ICA of complex valued signals: a fast and robust deflationary algorithm

Ella Bingham and Aapo Hyvärinen
Neural Networks Research Centre, Helsinki University of Technology
P.O. Box 5400, FIN-02015 HUT, Finland
Ella.Bingham@hut.fi, Aapo.Hyvarinen@hut.fi

Abstract

Separation of complex valued signals is a frequently arising problem in signal processing. In this article it is assumed that the original, complex valued source signals are mutually statistically independent, and the problem is solved by the independent component analysis (ICA) model. ICA is a statistical method for transforming an observed multidimensional random vector into components that are mutually as independent as possible. In this article, a fast fixed-point type algorithm that is capable of separating complex valued, linearly mixed source signals is presented and its computational efficiency is shown by simulations. We also present a theorem on the local consistency of the estimator given by the algorithm.
Keywords: Independent Component Analysis, Complex valued signals, Deflationary separation

1 Introduction

Separation of complex valued signals is a frequently arising problem in signal processing: frequency-domain implementations involving complex valued signals have advantages over time-domain implementations. Especially in the separation of convolutive mixtures it is a common practice to Fourier transform the signals, which results in complex valued signals.

In this article we present an algorithm for the separation of complex valued signals. Our framework is Independent Component Analysis [2], [6], and the ICA model used in this paper is

$$\mathbf{x} = \mathbf{A}\mathbf{s} \tag{1}$$

where $\mathbf{x} = (x_1, \ldots, x_m)$ is the vector of observed signals, $\mathbf{s} = (s_1, \ldots, s_n)$ is the vector of statistically independent latent variables called the independent components, and $\mathbf{A}$ is an unknown constant mixing matrix. The above model is identifiable under the following fundamental restrictions [2]: at most one of the independent components s_j may be Gaussian, and the matrix $\mathbf{A}$ must be of full column rank.

A fast fixed point algorithm (FastICA) for the separation of linearly mixed independent source signals was presented by Hyvärinen and Oja [3], [4]. In this paper we show how the FastICA algorithm can be extended to complex valued signals. Both $\mathbf{s}$ and $\mathbf{x}$ in model (1) assume complex values. For simplicity, $n = m$. The mixing matrix $\mathbf{A}$ is of full rank and it may be complex as well, but this is optional. A necessary preprocessing of the data $\mathbf{x}$ is whitening, which can always be accomplished by e.g. Principal Component Analysis [2]. We assume that the signals s_j are white, i.e. they are zero mean and of unit variance and the real and imaginary parts of s_j are uncorrelated and of equal variance. The whitening can be equivalently expressed as [8] $E\{\mathbf{s}\mathbf{s}^H\} = \mathbf{I}$ and $E\{\mathbf{s}\mathbf{s}^T\} = \mathbf{O}$. This kind of whitening of s_j is quite realistic in practical problems, because when separating complex valued signals using ICA it is impossible to retain the phases of the signals — the independent components are found up to permutation, signs and scaling by a complex constant.

Algorithms for independent component analysis of complex valued signals are also presented in [1] and [7]. Both of these algorithms are computationally more intensive than our algorithm, and no theoretical justification on consistency is given. In contrast, we present a theorem on the local consistency of the estimator given by our algorithm, and show its computational efficiency by simulations. Our algorithm is also more robust against outliers than kurtosis-based ICA algorithms (see [3] for a discussion on robust estimators for ICA). Also, our algorithm is capable of deflationary separation of the independent component signals; it is possible to estimate only one or some of the independent components, which useful if the exact number of independent components is not known beforehand. In deflationary separation the components tend to separate in the order of decreasing non-Gaussianity, which often equals decreasing "importance" of the components.

2 Contrast function

Now we generalize the ICA framework in [3], [4] and [5] for complex valued signals. One might make a distinction between "top-down" and "bottom-up" approaches to ICA [5]. In the top-down approach, independence is measured by such measures as mutual information which is often approximated by using cumulants. This may result in non-robust contrast functions and burdensome computations. We choose here the bottom-up approach, where the higher order statistics are implicitly embedded into the algorithm by arbitrary nonlinearities. We start from an arbitrary nonlinear contrast function and show that its extrema coincide with the independent components. This bottom-up approach is computationally simple, and the nonlinearity can be chosen quite freely to optimize e.g. the statistical behavior of the estimator.

Our contrast function is

$$J_G(\mathbf{w}) = E\{G(|\mathbf{w}^H\mathbf{x}|^2)\} \tag{2}$$

where $G : \mathbb{C}^n \to \mathbb{R}$ is a smooth even function, $\mathbf{w}$ is an n-dimensional complex weight vector and $E\{|\mathbf{w}^H\mathbf{x}|^2\} = 1$.

It is highly preferable that the estimator given by the contrast function is robust against outliers. The more slowly G grows as its argument increases, the more robust is the estimator.

For the choice of G we propose now three different functions:

$$G_1(y) = \sqrt{a_1 + y}, \quad a_1 \approx 0.1 \tag{3}$$
$$G_2(y) = \log(a_2 + y), \quad a_2 \approx 0.1 \tag{4}$$
$$G_3(y) = \frac{1}{2}y^2. \tag{5}$$

Of the above functions, G_1 and G_2 grow more slowly than G_3 and thus they give more robust estimators. G_3 is motivated by kurtosis: for complex random variables, kurtosis may be defined e.g. as [8] $\mathrm{kurt}(y) = E\{|y|^4\} - 2(E\{|y|^2\})^2 - |E\{y^2\}|^2 = E\{|y|^4\} - 2$ where y is white.

In [5] it was stated that any nonlinear learning function G divides the space of probability distributions into two half-spaces. Independent components can be estimated by either maximizing or minimizing $E\{G(|\mathbf{w}^H\mathbf{x}|^2)\}$, depending on which half-space their distribution lies in. In [5] a theorem for real valued signals was presented that distinguished between maximization and minimization and gave the exact conditions for convergence. In the following we show how this idea can be generalized to complex valued random variables. We have the following theorem on the local consistency of the estimators, the proof of which is omitted due to space limitations:

Theorem. Assume that the input data follows the model (1). The observed variables x_k, $k = 1, \ldots, n$ in $\mathbf{x}$ are prewhitened using $E\{\mathbf{x}\mathbf{x}^H\} = \mathbf{I}$. The independent component variables s_k, $k = 1, \ldots, n$ in $\mathbf{s}$ are zero mean and have unit variances and uncorrelated real and imaginary parts of equal variances. Also, $G : \mathbb{C}^n \to \mathbb{R}$ is a sufficiently smooth even function. Then the local maxima (resp. minima) of $E\{G(|\mathbf{w}^H\mathbf{x}|^2)\}$ under the constraint $E\{|\mathbf{w}^H\mathbf{x}|^2\} = ||\mathbf{w}||^2 = 1$ include those rows $\mathbf{a}_k$ of the inverse of the mixing matrix $\mathbf{A}$ such that the corresponding independent components s_k satisfy

$$E\{g(|s_k|^2) + |s_k|^2 g'(|s_k|^2) - |s_k|^2 g(|s_k|^2)\} < 0 \quad (> 0, \text{ resp.}). \tag{6}$$

where $g()$ is the derivative of $G()$ and $g'()$ is the derivative of $g()$. The same is true for the points $-\mathbf{a}_k$. $\square$

3 Fixed-point algorithm

We shall derive the fixed-point algorithm for one unit. We assume that the source signals s_j are whitened, i.e. they are zero mean and have unit variances and uncorrelated real and imaginary parts of equal variances. This can be equivalently expressed as $E\{\mathbf{s}\mathbf{s}^H\} = \mathbf{I}$ and $E\{\mathbf{s}\mathbf{s}^T\} = \mathbf{O}$. The observed variable $\mathbf{x}$ is whitened so that it also obeys $E\{\mathbf{x}\mathbf{x}^H\} = \mathbf{I}$.

According to the Kuhn-Tucker conditions, the optima of the contrast function $E\{G(|\mathbf{w}^H\mathbf{x}|^2)\}$ under the constraint $E\{|\mathbf{w}^H\mathbf{x}|^2\} = ||\mathbf{w}||^2 = 1$ are obtained at points where

$$\nabla E\{G(|\mathbf{w}^H\mathbf{x}|^2)\} - \beta\nabla E\{|\mathbf{w}^H\mathbf{x}|^2\} = 0 \tag{7}$$

where $\beta \in \mathbb{R}$ and the gradient is computed with respect to real and imaginary parts of w separately. The Newton method is used to solve (7). The Jacobian matrix of $\nabla E\{G(|\mathbf{w}^H\mathbf{x}|^2)\}$ can be approximated as

$$\begin{aligned}\nabla^2 E\{G(|\mathbf{w}^H\mathbf{x}|^2)\} &= 2E\{(\nabla^2|\mathbf{w}^H\mathbf{x}|^2)g(|\mathbf{w}^H\mathbf{x}|^2) + 2(\nabla|\mathbf{w}^H\mathbf{x}|^2)(\nabla|\mathbf{w}^H\mathbf{x}|^2)^T g'(|\mathbf{w}^H\mathbf{x}|^2)\} \\ &\approx 2E\{g(|\mathbf{w}^H\mathbf{x}|^2) + |\mathbf{w}^H\mathbf{x}|^2 g'(|\mathbf{w}^H\mathbf{x}|^2)\}\mathbf{I} \qquad (8)\end{aligned}$$

where the approximation was done by separating the expectations. Also, $E\{\mathbf{x}\mathbf{x}^T\} = \mathbf{O}$ (which follows straightforward from the whitening $E\{\mathbf{s}\mathbf{s}^T\} = \mathbf{O}$) was used.

The Jacobian matrix of the other term in (7), $\beta\nabla E\{|\mathbf{w}^H\mathbf{x}|^2\}$ is $\beta\nabla^2 E\{|\mathbf{w}^H\mathbf{x}|^2\} = 2\beta\mathbf{I}$. The total approximative Jacobian of (7) is now

$$\mathbf{J} = 2\,(\,E\{g(|\mathbf{w}^H\mathbf{x}|^2) + |\mathbf{w}^H\mathbf{x}|^2 g'(|\mathbf{w}^H\mathbf{x}|^2)\} - \beta\,)\,\mathbf{I} \qquad (9)$$

which is diagonal and thus easy to invert. We obtain the following approximative Newton iteration:

$$\mathbf{w}^+ = \mathbf{w} - \frac{E\{\mathbf{x}(\mathbf{w}^H\mathbf{x})^* g(|\mathbf{w}^H\mathbf{x}|^2)\} - \beta\mathbf{w}}{E\{g(|\mathbf{w}^H\mathbf{x}|^2) + |\mathbf{w}^H\mathbf{x}|^2 g'(|\mathbf{w}^H\mathbf{x}|^2)\} - \beta} \qquad (10)$$

$$\mathbf{w}_{new} = \frac{\mathbf{w}^+}{||\mathbf{w}^+||} \qquad (11)$$

If we multiply both sides of (10) by $\beta - E\{g(|\mathbf{w}^H\mathbf{x}|^2) + |\mathbf{w}^H\mathbf{x}|^2 g'(|\mathbf{w}^H\mathbf{x}|^2)\}$, the fixed-point algorithm simplifies to

$$\mathbf{w}^+ = E\{\mathbf{x}(\mathbf{w}^H\mathbf{x})^* g(|\mathbf{w}^H\mathbf{x}|^2)\} - E\{g(|\mathbf{w}^H\mathbf{x}|^2) + |\mathbf{w}^H\mathbf{x}|^2 g'(|\mathbf{w}^H\mathbf{x}|^2)\}\mathbf{w} \qquad (12)$$

$$\mathbf{w}_{new} = \frac{\mathbf{w}^+}{||\mathbf{w}^+||} \qquad (13)$$

The one-unit algorithm can be extended to the estimation of the whole ICA transformation $\mathbf{s} = \mathbf{W}^H\mathbf{x}$. To prevent different neurons from converging to the same maxima, the outputs $\mathbf{w}_1^H\mathbf{x}, \ldots, \mathbf{w}_n^H\mathbf{x}$ are decorrelated after every iteration, using e.g. a Gram-Schmidt-like decorrelation [3], [4].

Sometimes it is preferable to to estimate all the independent components simultaneously, and use a symmetric decorrelation. This can be accomplished e.g. by

$$\mathbf{W} = \mathbf{W}(\mathbf{W}^H\mathbf{W})^{-1/2} \qquad (14)$$

where $\mathbf{W} = (\mathbf{w}_1 \cdots \mathbf{w}_n)$ is the matrix of the vectors.

4 Simulation results

Complex signals were separated to test the performance of the fast fixed-point algorithm and the Theorem. The data were artificially generated complex random signals $s_j = r_j(\cos\phi_j + i\sin\phi_j)$ where r_j were drawn from different distributions and the ϕ_j was uniformly distributed on $[0, 2\pi]$. There were a total of 8 complex random signals and 50 000 samples per signal at each trial. Each signal was normalized to unit variance before mixing with a randomly generated complex mixing matrix $\mathbf{A}$. The mixed signals were whitened before using the fixed point algorithm. A complex unmixing matrix $\mathbf{W}$ was sought so that $\mathbf{s} = \mathbf{W}^H\mathbf{x}$. The result of the separation can be measured by $\mathbf{W}^H\mathbf{A}$. It should converge to a matrix where in each row and each column there is one nonzero element $v \in \mathbb{C}$ of unit modulus; i.e. in the end, $|\mathbf{W}^H\mathbf{A}|$ should be a permutation matrix. Our error measure is the sum of squared deviation of $|\mathbf{W}^H\mathbf{A}|$ from a permutation matrix. All three contrast functions (3)–(5) were successful in that the Theorem was always fulfilled and $|\mathbf{W}^H\mathbf{A}|$ converged to a permutation matrix in about 6 steps. Figure 1 shows the convergence using G_2.

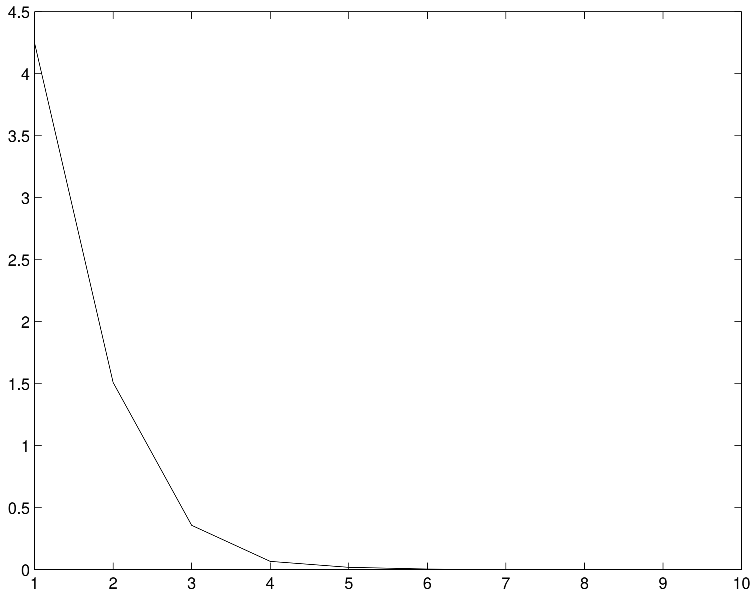

Figure 1: Convergence of the fixed-point algorithm using contrast function $G_2(y) = \log(a_2 + y)$; average result over 10 runs. About 6 iteration steps were needed for convergence.

5 Conclusion

We have presented a fixed-point type algorithm for the separation of linearly mixed, complex valued signals in the ICA framework. Our algorithm is based on a deflationary separation of independent components. The algorithm is robust and computationally simple, and the estimator given by the algorithm is locally consistent. We have also shown the computational efficiency of the algorithm by simulations.

References

[1] A. D. Back and A. C. Tsoi. Blind deconvolution of signals using a complex recurrent network. In *Neural Networks for Signal Processing 4, Proceedings of the 1994 IEEE Workshop*,

J. Vlontzos, J. Hwang and E. Wilson (Eds), pp. 565-574, IEEE Press, 1994.

[2] P. Comon. Independent Component Analysis — a new concept? *Signal Processing*, 36:287-314, 1994.

[3] A. Hyvärinen. Fast and Robust Fixed-Point Algorithms for Independent Component Analysis. *IEEE Transactions on Neural Networks*, Vol. 10, No. 3, pp. 626-634, May 1999.

[4] A. Hyvärinen, E. Oja. A Fast Fixed-Point Algorithm for Independent Component Analysis. *Neural Computation*, 9:1483–1492, 1997.

[5] A. Hyvärinen, E. Oja. Independent component analysis by general nonlinear Hebbian-like learning rules. *Signal Processing*, 64:301–313, 1998.

[6] C. Jutten and J. Herault. Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture. *Signal Processing*, 24:1-10, 1991.

[7] E. Moreau and O. Macchi. Complex Self-Adaptive Algorithms for Source Separation Based on High Order Contrasts. In *Proc. VII European Signal Processing Conference (EUSIPCO'94)*, Vol. II, pages 1157-1160, Edinburgh, Scotland, September 1994.

[8] C. W. Therrien. *Discrete Random Signals and Statistical Signal Processing.* Prentice-Hall, 1992.