Exercises* on Independent Component Analysis

Laurenz Wiskott
Institut für Neuroinformatik
Ruhr-Universität Bochum, Germany, EU

4 February 2017

Contents

Several of my exercises (not necessarily on this topic) were inspired by papers and textbooks by other authors. Unfortunately, I did not document that well, because initially I did not intend to make the exercises publicly available, and now I cannot trace it back anymore. So I cannot give as much credit as I would like to. The concrete versions of the exercises are certainly my own work, though.

In cases where I reuse an exercise in different variants, references may be wrong for technical reasons.

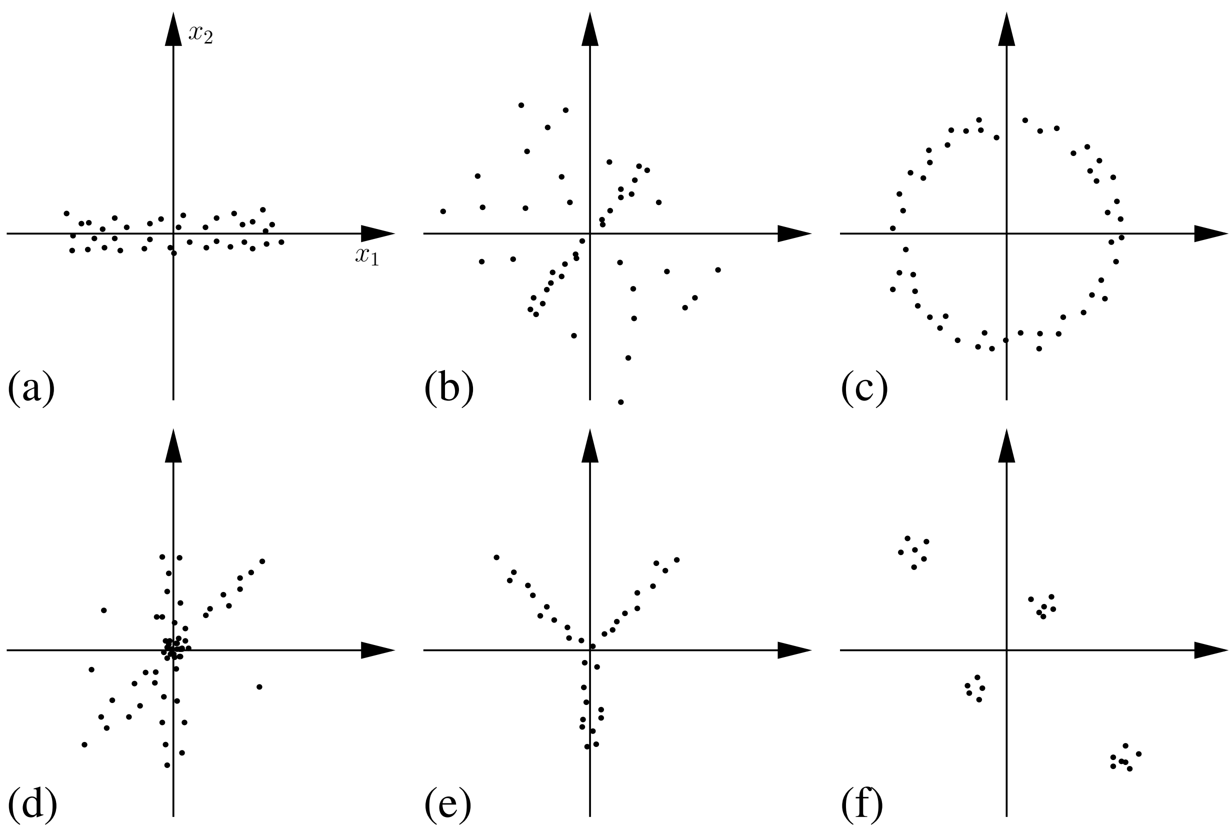
*These exercises complement my corresponding lecture notes available at https://www.ini.rub.de/PEOPLE/wiskott/Teaching/Material/, where you can also find other teaching material such as programming exercises. The table of contents of the lecture notes is reproduced here to give an orientation when the exercises can be reasonably solved. For best learning effect I recommend to first seriously try to solve the exercises yourself before looking into the solutions.

1 Intuition

1.1 Mixing and unmixing

1.1.1 Exercise: Guess independent components and distributions from data

Decide whether the following distributions can be linearly separated into independent components. If yes, sketch the (not necessarily orthogonal) axes onto which the data must be projected to extract the independent components. Draw on these axes also the marginal distributions of the corresponding components.



1.2 How to find the unmixing matrix?

1.3 Sources can only be recovered up to permutation and rescaling

1.4 Whiten the data first

1.5 A generic ICA algorithm

2 Formalism based on cumulants

2.1 Moments and cumulants

2.1.1 Exercise: Moments and cumulants

Mixed cumulants κ can be written as sums of products of mixed moments and *vice versa*. More intuitive is the latter. Since cumulants represent a distribution only in exactly one order, since the lower order moments have been subtracted off, it is easy to imagine that a moment can be written by summing over all possible combinations of cumulants that add up exactly to the order of the moment, for instance

$$\begin{aligned}\langle X_1 X_2 X_3 \rangle &= \kappa(X_1, X_2, X_3) \\ &\quad + \kappa(X_1, X_2)\kappa(X_3) + \kappa(X_1, X_3)\kappa(X_2) + \kappa(X_2, X_3)\kappa(X_1) \\ &\quad + \kappa(X_1)\kappa(X_2)\kappa(X_3)\,. \end{aligned} \tag{1}$$

The general rule is

$$\langle X_1 \cdots X_N \rangle = \sum_{\pi} \prod_{B \in \pi} \kappa(X_i : i \in B)\,, \tag{2}$$

with π going through the list of all possible partitions of the N variables into disjoint blocks and B indicating the blocks within one partition.

Hint: You cannot assume zero-mean data here.

Hint: In the following it is convenient to write M_{123} and C_{123} etc. instead of $\langle X_1 X_2 X_3 \rangle$ and $\kappa(X_1, X_2, X_3)$ etc.

1. Write with the help of equation (2) all mixed moments up to order four as a sum of products of cumulants, like in equation (1).
2. From the equations found in part 1 derive expressions for the cumulants up to order three written as sums of products of moments.

2.1.2 Exercise: Moments and cumulants of concrete distributions

1. What can you say in general about moments and cumulants of symmetric distributions (even functions) of a single variable?
2. Calculate all moments up to order ten and all cumulants up to order four for the following distributions.

   Hint: First derive a closed form or a recursive formula for $\langle x^n \rangle$ and then insert the different values for n.

(a) Uniform distribution (D: Gleichverteilung):

$$p(x) := \begin{cases} 1/2 & \text{if } x \in [-1, +1] \\ 0 & \text{otherwise} \end{cases} . \tag{1}$$

(b) Laplace- or double exponential distribution (D: doppelt exponentielle Verteilung):

$$p(x) := \frac{\exp(-|x|)}{2} . \tag{2}$$

2.1.3 Exercise: Moments and cumulants of scaled distributions

Assume all moments and cumulants of a scalar random variable X are given. How do moments and cumulants change if the variable is scaled by a factor s around the origin, i.e. simply multiplied by s ? Prove your result.

Hint: Cumulants can always be written as a sum of products of moments of identical overall order.

2.1.4 Exercise: Moments and cumulants of shifted distributions

Assume all moments and cumulants of a (non zero-mean) scalar random variable X are known. How do the first four moments and the first three cumulants change if the variable is shifted by m?

Hint: See exercise 2.1.1 for the cumulants of non-zero-mean data.

2.1.5 Exercise: Kurtosis is additive

Kurtosis for a zero-mean random variable x is defined as

$$\mathrm{kurt}(x) := \langle x^4 \rangle - 3\langle x^2 \rangle^2 . \tag{1}$$

Show that for two statistically independent zero-mean random variables x und y

$$\mathrm{kurt}(x+y) = \mathrm{kurt}(x) + \mathrm{kurt}(y) \tag{2}$$

holds. Make clear where you use which argument for simplifications.

2.1.6 Exercise: Moments and cumulants are multilinear

1. Show that cross-moments, such as $\langle x_1 x_2 x_3 \rangle$, are multilinear, i.e. linear in each of their arguments, e.g.

   $$\langle (ax_1 + bx'_1) x_2 x_3 ... \rangle \quad = \quad a\langle x_1 x_2 x_3 ... \rangle + b\langle x'_1 x_2 x_3 ... \rangle \tag{1}$$

   with a and b being constants and x'_1 being another random variable.

2. Show that cross-cumulants, such as $\kappa(x_1, x_2, x_3)$, are multilinear, i.e. linear in each of their arguments.

3. As we have seen in exercise 2.1.5, the kurtosis is additive for two statistically independent zero-mean random variables x und y, i.e.

   $$\mathrm{kurt}(x+y) = \mathrm{kurt}(x) + \mathrm{kurt}(y) . \tag{2}$$

   Why is statistical independence required for the additivity of kurtosis of the signals while it is not for the multilinearity of cross-cumulants.

2.1.7 Exercise: Mixing statistically independent sources

Given some scalar and statistically independent random variables (signals) s_i with zero mean, unit variance, and a value a_i for the kurtosis that lies between $-a$ and $+a$, with arbitrary but fixed value of $0 < a$. The s_i shall be mixed like

$$x := \sum_i w_i s_i \tag{1}$$

with constant weights w_i.

1. Which constraints do you have to impose on the weights w_i to guarantee that the mixture has unit variance as well?
2. Prove that the kurtosis of an equally weighted mixture ($w_i = w_j \forall i, j$) of N signals converges to zero as N goes to infinity.

   Hints: (i) For the kurtosis and two statistically independent random variables s_1 and s_2

   $$\mathrm{kurt}(s_1 + s_2) = \mathrm{kurt}(s_1) + \mathrm{kurt}(s_2) \tag{2}$$

   holds. (ii) Use the constraint from part 1.

2.2 Cross-cumulants of statistically independent components are zero

2.3 Components with zero cross-cumulants are statistically independent

2.4 Rotated cumulants

2.5 Contrast function

2.6 Givens-rotations

2.7 Optimizing the contrast function

2.8 The algorithm