Separation Theorem for $\mathbb{K}$-Independent Subspace Analysis with Sufficient Conditions

Zoltán Szabó, Barnabás Póczos, and András Lőrincz

Department of Information Systems, Eötvös Loránd University,
Pázmány P. sétány 1/C, Budapest H-1117, Hungary
Research Group on Intelligent Information Systems
Hungarian Academy of Sciences
szzoli@cs.elte.hu, pbarn@cs.elte.hu, lorincz@inf.elte.hu,
WWW home page: http://nipg.inf.elte.hu

Abstract. Here, a Separation Theorem about $\mathbb{K}$-Independent Subspace Analysis ($\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$ real or complex), a generalization of $\mathbb{K}$-Independent Component Analysis ($\mathbb{K}$-ICA) is proven. According to the theorem, $\mathbb{K}$-ISA estimation can be executed in two steps under certain conditions. In the first step, 1-dimensional $\mathbb{K}$-ICA estimation is executed. In the second step, optimal permutation of the $\mathbb{K}$-ICA elements is searched for. We present sufficient conditions for the $\mathbb{K}$-ISA Separation Theorem. Namely, we shall show that (i) spherically symmetric sources (both for real and complex cases), as well as (ii) real 2-dimensional sources invariant to 90° rotation, among others, satisfy the conditions of the theorem.

1 Introduction

(Real) Independent Component Analysis ($\mathbb{R}$-ICA) [1,2] aims to recover linearly or non-linearly mixed independent and hidden sources. There is a broad range of applications for $\mathbb{R}$-ICA, such as blind source separation and blind source deconvolution [3], feature extraction [4], denoising [5]. Particular applications include, e.g., the analysis of financial data [6], data from neurobiology, fMRI, EEG, and MEG (see, e.g., [7,8] and references therein). For a recent review on $\mathbb{R}$-ICA see [9].

Original $\mathbb{R}$-ICA algorithms are 1-dimensional in the sense that all sources are assumed to be independent real valued random variables. However, applications where not all, but only certain groups of the sources are independent may have high relevance in practice. In this case, independent sources can be multi-dimensional. For example, consider the generalization of the cocktail-party problem, where independent groups of people are talking about independent topics, or that more than one group of musicians are playing at the party. The separation task requires an extension of $\mathbb{R}$-ICA, which can be called (Real) Independent Subspace Analysis ($\mathbb{R}$-ISA) [10], Multi-Dimensional Independent Component Analysis (MICA) [11], Group ICA [12], and Independent Vector Analysis (IVA) [13]. Throughout the paper, we shall use the first abbreviation. An important application for $\mathbb{R}$-ISA is, e.g., the processing of EEG-fMRI data [14].

Efforts have been made to develop $\mathbb{R}$-ISA algorithms [11,14,15,16,17,18,12]. Related theoretical problems concern mostly the estimation of entropy or mutual information. In this context, entropy estimation by Edgeworth expansion [14] has been extended to more than 2 dimensions and has been used for clustering and mutual information testing [19]. k-nearest neighbors and geodesic spanning tree methods have been applied in [17] and [18] for the $\mathbb{R}$-ISA problem. Other recent approaches search for independent subspaces via kernel methods [16] and joint block diagonalization [12].

Beyond the case of real numbers, the search for complex components (Complex Independent Component Analysis, $\mathbb{C}$-ICA) assumes more and more practical relevance. Such problems include, beyond others, (i) communication systems, (ii) biomedical signal processing, e.g., processing of (f)MRI and EEG data, brain modelling, (iii) radar applications, (iv) frequency domain methods (e.g., convolutive models).

There is a large number of existing $\mathbb{C}$-ICA procedures [20,21,22,23,2,24,25,26,27,28,29,30,31,32,33,34,35,36]. Maximum likelihood principle and complex recurrent neural network are used in [20], and [21], respectively. The APEX algorithm is based on Hebbian learning [22]. Complex FastICA algorithm can be found in [23]. More solutions are based on cumulants: e.g., [2], the JADE algorithm [24,25], its higher order variants [26], and the EASI algorithm family [27]. 'Rigid-body' learning theory is used in [28]. The SOBI algorithm [29] searches for joint diagonalizer matrix, its refined version, the WASOBI method [30] approximates by means of weighted nonlinear least squares. There are complex variants of the infomax technique, such as the a split-complex [31] and the fully-complex infomax [32] procedures. Minimax Mutual Information [33] and strong-uncorrelating transforms [34,35,36] make further promising directions.

An important observation of previous computer studies [11,37] is that general $\mathbb{R}$-ISA solver algorithms are not more efficient, in fact, sometimes produce lower quality results than simple $\mathbb{R}$-ICA algorithm superimposed with searches for the optimal permutation of the components. This observation led to the present theoretical work and to some computer studies that have been published elsewhere [38,39]. We treat both the real and the complex cases.

This technical report is constructed as follows: Section 2 introduces complex random variables. In Section 3 the $\mathbb{K}$-ISA task is described. Section 4 contains our Separation Theorem for the $\mathbb{K}$-ISA task. Sufficient conditions for the theorem are provided in Section 5. Conclusions are drawn in Section 6.

2 Basic Concepts: Matrices, Complex Random Variables

We introduce the basic concepts for using complex random variables. Excellent review can be found in [40].

$\boldsymbol{B}^T$ is the transposed of matrix $\boldsymbol{B} \in \mathbb{C}^{L\times L}$. Complex conjugation is denoted by a bar it concerns all elements of a matrix. The transposed complex conjugate of matrix $\boldsymbol{B}$ is the adjoint matrix $\boldsymbol{B}^* = \bar{\boldsymbol{B}}^T$. Matrix $\boldsymbol{B} \in \mathbb{C}^{L\times L}$ is called *unitary* if $\boldsymbol{B}\boldsymbol{B}^* = \boldsymbol{I}_L$, *orthogonal* if $\boldsymbol{B}\boldsymbol{B}^T = \boldsymbol{I}_L$, where $\boldsymbol{I}_L$ is the L-dimensional identity matrix. The sets of $L \times L$ dimensional unitary and orthogonal matrices are denoted by $\mathcal{U}^D$ and $\mathcal{O}^L$, respectively.

A *complex-valued random variable* $\boldsymbol{u} \in \mathbb{C}^L$ (shortly complex random variable) is defined as a random variable of the form $\boldsymbol{u} = \boldsymbol{u}_R + i\boldsymbol{u}_I$, where the real and imaginary parts of $\boldsymbol{u}$, i.e., $\boldsymbol{u}_R$ and $\boldsymbol{u}_I \in \mathbb{R}^L$ are real random variables, $i = \sqrt{-1}$. Expectation value of complex random variables is $E[\boldsymbol{u}] = E[\boldsymbol{u}_R] + iE[\boldsymbol{u}_I]$, and the variable can be characterized in second order by its *covariance matrix* $cov[\boldsymbol{u}] = E[(\boldsymbol{u} - E[\boldsymbol{u}])(\boldsymbol{u} - E[\boldsymbol{u}])^*]$ and by its *pseudo-covariance matrix* $pcov[\boldsymbol{u}] = E[(\boldsymbol{u} - E[\boldsymbol{u}])(\boldsymbol{u} - E[\boldsymbol{u}])^T]$. Complex random variable $\boldsymbol{u}$ is called *full*, if $cov[u]$ is positive definite. Throughout this paper all complex variables are assumed to be full (that is, they are not concentrated in any lower dimensional complex subspace).

3 The $\mathbb{K}$-ISA Model

First, Section 3.1 introduces the $\mathbb{K}$-ISA task. Section 3.2 is about the ambiguities of the problem. Section 3.3 defines the entropy based cost function of the $\mathbb{K}$-ISA task.

3.1 The $\mathbb{K}$-ISA Equations

We define the complex (and real) ISA task. The two models are treated together by using the joint notation $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$. Assume that we have M hidden and independent *components* (random variables) and that only the mixture of them is available for observation:

$$\boldsymbol{z}(t) = \boldsymbol{A}\boldsymbol{s}(t), \tag{1}$$

where $\boldsymbol{s}(t) = \left[\boldsymbol{s}^1(t); \ldots; \boldsymbol{s}^M(t)\right]$ is a vector concatenated of components $\boldsymbol{s}^m(t) \in \mathbb{K}^d$. For fix m, $\boldsymbol{s}^m(t)$ is i.i.d. (independent identically distributed) in t, $\boldsymbol{s}^i$ is independent from $\boldsymbol{s}^j$, if $i \neq j$. The total dimension of the components is $D := Md$. Thus, $\boldsymbol{s}(t), \boldsymbol{z}(t) \in \mathbb{K}^D$. Matrix $\boldsymbol{A} \in \mathbb{K}^{D\times D}$ is the so called *mixing matrix* which, according to our assumptions, is invertible.

The goal of the $\mathbb{K}$-ISA problem is to estimate the original source $\boldsymbol{s}(t)$ and the unknown mixing matrix $\boldsymbol{A}$ (or its inverse $\boldsymbol{W}$, which is called the *separation matrix*) by using observations $\boldsymbol{z}(t)$ only. We talk about complex ISA task if $\mathbb{K} = \mathbb{C}$ ($\mathbb{C}$-ISA), and real ISA task if $\mathbb{K} = \mathbb{R}$ ($\mathbb{R}$-ISA). If $d = 1$, then complex ICA ($\mathbb{C}$-ICA), and real ICA ($\mathbb{R}$-ICA) tasks are obtained.

3.2 Ambiguities of the $\mathbb{K}$-ISA Model

Identification of the $\mathbb{K}$-ISA model is ambiguous. However, there are obvious ambiguities of the model: hidden components can be determined up to permutation of subspaces and invertible transformation within the subspaces. Further details concerning ambiguities can be found here: $\mathbb{R}$-ICA [41], $\mathbb{R}$-ISA [42], $\mathbb{C}$-ICA [41,40,42,43], $\mathbb{C}$-ISA (see Appendix A).

Ambiguities within subspaces can be lessened. Namely, given our assumption on the invertibility of matrix $\boldsymbol{A}$, we can assume without any loss of generality that both the sources and the observation are *white*, that is,

$$E[\boldsymbol{s}] = \boldsymbol{0}, cov\,[\boldsymbol{s}] = \boldsymbol{I}_D, \tag{2}$$

$$E[\boldsymbol{z}] = \boldsymbol{0}, cov\,[\boldsymbol{z}] = \boldsymbol{I}_D. \tag{3}$$

Below, we treat real and complex cases separately:

Real case: It then follows that the mixing matrix $\boldsymbol{A}$ and thus the separation matrix $\boldsymbol{W} = \boldsymbol{A}^{-1}$ are orthogonal:

$$\boldsymbol{I}_D = cov[\boldsymbol{z}] = E\left[\boldsymbol{z}\boldsymbol{z}^T\right] = \boldsymbol{A}E\left[\boldsymbol{s}\boldsymbol{s}^T\right]\boldsymbol{A}^T = \boldsymbol{A}\boldsymbol{I}_D\boldsymbol{A}^T = \boldsymbol{A}\boldsymbol{A}^T. \tag{4}$$

The ambiguity of the ISA task is decreased by Eqs. (2)–(3): Now, $\boldsymbol{s}^m$ sources are determined up to permutation *and* orthogonal transformation.

Complex case: It then follows that the mixing matrix $\boldsymbol{A}$ and thus the separation matrix $\boldsymbol{W} = \boldsymbol{A}^{-1}$ are unitary:

$$\boldsymbol{I}_D = cov[\boldsymbol{z}] = E\left[\boldsymbol{z}\boldsymbol{z}^*\right] = \boldsymbol{A}E\left[\boldsymbol{s}\boldsymbol{s}^*\right]\boldsymbol{A}^* = \boldsymbol{A}\boldsymbol{I}_D\boldsymbol{A}^* = \boldsymbol{A}\boldsymbol{A}^*. \tag{5}$$

Thus, components $\boldsymbol{s}^m$ are determined up to permutation *and* unitary transformation within the subspace.

3.3 The $\mathbb{K}$-ISA Cost Function

Now we sketch how to transcribe the $\mathbb{K}$-ISA task into the minimization of sum of multi-dimensional entropies for orthogonal matrices (in the real case) and for unitary matrices (in the complex case). We shall use these formulations of the $\mathbb{K}$-ISA task to prove the real and complex versions of the Separation Theorem (Section 4).

3.3.1 Real Case The $\mathbb{R}$-ISA task can be viewed as the minimization of mutual information between the estimated components:

$$I\left(\boldsymbol{y}^1,\ldots,\boldsymbol{y}^M\right) := \int f(\boldsymbol{v})\ln\left[\frac{f(\boldsymbol{v})}{\prod_{i=1}^M f_m(v_m)}\right]\mathrm{d}\boldsymbol{v} \tag{6}$$

on the orthogonal group ($\boldsymbol{W} \in \mathcal{O}^D$), where $\boldsymbol{y} = \boldsymbol{W}\boldsymbol{z}$, $\boldsymbol{y} = \left[\boldsymbol{y}^1;\ldots;\boldsymbol{y}^M\right]$, f and f_m are density functions of $\boldsymbol{y}$ and marginals $\boldsymbol{y}^m$, respectively. This cost function I is equivalent to the minimization of the sum of d-dimensional entropies, because

$$I\left(\boldsymbol{y}^1,\ldots,\boldsymbol{y}^M\right) = \sum_{m=1}^M H\left(\boldsymbol{y}^m\right) - H(\boldsymbol{y}) \tag{7}$$

$$= \sum_{m=1}^M H\left(\boldsymbol{y}^m\right) - H(\boldsymbol{W}\boldsymbol{z}) \tag{8}$$

$$= \sum_{m=1}^M H\left(\boldsymbol{y}^m\right) - [H(\boldsymbol{z}) + \ln(|\det(\boldsymbol{W})|)]. \tag{9}$$

Here, H is Shannon's (multi-dimensional) differential entropy defined with logarithm of base e, $|\cdot|$ denotes absolute value, 'det' stands for determinant. In the second equality, the $\boldsymbol{y} = \boldsymbol{W}\boldsymbol{z}$ relation was exploited, and the

$$H(\boldsymbol{W}\boldsymbol{z}) = H(\boldsymbol{z}) + \ln\left(|\det(\boldsymbol{W})|\right) \tag{10}$$

rule describing transformation of the differential entropy [44] was used. $\det(\boldsymbol{W}) = 1$ because of the orthogonality of $\boldsymbol{W}$, so $\ln(|\det(\boldsymbol{W})|) = 0$. The $H(\boldsymbol{z})$ term of the cost is constant in $\boldsymbol{W}$, therefore the $\mathbb{R}$-ISA task is equivalent to the minimization of the cost function

$$J(\boldsymbol{W}) := \sum_{m=1}^M H\left(\boldsymbol{y}^m\right) \to \min_{\boldsymbol{W}\in\mathcal{O}^D}. \tag{11}$$

3.3.2 Complex Case Similarly, the $\mathbb{C}$-ISA task can be viewed as the minimization of mutual information between the estimated components [see Eq. (6)], but on the unitary group ($\boldsymbol{W} \in \mathcal{U}^D$). Here, the Shannon entropy of random variable $\mathbb{C}^L \ni \boldsymbol{u}$ ($\boldsymbol{y}^m$, or $\boldsymbol{y}$) is the entropy of $\varphi_v(\boldsymbol{u}) \in \mathbb{R}^{2L}$, where

$$\varphi_v : \mathbb{C}^L \ni \boldsymbol{u} \mapsto \boldsymbol{u} \otimes \begin{bmatrix}\Re(\cdot)\\ \Im(\cdot)\end{bmatrix} \in \mathbb{R}^{2L}. \tag{12}$$

That is, $H(\boldsymbol{u}) := H[\varphi_v(\boldsymbol{u})]$. Here: $\otimes$ is the Kronecker product, $\Re$ stands for the real part, $\Im$ for the imaginary part, subscript 'v' for vector. One can neglect the last term of the $H(\boldsymbol{y})$ cost function [see, Eq. (7)] during optimization (alike in the real case). To see this, consider the mapping

$$\varphi_M : \mathbb{C}^{L\times L} \ni \boldsymbol{M} \mapsto \boldsymbol{M} \otimes \begin{bmatrix} \Re(\cdot) & -\Im(\cdot) \\ \Im(\cdot) & \Re(\cdot) \end{bmatrix} \in \mathbb{R}^{2L\times 2L}, \tag{13}$$

where subscript 'M' indicate matrices. Known properties of mappings φ_v, φ_M are as follows [45]:

$$\det[\varphi_M(\boldsymbol{M})] = |\det(\boldsymbol{M})|^2, \tag{14}$$

$$\varphi_v(\boldsymbol{M}\boldsymbol{v}) = \varphi_M(\boldsymbol{M})\varphi_v(\boldsymbol{v}). \tag{15}$$

In words: (14) describes transformation of determinant, (15) expresses preservation of operation for matrix-vector multiplication.[1] The following relation holds for the entropy transformation of complex variables:

Lemma 1 (Transformation of entropy for complex variables). *Let $\boldsymbol{u} \in \mathbb{C}^L$ denote a random variable and let $\boldsymbol{V} \in \mathbb{C}^{L\times L}$ be a matrix. Then*

$$H(\boldsymbol{V}\boldsymbol{u}) = H(\boldsymbol{u}) + \ln\left(|\det(\boldsymbol{V})|^2\right) \tag{16}$$

Proof.

$$H(\boldsymbol{V}\boldsymbol{u}) = H[\varphi_v(\boldsymbol{V}\boldsymbol{u})] = H[\varphi_M(\boldsymbol{V})\varphi_v(\boldsymbol{u})] = H[\varphi_v(\boldsymbol{u})] + \ln(|\det[\varphi_M(\boldsymbol{V})]|) = H[\varphi_v(\boldsymbol{u})] + \ln(|\det \boldsymbol{V}|^2) \tag{17}$$

$$= H(\boldsymbol{u}) + \ln(|\det \boldsymbol{V}|^2) \tag{18}$$

The above steps can be justified as follows:

1. the first equation uses the definition of entropy for complex variables,
2. then we used property (15),
3. transformed the entropy of random variables in $\mathbb{R}^{2L}$ [see, Eq. (10)].
4. exploited (14), and
5. applied the definition of entropy for complex variables again.

$\square$

Thus $H(\boldsymbol{y}) = H(\boldsymbol{W}\boldsymbol{z}) + \ln(1) = H(\boldsymbol{z})$, where unitarity of $\boldsymbol{W}$ is exploited. Further, $H(\boldsymbol{z})$ is not dependent of matrix $\boldsymbol{W}$ and thus term $H(\boldsymbol{y})$ can be neglected during the course of optimization. We conclude that the $\mathbb{C}$-ISA task can be written as the minimization of sum of multi-dimensional entropies. The cost function to be optimized within unitary matrices:

$$J(\boldsymbol{W}) = \sum_{m=1}^{M} H(\boldsymbol{y}^m) \to \min_{\boldsymbol{W}\in\mathcal{U}^D}. \tag{19}$$

4 The $\mathbb{R}$-ISA and $\mathbb{C}$-ISA Separation Theorem

The main result of this work is that the $\mathbb{K}$-ISA task may be accomplished in two steps under certain conditions. In the first step $\mathbb{K}$-ICA is executed. The second step is search for the optimal permutation of the $\mathbb{K}$-ICA components. Section 4.1 is about the real, whereas Section 4.2 is about the complex case.

4.1 The $\mathbb{R}$-ISA Separation Theorem

We shall rely on entropy inequalities (Section 4.1.1). Connection to the $\mathbb{R}$-ICA cost function is derived in Section 4.1.2. Finally, Section 4.1.3 contains the proof of our theorem.

[1] Note that this connection allows one to reduce the $\mathbb{C}$-ISA task (and thus the $\mathbb{C}$-ICA task) to a $\mathbb{R}$-ISA task directly. According to our experiences, however, methods that rely on the $\mathbb{C}$-Separation Theorem that we present here are much more efficient.

4.1.1 EPI-type Relations (Real Case)

First, consider the so called Entropy Power Inequality (EPI)

$$e^{2H\left(\sum_{i=1}^{L} u_i\right)} \geq \sum_{i=1}^{L} e^{2H(u_i)}, \tag{20}$$

where $u_1, \ldots, u_L \in \mathbb{R}$ denote continuous random variables (The name of this inequality is $\mathbb{R}$-EPI, because we shall need its complex variant later). This inequality holds for example, for independent continuous variables [44].

Let $\|\cdot\|$ denote the Euclidean norm. That is, for $\boldsymbol{w} \in \mathbb{R}^L$

$$\|\boldsymbol{w}\|^2 := \sum_{i=1}^{L} w_i^2, \tag{21}$$

where w_i is the i^{th} coordinate of vector $\boldsymbol{w}$. The surface of the L-dimensional unit sphere shall be denoted by $S^L(\mathbb{R})$:

$$S^L(\mathbb{R}) := \{\boldsymbol{w} \in \mathbb{R}^L : \|\boldsymbol{w}\| = 1\}. \tag{22}$$

If $\mathbb{R}$-EPI is satisfied [on $S^L(\mathbb{R})$] then a further inequality holds:

Lemma 2. *Suppose that continuous random variables $u_1, \ldots, u_L \in \mathbb{R}$ satisfy the following inequality*

$$e^{2H\left(\sum_{i=1}^{L} w_i u_i\right)} \geq \sum_{i=1}^{L} e^{2H(w_i u_i)}, \forall \boldsymbol{w} \in S^L(\mathbb{R}). \tag{23}$$

This inequality will be called the $\mathbb{R}$-w-EPI *condition. Then*

$$H\left(\sum_{i=1}^{L} w_i u_i\right) \geq \sum_{i=1}^{L} w_i^2 H\left(u_i\right), \forall \boldsymbol{w} \in S^L(\mathbb{R}). \tag{24}$$

Note 1. $\mathbb{R}$-w-EPI holds, for example, for independent variables u_i, because independence is not affected by multiplication with a constant.

Proof. Assume that $\boldsymbol{w} \in S^L(\mathbb{R})$. Applying ln on condition (23), and using the monotonicity of the ln function, we can see that the first inequality is valid in the following inequality chain

$$2H\left(\sum_{i=1}^{L} w_i u_i\right) \geq \ln\left(\sum_{i=1}^{L} e^{2H(w_i u_i)}\right) = \ln\left(\sum_{i=1}^{L} e^{2H(u_i)} \cdot w_i^2\right) \geq \sum_{i=1}^{L} w_i^2 \cdot \ln\left(e^{2H(u_i)}\right) = \sum_{i=1}^{L} w_i^2 \cdot 2H(u_i). \tag{25}$$

Then,

1. we used the relation [44]:
$$H(w_i u_i) = H(u_i) + \ln\left(|w_i|\right) \tag{26}$$
for the entropy of the transformed variable. Hence
$$e^{2H(w_i u_i)} = e^{2H(u_i) + 2\ln(|w_i|)} = e^{2H(u_i)} \cdot e^{2\ln(|w_i|)} = e^{2H(u_i)} \cdot w_i^2. \tag{27}$$
2. In the second inequality, we exploited the concavity of ln. □

4.1.2 Connection to the Cost Function of the $\mathbb{R}$-ICA Task

Now we shall use Lemma 2 to proceed. The $\mathbb{R}$-ISA Separation Theorem will be a corollary of the following claim:

Proposition 1. *Let $\boldsymbol{y} = \left[\boldsymbol{y}^1; \ldots; \boldsymbol{y}^M\right] = \boldsymbol{y}(\boldsymbol{W}) = \boldsymbol{W}\boldsymbol{s}$, where $\boldsymbol{W} \in \mathcal{O}^D$, $\boldsymbol{y}^m$ is the estimation of the m^{th} component of the $\mathbb{R}$-ISA task. Let y_i^m be the i^{th} coordinate of the m^{th} component. Similarly, let s_i^m stand for the i^{th} coordinate of the m^{th} source. Let us assume that the $\boldsymbol{s}^m$ sources satisfy condition (24). Then*

$$\sum_{m=1}^{M} \sum_{i=1}^{d} H\left(y_i^m\right) \geq \sum_{m=1}^{M} \sum_{i=1}^{d} H\left(s_i^m\right). \tag{28}$$

Proof. Let us denote the $(i,j)^{th}$ element of matrix $\boldsymbol{W}$ by $W_{i,j}$. Coordinates of $\boldsymbol{y}$ and $\boldsymbol{s}$ will be denoted by y_i and s_i, respectively. Further, let $\mathcal{G}^1,\ldots,\mathcal{G}^M$ denote the indices of the $1^{st},\ldots,M^{th}$ subspaces, i.e., $\mathcal{G}^1 := \{1,\ldots,d\},\ldots,\mathcal{G}^M := \{D-d+1,\ldots,D\}$. Now, writing the elements of the i^{th} row of matrix multiplication $\boldsymbol{y}=\boldsymbol{W}\boldsymbol{s}$, we have

$$y_i = \sum_{j\in\mathcal{G}^1} W_{i,j}s_j + \ldots + \sum_{j\in\mathcal{G}^M} W_{i,j}s_j \tag{29}$$

and thus,

$$\begin{aligned}
&H(y_i) = \\
&= H\left(\sum_{j\in\mathcal{G}^1} W_{i,j}s_j + \ldots + \sum_{j\in\mathcal{G}^M} W_{i,j}s_j\right) &(30)\\
&= H\left(\left(\sum_{l\in\mathcal{G}^1} W_{i,l}^2\right)^{\frac{1}{2}} \frac{\sum_{j\in\mathcal{G}^1} W_{i,j}s_j}{\left(\sum_{l\in\mathcal{G}^1} W_{i,l}^2\right)^{\frac{1}{2}}} + \ldots + \left(\sum_{l\in\mathcal{G}^M} W_{i,l}^2\right)^{\frac{1}{2}} \frac{\sum_{j\in\mathcal{G}^M} W_{i,j}s_j}{\left(\sum_{l\in\mathcal{G}^M} W_{i,l}^2\right)^{\frac{1}{2}}}\right) &(31)\\
&\geq \left(\sum_{l\in\mathcal{G}^1} W_{i,l}^2\right) H\left(\frac{\sum_{j\in\mathcal{G}^1} W_{i,j}s_j}{\left(\sum_{l\in\mathcal{G}^1} W_{i,l}^2\right)^{\frac{1}{2}}}\right) + \ldots + \left(\sum_{l\in\mathcal{G}^M} W_{i,l}^2\right) H\left(\frac{\sum_{j\in\mathcal{G}^M} W_{i,j}s_j}{\left(\sum_{l\in\mathcal{G}^M} W_{i,l}^2\right)^{\frac{1}{2}}}\right) &(32)\\
&= \left(\sum_{l\in\mathcal{G}^1} W_{i,l}^2\right) H\left(\sum_{j\in\mathcal{G}^1} \frac{W_{i,j}}{\left(\sum_{l\in\mathcal{G}^1} W_{i,l}^2\right)^{\frac{1}{2}}} s_j\right) + \ldots + \left(\sum_{l\in\mathcal{G}^M} W_{i,l}^2\right) H\left(\sum_{j\in\mathcal{G}^M} \frac{W_{i,j}}{\left(\sum_{l\in\mathcal{G}^M} W_{i,l}^2\right)^{\frac{1}{2}}} s_j\right) &(33)\\
&\geq \left(\sum_{l\in\mathcal{G}^1} W_{i,l}^2\right) \sum_{j\in\mathcal{G}^1} \left(\frac{W_{i,j}}{\left(\sum_{l\in\mathcal{G}^1} W_{i,l}^2\right)^{\frac{1}{2}}}\right)^2 H(s_j) + \ldots + \left(\sum_{l\in\mathcal{G}^M} W_{i,l}^2\right) \sum_{j\in\mathcal{G}^M} \left(\frac{W_{i,j}}{\left(\sum_{l\in\mathcal{G}^M} W_{i,l}^2\right)^{\frac{1}{2}}}\right)^2 H(s_j) &(34)\\
&= \sum_{j\in\mathcal{G}^1} W_{i,j}^2 H(s_j) + \ldots + \sum_{j\in\mathcal{G}^M} W_{i,j}^2 H(s_j) &(35)
\end{aligned}$$

The above steps can be justified as follows:

1. (30): Eq. (29) was inserted into the argument of H.
2. (31): New terms were added for Lemma 2.
3. (32): Sources $\boldsymbol{s}^m$ are independent of each other and this independence is preserved upon mixing *within* the subspaces, and we could also use Lemma 2, because $\boldsymbol{W}$ is an orthogonal matrix.
4. (33): Nominators were transferred into the $\sum_j$ terms.
5. (34): Variables $\boldsymbol{s}^m$ satisfy condition (24) according to our assumptions.
6. (35): We simplified the expression after squaring.

Using this inequality, summing it for i, exchanging the order of the sums, and making use of the orthogonality of matrix $\boldsymbol{W}$, we have

$$\begin{aligned}
\sum_{i=1}^{D} H(y_i) &\geq \sum_{i=1}^{D}\left(\sum_{j\in\mathcal{G}^1} W_{i,j}^2 H(s_j) + \ldots + \sum_{j\in\mathcal{G}^M} W_{i,j}^2 H(s_j)\right) &(36)\\
&= \sum_{j\in\mathcal{G}^1}\left(\sum_{i=1}^{D} W_{i,j}^2\right) H(s_j) + \ldots + \sum_{j\in\mathcal{G}^M}\left(\sum_{i=1}^{D} W_{i,j}^2\right) H(s_j) &(37)\\
&= \sum_{j=1}^{D} H(s_j). &(38)
\end{aligned}$$

□

Note 2. The proof holds for components with different dimensions. This is also true for the following theorem.

4.1.3 Proof of the $\mathbb{R}$-ISA Separation Theorem

Having this proposition, now we present our main theorem.

Theorem 1 (Separation Theorem for $\mathbb{R}$-ISA). *Presume that the $\boldsymbol{s}^m$ sources of the $\mathbb{R}$-ISA model satisfy condition (24), and that the $\mathbb{R}$-ICA cost function $J(\boldsymbol{W}) = \sum_{m=1}^{M}\sum_{i=1}^{d} H(y_i^m)$ has minimum ($\boldsymbol{W} \in \mathcal{O}^D$). Then it is sufficient to search for the minimum of the $\mathbb{R}$-ISA task as a permutation of the solution of the $\mathbb{R}$-ICA task. Using the concept of separation matrices, it is sufficient to explore forms*

$$\boldsymbol{W}_{\mathbb{R}\text{-}ISA} = \boldsymbol{P}\boldsymbol{W}_{\mathbb{R}\text{-}ICA}, \tag{39}$$

where $\boldsymbol{P}\left(\in \mathbb{R}^{D\times D}\right)$ is a permutation matrix to be determined.

Proof. $\mathbb{R}$-ICA minimizes the l.h.s. of Eq. (28), that is, it minimizes $\sum_{m=1}^{M}\sum_{i=1}^{d} H\left(y_i^m\right)$. The set of minima is invariant to permutations and to changes of the signs. Also, according to Proposition 1, $\{s_i^m\}$, i.e., the coordinates of the $\boldsymbol{s}^m$ components of the $\mathbb{R}$-ISA task belong to the set of the minima. □

4.2 The $\mathbb{C}$-ISA Separation Theorem

The proof of the complex case is similar to the proof of the real case. The difference is in the EPI-type relations that we apply. Procedure: We define a $\mathbb{C}$-EPI property and then a $\mathbb{C}$-w-EPI relation starting from the vector variant of the $\mathbb{R}$-EPI relation. Then the proof relies on analogous steps with the real case, that we detail here for the sake of completeness.

4.2.1 EPI-type Relations (Complex Case)

Let us consider the vector variant of the $\mathbb{R}$-EPI relation.

Lemma 3 (vector-EPI). *For independent (finite covariance random variables) $\boldsymbol{u}_1,\dots,\boldsymbol{u}_L \in \mathbb{R}^q$ holds [46] that*

$$e^{2H\left(\sum_{i=1}^{L}\boldsymbol{u}_i\right)/q} \geq \sum_{i=1}^{L} e^{2H(\boldsymbol{u}_i)/q}. \tag{40}$$

Let us define a similar property for complex random variables:

Definition 1 ($\mathbb{C}$-EPI). *We say that random variables $u_1,\dots,u_L \in \mathbb{C}$ satisfy relation $\mathbb{C}$-EPI if*

$$e^{H\left(\sum_{i=1}^{L} u_i\right)} \geq \sum_{i=1}^{L} e^{H(u_i)}. \tag{41}$$

Note 3. This holds for independent random variables $u_1,\dots,u_L \in \mathbb{C}$, because according to vector-EPI ($q=2$)

$$e^{2H\left(\sum_{i=1}^{L} u_i\right)/2} \geq \sum_{i=1}^{L} e^{2H(u_i)/2}. \tag{42}$$

We need to following lemma:

Lemma 4. *Let us assume that random variables $u_1,\dots,u_L \in \mathbb{C}$ satisfy condition*

$$e^{H(\sum_{i=1}^{L} w_i u_i)} \geq \sum_{i=1}^{L} e^{H(w_i u_i)} \quad \forall \boldsymbol{w} = [w_1;\dots;w_L] \in S^L(\mathbb{C}) \tag{43}$$

that we shall call condition $\mathbb{C}$-w-EPI. Here, $S^L(\mathbb{C})$ denotes the L-dimensional complex unit sphere, that is

$$S^L(\mathbb{C}) := \left\{\boldsymbol{w} = [w_1;\dots;w_L] \in \mathbb{C}^L : \sum_{i=1}^{L} |w_i|^2 = 1\right\}. \tag{44}$$

Then

$$H\left(\sum_{i=1}^{L} w_i u_i\right) \geq \sum_{i=1}^{L} |w_i|^2 H(u_i) \quad \forall \boldsymbol{w} \in S^L(\mathbb{C}). \tag{45}$$

Proof. Assume that $\boldsymbol{w} \in S^L(\mathbb{C})$. Applying ln on condition (43), and using the monotonicity of the ln function, we can see that the first inequality is valid in the following inequality chain

$$H\left(\sum_{i=1}^{L} w_i u_i\right) \geq \ln\left(\sum_{i=1}^{L} e^{H(w_i u_i)}\right) = \ln\left(\sum_{i=1}^{L} e^{H(u_i)} \cdot |w_i|^2\right) \geq \sum_{i=1}^{L} |w_i|^2 \cdot \ln\left(e^{H(u_i)}\right) = \sum_{i=1}^{L} |w_i|^2 \cdot H(u_i). \tag{46}$$

Then,

1. we used the relation:
$$H(wu) = H(u) + \ln\left(|w|^2\right) \quad (w, u \in \mathbb{C}) \tag{47}$$
for the entropy of the transformed variable (see Lemma 1). Hence
$$e^{H(w_i u_i)} = e^{H(u_i) + \ln\left(|w_i|^2\right)} = e^{H(u_i)} \cdot e^{\ln\left(|w_i|^2\right)} = e^{H(u_i)} \cdot |w_i|^2. \tag{48}$$
2. In the second inequality, we exploited the concavity of ln. □

4.2.2 Connection to the Cost Function of the $\mathbb{C}$-ICA Task

Now we shall use Lemma 4 to proceed. The $\mathbb{C}$-ISA Separation Theorem will be a corollary of the following claim:

Proposition 2. *Let $\boldsymbol{y} = \left[\boldsymbol{y}^1; \ldots; \boldsymbol{y}^M\right] = \boldsymbol{y}(\boldsymbol{W}) = \boldsymbol{W}\boldsymbol{s}$, where $\boldsymbol{W} \in \mathcal{U}^D$, $\boldsymbol{y}^m$ is the estimation of the m^{th} component of the $\mathbb{C}$-ISA task. Let y_i^m be the i^{th} complex coordinate of the m^{th} component. Similarly, let s_i^m stand for the i^{th} coordinate of the m^{th} source. Let us assume that the $\boldsymbol{s}^m$ sources satisfy condition (45). Then*

$$\sum_{m=1}^{M} \sum_{i=1}^{d} H\left(y_i^m\right) \geq \sum_{m=1}^{M} \sum_{i=1}^{d} H\left(s_i^m\right). \tag{49}$$

Proof. Let us denote the $(i,j)^{th}$ element of matrix $\boldsymbol{W}$ by $W_{i,j}$. Coordinates of $\boldsymbol{y}$ and $\boldsymbol{s}$ will be denoted by y_i and s_i, respectively. Let $\mathcal{G}^1, \ldots, \mathcal{G}^M$ denote the indices belonging to the $1^{st}, \ldots, M^{th}$. subspaces, that is, $\mathcal{G}^1 := \{1, \ldots, d\}, \ldots, \mathcal{G}^M := \{D-d+1, \ldots, D\}$. Now, writing the elements of the i^{th} row of matrix multiplication $\boldsymbol{y} = \boldsymbol{W}\boldsymbol{s}$, we have

$$y_i = \sum_{j \in \mathcal{G}^1} W_{i,j} s_j + \ldots + \sum_{j \in \mathcal{G}^M} W_{i,j} s_j \tag{50}$$

and thus,

$$\begin{aligned}
&H\left(y_i\right) = \\
&= H\left(\sum_{j \in \mathcal{G}^1} W_{i,j} s_j + \ldots + \sum_{j \in \mathcal{G}^M} W_{i,j} s_j\right) && (51)\\
&= H\left(\left(\sum_{l \in \mathcal{G}^1} |W_{i,l}|^2\right)^{\frac{1}{2}} \frac{\sum_{j \in \mathcal{G}^1} W_{i,j} s_j}{\left(\sum_{l \in \mathcal{G}^1} |W_{i,l}|^2\right)^{\frac{1}{2}}} + \ldots + \left(\sum_{l \in \mathcal{G}^M} |W_{i,l}|^2\right)^{\frac{1}{2}} \frac{\sum_{j \in \mathcal{G}^M} W_{i,j} s_j}{\left(\sum_{l \in \mathcal{G}^M} |W_{i,l}|^2\right)^{\frac{1}{2}}}\right) && (52)\\
&\geq \left(\sum_{l \in \mathcal{G}^1} |W_{i,l}|^2\right) H\left(\frac{\sum_{j \in \mathcal{G}^1} W_{i,j} s_j}{\left(\sum_{l \in \mathcal{G}^1} |W_{i,l}|^2\right)^{\frac{1}{2}}}\right) + \ldots + \left(\sum_{l \in \mathcal{G}^M} |W_{i,l}|^2\right) H\left(\frac{\sum_{j \in \mathcal{G}^M} W_{i,j} s_j}{\left(\sum_{l \in \mathcal{G}^M} |W_{i,l}|^2\right)^{\frac{1}{2}}}\right) && (53)\\
&= \left(\sum_{l \in \mathcal{G}^1} |W_{i,l}|^2\right) H\left(\sum_{j \in \mathcal{G}^1} \frac{W_{i,j}}{\left(\sum_{l \in \mathcal{G}^1} |W_{i,l}|^2\right)^{\frac{1}{2}}} s_j\right) + \ldots + \left(\sum_{l \in \mathcal{G}^M} |W_{i,l}|^2\right) H\left(\sum_{j \in \mathcal{G}^M} \frac{W_{i,j}}{\left(\sum_{l \in \mathcal{G}^M} |W_{i,l}|^2\right)^{\frac{1}{2}}} s_j\right) && (54)\\
&\geq \left(\sum_{l \in \mathcal{G}^1} |W_{i,l}|^2\right) \sum_{j \in \mathcal{G}^1} \left|\frac{W_{i,j}}{\left(\sum_{l \in \mathcal{G}^1} |W_{i,l}|^2\right)^{\frac{1}{2}}}\right|^2 H\left(s_j\right) + \ldots + \left(\sum_{l \in \mathcal{G}^M} |W_{i,l}|^2\right) \sum_{j \in \mathcal{G}^M} \left|\frac{W_{i,j}}{\left(\sum_{l \in \mathcal{G}^M} |W_{i,l}|^2\right)^{\frac{1}{2}}}\right|^2 H\left(s_j\right) && (55)\\
&= \sum_{j \in \mathcal{G}^1} |W_{i,j}|^2 H\left(s_j\right) + \ldots + \sum_{j \in \mathcal{G}^M} |W_{i,j}|^2 H\left(s_j\right) && (56)
\end{aligned}$$

The above steps can be justified as follows:

1. (51): Eq. (50) was inserted into the argument of H.
2. (52): New terms were added for Lemma 4.
3. (53): Sources $\boldsymbol{s}^m$ are independent of each other and this independence is preserved upon mixing *within* the subspaces, and we could also use Lemma 4, because $\boldsymbol{W}$ is a unitary matrix.
4. (54): Nominators were transferred into the $\sum_j$ terms.
5. (55): Variables $\boldsymbol{s}^m$ satisfy condition (45) according to our assumptions.
6. (56): We simplified the expression after squaring.

Using this inequality, summing it for i, exchanging the order of the sums, and making use of the unitary property of matrix $\boldsymbol{W}$, we have

$$\sum_{i=1}^{D} H(y_i) \geq \sum_{i=1}^{D} \left(\sum_{j\in\mathcal{G}^1} |W_{i,j}|^2 H(s_j) + \ldots + \sum_{j\in\mathcal{G}^M} |W_{i,j}|^2 H(s_j) \right) \tag{57}$$

$$= \sum_{j\in\mathcal{G}^1} \left(\sum_{i=1}^{D} |W_{i,j}|^2 \right) H(s_j) + \ldots + \sum_{j\in\mathcal{G}^M} \left(\sum_{i=1}^{D} |W_{i,j}|^2 \right) H(s_j) \tag{58}$$

$$= \sum_{j=1}^{D} H(s_j). \tag{59}$$

□

Note 4. The proof of the proposition is similar when the dimensions of the subspaces are not constrained to be equal. The situation is the same in the next theorem.

4.2.3 Proof of the $\mathbb{C}$-ISA Separation Theorem

Having this proposition, now we present our main theorem.

Theorem 2 (Separation Theorem for $\mathbb{C}$-ISA). *Presume that the $\boldsymbol{s}^m$ sources of the $\mathbb{C}$-ISA model satisfy condition (45), and that $J(\boldsymbol{W}) = \sum_{m=1}^{M}\sum_{i=1}^{d} H(y_i^m), (\boldsymbol{W} \in \mathcal{U}^D)$, i.e., the $\mathbb{C}$-ICA cost function has minimum. Then it is sufficient to search for the minimum of the $\mathbb{C}$-ISA task ($\boldsymbol{W}_{\mathbb{C}\text{-}ISA}$) as a permutation of the solution of the $\mathbb{C}$-ICA task ($\boldsymbol{W}_{\mathbb{C}\text{-}ICA}$). That it, it is sufficient to search in the form*

$$\boldsymbol{W}_{\mathbb{C}\text{-}ISA} = \boldsymbol{P}\boldsymbol{W}_{\mathbb{C}\text{-}ICA}, \tag{60}$$

where $\boldsymbol{P} \left(\in \mathbb{R}^{D\times D}\right)$ is the permutation matrix to be determined.

Proof. $\mathbb{C}$-ICA minimizes the l.h.s. of Eq. (49), that is, it minimizes $\sum_{m=1}^{M}\sum_{i=1}^{d} H(y_i^m)$. The set of minima is invariant for permutations and for multiplication of the coordinates by numbers with unit absolute value, and according to Proposition 2 $\{s_i^m\}$ (i.e., the coordinates of the $\mathbb{C}$-ISA task) is among the minima.

We can disregard multiplications with unit absolute values, because unitary ambiguity within subspaces are present in the $\mathbb{C}$-ISA task. □

5 Sufficient Conditions of the Separation Theorem

In the Separation Theorem, we assumed that relations (24) and (45) are fulfilled for the $\boldsymbol{s}^m$ sources in the real and complex cases, respectively. Here, we shall provide sufficient conditions when these inequalities are fulfilled.

5.1 Real Case

5.1.1 $\mathbb{R}$-w-EPI

According to Lemma 2, if the $\mathbb{R}$-w-EPI property [i.e., (23)] holds for sources $\boldsymbol{s}^m$, then inequality (24) holds, too.

5.1.2 Real Spherically Symmetric Sources

Definition 2 (real spherically symmetric variable). *A random variable $\boldsymbol{u} \in \mathbb{R}^d$ is called real spherically symmetric (or shortly $\mathbb{R}$-spherical), if its density function is not modified by any rotation. Formally, if*

$$\boldsymbol{u} \stackrel{\text{distr}}{=} \boldsymbol{O}\boldsymbol{u}, \quad \forall \boldsymbol{O} \in \mathcal{O}^d, \tag{61}$$

where $\stackrel{\text{distr}}{=}$ denotes equality in distribution.

A $\mathbb{R}$-spherical random variable has a density function (under mild conditions) and this density function takes constant values on concentric spheres around the origin. We shall make use of the following well-known properties of spherically symmetric variables [47,48]:

Lemma 5 (Identical distribution of 1-dimensional projections - Real case). *Let $\boldsymbol{v}$ denote a d-dimensional variable, which is $\mathbb{R}$-spherically symmetric. Then the projection of $\boldsymbol{v}$ onto lines through the origin have identical univariate distribution.*

Lemma 6 (Momenta - Real case). *The expectation value and the variance of a d-dimensional $\boldsymbol{v}$ $\mathbb{R}$-spherically symmetric variable are*

$$E[\boldsymbol{v}] = \boldsymbol{0}, \tag{62}$$

$$cov[\boldsymbol{v}] = c(onstant) \cdot \boldsymbol{I}_d. \tag{63}$$

Now we are ready to claim the following theorem.

Proposition 3. *For spherically symmetric sources $\boldsymbol{s}^m$ $(m = 1, \ldots, M)$ with finite covariance Eq. (24) holds. Further, the stronger $\mathbb{R}$-w-EPI property [Eq. (23)] also holds and with equality between the two sides [$\forall \boldsymbol{w} \in S^d(\mathbb{R})$].*

Proof. Here, we show that the $\mathbb{R}$-w-EPI property is fulfilled with equality for $\mathbb{R}$-spherical sources. According to (62)–(63), spherically symmetric sources $\boldsymbol{s}^m$ have zero expectation values and up to a constant multiplier they also have identity covariance matrices:

$$E[\boldsymbol{s}^m] = \boldsymbol{0}, \tag{64}$$

$$cov[\boldsymbol{s}^m] = c^m \cdot \boldsymbol{I}_d. \tag{65}$$

Note that our constraint on the $\mathbb{R}$-ISA task, namely that covariance matrices of the $\boldsymbol{s}^m$ sources should be equal to $\boldsymbol{I}_d$, is fulfilled up to constant multipliers.

Let $P_{\boldsymbol{w}}$ denote the projection to straight line with direction $\boldsymbol{w} \in S^d(\mathbb{R})$, which crosses the origin, i.e.,

$$P_{\boldsymbol{w}} : \mathbb{R}^d \ni \boldsymbol{u} \mapsto \sum_{i=1}^{d} w_i u_i \in \mathbb{R}. \tag{66}$$

In particular, if $\boldsymbol{w}$ is chosen as the canonical basis vector $\boldsymbol{e}_i$ (all components are 0, except the i^{th} component, which is equal to 1), then

$$P_{\boldsymbol{e_i}}(\boldsymbol{u}) = u_i. \tag{67}$$

In this interpretation $\mathbb{R}$-w-EPI [see Eq. (23)] is concerned with the entropies of the projections of the different sources onto straight lines crossing the origin. The l.h.s. projects to $\boldsymbol{w}$, whereas the r.h.s. projects to the canonical basis vectors. Let $\boldsymbol{u}$ denote an arbitrary source, i.e., $\boldsymbol{u} := \boldsymbol{s}^m$. According to Lemma 5, distribution of the spherical $\boldsymbol{u}$ is the same for all such projections and thus its entropy is identical. That is,

$$\sum_{i=1}^{d} w_i u_i \stackrel{\text{distr}}{=} u_1 \stackrel{\text{distr}}{=} \ldots \stackrel{\text{distr}}{=} u_d, \quad \forall \boldsymbol{w} \in S^d(\mathbb{R}), \tag{68}$$

$$H\left(\sum_{i=1}^{d} w_i u_i\right) = H(u_1) = \ldots = H(u_d), \quad \forall \boldsymbol{w} \in S^d(\mathbb{R}). \tag{69}$$

Thus:

– l.h.s. of $\mathbb{R}$-w-EPI: $e^{2H(u_1)}$.
– r.h.s. of $\mathbb{R}$-w-EPI:

$$\sum_{i=1}^{d} e^{2H(w_i u_i)} = \sum_{i=1}^{d} e^{2H(u_i)} \cdot w_i^2 = e^{2H(u_1)} \sum_{i=1}^{d} w_i^2 = e^{2H(u_1)} \cdot 1 = e^{2H(u_1)} \tag{70}$$

At the first step, we used identity (27) for each of the terms. At the second step, (69) was exploited. Then term $e^{H(u_1)}$ was pulled out and we took into account that $\boldsymbol{w} \in S^d(\mathbb{R})$.

□

Note 5. We note that sources of spherically symmetric distribution have already been used in the context of $\mathbb{R}$-ISA in [10]. In that work, a generative model was assumed. According to the assumption, the distribution of the norms of sample projections to the subspaces were independent. This way, the task was restricted to spherically symmetric source distributions, which is a special case of the general $\mathbb{R}$-ISA task.

Note 6. Spherical variables as well as their non-degenerate affine transforms, the so called elliptical variables (which are equivalent to spherical ones from the point of view of $\mathbb{R}$-ISA) are thoroughly treated in [47,48].

5.1.3 Sources Invariant to 90° Rotation In the previous section, we have seen that random variables with density functions invariant to orthogonal transformations ($\mathbb{R}$-spherical variables) satisfy the conditions of the $\mathbb{R}$-ISA Separation Theorem. For mixtures of 2-dimensional components ($d = 2$), invariance to 90° rotation suffices. First, we observe that:

Note 7. In the $\mathbb{R}$-ISA Separation Theorem, it is sufficient if some orthogonal transformation of the $\boldsymbol{s}^m$ sources, $\boldsymbol{C}^m \boldsymbol{s}^m$ ($\boldsymbol{C}^m \in \mathcal{O}^d$) satisfy the condition (24). In this case, the $\boldsymbol{C}^m \boldsymbol{s}^m$ variables are extracted by the permutation search after the $\mathbb{R}$-ICA transformation. Because the $\mathbb{R}$-ISA identification has ambiguities up to orthogonal transformation in the respective subspaces, this is suitable. In other words, for the $\mathbb{R}$-ISA identification the existence of an Orthonormal Basis (ONB) for each $\boldsymbol{u} := \boldsymbol{s}^m \in \mathbb{R}^d$ components is sufficient, on which the

$$h : \mathbb{R}^d \ni \boldsymbol{w} \mapsto H[\langle \boldsymbol{w}, \boldsymbol{u} \rangle] \tag{71}$$

function takes its minimum. [Here, the $\langle \boldsymbol{w}, \boldsymbol{u} \rangle := \sum_{i=1}^{d} w_i u_i$ random variable is the projection of $\boldsymbol{u}$ to the direction $\boldsymbol{w} \in S^d(\mathbb{R})$.] In this case, the entropy inequality (24) is met with equality on the elements of the ONB.

Now we present our result concerning to the $d = 2$ case.

Proposition 4. *Let us suppose, that the density function f of random variable $\boldsymbol{u} = (u_1, u_2)(= \boldsymbol{s}^m) \in \mathbb{R}^2$ exhibits the invariance*

$$f(u_1, u_2) = f(-u_2, u_1) = f(-u_1, -u_2) = f(u_2, -u_1) \quad \left(\forall \boldsymbol{u} \in \mathbb{R}^2\right), \tag{72}$$

that is, it is invariant to 90° rotation. If function $h(\boldsymbol{w}) = H[\langle \boldsymbol{w}, \boldsymbol{u} \rangle]$ has minimum on the set $\{\boldsymbol{w} \geq \boldsymbol{0}\} \cap S^2(\mathbb{R})$, it also has minimum on an ONB. [2] *Consequently, the $\mathbb{R}$-ISA task can be identified by the use of the $\mathbb{R}$-ISA Separation Theorem.*

Proof. Let

$$\boldsymbol{R} := \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \tag{73}$$

denote the matrix of 90° ccw rotation. Let $\boldsymbol{w} \in S^2(\mathbb{R})$. $\langle \boldsymbol{w}, \boldsymbol{u} \rangle \in \mathbb{R}$ is the projection of variable $\boldsymbol{u}$ onto $\boldsymbol{w}$. The value of the density function of the random variable $\langle \boldsymbol{w}, \boldsymbol{u} \rangle$ in $t \in \mathbb{R}$ (we move t in direction $\boldsymbol{w}$) can be calculated by integration starting from the point $\boldsymbol{w}t$, in direction perpendicular to $\boldsymbol{w}$

$$f_{y=y(\boldsymbol{w})=\langle \boldsymbol{w}, \boldsymbol{u} \rangle}(t) = \int_{\boldsymbol{w}^\perp} f(\boldsymbol{w}t + \boldsymbol{z}) \mathrm{d}\boldsymbol{z}. \tag{74}$$

Using the supposed invariance of f and the relation (74) we have

$$f_{y(\boldsymbol{w})} = f_{y(\boldsymbol{R}\boldsymbol{w})} = f_{y(\boldsymbol{R}^2\boldsymbol{w})} = f_{y(\boldsymbol{R}^3\boldsymbol{w})}, \tag{75}$$

[2] Relation $\boldsymbol{w} \geq \boldsymbol{0}$ concerns each coordinates.

where '=' denotes the equality of functions. Consequently, it is enough to optimize h on the set $\{\boldsymbol{w} \geq \mathbf{0}\}$. Let $\boldsymbol{w}_{min}$ be the minimum of function h on the set $S^2(\mathbb{R}) \cap \{\boldsymbol{w} \geq \mathbf{0}\}$. According to Eq. (75), h takes constant and minimal values in the

$$\{\boldsymbol{w}_{min}, \boldsymbol{R}\boldsymbol{w}_{min}, \boldsymbol{R}^2\boldsymbol{w}_{min}, \boldsymbol{R}^3\boldsymbol{w}_{min}\}$$

points. $\{\boldsymbol{v}_{min}, \boldsymbol{R}\boldsymbol{v}_{min}\}$ is a suitable ONB in Note 7. □

Note 8. A special case of the requirement (72) is invariance to permutation and sign changes, that is

$$f(\pm u_1, \pm u_2) = f(\pm u_2, \pm u_1). \tag{76}$$

In other words, there exists a function $g : \mathbb{R}^2 \to \mathbb{R}$, which is symmetric in its variables and

$$f(\boldsymbol{u}) = g(|u_1|, |u_2|). \tag{77}$$

The domain of Proposition (4) includes

1. the formerly presented $\mathbb{R}$-spherical variables,
2. or more generally, variables with density function of the form

$$f(\boldsymbol{u}) = g\left(\sum_i |u_i|^p\right) \quad (p > 0). \tag{78}$$

   In the literature *essentially* these variables are called $L^p(\mathbb{R})$*-norm sphericals* (for $p > 1$). Here, we use the $L^p(\mathbb{R})$*-norm spherical* denomination in a slightly extended way, for $p > 0$.

5.1.4 Takano's Dependency Criterion

We have seen that the $\mathbb{R}$-w-EPI property is sufficient for the $\mathbb{R}$-ISA Separation Theorem. In [49], sufficient condition is provided to satisfy the EPI condition. The condition is based on the dependencies of the variables and it concerns the 2-dimensional case. The constraint of $d = 2$ may be generalized to higher dimensions. We are not aware of such generalizations.

We note, however, that $\mathbb{R}$-w-EPI requires that $\mathbb{R}$-EPI be satisfied on the surface of the unit sphere. Thus it is satisfactory to consider the intersection of the conditions detailed in [49] on surface of the unit sphere.

5.1.5 Summary of Sufficient Conditions (Real Case)

Here, we summarize the presented sufficient conditions of the $\mathbb{R}$-ISA Separation Theorem. We have proven, that the requirement described by Eq. (24) for the $\boldsymbol{s}^m$ sources is sufficient for the theorem. This holds if the (23) $\mathbb{R}$-w-EPI condition is fulfilled. The stronger $\mathbb{R}$-w-EPI is valid for

1. sources satisfying Takano's weak dependency criterion,
2. $\mathbb{R}$-spherical sources (with equality),
3. sources invariant to 90° rotation (for $d = 2$). Specially, (i) variables invariant to permutation and sign changes, and (ii)$L^p(\mathbb{R})$-norm spherical variables belong to this family.

These results are summarized schematically in Table 1.

5.2 Complex Case

We provide sufficient conditions that fulfill (45).

5.2.1 $\mathbb{C}$-w-EPI

According to Lemma 4, if the $\mathbb{C}$-w-EPI property [i.e., (43)] holds for sources $\boldsymbol{s}^m$, then inequality (45) holds, too.

Table 1. Sufficient conditions for the $\mathbb{R}$-ISA Separation Theorem.

$$
\begin{tikzcd}
 & \text{invariance to } 90^\circ \text{ rotation } (d=2) \arrow[ddd, Rightarrow, "\text{(with `=' for a suitable ONB)}"] \arrow[dr, "\text{specially}"] & \\
 & & \text{invariance to sign and permutation} \arrow[d, "\text{specially}"] \\
 & & L^p(\mathbb{R})\text{-norm spherical } (p>0) \\
\begin{array}{c}\text{Takano's dependency}\\ (d=2)\end{array} \arrow[r, Rightarrow] & \mathbb{R}\text{-w-EPI} \arrow[d, Rightarrow] & \mathbb{R}\text{-spherical symmetry (or elliptical)} \arrow[l, Rightarrow, "{[\text{with `=' for all } \boldsymbol{w} \in S^d(\mathbb{R})]}"'] \arrow[u, "\text{generalization for } d=2"] \\
 & \begin{array}{c}\text{Equation (24): sufficient}\\ \text{for the } \mathbb{R}\text{-Separation Theorem}\end{array} &
\end{tikzcd}
$$

5.2.2 Complex Spherically Symmetric Sources

A complex random variable is complex spherically symmetric, or $\mathbb{C}$-spherical, for short, if its density function – which exists under mild conditions – is constant on concentric complex spheres. We shall show that (45) as well as the stronger (43) $\mathbb{C}$-w-EPI relations are fulfilled. We need certain definitions and some basic features to prove the above statement. Thus, below we shall elaborate on complex sphericals [45].

Definition 3 ($\mathbb{C}$-spherical variable). *A random variable $\boldsymbol{v} \in \mathbb{C}^d$ is called* $\mathbb{C}$-spherical, *if $\boldsymbol{u} = \varphi_v(\boldsymbol{v}) \in \mathbb{R}^{2d}$ $\mathbb{R}$-spherical [45]. Equivalent definition for $\mathbb{C}$-sphericals is that they are invariant to unitary transformations. Formally, if*

$$\boldsymbol{v} \stackrel{\text{distr}}{=} \boldsymbol{U}\boldsymbol{v}, \quad \forall \boldsymbol{U} \in \mathcal{U}^d. \tag{79}$$

We need two basic properties of $\mathbb{C}$-sphericals [45] to prove the theorem. These are analogous to Lemma 5 and Lemma 6:

Lemma 7 (Identical distribution of 1-dimensional projections - Complex case). *Projections of $\mathbb{C}$-spherical variables $\boldsymbol{u} \in \mathbb{C}^d$ onto any unit vectors in $S^d(\mathbb{C})$ have identical distributions. Formally, for $\forall$ $\boldsymbol{w}_1, \boldsymbol{w}_2 \in S^d(\mathbb{C})$*

$$\boldsymbol{w}_1^*\boldsymbol{u} \stackrel{\text{distr}}{=} \boldsymbol{w}_2^*\boldsymbol{u} \in \mathbb{C}. \tag{80}$$

Lemma 8 (Momenta - Complex case). *For $\mathbb{C}$-spherical variable $\boldsymbol{u} \in \mathbb{C}^d$:*

$$E[\boldsymbol{u}] = \boldsymbol{0}, \tag{81}$$
$$cov[\boldsymbol{u}] = c \cdot \boldsymbol{I}_d. \tag{82}$$

We claim the following:

Proposition 5. *$\mathbb{C}$-spherical sources $\boldsymbol{s}^m \in \mathbb{C}^d$ $(m = 1, \ldots, M)$ with finite covariances satisfy condition (45) of the $\mathbb{C}$-ISA Separation Theorem. Further, they satisfy $\mathbb{C}$-w-EPI (with equality).*

Proof. According to (81) and (82) for $\mathbb{C}$-spherical components $\boldsymbol{s}^m \in \mathbb{C}^d$: $E[\boldsymbol{s}^m] = \boldsymbol{0}$, $cov[\boldsymbol{s}^m] = c^m \cdot \boldsymbol{I}_d$. Note that our constraint on the $\mathbb{C}$-ISA task, namely that covariance matrices of the $\boldsymbol{s}^m$ sources should be equal to identity, is fulfilled up to constant multipliers.

Let $P_{\boldsymbol{w}}$ denote the projection to straight line with direction $\boldsymbol{w} \in S^d(\mathbb{C})$, which crosses the origin, i.e.,

$$P_{\boldsymbol{w}} : \mathbb{C}^d \ni \boldsymbol{u} \mapsto \boldsymbol{w}^* \cdot \boldsymbol{u} = \sum_{i=1}^{d} \bar{w}_i u_i \in \mathbb{C}. \tag{83}$$

The left and right hand sides of condition (45) correspond to projection onto vector $\bar{\boldsymbol{w}}$, and projections onto vectors $\boldsymbol{e}_i = [0; ...; 0; 1; 0; ...]$ (1 in the i^{th} position and 0s otherwise), respectively. $\bar{\boldsymbol{w}} \in S^d(\mathbb{C}) \Leftrightarrow \boldsymbol{w} \in S^d(\mathbb{C})$, because conjugation preserves length. Given property (80), the distribution and thus the entropy of these projections are equal. That is (let $\boldsymbol{u}$ denote an arbitrary source, i.e., $\boldsymbol{u} := \boldsymbol{s}^m$),

$$\sum_{i=1}^{d} w_i u_i \stackrel{\text{distr}}{=} u_1 \stackrel{\text{distr}}{=} \ldots \stackrel{\text{distr}}{=} u_d, \quad \forall \boldsymbol{w} \in S^d(\mathbb{C}), \tag{84}$$

$$H\left(\sum_{i=1}^{d} w_i u_i\right) = H(u_1) = \ldots = H(u_d), \quad \forall \boldsymbol{w} \in S^d(\mathbb{C}). \tag{85}$$

Thus:

- l.h.s. of $\mathbb{C}$-w-EPI: $e^{H(u_1)}$.
- r.h.s. of $\mathbb{C}$-w-EPI:

$$\sum_{i=1}^{d} e^{H(w_i u_i)} = \sum_{i=1}^{d} e^{H(u_i)} \cdot |w_i|^2 = e^{H(u_1)} \sum_{i=1}^{d} |w_i|^2 = e^{H(u_1)} \cdot 1 = e^{H(u_1)} \tag{86}$$

At the first step, we used identity (48) for each of the terms. At the second step, (85) was exploited. Then term $e^{H(u_1)}$ was pulled out and we took into account that $\boldsymbol{w} \in S^d(\mathbb{C})$.

□

5.2.3 Summary of Sufficient Conditions (Complex Case) Here, we summarize the presented sufficient conditions of the $\mathbb{C}$-ISA Separation Theorem. We have proven, that the requirement described by Eq. (45) for the $\boldsymbol{s}^m$ sources is sufficient for the theorem. This holds if the (43) $\mathbb{C}$-w-EPI condition is fulfilled. The stronger $\mathbb{C}$-w-EPI is valid for $\mathbb{C}$-spherically symmetric variables.

These results are summarized schematically in Table 2.

Table 2. Sufficient conditions for the $\mathbb{C}$-ISA Separation Theorem.

$$\mathbb{C}\text{-spherical symmetry} \xRightarrow{[\text{with '='} \text{ for all } \boldsymbol{w} \in S^d(\mathbb{C})]} \mathbb{C}\text{-w-EPI} \Longrightarrow \begin{array}{c}\text{Equation (45): sufficient} \\ \text{for the } \mathbb{C}\text{-Separation Theorem}\end{array}$$

6 Conclusions

In this paper a Separation Theorem, a decomposition principle, was presented for the $\mathbb{K}$-Independent Subspace Analysis ($\mathbb{K}$-ISA) problem. If the conditions of the theorem are satisfied then the $\mathbb{K}$-ISA task can be solved in 2 steps. The first step is concerned with the search for 1-dimensional independent components. The second step corresponds to a combinatorial problem, the search for the optimal permutation. We have shown that spherically symmetric sources (for the real and the complex cases, too) satisfy the conditions of the theorem. For the real case and for 2-dimensional sources ($d = 2$) invariance to 90° rotation, or the Takano's dependency criterion is sufficient for the separation.

These results underline our experiences that the presented 2 step procedure for solving the $\mathbb{K}$-ISA task may produce higher quality subspaces than sophisticated search algorithms [17].

Finally we mention that the possibility of this two step procedure (for the real case) was first noted in [11].

7 Appendix

A Uniqueness of $\mathbb{C}$-ISA

Here we provide ambiguities of the $\mathbb{C}$-ISA task. The derivation is similar to that of [42], slight modification is used through mappings φ_v, φ_M [see, Eq. (12) and Eq. (13)].

Notations that we need: Let $D = dM$. Let $Gl(L, \mathbb{K})$ denote the set of invertible matrices in $\mathbb{K}^{L\times L}$. Let us decompose matrix $\boldsymbol{V} \in \mathbb{C}^{D\times D}$ into $d \times d$ blocks: $\boldsymbol{V} = \left[\boldsymbol{V}^{i,j}\right]_{i,j=1,\ldots,M}$ ($\boldsymbol{V}^{i,j} \in \mathbb{C}^{d\times d}$). We say that matrix $\boldsymbol{V}$ is a $d \times d$ *block-permutation matrix*, if there is exactly one index j for $\forall$ i and exactly one i for $\forall$ j ($i, j \in \{1,\ldots,M\}$), that $\boldsymbol{V}^{ij} \neq \boldsymbol{0}$, and further, this block can be inverted. Matrices $\boldsymbol{B}, \boldsymbol{C} \in \mathbb{C}^{D\times D}$ are *d-equivalent* (notation: $\boldsymbol{B} \sim_d \boldsymbol{C}$), if $\boldsymbol{B} = \boldsymbol{C}\boldsymbol{L}$, where $\boldsymbol{L} \in \mathbb{C}^{D\times D}$ is a $d \times d$ block-permutation matrix.[3] Stochastic variable $\mathbb{C}^D \ni \boldsymbol{u} = [\boldsymbol{u}^1;\ldots;\boldsymbol{u}^M]$ is called *d-independent*, if its parts $\boldsymbol{u}^1,\ldots,\boldsymbol{u}^M \in \mathbb{C}^d$ are independent. Using $\boldsymbol{L}$: if $\boldsymbol{u}$ is d-independent, then $\boldsymbol{L}\boldsymbol{u}$ is that, too. Stochastic variable $\boldsymbol{u} \in \mathbb{R}^L$ is called *normal*, if every coordinate is normal. Stochastic variable $\boldsymbol{u} \in \mathbb{C}^L$ is called normal, if both $\Re(\boldsymbol{u})$ and $\Im(\boldsymbol{u})$ are normal. Matrix $\boldsymbol{B} \in \mathbb{C}^{D\times D}$ is called *d-admissible*, if for decomposition $\boldsymbol{B} = \left[\boldsymbol{B}^{i,j}\right]_{i,j=1,\ldots,M}$ ($\boldsymbol{B}^{i,j} \in \mathbb{C}^{d\times d}$) all $\boldsymbol{B}^{i,j}$ blocks are either invertible or indentically 0. (Note: Choosing the coordinates of matrix $\boldsymbol{B}$ from a continuous distribution, the matrix is d-admissible with probability 1.).

Known properties of φ_M, φ_v beyond [(14), (15)] [45] are:

$$\varphi_M(\boldsymbol{M}) \text{ nonsingular (singular)} \Leftrightarrow \boldsymbol{M} \text{ nonsingular (singular)}, \tag{87}$$

$$\varphi_v(\boldsymbol{v}_1 + \boldsymbol{v}_2) = \varphi_v(\boldsymbol{v}_1) + \varphi_v(\boldsymbol{v}_2). \tag{88}$$

To prove our statement, we use the following corollary of the Multivariate Skitovitch-Darmois theorem:

Lemma 9 (Corollary 3.3 in [42]). *Let $\boldsymbol{w}_1 = \sum_{m=1}^M \boldsymbol{B}^m\boldsymbol{u}^m$ and $\boldsymbol{w}_2 = \sum_{m=1}^M \boldsymbol{C}^m\boldsymbol{u}^m$, where $\boldsymbol{u}^m$ are independent random variables from $\mathbb{R}^d$, matrices $\boldsymbol{B}^m, \boldsymbol{C}^m \in \mathbb{R}^{d\times d}$ are zeros, or they belong to $GL(d,\mathbb{R})$. Then, $\boldsymbol{u}^m$ belonging to $\boldsymbol{B}^m\boldsymbol{C}^m \neq \boldsymbol{0}$ are normal, provided that $\boldsymbol{w}_1$ and $\boldsymbol{w}_2$ are independent.*

Theorem 3 (Ambiguities of $\mathbb{C}$-ISA). *Let $Gl(D,\mathbb{C}) \ni \boldsymbol{B} = [\boldsymbol{B}^{i,j}]_{i,j=1..M}$ ($\boldsymbol{B}^{i,j} \in \mathbb{C}^{d\times d}$) d-admissible and $\boldsymbol{s} = [\boldsymbol{s}^1;\ldots;\boldsymbol{s}^M]$ d-independent $D = dM$-dimensional variable, and none of the variables $\boldsymbol{s}^m \in \mathbb{C}^d$ be normal. If $\boldsymbol{B}\boldsymbol{s}$ is again d-independent, then $\boldsymbol{B}$ is d-equivalent to the identity, that is $\boldsymbol{B} \sim_d \boldsymbol{I}_D$.*

Proof. Indirect. Let us assume that $\boldsymbol{B}\boldsymbol{s}$ is d-independent, nonetheless $\boldsymbol{B} \sim_d \boldsymbol{I}_D$ does not hold. Then there is a column index j and there are row indices $i_1 \neq i_2$ for which $\boldsymbol{B}^{i_1,j}, \boldsymbol{B}^{i_2,j} \neq \boldsymbol{0}$ (and because $\boldsymbol{B}$ d-admissible, thus they are invertible).[4] Let us take the parts that correspond to indices i_1, i_2 off from $\boldsymbol{B}\boldsymbol{s}$:

$$\mathbb{C}^d \ni \boldsymbol{y}^{i_1} = \boldsymbol{B}^{i_1,j}\boldsymbol{s}^j + \sum_{m\in\{1,\ldots,M\}\setminus j} \boldsymbol{B}^{i_1,m}\boldsymbol{s}^m \tag{89}$$

$$\mathbb{C}^d \ni \boldsymbol{y}^{i_2} = \boldsymbol{B}^{i_2,j}\boldsymbol{s}^j + \sum_{m\in\{1,\ldots,M\}\setminus j} \boldsymbol{B}^{i_2,m}\boldsymbol{s}^m \tag{90}$$

Applying φ_v, and using properties (88) and (15) we have:

$$\mathbb{R}^{2d} \ni \varphi_v(\boldsymbol{y}^{i_1}) = \varphi_M(\boldsymbol{B}^{i_1,j})\varphi_v(\boldsymbol{s}^j) + \sum_{m\in\{1,\ldots,M\}\setminus j} \varphi_M(\boldsymbol{B}^{i_1,m})\varphi_v(\boldsymbol{s}^m) \tag{91}$$

$$\mathbb{R}^{2d} \ni \varphi_v(\boldsymbol{y}^{i_2}) = \varphi_M(\boldsymbol{B}^{i_2,j})\varphi_v(\boldsymbol{s}^j) + \sum_{m\in\{1,\ldots,M\}\setminus j} \varphi_M(\boldsymbol{B}^{i_2,m})\varphi_v(\boldsymbol{s}^m) \tag{92}$$

Taking advantage of (87): invertibility of $\boldsymbol{B}^{i_1,j}, \boldsymbol{B}^{i_2,j}$ is inherited to $\varphi_M(\boldsymbol{B}^{i_1,j}), \varphi_M(\boldsymbol{B}^{i_2,j})$. Similarly, matrices $\varphi_M(\boldsymbol{B}^{i,m})$ ($i \in \{i_1, i_2\}, m \neq j$) are either zero or they are invertible, according to their ancestor $\boldsymbol{B}^{i,m}$, whether it is zero or invertible. If $\boldsymbol{s}^m \in \mathbb{C}^d$ are independent then variables $\varphi_v(\boldsymbol{s}^m) \in \mathbb{R}^{2d}$ are also independent. Thus, as a result of Lemma 9, $\varphi_v(\boldsymbol{s}^j)$ is normal, meaning – by definition – that $\boldsymbol{s}^j$ is also normal: a contradiction. $\square$

[3] Note: this is an equivalence relation, indeed, because the set of $\boldsymbol{L}$s that satisfy the conditions is closed for inversion and for multiplication.

[4] Reasoning: if for all j there is at most one block (submatrix), which is non-zero, then: (a) for all j there is exactly one block, which is non-zero and then $\boldsymbol{B} \sim_d \boldsymbol{I}_D$, which is a contradiction, or, (b) there is a j index for which submatrix $\boldsymbol{B}^{i,j}$ has only zeros, and then the invertibility of $\boldsymbol{B}$ is not fulfilled.

Note 9. [40] has shown an interesting result: for the complex case and for $d = 1$ ($\mathbb{C}$-ICA task) certain normal sources can be separated. This result (and thus Theorem 3) may be extended to $d > 1$, too, but we are not aware of such generalization.

References

1. Jutten, C., Herault, J.: Blind separation of sources: An adaptive algorithm based on neuromimetic architecture. Signal Processing **24** (1991) 1–10 Cited on page(s): 1
2. Comon, P.: Independent component analysis, a new concept? Signal Processing **36** (1994) 287–314 Cited on page(s): 1
3. Bell, A.J., Sejnowski, T.J.: An information maximisation approach to blind separation and blind deconvolution. Neural Computation **7** (1995) 1129–1159 Cited on page(s): 1
4. Bell, A.J., Sejnowski, T.J.: The 'independent components' of natural scenes are edge filters. Vision Research **37** (1997) 3327–3338 Cited on page(s): 1
5. Hyvärinen, A.: Sparse code shrinkage: Denoising of nongaussian data by maximum likelihood estimation. Neural Computation **11** (1999) 1739–1768 Cited on page(s): 1
6. Kiviluoto, K., Oja, E.: Independent component analysis for parallel financial time series. In: Proceedings of ICONIP'98. Volume 2. (1998) 895–898 Cited on page(s): 1
7. Makeig, S., Bell, A.J., Jung, T.P., Sejnowski, T.J.: Independent component analysis of electroencephalographic data. In: Proceedings of NIPS. Volume 8. (1996) 145–151 Cited on page(s): 1
8. Vigário, R., Jousmaki, V., Hamalainen, M., Hari, R., Oja, E.: Independent component analysis for identification of artifacts in magnetoencephalographic recordings. In: Proceedings of NIPS. Volume 10. (1997) 229–235 Cited on page(s): 1
9. Choi, S., Cichocki, A., Park, H.M., Lee, S.Y.: Blind source separation and independent component analysis. Neural Inf. Proc. Letters and Reviews **6** (2005) 1–57 Cited on page(s): 1
10. Hyvärinen, A., Hoyer, P.O.: Emergence of phase and shift invariant features by decomposition of natural images into independent feature subspaces. Neural Computation **12** (2000) 1705–1720 Cited on page(s): 1, 11
11. Cardoso, J.: Multidimensional independent component analysis. In: Proceedings of ICASSP'98, Seattle, WA. (1998) Cited on page(s): 1, 2, 14
12. Theis, F.J.: Blind signal separation into groups of dependent signals using joint block diagonalization. In: Proc. ISCAS 2005, Kobe, Japan (2005) 5878–5881 Cited on page(s): 1
13. Kim, T., Eltoft, T., Lee, T.W.: Independent vector analysis: An extension of ICA to multivariate components. In: Independent Component Analysis and Blind Signal Separation. Volume 3889 of LNCS., Springer (2006) 165–172 Cited on page(s): 1
14. Akaho, S., Kiuchi, Y., Umeyama, S.: MICA: Multimodal independent component analysis. In: Proceedings of IJCNN. (1999) 927–932 Cited on page(s): 1
15. Vollgraf, R., Obermayer, K.: Multi-dimensional ICA to separate correlated sources. In: Proceedings of NIPS. Volume 14. (2001) 993–1000 Cited on page(s): 1
16. Bach, F.R., Jordan, M.I.: Finding clusters in independent component analysis. In: Proceedings of ICA2003. (2003) 891–896 Cited on page(s): 1
17. Póczos, B., Lőrincz, A.: Independent subspace analysis using k-nearest neighborhood distances. Artificial Neural Networks: Formal Models and their Applications - ICANN 2005, pt 2, Proceedings **3697** (2005) 163–168 Cited on page(s): 1, 14
18. Póczos, B., Lőrincz, A.: Independent subspace analysis using geodesic spanning trees. In: Proc. of Int. Conf. on Machine Learing (ICML). (2005) 673–680 Cited on page(s): 1
19. Van Hulle, M.M.: Edgeworth approximation of multivariate differential entropy. Neural Computation **17** (2005) 1903–1910 Cited on page(s): 1
20. Anemüller, J., Sejnowski, T.J., Makeig, S.: Complex independent component analysis of frequency-domain electroencephalographic data. Neural Networks **16** (2003) 1311–1323 Cited on page(s): 1
21. Back, A.D., Tsoi, A.C.: Blind deconvolution of signals using a complex recurrent network. In: Neural Networks for Signal Processing 4, Proc. of the 1994 IEEE Workshop, IEEE Press (1994) 565–574 Cited on page(s): 1
22. Fiori, S.: Blind separation of circularly distributed sources by neural extended APEX algorithm. Neurocomputing Letters **34** (2000) 239–252 Cited on page(s): 1
23. Bingham, E., Hyvärinen, A.: A fast fixed-point algorithm for independent component analysis of complex-valued signals. International Journal of Neural Systems **10** (2000) 1–8 Cited on page(s): 1
24. Cardoso, J.F., Souloumiac, A.: Blind beamforming for non-gaussian signals. IEE-Proceedings-F **140** (1993) 362–370 Cited on page(s): 1
25. Cardoso, J.F.: High-order contrasts for independent component analysis. Neural Computation **11** (1999) 157–192 Cited on page(s): 1
26. Moreau, E.: An any order generalization of JADE for complex source signals. In: Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'2001). Volume 5., Salk Lake City, Utah, USA (2001) 2805–2808 Cited on page(s): 1

27. Cardoso, J.F., Laheld, B.: Equivariant adaptive source separation. IEEE Transactions on Signal Processing (1996) Cited on page(s): 1
28. Fiori, S.: Complex-weighted one-unit 'rigid-bodies' learning rule for independent component analysis. Neural Processing Letters **15** (2002) 275–282 Cited on page(s): 1
29. Belouchrani, A., Abed-Meraim, K., Cardoso, J., Moulines, E.: A blind source separation technique using second-order statistics. IEEE Transactions on Signal Processing **45** (1997) 434–444 Cited on page(s): 1
30. Yeredor, A.: Blind separation of gaussian sources via second-order statistics with asymptotically optimal weighting. IEEE Signal Processing Letters **7** (2000) Cited on page(s): 1
31. Smaragdis, P.: Blind separation of convolved mixtures in the frequency domain. Neurocomputing **22** (1998) 21–34 Cited on page(s): 1
32. Calhoun, V.D., Adali, T.: Complex infomax: Convergence and approximation of infomax with complex nonlinearities. In: NNSP. (2002) Cited on page(s): 1
33. Xu, J.W., Erdogmus, D., Rao, Y.N., Príncipe, J.C.: Minimax mutual information approach for ICA of complex-valued linear mixtures. In: ICA. (2004) 311–318 Cited on page(s): 1
34. Eriksson, J., Koivunen, V.: Complex-valued ICA using second order statistics. 2004 IEEE Workshop on Machine Learning for Signal Processing (2004) Cited on page(s): 1
35. Douglas, S.C., Eriksson, J., Koivunen, V.: Equivariant algorithms for estimating the strong-uncorrelating transform in complex independent component analysis. In: Independent Component Analysis and Blind Signal Separation. (2006) 57–65 Cited on page(s): 1
36. Douglas, S.C., Eriksson, J., Koivunen, V.: Fixed-point complex ICA algorithms for the blind separation of sources using their real or imaginary components. In: Independent Component Analysis and Blind Signal Separation. (2006) 343–351 Cited on page(s): 1
37. Póczos, B., Takács, B., Lőrincz, A.: Independent subspace analysis on innovations. Machine Learning: ECML 2005, Proceedings **3720** (2005) 698–706 Cited on page(s): 2
38. Szabó, Z., Póczos, B., Lőrincz, A.: Cross-entropy optimization for independent process analysis. In: Independent Component Analysis and Blind Signal Separation. Volume 3889 of LNCS., Springer (2006) 909–916 Cited on page(s): 2
39. Szabó, Z., Lőrincz, A.: Real and complex independent subspace analysis by generalized variance. In: ICA Research Network Workshop 2006. (2006) Cited on page(s): 2
40. Eriksson, J.: Complex random vectors and ICA models: Identifiability, uniqueness and separability. IEEE Transactions on Information Theory **52** (2006) Cited on page(s): 2, 15
41. Eriksson, J.: Contributions to Theory and Algorithms of Independent Component Analysis and Signal Separation. PhD thesis, Helsinki University of Technology (2004) Cited on page(s): 2
42. Theis, F.: Uniqueness of complex and multidimensional independent component analysis. Signal Processing **84** (2004) 951–956 Cited on page(s): 2, 15
43. Theis, F.: Uniqueness of real and complex linear independent component analysis revisited. In: Proc. EUSIPCO 2004, Vienna, Austria (2004) 1705–1708 Cited on page(s): 2
44. Cover, T., Thomas, J.: Elements of information theory. John Wiley and Sons, New York, USA (1991) Cited on page(s): 3, 5
45. Krishnaiah, P., Lin, J.: Complex elliptically symmetric distributions. Communications in Statistics **15** (1986) 3693–3718 Cited on page(s): 4, 13, 15
46. Taneja, I.J.: Generalized Information Measures and Their Applications. on-line book: www.mtm.ufsc.br/ taneja/book/book.html (2001) Cited on page(s): 7
47. Fang, K.T., Kotz, S., Ng, K.W.: Symmetric multivariate and related distributions. Chapman and Hall (1990) Cited on page(s): 10, 11
48. Frahm, G.: Generalized elliptical distributions: Theory and applications. PhD thesis, University of Köln (2004) Cited on page(s): 10, 11
49. Takano, S.: The inequalities of Fisher information and entropy power for dependent variables. Proceedings of the 7th Japan-Russia Symposium on Probability Theory and Mathematical Statistics (1995) Cited on page(s): 12