“Blind Source Separation in Astrophysical Images using Entropy related measures”

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)
Blind Source Separation (BSS)  
General Statement of the problem

The seminal work on blind source separation was done by Jutten, Herault and Guerin (1991) [1]. During the last two decades, many algorithms for source separation were introduced, specially for the case of independent sources leading to the so called Independent Component Analysis (ICA) [2]. Generally speaking the purpose of BSS is to obtain the best estimates of P input signals ($s$) from their M observed linear mixtures ($x$).

The Linear Mixing Model:

$$x(t) = As(t) + n(t)$$

Sources signals are assumed with zero-mean and unit-variance. We consider here the over determined case (M >= P)

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

$$\hat{s} = A^\dagger x$$

Where $A^\dagger$ is the Moore-Penrose inverse matrix

**Note 1:** Usually, matrix A is unknown and a “Blind” technique is required.

**Note 2:** 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) [3]. He has shown that: if at most one source is Gaussian then ICA problem can be solved, he also explained the permutation indeterminacy, etc.
• Many algorithms were developed by by using the concept of contrast functions (objective functions to be minimized) mainly based on approximations to Mutual Information-MI measure, which is defined as follows through the Kullback-Leibler distance ([2], [4]):

\[ I(\hat{s}) = \int p(\hat{s}) \log \frac{p(\hat{s})}{\prod_i p(\hat{s}_i)} d\hat{s} \]

Note that, if all source estimates \( \hat{S}_i \) are independent, then \( p(\hat{s}) = \prod_i p(\hat{s}_i) \) and \( I(\hat{s}) = 0 \)

Existing ICA/BSS algorithms

By minimizing Mutual Information
• P. Comon algorithm (1994) [3];
• InfoMax (1995) by Sejnowski et al ([2], [4]);
• FastIca (1999) by Hyvärinen ([2], [4]);
• R. Boscolo algorithm (2004) [5];
• and many others ([2], [4]).
DCA (Dependent Component Analysis)
How can we separate Dependent Sources?

• Few algorithms for dependent sources were reported in the literature. Cichocki et al. (2000) [6] have approached the separation of acoustic signals by exploiting their time correlations. Bedini et al. (2005) [7] 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 (A. Hyvärinen [2], P. Comon [3]).

• As we have experimentally demonstrated in a recent paper (Caiafa et al. 2006, [8]), 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 with high degree of cross correlation (see examples of speech signals in Caiafa et al. (SPARS05 – 2005) [9], Hyperspectral images in Caiafa et. al (EUSIPCO06 - 2006) [10], and dependent signals taken from satellite images in Caiafa et al. – 2006 [8].
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 usually 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:

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

$$ H_{GM}(y) = -\frac{1}{N^2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \frac{1}{h^2} \Phi \left( \frac{y(j)-y(i)}{h^2/2} \right) + \frac{2}{N} \sum_{i=0}^{N-1} \frac{1}{\sqrt{h^2+1}} \Phi \left( \frac{y(i)}{\sqrt{h^2+1}} \right) - \frac{1}{2\sqrt{\pi}} $$

**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) ([8])

---

MEDYFINOL 2006 – Mar del Plata, Argentina, September 4-8, 2006

6
The astrophysical problem
The Planck Surveyor Satellite mission

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: \( \hat{s} = Dx \)

• 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:

\[
\hat{s} = \tilde{D} \tilde{x} \quad \text{with} \quad \tilde{x} = \Lambda^{-\frac{1}{2}} V^T x \quad \text{(Karhunen Loeve Transformation)}
\]

• Covariance Matrices are:

\[
\begin{cases}
E[\tilde{x}\tilde{x}^T] = R_{\tilde{x}\tilde{x}} = I \\
E[\tilde{s}\tilde{s}^T] = R_{\tilde{s}\tilde{s}} = \tilde{D}\tilde{D}^T
\end{cases}
\]

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

\( \tilde{d}(\theta_0, \theta_1) = [\sin(\theta_0) \cos(\theta_1) \quad \sin(\theta_0) \sin(\theta_1) \quad \cos(\theta_0)]^T \)

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

\( \hat{s}_i = \tilde{d}_i^T (\theta_0^i, \theta_1^i) \cdot \tilde{x} \)
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 \((\theta_0, \theta_1)\). This set of parameters are associated with Minimum Entropy sources (SYN and DUST). See Figure below.

**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.
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).

![Images of CMB, SYN, DUST, Estimated CMB, Estimated SYN, Estimated DUST with their respective SIR values: SIR = 13.6 dB for Estimated CMB, SIR = 31.9 dB for Estimated SYN, SIR = 21.4 dB for Estimated DUST.]

Correlations:
\[
\begin{align*}
\text{CMB - SYN} & \rightarrow E[s_0 s_1] = -0.012 \\
\text{CMB - DUST} & \rightarrow E[s_0 s_2] = +0.149 \\
\text{SYN - DUST} & \rightarrow E[s_1 s_2] = -0.373
\end{align*}
\]

Note: We are indebted to the Planck teams in Bologna and Trieste, Italy, for supplying us with the maps we used for our simulations.
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 percentage of images successfully separated (SIR > 8dB) are: 91%, 93% and 71% for GM, SE and FastICA respectively.

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

References