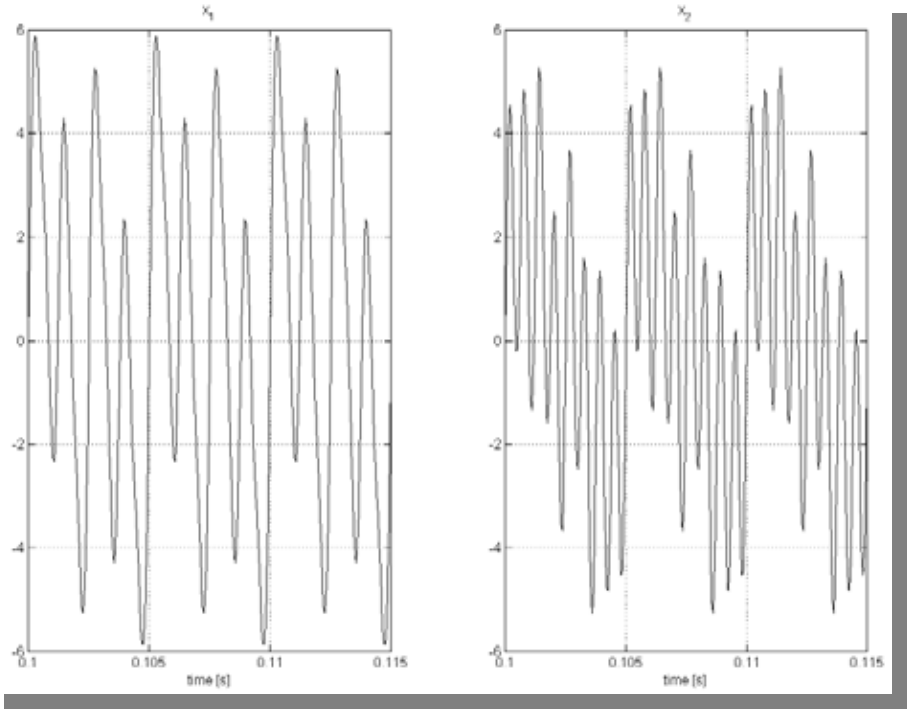




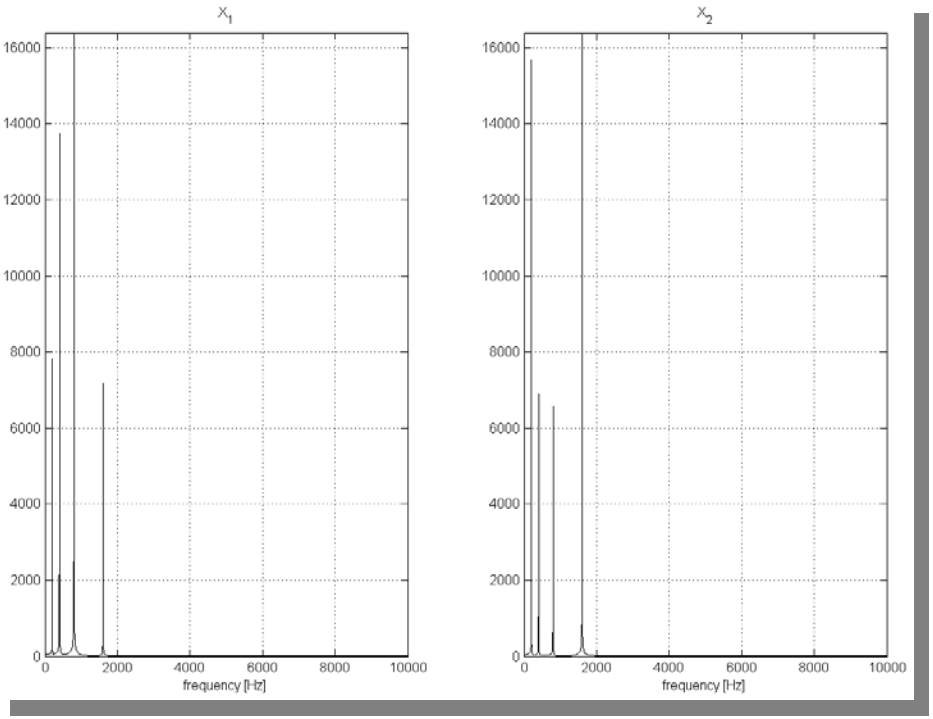


Blind separation of four sine signals from two mixtures

Two mixed signals



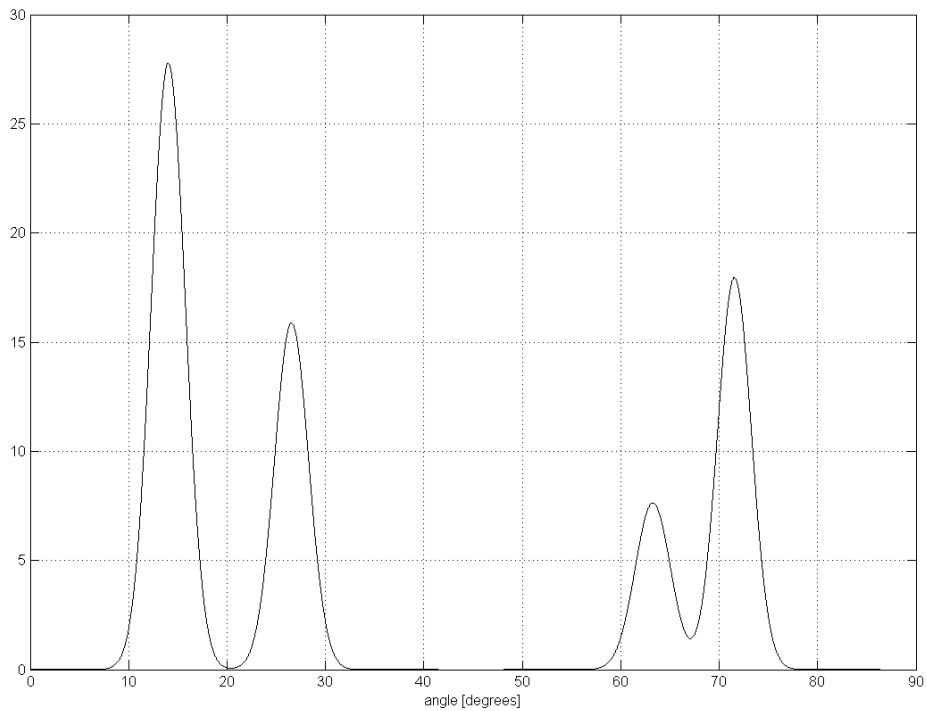
TIME DOMAIN



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

Linear programming based estimation of the sources using estimated mixing matrix **A**

$$\begin{bmatrix} \mathbf{x}_r(\omega) \\ \mathbf{x}_i(\omega) \end{bmatrix} = \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{A} \end{bmatrix} \begin{bmatrix} \mathbf{s}_r(\omega) \\ \mathbf{s}_i(\omega) \end{bmatrix}$$

or:

$$\overline{\mathbf{x}}(\omega) = \overline{\mathbf{A}}\,\overline{\mathbf{s}}(\omega)$$

$\mathbf{s}_r(\omega)$ and $\mathbf{s}_i(\omega)$ are not necessarily nonnegative. Thus, constraint $\overline{\mathbf{s}}(\omega) \geq \mathbf{0}$ required by linear program is not satisfied. In such a case it is customary to introduce dummy variables: $\mathbf{u},\mathbf{v} \geq \mathbf{0}$, such that $\overline{\mathbf{s}}(\omega) = \mathbf{u} - \mathbf{v}$.

Blind separation of four sine signals from two mixtures

Introducing:

$$\mathbf{z}(\omega)=\begin{pmatrix}\mathbf{u}\\ \mathbf{v}\end{pmatrix} \qquad \tilde{\mathbf{A}}=\begin{bmatrix}\overline{\mathbf{A}} & -\overline{\mathbf{A}}\end{bmatrix}$$

yields:

$$\hat{\mathbf{z}}(\omega)=\arg\min_{\mathbf{z}(\omega)}\sum_{m=1}^{4M} z_m(\omega) \quad \text{s.t.} \quad \tilde{\mathbf{A}}\mathbf{z}(\omega)=\overline{\mathbf{x}}$$

$$\mathbf{z}(\omega)\geq\mathbf{0}$$

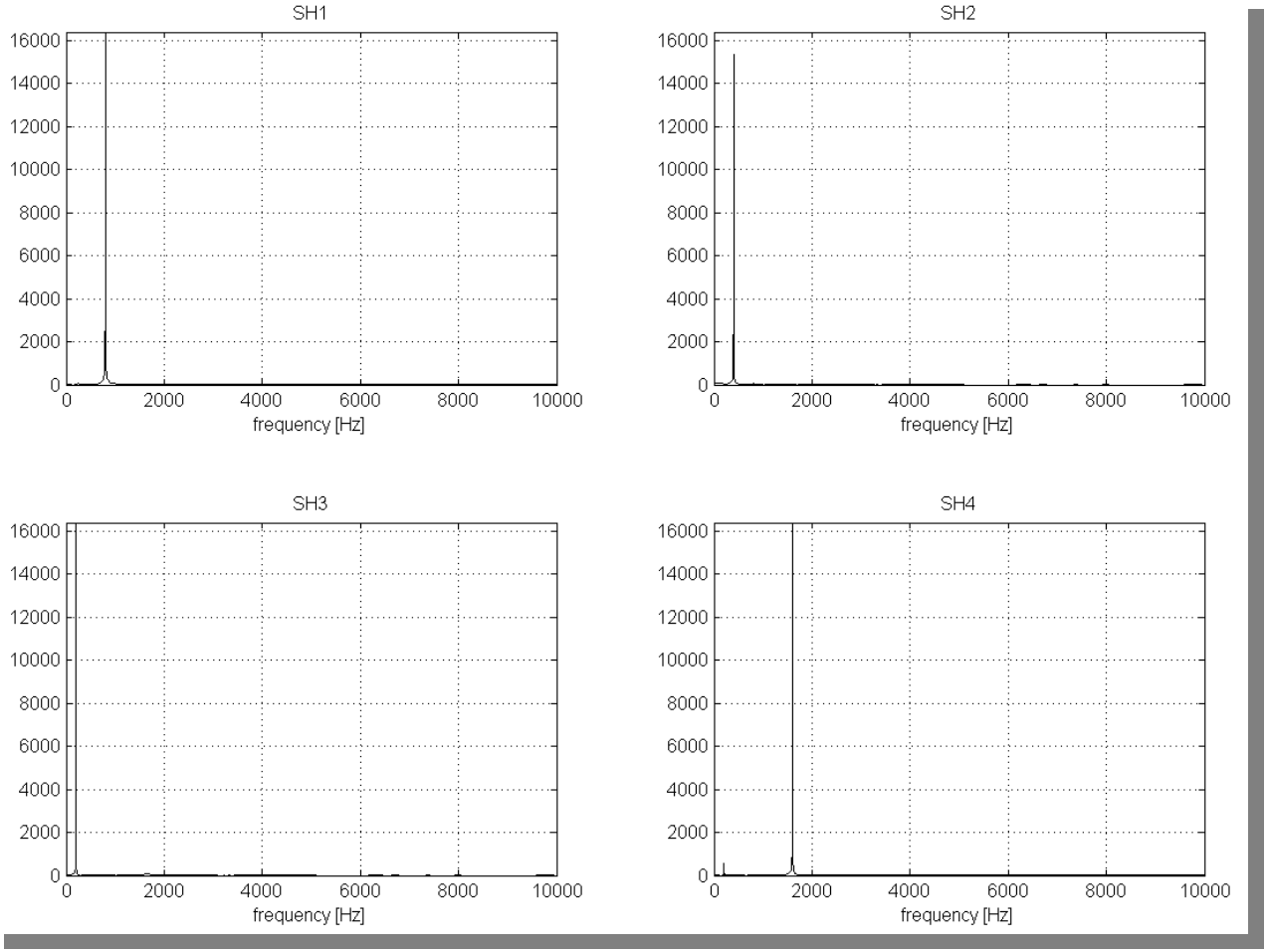
We obtain $\overline{\mathbf{s}}(\omega)$ from $\hat{\mathbf{z}}(\omega)$ as:

$$\overline{\mathbf{s}}(\omega)=\hat{\mathbf{u}}-\hat{\mathbf{v}}$$

and s(t) as:

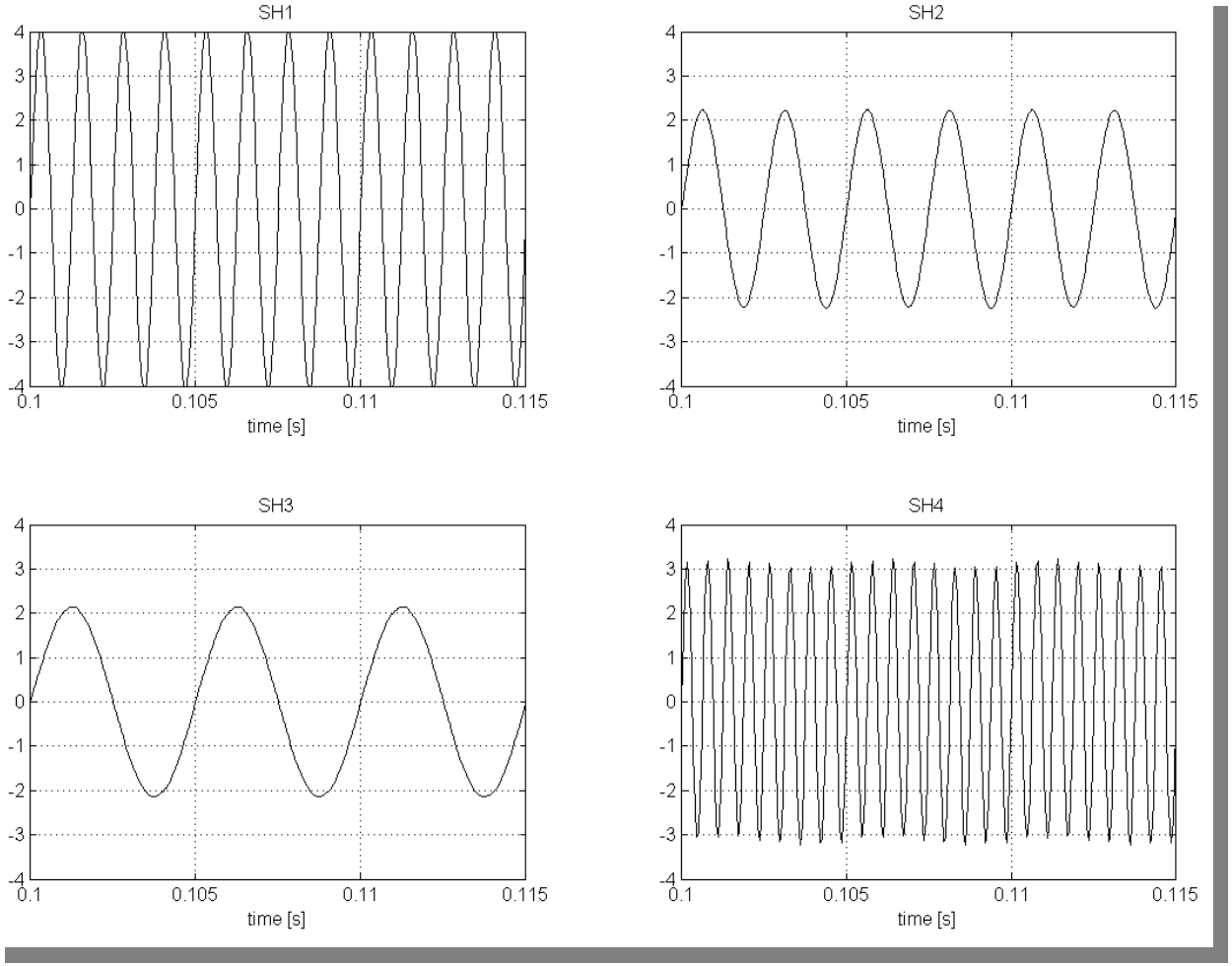
$$\hat{s}_m(t)=IDFT\left[\hat{s}_m(\omega)\right] \qquad \forall\, m=1,\ldots,M$$

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

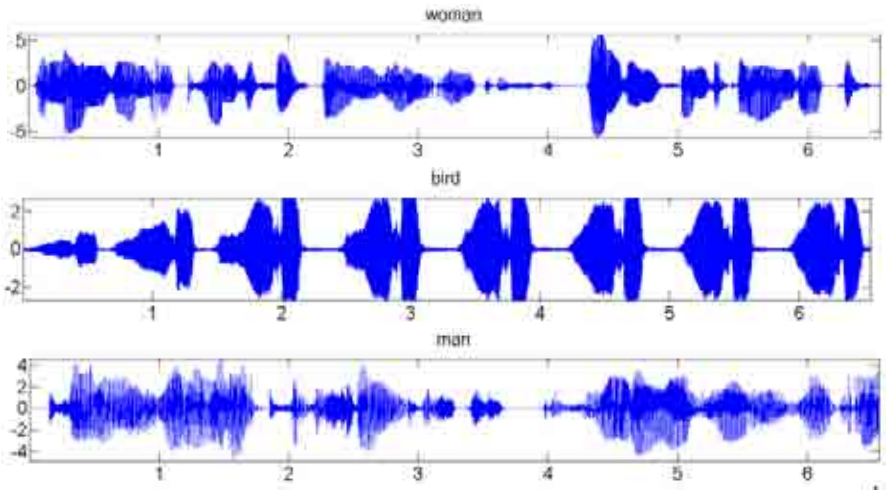


Estimated sources in TIME DOMAIN

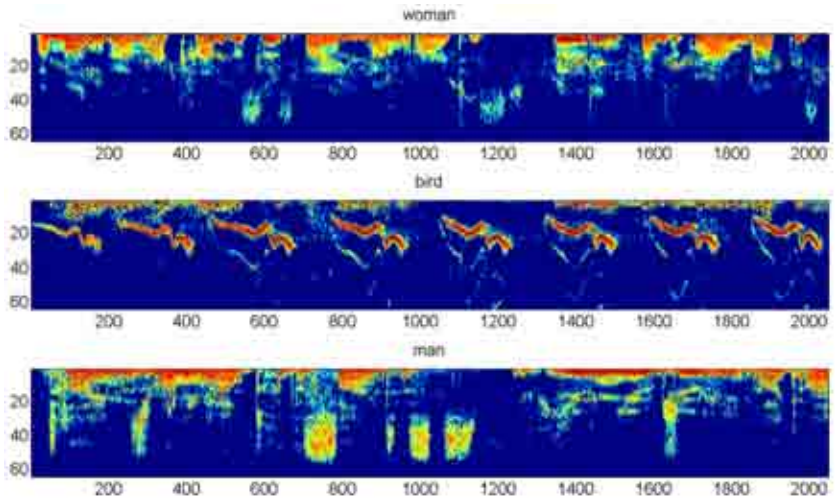
Blind separation of three sounds from two mixtures

Blind separation of three sounds from two mixtures

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



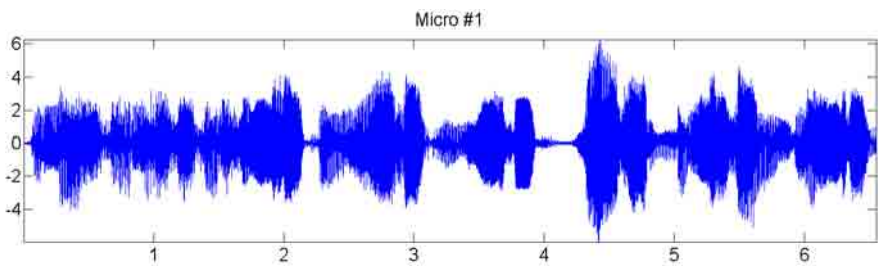
Time domain waveforms

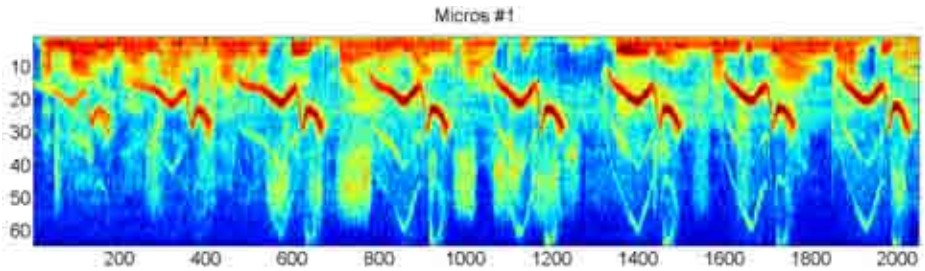


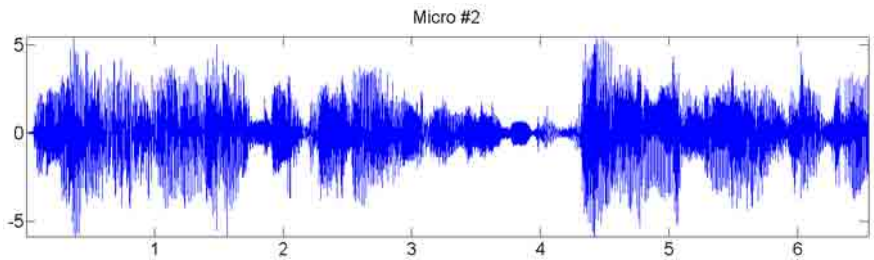
Time-frequency representations

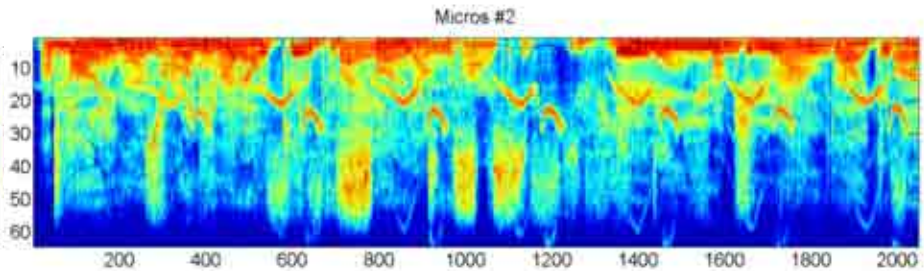
Blind separation of three sounds from two mixtures

Two mixtures of sounds:







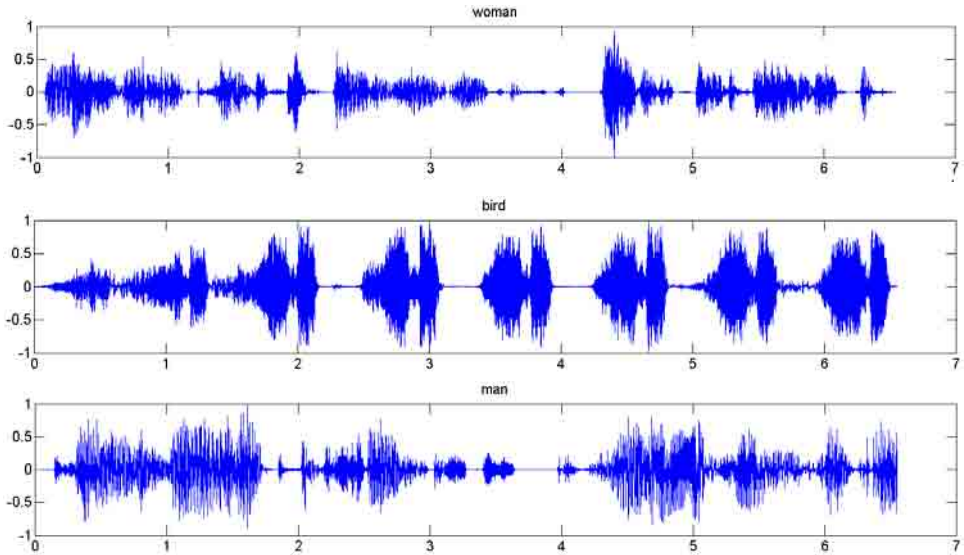


Time domain waveforms

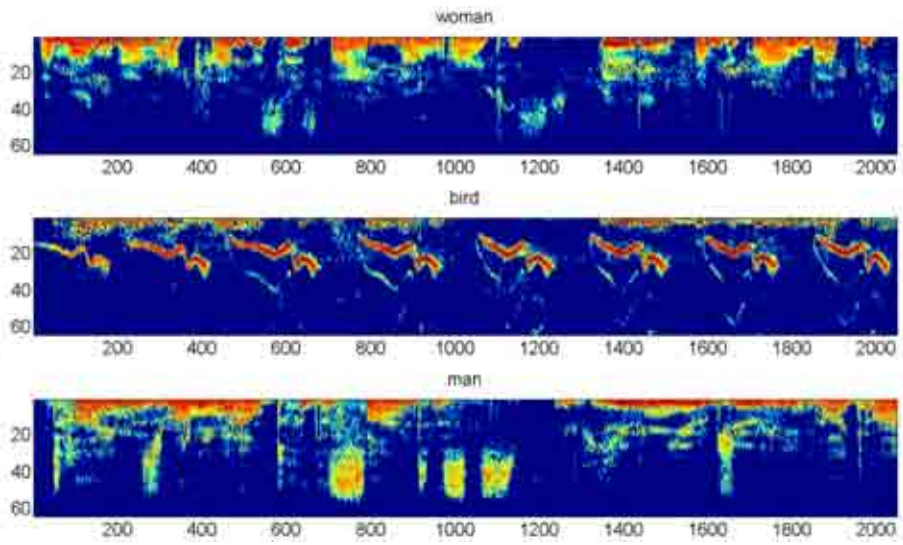
Time-frequency representations

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:



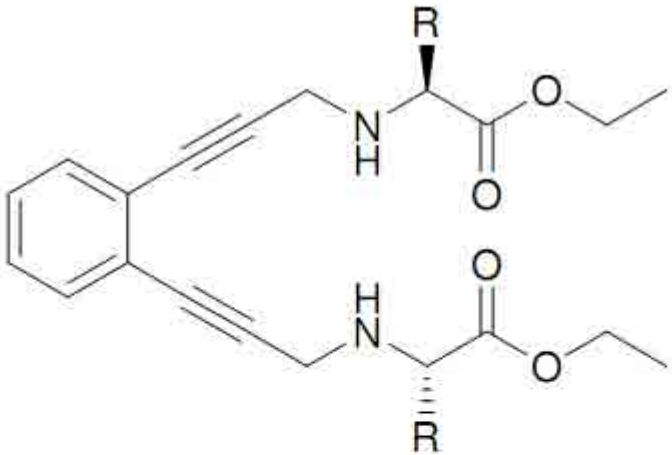
Time domain waveforms



Time-frequency representations

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

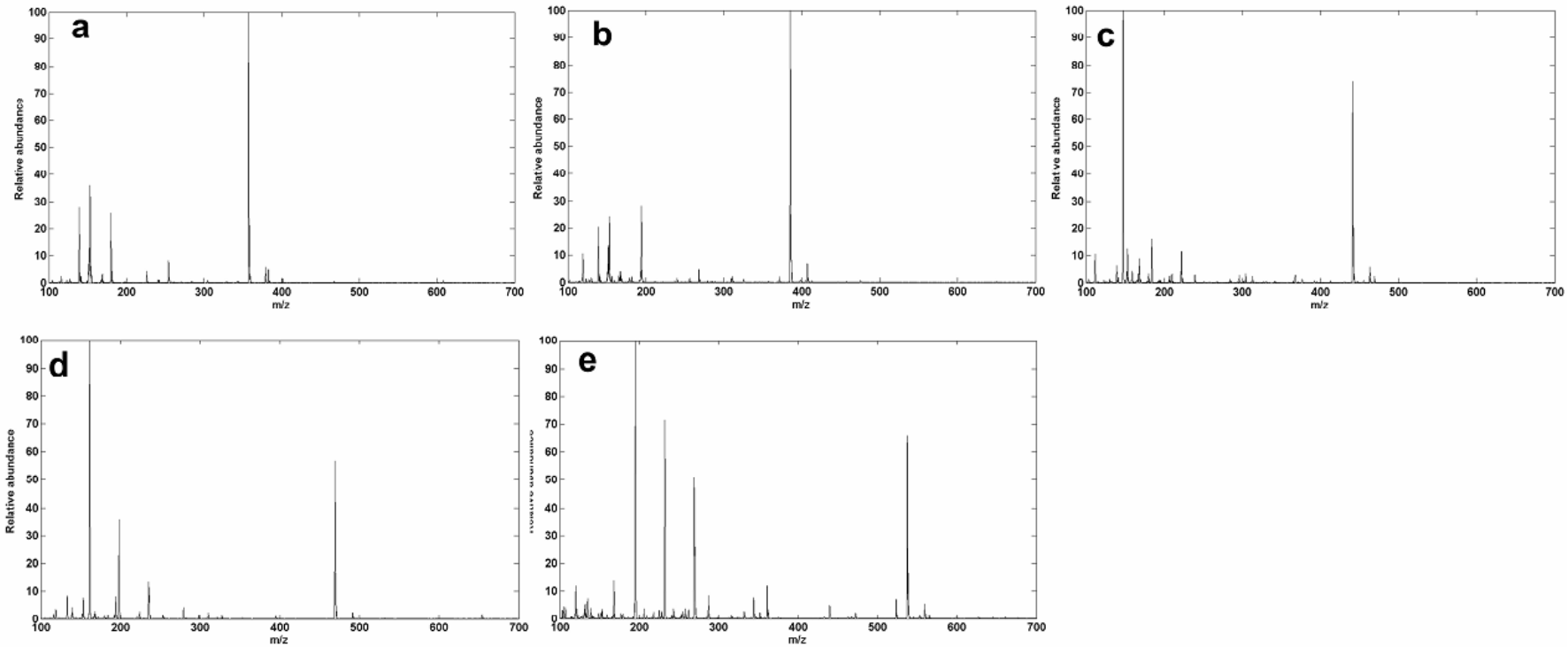
I. Kopriva, I. Jerić (2010). Blind separation of analytes in nuclear magnetic resonance spectroscopy and mass spectrometry: sparseness-based robust multicomponent analysis, **Analytical Chemistry 82**:1911-1920.
I. Kopriva, I. Jerić, V. Smrečki (2009). Extraction of multiple pure component ^{1}H and ^{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**.**



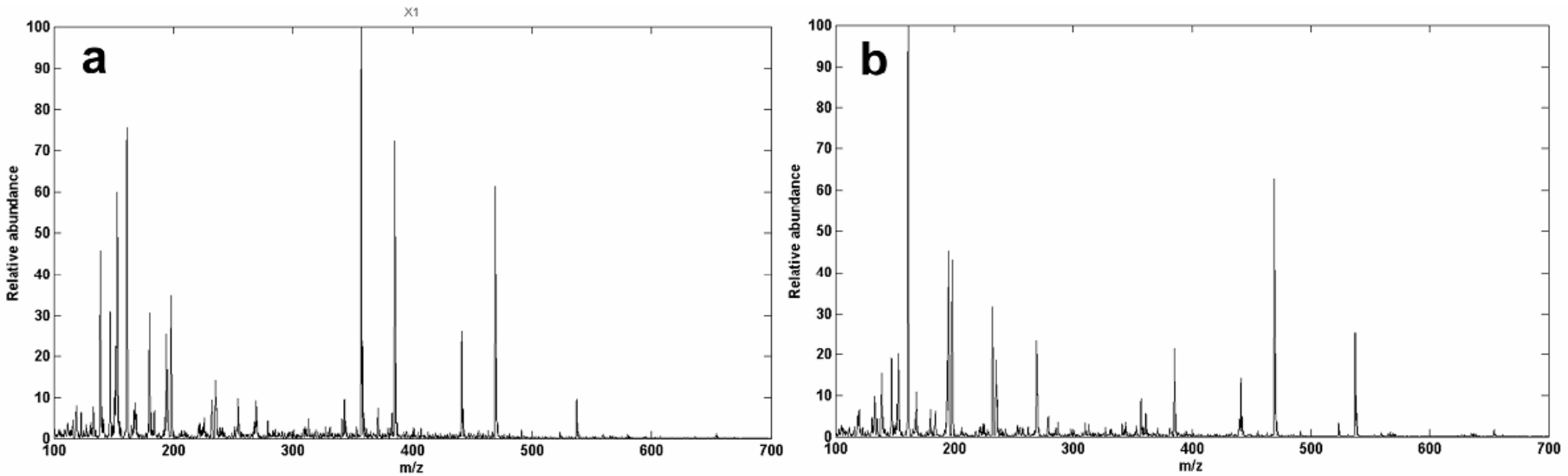
5 R=H
6 R=CH_3
7 R=$CH(CH_3)_2$
8 R=$CH_2CH(CH_3)_3$
9 R=$CH_2C_6H_5$

Figure S-1.

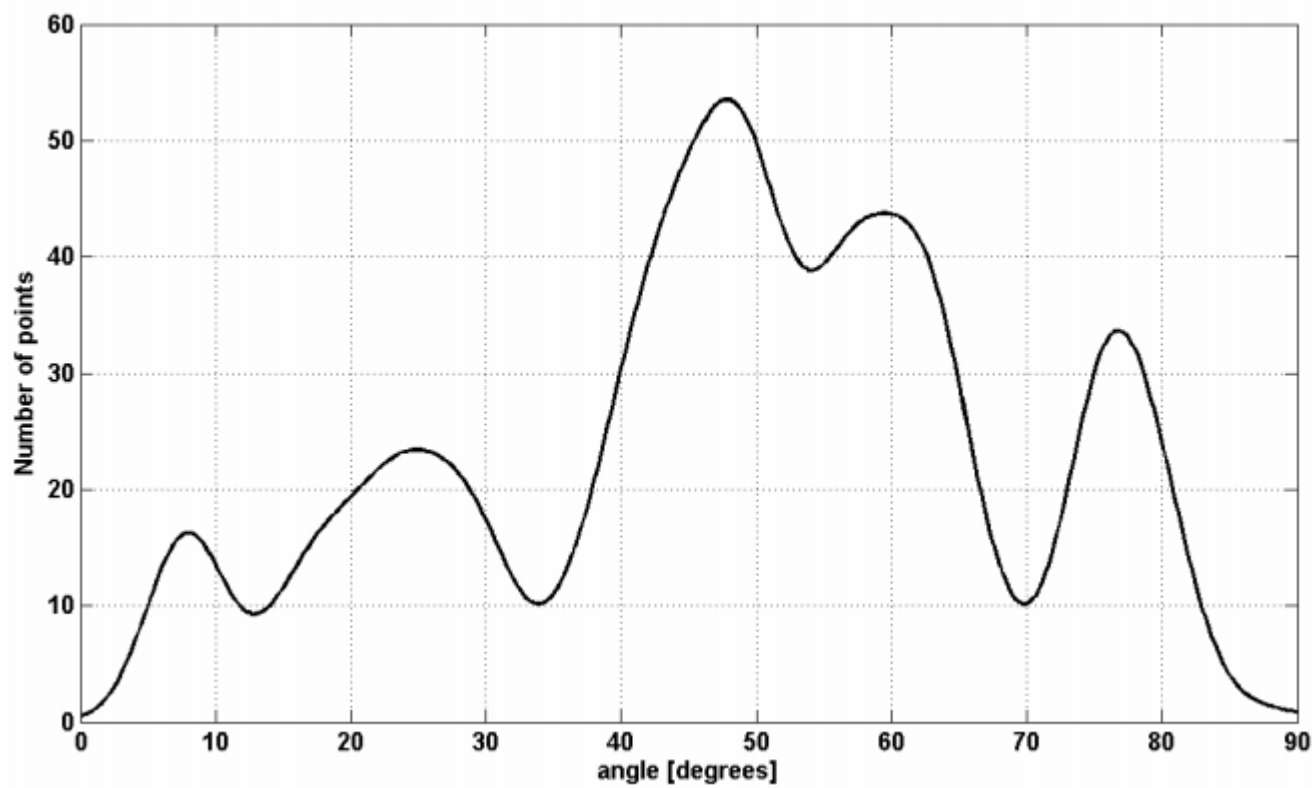
Chemical structure of five pure components.



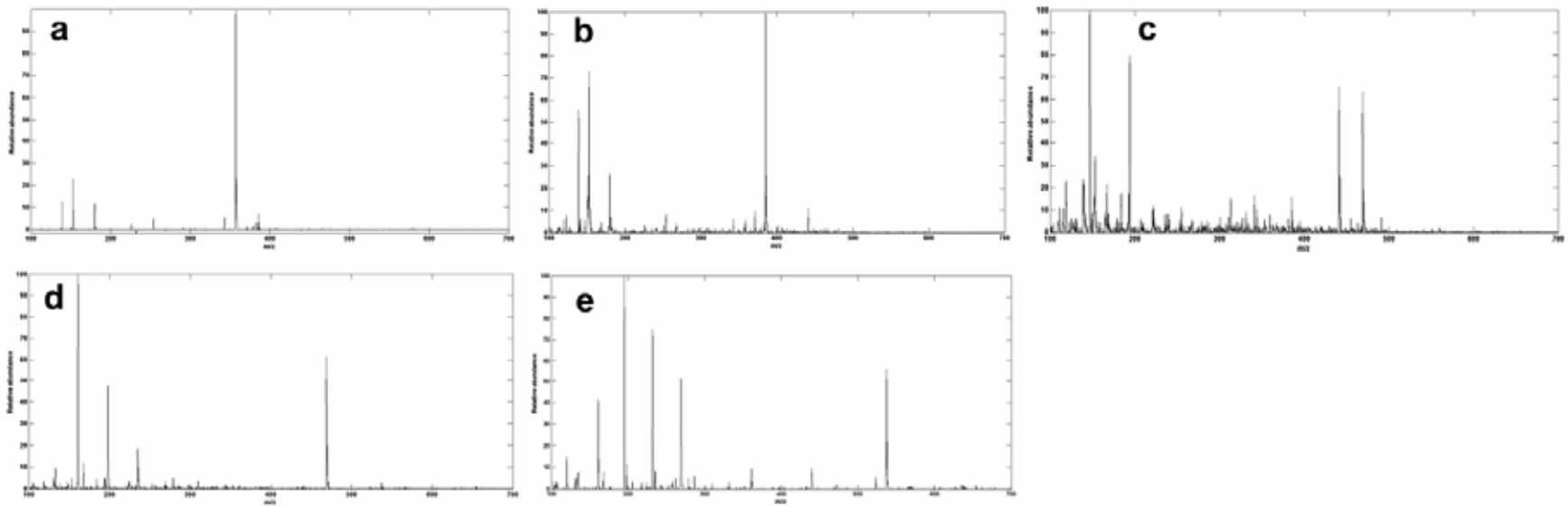
Mass spectra of five pure components.



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		$\mathbf{An_5}$	$\mathbf{An_6}$	$\mathbf{An_7}$	$\mathbf{An_8}$	$\mathbf{An_9}$
a	$\mathbf{An_5}$	**1**	0.1268	0.0456	0.0266	0.0075
	$\mathbf{An_6}$	0.1268	**1**	0.0321	0.0332	0.0379
	$\mathbf{An_7}$	0.0456	0.0321	**1**	0.0134	0.0030
	$\mathbf{An_8}$	0.0265	0.0332	0.0134	**1**	0.0029
	$\mathbf{An_9}$	0.0075	0.0379	0.0030	0.0029	**1**
b	$\mathbf{\hat{A}n_5}$	**0.9038**	0.0305	0.0044	0.0002	0.0120
	$\mathbf{\hat{A}n_6}$	0.3162	**0.8294**	0.1198	0.0325	0.0043
	$\mathbf{\hat{A}n_7}$	0.0959	0.2334	**0.7275**	0.2009	0.0038
	$\mathbf{\hat{A}n_8}$	0.0043	0.0038	0.0124	**0.9736**	0.0293
	$\mathbf{\hat{A}n_9}$	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, Ref. 83, 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: $$\mathbf{S}^{(k+1)} = \arg\min_{s_{mt}\geq 0} D_{\mathbf{S}}\left(\mathbf{X}\middle\|\mathbf{A}^{(k)}\mathbf{S}\right)_{\mathbf{S}^{(k)}}$$

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.

Nonnegative matrix factorization

Without additional constraints original Lee-Seung NMF algorithm does not yield unique solution. Generalization that involves sparseness constraints is given in:

$$D\left(\mathbf{X}\|\mathbf{AS}\right)=\frac{1}{2}\|\mathbf{X}-\mathbf{AS}\|_2^2+\alpha_{\mathbf{S}}J_{\mathbf{S}}(\mathbf{S})+\alpha_{\mathbf{A}}J_{\mathbf{A}}(\mathbf{A})$$

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. $\alpha_{\mathbf{S}}$ and $\alpha_{\mathbf{A}}$ are regularization terms. Gradient components in matrix form are

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

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

Nonnegative matrix factorization

By choosing learning rates proposed by Lee and Seung (they ensure nonnegativity)

$$\eta_{nm} = \frac{a_{nm}}{\left[\mathbf{A}\mathbf{S}\mathbf{S}^{\mathrm{T}}\right]_{nm}} \qquad \eta_{mt} = \frac{s_{mt}}{\left[\mathbf{A}^{\mathrm{T}}\mathbf{A}\mathbf{S}\right]_{mt}}$$

Multiplicative learning rules are obtained

$$a_{nm} \leftarrow a_{nm} \frac{\left[\left[\mathbf{X}\mathbf{S}^{\mathrm{T}}\right]_{nm} - \alpha_{\mathbf{A}} \dfrac{\partial J_{\mathbf{A}}(\mathbf{A})}{\partial a_{nm}}\right]_{+}}{\left[\mathbf{A}\mathbf{S}\mathbf{S}^{\mathrm{T}}\right]_{nm} + \varepsilon} \qquad s_{mt} \leftarrow s_{mt} \frac{\left[\left[\mathbf{A}^{\mathrm{T}}\mathbf{X}\right]_{mt} - \alpha_{\mathbf{S}} \dfrac{\partial J_{\mathbf{S}}(\mathbf{S})}{\partial s_{mt}}\right]_{+}}{\left[\mathbf{A}^{\mathrm{T}}\mathbf{A}\mathbf{S}\right]_{mt} + \varepsilon}$$

where $[x]_+ = \max\{\varepsilon, x\}$ with small ε. In a case of sparseness constraints derivatives in above expressions are equal to 1.

Nonnegative matrix factorization

NMF through minimization of Froebenius norm is optimal when data are corrupted by additive Gaussian noise. Squared Euclidean norm-based cost function is equivalent to maximization of likelihood:

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

Another cost function that is used most often for NMF is Kullback-Leibler divergence, also called I-divergence

$$D\left(\mathbf{X}\middle\|\mathbf{AS}\right)=\sum_{nt}\left(x_{nt}\ln\frac{x_{nt}}{\left[\mathbf{AS}\right]_{nt}}-x_{nt}+\left[\mathbf{AS}\right]_{nt}\right)$$

It can be shown that minimization of Kullback-Leibler divergence is equivalent to the maximization of the Poisson likelihood

$$L\left(\mathbf{X}\middle|\mathbf{A},\mathbf{S}\right)=\prod_{nt}\left(\frac{\left[\mathbf{AS}\right]_{nt}}{x_{nt}!}\exp\left(-\left[\mathbf{AS}\right]_{nt}\right)\right)$$

Nonnegative matrix factorization

Calculating gradients of I-divergence cost function w.r.t. a_{nm} and s_{mt} the following learning rules in MATLAB notation are obtained

$$\mathbf{S}^{(k+1)} = \left(\mathbf{S}^{(k)} \otimes \left(\mathbf{A}^{\mathrm{T}} \left(\mathbf{X} \oslash \left(\mathbf{A}\mathbf{S}^{(k)} \right) \right) \right)^{.[\omega]} \right)^{.[1+\alpha_{\mathbf{S}}]}$$

$$\mathbf{A}^{(k+1)} = \left(\mathbf{A}^{(k)} \otimes \left(\left(\mathbf{X} \oslash \left(\mathbf{A}^{(k)} \mathbf{S} \right) \right) \mathbf{S}^{\mathrm{T}} \right)^{.[\omega]} \right)^{.[1+\alpha_{\mathbf{A}}]}$$

where $\otimes$ denotes component-wise multiplication, and $\oslash$ denotes component-wise division. Relaxation parameter $\omega \in (0,2]$ provides improvement of the convergence, while $\alpha_{\mathbf{S}} \geq 0$ and $\alpha_{\mathbf{A}} \geq 0$ are sparseness constraints that are typically confined in the interval [0.001, 0.005].

Nonnegative matrix factorization

In order to obtain NMF algorithms optimal for different statistics of data and noise the α-divergence cost function can be used

$$D\left(\mathbf{X}\middle\|\mathbf{AS}\right)=\frac{1}{\alpha(\alpha-1)}\sum_{nt}\left(x_{nt}^{\alpha}\left[\mathbf{AS}\right]_{nt}^{1-\alpha}-\alpha x_{nt}+(\alpha-1)\left[\mathbf{AS}\right]_{nt}\right)$$

I-divergence is obtained in the limit when $\alpha\rightarrow 1$ and dual Kullback-Leibler divergence when $\alpha\rightarrow 0$. Using MATLAB notation the following update rules are obtained for $\alpha\neq 0,1$.

$$\mathbf{S}\leftarrow\left(\mathbf{S}.*\left(\mathbf{A}^{\mathrm{T}}*\left(\mathbf{X}./\left[\mathbf{AS}\right]_{+}\right)_{.}^{\alpha}\right)_{.}^{\omega/\alpha}\right)_{.}^{1+\alpha_{\mathbf{S}}}$$

$$\mathbf{A}\leftarrow\left(\mathbf{A}.*\left(\left(\mathbf{X}./\left[\mathbf{AS}\right]_{+}\right)_{.}^{\alpha}\mathbf{S}^{\mathrm{T}}\right)_{.}^{\omega/\alpha}\right)_{.}^{1+\alpha_{\mathbf{A}}}$$

$$\mathbf{A}\leftarrow\mathbf{A}*diag\left(1./sum(\mathbf{A},1)\right)$$

Hierarchical ALS NMF

Local or hierarchical ALS NMF algorithms were recently derived. They are biologically plausible and employ minimization of the global cost function to learn the mixing matrix and minimization of set of local cost functions to learn the sources. Global cost function can for example be squared Euclidean norm:

$$D\left(\mathbf{X}\|\mathbf{AS}\right)=\frac{1}{2}\left\|\mathbf{X}-\mathbf{AS}\right\|_2^2+\alpha_{\mathbf{S}}J_{\mathbf{S}}(\mathbf{S})+\alpha_{\mathbf{A}}J_{\mathbf{A}}(\mathbf{A})$$

Local cost functions can be also squared Euclidean norms

$$D^{(m)}\left(\mathbf{X}^{(m)}\|\mathbf{a}_m\underline{\mathbf{s}}_m\right)=\frac{1}{2}\left\|\mathbf{X}^{(m)}-\mathbf{a}_m\underline{\mathbf{s}}_m\right\|_2^2+\alpha_{\mathbf{s}}^{(m)}J_{\mathbf{S}}(\underline{\mathbf{s}}_m)+\alpha_{\mathbf{a}}^{(m)}J_{\mathbf{a}}(\mathbf{a}_m)\quad m=1,\ldots,M$$

$$\mathbf{X}^{(m)}=\mathbf{X}-\sum_{j\neq m}\mathbf{a}_j\underline{\mathbf{s}}_j$$

Hierarchical ALS NMF

Minimization of above cost functions in ALS manner with sparseness constraints imposed on **A** and/or **S** yields

$$\left\{ \underline{\mathbf{s}}_m \leftarrow \left[\mathbf{a}_m^{\mathrm{T}} \mathbf{X}^{(m)} - \alpha_{\mathbf{s}}^{(m)} \mathbf{1}_{1\times T} \right]_+ \right\}_{m=1}^{M}$$

$$\mathbf{A} \leftarrow \left[\left(\mathbf{X}\mathbf{S}^{\mathrm{T}} - \alpha_{\mathbf{A}} \mathbf{1}_{N\times M} \right) \left(\mathbf{S}\mathbf{S}^{\mathbf{T}} + \lambda \mathbf{I}_M \right)^{-1} \right]_+$$

$$\left\{ \mathbf{a}_m \leftarrow \mathbf{a}_m \Big/ \left\| \mathbf{a}_m \right\|_2 \right\}_{m=1}^{M}$$

where $\mathbf{I}_{1\times T}$ is an $M\times M$ identity matrix, $\mathbf{1}_{1\times T}$ and $\mathbf{1}_{N\times M}$ are row vector and matrix with all entries equal to one and $[\xi]_+ = \max\{\varepsilon, \xi\}$ (e.g., $\varepsilon = 10^{-16}$).

Regularization constant λ changes as a function of the iteration index as $\lambda_k = \lambda_0 \exp(-k/\tau)$ (with λ_0 = 100 and τ = 0.02 in the experiments).

Multilayer NMF

Significant improvement in the performance of the NMF algorithms is obtained when they are applied in the multilayer mode, whereas sequential decomposition of the nonnegative matrices is performed as follows.

In the first layer, the basic approximation decomposition is performed:

$$\mathbf{X} \cong \mathbf{A}^{(1)}\mathbf{S}^{(1)} \in \mathbb{R}_{0+}^{N\times T}$$

In the second layer result from the first layer is used to build up new input data matrix for the second layer $\mathbf{X} \leftarrow \mathbf{S}^{(1)} \in \mathbb{R}_{0+}^{M\times T}$. This yields $\mathbf{X}^{(1)} \cong \mathbf{A}^{(2)}\mathbf{S}^{(2)} \in \mathbb{R}_{0+}^{M\times T}$.

After *L* layers data decomposes as follows $\mathbf{X} \cong \mathbf{A}^{(1)}\mathbf{A}^{(2)}\cdots\mathbf{A}^{(L)}\mathbf{S}^{(L)}$

Multi-start initialization for NMF algorithms

Combined optimization of the cost function $D(\mathbf{X}||\mathbf{AS})$ with respect to $\mathbf{A}$ and $\mathbf{S}$ is nonconvex optimization problem. Hence, some strategy is necessary to decrease probability that optimization process will get stuck in some local minima. Such procedure is outlined with the following pseudo code: Select R-number of restarts, K_i number of alternating steps, K_f number of final alternating steps.

for $r = 1,\ldots,R$ **do**

Initialize randomly $\mathbf{A}^{(0)}$ and $\mathbf{S}^{(0)}$

$\{\mathbf{A}^{(r)},\mathbf{S}^{(r)}\} \leftarrow$ nmf_algorithm$(\mathbf{X},\mathbf{A}^{(0)},\mathbf{S}^{(0)},K_i)$;

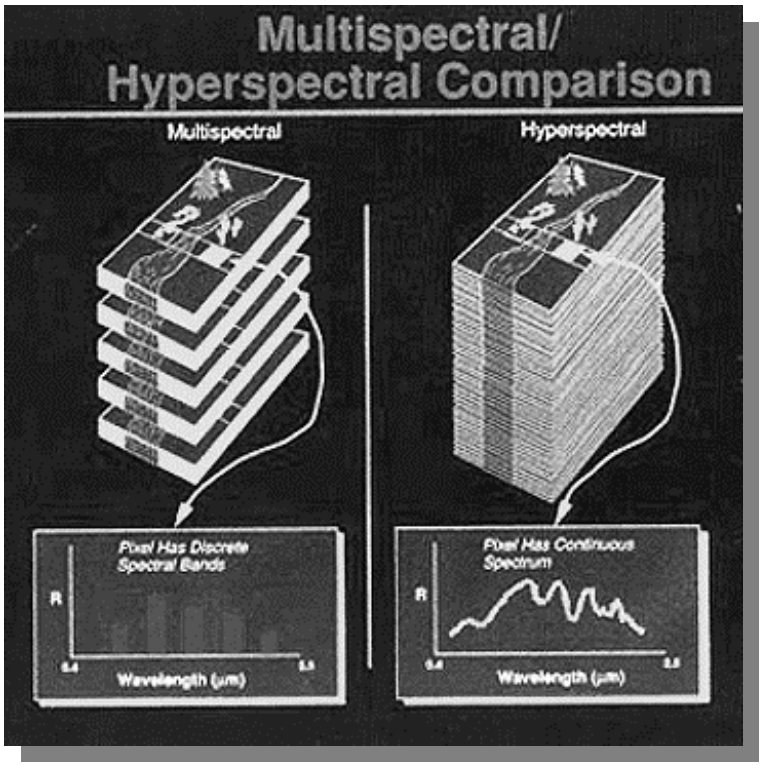
compute $d=D(\mathbf{X}||\mathbf{A}^{(r)}\mathbf{S}^{(r)})$;

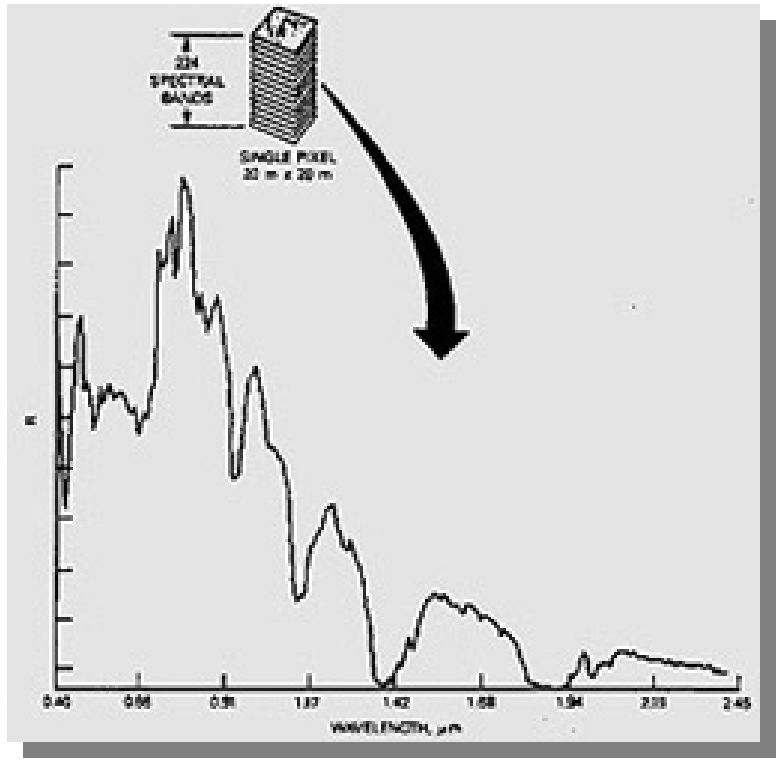
end

$r_{min}=\text{argmin}_{1\le n\le R} d_r$;

$\{\mathbf{A},\mathbf{S}\} \leftarrow$nmf_algorithm$(\mathbf{X},\mathbf{A}^{(rmin)},\mathbf{S}^{(rmin)},K_f)$;

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:

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

$$\left[\mathbf{a}_1 \quad \mathbf{a}_2 \quad \ldots \quad \mathbf{a}_M\right] \equiv \mathbf{A}$$

$$\left[\mathbf{s}_1 \quad \mathbf{s}_2 \quad \ldots \quad \mathbf{s}_M\right]^T \equiv \mathbf{S}$$

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

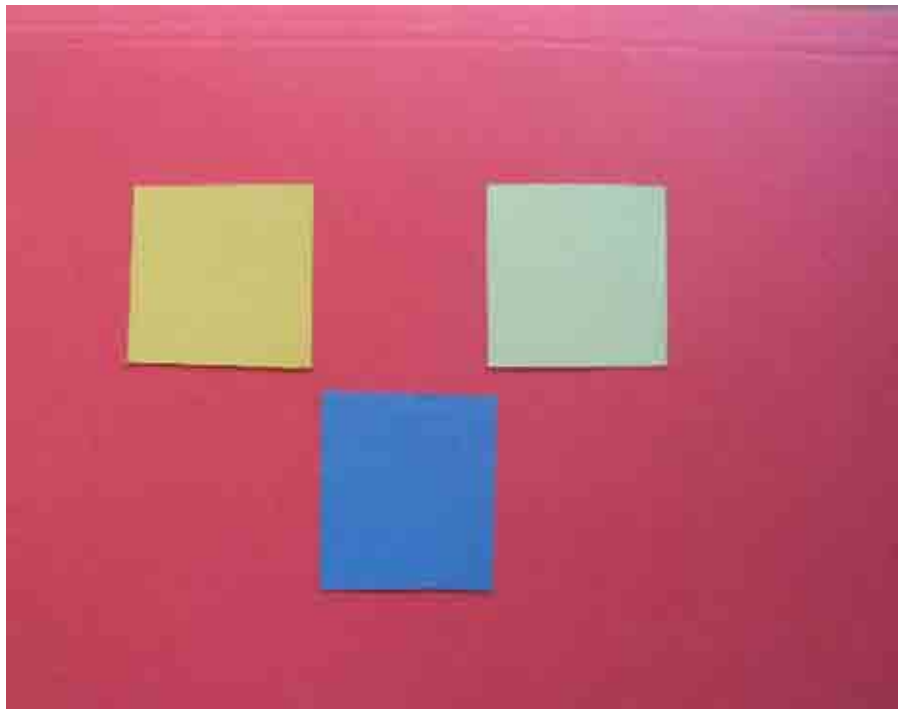
Spectral similarity between the sources s_m and s_n implies that corresponding column vectors are close to collinear i.e. $\mathbf{a}_m \cong c\mathbf{a}_n$.

Contribution at certain pixel location t is: $\mathbf{a}_m s_{mt} + \mathbf{a}_n s_{nt} \simeq c\mathbf{a}_n s_{mt} + \mathbf{a}_n s_{nt}$. 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. $s_m(t)*s_n(t)\approx\delta_{nm.}$. 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

Assuming unit L_2-norm of $\boldsymbol{a}_m$ we can parameterize column vectors in 3D space by means of azimuth and elevation angles

$$\mathbf{a}_m = [\cos(\varphi_m)\sin(\theta_m) \;\; \sin(\varphi_m)\sin(\theta_m) \;\; \cos(\theta_m)]^{\mathrm{T}}$$

Due to nonnegativity constraints both angles are confined in $[0,\pi/2]$. Now estimation of **A** and M is obtained by means of data clustering algorithm:

- We remove all data points close to the origin for which applies: $\left\{\left|\mathbf{x}(t)\right|_2 \le \varepsilon\right\}_{t=1}^{T}$ where ε represents some predefined threshold.

- Normalize to unit L_2-norm remaining data points $\mathbf{x}(t)$, i.e., $\left\{\mathbf{x}(t) \to \mathbf{x}(t) \big/ \left|\mathbf{x}(t)\right|_2\right\}_{t=1}^{\bar{T}}$

Unsupervised segmentation of multispectral images

• Calculate function $f(\mathbf{a})$:

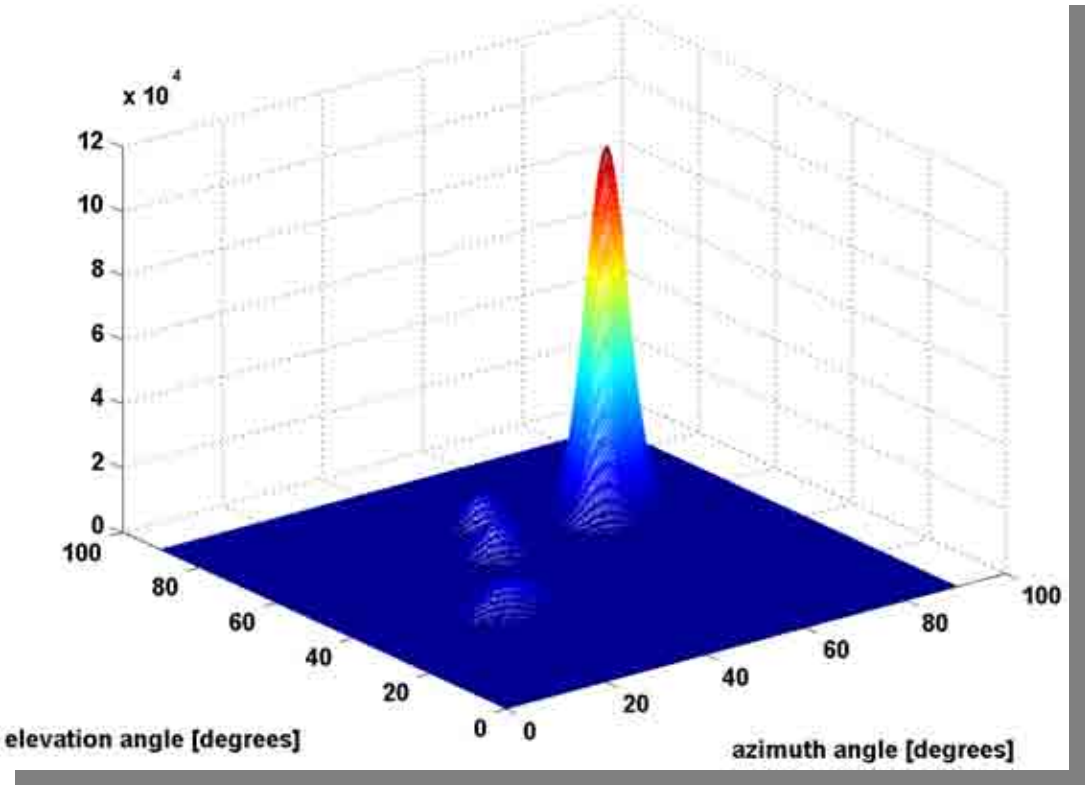
$$f(\mathbf{a}) = \sum_{t=1}^{\bar{T}} \exp\left(-\frac{d^2(\mathbf{x}(t), \mathbf{a})}{2\sigma^2} \right)$$

where $d(\mathbf{x}(t), \mathbf{a}) = \sqrt{1 - (\mathbf{x}(t) \cdot \mathbf{a})^2}$ and $(\mathbf{x}(t) \cdot \mathbf{a})$ denotes inner product. Parameter σ is called dispersion. If set to sufficiently small value, in our experiments this turned out to be $\sigma \approx 0.05$, the value of the function $f(\mathbf{a})$ will approximately equal the number of data points close to $\mathbf{a}$. Thus by varying mixing angles $0 \leq \varphi, \theta \leq \pi/2$ we effectively cluster data.

• Number of peaks of the function $f(\mathbf{a})$ corresponds with the estimated number of materials M. Locations of the peaks correspond with the estimates of the mixing angles $\left\{ \left(\hat{\varphi}_m, \hat{\theta}_m \right) \right\}_{m=1}^{\hat{M}}$, i.e., mixing vectors $\left\{ \hat{\mathbf{a}}_m \right\}_{m=1}^{\hat{M}}$.

Unsupervised segmentation of multispectral images

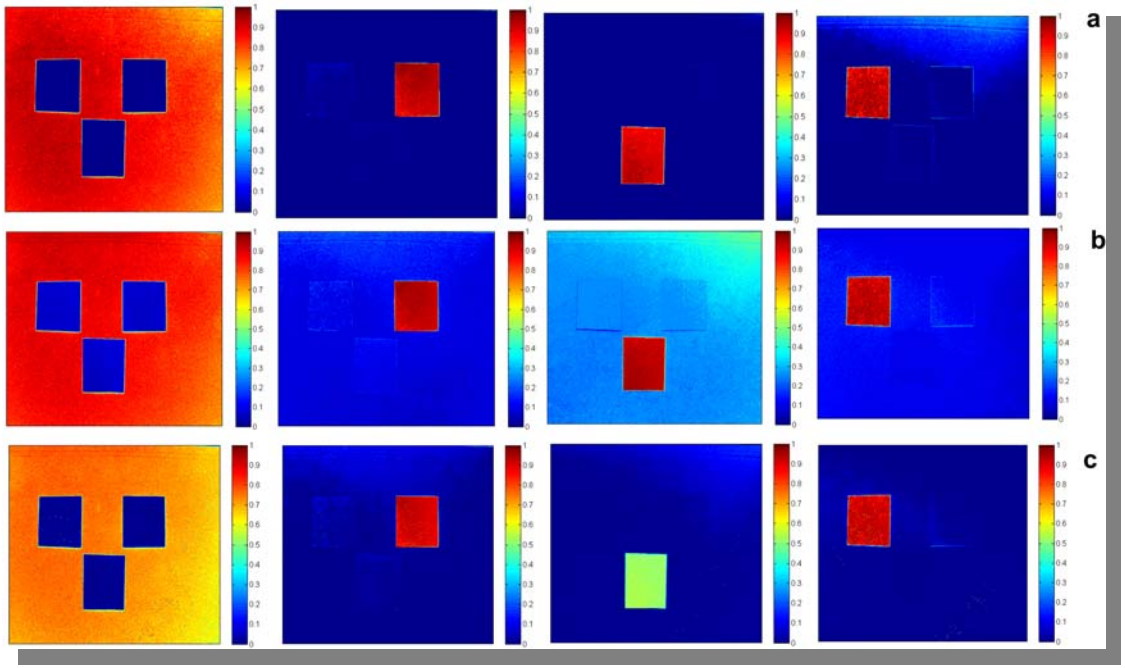
For shown experimental RGB image clustering function is obtained as:



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:

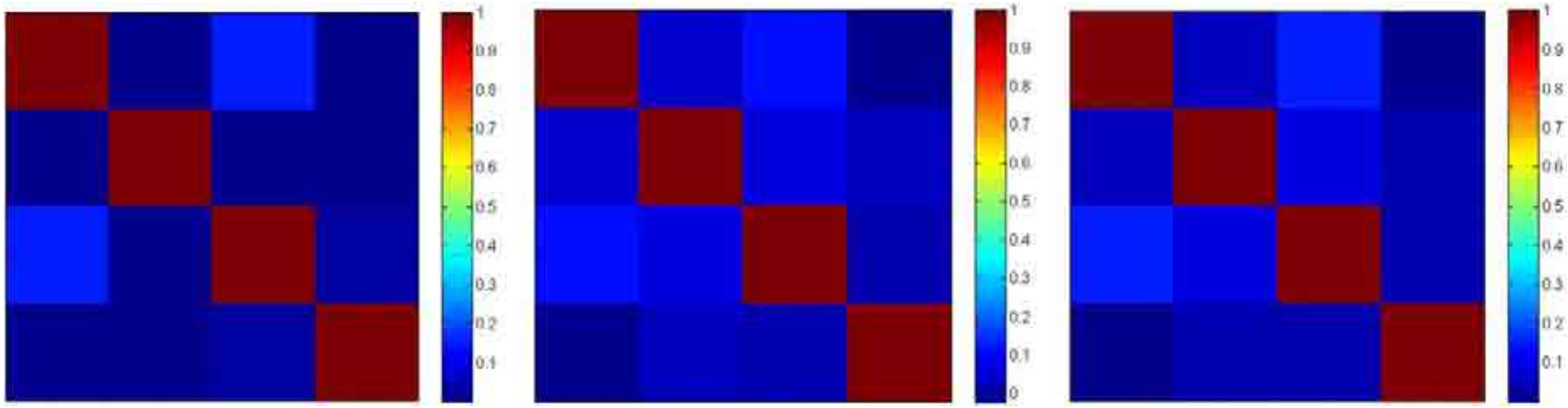


a) 25 layers HALS NMF; b) Interior point method; c) Linear programming.

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.