Lecture III

Independent component analysis (ICA) for linear static problems: information-theoretic approaches

Ivica Kopriva

Ruđer Bošković Institute

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

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)

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. <u>Independent component analysis (ICA)</u> based separation of two uniformly distributed signals (scatter plots of true sources, mixtures and estimated sources). <u>ICA based separation of two images</u> (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 ($0<p\leq1$) 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 ($0<p\leq1$) 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.**

Information theoretic ICA

Maximization of nongaussianity. Due to the central limit theorem mixed signals **x** are close to Gaussian distribution. Because in ICA source signals **s** are non-Gaussian by assumption any method that maximizes distance from Gaussianity should lead to the separation of source signals.

The measure of the non-Gaussianity is normalized FO cumulant called kurtosis. Thus, method that maximizes contrast function based on kurtosis criterion is an ICA algorithm[1]

$$\mathbf{W} = \operatorname{argmax} \Phi(\mathbf{W}) = \sum_{i=1}^{N} C_4^2(y_i)$$

Kurtosis approximates (small deviation from non-Gaussianity) negentropy of standardized (zero mean and unit variance) non-Gaussian symmetrically distributed random process. Disadvantage of the kurtosis based ICA algorithms is high sensitivity on outliers.

[1]P. Common, “Independent Component Analysis – a new concept ?”, Signal Processing,36(3):287-314.

Information theoretic ICA

ICA by maximization of absolute value of kurtosis. Non-gaussianity measure based on absolute value of the kurtosis is equivalent to approximation of negentropy based on an assumption of small deviation from Gaussianity for symmetrically distributed random process with zero mean and unit variance. Thus, it is assumed that measured data $\mathbf{x}$ are whitened i.e. $\mathbf{z}=\mathbf{V}\mathbf{x}$ where $\mathbf{V}$ represents whitening transform.

We want to find direction vector $\mathbf{w}$ such that absolute value of the kurtosis of $y=\mathbf{w}^T\mathbf{z}$ be maximal. Kurtosis of y can be approximated as $\kappa(y)=E\{y^4\}$. Gradient of $\kappa(y)$ is obtained as

$$\Delta\mathbf{w}\propto 4\,\mathrm{sign}(\mathbf{w}^T\mathbf{z})E\{\mathbf{z}(\mathbf{w}^T\mathbf{z})^3\}.$$

Since we are interesting in direction in which y becomes maximally non-Gaussian (and want to keep y with unit variance) we obtain absolute value of the kurtosis based ICA algorithm that maximizes departure from normality (Gaussian distribution) as:

$$\mathbf{w} \leftarrow \Delta\mathbf{w}/\|\Delta\mathbf{w}\|^2$$

This algorithm is: sensitive to outliers (kurtosis as a contrast function), sensitive to choice of initial value for $\mathbf{w}$ and leads to locally optimal solution.

Information theoretic ICA

ICA by maximization of negentropy – the FastICA algorithm[2,3].

$$\mathbf{w} = \arg\max J(\mathbf{y}) = H(\mathbf{y}_{gauss}) - H(\mathbf{y})$$

Because negentropy is difficult to estimate it has to be approximated. Assuming small deviation from Gaussianity as well random process with zero mean and unit variance the approximation is obtained as

$$J(y) \propto \frac{1}{12} E\{y^3\}^2 + \frac{1}{48} \kappa(y)^2$$

Above approximation is based on third and fourth order polynomial nonlinearities. It does not characterize non-Gaussian distributions rich enough: only third and fourth order moments are generated in the approximation (that is because *J*(y) has been derived assuming small deviation from Gaussianity).

[2] A. Hyvärinen and E. Oja, "A fast fixed-point algorithm for independent component analysis," Neural Computation, vol. 9, pp. 1483-1492, 1997.

[3] A. Hyvärinen, J. Karhunen, E. Oja, *Independent Component Analysis*, John Wiley, 2001.

Information theoretic ICA

However, it is possible to use some non-quadratic function G(y) with the Taylor expansion that contains many higher order terms in which case $E\{G(\mathrm{y})\}$ will generate many higher order moments and, thus, characterize better non-Gaussian nature of random process y. Thus, we can use the following approximation for negentropy of y:

$$J(y) \propto \left[\left\langle \{G(y)\} \right\rangle - \left\langle \{G(\nu)\} \right\rangle \right]^2$$

where ν is Gaussian process with zero mean and unit variance. The following choice for G has proved very useful[3]:

$$G(y) = \frac{1}{a} \log\cosh(ay)$$

Assuming unit variance of y: $E\{(\mathbf{w}^T\mathbf{z})^2\}=\|\mathbf{w}\|^2=1$ the gradient w.r.t. **w** is obtained as: $\Delta\mathbf{w}\propto\gamma\, E\{\mathbf{z}g(\mathbf{w}^T\mathbf{z})\}$. $\gamma = E\{G(\mathbf{w}^T\mathbf{z})\} - E\{G(\nu)\}$ and $g=\partial G(\mathrm{y})/\partial \mathrm{y}$. For above choice of G it follows that $g(\mathrm{y})=\tanh(a\mathrm{y})$. That choice matches well super-Gaussian processes. Cubic function $g(\mathrm{y})=y^3$ is a good choice for sub-Gaussian processes.

Information theoretic ICA

Convergence of gradient based method is slow. Instead, Lagrangian can be formulated with explicit constraint on $\|\mathbf{w}\|^2=1$:

$$L(\mathbf{w}^{\mathrm{T}}\mathbf{z}) = J(y) + \lambda(\|\mathbf{w}\|^2 - 1)$$

where λ represents Lagrange multiplier. Derivation of Lagrangian yields:

$$E\left\{\mathbf{z}g(\mathbf{w}^{\mathrm{T}}\mathbf{z})\right\} + \lambda\mathbf{w} = \mathbf{0}$$

Approximate Newton method is used to solve above equation for **w** and obtain fixed point update:

$$\mathbf{w} \leftarrow \left\langle \mathbf{z}g\left(\mathbf{w}^{\mathrm{T}}\mathbf{z}\right)\right\rangle - \left\langle \mathbf{z}g'\left(\mathbf{w}^{\mathrm{T}}\mathbf{z}\right)\right\rangle \mathbf{w}$$

$$\mathbf{w} \leftarrow \mathbf{w} / \|\mathbf{w}\|$$

where $g'(y)=dg(y)/dy$. The FastICA algorithm is sequential algorithm i.e. it separates source signals one by one. This makes it very useful for application on high dimensional problems such as hyperspectral images or fMRI data sets that could have few hundreds of sensors. The fact that FastICA algorithm is based on the negentropy approximation makes it less accurate than some other ICA methods that will be introduced soon. MATLAB code for FastICA method can be downloaded from: http://www.cis.hut.fi/projects/ica/fastica/ .

Information theoretic ICA

Because mutual information I(y_1,y_2,..,y_N) between components of **y=Wx** represents direct measure of statistical (in)dependence it is wise to develop ICA algorithms directly form minimizing I(y_1,y_2,..,y_N).[4] Using definition for mutual information and relations for entropy of the line transformation we can write:

$$I(y_1, y_2, .., y_N) = \sum_{i=1}^{N} H(y_i) - H(\mathbf{x}) - \log\left|\det \mathbf{W}\right|$$

which is equivalent to the contrast functions for ML and *Infomax* ICA algorithms. If logarithm of the probability density function is approximated with $G_i(y_i)=log p_i(y_i)$ above expression becomes:

$$I(y_1, y_2, .., y_N) = \sum_{i=1}^{N} E(G_i(y_i)) - \log\left|\det \mathbf{W}\right| - H(\mathbf{x})$$

which is equivalent to contrast function optimized by FastICA algorithm. So mutual information could be seen as a framework for rigorous justification of the ML, Infomax and FastICA ICA algorithms.

[4]D. Erdogmus, K. E. Hild II, Y. N. Rao and J.C. Principe, “Minimax Mutual Information Approach for Independent Component Analysis,” Neural Computation, vol. 16, No. 6, pp. 1235-1252, June, 2004.

Information theoretic ICA

ICA by maximum likelihood (ML)[5]. Likelihood of the noise free ICA model **x=As** follows from Bayes rule and maximum *a* posteriori (MAP) probability:

$$p(\mathbf{A},\mathbf{s}|\mathbf{x}) = \frac{p\left(\mathbf{x}|\mathbf{A},\mathbf{s}\right)p(\mathbf{A})\,p(\mathbf{s})}{p(\mathbf{x})}$$

$p(\mathbf{x})$ is a constant and does not influence MAP probability. Uninformative *prior* on **A** is assumed (uniform distribution): $p(\mathbf{A})$ can be considered a constant as well. Thus above expression simplifies to:

$$p(\mathbf{s}|\mathbf{x}) \propto p\left(\mathbf{x}|\mathbf{A},\mathbf{s}\right)p(\mathbf{s})$$

Thus, MAP approach is equivalent to maximizing likelihood $p(\mathbf{x}|\mathbf{A},\mathbf{s})$ under some *prior* on **s**: $p(\mathbf{s})$. Assuming linear mixture model **x**=**As** it follows $p\,(\mathbf{x}|\mathbf{A},\mathbf{s})=p(\mathbf{s})/|\det(\mathbf{A})|=\det(\mathbf{W})|p(\mathbf{s})$.

[5] 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

Assuming statistically independent sources likelihood of the noise free ICA model **x=As** is finally obtained 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 $\mathbf{W}=[\mathbf{w}_1\ \mathbf{w}_2\ \ldots\ \mathbf{w}_N]=\mathbf{A}^{-1}$. 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:

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

[5] 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:

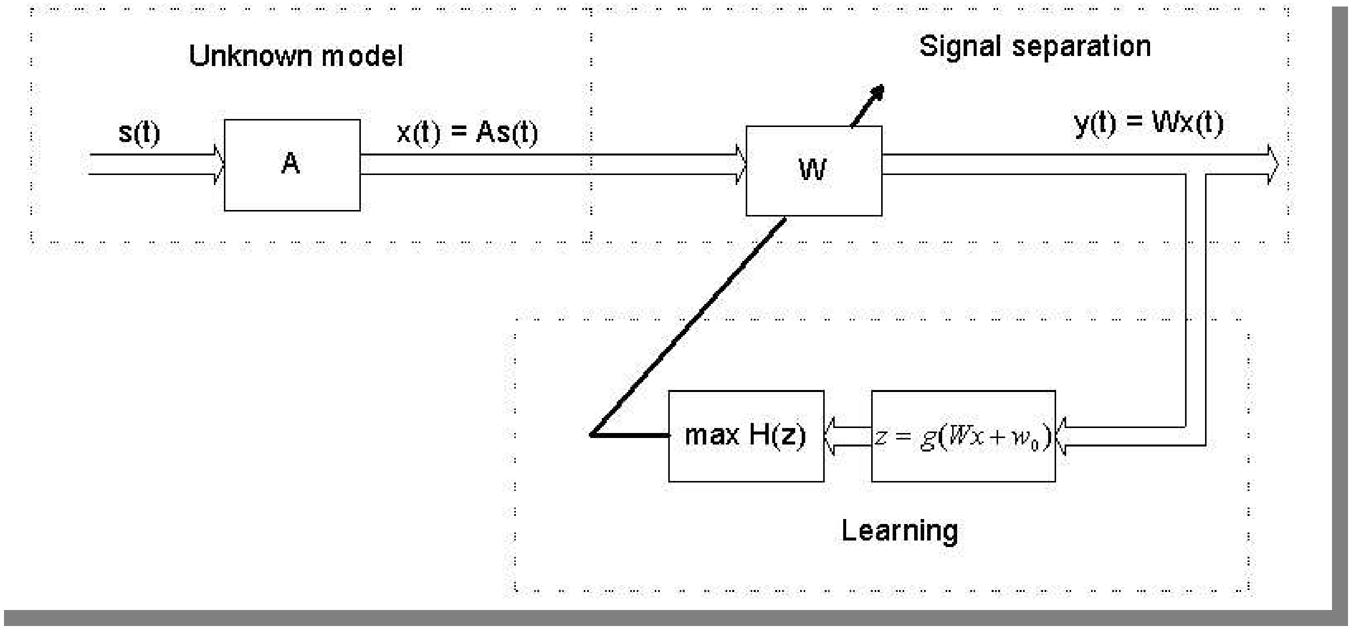
$$\mathbf{W}(k+1) = \mathbf{W}(k) + \eta\left[\mathbf{I} - E\left\{\varphi(\mathbf{y})\mathbf{y}^{\mathrm{T}}\right\}\right]\mathbf{W}(k)$$

ML adaptive ICA algorithm is obtained by dropping expectation:

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

Information theoretic ICA

ICA by information maximization[6]. This Bell-Sejnowski method is formulated by maximizing information transfer through nonlinear network:



[6]A. J. Bell and T. J. Sejnowski, “An information-maximization approach to blind separation and blind deconvolution,” *Neural Comp.* 7, 1129-1159, 1995.

Information theoretic ICA

Mutual information I(**z;x**) is formulated as a difference between joint and conditional entropies:

$$\mathrm{I}(\mathbf{z};\mathbf{x}) = \mathrm{H}(\mathbf{z}) \quad - \quad \mathrm{H}(\mathbf{z}/\mathbf{x})$$

Because there is no uncertainty about **z** due to presence of some other source H(**z**/**x**)=-∞. Therefore:

$$\Delta\mathbf{W} = \arg\max I(\mathbf{z};\mathbf{x}) \approx \arg\max H(\mathbf{z})$$

We formulate mutual information between components of **z** using Kullback-Leibler divergence

$$\mathrm{MI}(\mathbf{z}) = \mathrm{D}\left(p(\mathbf{z}), \prod_{i=1}^{N} p_i(z_i) \right) = \int p(\mathbf{z}) \log \frac{p(\mathbf{z})}{\prod_{i=1}^{N} p_i(z_i)} \mathbf{dz}$$

and mutual information I(**z;x**) becomes:

$$\mathrm{I}(\mathbf{z};\mathbf{x}) = \sum_{\mathrm{i}=1}^{\mathrm{N}} \mathrm{H}_i(\mathrm{z}_\mathrm{i}) \quad - \quad \mathrm{MI}(\mathbf{z})$$

Information theoretic ICA

It follows:

$$\Delta\mathbf{W} = \arg\max I(\mathbf{z};\mathbf{x}) \approx \arg\max H(\mathbf{z}) \approx \min MI(\mathbf{z})$$

Because KL distance is convex nonnegative function equal to zero only when

$$p(\mathbf{z}) = \prod_{i=1}^{N} p(z_i)$$

it follows that maximizing H($\mathbf{z}$) will make component of $\mathbf{z}$ mutually independent. Because $z_i = g_i(y_i)$ and $g_i()$ is invertible this will also lead to mutually independent components of $\mathbf{y}$.

$$\Delta\mathbf{W} = \frac{\partial H(\mathbf{z})}{\partial\mathbf{W}} = \left[\left(\mathbf{W}^T\right)^{-1} - E\left\{\varphi(\mathbf{y})\mathbf{x}^T\right\}\right]$$

After correcting Eucledian gradient $\Delta\mathbf{W}$ by metric tensor $\mathbf{W}^{\mathrm{T}}\mathbf{W}$ the batch *Infomax* ICA learning rule is obtained:

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

Information theoretic ICA

and on-line (adaptive) version is obtained by dropping expectation term:

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

Infomax and ML ICA learning rules are equivalent. The central problem is that optimal value of $\varphi(\mathbf{y})$ requires knowledge of the probability density of the source signals:

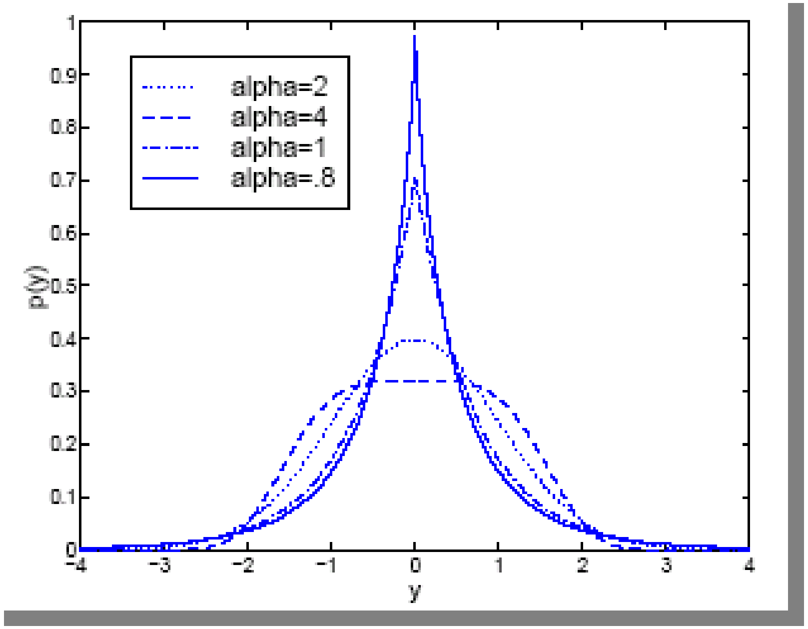
$$\varphi_i=-\frac{1}{p_i}\frac{dp_i}{dy_i}$$

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

Information theoretic ICA

Flexible nonlinearity[7,8] 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.

[7]S. Choi, A. Cihcocki, S. Amari, "Flexible Independent Component Analysis," Journal VLSI, KAP, 2000.
[8]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 α_i=1 because speech and music are super-Gaussian signals. If source signals are various communication signals α_i can be set to α_i=2.5 or α_i=3 because communication signals are sub-Gaussian signals.

Alternative way is to estimate α_i adaptively from data[8].

Information theoretic ICA

In [9] an approach to the estimation of the score functions from data is proposed. It is based on the estimation of the probability density function from data using Gaussian kernel estimator.

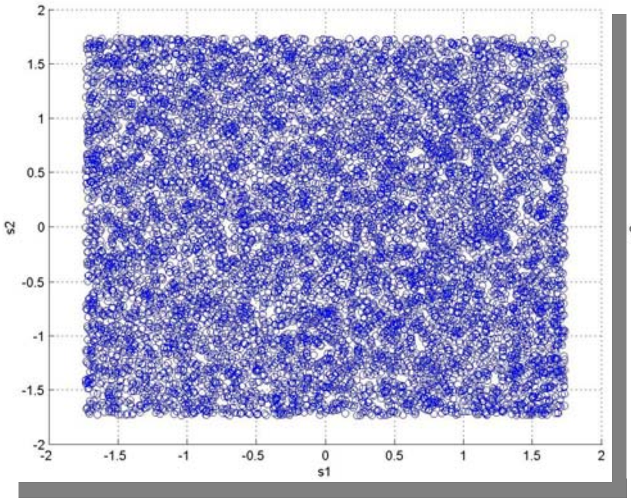
$$\hat{p}_i(y_i(t), \mathbf{y}_i) = \frac{1}{T}\sum_{tt=1}^{T} G\left(y_i(t) - y_i(tt), \sigma^2\mathbf{I}\right)$$

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

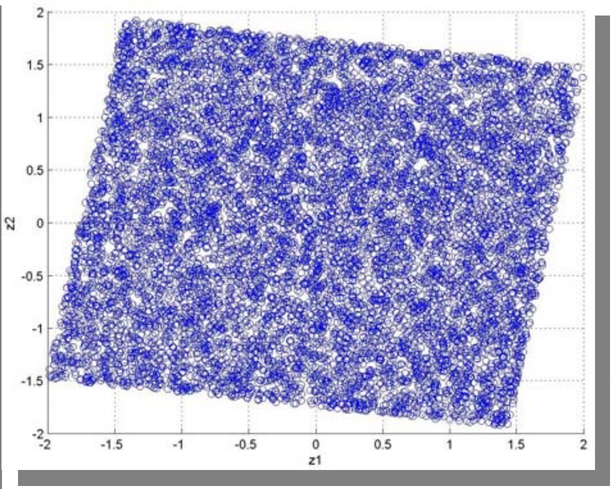
$$G\left(y_i(t), \sigma^2\mathbf{I}\right) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{y_i^2(t)}{2\sigma^2}\right)$$

[9]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. .

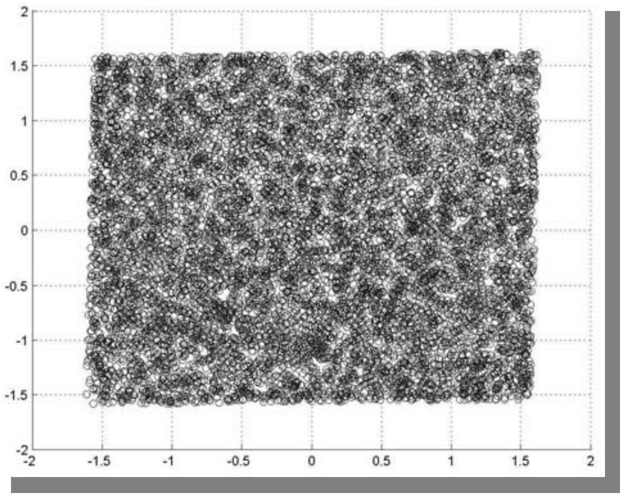
Scatter diagrams of PCA and ICA extracted signals



Source signals



PCA extracted signalsi



ICA extracted signals (min *MI*(**y**)).

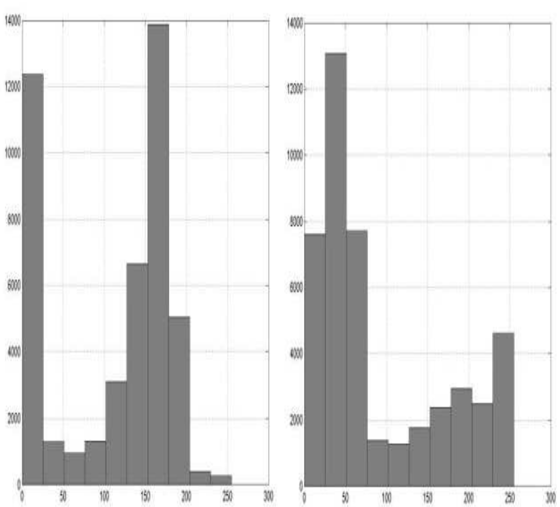
Homework: - reproduce these results by the ICA algorithm of choice.

PCA

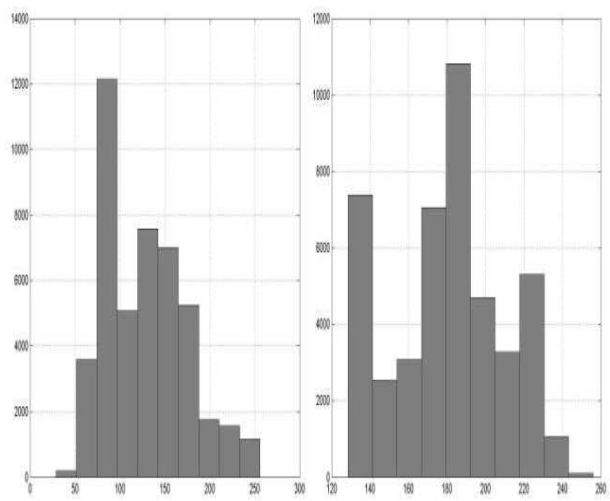
ICA (min $MI(\mathbf{y})$).

Homework: - reproduce these results by the ICA algorithm of choice.

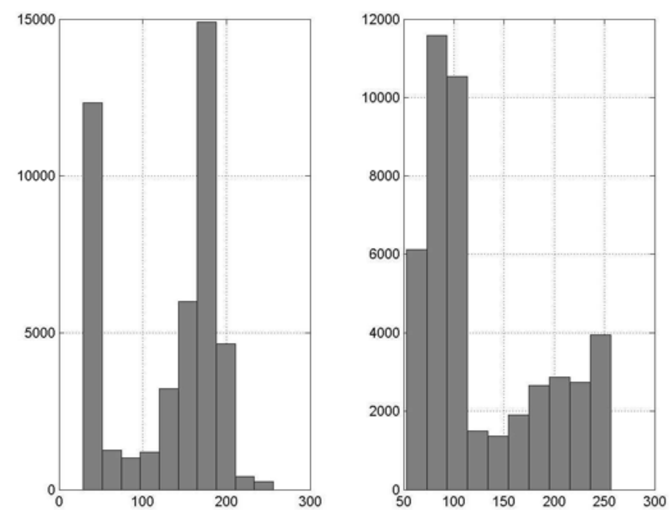
Histograms of the PCA and ICA extracted images



Source images



PCA extracted images



ICA extracted images

Applications of memoryless ICA

- Communications (Multiple Access Interference cancellation in DS-CDMA systems)
- Multispectral and hyperspectral remote sensing
- fMRI Imaging
- Image sharpening in the presence of atmospheric turbulence

ICA in Communications

- Signals are man-made with known properties. Transmission in short bursts allows use of batch mode algorithms.

- ICA(BSS) is similar to the methods which rely on signal properties only: constant modulus, finite alphabet, cyclostationarity.

- BSS relies on non-Gaussianity and statistical independence.

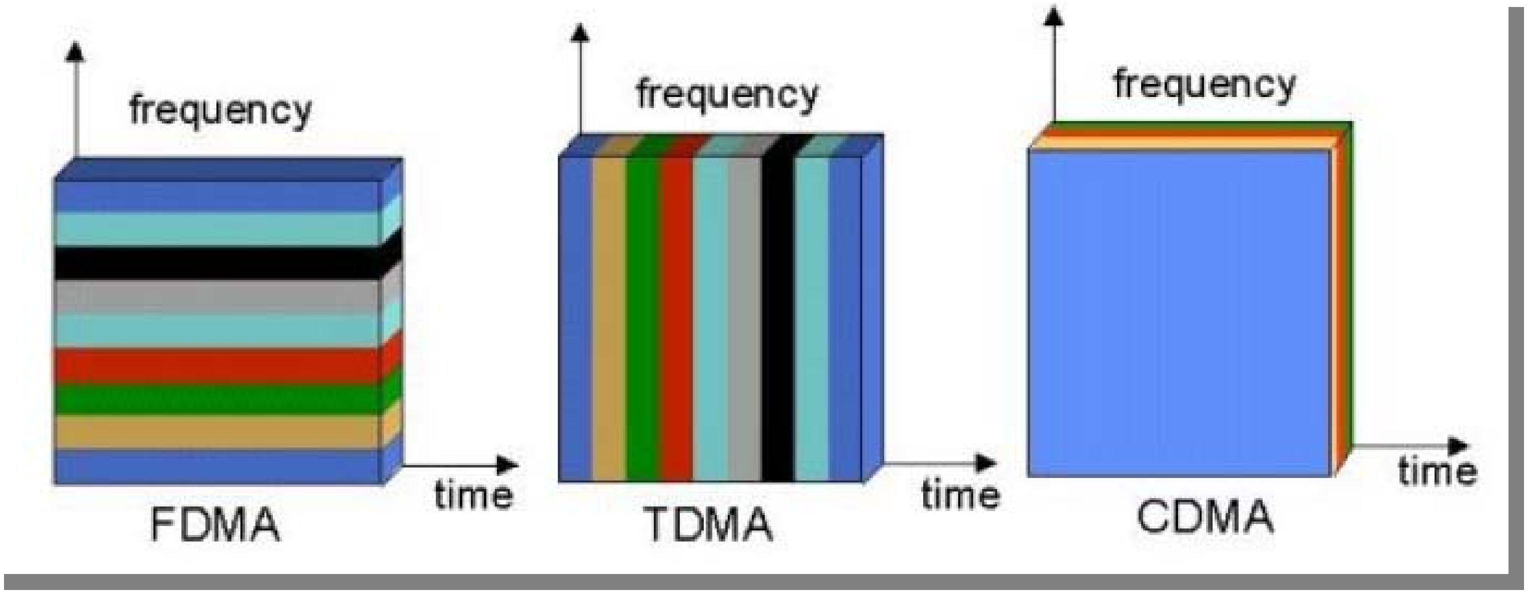
- BSS is enabling technology to increase the capacity of the cellular system in the uplink.

- BSS could be a basis for interference rejection.

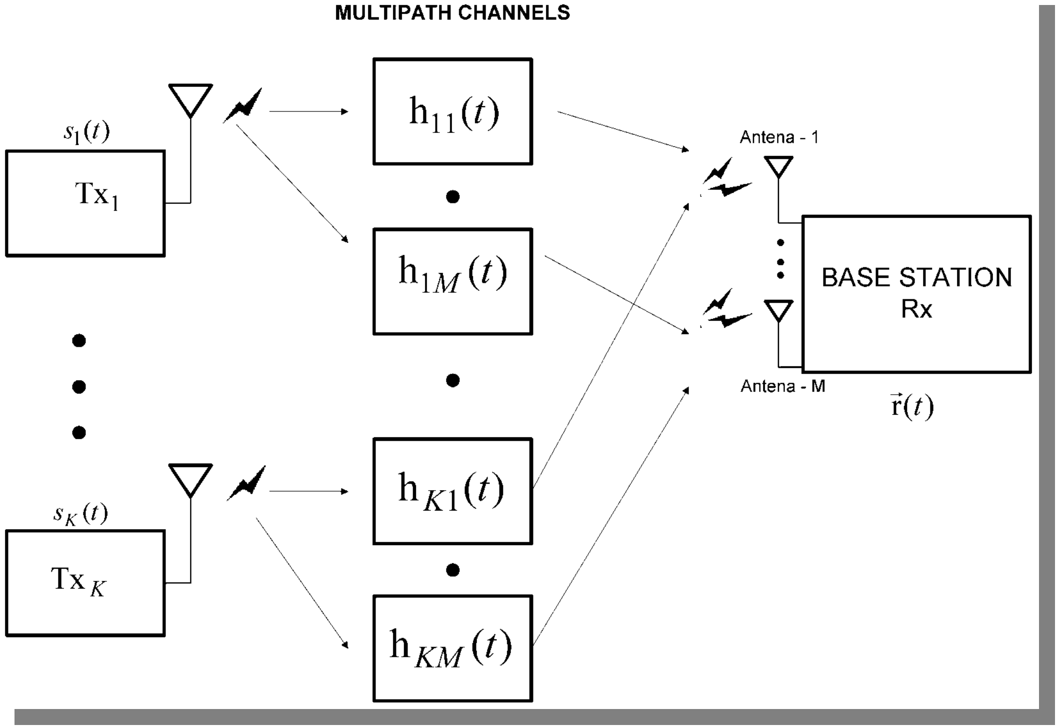
Interference suppression in Direct Sequence (DS) Code Division Multiple Access (CDMA) Systems

http://www.cis.hut.fi/karthik/tech.shtml

Unlike in FDMA and TDMA multiuser schemes in CDMA there is no disjoint division in frequency or time spaces. Each user occupies the same frequency band simultaneously.



Interference suppression in Direct Sequence (DS) Code Division Multiple Access (CDMA) Uplink Systems[10,11]



[10]T. Ristaniemi, J. Joutensalo, ‘Advanced ICA-based receivers for block fading DS-CDMA channels’, Signal. Processing, vol. 85, no.3, 2002, 417-431.
[11]A. Hyvarinen, J. Karhunen and E. Oja, *Chapter 23* in *Independent Component Analysis*, J. Wiley, 2001.

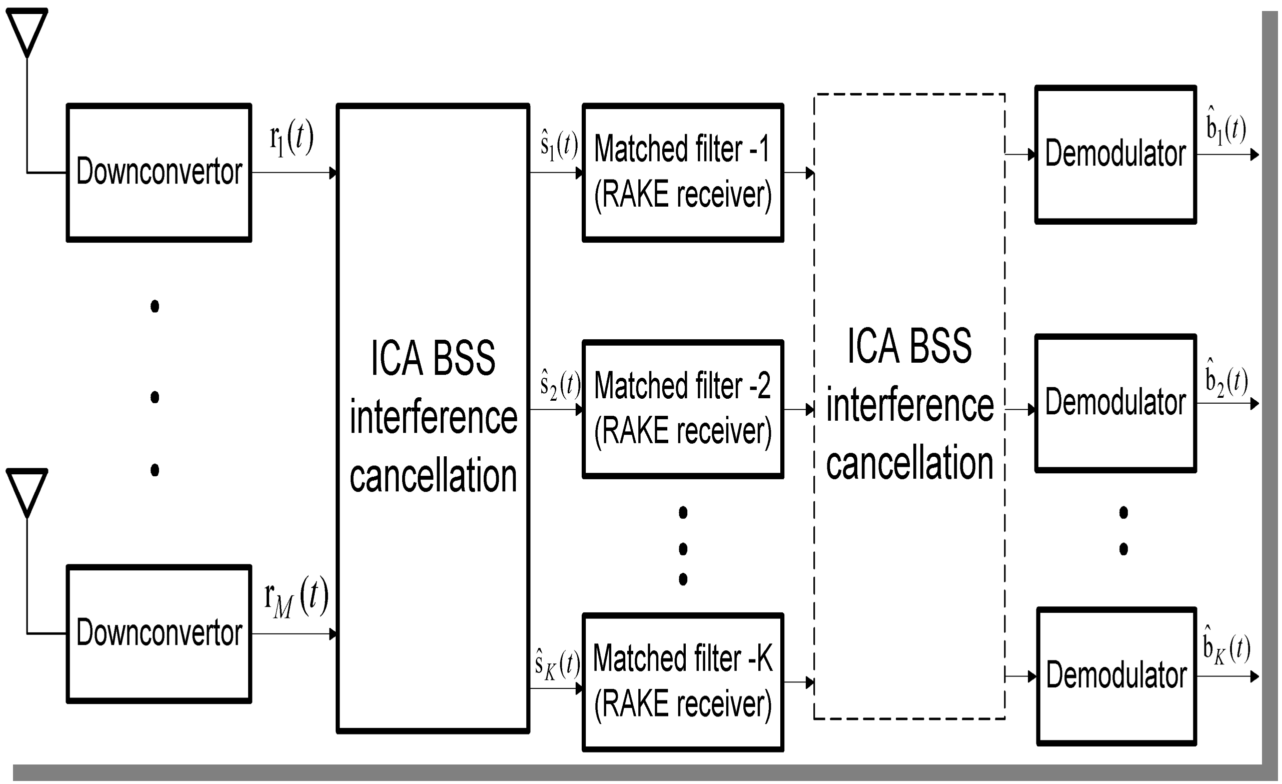
Interference suppression in block DS-CDMA uplink systems[10,11]

$$r_m(t) = \sum_{l=1}^{L} h_{ml} \sum_{n=1}^{N} \sum_{k=1}^{K} b_{nk} a_k \left(t - nT - d_l\right) + v(t)$$

b_{nk} - n^{th} data symbol of the k-th user
d_l - path delays
a_k - k^{th} user binary chip sequence
h_{kl} - channel impulse response coefficients
$v(t)$ - AWGN
m=1,..M (number of antennas)
K – number of users
N – number of symbols
L- number of paths

Interference suppression in uplink CDMA systems

- ICA can be used as additional element in RAKE or MMSE receivers to additionally increase interference suppression capability.



- Combining ICA with RAKE or MMSE receivers alleviates permutation indeterminacy problems because users are identified by their own codes.

Interference suppression in block DS-CDMA downlink systems

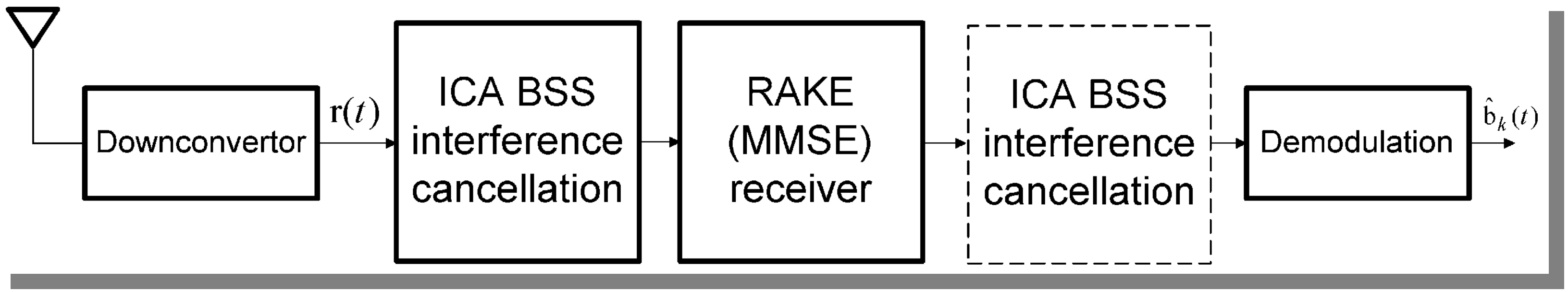
$$r(t) = \sum_{n=1}^{N}\sum_{k=1}^{K} b_{nk} \sum_{l=1}^{L} h_{kl} a_k \left(t - nT - d_l\right) + v(t)$$

b_{nk} - n^{th} data symbol of the k-th user
d_l - path delays
a_k - k^{th} user binary chip sequence
h_{kl} - channel impulse response coefficients
$v(t)$ - AWGN
K – number of users
N – number of symbols
L- number of paths

With chip-rate sampling and collecting data in symbol vector form yields the vector signal representation:

$$\mathbf{r}_n = \mathbf{G}\mathbf{b}_n + \mathbf{v}_n \quad n=1,2,..N$$

Interference suppression in block DS-CDMA downlink systems[11]



$$\mathbf{r}_m = \mathbf{G}\mathbf{b}_m + \mathbf{n}_m$$

For BER=10^{-2} ICA-RAKE receivers gives 8 dB gain in relation to RAKE receiver.

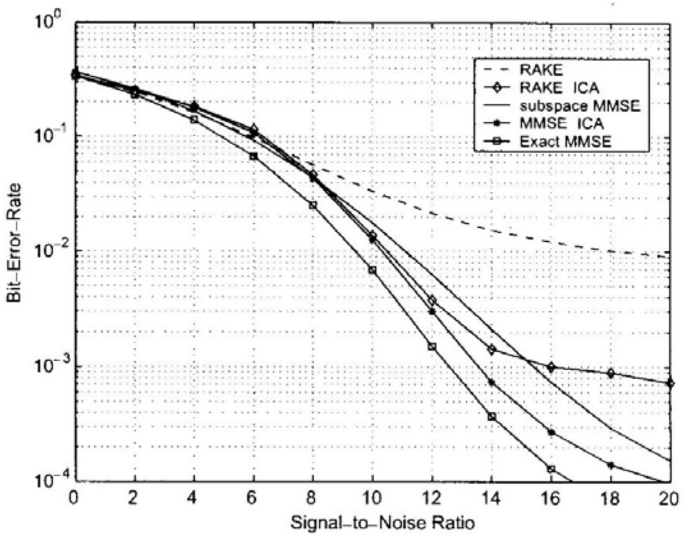


Fig. 3. Bit-error-rate as a function of signal-to-noise ratio $(0,\ldots,20)$ in an equal energy two-path Rayleigh block fading channel. The system includes $K = 3$ users with equal strength.

ICA and fMRI signal processing[12,13]

- 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.

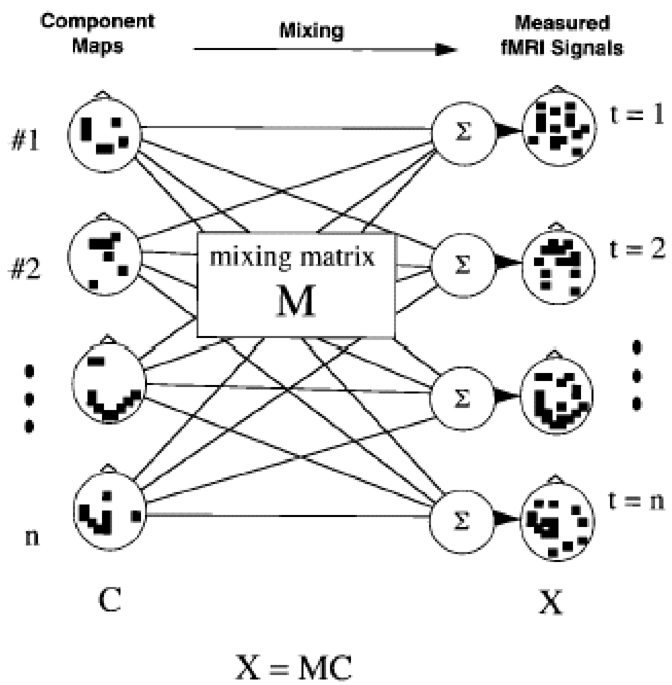


Figure 3.
fMRI data as a mixture of independent components. The mixing matrix M specifies the relative contribution of each component at each time point. ICA finds an *unmixing* matrix that separates the observed component mixtures into the independent component maps and time courses.

[12]McKeown, et. all, “Analysis of fMRI Data by Blind Separation Into Independent Spatial Components,” Human Brain Mapping 6: 160-188 (1998).

[13]M. J. McKewon, et. all, “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 $\boldsymbol{W}$.

Where $\boldsymbol{X}$ 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 jth voxel of the ith component.

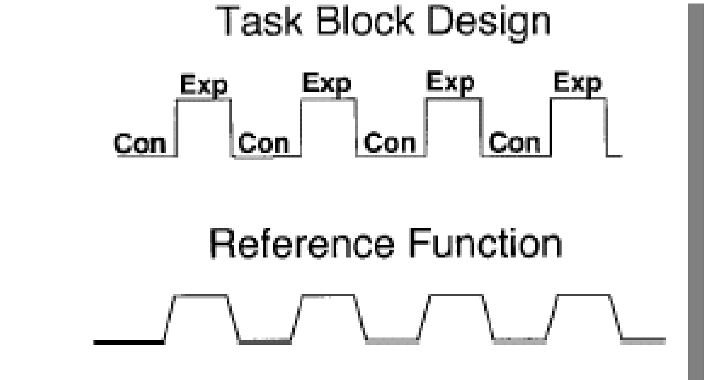
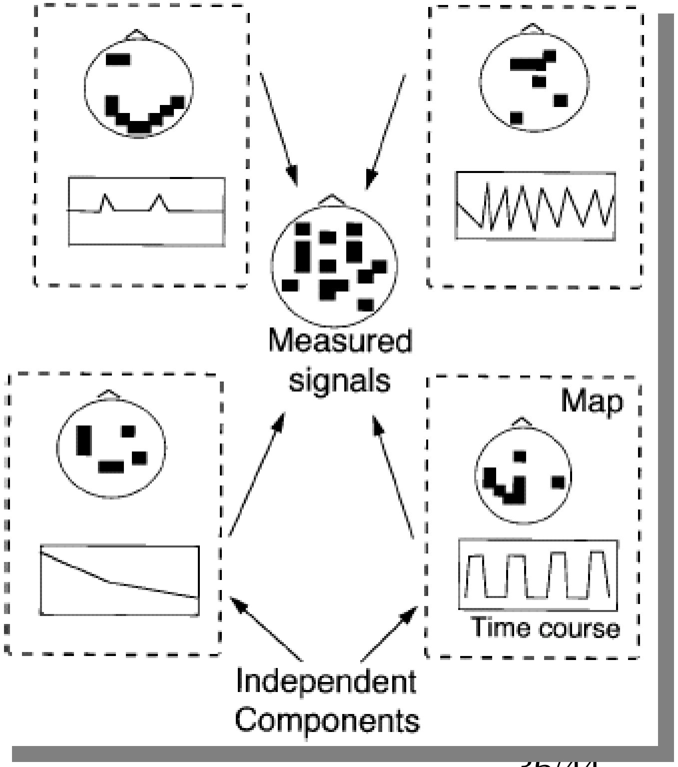


Figure 1.



ICA and fMRI signal processing[13]

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

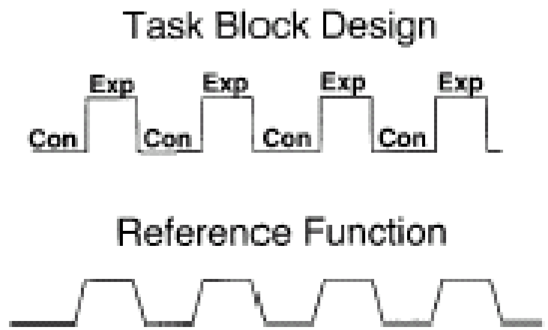


Figure 1.
BOLD signal complexity and task reference function. a: Time courses of 10 randomly selected voxels from a 6-min fMRI trial of the Stroop color-naming task illustrate the typical complexity of BOLD signals. b: Convolving an a priori estimate of the hemodynamic response function with the square-wave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

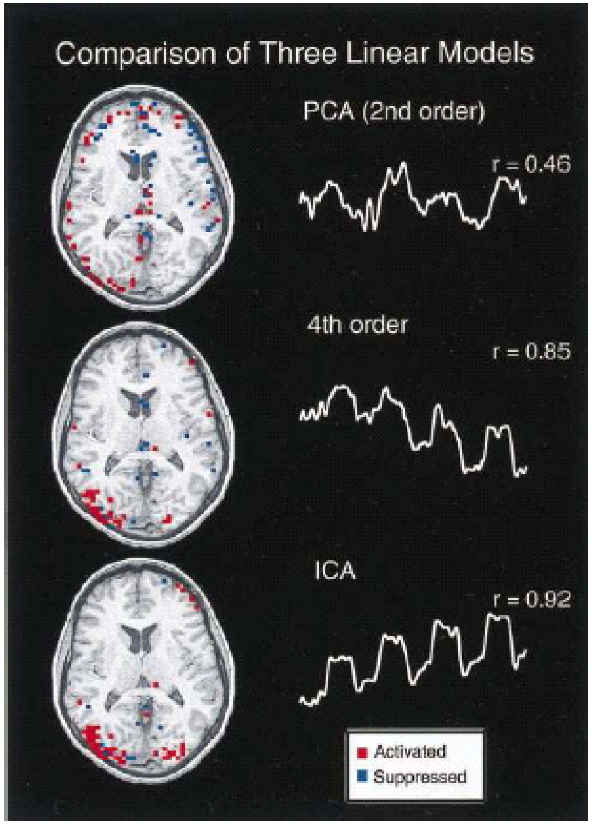


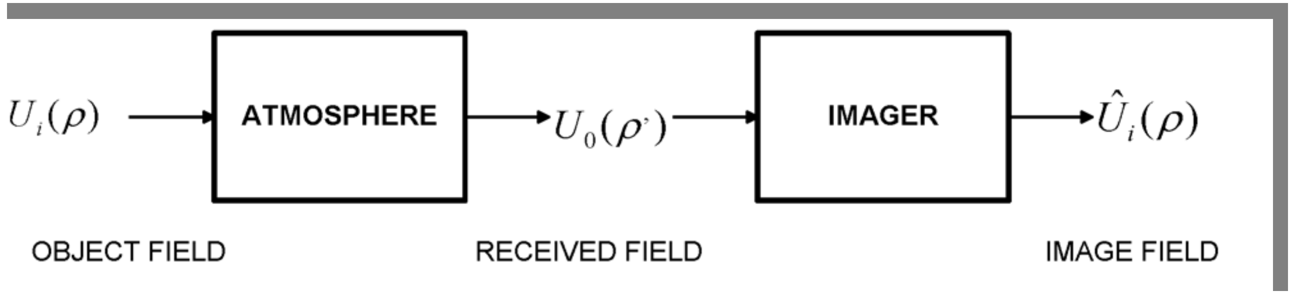
FIG. 2. Comparison of three linear models for analyzing fMRI data. PCA and two versions of ICA were used to linearly separate the data into partially spatially independent maps. The most consistently task-related component determined by each of the three methods from the first trial are shown, along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

Applications of ICA

Imaging in atmospheric turbulence

ICA and image sharpening in the atmospheric turbulence[14]

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.



By Hygens-Fresnel principle

$$\hat{U}_i'(\alpha_2) = \int d\alpha_1 U_0(\alpha_1) p_l(\alpha_2 - \alpha_1) \qquad \alpha_i = \rho_i / L \text{ angular coordinate}$$

The point spread function

$$p_l(\alpha) = \lambda^{-2} \int_{R_2} d\rho' \exp\left\{\chi\left(\rho',0\right) + i\left[\Phi\left(\rho',0\right) + k\rho'\alpha\right]\right\}$$

$$\chi\left(\rho',0\right) \quad \text{– log– amplitude of atmosperic turbulence}$$

$$\Phi\left(\rho',0\right) \quad \text{– phase of the atmospheric turbulence}$$

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

Physics of the propagation of the quasi-monochromatic EM field

Due to the Taylor’s frozen hypothesis

$$n_m(x, y, t) = n_m(x - v_x(t - t_0), y - v_y(t - t_0), t_0)$$

- **Interpretation:** Source of turbulence can be placed at some reference temporal location $\rho_i(t_0)$ such that relative turbulence contribution for $t>t_0$ can be taken into account as

$$\chi(\rho', 0, t) = \Delta\chi(\rho', 0, t) + \chi(\rho, 0, t_0)$$

$$\Phi(\rho', 0, t) = \Delta\Phi(\rho', 0, t) + \Phi(\rho, 0, t_0)$$

Physics of the propagation of the quasi-monochromatic EM field

The ICA framework

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

The mixing coefficients

$$a_{kn} = L^{-2} A_{sn} \exp\left[2\Delta\chi(\rho_n', 0, t_k)\right]$$

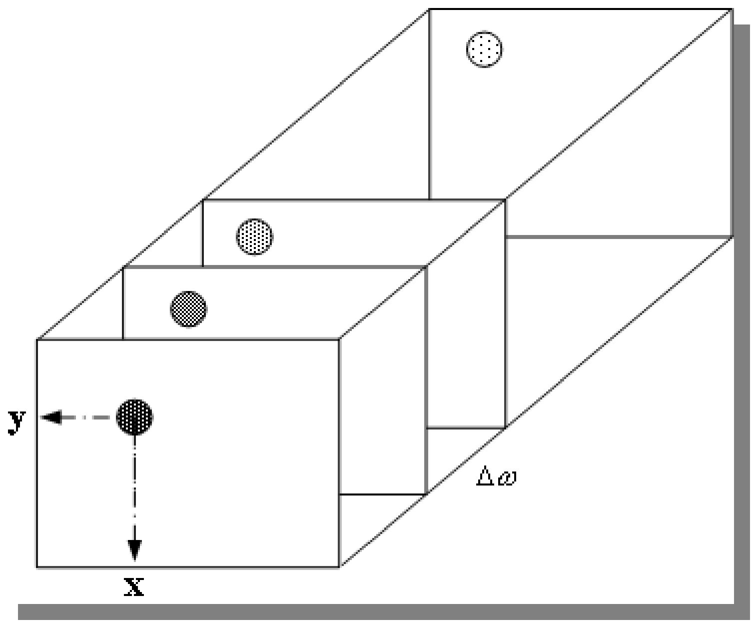
The source signals

$$I_{on}(\alpha, t_0) \;=\; I_{0n}(t_0)\left|p_l(\alpha, t_0)\right|^2$$

ICA representation of the image sequence

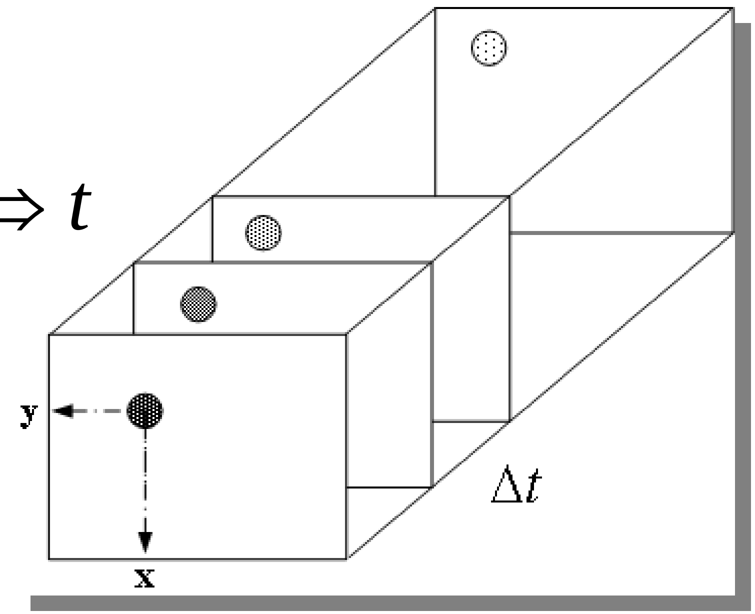
$$\mathbf{I}_i(x,y) = \mathbf{A}\mathbf{I}_{\mathbf{o}}(x,y) \;\; + \;\; \nu(x,y)$$

Image cube for multispectral imaging



$$\omega \Leftrightarrow t$$

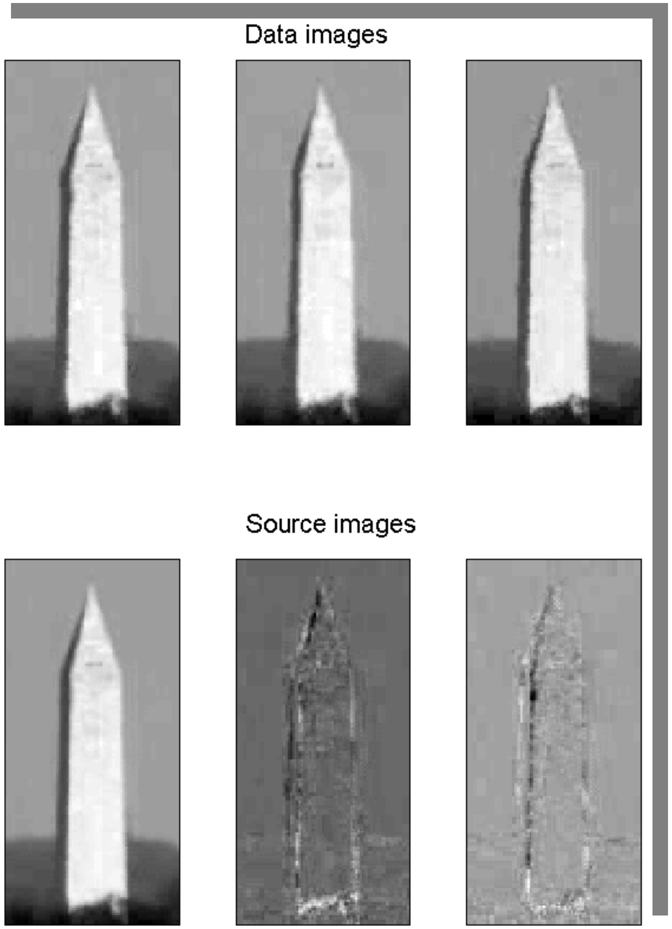
Image cube for video sequence



- ❑ Mixing matrix nonsingularity condition is fulfilled by ensuring that selected frames have nonzero mutual information (measured by the Kullback divergence).

Experimental results

- Three randomly selected frames with nonzero mutual information



Experimental results

- ❑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).