COMPLEX ICA FOR CIRCULAR AND NON-CIRCULAR SOURCES

Jan Eriksson, Anne-Mari Seppola, and Visa Koivunen*

SMARAD CoE, Signal Processing Laboratory
Department of Electrical Engineering
Helsinki University of Technology, FIN-02015 HUT, Finland
Tel: +358-9-451 2453, Fax: +358-9-452 3614
E-mail: {jan.eriksson,visa.koivunen}@hut.fi

ABSTRACT

In this paper we propose an information theory based generic method for complex Independent Component Analysis (ICA). Expressions for the complex score function are derived. The method exploits the full second order structure of complex signals. It combines a preprocessing step called the strong-uncorrelating transform (SUT) [10] with ICA methods that use the proposed complex score function. The method is capable of separating circular or non-circular and symmetric or asymmetric source distributions from complex mixtures. It allows the separation of such signals with relatively simple modifications to existing methods for *real*-valued signals. The performance of the proposed method is compared to the standard complex JADE [6] and FastICA [3] algorithms in a simulation.

1. INTRODUCTION

Independent component analysis (ICA) [8] is already a relatively established signal processing and data analysis technique. It may be used, for example, in blind source separation (BSS) and identifying or equalizing instantaneous Multiple-Input Multiple-Output (I-MIMO) systems. It has found applications e.g. in wireless communications, biomedical signal processing and data mining (see [13] for references). ICA for separating complex-valued sources is needed, e.g., for convolutive source separation in the frequency domain, or for performing source separation on complex-valued data, such as magnetic resonance imaging, radar or communication data. In instantaneous complex-valued ICA problem

$$\bar{x} = \boldsymbol{A}\bar{s}, \tag{1}$$

the goal is to recover the original source signal vectors $\bar{s}$ from the observation vectors $\bar{x}$ blindly without explicit knowledge of the sources or the linear mixing system $\boldsymbol{A}$. ICA is based on a crucial assumption that the underlying unknown source signals are statistically independent. Recent textbooks provide excellent tutorial material and extensive review on ICA methods [7, 13].

Theoretical conditions on separation of real-valued signals are now well-known [8, 9]. Even though algorithms for separation of complex-valued signals have been developed, see for example [2, 3, 5, 6, 8, 11], the most work is done for circular complex random vectors with symmetric distribution. The conditions when the separation is possible were established only recently [9]. These conditions were based on a careful characterization of the second order statistical properties of non-circular complex vectors. The results show that although theoretically the model is separable under rather mild conditions, the existing complex ICA separation algorithms make implicit additional assumptions about the model, and therefore perform rather poorly in many cases that do not satisfy these assumptions. For instance, cumulant-based methods [6, 8] use only even order cumulants and most information theoretic approaches [2, 3, 11] assume circular sources. In this paper we present a generic information theoretic method that does not require such assumptions. Moreover, we show how existing real-valued ICA algorithms can be used to solve the complex ICA problem.

The paper is organized as follows. First some relevant second order properties of complex random vectors (r.vc.'s) are reviewed. A general framework for information theoretic ICA methods is proposed and an expression for the complex score function is derived. Then a generic complex ICA algorithm is derived in detail, and a simulation example is presented. Finally, some concluding remarks are given.

2. SECOND-ORDER STATISTICS OF COMPLEX R.VC.'S

Let us begin with some definitions and notations. The *modulus* of a complex number $z = z_R + \jmath z_I \in \mathbb{C}$ is denoted $|z| = \sqrt{z^* z} = \sqrt{z_R^2 + z_I^2}$, where the superscript * denotes the complex *conjugate*, $z^* = z_R - \jmath z_I$, and $\jmath = \sqrt{-1}$ is the imaginary unit. The real part of a p-dimensional complex vector $\bar{\mathbf{z}} = (z_1 \ z_2 \ \cdots \ z_p)^T \in \mathbb{C}^p$, where T is the ordinary *transpose*, is denoted by $\bar{\mathbf{z}}_R = \mathrm{Re}\{\bar{\mathbf{z}}\}$ and the imaginary part by $\bar{\mathbf{z}}_I = \mathrm{Im}\{\bar{\mathbf{z}}\}$. A complex matrix $\boldsymbol{C} \in \mathbb{C}^{p\times p}$ is termed *symmetric* if $\boldsymbol{C}^T = \boldsymbol{C}$ and *Hermitian* if $\boldsymbol{C}^H = \boldsymbol{C}$, where the superscript H denotes the *Hermitian transpose*, $\boldsymbol{C}^H = (\boldsymbol{C}^*)^T$. Furthermore, a matrix $\boldsymbol{C}$ is *orthogonal* if $\boldsymbol{C}^T\boldsymbol{C} = \boldsymbol{C}\boldsymbol{C}^T = \boldsymbol{I}$ and *unitary* if $\boldsymbol{C}^H\boldsymbol{C} = \boldsymbol{C}\boldsymbol{C}^H = \boldsymbol{I}$, where $\boldsymbol{I}$ denotes the identity matrix.

A p-variate complex random vector (r.vc.) $\bar{x}$ is defined as a r.vc. of the form

$$\bar{x} = \bar{x}_R + \jmath\bar{x}_I, \tag{2}$$

where $\bar{x}_R$ and $\bar{x}_I$ are p-variate real r.vc.'s. The *expectation* $\mathrm{E}[\cdot]$ of a complex r.vc. $\bar{x}$ is defined as

$$\mathrm{E}_{\bar{x}}[\bar{x}] = \mathrm{E}_{\bar{x}_R}[\bar{x}_R] + \jmath\,\mathrm{E}_{\bar{x}_I}[\bar{x}_I], \tag{3}$$

and the *covariance matrix* $\mathrm{cov}[\bar{x}]$ is given by

$$\mathrm{cov}[\bar{x}] \triangleq \mathrm{E}_{\bar{x}}[(\bar{x} - \mathrm{E}_{\bar{x}}[\bar{x}])(\bar{x} - \mathrm{E}_{\bar{x}}[\bar{x}])^H]. \tag{4}$$

All r.vc.'s in this paper are assumed to have *second-order statistics*, i.e. the covariance matrix is finite. Finally, a r.vc. $\bar{x}$ is called *full*, if its covariance matrix is positive definite, and thus $\mathrm{cov}[\bar{x}]$ is invertible.

The probabilistic structure of a p-dimensional complex r.vc. $\bar{x} = \bar{x}_R + \jmath\bar{x}_I$ is equivalently given by the probabilistic structure of a real $2p$-dimensional r.vc. $(\bar{x}_R^T \ \bar{x}_I^T)^T$. Hence, all the probabilistic concepts for the complex r.vc.'s can be defined using the corresponding concepts of the real r.vc.'s. For

*This work was supported in part by the Academy of Finland.

instance, the probability density function (p.d.f.) of a complex r.vc. $\bar{x}$ is defined as $f_{\bar{x}}(\bar{\mathbf{z}}) \triangleq f_{(\bar{x}_R^T, \bar{x}_I^T)^T}\big((\bar{\mathbf{z}}_R^T, \bar{\mathbf{z}}_I^T)^T\big)$. Thus the full second order description of $\bar{x}$ is given by $\mathrm{cov}\big[\bar{x}_R\big]$, $\mathrm{cov}\big[\bar{x}_I\big]$ and the cross covariance between $\bar{x}_R$ and $\bar{x}_I$. It is easily seen that the complex covariance matrix of the expression (4) does not completely define the second order statistics. However, the covariance matrix $\mathrm{cov}\big[\bar{x}\big]$ together with the *pseudo-covariance* matrix

$$\mathrm{pcov}\big[\bar{x}\big] \triangleq \mathrm{E}_{\bar{x}}\big[(\bar{x} - \mathrm{E}_{\bar{x}}\big[\bar{x}\big])(\bar{x} - \mathrm{E}_{\bar{x}}\big[\bar{x}\big])^T\big] \tag{5}$$

gives the full description of the second order statistics of a complex r.vc. $\bar{x}$ [14]. If a r.vc. has a zero pseudo-covariance, it is called *second-order circular* (proper).

3. COMPLEX ICA BASED ON MUTUAL INFORMATION

Most information theoretic real ICA methods [13] are essentially based on finding a matrix $\boldsymbol{W}$ minimizing the *mutual information* (m.i.) between the joint and the marginal distribution,

$$\mathrm{MI}(\bar{y}) \triangleq \int f_{\bar{y}}(\bar{\mathbf{z}}) \log \frac{f_{\bar{y}}(\bar{\mathbf{z}})}{\prod_{k=1}^{p} f_{y_k}(z_k)} d\bar{\mathbf{z}} = \mathrm{E}_{\bar{y}}\big[\log \frac{f_{\bar{y}}(\bar{y})}{\prod_{k=1}^{p} f_{y_k}(y_k)}\big], \tag{6}$$

of the output signal $\bar{y} = \boldsymbol{W}\bar{x}$. The differences in methods are essentially in how the minimization and the unknown p.d.f.'s are handled. Usually, the minimization is done using gradient-based method which require differentiating $\mathrm{MI}(\boldsymbol{W}\bar{x})$ with respect to $\boldsymbol{W}$. This consideration shows that the optimal gradient is a function of the *score functions*, i.e. negative first derivative of the logarithm of the p.d.f. Since the p.d.f.'s are unknown, the score functions are either estimated, approximated by some easily computable statistics (e.g. cumulants), or replaced by some appropriate fixed functions.

The apparent difficulty of extending the above generic information theoretic approach to complex signals comes from the fact that the m.i. of Eq. (6) for complex r.vc.'s is real valued and therefore it is *not complex differentiable* unless it is a *constant* in the neighborhood of $\boldsymbol{W}\bar{x}$. The same applies to the density function. However, it is actually *real differentiation* we are interested in. This can be conveniently done without separating real and imaginary parts with the following complex (partial) differential operators [4] (see also [15]):

$$\frac{\mathrm{d}\, f}{\mathrm{d}\, z} \triangleq \frac{1}{2}\Big(\frac{\partial f}{\partial z_R} - \jmath\frac{\partial f}{\partial z_I}\Big) \text{ and } \frac{\mathrm{d}\, f}{\mathrm{d}\, z^*} \triangleq \frac{1}{2}\Big(\frac{\partial f}{\partial z_R} + \jmath\frac{\partial f}{\partial z_I}\Big), \tag{7}$$

where the operators on the right hand sides are ordinary real partial derivatives and $f : \mathbb{C} \to \mathbb{C}$ is a real-differentiable function. It should be noted that these operators are not (complex) partial derivatives in the normal sense, although they do satisfy the product rule, $\frac{\mathrm{d}\, z}{\mathrm{d}\, z} = \frac{\mathrm{d}\, z^*}{\mathrm{d}\, z^*} = 1$, and $\frac{\mathrm{d}\, z^*}{\mathrm{d}\, z} = \frac{\mathrm{d}\, z}{\mathrm{d}\, z^*} = 0$. However, they do not satisfy the normal chain rule. Therefore, one should apply them with caution.

Now it is known [4] that the optimal gradient of a real-valued function of complex vector argument is given by the vector operator corresponding to the conjugate operator $\frac{\mathrm{d}}{\mathrm{d}\, z^*}$. Hence, the complex matrix gradient of $\mathrm{MI}(\bar{y})$, $\bar{y} = \boldsymbol{W}\bar{x}$, is computed as follows:

$$\nabla \mathrm{MI}(\bar{y}) \triangleq \frac{\mathrm{d}\, \mathrm{MI}(\boldsymbol{W}\bar{x})}{\mathrm{d}\, \boldsymbol{W}^*}(\boldsymbol{W}) = \big(\mathrm{E}_{\bar{y}}\big[\varphi_{\bar{y}}(\bar{y})\bar{y}^H\big] - I\big)\big(\boldsymbol{W}^{-1}\big)^H, \tag{8}$$

where $\varphi_{\bar{y}}(\bar{\mathbf{z}}) = \big(\varphi_{y_1}(z_1), \ldots, \varphi_{y_p}(z_p)\big)^T$ is a vector of the functions

$$\begin{aligned}\varphi_y(z) &= -\frac{\mathrm{d}\log f_y}{\mathrm{d}\, z^*}(z) = -\frac{1}{2}\frac{\frac{\partial f_y}{\partial z_R}(z) + \jmath\frac{\partial f_y}{\partial z_I}(z)}{f_y(z)} \\ &= -\frac{1}{2}\Big(\frac{\partial \log f_y}{\partial z_R}(z) + \jmath\frac{\partial \log f_y}{\partial z_I}(z)\Big).\end{aligned} \tag{9}$$

These functions are *complex score functions*. This gives the correct form for nonlinearities in complex information theoretic ICA algorithms. Notice that the real and imaginary parts of the score function in Eq. (9) are not simply functions of real and imaginary parts of the argument unless f_y factorizes, i.e. unless real and imaginary parts of the r.v. y are independent. Therefore, in the complex Infomax algorithm [5] for example, the complex `tanh` as the nonlinearity performs better in general than the (split) nonlinearity of the (real) `tanh` function of a real argument applied to the real and imaginary parts separately. This was noticed in [5], where it was considered to be due to the fact that the complex hyperbolic tangent function is analytic.

The standard gradient descent update corresponding to the gradient (8) is described by

$$\boldsymbol{W}_{k+1} = \boldsymbol{W}_k - \mu\nabla \mathrm{MI}(\boldsymbol{W}_k\bar{x}), \tag{10}$$

where μ is a small constant determining the learning rate of the algorithm. This basic algorithm seems to have rather poor convergence properties. It is possible to improve the convergence [5] by using an ad hoc extension of the real-valued natural gradient [1] to the complex field. This is accomplished by multiplying from the right the gradient (8) by $\boldsymbol{W}^H\boldsymbol{W}$ instead of $\boldsymbol{W}^T\boldsymbol{W}$ that is found to be the correct term for the real case. However, it seems that this algorithm may still converge to a local minimum. In the real case, this is usually avoided by first applying the *whitening transform* to data, which reduces the unknown parameter to be just an *orthogonal matrix* instead of an invertible matrix [8]. Then the gradient search is constrained to the orthogonal matrices. In the rest of the paper, we describe a similar method for complex ICA. It turns out that in the complex case the problem can be simplified even more than in the real case.

4. STRONG-UNCORRELATING TRANSFORM

For a general complex r.vc. $\bar{x}$, the necessary and sufficient condition for marginal random variables (r.v.'s) to be *uncorrelated* [14] is that both the covariance and the pseudo-covariance matrices are diagonal. It was shown in [9] that any complex r.vc. with invertible covariance matrix can be linearly transformed such that it has uncorrelated marginal r.v.'s. Specifically the following theorem was proved [9]:

Theorem 1. *Any full complex p-dimensional r.vc. $\bar{x}$ can be transformed by using a nonsingular square matrix $\boldsymbol{C}^{-1}$ such that the r.vc. $\bar{s} = (s_1, \ldots, s_p)^T = \boldsymbol{C}^{-1}\bar{x}$ has the following properties:*

(i) $\mathrm{cov}\big[\bar{s}\big] = \boldsymbol{I}$

(ii) $\mathrm{pcov}\big[\bar{s}\big] = \mathrm{diag}(\boldsymbol{\lambda}\big[\bar{s}\big])$, *where* $\boldsymbol{\lambda}\big[\bar{s}\big] = (\lambda_1, \ldots, \lambda_p)^T$ *denotes a vector such that* $1 \geq \lambda_1 \geq \cdots \geq \lambda_p \geq 0$, $\mathrm{cov}\big[\mathrm{Re}\big\{s_k\big\}, \mathrm{Im}\big\{s_k\big\}\big] = 0$, $k = 1, \ldots, p$, *and* $\lambda_k = \mathrm{cov}\big[\mathrm{Re}\big\{s_k\big\}\big] - \mathrm{cov}\big[\mathrm{Im}\big\{s_k\big\}\big]$.

Also $\mathrm{cov}\big[\bar{x}\big] = \boldsymbol{C}\boldsymbol{C}^H$ *and* $\mathrm{pcov}\big[\bar{x}\big] = \boldsymbol{C}\,\mathrm{diag}(\boldsymbol{\lambda}\big[\bar{s}\big])\boldsymbol{C}^T$.

The matrix $\boldsymbol{C}^{-1}$ in Theorem 1 is called the *strong-uncorrelating transform* (SUT). It is not necessarily unique, i.e. there may exist several SUT matrices for a given r.vc. $\bar{x}$. However, it can be shown [9] that the vector $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_p)^T$ in Theorem 1, called the *spectrum*, is

unique for any given r.vc. $\bar{x}$. Therefore, the second order statistics may be equivalently described with the SUT $\boldsymbol{C}^{-1}$ and the spectrum $\boldsymbol{\lambda}$. A zero spectrum vector corresponds to a second order circular r.vc., and the spectrum of all ones corresponds to a complex r.vc. that is a complex mixture of real r.vc.'s. The components λ_k of the spectrum are called *spectral coefficients*.

It is straightforward to verify that a strong-uncorrelating transform matrix may be found by the following procedure:

(i) Find the usual whitening transform $\boldsymbol{H} = \operatorname{cov}\big[\bar{x}\big]^{-\frac{1}{2}}$, i.e. the inverse of the matrix square root of $\operatorname{cov}\big[\bar{x}\big]$.

(ii) Any symmetric matrix $\boldsymbol{B}$ has a special form of SVD known as *Takagi's factorization* (or symmetric SVD). The factorization is given as $\boldsymbol{B} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^T$, where $\boldsymbol{U}$ is unitary and $\boldsymbol{\Lambda}$ is a diagonal matrix with real non-decreasing nonnegative main diagonal entries. Hence, find the factorization for pseudo-covariance matrix of whitened data, i.e. $\operatorname{pcov}\big[\boldsymbol{H}\bar{x}\big] = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^T$.

(iii) Set $\boldsymbol{C}^{-1} = \boldsymbol{U}^H\boldsymbol{H}$.

For the estimation of the SUT, the first step can be performed with the standard tools in any major numerical software such as Matlab. A fast and efficient method for Takagi's factorization can be found in [12].

5. GENERIC COMPLEX ICA METHOD

The model of Eq. (1) is called *separable* if for every complex matrix $\boldsymbol{W}$ such that $\boldsymbol{W}\bar{x}$ has m independent components, we have $\boldsymbol{\Lambda}\boldsymbol{P}\bar{s} = \boldsymbol{W}\bar{x}$ for some diagonal matrix $\boldsymbol{\Lambda}$ with nonzero complex diagonal elements and permutation matrix $\boldsymbol{P}$. Such matrices $\boldsymbol{W}$ are called *separating matrices*. It was recently shown [9] that the complex ICA model with a full column rank matrix is separable if there are no two sources with equal spectral coefficients and normally distributed real and imaginary parts.

Now the use of the SUT as an ICA technique is summarized in the following theorem [10].

Theorem 2. *Suppose that in the spectrum $\boldsymbol{\lambda}$ of the sources $\bar{s}$ in the ICA model (1) with equal number p of sources and sensors, each spectral coefficient has the multiplicity one, i.e. the spectrum is distinct. Then any strong-uncorrelating transform of the mixture $\bar{x}$ is a separating matrix.*

It should be noted that the requirement that there are as many sources as sensors in Theorem 2 is just a technical limitation. It may be changed to the condition that there are at least as many sensors as sources by noticing that the first part of the SUT calculation is the ordinary whitening transform based on a nonnegative definitive covariance matrix.

Although the SUT can separate all mixtures with a distinct spectrum, in real engineering applications the different source r.v.'s might have the same statistics (the same spectral coefficients). However, in this situation, the SVD theorem can be used to show [10] that the SUT matrix is essentially unique up to *real orthogonal* matrices corresponding to source r.v.'s with the same nonzero spectral coefficients. For second order circular source r.v.'s, the SUT is essentially unique up to a unitary matrix. Thus the SUT can be used as a preprocessing tool in a general complex ICA algorithm [10]:

(i) Apply SUT $\boldsymbol{C}^{-1}$ on the observed mixture $\bar{x}$: $\bar{y} = \boldsymbol{C}^{-1}\bar{x}$.

(ii) Decompose the vector $\bar{y} = (\bar{y}_1^T, \ldots, \bar{y}_k^T)^T$ to subvectors such that marginal r.v.'s in each subvector $\bar{y}_l$ have the same spectral coefficients.

(iii) For each subvector with nonzero spectral coefficient, find a real orthonormal matrix and for the subvector with zero spectral coefficient find a unitary matrix that transforms the submatrix to independent components.

The advantage of the above generic algorithm is that it may transform the original problem to several problems of smaller dimension. The computational complexity is also greatly reduced, since in most cases one needs to deal only with *real-valued* matrices. The estimation of SUT was described earlier. In the rest of the paper we describe in detail how to perform the steps (ii) and (iii) of the algorithm.

6. DECOMPOSING THE STRONGLY UNCORRELATED R.VC.

After performing the SUT, one needs to partition the spectrum vector $\bar{y}$ into subvectors according to its spectral coefficients. However, we only have the estimated spectral coefficient vector $\hat{\boldsymbol{\lambda}}$, and its components are not likely to be equal although their theoretical counterparts might be. We propose the following generic algorithm for deciding the vector decomposition.

(i) Set $k = 1$.

(ii) Find the number l_k of the estimated spectral coefficients in $\hat{\boldsymbol{\lambda}}$ corresponding to zero.

(iii) If $\sum_{n=1}^{k} l_n = p$, quit. Otherwise remove l_k smallest numbers from $\hat{\boldsymbol{\lambda}}$, and increase k by one. Find the number l_k of the estimated spectral coefficients in the remaining vector $\hat{\boldsymbol{\lambda}}$ corresponding to the next smallest spectral coefficient. Repeat the step.

The step (ii) can be done parametrically using the standard decision theoretic methods. Also by noticing that nonnegative definite matrix $\operatorname{pcov}\big[\bar{x}\big]\operatorname{pcov}\big[\bar{x}\big]^H$ has eigenvalues that are squared values of the spectral coefficients, we see that methods designed to estimate the number of signals from e.g. a covariance matrix can be used. For example, the well-known information theoretic criteria [16] based on minimum description length (MDL) may be used.

We are currently developing reliable methods to perform step (iii) that will be reported in a future work. However, one may also skip the step (iii) altogether, and take all the non-circular signals to be a single subvector. The non-circular subvector is necessaryly a real orthogonal matrix away from the true solution, so the methods described for the individual subvectors with the same nonzero spectral coefficient apply also for this case. This approach is used in the experiment of this paper.

7. INDEPENDENT COMPONENTS OF R.VC.'S WITH A UNIFORM SPECTRUM

Finally we describe how to separate components from each subvector $\bar{y}_l$. If the spectral coefficient of the subvector is positive, one needs to find a *real orthogonal* matrix to achieve separation. Then if one is using methods relying on gradient descent-type of optimization of the m.i., the gradient should be naturally real-valued. This is achieved simply by taking the real part of of the expression inside the expectation operator of Eq. (8), i.e., one calculates:

$$\operatorname{Re}\big\{\varphi_{\bar{y}}(\bar{y})\bar{y}^H\big\} = \operatorname{Re}\big\{\varphi_{\bar{y}}(\bar{y})\big\}\operatorname{Re}\big\{\bar{y}\big\}^T + \operatorname{Im}\big\{\varphi_{\bar{y}}(\bar{y})\big\}\operatorname{Im}\big\{\bar{y}\big\}^T. \tag{11}$$

Therefore, by Eq. (9), the gradient of Eq. (8) is essentially a sum of two gradients of the *real*-valued information theoretic ICA. Hence, any gradient method for orthogonal real ICA can be used to solve this part of in our generic ICA method using real arithmetic only.

For the subvector with the zero spectral coefficient, a unitary matrix should be found. This is the case assumed by most approaches [2, 3, 5, 11] in literature, and hence any such algorithm may be used. However, one can now take

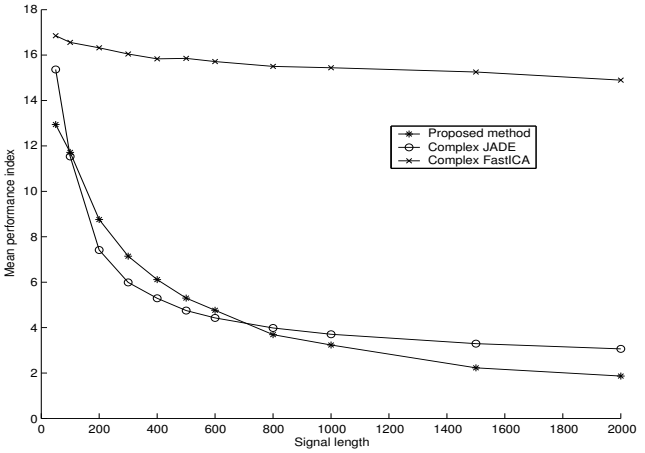

Figure 1: The separation performance of the three methods in terms of the signal length for two circular and three non-circular sources. Complex FastICA is unable to separate the signals although the proposed method that relies partially on real FastICA shows reliable performance.

advantage of the fact that the real and imaginary parts have the same variance. Hence, the first two terms of the Taylor series expansions of the real and imaginary parts of the score function (9) are equal, and it is reasonable to use the same function for the both parts. This is equivalent to assuming that the r.v. is strictly circular, i.e. its density does not depend on the phase information.

8. SIMULATION EXAMPLE

In this section, we present a simulation example using the proposed method for complex ICA. The method is compared to the following widely used ICA methods: the complex JADE algorithm [6] and the complex FastICA algorithm [3]. The quality of separation was measured with the *performance index* [1], $\mathrm{PI} \triangleq \sum_{k=1}^{p}\big(\sum_{l=1}^{p}|c_{kl}|/\max_n|c_{kn}|-1\big)+\sum_{l=1}^{p}\big(\sum_{k=1}^{p}|c_{kl}|/\max_n|c_{nl}|-1\big)$, where $(c_{kl})=\boldsymbol{WA}$. It is zero for a perfect separation. In our generic method, we used the MDL criterion for selecting the circular and non-circular subvectors. The circular subvector was then separated with the complex JADE, and the non-circular subvector with (symmetric) real FastICA [13] with the modifications given by Eq. (9) and Eq. (11). We used `pow3` and `tanh` nonlinearities for real and imaginary parts, respectively. The argument for the both functions was simply the sum of real and imaginary parts of the complex argument, i.e. the Eq. (11) written as $(\bar{y}_R+\bar{y}_I)^3\bar{y}_R^T+\tanh(\bar{y}_R+\bar{y}_I)\bar{y}_I^T$.

Five signals with different lengths were generated for the simulation. The mixing matrix was randomly generated for each of 1000 runs for each signal length. Two second order circular signals were generated with one having standard normal and the other Uniform(0,1) distributed real and imaginary parts. Two non-circular sources were generated as the circular ones but the real part was multiplied by two. Finally, the fifth (non-circular) source had Rayleigh(1) distributed real part and Uniform(0,1) distributed imaginary part. The separation results are shown in Figure 1. It is seen that complex FastICA [3] is unable to separate the sources. The proposed method shows the best performance, which improves as the signal length increases. For a relatively short signals, JADE outperforms the proposed method. This is mainly due to the fact that the MDL criterion requires a large sample size to work reliably.

9. CONCLUSION

In this paper we have described a novel generic method for complex ICA and derived the complex score function. The method does not need any additional a priori assumptions on the type of source signals beyond second order statistics. Moreover, the SUT preprocessing transformation used significantly reduces the original problem and allows the use of ICA methods designed for real-valued signals.

REFERENCES

[1] S. Amari, A. Cichocki, and H. Yang. A new learning algorithm for blind signal separation. In D. Touretsky, M. Mozer, and H. Hasselmo, editors, *Advances in Neural Information Processing Systems 8*, pages 757–763. MIT Press, 1996.

[2] J. Anemüller, T.J. Sejnowski, and S. Makeig. Complex independent component analysis of frequency-domain electroencephalographic data. *Neural Networks*, 16(9):1311–1323, November 2003.

[3] E. Bingham and A. Hyvärinen. A fast fixed-point algorithm for independent component analysis of complex valued signals. *Int. J. Neural Systems*, 10(1):1–8, February 2000.

[4] D.H. Brandwood. A complex gradient operator and its application in adaptive array theory. *IEE Proc.-Pts. F and H*, 130(1):11–16, February 1983.

[5] V. Calhoun and T. Adali. Complex Infomax: Convergence and approximation of Infomax with complex nonlinearities. In *Proc. NNSP 2002*, pages 307–316, Martigny, Switzerland, September 2002.

[6] J-F. Cardoso. An efficient technique for the blind separation of complex sources. In *Proc. HOS'93*, pages 275–279, South Lake Tahoe, CA, June 1993.

[7] A. Cichocki and S. Amari. *Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications*. John Wiley & Sons, 2002.

[8] P. Comon. Independent component analysis, a new concept? *Signal Processing*, 36(3):287–314, April 1994.

[9] J. Eriksson and V. Koivunen. Complex random vectors and ICA models: Identifiability, uniqueness and separability. *IEEE Trans. Inform. Theory*. Submitted (March 2004), see also the PhD thesis: `http://lib.hut.fi/Diss/2004/isbn951227227X/`.

[10] J. Eriksson and V. Koivunen. Complex-valued ICA using second order statistics. In *Proc. MLSP 2004*, pages 183–191, Sao Luis, Brazil, October 2004.

[11] S. Fiori. Extended Hebbian learning for blind separation of complex-valued sources. *IEEE Trans. Circuits Syst. II*, 50(4):195–202, April 2003.

[12] C. Guo and S. Qiao. A stable lanczos tridiagonalization of complex symmetric matrices. Technical Report CAS 03-08-SQ, Department of Computing and Software, McMaster University, June 2003.

[13] A. Hyvärinen, J. Karhunen, and E. Oja. *Independent Component Analysis*. John Wiley & Sons, 2001.

[14] F.D. Neeser and J.L. Massey. Proper complex random processes with applications to information theory. *IEEE Trans. Inform. Theory*, 39(4):1293–1302, July 1993.

[15] A. van den Bos. Complex gradient and Hessian. *IEE Proc.-Vis. Image Signal Process.*, 141(6):380–382, December 1994.

[16] M. Wax and T. Kailath. Detection of signals by information theoretic criteria. *IEEE Trans. Acoust., Speech, Signal Processing*, 33(2):387–392, April 1985.