Complex-valued ICA based on a pair of generalized covariance matrices

Esa Ollila\textsuperscript{a,\!*}, Hannu Oja\textsuperscript{b}, Visa Koivunen\textsuperscript{a}

\textsuperscript{a}Signal Processing Laboratory, Helsinki University of Technology, P.O. Box 3000, FIN-02015 HUT, Finland
\textsuperscript{b}Tampere School of Public Health, University of Tampere, FIN-33014 University of Tampere, Finland

Received 2 April 2007; received in revised form 3 January 2008; accepted 4 January 2008
Available online 15 January 2008

Abstract

It is shown that any pair of scatter and spatial scatter matrices yields an estimator of the separating matrix for complex-valued independent component analysis (ICA). Scatter (resp. spatial scatter) matrix is a generalized covariance matrix in the sense that it is a positive definite hermitian matrix functional that satisfies the same affine (resp. unitary) equivariance property as does the covariance matrix and possesses an additional IC-property, namely, it reduces to a diagonal matrix at distributions with independent marginals. Scatter matrix is used to decorrelate the data and the eigenvalue decomposition of the spatial scatter matrix is used to find the unitary mixing matrix of the uncorrelated data. The method is a generalization of the FOBI algorithm, where a conventional covariance matrix and a certain fourth-order moment matrix take the place of the scatter and spatial scatter matrices, respectively. Emphasis is put on estimators employing robust scatter and spatial scatter matrices. The proposed approach is one among the computationally most attractive ones, and a new efficient algorithm that avoids decorrelation of the data is also proposed. Moreover, the method does not rely upon the commonly made assumption of complex circularity of the sources. Simulations and examples are used to confirm the reliable performance of our method.

c \copyright 2008 Elsevier B.V. All rights reserved.

1. Introduction

Independent component analysis (ICA) is a relatively recent technique of multivariate data analysis; see e.g. Comon (1994), Cardoso (1999), Hyvärinen et al. (2001), Cichocki and Amari (2002) and their bibliographies. The complex-valued ICA model is used e.g. for convolutive source separation in the frequency domain and for source separation of complex-valued data arising e.g. in magnetic resonance imaging or antenna array signal processing for wireless communications and radar applications. Complex-valued ICA is based on modelling the observed complex random variables (r.v.a.’s) $z_1, \ldots, z_m$ as linear mixtures of the unobserved \textit{mutually independent complex source signals} (also called \textit{independent components}, ICs) $s_1, \ldots, s_d$, that is,

$$z_i = a_{i1}s_1 + \cdots + a_{id}s_d, \quad i = 1, \ldots, m \quad (m \geq d).$$

\!* Corresponding author. Tel.: +358 9 451 3211; fax: +358 9 452 3614.
E-mail address: esollila@wooster.hut.fi (E. Ollila).

0167-9473/S - see front matter \copyright 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2008.01.001
The mixing coefficients $a_{ij} \in \mathbb{C}$ as well as the distributions of the sources $s_i$ are assumed to be unknown. The goal is then to find the inverse (linear) transformation that recovers the original ICs. Typically, the number of observed variables $m$ is assumed to be equal to that of latent variables $d$, which we also assume hereafter.

It is common to most ICA algorithms, such as complex FastICA (Bingham and Hyvärinen, 2000), FOBI (Cardoso, 1989), JADE (Cardoso and Souloumiac, 1993), InfoMax (Adali et al., 2004) and GUT (Ollila and Koivunen, submitted for publication) to work with preprocessed, whitened (i.e. decorrelated) data since the whitening transformation simplifies the ICA problem to that of finding a unitary mixing matrix of the whitened mixture. Pre-whitening also stabilizes optimization algorithms. To be more specific, if $\mathbf{z}$ represents the vector of mixtures and $\mathbf{B}$ is some whitening matrix, then a unitary linear transformation of the whitened mixture $\mathbf{v} = \mathbf{Bz}$ can separate the sources (up to some indeterminacies discussed in Section 3), i.e. for some unitary matrix $U$ (i.e. $U^H U = I$)

$$U^H \mathbf{v} = (U^H \mathbf{B}) \mathbf{z} = \mathbf{s},$$

where $\mathbf{s}$ represents the vector of sources. The above superscript $^H$ denotes hermitian transpose whereas superscript $^T$ is used to denote the transpose operator. Eq. (1) implies that $\mathbf{W} = U^H \mathbf{B}$ is a separating matrix, that is, a linear transformation that maps the vector of mixtures $\mathbf{z}$ back to vector of ICs $\mathbf{s}$. Computation of the whitening matrix $\mathbf{B}$ is a straightforward task, commonly solved by computing the inverse of the square-root matrix of the covariance matrix $\mathbf{C}(\mathbf{z})$ of $\mathbf{z}$. The hard part of the ICA problem is finding the unitary matrix $U$. Many ICA methods find matrix $U$ as the maximizer of some contrast function (Comon, 1994) $f(U^H \mathbf{v})$, which often leads to a difficult optimization task. A different approach was taken by Cardoso. He proposed a simple and elegant solution, called Fourth-Order Blind Identification (FOBI) (Cardoso, 1989), which solves $U$ as the matrix of eigenvectors of the kurtosis matrix $\mathbf{K}(\mathbf{v}) = \mathbf{C}(\|\mathbf{v}\|^{4})$, where $\|\cdot\|$ denotes the norm of $\mathbf{v}$, defined as $\|\mathbf{v}\| = \sqrt{\mathbf{v}^H \mathbf{v}}$.

In this paper the FOBI method is generalized. We show that the separating matrix can be found by employing any scatter and spatial scatter matrix with independent component (IC-)property in place of the covariance matrix $\mathbf{C}(\mathbf{z})$ and kurtosis matrix $\mathbf{K}(\mathbf{v})$ in the FOBI algorithm, respectively. We call our estimator DOGMA (Diagonalization Of Generalized covariance MAtrices) since it is the data transformation which diagonalizes the given pair of scatter and spatial scatter matrices. Scatter matrix is a generalized covariance matrix in the sense that it is also a positive definite hermitian matrix functional and equivariant under affine linear transformations. Spatial scatter matrix is a broader notion of scatter matrix for which the requirement of affine equivariance is replaced by the weaker assumption of equivariance under unitary linear transformations. Hence every scatter matrix is also a spatial scatter matrix. For the pair of scatter and spatial scatter matrices we require the IC-property, namely, that they reduce to a diagonal matrix at distributions with independent marginals. Since a scatter matrix and spatial scatter matrix do not generally possess the IC-property, we show that a symmetrized version of the estimator automatically has this property. From a little different point of view, this approach for the real-valued ICA model was also studied by Oja et al. (2006); see also Taskinen et al. (2007).

DOGMA estimators constitute a large family of estimators of the separating matrix. Depending on the choices of the scatter and spatial scatter matrices, estimators may have very different statistical properties (convergence, limiting distributions, robustness, efficiency, computation, etc.). Our focus is on DOGMA estimators employing robust scatter and spatial scatter matrices which avoid the assumption of finite fourth-order moments required by FOBI. We also show that DOGMA estimators possess the desirable property of equivariance advocated by Cardoso (1994) and Cardoso and Laheld (1996). Our approach is computationally attractive and it relaxes the commonly made assumption of circularity of the sources. We also derive a computationally efficient, improved algorithm/formulation of the DOGMA method that avoids whitening of the data provided that the employed spatial scatter matrix is a scatter matrix or the so-called weighted spatial scatter matrix. We demonstrate that this improved algorithm can be used for computing FOBI efficiently in three lines of Matlab code. A drawback of our method is that it requires that sources do not have identical distribution.

The paper is organized as follows. Section 2 reviews the properties and characteristics of complex random variables. Section 3 reviews the complex-valued ICA model and some common assumptions in more detail. The topic of generalized covariance matrices is addressed in Section 4. In Sections 4.1 and 4.2, the concepts of scatter and spatial scatter matrix are defined and examples of robust scatter matrices and spatial scatter matrices, such as (complex-valued) $M$-estimators of scatter matrix and the spatial sign covariance matrix are given. The IC-property and symmetrized estimators of scatter and spatial scatter matrices are defined in Section 4.3. The proposed method...
is introduced and studied in Section 5. The FOBI algorithm and its pros and cons are discussed in Section 5.1. In Section 5.2, the DOGMA method, which is a generalization of FOBI, based on a pair of scatter matrix and spatial scatter matrix is defined and its properties, such as (ICA-)equivariance, are established as well. In Section 5.3, we derive alternative formulations of the DOGMA method. Section 6 contains simulation studies and Section 7 contains a communications example. Section 8 concludes.

2. Complex random variables: Preliminaries

The mean of a complex r.v. $z = x + jy$ is defined as $E[z] = E[x] + jE[y]$, where $j = \sqrt{-1}$ denotes the imaginary unit. For the simplicity of notation, assume $E[z] = 0$, otherwise replace $z$ by $z - E[z]$ in the definitions below. The variance and pseudo-variance of $z$ are defined as

$$\text{var}(z) \overset{\text{def}}{=} E\left[ |z|^2 \right] = E\left[ x^2 \right] + E\left[ y^2 \right]$$

and

$$\text{pvar}(z) \overset{\text{def}}{=} E\left[ z^2 \right] = E\left[ x^2 \right] - E\left[ y^2 \right] + 2jE[x y],$$

where $|z| = \sqrt{x^2 + y^2}$ is called the modulus of $z$ and superscript * denotes complex conjugate. R.v. $z$ is said to be second-order circular if $\text{pvar}(z) = 0$, which implies that $x = \text{Re}(z)$ and $y = \text{Im}(z)$ are uncorrelated with equal variances. Kurtosis of $z$ is defined as

$$\kappa(z) \overset{\text{def}}{=} \frac{E\left[ |z|^4 \right]}{E\left[ |z|^2 \right]^2} - 2.$$

This definition of kurtosis is meaningful for second-order circular r.v.a.’s, and it possesses the property that $\kappa(z) = 0$ if $z$ possesses standard zero-mean (circular) complex Gaussian density $f(z) = (\pi \sigma)^{-1}e^{-|z|^2/2\sigma^2}$, where $\sigma^2$ denotes the variance.

R.v. $z$ is said to be symmetric if $z^d = -z$, where $^d$ should be read as ‘has the same distribution as’. A restricted notion of symmetry is that of circular symmetry (or sphericity): a r.v. $z$ is said to be circular if $z^d = e^{i\theta}z$ for all $\theta \in \mathbb{R}$. Note that circularity implies second-order circularity if $z$ has finite variance.

The complex covariance matrix of a zero-mean complex random vector (r.v.) $Z = (z_1, \ldots, z_d)^T$ is defined as

$$C(z) \overset{\text{def}}{=} E\left[ z z^H \right].$$

We assume that $Z$ is non-degenerate in any subspace of $\mathbb{C}^d$ so that $C = C(z)$ is complex positive definite, hermitian (i.e. $C^H = C$) $d \times d$ matrix, denoted by $C \in \text{PDH}(d)$. Kurtosis matrix of $Z$ is defined as

$$K(z) \overset{\text{def}}{=} C(\|z\|z) = E\left[ \|z\|^2 z z^H \right].$$

The matrix $K(z)$ derives its name from the property that if $z$ has independent components of unit-variance, then $K(z) = \text{diag}(\kappa_1, \ldots, \kappa_d) + (d + 1)I$, where $\kappa_i = \kappa(z_i)$ and $I$ denotes the identity matrix. We also use the notation $C(F)$ (and similarly $K(F)$), where $F$ denotes the cumulative distribution function (c.d.f) of $Z$, to highlight that $C(z)$ is a statistical functional, namely, a mapping $C : F \to \text{PDH}(d)$ which sends an arbitrary distribution function $F$ to $C = C(F) \in \text{PDH}(d)$.

3. Complex-valued ICA model

In matrix form, the complex-valued ICA model becomes

$$Z = A S = A_1 s_1 + \cdots + A_d s_d$$

where r.v.’s $S = (s_1, \ldots, s_d)^T$ and $Z = (z_1, \ldots, z_d)^T$ contain the sources and their mixtures, respectively, and $A = (a_1 \cdots a_d)$ represents the unknown complex $d \times d$ mixing matrix that is assumed to be of full 
Table 1

<table>
<thead>
<tr>
<th>Assumption</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1:</td>
<td>$s_i$’s have zero mean</td>
</tr>
<tr>
<td>A2:</td>
<td>$s_i$’s have finite variance</td>
</tr>
<tr>
<td>A3:</td>
<td>$s_i$’s have finite kurtosis</td>
</tr>
<tr>
<td>A4:</td>
<td>$s_i$’s have distinct kurtosis</td>
</tr>
<tr>
<td>A5:</td>
<td>$s_i$’s are circular</td>
</tr>
<tr>
<td>A6:</td>
<td>At most one of $s_i$ is circular</td>
</tr>
<tr>
<td>A7:</td>
<td>$s_i$’s are symmetric</td>
</tr>
</tbody>
</table>

Some common additional assumptions on the distributions of the sources $s_i$

rank, i.e. $a_1, \ldots, a_d$ form a set of linearly independent vectors. Each source r.v. $s_i$ is assumed to be non-degenerate and at most one source can possess circular Gaussian distribution. Suppose we have a random sample $z_1, \ldots, z_n$ from the ICA model and write $Z_n = (z_1 \cdots z_n)$ for the $d \times n$ data matrix. The aim is then to find the estimator $\hat{W}$ of the separating matrix (also called demixing matrix)

$$W = (w_1 \cdots w_d)^T \in \mathbb{C}^{d \times d}$$

which transforms $z$ back to independent components via $s = Wz$ (up to some indeterminacies discussed in the next paragraph).

Define ambiguity matrix (called non-mixing matrix by Cardoso (1999)) as

$$G \defeq PLD,$$

where $P$ is any $d \times d$ permutation matrix (obtained by permuting the rows of $d \times d$ identity matrix), $D$ is any scaling matrix (a diagonal matrix with positive real diagonal elements) and $L$ is any phase-shift matrix, i.e. $L = \text{diag}(e^{i\theta_1}, \ldots, e^{i\theta_d})$, where $\theta_i \in \mathbb{R}$. Hence $G$ has only one non-zero complex value in each row and in each column. Let $\mathcal{G}$ denote the set of all ambiguity matrices $G$. The matrix $G$ inherits its name from the fundamental indeterminacy of the ICA model: the mixture $z$ does not have a unique representation since for any $G \in \mathcal{G}$, $z = (AG^{-1})Gs$, where $Gs$, called the copy (Cardoso, 1999) of $s$, has independent components as well. Hence it is possible to identify $A$ (resp. $W$) only up to scaling, phase-shift and permutation of its column vectors $a_i$ (resp. row vectors $w_i$). The assumption of at most one circular Gaussian source is required for $A$ (and $W$) to be identifiable up to the fundamental indeterminacy (Eriksson and Koivunen, 2006). Note, however, that sources can have non-identical non-circular Gaussian distribution.

Hence separation of the ICs from their mixtures should be understood as the determination of a matrix $W$, called the separating matrix, that satisfies $WA = G$ for any $G \in \mathcal{G}$, that is, sources are separated when $\hat{s} = Wz = Gs$, i.e. $\hat{s}$ is a copy of $s$. This indeterminacy also leads us to introduce the following definition.

Definition 1. Two matrices $B$ and $C$ are said to be fundamentally equal, denoted by $B \equiv_{\text{ica}} C$, if $B = GC$ for some $G \in \mathcal{G}$.

In other words, $W$ is a separating matrix if it is fundamentally equal to $A^{-1}$.

Some common additional simplifying assumptions on moments and/or on the functional form of the distributions of the sources imposed by many ICA methods are summarized in Table 1. Note that A1 means that $E[s] = E[z] = 0$. Consequently, the most basic and often necessary preprocessing common to most ICA algorithms is to center $z$, i.e. subtract its mean vector $E[z]$ so that A1 holds. Since we do not wish to impose any finite moment assumptions on the sources, we assume that $z$ is centered in the sense that $T(z) = T(s) = 0$ holds, where $T(F) \in \mathbb{C}^d$ denotes some appropriate location vector functional satisfying $T(Az + b) = AT(z) + b$ for any non-singular matrix $A \in \mathbb{C}^{d \times d}$ and $b \in \mathbb{C}^d$. Naturally, the mean vector $T(z) = E[z]$ serves as an example of a location vector functional. Note that A4 implies that $A4'$: $s_i$'s have non-identical distribution.
4. Generalized covariance matrices

4.1. Scatter matrix

**Definition 2.** Functional \( C(F) \in \text{PDH}(d) \) is said to be a scatter matrix if it is affine equivariant in the sense that for any non-singular matrix \( A \in \mathbb{C}^{d \times d} \), \( C(F_A) = AC(F)A^H \), where \( F_A \) and \( F \) represent the distribution functions of \( z \) and \( Az \), respectively. We also write \( C(\mathbf{z}) \) instead of \( C(F) \) when \( \mathbf{z} \) is distributed according to \( F \).

Let \( F_n \) denote the empirical distribution function associated with the data set \( Z_n \). Then \( \hat{C} = C(F_n) \), which we also denote by \( \hat{C} = C(\mathbf{Z}_n) \), represents the estimator associated with scatter matrix \( C(F) \). Affine equivariance now implies that for any non-singular \( A \in \mathbb{C}^{d \times d} \), the scatter matrix estimator for the transformed data \( AZ_n = (AZ_1 \cdots AZ_n) \) is \( A\hat{C}A^H \), i.e. \( C(AZ_n) = AC(Z_n)A^H \). Classical example of a scatter matrix is the covariance matrix \( C(F) \). The corresponding estimator is the sample covariance matrix \( \hat{C} = \hat{C}(F_n) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{z}_i \mathbf{z}_i^H \). Note that the kurtosis matrix \( K(F) \) is not a scatter matrix since it does not satisfy the property of affine equivariance.

Since the covariance matrix suffers from poor robustness properties, one can consider weighted covariance matrix

\[
\hat{C}_{\psi,C}(F) = E\left[ \psi(\mathbf{z}^H C(F)^{-1} \mathbf{z}) \mathbf{z} \mathbf{z}^H \right],
\]

where \( \psi(\cdot) \) is any real-valued weighting function on \([0, \infty)\) and \( C(F) \) is any scatter matrix, e.g. the covariance matrix \( \hat{C}(F) \). It is then easy to verify that \( \hat{C}_{\psi,C}(F) \) is a scatter matrix. For example, choices \( \psi(x) = x \) and \( C(F) = \hat{C}(F) \), yield a matrix

\[
K_{\hat{C}}(F) = E\left[ (\mathbf{z}^H \hat{C}(F)^{-1} \mathbf{z}) \mathbf{z} \mathbf{z}^H \right],
\]

which is an affine equivariant version of the kurtosis matrix \( K(F) \). The chosen \( \psi(\cdot) \) function determines the properties of the weighted covariance matrix and a robust “psi-function” should give a “robustified” version of the covariance matrix.

For real-valued data, a well-established improved idea of the weighted covariance matrix is the \( M \)-functionals of scatter (Maronna, 1976; Huber, 1981; Maronna et al., 2006). In fact, weighted covariance matrix is often said to be a “1-step \( M \)-estimator”. \( M \)-functionals of scatter can be easily extended for complex-valued data (Ollila and Koivunen, 2003b,a,c). The complex \( M \)-functional of scatter \( C_{\psi,c}(F) \in \text{PDH}(d) \) is defined as the solution of an implicit equation

\[
C_{\psi,c}(F) = E\left[ \psi(\mathbf{z}^H C_{\psi,c}(F)^{-1} \mathbf{z}) \mathbf{z} \mathbf{z}^H \right],
\]

where \( \psi(\cdot) \) is any real-valued weighting function on \([0, \infty)\). The choice \( \psi(x) = 1 \) yields the conventional covariance matrix. Using a robust “phi-function”, however, we obtain a robust \( M \)-functional. Some examples of robust \( \psi \)-functions are given below.

- **Complex Tyler’s \( M \)-estimator** of scatter (Kent, 1997; Ollila and Koivunen, 2003b) utilizes \( \psi(\cdot) \) function of the form \( \psi(x) = d/x \) (where \( d \) is the dimension of \( \mathbf{z} \)).
- **Complex Huber’s \( M \)-estimator of scatter** utilizes Huber’s \( \psi(\cdot) \) function

\[
\psi(x) = \begin{cases} 1/\beta, & \text{for } x \leq c^2 \\ c^2/x\beta, & \text{for } x > c^2 \end{cases}
\]

where the cut-off point \( c \) is defined so that \( q = F_{\chi_d^2}(2c^2) \) for a chosen \( q \) and the scaling factor \( \beta = F_{\chi_{2d+1}^2}(2c^2) + c^2(1-q)/d \), where \( F_{\chi_d^2} \) denotes the c.d.f. of chi-squared distribution with \( d \) degrees of freedom. The scaling factor is chosen so that \( C_{\psi}(F) = C \) when \( F \) equals the circular complex \( d \)-variate Gaussian distribution. In our simulations we use \( q = 0.9 \).

The \( M \)-estimator of scatter \( \hat{C} = C_{\psi,c}(F_n) \in \text{PDH}(d) \) is then the solution of the equation

\[
\hat{C} = \frac{1}{n} \sum_{i=1}^{n} \psi(\mathbf{z}_i^H \hat{C}^{-1} \mathbf{z}_i) \mathbf{z}_i \mathbf{z}_i^H.
\]

As in the real case, a simple iterative algorithm can be used to compute the estimator. See Ollila and Koivunen (2003b).
4.2. Spatial scatter matrix

Definition 3. Functional $C(F) \in \text{PDH}(d)$ is called spatial scatter matrix if it is unitary equivariant in the sense that $C(FA) = AC(F)A^H$ for any unitary matrix $A \in \mathbb{C}^{d \times d}$.

Remark 1. Spatial scatter matrix is a broader notion of scatter matrix because the requirement of affine equivariance is relaxed by the weaker assumption of equivariance under unitary linear transformations. Hence every scatter matrix is also a spatial scatter matrix.

We define a weighted spatial scatter matrix as

$$S_\psi(F) \overset{\text{def}}{=} C_\psi,l(F) = E \left[ \psi(\|z\|^2)zz^H \right].$$

Some examples of $\psi$-functions are:

- $\psi(x) = x$ which yields the kurtosis matrix $K(F)$.
- $\psi(x) = x^{-1}$ which yields the sign covariance matrix (SCM),
  $$\hat{C}(\|z\|^{-1}z) = E \left[ \|z\|^{-2}zz^H \right],$$
  used e.g. by Visuri et al. (2001).

This weight function of SCM is equivalent (up to scaling and translation) with the location score function of the chi-squared distribution with $d$ degrees of freedom, i.e. $\psi(x) = -\frac{d}{dx} \log f_{\chi_d^2}(x) = x^{-1}$. For the ICA model, the SCM weighting can be motivated from the point of view that r.v.a.’s $\|v\|^2$ and $\chi_d^2$ have the same first-order properties, where $v$ is the whitened mixture. Naturally some other weight could be considered. For example, Cardoso (1989) also suggested the use of $\psi(x) = x^2$ giving a sixth-order moment matrix. This weight has the disadvantage of being extremely sensitive to measurement errors. Naturally, we could construct e.g. a weighted spatial scatter matrix using Huber’s $\varphi$-function.

4.3. IC-property and symmetrized estimators

An important property of a scatter or spatial scatter matrix that we require for purposes of ICA is the independent component (IC-)property (Oja et al., 2006) defined below.

Definition 4. Scatter matrix or spatial scatter matrix functional $C(F)$ is said to possess IC-property if $C(F)$ is a diagonal matrix when $F$ is the c.d.f. of a random vector with independent components.

Clearly, for centered (zero-mean) sources $s$, covariance matrix $C(\cdot)$ and kurtosis matrix $K(\cdot)$ possess IC-property. However, $M$-estimator $C_{\psi}(\cdot)$, weighted covariance matrix $C_{\psi,C}(\cdot)$ or weighted spatial scatter matrix $S_{\psi}(\cdot)$ do not generally possess IC-property. The next theorem points out the obvious condition on the sources under which any scatter or spatial scatter possess IC-property.

Lemma 1. Let $C(F)$ denote any scatter or spatial scatter matrix. Then $C(s)$ possesses IC-property when $s$ has independent components that satisfy Assumption A7.

The Assumption A7 (that sources $s_i$ possess symmetric distribution) may not hold in many applications. However, Oja et al. (2006) have shown that a “symmetrization” of the estimator guarantees IC-property. In a similar manner, we now show that a symmetrized version of a scatter matrix or spatial scatter matrix automatically possesses IC-property without assuming A7.

Lemma 2. Assume $z_1$ and $z_2$ are two independent copies of $z \sim F$. Then any scatter or spatial scatter matrix $C(F)$ yields a symmetrized version $C_{\text{sym}}(F) = C(F_{z_1-z_2})$ that has IC-property.
For example, the symmetrized weighted spatial scatter matrix is defined as
\[
S_{\psi}(z) \overset{\text{def}}{=} S_{\psi}(z_1 - z_2) = E \left[ \psi(\|z_1 - z_2\|^2)(z_1 - z_2)(z_1 - z_2)^H \right],
\]
where \(z_1\) and \(z_2\) are independent copies of \(z \sim F\). We define the respective estimator as
\[
\hat{S}_{\psi} = \frac{2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \psi(\|z_i - z_j\|^2)(z_i - z_j)(z_i - z_j)^H.
\]
In a similar manner, one can construct a symmetrized version of any weighted covariance matrix \(C_{\psi,C}(F)\). Symmetrized \(M\)-estimator \(C_{\psi}(F) \in \text{PDH}(d)\) of scatter is defined as the solution of
\[
C_{\psi}(F) = E \left[ \varphi \left( [z_1 - z_2]^H C_{\psi}(F)^{-1} [z_1 - z_2] \right) (z_1 - z_2)(z_1 - z_2)^H \right].
\]
For symmetrized Huber’s \(M\)-estimator we define the cut-off point \(c\) needed in Huber’s \(\varphi\)-function as \(q = F_{\chi^2_d}^{\phi}(c^2)\) for a chosen \(q\) and the scaling factor as \(\beta = 2F_{\chi^2_{d+1}}^{\phi}(c^2) + c^2(1 - q)/d\). The scaling factor is again chosen to guarantee that \(C_{\psi}(F) = C\) when \(F\) equals the circular complex \(d\)-variate Gaussian distribution. In our simulations we use \(q = 0.9\).

The drawback of the symmetrized version of an estimator is the increased computational load since symmetrization increases the number of “observations” by power of two. The advantage of using symmetrized estimators is that they possess IC-property without centering the data.

5. ICA based on a pair of generalized covariance matrices

5.1. FOBI method

Assume that \(s\) satisfies A1 and A2. The whitening matrix \(B = B(z) \in \mathbb{C}^{d \times d}\) is computed as the square-root matrix of \(C^{-1}(z)\), i.e. \(B^H B = C^{-1}\). There exists many methods to compute \(B\) (e.g. Cholesky decomposition), but for our purposes, we do not need to specify any particular one. Then the whitened mixture \(v = Bz\) is uncorrelated, i.e. \(C(v) = I\), and also follows the ICA model:
\[
v = \hat{A}s \quad \text{with} \quad \hat{A} = BA.
\]
It can be shown e.g. Cardoso (1999) and Hyvärinen et al. (2001) that the separating matrix of the whitened mixture is a unitary matrix and thus \(W = U^H B\) gives a separating matrix for the original mixture for some unitary matrix \(U \in \mathbb{C}^{d \times d}\). Cardoso (1989) has shown that, if \(U = U(v)\) is the matrix of eigenvectors of \(K(v)\), then the FOBI functional \(W(z) = U(v)^H B(z)\) is a separating matrix if A1 and A4 hold.

FOBI is perhaps the simplest method to solve the ICA problem proposed thus far. Since the FOBI functional can be computed via standard matrix decompositions operating on matrices \(C(z)\) and \(K(v)\), it is also computationally (among) the most efficient approach to ICA. It has some drawbacks however. First of all, the necessity of A4 (which also implies A4’) restricts the applicability of the method to some extent since sources with identical distributions (and hence with identical kurtosis) may occur frequently in some applications. Second of all, Assumption A3 clearly confines permissible distributions of the sources the method can separate since fourth-order moments do not exist for many distributions. Third of all, the method is not robust, since the covariance matrix (used for whitening) and the kurtosis matrix are highly non-robust.

In order to separate sources with identical distribution and identical kurtosis, FOBI was later generalized to JADE (Cardoso and Souloumiac, 1993; Cardoso, 1999) which is based on joint diagonalization of several cumulant matrices. However, JADE still relies on A3, is not robust, and has the disadvantage that simplicity and computational efficiency of the FOBI algorithm are lost.

5.2. DOGMA method

Algorithm: Matrix \(W(z) \in \mathbb{C}^{d \times d}\) based on any scatter matrix \(C_1(F)\) and spatial scatter matrix \(C_2(F)\) with the IC-property is calculated as follows:
(a) Calculate the square-root matrix $B_1(z)$ of $C_1(z)^{-1}$, so $B_1(z)H B_1(z) = C_1(z)^{-1}$, and the whitened data $v = B_1(z)z$.

(b) Calculate the EVD of $C_2(F)$ of the whitened data:

$$C_2(v) = U_2(v)A_2(v)U_2(v)^H,$$

where $A_2(v)$ is a diagonal matrix of eigenvalues of $C_2(v)$ and $U_2(v)$ is a unitary matrix with the respective eigenvectors as columns.

(c) Set $W(z) = U_2(v)^H B_1(z)$.

Remark 2. Clearly, the algorithm is a generalization of FOBI which corresponds to the DOGMA functional $W(z)$ with choices $C_1(F) = C(F)$ and $C_2(F) = K(F)$. Note that IC-property is required which implies for example that the SCM $C((z)^{-1}z)$ cannot be employed as the choice of $C_2(z)$ unless A7 holds. Naturally, by Lemma 2, the symmetrized version of the SCM $C((z)^{-1}_1 - (z)^{-1}_2)$ has the IC-property without assuming A7.

Remark 3. Observe that $W(z) = U_2(v)^H B_1(z)$ simultaneously diagonalizes $C_1(z)$ and $C_2(z)$, namely, for transformed data $\hat{s} = W(z)z$ it holds that

$$C_1(\hat{s}) = I \quad \text{and} \quad C_2(\hat{s}) = A_2(v).$$

Hence we call the above algorithm DOGMA (Diagonalization Of Generalized covariance MAtrices).

We also use the notation $W(F)$ instead of $W(z)$ when $z$ is distributed according to $F$. Then $\hat{W} = W(F_n)$, also denoted by $\hat{W} = W(Z_n)$, is the estimator corresponding to functional $W(F)$, i.e. obtained by the above algorithm using the corresponding estimators $\hat{C}_1 = C_1(F_n)$ and $\hat{C}_2 = C_2(F_n)$ in place of $C_1(F)$ and $C_2(F)$.

Our first task is to resolve the explicit form of the EVD in (3).

**Theorem 1.** Assume $z = A$ follows the ICA model and that

**A8:** $C_1(s)$ and $C_2(s)$ exist.

Then

$$A_2(v) = C_2\left(\frac{C_1(s)^{-1/2}s}{2}\right) \quad \text{and} \quad U_2(v) = B_1(z)AC_1(s)^{1/2},$$

where $C_1(s)^{1/2}$ denotes the diagonal matrix with square roots of the elements of $C_1(s)$ on the diagonal and $C_1(s)^{-1/2}$ is its inverse.

**Theorem 1** shows that $U_2(v)$ can be factored (up to irrelevant right multiplication by a scaling matrix $C_1(s)^{1/2}$) to a product of the whitening matrix $B_1(z)$ and the mixing matrix $A$. The next corollary is an immediate consequence of this result.

**Corollary 1.** Assume $z = A$ follows the ICA model and that A8 and

**A9:** $A_2(v) = C_2\left(\frac{C_1(s)^{-1/2}s}{2}\right) = \text{diag}(\delta_1, \ldots, \delta_d)$ is such that $\delta_i \neq \delta_j$ for all $i \neq j$

hold. Then $W(z) = U_2(v)^H B_1(z)$ is a separating matrix, i.e. $W(z) \overset{\text{IC}}{=} \Lambda^{-1}$.

As an example, consider the FOBI functional. Then A8 is equivalent with A3. Assumption A9 states that none of the eigenvalues of $C_2(v)$ should be equal. For the FOBI method, the eigenvalues are easily calculated:

$$A_2(v) = K\left(C(s)^{-1/2}\right) = C(s)^{-1/2}E\left[(s^H C(s)^{-1}s)ss^H\right]C(s)^{-1/2}$$

$$= \text{diag}(\kappa_1, \ldots, \kappa_d) + (d + 1)I,$$

where $\kappa_i = \kappa(s_i)$ denotes the kurtosis of $s_i$. Thus A9 is equivalent with A4 for the case of FOBI.

We would like to emphasize that the DOGMA algorithm contains a *built-in warning*: since the matrix of eigenvalues $A_2(v)$ is also necessarily extracted when computing the EVD of $C_2(v)$ in step (b) of the algorithm, detection of two close eigenvalues is an indication that the corresponding sources may not be reliably separated. Also, we note that Assumption A9 is needed to separate all the sources. Inspection of the proof of **Corollary 1** shows that $W(z)$ is not able to separate the sources that have identical eigenvalues, but the rest of the sources are separated. This important fact is more formally stated in the next corollary.
Corollary 2. Assume \( z = A s \) as follows the ICA model and A8 holds. Decompose \( s = (s_1^T, s_2^T)^T \) so that each source \( s_i \) in \( s_1 \) (resp. in \( s_2 \)) has eigenvalue \( \delta_i = |C_2(C_1(s)^{-1/2}s)|_{ii} \) of multiplicity 1 (resp. > 1). Decompose \( \hat{s} = (s_1^T, s_2^T)^T = W(z)z \) accordingly. Then \( \hat{s}_1 \) is a copy of \( s_1 \).

A desirable property of equivariance for ICA was advocated by Cardoso (1994) and Cardoso and Laheld (1996). We may formulate this equivariance property for separating matrix functional/estimator as follows.

Definition 5. Separating matrix \( W(z) \) is said to be ICA-equivariant if

\[
W(Qz) \stackrel{\text{ica}}{=}= W(z)Q^{-1}, \quad \forall Q \in \mathbb{C}^{d \times d} \text{ non-singular}.
\]

Analogously, the estimator \( \hat{W} = W(Z_n) \) is said to be ICA-equivariant if

\[
W(QZ_n) \stackrel{\text{ica}}{=}= W(Z_n)Q^{-1}, \quad \forall Q \in \mathbb{C}^{d \times d} \text{ non-singular}
\]

where \( QZ_n = (Qz_1 \cdots Qz_n) \) denotes the transformed data set.

Since data matrix \( Z_n \in \mathbb{C}^{d \times n} \) is a realisation from the ICA model, it can be factored as \( Z_n = AS_n \) with \( S_n = (s_1 \cdots s_n) \in \mathbb{C}^{d \times n} \), where \( s_1, \ldots, s_n \) is a random sample distributed as \( s \). Then we see that ICA-equivariant estimator \( \hat{W} \) has the natural property that source estimates are independent of \( A \), because \( \hat{S}_n = W(Z_n)Z_n = GW(S_n)S_n \) for some \( G \in G \). Thus, the estimated sources \( \hat{S}_n \) depend only on realisation \( S_n \) but not on the mixing matrix \( A \).

Next we show that DOGMA functional/estimator satisfies the desirable feature of ICA-equivariance.

Theorem 2. Assume \( z = A s \) as follows the ICA model and A8 and A9 hold. Then the DOGMA separating matrix \( W(z) \) and the corresponding estimator \( \hat{W} = W(Z_n) \) are ICA-equivariant.

Then Theorem 2 also means that the FOBI functional/estimator is ICA-equivariant. Note that the JADE estimator (the extension of FOBI), loses the desirable ICA-equivariance property, since the set of cumulant matrices cannot be exactly jointly diagonalized. Also the complex FastICA estimator is not ICA-equivariant.

5.3. Alternative formulations of the method

Since a scatter matrix is also a spatial scatter matrix, \( W(z) \) can be based on a pair of scatter matrices \( C_1(F) \) and \( C_2(F) \). Naturally, \( C_2(F) \) used in step (b) of the DOGMA algorithm need not be a scatter matrix, but if it is, \( W(z) \) will have some interesting properties which do not apply in the case that \( C_2(F) \) is strictly spatial. First we note that eigenvalue matrix \( A_2(v) \) can be written as

\[
A_2(v) = C_2 \left( C_1(s)^{-1/2}s \right) = C_1(s)^{-1}C_2(s)
\]

provided that \( C_2(F) \) is a scatter matrix. Furthermore, the functional \( W(z) \) can be formulated in a manner which avoids the data whitening step as is shown next.

Theorem 3. Assume \( C_2(F) \) is a scatter matrix with IC-property and that \( z \) follows the ICA model and A8 holds. Then \( W(z)^H \) and \( A_2(v) \) correspond, respectively, to the matrix of eigenvectors and diagonal matrix of eigenvalues of \( C_1(z)^{-1}C_2(z) \). Consequently, \( W(z)^H \) and \( A_2(v)^{-1} \) are the matrix of eigenvectors and diagonal matrix of eigenvalues of \( C_2(z)^{-1}C_1(z) \).

Remark 4. Hence, by Theorem 3 and Corollary 1, hermitian transpose of an eigenvector matrix of \( C_1^{-1}(z)C_2(z) \) is fundamentally equal to \( A^{-1} \). For purposes of ICA, calculation of \( W(z) \) is thus tantamount to calculation of the eigenvectors of \( C_1(z)^{-1}C_2(z) \) provided that \( C_2(F) \) is a scatter matrix. This is clearly a more efficient method to calculate \( W(z) \) since whitening of the data (and thus calculation of the whitening matrix) is avoided. Theorem 3 also reveals an important insight that statistical properties (e.g. consistency, limiting distribution, statistical efficiency, robustness) of the DOGMA estimator \( \hat{W} \) may be derived rather directly from those of \( \hat{C}_1 \) and \( \hat{C}_2 \). Furthermore, due to ICA-equivariance of \( \hat{W} \), one may consider without any loss of generality the case that \( A = I \) (i.e. that the mixing matrix equals the identity matrix).
Remark 5. Theorem 3 also shows surprising symmetry feature: the separating matrix obtained when $C_2$ is used in place of $C_1$ in the whitening step (a) and $C_1$ in place of $C_2$ in step (b) is fundamentally equal to $W(z)$. In other words, the order of scatter matrices $C_1$ and $C_2$ given as input to the DOGMA algorithm is irrelevant.

Theorem 3 does not apply for FOBI since the kurtosis matrix $K(F)$ is not a scatter matrix. Let us now concentrate on the case that $C_2(F)$ is any weighted spatial matrix $S_\psi(F)$ with IC-property (such as kurtosis matrix), i.e. $C_2(F) = S_\psi(F)$. Then, $A_2(v) = S_\psi(C_1(s)^{-1/2}s)$ becomes

$$A_2(v) = C_1(s)^{-1/2} E \left[ \psi(s^H C_1(s)^{-1}s)s^H \right] C_1(s)^{-1/2} = C_1(s)^{-1} \psi, C_1(s),$$

where the last equality follows since $C_\psi, C_1(F)$ has IC-property. This suggests that the corresponding DOGMA functional $W(F)$ can be calculated by a method similar to Theorem 3.

Theorem 4. Let $C_2(F)$ be any weighted spatial scatter matrix $S_\psi(F)$ with IC-property. Assume $z$ follows the ICA model and $A_8$ holds. Then $W(z)^H$ and $A_2(v)$ correspond, respectively, to the matrix of eigenvectors and diagonal matrix of eigenvalues of $C_1(z)^{-1} \psi, C_1(z)$.

Theorem 4 applies for FOBI in which case $C_1(F) = C(F), C_2(F) = K(F)$ and $C_\psi, C_1(F) = K_C(F)$. This simpler, computationally improved FOBI algorithm proceeds as follows:

(a) Calculate $Q(z) = C(z)^{-1}$.
Matlab code: $Q = (Z*Z'/n)/eye(d); \% Z$ is d times n data matrix
(b) Calculate $K_C(z) = E [z^H Q(z) z z^H]$.
Matlab code: $K = ones(d,1)*sum(conj(Z).*Q*Z).*Z*Z'/n$;
(c) Calculate $W(z)^H$ as the matrix of eigenvectors of $Q(z) K_C(z)$.
Matlab code: $[w L] = eig(Q*K); \% W = w'$;

Note that implementation of the improved FOBI algorithm requires only three lines of Matlab code. This is in sharp contrast to most of the ICA methods proposed thus far. Naturally, the above improved algorithm for FOBI can be adopted to any other DOGMA estimator based on a scatter matrix $C_1(F)$ and a weighted spatial scatter matrix $S_\psi(F)$ as the choice of $C_2(F)$. Besides algorithmic advantages, Theorem 4 provides more accessible formulation for purposes of deriving statistical properties of a DOGMA estimator employing a weighted spatial scatter matrix. In fact, we have derived influence functions, asymptotic covariance structure and efficiencies of the FOBI estimator based on the above formulation for FOBI. These will be reported in a separate paper.

6. Performance studies

6.1. Assessing the quality of separation

The performance of separation is often investigated via the “interference matrix”

$$\hat{G} = (\hat{g}_1 \cdots \hat{g}_d)^T = \hat{W}A,$$

where $\hat{W} = (\hat{w}_1 \cdots \hat{w}_d)^T = W(Z_n)$ is the separating matrix estimator and $Z_n = AS_n = (A_1 \cdots A_n)$ is a random sample from the ICA model. Due to indeterminacy of ICA, perfect separation implies that $\hat{G} = G$ for some $G \in \mathcal{G}$. Since $\hat{g}_i = \hat{w}_i^T z = \hat{g}_i^T s = \sum_{j=1}^d \hat{g}_{ij}s_j$, the values $\hat{g}_{ij}$ and $E[|\hat{g}_{ij}|^2]$ for $i \neq j$ represent the magnitude and the average power of interference of the $j$th source in the estimated $i$th source signal. If the estimator $\hat{W}$ is ICA-equivariant, then $\hat{G}$ does not depend on $A$ since $\hat{G} = W(Z_n)A = W(S_n)A^{-1}A = W(S_n)$, that is, $\hat{G}$ is simply the separating matrix estimator calculated from (unmixed) source signals $S_n = (s_1 \cdots s_n)$ and hence one could set $A = I$ in simulations without any loss of generality.

The quality of separation is assessed by calculating the widely used performance index (PI) (Amari et al., 1996)

$$\text{PI}(\hat{G}) = \frac{1}{2d(d-1)} \left\{ \sum_{i=1}^d \left( \sum_{j=1}^d \frac{|g_{ij}|}{\max_{\ell} |g_{i\ell}|} - 1 \right) + \sum_{j=1}^d \left( \sum_{i=1}^d \frac{|g_{ij}|}{\max_{\ell} |g_{\ell j}|} - 1 \right) \right\}.$$
Table 2

<table>
<thead>
<tr>
<th>Estimator</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>E1:</td>
<td>FOBI</td>
</tr>
<tr>
<td>E2:</td>
<td>DOGMA using covariance matrix and sign covariance matrix (SCM)</td>
</tr>
<tr>
<td>E3:</td>
<td>DOGMA using covariance matrix and symmetrized SCM</td>
</tr>
<tr>
<td>E4:</td>
<td>DOGMA using covariance matrix and symmetrized Tyler’s M-estimator</td>
</tr>
<tr>
<td>E5:</td>
<td>DOGMA using Tyler’s M-estimator and Huber’s M-estimator</td>
</tr>
<tr>
<td>E6:</td>
<td>DOGMA using Tyler’s M-estimator and symmetrized Huber’s M-estimator</td>
</tr>
</tbody>
</table>

Under perfect separation PI(\(\hat{G}\)) = 0. When the estimator \(\hat{W}\) fails to separate the sources, the value of PI(\(\hat{G}\)) increases. PI is scaled so that the maximum value is 1. Since PI depends on \(\hat{W}\) only via \(\hat{G}\), it is independent of the value of \(A\) if \(\hat{W}\) is ICA-equivariant.

6.2. Simulation set-up

In our simulation studies we employ DOGMA estimators listed in Table 2. We compare the results to the widely used complex FastICA (Bingham and Hyvarinen, 2000) with deflationary approach and contrast \(G_2(x) = \log(0.1 + x)\) and JADE (Cardoso, 1999).

In our simulations \(m = 300\) simulated samples \(S_n\) of the source signals \(s\) were generated using several different sample lengths ranging from \(n = 50\) to \(n = 2000\). Each sample \(S_n\) was mixed by a randomly generated mixing matrix \(A\) yielding the data set \(Z_n = AS_n\) and then PI(\(\hat{G}\)) was calculated for each estimator. Note that the mixing matrix needs to be generated randomly for each sample for the sake of fair comparison since complex FastICA and JADE are not ICA-equivariant and hence PI for these estimators is dependent on \(A\).

In Simulation (A), source r.v. \(s\) consists of \(d = 4\) circular (or second-order circular) zero-mean and unit-variance source signals: \(s_1\) has second-order circular complex uniform distribution (namely, its real part and imaginary part are independent with identical uniform distribution), \(s_2\) has circular complex Gaussian distribution, \(s_3\) has circular complex normal distribution, \(s_4\) has circular exponential distribution (namely, its modulus has exponential distribution and phase has uniform distribution on \((0, 2\pi)\)). Since all sources are symmetric (i.e. they satisfy Assumption A7), all the scatter and spatial scatter matrix estimators E1–E6 listed in Table 2 possess IC-property.

In Simulation (B), we consider the effect of outliers on estimators: 1% of the data vectors in each sample \(Z_n\) generated in Simulation (A) are replaced by outliers as follows. Let \(z_{i,\text{max}}\) denote the value of generated \(z_i\)’s in \(Z_n\) with largest modulus \((i = 1, \ldots, 4)\). We then generate i.i.d. r.v.a.’s \(u_1, \ldots, u_4\) possessing uniform distribution on \([1, 5]\) and multiply it by \(-1\) or \(1\) with equal probability \(\frac{1}{2}\) which creates a r.v.a. denoted by ±\(u_i\). Then an outlier data vector is \(z_{\text{out}} = (±u_1z_{i,\text{max}}, \ldots, ±u_4z_{i,\text{max}})^T\). We replace 1% of data vectors in \(Z_n\) by outliers generated as \(z_{\text{out}}\). Note that the outlying value \(z_{i,\text{out}} = ±u_i z_{i,\text{max}}\) lies on the same or opposite direction of \(z_{i,\text{max}}\) (i.e. phase of \(z_{i,\text{out}}\) is ± times the phase of \(z_{i,\text{max}}\)) but its magnitude is at least as big but at most 5 times larger than the magnitude of \(z_{i,\text{max}}\).

Note that robustness of an estimator is an important issue in complex-valued ICA since at most one source can possess conventional circular Gaussian distribution. Thus a separating matrix estimator should work reliably for all types of source distributions.

6.3. Results

Results for Simulation (A) are depicted in Fig. 1. The median values of the calculated PIs of the samples reveal that FastICA is performing the best which is not surprising since it is designed for circular sources. Also JADE is performing well and so are the estimators E5 and E6 which are just a bit behind it. Surprisingly estimators E2, E3, E4 and FOBI have very similar performance, but they do not quite reach the level of E5 and E6. Fig. 1 also depicts boxplots (or, box-whisker plots) of PI values for each estimator at sample length \(n = 1000\). Recall that the line that cuts the box is at the median value and boundaries of the box at the lower quartile and upper quartile values. Outliers are PI values that go beyond the ends of the whiskers. The boxplots demonstrate that E5 and E6 have the smallest extent of PI values and in this sense they are performing better than FastICA and JADE.
Fig. 1. Results for Simulation (A): Median performance index (first column) of various methods as the function of the sample length and boxplots of PIs (second column) for sample length \( n = 1000 \). The acronyms E2 etc. refer to DOGMA estimators listed in Table 2. All estimators are able to separate the sources. FastICA which is designed for circular sources has the best performance.

Fig. 2. Results for Simulation (B): Median performance index (first column) of various methods as the function of the sample length and boxplots of PIs (second column) for sample length \( n = 1000 \). The acronyms E2 etc. refer to DOGMA estimators listed in Table 2. Only the DOGMA estimators E5 and E6 which employ a pair of robust scatter matrices are not affected by outliers and are able to separate the sources. Results for Simulation (B) are shown in Fig. 2. Only the methods E5 and E6 which employ a pair of robust scatter matrices are not affected by outliers and are seen to perform very well. Methods, which use the conventional, non-robust covariance matrix for whitening, \( i.e. \) FOBI, FastICA, JADE and methods E2, E3 and E4 fail completely due to the effect of outliers. The boxplots show that E5 has the best performance considering the median value and the extent of PI values.

7. Communications example

In this section, we consider a practical sensor array signal processing application — a hard core topic in signal processing; see e.g. Krim and Viberg (1996) for an excellent review and text-books by Johnson and Dudgeon (1993) and Van Trees (2002) and their bibliographies. In sensor array signal processing, a group of sensors located at distinct spatial locations are deployed to measure propagating wavefields through a medium (for example air or water). The multichannel sensor outputs are then processed to obtain knowledge of the parameters of interest. Application areas that use arrays include communications, radar, sonar, seismology, biomedicine and astronomy. Sensor array processing deals with complex-valued data. Many pioneering works of ICA, such as Cardoso’s FOBI and JADE algorithms, were originally derived for the sensor array model. In fact, the real-valued versions of the algorithms were considered later.
when the generic real-valued ICA model was brought into use. In this sense, with the example studied in this section, we are in essence at the origin of ICA.

We start by formulating the sensor array signal processing problem which is explained in more detail in Krim and Viberg (1996), Johnson and Dudgeon (1993) and Van Trees (2002). Consider \( m \) far-field and narrowband signals \( s_1(t), \ldots, s_m(t) \) impinging on an array of \( d \) sensors \( (m \leq d) \) from distinct Direction-of-Arrivals (DOAs) \( \theta_1, \ldots, \theta_m \). The sensor array output \( z(t) \) at time instant \( t \) can be modelled as

\[
  z(t) = A(\theta)s(t) + n(t)
\]

where \( \theta = (\theta_1, \ldots, \theta_m)^T \), \( s(t) = (s_1(t), \ldots, s_m(t))^T \) is the \( m \)-vector of signal waveforms, \( A(\theta) = (a(\theta_1) \cdots a(\theta_m)) \) is the complex \( d \times m \) array response matrix related to DOAs and \( n(t) \) is the \( d \)- variate sensor thermal noise vector commonly modelled as complex circular Gaussian random vector with covariance matrix \( \mathcal{C}(n(t)) = \sigma_n^2 I \). Each vector \( a(\theta_1) \) corresponds to a point in the known array manifold (array transfer function, steering vector) \( a(\theta) \) representing the response of the array to a wavefield with DOA \( \theta \). For example, in case of uniform linear array (ULA),

\[
  a(\theta) = (1, e^{-j\omega}, \ldots, e^{-j(d-1)\omega})^T,
\]

where \( \omega = 2\pi(\delta/\lambda)\sin(\theta) \) depends on the signal wavelength \( \lambda \), the DOA \( \theta \in [-\pi/2, \pi/2] \) of the signal w.r.t. broadside, and the sensor spacing \( \delta \). An example of a multiantenna sensing system with ULA configuration is depicted in Fig. 3. Based on the measured data, denoted by \( z_1, \ldots, z_n \) (samples at discrete time instants), the goal – which depends largely on the application at hand – is e.g. to estimate the source signals, the DOAs, to enhance the signal-of-interest (SOI) from certain direction and attenuate undesired signals from other directions, locate the position of sources, etc.

The assumption that sampled source signals are statistically independent is often realistic and plausible for physically separated emitters. Furthermore, under the high signal-to-noise ratio (SNR) \( 10 \log_{10}[\sigma_s^2/\sigma_n^2] \) (dB) assumption (e.g. SNR \( \geq 20 \) dB or larger), we essentially have the linear (“noise-free”) complex-valued ICA model, with the distinction that mixing matrix \( A = A(\theta) \) possesses a specific known structure, e.g. Vandermonde structure in case of ULA. Identifying the mixing matrix \( A(\theta) \) is naturally equivalent with identifying the DOAs of the sources, which subsequently allows for the estimation of source signals. Many efficient DOA estimation techniques have been proposed, e.g. ESPRIT, MUSIC, root-MUSIC to name only a few. These methods assume that the array manifold is known and they are often limited for certain array geometries. Due to the lack of knowledge about array geometry and sensor locations, the array manifold \( a(\theta) \) is unknown and we are essentially left with the complex-valued ICA model with unknown mixing matrix \( A \). Methods which assume no knowledge about array manifold are often called blind array processors.

We now explore performance of our method with a communications example on synthetic data using the Matlab technical computing environment. Three independent random signals, a BPSK signal, 8-QAM signal and circular Gaussian signal of equal power \( \sigma_s^2 \) are impinging on \( d = 3 \) sensor ULA with half a wavelength interelement spacing.
We then estimated the source signals via JADE and the robust DOGMA method E5 listed in Table 2. Shown in Fig. 4 are the signal constellations obtained by JADE and E5. As can be seen, both methods are able to separate the sources: BPSK, 8-QAM and circular Gaussian source signals are clearly discernible on the left, middle and right plots. Naturally, due to ICA ambiguities, sources are estimated only up to a phase (rotation), sign and scale. PI was 0.0252 for E5, 0.0165 for JADE and 0.0170 for FastICA.

To illustrate the reliable performance of the DOGMA method E5 under contamination, four observations were replaced by an outlier $z_{\text{out}}$ generated as in Simulation (B). Hence only 0.2% of the data is contaminated. Shown in Fig. 5 are the estimated source signal constellations obtained by JADE and E5. As can be seen, only E5 is able to separate the sources and is unaffected by outliers, which are clearly detected in the plots. JADE method on the other hand fails completely: BPSK and 8-QAM sources are no longer discernible in the plots. The reliable performance of E5 is evident in the calculated PI values: 0.027 for E5, 0.7024 for JADE and 0.4804 for FastICA.

8. Conclusions and plans for future work

In this paper FOBI algorithm was generalized to an algorithm called DOGMA which employs any pair of scatter and spatial scatter matrices with IC-property in place of the covariance and kurtosis matrices in the FOBI algorithm. Our approach yields a large family of estimators, which – depending on the choices of the scatter and spatial matrix – can have largely different statistical properties. For example, DOGMA estimators that employed a pair of robust scatter and spatial scatter matrices (E5 and E6) were shown to outperform the original FOBI estimator in our simulation set-up with or without outliers. The performance of these robust DOGMA estimators was unaffected by outliers whereas FOBI completely failed to separate the sources. This behaviour was detected also in the communications example of Section 7. DOGMA estimators provide a computationally attractive class of estimators since they avoid
difficult optimization tasks (common to most ICA algorithms) and are solely based on straightforward matrix computations. The statistical properties (influence functions, consistency, limiting distributions and efficiencies) of DOGMA estimators are currently under our investigation and will be reported in a separate paper.

A drawback of the method is that it requires Assumption A9, namely, that the eigenvalues $\Lambda_2(v)$ are distinct. Since the eigenvector matrix $U_2(v)$ of $C_2(v)$ is unique up to right multiplication by a unitary matrix corresponding to sources with the same eigenvalue, we could use the following approach to separate sources with identical eigenvalue. Suppose that $s_I = (s_{i_1}, \ldots, s_{i_m})^T$ contain the set of sources with identical eigenvalue $\delta = [A_2(v)]_{ii} = [C_2(C_1(s)^{-1/2}s)]_{ii} \forall i \in I = \{i_1, \ldots, i_m\}$ and let $U_{2,I}(v) = (u_{i_1}(v) \cdots u_{i_m}(v))$ be the $d \times m$ matrix with corresponding eigenvectors. Then $\hat{s}_I = U_{2,I}(v)^H B_1(z)$ is a copy of $s_I$ up to a unitary matrix. Then any ICA method that does not assume A4' and that is designed for whitened sources (such as FastICA) can be used to find the unitary factor that recovers the sources $s_I$ from $\hat{s}_I$. Naturally, for finite samples, the estimated eigenvalues $\hat{\delta}_i$ are not exactly equal and hence one needs a reliable statistical procedure/test for guiding the decision concerning equality of eigenvalues.

A more practical extension of the DOGMA method (which avoids Assumption A9) is obtained by employing the joint diagonalization approach used in JADE algorithm of Cardoso and Souloumiac (1993). We are currently working towards an algorithm that generalizes the JADE method. It was assumed that data is centered by an appropriate affine equivariant location functional $T(z)$, so that $T(z) = T(s) = 0$ holds. This is a requisite preprocessing step needed by DOGMA methods and basically all ICA methods in general. For example, in the case of FOBI, the mean vector $T(z) = E[z]$ is used for centering. For robustness purposes, it is sensible to employ robust location functional $T(\cdot)$ instead of the mean vector. Naturally, if one uses symmetrized $M$-estimators, then centering is not needed. In the case of $M$-estimators, location vector and scatter matrix can be estimated jointly. The joint $M$-functionals of location vector $T(F) \in \mathbb{C}^d$ and scatter matrix $C(F) \in PDH(d)$ are the joint solutions to a pair of implicit equations

$$T(F) = E[\psi(r)z]/E[\psi(r)]$$

(4)
(5)
\[ C(F) = \mathbb{E}[\psi(r^2)(z - T(F))(z - T(F))^H] , \]
where \( r^2 = (z - T(F))^H C(F)^{-1} (z - T(F)) \) and \( \psi \) and \( \varphi \) are real-valued functions on \([0, \infty)\). For example, complex Huber’s \( M \)-estimators of location and scatter are obtained using \( \psi(\cdot) \) as defined in (2) and \( \psi \)-function, defined as \( \psi(x) = 1 \), if \( r \leq c \), \( \psi(x) = c/r \), if \( r > c \).

Acknowledgement
The first author would like to thank the Academy of Finland for supporting this research.

Appendix. Proofs

Proof of Lemma 1. Since \( s \) has ICs that satisfy A7, it follows that \( Js^d_s = s \) for all diagonal matrices \( J \) with diagonal elements \( \pm 1 \). By affine/unitary equivariance, \( C(Js) = JC(s)J \), which implies that \( [C(s)]_{ij} = -[C(s)]_{ji} \) for all \( i \neq j \). Thus \( C(s) \) is a diagonal matrix. \( \square \)

Proof of Lemma 2. Assume \( s \) has independent components. Then the components of r.v. \( s_1 - s_2 \) are independent as well and they satisfy A7. Thus \( C(F_{s_1 - s_2}) \) possesses independence property by Lemma 1. \( \square \)

Proof of Theorem 1. Since \( C_1(F) \in \text{PDH}(d) \) has IC-property, \( C_1(s) \) is a diagonal matrix with real positive diagonal elements. Write \( C_1(s)^{1/2} \) for the diagonal matrix with square roots of the diagonal elements of \( C_1(s) \) on the diagonal and \( C_1(s)^{-1/2} \) for its inverse. The whitened mixture \( v = B_1(z)v \) follows the ICA model \( v = \tilde{A}s \), where \( \tilde{A} = B_1(z)A \) is the new mixing matrix. Then, from
\[ I = C_1(v) = C_1(\tilde{A}s) = \tilde{A}C_1(s)\tilde{A}^H = (\tilde{A}C_1(s)^{1/2})(\tilde{A}C_1(s)^{1/2})^H , \]
we observe that \( \tilde{A}C_1(s)^{1/2} \) is a unitary matrix. Since \( C_2(F) \) is unitary equivariant and has IC-property, it follows that
\[ C_2(v) = C_2(\tilde{A}s) = C_2((\tilde{A}C_1(s)^{1/2})C_1(s)^{-1/2}s) \]
\[ = \tilde{A}C_1(s)^{1/2}C_2(C_1(s)^{-1/2}s)(\tilde{A}C_1(s)^{1/2})^H . \]
Since \( \tilde{A}C_1(s)^{1/2} \) is a unitary matrix and \( C_2(C_1(s)^{-1/2}s) \) is a diagonal matrix with positive real diagonal elements, it follows that (A.1) is the EVD (3), namely, \( U_2(v) = \tilde{A}C_1(s)^{1/2} \) and \( A_2(v) = C_2(C_1(s)^{-1/2}s) \). \( \square \)

Proof of Corollary 1. Since the eigenvalues \( \delta_1, \ldots, \delta_d \) are distinct due to A9, the matrix of eigenvectors \( U_2(v) = B_1(z)AC_1(s)^{1/2} \) is unique up to right multiplication by a phase-shift matrix \( L \) and a permutation matrix \( P \). This then means that \( U_2(v)^{-1} = U_2(v)^H \) is fundamentally equal to \( A^{-1}B_1(z)^{-1} \), and hence \( W(z) = U_2(v)^H B_1(z) \) is fundamentally equal to \( A^{-1} \), i.e. it is a separating matrix. \( \square \)

Proof of Corollary 2. Suppose (for simplicity of presentation) that an eigenvalue of \( C_2(v) \) has multiplicity \( m \), but all others are distinct, so that \( s_1 \in \mathbb{C}^{d-m} \) and \( s_2 \in \mathbb{C}^m \). Decompose \( W(z), U_2(v) \) and \( A^{-1} \) accordingly:
\[ W(z) = \begin{pmatrix} W_{(1)}(z) \\ W_{(2)}(z) \end{pmatrix} , \quad U_2(v)^H = \begin{pmatrix} U_{(1)}(v)^H \\ U_{(2)}(v)^H \end{pmatrix} \quad \text{and} \quad A^{-1} = \begin{pmatrix} A_{(1)}^{-1} \\ A_{(2)}^{-1} \end{pmatrix} . \]
In the above formulae, the block matrix on top is a \((d-m) \times d\) and block matrix on bottom is a \(m \times d\) matrix. It follows from the uniqueness properties of EVD that \( U_{(1)}(v)^H \) is unique up to left multiplication by \((d-m) \times (d-m)\) phase-shift matrix \( L \) and permutation matrix \( P \), and \( U_{(2)}(v)^H \) is unique up to left multiplication by a unitary matrix \( Q \in \mathbb{C}^{m \times m} \). By Theorem 1, we have that \( W(z) = U_2(v)^H B_1(z) = C_1(s)^{-1/2}A^{-1} \), and hence, \( W_{(1)}(z) = U_{(1)}(v)^H B_1(z) \) is fundamentally equal to \( A_{(1)}^{-1} \). \( \square \)

Proof of Theorem 2. Write \( \tilde{z} = Qz \). Recall that \( W(z) = U_2(B_1(z)z)^H B_1(z) \) and hence \( W(\tilde{z}) = U_2(B_1(Qz)z)^H B_1(Qz) \). Since \( C_1(z)^{-1} = B_1(z)^H B_1(z) \) and \( C_1(\tilde{z})^{-1} = (Q^{-1})^H C_1(z)^{-1} Q^{-1} \) due to affine equivariance, a square-root matrix of \( C_1(\tilde{z})^{-1} \) is \( B_1(z) = B_1(z)Q^{-1} \). Thus
\[ W(\tilde{z}) = U_2(B_1(z)Q^{-1} Qz)^H B_1(z)Q^{-1} = W(z)Q^{-1} . \]
The property for the estimator $W(Z_n)$ follows in an analogous manner. □

**Proof of Theorem 3.** The first step is to note that affine equivariance of $C_2$ implies that $C_2(v) = C_2(B_1(z)v) = B_1(z)C_2(z)B_1(z)^H$. The second step is to note that EVD of $C_2(F)$ implies that $C_2(v)U_2(v) = U_2(v)\Lambda_2(v)$. Combining these results and recalling that $B_1(z)^H B_1(z) = C_1^{-1}(z)$, we have that

$$B_1(z)C_2(z)B_1(z)^H U_2(v) = U_2(v)\Lambda_2(v)$$

$$\Leftrightarrow C_1(z)^{-1}C_2(z)B_1(z)^H U_2(v) = [B_1(z)^H U_2(v)]\Lambda_2(v).$$

Thus $W(z)^H$ corresponds to the matrix of eigenvectors of $C_1(z)^{-1}C_2(z)$ and $\Lambda_2(v)$ is the diagonal matrix of eigenvalues of $C_1(z)^{-1}C_2(z)$. The next claim follows at once since matrix $Q^{-1}$ (say) has the same set of eigenvectors as $Q$ and eigenvalues that are reciprocals of the eigenvalues of $Q$. □

**Proof of Theorem 4.** The first step is to observe that

$$S_\psi(v) = S_\psi(B_1(z)v) = E\left[\psi\left(z^H B_1(z)^H B_1(z)v\right) B_1(z)zz^H B_1(z)^H\right]$$

$$= B_1(z)C_\psi(C_1(z)B_1(z))^H,$$

where we used that $C_1(z)^{-1} = B_1(z)^H B_1(z)$. The remaining steps are as in the proof of Theorem 3. □

**References**


