SCATTER MATRICES WITH INDEPENDENT BLOCK PROPERTY AND ISA

Klaus Nordhausen, and Hannu Oja

School of Health Sciences,
University of Tampere
FIN-33014 Tampere, Finland

ABSTRACT

In independent component analysis (ICA) it is often assumed that the p components of the observation vector are linear combinations of p underlying independent components. Two scatter matrices having the so called independence property can then be used to recover the independent components. The assumption of (exactly) p independent components is however often criticized, and several alternative and more realistic models have been suggested. One of these models is the independent subspace model where it is assumed that the p-variate observed vectors are based on k independent subvectors of lengths $p_1,\ldots,p_k$, $p_1+\ldots+p_k=p$. In independent subspace analysis (ISA) the aim is to recover these subvectors. In this paper we describe a solution to ISA which is based on the use of three scatter matrices with the independent block property.

1. INTRODUCTION

In recent years independent component analysis (ICA) has become a popular technique to analyze multivariate data. In the simplest formulation of the independent component model, it is assumed that the components of a p-variate observed random vector $\mathbf{x}$ are linear combinations of the components of an unobserved random vector $\mathbf{s}$ such that the p components of $\mathbf{s}$ are independent. We can then write

$$\mathbf{x}=\Omega\mathbf{s}$$

where Ω is a full-rank $p\times p$ mixing matrix. The main goal in ICA is to find an estimate for any unmixing matrix Γ such that $\mathbf{z}=\Gamma\mathbf{x}$ has independent components. Naturally $\Gamma=\Omega^{-1}$ is one possible choice of the unmixing matrix. There are many estimation algorithms such as FOBI, fastICA, and JADE to solve the problem. For an overview see [6]. Quite recently, an approach based on the use of two scatter matrices with the independence property was proposed, see [8–10].

The assumption that all components have to be independent is often criticized, however, and several alternative models have been suggested. One can assume, for example, that the p-vector $\mathbf{s}$ consists of k subvectors $\mathbf{s}_1$, ..., $\mathbf{s}_k$ which are independent. The model is then called the multivariate independent component model [1, 13] or independent subspace (IS) model (e.g. in [4, 5, 14]) or ISA model [12]. In independent subspace analysis (ISA) one then tries to find an estimate for an unmixing matrix to separate the independent subvectors. In this paper we show how three scatter matrices with block independence property can be used to solve the ISA problem.

The paper is organized as follows. In Section 2 we recall the concept of a scatter matrix and its main properties. We show how two scatter matrices with the independence property can be used to solve the IC problem. Then in Section 3 we introduce the independent subspace (IS) model as given in Theis [14]. Three scatter matrices with the block independence property are then used to solve the IS problem. The approach was already proposed in [7] but several results are proven first time here. A new performance index for the comparison of different independent subspace estimates is proposed as well and then used in the simulation study in Section 4.

Throughout the paper we use the following notation: A $p\times p$-matrix $\mathbf{U}$ is an orthogonal matrix ($\mathbf{U}'\mathbf{U}=\mathbf{U}\mathbf{U}'=\mathbf{I}_p$), $\mathbf{J}$ is a sign-change matrix (a diagonal matrix with diagonal elements ±1), $\mathbf{D}$ is a rescaling matrix (a diagonal matrix with positive diagonal elements), and $\mathbf{P}$ is a permutation matrix (obtained from $\mathbf{I}_p$ by permuting its rows or columns). We also write

$$\mathscr{C}=\{\mathbf{C}\ :\ \mathbf{C}=\mathbf{JPD}\text{ for some }\mathbf{J},\mathbf{P},\text{ and }\mathbf{D}\},$$

that is, if $\mathbf{C}\in\mathscr{C}$ then $\mathbf{C}$ has exactly one non-zero element in each row and each column. Note also that if $\mathbf{C}=\mathbf{JPD}$ then $\mathbf{CC}'=\mathbf{C}'\mathbf{C}=\mathbf{D}^2$ (diagonal). As usual, throughout the paper we denote the L_2 matrix norm $||\cdot||$ which is defined as

$$||\mathbf{A}||=tr(\mathbf{AA}')^{1/2}.$$

2. SCATTER MATRICES AND IC FUNCTIONALS

Let $\mathbf{x}$ be a p-variate random vector with cumulative distribution function $F_{\mathbf{x}}$. Its multivariate variation can be described using a so called scatter functional:

Definition 1 *A $p\times p$-matrix valued functional $\mathbf{S}(F)$ is a scatter matrix functional if it is symmetric, positive definite, and affine equivariant in the sense that*

$$\mathbf{S}(F_{\mathbf{Ax+b}})=\mathbf{AS}(F_{\mathbf{x}})\mathbf{A}'$$

for all full-rank $p\times p$ matrices $\mathbf{A}$ and for all p-vectors $\mathbf{b}$.

The regular covariance matrix $\mathbf{COV}(F)$ is naturally a scatter matrix, and there are many general families of scatter matrix functionals (M-functionals, S-functionals, an so on) proposed in the literature. Note also that if $\mathbf{S}(F)$ is a scatter matrix then $c\cdot\mathbf{S}(F)$ is a scatter matrix as well for all $c>0$. A general practice then is that the scatter matrices are scaled so that $\mathbf{S}(F)=\mathbf{I}_p$ if F is the cdf of an $N_p(\mathbf{0},\mathbf{I}_p)$ distribution. The affine equivariance property also implies that if $\mathbf{x}$ has an elliptically symmetric distribution then all scatter matrices are proportional to the covariance matrix.

A scatter matrix functional $\mathbf{S}(F)$ has the *independence property* if $\mathbf{S}(F_{\mathbf{x}})$ is a diagonal matrix for all $\mathbf{x}$ having independent components. The covariance matrix $\mathbf{COV}(F)$ serves as the first example with this property. Another example is the scatter matrix based on fourth moments: If $\mathbf{E}(F_{\mathbf{x}}) = \mu$, $\mathbf{COV}(F_{\mathbf{x}}) = \Sigma$, and $\mathbf{z} = \Sigma^{-1/2}(\mathbf{x} - \mu)$, then $\mathbf{COV}_4(F)$ is defined by

$$\mathbf{COV}_4(F_{\mathbf{x}}) = \frac{1}{p+2}\Sigma^{1/2}\mathbf{E}\left(\mathbf{z}\mathbf{z}'\mathbf{z}\mathbf{z}'\right)\Sigma^{1/2}.$$

(Here $\Sigma^{1/2}$ and $\Sigma^{-1/2}$ are taken to be symmetric.) It is easy to see that $\mathbf{COV}_4(F)$ is affine equivariant and possesses the independence property. General families of scatter matrices such as M-functionals and S-functionals are designed for elliptical distributions. Scatter matrices then typically do not have the independence property. For any scatter matrix $\mathbf{S}(F)$, one can, however, find a corresponding "symmetrized" scatter matrix with the independence property just by defining

$$\mathbf{S}_{sym}(F_{\mathbf{x}}) = \mathbf{S}(F_{\mathbf{x}_1 - \mathbf{x}_2})$$

where $\mathbf{x}_1$ and $\mathbf{x}_2$ are two independent copies of $\mathbf{x}$. (It is an open question whether all scatter matrices with the independent property can be formulated in this way.) The regular covariance matrix $\mathbf{COV}$ and the scatter matrix $\mathbf{COV}_4$ for example can be reformulated as a symmetrized scatter matrix.

Next we define what we mean by the independent component functional.

Definition 2 *We say that a $p \times p$-matrix valued functional $\Gamma(F_x)$ is a independent component (IC) functional if*

(i) in the IC model $\mathbf{x} = \Omega\mathbf{s}$,

$$\Gamma(F_x)\Omega^{-1} = \mathbf{C} \ \textit{for some} \ \mathbf{C} \in \mathscr{C}.$$

(ii) $\Gamma(F_{\mathbf{Ax+b}}) = \Gamma(F_x)\mathbf{A}^{-1}$ for all full-rank $p \times p$ matrices $\mathbf{A}$ and for all p-vectors $\mathbf{b}$.

Two scatter matrices with the independence property can be used to find an IC functional as follows.

Theorem 1 *Let $\mathbf{S}_1$ and $\mathbf{S}_2$ be two different scatter matrix functionals having the independence property. Define the $p \times p$ matrix-valued functional Γ (and the $p \times p$ diagonal matrix functional Λ) by*

$$\Gamma\mathbf{S}_1\Gamma' = \mathbf{I}_p \ \textit{and} \ \Gamma\mathbf{S}_2\Gamma' = \Lambda$$

where the diagonal elements of Λ are in a decreasing order. Then Γ is an IC functional in the submodel $\mathbf{x} = \Omega\mathbf{s}$ where $\mathbf{S}_1(F_{\mathbf{s}})^{-1}\mathbf{S}_2(F_{\mathbf{s}})$ has distinct diagonal elements.

Proof As $\mathbf{S}_1$ and $\mathbf{S}_2$ are affine equivariant, the definition of Γ implies that $\Gamma(F_{\mathbf{Ax+b}}) = \Gamma(F_{\mathbf{x}})\mathbf{A}^{-1}$ for all full-rank $\mathbf{A}$ and all $\mathbf{b}$. Thus (ii) in Definition 2 is true. But then, in the submodel considered, $\Gamma(F_{\mathbf{x}})\Omega^{-1} = \Gamma(F_{\mathbf{s}})$. As both $\mathbf{S}_1(F_{\mathbf{s}})$ and $\mathbf{S}_1(F_{\mathbf{s}})$ are diagonal, $\Gamma(F_{\mathbf{s}}) = \mathbf{C}$ for some $\mathbf{C} \in \mathscr{C}$ and also (i) in Definition 2 is true.

Note that, if Γ is an IC functional, then the random variable $\mathbf{z} = \Gamma(F_{\mathbf{x}})(\mathbf{x} - \mathbf{E}(\mathbf{x}))$ has independent components and is affine invariant in the sense that

$$\Gamma(F_{\mathbf{x}})(\mathbf{x} - \mathbf{E}(\mathbf{x})) = \Gamma(F_{\mathbf{Ax+b}})((\mathbf{Ax+b}) - \mathbf{E}(\mathbf{Ax+b})).$$

Next we define

$$\mathbf{H}_g(F_{\mathbf{x}}) = \mathbf{E}(\mathbf{z}g(\mathbf{z})')\left(\mathbf{E}(\mathrm{diag}(\mathbf{z}g(\mathbf{z})'))\right)^{-2}\mathbf{E}(g(\mathbf{z})\mathbf{z}')$$

where $\mathbf{g}(\mathbf{x}) = (g_1(x_1), \ldots, g_p(x_p))'$ is a p-variate score function and $\mathbf{z} = \Gamma(F_{\mathbf{x}})(\mathbf{x} - \mathbf{E}(\mathbf{x}))$. Then also $\mathbf{H}_g(F_{\mathbf{x}})$ is affine invariant, and, in the IC model, $\mathbf{H}_g(F_{\mathbf{x}}) = \mathbf{I}_p$. Then an IC functional can be used to build scatter functionals with the independence property as follows.

Theorem 2 *Let Γ be an IC functional. Then*

$$\mathbf{S} = (\Gamma'\Gamma)^{-1} \ \textit{and} \ \mathbf{S}_{\mathbf{g}} = \Gamma^{-1}\mathbf{H}_g(\Gamma^{-1})'.$$

are scatter functionals with the independence property.

Proof (i) As $\Gamma(F_{\mathbf{Ax+b}}) = \Gamma(F_{\mathbf{x}})\mathbf{A}^{-1}$ for all full-rank $\mathbf{A}$ and all $\mathbf{b}$ then also $\mathbf{S}(F_{\mathbf{Ax+b}}) = \mathbf{AS}(F_{\mathbf{x}})\mathbf{A}'$ for all full-rank $\mathbf{A}$ and all $\mathbf{b}$. Next note that $\mathbf{H}_g(F_{\mathbf{Ax+b}}) = \mathbf{H}_g(F_{\mathbf{x}})$ for all full-rank $\mathbf{A}$ and all $\mathbf{b}$. Thus both $\mathbf{S}$ and $\mathbf{S}_g$ are scatter functionals. (ii) Next assume that $\mathbf{x}$ has independent components. Then $\Gamma(F_{\mathbf{x}}) = \mathbf{C} = \mathbf{JPD}$ for some $\mathbf{J}$, $\mathbf{P}$, and $\mathbf{D}$. But then $\mathbf{S}(F_{\mathbf{x}}) = \mathbf{D}^{-2}$ is diagonal. As $\mathbf{H}_g(F_{\mathbf{x}}) = \mathbf{I}_p$, also $\mathbf{S}_g(F_{\mathbf{x}}) = \mathbf{D}^{-2}$. Thus both $\mathbf{S}$ and $\mathbf{S}_g$ possess the independence property.

A whole family of scatter matrix functionals with the independence property is thus given by

$$\mathbf{S}_{\mathbf{g}}(F) = \Gamma(F)^{-1}\mathbf{H}_g(F)(\Gamma(F)^{-1})'$$

where g is the score function. Note that if Γ is based on $\mathbf{S}_1$ and $\mathbf{S}_2$ then, in the IC model, $\mathbf{S}_g(F_{\mathbf{x}}) = \mathbf{S}_1(F_{\mathbf{x}})$. If $\mathbf{x}$ does not obey the IC model then we can expect that $\mathbf{S}_g(F_{\mathbf{x}}) \neq \mathbf{S}_1(F_{\mathbf{x}})$.

3. THREE SCATTER MATRICES AND ISA

3.1 An algorithm for the ISA problem

The independence subspace model is obtained if the standardized vector $\mathbf{z}$ has independent subvectors. Write then $\mathbf{s}_i$ for the independent p_i-subvectors, $i = 1,...,k$, and $\mathbf{s} = (\mathbf{s}_1', \ldots, \mathbf{s}_k')'$. Write also $p = p_1 + \ldots + p_k$. We also require that the subvectors $\mathbf{s}_i$ are *irreducible*, which means that they cannot be further transformed and decomposed into independent subvectors.

For independent subspace analysis we need the new concept of the block independence property. A scatter matrix $\mathbf{S}(F)$ has the *block independence property* if, for all k and for all $\mathbf{s} = (\mathbf{s}_1', \ldots, \mathbf{s}_k')'$ as described above, $\mathbf{S}(F_{\mathbf{s}})$ is block diagonal with block sizes $p_1, \ldots, p_k$. All the scatter matrices with the independence property discussed in Section 2 have also the block independence property. (Naturally the block independence property implies the independence property. We do not know, however, whether the independence property implies the block independence property.)

Let $\mathbf{S}_1$, $\mathbf{S}_2$, and $\mathbf{S}_3$ be three scatter matrix functionals having the block independence property. If $\mathbf{s} = (\mathbf{s}_1', \ldots, \mathbf{s}_k')'$ has independent subvectors then $\mathbf{S}_i(F_{\mathbf{s}})$, $i = 1,2,3$, are all block diagonal with block sizes $p_1, ..., p_k$.

Theorem 3 *Assume that $\mathbf{s} = (\mathbf{s}_1', \ldots, \mathbf{s}_k')'$ has independent subvectors with dimensions $p_1, ..., p_k$, and that $\mathbf{x} = \Omega\mathbf{s}$ where Ω is a full-rank $p \times p$ matrix. Let $\mathbf{S}_1$, $\mathbf{S}_2$, and $\mathbf{S}_3$ be three scatter matrix functionals having the block independence property, and let Γ be the IC functional based on $\mathbf{S}_1$ and $\mathbf{S}_2$.*

Assume that $\mathbf{S}_1(F_\mathbf{s})^{-1}\mathbf{S}_2(F_\mathbf{s})$ has distinct eigenvalues. Then there exists a permutation matrix $\mathbf{P}$ such that, if $\mathbf{z} = \mathbf{P}\Gamma\mathbf{x}$ then $\mathbf{S}_1(F_\mathbf{z}) = \mathbf{I}_p$, $\mathbf{S}_2(F_\mathbf{z})$ is diagonal, and $\mathbf{S}_3(F_\mathbf{z})$ is block-diagonal with block sizes $p_1,...,p_k$. Then

$$\begin{pmatrix} \mathbf{z}_1 \\ \mathbf{z}_2 \\ \cdots \\ \mathbf{z}_k \end{pmatrix} = \begin{pmatrix} \mathbf{A}_1\mathbf{s}_1 \\ \mathbf{A}_2\mathbf{s}_2 \\ \cdots \\ \mathbf{A}_k\mathbf{s}_k \end{pmatrix}$$

for some full-rank matrices $\mathbf{A}_1,\mathbf{A}_2,...,\mathbf{A}_k$ of sizes $p_1\times p_1$, $p_2\times p_2$, ..., $p_k\times p_k$.

Proof As Γ is affine equivariant we can, without loss of generality, assume that $\Omega = \mathbf{I}_p$. Then both $\mathbf{S}_1 = \mathbf{S}_1(F_\mathbf{x})$ and $\mathbf{S}_2 = \mathbf{S}_2(F_\mathbf{x})$ are block-diagonal, and there is a block-diagonal $\Gamma^* = diag(\Gamma_1^*,...,\Gamma_k^*)$ such that $\Gamma^*\mathbf{S}_1(\Gamma^*)' = \mathbf{I}_p$ and $\Gamma^*\mathbf{S}_2(\Gamma^*)' = \Lambda^*$ (diagonal). Then $\Gamma = \mathbf{P}'\Gamma^*$ where permutation matrix $\mathbf{P}$ is chosen so that the diagonal element of $\mathbf{P}'\Lambda^*\mathbf{P}$ are in a decreasing order. If, with these choices, $\mathbf{z} = \mathbf{P}\Gamma\mathbf{x} = \Gamma^*\mathbf{x}$, then $\mathbf{S}_1(F_\mathbf{z}) = \mathbf{I}_p$, $\mathbf{S}_2(F_\mathbf{z})$ is diagonal, and $\mathbf{S}_3(F_\mathbf{z})$ is block-diagonal with block sizes $p_1,...,p_k$. The final step is to note that one can choose $\mathbf{A}_i = \Gamma_i^*$, $i = 1,...,k$.

Remark 1 *If in the IS model $\mathbf{x} = \Omega\mathbf{s}$, the independent components are standardized so that $\mathbf{S}_1(F_\mathbf{s}) = \mathbf{I}_p$ then the IS components in Theorem 3 satisfy*

$$\begin{pmatrix} \mathbf{z}_1 \\ \mathbf{z}_2 \\ \cdots \\ \mathbf{z}_k \end{pmatrix} = \begin{pmatrix} \mathbf{U}_1\mathbf{s}_1 \\ \mathbf{U}_2\mathbf{s}_2 \\ \cdots \\ \mathbf{U}_k\mathbf{s}_k \end{pmatrix}$$

for some orthogonal matrices $\mathbf{U}_1,\mathbf{U}_2,...,\mathbf{U}_k$ of sizes $p_1\times p_1$, $p_2\times p_2$, ..., $p_k\times p_k$. Thus in this case the IS components are the same up to an orthogonal transformation $\mathbf{U} = diag(\mathbf{U}_1,...,\mathbf{U}_k)$.

Based on the above result we now propose the following Algorithm 1 for the ISA problem with known block sizes $p_1,...,p_k$. For the algorithm we need the following notation. We write $\text{diag}_{p_1,...,p_k}(\mathbf{S})$ for a block diagonal matrix with block sizes $p_1,...,p_k$ and with block elements given by the same elements in $\mathbf{S}$. Matrix $\text{off}_{p_1,...,p_k}(\mathbf{S})$ is defined by

$$\text{off}_{p_1,...,p_k}(\mathbf{S}) = \mathbf{S} - \text{diag}_{p_1,...,p_k}(\mathbf{S}).$$

Algorithm 1 An algorithm for ISA problem with known block sizes $p_1,...,p_k$

[(1)] Find $\Gamma(\mathbf{X})$ based on $\mathbf{S}_1(\mathbf{X})$ and $\mathbf{S}_2(\mathbf{X})$ (Theorem 1)
[(2)] $\mathbf{Z} \leftarrow (\mathbf{X} - \mathbf{1}_n\bar{\mathbf{x}}')\Gamma(\mathbf{X})'$
[(3)] Calculate
$\mathbf{C}_3(\mathbf{Z}) = (\text{diag}(\mathbf{S}_3(\mathbf{Z}))^{-1/2}\mathbf{S}_3(\mathbf{Z})\,(\text{diag}(\mathbf{S}_3(\mathbf{Z}))^{-1/2}$
[(4)] Find $\mathbf{P} = \mathbf{P}(\mathbf{Z})$ to minimize
$||\text{off}_{p_1,...,p_k}(\mathbf{P}\mathbf{C}_3(\mathbf{Z})\mathbf{P}')||^2$
[(5)] Return unmixing matrix estimate $\mathbf{P}(\mathbf{Z})\Gamma(\mathbf{X})$

Note that the solution $\mathbf{P}(\mathbf{Z})\Gamma(\mathbf{X})$ is affine equivariant in the following sense. If $\mathbf{X}^* = \mathbf{A}\mathbf{X}$ then

$$\begin{aligned} \mathbf{Z}^* &= \Gamma(\mathbf{X}^*)\mathbf{X}^* = \Gamma(\mathbf{A}\mathbf{X})(\mathbf{A}\mathbf{X}) \\ &= \Gamma(\mathbf{X})\mathbf{A}^{-1}\mathbf{A}\mathbf{X} = \Gamma(\mathbf{X})\mathbf{X} = \mathbf{Z} \end{aligned}$$

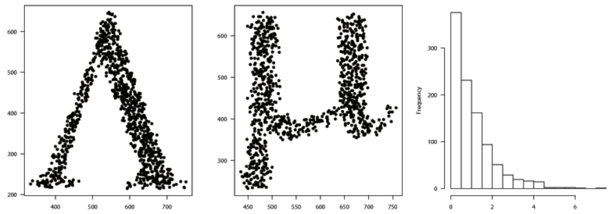

Figure 1: The observed distributions of the three sources.

and therefore $\mathbf{P}(\mathbf{Z}^*)\Gamma(\mathbf{X}^*) = \mathbf{P}(\mathbf{Z})\Gamma(\mathbf{X})\mathbf{A}^{-1}$ and

$$\mathbf{P}(\mathbf{Z}^*)\Gamma(\mathbf{X}^*)\mathbf{X}^* = \mathbf{P}(\mathbf{Z})\Gamma(\mathbf{X})\mathbf{X}.$$

3.2 Comparison of the ISA solutions

We thus a have a family of solutions for the ISA problem based on the three scatter matrices $\mathbf{S}_1$, $\mathbf{S}_2$ and $\mathbf{S}_3$ with the block independence property. How then to compare different choices of matrices $\mathbf{S}_1$, $\mathbf{S}_2$ and $\mathbf{S}_3$? For the comparison, we first write the unmixing matrix as

$$\mathbf{P}(\mathbf{Z})\Gamma(\mathbf{X}) = \begin{pmatrix} \mathbf{B}_1 \\ \mathbf{B}_2 \\ \cdots \\ \mathbf{B}_k \end{pmatrix}$$

where $\mathbf{B}_i = \mathbf{B}_i(\mathbf{X})$ is a $p_i\times p$ transformation matrix estimate to the ith subvector, $i = 1,...,k$. The solution can also be given by projection matrices $\mathbf{P}_i = \mathbf{B}_i'(\mathbf{B}_i\mathbf{B}_i')^{-1}\mathbf{B}_i$, and $\mathbf{P}_i(\mathbf{X})$ then estimates the corresponding population value $\mathbf{P}_i$ with rank p_i, $i = 1,...,k$. Naturally, we require that

$$\mathbf{P}_1(\mathbf{X}) + ... + \mathbf{P}_k(\mathbf{X}) = \mathbf{P}_1 + ... + \mathbf{P}_k = \mathbf{I}_p,$$

and

$$\mathbf{P}_i(\mathbf{X})\mathbf{P}_j(\mathbf{X}) = \mathbf{P}_i\mathbf{P}_j = \mathbf{0}, \text{ if } i \neq j.$$

We then suggest the following criterion for the comparison of the estimates.

Definition 3 *In the ISA problem, the average distance between the estimate $(\mathbf{P}_1(\mathbf{X}),...,\mathbf{P}_k(\mathbf{X}))$ and the population value $(\mathbf{P}_1,...,\mathbf{P}_k)$ is given by*

$$D^2 = \frac{1}{2p}\min_{(\alpha_1,...,\alpha_k)}\sum_{i=1}^k ||\mathbf{P}_{\alpha_i}(\mathbf{X}) - \mathbf{P}_i||^2$$

where $(\alpha_1,...,\alpha_k)$ goes through all the permutations of $(1,...,k)$.

The index D^2 is motivated by the subspace distance introduced in [2]. Clearly, $0 \leq D^2 \leq 1$ and, under perfect separation, $D^2 = 0$. The strict upper limit depends on the values $p_1,...,p_k$. The index is also invariant in the sense that the values of the index for the unmixing matrix estimates

$$\begin{pmatrix} \mathbf{B}_1 \\ \mathbf{B}_2 \\ \cdots \\ \mathbf{B}_k \end{pmatrix} \text{ and } \begin{pmatrix} \mathbf{A}_1\mathbf{B}_1 \\ \mathbf{A}_2\mathbf{B}_2 \\ \cdots \\ \mathbf{A}_k\mathbf{B}_k \end{pmatrix}$$

are the same for all full-rank matrices $\mathbf{A}_1,\mathbf{A}_2,...,\mathbf{A}_k$ of sizes $p_1\times p_1$, $p_2\times p_2$, ..., $p_k\times p_k$.

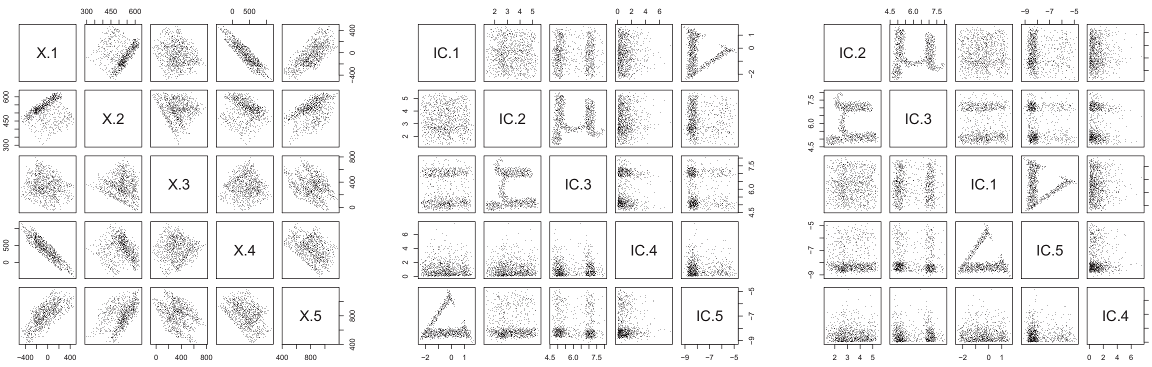

Figure 2: Estimation procedure: The original data set (left panel), transformed data set (middle panel), and the transformed and permuted data set (right panel). The transformation is based on the symmetrized Huber's matrix and Dümbgen's shape matrix, and the permutation is based on $\mathbf{S}_3^*$.

Remark 2 *Consider the model in Theorem 3. Let* $\mathbf{B}$ *be an unmixing matrix estimate, and write* $\mathbf{G} = \mathbf{B}\Omega^{-1}$ *for the corresponding gain matrix. Roughly speaking, the estimate* $\mathbf{B}$ *is good if* $\mathbf{G}$ *is close to a matrix* $\mathbf{A} = diag(\mathbf{A}_1, \ldots \mathbf{A}_k)$ *where* $\mathbf{A}_1, \ldots, \mathbf{A}_k$ *are full rank matrices of sizes* $p_1 \times p_1$, ..., $p_k \times p_k$*. If* $p_1 = \ldots p_k = d$ *and*

$$\mathbf{G} = \begin{pmatrix} \mathbf{G}_{11} & \mathbf{G}_{12} & \ldots & \mathbf{G}_{1k} \\ \mathbf{G}_{21} & \mathbf{G}_{22} & \ldots & \mathbf{G}_{2k} \\ \ldots & \ldots & \ldots & \ldots \\ \mathbf{G}_{k1} & \mathbf{G}_{k2} & \ldots & \mathbf{G}_{kk} \end{pmatrix}$$

then, in this special case, an extension of the Amari index [12, 14] is

$$\frac{1}{2k(k-1)} \left[\sum_{i=1}^{k} \left(\frac{\sum_{j=1}^{k} ||\mathbf{G}_{ij}||_1}{\max_j ||\mathbf{G}_{ij}||_1} \right) + \sum_{j=1}^{k} \left(\frac{\sum_{i=1}^{k} ||\mathbf{G}_{ij}||_1}{\max_i ||\mathbf{G}_{ij}||_1} \right) \right]$$

but now with L_1 *matrix norm* $||\cdot||_1$. *(*L_1 *matrix norm is the sum of the absolute values of the elements of the matrix.) Unfortunately, the value of the extended Amari index is not necessarily the same for* $\mathbf{G}$ *and* $\mathbf{AG}$ *with* $\mathbf{A} = diag(\mathbf{A}_1, \ldots, \mathbf{A}_k)$*. The value is not even necessary the same for* $\mathbf{G}$ *and* $\mathbf{UG}$ *with orthogonal* $\mathbf{U} = diag(\mathbf{U}_1, \ldots, \mathbf{U}_k)$*. If one uses the* L_2 *matrix norm, then* $\mathbf{G} = \mathbf{UG}$ *for all orthogonal* $\mathbf{U} = diag(\mathbf{U}_1, \ldots, \mathbf{U}_k)$.

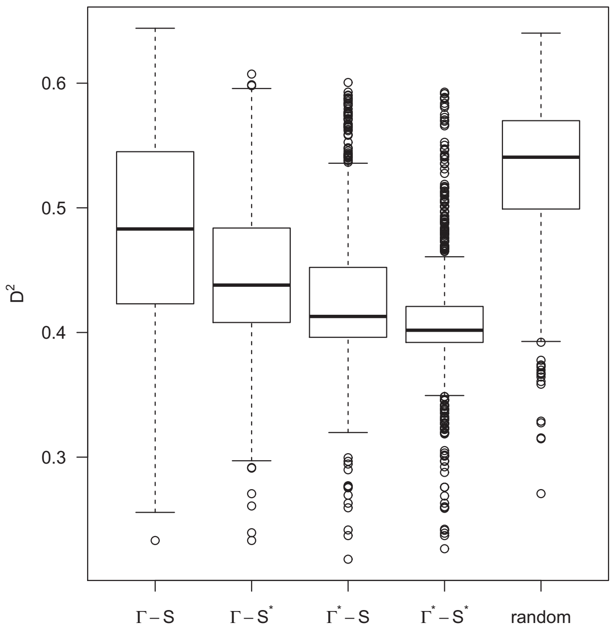

Figure 3: Boxplots for observed values of D^2 for the four estimates Γ-$\mathbf{S}$, Γ-$\mathbf{S}^*$, Γ^*-$\mathbf{S}$, and Γ^*-$\mathbf{S}^*$ different scatter matrix combinations with a random estimate as reference.

4. EXAMPLE

To illustrate our approach, we considered four different estimates $\mathbf{P}(\mathbf{Z})\Gamma(\mathbf{X})$ where $\Gamma(\mathbf{X})$ is based on $\mathbf{S}_1(\mathbf{X})$ and $\mathbf{S}_2(\mathbf{X})$ and $\mathbf{P}(\mathbf{Z})$ is based on $\mathbf{S}_3(\mathbf{Z})$. The idea then is that all three scatter matrices estimate a different type of nonlinear dependency between the components. We then have

(i) two choices of Γ:

- Γ is based on $\mathbf{COV}$ and $\mathbf{COV}_4$ (FOBI)
- Γ^* is based on the symmetrized Huber scatter matrix [11] and Dümbgen's shape matrix [3]. This combination is highly robust and was for example used for robust ICA in [8].

(ii) two choices of $\mathbf{S}_3$:

- $\mathbf{S}_3$ is $\mathbf{S}_g$ where $g_j(z_{ij})$ is the rank of z_{ij} among $z_{1j}, \ldots, z_{nj}$.
- $\mathbf{S}_3^*$ is $\mathbf{S}_g$ where $g_j(z_{ij}) = I(z_{ij} \geq q_{3j}) - I(z_{ij} < q_{3j})$ and q_{3j} is the observed third quartile of the jth component of $\mathbf{z}$.

The data in our simulations come from a 5-variate distribution with two bivariate subvectors $\mathbf{s}_1$ and $\mathbf{s}_2$ and one univariate component s_3. The bivariate sources have the bivariate densities with shapes of the Greek letters Λ and μ, and the third univariate source follows an exponential distribution. In Figure 1 a random sample of size $n = 1000$ was taken to illustrate the distributions of the components. Figure 2 shows the stages of the estimation procedure: Using the original mixed data set $\mathbf{X}$ in left panel, transformation $\Gamma^*(\mathbf{X})$ gives the transformed data $\mathbf{Z}$ in the middle panel, and, finally, the permutation $\mathbf{P}(\mathbf{Z})$ based on $\mathbf{S}_3^*(\mathbf{Z})$ gives the blocked data in the right panel.

As our estimation procedure is affine equivariant, it is not a restriction to consider only one choice of Ω, say, $\Omega = \mathbf{I}_p$. For all four methods, Γ-$\mathbf{S}$, Γ-$\mathbf{S}^*$, Γ^*-$\mathbf{S}$, and Γ^*-$\mathbf{S}^*$, we then computed the performance index D^2 in 1000 repetitions with sample size $n = 1000$. The boxplots for the results are given in Figure 3. Natural reference values are given by the observed values of the performance indices for an estimate which is just a "random" 5×5 unmixing matrix. For this data set, the symmetrized Huber and Dümbgen's matrices at first stage (Γ^*), and the scatter matrix $\mathbf{S}_3^*$ provides the best estimate of the unmixing matrix. (Note that $\mathbf{S}_3^*$ measures the dependence in a very strange, non-symmetric, way.)

5. CONCLUSION

In this paper we propose a new approach for the estimation of the unmixing matrix in the ISA problem which is partly based on some earlier results presented in [7]. The procedure presented here is based on three scatter matrices with the block independence property. First two scatter matrices are used to transform the data, and the third scatter matrix then finds a correct permutation for the components of the transformed data, assuming the dimensions of the independent subspaces are known. A new index for the comparison of equivariant estimates is also proposed. The approach is illustrated with a small example and the procedures with different choices of scatter matrices are compared with a simulation study.

REFERENCES

[1] J.F. Cardoso, "Multidimensional independent component analysis," in *Proc. of ICASSP 1998*, Seattle. 1998.

[2] L.J. Crone and D.S. Crosby, "Statistical applications of a metric on subsapces to satellite meterology," *Technometrics*, vol. 37, pp. 324–328, 1995.

[3] L. Dümbgen, "On Tyler's M-functional of scatter in high dimension," *Annals of the Institute of Statistical Mathematics*, vol. 50, pp. 471–491, 1998.

[4] H. Gutch and F. Theis, "Independent subspace analysis is unique, given irreducibility," in M. Davies, C. James, S. Abdallah, and M. Plumbley (Eds.) *Independent Components Analysis and Signal Separation*, pp. 49–56, Heidelberg: Springer. 2007.

[5] H. Gutch and F. Theis, "Uniqueness of linear factorization into independent subspaces," submitted.

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

[7] K. Nordhausen and H. Oja, "Independent subspace analysis using three scatter matrices," *Austrian Journal of Statistics*, vol. 40, pp.93–101, 2011.

[8] K. Nordhausen, H. Oja, and E. Ollila, "Robust independent component analysis based on two scatter matrices," *Austrian Journal of Statistics*, vol. 37, pp. 91–100, 2008.

[9] H. Oja, S. Sirkiä, and J. Eriksson, "Scatter matrices and independent component analysis," *Austrian Journal of Statistics*, vol. 35, pp. 175–189, 2006.

[10] E. Ollila, H. Oja, and V. Koivunen, "Complex-valued ICA based on a pair of generalized covariance matrices," *Computational Statistics & Data Analysis*, vol. 52, pp. 3789–3805, 2008.

[11] S. Sirkiä, S. Taskinen, and H. Oja, "Symmetrised M-estimators of scatter," *Journal of Multivariate Analysis*, vol. 98, pp. 1611–1629, 2007.

[12] Szabó, Z., Póczos, B., and Lőrincz, A., "Undercomplete Blind Subspace Deconvolution," *Journal of Machine Learning Research*, vol. 8, pp. 1063–1095, 2007.

[13] F. Theis, "Uniqueness of complex and multidimensional independent component analysis," *Signal Processing*, vol. 84, pp. 951–956, 2004.

[14] F. Theis, "Towards a general independent subspace analysis," in *Proc. NIPS 2006*. 2007.