Separation Theorem for Independent Subspace Analysis

Zoltán Szabó, Barnabás Póczos, and András Lőrincz

Department of Information Systems
Eötvös Loránd University, Budapest, 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 Independent Subspace Analysis (ISA), a generalization of Independent Component Analysis (ICA) is proven. According to the theorem, ISA estimation can be executed in two steps under certain conditions. In the first step, 1-dimensional ICA estimation is executed. In the second step, optimal permutation of the ICA elements is searched for. We shall show that elliptically symmetric sources, among others, satisfy the conditions of the theorem.

1 Introduction

Independent Component Analysis (ICA) [1,2] aims to recover linearly or non-linearly mixed independent and hidden sources. There is a broad range of applications for 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 ICA see [9].

Original ICA algorithms are 1-dimensional in the sense that all sources are assumed to be independent real valued stochastic 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 ICA, which can be called Independent Subspace Analysis (ISA) or, alternatively, Multi-Dimensional Independent Component Analysis (MICA) [10,11]. Throughout the paper, we shall use the former abbreviation. An important application for ISA is, e.g., the processing of EEG-fMRI data [12].

Efforts have been made to develop ISA algorithms [10,12,13,14,15,16]. Related theoretical problems concern mostly the estimation of entropy and/or mutual information. In this context, entropy estimation by Edgeworth expansion [12] can deal with relatively high dimensions and such estimations have been used for clustering and mutual information testing [17]. k-nearest neighbor and geodesic spanning trees methods have been applied in [15] and [16] for the ISA problem. Another recent approach searches for independent subspaces via kernel methods [14].

An important observation of previous computer studies [10,18] is that general ISA solver algorithms are not more efficient, in fact, sometimes produce lower quality results than simple 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 will be published elsewhere [19].

This technical report is constructed as follows: In Section 2 the ISA task is described. Section 3 contains our separation theorem for the ISA task. Sufficient conditions for the theorem are provided in Section 4. Conclusions are drawn in Section 5.

2 The ISA Model

The generative model of mixed independent multi-dimensional sources (independent subspace analysis, ISA) is the following. We assume that there are M pieces of hidden d-dimensional sources: $\mathbf{s}^m$ $(m = 1, \ldots, M)$. The linear transformation

$$\mathbf{z} = \mathbf{A}\mathbf{s} \tag{1}$$

of their concatenated form

$$\mathbf{s} := \left[\left(\mathbf{s}^1\right)^T, \ldots, \left(\mathbf{s}^M\right)^T\right]^T \tag{2}$$

is available for observation only. Here, superscript T denotes transposition, the total dimension of the sources is $D := d \cdot M$ and thus, $\boldsymbol{s} \in \mathbb{R}^D$ and $\boldsymbol{A} \in \mathbb{R}^{D \times D}$, and $\boldsymbol{z} \in \mathbb{R}^D$. In what follows, we shall assume that *mixing matrix* $\boldsymbol{A}$ is invertible. The ISA task is to estimate the unknown matrix $\boldsymbol{A}$ (or its inverse, the so-called *separation matrix* $\boldsymbol{W}$) and the original sources by means of the observations $\boldsymbol{z}(t)$. The special case of $d = 1$ corresponds to the ICA task.

Given our assumption on the invertibility of matrix $\boldsymbol{A}$, we can assume without any loss of generality that both the sources and the observations are *white*, that is,

$$E[\boldsymbol{s}] = \boldsymbol{0}, E\left[\boldsymbol{s}\boldsymbol{s}^T\right] = \boldsymbol{I}_D, \tag{3}$$

$$E[\boldsymbol{z}] = \boldsymbol{0}, E\left[\boldsymbol{z}\boldsymbol{z}^T\right] = \boldsymbol{I}_D, \tag{4}$$

where $\boldsymbol{I}_D$ is the D-dimensional identity matrix, $E[\cdot]$ denotes the expectation value operator. It then follows that the mixing matrix $\boldsymbol{A}$ and thus the separation matrix $\boldsymbol{W} = \boldsymbol{A}^{-1}$ are orthogonal:

$$\boldsymbol{I}_D = E[\boldsymbol{z}\boldsymbol{z}^T] = \boldsymbol{A}E[\boldsymbol{s}\boldsymbol{s}^T]\boldsymbol{A}^T = \boldsymbol{A}\boldsymbol{I}_D\boldsymbol{A}^T = \boldsymbol{A}\boldsymbol{A}^T. \tag{5}$$

The ambiguity of the ISA task is decreased by Eqs. (3)-(4): Now, sources are determined up to permutation of the sources *and* orthogonal transformation of the subspaces belonging to the sources. For more details on this subject, see [20].

The ISA task can be viewed as the minimization of mutual information between the components. That is, we should minimize the cost function

$$J(\boldsymbol{W}) := \sum_{m=1}^{M} H(\boldsymbol{y}^m) \tag{6}$$

in the space of $D \times D$ orthogonal matrices, where $\boldsymbol{y} = \boldsymbol{W}\boldsymbol{z}$, $\boldsymbol{y}^T = \left[(\boldsymbol{y}^1)^T, \ldots, (\boldsymbol{y}^M)^T\right]$, $\boldsymbol{y}^m$ $(m = 1, \ldots, M)$ are the estimated components and H is Shannon's (multi-dimensional) differential entropy (for more details, see, e.g., [15]).

3 The ISA Separation Theorem

The main result of this work is that the ISA task may be accomplished in two steps under certain conditions. In the first step ICA is executed. The second step is search for the optimal permutation of the ICA components.

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

$$2^{2H\left(\sum_{i=1}^{K} u_i\right)} \geq \sum_{i=1}^{K} 2^{2H(u_i)}, \tag{7}$$

where $u_1, \ldots, u_K \in \mathbb{R}$ denote continuous stochastic variables. This inequality holds for example, for independent continuous variables [21].

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

$$\|\boldsymbol{w}\|^2 := \sum_{i=1}^{K} w_i^2, \tag{8}$$

where w_i is the i^{th} coordinate of vector $\boldsymbol{w}$. The surface of the unit sphere in K dimensions shall be denoted by S^K:

$$S^K := \{\boldsymbol{w} \in \mathbb{R}^K : \|\boldsymbol{w}\| = 1\}. \tag{9}$$

If EPI is satisfied (on S^K) then a further inequality holds:

Lemma 1. *Suppose that continuous stochastic variables $u_1, \ldots, u_K \in \mathbb{R}$ satisfy the following inequality*

$$2^{2H\left(\sum_{i=1}^{K} w_i u_i\right)} \geq \sum_{i=1}^{K} 2^{2H(w_i u_i)}, \forall \boldsymbol{w} \in S^K. \tag{10}$$

This inequality will be called the w-EPI *condition. Then*

$$H\left(\sum_{i=1}^{K} w_i u_i\right) \geq \sum_{i=1}^{K} w_i^2 H\left(u_i\right), \forall \boldsymbol{w} \in S^K. \tag{11}$$

Note 1. w-EPI holds, for example, for independent variables u_i, because independence is not affected by multiplication with a constant.

Proof. Assume that $\boldsymbol{w} \in S^K$. Applying $\log_2$ on condition (10), and using the monotonicity of the $\log_2$ function, we can see that the first inequality is valid in the following inequality chain

$$2H\left(\sum_{i=1}^{K} w_i u_i\right) \geq \log_2\left(\sum_{i=1}^{K} 2^{2H(w_i u_i)}\right) = \log_2\left(\sum_{i=1}^{K} 2^{2H(u_i)} \cdot w_i^2\right) \geq \sum_{i=1}^{K} w_i^2 \cdot \log_2\left(2^{2H(u_i)}\right) = \sum_{i=1}^{K} w_i^2 \cdot 2H(u_i). \quad (12)$$

Then,

1. we used the relation [21]:

$$H(w_i u_i) = H(u_i) + \log_2(|w_i|) \quad (13)$$

   for the entropy of the transformed variable. Hence

$$2^{2H(w_i u_i)} = 2^{2H(u_i)+2\log_2(|w_i|)} = 2^{2H(u_i)} \cdot 2^{2\log_2(|w_i|)} = 2^{2H(u_i)} \cdot w_i^2. \quad (14)$$

2. In the second inequality, we utilized the concavity of $\log_2$. □

Now we shall use Lemma 1 to proceed. The separation theorem will be a corollary of the following claim:

Proposition 1. *Let $\mathcal{O}^D$ denote the space of the $D \times D$ orthogonal matrices, let $\boldsymbol{y} = \left[\left(\boldsymbol{y}^1\right)^T, \ldots, \left(\boldsymbol{y}^M\right)^T\right]^T = \boldsymbol{y}(\boldsymbol{W}) = \boldsymbol{W}\boldsymbol{s}$, where $\boldsymbol{W} \in \mathcal{O}^D$, $\boldsymbol{y}^m$ is the estimation of the m^{th} component of the ISA task. Let y_i^m be the i^{th} coordinate of this 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* (11)*. Then*

$$\sum_{m=1}^{M}\sum_{i=1}^{d} H\left(y_i^m\right) \geq \sum_{m=1}^{M}\sum_{i=1}^{d} H\left(s_i^m\right). \quad (15)$$

Proof. Let us denote the $(i,j)^{th}$ element of matrix $\boldsymbol{W}$ by $W_{i,j}$. For the sake of simplicity, coordinates of $\boldsymbol{y}$ and $\boldsymbol{s}$ will be denoted by y_i and s_i, respectively. Now, writing the elements of the i^{th} row of matrix multiplication $\boldsymbol{y} = \boldsymbol{W}\boldsymbol{s}$, we have

$$y_i = (W_{i,1}s_1 + \ldots + W_{i,d}s_d) + \ldots + (W_{i,D-d+1}s_{D-d+1} + \ldots + W_{iD}s_D) \quad (16)$$

and thus,

$$H(y_i) =$$

$$= H\left(\sum_{j=1}^{d} W_{i,j}s_j + \ldots + \sum_{j=D-d+1}^{D} W_{i,j}s_j\right) \quad (17)$$

$$= H\left(\left(\sum_{l=1}^{d} W_{i,l}^2\right)^{\frac{1}{2}} \frac{\sum_{j=1}^{d} W_{i,j}s_j}{\left(\sum_{l=1}^{d} W_{i,l}^2\right)^{\frac{1}{2}}} + \ldots + \left(\sum_{l=D-d+1}^{D} W_{i,l}^2\right)^{\frac{1}{2}} \frac{\sum_{j=D-d+1}^{D} W_{i,j}s_j}{\left(\sum_{l=D-d+1}^{D} W_{i,l}^2\right)^{\frac{1}{2}}}\right) \quad (18)$$

$$\geq \left(\sum_{l=1}^{d} W_{i,l}^2\right) H\left(\frac{\sum_{j=1}^{d} W_{i,j}s_j}{\left(\sum_{l=1}^{d} W_{i,l}^2\right)^{\frac{1}{2}}}\right) + \ldots + \left(\sum_{l=D-d+1}^{D} W_{i,l}^2\right) H\left(\frac{\sum_{j=D-d+1}^{D} W_{i,j}s_j}{\left(\sum_{l=D-d+1}^{D} W_{i,l}^2\right)^{\frac{1}{2}}}\right) \quad (19)$$

$$= \left(\sum_{l=1}^{d} W_{i,l}^2\right) H\left(\sum_{j=1}^{d} \frac{W_{i,j}}{\left(\sum_{l=1}^{d} W_{i,l}^2\right)^{\frac{1}{2}}} s_j\right) + \ldots + \left(\sum_{l=D-d+1}^{D} W_{i,l}^2\right) H\left(\sum_{j=D-d+1}^{D} \frac{W_{i,j}}{\left(\sum_{l=D-d+1}^{D} W_{i,l}^2\right)^{\frac{1}{2}}} s_j\right) \quad (20)$$

$$\geq \left(\sum_{l=1}^{d} W_{i,l}^2\right) \sum_{j=1}^{d} \left(\frac{W_{i,j}}{\left(\sum_{l=1}^{d} W_{i,l}^2\right)^{\frac{1}{2}}}\right)^2 H(s_j) + \ldots + \left(\sum_{l=D-d+1}^{D} W_{i,l}^2\right) \sum_{j=D-d+1}^{D} \left(\frac{W_{i,j}}{\left(\sum_{l=D-d+1}^{D} W_{i,l}^2\right)^{\frac{1}{2}}}\right)^2 H(s_j) \quad (21)$$

$$= \sum_{j=1}^{d} W_{i,j}^2 H(s_j) + \ldots + \sum_{j=D-d+1}^{D} W_{i,j}^2 H(s_j) \quad (22)$$

The above steps can be justified as follows:

1. (17): Eq. (16) was inserted into the argument of H.
2. (18): New terms were added for Lemma 1.
3. (19): 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 1, because $\boldsymbol{W}$ is an orthogonal matrix.
4. (20): Nominators were transferred into the $\sum_j$ terms.
5. (21): Variables $\boldsymbol{s}^m$ satisfy Lemma 1 according to our conditions.
6. (22): 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

$$\sum_{i=1}^{D} H(y_i) \geq \sum_{i=1}^{D}\left(\sum_{j=1}^{d} W_{i,j}^2 H\left(s_j\right) + \ldots + \sum_{j=D-d+1}^{D} W_{i,j}^2 H\left(s_j\right)\right) \tag{23}$$

$$= \sum_{j=1}^{d}\left(\sum_{i=1}^{D} W_{i,j}^2\right) H\left(s_j\right) + \ldots + \sum_{j=D-d+1}^{D}\left(\sum_{i=1}^{D} W_{i,j}^2\right) H\left(s_j\right) \tag{24}$$

$$= \sum_{j=1}^{D} H(s_j). \tag{25}$$

□

Having this proposition, now we present our main theorem.

Theorem 1 (Separation theorem for ISA). *Presume that the $\boldsymbol{s}^m$ sources of the ISA model satisfy Condition (11). Then the ISA task can be executed in two steps. In the first step, ICA preprocessing is executed on observed data $\boldsymbol{z}$. If the $\boldsymbol{W}_{ICA}$ solution is unique (up to permutation and the sign of the coordinates), then the same matrix is also the $\boldsymbol{W}$ separation matrix of the ISA task (up to permutation and sign of the coordinates). Therefore, it is satisfactory to search for the separation matrix of the ISA task in the following form*

$$\boldsymbol{W} = \boldsymbol{P}\boldsymbol{W}_{ICA}, \tag{26}$$

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

Proof. ICA minimizes the l.h.s. of Eq. (15), that is, it minimizes $\sum_{m=1}^{M}\sum_{i=1}^{d} H\left(y_i^m\right)$. If the solution of the ICA task is unique (up to permutation and the sign of the coordinates), then the location of the minimum is also unique (up to permutation and the sign of the coordinates). However, the minimum, according to Proposition 1, can be achieved only at $\{s_i^m\}$ (up to permutation and the sign of the coordinates). □

4 Sufficient Conditions of the Separation Theorem

In the separation theorem, we assumed that relation (11) is fulfilled for the $\boldsymbol{s}^m$ sources. Here, we shall provide sufficient conditions when this inequality is fulfilled.

4.1 w-EPI

According to Lemma 1, if the w-EPI property (i.e., (10)) holds for sources $\boldsymbol{s}^m$, then inequality (11) holds, too.

4.2 Elliptically Symmetric Sources

A stochastic variable is elliptically symmetric, or elliptical, for short, if its density function – which exists under mild conditions – is constant on elliptic surfaces.[1] We shall show that (11) as well as the stronger (10) w-EPI relations are fulfilled. We need certain definitions and some basic features to prove the above statement. Thus, below we shall elaborate on spherical (spherically symmetric) and elliptically symmetric stochastic variables [22,23].

[1] They are often called elliptically contoured stochastic variables.

Basic Definitions

Definition 1. *(Characteristic function) The characteristic function of stochastic variable* $\boldsymbol{v} \in \mathbb{R}^d$ *is defined by the mapping*

$$\mathbb{R}^d \ni \boldsymbol{t} \mapsto \varphi_{\boldsymbol{v}}(\boldsymbol{t}) := E[\exp(i\boldsymbol{t}^T\boldsymbol{v})], \tag{27}$$

where $i = \sqrt{-1}$ *and* exp *is the exponential function.*

Spherically symmetric variables can be introduced in different ways that, together, provide the view that we need here.

Definition 2 (Spherically symmetric variable around $\boldsymbol{\mu}$). *A stochastic variable* $\boldsymbol{v} \in \mathbb{R}^d$ *is called spherically symmetric around* $\boldsymbol{\mu}$*, if:*

1. *its density function is not modified by any rotation around* $\boldsymbol{\mu}$*. Formally, if*

$$\boldsymbol{v} - \boldsymbol{\mu} \overset{\text{distr}}{=} \boldsymbol{O}\left(\boldsymbol{v} - \boldsymbol{\mu}\right), \quad \forall \boldsymbol{O} \in \mathcal{O}^d, \tag{28}$$

   where $\overset{\text{distr}}{=}$ *denotes equality in distribution.*
2. *its characteristic function with some* $\phi : [0, \infty) \to \mathbb{R}$ *assumes the following form*

$$\varphi_{\boldsymbol{v}-\boldsymbol{\mu}}(\boldsymbol{t}) = \phi\left(\boldsymbol{t}^T\boldsymbol{t}\right). \tag{29}$$

   Function ϕ *is called the* characteristic generator *of* $\boldsymbol{v}$*.*
3. *it has the following stochastic representation*

$$\boldsymbol{v} \overset{\text{distr}}{=} \boldsymbol{\mu} + r\boldsymbol{u}^{(d)}, \tag{30}$$

   where
   (a) $\boldsymbol{\mu} \in \mathbb{R}^d$*: is a constant vector,*
   (b) $\boldsymbol{u}^{(d)}$*: is a stochastic variable of uniform distribution over* S^d*,*
   (c) r*: is a non-negative scalar stochastic variable, which is independent of* $\boldsymbol{u}^{(d)}$*.*

We shall make use of the following well known property of spherically symmetric variables:

Proposition 2. *Let* $\boldsymbol{v}$ *denote a d-dimensional variable, which is spherically symmetric around* $\boldsymbol{\mu}$*. Then the projection of* $\boldsymbol{v} - \mu$ *onto lines through the origin have identical univariate distribution.*

Affine transforms of spherically symmetric variables take us to the concept of elliptically symmetric variables. We shall be interested in the case, when the affine transform is bijective. Then the following definitions are equivalent:

Definition 3 (Elliptically symmetric variable around $\boldsymbol{\mu}$). *A stochastic variable* $\boldsymbol{e} \in \mathbb{R}^d$ *is called elliptically symmetric around* $\boldsymbol{\mu}$*, if:*

1. *there exists* $\boldsymbol{\mu} \in \mathbb{R}^d$ *and an invertible* $\boldsymbol{\Lambda} \in \mathbb{R}^{d\times d}$ *such that*

$$\boldsymbol{e} = \boldsymbol{\mu} + \boldsymbol{\Lambda}\boldsymbol{v}, \tag{31}$$

   where $\boldsymbol{v}$ *is a d-dimensional stochastic variable, which is spherically symmetric around* $\mathbf{0}$*. In this case, the characteristic function of* $\boldsymbol{e}$ *is*

$$\varphi_{\boldsymbol{e}}(\boldsymbol{t}) = \exp\left(i\boldsymbol{t}^T\boldsymbol{\mu}\right)\phi_{\boldsymbol{v}}\left(\boldsymbol{t}^T\boldsymbol{\Sigma}\boldsymbol{t}\right), \tag{32}$$

   where $\boldsymbol{\Sigma} := \boldsymbol{\Lambda}\boldsymbol{\Lambda}^T$ *and* $\phi_{\boldsymbol{v}}$ *is the characteristic function of* $\boldsymbol{v}$*.*
2. *there exists vector* $\boldsymbol{\mu} \in \mathbb{R}^d$*, positive definite symmetric matrix* $\boldsymbol{\Sigma} \in \mathbb{R}^{d\times d}$*, and function* $\phi : [0, \infty) \to \mathbb{R}$ *such, that the characteristic function of* $\boldsymbol{e} - \boldsymbol{\mu}$ *is*

$$\varphi_{\boldsymbol{e}-\boldsymbol{\mu}}(\boldsymbol{t}) = \phi\left(\boldsymbol{t}^T\boldsymbol{\Sigma}\boldsymbol{t}\right). \tag{33}$$

   This property will be denoted as $\boldsymbol{e} \sim E_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi)$*.* ϕ *will be called the* characteristic generator *of variable* $\boldsymbol{e}$*.*
3. $\boldsymbol{e}$ *has stochastic representation of the form*

$$\boldsymbol{e} \overset{\text{distr}}{=} \boldsymbol{\mu} + r\boldsymbol{\Lambda}\boldsymbol{u}^{(d)} \tag{34}$$

   where $\boldsymbol{\Lambda} \in \mathbb{R}^{d\times d}$ *is an invertible matrix and*
   (a) $\boldsymbol{\mu} \in \mathbb{R}^d$*: is a constant vector,*
   (b) $\boldsymbol{u}^{(d)}$*: stochastic variable with uniform distribution on* S^d*,*
   (c) r*: non-negative scalar stochastic variable, which is independent from* $\boldsymbol{u}^{(d)}$*.*

Here: $\boldsymbol{\mu}$*,* $\boldsymbol{\Sigma}$*, and* r *are called the* location vector*, the* dispersion matrix*, and the* generating variate*, respectively.*

Basic Properties Here, we list important properties of an elliptic variable $\boldsymbol{e} \sim E_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi)$.

1. Density function: if $\boldsymbol{e}$ has a density function, then it assumes the form

$$f_{\boldsymbol{e}}(\boldsymbol{x}) = |\boldsymbol{\Lambda}|^{-\frac{1}{2}} \cdot g\left((\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{\Lambda}^{-1} (\boldsymbol{x}-\boldsymbol{\mu})\right), \quad \boldsymbol{x} \neq \boldsymbol{\mu} \tag{35}$$

where

$$\int_0^\infty \frac{\pi^{\frac{d}{2}}}{\Gamma\left(\frac{d}{2}\right)} t^{\frac{d}{2}-1} g(t) \mathrm{d}t = 1 \tag{36}$$

and $g : [0, \infty) \to \mathbb{R}$ is a non-negative function. Here, Γ denotes the *gamma function* defined as

$$\Gamma(a) := \int_0^\infty t^a \exp(-t) \mathrm{d}t \quad (a > 0). \tag{37}$$

One can show that condition (36) on g is necessary and sufficient for making (35) a density function. For the existence of the density function it is sufficient if variable r is absolutely continuous. Then function g has an explicit form, see [23].

2. Momenta: we consider the expectation value and the variance

$$Var[\boldsymbol{e}] := E\left[(\boldsymbol{e} - E[\boldsymbol{e}])(\boldsymbol{e} - E[\boldsymbol{e}])^T\right] \tag{38}$$

of variable $\boldsymbol{e}$. They exist iff the respective momenta of r are finite. Then, supposing that $E\left[r^2\right]$ is finite, we have

$$E[\boldsymbol{e}] = \boldsymbol{\mu} \tag{39}$$

$$Var[\boldsymbol{e}] = \frac{E[r^2]}{d} \boldsymbol{\Sigma} = -\phi'(0) \boldsymbol{\Sigma}. \tag{40}$$

In what follows, we assume that $E\left[r^2\right]$ is finite.

Elliptical Sources Now we are ready to claim the following theorem.

Proposition 3. *Elliptical sources $\boldsymbol{s}^m$ $(m = 1, \ldots, M)$ with finite covariances satisfy condition* (11) *of the ISA separation theorem. Further, they satisfy w-EPI (with equality).*

Proof. Let $\boldsymbol{s}^m \sim E_d(\boldsymbol{\mu}^m, \boldsymbol{\Sigma}^m, \phi^m)$ $(m = 1, \ldots, M)$ denote elliptical sources. Let us normalize each of them as

$$\boldsymbol{y} \mapsto (\boldsymbol{\Sigma}^m)^{-\frac{1}{2}} (\boldsymbol{y} - \boldsymbol{\mu}^m). \tag{41}$$

So, it is satisfactory to prove the theory for spherically symmetric sources. In what follows, $\boldsymbol{s}^{\boldsymbol{m}}$ denotes these spherically symmetric sources. According to (39)-(40), spherically symmetric sources $\boldsymbol{s}^m$ have zero expectation values and up to a constant multiplier they also have identity covariance matrices:

$$E[\boldsymbol{s}^m] = \boldsymbol{0} \tag{42}$$

$$Var[\boldsymbol{s}^m] = c^m \cdot \boldsymbol{I}_d \tag{43}$$

Note that our constraint on the 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 $\boldsymbol{w} \in S^K$, which crosses the origin, i.e.,

$$P_{\boldsymbol{w}} : \mathbb{R}^d \ni \boldsymbol{u} \mapsto \sum_{i=1}^d w_i u_i \in \mathbb{R}. \tag{44}$$

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

$$P_{\boldsymbol{e}_i}(\boldsymbol{u}) = u_i. \tag{45}$$

In this interpretation, (10) and w-EPI are 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 proposition 2, distribution of the spherical $\boldsymbol{u}$ is the same for all such projections and thus its entropy is identical. That is,

$$\sum_{i=1}^{d} w_i u_i \stackrel{\text{distr}}{=} u_1 \stackrel{\text{distr}}{=} \ldots \stackrel{\text{distr}}{=} u_d, \quad \forall \boldsymbol{w} \in S^K, \tag{46}$$

$$h := H\left(\sum_{i=1}^{d} w_i u_i\right) = H(u_1) = \ldots = H(u_d), \quad \forall \boldsymbol{w} \in S^K. \tag{47}$$

Thus:

- l.h.s. of w-EPI: $2^{2H(u_1)}$.
- r.h.s. of w-EPI:

$$\sum_{i=1}^{K} 2^{2H(w_i u_i)} = \sum_{i=1}^{K} 2^{2H(u_i)} \cdot w_i^2 = 2^{2H(u_1)} \sum_{i=1}^{K} w_i^2 = 2^{2H(u_1)} \cdot 1 = 2^{2H(u_1)} \tag{48}$$

At the first step, we used identity (14) for each of the terms. At the second step, (47) was utilized. Then term $2^{H(u_1)}$ was pulled out and we took into account that $\boldsymbol{w} \in S^K$.

□

Note 2. We note that sources of spherically symmetric distribution have already been used in the context of ISA in [11]. 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 ISA task.

4.3 Takano's Dependency Criterium

We have seen that the w-EPI property is sufficient for the ISA separation theorem. In [24], 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 w-EPI requires that EPI be satisfied on the surface of the unit sphere. Thus it is satisfactory to consider the intersection of the conditions detailed in [24] on surface of the unit sphere.

5 Conclusions

In this paper a separation theorem was presented for the independent subspace analysis (ISA) problem. If the conditions of the theorem are satisfied then the 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 elliptically symmetric sources satisfy the conditions of the theorem. We have also noted that the mixture of 2-dimensional sources can also satisfy the theorem provided that the sources satisfy Takano's dependency criterium.

These results underline our experiences that the presented 2 step procedure for solving the ISA task may produce higher quality subspaces than sophisticated search algorithms [19].

Finally we mention that the possibility of this two step procedure was first noted in [10].

References

1. Jutten, C., Herault, J.: Blind Separation of Sources: An Adaptive Algorithm Based on Neuromimetic Architecture. Signal Processing **24** (1991) 1–10
2. Comon, P.: Independent Component Analysis, a New Concept? Signal Processing **36** (1994) 287–314
3. Bell, A.J., Sejnowski, T.J.: An Information Maximisation Approach to Blind Separation and Blind Deconvolution. Neural Computation **7** (1995) 1129–1159

4. Bell, A.J., Sejnowski, T.J.: The ‘Independent Components’ of Natural Scenes are Edge Filters. Vision Research **37** (1997) 3327–3338
5. Hyvärinen, A.: Sparse Code Shrinkage: Denoising of Nongaussian Data by Maximum Likelihood Estimation. Neural Computation **11** (1999) 1739–1768
6. Kiviluoto, K., Oja, E.: Independent Component Analysis for Parallel Financial Time Series. In: Proceedings of ICONIP’98. Volume 2. (1998) 895–898
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
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
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
10. Cardoso, J.: Multidimensional Independent Component Analysis. In: Proceedings of ICASSP’98, Seattle, WA. (1998)
11. Hyvärinen, A., Hoyer, P.: Emergence of Phase and Shift Invariant Features by Decomposition of Natural Images into Independent Feature Subspaces. Neural Computation **12** (2000) 1705–1720
12. Akaho, S., Kiuchi, Y., Umeyama, S.: MICA: Multimodal Independent Component Analysis. In: Proceedings of IJCNN. (1999) 927–932
13. Vollgraf, R., Obermayer, K.: Multi-Dimensional ICA to Separate Correlated Sources. In: Proceedings of NIPS. Volume 14. (2001) 993–1000
14. Bach, F.R., Jordan, M.I.: Finding Clusters in Independent Component Analysis. In: Proceedings of ICA2003. (2003) 891–896
15. Póczos, B., Lőrincz, A.: Independent Subspace Analysis Using k-Nearest Neighborhood Distances. In Duch, W., Kacprzyk, J., Oja, E., Zadrozny, S., eds.: Artificial Neural Networks: Formal Models and Their Applications, ICANN 2005. LNCS 3697, Springer-Verlag (2005) 163–168
16. Póczos, B., Lőrincz, A.: Independent Subspace Analysis Using Geodesic Spanning Trees. In Raedt, L.D., Wrobel, S., eds.: Proceedings of the 22nd International Conference on Machine Learning. (2005) 673–680
17. Hulle, V., M., M.: Edgeworth Approximation of Multivariate Differential Entropy. Neural Computation **17** (2005) 1903–1910
18. Póczos, B., Takács, B., Lőrincz, A.: Independent Subspace Analysis on Innovations. In: ECML, Porto, Portugal. (2005) 698–706
19. Szabó, Z., Póczos, B., Lőrincz, A.: Cross-Entropy Optimization for Independent Process Analysis. (2005) submitted.
20. Theis, F.J.: Uniqueness of Complex and Multidimensional Independent Component Analysis. Signal Processing **84** (2004) 951–956
21. Cover, T., Thomas, J.: Elements of Information Theory. John Wiley and Sons, New York, USA (1991)
22. Fang, K.T., Kotz, S., Ng, K.W.: Symmetric Multivariate and Related Distributions. Chapman and Hall (1990)
23. Frahm, G.: Generalized Elliptical Distributions: Theory and Applications. PhD thesis, University of Köln (2004)
24. 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)