Independent Component Analysis

MVA Project – One page draft

Nathan de Lara   Florian Tilquin   Vincent Vidal

Abstract

We decided to work on the Independent Component Analysis problem. We intent to implement, apply and compare several algorithms while being certain to understand the link between the likelihood maximisation and the mutual information.

1 Definition of the problem

The general problem can be formalised this way. Suppose we have some random variables $x \in \mathbb{R}^p$ which correspond to a mix of some primitive sources $s \in \mathbb{R}^n$. The aim is to extract from x every source s_i. To do so, we will suppose here that:
– the sources are independents
– the mix is linear and instantaneous
– at most one source has a gaussian distribution.
We will write:

$$x = As \quad \text{and} \quad y = Wx, \tag{1}$$

where A is the mixing matrix, W the separation matrix and y the estimation of the sources. The goal is then to find the matrix W that maximise the independence of y. As a measure of independence, we will consider, for theoretical purpose, the mutual information:

$$\mathrm{I}(Y) = \int_{\mathbb{R}^p} P(Y) \log \frac{P(Y)}{\Pi_i P_i(Y_i)} \mathrm{d}Y. \tag{2}$$

However, as it is too hard to compute, we will consider other contrast functions, invariant by permutation, scaling on coordinates and maximal for independant ones.

2 Algorithms for ICA

2.1 Hérault and Jutten (HJ) algorithm

This method is based on the neural network principle. We write $W = \left(I_n + \widetilde{W}\right)^{-1}$ and for a pair of given functions $(f,\, g)$, we adapt the $\widetilde{W}$ as follows:

$$\widetilde{W}_{ij} = f(y_i)g(y_j). \tag{3}$$

2.2 EASI algorithm

For a given contrast function ψ and a cost function J_ψ, we iterate as follows:

$$W_{t+1} = (I_n - \lambda_t \nabla J_\psi(y_t))\, W_t. \tag{4}$$

The EASI algorithm correspond to the functions $\psi(y) = \sum |y_i|^4$ and $J_\psi(y) = \mathbb{E}[\, \psi(y)\,]$.

2.3 Jade algorithm

Several methods are based on the cumulants. The aim here is to annul all the cross cumulants of order 4. Thus, we diagonalize the cumulant tensor which is equivalent to minimise the following contrast function:

$$c\,(x) = \sum_{i,k,l} |\mathrm{Cum}\,(x_i, x_i^*, x_k, x_l)|^2\,. \tag{5}$$

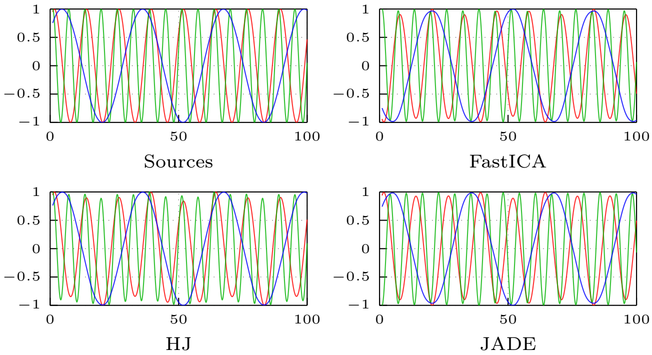

Figure 1: Results of the main ICA algorithms on simulated data. Each colour corresponds to a signal. The "Sources" graph shows the unmixed signals and the other ones the results of the algorithms.

2.4 FastICA algorithm

The FastICA algorithm is based on the information theory. We want here to maximise the marginal non-gaussianity on the whitened data, relying on a non linear quadratic function f with the following rule:

$$\widetilde{W}_{t+1} = \mathbb{E}[\, X.f(W_t^\mathsf{T} X)^\mathsf{T}\,] - \mathbb{E}[\, f''(W_t^\mathsf{T} X)\,]\, W_t, \tag{6}$$

with W_t the normalise vector of $\widetilde{W}_t$. In our case, we will use $f(x) = \frac{x^4}{4}$. We may use $f(x) = \log\cosh x$ or $f(x) = \exp\left(-\frac{x^2}{2}\right)$ too.

3 Results

First, we decided to apply the main algorithm to simple simulated data. For three sinusoidal signals of arbitrary frequencies and phases, we obtain the results in figure 1, which are good for all methods. We only have to remark that FastICA doesn't retrieve the sign of the signal, which is irrelevant.

4 What's next

We intent now to implement our own version of some of the algorithms and apply it to different types of data such as electroencephalogram, financial data and images. We will also look into more recent methods such as kernel ICA.