AUSTRIAN JOURNAL OF STATISTICS
Volume 40 (2011), Number 1 & 2, 93-101

# Independent Subspace Analysis Using Three Scatter Matrices

Klaus Nordhausen and Hannu Oja
University of Tampere, Finland

**Abstract:** In independent subspace analysis (ISA) one assumes that the components of the observed random vector are linear combinations of the components of a latent random vector with independent subvectors. The problem is then to find an estimate of a transformation matrix to recover the independent subvectors. Regular independent component analysis (ICA) is a special case. In this paper we show how three scatter matrices with the so called block independence property can be used in independent subspace analysis. The procedure is illustrated with a small example.

**Keywords:** Independent Component Analysis, Source Separation.

## 1 Introduction

In recent years independent component analysis (ICA) has become a popular technique to analyze multivariate data. The independent component model simply assumes 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

$$\mathbf{x} = \mathbf{A}\mathbf{s}\,,$$

where $\mathbf{A}$ is a full-rank $p \times p$ mixing matrix. The main goal in ICA is to find an estimate for any unmixing matrix $\mathbf{W}$ such that $\mathbf{z} = \mathbf{W}\mathbf{x}$ has independent components. If $\mathbf{W}$ is an unmixing matrix and $\mathbf{W}^*$ is obtained from $\mathbf{W}$ by permuting its rows and/or multiplying its rows by nonzero constants, then naturally also $\mathbf{W}^*$ is an unmixing matrix. These are the only indeterminacies if $\mathbf{s}$ has at most one Gaussian component. There are many algorithms such as FOBI, fastICA, and JADE to solve the problem. For an overview see Hyvärinen, Karhunen, and Oja (2001). Quite recently, an approach based on the use of two scatter matrices with the independence property was proposed, see Oja, Sirkiä, and Eriksson (2006); Nordhausen, Oja, and Ollila (2008); Ollila, Oja, and Koivunen (2008).

In practical applications of ICA the independent components often have a real physical interpretation, and ICA has therefore been used as a tool for dimension reduction. The assumption that all components have to be independent has often been criticized, however, and several alternative model assumptions have been suggested. One can assume, for example, that the $p$-vector $\mathbf{s}$ consists of $k$ subvectors $\mathbf{s}_1, \ldots, \mathbf{s}_k$ which are independent. The model is then called the multivariate independent component model or independent subspace model. In independent subspace analysis (ISA) one then tries to find 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 will introduce the independent subspace (IS) model as given in Theis (2007). Three scatter matrices with the block independence property are then used to solve the IS problem. The theory is illustrated with an example 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 $\pm 1$), $\mathbf{D}$ 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).

# 2 Two Scatter Matrices and ICA

Let $\mathbf{x}$ be a $p$-variate random vector with cumulative distribution function $F_{\mathbf{x}}$. A $p \times p$-matrix valued functional $\mathbf{S}(F)$ is a *scatter matrix* or *scatter matrix functional* if it is symmetric, positive definite, and affine equivariant in the sense that

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

for all full-rank $p \times p$ matrices $\mathbf{A}$ and for all $p$-vectors $\mathbf{b}$. The regular covariance matrix $\mathrm{cov}(F)$ 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. If $\mathbf{x}$ has an elliptically symmetric distribution then all scatter matrices are proportional to the covariance matrix.

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 $\mathrm{cov}(F)$ serves as the first example with this property. Another example is the scatter matrix based on fourth moments, namely,

$$\mathrm{cov}_4(F_{\mathbf{x}}) = \frac{1}{p+2}\mathbf{E}\left((\mathbf{x}-\mathbf{E}(\mathbf{x}))'\mathrm{cov}^{-1}(\mathbf{x})(\mathbf{x}-\mathbf{E}(\mathbf{x}))(\mathbf{x}-\mathbf{E}(\mathbf{x}))(\mathbf{x}-\mathbf{E}(\mathbf{x}))'\right) .$$

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 symmetrized version with the independence property 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 $\mathbf{x}$.

Let $\mathbf{S}_1$ and $\mathbf{S}_2$ be two scatter matrix functionals having the independence property. Then we can write the independent component model as

$$\mathbf{x} = \mathbf{\Omega}\mathbf{z} ,$$

where $\mathbf{z}$ is now standardized so that

$$\mathbf{S}_1(F_{\mathbf{z}}) = \mathbf{I}_p \quad \text{and} \quad \mathbf{S}_2(F_{\mathbf{z}}) = \mathbf{\Lambda} .$$

If the diagonal elements of $\mathbf{\Lambda}$ are strictly ordered, $\lambda_1 > \cdots > \lambda_p > 0$, then the *mixing matrix* $\mathbf{\Omega}$ and *unmixing matrix* $\mathbf{\Gamma} = \mathbf{\Omega}^{-1}$ are uniquely defined up to sign changes of the columns and rows, respectively. If $\mathbf{S}_1 = \mathbf{S}_1(F_{\mathbf{x}})$ and $\mathbf{S}_2 = \mathbf{S}_2(F_{\mathbf{x}})$ are given then the ICA solution $\mathbf{\Gamma}$ (and $\mathbf{\Lambda}$) solves

$$\mathbf{\Gamma}\mathbf{S}_1\mathbf{\Gamma}' = \mathbf{I}_p \quad \text{and} \quad \mathbf{\Gamma}\mathbf{S}_2\mathbf{\Gamma}' = \mathbf{\Lambda}\,.$$

Note that $\mathbf{\Gamma}$ and $\mathbf{\Lambda}$ give the eigenvectors and eigenvalues of $\mathbf{S}_1^{-1}\mathbf{S}_2$. Of course, any matrix

$$\mathbf{\Gamma}^* = \mathbf{DJP\Gamma}$$

is then also an unmixing matrix (and an ICA solution), that is, $\mathbf{\Gamma}^*\mathbf{x}$ has independent components as well.

We end this section with some supplementary notes on the above unmixing matrix functional $\mathbf{\Gamma} = \mathbf{\Gamma}(F)$. The transformation $\mathbf{z} = \mathbf{\Gamma}(F_{\mathbf{x}})\mathbf{x}$ is invariant (up to sign changes) in the sense that

$$\mathbf{\Gamma}(F_{\mathbf{Ax}})(\mathbf{Ax}) = \mathbf{J\Gamma}(F_{\mathbf{x}})\mathbf{x}$$

for some sign-change matrix $\mathbf{J}$. The sign-change matrix can be fixed by requiring, for example, that the mean is greater than the median for each component of $\mathbf{J\Gamma}(F_{\mathbf{x}})\mathbf{x}$. (For invariant coordinate selection based on two shape matrices in the general case, see also Tyler, Critchley, Dümbgen, and Oja (2009).) Then a whole family of scatter matrices with the independence property is given by

$$\mathbf{S}_{\mathbf{g}}(\mathbf{x}) = \mathbf{E}((\mathbf{x}-\mathbf{E}(\mathbf{x}))\mathbf{g}(\mathbf{z})')[\mathbf{E}(\text{diag}(\mathbf{zg}(\mathbf{z})'))]^{-2}\mathbf{E}(\mathbf{g}(\mathbf{z})(\mathbf{x}-\mathbf{E}(\mathbf{x}))')\,,$$

where $\mathbf{g}(\mathbf{x}) = (g_1(x_1),\ldots,g_p(x_p))'$ is a $p$-variate score function and $\mathbf{z} = \mathbf{\Gamma}(F_{\mathbf{x}})(\mathbf{x} - \mathbf{E}(\mathbf{x}))$. In the IC model, $\mathbf{S}_g(F_{\mathbf{x}}) = \mathbf{S}_1(F_{\mathbf{x}})$.

# 3 Three Scatter Matrices and ISA

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

$$\mathbf{S}_i(F_{\mathbf{z}}) = \text{diag}(\mathbf{S}_{i1}(F_{\mathbf{z}}),\ldots,\mathbf{S}_{ik}(F_{\mathbf{z}}))\,, \qquad i = 1,2,3$$

are all block diagonal. We can now write the independent subspace (IS) model as

$$\mathbf{x} = \mathbf{\Omega}\mathbf{z}\,,$$

where now $\mathbf{z}$ is standardized so that

$$\mathbf{S}_1(F_{\mathbf{z}}) = \mathbf{I}_p \quad \text{and} \quad \mathbf{S}_2(F_{\mathbf{z}}) = \boldsymbol{\Lambda},$$

and

$$\mathbf{S}_3(F_{\mathbf{z}}) = \text{diag}(\mathbf{S}_{31}(F_{\mathbf{z}}), \ldots, \mathbf{S}_{3k}(F_{\mathbf{z}}))$$

is block diagonal with block sizes $p_1, \ldots, p_k$. For uniqueness, $\boldsymbol{\Lambda} = \text{diag}(\boldsymbol{\Lambda}_1, \ldots, \boldsymbol{\Lambda}_k)$ is supposed to be a diagonal matrix such that all the diagonal elements are distinct and that the diagonal elements in $\boldsymbol{\Lambda}_i$ satisfy $\lambda_{i1} > \cdots > \lambda_{ip_i} > 0$, $i = 1, \ldots, k$, and $\lambda_{i1} > \cdots > \lambda_{k1}$.

If $\mathbf{S}_1 = \mathbf{S}_1(F_{\mathbf{x}})$ and $\mathbf{S}_2 = \mathbf{S}_2(F_{\mathbf{x}})$ are given then one again first finds a ICA solution $\boldsymbol{\Gamma}$ (and $\boldsymbol{\Lambda}$) by solving

$$\boldsymbol{\Gamma}\mathbf{S}_1\boldsymbol{\Gamma}' = \mathbf{I}_p \quad \text{and} \quad \boldsymbol{\Gamma}\mathbf{S}_2\boldsymbol{\Gamma}' = \boldsymbol{\Lambda}.$$

Then a permutation matrix $\mathbf{P}$ is found such that $\mathbf{S}_3(F_{\mathbf{P}\boldsymbol{\Gamma}\mathbf{x}})$ is block diagonal. Our solution of the IS problem is then the unmixing matrix $\mathbf{W} = \mathbf{P}\boldsymbol{\Gamma}$.

Assume that $\mathbf{W} = (\mathbf{W}_1', \ldots, \mathbf{W}_k')'$ is any solution in the IS problem, that is, $\mathbf{W}_i\mathbf{x}$ are independent $p_i$-subvectors, $i = 1, \ldots, k$. If $\mathbf{A}_i$ is a full-rank $p_i \times p_i$-matrix, $i = 1, \ldots, k$, then naturally also

$$\mathbf{W}^* = ((\mathbf{A}_1\mathbf{W}_1)', \ldots, (\mathbf{A}_k\mathbf{W}_k)')'$$

is an IS solution. Also $\mathbf{W} = (\mathbf{W}_{\alpha_1}', \ldots, \mathbf{W}_{\alpha_k}')'$ is a solution for any permutation $(\alpha_1, \ldots, \alpha_k)$ of $(1, \ldots, k)$. These are the only indeterminacies in this model when at most one component is Gaussian. See Theis (2004, 2007) and Gutch and Theis (2007, 2010).

# 4 Example

In our example a random vector $\mathrm{s}$ consists of three subvectors $\mathrm{s}_1$, $\mathrm{s}_2$, and $\mathrm{s}_3$, with dimensions 1, 2, and 2. The univariate component $\mathrm{s}_1$ has an exponential distribution, and the densities of bivariate subvectors $\mathrm{s}_2$ and $\mathrm{s}_3$ have the shapes of Greek letters $\boldsymbol{\mu}$ and $\boldsymbol{\Lambda}$. A sample of size 1000 was drawn from these three independent sources, see Figure 1 for the observed distributions of $\mathrm{s}_1$, $\mathrm{s}_2$, and $\mathrm{s}_3$.

The sources (vector $\mathrm{s} = (\mathrm{s}_1, \mathrm{s}_2', \mathrm{s}_3')'$) were mixed by a $5 \times 5$ mixing matrix $\boldsymbol{\Omega}$ where, for this example, the elements of $\boldsymbol{\Omega}$ were independently sampled from a uniform distribution on $[0, 1]$. (In fact, due to the invariance of our method the estimated independent subvectors do not depend on the choice of the mixing matrix $\boldsymbol{\Omega}$.) The mixed data is presented in Figure 2 and shows no clear structure. Similarly the principal components as shown in Figure 3 do not reveal the structure we are looking for.

For our independent subspace analysis we compared two choices of $\mathbf{S}_1$ and $\mathbf{S}_2$. First, we used the FOBI functional based on the choices $\mathbf{S}_1 = \text{cov}$ and $\mathbf{S}_2 = \text{cov}_4$. Second, we used Dümbgen's shape matrix ($\mathbf{S}_1$) and the symmetrized Huber matrix ($\mathbf{S}_2$). For descriptions and details on these robust choices of scatter matrices, see Nordhausen et al. (2008). For scatterplots of the data after these two ICA transformations, see Figure 4 and Figure 5.

Figure 4 shows that FOBI did not yield good estimates of the subspaces. However, the second and fourth variables seem to provide a (bad) estimate for the $\boldsymbol{\Lambda}$ subspace, and the

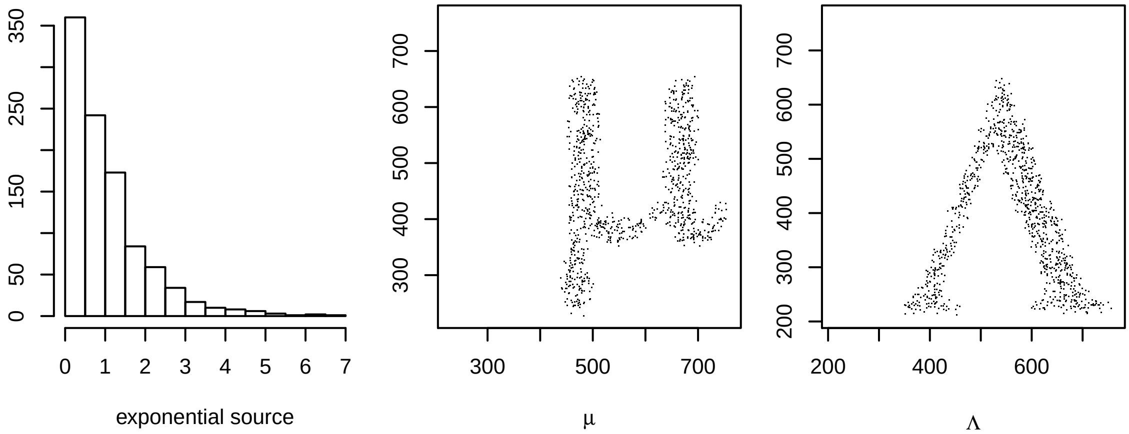


Figure 1: The observed distributions of the three sources ($\mathbf{s}_1$, $\mathbf{s}_2$, and $\mathbf{s}_3$).

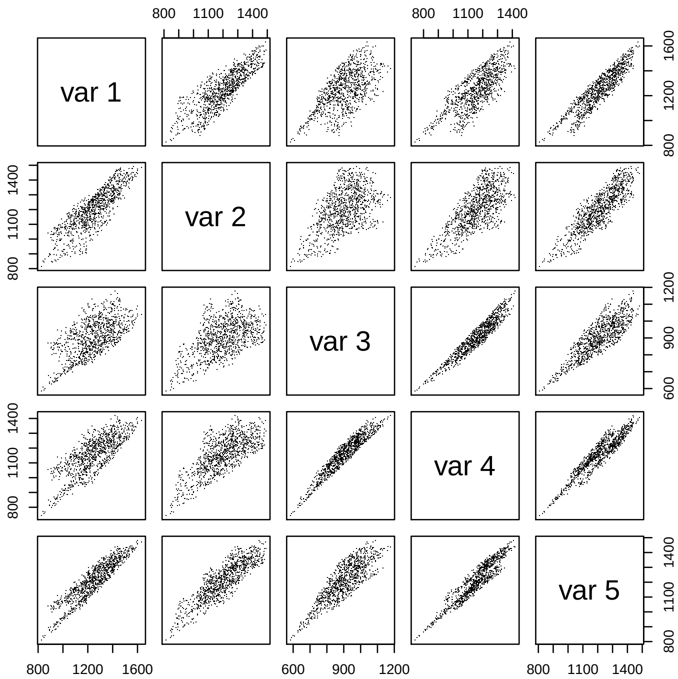


Figure 2: Scatterplot of the sources mixed using a random matrix.

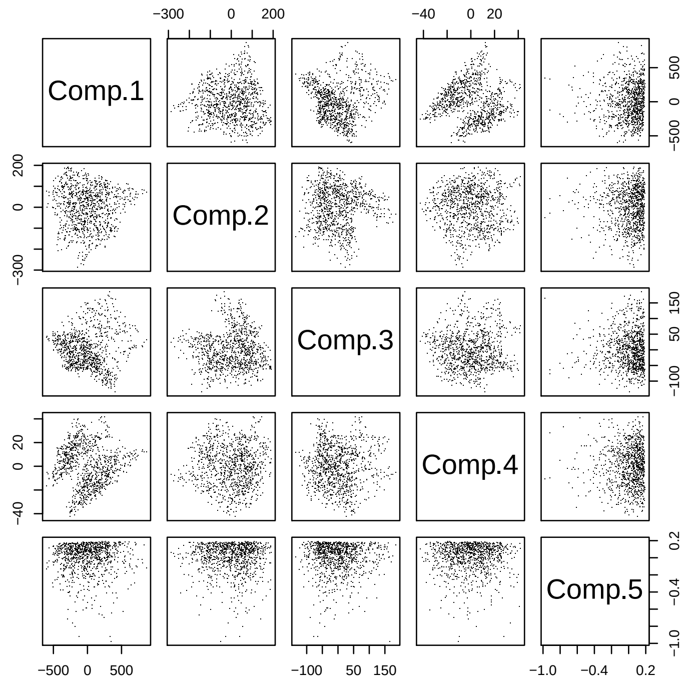


Figure 3: Scatterplot of the principal components based on the mixed data.

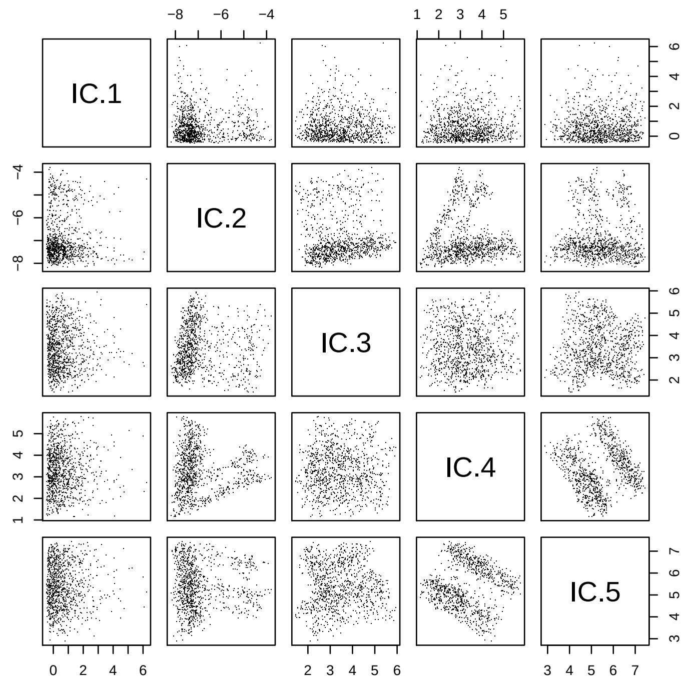


Figure 4: Scatterplot of the data after the ICA transformation based on FOBI (regular covariance matrix and matrix of fourth moments).

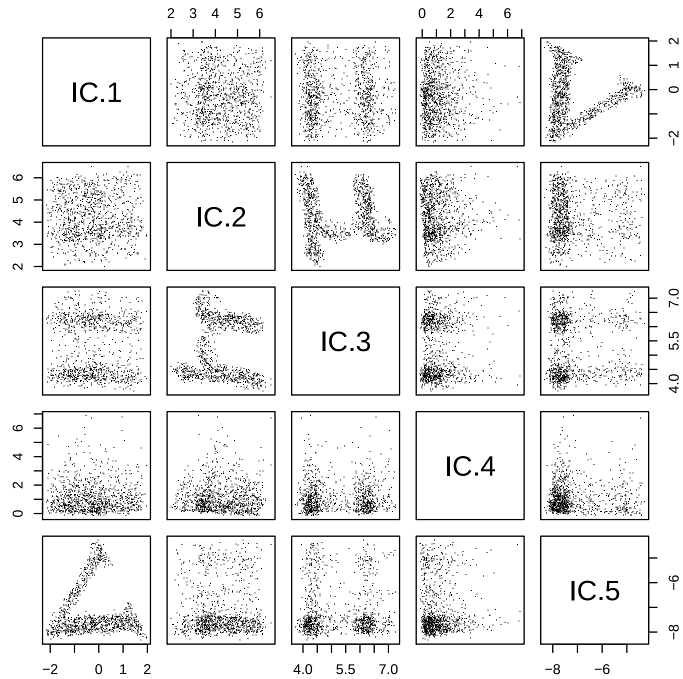


Figure 5: Scatterplot after the ICA transformation based on Dümbgen's shape matrix and symmetrized Huber matrix.

third and fifth variables the $\boldsymbol{\mu}$ subspace. The use of Dümbgen's shape matrix and the symmetrized Huber matrix seemed to work much better here: The first and fifth transformed variables approximately correspond to the $\boldsymbol{\Lambda}$ subspace, the second and third variables the $\boldsymbol{\mu}$ subspace, and the fourth variable the "exponential" subspace. The subspaces should then be found and confirmed by a third scatter matrix. We try two different scatter matrices from the family $\{\mathbf{S}_\mathbf{g}\}$ discussed in Section 2. Scatter matrix $\mathbf{S}_r$ uses $g_j(z_{ij})$ which is the rank of $z_{ij}$ among $z_{1j}, \ldots, z_{nj}$. Scatter matrix $\mathbf{S}_{q_3}$ uses the asymmetric function

$$g_j(z_{ij}) = I(z_{ij} \geq q_{3j}) - I(z_{ij} < q_{3j}),$$

where $q_{3j}$ is the third quartile of $z_{1j}, \ldots, z_{nj}$.

The correlation matrices based on $\mathbf{S}_r$ and $\mathbf{S}_{q_3}$ for the components found by FOBI are

$$\begin{pmatrix} 1.00 & -0.05 & -0.01 & 0.01 & -0.01 \\ -0.05 & 1.00 & 0.16 & \mathbf{0.08} & -0.07 \\ -0.01 & 0.16 & 1.00 & 0.01 & \mathbf{0.00} \\ 0.01 & \mathbf{0.08} & 0.01 & 1.00 & 0.02 \\ -0.01 & -0.07 & \mathbf{0.00} & 0.02 & 1.00 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 1.00 & 0.05 & -0.01 & -0.02 & 0.01 \\ 0.05 & 1.00 & -0.09 & \mathbf{-0.27} & 0.12 \\ -0.01 & -0.09 & 1.00 & -0.01 & \mathbf{-0.20} \\ -0.02 & \mathbf{-0.27} & -0.01 & 1.00 & -0.08 \\ 0.01 & 0.12 & \mathbf{-0.20} & -0.08 & 1.00 \end{pmatrix}.$$

Recall that the FOBI transformation is not that good. If we use $\mathbf{S}_r$ as a third scatter matrix, we do not find the block structure. Surprisingly, $\mathbf{S}_{q_3}$ as a third scatter matrix seems to find the structure; the highest correlations show the blocks weakly suggested by Figure 4.

Consider next the components resulting from the transformation based on Dümbgen's shape matrix and symmetrized Huber matrix. See Figure 5. The two correlation matrices are now

$$\begin{pmatrix} 1.00 & 0.00 & -0.02 & 0.01 & \mathbf{0.08} \\ 0.00 & 1.00 & \mathbf{-0.13} & -0.01 & 0.00 \\ -0.02 & \mathbf{-0.13} & 1.00 & -0.00 & 0.03 \\ 0.01 & -0.01 & -0.00 & 1.00 & 0.01 \\ \mathbf{0.08} & 0.00 & 0.03 & 0.01 & 1.00 \end{pmatrix} \text{ and } \begin{pmatrix} 1.00 & 0.05 & -0.10 & 0.01 & \mathbf{-0.41} \\ 0.05 & 1.00 & \mathbf{-0.24} & -0.02 & 0.02 \\ -0.10 & \mathbf{-0.24} & 1.00 & 0.04 & -0.00 \\ 0.01 & -0.02 & 0.04 & 1.00 & 0.01 \\ \mathbf{-0.41} & 0.02 & -0.00 & 0.01 & 1.00 \end{pmatrix}.$$

Both scatter matrices find the block structure, and one obtains again really high block correlations with $\mathrm{S}_{q_3}$. In this example and with these scatter matrices it is not difficult to choose a permutation matrix to obtain the final result. Our preliminary results suggest that one should be careful in selecting the scatter matrices. The two scatter matrices for the ICA step should not be too similar. The matrix $\mathrm{S}_{q_3}$ with asymmetrical $g$-function seemed to be a good choice for the third scatter matrix. We believe that there is no global best triple of scatter matrices but that the best combination depends on the actual underlying distributions. Hence in a real data situation we would advise to try several different combinations.

# 5 Summary

In this paper we show how three scatter matrices with the independence block property can be used to perform ISA. Our independent subspace model assumptions are a bit more restrictive than Theis's definition (Theis, 2007) as he does not assume that the diagonal values of $\boldsymbol{\Lambda}$ are distinct. In the literature, it is often assumed that $p_1 = \cdots = p_k = K$. This is known as K-ISA, see for example Cardoso (1998); Szabo and Lörincz (2001); Poczos and Lörincz (2005); Theis (2007). The example showed that our approach requires still further investigation. The choice of scatter functionals had a strong impact on the results. It seems that the scatter functionals should measure different types of dependencies, not only linear. Furthermore the procedure should automatically decide what are the number and dimensions of the subvectors, and then permute the components accordingly. Our future work is to compare our approach to the other recently introduced methods such as SJADE (Theis, 2007), fastISA (Hyvärinen and Köster, 2006) and other approaches (Szabo and Lörincz, 2001; Poczos and Lörincz, 2005).

## Acknowledgement

This research work was supported by the Academy of Finland. Fabian Theis kindly provided functions to draw samples from the distributions with the letter shapes used in our example.

# References

Cardoso, J.-F. (1998). Multidimensional independent component analysis. In *Proceedings of ICASSP 1998* (p. 1941-1944). Seattle.

Gutch, H. W., and Theis, F. J. (2007). Independent subsapce analysis is unique, given irreducibility. In M. E. Davies, C. C. James, S. A. Abdallah, and M. D. Plumbley (Eds.), *Independent Component Analysis and Signal Separation* (p. 49-56). Heidelberg: Springer.

Gutch, H. W., and Theis, F. J. (2010). *Uniqueness of linear factorizations into independent subspaces.* (submitted)

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

Hyvärinen, A., and Köster, U. (2006). FastISA: A fast fixed-point algorithm for independent subspace analysis. In *Proceedings of the European Symposium on Artificial Neural Networks.* Bruges, Belgium.

Nordhausen, K., Oja, H., and Ollila, E. (2008). Robust independent component analysis based on two scatter matrices. *Austrian Journal of Statistics*, *37*, 91-100.

Oja, H., Sirkiä, S., and Eriksson, J. (2006). Scatter matrices and independent component analysis. *Austrian Journal of Statistics*, *35*, 175-189.

Ollila, E., Oja, H., and Koivunen, V. (2008). Complex-valued ICA based on a pair of generalized covariance matrices. *Computational Statistics & Data Analysis*, *52*, 3789-3805.

Poczos, B., and Lörincz, A. (2005). Independent subspace analysis using k-nearest neighborhood distances. In *Proceedings of ICANN 2005, Vol. 3696 of LNCS* (p. 163-168). Warsaw: Springer.

Szabo, Z., and Lörincz, A. (2001). Real and complex independent subspace analysis by generalized variance. In *ICA Research Network International Workshop (ICARN 2006)* (p. 85-88).

Theis, F. (2004). Uniqueness of complex and multidimensional independent component analysis. *Signal Processing*, *84*, 951-956.

Theis, F. (2007). Towards a general independent subspace analysis. In B. Schölkopf, J. Platt, and T. Hoffman (Eds.), *Proceedings of NIPS 2006* (p. 1361-1368). Cambridge, MA: MIT Press.

Tyler, D. E., Critchley, F., Dümbgen, L., and Oja, H. (2009). Invariant coordinate selection. *Journal of Royal Statistical Society, Series B*, *71*, 549-592.

Authors' address:
Klaus Nordhausen and Hannu Oja
School of Health Sciences
University of Tampere
FIN-33014 Tampere
Finland

E-mails: klaus.nordhausen@uta.fi and hannu.oja@uta.fi