Uniqueness Analysis of Non-Unitary Matrix Joint Diagonalization [1]

Martin Kleinsteuber and Hao Shen

November 8, 2018

Abstract

Matrix Joint Diagonalization (MJD) is a powerful approach for solving the Blind Source Separation (BSS) problem. It relies on the construction of matrices which are diagonalized by the unknown demixing matrix. Their joint diagonalizer serves as a correct estimate of this demixing matrix only if it is uniquely determined. Thus, a critical question is under what conditions a joint diagonalizer is unique. In the present work we fully answer this question about the identifiability of MJD based BSS approaches and provide a general result on uniqueness conditions of matrix joint diagonalization. It unifies all existing results which exploit the concepts of *non-circularity*, *non-stationarity*, *non-whiteness*, and *non-Gaussianity*. As a corollary, we propose a solution for complex BSS, which can be formulated in a closed form in terms of an eigenvalue and a singular value decomposition of two matrices.

Index Terms

Non-unitary joint diagonalization, uniqueness analysis, Complex Blind Source Separation (BSS), Second-Order Statistics (SOS), Higher-Order Statistics (HOS).

1 Introduction

Joint diagonalization of a set of square matrices is a prominent algorithmic paradigm for solving the Blind Source Separation (BSS) problem [1]. One critical task of the Matrix Joint Diagonalization (MJD) approach is to construct a set of square matrices, so that there exists a unique joint diagonalizer, which serves as a correct demixing matrix of the BSS problem. Although uniqueness and solvability conditions of the BSS problem are well known in the framework of Independent Component Analysis (ICA), cf. [2], practical identifiability of MJD based BSS approaches has not been systematically studied yet and is investigated thoroughly in this work.

By imposing the assumption that source signals are mutually statistically independent, it is well known that source signals can only be extracted up to arbitrary scaling and permutation. The statistical independence assumption leads to the celebrated technique of Independent Component Analysis (ICA) [2]. One fundamental question is: *Under what conditions on the sources can the mixing process be uniquely identified up to permutation and scaling?* General

[1]M. Kleinsteuber and H. Shen are with the Department of Electrical Engineering and Information Technology, Technische Universität München, München, Germany. e-mail: (see http://www.gol.ei.tum.de).

Authors are listed in alphabetical order due to equal contribution.

identifiability results of the ICA problem have been developed based on either the Darmois-Skitovitch theorem, cf. [2, 3], or diagonalization of the Hessian of the characteristic function, cf. [4]. They provide a theoretic ground for developing ICA algorithms that minimize the so-called *contrast functions*, cf. [5, 6, 7]. Many popular ICA contrast functions originate from information theory, such as the mutual information [8] or the differential entropy [9]. Unfortunately, performance of these contrast function based approaches depends significantly on a correct estimation of the distribution of the sources, which is often an infeasible undertaking in application scenarios. Although there have been alternative non-parametric approaches developed to cope with this difficulty, cf. [10, 11], these methods go along with a high computational burden.

One simple approach to overcome the aforementioned difficulty is to utilize some additional properties of the sources for the separation task. Although the standard ICA model does not make any assumption on the temporal structure of the sources, temporal information is richly available in many real applications, and has been extensively exploited in developing efficient ICA algorithms, cf. [12, 13]. Specifically, these approaches utilize only selected second-order statistics (SOS) or higher-order statistics (HOS) of the observations, and often result in a form of a joint diagonalization problem of a finite number of matrices. These matrix joint diagonalization based methods are known as the *tensorial* BSS approaches. In parallel to the general ICA theory, the present work aims to answer the following critical question: *Under what conditions on the matrices, which are constructed for joint diagonalization, can we identify the mixing process uniquely up to permutation and scaling?*

Early works on Matrix Joint Diagonalization in BSS are restricted to unitary transformations, as a whitening process on the observed mixtures is often used as a preprocessing step, cf. [14]. However, it has been shown in [15] that linear BSS via a Unitary Joint Diagonalization (UJD) approach may have a serious limit of degraded performance in the presence of additive noise. In particular, many criterions for joint diagonalization can be significantly distorted by the whitening process, cf. [16, 17]. To avoid such a limit of UJD, a natural generalization of UJD, known as Non-Unitary Joint Diagonalization (NUJD), has been proposed and studied with dramatically increasing attention, cf. [18, 19, 20].

Non-stationarity is a common property, which describes the temporal structure of a signal. One simple assumption that can be employed for BSS in this context is that the covariance matrix of the sources varies over time. By exploiting this property, the signals can be separated via a joint diagonalization of a finite set of covariance matrices within different time intervals [21]. Identifiability conditions of this approach are developed in [14] for the UJD case, and in [22] under a limited NUJD setting, where only the real valued ICA problems are considered. Similar approaches employ also cyclo-stationarity of the sources [23, 24], or time-frequency distributions at different time frequency points [25].

Another simple temporal concept used in BSS is the non-whiteness of sources. Pioneering works in [26, 27] show that real valued source signals with distinct spectral density functions are blindly identifiable by using only the autocorrelation of the observations. Similar results in [28] show that stationary colored complex signals can be blindly separated by using a set of autocorrelation matrices. When source signals are both stationary and white, it requires more knowledge about the signals, such as higher-order statistics [14]. In particular, third- and fourth-order cumulants have been used and demonstrate their success in solving BSS problems, cf. [29, 30, 31]. In practice, higher order statistics are often rearranged in matrix form, so that the matrix joint diagonalization approach is applicable.

All the aforementioned statistical properties can be used for separating both real- and complex-valued signals. If the sources are second-order non-circular and the values of the circularity coefficients are distinct, complex BSS can be solved effectively by a joint diagonalization of only one covariance matrix and one pseudo-covariance matrix, cf. [32]. The corresponding method is known as Strong Uncorrelating Transform (SUT) [33]. Unfortunately, a solution given by SUT does not in general yield a correct demixing of the sources in real applications, where noise is commonly present. Recently, generalized SUT approaches have been proposed independently in [34] and [35] to jointly diagonalize both covariance and pseudo-covariance matrices. In particular, the work in [35] demonstrates that in the presence of noise, this generalized approach outperforms the state-of-the-art MJD approaches in terms of recovery quality.

To summarize, rich literature is available in the community on developing the matrix joint diagonalization based BSS methods. Existing identifiability results are mainly focused on the SOS based approaches. However, identifiability analysis for the HOS methods has not been addressed systematically. In this work, we derive the uniqueness conditions of the NUJD setting. It leads us to the most general result so far on identifiability conditions for the HOS based BSS methods, and an *algebraic solution*, i.e. a solution that only involves Eigenvalue Decompositions (EVD) or Singular Value Decompositions (SVD). Furthermore, it also provides a rigorous analysis on the convergence properties of existing iterative algorithms [36]. This is due to the fact that isolated critical points of functions can be identified which measure the degree of joint diagonality. This issue is not discussed further in this paper and is subject matter of ongoing work of the authors.

The paper is organized as follows. Section 2 gives a setting of the complex BSS problem and motivates the non-unitray joint diagonalization approach as a solution to BSS. In Section 3 we derive necessary and sufficient conditions for the uniqueness of non-unitary joint diagonalization. In Section 4 this uniqueness result is used to analyze the identifiability of tensorial BSS methods and to propose a new algebraic solution which generalizes the SUT approach and is able to separate non-circular signals with indistinct circularity coefficients.

2 Complex BSS and Matrix Joint Diagonalization

In this section we review the complex linear BSS problem to make this work self-contained, together with several second- and higher-order statistics based BSS approaches. Thereafter, we introduce a non-unitary joint diagonalization approach which is general enough to unify all the existing approaches in the literature.

2.1 Notations

We denote by $(\cdot)^{\mathsf{T}}$ the matrix transpose, by $(\cdot)^{\mathsf{H}}$ the Hermitian transpose, and by $(\cdot)^*$ the (entry-wise) complex conjugate. Furthermore, $|z| = \sqrt{zz^*}$, $\Re z$ and $\Im z$

denotes the modulus, the real part and the imaginary part of $z \in \mathbb{C}$, respectively. The complex unit is denoted by $\mathrm{i} := \sqrt{-1}$. $(\cdot)^\dagger$ stands optionally for either the matrix transpose or the Hermitian transpose. Matrices are denoted with capital Roman and Greek letters, e.g. A, Ω. Vectors are in lower case bold face, e.g. $\boldsymbol{s}, \boldsymbol{\omega}$. The expectation value of a random variable is denoted by $\mathbb{E}[\cdot]$.

By $Gl(m)$ we denote the set of all invertible $(m \times m)$-matrices. I_m is the $(m \times m)$-identity matrix, and the sets of all unitary and real orthogonal $(m \times m)$-matrices are defined as

$$U(m) := \{X \in Gl(m) | X^\mathsf{H} X = I_m\} \quad \text{and} \tag{1}$$

$$O(m) := U(m) \cap \mathbb{R}^{m \times m}, \tag{2}$$

respectively. The set of all complex orthogonal $(m \times m)$-matrices is given by

$$O(m, \mathbb{C}) := \{X \in Gl(m) | X^\mathsf{T} X = I_m\}. \tag{3}$$

For $C \in \mathbb{C}^{m \times m}$ and $X \in Gl(m)$, we define the linear transformations

$$C \mapsto X C X^\mathsf{T}, \tag{4a}$$

$$C \mapsto X C X^\mathsf{H}, \tag{4b}$$

as the transpose congruence transform and Hermitian congruence transform, respectively. Finally, we denote by $\oplus$ the exclusive disjunction operator.

2.2 Properties of Complex Signals

In this work we model a complex signal $s(t) = x(t) + \mathrm{i}y(t)$ as a complex stochastic process indexed by the variable t with real $x(t)$ and $y(t)$. Let $[s(t_1), \ldots, s(t_n)]^\mathsf{T}$ be an n-dimensional induced random vector of the signal $s(t)$.

2.2.1 Stationarity

A signal $s(t)$ is said to be *completely stationary* if the joint probability distribution of $[s(t_1), \ldots, s(t_n)]^\mathsf{T}$ is identical to the joint probability distribution of $[s(t_1 - \tau), \ldots, s(t_n - \tau)]^\mathsf{T}$ for any n, cf. [37]. A real signal $x(t)$ is said to be *weakly stationary*, if the following holds:

(i) $\mathbb{E}[x(t)] = \mathbb{E}[x(t + \tau)]$ for all $\tau \in \mathbb{R}$ and

(ii) $\mathbb{E}[x(t_1)x(t_2)] = \mathbb{E}[x(t_1 + \tau)x(t_2 + \tau)]$.

The first property states that the mean of the signal is constant, and the second that the correlation only depends on the time difference $t_1 - t_2$.

2.2.2 Circularity

A complex signal $s(t) = x(t) + \mathrm{i}y(t)$ is said to be *(weakly) circular*, if $s(t)$ and $\mathrm{e}^{\mathrm{i}\alpha} s(t)$ have the same probability distribution. The circularity assumption implies $\mathbb{E}[s(t)^2] = \mathrm{e}^{2\mathrm{i}\alpha}\mathbb{E}[s(t)^2]$ for all α, i.e. $\mathbb{E}[s(t)^2] = 0$. Given a complex signal $s(t)$ with a bounded variance, i.e. $\mathbb{E}[|s(t)|^2] < \infty$, the following quantity

$$\lambda_{s(t)} := \frac{|\mathbb{E}[s(t)^2]|}{\mathbb{E}[|s(t)|^2]} \tag{5}$$

is referred to as the *circularity coefficient* of $s(t)$. The definition of circularity can be extended straightforwardly to the case of multiple signals. Let $\boldsymbol{s}(t) \in \mathbb{C}^m$ be a vector consisting of m signals. Then $\boldsymbol{s}(t)$ is circular if $\boldsymbol{s}(t)$ and $e^{i\alpha}\boldsymbol{s}(t)$ have the same probability distribution.

A signal $s(t)$ is said to be *completely circular* if the induced random vector $[s(t_1), \ldots, s(t_n)]^\mathsf{T}$ is circular for all $n \in \mathbb{N}$. A signal is *circular of order* n if the induced vectors of order lower or equal to n are circular, cf. [38].

2.2.3 Whiteness

A real signal $x(t)$ is said to be *white* if

(i) $\mathbb{E}[x(t)] = 0$;

(ii) $\mathbb{E}[x(t_1)x(t_2)] = c\delta(t_1 - t_2)$,

where δ is the *Kronecker delta* function and c some positive constant. We refer to [39] for generalization of the concept of whiteness to higher order.

2.3 Complex Linear BSS Model

Let $\boldsymbol{s}(t) = [s_1(t), \ldots, s_m(t)]^\mathsf{T}$ be an m-dimensional mutually statistically independent complex signal. The noise-free instantaneous linear complex BSS model is given by

$$\boldsymbol{w}(t) = A\boldsymbol{s}(t), \tag{6}$$

where $A \in Gl(m)$ is the mixing matrix and $\boldsymbol{w}(t) = [w_1(t), \ldots, w_m(t)]^\mathsf{T}$ presents m observed linear mixtures of $\boldsymbol{s}(t)$. Without loss of generality, we assume that the sources $\boldsymbol{s}(t)$ have zero mean, i.e. $\mathbb{E}[\boldsymbol{s}(t)] = 0$, cf. [1].

The task of the linear complex BSS problem (6) is to recover the source signals $\boldsymbol{s}(t)$ by estimating the mixing matrix A or its inverse A^{-1} only based on the observations $\boldsymbol{w}(t)$ via the demixing model

$$\boldsymbol{y}(t) = X^\mathsf{H}\boldsymbol{w}(t), \tag{7}$$

where $X^\mathsf{H} \in Gl(m)$ is the demixing matrix, which is an estimation of A^{-1}, and $\boldsymbol{y}(t)$ represents the corresponding extracted signals. The statistical independence assumption provides various statistical properties of sources to identify the demixing matrix. The widely used properties include *non-circularity*, *non-stationarity*, *non-whiteness*, and *non-Gaussianity*.

2.4 Second-Order Statistics Based ICA Approaches

In this subsection, we briefly review the second-order statistics based ICA approaches and motivate our general approach of joint diagonalization.

Given the mixing model (6), the covariance matrix of the observations $\boldsymbol{w}(t)$ is computed as

$$C_{\boldsymbol{w}}(t) := \mathbb{E}[\boldsymbol{w}(t)\boldsymbol{w}^\mathsf{H}(t)] = A\underbrace{\mathbb{E}[\boldsymbol{s}(t)\boldsymbol{s}^\mathsf{H}(t)]}_{=:C_{\boldsymbol{s}}(t)}A^\mathsf{H} \tag{8}$$

where the covariance matrix of the sources $C_{\boldsymbol{s}}(t)$ is diagonal and non-negative following the statistical independence assumption. When the source signals are assumed to be non-stationary, i.e. $C_{\boldsymbol{w}}(t_i) \neq C_{\boldsymbol{w}}(t_j)$ for $t_i \neq t_j$, the demixing matrix is expected to be identifiable via a joint diagonalization of a set of covariance matrices at different times.

In order to separate stationary but non-white signals, one possibility is to use the non-zero autocorrelations at different time instances t_1 and t_2 with $t_1 \neq t_2$, namely

$$\widetilde{C}_{\boldsymbol{w}}(t_1, t_2) := \mathbb{E}[\boldsymbol{w}(t_1)\boldsymbol{w}^{\mathsf{H}}(t_2)] = A\widetilde{C}_{\boldsymbol{s}}(t_1, t_2)A^{\mathsf{H}}. \tag{9}$$

Note that, although the autocorrelation matrix $\widetilde{C}_{\boldsymbol{s}}(t_1, t_2)$ of the sources is still diagonal, it is not real in general. In other words, the autocorrelation matrix of the observations is generally *not* a Hermitian matrix and consequently *not* positive definite either. Similarly as above, the demixing matrix is expected to be identifiable via a joint diagonalization of a set of autocorrelation matrices with different time pairs.

If the signals have a non-trivial imaginary part, additional properties can be employed for BSS. Besides the standard covariance matrix (8), a similar statistical quantity of complex valued signals, known as *pseudo-covariance matrix*, is defined as

$$R_{\boldsymbol{w}}(t) := \mathbb{E}[\boldsymbol{w}(t)\boldsymbol{w}^{\mathsf{T}}(t] = AR_{\boldsymbol{s}}(t)A^{\mathsf{T}}. \tag{10}$$

The works in [32, 33] have shown that, if the sources are all non-circular with distinct circularity coefficients (5), i.e. distinct diagonal entries of $R_{\boldsymbol{s}}(t)$, the demixing matrix can be successfully identified by jointly diagonalizing both the covariance and the pseudo-covariance matrix. The resulting algebraic solution, referred to as *Strong Uncorrelating Transform*, provides a simple answer to the complex BSS problem. However, it fails in separating non-circular signals with same circularity coefficients. To overcome this problem, one can either utilize iterative contrast function based algorithms or employ some additional information, as for example the *pseudo-autocorrelation matrix* of the signals, which is defined as

$$\widetilde{R}_{\boldsymbol{w}}(t_1, t_2) := \mathbb{E}[\boldsymbol{w}(t_1)\boldsymbol{w}^{\mathsf{T}}(t_2)] = A\widetilde{R}_{\boldsymbol{s}}(t_1, t_2)A^{\mathsf{T}}. \tag{11}$$

Note that both the pseudo-covariance and pseudo-autocorrelation matrix are complex symmetric. Recent work in [40] considers the problem of jointly diagonalizing a set of both auto-correlation and pseudo-autocorrelation matrices. The identifiability results for this approach are still lacking in the literature and follow from our main result in Section 3.

2.5 Higher-Order Statistics (Tensor) Based ICA Approaches

In many real applications, second-order statistics may not be sufficient to accomplish the task of separation. In these situations higher-order statistics can be exploited. For example, statistically independent non-Gaussian signals can be blindly separated by using the fourth-order [14], or higher-order cumulants, cf. [41, 42].

Recalling the model as given in (6), the k-th order cumulant tensor of the sources $\boldsymbol{s}(t)$, denoted by $\mathcal{C}^{(k)}_{\boldsymbol{s},\boldsymbol{\iota}} \in (\mathbb{C}^m)^k$, is defined with its $(i_1, \ldots, i_k)$-th entry by

$$
\begin{aligned}
(\mathcal{C}^{(k)}_{\boldsymbol{s},\boldsymbol{\iota}})_{i_1 \ldots i_k} &:= \operatorname{cum}\left(s^{(*)}_{i_1}(t), \ldots, s^{(*)}_{i_k}(t)\right) \\
&= \sum_{p=1}^{k} (-1)^{p-1} (p-1)!\, \mathbb{E}\Big[\prod_{q \in J_1} s^{(*)}_q(t)\Big] \cdot \ldots \cdot \mathbb{E}\Big[\prod_{q \in J_p} s^{(*)}_q(t)\Big],
\end{aligned} \tag{12}
$$

where $\boldsymbol{\iota} = [\iota_1, \ldots, \iota_k] \in \{0,1\}^k$ is a binary vector which enables or disables complex conjugate in each dimension, i.e.

$$
\iota_i = \begin{cases} 0 & \text{no complex conjugate,} \\ 1 & \text{complex conjugate.} \end{cases} \tag{13}
$$

The summation in (12) involves all possible partitions $\{J_1, \ldots, J_p\}$ $(1 \leq p \leq k)$ of the indices $\{i_1, \ldots, i_k\}$. We refer to [41, 43] for further details regarding higher-order cumulant tensors.

Now, by varying two selected indices, say (i_p, i_q), with all other indices fixed, we obtain one *cumulant matrix* or *cumulant slice* of $\mathcal{C}^{(k)}_{\boldsymbol{s},\boldsymbol{\iota}}$, denoted by $(\mathcal{C}^{(k)}_{\boldsymbol{s},\boldsymbol{\iota}})_{\{p,q\}} \in \mathbb{C}^{m \times m}$. The assumption that sources are mutually statistically independent implies that all off-diagonal entries of the cumulant tensors of any order must be zero, i.e., the cumulant matrices $(\mathcal{C}^{(k)}_{\boldsymbol{s},\boldsymbol{\iota}})_{\{p,q\}}$ are diagonal for all $p \neq q$. Multilinear properties of the cumulant tensors lead to

$$
(\mathcal{C}^{(k)}_{\boldsymbol{w},\boldsymbol{\iota}})_{\{p,q\}} = A (\mathcal{C}^{(k)}_{\boldsymbol{s},\boldsymbol{\iota}})_{\{p,q\}} A^{\dagger}, \tag{14}
$$

where $(\cdot)^{\dagger}$ is determined by the construction (12). Similarly, by exploiting higher-order non-stationarity or higher-order circularity of the sources, the ICA problem is formulated as jointly diagonalizing a set of slices of the cumulant tensors via either Hermitian congruence or transpose congruence. Note, that up to date only specific cumulant tensors have been considered for BSS [42] via joint diagonalization, that end up with $(\cdot)^{\dagger}$ being Hermitian conjugate in Equation (14).

In this work, we answer the question on the identifiability of the BSS problem based on the joint diagonalization of a finite set of higher-order statistics matrices.

2.6 A Unified NUJD Approach

We summarize the above observations in a unified approach for non-unitary joint diagonalization. Let $\{C_i\}_{i=1}^{n}$ be a set of $m \times m$ complex matrices, constructed by

$$
C_i = A \Omega_i A^{\dagger_i}, \quad i = 1, \ldots n, \tag{15}
$$

where $\Omega_i = \operatorname{diag}\left(\omega_{i1}, \ldots, \omega_{im}\right) \in \mathbb{C}^{m \times m}$ and $\Omega_i \neq 0$. Note, that Equation (15) allows mixtures of both Hermitian congruence and transpose congruence transformations. The task is to find a matrix $X \in Gl(m)$ such that the matrices

$$
\left\{ X^{\mathsf{H}} C_i (X^{\mathsf{H}})^{\dagger_i} \,\middle|\, i = 1, \ldots, n \right\} \tag{16}
$$

are simultaneously diagonalized. Our present work concentrates on developing the uniqueness conditions on the set of Ω_i's, such that the matrix A is identifiable up to permutation and diagonal scaling.

Note that in the Hermitian congruence case, $X^{\mathsf{H}}C_iX$, $i = 1, \ldots, n$, are diagonal if and only if X simultaneously diagonalizes the Hermitian and the skew-Hermitian part of the C_i. Namely, by considering the real and the imaginary part of Ω_i in Equation (15) that corresponds to $C_i = A\Omega_i A^{\mathsf{H}}$ separately, we can construct two Hermitian matrices as

$$C_i' = A\Re\Omega_i A^{\mathsf{H}} \qquad \text{and} \tag{17}$$

$$C_i'' = A\Im\Omega_i A^{\mathsf{H}}. \tag{18}$$

Therefore, without loss of generality, we study an equivalent formulation of Problem (15) by restricting Ω_i to be real diagonal whenever $(\cdot)^{\dagger}$ is the Hermitian transpose.

Clearly, the mixing matrix can only be identified up to permutation and scaling. We define the set of all column-wise permutated diagonal $(m \times m)$-matrices by

$$\begin{aligned}\mathcal{G}(m) := \{DP \mid\ & D \in Gl(m) \text{ is diagonal and} \\ & P \text{ is a permutation matrix}\}.\end{aligned} \tag{19}$$

Since $\mathcal{G}(m)$ admits a matrix group structure, we can define the following equivalence class on $\mathbb{C}^{m\times m}$, cf. [44].

Definition 1 (Essential Equivalence)**.** *Let $X, Y \in Gl(m)$, then X is said to be* essentially equivalent *to Y, and vice versa, if there exists $E \in \mathcal{G}(m)$ such that*

$$X = YE. \tag{20}$$

Moreover, we say that the solution of a matrix equation is essentially unique*, if the equation admits a unique solution on the set of equivalence classes.*

Since

$$X^{\mathsf{H}}C_i(X^{\mathsf{H}})^{\dagger_i} = (X^{\mathsf{H}}A)\Omega_i(X^{\mathsf{H}}A)^{\dagger_i}, \tag{21}$$

we assume without loss of generality for further studies that the $C_i = \Omega_i$, $i = 1, \ldots, n$, are already diagonal. Thus, the identifiability analysis is restricted to investigating under what conditions the unit equivalence class $\mathcal{G}(m)$ admits the only solutions to the simultaneous diagonalization problem (16).

3 Uniqueness of Non-Unitary Joint Diagonalzition

In this section we present the main results on the uniqueness analysis of the NUJD problem given by Equations (15) and (16). In contrast to existing results on joint diagonalization, we do not assume the matrices to be real as in [22], positive definite as in [18], nor do we restrict the number of matrices to two as in [45, 46]. For the sake of readability, we outsource the proofs of the results to the appendix.

The identifiability results require a notion of collinerarity for diagonal matrices. Let Z_i, $i = 1, \ldots, n$, denote n complex diagonal $(m \times m)$-matrices with the diagonal entries $z_{i1}, \ldots, z_{im}$. For a fixed diagonal position k, we denote by $\boldsymbol{z}_k := [z_{1k}, \ldots, z_{nk}]^{\mathsf{T}} \in \mathbb{C}^n$ the vector consisting of the k-th diagonal element of each matrix, respectively. Recall that the cosine of the complex angle between two vectors $\boldsymbol{v}, \boldsymbol{w} \in \mathbb{C}^n$ is computed as

$$c(\boldsymbol{v}, \boldsymbol{w}) := \begin{cases} \frac{\boldsymbol{v}^{\mathsf{H}} \boldsymbol{w}}{\|\boldsymbol{v}\| \|\boldsymbol{w}\|} & \text{if } \boldsymbol{v} \neq 0 \wedge \boldsymbol{w} \neq 0, \\ 1 & \text{otherwise.} \end{cases} \tag{22}$$

where $\|\boldsymbol{v}\|$ denotes the Euclidean norm of a vector $\boldsymbol{v}$. We measure the collinearity of a set of n complex diagonal $(m \times m)$-matrices by means of the complex angle of the vectors formed by stacking the entries at corresponding positions together. Precisely, the collinearity measure for the set of Z_i's is defined by

$$\rho(Z_1, \ldots, Z_n) := \max_{1 \leq k < l \leq m} |c(\boldsymbol{z}_k, \boldsymbol{z}_l)|. \tag{23}$$

Note, that $0 \leq \rho \leq 1$ and that $\rho = 1$ if and only if there exists a complex scalar ω and a pair $\boldsymbol{z}_k, \boldsymbol{z}_l$, $k \neq l$ so that $\boldsymbol{z}_k = \omega \boldsymbol{z}_l$.

Our first result deals with the simple situations where either only purely Hermitian or purely complex symmetric matrices are involved. The techniques used for deriving the uniqueness conditions for both cases are adapted from the work in [22], where only real symmetric matrices are considered.

Theorem 1. *(a) Let $\Omega_i \in \mathbb{C}^{m \times m}$, for $i = 1, \ldots, n$, be diagonal, and let $X \in Gl(m)$ so that $X^{\mathsf{H}} \Omega_i X^*$ is diagonal as well. Then X is essentially unique if and only if $\rho(\Omega_1, \ldots, \Omega_n) < 1$.*

(b) Let $\Omega_i \in \mathbb{R}^{m \times m}$, for $i = 1, \ldots, n$, be diagonal, and let $X \in Gl(m)$ so that $X^{\mathsf{H}} \Omega_i X$ is diagonal as well. Then X is essentially unique if and only if $\rho(\Omega_1, \ldots, \Omega_n) < 1$.

The above theorem answers the identifiability question of complex BSS by means of matrix joint diagonalization approaches, when either purely complex symmetric, or purely Hermitian matrices are involved. For the situations with a mixture of Hermitian and complex symmetric matrices, we continue by firstly considering the case of simultaneously diagonalizing *one* complex symmetric and *one* Hermitian matrix. The following theorem generalizes a result in [47] and [48], wherein the authors require positive definiteness of the Hermitian matrix.

Theorem 2. *Let $X \in Gl(m)$ and let Ω_1 be a complex and Ω_2 a real diagonal matrix such that*

$$X^{\mathsf{H}} \Omega_1 X^* \text{ and } X^{\mathsf{H}} \Omega_2 X \text{ are diagonal.} \tag{24}$$

Then X is essentially unique if and only if

$$|\omega_{1k}||\omega_{2l}| \neq |\omega_{1l}||\omega_{2k}|, \tag{25}$$

holds for all pairs (k, l) with $k \neq l$.

Finally, by considering a mixture of multiple Hermitian and complex symmetric matrices, the following theorem completes our answer to the uniqueness analysis to the NUJD problem.

Theorem 3. *Let $C_i = X^{\mathsf{H}} \Omega_i X^*$ for $i = 1, \ldots, s$ and $C'_j = X^{\mathsf{H}} \Omega'_j X$ for $j = 1, \ldots, h$ be diagonal. Moreover, let*

$$\rho(\Omega_1, \ldots, \Omega_s) = \rho(\Omega'_1, \ldots, \Omega'_h) = 1, \tag{26}$$

then X is essentially unique if and only if there exists no *pair (k, l) with $k \neq l$, such that the following two conditions hold:*

$$\textit{(i)} \quad |c(\boldsymbol{\omega}_k, \boldsymbol{\omega}_l)| = |c(\boldsymbol{\omega}'_k, \boldsymbol{\omega}'_l)| = 1 \tag{27a}$$

$$\textit{(ii)} \quad \|\boldsymbol{\omega}_k\| \|\boldsymbol{\omega}_l\| = \|\boldsymbol{\omega}'_k\| \|\boldsymbol{\omega}'_l\|. \tag{27b}$$

4 Applications to Complex BSS

In this section, we firstly apply the uniqueness results from the previous section to the NUJD based complex BSS methods. The second application of the uniqueness results focuses on the development of algebraic solutions, i.e. solutions that only involve eigenvalue or singular value decompositions. Although the algebraic approaches are in general less powerful and less robust to noise and estimation errors than their iterative counterparts, cf. [35], these methods are of particular interest, as they provide simple, efficient solutions based on various powerful eigensolvers, cf. [45, 49, 50].

4.1 Identifiability of Complex BSS

From the main results developed in Section 3, any existing identifiability result of complex BSS follows straightforwardly. However, to the best of the authors' knowledge there are no general results, which unify HOS based NUJD approaches.

Let $\boldsymbol{t} := [t_1, \ldots, t_m]^\mathsf{T} \in \mathbb{R}^m$ be a set of time instances for each observed signal $w_i(t)$, we define the k-th order auto-cumulant tensor of the observations $\boldsymbol{w}(t)$, cf. [43], denoted by $\mathcal{C}^{(k)}_{\boldsymbol{w},\boldsymbol{\iota}}(\boldsymbol{t})$, with its $(i_1, \ldots, i_k)$-th entry

$$\left(\mathcal{C}^{(k)}_{\boldsymbol{w},\boldsymbol{\iota}}(\boldsymbol{t})\right)_{i_1 \ldots i_k} := \operatorname{cum}(w^{(*)}_{i_1}(t_{i_1}) \cdot \ldots \cdot w^{(*)}_{i_k}(t_{i_k})), \tag{28}$$

Similarly as in Equation (14), the (p, q)-th slice of the k-th auto-cumulant tensor with a set of given time $\boldsymbol{t}$ is computed as

$$\left(\mathcal{C}^{(k)}_{\boldsymbol{w},\boldsymbol{\iota}}(\boldsymbol{t})\right)_{\{p,q\}} := A \left(\mathcal{C}^{(k)}_{\boldsymbol{s},\boldsymbol{\iota}}(\boldsymbol{t})\right)_{\{p,q\}} A^\dagger. \tag{29}$$

The identifiability of the complex BSS problem via jointly diagonalizing a set of higher-order cumulant matrices is summarized as follows.

Theorem 4 (Identifiability of Complex ICA)**.** *Given the complex linear BSS model as in* (6) *and a set of time instances $\boldsymbol{t}_i := [t_{i1}, \ldots, t_{im}]^\mathsf{T}$ for $i = 1, \ldots, T$, then the joint diagonalizer of the set*

$$\left\{ \left(\mathcal{C}^{(k)}_{\boldsymbol{w},\boldsymbol{\iota}}(\boldsymbol{t_i})\right)_{\{p,q\}} \right\}_{\substack{i=1,\ldots,T \\ k=2,\ldots,K \\ p,q=1,\ldots,m}} \tag{30}$$

is essentially unique and solves the BSS problem up to permutation and scaling, if and only if the diagonal matrices

$$\left\{ \left(\mathcal{C}^{(k)}_{\boldsymbol{s},\boldsymbol{\iota}}(\boldsymbol{t_i})\right)_{\{p,q\}} \right\}_{\substack{i=1,\ldots,T \\ k=2,\ldots,K \\ p,q=1,\ldots,m}} \tag{31}$$

fulfill one of the following three conditions:

(i) $\rho(\mathcal{S}) < 1$*, where* $\mathcal{S}$ *denotes the set of cumulant matrices constructed via transpose congruence, i.e.*

$$\mathcal{S} := \left\{ \left(\mathcal{C}^{(k)}_{\boldsymbol{s},\boldsymbol{\iota}}(\boldsymbol{t_i}) \right)_{\{p,q\}} \middle| \iota_p \oplus \iota_q = 0 \right\}; \tag{32}$$

(ii) $\rho(\mathcal{H}) < 1$*, where* $\mathcal{H}$ *is the set of cumulant matrices constructed via Hermitian congruence, i.e.*

$$\mathcal{H} := \left\{ \left(\mathcal{C}^{(k)}_{\boldsymbol{s},\boldsymbol{\iota}}(\boldsymbol{t_i}) \right)_{\{p,q\}} \middle| \iota_p \oplus \iota_q = 1 \right\}; \tag{33}$$

(iii) When both the previous two conditions are violated, Equation (27) *still holds.*

Example 1 (Fourth-Order Cumulants)**.** *Recall the complex BSS model* (6)*, the fourth-order cumulant of a subset of chosen sources* $(s_{i_1}, s_{i_2}, s_{i_3}, s_{i_4})$ *is computed explicitly as*

$$\begin{aligned} \operatorname{cum}(s_{i_1}, s_{i_2}, s_{i_3}, s_{i_4}) = & \mathbb{E}[s_{i_1}(t)s_{i_2}(t)s_{i_3}(t)s_{i_4}(t)] \\ & - \mathbb{E}[s_{i_1}(t)s_{i_2}(t)]\mathbb{E}[s_{i_3}(t)s_{i_4}(t)] \\ & - \mathbb{E}[s_{i_1}(t)s_{i_3}(t)]\mathbb{E}[s_{i_2}(t)s_{i_4}(t)] \\ & - \mathbb{E}[s_{i_1}(t)s_{i_4}(t)]\mathbb{E}[s_{i_2}(t)s_{i_3}(t)]. \end{aligned} \tag{34}$$

By taking into account all possible combinations of complex conjugate on each component, we have three different fourth-order cumulant tensors

$$(\mathcal{C}^{(4)}_{\boldsymbol{w},\boldsymbol{\iota}_1})_{i_1 \ldots i_4} := \operatorname{cum}(w_{i_1}, w_{i_2}, w_{i_3}, w_{i_4}), \tag{35a}$$

$$(\mathcal{C}^{(4)}_{\boldsymbol{w},\boldsymbol{\iota}_2})_{i_1 \ldots i_4} := \operatorname{cum}(w^*_{i_1}, w_{i_2}, w_{i_3}, w_{i_4}), \tag{35b}$$

$$(\mathcal{C}^{(4)}_{\boldsymbol{w},\boldsymbol{\iota}_3})_{i_1 \ldots i_4} := \operatorname{cum}(w^*_{i_1}, w^*_{i_2}, w_{i_3}, w_{i_4}). \tag{35c}$$

Current works in the BSS literature only focus on the cases, where source signals are assumed to be harmonic*, i.e. the quantity* (35c) *does not vanish, while the other two are equal to zero, cf. [43]. Theorem 3 in [27] presents a result on the identifiability of separating harmonic sources using only the* 4*-th order cumulants* (35c)*. Certainly, when the sources are* non-harmonic*, i.e. all possible fourth-order cumulants* (35) *do not vanish, then the BSS problem can be still solvable via a joint diagonalization of fourth-order cumulant matrices, even though the conditions given in [27] are violated.*

4.2 Algebraic Solutions to Complex BSS

In this subsection, we investigate a particularly simple solution to the complex BSS problem. It is given in closed form in terms of an eigenvalue and a singular value decomposition of two matrices. We refer to such solutions as *algebraic solutions*. These methods are of high interest, since the existence of fast eigensolvers turns them into very fast solvers for BSS. Algebraic solutions exist if either both matrices are complex symmetric or Hermitian, and at least one is invertible, cf. [45]. For the mixed case, the strong uncorrelating transform

(SUT), where a combination of an eigenvalue decomposition and a Takagi factorization is used, provides an algebraic solution only if the Hermitian matrix is positive definite. In this subsection, we extend this approach and investigate the situation of separating non-circular signals with non-distinct circularity coefficients, cf. [51].

Lemma 1. *Let $C_1, C_2 \in Gl(m)$ be one complex symmetric and one Hermitian matrix, respectively, constructed by*

$$C_1 := A\Omega_1 A^{\mathsf{T}} \qquad \textit{and} \tag{36}$$

$$C_2 := A\Omega_2 A^{\mathsf{H}}, \tag{37}$$

where $A \in Gl(m)$, Ω_1 is complex diagonal, and Ω_2 is real diagonal. Let $C_2 = U\Sigma U^{\mathsf{T}}$ be the Takagi factorization of C_2. Then,

(i) the matrix $\widetilde{C}_1 := \Sigma^{-1/2}U^{\mathsf{H}}C_1 U\Sigma^{-1/2}$ admits a matrix factorization of the form $\widetilde{C}_1 = V\Lambda V^{\mathsf{H}}$, where $V \in O(m)$ and Λ is diagonal;

(ii) the matrix $X := U\Sigma^{-1/2}V^*$ satisfies

$$X^{\mathsf{H}}C_2 X^* = I \quad \textit{and} \quad X^{\mathsf{H}}C_1 X \textit{ is diagonal.} \tag{38}$$

As the complex symmetric matrix C_2 reflects the pseudo second-order statistics of complex signals, we name the matrix X *Pseudo-Uncorrelating Transform (PUT)* in referring its connection to SUT. A straightforward computation shows that the matrix V consists of the eigenvectors of $\widetilde{C}_1\widetilde{C}_1^{\mathsf{T}}$, as

$$\widetilde{C}_1\widetilde{C}_1^{\mathsf{T}} = V\Lambda V^{\mathsf{H}}V^*\Lambda V^{\mathsf{T}} = V\Lambda^2 V^{\mathsf{T}}. \tag{39}$$

Thus, if W is a matrix such that $\widetilde{C}_1\widetilde{C}_1^{\mathsf{T}} = W\Lambda' W^{-1}$ and if the eigenvalues Λ' are pairwise distinct, it follows by the uniqueness of the EVD, that $V = W(W^{\mathsf{T}}W)^{-1/2}DP$, where P is a permutation and D is diagonal with entries being ± 1. Ultimately, we summarize the procedure for computing the PUT in Algorithm 1.

Algorithm 1. *Pseudo-Uncorrelating Transform (PUT)*

Step 1: Construct C_1, C_2 from the observations $\boldsymbol{w}(t)$, where C_1 and C_2 are constructed via Hermitian congruence and matrix congruence, respectively;
Step 2: Compute the Takagi factorization of $C_2 = U\Sigma U^{\mathsf{T}}$;
Step 3: Let $\widetilde{C}_1 := \Sigma^{-1/2}U^{\mathsf{H}}C_1 U\Sigma^{-1/2}$, compute EVD of $\widetilde{C}_1\widetilde{C}_1^{\mathsf{T}} = W\Lambda W^{-1}$;
Step 4: Compute $V = W(W^{\mathsf{T}}W)^{-1/2}$;
Step 5: Compute the PUT matrix $X = U\Sigma^{-1/2}V^*$;

Remark 1. *When the matrix C_1 is Hermitian and positive definite, i.e. C_1 being the covariance matrix of the observations, then the entries of Λ in* (39) *are simply the reciprocal of the circularity coefficients of sources. Our result coincides with the identifiability condition of SUT, cf. theorem 2 in [33].*

Remark 2. *The second observation is that the SUT of an arbitrary pair of one positive definite Hermitian and one complex symmetric matrix does always exist, cf. [48]. In contrast, the existence of the PUT matrix is not guaranteed for an arbitrary pair of a complex symmetric and a (general) Hermitian matrix. However, existence of SUT implies the applicability of PUT on an arbitrary pair of positive definite Hermitian and complex symmetric matrix. In other words, PUT can be considered as a generalization of SUT.*

Corollary 1. *For an arbitrary pair of one Hermitian positive definite and one non-singular complex symmetric matrix, a PUT matrix always exists.*

Finally, we characterize the applicability of PUT as an effective BSS technique. Recall the complex linear BSS model as in (6), let $\boldsymbol{t} := [t_1, \ldots, t_m]^\mathsf{T} \in \mathbb{R}^m$ represent m time instances of individual observations, and denote by $\widetilde{C}_{\boldsymbol{s}}(\boldsymbol{t})$ and $\widetilde{R}_{\boldsymbol{s}}(\boldsymbol{t})$ the autocorrelation and pseudo-autocorrelation matrix of the sources $\boldsymbol{s}(t)$, respectively. Their (i,j)-th entries are computed as

$$\left(\widetilde{C}_{\boldsymbol{s}}(\boldsymbol{t})\right)_{ij} := \mathbb{E}[s_i(t_i)s_j^*(t_j)], \tag{40}$$

and

$$\left(\widetilde{R}_{\boldsymbol{s}}(\boldsymbol{t})\right)_{ij} := \mathbb{E}[s_i(t_i)s_j(t_j)]. \tag{41}$$

Corollary 2. *If one of the two conditions:*

(i) $\left|\Re\left(\widetilde{C}_{\boldsymbol{s}}(\boldsymbol{t})\right)_{ii}\right|\left|\left(\widetilde{R}_{\boldsymbol{s}}(\boldsymbol{t})\right)_{jj}\right| \neq \left|\Re\left(\widetilde{C}_{\boldsymbol{s}}(\boldsymbol{t})\right)_{jj}\right|\left|\left(\widetilde{R}_{\boldsymbol{s}}(\boldsymbol{t})\right)_{ii}\right|,$

(ii) $\left|\Im\left(\widetilde{C}_{\boldsymbol{s}}(\boldsymbol{t})\right)_{ii}\right|\left|\left(\widetilde{R}_{\boldsymbol{s}}(\boldsymbol{t})\right)_{jj}\right| \neq \left|\Im\left(\widetilde{C}_{\boldsymbol{s}}(\boldsymbol{t})\right)_{jj}\right|\left|\left(\widetilde{R}_{\boldsymbol{s}}(\boldsymbol{t})\right)_{ii}\right|,$

is fulfilled for all pairs (i,j) with $i \neq j$, then the joint diagonalizer of one Hermitian and one complex symmetric matrix, constructed correspondingly from the observations $\mathbf{w}(t)$ via PUT, is essentially unique and solves the BSS problem up to permutation and scaling.

5 Conclusions

In this work, we study the problem of simultaneously diagonalizing a set of complex square matrices, and provide a thorough uniqueness analysis of the problem. In particular, we focus on its application in the problem of complex linear BSS. Our work not only characterizes a general result on identifiability conditions of the MJD based BSS methods, but also derives a generalized algebraic BSS solution, i.e. the PUT algorithm. Furthermore, the present results may also have impact in the areas of beamforming [52], and direction of arrival estimation [53], where matrix joint diagonalization approaches play an important role.

Acknowledgement

This work has partially been supported by the Cluster of Excellence *CoTeSys* - Cognition for Technical Systems, funded by the German research foundation (DFG).

.1 Proof of Theorem 1

(a) First, consider the case $m = 2$ and let

$$X = \begin{bmatrix} x_1 & x_2 \\ x_3 & x_4 \end{bmatrix} \in Gl(2). \tag{42}$$

Then $X^{\mathsf{H}}\Omega_i X^*$ is diagonal for $i = 1, \ldots, n$, if and only if

$$\omega_{i1}^* x_1 x_2 + \omega_{i2}^* x_3 x_4 = 0, \tag{43}$$

for $i = 1, \ldots, n$. The corresponding system of linear equations reads as

$$\begin{bmatrix} \omega_{11} & \omega_{21} & \ldots & \omega_{n1} \\ \omega_{12} & \omega_{22} & \ldots & \omega_{n2} \end{bmatrix}^{\mathsf{H}} \begin{bmatrix} x_1 x_2 \\ x_3 x_4 \end{bmatrix} = 0, \tag{44}$$

which only has a unique trivial solution if and only if the coefficient matrix has rank 2. This is equivalent to $\rho(\Omega_1, \ldots, \Omega_n) < 1$. The trivial solution, i.e. $x_1 x_2 = x_3 x_4 = 0$, together with the invertibility of X yields that either $x_1 = 0$ and $x_4 = 0$, or, $x_2 = 0$ and $x_3 = 0$. This, in turn, is equivalent to $X \in \mathcal{G}(2)$.

Consider now the case $m > 2$. If $\rho = 1$ then there exists a pair (k, l) such that $|c(\boldsymbol{\omega}_k, \boldsymbol{\omega}_l)| = 1$ and the same argument as above shows that $\rho = 1$ implies the non-uniqueness of the joint diagonalizer. For the reverse direction of the statement, assume that the joint diagonalizer X is not in $\mathcal{G}(m)$. We have to show that this implies $\rho = 1$.

Now assume first that one of the Ω_i's, say Ω_1, is invertible. Then

$$X^{\mathsf{H}}\Omega_i X^* (X^{\mathsf{H}}\Omega_1 X^*)^{-1} = X^{\mathsf{H}}\Omega_i \Omega_1^{-1} (X^{\mathsf{H}})^{-1}, \tag{45}$$

for $i = 1, \ldots, n$, gives the simultaneous eigendecomposition of the diagonal matrices $\Omega_i \Omega_1^{-1}$. Since $X \notin \mathcal{G}(m)$, there exists a pair (k, l) with $k \neq l$ such that

$$\frac{\omega_{ik}}{\omega_{1k}} = \frac{\omega_{il}}{\omega_{1l}}, \tag{46}$$

which is equivalent to $|c(\boldsymbol{\omega}_k, \boldsymbol{\omega}_l)| = 1$ and hence $\rho(\Omega_1, \ldots, \Omega_n) = 1$. If all the Ω_i's are singular, we distinguish between two cases. Firstly, assume that there is a position on the diagonals, say k, where all $\omega_{ik} = 0$. Then $|c(\boldsymbol{\omega}_k, \boldsymbol{\omega}_l)| = 1$ holds true for any $k \neq l$ and thus $\rho = 1$. Secondly, if there is no common position where all the Ω_i's have a zero entry, there exists an invertible linear combination, say Ω_0, which can also be diagonalized via the same transformations. Then by considering a new set $\{\Omega_i\}_{i=0}^n$, the same argument as from (45) to (46) for the invertible case applies by replacing Ω_1 with Ω_0. This completes the proof for part (a).

(b) For $m = 2$, the condition that $X^{\mathsf{H}}\Omega_i X$ is diagonal for all $i = 1, \ldots, n$ leads to the system of linear equations

$$\begin{bmatrix} \omega_{11} & \omega_{21} & \ldots & \omega_{n1} \\ \omega_{12} & \omega_{22} & \ldots & \omega_{n2} \end{bmatrix}^{\mathsf{T}} \begin{bmatrix} x_1 x_3^* \\ x_2 x_4^* \end{bmatrix} = 0, \tag{47}$$

which admits a non-trivial solution if and only if $\rho(\Omega_1, \ldots, \Omega_n) = 1$. Now, $x_1 x_3^* = x_2 x_4^* = 0$ together with the invertibility of X implies that X is essentially unique. The case for $m > 2$ is now just as in Section .1 and is omitted here.

.2 Proof of Theorem 2

We prove an equivalent formulation of Theorem 2. Namely, a matrix $X \in Gl(m) \setminus \mathcal{G}(m)$ that fulfills condition (24) exists, if and only if there exists a pair (k, l) with $k \neq l$ such that

$$|\omega_{1k}||\omega_{2l}| = |\omega_{1l}||\omega_{2k}|. \tag{48}$$

Firstly, consider the case $m = 2$. From Equations (43) and (47) we see that the condition (24) is equivalent to

$$\begin{cases} \omega_{11}^* x_1 x_2 + \omega_{12}^* x_3 x_4 = 0 \\ \omega_{21} x_1 x_2^* + \omega_{22} x_3 x_4^* = 0. \end{cases} \tag{49}$$

Assume now that $X \in Gl(2) \setminus \mathcal{G}(2)$ and, without loss of generality $|x_1 x_2| \neq 0$. Then either $|x_3 x_4| \neq 0$ and Equation (49) yields

$$|\omega_{11}| = |\omega_{12}| \frac{|x_3 x_4|}{|x_1 x_2|}, \quad |\omega_{21}| = |\omega_{22}| \frac{|x_3 x_4|}{|x_1 x_2|}, \tag{50}$$

or $|x_3 x_4| = 0$. Both cases imply Equation (48).

For the other direction, let Equation (48) hold true. We construct explicitly a common diagonalizer in $Gl(2) \setminus \mathcal{G}(2)$. The case when either $\Omega_1 = 0$ or $\Omega_2 = 0$ is trivial and not further discussed. Equation (48) implies

$$\Omega_1 = r \begin{bmatrix} \exp(\mathrm{i}\varphi_1) & \\ & \exp(\mathrm{i}\varphi_2) \end{bmatrix} \Omega_2, \tag{51}$$

with suitable $\varphi_i \in [0, 2\pi)$ and $r > 0$. Firstly, assume that one, and hence both, matrices Ω_1 and Ω_2 are not invertible. We choose without loss of generality ω_{22} to be 0. Equation (49) now implies $x_1 x_2 = 0$, but x_3 and x_4 can be chosen arbitrarily. Indeed, it is easily checked that in this case,

$$X := \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \tag{52}$$

is a common diagonalizer. Assume now that both, Ω_1 and Ω_2 are invertible. Then it is straightforwardly verified that

$$X := \Theta \Omega_2^{-1/2} \begin{bmatrix} \exp(-\frac{\mathrm{i}}{2}\varphi_1) & \\ & \exp(-\frac{\mathrm{i}}{2}\varphi_2) \end{bmatrix} \tag{53}$$

is a common diagonalizer for any real orthogonal matrix $\Theta \in O(2)$.

Now, let $m > 2$. If Equation (48) holds true, then the case for $m = 2$ applies and the diagonalizer is not essentially unique.

For the reverse direction, we assume firstly that both Ω_1 and Ω_2 are not invertible. Then either there exists an index pair (k, l) with $k \neq l$, such that Equation (48) holds true (with zeros on both sides of the equation) and it follows again from the case $m = 2$ that the diagonalizer is not essentially unique. Or, Ω_1 and Ω_2 both have at most one zero diagonal entry at different positions. This case will be treated at the end of the proof.

Let us now consider the case where Ω_2 is invertible. Assume that the diagonalizer is not essentially unique, i.e. that X in Equation (24) (and hence

X^{H} and X^*) differs from a product of a diagonal and a permutation matrix. The uniqueness of the QR-decomposition of the invertible matrix X, i.e. $X = Q_X R_X$, allows by further decomposing $R_X = D_X N_X$ with $D_X := \mathrm{ddiag}(R_X)$ and $N_X := D_X^{-1} R_X$ the unique factorization

$$X = Q_X D_X N_X \tag{54}$$

with unitary Q_X, positive real diagonal D_X, and N_X being upper triangular with $\mathrm{ddiag}(N_X) = I_m$. Here, $\mathrm{ddiag}(N_X)$ forms a diagonal matrix, whose diagonal entries are just those of N_X.

Using this decomposition, X is not in $\mathcal{G}(m)$ if and only if either $N_X \neq I_n$ or Q_X is not a product of a permutation matrix and a diagonal phase shift matrix. By a diagonal phase shift matrix, we mean all diagonal matrices in $U(m)$. Using the invertibility assumption on Ω_2,

$$\begin{aligned} Z : &= (X^{\mathsf{H}}\Omega_2 X)^{-1} X^{\mathsf{H}}\Omega_1 X^* \\ &= X^{-1}\Omega_2^{-1}\Omega_1 X^* \\ &= N_X^{-1} D_X^{-1} Q_X^{\mathsf{H}} \Omega_2^{-1}\Omega_1 Q_X^* D_X^* N_X^* \end{aligned} \tag{55}$$

is diagonal. This yields

$$D_X^{-1} Q_X^{\mathsf{H}} \Omega_2^{-1}\Omega_1 Q_X^* D_X^* = N_X Z (N_X^*)^{-1}, \tag{56}$$

where the matrix is symmetric on the left hand side and upper triangular on the right hand side. This leads us to two conclusions, namely that

$$N_X Z (N_X^*)^{-1} \quad \text{is diagonal} \tag{57}$$

and

$$D_X^{-1} Q_X^{\mathsf{H}} \Omega_2^{-1}\Omega_1 Q_X^* D_X^* \quad \text{is diagonal.} \tag{58}$$

Since $D_X = D_X^*$ is real and diagonal, the last Equation implies that

$$\widetilde{R} = Q_X^{\mathsf{H}} \Omega_2^{-1}\Omega_1 Q_X^* \quad \text{is diagonal} \tag{59}$$

and hence

$$\widetilde{R} = N_X Z (N_X^*)^{-1}, \tag{60}$$

Let us have a closer look at Equation (59). By introducing suitable diagonal phase shift matrices Φ_1 and Φ_2 we have

$$\Phi_1 Q_X^{\mathsf{H}} \Phi_2 \begin{bmatrix} |\omega_{11}/\omega_{21}| & & \\ & \ddots & \\ & & |\omega_{1m}/\omega_{2m}| \end{bmatrix} \Phi_2 Q_X^* \Phi_1 = R \tag{61}$$

where R is diagonal with real and nonnegative entries. Note that Equation (61) gives a Takagi factorization of R. If Q_X differs from a product of a permutation matrix and a phase shift matrix, the uniqueness of the Takagi factorization now implies that (at least) two diagonal entries have to coincide and consequently Equation (48) follows.

Assume now that N_X differs from the identity, and let its (k,l)-th entry, say z, differ from 0. Now $\widetilde{R} = \Phi_1^{*2} R$ and consequently Equation (60) yields

$$(N_X)^{-1} \Phi_1^{*2} R N_X = Z. \tag{62}$$

Note that, by the special structure of N_X, namely upper triangular with ones on the diagonal, this immediately implies $Z = \Phi_1^{*2} R$. Thus, the (k,l)-th entry of equation (62) reads as

$$z(\Phi_1^{*2} R)_{kk} = z(\Phi_1^{*2} R)_{ll}. \tag{63}$$

Taking absolute values, this implies $|R_{kk}| = |R_{ll}|$ for the corresponding diagonal entries of R and Equation (48) follows.

Now let us get back to the case where exactly one diagonal entry of Ω_2 is zero and the corresponding diagonal entry of Ω_1 differs from zero. Since Equation (24) is equivalent to

$$\Pi_1 X^{\mathsf{H}} \Pi_2 \Pi_2^{\mathsf{T}} \Omega_1 \Pi_2 \Pi_2^{\mathsf{T}} X^* \Pi_1^{\mathsf{T}} \text{ is diagonal,} \tag{64}$$

and

$$\Pi_1 X^{\mathsf{H}} \Pi_2 \Pi_2^{\mathsf{T}} \Omega_2 \Pi_2 \Pi_2^{\mathsf{T}} X \Pi_1^{\mathsf{T}} \text{ is diagonal,} \tag{65}$$

for any permutation matrices Π_1, Π_2, we assume without loss of generality that

$$X = \left[\begin{array}{c|c} \widetilde{X} & \boldsymbol{x}_1 \\ \hline \boldsymbol{x}_2^{\mathsf{H}} & x \end{array}\right], \Omega_1 = \left[\begin{array}{c|c} \widetilde{\Omega}_1 & 0 \\ \hline 0 & \omega_m \end{array}\right], \Omega_2 = \left[\begin{array}{c|c} \widetilde{\Omega}_2 & 0 \\ \hline 0 & 0 \end{array}\right], \tag{66}$$

where $\omega_m \neq 0$ and $\widetilde{X}, \widetilde{\Omega}_2 \in Gl(m-1)$. Now

$$X^{\mathsf{H}} \Omega_2 X = \left[\begin{array}{c|c} \widetilde{X}^{\mathsf{H}} \widetilde{\Omega}_2 \widetilde{X} & \widetilde{X}^{\mathsf{H}} \widetilde{\Omega}_2 \boldsymbol{x}_1 \\ \hline \star & \star \end{array}\right], \tag{67}$$

and Equation (24) together with the invertibility assumption on $\widetilde{X}$ and $\widetilde{\Omega}_2$ implies $\boldsymbol{x}_1 = 0$ and $x \neq 0$. Thus

$$X^{\mathsf{H}} \Omega_1 X^* = \left[\begin{array}{c|c} \widetilde{X}^{\mathsf{H}} \widetilde{\Omega}_1 \widetilde{X}^* + \omega_m \boldsymbol{x}_2 \boldsymbol{x}_2^{\mathsf{T}} & x \omega_m \boldsymbol{x}_2 \\ \hline \star & \star \end{array}\right], \tag{68}$$

and since $x \neq 0$ and $\omega_m \neq 0$, Equation (24) yields that $\boldsymbol{x}_2 = 0$. Hence, we just showed that if Ω_1 and Ω_2 are structured as in Equation (66), X can only be a common diagonalizer if

$$X = \left[\begin{array}{c|c} \widetilde{X} & 0 \\ \hline 0 & x \end{array}\right]. \tag{69}$$

Now, it is clear that $X \in Gl(m) \setminus \mathcal{G}(m)$ if and only if $\widetilde{X} \in Gl(m-1) \setminus \mathcal{G}(m-1)$, and we reduced the problem to the invertible case treated above. This concludes the proof of the theorem.

.3 Proof of Theorem 3

Again, we firstly consider the case $m = 2$. Assumption (26) is equivalent to Condition (27a) and due to the fact that Equations (44) and (47) have both nontrivial solutions, say

$$[x_1 x_2, x_3 x_4]^{\mathsf{T}} \text{ and } [x_1 x_2^*, x_3 x_4^*]^{\mathsf{T}}. \tag{70}$$

Thus, we have

$$\begin{cases} x_1 x_2 \boldsymbol{\omega}_1 + x_3 x_4 \boldsymbol{\omega}_2 = 0, \\ x_1 x_2^* \boldsymbol{\omega}_1 + x_3 x_4^* \boldsymbol{\omega}_2 = 0 \end{cases} \tag{71}$$

and, by taking absolute values,

$$\begin{cases} |x_1 x_2| \|\boldsymbol{\omega}_1\| = |x_3 x_4| \|\boldsymbol{\omega}_2\|, \\ |x_1 x_2^*| \|\boldsymbol{\omega}_1\| = |x_3 x_4^*| \|\boldsymbol{\omega}_2\| \end{cases} \tag{72}$$

and Condition (27b) follows.

To see the reverse direction, let Condition (27b) hold true. If $\|\boldsymbol{\omega}_1\| = \|\boldsymbol{\omega}_2\| = 0$ or if $\|\boldsymbol{\omega}_1'\| = \|\boldsymbol{\omega}_2'\| = 0$, the non-uniqueness of X follows from Theorem 1. Otherwise, (ii) implies (after a possible renumeration)

$$\begin{aligned} \boldsymbol{\omega}_2 &= r \mathrm{e}^{\mathrm{i}\varphi_1} \boldsymbol{\omega}_1 \\ \boldsymbol{\omega}_2' &= r \boldsymbol{\omega}_1', \end{aligned} \tag{73}$$

with $r > 0$ and $\varphi_1 \in [0, 2\pi)$. Using Equation (71), we find an explicit diagonalizer that is not in $\mathcal{G}(m)$, namely

$$\begin{aligned} x_1 &= \widetilde{r} \exp(\tfrac{\mathrm{i}}{2}\varphi_1), \quad x_2 = \tfrac{1}{r}, \\ x_3 &= \tfrac{1}{\widetilde{r}} \exp(\tfrac{\mathrm{i}}{2}\varphi_1), \quad x_4 = -1, \end{aligned} \tag{74}$$

where $\widetilde{r} \neq 0$ can be chosen arbitrarily such that X is invertible.

Let us consider now the case $m > 2$. If there exists a pair (k, l) with $k \neq l$, such that Conditions (27) hold, then we can use the above argument for the corresponding (2×2)-sub matrix and conclude that the common diagonalizer is not essentially unique. Now, let $X \in Gl(m) \setminus \mathcal{G}(m)$. Assume for the moment that at least one per C_i's and C_j''s is invertible, say, C_1 and C_1'. This implies that $C_i C_j' C_1^{-1} C_1'^{-1} = X^{\mathsf{H}} \Omega_i \Omega_j' \Omega_1^{-1} \Omega_1'^{-1} X^{-1}$, for $i = 1, \ldots, s$ and $j = 1, \ldots, h$, is a simultaneous eigendecomposition. Since $X \in Gl(m) \setminus \mathcal{G}(m)$, there must be an index pair (k, l) with $k \neq l$, such that

$$\frac{\omega_{ik} \omega_{jk}'}{\omega_{1k} \omega_{1k}'} = \frac{\omega_{il} \omega_{jl}'}{\omega_{1l} \omega_{1l}'} \tag{75}$$

for all $i = 1, \ldots, s$ and $j = 1, \ldots, h$. This yields $|c(\boldsymbol{\omega}_k, \boldsymbol{\omega}_l)| = |c(\boldsymbol{\omega}_k', \boldsymbol{\omega}_l')| = 1$ and hence Equation (i) follows. If none of the C_i is invertible, the same argument as in Theorem 1 applied to both sets $\{C_i\}_{i=1}^s$ and $\{C_j'\}_{j=1}^h$ individually yields the same conclusion as in (27a).

Hence, by permuting k and l if necessary, there exist $z_1, z_2 \in \mathbb{C}$ such that

$$\boldsymbol{\omega}_k = z_1 \boldsymbol{\omega}_l, \text{ and } \boldsymbol{\omega}_k' = z_2 \boldsymbol{\omega}_l'. \tag{76}$$

On the other hand, by Theorem 2, we obtain

$$|\omega_{ik}| |\omega_{jl}'| = |\omega_{il}| |\omega_{jk}'| \tag{77}$$

for all $i = 1, \ldots, s$ and $j = 1, \ldots, h$, and hence $|z_1| = |z_2|$. Equation (76) now yields Equation (27b) and the proof is complete.

.4 Proof of Lemma 1

(i) The construction of C_2 as in Equations (36) and (37) implies

$$A\Omega_2 A^{\mathsf{T}} = U\Sigma U^{\mathsf{T}}. \tag{78}$$

As diagonal entries of Σ are all positive, Equation (78) is equivalent to

$$\Sigma^{-1/2} U^{\mathsf{H}} A\Omega_2 A^{\mathsf{T}} U^{*} \Sigma^{-1/2} = I_m. \tag{79}$$

By inserting $\Omega_2 = (\Omega_2^{1/2})^2$ into the above equation, it can be seen that $V := \Sigma^{-1/2} U^{\mathsf{H}} A \Omega_2^{1/2}$ is complex orthogonal. Now, $A = U\Sigma^{1/2} V \Omega_2^{-1/2}$, and thus Equations (36) and (37) yield

$$C_1 = A\Omega_1 A^{\mathsf{H}} = U\Sigma^{1/2} V \underbrace{\Omega_2^{-1/2}\Omega_1\Omega_2^{-\mathsf{H}/2}}_{=:\Lambda} V^{\mathsf{H}} \Sigma^{1/2} U^{\mathsf{H}}, \tag{80}$$

where Λ is diagonal. Then, Equation (80) is equivalent to

$$\Sigma^{-1/2} U^{\mathsf{H}} C_1 U \Sigma^{-1/2} = V\Lambda V^{\mathsf{H}}. \tag{81}$$

(ii) It is straightforward to verify that

$$X^{\mathsf{H}} C_1 X = V^{\mathsf{T}} \Sigma^{-1/2} U^{\mathsf{H}} C_1 U \Sigma^{-1/2} V^{*} = \Lambda, \tag{82}$$

and

$$X^{\mathsf{H}} C_2 X^{*} = V^{\mathsf{T}} \Sigma^{-1/2} U^{\mathsf{H}} C_2 U^{*} \Sigma^{-1/2} V = I_m. \tag{83}$$

References

[1] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. New York: Wiley, 2001.

[2] P. Comon, "Independent component analysis, a new concept?" *Signal Processing*, vol. 36, no. 3, pp. 287–314, 1994.

[3] F. J. Theis, "Uniqueness of complex and multidimensional independent component analysis," *Signal Processing*, vol. 84, no. 5, pp. 951–956, 2004.

[4] ——, "A new concept for separability problems in blind source separation," *Neural Computation*, vol. 16, no. 9, pp. 1827–1850, 2004.

[5] A. Hyvärinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Transactions on Neural Networks*, vol. 10, no. 3, pp. 626–634, 1999.

[6] H. Shen, M. Kleinsteuber, and K. Hüper, "Local convergence analysis of FastICA and related algorithms," *IEEE Transactions on Neural Networks*, vol. 19, no. 6, pp. 1022–1032, 2008.

[7] H. Shen, K. Hüper, and M. Kleinsteuber, "On FastICA algorithms and some generalisations," in *Numerical Linear Algebra in Signals, Systems and Control*, ser. Lecture Notes in Electrical Engineering, P. Van Dooren, S. P. Bhattacharyya, R. H. Chan, V. Olshevsky, and A. Routray, Eds. Springer Netherlands, 2011, vol. 80, pp. 403–432.

[8] D.-T. Pham, "Mutual information approach to blind separation of stationary sources," *IEEE Transactions on Information Theory*, vol. 48, no. 7, pp. 1935–1946, 2002.

[9] F. Vrins, D.-T. Pham, and M. Verleysen, "Mixing and non-mixing local minima of the entropy contrast for blind source separation," *IEEE Transactions on Information Theory*, vol. 53, no. 3, pp. 1030–1042, 2007.

[10] R. Boscolo, H. Pan, and V. P. Roychowdhury, "Independent component analysis based on nonparametric density estimation," *IEEE Transactions on Neural Networks*, vol. 15, no. 1, pp. 55–65, 2004.

[11] H. Shen, S. Jegelka, and A. Gretton, "Fast kernel-based independent component analysis," *IEEE Transactions on Signal Processing*, vol. 59, no. 9, pp. 3498–3511, 2009.

[12] N. Murata, S. Ikeda, and A. Ziehe, "An approach to blind source separation based on temporal structure of speech signals," *Neurocomputing*, vol. 41, no. 1-4, pp. 1–24, 2001.

[13] A. Hyvärinen, "Blind source separation by nonstationarity of variance: A cumulant-based approach," *IEEE Transactions on Neural Networks*, vol. 12, no. 6, pp. 1471–1474, 2001.

[14] J.-F. Cardoso and A. Souloumiac, "Blind beamforming for non Gaussian signals," *The IEE Proceedings of F*, vol. 140, no. 6, pp. 363–370, 1993.

[15] J.-F. Cardoso, "On the performance of orthogonal source separation algorithms," in *Proceedings of the* 9^{th} *European Signal Processing Conference*, 1994, pp. 776–779.

[16] A. Yeredor, "Non-orthogonal joint diagonalization in the least-squares sense with application in blind source separation," *IEEE Transactions on Signal Processing*, vol. 50, no. 7, pp. 1545–1553, 2002.

[17] A. Souloumiac, "Joint diagonalization: Is non-orthogonal always preferable to orthogonal?" in *Proceedings of the* 3^{rd} *IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing*, 2009, pp. 305–308.

[18] D.-T. Pham, "Joint approximate diagonalization of positive definite Hermitian matrices," *SIAM Journal on Matrix Analysis and Applications*, vol. 22, no. 4, pp. 1136–1152, 2001.

[19] A. Ziehe, P. Laskov, G. Nolte, and K.-R. Müller, "A fast algorithm for joint diagonalization with non-orthogonal transformations and its application to blind source separation," *Journal of Machine Learning Research*, vol. 5, pp. 777–800, July 2004.

[20] A. Souloumiac, "Nonorthogonal joint diagonalization by combining Givens and hyperbolic rotations," *IEEE Transactions on Signal Processing*, vol. 57, no. 6, pp. 2222–2231, 2009.

[21] D.-T. Pham and J.-F. Cardoso, "Blind separation of instantaneous mixtures of nonstationary sources," *IEEE Transactions on Signal Processing*, vol. 49, no. 9, pp. 1837–1848, 2001.

[22] B. Afsari, "Sensitivity analysis for the problem of matrix joint diagonalization," *SIAM Journal of Matrix Analysis and Application*, vol. 30, no. 3, pp. 1148–1171, 2008.

[23] K. Abed-Meraim, Y. Xiang, J. H. Manton, and Y. Hua, "Blind source sseparation using second-order cyclostationary statistics," *IEEE Transactions on Signal Processing*, vol. 49, no. 4, pp. 694–701, 2001.

[24] W.-J. Zeng, X.-L. Li, X.-D. Zhang, and X. Jiang, "An improved signal-selective direction finding algorithm using second-order cyclic statistics," in *Proceedings of the* 34^{th} *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2009, pp. 2141–2144.

[25] A. Belouchrani, K. Abed-Meraim, M. G. Amin, and A. M. Zoubir, "Blind separation of nonstationary sources," *IEEE Signal Processing Letters*, vol. 11, no. 7, pp. 605–608, 2004.

[26] L. Tong, V. C. Soon, Y. F. Huang, and R. Liu, "AMUSE: A new blind identification algorithms," in *Proceedings of IEEE International Symposium on Circuits and Systems*, vol. 3, 1990, pp. 1784–1787.

[27] L. Tong, R.-w. Liu, V. C. Soon, and Y.-F. Huang, "Indeterminacy and identifiability of blind identification," *IEEE Transactions on Circuits and Systems*, vol. 38, no. 5, pp. 499–509, 1991.

[28] A. Aïssa-El-Bey, K. Abed-Meraim, Y. Grenier, and Y. Hua, "A general framework for second-order blind separation of stationary colored sources," *Signal Processing*, vol. 88, no. 9, pp. 2123–2137, 2008.

[29] L. de Lathauwer, B. de Moor, and J. Vandewalle, "Independent component analysis and (simultaneous) third-order tensor diagonalization," *IEEE Transactions on Signal Processing*, vol. 49, no. 10, pp. 2262–2271, 2001.

[30] J.-F. Cardoso, "Source separation using higher order moments," in *Proceedings of the* 13^{th} *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 1989, pp. 2109–2112.

[31] A. K. Nandi and V. Zarzoso, "Fourth-order cumulant based blind source separation," *IEEE Signal Processing Letters*, vol. 3, no. 12, pp. 312–314, 1996.

[32] L. de Lathauwer and B. de Moor, "On the blind separation of non-circular sources," in *Proceedings of the* 11^{th} *European Signal Processing Conference*, 2002, pp. 99–102.

[33] J. Eriksson and V. Koivunen, "Complex-valued ICA using second order statistics," in *Proceedings of the* 14^{th} *IEEE International Workshop on Machine Learning for Signal Processing*, 2004, pp. 183–191.

[34] T. Trainini, X.-L. Li, E. Moreau, and T. Adalı, "A relative gradient algorithm for joint decompositions of complex matrices," in *Proceedings of the* 18^{th} *European Signal Processing Conference*, 2010, pp. 1073–1076.

[35] H. Shen and M. Kleinsteuber, "Complex blind source separation via simultaneous strong uncorrelating transform," in *Lecture Notes in Computer Science, Proceedings of the* 9^{th} *International Conference on Latent Variable Analysis and Signal Separation*, vol. 6365. Berlin/Heidelberg: Springer-Verlag, 2010, pp. 287–294.

[36] H. Shen and K. Hüper, "Block Jacobi-type methods for non-orthogonal joint diagonalisation," in *Proceedings of the* 34^{th} *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2009, pp. 3285–3288.

[37] M. B. Priestley, *Non-linear and Non-stationary Time Series Analysis.* Academic Press, 1988.

[38] P. O. Amblard, M. Gaeta, and J. L. Lacoume, "Statistics for complex variables and signals - part ii: Signals," *Signal Processing*, vol. 53, no. 1, pp. 15–25, 1996.

[39] P. Bondon, P. L. Combettes, and B. Picinbono, "Volterra filtering and higher order whiteness," *IEEE Transactions on Signal Processing*, vol. 43, no. 9, pp. 2209–2212, 1995.

[40] X.-L. Li and T. Adalı, "Blind separation of noncircular correlated sources using Gaussian entropy rate," *IEEE Transactions on Signal Processing*, vol. 59, no. 6, pp. 2969–2975, 2011.

[41] P. Comon and C. Jutten, Eds., *Handbook of Blind Source Separation: Independent Component Analysis and Applications.* Academic Press Inc., 2010.

[42] X.-L. Li, T. Adalı, and M. Anderson, "Joint blind source separation by generalized joint diagonalization of cumulant matrices," *Signal Processing*, vol. 91, no. 10, pp. 2314–2322, 2011.

[43] J. M. Mendel, "Tutorial on higher-order statistics (spectra) in signal processing and system theory: Theoretical results and some applications," *Proceedings of the IEEE*, vol. 79, no. 3, pp. 278–305, 1991.

[44] A. Belouchrani, K. A. Meraim, J.-F. Cardoso, and E. Moulines, "A blind source separation technique based on second order statistics," *IEEE Transactions on Signal Processing*, vol. 45, no. 2, pp. 434–444, 1997.

[45] L. Parra and P. Sajda, "Blind source separation via generalized eigenvalue decomposition," *The Journal of Machine Learning Research*, vol. 4, no. 7-8, pp. 1261–1269, 2004.

[46] E. Ollila and V. Koivunen, "Complex ICA using generalized uncorrelating transform," *Signal Processing*, vol. 89, no. 4, pp. 365–377, 2009.

[47] J. Eriksson and V. Koivunen, "Complex random vectors and ICA models: Identifiability, uniqueness, and separability," *IEEE Transactions on Information Theory*, vol. 52, no. 3, pp. 1017–1029, 2006.

[48] R. Benedetti and P. Cragnolini, "On simultaneous diagonalization of one Hermitian and one symmetric form," *Linear Algebra and its Applications*, vol. 57, pp. 215–226, February 1984.

[49] A. Yeredor, "Performance analysis of the strong uncorrelating transformation in blind separation of complex-valued sources," *IEEE Transactions on Signal Processing*, vol. 60, no. 1, pp. 478–483, 2012.

[50] M. Kleinsteuber, "A sort-Jacobi algorithm for semisimple Lie algebras," *Linear Algebra and its Applications*, vol. 430, no. 1, pp. 155–173, 2009.

[51] H. Shen and M. Kleinsteuber, "Algebraic solutions to complex blind source separation," in *Proceedings of the* 10^{th} *International Conference on Latent Variable Analysis and Signal Separation*, ser. Lecture Notes in Computer Science, F. Theis, A. Cichocki, A. Yeredor, and M. Zibulevsky, Eds., vol. 7191. Springer Berlin/Heidelberg, 2012, pp. 74–81.

[52] X. Huang, H.-C. Wu, and J. C. Principe, "Robust blind beamforming algorithm using joint multiple matrix diagonalization," *IEEE Sensors Journal*, vol. 7, no. 1, pp. 130–136, 2007.

[53] W.-J. Zeng, X.-L. Li, and X.-D. Zhang, "Direction-of-arrival estimation based on the joint diagonalization structure of multiple fourth-order cumulant matrices," *IEEE Signal Processing Letters*, vol. 16, no. 3, pp. 164–167, 2009.