MaxEnt 2006

Twenty sixth International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering CNRS, Paris, France, July 8-13, 2006

"A Minimax Entropy Method for Blind Separation of Dependent Components in Astrophysical Images"

Cesar Caiafa

ccaiafa@fi.uba.ar

Laboratorio de Sistemas Complejos. Facultad de Ingenieria, Universidad de Buenos Aires, Argentina

In collaboration with Ercan E. Kuruoglu (ISTI-CNR, Italy) and Araceli N. Proto (CIC, Argentina)

Summary

1-Introduction

- 1.1- Statement of the BSS problem
- 1.2- Independent Sources case (ICA)
- 1.3- Dependent Sources case (DCA)

2- Entropic measures

- 2.1- Shannon Entropy (SE) and Gaussianity Measure (GM)
- 2.2- Parzen Windows based calculations

3- The MiniMax Entropy algorithm for separation of astrophysical images

3.1- The Planck Surveyor Satellite mission

3.2- Description of the MiniMax Entropy method

4- Experimental results

- 2.1- Noiseless case
- 2.2- Robustness against noise

5- Conclusions

Blind Source Separation (BSS) General Statement of the problem

The seminal work on blind source separation is by Jutten, Herault and Guerin (1988). 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 P input signals (s) from their M observed linear mixtures (x).



Sources signals are assumed with zero-mean and unit-variance. We consider here the overdetermined case $(M \ge P)$

In the noiseless case (n=0), obtaining sources estimates (\hat{s}) is a linear problem:

$$\mathbf{\hat{s}} = A^{\dagger} \mathbf{x}$$
 Where A^{\dagger} is the Moore-Penrose inverse matrix

<u>Note</u>: When noise is present, a non-linear estimator is required.

Independent Sources (ICA)

• 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 distance:



By minimizing Mutual Information

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

By exploiting the time structure of sources Second and High Order statistics (SOS-HOS)

- •AMUSE (1990) by L. Tong et al;
- JadeTD (2002) by . Georgiev et al (based on the JADE algorithm Cardoso (1993))
- SOBI (1993) by A. Belouchrani et al;
- EVD (2001) by P . Georgiev and A. Cichocki; and others.

DCA (Dependent Component Analysis) How can we separate Dependent Sources?

• Few algorithms for dependent sources were reported in the literature. Cichocki et al. (2000) have approached the separation of acoustic signals by exploiting their time correlations. Bedini et al. (2005) have developed an algorithm based on 2nd order statistics at different time lags for astrophysical images.

• In ICA context, many authors have shown that minimizing MI of sources is equivalent to minimize the Entropy of the non-Gaussian source estimates. It is a consequence of Central Limit Theorem (P. Comon, A. Hyvärinen).



• As we have experimentally demonstrated in a recent paper (Caiafa et al. 2006), when sources are allowed to be dependent, the minimization of the entropies of the non-Gaussian source estimates remains as an useful tool for the separation, while the minimization of MI fails.

• We introduce the term **DCA** (Dependent Component Analysis) for a method which obtains the non-Gaussian source estimates by minimizing their entropies allowing them to be cross correlated (dependent).

• This DCA method has demonstrated to be effective on several real world signals exhibiting even high degree of cross correlation (see examples of speech signals in Caiafa et al. (SPARS05) – 2005, Hyperspectral images in Caiafa et. al (EUSIPCO06 - 2006), and dependent signals taken from satellite images in Caiafa et al. (Signal Processing) in press (2006)).

Entropic measures

Considering a continuous random variable y (with zero-mean and unit-variance), we define the following Entropic measures:

Shannon Entropy (SE): $H_{SE}(y) = -\int p(y) \log[p(y)] dy$ Gaussianity Measure (GM): $H_{GM}(y) = -\int [p(y) - \Phi(y)]^2 dy$

with the Gaussian pdf defined as ussually by: $\Phi(y) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}y^2\right]$

By the Central Limit Theorem (CLT) effect, a linear combination of independent variables has a higher Entropic measure (SE and GM) value than individual variables.

Generalizations of the CLT for dependent variables allows us to base our method in these two measures.

Calculation of Entropic Measures by using Parzen Windows

• Given a set of N samples of the variable y: y(0), y(1),..., y(N-1), Parzen windows is a non parametric technique for the estimation of the corresponding pdf:

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

where: $\Phi(y)$ 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

• Shannon Entropy and Gaussianity Measure can be written in terms of data samples:

$$H_{SE}(y) = -\frac{1}{N} \sum_{j=0}^{N-1} \log \left[\frac{1}{N} \sum_{i=0}^{N-1} \frac{1}{h} \Phi\left(\frac{y(j) - y(i)}{h}\right) \right]$$
(Erdogmus et al. (2004)

$$H_{GM}(y) = -\frac{1}{N^2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \frac{1}{h\sqrt{2}} \Phi\left(\frac{y(j) - y(i)}{h\sqrt{2}}\right) + \frac{2}{N} \sum_{i=0}^{N-1} \frac{1}{\sqrt{h^2 + 1}} \Phi\left(\frac{y(i)}{\sqrt{h^2 + 1}}\right) - \frac{1}{2\sqrt{\pi}} \quad \text{(Caiafa et al. (2006))}$$

Notes:

• The advantage of having an analytical expressions for these measures, is that we are able to analytically calculate derivatives for searching the local maxima.

• Parzen window estimation technique also allows us to implement the calculations in a fast way by calculating convolutions through the Fast Fourier Transform (FFT) (Silverman (1985))

The astrophysical problem The Planck Surveyor Satellite mission



Sensor Measurements at different center frequencies: (100 GHz, 70 GHz, 44 GHz and 30 GHz)

Planck Telescope (on a satellite)



SOURCES: - CMB (Cosmic microwave Background) - DUST (Thermal Dust) - SYN (Galactic Synchrotron)

Assumptions:

A1: CMB images are Gaussian, DUST and SYN images are non-Gaussian.
A2: CMB-DUST and CMB-SYN are uncorrelated pairs. (DUST-SYN are usually correlated)
A3: We consider low level noise (source estimates can be obtained as linear combination of mixtures)

Objective: To obtain estimates of CMB, DUST and SYN images (sources) by using the available measurements (mixtures).

The MiniMax Entropy algorithm for the astrophysical case

• By using the low level noise assumption (A3), the source estimates are: $\hat{\mathbf{s}} = D\mathbf{x}$

• In order to enforce source estimates to have unit-variance, we first apply a whitening (or sphering) filter and we define a new separating matrix which can be parameterized with spherical coordinates:

$$\hat{\mathbf{S}} = \widetilde{D}\widetilde{\mathbf{X}} \quad \text{with} \quad \widetilde{\mathbf{X}} = \underbrace{\Lambda^{-\frac{1}{2}}V^{T}}_{\text{(Non zero eigenvalues, and eigenvectors)}} \text{(Karhunen Loeve Transformation)} \\ \cdot \text{Covariance Matrices are:} \quad \begin{cases} E[\widetilde{\mathbf{x}}\widetilde{\mathbf{x}}^{T}] = R_{\widetilde{\mathbf{x}}\widetilde{\mathbf{x}}} = I \\ E[\widehat{\mathbf{s}}\widehat{\mathbf{s}}^{T}] = R_{\widehat{\mathbf{s}}\widetilde{\mathbf{s}}} = \widetilde{D}\widetilde{D}^{T} \end{cases}$$

- Then, each row of matrix \tilde{D} has unit-norm and therefore can be parameterized by using spherical coordinates:

$$\widetilde{\mathbf{d}}(\theta_0, \theta_1) = \begin{bmatrix} \sin(\theta_0) \cos(\theta_1) & \sin(\theta_0) \sin(\theta_1) & \cos(\theta_0) \end{bmatrix}^T$$

• And every source estimate can be obtained by identifying the appropriate points in the parameter space

$$\hat{\mathbf{s}}_i = \widetilde{\mathbf{d}}_{\mathbf{i}}^{\mathrm{T}}(\theta_0^i, \theta_1^i) \cdot \widetilde{\mathbf{x}}$$

The MiniMax Entropy method steps

Minimum Entropy STEP: We seek for the local minima of the Entropic measure (SE or GM) as a function of the separating parameters (θ_0, θ_1) . These set of parameter are associated with Minimum Entropy sources (SYN and DUST). See Figure.

Maximum Entropy STEP: We seek for the maximum of the Entropic measure (SE or GM) which is associated with the only Gaussian source (CMB). See Figure.



Using uncorrelateness for enhancing CMB estimate

After the local minima were identified (vectors \mathbf{d}_1 and \mathbf{d}_2 corresponding to SYN and DUST) we can determine the vector \mathbf{d}_0 (CMB) by using the assumption A2 instead of using the Maximum Entropy step.



Experimental Results on simulated data Example of the Noiseless case (using Shannon Entropy)

We have synthetically generated the mixture from simulated CMB, SYN and DUST images (256x256 pixels).



CMB



SYN

DUST



Estimated CMB

SIR = 13.6 dB



Estimated SYN

SIR = 31.9 dB



SIR = 21.4 dB

Mixture 0Mixture 1Mixture 2Mixture 3Mixture 0Mixture 1Mixture 2Mixture 3Correlations: $CMB - SYN \rightarrow E[s_0s_1] = -0.012$
 $CMB - DUST \rightarrow E[s_0s_2] = +0.149$
 $SYN - DUST \rightarrow E[s_1s_2] = -0.373$

Experimental Results on simulated data Comparison with FastICA

The following table presents the results of applying our method (with SE and GM as entropic measures) together with the results of FastICA for a set of 15 patches.

TABLE 1: Results for DCA with GM and SE entropic measures and FastICA. SIR levels under 8dB are highlighted. The percentageof images sucesfully separated (SIR > 8dB) are: 91%, 93% and 71% for GM, SE and FastICA respectively.

	SIR levels (dB) with GM				SIR levels (dB) with SE				SIR levels (dB) with FastICA			
Patch	CMB	SYN	DUST		CMB	SYN	DUST		СМВ	SYN	DUST	
1	15.8	10.0	13.9		15.2	13.2	15.6		22.9	13.0	10.2	
2	12.0	7.1	22.1		13.6	31.9	21.4		8.3	22.1	6.4	
3	23.3	2.4	17.3		23.1	2.2	17.4		25.4	13.6	9.8	
4	9.8	15.4	8.8		10.3	13.8	9.8		8.7	6.9	2.9	
5	12.5	15.2	20.6		19.7	23.6	20.1		17.5	17.9	22.8	
6	14.9	11.5	14.5		18.4	13.0	16.1		20.8	14.8	10.7	
7	16.5	11.7	9.6		16.0	14.5	10.3		16.8	15.6	3.0	
8	42.3	6.4	15.6		52.1	6.7	15.3		-1.9	5.6	7.6	
9	7.8	14.7	8.6		11.8	24.1	9.3		5.0	2.6	1.5	
10	14.2	13.3	13.8		20.8	22.2	19.4		19.8	16.9	27.0	
11	15.8	10.1	14.0		15.3	13.2	14.7		22.7	13.5	10.3	
12	20.7	26.6	17.8		21.0	22.5	24.4		28.2	19.1	9.6	
13	8.7	11.4	17.4		10.6	6.2	12.6		4.1	1.0	4.8	
14	11.9	18.6	21.3		13.4	23.6	16.9		13.2	8.5	15.3	
15	12.4	15.2	21.0		19.6	23.8	19.9		17.5	17.9	22.8	
Mean	15.9	12.6	15.7		18.7	17.0	16.2		15.3	12.6	11.0	

Robustness against noise

We have analyzed the sensitivity of the separation matrix estimation to Gaussian noise. As the level of noise is increased the Shannon Entropy (and the Gaussianity Measure) surfaces tends to be flatter and local extrema are more difficult to be detected.



Shannon Entropy 2D-contour plots for different levels of SNR (infinity, 40dB and 20dB)

Conclusions

• Shannon Entropy (SE) and Gaussianity Measure (GM) have proved to be useful for separating dependent sources.

• A new algorithm based on these Entropic Measures was developed for the separation of potentially dependent astrophysical sources showing better performance than the classical ICA approach (FastICA).

• Our technique was demonstrated to be reasonably robust to low level additive Gaussian noise.

Discussion about future directions

- The theoretical basement for Minimum Entropy methods is an open issue for dependent source case.
- An extension to a noisy model should be investigated. The present technique provides an estimation of the separating matrix but a non linear estimator should be developed for recovering sources.
- Separation of other source of radiation in astrophysical images need to be investigated.
- This technique should be tested also for the separation of sources from real mixtures (when available).