Seminar at the “Signal & Images Laboratory of ISTI”- CNR, Pisa, Italy.

November 2005

“Separation of statistical dependent sources using a measure of non-Gaussianity”

Cesar Caiafa[1]

ccaiafa@fi.uba.ar

[1]Phd student, Laboratorio de Sistemas Complejos.
Facultad de Ingenieria, Universidad de Buenos Aires, Argentina
http://www.fi.uba.ar/laboratorios/lsc/

Summary

1- Introduction

1.1- Statement of the BSS problem
1.2- Independent Sources case (ICA) (some famous algorithms)
1.3- Dependent Sources case (MI, Non Gaussianity, Minimum Entropy Methods, etc)

2- The MaxNG (Maximum Non-Gaussianity) algorithm

2.1- A measure of Non Gaussianity based on the L^2 euclidean distance.
2.2- Parzen Windows estimation of probability density functions
2.3- Local Maxima search
2.4- Computational complexity

3- Examples / Experimental Results

2.1- MaxNG versus MinMI
2.2- MaxNG performance
2.3- Comparison with some famous ICA/BSS algorithms

4- Conclusions

Blind Source Separation (BSS)
General Statement of the problem

During the last two decades, many algorithms for source separation were introduced, specially for the case of independent sources reaching to the so called **Independent Component Analysis (ICA)**. Generally speaking the purpose of BSS is to obtain the best estimates of M input signals (**s**) from their M observed linear mixtures (**x**) .

The Linear Mixing Model:

$$\overset{\text{sources}}{s = \begin{bmatrix} s_0 \\ s_1 \\ \vdots \\ s_{M-1} \end{bmatrix}} \qquad \overset{\text{mixtures}}{x = \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_{M-1} \end{bmatrix}}$$

$$\underset{\substack{\swarrow \\ \text{mixtures}}}{x(t)} = \underset{\substack{\downarrow \\ \text{Mixing} \\ \text{matrix (MxM)}}}{A}\underset{\substack{\searrow \\ \text{sources}}}{s(t)}$$

As usually, it is assumed sources signals with zero-mean and unit-variance.

Obtaining sources estimates ($\hat{s}$) is a linear problem, we look for a separating matrix D such that

$\hat{s} = Dx$ where $D = PA^{-1}$ with P being a permutation matrix

Where $\hat{s}$ is composed by permuted and/or sign changed versions of s(t) entries.

<u>Note</u>: A more complete model should consider additive noise and non-square matrix A

Independent Sources case

- A precise mathematical framework for ICA (noiseless case) was stated by P. Comon (1994). He has shown that if at most one source is Gaussian then ICA problem can be solved, has explained the permutation indeterminacy, etc.

- Many algorithms were developed by researches using the concept of contrast functions (objective functions to be minimized) mainly based on approximations to Mutual Information-MI measure is defined as follows through the Kullback-Leibler distan

$$I(\mathbf{y}) = \int p(\mathbf{y}) \log \frac{p(\mathbf{y})}{\prod_i p(y_i)} d\mathbf{y}$$

Joint density

Marginal density

Note that, if all source estimate y_i are independent, then $p(\mathbf{y}) = \prod_i p(y_i)$ and I(y)=0

- ICA can be interpreted as an extension of Principal Component Analysis (PCA). In PCA we seek for the sources that are uncorrelated (not necessarily independent) that concentrate as much as possible the energy of signals in the principal components. ICA requires more than decorrelation of data, it requires the independence of sources.

Some famous ICA/BSS algorithms

Some ICA algorithms that minimize Mutual Information of source estimates

- **P. Comon algorithm (1994);**
- **InfoMax (1995)** by Sejnowski et al;
- **FastIca (1999)** by Hyvärinen;
- **R. Boscolo algorithm (2004);**
- and others.

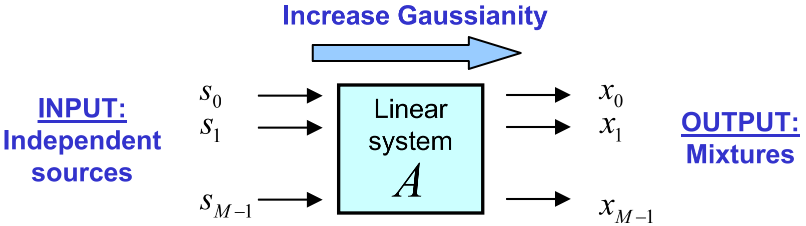
Some BSS algorithms that exploit the time correlation of sources (based on second order statistics (SOS) or higher order statistics (HOS))

- **AMUSE (Algorithm for Multiple Unknown Signals Extraction) (1990)** by L. Tong et al**;**
- **JadeTD (Joint Approximate Diagonalization Eigenmatrices with Time Delay) (2002)** by **.** Georgiev et al (based on the JADE algorithm – Cardoso (1993))
- **SOBI (Second Order Blind Identification) (1993)** by A. Belouchrani et al;
- **EVD (Eigenvalue Decomposition) (2001)** by P . Georgiev and A. Cichocki;
- and others.

Note: This is not a complete list. There are a lot of algorithms in the literature

Mutual Information and non-Gaussianity relationship
How can we approach the Dependent Sources case?

- In ICA context, many authors have shown that minimizing MI of sources is equivalent to maximize the Non Gaussianity of source estimates. It's a consequence of Central Limit Theorem (P. Comon, A. Hyvärinen).



- But, what happen when the sources have some dependence degree and are correlated? The MI minimization may be not a good strategy. Cardoso has presented (2003) a meaningful decomposition of MI measure

$$I(\mathbf{y}) = C(\mathbf{y}) - \sum_{i=0}^{M-1} G(y_i) + k$$

Mutual Information

Overall correlation of **y** entries

Non-Gaussianity measure for y_i

constant

In case of imposing independence of sources, then C(y)=0 and clearly, it is equivalent minimize MI to maximize each of $G(y_i)$. **But if sources are allowed to be correlated maximizing non-Gaussianity is not the same as Minimizing MI.**

Maximum Non-Gaussianity and Minimum Entropy related methods

- Since the Gaussian distribution has the maximum entropy over all distributions with the same variance (C. Shanon-1948), **to maximize Non-Gaussianity means to Minimize Entropy**.

- Some measures of Non-Gaussianity have been used in the past in the framework of Projection Pursuit (PP) (J. H. Friedman - 1974). In PP lower-dimensional projection are looked for in order to analyze higher-dimensional data. A projection became more interesting as less Gaussian it is (more structured).

- Experimental results, as the ones presented here, show that Maximizing non-Gaussianity (minimize entropy) is useful for separate dependent sources. Other authors have mentioned the power of Minimum Entropy Methods (like D. Donoho - 1981) for dependent cases in different applications (i.e. deconvolution).

- The theoretical basement for the Maximum Non-Gaussianity method (Minimum Entropy) for dependent sources remains as an open issue. It is not clear at this time what are the minimum conditions that dependent sources must satisfy in order to guarantee the separation through a minimum entropy method.

A measure of Non Gaussianity based on the L^2 euclidean distance

Considering a continuous random variable y (with zero-mean and unit-variance), we define our non-gaussianity measure of a probability density function (pdf) p_y denoted by $\Gamma(p_y)$, as following:

$$\Gamma(p_y) = \int [\Phi(y) - p_y(y)]^2 \, dy$$

Integrated Squared Error

with $\Phi(y) = N(0,1) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}y^2\right)$

- $\Gamma(p_y)$ is the Euclidean distance in the space of probability density functions (with zero-mean and unit-variance).

- It provides a bounded positive number: $0 \le \Gamma(p_y) < \infty$

Estimating a probability density function (pdf) by Parzen Windows

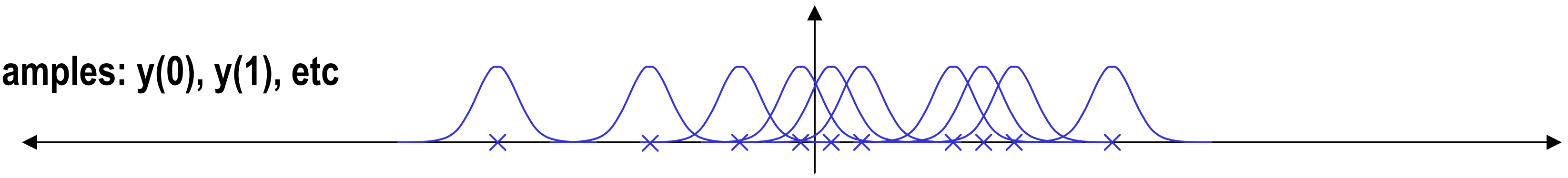
- We want to estimate a pdf from a set of N samples of the variable y: y(0), y(1),.., y(N-1).
- Parzen windows is a non parametric estimation technique and has the following form:

$$\widehat{p}_y(y) = \frac{1}{Nh} \sum_{i=0}^{N-1} \Phi\left(\frac{y - y(i)}{h}\right)$$

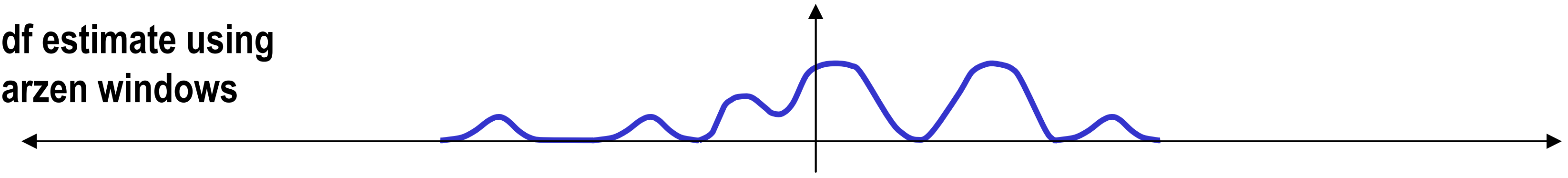
Where: Φ is a window function (or kernel), for example a Gaussian function, and

h is as the parameter which affects the width and height of the windows functions

Samples: y(0), y(1), etc



Pdf estimate using Parzen windows



Analytical form of Non-Gaussianity measure

Using Parzen windows and properties of the convolution of Gaussian functions, we finally reach to the following formulae for our Non-Gaussianity measure.

$$\Gamma(p_y) = \underbrace{\frac{1}{2\sqrt{\pi}}}_{\text{constant}} - \frac{2}{N\sqrt{h^2+1}} \underbrace{\sum_{i=0}^{N-1} \Phi\left(\frac{y(i)}{\sqrt{h^2+1}}\right)}_{\text{Complexity : O(N)}} + \frac{1}{N^2 h\sqrt{2}} \underbrace{\sum_{\substack{i=0\\j=0}}^{\substack{N-1\\N-1}} \Phi\left(\frac{y(j)-y(i)}{\sqrt{2}h}\right)}_{\text{Complexity : O}(\text{N}^2)}$$

<u>Notes</u>:

- **The advantage of having an analytical expression of the measure, is that we are able to analytically calculate derivatives for searching the local maxima.**

- **Additionally this expression has good properties, is continuous, is a linear combination of Gaussians, etc**

- **We can take advantage on these properties in order to obtain a fast calculation of local maxima.**

The MaxNG Algorithm

- Source estimate vector has the linear form: $\mathbf{y}(t) = D\mathbf{x}(t)$

- We propose to determine matrix D row by row maximizing the measure of non-Gaussianity for each y_i for a fixed unit-variance. In other words we need to find M different local maxima of the non-Gaussianity measure (p_y) using a parameterization of the separating matrix.

- In order to enforce the unit-variance of y_i we propose to do two things (Pre-processing):

1- To apply a spatial whitening (or sphering) technique using the well known Karhunen-Loeve transformation.

$$\tilde{\mathbf{x}} = \underbrace{V\Lambda^{-1/2}V^T}_{\text{KLT}}\mathbf{x} \quad \Longrightarrow \quad E\left[\tilde{x}\tilde{x}^T\right]=R_{\tilde{x}\tilde{x}}=I$$

Whitened data

KLT
(eigenvectors, eigenvalues)

Original data (mixtures)

2- Parameterize each row of new separating matrix $\widetilde{D}$ as unit norm vectors using hyperspheric coordinates.

$$\mathbf{y} = \widetilde{D}\tilde{\mathbf{x}} \quad \Longrightarrow \quad E\left[yy^T\right]=R_{yy}=\widetilde{D}\widetilde{D}^T$$

For example for M=3 a general parameterization of a row of matrix $\widetilde{D}$ is:

$$\left[\sin(\theta_0)\cos(\theta_1) \quad \sin(\theta_0)\sin(\theta_1) \quad \cos(\theta_0)\right]$$

- In general, for estimating M sources from M mixtures we need to search for M local maxima in a (M-1)-dimensional space

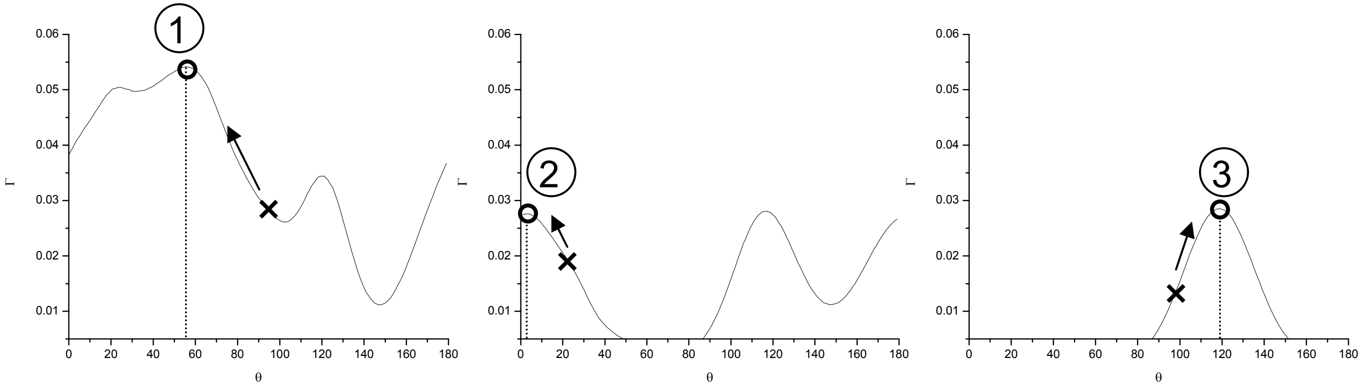
Local Maxima search

- In order to maximize the Non-Gaussianity measure we implement a search based on the Gradient Ascend Algorithm. Fortunately an exact expression of the gradient can be directly obtained.

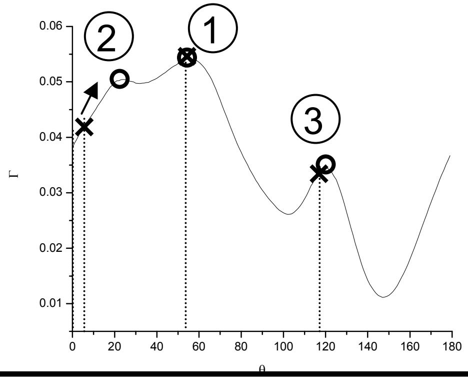
- In order to obtain M different local maxima we run the gradient ascend search many times from different starting points. After each local maximum is found, we need to remove it from the objective function in order to avoid to reach it again. This procedure is called Deflation or non-Gaussian structure removal.

Example of finding local maxima in with M=2

Step 1:
Random
starting points
+ deflation

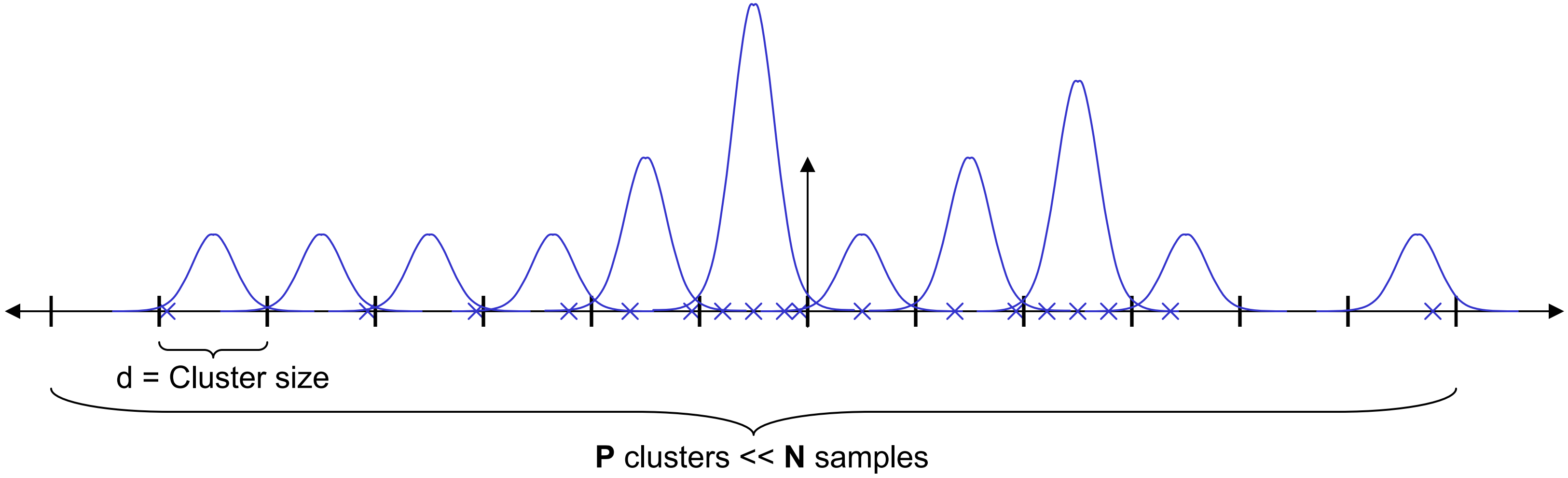


Step 2:
Previous solutions as
Starting points and
NO deflation



Computational complexity

- One of the main issues addressed in this approach is how to reach to a reasonable computational complexity in order to allow fast algorithms.
- A common technique in Parzen windows methods is to arrange the data set in clusters.



- Summations of **N** elements are reduced to summations of **P** elements

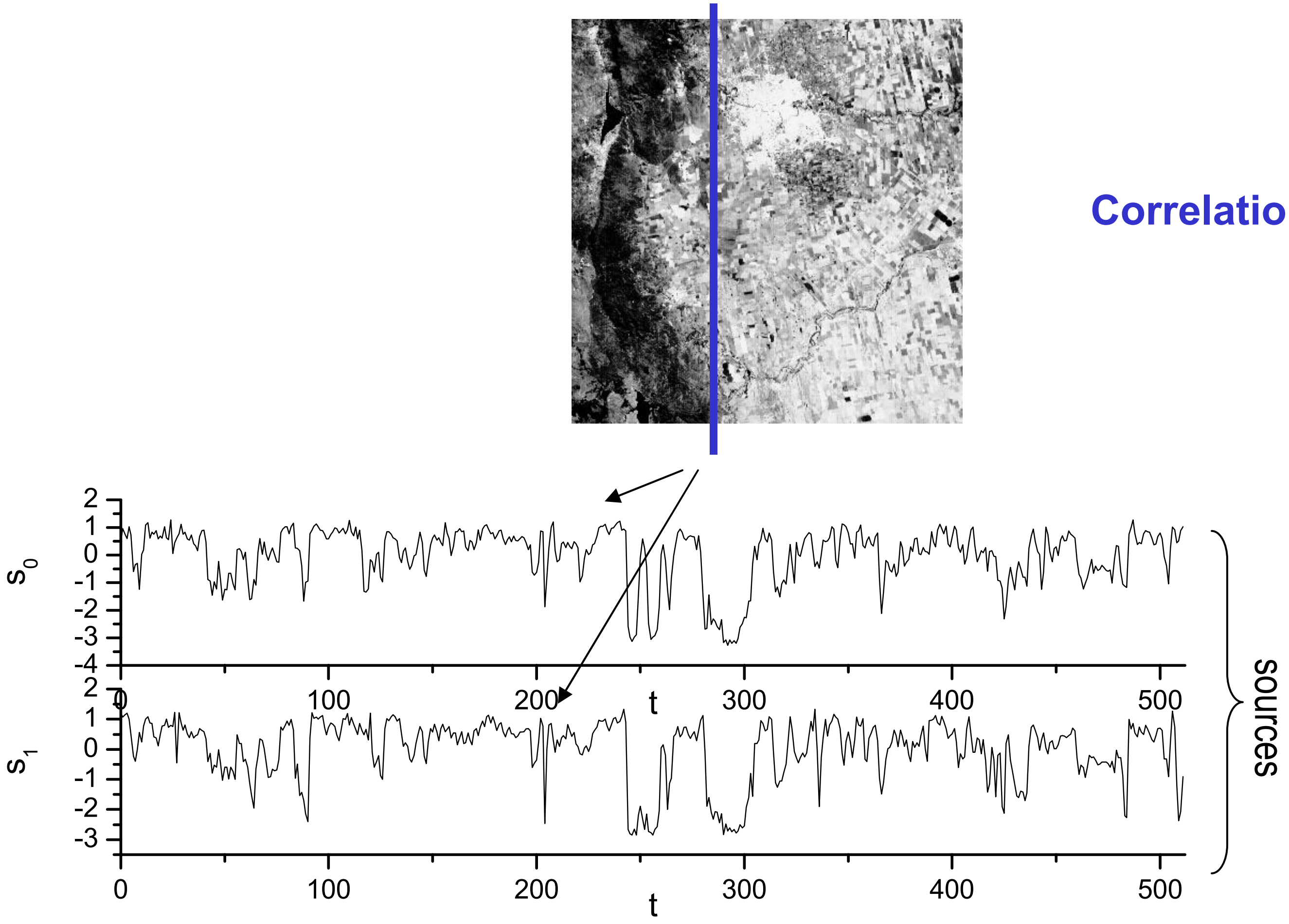
$$\sum_{i=0}^{N-1} \Phi(x - x_i) \approx \sum_{n=0}^{P-1} f(n)\Phi(x - x_n)$$

- Additionally double summations (O(**N**^{2})) can be written as convolutions and calculated in the Fourier domain (FFT).

$$\sum_{i=0}^{N-1}\sum_{j=0}^{N-1} \Phi(x_i - x_j) \approx \sum_{m=0}^{P-1} f(m)\left(\sum_{n=0}^{P-1} f(n)\Phi(n-m)\right) = \sum_{m=0}^{P-1} f(m)(f * \Phi)(m)$$

Some experimental results
Example: High correlated data

Sources extracted from two adjacent columns in a satellite image

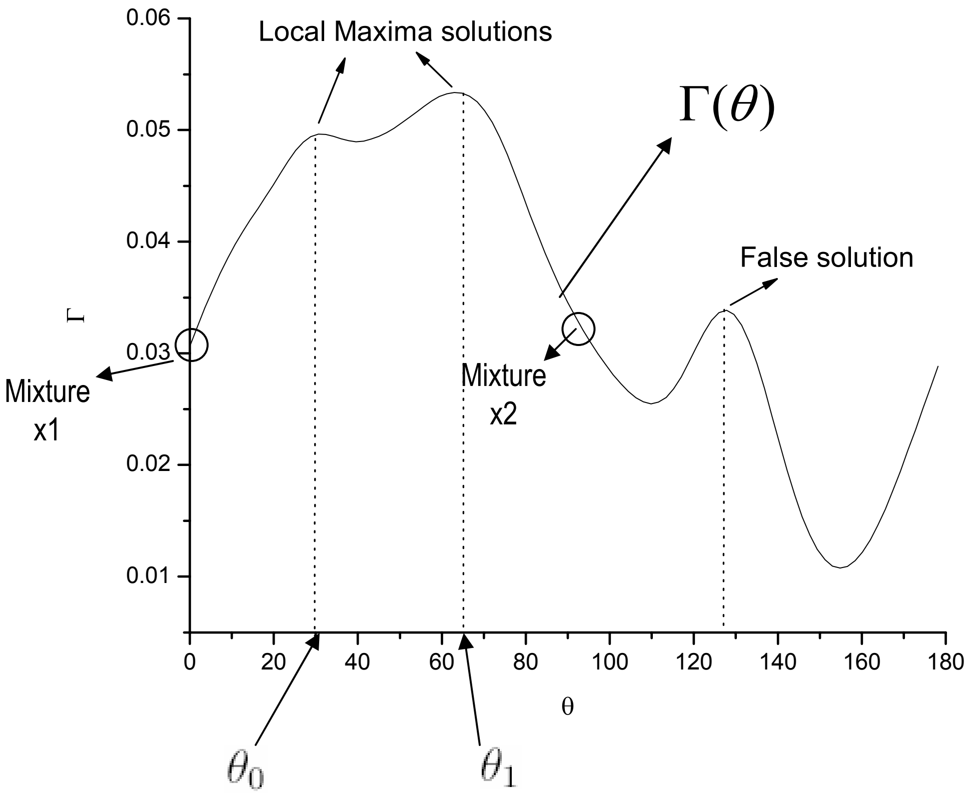


Correlation coefficient = 0.81

Mixing matrix

$$A = \begin{bmatrix} \frac{1}{\sqrt{10}} & \frac{3}{\sqrt{10}} \\ \frac{2}{\sqrt{10}} & \frac{1}{\sqrt{10}} \end{bmatrix}$$

Some experimental results (cont)
Non Gaussianity measure estimates



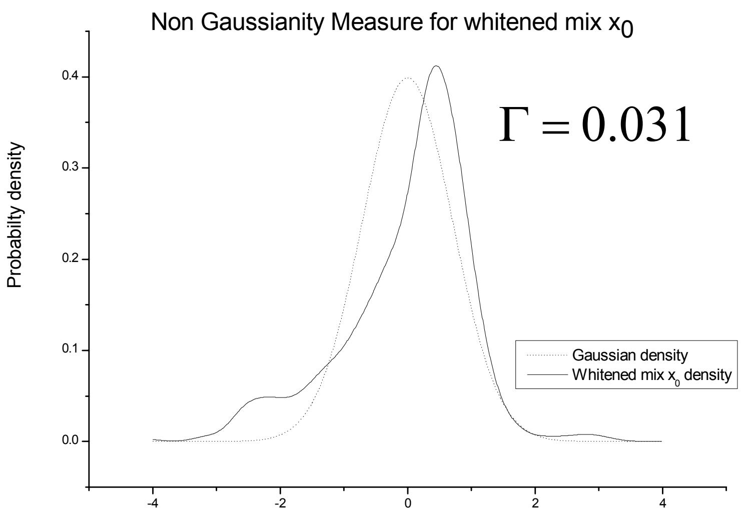
Separating matrix parameterization

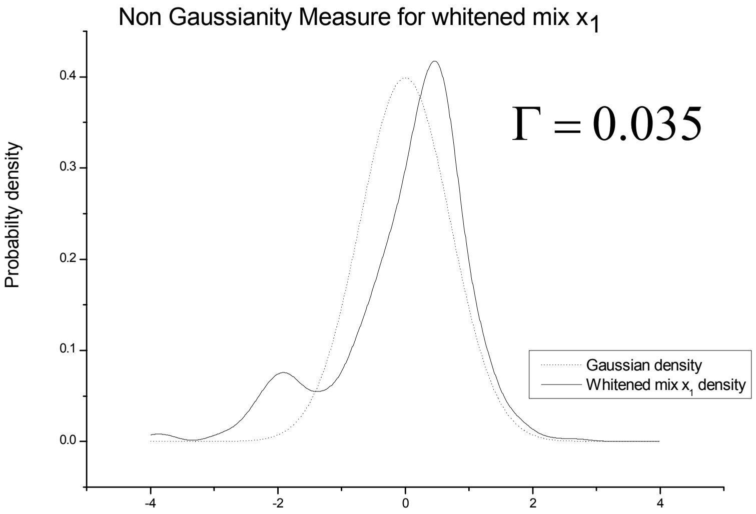
$$\widetilde{D} = \begin{bmatrix} \cos(\theta_0) & \sin(\theta_0) \\ \cos(\theta_1) & \sin(\theta_1) \end{bmatrix}$$

Important: some false solutions can be reached by the algorithm.

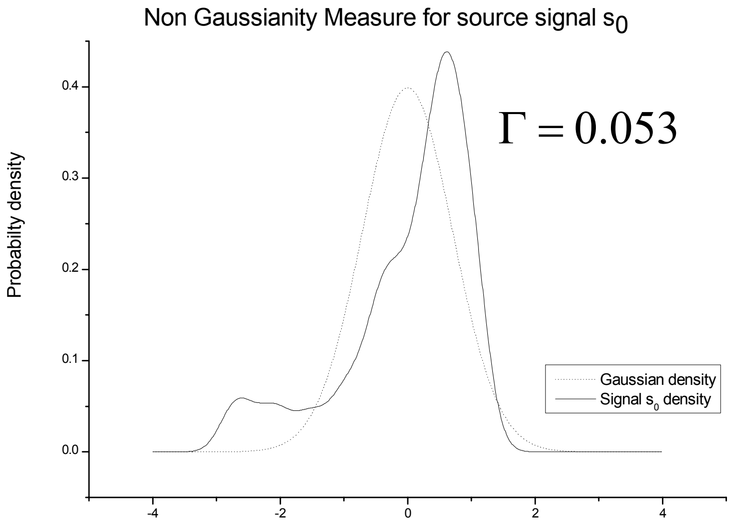
Some experimental results (cont) – pdf comparison

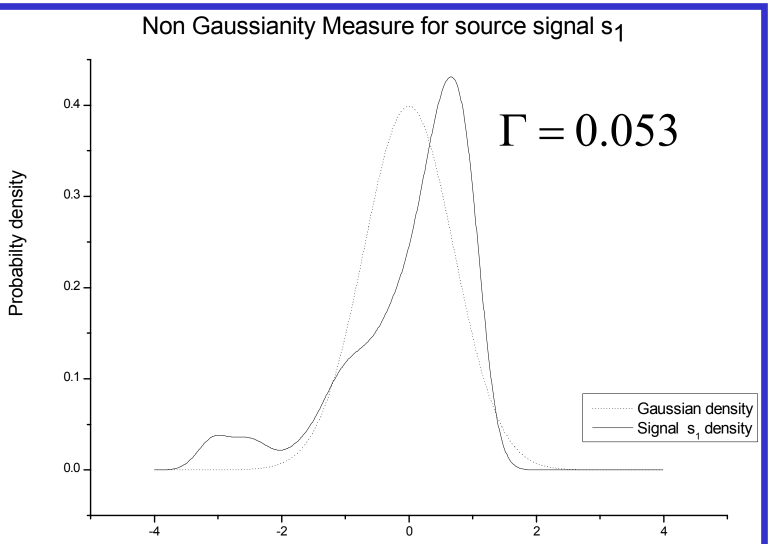
Whitened mixtures





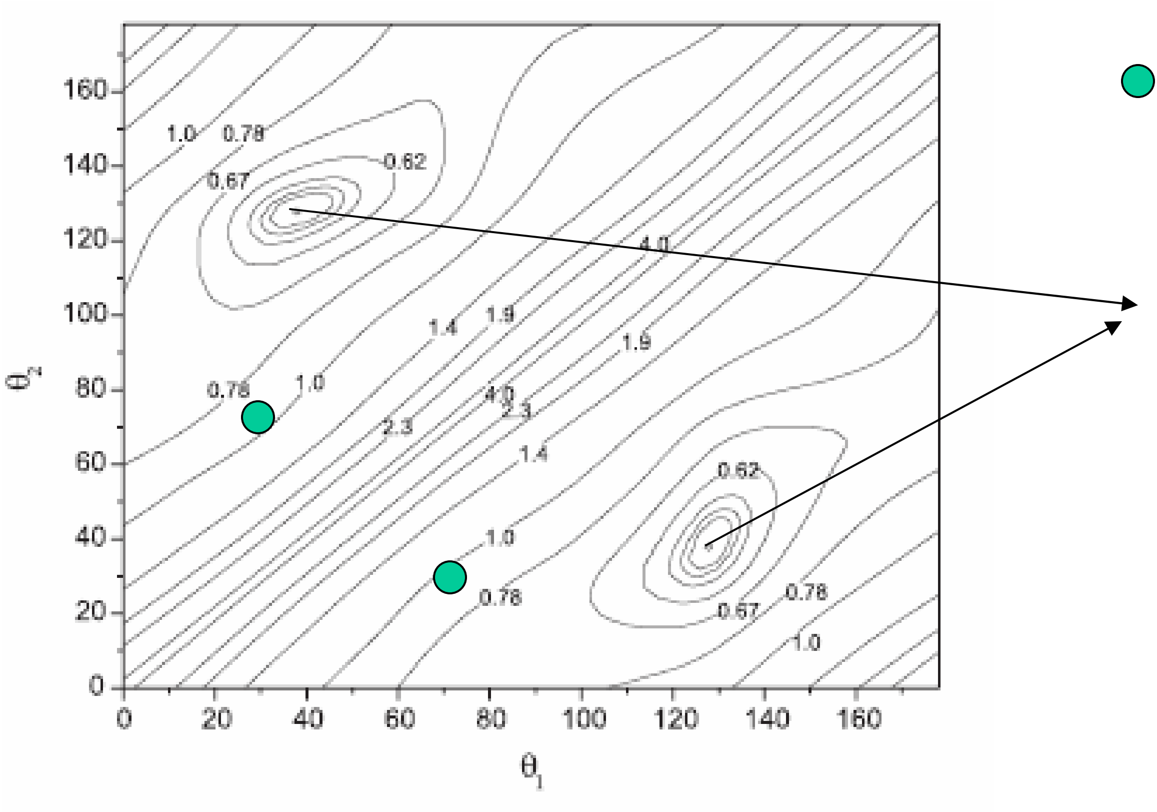
Sources





Some experimental results (cont)
Minimum Mutual Information (MI) estimates

Using a calculation of Mutual Information based on R. Boscolo work we obtained the following for the same example. Mutual Information is a function of two angles. $I(\theta_1, \theta_2)$



● Real position of angles

Minima are not located in the real positions (Wrong estimate)

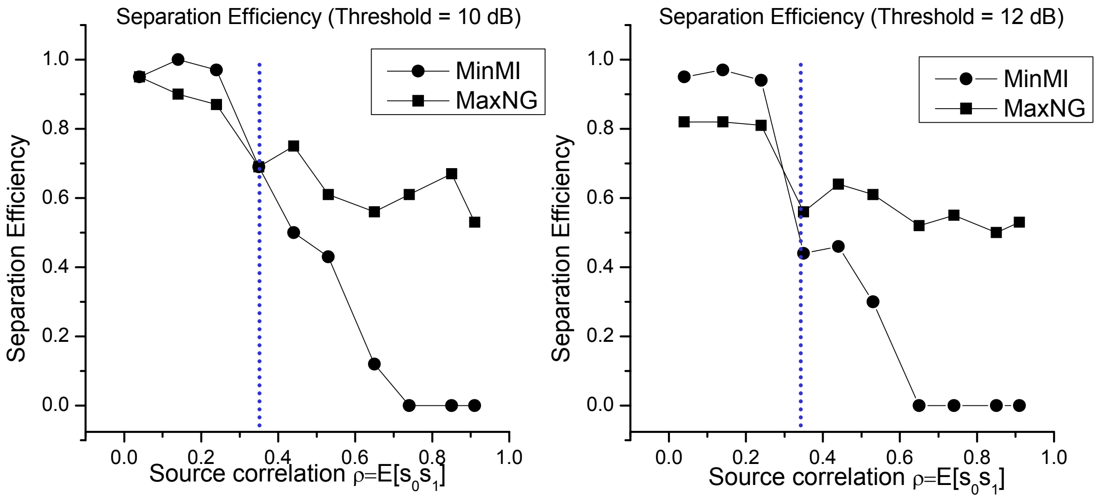
Some experimental results (cont)
Experiment 1: MaxNG vs Min MI comparison

A total of 300 simulations for different sources and different levels of dependence were done. Original sources where extracted from pixel columns of various satellite images. Selecting different column offsets between signals, we have a control of the level of dependence. A signal length of **N=512** was chosen. Using a known mixing matrix A, mixtures were generated and estimation of sources using MaxNG and Min MI were performed.

Error: $\mathbf{e} = \hat{\mathbf{s}} - \mathbf{s}$ Signal To Interference ratio $SIR_i = -10\log_{10}(var(e_i))$

In general, SIR levels below 10-12 dB threshold are indicative of a failure in obtaining the desired source separation.

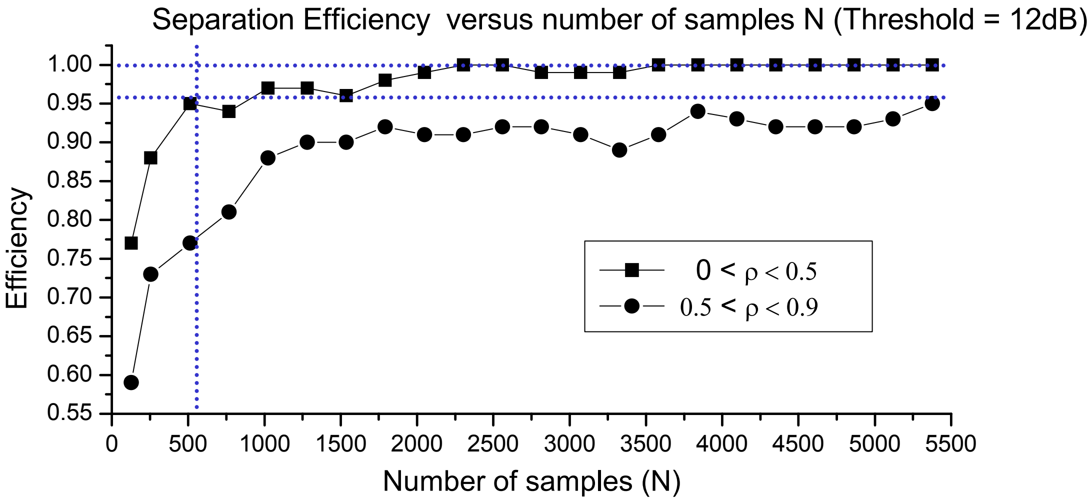
$$Efficiency\ (\%) = \frac{\text{Number of cases with MinSIR above threshold}}{\text{Total number of cases}}$$



Some experimental results (cont)
Experiment 2: MaxNG algorithm efficiency versus data sample size N

Correlation of sources $\rho = E[s_0 s_1]$

We have applied the algorithm for data sample size from **N=128** to **N=5376**, with a step size N=128. Efficiency was calculated averaging a **total of 600** separation cases for each N.



Some experimental results (cont)
Experiment 3: Comparison with other algorithms

We have compared the results of our MaxNG algorithm against the results obtained through the application of some classical BSS/ICA methods like: AMUSE, EVD2, SOBI, JADE-opt, FPICA, Pearson-opt (ICALAB software package).

Case 1: Speech signals. Two speakers say the same sentence. These signals were extracted from the ICALAB benchmark example named halo10.mat. These signals exhibit a slight level of correlation, in our case was: **$E[s_0 s_1]=-0.049$**. The number of used data samples was N=6000.

Case 2: Satellite signals. Two pixel columns were extracted from an optical satellite image. These two columns were 2 pixels apart one from the other in the original image, therefore they are highly correlated, the coefficient correlation was: **$E[s_0 s_1]=0.818$** which is a very high value. The number of used data samples was N=5960.

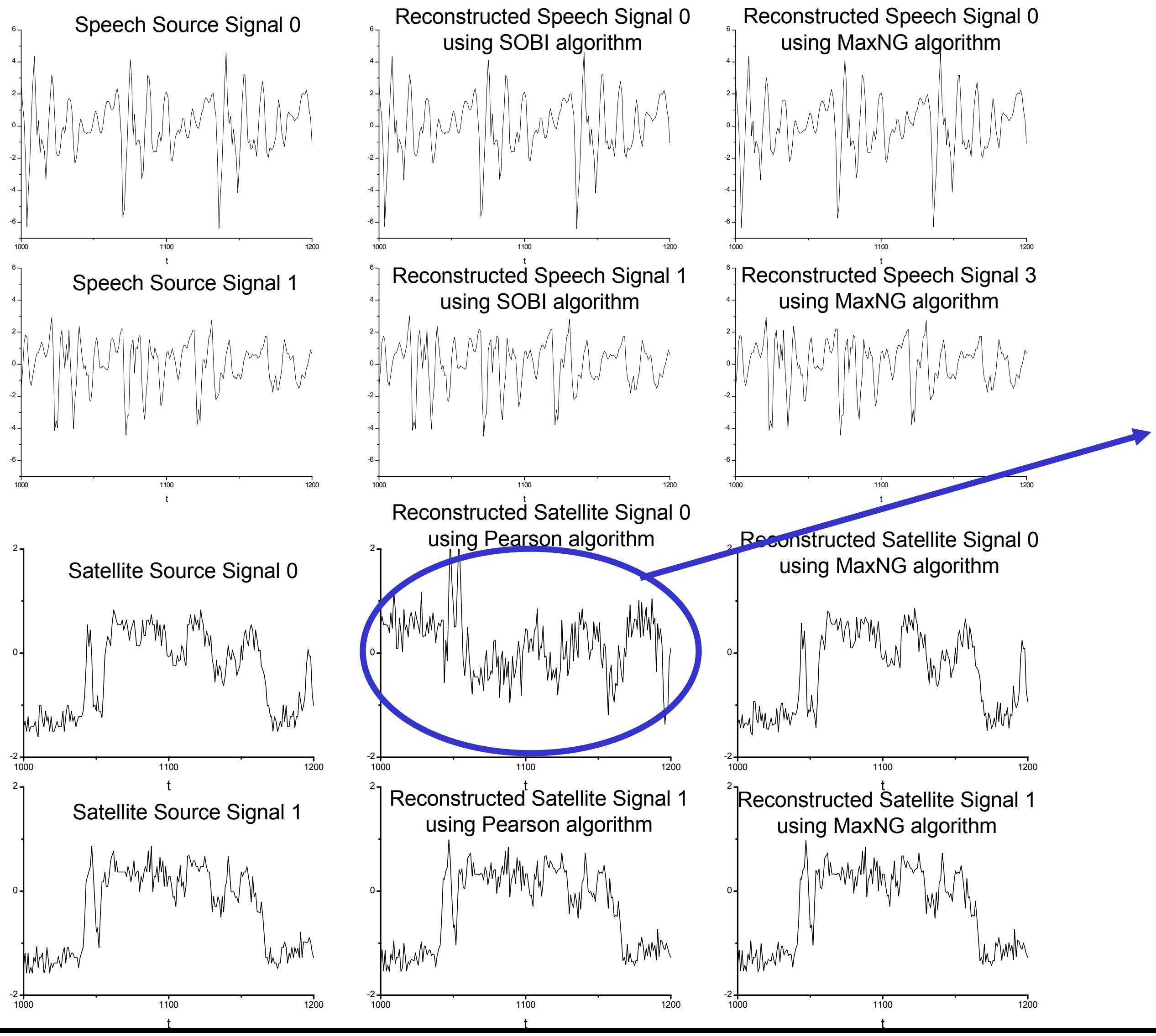
Example 1: speech signals	AMUSE	EVD2	SOBI	JADE	FPICA Gauss	Pearson	MaxNG
SIR signal 0	39.59	49.45	63.97	11.41	31.42	25.83	25.20
SIR signal 1	28.36	32.27	31.34	10.57	61.11	22.07	57.27
MeanSIR	33.97	40.86	**47.66**	10.99	46.27	23.95	**41.24**

Example 2: satellite signals	AMUSE	EVD2	SOBI	JADE	FPICA Gauss	Pearson	MaxNG
SIR signal 0	9.80	9.92	0.11	9.83	9.57	2.95	20.29
SIR signal 1	10.42	10.30	0.11	10.39	10.68	19.92	20.40
MeanSIR	10.11	10.11	0.11	10.11	10.12	**11.43**	**20.34**

Visual comparison of results

Speech Signals

Satellite Signals



Speech Source Signal 0

Reconstructed Speech Signal 0 using SOBI algorithm

Reconstructed Speech Signal 0 using MaxNG algorithm

Speech Source Signal 1

Reconstructed Speech Signal 1 using SOBI algorithm

Reconstructed Speech Signal 3 using MaxNG algorithm

Satellite Source Signal 0

Reconstructed Satellite Signal 0 using Pearson algorithm

Reconstructed Satellite Signal 0 using MaxNG algorithm

Satellite Source Signal 1

Reconstructed Satellite Signal 1 using Pearson algorithm

Reconstructed Satellite Signal 1 using MaxNG algorithm

Wrong Estimate

Signal & Images Laboratory of ISTI, CNR, Pisa, Italy. 2005

Conclusions

- Non-Gaussianity measures are useful for separating dependent sources.
- A new algorithm called MaxNG is proposed showing good performance for independent as well as for dependent sources.

Discussion about future directions

- The theoretical basement for Minimum Entropy (MaxNG) methods is an open issue. It seems to be a powerful and general approach but precise conditions on signals are not available
- An extension to a noisy model should be investigated. It is well known that in this case, the estimation of sources is a non-linear estimation problem.
- An extension to non-square mixing matrix should be investigated.