



Testing blind separability of complex Gaussian mixtures



David Ramírez^{a,*}, Peter J. Schreier^a, Javier Vía^b, Ignacio Santamaría^b

^a Signal and System Theory Group, Universität Paderborn, Paderborn, Germany

^b Communications Engineering Dept., University of Cantabria, Santander, Spain

ARTICLE INFO

Article history:

Received 11 February 2013

Received in revised form

23 June 2013

Accepted 22 August 2013

Available online 30 August 2013

Keywords:

Complex independent component analysis (ICA)

Circularity coefficients

Generalized likelihood ratio test (GLRT)

Hypothesis test

Maximum likelihood (ML) estimation

Wilks' theorem

ABSTRACT

The separation of a complex mixture based solely on second-order statistics can be achieved using the Strong Uncorrelating Transform (SUT) if and only if all sources have distinct circularity coefficients. However, in most problems we do not know the circularity coefficients, and they must be estimated from observed data. In this work, we propose a detector, based on the generalized likelihood ratio test (GLRT), to test the separability of a complex Gaussian mixture using the SUT. For the separable case (distinct circularity coefficients), the maximum likelihood (ML) estimates are straightforward. On the other hand, for the non-separable case (at least one circularity coefficient has multiplicity greater than one), the ML estimates are much more difficult to obtain. To set the threshold, we exploit Wilks' theorem, which gives the asymptotic distribution of the GLRT under the null hypothesis. Finally, numerical simulations show the good performance of the proposed detector and the accuracy of Wilks' approximation.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

The blind separation of a linear mixture of complex independent sources is an important problem with a range of applications, e.g. in biomedical image analysis. See [1–3], and references therein. The Strong Uncorrelating Transform (SUT) allows blind separation based solely on second-order statistics [4–6], provided that these sources correlate with their complex conjugates and that the strengths of these correlations differ from source to source. A complex random variable x that correlates with its complex conjugate x^* has a nonzero complementary covariance $E[x^2]$ and is called improper or noncircular.

The circularity coefficient $k = |E[x^2]|/E[|x|^2]^2$ takes values between 0 and 1 and measures how noncircular or improper a random variable is. This may be illustrated by

the density contours of a univariate complex Gaussian random variable. These contours are ellipses, and the shape of these ellipses is controlled by the circularity coefficients [7]. If a Gaussian random variable has circularity coefficient $k=0$, then its probability density contours are circular [8–10]; if it has circularity coefficient $k=1$, then its probability density contours degenerate into a line in the complex plane.

The circularity coefficients are invariant to linear transformations. Thus, a linear mixture of complex sources has the same set of circularity coefficients as the original sources. This invariance property is exploited by the SUT for blind separation. A necessary and sufficient condition for separability using the SUT is that all circularity coefficients of the sources are distinct. It thus makes intuitive sense that separation of the mixture should be easier if the circularity coefficients are more clearly separated, and it should become more difficult if the circularity coefficients are more clustered. This intuition is supported theoretically by [11].

In practice, the circularity coefficients are not known a priori and must be estimated from the observed data.

* Corresponding author. Tel.: +49 525 160 4367.

E-mail addresses: david.ramirez@sst.upb.de (D. Ramírez).

peter.schreier@sst.upb.de (P.J. Schreier).

jvia@gtas.dicom.unican.es (J. Vía),

nacho@gtas.dicom.unican.es (I. Santamaría).

We are thus confronted with the question whether or not a mixture is separable, based on a given set of observations. This paper deals with this problem by deriving a generalized likelihood ratio test (GLRT) to decide whether a mixture of complex-valued improper signals is separable or not. The test boils down to testing whether all circularity coefficients are distinct or whether there are circularity coefficients with multiplicity greater than one. This paper extends preliminary results reported at a conference [12], where we did not include any details of the rather lengthy proofs.

The structure of our paper is as follows. In Section 2, we review how the SUT enables ICA of complex sources. In Section 3, we formally define our hypothesis testing problem, and in Section 4, we derive the GLRT. Finally, Section 5 presents simulation results that illustrate the performance of our detector.

1.1. Notation

In this paper we use bold-face upper-case letters to denote matrices, with elements $x_{k,l}$ or $[\mathbf{X}]_{k,l}$; bold-face lower-case letters for column vectors, and light-face lower case letters for scalar quantities. The superscripts $(\cdot)^T$ and $(\cdot)^H$ denote transpose and Hermitian transpose, respectively. The determinant and trace of a matrix \mathbf{A} will be denoted, respectively, as $\det(\mathbf{A})$ and $\text{tr}(\mathbf{A})$. The notation $\mathbf{A} \in \mathbb{C}^{M \times N}$ ($\mathbf{A} \in \mathbb{R}^{M \times N}$) will be used to denote that \mathbf{A} is a complex (real) matrix of dimension $M \times N$. For vectors, the notation $\mathbf{x} \in \mathbb{C}^M$ ($\mathbf{x} \in \mathbb{R}^M$) denotes that \mathbf{x} is a complex (real) vector of dimension M , and $\mathbf{x} \sim \mathcal{CN}(\boldsymbol{\mu}, \mathbf{R})$ indicates that \mathbf{x} is a complex circular Gaussian random vector of mean $\boldsymbol{\mu}$ and covariance matrix \mathbf{R} . The expectation operator will be denoted as $E[\cdot]$. The notation \mathbf{I}_L is used to denote the identity matrix of size $L \times L$, whereas $\mathbf{I}_{L \times P}$ is a $L \times P$ matrix with ones in the main diagonal and zeros elsewhere. The matrix $\mathbf{0}_{L \times P}$ denotes the zero matrix of size $L \times P$. We use $\mathbf{A}^{1/2}$ to denote the positive semidefinite square root matrix of the positive semidefinite matrix \mathbf{A} . Finally, $\text{diag}(\mathbf{A})$ is a diagonal matrix formed by the main diagonal of \mathbf{A} and $\text{diag}(\mathbf{a})$ is a diagonal matrix formed by the vector \mathbf{a} .

2. ICA from second-order statistics

In this section, we present a review of independent component analysis (ICA) of complex sources based solely on second-order statistics (SOS). This technique is based on the SUT [4–6]. Let us consider the instantaneous noiseless linear complex ICA model

$$\mathbf{x} = \mathbf{A}\mathbf{s}, \quad (1)$$

where $\mathbf{x} \in \mathbb{C}^P$ are the measurements, $\mathbf{A} \in \mathbb{C}^{P \times P}$ is the *unknown* mixing matrix, assumed to have full rank, and $\mathbf{s} \in \mathbb{C}^P$ are zero-mean sources, which are assumed to be independent. Note that there is the same number of sources and measurements. This is a safe assumption for overdetermined problems since we can always apply a dimensionality reduction technique based on principal component analysis (PCA). On the other hand, the case of fewer measurements than sources can be ignored since there exists no solution using only SOS.

The idea behind ICA is to recover \mathbf{s} without knowledge of \mathbf{A} , utilizing only the linearity of the model and the independence of the sources. For the linear model (1), the sources are recovered as

$$\hat{\mathbf{s}} = \mathbf{B}\mathbf{x}, \quad (2)$$

where \mathbf{B} is the separating matrix. Since the technique is based only on the independence of the sources, there exist some ambiguities. Any scaling of \mathbf{s} , i.e., multiplication with a diagonal matrix, and any reordering of the components of \mathbf{s} , i.e., multiplication with a permutation matrix, preserves independence. Hence, we can obtain \mathbf{B} only up to a multiplication with a monomial matrix, which is the product of a permutation and a diagonal matrix.

Typically, ICA for real sources is based on higher-order statistics, and if there is more than one Gaussian source, it is only possible to recover \mathbf{s} if the sources have some temporal (sample-to-sample) correlation with different autocorrelation functions [13]. Temporally uncorrelated complex sources, on the other hand, may be separated based on SOS, provided that these satisfy certain conditions. For complex random vectors, all the SOS information is contained in two matrices: the covariance matrix $\mathbf{R}_{ss} = E[\mathbf{s}\mathbf{s}^H]$ and the complementary covariance matrix $\tilde{\mathbf{R}}_{ss} = E[\mathbf{s}\mathbf{s}^T]$ [6]. The assumption of independent sources implies a diagonal structure for both the covariance matrix \mathbf{R}_{ss} and the complementary covariance matrix $\tilde{\mathbf{R}}_{ss}$. Moreover, taking into account the ambiguities of the ICA problem, we may even make the stronger assumptions that $\mathbf{R}_{ss} = \mathbf{I}$ and $\tilde{\mathbf{R}}_{ss} = \mathbf{K}$, where $\mathbf{K} = \text{diag}(k_1, \dots, k_p)$ and $1 \geq k_1 \geq \dots \geq k_p \geq 0$. The diagonal elements k_i are the so-called *circularity coefficients* [5], which we will derive momentarily. Under these assumptions, the covariance matrix of the measurements is

$$\mathbf{R}_{xx} = E[\mathbf{x}\mathbf{x}^H] = \mathbf{A}\mathbf{R}_{ss}\mathbf{A}^H = \mathbf{A}\mathbf{A}^H, \quad (3)$$

and the complementary covariance matrix is

$$\tilde{\mathbf{R}}_{xx} = E[\mathbf{x}\mathbf{x}^T] = \mathbf{A}\tilde{\mathbf{R}}_{ss}\mathbf{A}^T = \mathbf{A}\mathbf{K}\mathbf{A}^T. \quad (4)$$

To recover \mathbf{s} , the separating matrix \mathbf{B} must simultaneously diagonalize \mathbf{R}_{xx} and $\tilde{\mathbf{R}}_{xx}$, i.e., both $\mathbf{B}\mathbf{R}_{xx}\mathbf{B}^H$ and $\mathbf{B}\tilde{\mathbf{R}}_{xx}\mathbf{B}^T$ must be diagonal. To this end, we first compute the coherence matrix

$$\mathbf{C} = \mathbf{R}_{xx}^{-1/2}\tilde{\mathbf{R}}_{xx}(\mathbf{R}_{xx}^*)^{-H/2} = \mathbf{R}_{xx}^{-1/2}\tilde{\mathbf{R}}_{xx}\mathbf{R}_{xx}^{-T/2}, \quad (5)$$

which appears in the canonical correlation analysis (CCA) [14] of the vectors \mathbf{x} and \mathbf{x}^* [6]. Then we obtain the Takagi factorization [15] of \mathbf{C} , which is a special singular value decomposition for complex and symmetric (not Hermitian symmetric) matrices $\mathbf{C} = \mathbf{C}^T$:

$$\mathbf{C} = \mathbf{F}\mathbf{K}\mathbf{F}^T, \quad (6)$$

where $\mathbf{F} \in \mathbb{C}^{P \times P}$ is a unitary matrix and $\mathbf{K} = \text{diag}(k_1, \dots, k_p)$ is a diagonal matrix that contains the circularity coefficients. These circularity coefficients are the canonical correlations between \mathbf{x} and \mathbf{x}^* . The separating matrix is now given by the SUT

$$\mathbf{B} = \mathbf{F}^H\mathbf{R}_{xx}^{-1/2}. \quad (7)$$

The complex ICA model is separable if and only if all circularity coefficients are distinct [4,5]. Hence, it is possible to

separate more than one complex Gaussian source provided that they all have distinct circularity coefficients. Sources corresponding to identical circularity coefficients will end up in a unitary residual mixture that cannot be separated.

3. Problem formulation

In practice, we typically do not know the true circularity coefficients and have to work with their estimates, which are affected by estimation errors. Hence, even if some of the true circularity coefficients are identical, it is very likely that their estimates are not. This raises the question of how to decide whether a complex mixture can be separated using the SUT, based on estimated circularity coefficients. This question may be cast as a hypothesis test, assuming Gaussian signals. This assumption leads to a tractable analysis and useful detectors. Moreover, a Gaussian can also be seen as a worst case [16] or least informative (maximum entropy) [17] distribution.

We formulate the test as follows. Given a finite set of observations $\{\mathbf{x}_n\}_{n=0}^{N-1}$,

- \mathcal{H}_1 : The model is separable using the SUT,
 \mathcal{H}_0 : The model is not separable using the SUT. (8)

Let us now formally express this test. We know the model is not separable if two or more circularity coefficients are equal, that is, two or more elements of $\tilde{\mathbf{R}}_{ss}$ are equal. Let \mathbb{D}_+ be the set of $P \times P$ diagonal matrices with elements in $[0, 1]$, and define \mathbb{D}_{2+} as that subset of \mathbb{D}_+ where at least two of the diagonal entries are identical. Then, the hypothesis test can be written as

- \mathcal{H}_1 : $\tilde{\mathbf{R}}_{ss} \in \mathbb{D}_+$,
 \mathcal{H}_0 : $\tilde{\mathbf{R}}_{ss} \in \mathbb{D}_{2+}$. (9)

Under the alternative hypothesis, the complementary covariance matrix is constrained to be diagonal with elements in $[0, 1]$. Under the null hypothesis, the complementary covariance matrix satisfies the same constraints, but it has *at least* two repeated entries. Both hypotheses are composite since the circularity coefficients are unknown, and they remain composite even after applying invariance techniques to reduce the number of unknown parameters [18–20]. Using the Gaussianity assumption, the hypothesis test (9) becomes

- \mathcal{H}_1 : $\mathbf{x} \sim \mathcal{CN}(\mathbf{0}, \mathbf{R}_{xx}^{(1)})$,
 \mathcal{H}_0 : $\mathbf{x} \sim \mathcal{CN}(\mathbf{0}, \mathbf{R}_{xx}^{(0)})$, (10)

where the augmented covariance matrices [6] are

$$\mathbf{R}_{xx}^{(i)} = E[\mathbf{x}\mathbf{x}^H] = \begin{bmatrix} \mathbf{R}_{xx}^{(i)} & \tilde{\mathbf{R}}_{xx}^{(i)} \\ \tilde{\mathbf{R}}_{xx}^{(i)*} & \mathbf{R}_{xx}^{(i)*} \end{bmatrix}, \quad i = \{0, 1\}, \quad (11)$$

and $\mathbf{x} = [\mathbf{x}^T \ \mathbf{x}^{H*}]^T$ is an augmented vector constructed from placing \mathbf{x} on top of its complex conjugate \mathbf{x}^* . Recalling the results from the previous section, the covariance and complementary covariance matrices are

$$\mathbf{R}_{xx}^{(i)} = \mathbf{A}\mathbf{A}^H, \quad \tilde{\mathbf{R}}_{xx}^{(i)} = \mathbf{A}\mathbf{K}_i\mathbf{A}^T, \quad (12)$$

with diagonal

$$\mathbf{K}_1 \in \mathbb{D}_+, \quad \mathbf{K}_0 \in \mathbb{D}_{2+}. \quad (13)$$

We are therefore testing the covariance structure of the augmented vector \mathbf{x} .

4. Derivation of the GLRT

To solve the hypothesis test (10), we propose a generalized likelihood ratio test (GLRT), which usually results in a simple detector with good performance [21]. The generalized likelihood ratio is [21]

$$\mathcal{G} = \frac{\max_{\mathbf{A}, \mathbf{K}_0 \in \mathbb{D}_{2+}} p(\mathbf{X}; \mathbf{A}, \mathbf{K}_0)}{\max_{\mathbf{A}, \mathbf{K}_1 \in \mathbb{D}_+} p(\mathbf{X}; \mathbf{A}, \mathbf{K}_1)}, \quad (14)$$

where the data matrix is $\mathbf{X} = [\mathbf{x}_0, \dots, \mathbf{x}_{N-1}]$, and $p(\cdot)$ denotes the probability density function of the observations. The first step in the derivation is to find the maximum likelihood (ML) estimates of the unknown parameters under each hypothesis. In order to do that, let us introduce the log-likelihood¹

$$\log p(\mathbf{X}; \mathbf{R}_{xx}^{(i)}) = -\log \det(\mathbf{R}_{xx}^{(i)}) - \text{tr}(\mathbf{R}_{xx}^{(i)-1} \hat{\mathbf{R}}), \quad (15)$$

where the augmented sample covariance matrix is

$$\hat{\mathbf{R}} = \frac{1}{N} \mathbf{X}\mathbf{X}^H = \begin{bmatrix} \hat{\mathbf{R}} & \hat{\mathbf{R}} \\ \hat{\mathbf{R}}^* & \hat{\mathbf{R}}^* \end{bmatrix} \in \mathcal{R}.$$

Here, we have introduced the augmented data matrix $\mathbf{X} = [\mathbf{X}^T \ \mathbf{X}^{H*}]^T$, and \mathcal{R} denotes the set of augmented covariance matrices without further structure. In the next subsections we find the ML estimates for the unknown parameters and derive a closed-form GLRT.

4.1. ML estimates under the alternative hypothesis

We are only interested in the maximum value of the likelihood, attained by substituting the unknown parameters with their maximum likelihood estimates, rather than the estimates themselves. We also note that the augmented covariance matrix under the alternative hypothesis does not have further structure beyond being an augmented covariance matrix. This means that

$$\max_{\mathbf{A}, \mathbf{K}_1 \in \mathbb{D}_+} p(\mathbf{X}; \mathbf{A}, \mathbf{K}_1) = \max_{\mathbf{R}_{xx}^{(1)} \in \mathcal{R}} p(\mathbf{X}; \mathbf{R}_{xx}^{(1)}). \quad (16)$$

For this, we only need to obtain the ML estimate of an augmented covariance matrix, which is presented next.

Lemma 1. *The ML estimate of $\mathbf{R}_{xx}^{(1)}$ is*

$$\hat{\mathbf{R}}_{xx}^{(1)} = \hat{\mathbf{R}}. \quad (17)$$

Proof. The proof can be found, for instance, in [6]. \square

Inserting this ML estimate into (15), the compressed log-likelihood becomes

$$\log p(\mathbf{X}; \hat{\mathbf{R}}_{xx}^{(1)}) = -\log \det(\hat{\mathbf{R}}), \quad (18)$$

¹ For the sake of notational simplicity, we will omit additive and multiplicative constants that do not depend on the data.

which can be rewritten as

$$\log p(\mathbf{X}; \hat{\mathbf{R}}_{xx}^{(1)}) = -\log \det(\hat{\mathbf{R}}) - \frac{1}{2} \sum_{i=1}^P \log(1 - \hat{k}_i^2). \quad (19)$$

In this expression, \hat{k}_i are the estimated circularity coefficients, which are given by the singular values of the estimated coherence matrix

$$\hat{\mathbf{C}} = \hat{\mathbf{R}}^{-1/2} \hat{\mathbf{R}} \hat{\mathbf{R}}^{-T/2}. \quad (20)$$

4.2. ML estimates under the null hypothesis

The ML estimates under the null hypothesis are much more difficult to obtain. This is because we do not know the number of circularity coefficients with multiplicity greater than one, and we do not know their multiplicities, either. Nevertheless, it is well known that fewer constraints lead to greater likelihood, so we can write

$$\max_{\mathbf{A}, \mathbf{K}_0 \in \mathbb{D}_{2+}} p(\mathbf{X}; \mathbf{A}, \mathbf{K}_0) = \max_{\mathbf{A}, \mathbf{K}_0 \in \cup_i \mathbb{D}_2^i} p(\mathbf{X}; \mathbf{A}, \mathbf{K}_0), \quad (21)$$

where $\mathbb{D}_2^i, i = 1, \dots, P-1$, denotes the subset of \mathbb{D}_{2+} where the i th and $(i+1)$ th entries are identical, and $\cup_i \mathbb{D}_2^i$ denotes the union of those sets. That is, the maximum is achieved when exactly two circularity coefficients are identical, while the others may vary. This is the worst case for \mathcal{H}_0 , because it is closest to \mathcal{H}_1 . The test would obviously perform better if we knew in advance the multiplicity for each repeated circularity coefficient. We may simplify the above equation to obtain

$$\max_{\mathbf{A}, \mathbf{K}_0 \in \mathbb{D}_{2+}} p(\mathbf{X}; \mathbf{A}, \mathbf{K}_0) = \max_i \max_{\mathbf{A}, \mathbf{K}_0 \in \mathbb{D}_2^i} p(\mathbf{X}; \mathbf{A}, \mathbf{K}_0). \quad (22)$$

The unknown parameters in this equation are the circularity coefficients and the mixing matrix. We will proceed as follows. We will establish the perhaps unsurprising result that the ML estimates of the circularity coefficients are the sample circularity coefficients with the exception of the repeated coefficient, whose ML estimate is given by the average of the two corresponding estimated coefficients. Furthermore, rather than estimating the mixing matrix itself, we will estimate a linear transformation of it, which simplifies the derivations. To be specific, we estimate the product of the mixing matrix and the SUT, which results in an ML estimate of the inverse mixing matrix given by the SUT, with the i th and $(i+1)$ th row scaled by some constants depending on the sample circularity coefficients. In (22), we first obtain the ML estimates for a fixed i , and then vary i from 1 to $P-1$.

We shall start with the augmented covariance matrix, given by

$$\mathbf{R}_{xx}^{(0)} = \underbrace{\begin{bmatrix} \mathbf{A} & \mathbf{0}_p \\ \mathbf{0}_p & \mathbf{A}^* \end{bmatrix}}_{\mathbf{A}} \underbrace{\begin{bmatrix} \mathbf{I} & \mathbf{K}_0 \\ \mathbf{K}_0 & \mathbf{I} \end{bmatrix}}_{\mathbf{K}_0} \underbrace{\begin{bmatrix} \mathbf{A}^H & \mathbf{0}_p \\ \mathbf{0}_p & \mathbf{A}^T \end{bmatrix}}_{\mathbf{A}^H},$$

where $\mathbf{K}_0 = \text{diag}(k_1, \dots, k_i, k_i, \dots, k_p)$. This allows us to rewrite the log-likelihood as

$$\log p(\mathbf{X}; \mathbf{A}, \mathbf{K}_0) = \log \det(\mathbf{A}^{-1} \mathbf{A}^{-H}) - \log \det(\mathbf{K}_0) - \text{tr}(\mathbf{K}_0^{-1} \mathbf{S}), \quad (23)$$

where

$$\mathbf{S} = \begin{bmatrix} \mathbf{S} & \tilde{\mathbf{S}} \\ \mathbf{S}^* & \mathbf{S}^* \end{bmatrix} = \begin{bmatrix} \mathbf{A}^{-1} \hat{\mathbf{R}} \mathbf{A}^{-H} & \mathbf{A}^{-1} \hat{\mathbf{R}} \mathbf{A}^{-T} \\ \mathbf{A}^{-*} \hat{\mathbf{R}}^* \mathbf{A}^{-H} & \mathbf{A}^{-*} \hat{\mathbf{R}}^* \mathbf{A}^{-T} \end{bmatrix}$$

is the augmented covariance matrix of $\mathbf{A}^{-1} \mathbf{x}$. We use Takagi's factorization of the sample coherence matrix

$$\hat{\mathbf{C}} = \hat{\mathbf{R}}^{-1/2} \hat{\mathbf{R}} \hat{\mathbf{R}}^{-T/2} = \hat{\mathbf{K}} \hat{\mathbf{F}} \hat{\mathbf{F}}^T, \quad (24)$$

where $\hat{\mathbf{K}}$ is a diagonal matrix containing the sample circularity coefficients. We also introduce the matrix $\mathbf{W}^{-1} = \hat{\mathbf{B}} \mathbf{A} = \hat{\mathbf{F}}^H \hat{\mathbf{R}}^{-1/2} \mathbf{A}^{-1}$, which may be seen as the residual matrix, i.e., the combined effect of mixing and separating matrices. In case of perfect separation, it should be a monomial matrix. Then, (23) may be rewritten as

$$\log p(\mathbf{X}; \mathbf{W}, \mathbf{K}_0) = -\log \det(\hat{\mathbf{R}}) + \log \det(\mathbf{W} \mathbf{W}^H) - \frac{1}{2} \log \det(\mathbf{K}_0) - \frac{1}{2} \text{tr}(\mathbf{K}_0^{-1} \mathbf{S}). \quad (25)$$

We now need to maximize (25) with respect to the transformed parameters \mathbf{W} and \mathbf{K}_0 .

First, we find the ML estimate of \mathbf{K}_0 , for which the repeated entry occurs in \mathbf{K}_0 at positions i and $i+1$. Taking into account the problem invariances, we may relax the restrictions that \mathbf{K}_0 must have 1s on the main diagonal and that the diagonal elements of \mathbf{K}_0 must be in $[0, 1]$, since these can also be imposed when estimating \mathbf{A} (or \mathbf{W}). The ML estimate is presented in the following lemma.

Lemma 2. The ML estimate of \mathbf{K}_0 is given by

$$\hat{\mathbf{K}}_0 = \begin{bmatrix} \hat{\mathbf{I}} & \hat{\mathbf{K}}_0 \\ \hat{\mathbf{K}}_0^* & \hat{\mathbf{I}}^* \end{bmatrix}, \quad (26)$$

where $\hat{\mathbf{I}}$ and $\hat{\mathbf{K}}_0$ are diagonal matrices composed by the diagonal elements of \mathbf{S} and $\tilde{\mathbf{S}}$, respectively, with the exception of the i th and $(i+1)$ th entries of either matrix, which are replaced by their respective averages. That is,

$$\hat{\mathbf{I}} = \text{diag}(s_{11}, \dots, \gamma, \gamma, \dots, s_{p,p}), \quad \hat{\mathbf{K}}_0 = \text{diag}(\tilde{s}_{11}, \dots, \tilde{\gamma}, \tilde{\gamma}, \dots, \tilde{s}_{p,p}), \quad (27)$$

with the averages given by

$$\gamma = \frac{1}{2}(s_{i,i} + s_{i+1,i+1}), \quad \tilde{\gamma} = \frac{1}{2}(\tilde{s}_{i,i} + \tilde{s}_{i+1,i+1}). \quad (28)$$

Proof. Let us introduce a simple permutation of the elements of \mathbf{x} (which does not modify the value of the likelihood), such that

$$\bar{\mathbf{K}} = \mathbf{P} \mathbf{K} \mathbf{P}^T = \begin{bmatrix} \bar{\mathbf{K}}_{1,1} & \mathbf{0}_2 & \cdots & \mathbf{0}_2 \\ \mathbf{0}_2 & \bar{\mathbf{K}}_{2,2} & \cdots & \mathbf{0}_2 \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_2 & \mathbf{0}_2 & \cdots & \bar{\mathbf{K}}_{p,p} \end{bmatrix} \quad (29)$$

where

$$\bar{\mathbf{K}}_{ii} = E \left[\begin{bmatrix} s_i \\ s_i^* \end{bmatrix} \begin{bmatrix} s_i^* & s_i \end{bmatrix} \right], \quad (30)$$

and, consequently,

$$\bar{\mathbf{S}} = \begin{bmatrix} \bar{\mathbf{S}}_{1,1} & \bar{\mathbf{S}}_{1,2} & \cdots & \bar{\mathbf{S}}_{1,P} \\ \bar{\mathbf{S}}_{2,1} & \bar{\mathbf{S}}_{2,2} & \cdots & \bar{\mathbf{S}}_{2,P} \\ \vdots & \vdots & \ddots & \vdots \\ \bar{\mathbf{S}}_{P,1} & \bar{\mathbf{S}}_{P,2} & \cdots & \bar{\mathbf{S}}_{P,P} \end{bmatrix} = \mathbf{P} \underline{\mathbf{S}} \mathbf{P}^T. \quad (31)$$

Therefore, (25) becomes

$$\log p(\mathbf{X}; \mathbf{A}, \hat{\mathbf{K}}_0) = -\log \det(\hat{\mathbf{R}}) + \log \det(\mathbf{W}\mathbf{W}^H) - \frac{1}{2} \log \det(\hat{\mathbf{K}}_0) - \frac{1}{2} \text{tr}(\hat{\mathbf{K}}_0^{-1} \bar{\mathbf{S}}). \quad (32)$$

The likelihood in (32) is maximized with respect to $\bar{\mathbf{K}}$ under the constraint $\bar{\mathbf{K}}_{i,i} = \bar{\mathbf{K}}_{i+1,i+1}$ by the block-diagonal matrix

$$\hat{\bar{\mathbf{K}}} = \text{diag}(\bar{\mathbf{S}}_{1,1}, \dots, \Gamma, \Gamma, \dots, \bar{\mathbf{S}}_{P,P}), \quad (33)$$

where

$$\Gamma = \frac{1}{2}(\bar{\mathbf{S}}_{i,i} + \bar{\mathbf{S}}_{i+1,i+1}) \quad (34)$$

The proof is concluded after inverting the permutation. \square

Using (26) the compressed likelihood is

$$\log p(\mathbf{X}; \mathbf{W}, \hat{\mathbf{K}}_0) = -\log \det(\hat{\mathbf{R}}) + \log \det(\mathbf{W}\mathbf{W}^H) - \frac{1}{2} \log \det(\hat{\mathbf{K}}_0), \quad (35)$$

which has to be maximized with respect to \mathbf{W} . Here, we only give the result and relegate the proof to an appendix.

Lemma 3. *The ML estimate of \mathbf{W} is given by*

$$\hat{\mathbf{W}} = \text{diag} \left(\underbrace{1, \dots, 1}_{i-1 \text{ ones}}, \sqrt{1+\chi}, \sqrt{1-\chi}, \underbrace{1, \dots, 1}_{P-i-1 \text{ ones}} \right) \quad (36)$$

where

$$\chi = \frac{2 - (\hat{k}_i^2 + \hat{k}_{i+1}^2) - 2\sqrt{(1-\hat{k}_i^2)(1-\hat{k}_{i+1}^2)}}{(\hat{k}_i^2 - \hat{k}_{i+1}^2)}. \quad (37)$$

Proof. See Appendix A. \square

Using the ML estimate $\hat{\mathbf{W}}$, it is easy to show that the ML estimate of the inverse mixing matrix—the separation matrix—is

$$\hat{\mathbf{A}}^{-1} = \hat{\mathbf{W}} \hat{\mathbf{F}}^H \hat{\mathbf{R}}^{-1/2}. \quad (38)$$

Hence, the ML estimate of the inverse mixing matrix is just the estimated SUT, but with its i th and $(i+1)$ th rows scaled.

Finally, let $\hat{h}_i = 1 - \hat{k}_i$ and insert the ML estimates into the likelihood. Then, after some tedious but straightforward algebra, the compressed likelihood becomes

$$\log p(\mathbf{X}, \hat{\mathbf{W}}, \hat{\mathbf{K}}_0) = \max_i \left\{ -\log \det(\hat{\mathbf{R}}) - \sum_{\substack{l=1 \\ l \neq i, i+1}}^P \log \hat{h}_l + \log \left(\frac{2}{\hat{h}_i \hat{h}_{i+1} - \hat{k}_i \hat{k}_{i+1} + 1} \right) \right\}. \quad (39)$$

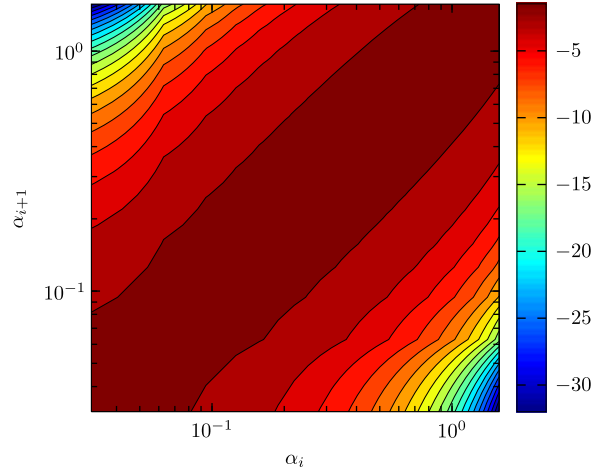


Fig. 1. Plot of the contours of \mathcal{G} (in linear scale) as a function of α_i and α_{i+1} . (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

4.3. A closed-form GLRT

Using the transformed circularity coefficients \hat{h}_i , the compressed likelihood under \mathcal{H}_1 , given by (19), may be rewritten as

$$\log p(\mathbf{X}; \hat{\mathbf{R}}_{xx}^{(1)}) = -\log \det(\hat{\mathbf{R}}) - \sum_{i=1}^P \log \hat{h}_i. \quad (40)$$

The GLRT is finally obtained by inserting the expressions for the compressed likelihoods, given by (39) and (40), into (14). This yields the GLRT

$$\mathcal{G} = \max_i \left\{ f(\hat{k}_i^{-1}, \hat{k}_{i+1}^{-1}) - f(\hat{k}_i, \hat{k}_{i+1}) \right\} \underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\geq}} \eta, \quad (41)$$

where

$$f(a, b) = [(1-a^2)(1-b^2)]^{-1/2}. \quad (42)$$

An alternative form of the GLRT may be derived as follows. Because \hat{k}_i , which is the canonical correlation between \hat{s}_i and \hat{s}_i^* , is bounded between 0 and 1, it may also be expressed as $\hat{k}_i = \cos(\alpha_i)$. The angle α_i is the estimated principal angle [22], i.e., the angle between \hat{s}_i and \hat{s}_i^* . The GLRT may be alternatively expressed in terms of the principal angles as

$$\mathcal{G} = \max_i \left\{ \frac{\cos(\alpha_i - \alpha_{i+1}) + \cos(\alpha_i + \alpha_{i+1}) - 2}{\cos(\alpha_i - \alpha_{i+1}) - \cos(\alpha_i + \alpha_{i+1})} \right\} \underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\geq}} \eta. \quad (43)$$

This shows that the GLR is only a function of the sum and difference of principal angles.

The GLR is depicted in Fig. 1, which shows a plot of \mathcal{G} as a function of α_i and α_{i+1} for some fixed i . Different colors denote the critical regions for different values of the threshold. As can be seen in the figure, large values of \mathcal{G} are obtained for $\alpha_i \approx \alpha_{i+1}$, which agrees with the fact that \mathcal{H}_0 is selected for large values of \mathcal{G} . The largest value of the GLR is $\mathcal{G} = -1$ for $\alpha_i = \alpha_{i+1}$.

4.4. Asymptotic distribution of the GLR

In order to set the threshold η for the GLRT, we need to determine the probability of false alarm p_{fa} . For a

hypothesis test with composite hypotheses, the probability of false alarm is determined by the worst-case selection of the unknown parameters [18]. In our case this happens when exactly two circularity coefficients are different. Hence,

$$p_{fa} = \sup_{\mathbf{K}_0 \in \mathbb{D}_{2+}} p_{fa}(\mathbf{K}_0), \quad (44)$$

where $p_{fa}(\mathbf{K}_0)$ denotes the probability of false alarm for a fixed matrix \mathbf{K}_0 . As one might expect, deriving the exact distribution of the GLR is very difficult. Instead, we employ Wilks' Theorem [23], which states that, under some regularity conditions, the log-GLR under \mathcal{H}_0 has an asymptotic χ^2_ρ distribution with ρ degrees of freedom, where ρ is the difference between the number of unknown parameters under \mathcal{H}_1 and \mathcal{H}_0 .

The number of free parameters under \mathcal{H}_1 is $2P^2 + P$, accounting for $2P^2$ real elements in \mathbf{A} , and P circularity coefficients. Under \mathcal{H}_0 , there is one repeated circularity coefficient, so there are only $P-1$ degrees of freedom for choosing the circularity coefficients. Moreover, there is a *unitary* residual mixture of the two sources that have the same circularity coefficient, which takes away one further degree of freedom. In total, there are $2P^2 + P - 2$ degrees of freedom under \mathcal{H}_0 . Wilks' Theorem now says that, as the number of observations tends to infinity, $N \rightarrow \infty$, the log-GLR is distributed as²

$$-2 \log \mathcal{G} = 2N \log \left(2 \frac{\hat{h}_i \hat{h}_{i+1} - \hat{k}_i \hat{k}_{i+1} + 1}{\hat{h}_i \hat{h}_{i+1}} \right) \stackrel{a}{\sim} \chi^2_2. \quad (45)$$

We can use this asymptotic distribution to determine the threshold for a given probability of false alarm.

One final comment is in order. Other works, e.g., [24,25], obtain the threshold through simulations. At first, this might seem useless, but in problems where the distribution of the statistic under \mathcal{H}_0 does not depend on unknown parameters, it does, in fact, make sense. Nevertheless, for our particular problem, the distribution under the null hypothesis depends on unknown parameters, so we may not use this approach.

5. Numerical results

In this section, we evaluate the performance of the proposed GLRT using Monte Carlo simulations. First, we present receiver operating characteristic (ROC) curves and probability of missed detection. Second, we evaluate the accuracy of Wilks' approximation of the null distribution for a finite number of samples.

5.1. Performance of the GLRT

We consider $P=4$ complex Gaussian sources. In the first four examples, the circularity coefficients under \mathcal{H}_1 are 0.9, 0.6, 0.35 and 0.1. Under \mathcal{H}_0 , we consider different numbers and orders of repeated circularity coefficients:

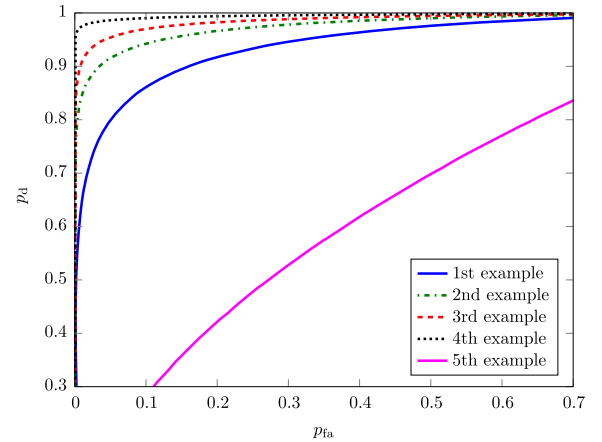


Fig. 2. ROC curves for five different examples with $P=4$ sources and $N=250$ samples.

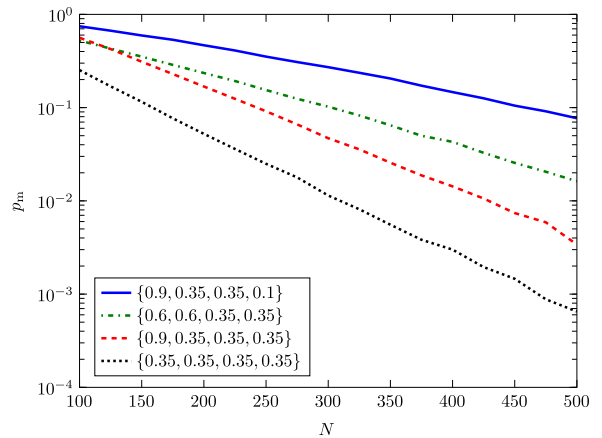


Fig. 3. Probability of missed detection vs. number of samples in the first four examples. The number of sources is $P=4$, and we consider a fixed probability of false alarm $p_{fa} = 0.01$.

1. 0.9, 0.35, 0.35 and 0.1 (two identical circularity coefficients).
2. 0.6, 0.6, 0.35 and 0.35 (two pairs of circularity coefficients).
3. 0.9, 0.35, 0.35 and 0.35 (three identical circularity coefficients).
4. 0.35, 0.35, 0.35 and 0.35 (four identical circularity coefficients).

Finally, in the fifth example, the circularity coefficients under \mathcal{H}_1 are 0.9, 0.6, 0.35 and 0.25 and 0.9, 0.6, 0.35 and 0.35 under \mathcal{H}_0 . That is, under \mathcal{H}_1 the circularity coefficients are fairly close.

Fig. 2 shows the receiver operating characteristic (ROC) curves for $N=250$ samples. As there is no other existing test for this problem, we cannot compare our results with any competitor. We notice that Example 4 performs best since all circularity coefficients are equal and, therefore, the two hypotheses are most separated. We also notice that three identical circularity coefficients are easier to detect than two pairs of identical circularity coefficients. Moreover, the case with fairly close circularity coefficients

² We now have to take into account the previously ignored additive and multiplicative constant terms.

performs the worst. Of course, these statements only hold for this particular setup, and the results will change depending on the specific values of the circularity coefficients, in particular, how much separated they are. Fig. 3 shows the probability of missed detection p_m vs. number of samples for a fixed probability of false alarm $p_{fa} = 0.01$, where similar conclusions can be drawn. Most notably, the slope or error exponent [17] differs among the four examples. This supports our intuition about the separation between the hypotheses, since the error exponent is given by the Kullback–Leibler divergence [17], which can be considered a “distance” between probability density functions.

5.2. Distribution of the GLR under the null hypothesis

In this section, we analyze Wilks’ χ^2 approximation of the GLR distribution under \mathcal{H}_0 . We consider $P=4$ signals with circularity coefficients 0.9, 0.35, 0.35 and 0.1. In this example, exactly two circularity coefficients are identical, which is the worst case in terms of probability of false alarm. First, we consider a large number of samples, $N=1000$, and obtain the empirical cumulative distribution function (ECDF) and the χ^2_2 approximation. As can be seen in Fig. 4, the approximation is very good.

While the χ^2 approximation is an asymptotic result, it works reasonably well even in the cases of small to moderate number of samples. This can be seen in Fig. 5a, which compares the empirically determined threshold required to achieve $p_{fa} = 0.1$ with the threshold obtained using Wilks’ theorem, for different sample sizes N . Nevertheless, one might be concerned that errors in the selection of the threshold may lead to larger than desired probabilities of false alarm. This concern is unfounded, though, as can be verified in Fig. 5b: The actual value of the probability of false alarm is smaller than the asymptotic value, yet obviously at the expense of a smaller probability of detection, as can be seen in Fig. 5c.

6. Conclusions

ICA of complex sources based on second-order statistics can be achieved using the strong uncorrelating transform if and only if all sources have distinct circularity coefficients. However, in most practical applications, we do not know the circularity coefficients. We have developed a generalized likelihood ratio test (GLRT) for separability. The maximum likelihood estimates under the alternative hypothesis “all circularity coefficients are distinct” are straightforward, but under the null hypothesis, their derivation is rather involved. The threshold selection for the proposed GLRT is a difficult problem since (1) we cannot obtain the theoretical distributions and (2) even after applying invariance techniques \mathcal{H}_0 remains composite, which prevents the use of simulations to select the threshold. Hence, we have used Wilks’ theorem to derive the asymptotic distribution of the statistic under the null hypothesis. Finally, simulation results show the good performance of the detector and the accuracy of the approximation of the distribution under \mathcal{H}_0 , even for small and moderate number of samples.

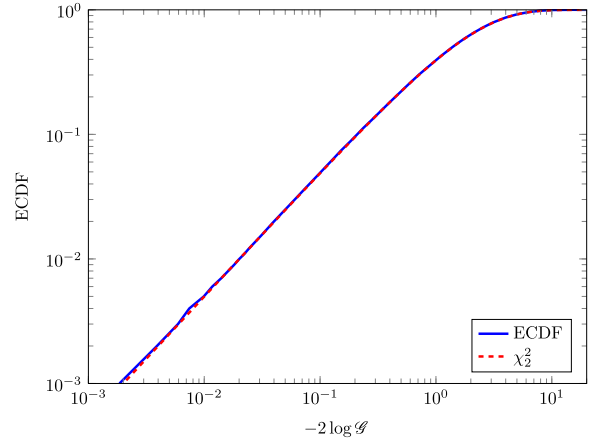


Fig. 4. Empirical cumulative distribution function (ECDF) and χ^2 approximation for an example with $P=4$ sources with circularity coefficients {0.9, 0.35, 0.35, 0.1}, and $N=1000$ samples.

Acknowledgments

The work of P. Schreier was supported by the Alfried Krupp von Bohlen und Halbach Foundation, under its program “Return of German scientists from abroad”. The work of I. Santamaría and J. Vía was supported by the Spanish Government, Ministerio de Ciencia e Innovación (MICINN), under project COSIMA (TEC2010-19545-C04-03) and project COMONSENS (CSD2008-00010, CONSOLIDER-INGENIO 2010 Program).

Appendix A. Proof of lemma 3

One of the indeterminacies of the ICA problem is the energy of the sources. In the problem statement we included a unit-energy constraint for the sources. However, in the ML estimation of the covariance matrices we drop that constraint since it is easier to constrain the norm of the columns of the mixing matrix. After some manipulations, these constraints are equivalent to

$$s_{l,l} = \|\mathbf{w}_l\|^2 = 1, \quad \tilde{s}_{l,l} = \sum_{m=1}^P \hat{k}_m w_{l,m}^2 \in [0, 1], \quad (A.1)$$

for $l \neq i, i+1$, and

$$\gamma = \frac{1}{2}(\|\mathbf{w}_i\|^2 + \|\mathbf{w}_{i+1}\|^2) = 1, \quad \tilde{\gamma} = \frac{1}{2} \sum_{l=i}^{i+1} \sum_{m=1}^P \hat{k}_m w_{l,m}^2 \in [0, 1], \quad (A.2)$$

where \mathbf{w}_l denotes the l th row of the matrix \mathbf{W} . Without loss of generality, we introduce a permutation such that $\mathbf{W} = [\mathbf{W}_A \quad \mathbf{W}_B]^T$ with

$$\mathbf{W}_A = \begin{bmatrix} \mathbf{w}_1 \\ \vdots \\ \mathbf{w}_{i-1} \\ \mathbf{w}_{i+2} \\ \vdots \\ \mathbf{w}_P \end{bmatrix}, \quad \mathbf{W}_B = \begin{bmatrix} \mathbf{w}_i \\ \mathbf{w}_{i+1} \end{bmatrix}. \quad (A.3)$$

That is, we group together all rows of \mathbf{W} that correspond to sources with distinct circularity coefficients. With this

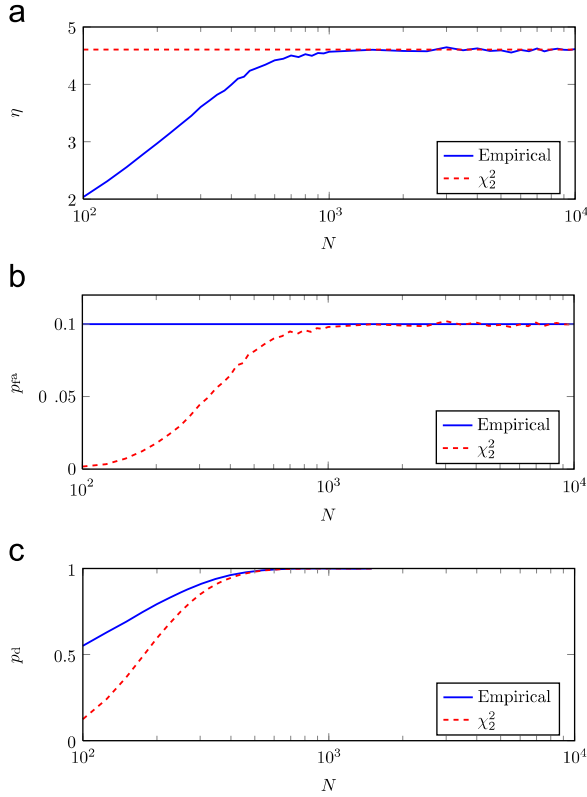


Fig. 5. Comparison between empirical results and Wilks' approximation. (a) Threshold required to achieve $p_{fa} = 0.1$ for different number of samples using the χ^2 approximation and the empirically determined value. (b) Probability of false alarm for different number of samples using the χ^2 approximation and the empirically determined value. (c) Probability of detection for different number of samples using the threshold obtained through the χ^2 approximation and the empirically determined value.

permutation the constraints become

$$\text{diag}(\mathbf{W}_A \mathbf{W}_A^H) = \mathbf{I}_{P-2}, \quad [\mathbf{W}_A \hat{\mathbf{K}} \mathbf{W}_A^T]_{l,l} \in [0, 1], \quad (\text{A.4})$$

$$\frac{1}{2} \text{tr}(\mathbf{W}_B \mathbf{W}_B^H) = 1, \quad \frac{1}{2} \text{tr}(\mathbf{W}_B \hat{\mathbf{K}} \mathbf{W}_B^T) \in [0, 1], \quad (\text{A.5})$$

where $l = 1, \dots, P-2$. Now we may write the compressed likelihood in (35) as

$$\begin{aligned} \log p(\mathbf{X}; \mathbf{W}_A, \mathbf{W}_B, \hat{\mathbf{K}}_0) &= -\log \det(\hat{\mathbf{R}}) + \log \det(\mathbf{W} \mathbf{W}^H) \\ &\quad - \frac{1}{2} \log \det[\mathbf{I}_{P-2} - \text{diag}(\mathbf{W}_A \hat{\mathbf{K}} \mathbf{W}_A^T)] - \log \left[1 - \frac{1}{4} \text{tr}^2(\mathbf{W}_B \hat{\mathbf{K}} \mathbf{W}_B^T) \right]. \end{aligned} \quad (\text{A.6})$$

Lemma 4. The likelihood given by (A.6) subject to the constraints (A.4) and A.5) is maximized for $\mathbf{W}_B \mathbf{W}_B^H = \mathbf{0}_{2 \times P}$.

Proof. The expression for the determinant of a block matrix allows us to write

$$\det(\mathbf{W} \mathbf{W}^H) = \det(\mathbf{W}_A \mathbf{W}_A^H) \det(\mathbf{W}_B \mathbf{W}_B^H) \det(\mathbf{I}_P - \mathbf{Q}), \quad (\text{A.7})$$

where

$$\mathbf{Q} = (\mathbf{W}_B \mathbf{W}_B^H)^{-1/2} \mathbf{W}_B \mathbf{W}_A^H (\mathbf{W}_A \mathbf{W}_A^H)^{-1} \mathbf{W}_A \mathbf{W}_B^H (\mathbf{W}_B \mathbf{W}_B^H)^{-1/2}. \quad (\text{A.8})$$

The likelihood in (A.6) is maximized when this determinant is maximized. Because \mathbf{Q} is positive definite, any solution that does not satisfy $\mathbf{W}_B \mathbf{W}_B^H = \mathbf{0}_{2 \times P}$ would decrease the determinant. The proof now follows since the remaining terms do not depend on the inner products of the rows of \mathbf{W} . \square

Using Lemma 4, the ML estimation problem aims at maximizing

$$\begin{aligned} \log p(\mathbf{X}; \mathbf{W}_A, \mathbf{W}_B, \hat{\mathbf{K}}_0) &= -\log \det(\hat{\mathbf{R}}) + \log \det(\mathbf{W}_A \mathbf{W}_A^H) \\ &\quad + \log \det(\mathbf{W}_B \mathbf{W}_B^H) - \frac{1}{2} \log \det[\mathbf{I}_{P-2} - \text{diag}(\mathbf{W}_A \hat{\mathbf{K}} \mathbf{W}_A^T)] \\ &\quad - \log \left[1 - \frac{1}{4} \text{tr}^2(\mathbf{W}_B \hat{\mathbf{K}} \mathbf{W}_B^T) \right], \end{aligned} \quad (\text{A.9})$$

subject to

$$\text{diag}(\mathbf{W}_A \mathbf{W}_A^H) = \mathbf{I}_{P-2}, \quad [\mathbf{W}_A \hat{\mathbf{K}} \mathbf{W}_A^T]_{l,l} \in [0, 1], \quad (\text{A.10})$$

$$\text{tr}(\mathbf{W}_B \mathbf{W}_B^H) = 2, \quad \text{tr}(\mathbf{W}_B \hat{\mathbf{K}} \mathbf{W}_B^T) \in [0, 2], \quad (\text{A.11})$$

$$\mathbf{W}_B \mathbf{W}_A^H = \mathbf{0}_{2 \times P}, \quad (\text{A.12})$$

where $l = 1, \dots, P-2$. It is clear that the above optimization problem is separable except for the constraint $\mathbf{W}_B \mathbf{W}_A^H = \mathbf{0}_{2 \times P}$. Thus, we will first ignore that constraint and optimize with respect to \mathbf{W}_A , but then enforce it in the optimization with respect to \mathbf{W}_B .

Lemma 5. The ML estimate of \mathbf{W}_A is $\hat{\mathbf{W}}_A = \mathbf{I}_{P-2 \times P}$.

Proof. We have to maximize

$$\begin{aligned} \log p(\mathbf{X}; \mathbf{W}_A, \mathbf{W}_B, \hat{\mathbf{K}}_0) &= \log \det(\mathbf{W}_A \mathbf{W}_A^H) \\ &\quad - \frac{1}{2} \log \det[\mathbf{I}_{P-2} - \text{diag}(\mathbf{W}_A \hat{\mathbf{K}} \mathbf{W}_A^T)], \end{aligned} \quad (\text{A.13})$$

subject to

$$\text{diag}(\mathbf{W}_A \mathbf{W}_A^H) = \mathbf{I}_{P-2}, \quad [\mathbf{W}_A \hat{\mathbf{K}} \mathbf{W}_A^T]_{l,l} \in [0, 1]. \quad (\text{A.14})$$

We begin by noting that \mathbf{W}_A must be real since $[\mathbf{W}_A \hat{\mathbf{K}} \mathbf{W}_A^T]_{l,l} \in [0, 1]$. Because of this and $\hat{k}_l \in [0, 1]$, any solution that satisfies $\text{diag}(\mathbf{W}_A \mathbf{W}_A^H) = \mathbf{I}_{P-2}$ also fulfils $[\mathbf{W}_A \hat{\mathbf{K}} \mathbf{W}_A^T]_{l,l} \in [0, 1]$. The second constraint in (A.14) can therefore be dropped. We now consider the two terms in (A.13) separately. First observe that, by Hadamard's inequality [26], the determinant is maximized when $\mathbf{W}_A \mathbf{W}_A^T$ is diagonal, and using the first constraint in (A.14) yields $\hat{\mathbf{W}}_A = \mathbf{I}_{P-2 \times P}$. On the other hand, the second term in (A.13) is Schur-convex [26], and it is maximized when $\mathbf{W}_A \hat{\mathbf{K}} \mathbf{W}_A^T$ is diagonal [26]. Now we use both the constraint (A.14) and the fact that $\hat{\mathbf{K}}_p$ is diagonal to conclude $\hat{\mathbf{W}}_A = \mathbf{I}_{P-2 \times P}$. Since both terms in (A.13) are maximized for the same \mathbf{W}_A , we have the solution to the optimization problem. \square

So far, the compressed likelihood is

$$\log p(\mathbf{X}; \hat{\mathbf{W}}_A, \mathbf{W}_B, \hat{\mathbf{K}}_0) = -\log \det(\hat{\mathbf{R}}) + \log \det(\mathbf{W}_B \mathbf{W}_B^H)$$

$$-\frac{1}{2} \sum_{\substack{l=1 \\ l \neq i, i+1}}^P \log(1-\hat{k}_l^2) - \log \left[1 - \frac{1}{4} \text{tr}^2(\mathbf{W}_B \hat{\mathbf{K}} \mathbf{W}_B^T) \right]. \quad (\text{A.15})$$

The next lemma presents the ML estimate of \mathbf{W}_B .

Lemma 6. *The ML estimate of \mathbf{W}_B is given by*

$$\hat{\mathbf{W}}_B = \begin{bmatrix} 0 & \cdots & 0 & \sqrt{1+\chi} & 0 \\ 0 & \cdots & 0 & 0 & \sqrt{1-\chi} \end{bmatrix}, \quad (\text{A.16})$$

where

$$\chi = \frac{2 - (\hat{k}_i^2 + \hat{k}_{i+1}^2) - 2\sqrt{(1-\hat{k}_i^2)(1-\hat{k}_{i+1}^2)}}{(\hat{k}_i^2 - \hat{k}_{i+1}^2)}. \quad (\text{A.17})$$

Proof. The ML estimate of \mathbf{W}_B is the solution to the optimization problem

$$\begin{aligned} & \underset{\mathbf{W}_B}{\text{maximize}} \quad \log \det(\mathbf{W}_B \mathbf{W}_B^H) - \log \left[1 - \frac{1}{4} \text{tr}^2(\mathbf{W}_B \hat{\mathbf{K}} \mathbf{W}_B^T) \right], \\ & \text{subject to} \quad \text{tr}(\mathbf{W}_B \mathbf{W}_B^H) = 2, \\ & \quad \text{tr}(\mathbf{W}_B \hat{\mathbf{K}}_p \mathbf{W}_B^T) \in [0, 2], \\ & \quad \mathbf{W}_B \hat{\mathbf{W}}_A^H = \mathbf{0}_{2 \times P}. \end{aligned} \quad (\text{A.18})$$

The last constraint directly implies $\mathbf{W}_B = [\mathbf{0}_{2 \times P-2} \quad \bar{\mathbf{W}}_B]$, which yields the simplified optimization problem

$$\begin{aligned} & \underset{\bar{\mathbf{W}}_B}{\text{maximize}} \quad \log \det(\bar{\mathbf{W}}_B \bar{\mathbf{W}}_B^H) - \log \left[1 - \frac{1}{4} \text{tr}^2(\bar{\mathbf{W}}_B \hat{\mathbf{K}}_r \bar{\mathbf{W}}_B^T) \right], \\ & \text{subject to} \quad \text{tr}(\bar{\mathbf{W}}_B \bar{\mathbf{W}}_B^H) = 2, \\ & \quad \text{tr}(\bar{\mathbf{W}}_B \hat{\mathbf{K}}_r \bar{\mathbf{W}}_B^T) \in [0, 2], \end{aligned} \quad (\text{A.19})$$

where $\hat{\mathbf{K}}_r = \text{diag}(\hat{k}_i, \hat{k}_{i+1})$ is a (reduced) diagonal matrix containing the estimated circularity coefficients of the identically distributed sources. Following arguments similar to those in the previous proof, $\bar{\mathbf{W}}_B$ must be a real matrix and we can drop the last constraint in (A.19). Using the eigenvalue decomposition (EVD) of $\bar{\mathbf{W}}_B^H \bar{\mathbf{W}}_B = \bar{\mathbf{W}}_B^T \bar{\mathbf{W}}_B$ and the results in [27], it is easy to show that $\bar{\mathbf{W}}_B^T \bar{\mathbf{W}}_B$ must be diagonal to maximize the likelihood. Hence, letting $\bar{\mathbf{W}}_B^T \bar{\mathbf{W}}_B = \text{diag}(1+\chi, 1-\chi)$, the ML estimation problem reduces to

$$\begin{aligned} & \underset{\chi}{\text{maximize}} \quad \log(1-\chi^2) - \log \left[1 - \frac{1}{4} (\hat{k}_i + \hat{k}_{i+1} + (\hat{k}_i - \hat{k}_{i+1})\chi)^2 \right], \\ & \text{subject to} \quad 0 \leq \chi \leq 1. \end{aligned} \quad (\text{A.20})$$

Ignoring the constraint and taking the derivative of the objective function with respect to χ , the solution is one of the two roots of a second-degree polynomial. Finally, it is easy to check that only one of these roots fulfils the constraint and it is indeed a maximum. \square

References

- [1] V.D. Calhoun, T. Adali, G.D. Pearlson, P.C. van Zijl, J.J. Pekar, Independent component analysis of fMRI data in the complex domain, *Magnetic Resonance in Medicine* 48 (1) (2002) 180–192.
- [2] T. Adali, P.J. Schreier, L.L. Scharf, Complex-valued signal processing: the proper way to deal with impropriety, *IEEE Transactions on Signal Processing* 59 (11) (2011) 5101–5125.
- [3] V.D. Calhoun, T. Adali, Complex ICA of brain imaging data, *IEEE Signal Processing Magazine* 24 (5) (2007) 136–139.
- [4] L. DeLathauwer, B. DeMoor, On the blind separation of non-circular sources, in: *European Signal Process. Conf. (EUSIPCO)*, Toulouse, France, 2002.
- [5] J. Eriksson, V. Koivunen, Complex random vectors and ICA models: identifiability, uniqueness, and separability, *IEEE Transactions on Information Theory* 52 (3) (2006) 1017–1029.
- [6] P.J. Schreier, L.L. Scharf, *Statistical Signal Processing of Complex-Valued Data*, Cambridge University Press, 2010.
- [7] E. Ollila, On the circularity of a complex random variable, *IEEE Signal Processing Letters* 15 (2008) 841–844.
- [8] F.D. Neeser, J.L. Massey, Proper complex random processes with applications to information theory, *IEEE Transactions on Information Theory* 39 (4) (1993) 1293–1302.
- [9] B. Picinbono, Second-order complex random vectors and normal distributions, *IEEE Transactions on Signal Processing* 44 (10) (1996) 2637–2640.
- [10] B. Picinbono, P. Bondon, Second-order statistics of complex signals, *IEEE Transactions on Signal Processing* 45 (2) (1997) 411–420.
- [11] A. Yeredor, Performance analysis of the strong uncorrelating transformation in blind separation of complex-valued sources, *IEEE Transactions on Signal Processing* 60 (1) (2012) 478–483.
- [12] D. Ramírez, P.J. Schreier, J. Via, I. Santamaría, GLRT for testing separability of a complex-valued mixture based on the strong uncorrelating transform, in: *2012 IEEE Int. Work. Machine Learning for Signal Process.*, Santander, Spain, 2012.
- [13] A. Cichocki, S.I. Amari, *Adaptive Blind Signal and Image Processing*, John Wiley, 2002.
- [14] H. Hotelling, Relations between two sets of variates, *Biometrika* 28 (1936) 321–377.
- [15] R.A. Horn, C.R. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, UK, 1985.
- [16] J. Villares, G. Vázquez, The Gaussian assumption in second-order estimation problems in digital communications, *IEEE Transactions on Signal Processing* 55 (10) (2007) 4994–5002.
- [17] T.M. Cover, J.A. Thomas, *Elements of Information Theory*, Wiley-Interscience, 2006.
- [18] L.L. Scharf, *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis*, Addison-Wesley, 1991.
- [19] T.S. Ferguson, *Mathematical Statistics*, Academic Press, New York, 1967.
- [20] E.L. Lehmann, *Testing Statistical Hypotheses*, Springer texts in Statistics, 1986.
- [21] K.V. Mardia, J.T. Kent, J.M. Bibby, *Multivariate Analysis*, Academic, New York, 1979.
- [22] L.L. Scharf, C.T. Mullis, Canonical coordinates and the geometry of inference rate and capacity, *IEEE Transactions on Signal Processing* 48 (3) (2000) 824–831.
- [23] S.M. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory*, Prentice Hall, 1998 Vol. II.
- [24] A.T. Walden, P. Rubin-Delanchy, On testing for impropriety of complex-valued Gaussian vectors, *IEEE Transactions on Signal Processing* 57 (3) (2009) 825–834.
- [25] D. Ramírez, J. Via, I. Santamaría, L.L. Scharf, Detection of spatially correlated Gaussian time series, *IEEE Transactions on Signal Processing* 58 (10) (2010) 5006–5015.
- [26] E. Jorswieck, H. Boche, Majorization and matrix-monotone functions in wireless communications, *Foundations and Trends in Communications and Information Theory* 3 (6) (2006) 553–701.
- [27] E. Oja, H. Ogawa, J. Wangviwattana, Principal component analysis by homogeneous neural networks, Part I: the weighted subspace criterion, *IEICE Transactions on Information and Systems* E75-D (1992) 366–375.