Estimating Non-Gaussian Subspaces by Characteristic Functions

Motoaki Kawanabe[1] and Fabian J. Theis[2]

[1] Fraunhofer FIRST.IDA, Germany
[2] Institute of Biophysics, University of Regensburg, Germany
nabe@first.fhg.de, fabian@theis.name

Abstract. In this article, we consider high-dimensional data which contains a low-dimensional non-Gaussian structure contaminated with Gaussian noise and propose a new method to identify the non-Gaussian subspace. A *linear dimension reduction* algorithm based on the fourth-order cumulant tensor was proposed in our previous work [4]. Although it works well for sub-Gaussian structures, the performance is not satisfactory for super-Gaussian data due to outliers. To overcome this problem, we construct an alternative by using Hessian of characteristic functions which was applied to (multidimensional) independent component analysis [10,11]. A numerical study demonstrates the validity of our method.

1 Introduction

Recently enormous amount of data with a huge number of features have been stored and are to be analyzed. In most real-world applications, the 'signal' or 'information' is typically contained only in a low-dimensional subspace of the high-dimensional data, thus dimensionality reduction is a useful preprocessing for further data analysis. Here we make an assumption on the data: the high-dimensional data $\boldsymbol{x} \in \mathbb{R}^d$ is a sum of low-dimensional non-Gaussian components ('signal') $\boldsymbol{s} \in \mathbb{R}^m$ ($m < d$) and a Gaussian noise $\boldsymbol{n} \sim N(\boldsymbol{0}, \Gamma)$,

$$\boldsymbol{x} = A\boldsymbol{s} + \boldsymbol{n} \tag{1}$$

where A is a $d \times m$ full rank matrix indicating the non-Gaussian subspace and $\boldsymbol{s}$ and $\boldsymbol{n}$ are assumed to be independent. Under this modeling assumption, therefore, the tasks are to estimate the relevant non-Gaussian subspace and to recover the low-dimensional non-Gaussian structures by *linear dimension reduction*. Although our goal is dimension reduction, we want to emphasize that we do *not* assume the Gaussian components to be of *smaller* order of magnitude than the signal components. This setting therefore excludes the use of common linear and non-linear dimensionality reduction methods such as PCA, Isomap [9] and LLE [8].

If the non-Gaussian components s_i's are mutually independent, the model turns out to be the under-complete noisy ICA, and there exist algorithms to extract the independent components in the presence of Gaussian noise [7]. However, this is often a too strict assumption on the practical data.

In contrast, Projection Pursuit (PP) [3,5] or FastICA in the deflation mode [6,7] can also extract dependent non-Gaussian structures by maximizing a prefixed non-Gaussianity index which contains higher order information. Recently two procedures

J. Rosca et al. (Eds.): ICA 2006, LNCS 3889, pp. 157–164, 2006.

have been developped in the same spirit of PP/FastICA. Non-Gaussian Component Analysis (NGCA) [1] was built upon a general semi-parametric framework in mathematical statistics, while the other [4] is a modification of an ICA algorithm (JADE [2]) to the dimension reduction problem. In this paper, we will propose an alternative of the second algorithm with Hessian of characteristic functions which was applied to (multidimensional) independent component analysis [10,11]. In comparison with the fourth-order cumulant, characteristic functions yield more robust and efficient method when data contain super-Gaussian structures.

2 Mathematical Preliminaries

Since the decomposition (1) is not uniquely determined, we will further transform the model to reduce indeterminacies. The noise term $\boldsymbol{n}$ can be decomposed into two independent parts as $\boldsymbol{n} = \boldsymbol{n}_1 + \boldsymbol{n}_2$, where $\boldsymbol{n}_1 = A\boldsymbol{\eta} \in \mathrm{Range}(A)$ and $\boldsymbol{n}_2$ is restricted in the $(d-m)$-dimensional complementary subspace s.t. $\mathrm{Cov}(\boldsymbol{n}_1, \boldsymbol{n}_2) = 0$. Thus, we get a representation with less indeterminacies

$$\boldsymbol{x} = A\widetilde{\boldsymbol{s}} + \boldsymbol{n}_2, \tag{2}$$

where $\widetilde{\boldsymbol{s}} := \boldsymbol{s} + \boldsymbol{\eta}$ and the noise term $\boldsymbol{n}_2$ distributes with a $(d-m)$-dimensional degenerated Gaussian. We remark that we can only recover $\widetilde{\boldsymbol{s}}$, the signal with contaminated noise in the non-Gaussian subspace $\mathrm{Range}(A)$. By changing the symbols as $A \to A_N$, $\widetilde{\boldsymbol{s}} \to \boldsymbol{s}_N$ and $\boldsymbol{n}_2 \to A_G\boldsymbol{s}_G$, we will consider

$$\boldsymbol{x} = A_N\boldsymbol{s}_N + A_G\boldsymbol{s}_G \tag{3}$$

as our model fomulation, where A_G indicates the subspace and $\boldsymbol{s}_G$ denotes a $(d-m)$-dimensional Gaussian random vector. Independence of $\boldsymbol{s}_N$ and $\boldsymbol{s}_G$ implies that the non-Gaussian subspace and the Gaussian noise components are orthogonal with the metric Σ^{-1}, i.e. $A_N^\top \Sigma^{-1} A_G = 0$, where $\Sigma := \mathrm{Cov}(\boldsymbol{x})$.

Let $(B_N^\top, B_G^\top)^\top$ be the inverse matrix of (A_N, A_G). Then, the submatrices B_N and B_G extract the non-Gaussian and the Gaussian parts of the data $\boldsymbol{x}$, i.e. $B_N\boldsymbol{x} = \boldsymbol{s}_N$ and $B_G\boldsymbol{x} = \boldsymbol{s}_G$. The primal goal of dimension reduction in this paper is estimating the linear mapping B_N onto the non-Gaussian subspace in order to project out the irrelevant Gaussian components $\boldsymbol{s}_G$ and obtain the non-Gaussian signals $\boldsymbol{s}_N = B_N\boldsymbol{x}$. We remark that other matrices B_G, A_N and A_G can also be determined automatically, once B_N is derived. From independence of $\boldsymbol{s}_N$ and $\boldsymbol{s}_G$, the density function of $\boldsymbol{x}$ can be expressed as a product of the non-Gaussian and the Gaussian components

$$p(\boldsymbol{x}) = g(B_N\boldsymbol{x})\phi_L(B_G\boldsymbol{x}), \tag{4}$$

where g is an unknown function describing the density of $\boldsymbol{s}_N$ and ϕ_L is the Gaussian density with covariance L.

There still remain trivial indeterminacies in the model (3)

$$\boldsymbol{x} = (A_N C_1)(C_1^{-1}\boldsymbol{s}_N) + (A_G C_2)(C_2^{-1}\boldsymbol{s}_G), \tag{5}$$

where C_1 and C_2 are m- and $(d-m)$-dimensional square invertible matrices, respectively. Because of the indeterminacies (5) we should evaluate the results by $\mathcal{I} = \mathrm{Range}(B_N^\top)$ (called non-Gaussian index space here) rather than B_N itself. We recently proved that the decomposition (3) is unique up to this indeterminacies, if we assume that the dimension m of the non-Gaussian subspace is correct [12].

3 Joint Low-Rank Approximation Method

3.1 Dimension Reduction by Using Fourth-Order Cumulant Tensor

In our previous work [4], we propose a procedure for estimating the non-Gaussian subspace $\mathcal{I}$ based on the fourth-order cumulant tensor

$$\begin{aligned}&\mathrm{cum}(x_i, x_j, x_k, x_l)\\ &:= \mathbb{E}[x_i x_j x_k x_l] - \mathbb{E}[x_i x_j]\mathbb{E}[x_k x_l] - \mathbb{E}[x_i x_k]\mathbb{E}[x_j x_l] - \mathbb{E}[x_i x_l]\mathbb{E}[x_j x_k].\end{aligned}$$

The method was inspired by the JADE algorithm [2] for ICA which uses this tensor.

As is used in the JADE algorithm, we also apply the whitening transformation $\boldsymbol{z} = V^{-1/2}\boldsymbol{x}$ as preprocessing, where $V = \mathrm{Cov}[\boldsymbol{x}]$. Let us define the matrices

$$W_N := B_N V^{1/2}, \qquad W_G := B_G V^{1/2},$$

which are the linear transformations from the sphered data to the factors $\boldsymbol{s} = (\boldsymbol{s}_N^\top, \boldsymbol{s}_G^\top)^\top$. We remark that the non-Gaussian index space can be expressed as

$$\mathcal{I} = \mathrm{Range}(B_N^\top) = V^{-1/2}\mathrm{Range}(W_N^\top).$$

and therefore, it is enough to estimate the matrix W_N. Without loss of generality, we can assume that $\mathrm{Cov}[\boldsymbol{s}] = I$. Then, $(W_N^\top, W_G^\top)$ becomes an orthogonal matrix.

The method proposed in [4] rests on the fact that the cumulant tensor of the sources $(\boldsymbol{s}_N, \boldsymbol{s}_G)$ has simple structure. Let us order the sources as $\boldsymbol{s}_N = (s_1, \ldots, s_m)$ and $\boldsymbol{s}_G = (s_{m+1}, \ldots, s_d)$. The cumulant tensor $\mathrm{cum}(s_i, s_j, s_k, s_l)$ takes 0, unless $1 \le i, j, k, l \le m$ (i.e. all components should belong to the non-Gaussian part). Let $Q^{(kl)}$ be the matrix whose (i, j) element is $\mathrm{cum}(z_i, z_j, z_k, z_l)$ for all $1 \le k, l \le d$ and W° be a d-dimensional orthogonal matrix which recovers the sources, i.e. $\boldsymbol{s} = W^\circ \boldsymbol{z}$. Then, it can be proven that, for all (k, l),

$$W^\circ Q^{(kl)} (W^\circ)^\top = \begin{pmatrix} * & 0 \\ 0 & 0 \end{pmatrix}$$

holds, that is, all components which are not contained in the $m \times m$ submatrix $*$ vanish after the similar transformation by W°. This fact implies that we can estimate the transformation W_N° to the non-Gaussian components $\boldsymbol{s}_N$ by maximizing the Frobenius norms of the $m \times m$ submatrices corresponding to the non-Gaussian subspace

$$\mathcal{L}(W_N) = \sum_{k,l=1}^{d} \|W_N Q^{(kl)} W_N^\top\|_{\mathrm{Fro}}^2 = \sum_{k,l=1}^{d} \sum_{i',j'=1}^{m} \mathrm{cum}(y_{i'}, y_{j'}, z_k, z_l)^2 \qquad (6)$$

w.r.t. W_N s.t. $W_N W_N^\top = I_m$, where $\boldsymbol{y}_N = (y_1, \ldots, y_m)^\top = W_N \boldsymbol{z}$ denotes the reconstructed non-Gaussian components by W_N and $\|\cdot\|_{\mathrm{Fro}}^2$ is Frobenius norm of matrices. The contrast function (6) was optimized by iterative eigenvalue decomposition in [4],

$$W_N^{(t+1)} \sum_{k,l=1}^{d} \widehat{Q}^{(kl)} \{W_N^{(t)}\}^\top W_N^{(t)} \widehat{Q}^{(kl)} = \Lambda W_N^{(t+1)}, \tag{7}$$

where $\widehat{Q}^{(kl)}$ is the empirical correspondent of the matrix $Q^{(kl)}$ and $W_N^{(t)}$ is the t-step estimator.

The algorithm works well for sub-Gaussian structures. However, due to outliers it performs worse when the data contains heavy-tailed structures. In the remaining of this section, we will introduce joint low-rank approximation (JLA) of matrices by generalizing this method and show its global consistency. A novel algorithm using Hessian of the characteristic function will be proposed as an example.

3.2 Joint Low-Rank Approximation of Matrices

Let us consider the ideal situation as the discussion with the expected cumulant tensor in the previous section. Suppose that K complex matrices $M_1, \ldots, M_K$ can be simultaneously transformed into

$$W^\circ M_k (W^\circ)^\top = \begin{pmatrix} * & 0 \\ 0 & 0 \end{pmatrix}, \qquad k = 1, \ldots, K, \tag{8}$$

that is, all components of all the transformed matrices vanish except for those in $m \times m$ submatrices indicated by $*$, where W° is a d-dimensional orthogonal matrix. Let W_N° be the $m \times d$ matrix composed of the first m rows of W°. We remark that $W_N^\circ (W_N^\circ)^\top = I_m$. The goal here is to estimate the mapping W_N° as before.

Let us consider the contrast function

$$\mathcal{L}(W_N) = \sum_{k=1}^{K} \|W_N M_k W_N^\top\|_{\mathrm{Fro}}^2, \tag{9}$$

where Frobenius norm $\|C\|_{\mathrm{Fro}}^2 = \mathrm{tr}(CC^*)$ in complex case. We can show that the desired mapping W_N° can be obtained up to an orthogonal matrix by maximizing the contrast function $\mathcal{L}(W_N)$.

Theorem 1. *The objective function $\mathcal{L}(W_N)$ is maximal at $W_N = U W_N^\circ$, where W_N° is the first $m \times d$ submatrix of W° defined by Eq. (8) and U is an m-dimensional orthogonal matrix.*

Proof. We remark that Frobenius norm $\|W M_k W^\top\|_{\mathrm{Fro}}^2$ is unchanged for any orthogonal matrix W, i.e. $\|W M_k W^\top\|_{\mathrm{Fro}}^2 = \|M_k\|_{\mathrm{Fro}}^2 = \|W^\circ M_k (W^\circ)^\top\|_{\mathrm{Fro}}^2$. From the property (8) of the matrix W°, we get

$$\|W^\circ M_k (W^\circ)^\top\|_{\mathrm{Fro}}^2 = \left\| \begin{pmatrix} W_N^\circ M_k (W_N^\circ)^\top & 0 \\ 0 & 0 \end{pmatrix} \right\|_{\mathrm{Fro}}^2 = \left\| W_N^\circ M_k (W_N^\circ)^\top \right\|_{\mathrm{Fro}}^2,$$

where we divided W° into two submatrices W_N° and W_G°. On the other hand, for a general orthogonal matrix W,

$$\begin{aligned}\|WM_kW^\top\|_{\mathrm{Fro}}^2 &= \left\| \begin{pmatrix} W_NM_kW_N^\top & W_NM_kW_G^\top \\ W_GM_kW_N^\top & W_GM_kW_G^\top \end{pmatrix} \right\|_{\mathrm{Fro}}^2 \\ &= \|W_NM_kW_N^\top\|_{\mathrm{Fro}}^2 + \|W_NM_kW_G^\top\|_{\mathrm{Fro}}^2 \\ &\qquad + \|W_GM_kW_N^\top\|_{\mathrm{Fro}}^2 + \|W_GM_kW_G^\top\|_{\mathrm{Fro}}^2 \\ &\geq \|W_NM_kW_N^\top\|_{\mathrm{Fro}}^2,\end{aligned}$$

where W was also divided into two submatrices W_N and W_G. Since the inequality is strict if and only if one of the three terms is non-zero, we notice that all global maxima were already found. Therefore, for all orthogonal matrices U, finally we get

$$\|UW_N^\circ M_k(W_N^\circ)^\top U^\top\|_{\mathrm{Fro}}^2 = \|W_N^\circ M_k(W_N^\circ)^\top\|_{\mathrm{Fro}}^2 \geq \|W_NM_kW_N^\top\|_{\mathrm{Fro}}^2. \qquad \square$$

For simplicity, we further assume that $M_k^\top = M_k$, as is the case with our algorithms. By differentiating the criterion $\mathcal{L}$ under the constraint $W_NW_N^\top = I_m$, we get

$$W_N \sum_{k=1}^{K} \mathcal{M}_k(W_N) = \Lambda W_N, \tag{10}$$

where

$$\mathcal{M}_k(W_N) := M_kW_N^\top W_NM_k^* + M_k^*W_N^\top W_NM_k \tag{11}$$

is a $d \times d$ matrix depeding on W_N and Lagrange multipliers Λ is assumed to be diagonal without loss of generality. As the algorithm with the cumulant tensor, the maximization of the contrast function (9) can be solved by iterating the eigenvalue problem

$$W_N^{(t+1)} \sum_{k=1}^{K} \widehat{\mathcal{M}}_k(W_N^{(t)}) = \Lambda W_N^{(t+1)} \tag{12}$$

where $\widehat{\mathcal{M}}_k$ is the empirical correspondent of the matrix $\mathcal{M}_k$ and $W_N^{(t)}$ is the t-step estimator.

3.3 Dimension Reduction by Using Characteristic Functions

In [10] and [11], Hessians of the characteristic function were used for (multidimensional) independent component analysis. Since they satisfy the property (8) under our model assumption as we will show, they can also be used as the matrices M_k in the joint low-rank approximation procedure. The characteristic function of the random variable $\boldsymbol{z}$ can be defined by $\widehat{Z}(\boldsymbol{\zeta}) := \mathbb{E}[\exp(i\boldsymbol{\zeta}^\top \boldsymbol{z})]$. Let $W^\circ = ((W_N^\circ)^\top, (W_G^\circ)^\top)^\top$ be an orthogonal matrix s.t. $\boldsymbol{s} = W^\circ \boldsymbol{z}$. Then, the characteristic function can be expressed as

$$\widehat{Z}(\boldsymbol{\zeta}) = \widehat{S}(W^\circ\boldsymbol{\zeta}) = \widehat{S}_N(W_N^\circ\boldsymbol{\zeta}) \exp\left(-\frac{1}{2}\|W_G^\circ\boldsymbol{\zeta}\|^2\right), \tag{13}$$

where $\widehat{S}$ and $\widehat{S}_N$ are the characteristic functions of $\boldsymbol{s}$ and $\boldsymbol{s}_N$, respectively. Therefore, if $\log \widehat{Z}(\boldsymbol{\zeta})$ exists, the Hessian of $\log \widehat{Z}(\boldsymbol{\zeta})$ becomes

$$H_{\log \widehat{Z}}(\boldsymbol{\zeta}) := \frac{\partial^2}{\partial \boldsymbol{\zeta} \partial \boldsymbol{\zeta}^\top} \log \widehat{Z}(\boldsymbol{\zeta}) = (W^\circ)^\top \begin{pmatrix} \frac{\partial^2}{\partial \boldsymbol{\xi}_N \partial \boldsymbol{\xi}_N^\top} \log \widehat{S}_N(W_N^\circ \boldsymbol{\zeta}) & 0 \\ 0 & -I_{d-m} \end{pmatrix} W^\circ, \tag{14}$$

where $\boldsymbol{\xi}_N = W_N^\circ \boldsymbol{\zeta}$. For K selected vectors $\boldsymbol{\zeta}_1, \ldots, \boldsymbol{\zeta}_K \in \mathbb{R}^d$, each matrix $M_k := H_{\log \widehat{Z}}(\boldsymbol{\zeta}_k) + I_d$ satisfies the property (8).

Suppose that samples $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ are given. The algorithm with Hessians of the characteristic function is summarized as follows.

Algorithm

1. Sphere the data $\{\boldsymbol{x}_i\}_{j=1}^n$ by $\widehat{\boldsymbol{z}}_j = \widehat{V}^{-1/2} \boldsymbol{x}_j$, where $\widehat{V} = \widehat{\mathrm{Cov}}[\boldsymbol{x}]$.
2. Calculate the Hessian $\widehat{M}_k := \widehat{H}_{\log \widehat{Z}}(\boldsymbol{\zeta}_k) + I_d$ at selected vectors $\boldsymbol{\zeta}_k$ from the empirical characteristic function $\widehat{Z}_{\mathrm{emp}}(\boldsymbol{\zeta}) = \frac{1}{n} \sum_{j=1}^n \exp(i \boldsymbol{\zeta}^\top \widehat{\boldsymbol{z}}_j)$.
3. Compute m eigenvectors with largest absolute eigenvalues.

$$W_N^{(0)} \sum_{k=1}^K \left\{ \mathrm{Re}(\widehat{M}_k) + \mathrm{Im}(\widehat{M}_k) \right\} = \Lambda W_N^{(0)}$$

4. Solve the following eigenvalue problem until the matrix $W_N^{(t)}$ converges.

$$W_N^{(t+1)} \sum_{k=1}^K \widehat{\mathcal{M}}_k(W_N^{(t)}) = \Lambda W_N^{(t+1)}$$

The symbols with hat denote the empirical versions of the corresponding quantities, for example, $\widehat{\mathrm{Cov}}$ is the sample covariance. $\mathrm{Re(M)}$ and $\mathrm{Im(M)}$ are the real and the imaginary parts of a matrix M. The matrix $\widehat{\mathcal{M}}_k(W_N)$ is calsulated from $\widehat{M}_k$ by Eq. (11).

4 Numerical Results

For testing our algorithm, we performed numerical experiments using various synthetic data used in [1]. Each data set includes 1000 samples in 10 dimension. Each sample consists of 8-dimensional independent standard Gaussian and 2 non-Gaussian components as follows.

(A) Simple: 2-dimensional independent Gaussian mixtures with density of each component given by $\frac{1}{2}\phi_{-3,1}(x) + \frac{1}{2}\phi_{3,1}(x)$.

(B) Dependent super-Gaussian: 2-dimensional isotropic distribution with density proportional to $\exp(-\|x\|)$.

(C) Dependent sub-Gaussian: 2-dimensional isotropic uniform with constant positive density for $\|x\| \leq 1$ and 0 otherwise.

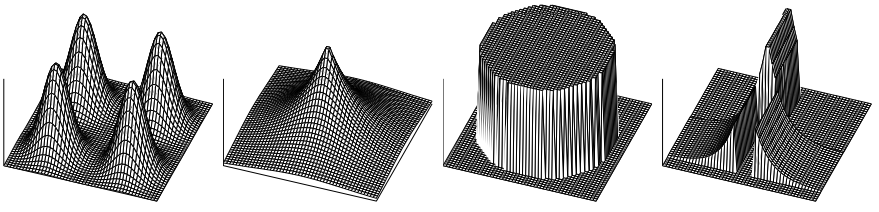
(D) Dependent super- and sub-Gaussian: 1-dimensional Laplacian with density proportional to $\exp(-|x_{Lap}|)$ and 1-dimensional dependent uniform $U(c, c+1)$, where $c = 0$ for $|x_{Lap}| \leq \log 2$ and $c = -1$ otherwise.

The profiles of the density functions of the non-Gaussian components in the above data sets are described in Fig. 1. The mean and standard deviation of samples are normalized to zero and one in a component-wise manner.

Besides the proposed algorithm, we applied for reference the following four methods in the experiments: FastICA with 'pow3' or 'tanh' index (denoted by FIC3 and FICt, respectively), JADE and joint low-rank approximation (JLA) algorithms with the fourth-order cumulant tensor and Hessian of the characteristic function (denoted by JLA4 and JLAH, respectively). In JLA with Hessian, 1000 vectors $\boldsymbol{\zeta}$ were randomly chosen and 10% of them with high norm $\|M_k\|_{\mathrm{Fro}}$ were taken in the contrast function. We remark that we did not include the better method [1], because the main purpose of the experiments is compareing the two JLA algorithm. Further research is necessary to improve the algorithm. In FastICA and JLA with the cumulant tensor, additionally 9 runs from random initial matrices were also carried out and the optimum among these 10 solutions were chosen to avoid local optima.

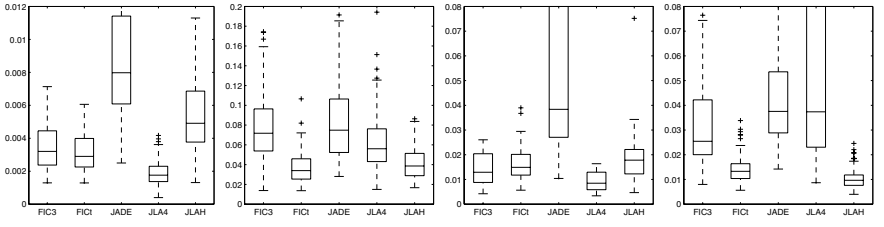
Fig. 2 shows boxplots of the error criterion

$$\mathcal{E}(\widehat{\mathcal{I}}, \mathcal{I}) = \frac{1}{m} \|(I_d - \Pi_{\mathcal{I}})\Pi_{\widehat{\mathcal{I}}}\|^2_{\mathrm{Fro}}, \tag{15}$$



(A) (B) (C) (D)

Fig. 1. Densities of non-Gaussian components. The datasets are: (a) 2D independent Gaussian mixtures, (b) 2D isotropic super-Gaussian, (c) 2D isotropic uniform and (d) dependent 1D Laplacian + 1D uniform.



(A) (B) (C) (D)

Fig. 2. Boxplots of the error criterion $\mathcal{E}(\widehat{\mathcal{I}}, \mathcal{I})$. Algorithms are FIC3, FICt, JADE, JLA4 and JLAH (from left to right).

obtained from 100 runs, where $\Pi_{\mathcal{I}}$ (resp. $\Pi_{\widehat{\mathcal{I}}}$) is the projection matrix onto the true non-Gaussian subspace $\mathcal{I}$ (resp. the estimated one $\widehat{\mathcal{I}}$).

Although we did not prove theoretically, JADE could find the non-Gaussian subspace $\mathcal{I}$ in all these examples. Unfortunately, the performance of the proposed algorithm JLAH was worse than that of the previous version JLA4 for the simple data (A) and was on par for the sub-Gaussian data (C). However, for data (B) and (D) which contain super-Gaussian structures, the Hessian version JLAH outperformed the cumulant one JLA4. Moreover, JLAH was much more robust than JLA4. The proposed algorithm (JLAH) missed only one case, while the latter (JLA4) failed to estimate the index space $\mathcal{I}$ many times ((B)7%, (C)21% and (D)30%).

5 Conclusions

In this paper, we proposes a new *linear* method to identify the non-Gaussian subspace based on Hessian of the characteristic function. In our numerical experiments, the proposed algorithm was more robust and efficient than the previous version with the cumulant tensor when data contain super-Gaussian structures. Global consistency of the method was also proved in a more general framework. Further research should be done on selection of the vectors $\boldsymbol{\zeta}_k$ to improve its performance. Other examples of joint low-rank approximation procedure can also be interesting.

References

1. G. Blanchard, M. Kawanabe, M. Sugiyama, V. Spokoiny and K.-R. Müller. In search of non-Gaussian components of a high-dimensional distribution. submitted to *Journal of Machine Learning Research*.
2. J.-F. Cardoso and A. Souloumiac. Blind beamforming for non Gaussian signals. *IEE Proceedings-F*, 140(6):362–370, 1993.
3. J.H. Friedman and J.W. Tukey. A projection pursuit algorithm for exploratory data analysis. *IEEE Transactions on Computers*, 23(9):881–890, 1975.
4. M. Kawanabe. Linear dimension reduction based on the fourth-order cumulant tensor. *Proc. ICANN 2005*, LNCS 3697: 151-156, Warsaw, Poland, 2005.
5. P.J. Huber. Projection pursuit. *The Annals of Statistics*, 13:435–475, 1985.
6. A. Hyvärinen. Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions on Neural Networks*, 10(3):626–634, 1999.
7. A. Hyvärinen, J. Karhunen and E. Oja. *Independent Component Analysis*. Wiley, 2001.
8. S. Roweis and L. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500):2323–2326, 2000.
9. J.B. Tenenbaum, V. de Silva and J.C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323, 2000.
10. F.J. Theis. A new concept for separability problems in blind source separation. *Neural Computation*, 16: 1827-1850, 2004
11. F.J. Theis. Multidimensional independent component analysis using characte ristic functions. *Proc. EUSIPCO 2005*, Antalya, Turkey, 2005.
12. F.J. Theis and M. Kawanabe. Uniqueness of non-Gaussian component analysis. submitted to ICA 2006.