Complex ICA using generalized uncorrelating transform

Esa Ollila\(^{a,b,*}\), Visa Koivunen\(^b\)

\(^a\) Department of Mathematical Sciences, University of Oulu, P.O. Box 3000, FIN 90014, Oulu, Finland
\(^b\) Signal Processing Laboratory, Helsinki University of Technology, Finland

**Abstract**

An extension of the whitening transformation for complex random vectors, called the generalized uncorrelating transformation (GUT), is introduced. GUT is a generalization of the strong-uncorrelating transform [J. Eriksson, V. Koivunen, Complex-valued ICA using 2nd-order statistics, in: Proceedings of the IEEE Workshop on Machine Learning for Signal Processing (MLSP’04), Sao Luis, Brazil, 2004] based upon generalized estimators of the covariance and pseudo-covariance matrix, called the scatter matrix and spatial pseudo-scatter matrix, respectively. Depending on the selected scatter and spatial pseudo-scatter matrix, GUT estimators can have largely different statistical properties. Special emphasis is put on robust GUT estimators. We show that GUT is a separating matrix estimator for complex-valued independent component analysis (ICA) when at most one source random variable possesses circularly symmetric distribution and sources do not have identical distribution. In the context of ICA, our approach is computationally attractive as it is based on straightforward matrix computations. Simulations and examples are used to confirm reliable performance of our method.

**Keywords:**

Blind source separation
Independent component analysis
Non-circular complex random vector
Robustness
Whitening transform

1. Introduction

Non-circularity (or non-properness) of complex random vectors (r.v.’s) has attained considerable research interest during the last decade (see e.g. [1–9]). Non-circular complex random variables arise in communications and radar signals with real-valued modulations such as binary phase-shift keying (BPSK) or GMSK. For circular r.v. \( z \in \mathbb{C}^d \), the conventional whitening transform \( v = Bz \) where \( B \) is a whitening matrix uncorrelates the data. If \( z \) has non-circular components, then the conventional covariance matrix do not completely describe the 2nd-order properties, and consequently, the conventional whitening transformation is not sufficient in removing the correlations between the components of \( z \).

A transformation that maps \( z \) to (strongly)uncorrelated components is achieved via the SUT (strong-uncorrelating transform) \([6,10]\) \( W \in \mathbb{C}^{d \times d} \), which is a data transformation that jointly diagonalizes the covariance and pseudo-covariance matrix of the transformed data \( s = Wz \). We define GUT (generalized uncorrelating transformation) as a data transformation that jointly diagonalizes generalized estimators of the covariance and pseudo-covariance matrix, called the scatter matrix and spatial pseudo-scatter matrix. Depending on the choices of the scatter and spatial pseudo-scatter matrix, the obtained GUT matrix can have largely different statistical properties (convergence, limiting distributions, robustness, efficiency). Our focus is on robust GUT matrices (those that employ a pair of robust scatter and spatial pseudo-scatter matrix). Important properties of the GUT matrices, such as uniqueness and equivariance under invertible linear transformations, and different algorithms and approaches to compute the GUT matrix are derived.

De Lathauwer and De Moor [5] and Eriksson and Koivunen [6,10], showed that SUT matrix is a separating...
matrix for complex-valued independent component analysis (ICA) model provided that at most one of the sources possess circular distribution and sources do not have identical distribution. This is shown to be true for the whole class of GUT matrices. For purposes of ICA, however, we need to require that the selected pair of scatter and spatial pseudo-scatter matrix possess IC-property (independent component), namely, that they reduce to a diagonal matrix at distributions with independent marginals. Since a scatter and spatial pseudo-scatter matrix do not generally possess IC-property, we highlight that a symmetrized version of the estimator always possess this property. Many ICA methods implicitly or explicitly assume circularity of the sources and may perform poorly under violations of this assumption. Our approach for ICA is computationally attractive as it avoids high complexity iterative optimization and is solely based on straightforward matrix computations; in fact, we illustrate that an efficient implementation of the GUT algorithm for ICA requires only three lines of Matlab code. Moreover, a robust GUT matrix provides a robust separating matrix estimator (i.e. insensitive to outliers in the data set).

The paper is organized as follows. Section 2 reviews notations and the Takagi factorization of a complex symmetric matrix. Important statistics (circularity, covariances and pseudo-covariances, SUT) based on complex r.v.'s are reviewed as well. In Section 3.2, the GUT matrix is formally defined, its properties (existence, uniqueness, equivariance) are established as well as a general algorithm for its computation. Also, alternative, simpler methods for its computation under the assumptions of distinct circularity coefficients are derived. In Section 4, application to complex-valued ICA is considered with extensive simulation studies presented in Section 5. Section 6 contains a communications example and Section 7 concludes. Appendix A reviews M-estimators of scatter and proofs of the results are given in Appendix B.

2. Complex random numbers and variables: preliminaries

2.1. Notations

Superscripts (·)\textsuperscript{H}, (·)\textsuperscript{T}, (·)* stand for the Hermitian transpose, transpose and complex conjugate, respectively. Symbol \|·\| denotes the modulus of a complex number (\|z\| = \sqrt{z\overline{z}}), \|·\| denotes the norm of a complex vector (\|z\| = \sqrt{z^\dagger z}) and symbol \sim_d reads “has the same distribution as”. Recall that every non-zero complex number has a unique (polar) representation, z = |z| e^\i \theta, where \(-\pi \leq \theta < \pi\) is called the argument of z, denoted \theta = arg(z). As a convention, for \(z = 0\), we define arg(0) = 0.

Complex matrix \(A \in \mathbb{C}^{d \times d}\) is Hermitian (resp. symmetric) if \(A^\dagger = A\) (resp. \(A^T = A\)) and unitary (resp. orthogonal) if \(A^\dagger A = I\) (resp. \(A^T A = I\)), where \(I\) denotes the identity matrix. By PDH(d) and CS(d) we denote the set of \(d \times d\) positive definite Hermitian and complex symmetric matrices, respectively. By \([A]_{ij}\) we denote the \((i,j)\)th component of \(A\) and notation \(A = \text{diag}(a_i)\) implies that \(A\) is a diagonal matrix with diagonal elements \(a_1, \ldots, a_d\). Two \(d \times d\) matrices \(B\) and \(A\) are said to be essentially equal, denoted by \(B = A\), if \(B = \beta A\) for some \(d \times d\) ambiguity matrix \(\beta\), defined as a complex matrix with only one non-zero complex value in each row and in each column. Matrix \(B \in \mathbb{C}^{d \times d}\) is a square-root matrix of \(A \in \mathbb{C}^{d \times d}\) if \(B^B B = A\). The principal square-root matrix of \(A \in \mathbb{C}^{d \times d}\) is a matrix, denoted by \(A^{1/2} \in \mathbb{C}^{d \times d}\), such that \(A = (A^{1/2})^2\). If \(A\) is non-singular, then \(A^{-1/2}\) denotes the inverse of \(A^{1/2}\) (the principal square root of \(A^{-1}\)).

2.2. Takagi factorization

Complex matrix \(A \in \mathbb{C}^{d \times d}\) has a singular value decomposition (SVD)

\[
A = GAV^\dagger = \sum_{i=1}^{d} \lambda_i g_i v_i^\dagger, \tag{1}
\]

where \(G\) (resp. \(V\)) is a unitary matrix of left (resp. right) singular vectors \(g_i\) (resp. \(v_i\)) as columns and \(A = \text{diag}(\lambda_i)\) is a non-negative diagonal matrix of singular values \(\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d \geq 0\).

If \(A \in \text{CS}(d)\), then it possesses a special form of SVD, called Takagi factorization (cf. [11]), where the right singular vectors are complex conjugates of the left singular vectors, namely

\[
A = U_{\text{T}} U_{\text{T}}^\dagger = \sum_{i=1}^{d} \lambda_i u_i u_i^\dagger, \tag{2}
\]

where the Takagi factor \(U\) is a unitary matrix of Takagi vectors \(u_i\), as columns. Takagi factorization is computationally efficient as only one set of singular vectors (instead of two sets of singular vectors as in the SVD (1) of a general matrix) needs to be computed. In Takagi factorization (2) of \(A \in \text{CS}(d)\), matrix \(A\) equals the singular values of \(A\), but Takagi vectors \(u_i\) are not necessarily equal to the left singular vectors \(g_i\) of \(A\). Next we show that if the distinct singular values are assumed then \(u_i\) equals \(g_i\) up to a phase shift.

Lemma 1. If the singular values of \(A \in \text{CS}(d)\) are distinct, then \(U = G_{\text{T}} \text{L}_{1/2}\), where \(L = G_{\text{T}}^H V^\dagger = \text{diag}(e^{\i \theta_i})\), \(\theta_i \in \mathbb{R}\), is a phase-shift matrix.

Note also that \(A^* = U L^2 U^\dagger\), i.e. Takagi factor \(U\) and the squared singular values \(\lambda_1^2, \ldots, \lambda_d^2\) correspond to the matrix of eigenvectors and eigenvalues of the positive semidefinite Hermitian matrix \(A^* A\) [11, Corollary 4.4.4]. This result does not imply (due to the lack of uniqueness of EVD) that every orthonormal set of eigenvectors of \(A^* A\) gives identifiable Takagi factor \(U\), unless singular values (the eigenvalues of \(A^* A\)) are assumed to be distinct.

Lemma 2 (Horn and Johnson [11, Corollary 4.4.5]). If the singular values of \(A \in \text{CS}(d)\) are distinct, then \(U = GL\), where \(Q\) is the unitary matrix of eigenvectors of \(A^* A\) and \(L = \text{diag}(e^{\i \theta_i})\), \(\theta_i = \frac{1}{2} \text{arg}([Q^H A^*]^*)_{ii}\).

If the assumption of distinct singular values does not hold (as required by Lemmas 1 and 2), then specialized
numerical code (e.g. [12]) is needed to compute the Takagi factorization.

2.3. Complex r.v.: preliminaries

Let \( \mathbf{z} = (z_1, \ldots, z_d)^T = \mathbf{x} + j \mathbf{y} \in \mathbb{C}^d \)

denote the observed complex r.v. with zero mean, i.e.
\[ \mathbb{E} [\mathbf{z}] = \mathbb{E} \mathbf{x} + j \mathbb{E} \mathbf{y} = \mathbf{0}, \]
where each component \( z_i = x_i + j y_i \)

is assumed to be non-degenerate. The variance \( \sigma_i^2 = \mathbb{E} (z_i^2) > 0 \) of \( z_i \) is defined as
\[ \sigma_i^2 (z_i) = \mathbb{E} [|z_i|^2] = \mathbb{E} [\chi_i^2] + \mathbb{E} [\gamma_i^2] \]

and the pseudo-variance \( \tau_i = \tau(z_i) \in \mathbb{C} \) of \( z_i \) is defined as
\[ \tau(z_i) = \mathbb{E} (z_i^2) = \mathbb{E} [\chi_i^2] - \mathbb{E} [\gamma_i^2] + 2 \mathbb{E} [x_i y_i]. \]

Besides the 2nd-order moments of the components, also their correlations are of interest: the covariance and pseudo-covariance [1] between components \( z_i \) and \( z_j \) are defined as \( \text{cov}(z_i, z_j) = \mathbb{E} [z_i z_j^*] \) and \( \text{pcov}(z_i, z_j) = \mathbb{E} [z_i z_j^*] \).

Note that \( \sigma_i^2 = \text{cov}(z_i, z_i) \) and \( \tau_i = \text{pcov}(z_i, z_i) \). More compactly in matrix form, the whole 2nd-order description of \( \mathbf{z} \) is given by the covariance matrix \( \mathbb{C} (\mathbf{z}) = \mathbb{E} ||\mathbf{z}|^2| \) together with the pseudo-covariance matrix \( \mathbb{P} (\mathbf{z}) = \mathbb{E} ||\mathbf{z}|^2| \).

Observe that \( \mathbb{C} (\mathbf{z}) \in \mathbb{D} (d) \) (i.e. we assume that \( \mathbb{C} \) do not have a zero eigenvalue) and \( \mathbb{P} (\mathbf{z}) \in \mathbb{C} (d) \).

Characteristics of a complex r.v. can also be described via symmetry properties of its distribution. The most commonly made symmetry assumption in the statistical signal processing literature is that of circular symmetry. See e.g. [2]. Complex r.v. \( \mathbf{z} \) is said to be circular if \( \mathbf{z} = e^{j \Theta} \mathbf{z} \) for all \( \Theta \in \mathbb{R} \) and symmetric if \( \mathbf{z} = - \mathbf{z} \). Note that circularity implies symmetry. The r.v. \( \mathbf{z} \) is said to be 2nd-order circular [2] (or proper [11]) if \( \mathbb{P} (\mathbf{z}) = 0 \). Obviously, if \( \mathbf{z} \) is circular, then it is proper provided that the components of \( \mathbf{z} \) have finite variances.

The r.v. \( \mathbf{z} = (z_1, \ldots, z_d)^T \) is said to have uncorrelated components [11] if
\[ \text{cov}(z_i, z_j) = \text{pcov}(z_i, z_j) = 0, \quad \forall 1 \leq i \neq j \leq d \]

and strongly uncorrelated components if, in addition, \( z_i = x_i + j y_i \) are scaled such that (i) they have unit variance (i.e. \( \sigma_i^2 = 1 \)) and (ii) non-negative real-valued pseudo-variance (i.e. \( \arg(\tau_i) = 0 \)). Then, matrix \( \mathbf{W} \) is said to be SUT [16,10] if transformed data \( \mathbf{s} = \mathbf{Wz} \) has strongly uncorrelated components, i.e.
\[ \mathbb{C} (\mathbf{s}) = \mathbf{I} \quad \text{and} \quad \mathbb{P} (\mathbf{s}) = \text{diag}(\tau_i), \]

where \( \tau_i = \tau(z_i) > 0, \quad i = 1, \ldots, d \) are called the circularity coefficients of \( \mathbf{z} \). Note that for a circular r.v. \( \mathbf{z} \), the conventional whitening transform \( \mathbf{B} = \mathbb{C} (\mathbf{z})^{-1/2} \) is also the SUT, but this is not true for non-circular r.v.

3. Generalized uncorrelating transform

3.1. Definition

**Definition 1.** Let \( \mathbf{s} = \mathbf{Az} \) and \( \mathbf{v} = \mathbf{Uz} \) denote the non-singular linear and unitary transformations of \( \mathbf{z} \in \mathbb{C}^d \) for any non-singular \( \mathbf{A} \in \mathbb{C}^{d \times d} \) and unitary \( \mathbf{U} \in \mathbb{C}^{d \times d} \). Then, the matrix functional \( \mathbf{C} \in \mathbb{D} (d) \) is called a scatter matrix if \( \mathbb{C} (\mathbf{s}) = \mathbf{AC} (\mathbf{z}) \mathbf{A}^H \). Matrix functional \( \mathbf{P} \in \mathbb{C} (d) \) is called a pseudo-scatter matrix (resp. spatial pseudo-scatter matrix) if \( \mathbb{P} (\mathbf{s}) = \mathbf{AP} (\mathbf{z}) \mathbf{A}^T \) (resp. \( \mathbb{P} (\mathbf{v}) = \mathbf{UP} (\mathbf{z}) \mathbf{U}^T \)).

Spatial pseudo-scatter matrix is a broader notion than pseudo-scatter matrix since it requires equivariance only under unitary linear transformations, i.e. every pseudo-scatter is also a spatial pseudo-scatter matrix. **Weighted spatial pseudo-covariance matrix** \( \mathbb{P} (\mathbf{z}) = \mathbb{E} (\mathbf{v} (|\mathbf{z}|^2) \mathbf{z} \mathbf{z}^T) \),

where \( \phi (\cdot) \) is any real-valued weighting function on \([0, \infty)\), is an example of spatial pseudo-scatter matrix (but not of a pseudo-scatter matrix). Weights \( \phi (x) = x \) and \( \phi (x) = x^{-1} \) yield matrices called the **pseudo-kurtosis matrix** \( \mathbb{P}_{\text{kur}} (\mathbf{z}) = \mathbb{E} (\mathbf{v} (|\mathbf{z}|^2) \mathbf{z} \mathbf{z}^T) \)

and the **sign pseudo-covariance matrix** (SPM) \( \mathbb{P}_{\text{sign}} (\mathbf{z}) = \mathbb{E} (|\mathbf{z}|^2 \mathbf{z} \mathbf{z}^T) \)

respectively. These matrix functionals have natural Hermitian counterparts that have been shown to be useful in blind separation and array signal processing problems in [13–15]. These matrices also possess very different statistical (e.g. robustness) properties.

The covariance matrix and pseudo-covariance matrix serve as examples of a scatter matrix and pseudo-scatter matrix, respectively. More general family of scatter and pseudo-scatter matrices are the weighted covariance matrix and weighted pseudo-covariance matrix, defined as
\[ \mathbb{C}_{\phi, \mathbf{C}} (\mathbf{z}) = \mathbb{E} (\phi (|\mathbf{z}|^4) \mathbf{C} (\mathbf{z})^{-1} \mathbf{z} \mathbf{z}^T), \]

\[ \mathbb{P}_{\phi, \mathbf{C}} (\mathbf{z}) = \mathbb{E} (\phi (|\mathbf{z}|^2) \mathbf{C} (\mathbf{z})^{-1} \mathbf{z} \mathbf{z}^T), \]

respectively, where \( \phi (\cdot) \) is any real-valued weighting function on \([0, \infty)\) and \( \mathbf{C} \) is any scatter matrix, e.g. the covariance matrix. Note that the covariance matrix and the pseudo-covariance matrix are obtained with unit weight \( \phi = 1 \). An improved idea of the weighted covariance matrix are the complex M-functionals of scatter [16–18] defined in Appendix A.

**Definition 2.** Let \( \mathbf{C} \) be a scatter matrix and \( \mathbf{P} \) a spatial pseudo-scatter matrix. Matrix functional \( \mathbf{W} = \mathbf{W} (\mathbf{z}) \in \mathbb{C}^{d \times d} \) of \( \mathbf{z} \in \mathbb{C}^d \) is called the GUT if transformed data \( \mathbf{s} = \mathbf{Wz} \) satisfies
\[ \mathbb{C} (\mathbf{s}) = \mathbf{I} \quad \text{and} \quad \mathbb{P} (\mathbf{s}) = \Lambda, \]

where \( \Lambda = \mathbf{A} (\mathbf{s}) = \text{diag}(\lambda_i) \) is a real non-negative diagonal matrix, called the circularity matrix, and \( \lambda_i = ||\mathbf{P} (\mathbf{s})||_2 \geq 0 \) is called the \( i \)th circularity coefficient, \( i = 1, \ldots, d \).

The GUT matrix with choices \( \mathbb{C} = \mathbb{C} ' \) and \( \mathbb{P} = \mathbb{P} ' \) corresponds to the SUT. Essentially, \( \mathbf{W} \) is a data transformation that jointly diagonalizes the selected scatter and spatial pseudo-scatter matrix of the transformed data \( \mathbf{s} = \mathbf{Wz} \). Note that the pseudo-covariance matrix employed by SUT is a pseudo-scatter matrix, whereas in Definition 2, we only require \( \mathbf{P} \) to be a spatial pseudo-scatter matrix.
3.2. Properties: existence, uniqueness and equivariance

So far we have neglected the issue of existence of a GUT matrix. Corollary 4.1(b) of [11] states that GUT matrix always exists and can be calculated via following steps:

(a) Calculate the square-root matrix $B(z)$ of $C(z)^{-1}$, so $B(z)B(z)^T = C(z)^{-1}$, and the whitened data $v = B(z)z$.

(b) Calculate Takagi's factorization of $P$ for the whitened data $v$:

$$P(v) = U A U^T,$$

where $U = U(v) \in \mathbb{C}^{d \times d}$ is the Takagi factor of $P(v)$ and $A$ is the circularity matrix (i.e. the singular values of $P(v)$ are the circularity coefficients $\lambda_i = |P(v)|_{ii}$ appearing in (3)).

(c) Set $W(z) = U(v)^H B(z)$.

In step (a), data is whitened in the sense that $C(v) = I$. Naturally, if the selected scatter matrix is the covariance matrix, then the data is whitened in the conventional sense. Since the whitening transform $B$ is unique up to left-multiplication by a unitary matrix, GUT matrix $W = U^*B$ is also a whitening transform in the conventional sense but with an additional property that it diagonalizes the selected spatial pseudo-scatter matrix.

Next we turn into uniqueness properties of GUT. Write $W = (w_1 \ldots w_d)^T$, i.e. $w_i \in \mathbb{C}^d$ is the (transposed) $i$th row vector of $W$ (corresponding to $i$th circularity coefficient $\lambda_i$). Then consider the cases U1 and U2 listed below indicating increasing lack of uniqueness of the GUT matrix.

U1 Circularity coefficients are distinct, i.e. $\lambda_i \neq \lambda_j, \forall 1 \leq i \neq j \leq d$. Then, all rows of $W$ corresponding to non-zero circularity coefficients are unique up to a sign. A row vector of $W$ corresponding to a zero circularity coefficient is unique up to multiplication by $e^{\theta_i}$, $\theta \in \mathbb{R}$.

U2 A circularity coefficient has multiplicity $m$, say $\lambda = \lambda_1 = \ldots = \lambda_m$. If $\lambda > 0$, then $W = (W_1, W_2)^T$, where $W_i = (w_{1i} \ldots w_{di}) \in \mathbb{C}^{d \times m}$ is unique up to right multiplication by a $m \times m$ real orthogonal matrix. If $\lambda = 0$, then $W_i$ is unique up to right multiplication by a $m \times m$ unitary matrix.

If $W$ is a GUT matrix, then so is $OW$ for any permutation matrix $O$, i.e. the rows of $W$ can be shuffled to an arbitrary order. Naturally, it is always possible to get rid of this permutation ambiguity by defining a “labeling convention”. One obvious way is to arrange circularity coefficients in decreasing order, $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_d \geq 0$, which implies an ordering for the corresponding rows of $W$.

Next we show that GUT functional possess a natural equivariance property under invertible linear transformations.

**Theorem 1.** GUT functional $W$ is equivariant in the sense that if $s = Az$ denotes non-singular linear transformation of $z \in \mathbb{C}^d$ for any non-singular $A \in \mathbb{C}^{d \times d}$, then the GUT functional of $s$ is $W(s) = W(z)A^{-1}$.

Next theorem states that if the spatial pseudo-scatter matrix $P$ defining the GUT matrix $W$ is a pseudo-scatter matrix (instead of being strictly spatial), then the GUT functional can be formulated as a simple eigenvector computation task.

**Theorem 2.** Assume that $P$ is a pseudo-scatter matrix. Then

$$|E(z)|E(z)|W(z)|^H = W(z)|^H \Lambda(s)^2,$$

where $E(z) = C(z)^{-1}P(z)$, i.e. $W(z)|^H$ and $\Lambda(s)^2$ correspond, respectively, to the matrix of eigenvectors and diagonal matrix of eigenvalues of $E(z)|E(z)^*$. Theorem 2 applies, for example, in the case of SUT. It provides an important insight that statistical properties (e.g. consistency, limiting distribution, statistical efficiency, robustness) of the estimator $W$ may be derived rather directly from those of $C$ and $P$.

An interesting remark can be made about a GUT matrix that is based on a weighted spatial pseudo-covariance matrix $P_{\phi}(z)$. We point out that $P_{\phi}(v) = P_{\phi}(C(v)$ since $C(v) = I$. In light of the step (b) of the GUT algorithm, this means that GUT matrix based on $C(z)$ and $P_{\phi}(z)$ is the same as the GUT matrix based on $C(z)$ and $P_{\phi}(z)$. By Theorem 2 this means that the GUT matrix based on $C(z)$ and $P_{\phi}(z)$ correspond to the matrix of eigenvectors of $E(z)|E(z)^*$, where $E(z) = C(z)^{-1}P_{\phi}(z)$.

3.3. Algorithms

The computation of step (a) of the GUT algorithm is straightforward as there exists many methods to compute the whitening matrix $B(z)$ (e.g. as the principal square root, or, as the inverse of the Cholesky factor of $C(z)$) but for our purposes we do not need to specify any particular root. It is the computation of step (b) that requires discussion.

First we recall that the circularity coefficients $\lambda_1, \ldots, \lambda_d$ are the singular values of $P(v)$, or equivalently, the square roots of the eigenvalues of $P(v)|P(v)^*$. (cf. Section 2.1).

Under Assumption U2 the calculation of the Takagi factor $U$ needs specialized numerical code which often is not available in commercial software packages. However, under Assumption U1 (i.e. distinct circularity coefficients), methods of Lemma 1 or 2 can be used to compute the Takagi factor $U(v)$. Both methods can be implemented with two lines of Matlab code.

If in addition to U1, we assume that $P$ is a pseudo-scatter matrix (instead of being strictly spatial), then Theorem 2 provides a very simple, elegant and straightforward method to compute the GUT matrix.

**Corollary 1.** If $P$ is a pseudo-scatter matrix and U1 holds, then $W(z) = LDV(z)|^H$, where $V(z) \in \mathbb{C}^{d \times d}$ contains the eigenvectors of $E(z)|E(z)^*$ as columns, $D = diag(1/\sqrt{d_i})$ with $d_i = |V(z)|^H C(z) V(z)|^H > 0$ and $L = diag(e^{\theta_i})$ with $\theta_i = -\frac{1}{2} \arg(|V(z)|^H P(z)V(z)|^H)$.

In the nut shell, the above corollary states that $V(z)|^H$ is equal to $W$ up to scaling and permutation of its row
vectors, i.e. $W(z) = V(z)^H$. If $P$ is a pseudo-scatter matrix, U1 holds and the scalings of the rows of $W$ are immaterial (as is the case in the ICA application, cf. Corollary 2), then identifying $W(z) = U_1(v)^H B(z)$ with $V(z)^H$ is computationally highly attractive approach since it avoids calculation of the whitening matrix $B(z)$ and the Takagi factor $U_1(v)$ of $P(v)$. Also, as a side product we obtain the circularity coefficients $\lambda_i$'s since these are the square roots of the eigenvalues of $E(z)E(z)^*$. The algorithm thus proceeds as follows:

**Gut algorithm 2.** Assume that $P$ is a pseudo-scatter matrix, Assumption U1 holds and scalings of the rows of $W$ are immaterial.

(a') Calculate $E(z) = C(z)^{-1} P(z)$. Matlab: $E = \text{inv}(Z^*Z') * (Z*Z')$. 
(b') Calculate the matrix $V(z)$ containing the eigenvectors of $E(z)E(z)^*$. Matlab: $[V d] = eigs(E * \text{conj}(E))$.
(c') Set $W = V^H$. Matlab: $W = V^*$. 

The given Matlab code is to compute the SUT (under Assumption U1). As we can see only three lines of Matlab code is required to compute the SUT (up to arbitrary scaling of its rows) in an computationally efficient way.

### 4. Application to ICA

#### 4.1. Complex-valued ICA model

We assume that the observed r.v. $z = (z_1, \ldots, z_d)^T$ follows complex-valued ICA model

$$z = As,$$

where r.v. $s = (s_1, \ldots, s_d)^T$ contain the mutually independent unobserved complex source signals (or ICs) and $A$ represents the unknown complex $d \times d$ invertible mixing matrix. Each source r.v. $s_i$ is assumed to be non-degenerate and at most one source can possess complex circular Gaussian density $f(s) = (2\pi)^{-d/2} e^{-|s|^2/2}$ (but sources can have non-identical non-circular Gaussian distribution). Since it is possible to identify $A$ only up to scaling, phase shift and permutation of its column vectors [6], ICA should be understood as the determination of a matrix $B$, called the separating matrix, that is essentially equal to $A^{-1}$, i.e. $B = A^{-1}$. Thus sources are said to be separated when $\hat{s} = Bz$ is a copy of $s$, i.e. $\hat{s} = Ps$ for any ambiguity matrix $P$.

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 $E[s] = E[z] = 0$ and A3 implies A2. 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. Assumption A5 is needed by the GUT matrix to separate the sources. A5 is essentially negation of A4.

### 4.2. GUT is a separating matrix

For purposes of ICA, the selected $C$ and $P$ need to fulfill the so called IC-property [14,22].

#### Definition 3. Scatter matrix $C$ and spatial pseudo-scatter matrix $P$ are said to possess IC-property if $C(s)$ and $P(s)$ are diagonal matrices when r.v. $s$ has ICs.

It is easy to verify that the covariance matrix, pseudo-covariance matrix and pseudo-kurtosis matrix, for example, possess IC-property. However, SPM, or $M$-estimator of scatter, do not generally possess IC-property. It is easy to show [14], however, that under Assumption A6 ($s_j = s_j$), any scatter matrix or spatial pseudo-scatter matrix possesses IC-property.

We wish to point out if the scatter or spatial pseudo-scatter matrix does not possess IC-property, the corresponding symmetrized estimator always possesses the IC-property [14,22]. Let $z_1$ denote $z_i$ i.i.d r.v.'s distributed as $z$. Then, symmetrized versions of $C$ and $P$, $C(z_1 - z_2)$ and $P(z_1 - z_2)$, are also a scatter and spatial pseudo-scatter matrix, but with different statistical properties than the original non-symmetrized matrices $C(z)$ and $P(z)$. As an example, symmetrized weighted spatial pseudo-covariance matrix is

$$\varphi_1(z_1 - z_2) = E[\varphi(\|z_1 - z_2\|^2)/(z_1 - z_2)(z_1 - z_2)^T],$$

where $z_1$ and $z_2$ are i.i.d r.v.'s. The respective finite sample estimator can be calculated from the sample $z_1, \ldots, z_n$ as

$$\frac{2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \varphi(\|z_i - z_j\|^2)/(z_i - z_j)(z_i - z_j)^T.$$

In a similar manner, one can construct a symmetrized version of the weighted covariance matrix or $M$-estimator of scatter (cf. Appendix A). Clearly, the drawback of the symmetrized version of an estimator is the increased computational load since symmetrization increase the number of “observations” by the power of two.

Our first task is to resolve the explicit form of Takagi’s factorization (4) of $P(v)$. 

<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 are circular</td>
</tr>
<tr>
<td>A5</td>
<td>$s_i$’s are symmetric</td>
</tr>
</tbody>
</table>

| Table 1 Some common additional assumptions on the distributions of the sources $s_i$ |


Theorem 3. Suppose \( z = As \) follows ICA model and assume that

A7 \( C \) and \( P \) possess IC-property and \( C(s) \) and \( P(s) \) exist.

Then, the Takagi factor \( U(V) \) in Takagi's factorization (4) is

\[
U(V) = B(z)A(C(s))^{1/2}L(s),
\]

where \( \tilde{s} = C(s)^{-1/2}s \) and \( L(\tilde{s}) = \text{diag}(e^{j\theta}) \), \( \theta_i = \frac{1}{2} \text{arg}(|P(s)|_{ii}) \).

Furthermore, circularity coefficients are \( \lambda_i = |P(s)|_{ii} \), \( i = 1, \ldots, d \).

Note that, under Assumption A7, GUT matrix \( W(z) = U(V)^H B(z) \) necessarily exists (since \( C(z) = AC(s)A^H = B(z)^H B(z)^{-1} \) and \( U(V) \) exists). The main message of Theorem 3 is that the unitary matrix \( U(V) \) can be factored (up to irrelevant right multiplication by a scaling matrix and a phase-shift matrix) to a product of the whitening matrix \( B(z) \) and the true mixing matrix \( A \).

Corollary 2. Suppose that \( z = As \) follows ICA model. Under Assumption A7 and U1, any matrix \( \tilde{W}(z) \) that is essentially equal to GUT matrix \( W(z) \), i.e. \( \tilde{W}(z) = W(z) \), is a separating matrix.

It is in place to discuss the meaning of Assumption U1 in the ICA model. First, Assumption U1 indicates that sources have non-identical distribution since i.i.d. sources necessarily possess identical circularity coefficients by Theorem 3 since \( |P(s)|_{ii} = |P(s)|_{ij} \) for \( i \neq j \). Second, observe that U1 also implies A5.

Example 1. Consider the SUT functional \( (C = \ell, P = \rho) \). Then A7 is equivalent with A2, and \( \lambda_i = |\rho(s)|_{ii} = |\tau(s_i)| / |\sigma(s_i)| \). Thus U1 holds if and only if the moduli of the pseudo-variances of the whitened (unit variance) sources \( s_i \) are distinct.

Example 2. Consider the case that \( C = \ell \) and \( P = \rho_{\text{kur}} \). Then A7 holds if the 4th-order moments \( E[|s_i|^2 s_i^2] \) of \( s_i \)'s exist, and

\[
\lambda_i = |\rho_{\text{kur}}(s)|_{ii} = |E[(s_i s_i^H s_i s_i^H)]_{ii}| = |E[|s_i|^2 s_i^2]| + (d - 1) \tau(s_i).
\]

Hence the \( i \)th circularity coefficient \( \lambda_i \) is the modulus of a weighted sum of a 4th-order and 2nd-order moment of the \( i \)th whitened source \( s_i \). Therefore, \( \lambda_i = \lambda_j \) for \( i \neq j \), if the whitened sources \( s_i \) and \( s_j \) possess identical 4th-order and 2nd-order moments. However, it is not very likely that circularity coefficients coincide otherwise. This also means that circularity coefficients are more likely to be distinct than in the case of SUT (provided that the 4th-order moments of the sources exist).

It is important to observe that the GUT algorithm contains a built-in warning: since the circularity coefficients \( \lambda_1, \ldots, \lambda_d \) are also necessarily extracted detection of two close circularity coefficients is an indication that the corresponding sources may not be reliably separated. Also, Assumption U1 is needed to separate all the sources: GUT matrix is not able to separate the sources that have identical circularity coefficients, but the rest of the sources are separated. This fact is more formally stated in the next corollary.

Corollary 3. Assume \( z = As \) follows ICA model and A7 hold. Decompose \( s = (s_1^T, s_2^T)^T \) so that each source \( s_i \) in \( s_1 \) (resp. \( s_2 \)) has circularity coefficient \( \lambda_i \) of multiplicity 1 (resp. \( > 1 \)). Decompose \( \tilde{s} = (\tilde{s}_1^T, \tilde{s}_2^T)^T = W(z)z \) accordingly. Then \( s_1 \) is a copy of \( s_i \).

Finally, we point out that the equivariance property of the GUT functional established in Theorem 1 is a highly desirable property since an affine equivariant separating matrix estimator yields estimated sources that do not depend on the mixing matrix \( A \) but only on the realization of the sources; see [23], Section II-C for detailed discussion. Note that JADE estimator lacks affine equivariance property, since the set of cumulant matrices cannot be exactly jointly diagonalized.

5. Performance studies of ICA

5.1. Simulation setups

Let \( \tilde{B} \) denote an estimator of the separating matrix. The performance of the separation is often investigated via “interference matrix” \( \tilde{G} = B \tilde{A} \). Due to fundamental indeterminacy of ICA, perfect separation implies that \( \tilde{G} = \mathbb{E} \) for some ambiguity matrix \( \mathbb{E} \). The quality of the separation is then assessed by calculating the widely used performance index (PI) [24]

\[
\text{Pl}(\tilde{G}) = \frac{1}{2d(d-1)} \left( \sum_{i=1}^{d} \left( \sum_{j=1}^{d} \frac{|\tilde{g}_{ij}|}{\max_{l} |\tilde{g}_{il}|} - 1 \right) \right),
\]

where \( \tilde{g}_{ij} = |G|_{ij} \). Under perfect separation \( \text{Pl}(\tilde{G}) = 0 \). When the estimator fails to separate the sources, the value of PI increases. PI is scaled so that the maximum value is 1. If the separating matrix estimator \( \tilde{B} \) is equivariant, as is the case for GUT matrix, then \( \tilde{G} \) (and thus PI) does not depend on the mixing matrix \( A \), and hence one could set \( A = 1 \) in the simulations without any loss of generality.

In our simulation studies, GUT estimators employing the following choices of scatter matrix \( C \) and spatial pseudo-scatter matrix \( P \) are used: gut1 employs covariance matrix and pseudo-covariance matrix (i.e. SUT), gut2 employs covariance matrix and pseudo-kurtosis matrix, gut3 employs covariance matrix and SPM, gut4 employs Huber’s M-estimator of scatter (cf. Appendix A) and SPM, gut5 employs Tyler’s M-estimator of scatter (cf. Appendix A) and SPM. We compare the results to the jade [19], fobi [13], complex FastICA with deflationary approach and contrast \( G_{(x)}(x) = \log(0.1 + x) \) [20] (denoted as fica), complex fixed point algorithm using kurtosis based contrast and symmetric orthogonalization [7] (denoted as cfpa), fixed-point kurtosis maximization algorithm [8] (denoted as kmf) and DOGMA estimator [14] employing Tyler’s M-estimator and Huber’s M-estimator as the choices of scatter matrices (denoted as d1).
Throughout, the number of sources is $d = 4$ and $m = 300$ data sets $S_n$ of the source signals $s$ were generated using sample lengths $n = 500, 10,000$. Each sample $S_n$ was mixed by a randomly generated mixing matrix $A$ yielding the data set $Z_n$ of mixtures and then PI was calculated for each estimator.

We use two routes to generate a source that has a non-circular distribution. In the first approach, real part $\text{Re}(s)$ and imaginary part $\text{Im}(s)$ of the source r.v.a. $s$ are independently generated from a symmetric location-scale distribution with location parameter (symmetry center) equal to zero but with a different value for the scale parameter. Examples of distributions from symmetric location-scale family are uniform, normal (i.e. Gaussian), logistic, Cauchy, Laplacian, etc. Note that the larger is the difference between the scale parameters, the more non-circular is the source r.v.a. $s$.

In the second approach, generated source $s$ will have dependent real and imaginary part. We start with an r.v.a. $u$ whose distribution (on the complex plane) is uniform on an ellipse (with center at the origin) with a semimajor axis $a \in [1/\sqrt{2}, 1]$, semiminor axis $b = \sqrt{1 - a^2}$, and orientation $\theta \in (-\pi/2, \pi/2)$. Such an ellipse can be said to have unit scale (since $a^2 + b^2 = 1$) with eccentricity $e = \sqrt{2a^2 - 1} \in [0, 1]$ (i.e. $a = \sqrt{1 + e^2}/2$). A circle is a special case ($a = 1/\sqrt{2}$) that has zero eccentricity, while as the ellipse becomes more elongated (i.e. when $a \rightarrow 1$) the eccentricity approaches one. The r.v.a. $u$ can be generated as follows:

$$ u = a \cos \theta + b \sin \theta, \quad \theta \sim \text{Unif}(-\pi, \pi), $$

where $a = \frac{1}{2}(a + \sqrt{1 - a^2}) e^{i \theta}$ and $b = \frac{1}{2}(a - \sqrt{1 - a^2}) e^{i \theta}$. We denote $u \sim \text{Ell}(e, \theta)$. A random sample of length $n = 100$ from Ell(0.8, $\pi$/4) is shown in Fig. 1(a). We point out that $u$ has variance $\sigma^2(u) = \frac{1}{2}$ and pseudo-variance $\tau(u) = \frac{1}{2} \alpha^2 e^{2i \theta}$. If $e = 1$ and $\theta \in (-\pi/2, \pi/2)$ is arbitrary, then $u \sim \text{Ell}(1, \theta)$ has a uniform distribution on the unit complex sphere. Source r.v.a. $s$ is generated from $u$ as

$s = ru$

where $r$ is any real-valued positive r.v.a. independent of $u$. The r.v.a. $s$ is then said to have standardized complex elliptical (CE) distribution with eccentricity $e$ and orientation $\theta$. Note that the distribution of $r$ determines the distribution of $s$. For example:

- if $r^2 \sim \chi^2_2$ ($r^2$ has a chi-squared distribution with 2 degrees of freedom), then $s$ has scaled complex normal (CN) distribution, denote $s \sim \text{CN}(e, \theta)$, with variance $\sigma^2(s) = 1$ and pseudo-variance $\tau(s) = \alpha^2 e^{2i \theta}$. A random sample of length $n = 100$ from $\text{CN}(0.8, \pi/4)$ is shown in Fig. 1(b).
- if $r^2 \sim F_{2,v}$ ($F$-distribution with 2 and $v$ degrees of freedom), then $s$ is said to have scaled complex $t_v$ distribution, denoted $s \sim \text{Ct}_v(e, \theta)$ with variance $\sigma^2(s) = \frac{v}{2(v-2)}$ and pseudo-variance $\tau(s) = \frac{v}{2(v-2)} \alpha^2 e^{2i \theta}$. If $v = 1$, then $s$ is said to have complex Cauchy distribution, denoted $\text{Cau}(e, \theta)$.
- if $r \sim \text{Exp}(1)$ (exponential distribution with unit rate), then $s$ is said to have scaled complex exponential

![Fig. 1. (a) Random sample of length $n = 100$ from Ell($e, \theta$) with eccentricity $e = 0.8$ and orientation $\theta = \pi/4$. The semimajor axis has length $a = \sqrt{(1 + e^2)/2} = 0.9055$. (b) Random sample of length $n = 100$ from CN($e, \theta$) with eccentricity $e = 0.8$ and orientation $\theta = \pi/4$.](image)
distribution, denoted as $s \sim \text{CExp}(\epsilon, \vartheta)$ having variance $\sigma^2(s) = 1$ and pseudo-variance $\tau(s) = \epsilon^2 e^{2\epsilon}$.

The distribution of the source r.v.a.'s $s_1, \ldots, s_4$ for each simulation setting A–F are explained below:

**A** $\text{Re}(s_k) \sim \text{Unif}(-1, 1)$ and $\text{Im}(s_k) \sim \text{Unif}(-k, -k)$ are independent ($k = 1, \ldots, 4$). The first source $s_1$ is 2nd-order circular, but other sources are non-circular.

**A’** The 1% of the data vectors in each sample $Z_n$ generated in Simulation A are replaced by an outlier generated as $z_{out} = (b_1 u_1 z_{1,\text{max}}, \ldots, b_4 u_4 z_{4,\text{max}})^T$, where $z_{i,\text{max}}$ is the element of the $i$th row of $Z_n$ with largest modulus, $u_i \sim \text{Unif}(1, 5)$ and $b_i$ is a r.v.a. with value $-1$ or 1 with equal probability $\frac{1}{2}$. Note that $z_{out} = b_1 u_1 z_{1,\text{max}}$ points to same or opposite direction as $z_{i,\text{max}}$, but its magnitude is at least as big but at most 5 times larger than that of $z_{i,\text{max}}$.

**B** $\text{Re}(s_k) \sim \text{Logist}(1)$ (logistic distribution with unit standard deviation) and $\text{Im}(s_k) \sim \text{Logist}(0, k + 1)$ are independent ($k = 1, \ldots, 4$).

**C** $\text{Re}(s_1) \sim \text{Logist}(1)$ and $\text{Im}(s_1) \sim \text{Logist}(2)$ are independent, $\text{Re}(s_2) \sim \text{Cau}(1)$ (Cauchy distribution with unit scale) and $\text{Im}(s_2) \sim \text{Cau}(2)$ are independent, $s_3 \sim \text{CN}(\epsilon, \vartheta)$ where $\epsilon$ and $\vartheta$ are randomly generated (for each sample) from $\text{Unif}(0, 1)$ and $\text{Unif}(-\pi/2, \pi/2)$, respectively, $s_4 \sim \text{Ell}(1, 0)$.

**D** The modulus and argument of the source $s_k$ are independently generated from the Rayleigh(1) and $\text{Unif}(-\pi \kappa/4, \pi \kappa/4)$, respectively ($k = 1, \ldots, 4$). The last source $s_4$ has circular complex normal distribution and other sources are relatively close to normal. Hence the data is expected to be relatively difficult to separate.

**E** $z_k \sim \text{CN}(\epsilon_k, \vartheta_k)$, where $\epsilon_k$ and $\vartheta_k$ are randomly generated (for each sample) from $\text{Unif}(0, 1)$ and $\text{Unif}(-\pi/2, \pi/2)$, respectively ($k = 1, \ldots, 4$).

**F** $z_k \sim \text{Ell}(1, \epsilon_k)$, $s_2 \sim \text{CN}(\epsilon_k, \vartheta_k)$, $s_3 \sim \text{Cau}(\epsilon_k, \vartheta_k)$, $s_4 \sim \text{CExp}(\epsilon_4, \vartheta_4)$, where $\epsilon_k$ and $\vartheta_k$ ($k = 1, \ldots, 4$) are randomly generated (for each sample) from $\text{Unif}(0, 1)$ and $\text{Unif}(-\pi/2, \pi/2)$, respectively.

Robustness of an estimator is an important issue in complex-valued ICA since at most one source can possess

<table>
<thead>
<tr>
<th>Table 2</th>
<th>The median values of $-10 \log_{10}[\text{Pl}(\mathcal{G})]$ for Settings A–F and sample lengths $n = 500, 2000, 10000$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A</strong></td>
<td></td>
</tr>
<tr>
<td>$n$</td>
<td>gut1</td>
</tr>
<tr>
<td>500</td>
<td>12.1</td>
</tr>
<tr>
<td>2000</td>
<td>15.0</td>
</tr>
<tr>
<td>$10^4$</td>
<td>18.5</td>
</tr>
<tr>
<td><strong>A’</strong></td>
<td></td>
</tr>
<tr>
<td>$n$</td>
<td></td>
</tr>
<tr>
<td>$n$</td>
<td></td>
</tr>
<tr>
<td><strong>B</strong></td>
<td></td>
</tr>
<tr>
<td>$n$</td>
<td></td>
</tr>
<tr>
<td><strong>C</strong></td>
<td></td>
</tr>
<tr>
<td>$n$</td>
<td></td>
</tr>
<tr>
<td><strong>D</strong></td>
<td></td>
</tr>
<tr>
<td>$n$</td>
<td></td>
</tr>
<tr>
<td><strong>E</strong></td>
<td></td>
</tr>
<tr>
<td>$n$</td>
<td></td>
</tr>
<tr>
<td><strong>F</strong></td>
<td></td>
</tr>
<tr>
<td>$n$</td>
<td></td>
</tr>
</tbody>
</table>

The highest (resp. second highest) value obtained in each simulation run is written in bold face (resp. underlined).
conventional circular Gaussian distribution. Thus a separating matrix estimator should work reliably for all types of source distributions. The distribution of the sources in the setups satisfy A6, and hence all of the selected scatter and spatial pseudo-scatter estimators possess IC-property.

5.2. Results

The results for Settings A–F with sample lengths \( n = 500, 2000, 10,000 \) are given in Table 2.

In the case of Setting A, \( \text{fobi}, d1 \) and \( \text{fica} \) are performing poorly. This may be due to the fact that these methods do not exploit any non-circularity features of the signals. The best performance is attained by \( \text{jade} \) and \( \text{cfpa} \), but \( \text{kmf} \) do not fall far behind them. All the GUT estimators are performing very similarly and fairly well: \( \text{gut}1 \) (i.e. SUT) has slightly better performance than the others, but \( \text{gut}5 \) and \( \text{gut}2 \) are very close to it.

In the case of outliers (Setting A), only the GUT estimators \( \text{gut}4 \) and \( \text{gut}5 \) that employ robust scatter matrix together with robust spatial pseudo-scatter matrix are not affected by outliers and are shown to perform very well. Interestingly, \( \text{gut}3 \) although it uses the conventional non-robust covariance matrix for whitening but robust (in contrast to \( \text{gut}2 \) spatial pseudo-scatter matrix is performing decently and considerably better than \( \text{gut}2 \). Other estimators except the robust \( d1 \), however, fail completely due to the effect of outliers.

In the case of Setting B, the GUT estimators \( \text{gut}1 \) and \( \text{gut}5 \) obtain the best performance whereas \( \text{fobi} \) and \( d1 \) are performing rather poorly. It is interesting to note that \( \text{fica, cfpa, kmf} \) and \( \text{jade} \) have rather comparable performance. Among GUT estimators, \( \text{gut}2 \) stands out being far behind the others.

In the case of Setting C, \( \text{fica} \) attains the best performance at all sample lengths, but the second best performance is obtained by different algorithms at different sample lengths: here \( \text{gut}2 \) is working very well for small sample lengths, but \( \text{gut}5 \) and \( d1 \) outperform it when the sample length increases. \( \text{jade} \) is working very poorly, but also \( \text{fobi} \) and \( \text{gut}2 \) have rather poor performance.

In the case of Setting D, \( \text{gut}2 \) has the best performance and \( \text{gut}1 \) is just a bit behind it. All the GUT estimators are outperforming the other estimators, although their performance is quite unsatisfactory as well.

In the case of Setting E, \( \text{gut}1 \) and \( \text{gut}5 \) attain the best performance whereas \( \text{jade, cfpa} \) and \( \text{kmf} \) are totally failing to estimate the sources. \( \text{fica, fobi} \) and \( d1 \) have similar rather unsatisfactory performance. All of the GUT estimators show good performance.

In the case of Setting F, \( \text{fica} \) and \( d1 \) attain the best performance, others being rather far behind them.

Fig. 2. Estimated source signal constellations obtained by \( \text{jade} \) (first row) and GUT method \( \text{gut}5 \) (second row). BPSK, 8-QAM and circular Gaussian source signals are clearly discernible on the left, middle and right plots.
The robust estimators gut4 and gut5 show much better performance than the remaining group of GUT estimators.

6. Communications example

Three independent random signals—a BPSK signal, an 8-QAM signal and a circular Gaussian signal of equal power $\sigma_s^2$ are impinging on $d = 3$ element uniform linear array (ULA) with half a wavelength interelement spacing from direction-of-arrivals (DOAs) $-20^\circ$, $5^\circ$ and $35^\circ$. Note that BPSK and 8-QAM signals are 2nd-order non-circular. The above random communications signals are symmetric, and hence any pair of scatter and spatial pseudo-scatter matrices can be employed in the definition of the GUT matrix. The array outputs are corrupted by additive noise $n$ with a circular complex Gaussian distribution with covariance matrix $\mathbf{C}(n) = \sigma_n^2 \mathbf{I}$. The signal to noise ratio (SNR) is $10 \log_{10}(\sigma_s^2 / \sigma_n^2) = 20$ dB and the number of snapshots is $n = 2000$. The estimated source signal constellations obtained by jade and gut5 are shown in Fig. 2. Both of the methods were able to separate the sources as BPSK, 8-QAM and circular Gaussian source signals are clearly discernible. Table 3 shows the average performance of all the ICA estimators over 100 simulation runs. jade and kmf are performing the best but GUT methods are not far behind. fobi and d1, however, stand out as they do not quite reach the same level of performance as the others.

To illustrate the reliable performance of the robust GUT matrices under contamination, four observations were replaced by an outlier $z_{out}$ generated as in Simulation A. Hence only 0.2% of the data is contaminated. Fig. 3 depicts the estimated source signal constellations obtained by jade and gut5. As can be seen, only gut5 is able to separate the sources and is unaffected by outliers, which are clearly detected in the plots. jade on the other hand fails completely: BPSK and 8-QAM sources are no longer

| Table 3 |
| Mean values of $-10 \log_{10}(\text{Pl}(\hat{G}))$ computed from 100 array snapshot data sets without outliers (first row) and with outliers (second row) |
| gut1 | gut2 | gut3 | gut4 | gut5 | fica | jade | cfpa | kmf | d1 | fobi |
| 17.5 | 15.2 | 16.7 | 16.7 | 16.9 | 16.5 | 18.7 | 16.8 | 18.4 | 12.2 | 12.7 |
| 8.5  | 2.7  | 11.4 | 16.6 | 16.8 | 3.6  | 2.5  | 2.6  | 2.4  | 12.2 | 2.6  |

![Fig. 3. Estimated source signal constellations obtained by jade (first row) and GUT method gut5 (second row) for the data with four outliers. Signals that correspond to outliers are marked by triangle. Robust GUT method gut5 is unaffected by outliers, which are clearly detected in the plots whereas jade fails completely: outliers are able to destroy the patterns of BPSK and 8-QAM signals.](image)
discernible in the plots. The reliable performance of the robust GUT methods is evident in Table 3: only the robust GUT methods gut4 and gut5 and the robust DOGMA estimator d1 are able to separate the sources.

Table 4 gives average computation times. As can be seen gut1 (i.e. the SUT) and fobi are the fastest to compute whereas fica is the slowest. To compute fobi we used the fast algorithm of [14]. Also observe that occurrence of outliers severely increases the computation times of the iterative fixed-point algorithms fica, kmf and cfpa, whereas computation times for the other methods are only slightly affected by outliers.

7. Conclusions

In this paper, the elegant SUT method by Eriksson and Koivunen [6,10] and De Lathauwer and De Moor [5] was generalized. It was shown that depending on the selected scatter and spatial pseudo-scatter matrix, the associated GUT estimator can have largely different statistical (e.g. robustness) properties. In the context of complex-valued ICA, we showed that GUT matrix is a separating matrix provided that U1 holds. Special emphasis was put on robust GUT matrices. As our simulations demonstrated, GUT matrices (gut4 and gut5) that employed a pair of robust scatter and spatial pseudo-scatter matrix were unaffected by outliers whereas the other estimators except the robust DOGMA estimator d1 completely failed to separate the sources. It should be emphasized that GUT provides computationally attractive class of estimators (as Table 4 also demonstrates) for ICA since they avoid difficult optimization tasks (common to most ICA algorithms) and are solely based on straightforward matrix computations.

We also wish to point out that SUT is currently used as a prewhitening step in some ICA algorithms (e.g. in [7]). Instead of employing the SUT, perhaps more appropriate choice is to employ some robust GUT matrix in the prewhitening step in order to achieve robustness in the face of outliers in the sources.

Some estimators of scatter and spatial pseudo-scatter matrix do not possess IC-property and hence they cannot be used in ICA if sources do not have symmetric distributions. Symmetrized version of the estimator, however, possess IC-property, but is often computationally too expensive since it is calculated from the sample (of length \( n \)) of pairwise differences \( z_i - z_j \), \( i, j \in \{1, \ldots, n\}, i < j \). To relieve the computational burden, one possibility is to compute the symmetrized estimator from a much smaller subset of randomly selected pairwise differences instead of the whole sample of length \( n \).

These approaches will be a subject of a separate paper.

Acknowledgment

The authors would like to thank Hualiang Li for kindly providing us his Matlab program to compute the kmf method [8]. E. Ollila would like to thank the Academy of Finland for supporting this research.

Appendix A. M-estimators of scatter

M-estimators of multivariate scatter were first introduced in [25] for real data and later generalized in [16–18] for complex data. The complex M-functional of scatter \( \mathbf{C}_\varphi \in \text{PDH}(d) \) is defined as the solution of an implicit equation

\[
\mathbf{C}_\varphi(\mathbf{z}) = E[\varphi(\mathbf{z}^H \mathbf{C}_\varphi(\mathbf{z})^{-1} \mathbf{z}) \mathbf{z}^H],
\]

where \( \varphi(\cdot) \) is any real-valued weighting function on \([0, \infty)\). The weighted covariance matrix can be described as “1-step M-estimator”. Robustness of M-functional of scatter depends on the chosen weight function. Robust “phi-function” (a function \( \varphi(x) \) that descends to zero for large \( x \)) gives a robustified version of the covariance and pseudo-covariance matrix. Examples of robust weight functions are given at the end of the section.

M-estimators of scatter do not necessarily possess IC-property. Again it is possible to construct a symmetrized version that automatically has the IC-property. The symmetrized M-estimator of scatter \( \mathbf{C}_\varphi \in \text{PDH}(d) \) is defined as the solution of

\[
\mathbf{C}_\varphi(\mathbf{z}_1 - \mathbf{z}_2) = E[\varphi(\mathbf{z}_1 - \mathbf{z}_2)^H \mathbf{C}_\varphi(\mathbf{z}_1 - \mathbf{z}_2)^{-1} (\mathbf{z}_1 - \mathbf{z}_2)]\mathbf{z}_1 - \mathbf{z}_2\mathbf{z}_1^H],
\]

where \( \mathbf{z}_1 \) and \( \mathbf{z}_2 \) are i.i.d. r.v.’s distributed as \( \mathbf{z} \).

Some prominent robust weight functions are given next. Tyler’s M-estimator of scatter [16,26] utilizes weight function \( \varphi(x) = d/x \) and the Huber’s M-estimator of scatter utilizes weight function

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

where \( c \) is a tuning constant defined so that \( q = F_{c^2}(2c^2) \) for a chosen \( q \) and the scaling factor \( \beta = F_{2c^2+1}(2c^2)+c^2(1-q)/d \), where \( F_{c^2} \) denotes the c.d.f. of chi-squared distribution with \( d \) degrees of freedom. For symmetrized Huber’s M-estimator, tuning constant \( c \) is defined via \( q = F_{c^2}(2c^2) \) for a chosen \( q \) and the scaling factor as
\[ \beta = 2F_{\epsilon_{2\theta}}(C^2) + C^2(1 - q)/d. \]

In our simulations, we use \( q = 0.9 \). As in the real case, a simple iterative algorithm can be used to compute the above \( M \)-estimators [16].

Appendix B. Proofs

Proof of Lemma 1. The distinctness of singular values and symmetry \( A^T = A \), implies that \( G^T V^T = L \), where \( L = \text{diag}(\epsilon_i) \), \( \theta_i \in \mathbb{R} \) is a phase-shift matrix. This implies that the SVD of \( A \) can be written in a form

\[ A = \text{GALG}^T = GL^{1/2}AL^{1/2}G^T. \]

If we now write \( U = GL^{1/2} \), we observe that \( A = UAU^T \) is the Takagi factorization of \( A \) since \( U \) is unitary.

Proof of Theorem 1. Since \( C(s)^{-1} = A^{-1}LC(s)^{-1}A^{-1} \) by equivariance property of scatter matrix functional, and since \( C(s)^{-1} = B(s)A^{-1}B(s) \), a square-root matrix of \( C(s)^{-1} \) is \( B(s) = B(s)A^{-1} \). Thus \( W(s) = U(B(s)A^{-1})B(s) = U(B(s)A^{-1} Az)^iB(s)A^{-1} = W(s)A^{-1}. \]

Proof of Theorem 2. Since a pseudo-scatter matrix \( P \) is equivariant under invertible linear transformations, we have that \( P(v) = B(z)P(z)B(z)^T \). Then, by recalling Takagi’s factorization \( P(v) = U(v)A(s)U(v)^T \), we may write

\[ B(z)P(z)B(z)^T = U(v)A(s)U(v)^T. \]

Multiplying the above identity by \( B(z)^H \) from the left and by \( B(z)^{-1} \) from the right and recalling that \( B(z)^H B(z) = C^{-1}(z) \) and \( W(z) = U(z)B(z)^H \) gives the equation

\[ C(z)^{-1}P(z) = W(z)^H A(s)W(z)^{-1} \]

Denoting \( E(z) = C(z)^{-1}P(z) \), yields \( E(z)E(z)^* = W(z)^H A(s)^2 (W(z)^{-1})^H \). Hence \( E(z)E(z)^*W(z)^H = W(z)^H A(s)^2 \), that is, \( W(z)^H \) is the matrix of eigenvectors and \( A(s)^2 = \text{diag}(\lambda_1^2, \ldots, \lambda_m^2) \) is the diagonal matrix of eigenvalues of \( E(z)E(z)^* \).

Proof of Corollary 1. Since \( \hat{\lambda}_i^2 \)'s are the eigenvalues of \( E(z)E(z)^* \) by Theorem 2, Assumption U1 is equivalent with the statement that eigenvalues of \( E(z)E(z)^* \) are distinct. Hence, under U1, \( V(z) \) is unique up to a phase shift and scaling of its columns, and hence \( V(z)^H \) identifies \( W(z) \) up to left-multiplication by some scaling matrix \( D \) and phase-shift matrix \( L \), i.e., \( W(z) = LDV(z)^H \), or equivalently, \( V(z)^H = D^{-1}L^*W(z)^H \). Write \( \tilde{s} = V(z)^*z = D^{-1}L^*W(z)^*z = D^{-1}L^*s \). Then, using equivariance property of \( C \) and \( P \) and that \( C(s) = I \) and \( P(s) = A(s) \) (by definition of GUT), we get

\[ C(s) = D^{-1}L^*C(s)(D^{-1}L^*)^H = D^{-2}, \]
\[ P(s) = D^{-1}L^*P(s)(D^{-1}L^*)^T = (A(s)D^{-2})(L^*)^2. \]

Also note that \( C(s) = V(z)^H C(s)V(z) \) and \( P(s) = V(z)^H P(z)V(z)^* \) since \( \tilde{s} = V(z)^*z \) is an invertible linear transformation of \( z \). This shows that \( D = [V(z)^H C(s)V(z)]^{-1/2} \) and \( \Lambda = \text{diag}(\epsilon_i^2) \) with \( \theta_i = -\frac{1}{2} \text{arg}([V(z)^H P(z)V(z)]_{ii}). \]

Proof of Theorem 3. The whitened mixture \( \nu = B(z)z \) follows ICA model \( \nu = \Lambda s \), where \( \Lambda = B(z)A \) is the new mixing matrix. Then, from

\[ I = C(v) = C(\Lambda s) = \Lambda C(s)^H \]
\[ = [\Lambda C(s)^1/2][\Lambda C(s)^1/2]^H, \]

we observe that \( \Lambda C(s)^{1/2} \) is a unitary matrix. Since \( P(v) \) is unitary equivariant and possess IC-property, it follows that

\[ P(v) = P(\Lambda s) = P(\Lambda C(s)^{1/2} C(s)^{1/2} s) \]
\[ = [\Lambda C(s)^{1/2} P(s)]A(s)^{1/2} C(s)^{1/2} C(s)^{1/2} s, \]

where we wrote \( s = C(s)^{-1/2} s \) for the whitened source vector. Since \( P(s) = L(s)^2 D(s) \), where \( L(s) = \text{diag}(\epsilon_i^2) \) with \( \theta_i = \frac{1}{2} \text{arg}([P(s)]_{ii}) \) and \( D(s) = \text{diag}([P(s)]_{ii}) \), we may write

\[ P(v) = [\Lambda C(s)^{1/2} L(s)] D(s) [\Lambda C(s)^{1/2} L(s)]^H \]

(B.1)

Since \( \Lambda C(s)^{1/2} L(s) \) is a unitary matrix and \( D(s) \) is a real non-negative diagonal matrix, it follows that (B.1) is the Takagi’s factorization (4), namely, \( U(v) = \Lambda C(s)^{1/2} L(s) \) and \( A(s) = D(s). \)

Proof of Corollary 2. By Theorem 3, \( U(v)^H B(z) = L(s)^* C(s)^{-1/2} A^{-1} \), where \( L(s) \) is a phase-shift matrix and \( C(s)^{-1/2} \) is a scaling matrix. This means that under U1, \( W(z) = U(v)^H B(z) \) is essentially equal to \( A^{-1} \), i.e. it is a separating matrix. Hence any matrix \( W_0(z) \) (say) that is essentially equal to \( W(z) \) is also a separating matrix.

Proof of Corollary 3. Suppose (for simplicity of presentation) that a single circularity coefficient has multiplicity \( m \), say \( \lambda = \lambda_1 = \cdots = \lambda_m \) but remaining coefficients are distinct, so that \( s_1 \in \mathbb{C}^m \) and \( s_2 \in \mathbb{C}^{d-m} \). Decompose \( W(z) \) and \( A^{-1} \) accordingly:

\[ W(z) = \begin{pmatrix} W_1^T \\ W_2^T \end{pmatrix} \quad \text{and} \quad A^{-1} = \begin{pmatrix} A_1^{-1} \\ A_2^{-1} \end{pmatrix}. \]

If \( \lambda = 0 \), then by U1 and U2, \( W_1 = (w_1 \ldots w_m) \in \mathbb{C}^{d-m} \) is unique up to right multiplication by \( m \times m \) unitary matrix and \( W_2 \) is unique up to right multiplication by a real diagonal matrix with \( \pm 1 \) as diagonal elements. If \( \lambda > 0 \), then by U1 and U2, \( W_1 \) is unique up to right multiplication by an \( m \times m \) real orthogonal matrix and \( W_2 \) is unique up to right multiplication by a diagonal matrix having \( \pm 1 \) (resp. \( e^0, \beta \in \mathbb{R} \)) as its diagonal element if ith row of \( W_2 \) correspond to non-zero (resp. zero) coefficient. Since \( W(z) = L(s)^* C(s)^{-1/2} A^{-1} \) by Theorem 3, we must have that \( W_2 \) is essentially equal to \( A_2^{-1} \), i.e. \( s_2 \) is a copy of \( s_2 \).

References


