A fast fixed-point algorithm for independent component analysis 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
Ella.Bingham@hut.fi, Aapo.Hyvarinen@hut.fi
http://www.cis.hut.fi/projects/ica/

19th January 2000

Abstract

Separation of complex valued signals is a frequently arising problem in signal processing. For example, separation of convolutively mixed source signals involves computations on complex valued signals. 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. Also, the local consistency of the estimator given by the algorithm is proved.

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.

Independent component analysis (ICA) [1], [2] is a statistical model where the observed data is expressed as a linear 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 article is

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

where $\mathbf{x} = (x_1, \ldots, x_m)$ is the vector of observed random variables, $\mathbf{s} = (s_1, \ldots, s_n)$ 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 [1]: 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 [1] in the case $n = m$.)

A fast fixed point algorithm (FastICA) for the separation of linearly mixed independent source signals was presented by Hyvärinen and Oja [3], [4]. The FastICA algorithm is a computationally efficient and robust fixed-point type algorithm for independent component analysis and blind source separation.

In this article we show how the FastICA algorithm can be extended to complex valued signals. Both the independent component variables $\mathbf{s}$ and the observed variables $\mathbf{x}$ 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 [1]. We assume that the signals s_j are zero-mean and white, i.e. real and imaginary parts of s_j are uncorrelated and their variances are equal; this is quite realistic in practical problems.

Algorithms for independent component analysis of complex valued signals are also presented in [5] and [6]. Both of these algorithms are computationally more intensive than our algorithm, and no proofs of consistency are given in either of the references. In contrast, we prove 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 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.

This paper is organized as follows. We first go through some basic concepts of complex random variables in Section 2. We then discuss the indeterminacy that is inherent in estimating complex valued independent components (Section 3). In Section 4 we motivate our approach of ICA estimation and discuss the contrast function used in our algorithm. The fast fixed-point algorithm is presented in Section 5, and simulation results confirming the usefulness of the algorithm are shown in Section 6. Section 7 discusses connections to other ICA research. Finally, some conclusions are drawn in Section 8.

2 Basic concepts of complex random variables

A complex random variable may be represented as $y = u + iv$ where u and v are real-valued random variables. The density of y is $f(y) = f(u, v) \in \mathbb{R}^2$. The *expectation* of y is $E\{y\} = E\{u\} + iE\{v\}$. Two complex random variables y_1 and y_2 are *uncorrelated* if $E\{y_1 y_2^*\} = E\{y_1\}E\{y_2^*\}$, where y^* designates the complex conjugate of y. The *covariance matrix* of a zero-mean complex random vector $\mathbf{y} = (y_1, \ldots, y_n)$ is

$$E\{\mathbf{y}\mathbf{y}^H\} = \begin{bmatrix} C_{11} & \cdots & C_{1n} \\ \vdots & \ddots & \vdots \\ C_{n1} & \cdots & C_{nn} \end{bmatrix} \tag{2}$$

where $C_{jk} = E\{y_j y_k^*\}$ and $\mathbf{y}^H$ stands for the Hermitian of $\mathbf{y}$, that is, $\mathbf{y}$ transposed and conjugated. In our complex ICA model, all source signals s_j are zero-mean and they have unit variances and uncorrelated real and imaginary parts of equal variances. In short, these requirements are equivalent to $E\{\mathbf{s}\mathbf{s}^H\} = \mathbf{I}$ and $E\{\mathbf{s}\mathbf{s}^T\} = \mathbf{O}$. In the latter, the expectation of the outer product of a *complex* random vector *without* the conjugate is a null matrix. These assumptions imply that s_j must be strictly complex; that is, the imaginary part of s_j may not in general vanish.

A frequently encountered statistics in ICA is *kurtosis*, or fourth-order cumulant. For zero-mean, complex random variables it could be defined, for example, as [6], [7]

$$\mathrm{kurt}(y) = E\{|y|^4\} - E\{yy^*\}E\{yy^*\} - E\{yy\}E\{y^*y^*\} - E\{yy^*\}E\{y^*y\} \tag{3}$$

but the definitions vary with respect to the placement of conjugates (*) — actually, there are 2^4 ways to define the kurtosis [7]. We choose the definition in [8], where

$$\mathrm{kurt}(y) = E\{|y|^4\} - 2(E\{|y|^2\})^2 - |E\{y^2\}|^2 = E\{|y|^4\} - 2 \tag{4}$$

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.

3 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 this paper, we use the notation $\mathbf{s} = \mathbf{W}^H\mathbf{x}$ which is analogous to the notation in [4] but differs from the notation $\mathbf{s} = \mathbf{W}\mathbf{x}$ used in [3].

In the real case, 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 complex case there is an unknown phase v_j for each s_j: it is easily proved that

$$\mathbf{a}_j s_j = (v_j \mathbf{a}_j)(\frac{s_j}{v_j}), \quad |v_j| = 1,\ v_j \in \mathbb{C}. \tag{5}$$

If 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 does not change the distribution of s_j. Thus the distribution of $\mathbf{x}$ remains unchanged as well.

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 article.

4 Contrast function

4.1 Choice of the contrast function

Now we generalize the framework in [3], [4] and [9] for complex valued signals. One might make a distinction between "top-down" and "bottom-up" approaches to ICA [9]. 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 prove 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

$$J_G(\mathbf{w}) = E\{G(|\mathbf{w}^H\mathbf{x}|^2)\} \tag{6}$$

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 Formula (4) for the kurtosis of complex variables: if we choose $G(y) = y^2$, then $J_G(\mathbf{w}) = E\{|\mathbf{w}^H\mathbf{x}|^4\}$. Thus J essentially measures the kurtosis of $\mathbf{w}^H\mathbf{x}$, which is a classic measure in higher-order statistics.

Maximizing the sum of n one-unit contrast functions, and taking into account the constraint of decorrelation, one obtains the following optimization problem:

$$\begin{aligned} &\text{maximize } \sum_{j=1}^{n} J_G(\mathbf{w}_j) \text{ with respect to } \mathbf{w}_j,\ j = 1, \ldots, n \\ &\text{under constraint } E\{(\mathbf{w}_k^H\mathbf{x})(\mathbf{w}_j^H\mathbf{x})^*\} = \delta_{jk} \end{aligned} \tag{7}$$

where $\delta_{jk} = 1$ for $j = k$ and $\delta_{jk} = 0$ otherwise.

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, the derivatives g of which are also given:

$$G_1(y) = \sqrt{a_1 + y}, \quad g_1(y) = \frac{1}{2\sqrt{a_1 + y}} \tag{8}$$

$$G_2(y) = \log(a_2 + y), \quad g_2(y) = \frac{1}{a_2 + y} \tag{9}$$

$$G_3(y) = \frac{1}{2}y^2, \quad g_3(y) = y \tag{10}$$

where a_1 and a_2 are some arbitrary constants for which values $a_1 \approx 0.1$ and $a_2 \approx 0.1$ were chosen in this work. 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 (4).

4.2 Consistency

In [9], in the context of ICA on real-valued signals, 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 a function similar to (6), depending on which half-space their distribution lies in. In [9] 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 given in the Appendix:

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, $G : \mathbb{R}^+ \cup \{0\} \to \mathbb{R}$ 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

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

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

A special case of the theorem is when $g(y) = y$, $g'(y) = 1$. Condition (11) reads now

$$E\{|s_k|^2 + |s_k|^2 - |s_k|^2|s_k|^2\} = -E\{|s_k|^4\} + 2 < 0 \quad (> 0, \text{ resp.}). \tag{12}$$

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 (4) of s_k is positive.

5 Fixed-point algorithm

We now give the fixed-point algorithm for complex signals under the ICA data model (1). The algorithm searches for the extrema of $E\{G(|\mathbf{w}^H\mathbf{x}|^2)\}$. Details of the derivation are presented in the Appendix.

The algorithm requires a preliminary sphering or whitening of the data: the observed variable $\mathbf{x}_{old}$ is linearly transformed to a zero-mean variable $\mathbf{x} = \mathbf{Q}\mathbf{x}_{old}$, $\mathbf{x} = (x_{1r} + ix_{1i}, \ldots, x_{nr} + ix_{ni})$ such that $E\{\mathbf{x}\mathbf{x}^H\} = \mathbf{I}$. Whitening can always be accomplished by e.g. Principal Component Analysis [1].

The fixed-point algorithm for one unit is

$$\begin{aligned} \mathbf{w}^+ &= E\{\mathbf{x}(\mathbf{w}^H\mathbf{x})^* g(|\mathbf{w}^H\mathbf{x}|^2)\} - E\{g(|\mathbf{w}^H\mathbf{x}|^2) + |\mathbf{w}^H\mathbf{x}|^2 g'(|\mathbf{w}^H\mathbf{x}|^2)\}\mathbf{w} \\ \mathbf{w}_{new} &= \frac{\mathbf{w}^+}{||\mathbf{w}^+||}. \end{aligned} \tag{13}$$

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. A simple way to accomplish this is a deflation scheme based on a Gram-Schmidt-like decorrelation [3]: When we have estimated p independent components, or p vectors $\mathbf{w}_1, \ldots, \mathbf{w}_p$, we run the one-unit fixed-point algorithm for $\mathbf{w}_{p+1}$, and after every iteration step subtract from $\mathbf{w}_{p+1}$ the projections of the previously estimated p vectors, and then renormalize $\mathbf{w}_{p+1}$:

$$\begin{aligned} \mathbf{w}_{p+1} &= \mathbf{w}_{p+1} - \sum_{j=1}^{p} \mathbf{w}_j \mathbf{w}_j^H \mathbf{w}_{p+1} \\ \mathbf{w}_{p+1} &= \frac{\mathbf{w}_{p+1}}{||\mathbf{w}_{p+1}||}. \end{aligned} \tag{14}$$

The above decorrelation scheme is suitable for deflationary separation of the independent components. Sometimes it is preferable to estimate all the independent components simultaneously,

and use a symmetric decorrelation. This can be accomplished e.g. by

$$\mathbf{W} = \mathbf{W}(\mathbf{W}^H\mathbf{W})^{-1/2} \tag{15}$$

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

6 Simulation results

Complex signals were separated to test the performance of the fast fixed-point algorithm and the Theorem. Symmetric decorrelation scheme, presented in Formula (15) in the Appendix, was used in the algorithm. 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 $[-\pi, \pi]$, 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 $\mathbf{W}^H(\mathbf{Q}\mathbf{A})$. 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, $|\mathbf{W}^H(\mathbf{Q}\mathbf{A})|$ should be a permutation matrix. Our error measure is the sum of squared deviation of $|\mathbf{W}^H(\mathbf{Q}\mathbf{A})|$ from the nearest permutation matrix.

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

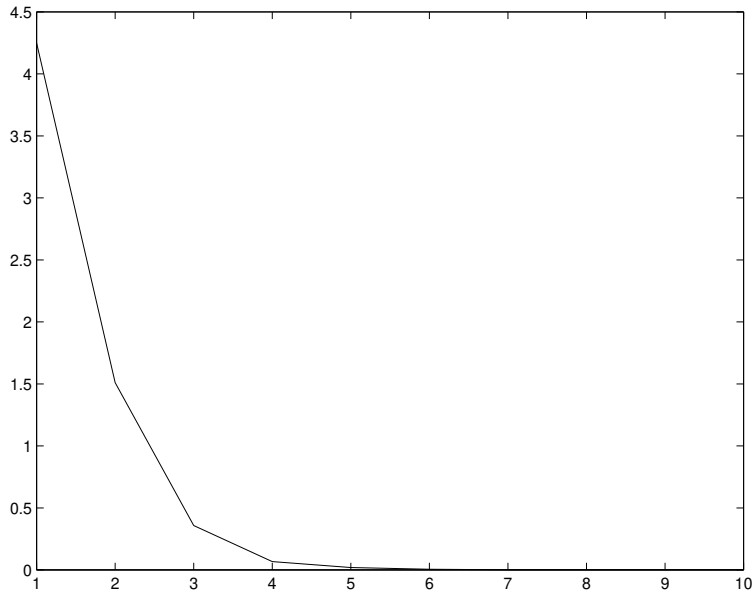

Figure 1: Convergence of the fixed-point algorithm using contrast function $G_2(y) = \log(a_2 + y)$; average result over 10 runs. About 6 iteration steps were needed for convergence.

7 Relation to subspace methods

Our complex ICA closely resembles independent subspace methods [10] and multidimensional ICA [11]. 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 [10] 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 [10]. 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.

8 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.

References

[1] Pierre Comon. Independent Component Analysis — a new concept? *Signal Processing*, 36:287–314, 1994.

[2] Christian Jutten and Jeanny Herault. Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture. *Signal Processing*, 24:1–10, 1991.

[3] Aapo Hyvärinen. Fast and robust fixed-point algorithms for Independent Component Analysis. *IEEE Transactions on Neural Networks*, 10(3):626–634, May 1999.

[4] Aapo Hyvärinen and Erkki Oja. A fast fixed-point algorithm for Independent Component Analysis. *Neural Computation*, 9:1483–1492, 1997.

[5] Andrew D. Back and Ah Chung Tsoi. Blind deconvolution of signals using a complex recurrent network. In J. Vlontzos, J. Hwang, and E. Wilson, editors, *Neural Networks for Signal Processing 4, Proceedings of the 1994 IEEE Workshop*, pages 565–574. IEEE Press, 1994.

[6] Eric Moreau and Odile Macchi. Complex self-adaptive algorithms for source separation based on higher order contrasts. In *Proc. VII European Signal Processing Conference (EUSIPCO'94)*, volume II, pages 1157–1160, Edinburgh, Scotland, September 1994.

[7] Chrysostomos L. Nikias and Athina P. Petropulu. *Higher-Order Spectra Analysis. A Nonlinear Signal Processing Framework.* Prentice-Hall, 1993.

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

[9] Aapo Hyvärinen and Erkki Oja. Independent Component Analysis by general nonlinear Hebbian-like learning rules. *Signal Processing*, 64:301–313, 1998.

[10] Aapo Hyvärinen and Patrik Hoyer. Emergence of phase and shift invariant features by decomposition of natural images into independent feature subspaces. *Neural Computation*, 2000. In press.

[11] Jean-François Cardoso. Multidimensional Independent Component Analysis. In *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP'98)*, Seattle, USA, 1998.

A Proof of Theorem

Denote by $H(\mathbf{w})$ the function to be minimized or maximized, $E\{G(|\mathbf{w}^H\mathbf{x}|^2)\}$. Make the orthogonal change of coordinates $\mathbf{z} = \mathbf{A}^H\mathbf{w}$, giving $H(\mathbf{z}) = E\{G(|\mathbf{z}^H\mathbf{s}|^2)\}$. When $\mathbf{w}$ coincides with one of the rows of $\mathbf{A}^{-1}$, we have $\mathbf{z} = (0, \ldots, 0, v, 0, \ldots, 0)$ — remember that $\mathbf{A}$ is orthogonal due to the prewhitening of $\mathbf{x}$. In the following we shall analyze the stability of such $\mathbf{z}$.

We now search for a Taylor expansion of H in the extrema. We do not use complex differentiation operators because H is in general not analytic and thus it cannot be expanded as a Taylor series in the complex form. The gradient of H with respect to $\mathbf{z}$ is

$$\nabla H(\mathbf{z}) = \begin{pmatrix} \frac{\partial}{\partial z_{1r}} \\ \frac{\partial}{\partial z_{1i}} \\ \vdots \\ \frac{\partial}{\partial z_{nr}} \\ \frac{\partial}{\partial z_{ni}} \end{pmatrix} H(\mathbf{z}) = 2 \begin{pmatrix} E\{\mathrm{Re}\{s_1(\mathbf{z}^H\mathbf{s})^*\}\, g(|\mathbf{z}^H\mathbf{s}|^2)\} \\ E\{\mathrm{Im}\{s_1(\mathbf{z}^H\mathbf{s})^*\}\, g(|\mathbf{z}^H\mathbf{s}|^2)\} \\ \vdots \\ E\{\mathrm{Re}\{s_n(\mathbf{z}^H\mathbf{s})^*\}\, g(|\mathbf{z}^H\mathbf{s}|^2)\} \\ E\{\mathrm{Im}\{s_n(\mathbf{z}^H\mathbf{s})^*\}\, g(|\mathbf{z}^H\mathbf{s}|^2)\} \end{pmatrix} \tag{16}$$

where $z_j = z_{jr} + iz_{ji}$ and $s_j = s_{jr} + is_{ji}$.

The Hessian of H is now a $2n \times 2n$ real matrix: denote ∇H as $(h_{R1}, h_{I1}, \ldots, h_{Rn}, h_{In})$ where

$$h_{Rj} = E\{\mathrm{Re}\{s_j(\mathbf{z}^H\mathbf{s})^*\}\, g(|\mathbf{z}^H\mathbf{s}|^2)\} \tag{17}$$

$$h_{Ij} = E\{\mathrm{Im}\{s_j(\mathbf{z}^H\mathbf{s})^*\}\, g(|\mathbf{z}^H\mathbf{s}|^2)\} \tag{18}$$

whence the Hessian of H is

$$\nabla^2 H(\mathbf{z}) = 2\begin{pmatrix} \frac{\partial h_{R1}}{\partial z_{1r}} & \frac{\partial h_{R1}}{\partial z_{1i}} & \cdots & \frac{\partial h_{R1}}{\partial z_{nr}} & \frac{\partial h_{R1}}{\partial z_{ni}} \\ \frac{\partial h_{I1}}{\partial z_{1r}} & \frac{\partial h_{I1}}{\partial z_{1i}} & \cdots & \frac{\partial h_{I1}}{\partial z_{nr}} & \frac{\partial h_{I1}}{\partial z_{ni}} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{\partial h_{Rn}}{\partial z_{1r}} & \frac{\partial h_{Rn}}{\partial z_{1i}} & \cdots & \frac{\partial h_{Rn}}{\partial z_{nr}} & \frac{\partial h_{Rn}}{\partial z_{ni}} \\ \frac{\partial h_{In}}{\partial z_{1r}} & \frac{\partial h_{In}}{\partial z_{1i}} & \cdots & \frac{\partial h_{In}}{\partial z_{nr}} & \frac{\partial h_{In}}{\partial z_{ni}} \end{pmatrix}. \tag{19}$$

Without loss of generality, it is enough to analyze the stability of the point $\mathbf{z} = v\mathbf{e}_1 = (v, 0, \ldots, 0)$, which corresponds to $\mathbf{w} = v\mathbf{a}_1$. Now $v = v_r + iv_i$ and $|\mathbf{z}^H\mathbf{s}|^2 = |s_1|^2$. Evaluating the gradient (16) at point $\mathbf{z} = v\mathbf{e}_1$, we get

$$\nabla H(v\mathbf{e}_1) = 2\begin{pmatrix} v_r E\{|s_1|^2\, g(|s_1|^2)\} \\ v_i E\{|s_1|^2\, g(|s_1|^2)\} \\ 0 \\ \vdots \\ 0 \end{pmatrix} \tag{20}$$

using the independence of s_j and the zero-mean and unit-variance properties of s_j.

For the Hessian at point $\mathbf{z} = v\mathbf{e}_1$ we use the independence of s_j and the assumptions $E\{\mathbf{s}\mathbf{s}^H\} = \mathbf{I}$ and $E\{\mathbf{s}\mathbf{s}^T\} = \mathbf{O}$, yielding

$$\nabla^2 H(v\mathbf{e}_1) =$$

$$2\begin{pmatrix} E\{|s_1|^2 g(|s_1|^2) + 2v_r^2|s_1|^4 g'(|s_1|^2)\} & 2v_r v_i E\{|s_1|^4 g'(|s_1|^2)\} & 0 & \ldots & 0 \\ 2v_r v_i E\{|s_1|^4 g'(|s_1|^2)\} & E\{|s_1|^2 g(|s_1|^2) + 2v_i^2|s_1|^4 g'(|s_1|^2)\} & 0 & \ldots & 0 \\ 0 & 0 & \alpha & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & \alpha \end{pmatrix} \tag{21}$$

where

$$\alpha = E\{g(|s_1|^2) + |s_1|^2 g'(|s_1|^2)\}. \tag{22}$$

Note that we do *not* assume that the real and imaginary parts of the same variable s_j are independent, even though we use the independence of s_j and s_k, $j \neq k$ as discussed in Section 2.

Now we make a small perturbation $\boldsymbol{\varepsilon} = (\varepsilon_{1r}, \varepsilon_{1i}, \ldots, \varepsilon_{nr}, \varepsilon_{ni})$ where ε_{jr} and ε_{ji} are the real and imaginary parts of $\varepsilon_j \in \mathbb{C}$ and evaluate the Taylor expansion of H:

$$\begin{aligned} H(v\mathbf{e}_1 + \boldsymbol{\varepsilon}) &= H(v\mathbf{e}_1) + \boldsymbol{\varepsilon}^T \nabla H(v\mathbf{e}_1) + \frac{1}{2}\boldsymbol{\varepsilon}^T \nabla^2 H(v\mathbf{e}_1)\boldsymbol{\varepsilon} + o(||\boldsymbol{\varepsilon}||^2) \\ &= H(\mathbf{e}_1) + 2(\varepsilon_{1r}v_r + \varepsilon_{1i}v_i)E\{|s_1|^2 g(|s_1|^2)\} + \\ &\quad \varepsilon_{1r}^2 E\{|s_1|^2 g(|s_1|) + 2v_r^2|s_1|^4 g'(|s_1|^2)\} + 4v_r v_i \varepsilon_{1r}\varepsilon_{1i} E\{|s_1|^4 g'(|s_1|^2)\} + \\ &\quad \varepsilon_{1i}^2 E\{|s_1|^2 g(|s_1|) + 2v_i^2|s_1|^4 g'(|s_1|^2)\} + \\ &\quad E\{g(|s_1|^2) + |s_1|^2 g'(|s_1|^2)\} \sum_{j>1} (\varepsilon_{jr}^2 + \varepsilon_{ji}^2) + o(||\boldsymbol{\varepsilon}||^2) \end{aligned} \tag{23}$$

Furthermore, due to the constraint $||\mathbf{w}|| = 1$ and thus $||v\mathbf{e}_1 + \boldsymbol{\varepsilon}|| = 1$ we get

$$2(\varepsilon_{1r}v_r + \varepsilon_{1i}v_i) = -\sum_{j=1}^{n} (\varepsilon_{jr}^2 + \varepsilon_{ji}^2). \tag{24}$$

Using this, we get

$$\begin{aligned} H(v\mathbf{e}_1 + \boldsymbol{\varepsilon}) &= H(v\mathbf{e}_1) + E\{g(|s_1|^2) + |s_1|^2 g'(|s_1|^2) - |s_1|^2 g(|s_1|^2)\} \sum_{j>1} (\varepsilon_{jr}^2 + \varepsilon_{ji}^2) + \\ &\quad 2(\varepsilon_{1r}v_r + \varepsilon_{1i}v_i)^2 E\{|s_1|^4 g'(|s_1|^2)\} + o(||\boldsymbol{\varepsilon}||^2) \end{aligned} \tag{25}$$

where the term of order $(\varepsilon_{1r}v_r + \varepsilon_{1i}v_i)^2$ is $o(||\boldsymbol{\varepsilon}||^2)$ according to (24), giving

$$\begin{aligned} H(v\mathbf{e}_1 + \boldsymbol{\varepsilon}) &= H(v\mathbf{e}_1) + E\{g(|s_1|^2) + |s_1|^2 g'(|s_1|^2) - |s_1|^2 g(|s_1|^2)\} \sum_{j>1} (\varepsilon_{jr}^2 + \varepsilon_{ji}^2) + \\ &\quad + o(||\boldsymbol{\varepsilon}||^2). \end{aligned} \tag{26}$$

Thus $\mathbf{z} = v\mathbf{e}_1$ is an extremum, and it is the maximum (minimum) if

$$E\{g(|s_1|^2) + |s_1|^2 g'(|s_1|^2) - |s_1|^2 g(|s_1|^2)\} < 0 \quad (> 0, \text{ resp.}). \quad \square \tag{27}$$

B Derivation of the algorithm

We shall derive the fixed-point algorithm for one unit. Let $w = w_r + iw_i$ and $x = x_r + ix_i$. For the ease of derivations, the algorithm updates the real and imaginary parts of w separately. We assume that the source signals s_j are white, i.e. they are zero-mean and have unit variances and

uncorrelated real and imaginary parts of equal variances, that is, $E\{\mathbf{s}\mathbf{s}^H\} = \mathbf{I}$ 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 $E\{G(|\mathbf{w}^H\mathbf{x}|^2)\}$ under the constraint $E\{|\mathbf{w}^H\mathbf{x}|^2\} = ||\mathbf{w}||^2 = 1$ are obtained at points where

$$\nabla E\{G(|\mathbf{w}^H\mathbf{x}|^2)\} - \beta \nabla E\{|\mathbf{w}^H\mathbf{x}|^2\} = 0 \tag{28}$$

where $\beta \in \mathbb{R}$ and the gradient is computed with respect to real and imaginary parts of w separately. The first term in (28) is

$$\nabla E\{G(|\mathbf{w}^H\mathbf{x}|^2)\} = \begin{pmatrix} \frac{\partial}{\partial w_{1r}} \\ \frac{\partial}{\partial w_{1i}} \\ \vdots \\ \frac{\partial}{\partial w_{nr}} \\ \frac{\partial}{\partial w_{ni}} \end{pmatrix} E\{G(|\mathbf{w}^H\mathbf{x}|^2)\} = 2 \begin{pmatrix} E\{\mathrm{Re}\{x_1(\mathbf{w}^H\mathbf{x})^*\}\, g(|\mathbf{w}^H\mathbf{x}|^2)\} \\ E\{\mathrm{Im}\{x_1(\mathbf{w}^H\mathbf{x})^*\}\, g(|\mathbf{w}^H\mathbf{x}|^2)\} \\ \vdots \\ E\{\mathrm{Re}\{x_n(\mathbf{w}^H\mathbf{x})^*\}\, g(|\mathbf{w}^H\mathbf{x}|^2)\} \\ E\{\mathrm{Im}\{x_n(\mathbf{w}^H\mathbf{x})^*\}\, g(|\mathbf{w}^H\mathbf{x}|^2)\} \end{pmatrix} \tag{29}$$

and the second term in (28) is

$$\nabla E\{|\mathbf{w}^H\mathbf{x}|^2\} = 2 \begin{pmatrix} \mathrm{Re}\{w_1\} \\ \mathrm{Im}\{w_1\} \\ \vdots \\ \mathrm{Re}\{w_n\} \\ \mathrm{Im}\{w_n\} \end{pmatrix} \tag{30}$$

where the assumption $E\{\mathbf{x}\mathbf{x}^H\} = \mathbf{I}$ was used.

The Newton method is used to solve (28). The Jacobian matrix of $\nabla E\{G(|\mathbf{w}^H\mathbf{x}|^2)\}$ as in (29) can be approximated as

$$\nabla^2 E\{G(|\mathbf{w}^H\mathbf{x}|^2)\} = 2E\{(\nabla^2|\mathbf{w}^H\mathbf{x}|^2)g(|\mathbf{w}^H\mathbf{x}|^2) + 2(\nabla|\mathbf{w}^H\mathbf{x}|^2)(\nabla|\mathbf{w}^H\mathbf{x}|^2)^T g'(|\mathbf{w}^H\mathbf{x}|^2)\} \tag{31}$$

$$\approx 2E\{g(|\mathbf{w}^H\mathbf{x}|^2) + |\mathbf{w}^H\mathbf{x}|^2 g'(|\mathbf{w}^H\mathbf{x}|^2)\}\mathbf{I} \tag{32}$$

where the approximation was done by separating the expectations. Also, $E\{\mathbf{x}\mathbf{x}^T\} = \mathbf{O}$ (which follows straightforward from $E\{\mathbf{s}\mathbf{s}^T\} = \mathbf{O}$) was used. The Jacobian matrix of $\beta\,\nabla E\{|\mathbf{w}^H\mathbf{x}|^2\}$ is, using (30),

$$\beta\,\nabla^2 E\{|\mathbf{w}^H\mathbf{x}|^2\} = 2\,\beta\,\mathbf{I}. \tag{33}$$

The total approximative Jacobian of (28) is now

$$\mathbf{J} = 2\,(\,E\{g(|\mathbf{w}^H\mathbf{x}|^2) + |\mathbf{w}^H\mathbf{x}|^2 g'(|\mathbf{w}^H\mathbf{x}|^2)\} - \beta\,)\,\mathbf{I} \tag{34}$$

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

$$\begin{aligned} \mathbf{w}^+ &= \mathbf{w} - \frac{E\{\mathbf{x}(\mathbf{w}^H\mathbf{x})^* g(|\mathbf{w}^H\mathbf{x}|^2)\} - \beta\mathbf{w}}{E\{g(|\mathbf{w}^H\mathbf{x}|^2) + |\mathbf{w}^H\mathbf{x}|^2 g'(|\mathbf{w}^H\mathbf{x}|^2)\} - \beta} \\ \mathbf{w}_{new} &= \frac{\mathbf{w}^+}{||\mathbf{w}^+||}. \end{aligned} \tag{35}$$

If we multiply both sides of (35) 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

$$\begin{aligned} \mathbf{w}^+ &= E\{\mathbf{x}(\mathbf{w}^H\mathbf{x})^* g(|\mathbf{w}^H\mathbf{x}|^2)\} - E\{g(|\mathbf{w}^H\mathbf{x}|^2) + |\mathbf{w}^H\mathbf{x}|^2 g'(|\mathbf{w}^H\mathbf{x}|^2)\}\mathbf{w} \\ \mathbf{w}_{new} &= \frac{\mathbf{w}^+}{||\mathbf{w}^+||}. \end{aligned} \tag{36}$$

Decorrelation schemes suitable for deflationary or symmetric separation of the independent components were presented in Section 5.