

Blind source separation using ICA

Guillaume Ausset

9 December 2016

All my code, report, poster and any other material will be tracked with Git. The repository is at <https://gitlab.com/aussetg/pgm-ica> but is private before the presentation, if you want access just ask me at guillaume@ausset.me and I will gladly add you.

1 Introduction of the Problem

It is quite common when measuring several signals at once to be unable a priori to isolate the different source signals and only be able to observe a mixture of those sources. When recording sound the signal recorded will be a mixture of different ambient sounds and the sound of interest, when acquiring EEGs eye movement and muscle twitching will be mixed with the main signal. One finds themselves before a problem of the form:

$$\mathbf{x} = A\mathbf{s}$$

Where \mathbf{x} are our observations, \mathbf{s} our unknown signal of origin and A the mixing matrix. The problem seems intractable but we will see that only making the hypothesis that the s_i are independent (and eventually some more simplifying hypothesis) make it tractable.

The problem then becomes:

- Find a measure of independence, or contrast function.
- Maximize that measure with respect to A .

Therefore for each measure of independence chosen we find a new method to solve ICA.

2 Methods that will be implemented

2.1 Fast-ICA

The performance of a given ICA method will depend on the quality of the measure of independence and this measure will also impact the computational complexity of the algorithm. We will therefore try to compare different methods.

The canonical method because of its extreme simplicity and speed is Fast-ICA [3], which chose as a contrast function the negentropy:

$$J(y) = H(y_{\text{gauss}}) - H(y)$$

with $H(y)$ the differential entropy, y_{gauss} a Gaussian variable with the same variance as y . And they give as an approximation if we consider that y is standardized:

$$J(y) \simeq [\mathbb{E}[G(y)] - \mathbb{E}[G(\nu)]]^2$$

with $\nu \sim \mathcal{N}(0, 1)$ and G a non-linear and non-quadratic function. This formulation enables the authors to give a very efficient algorithm to maximize J that I have implemented in Julia. See appendix for figures illustrating the different steps of FastICA on completely artificial data.

2.2 Kernel-ICA

The idea of [1] is to optimize the contrast function:

$$\rho_{\mathcal{F}} = \max_{f_1, f_2 \in \mathcal{F}} \frac{\text{cov}(f_1(x_1), f_2(x_2))}{(\text{var } f_1(x_1))^{1/2} (\text{var } f_2(x_2))^{1/2}}$$

Or more precisely a penalized version of that contrast function. Directly computation of this function seems intractable especially if \mathcal{F} is big enough, the authors therefore chose \mathcal{F} as a reproducing kernel Hilbert space in order to exploit the kernel "trick" so as to transport the problem from the space \mathcal{F} to \mathcal{X} the feature space:

$$\text{corr}(f_1(x_1), f_2(x_2)) = \text{corr}(\langle \Phi(x_1), f_1 \rangle, \langle \Phi(x_2), f_2 \rangle)$$

where Φ is the map associated to the kernel of the space and the problem is then a form of Kernel CCA which is tractable.

Kernel-ICA will be implemented in Julia and contrasted with Fast-ICA.

3 Possible extension

While I have not yet entirely read the papers and will concentrate on implementing Kernel-ICA first, I hope to be able to compare the two previous methods to a variational approach of ICA as in [4, 5, 2].

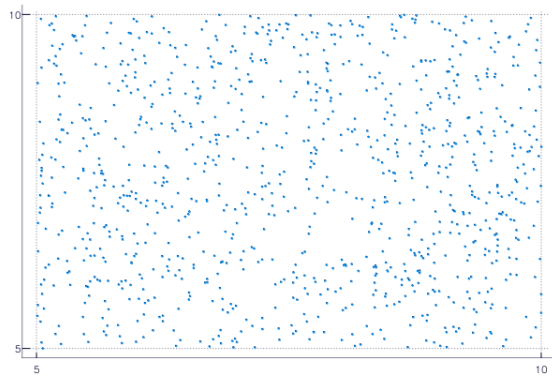
4 Results

We generate points uniformly on $[0, 10]^2$ and then mix them with:

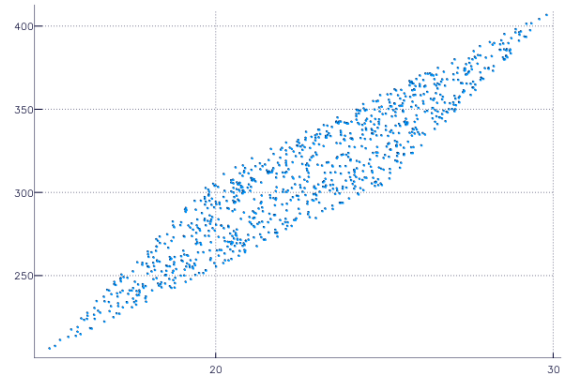
$$\begin{pmatrix} 1 & 2 \\ 21 & 20 \end{pmatrix}$$

And then use FastICA to retrieve the original components.

Note that it is completely normal that the retrieved components have different amplitudes, it's impossible to retrieve the amplitude and the sign. I, however, forgot to add back the mean.

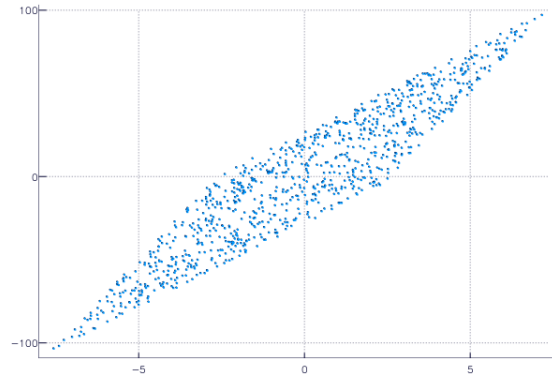


(a) s original

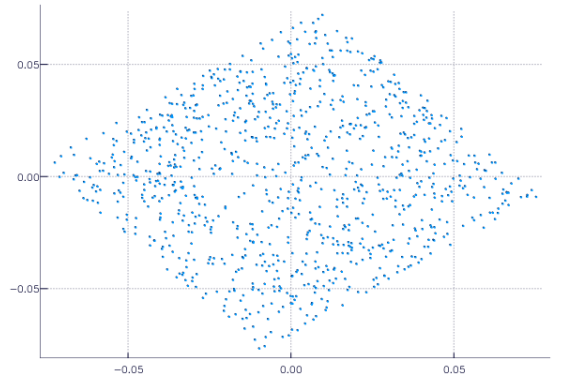


(b) $x = As$

Figure 1: Independent components and their mixture



(a) x centered



(b) x whitened

Figure 2: Preprocessing

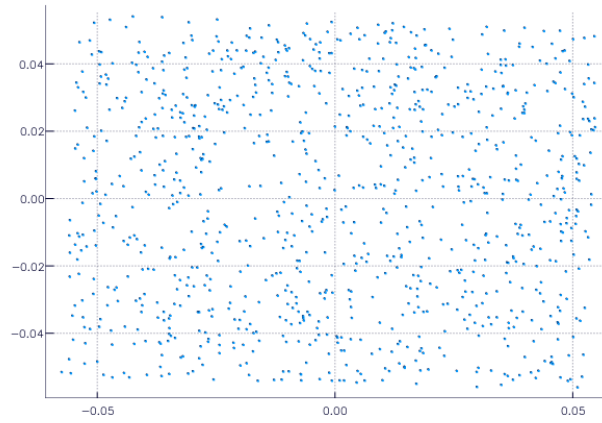


Figure 3: Final retrieved s

References

- [1] Francis R Bach and Michael I Jordan. “Kernel Independent Component Analysis”. In: *Journal of Machine Learning Research* 3 (2002), pp. 1–48. ISSN: 0003-6951. DOI: 10.1162/153244303768966085 (cit. on p. 2).
- [2] Rizwan A Choudrey. “Variational Methods for Bayesian Independent Component Analysis”. PhD thesis. 2002 (cit. on p. 2).
- [3] Aapo Hyvärinen and Erkki Oja. “Independent component analysis: algorithms and applications”. In: *Neural networks* 13.4-5 (2000), pp. 411–430. ISSN: 0893-6080. DOI: 10.1016/S0893-6080(00)00026-5. URL: <http://www.sciencedirect.com/science/article/pii/S0893608000000265> <http://www.ncbi.nlm.nih.gov/pubmed/10946390> (cit. on p. 1).
- [4] Neil D Lawrence and Christopher M Bishop. “Variational Bayesian independent component analysis”. In: *Univ of Cambridge Tech Report* 2 (2000) (cit. on p. 2).
- [5] Ole Winther and Kaare Brandt Petersen. “Bayesian independent component analysis: Variational methods and non-negative decompositions”. In: *Digital Signal Processing: A Review Journal* 17.5 (2007), pp. 858–872. ISSN: 10512004. DOI: 10.1016/j.dsp.2007.01.003 (cit. on p. 2).