Blind signal separation into groups of dependent signals using joint block diagonalization

Conference Paper · June 2005

DOI: 10.1109/ISCAS.2005.1465976 · Source: IEEE Xplore

1 author:

Fabian J Theis
Helmholtz Zentrum München
649 PUBLICATIONS **6,140** CITATIONS

SEE PROFILE

Blind signal separation into groups of dependent signals using joint block diagonalization

Fabian J. Theis
Institute of Biophysics, University of Regensburg
93040 Regensburg, Germany, Email: fabian@theis.name

***Abstract*— Multidimensional or group independent component analysis describes the task of transforming a multivariate observed sensor signal such that groups of the transformed signal components are mutually independent - however dependencies within the groups are still allowed. This generalization of independent component analysis (ICA) allows for weakening the sometimes too strict assumption of independence in ICA. It has potential applications in various fields such as ECG, fMRI analysis or convolutive ICA. Recently we could calculate the indeterminacies of group ICA, which finally enables us, also theoretically, to apply group ICA to solve blind source separation (BSS) problems. In this paper we introduce and discuss various algorithms for separating signals into groups of dependent signals. The algorithms are based on joint block diagonalization of sets of matrices generated using several signal structures.**

I. Introduction

In this work, we discuss *multidimensional blind source separation (MBSS)* i.e. the recovery of 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 solution for MBSS is *multidimensional independent component analysis (MICA)* — in section IV we will discuss other such conditions. The idea MICA is that we do not require full independence of the transform $\mathbf{y} := \mathbf{W}\mathbf{x}$ but only mutual independence of certain tuples $y_{i_1}, \dots, y_{i_2}$. 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 for the sake of simplicity we assume that they all have the same length k.

Multidimensional ICA has first been introduced by Cardoso [1] using geometrical motivations. Hyvärinen and Hoyer then presented a special case of multidimensional ICA which they called independent subspace analysis [2]; there the dependence within a k-tuple is explicitly modelled enabling the authors to propose better algorithms without having to resort to the problematic multidimensional density estimation.

II. Joint block diagonalization

Joint diagonalization has become an important tool in ICA-based BSS (used for example in JADE [3]) or in BSS relying on second-order time-decorrelation (for example in SOBI [4]). 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).

Introducing some notation, let us define for $r, s = 1, \dots, n$ the (r, s) *sub-k-matrix of* $\mathbf{W} = (w_{ij})$, denoted by $\mathbf{W}_{rs}^{(k)}$, to be the $k \times k$ submatrix of $\mathbf{W}$ ending at position (rk, sk). Denote $\mathrm{Gl}(n)$ the group of invertible $n \times n$ matrices. A matrix $\mathbf{W} \in \mathrm{Gl}(nk)$ is said to be a *k-scaling matrix* if $\mathbf{W}_{rs}^{(k)} = 0$ for $r \neq s$, and $\mathbf{W}$ is called a *k-permutation matrix* if for each $r = 1, \dots, n$ there exists precisely one s such that $\mathbf{W}_{rs}^{(k)}$ equals the $k \times k$ unit matrix.

Hence, fixing the block-size to k, JBD tries to find $\mathbf{E}$ such that $\mathbf{E}^\top \mathbf{M}_i \mathbf{E}$ is a k-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 [5] and references therein. In the following we will simply perform joint diagonalization (using for example the Jacobi-like algorithm from [6]) and then permute the columns of $\mathbf{E}$ to achieve block-diagonality — in experiments this turns out to be an efficient solution to JBD [5].

III. Multidimensional ICA (MICA)

Let $k, n \in \mathbb{N}$. We call an nk-dimensional random vector $\mathbf{y}$ *k-independent* if the k-dimensional random vectors $(y_1, \dots, y_k)^\top, \dots, (y_{nk-k+1}, \dots, y_{nk})^\top$ are mutually independent. A matrix $\mathbf{W} \in \mathrm{Gl}(nk)$ is called a *k-multidimensional ICA* of an nk-dimensional random vector $\mathbf{x}$ if $\mathbf{W}\mathbf{x}$ is k-independent. If $k = 1$, 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 nk-dimensional random vector $\mathbf{x}$ is given, and $\mathbf{A} \in \mathrm{Gl}(nk)$ and $\mathbf{s}$ are unknown. In the case of MICA $\mathbf{s}$ is assumed to be k-independent.

A. Indeterminacies

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

invariant under these transformations. In [7] we show that these are the only indeterminacies, given some additional weak restrictions to the model, namely that $\mathbf{A}$ has to be k-admissible and that $\mathbf{s}$ is not allowed to contain a Gaussian k-component.

As usual by preprocessing of the observations $\mathbf{x}$ by whitening we may also assume that $\mathrm{Cov}(\mathbf{x}) = \mathbf{I}$. 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.

B. MICA using Hessian diagonalization (MHICA)

We assume that $\mathbf{s}$ admits a $\mathcal{C}^2-$density $p_\mathbf{s}$. Using orthogonality of $\mathbf{A}$ we get $p_\mathbf{s}(\mathbf{s}_0) = p_\mathbf{x}(\mathbf{A}\mathbf{s}_0)$ for $\mathbf{s}_0 \in \mathbb{R}^{nk}$. Let $\mathbf{H}_f(\mathbf{x}_0)$ denote the Hessian of f evaluated at $\mathbf{x}_0$. It transforms like a 2-tensor so locally at $\mathbf{s}_0$ with $p_\mathbf{s}(\mathbf{s}_0) > 0$ we get

$$\mathbf{H}_{\ln p_\mathbf{s}}(\mathbf{s}_0) = \mathbf{H}_{\ln p_\mathbf{x} \circ \mathbf{A}}(\mathbf{s}_0) = \mathbf{A}\mathbf{H}_{\ln p_\mathbf{x}}(\mathbf{A}\mathbf{s}_0)\mathbf{A}^\top \quad (1)$$

The key idea now lies in the fact that $\mathbf{s}$ is assumed to be k-independent, so $p_\mathbf{s}$ factorizes into n groups depending only on k separate variables each. So $\ln p_\mathbf{s}$ is a sum of functions depending on k separate variables hence $\mathbf{H}_{\ln p_\mathbf{s}}(\mathbf{s}_0)$ is block-diagonal i.e. a k-scaling.

The algorithm, *multidimensional Hessian ICA (MHICA)*, now simply uses the block-diagonality structure from equation 1 and performs JBD of estimates of a set of Hessians $\mathbf{H}_{\ln p_\mathbf{s}}(\mathbf{s}_i)$ evaluated at different points $\mathbf{s}_i \in \mathbb{R}^{nk}$. Given slight restrictions on the eigenvalues, the resulting block diagonalizer then equals $\mathbf{A}^\top$ except for k-scaling and permutation. The Hessians are estimated using kernel-density approximation with a sufficiently smooth kernel, but other methods such as approximation using finite differences are possible, too. Density approximation is problematic, but in this setting due to the fact that we can use many Hessians we only need rough estimates. For more details on the kernel approximation we refer to the one-dimensional Hessian ICA algorithm from [8].

MHICA generalizes one-dimensional ideas proposed in [8], [9]. More generally, we could have also used characteristic functions instead of densities, which leads to a related algorithm, see [10] for the single-dimensional ICA case.

IV. Multidimensional time decorrelation

Instead of assuming k-independence of the sources in the MBSS problem, in this section we assume that $\mathbf{s}$ is a multivariate centered discrete WSS random process such that its *symmetrized autocovariances*

$$\bar{\mathbf{R}}_\mathbf{s}(\tau) := \frac{1}{2}\left(\mathbf{E}\left(\mathbf{s}(t+\tau)\mathbf{s}(t)^\top\right) + \mathbf{E}\left(\mathbf{s}(t)\mathbf{s}(t+\tau)^\top\right)\right) \quad (2)$$

are k-scalings for all τ. This models the fact that the sources are supposed to be block-decorrelated in the time-domain for all time-shifts τ.

A. Indeterminacies

Again $\mathbf{A}$ can only be found up to k-scaling and k-permutation because condition 2 is invariant under this transformation. One sufficient condition for identifiability is to have pairwise different eigenvalues of at least one $\mathbf{R}_\mathbf{s}(\tau)$, however generalizations are possible, see [4] for the case $k = 1$. Using whitening, we can again assume orthogonal $\mathbf{A}$.

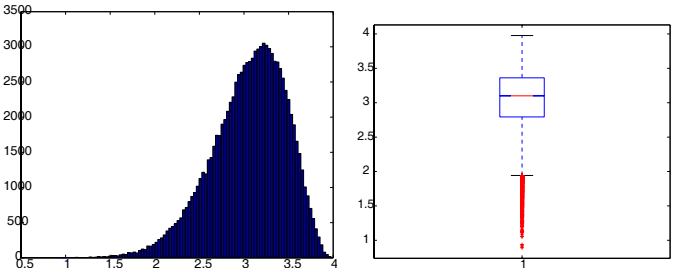


Fig. 1. Histogram and box plot of the multidimensional performance index $E^{(k)}(\mathbf{C})$ evaluated for $k = 2$ and $n = 2$. The statistics were calculated over 10^5 independent experiments using 4×4 matrices $\mathbf{C}$ with coefficients uniformly drawn out of $[-1, 1]$.

B. Multidimensional SOBI (MSOBI)

The idea of what we call *multidimensional second-order blind identification (MSOBI)* is now a direct extension of the usual SOBI algorithm [4]. Symmetrized autocovariances of $\mathbf{x}$ can easily be estimated from the data, and they transform as follows: $\bar{\mathbf{R}}_\mathbf{s}(\tau) = \mathbf{A}^\top \bar{\mathbf{R}}_\mathbf{x}(\tau)\mathbf{A}$. But $\bar{\mathbf{R}}_\mathbf{s}(\tau)$ is a k-scaling by assumption, so JBD of a set of such symmetrized autocovariance matrices yields $\mathbf{A}$ as diagonalizer (except for k-scaling and permutation).

Other researchers have worked on this problem in the setting of convolutive BSS — due to lack of space we want to refer to [11] and references therein.

V. Experimental results

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

A. Multidimensional Amari-index

In order to analyze algorithm performance, we consider the index $E^{(k)}(\mathbf{C})$ defined for fixed n, k and $\mathbf{C} \in \mathrm{Gl}(nk)$ as

$$\begin{aligned} E^{(k)}(\mathbf{C}) \quad = \quad & \sum_{r=1}^{n}\left(\sum_{s=1}^{n}\frac{\|\mathbf{C}_{rs}^{(k)}\|}{\max_i \|\mathbf{C}_{ri}^{(k)}\|} - 1\right) \\ & + \sum_{s=1}^{n}\left(\sum_{r=1}^{n}\frac{\|\mathbf{C}_{rs}^{(k)}\|}{\max_i \|\mathbf{C}_{is}^{(k)}\|} - 1\right). \end{aligned}$$

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 $nk \times nk$-matrix $\mathbf{C}$ generalizes the one-dimensional performance index introduced by Amari et al. [12] to block-diagonal matrices. It measures how much $\mathbf{C}$ differs from a permutation and scaling matrix in the sense of k-blocks, so it can be used to analyze algorithm performance:

Lemma 5.1: Let $\mathbf{C} \in \mathrm{Gl}(nk)$. $E^{(k)}(\mathbf{C}) = 0$ if and only if $\mathbf{C}$ is the product of a k-scaling and a k-permutation matrix.

Corollary 5.2: Consider the MBSS problem $\mathbf{x} = \mathbf{A}\mathbf{s}$ from section III respectively IV. An estimate $\hat{\mathbf{A}}$ of the mixing matrix solves the MBSS problem if and only if $E^{(k)}(\hat{\mathbf{A}}^{-1}\mathbf{A}) = 0$.

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

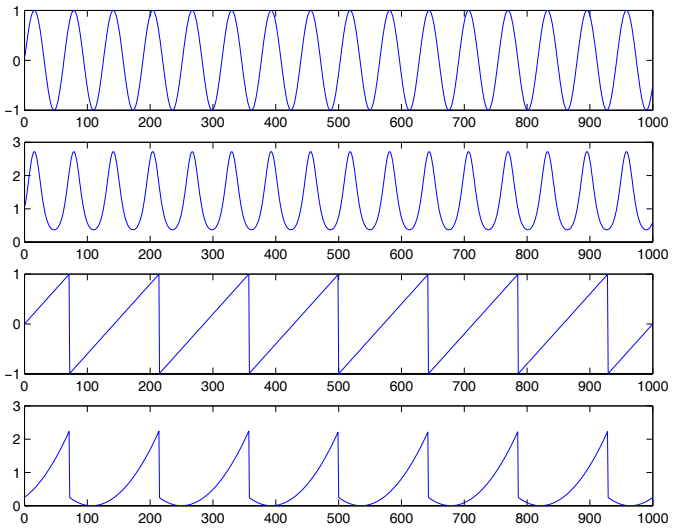


Fig. 2. Simulation, 4-dimensional 2-independent sources. Clearly the first and the second respectively the third and the fourth signal are dependent.

B. Simulations

We will discuss algorithm performance when applied to a 4-dimensional 2-independent toy signal. In order to see the performance of both MSOBI and MHICA we generate 2-independent sources with non-trivial autocorrelations. For this we use two independent generating signals, a sinusoid and a sawtooth given by

$$\mathbf{z}(t) := (\sin(0.1\,t), 2\lfloor 0.007\,t + 0.5 \rfloor - 1)^\top$$

for discrete time steps $t = 1, 2, \ldots, 1000$. We thus generated sources

$$\mathbf{s}(t) := (z_1(t), \exp(z_1(t)), z_2(t), (z_2(t) + 0.5)^2)^\top,$$

which are plotted in figure 2. 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.

$\mathbf{s}$ is mixed using a 4×4 matrix $\mathbf{A}$ with entries uniformly drawn out of $[-1, 1]$, and comparisons are made over 100 Monte-Carlo runs. We compare the two algorithms MSOBI (with 10 autocorrelation matrices) and MHICA (using 50 Hessians) with the ICA algorithms JADE and fastICA, where in the latter both the deflation and the symmetric approach was used. For each run we calculate the performance index $E^{(2)}(\hat{\mathbf{A}}^{-1}\mathbf{A})$ of the product of the mixing and the estimated separating matrix. Since the one-dimensional ICA algorithms are unable to use the group structure, for these we take the minimum of the index calculated over all row permutations of $\hat{\mathbf{A}}^{-1}\mathbf{A}$.

Figure 3 displays the result of the comparison. Clearly MHICA and MSOBI perform very well on this data, and MSOBI furthermore gives very robust estimates with the same error and negligibly small variance. JADE cannot separate

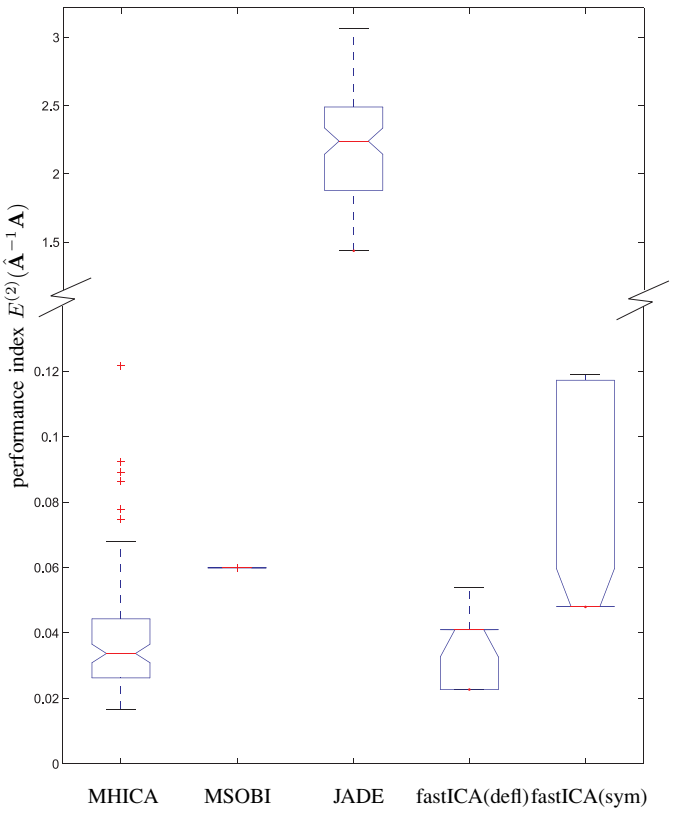
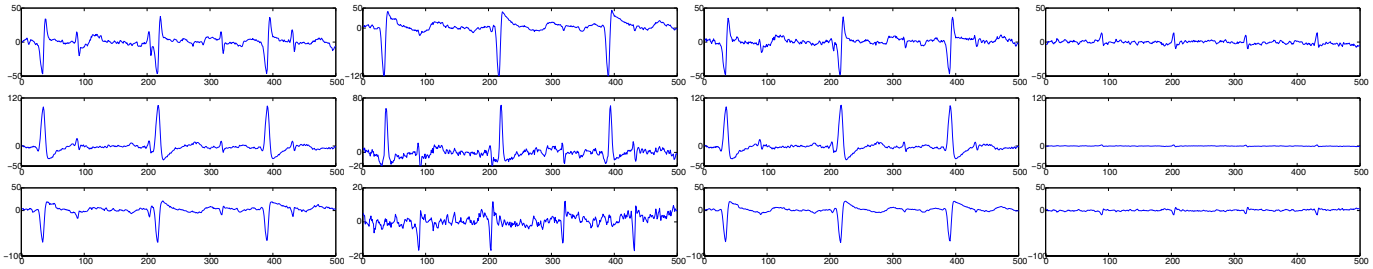


Fig. 3. Simulation, algorithm results. This notched boxed plot displays the performance index $E^{(2)}$ of the mixing-separating matrix $\hat{\mathbf{A}}^{-1}\mathbf{A}$ of each algorithm, sampled over 100 Monte-Carlo runs. The middle line of each column gives the mean, the boxes the 25th and 75th percentile. The deflationary fastICA algorithm only converged in 12% of all runs, the symmetric-approach based fastICA in 89% of all cases; the statistics are only given over successful runs. All other algorithms converged in all runs.

the data at all — it performs not much better than random choice of matrix, see figure 1; this is due to the fact that the cumulants of k-independent sources are not block-diagonal. FastICA only converges in 12% (deflation approach) respectively 89% (symmetric approach) of all cases. However, in the cases it converges it gives results comparable with the multidimensional algorithms. Apparently, especially the symmetric method seems to be able to use the weakened statistics to still find directions in the data.

C. Application to ECG data

Finally we illustrate how to apply the proposed algorithms to real-world data set. Following [1], we will show how to separate fetal ECG (FECG) recordings from the mother's ECG (MECG). The data set [13] consists of eight recorded signals with 2500 observations; the sampling frequency is misleadingly specified as 500 Hz (which would mean around 168 mother heartbeats per minute), it should be closer to around 250 Hz. We select the first three sensors cutaneously recorded on the abdomen of the mother. In order to save space and to compare the results with [1] we plot only the first 1000 samples, see figure 4(a).



(a) ECG recordings (b) extracted sources (c) MECG part (d) FECG part

Fig. 4. Fetal ECG example. (a) shows the ECG recordings. The underlying FECG (4 heartbeats) is partially visible in the dominating MECG (3 heartbeats). Figure (b) gives the extracted sources using MHICA with $k = 2$ and 500 Hessians. In (c) and (d) the projections of the mother sources (components 1 and 2 in (b)) respectively the fetal sources (component 3) onto the mixture space (a) are plotted.

Our goal is to extract an MECG and an FECG component; however it cannot be expected to find an only one-dimensional MECG due to the fact that projections of a three-dimensional vector (electric) field are measured. Hence modelling the data by a multidimensional BSS problem with $k = 2$ (but allowing for an additional one-dimensional component) makes sense. Application of MHICA (with 500 Hessians) and MSOBI (with 50 autocorrelation matrices) extracts a two-dimensional MECG component and a one-dimensional FECG component. After block-permutation we get the following estimated mixing matrices ($\mathbf{A}$ using MHICA and $\mathbf{A}'$ using MSOBI)

$$\mathbf{A} = \begin{pmatrix} 0.37 & 0.42 & -0.81 \\ -0.75 & 0.89 & -0.16 \\ 0.55 & -0.16 & 0.57 \end{pmatrix} \qquad \mathbf{A}' = \begin{pmatrix} 0.22 & 0.91 & -0.40 \\ -0.84 & 0.23 & -0.33 \\ 0.50 & 0.34 & 0.85 \end{pmatrix}.$$

The thus estimated sources using MHICA are plotted in figure 4(b). In order to compare the two mixing matrices, calculation of

$$\mathbf{A}^{-1}\mathbf{A}' = \begin{pmatrix} 0.85 & 1.02 & 0.64 \\ -0.23 & 1.11 & 0.35 \\ -0.01 & -0.08 & 0.98 \end{pmatrix}.$$

yields a somewhat visible block structure; the performance index is $E^{(2)}(\mathbf{A}^{-1}\mathbf{A}') = 1.12$. The block structure is not very dominant, which indicates that the two models — block independence versus time-block-decorrelation — are not fully equivalent.

A (scaling invariant) decomposition of the observed ECG data can be achieved by composing the extracted sources using only the relevant mixing columns. For example for the MECG part this means applying the projection $\Pi_M := (\mathbf{a}_1, \mathbf{a}_2, 0)\mathbf{A}^{-1}$ to the observations. This yields the projection matrices

$$\Pi_M = \begin{pmatrix} 0.52 & 0.38 & 0.84 \\ -0.10 & 1.08 & 0.17 \\ 0.34 & -0.27 & 0.41 \end{pmatrix} \qquad \Pi_F = \begin{pmatrix} 0.48 & -0.38 & -0.84 \\ 0.10 & -0.08 & -0.17 \\ -0.34 & 0.27 & 0.59 \end{pmatrix}$$

onto the mother respectively the fetal ECG using MHICA and

$$\Pi'_M = \begin{pmatrix} 0.78 & 0.21 & 0.45 \\ -0.18 & 1.17 & 0.36 \\ 0.47 & -0.44 & 0.05 \end{pmatrix} \qquad \Pi'_F = \begin{pmatrix} 0.22 & -0.21 & -0.45 \\ 0.18 & -0.17 & -0.36 \\ -0.47 & 0.44 & 0.95 \end{pmatrix}.$$

using MSOBI. The results of the first algorithm are plotted in figures 4 (c) and (d). The fetal ECG is most active at sensor 1 (as visual inspection of the observation confirms). When comparing the projection matrices with the results from [1], we get quite high similarity of the ICA-based results, and a modest difference with the projections of the time-based algorithm. Other one-dimensional ICA-based results on this data set are reported for example in [14].

VI. Conclusion

We have shown how the idea of joint block diagonalization as extension of joint diagonalization helps us to generalize ICA and time-structure based algorithms such as HICA and SOBI to the multidimensional ICA case. The thus defined algorithms are able to robustly decompose signals into groups of independent signals. In future work, besides more extensive experiments and tests with noise and outliers, we want to extend this result to a version of JADE using moments instead of cumulants, which preserve the block structure.

References

[1] J. Cardoso, "Multidimensional independent component analysis," in *Proc. of ICASSP '98*, Seattle, 1998.

[2] A. Hyvärinen and P. Hoyer, "Emergence of phase and shift invariant features by decomposition of natural images into independent feature subspaces," *Neural Computation*, vol. 12, no. 7, pp. 1705–1720, 2000.

[3] J.-F. Cardoso and A. Souloumiac, "Blind beamforming for non gaussian signals," *IEE Proceedings - F*, vol. 140, no. 6, pp. 362–370, 1993.

[4] A. Belouchrani, K. A. Meraim, J.-F. Cardoso, and E. Moulines, "A blind source separation technique based on second order statistics," *IEEE Transactions on Signal Processing*, vol. 45, no. 2, pp. 434–444, 1997.

[5] K. Abed-Meraim and A. Belouchrani, "Algorithms for joint block diagonalization," in *Proc. EUSIPCO 2004*, Vienna, Austria, 2004, pp. 209–212.

[6] J.-F. Cardoso and A. Souloumiac, "Jacobi angles for simultaneous diagonalization," *SIAM J. Mat. Anal. Appl.*, vol. 17, no. 1, pp. 161–164, Jan. 1995.

[7] F. Theis, "Uniqueness of complex and multidimensional independent component analysis," *Signal Processing*, vol. 84, no. 5, pp. 951–956, 2004.

[8] ——, "A new concept for separability problems in blind source separation," *Neural Computation*, vol. 16, pp. 1827–1850, 2004.

[9] J. Lin, "Factorizing multivariate function classes," in *Advances in Neural Information Processing Systems*, vol. 10, 1998, pp. 563–569.

[10] A. Yeredor, "Blind source separation via the second characteristic function," *Signal Processing*, vol. 80, no. 5, pp. 897–902, 2000.

[11] C. Févotte and C. Doncarli, "A unified presentation of blind separation methods for convolutive mixtures using block-diagonalization," in *Proc. ICA 2003*, Nara, Japan, 2003, pp. 349–354.

[12] S. Amari, A. Cichocki, and H. Yang, "A new learning algorithm for blind signal separation," *Advances in Neural Information Processing Systems*, vol. 8, pp. 757–763, 1996.

[13] B. D. M. (ed.), "DaISy: database for the identification of systems," *Department of Electrical Engineering, ESAT/SISTA, K.U.Leuven, Belgium*, Oct 2004. [Online]. Available: http://www.esat.kuleuven.ac.be/sista/daisy/

[14] L. D. Lathauwer, B. D. Moor, and J. Vandewalle, "Fetal electrocardiogram extraction by source subspace separation," in *Proc. IEEE SP / ATHOS Workshop on HOS*, Girona, Spain, 1995, pp. 134–138.