Independant Component Analysis

ENS CACHAN

Nathan de Lara, Florian Tilquin et Vincent Vidal

université PARIS-SACLAY

Master Mathématiques, Vision et Apprentissage

Problem statement

Let $x \in \mathbb{R}^p$ be some random variables whose components correspond to different mix of some primitive sources $s_i \in \mathbb{R}^n$. The aim is to retrieve an estimation y of every source s_i, given only x. We note A the mixing matrix and W the separation matrix such that:

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

We suppose that:

- the sources are independents
- the mix is linear and instantaneous
- at most one source has a Gaussian distribution.

Measure of independance

We want to find W that maximises the independence of $y = Wx$. The mutual information, defined in 2 is theoretically the best measure of independence:

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

but the information theory gives us other quantity to work with.
(... parler de la gaussianité/correlation/... ...)
We can show that:

$$\mathrm{I}(Y) + \sum_i \mathrm{G}(Y_i) = \mathrm{G}(Y) + \mathrm{C}(Y)\,. \tag{3}$$

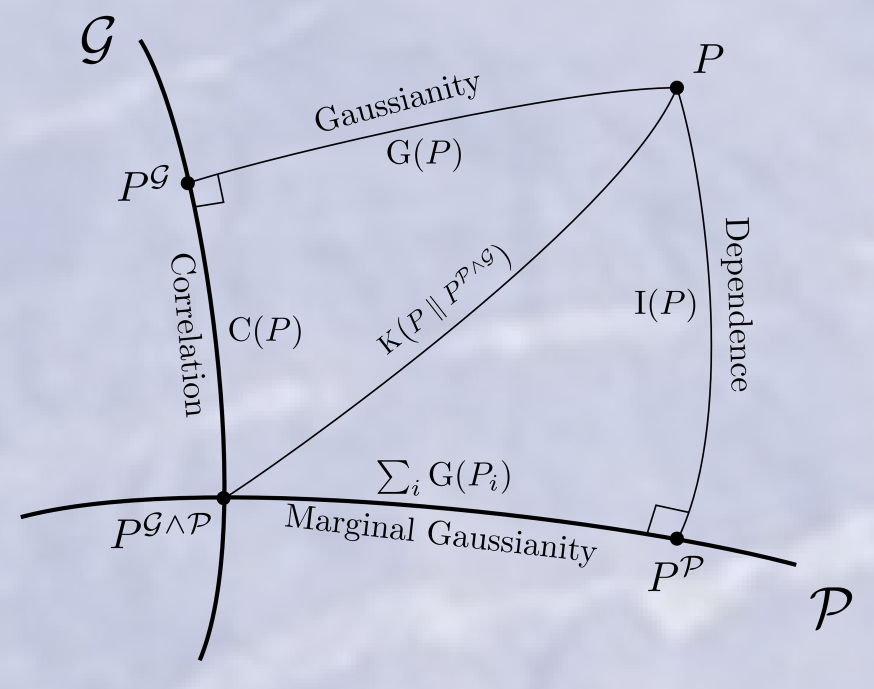

Figure: This figure illustrates equation 3. !!!Vince explique un peu les axes etc.!!!

Algorithms

Several algorithms have been developed to perform ICA, among those we can cite:

- HJ
- JADE
- FastICA
- KernelICA

Performance evaluation

One way to evaluate the performance of an ICA algorithm is to look at the estimated mixing matrix. However, the solution couple $(A, s)$of equation 1 is only defined up to a scaling factor and a permutation of the components. The "Amari distance" (which is not an actual distance), gives a criterion of proximity between two matrices regardless to these two issues and is then a convenient metrics to evaluate the performance of an algorithm. If U and V are two n-by-n matrices, the Amari distance is defined by:

$$d(U,V) = \frac{1}{2n}\sum_{i=1}^{n}\left(\frac{\sum_{j=1}^{n}|a_{ij}|}{\max_j |a_{ij}|} - 1\right) + \frac{1}{2n}\sum_{j=1}^{n}\left(\frac{\sum_{i=1}^{n}|a_{ij}|}{\max_i |a_{ij}|} - 1\right) \tag{4}$$

with $a_{ij} = (UV^{-1})_{ij}$. The closest it is to 0, the more U and V represent the same components.

Kernel ICA

The idea here is to have a contrast function defined by the correlation of a range of functions over a function vector space $\mathcal{F}$. Thus defining the $\mathcal{F}$-correlation :

$$\rho_{\mathcal{F}}(x_1, x_2) = \max_{f_1, f_2 \in \mathcal{F}} \operatorname{corr}(f_1(x_1), f_2(x_2))$$

For $\mathcal{F}$ "large enough", we have the following equivalence : $\rho_{\mathcal{F}}(x_1, x_2) = 0 \iff x_1$ & x_2 are independent, hence the "vector space contrast function" idea. We then use the *reproducing kernel Hilbert space* to obtain tractable computations over the vector space $\mathcal{F}$.

Modus operandi

We will thereby analyze the results of the algorithms in terms of Amari distance over large number of i.i.d generated sources from several different distributions.

Results

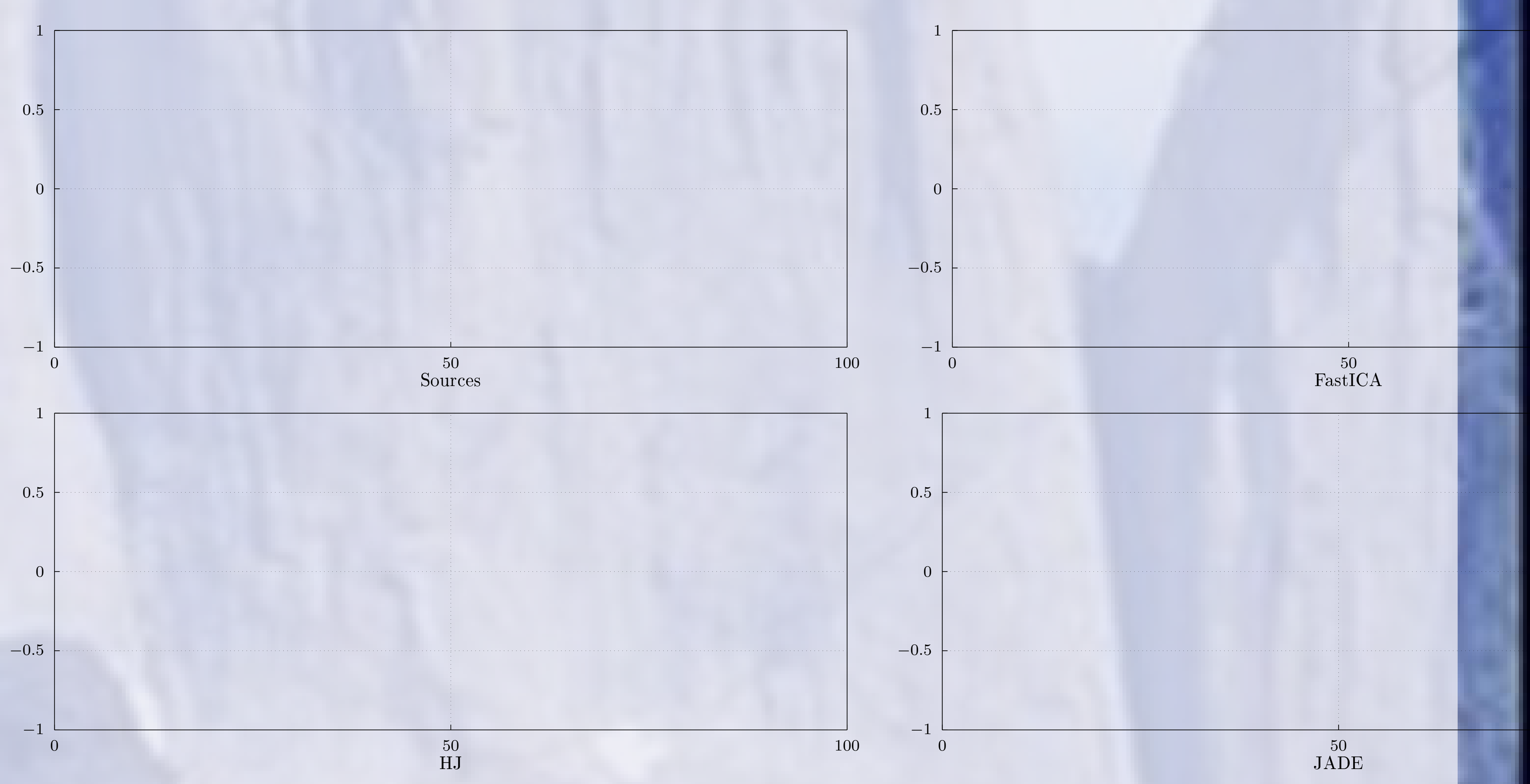

Figure: Results of the main ICA algorithms on simulated data. Each colour corresponds to a signal. The "Sources" graph shows the unmixed signal s and the other ones the results of the algorithms.