MULTIDIMENSIONAL INDEPENDENT COMPONENT ANALYSIS USING CHARACTERISTIC FUNCTIONS

Fabian J. Theis

Institute of Biophysics, University of Regensburg, 93040 Regensburg, Germany
phone: +49 941 943 2924, fax: +49 941 943 2479, email: fabian@theis.name

ABSTRACT

The goal of multidimensional independent component analysis (MICA) lies in the linear separation of data into statistically independent groups of signals. In this work, we give an elementary proof for the uniqueness of this problem in the case of equally sized subspaces, showing that the separation matrix is essentially unique except for row permutation and scaling. The proof is based on the reinterpretation of groupwise independence as factorization of the joint characteristic function. We then employ this property to propose a novel algorithm for robustly performing MICA. Simulation results demonstrate the reliability of our method.

1. INTRODUCTION

Multidimensional blind source separation (MBSS) denotes the problem of recovering underlying sources $\mathbf{S}$ from an observed mixture $\mathbf{X}$. As usual, $\mathbf{S}$ has to fulfill additional properties such as independence or diagonality of the autocovariances (if $\mathbf{S}$ possesses time structure). However in contrast to ordinary BSS, MBSS is more general as some source signals are allowed to possess common statistics. One possible model for MBSS is *multidimensional independent component analysis (MICA)* [3]. The idea MICA is that we do not require full independence of the transform $\mathbf{S} := \mathbf{W}\mathbf{X}$ but only mutual independence of certain tuples $S_{i_1}, \ldots, S_{i_2}$. If the size of all tuples is restricted to one, this reduces to ordinary ICA. In general, of course the tuples could have different sizes, but we will restrict ourselves to the simpler case of equally sized tuples of length m.

In this work, we first calculate the indeterminacies of MICA, extending results already presented in [9] and [11] for real and complex ICA. Based on the uniqueness proof we are then able to propose a novel MICA algorithm by reinterpreting the groupwise independence in terms of factorization properties of the characteristic functions. The algorithm is related to previous work in the ordinary ICA case [7, 13]. But instead of using blockwise joint diagonalization, we employ a generalization of the characteristic-function based algorithm proposed by [6].

2. MULTIDIMENSIONAL ICA

Introducing some notation, let us define for $r,s = 1,\ldots,n$ the (r,s) *m-submatrix of* $\mathbf{W} = (w_{ij})$, denoted by $\mathbf{W}^{(m)}_{rs}$, to be the $m \times m$ submatrix of $\mathbf{W}$ ending at position (rm, sm). Denote $\mathrm{Gl}(n)$ the group of invertible $n \times n$ matrices. A matrix $\mathbf{W} \in \mathrm{Gl}(mn)$ is said to be an *m-scaling matrix* if $\mathbf{W}^{(m)}_{rs} = 0$ for $r \neq s$, and $\mathbf{W}$ is called an *m-permutation matrix* if for each $r = 1,\ldots,n$ there exists precisely one s such that $\mathbf{W}^{(m)}_{rs}$ equals the $m \times m$ unit matrix and for each s there exists one r with that property. Let $m,n \in \mathbb{N}$. We call an mn-dimensional random vector $\mathbf{S}$ *m-independent* if the m-dimensional random vectors $(S_1,\ldots,S_m)^\top,\ldots,(S_{mn-m+1},\ldots,S_{mn})^\top$ are mutually independent. A matrix $\mathbf{W} \in \mathrm{Gl}(mn)$ is called a *m-multidimensional ICA* of an mn-dimensional random vector $\mathbf{X}$ if $\mathbf{W}\mathbf{X}$ is m-independent. If $m = 1$, this is the same as ordinary ICA.

Using MICA we want to solve the (noiseless) linear MBSS problem $\mathbf{X} = \mathbf{A}\mathbf{S}$, where the mn-dimensional random vector $\mathbf{X}$ is given, and $\mathbf{A} \in \mathrm{Gl}(mn)$ and $\mathbf{S}$ are unknown. In the case of MICA $\mathbf{S}$ is assumed to be m-independent.

3. INDETERMINACIES

Obvious indeterminacies of MICA are, similar to ordinary ICA, invertible transforms in $\mathrm{Gl}(m)$ in each tuple as well as the fact that the order of the independent m-tuples is not fixed. Indeed, if $\mathbf{A}$ is MBSS solution, then so is $\mathbf{ALP}$ with a m-scaling matrix $\mathbf{L}$ and a m-permutation $\mathbf{P}$, because m-independence is invariant under these transformations.

3.1 Uniqueness up to blockwise permutation and scaling

Let $\Pi(n, \mathrm{Gl}(m,\mathbb{R}))$ denote the group of all $mn \times mn$-matrices $\mathbf{A}$ such that for each k precisely one m-submatrix in the row $\mathbf{A}^{(m)}_{k,\cdot}$ and one m-submatrix in the column $\mathbf{A}^{(m)}_{\cdot,k}$ is invertible and the rest are zero. This is a subgroup of $\mathrm{Gl}(mn,\mathbb{R})$ and consists of all products of m-scaling and m-permutation matrices.

We want to show that m-independence under linear transformation only allows matrices from $\Pi(n, \mathrm{Gl}(m,\mathbb{R}))$, which proves separability of MICA. However, for the proof we need one more condition for $\mathbf{A}$: We call $\mathbf{A}$ *m-admissible* if for each $r,s = 1,\ldots,n$ the (r,s) m-submatrix of $\mathbf{A}$ is either invertible or zero. Note that this is not a strong restriction — if we randomly choose $\mathbf{A}$ with coefficients out of a continuous distribution, then with probability one we get an m-admissible matrix, because the non-m-admissible matrices $\subset \mathbb{R}^{m^2n^2}$ lie in a submanifold of dimension smaller than m^2n^2.

Theorem 3.1 (Uniqueness of MICA)**.** *Let* $\mathbf{S}$ *be an m-independent nm-dimensional square integrable random vector of positive definite covariance matrix having no normally distributed m-tuple* $(S_{rm-m+1},\ldots,S_{rm})^\top$*, and let* $\mathbf{A} \in \mathrm{Gl}(mn,\mathbb{R})$ *such that* $\mathbf{A}^{-1}$ *is m-admissible. Then* $\mathbf{AS}$ *is m-independent if and only if* $\mathbf{A} \in \Pi(n, \mathrm{Gl}(m,\mathbb{R}))$*.*

For the case $m = 1$ and $m = 2$ this theorem shows uniqueness of real- respectively complex-valued ICA, because every matrix is 1-admissible, and every complex matrix 2-admissible when interpreted as real-valued matrix. The condition prohibiting any normally-distributed m-tuple is sufficient as stated by the theorem; however necessity of this condition is not clear. In the one-dimensional case it is well known that a single normal source can be allowed and separability still holds [4]. This proof uses preprocessing by PCA to allow for an orthogonal matrix — a concept that seems to be difficult to extend to the case of arbitrary m.

The above theorem has been derived by Comon in the case $m = 1$ [4] from the Darmois-Skitovitch theorem [5, 8]. In [10] we have given an extension to multidimensional ICA (arbitrary m) also based on the D-S theorem. In the following we will present a proof that does not need this theorem. Furthermore, based on the proof we are able to propose a novel algorithm for MICA.

3.2 Proof

Since the statement of the theorem trivially holds for $n = 1$, we will assume $n \geq 2$.

Definition 3.2. *A function* $f : \mathbb{R}^{mn} \to \mathbb{C}$ *is said to be* m-separated *if there exist m-dimensional functions* $g_1,\ldots,g_n : \mathbb{R}^m \to \mathbb{C}$ *such* *that $f(x_1,\dots,x_{mn}) = g_1(x_1,\dots,x_m)\cdots g_n(x_{mn-m+1},\dots,x_{mn})$ for all $(x_1,\dots,x_{mn}) \in \mathbb{R}^{mn}$, in short $f \equiv g_1 \otimes \cdots \otimes g_n$.*

Note that the functions g_i are uniquely determined by f up to a scalar factor. Obviously the character and the density (if it exists) of an independent random vector are 1-separated or simply separated.

Remark 3.3. *If $f \in \mathscr{C}^2(\mathbb{R}^{mn},\mathbb{C})$ is m-separated, then $f\frac{\partial^2 f}{\partial x_i \partial x_j} - \frac{\partial f}{\partial x_i}\frac{\partial f}{\partial x_j} \equiv 0$ for $\lfloor \frac{i}{m} \rfloor \neq \lfloor \frac{j}{m} \rfloor$.*

If f is strictly positive, then the condition from remark 3.3 is equivalent to $\ln f$ having a blockwise-diagonal Hessian everywhere.

Lemma 3.4. *Let $\mathbf{X}$ be a m-dimensional random vector with twice continuously differentiable characteristic function $\widehat{\mathbf{X}}(\mathbf{x}) := \mathbf{E}(\exp i\mathbf{x}^\top \mathbf{X})$ satisfying*

$$\mathbf{C}\widehat{\mathbf{X}}^2 - \widehat{\mathbf{X}}\mathbf{H}_{\widehat{\mathbf{X}}} + \nabla\widehat{\mathbf{X}}(\nabla\widehat{\mathbf{X}})^\top \equiv 0. \tag{1}$$

for a constant matrix $\mathbf{C} \in \mathrm{Mat}(m \times m, \mathbb{C})$. Then $\mathbf{X}$ is normally distributed

Here ∇f denotes the gradient of f and $\mathbf{H}_f$ its Hessian.

Proof. We first show that the differential equation 1 locally at non-zeros of $\widehat{\mathbf{X}}$ has the solution $\exp g$, where g is a m-dimensional polynomial of degree ≤ 2. For this note that $\widehat{\mathbf{X}}(0) = 1$ by definition, so there exists an non-empty open set U containing 0 such that a complex logarithm log is defined on $\widehat{X}(U)$. Set $g := \log \widehat{X}|U$. Substituting $\exp g$ for $\widehat{X}$ in equation 1 yields equations

$$c_{ij}\exp(2g) - \exp(g)\left(\frac{\partial^2 g}{\partial x_i \partial x_j} + \frac{\partial g}{\partial x_i}\frac{\partial g}{\partial x_j}\right)\exp(g) + \frac{\partial g}{\partial x_i}\frac{\partial g}{\partial x_j}\exp(2g) \equiv 0$$

for $i,j \in \{1,\dots,m\}$, so $\partial^2 g/\partial x_i \partial x_j \equiv c_{ij}$. Hence g is a polynomial of degree ≤ 2, and $\widehat{\mathbf{X}} = \exp g \neq 0$ on all of $\overline{U}$. Therefore $\widehat{\mathbf{X}} \neq 0$ everywhere because of continuity.

The local argument from above then shows that $\widehat{\mathbf{X}}(\mathbf{x}) = \exp(\frac{1}{2}\sum_{ij} a_{ij}x_i x_j + \sum_i b_i x_i)$ everywhere, where we have already used $\widehat{\mathbf{X}}(0) = 1$. Moreover, from $\widehat{\mathbf{X}}(-\mathbf{x}) = \overline{\widehat{\mathbf{X}}(\mathbf{x})}$ we get $a_{ij} \in \mathbb{R}, a_{ij} = a_{ji}$ and $\mathbf{b} = i\mu$ with real $\mu \in \mathbb{R}^m$. And $|\widehat{X}| \leq 1$ shows that $\mathbf{A} = (a_{ij})$ is negative semidefinite. Altogether, with $\Gamma := -\mathbf{A}$ we get that

$$\widehat{\mathbf{X}}(\mathbf{x}) = \exp\left(i\mu^\top \mathbf{x} - \frac{1}{2}\mathbf{x}^\top \Gamma \mathbf{x}\right)$$

which means that $\mathbf{X}$ is normally distributed with mean μ and (possibly singular) covariance Γ. □

In the following, we will study the properties of m-separated functions under linear transformation.

Lemma 3.5. *Let $g_i \in \mathscr{C}^2(\mathbb{R}^m,\mathbb{C}), g_i \not\equiv 0$ and $\mathbf{B} \in \mathrm{Gl}(nm,\mathbb{R})$ such that $g_1 \otimes \cdots \otimes g_n(\mathbf{B}\mathbf{x})$ is m-separated. If $\mathbf{B}$ has two invertible blocks in the same row, i.e. if there exist indices l and $i \neq j$ with $\mathbf{B}^{(m)}_{li}, \mathbf{B}^{(m)}_{lj} \in \mathrm{Gl}(m,\mathbb{R})$, then g_l satisfies the differential equation 1.*

Proof. $f(\mathbf{x}) := g_1 \otimes \cdots \otimes g_n(\mathbf{B}\mathbf{x})$ is assumed to be m-separated, so by remark 3.3 we get for indices i,j from different blocks ($\lfloor \frac{i}{m} \rfloor \neq \lfloor \frac{j}{m} \rfloor$):

$$0 = \left(f\frac{\partial^2 f}{\partial x_i \partial x_j} - \frac{\partial f}{\partial x_i}\frac{\partial f}{\partial x_j}\right)(\mathbf{x})$$

The ingredients of this equation can be calculated as follows:

$$\begin{aligned}
\frac{\partial f}{\partial x_i}(\mathbf{x}) &= \sum_{k=1}^n g_1 \otimes \cdots \otimes \frac{\partial g_k}{\partial x_i} \otimes \cdots \otimes g_n(\mathbf{B}\mathbf{x}) \\
\frac{\partial f}{\partial x_i}\frac{\partial f}{\partial x_j}(\mathbf{x}) &= \sum_{k,l}(g_1 \otimes \cdots \otimes \frac{\partial g_k}{\partial x_i} \otimes \cdots \otimes g_n) \\
&\quad (g_1 \otimes \cdots \otimes \frac{\partial g_l}{\partial x_j} \otimes \cdots \otimes g_n)(\mathbf{B}\mathbf{x}) \\
\frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x}) &= \sum_k (g_1 \otimes \cdots \otimes \frac{\partial^2 g_k}{\partial x_i \partial x_j} \otimes \cdots \otimes g_n + \\
&\quad \sum_{l \neq k} g_1 \otimes \cdots \otimes \frac{\partial g_k}{\partial x_i} \otimes \cdots \otimes \frac{\partial g_l}{\partial x_j} \otimes \cdots \otimes g_n)(\mathbf{B}\mathbf{x})
\end{aligned}$$

Plugging this into the above equation yields

$$\begin{aligned}
0 &= \sum_k (g_1^2 \otimes \cdots \otimes g_k \frac{\partial^2 g_k}{\partial x_i \partial x_j} \otimes \cdots \otimes g_n^2 - \\
&\quad g_1^2 \otimes \cdots \otimes \frac{\partial g_k}{\partial x_i}\frac{\partial g_k}{\partial x_j} \otimes \cdots \otimes g_n^2)(\mathbf{B}\mathbf{x}) \\
&= \sum_k g_1^2 \otimes \cdots \otimes g_{k-1}^2 \otimes \left(g_k \frac{\partial^2 g_k}{\partial x_i \partial x_j} - \frac{\partial g_k}{\partial x_i}\frac{\partial g_k}{\partial x_j}\right) \otimes \\
&\quad g_{k+1}^2 \otimes \cdots \otimes g_n^2(\mathbf{B}\mathbf{x})
\end{aligned}$$

for $\mathbf{x} \in \mathbb{R}^{mn}$. We want to calculate the term in the brackets. For this note that $\frac{\partial}{\partial x_i} g_k(\mathbf{B}\mathbf{x}) = \mathbf{b}^{(k,i)\top}\nabla g_k \mid_{\mathbf{B}\mathbf{x}}$ with $\mathbf{b}^{(k,i)\top} := (b_{km-m+1,i},\dots,b_{km,i})$. So, the term in the brackets can be calculated as $(g_k \frac{\partial^2 g_k}{\partial x_i \partial x_j} - \frac{\partial g_k}{\partial x_i}\frac{\partial g_k}{\partial x_j})(\mathbf{B}\mathbf{x}) = \mathbf{b}^{(k,i)\top}(g_k H_{g_k} - \nabla g_k(\nabla g_k)^\top) \mid_{\mathbf{B}\mathbf{x}} \mathbf{b}^{(k,j)} =: h_{ijk}(\mathbf{B}\mathbf{x})$ and we get

$$0 = \sum_k g_1^2 \otimes \cdots \otimes g_{k-1}^2 \otimes h_{ijk} \otimes g_{k+1}^2 \otimes \cdots \otimes g_n^2(\mathbf{B}\mathbf{x})$$

$\mathbf{B}$ is invertible, so the whole function is zero:

$$0 \equiv \sum_k g_1^2 \otimes \cdots \otimes g_{k-1}^2 \otimes h_{ijk} \otimes g_{k+1}^2 \otimes \cdots \otimes g_n^2 \tag{2}$$

Choose $\mathbf{x} \in \mathbb{R}^{mn}$ with $g_k(x_{mk-m+1}, x_{mk}) \neq 0$ for $k = 1,\dots,n$. Evaluating equation 2 at $(x_1,\dots,x_{m(l-1)},\mathbf{y},x_{ml+1},\dots,x_{mn})$ for variable $\mathbf{y} \in \mathbb{R}^m$ and dividing the resulting equation by the constant $g_1^2(x_1,\dots,x_m)\cdots g_{l-1}^2(x_{ml-2m+1},\dots,x_{m(l-1)})$ $g_{l+1}^2(x_{ml+1},\dots,x_{m(l+1)})\cdots g_n^2(x_{mn-m+1},x_{mn})$ shows

$$h_{ijl}(\mathbf{y}) = -\left(\sum_{k \neq l} h_{ijk}(x_{mk-m+1},x_{mk})\right) g_l^2(\mathbf{y}) =: c_{ijl}g_l^2(\mathbf{y}) \tag{3}$$

for $\mathbf{y} \in \mathbb{R}^m$.

Now let $i \neq j$ be indices in $\{1,\dots,n\}$. Equation 3 for $(im-m+1, jm-m+1, l),\dots,(im, jm, l)$ can be gathered into a matrix to read

$$\mathbf{B}^{(m)\top}_{li}\left(g_l H_{g_l} - \nabla g_l(\nabla g_l)^\top\right)\mathbf{B}^{(m)}_{lj} \equiv \mathbf{C}g_l^2$$

If now the two m-submatrices of $\mathbf{B}$ in this equation are invertible, then

$$g_l H_{g_l} - \nabla g_l(\nabla g_l)^\top \equiv \mathbf{C}' g_l^2,$$

so g_l fulfills precisely the differential equation from lemma 3.4. □

Proof of theorem 3.1. Assume that $\mathbf{A}$ and hence $\mathbf{B} := \mathbf{A}^{-1} \notin \Pi(n, \mathrm{Gl}(m,\mathbb{R}))$. Then there exist indices l and $i \neq j$ such that the (l,i) and (l,j) m-submatrices of $\mathbf{B}$ are non-zero (hence in $\mathrm{Gl}(m,\mathbb{R})$ by m-admissability). Applying lemma 3.5 and then lemma 3.4 to the m-separated characteristic function $\widehat{\mathbf{S}}(\mathbf{B}\mathbf{x})$ therefore shows that $(S_{lm-m+1},\dots,S_{lm})^\top$ is normally distributed, which is a contradiction. □

4. AN MICA ALGORITHM USING CHARACTERISTIC FUNCTIONS

In this section, we derive an algorithm for performing MICA from the ideas presented in the proof above.

4.1 Joint block diagonalization

Joint diagonalization has become an important tool in ICA-based BSS or in BSS relying on second-order time-decorrelation. The task of (real) *joint diagonalization* is, given a set of commuting symmetric $n \times n$ matrices $\mathbf{M}_i$, to find an orthogonal matrix $\mathbf{E}$ such that $\mathbf{E}^\top \mathbf{M}_i \mathbf{E}$ is diagonal for all i.

In the following we will use a generalization of this technique as algorithm to solve MBSS problems. Instead of fully diagonalizing $\mathbf{M}_i$ in *joint block diagonalization (JBD)* we want to determine $\mathbf{E}$ such that $\mathbf{E}^\top \mathbf{M}_i \mathbf{E}$ is block-diagonal (after fixing the block-structure).

Fixing the block-size to m, JBD tries to find $\mathbf{E}$ such that $\mathbf{E}^\top \mathbf{M}_i \mathbf{E}$ is a m-scaling matrix. In practice due to estimation errors, such $\mathbf{E}$ will not exist, so we speak of approximate JBD and imply minimizing some error-measure on non-block-diagonality.

Various algorithms to actually perform JBD have been proposed, see [1] and references therein. In the following we will simply perform joint diagonalization (using for example the Jacobi-like algorithm from [2]) and then permute the columns of $\mathbf{E}$ to achieve block-diagonality — in experiments this turns out to be an efficient solution to JBD [1].

4.2 MICA using block-diagonalization of the Hessian of the characteristic function

We assume that the sources have existing non-singular covariance.

In the first step, we preprocess the observations $\mathbf{X}$ by whitening. Hence we may assume that both $\mathrm{Cov}(\mathbf{X}) = \mathbf{I}$ and $\mathrm{Cov}(\mathbf{S}) = \mathbf{I}$, the latter due to the scaling invariance of the BSS problem. Then $\mathbf{I} = \mathrm{Cov}(\mathbf{X}) = \mathbf{A}\,\mathrm{Cov}(\mathbf{S})\mathbf{A}^\top = \mathbf{A}\mathbf{A}^\top$ so $\mathbf{A}$ is orthogonal.

Consider now the characteristic function $\widehat{\mathbf{S}}$ of the sources $\mathbf{S}$. By assumption this function is twice continuously differentiable, then so is $\log \widehat{\mathbf{S}}$, well defined in a neighborhood $U \subset \mathbb{C}^{mn}$ of 0, because of $\widehat{\mathbf{S}}(0) = 1$. In lemma 3.4 and implicitly in 3.5, we used the fact that the character of an m-independent random vector is m-separated, and hence $\log \widehat{\mathbf{S}}$ is the sum of functions, depending on m separate variables each. Hence if we compute its Hessian $\mathbf{H}_{\log \widehat{\mathbf{S}}} : U \to \mathbb{C}^{(mn)\times(mn)}$, it is m-block-diagonal. This key observation has been used previously in [7, 9, 13] to separate mixtures in the case $m = 1$. The logarithmic character $\log \widehat{\mathbf{S}}$ is sometimes called second characteristic function, and we will use some of its properties in the following.

We note that the Hessian transforms like a 2-tensor. Using $\widehat{\mathbf{AS}}(\mathbf{x}) = \widehat{\mathbf{S}}(\mathbf{A}^\top \mathbf{x})$, we get locally at 0

$$\mathbf{H}_{\log \widehat{\mathbf{X}}}(\mathbf{x}) = \mathbf{H}_{\log \widehat{\mathbf{S}} \circ \mathbf{A}^\top}(\mathbf{x}) = \mathbf{A}^\top \mathbf{H}_{\log \widehat{\mathbf{S}}}(\mathbf{A}^\top \mathbf{x})\mathbf{A} \qquad (4)$$

The idea with respect to computation now lies in the fact that the above equation represents a m-block-factorization of $\mathbf{H}_{\log \widehat{\mathbf{X}}}$.

Characteristic-function based Multidimensional ICA (cfMICA) now simply uses the block-diagonality structure from equation 4 and performs JBD of estimates of a set of Hessians $\mathbf{H}_{\log \widehat{\mathbf{X}}}(\mathbf{x}_i)$ evaluated at different points $\mathbf{x}_i \in \mathbb{C}^{mn}$ sufficiently close to 0. Given slight restrictions to the eigenvalues, the resulting block diagonalizer then equals $\mathbf{A}^\top$ except for m-scaling and permutation. For uniqueness conditions we refer to [9, 13].

The characteristic function and its logarithmic Hessians can be estimated nicely from the data [13]: The expectation operator is denoted by $E(\mathbf{X}) \in \mathbb{R}^{mn}$. If N realizations i.e. samples $\mathbf{X}(1), \ldots, \mathbf{X}(N)$ of $\mathbf{X}$ are given, E is estimated by the sample mean $\frac{1}{N}\sum_i \mathbf{X}(i)$ as usual. The characteristic function itself can be consistently estimated by

$$\widehat{\mathbf{X}}(\mathbf{x}) \approx E(\exp(\mathbf{x}^\top \mathbf{X})) = \frac{1}{N}\sum_{i=1}^{n} \exp(\mathbf{x}^\top \mathbf{X}(i)).$$

Similary, its gradient can be estimated using

$$\nabla \widehat{\mathbf{X}}(\mathbf{x}) \approx E(\exp(\mathbf{x}^\top \mathbf{X})\mathbf{X}) = \frac{1}{N}\sum_{i=1}^{n} \exp(\mathbf{x}^\top \mathbf{X}(i))\mathbf{X}(i),$$

and its Hessian by

$$\mathbf{H}_{\widehat{\mathbf{X}}}(\mathbf{x}) \approx E(\exp(\mathbf{x}^\top \mathbf{X})\mathbf{X}\mathbf{X}^\top) = \frac{1}{N}\sum_{i=1}^{n} \exp(\mathbf{x}^\top \mathbf{X}(i))\mathbf{X}(i)\mathbf{X}(i)^\top.$$

The Hessian of $\log \widehat{\mathbf{X}}$ has entries $\left(\widehat{\mathbf{X}} \frac{\partial^2 \widehat{\mathbf{X}}}{\partial x_i \partial x_j} - \frac{\partial \widehat{\mathbf{X}}}{\partial x_i}\frac{\partial \widehat{\mathbf{X}}}{\partial x_j}\right)/\widehat{\mathbf{X}}^2$, so we can calculate it simply by

$$\mathbf{H}_{\log \widehat{\mathbf{X}}}(\mathbf{x}) = \frac{\mathbf{H}_{\widehat{\mathbf{X}}}(\mathbf{x})}{\widehat{\mathbf{X}}(\mathbf{x})} - \frac{\nabla \widehat{\mathbf{X}}(\mathbf{x})\left(\nabla \widehat{\mathbf{X}}(\mathbf{x})\right)^\top}{\widehat{\mathbf{X}}^2(\mathbf{x})}$$

and estimate it using the above sample approximations. For discussion of the noisy case, we refer to [13]; the results there can be easily extended to the MICA setting.

In previous work, [9] for ICA and [12] for MICA, we have proposed to estimate the mixing matrix by diagonalizing the Hessian of the logarithmic densities (*HICA* and *HMICA*) — those behave similar to the characteristic function. However, an albeit local multivariate density estimation is needed, and in general this problem is at least as difficult as MICA. In H(M)ICA we used kernel-based density approximation, which by some algebraic manipulation can enhanced in terms of speed quite considerably, but still seems to be unfeasible in high dimensions. Furthermore, the densities were required to have twice continuously differentiable densities, which is a condition not needed in cfMICA. Hence we expect cfMICA to be of broader applicability than HMICA.

4.3 Matlab implementation

In the experiments we use cfMICA and for comparison the density based Hessian MICA algorithm from [12]. Our software package, available at `http://fabian.theis.name/` implements all the details of the two algorithms. The package contains all the files needed to reproduce the results described in this paper.

5. EXPERIMENTAL RESULTS

In this section we demonstrate the validity of the proposed algorithms by applying them to both toy and real world data.

5.1 Multidimensional Amari-index

In order to analyze algorithm performance, we consider the index $E^{(m)}(\mathbf{C})$ defined for fixed n, k and $\mathbf{C} \in \mathrm{Gl}(mn)$ as $E^{(m)}(\mathbf{C}) = \sum_{r=1}^{n}\left(\sum_{s=1}^{n} \frac{\|\mathbf{C}_{rs}^{(m)}\|}{\max_i \|\mathbf{C}_{ri}^{(m)}\|} - 1\right) + \sum_{s=1}^{n}\left(\sum_{r=1}^{n} \frac{\|\mathbf{C}_{rs}^{(m)}\|}{\max_i \|\mathbf{C}_{is}^{(m)}\|} - 1\right)$. Here $\|.\|$ can be any matrix norm — we choose the operator norm $\|\mathbf{A}\| := \max_{|\mathbf{x}|=1} |\mathbf{A}\mathbf{x}|$. This *multidimensional performance index* of an $mn \times mn$-matrix $\mathbf{C}$ generalizes the one-dimensional performance index introduced by Amari to block-diagonal matrices. It measures how much $\mathbf{C}$ differs from a permutation and scaling matrix in the sense of m-blocks, so it can be used to analyze algorithm performance.

Note that if $\mathbf{C} \in \mathrm{Gl}(mn)$, then $E^{(m)}(\mathbf{C}) = 0$ if and only if $\mathbf{C}$ is the product of a m-scaling and a m-permutation matrix. Furthermore, if we consider the MBSS problem $\mathbf{X} = \mathbf{AS}$ from section 2.

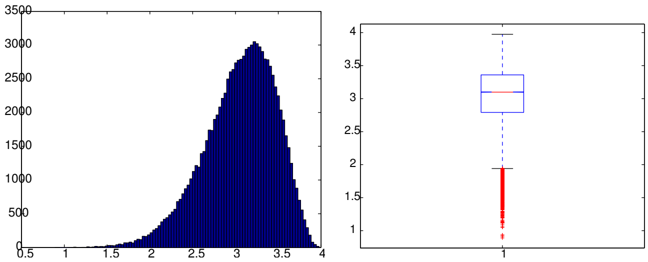

Figure 1: Histogram and box plot of the multidimensional performance index $E^{(m)}(\mathbf{C})$ evaluated for $m = 2$ and $n = 2$. The statistics were calculated over 10^5 independent experiments using 4×4 matrices $\mathbf{C}$ with coefficients uniformly drawn out of $[-1, 1]$.

Then an estimate $\tilde{\mathbf{A}}$ of the mixing matrix solves the MBSS problem if and only if $E^{(m)}(\tilde{\mathbf{A}}^{-1}\mathbf{A}) = 0$.

In order to be able to determine the scale of this index, figure 1 gives statistics of $E^{(m)}$ over randomly chosen matrices in the case $m = n = 2$. The mean is 3.05 and the median 3.10.

5.2 Simulations

We will discuss algorithm performance when applied to a 4-dimensional 2-independent toy signal. We use two independent generating signals, a sinusoid and a sawtooth given by $\mathbf{Z}(t) := (\sin(0.1t), 2\lfloor 0.007t + 0.5 \rfloor - 1)^\top$ for discrete time steps $t = 1, 2, \ldots, 1000$, and generate sources $\mathbf{S}(t) := (Z_1(t), \exp(Z_1(t)), Z_2(t), (Z_2(t) + 0.5)^2)^\top$. Their covariance is

$$\mathrm{Cov}(\mathbf{S}) = \begin{pmatrix} 0.50 & 0.57 & 0.01 & 0.01 \\ 0.57 & 0.68 & 0.01 & 0.01 \\ 0.01 & 0.01 & 0.33 & 0.33 \\ 0.01 & 0.01 & 0.33 & 0.42 \end{pmatrix}$$

so indeed $\mathbf{S}$ is not fully independent — it is only 2-independent by construction.

We perform 100 Monte-Carlo runs using the following parameters: The sources $\mathbf{S}$ are mixed using a 4×4-matrix $\mathbf{A}$ with coefficients drawn uniformly from $[-1, 1]$. The mixtures are separated using cfMICA, the density-based Hessian MICA (HMICA) and the ICA-only algorithms JADE and FastICA (with pow3-nonlinearity). 50 Hessians were used both in cfMICA and HMICA. The estimated mixing matrices are compared using the multidimensional Amari-index $E^{(m)}$ from above. Table 1 shows the results over the 100 iterations. Both cfMICA and HMICA could separate the data very well; JADE was not able to find separating matrices at all, and FastICA (after very slow convergence) found the correct matrix in 17% of all cases. Of course this was to be expected, since the sources are dependent. The two MICA algorithms performed comparably well, although the separation quality of HMICA was somewhat higher — albeit at an additional cost (which is not properly reflected by processing times alone, as Matlab is very fast with matrix computations). Apparently, the data satisfied the condition of $\mathscr{C}^2$-densities, so that cfMICA did not prove to be advantageous to HMICA. In higher dimensions and on more complex data, we expect this to change.

6. CONCLUSION

We have studied multidimensional ICA. We have first provided an elementary proof of uniqueness of the problem, implicitly relying on the fact that a random vector is m-independent if and only if the Hessian of its logarithmic characteristic function is m-block-diagonal everywhere. This generalization of the one-dimensional ICA case [7, 9, 13] can now be used to propose an extended multidimensional ICA algorithm based on joint block diagonalization of the second characteristic function. This direct extension of Yeredor's characteristic function algorithm [13] has the advantage of simple matrix estimation, especially in comparison to previous MICA algorithms. However, there exists no optimal choice of 'processing points' i.e. of points where to evaluate the Hessians. We are currently working on an extension of Eriksson's generalization of Yeredor's ideas [6], but still face some convergence problems, which will have to be resolved in the future.

	mean $E^{(m)}$	% of successful runs	mean time per run (s)
cfMICA	0.10 ± 0.067	100%	0.33 ± 0.043
HMICA	0.036 ± 0.013	100%	0.36 ± 0.044
JADE	2.2 ± 0.43	100%	0.025 ± 0.031
FastICA	0.043 ± 0.025	17%	2.1 ± 0.81

Table 1: Results: 4 two-dependent sources were randomly mixed and separated using the above algorithms. Means are taken over 100 iterations.

Acknowledgements

The author gratefully acknowledges partial financial support by the DFG (GRK 638) and the BMBF (project 'ModKog').

REFERENCES

[1] K. Abed-Meraim and A. Belouchrani. Algorithms for joint block diagonalization. In *Proc. EUSIPCO 2004*, pages 209–212, Vienna, Austria, 2004.

[2] Jean-François Cardoso and Antoine Souloumiac. Jacobi angles for simultaneous diagonalization. *SIAM J. Mat. Anal. Appl.*, 17(1):161–164, January 1995.

[3] J.F. Cardoso. Multidimensional independent component analysis. In *Proc. of ICASSP '98*, Seattle, 1998.

[4] P. Comon. Independent component analysis - a new concept? *Signal Processing*, 36:287–314, 1994.

[5] G. Darmois. Analyse générale des liaisons stochastiques. *Rev. Inst. Internationale Statist.*, 21:2–8, 1953.

[6] J. Eriksson and V. Koivunen. Characteristic-function-based independent component analysis. *Signal Processing*, 83:21952208, 2003.

[7] J.K. Lin. Factorizing multivariate function classes. In *Advances in Neural Information Processing Systems*, volume 10, pages 563–569, 1998.

[8] V.P. Skitovitch. On a property of the normal distribution. *DAN SSSR*, 89:217–219, 1953.

[9] F.J. Theis. A new concept for separability problems in blind source separation. *Neural Computation*, 16:1827–1850, 2004.

[10] F.J. Theis. Uniqueness of complex and multidimensional independent component analysis. *Signal Processing*, 84(5):951–956, 2004.

[11] F.J. Theis. Uniqueness of real and complex linear independent component analysis revisited. In *Proc. EUSIPCO 2004*, pages 1705–1708, Vienna, Austria, 2004.

[12] F.J. Theis. Blind signal separation into groups of dependent signals using joint block diagonalization. In *Proc. ISCAS 2005 accepted*, Kobe, Japan, 2005.

[13] A. Yeredor. Blind source separation via the second characteristic function. *Signal Processing*, 80(5):897–902, 2000.