FAST AND ROBUST DEFLATIONARY SEPARATION OF COMPLEX VALUED SIGNALS

Ella Bingham and Aapo Hyvärinen
Neural Networks Research Centre, Helsinki University of Technology,
P.O. Box 5400, FIN-02015 HUT, FINLAND
Tel: +358 9 4515282; fax: +358 9 4513277
e-mail: ella.bingham@hut.fi, aapo.hyvarinen@hut.fi

ABSTRACT

A fast and robust algorithm for the separation of complex valued signals is presented. It is assumed that the original, complex valued source signals are mutually statistically independent, and that the mixing process is linear. 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. Our fast, fixed-point type algorithm is capable of separating complex valued, linearly mixed source signals in a deflationary manner. The computational efficiency of the algorithm is shown by simulations. Also, a theorem on the local consistency of the estimator given by the algorithm is presented.

1 INTRODUCTION

1.1 The problem

In this paper we present an algorithm for the separation of complex valued signals. Our framework is Independent Component Analysis (ICA) [3], [9]. ICA is a statistical method where the observed data is expressed as a combination of underlying latent variables. The latent variables are assumed non-Gaussian and mutually independent. The task is to find out both the latent variables and the mixing process. 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. Comon [3] has presented the fundamental restrictions under which the above model is identifiable: at most one of the independent components s_j may be Gaussian, and the matrix $\mathbf{A}$ must be of full column rank. (The identifiability of the model is proved in [3] in the case $n = m$.)

1.2 Indeterminacy of the independent components

The independent components $\mathbf{s}$ in the ICA model (1) are found by searching for a matrix $\mathbf{W}$ such that $\mathbf{s} = \mathbf{W}^H\mathbf{x}$ up to some indeterminacies, which are discussed in the following.

In the case of real valued signals, a scalar factor $\alpha_j \in \mathbb{R}$, $\alpha_j \neq 0$ can be exchanged between s_j and a column $\mathbf{a}_j$ of $\mathbf{A}$ without changing the distribution of $\mathbf{x}$: $\mathbf{a}_j s_j = (\alpha_j \mathbf{a}_j)(\alpha_j^{-1} s_j)$. In other words, the order, the signs and the scaling of the independent components cannot be determined. Anyhow, the order of s_j may be chosen arbitrarily and it is a common practice to set $E\{s_j^2\} = 1$; thus only the signs of the independent components are indetermined.

Similarly in the case of complex valued signals there is an unknown phase v_j for each s_j: it is easily proved that

$$\mathbf{a}_j s_j = (v_j \mathbf{a}_j)(\frac{s_j}{v_j}), \quad |v_j| = 1, v_j \in \mathbb{C}. \tag{2}$$

From this indeterminacy it follows that it is impossible to retain the phases of s_j, and $\mathbf{W}^H\mathbf{A}$ is a matrix where in each row and each column there is one nonzero element $v_j \in \mathbb{C}$ that is of unit modulus. Note that the indeterminacy is an inherent property of complex ICA — it does not follow from the assumptions made in this paper.

In the light of this indeterminacy, spherically symmetric distributions are preferable. If a complex signal s_j has a spherically symmetric distribution, i.e. the distribution depends on the modulus of s_j only, the multiplication by a variable $v_j \in \mathbb{C}$ does not change the distribution of s_j. Thus the distribution of $\mathbf{x}$ remains unchanged as well.

1.3 Our approach

A fast fixed point algorithm (FastICA) for the separation of linearly mixed, real valued independent source signals was presented by Hyvärinen and Oja [5], [7]. The FastICA algorithm is a computationally efficient and robust fixed-point type batch algorithm for independent component analysis and blind source separation. For sphered real-valued data the one-unit FastICA algorithm has the following form:

$$\mathbf{w}_{new} = E\{\mathbf{x}g(\mathbf{w}^T\mathbf{x})\} - E\{g'(\mathbf{w}^T\mathbf{x})\}\mathbf{w} \qquad (3)$$

where the weight vector $\mathbf{w}$ is normalized to unit norm after every iteration. The nonlinearity g can be chosen quite freely, and e.g. $g(u) = \tanh(a_1 u)$, $a_1 \approx 1$, or $g(u) = u\exp(-u^2/2)$ are often used. Units using the algorithm in (3) can be combined into systems that estimate several independent components either one-by-one using deflation, or in parallel using symmetric decorrelation.

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, the number of independent component variables is the same as the number of observed linear mixtures, that is, $n = m$. 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 [3]. We assume that the signals s_j are zero-mean and white, i.e. the real and imaginary parts of s_j are uncorrelated and of equal variance; this can be summarized as $E\{\mathbf{s}\mathbf{s}^H\} = \mathbf{I}$ and $E\{\mathbf{s}\mathbf{s}^T\} = \mathbf{O}$. (The latter property is usually referred to as "circularity".) The above assumptions on s_j are 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, as discussed in Section 1.2.

1.4 Other approaches

Algorithms for independent component analysis of complex valued signals have been presented earlier, see [1], [4] and [10]. These earlier algorithms are either computationally more intensive than our algorithm or based on using kurtosis. Also, no theoretical justification on consistency is given in these references. 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 [5] 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 is 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

2.1 Choice of the contrast function

Our contrast function is

$$J_G(\mathbf{w}) = E\{G(|\mathbf{w}^H\mathbf{x}|^2)\} \qquad (4)$$

where $G : \mathbb{R}^+ \cup \{0\} \to \mathbb{R}$ is a smooth even function, $\mathbf{w}$ is an n-dimensional complex weight vector and $E\{|\mathbf{w}^H\mathbf{x}|^2\} = 1$. Finding the extrema of a contrast function is a well defined problem only if the function is real. For this reason our contrast functions operate on absolute values rather than on complex values. Remember also that we assumed a spherically symmetric distribution for the sources s (Section 1.2), and hence the use of absolute rather than complex values is not a severe restriction.

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 three different functions:

$$G_1(y) = \sqrt{a_1 + y},\ a_1 \approx 0.1, \qquad (5)$$
$$G_2(y) = \log(a_2 + y),\ a_2 \approx 0.1, \qquad (6)$$
$$G_3(y) = \frac{1}{2}y^2. \qquad (7)$$

Of these, 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 [11]

$$\text{kurt}(y) = E\{|y|^4\} - 2(E\{|y|^2\})^2 - |E\{y^2\}|^2 \qquad (8)$$
$$= E\{|y|^4\} - 2 \qquad (9)$$

where y is white; i.e. the real and imaginary parts of y are uncorrelated and their variances are equal. This definition of kurtosis is intuitive since it vanishes if y is Gaussian.

2.2 Consistency

Any nonlinear learning function G divides the space of probability distributions into two half-spaces. In the ICA of real valued signals, the independent components can be estimated by either maximizing or minimizing a function similar to (4), depending on which half-space their distribution lies in [8]. This idea can be generalized to complex valued signals. 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 $E\{\mathbf{x}\mathbf{x}^H\} = \mathbf{I}$ and the independent component variables s_k, $k = 1, \ldots, n$ in $\mathbf{s}$ have unit variances and uncorrelated real and imaginary parts of equal variances. Also, $G : \mathbb{R}^+ \cup \{0\} \to \mathbb{R}$ 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

$$E\{g(|s_k|^2) + |s_k|^2 g'(|s_k|^2) - |s_k|^2 g(|s_k|^2)\} < 0 \quad (> 0, \text{ resp.}). \tag{10}$$

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

A special case of the theorem is when $g(y) = y$, $g'(y) = 1$. Condition (10) reads now

$$E\{|s_k|^2 + |s_k|^2 - |s_k|^2|s_k|^2\} = -E\{|s_k|^4\} + 2 < 0 \quad (> 0, \text{ resp.}). \tag{11}$$

Thus the local maxima of $E\{G(|\mathbf{w}^H\mathbf{x}|^2)\}$ are found when $E\{|s_k|^4\} - 2 > 0$, i.e. the kurtosis (9) of s_k is positive.

3 FIXED-POINT ALGORITHM

We now give the fixed-point algorithm for complex signals under the ICA data model (1). The algorithm requires a preliminary sphering or whitening of the data such that $E\{\mathbf{x}\mathbf{x}^H\} = \mathbf{I}$. The algorithm searches for the extrema of $E\{G(|\mathbf{w}^H\mathbf{x}|^2)\}$ under the constraint $E\{|\mathbf{w}^T\mathbf{x}|^2\} = ||\mathbf{w}||^2 = 1$. Details of the derivation are omitted due to lack of space.

The fixed-point algorithm for one unit is

$$\begin{aligned} \mathbf{w}^+ &= E\{\mathbf{x}(\mathbf{w}^H\mathbf{x})^* g(|\mathbf{w}^H\mathbf{x}|^2)\} - \\ &\quad E\{g(|\mathbf{w}^H\mathbf{x}|^2) + |\mathbf{w}^H\mathbf{x}|^2 g'(|\mathbf{w}^H\mathbf{x}|^2)\}\mathbf{w} \\ \mathbf{w}_{new} &= \frac{\mathbf{w}^+}{||\mathbf{w}^+||} \end{aligned} \tag{12}$$

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 [5], [7].

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

$$\mathbf{W} = \mathbf{W}(\mathbf{W}^H\mathbf{W})^{-1/2} \tag{13}$$

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 for each signal j the radius r_j was drawn from a different distribution and the phase angle ϕ_j was uniformly distributed on $[0, 2\pi]$, which implied that real and imaginary parts of the signals were uncorrelated and of equal variance. These assumptions are quite realistic in practical problems. Also, each signal was normalized to unit variance. There were a total of 8 complex random signals and 50 000 samples per signal at each trial.

Source signals $\mathbf{s}$ were mixed using a randomly generated complex mixing matrix $\mathbf{A}$. The mixed signals $\mathbf{x}_{old} = \mathbf{A}\mathbf{s}$ were first whitened using $\mathbf{x} = \mathbf{Q}\mathbf{x}_{old}$ and then fed to 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 $\mathbf{W}^H(\mathbf{QA})$. 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, $|\mathbf{W}^H(\mathbf{QA})|$ should be a permutation matrix. Our error measure is the sum of squared deviation of $|\mathbf{W}^H(\mathbf{QA})|$ from the nearest permutation matrix.

All three contrast functions (5-7) were successful in that the Theorem was always fulfilled and $|\mathbf{W}^H(\mathbf{QA})|$ converged to a permutation matrix in about 6 steps. Figure 1 shows the convergence using G_2 (6).

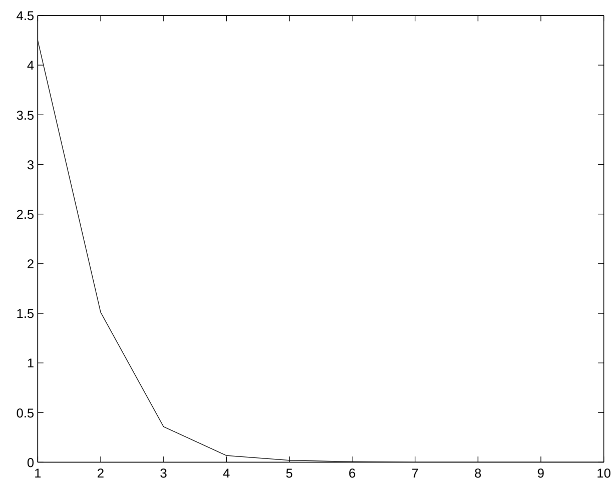

Figure 1: Convergence of the fixed-point algorithm using contrast function $G_2(y) = \log(a_2 + y)$, $a_2 = 0.1$; average result over 10 runs. About 6 iteration steps were needed for convergence.

5 RELATION TO SUBSPACE METHODS

Our complex ICA closely resembles independent subspace methods [6] and multidimensional ICA [2]. In both methods, the components s_j can be divided into m-tuples such that the components inside a given m-tuple may be dependent on each other but independent of other m-tuples. Each m-tuple corresponds to m basis vectors that are orthogonal after prewhitening. In [6] it was proposed that the distributions inside the m-tuples could be modeled by spherically symmetric distributions. This implies that the contrast function (for one subspace) should be of the form $E\{G(\sum_{j=1}^{m}(\mathbf{w}_j^T\mathbf{x})^2)\}$ where $\mathbf{w}_j^T\mathbf{w}_k = 0$, $j \neq k$.

In our complex ICA, the contrast function operates on $|\mathbf{w}^H\mathbf{x}|^2$ which may be expressed as $(\tilde{\mathbf{w}}^T\tilde{\mathbf{x}})^2 + (\tilde{\mathbf{w}'}^T\tilde{\mathbf{x}})^2$. Here $\mathbf{w} = (w_{1r} + iw_{1i}, \ldots, w_{nr} + iw_{ni})$, $\mathbf{x} = (x_{1r} + ix_{1i}, \ldots, x_{nr} + ix_{ni})$, $\tilde{\mathbf{w}} = (w_{1r}, w_{1i}, \ldots, w_{nr}, w_{ni})$, $\tilde{\mathbf{w}}' = (-w_{1i}, w_{1r}, \ldots, -w_{ni}, w_{nr})$ and $\tilde{\mathbf{x}} = (x_{1r}, x_{1i}, \ldots, x_{nr}, x_{ni})$. Thus the subspace is two-dimensional (real and imaginary parts of a complex number) and there are two orthogonal basis vectors: $\tilde{\mathbf{w}}^T\tilde{\mathbf{w}}' = 0$. In contrast to subspace methods, one of the basis vectors is determined straightforward from the other basis vector.

In independent subspace analysis, the independent subspace is determined only up to an orthogonal $m \times m$ matrix factor [6]. In complex ICA however, the indeterminacy is less severe: the sources are determined up to a complex factor v, $|v| = 1$.

It can be concluded that complex ICA is a restricted form of independent subspace methods.

6 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 against outliers 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.

7 *

References

[1] A. D. Back and A. C. Tsoi, "Blind deconvolution of signals using a complex recurrent network", Neural Networks for Signal Processing 4, Proceedings of the 1994 IEEE Workshop, J. Vlontzos, J. Hwang and E. Wilson (Eds), IEEE Press, 1994, pp. 565-574.

[2] J.-F. Cardoso, "Multidimensional Independent Component Analysis", Proceedings of the 1998 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'98), Seattle, USA, May 12–15, 1998, pp. 1941–1944.

[3] P. Comon, "Independent Component Analysis — a new concept?", Signal Processing, vol. 36, pp. 287-314, 1994.

[4] P. Comon and E. Moreau, "Improved contrast dedicated to blind separation in communications", Proceedings of the 1997 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'97), Munich, Germany, April 21–24, 1997, pp. 3453–3456.

[5] 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.

[6] A. Hyvärinen and P. Hoyer, "Emergence of phase and shift invariant features by decomposition of natural images into independent feature subspaces", Neural Computation (in press), 2000.

[7] A. Hyvärinen and E. Oja, "A Fast Fixed-Point Algorithm for Independent Component Analysis", Neural Computation, vol. 9, pp. 1483–1492, 1997.

[8] A. Hyvärinen and E. Oja, "Independent component analysis by general nonlinear Hebbian-like learning rules", Signal Processing, vol. 64, pp. 301–313, 1998.

[9] C. Jutten and J. Herault, "Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture", Signal Processing, vol. 24, pp. 1–10, 1991.

[10] E. Moreau and O. Macchi, "Complex Self-Adaptive Algorithms for Source Separation Based on High Order Contrasts". Proceedings of the VII European Signal Processing Conference (EUSIPCO'94), Vol. II, Edinburgh, Scotland, September 1994, pp. 1157–1160.

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