Independant Componant Analysis

Nathan de Lara, Florian Tilquin et Vincent Vidal

Master Mathématiques, Vision et Apprentissage, ENS Cachan

January 2, 2016

Definition of the problem

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 retrieve a estimation y of every source s_i, given only x. We not A the mixing matrix and W the separation matrix:

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

We suppose :

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

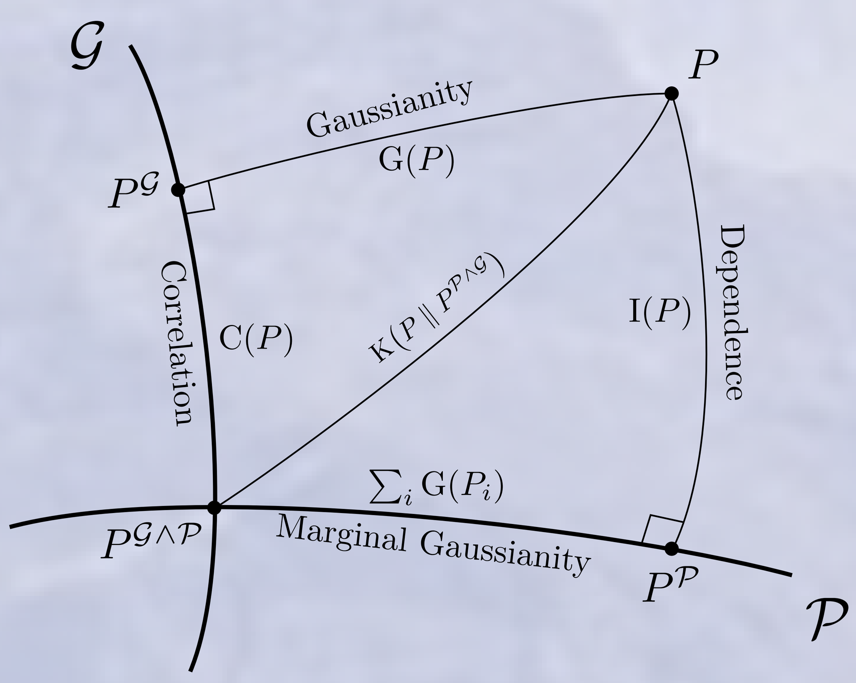
Measure of independance

We want to find W that maximise the independance of $y = Wx$.
The mutual information is theoretically the best measure:

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

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

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



Results

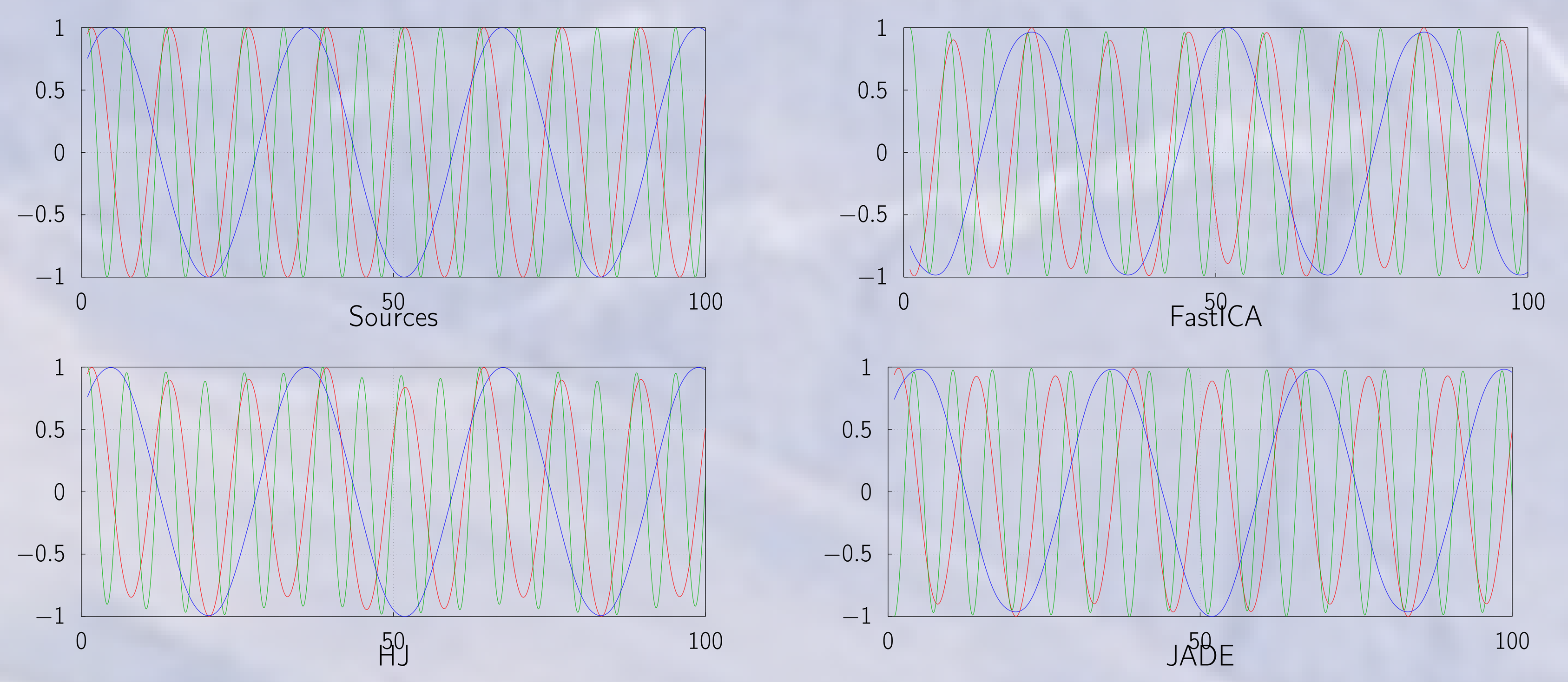


Figure: Results of the main ICA algorithms on simulated data. Each colour corresponds to a signal. The "Sources" graph shows the unmixed signal s and the other ones the results of the algorithms.

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 independant, hence the "vector space contrast function" idea. We then use the *reproducing kernel Hilbert space* to obtain tractable computations over the vector space $\mathcal{F}$.

Methods benchmark

We now wish to compare the different algorithms performing ICA we kept :

- HJ
- JADE
- FastICA
- KernelICA

One way to compare these algorithms is by their estimated mixing (ou unmixig) matrices : given numerous generated i.i.d. sources and a mixing matrix, all of these algorithms will return the best estimated mixing matrix according to their

Amari distance

Fortunately, the "Amari distance" (which is unfortunately not a distance), gives a criterion of proximity between two matrices regardless to these two problems (amplitude and permutation). If U and V are two n-by-n matrices, the Amari distance is defined by:

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

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

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.

Introduction

- some items and $\alpha = \gamma, \sum_i$
- some items
- some items
- some items

$\alpha = \gamma, \sum_i$

Introduction

- some items and $\alpha = \gamma, \sum_i$
- some items
- some items
- some items

$\alpha = \gamma, \sum_i$

Introduction

- some items and $\alpha = \gamma, \sum_i$
- some items
- some items
- some items

$\alpha = \gamma, \sum_i$

Introduction

- some items and $\alpha = \gamma, \sum_i$
- some items
- some items
- some items

$\alpha = \gamma, \sum_i$

Introduction

- some items
- some items
- some items
- some items

Introduction

- some items and $\alpha = \gamma, \sum_i$
- some items
- some items
- some items

$\alpha = \gamma, \sum_i$

Introduction

- some items and $\alpha = \gamma, \sum_i$
- some items
- some items
- some items

$\alpha = \gamma, \sum_i$