A comparison of algorithms for separation of synchronous subspaces

M. ALMEIDA[1,2*], J. BIOUCAS-DIAS[1], R. VIGÁRIO[2], and E. OJA[2]

[1] Instituto de Telecomunicações, Instituto Superior Técnico, Technical University of Lisbon, Av. Rovisco Pais, 1049-001 Lisbon, Portugal
[2] Department of Information and Computer Science, Aalto University, F1–00076 Aalto, Finland

Abstract. Independent Subspace Analysis (ISA) consists in separating sets (subspaces) of dependent sources, with different sets being independent of each other. While a few algorithms have been proposed to solve this problem, they are all completely general in the sense that they do not make any assumptions on the intra-subspace dependency. In this paper, we address the ISA problem in the specific context of Separation of Synchronous Sources (SSS), i.e., we aim to solve the ISA problem when the intra-subspace dependency is known to be perfect phase synchrony between all sources in that subspace. We compare multiple algorithmic solutions for this problem, by analyzing their performance on an MEG-like dataset.

Key words: phase-locking, synchrony, source separation, subspaces, Independent Component Analysis (ICA), Independent Subspace Analysis (ISA), magnetoencephalogram (MEG).

1. Introduction

In the human brain, it has been shown that synchrony [1] is associated with important phenomena. For example, there is evidence that autism, Parkinson's, and Alzheimer's disease are associated with a decrease in the synchrony of some brain regions, whereas some types of epilepsy are associated with an anomalous increase in synchrony [2].

It is also known that electrophysiological brain signals, such as electroencephalogram (EEG) or magnetoencephalogram (MEG) signals, are not a direct measurement of the activity of individual brain regions, but rather the result of a superposition of those regions' activities [3]. Due to the low frequencies involved (below 1 kHz), the approximation of a linear and instantaneous mixing process is valid [4].

Since synchrony is an important phenomenon in the human brain, and measurements are the result of a mixing process, we argue that it is important to perform Blind Source Separation (BSS) on these types of measurements, to be able to extract relevant networks of synchronous activity. If, as a first approximation, one considers those different networks to be independent, then finding the original networks can be cast as an Independent Subspace Analysis (ISA) task, where each one of the mutually independent subspaces has a specific type of internal dependency, namely, phase synchrony.

The goal of this paper is to compare the performance of various algorithms on a partial ISA task (defined in Subsec. 2.3) when the subspaces are composed of synchronous sources. Some of these algorithms are dedicated ISA algorithms, while others are Independent Component Analysis (ICA) algorithms which we nevertheless apply to the partial ISA task. These algorithms are compared on data which is created to mimic brain MEG recordings.

This paper is organized as follows. In Sec. 2 we provide an overview of the BSS problem, with specific focus on the ICA and ISA problems; the models of these problems are detailed there. In Sec. 3 we provide an overview of some of the algorithms available to perform ISA. Section 4 compares the performance of these algorithms on a set of MEG-like. These results are discussed in Sec. 5, and conclusions are drawn in Sec. 6.

2. Methodology

2.1. Blind Source Separation (BSS). One important subclass of signal processing problems is the topic of blind source separation (BSS). In BSS, one has access to a set of signals called measurements $\mathbf{y}(t)$, which are the result of a superposition of another set of signals $\mathbf{s}(t)$, which are called the sources, and which are not directly observable. The goal of BSS is to recover the sources using only the measurements.

Let $\mathbf{s}(t)$ denote a vector of N sources, $\mathbf{s}(t) \equiv (s_1(t), \ldots, s_N(t))^{\mathsf{T}}$. Suppose that these sources undergo a linear and instantaneous mixing process, described by a P-by-N mixing matrix $\mathbf{M}$, resulting in P mixed signals contained in the vector $\mathbf{y}(t) = \mathbf{M}\mathbf{s}(t)$, $\mathbf{y}(t) \equiv (y_1(t), \ldots, y_P(t))^{\mathsf{T}}$. The goal of (linear and instantaneous) BSS is therefore to recover the original sources $\mathbf{s}$ using only the observed mixtures $\mathbf{y}$. In this paper, we will limit ourselves to the case where $P = N$, *i.e.*, the case where the number of sources is equal to the number of observed signals[1].

In general, the BSS problem is ill-posed, in the sense that it has an infinity of solutions. To solve this problem, one must

---

*e-mail: malmeida@lx.it.pt

[1]The case $P > N$ is called the overdetermined case and can be solved using techniques similar to the ones described here by reducing the problem to an equivalent problem with a square mixing matrix [5]. The case $P < N$ is called the underdetermined case and is, in general, harder to solve [5].

assume extra conditions on the sources **s**, the mixing matrix **M**, or both. One possibility is to assume that, for each time instant t, the vector $\mathbf{s}(t)$ is an i.i.d. realization of a random vector whose components are statistically independent. This is the fundamental assumption of Independent Component Analysis (ICA) [5, 6]. Other possibilities include assuming that the mixing matrix and the sources are non-negative, which is known as Non-negative Matrix Factorization (NMF) [7, 8], or assuming that the sources have perfect phase synchrony, which leads to Separation of Synchronous Sources [9].

2.2. Independent Component Analysis (ICA). In ICA, the goal is to find a demixing matrix **W** such that the estimated sources, given by $\mathbf{x} \equiv \mathbf{W}\mathbf{y} = \mathbf{W}\mathbf{M}\mathbf{s}$, are equal to the original sources, which are assumed to be mutually independent. If no other information besides independency is used, some indeterminacies are unavoidable: the estimated sources are equal to the original sources only up to permutation and arbitrary scaling. Equivalently, the gain matrix, defined as $\mathbf{G} \equiv \mathbf{W}\mathbf{M}$, must be a permutation of the rows of a diagonal matrix.

In an informal way, one can say that the ICA problem is identifiable in the following sense: under mild assumptions, if $\mathbf{y} = \mathbf{M}\mathbf{s}$ and one finds a matrix **W** such that all components of $\mathbf{x} = \mathbf{W}\mathbf{y}$ are independent, then $\mathbf{x} = \mathbf{s}$ up to permutation and arbitrary scaling [5]. Unfortunately, independence is not easy to measure from a finite sample; therefore, most ICA algorithms replace that criterion with other contrast functions, such as entropy, kurtosis or lagged correlations. A good overview of ICA algorithms and theory can be found in [5, 6].

2.3. Independent Subspace Analysis (ISA). Independent Subspace Analysis (ISA) is a generalization of ICA, where one has independent random *vectors* instead of independent random (scalar) *variables*. Thus, when all the random vectors are uni-dimensional, ISA reduces to ICA.

Since ISA follows a different model than ICA, we likewise introduce new notation: let

$$\mathbf{s} \equiv \begin{bmatrix} \mathbf{s}^1 \\ \mathbf{s}^2 \\ \vdots \\ \mathbf{s}^K \end{bmatrix}, \quad \text{where} \quad \mathbf{s}^k \equiv \begin{bmatrix} s_1^k \\ \vdots \\ s_{N_k}^k \end{bmatrix}.$$

Note that **s** is a random vector composed of K random subvectors $\mathbf{s}^k$, $k = 1, \ldots, K$. The size of subvector $\mathbf{s}^k$ is N_k. The crucial assumption in ISA is that any two random subvectors are independent, i.e., $\mathbf{s}^k$ and $\mathbf{s}^l$ are independent for any $k \neq l$. Naturally, one must have $N_1 + \ldots + N_k = N$. Because realizations of the random subvector $\mathbf{s}^k$ span an N_k dimensional subspace, in the following for brevity $\mathbf{s}^k$ is called the k-th subspace of **s**.

A naïve approach to perform ISA would be to minimize the mutual information between the various subspaces:

$$\min I(\widehat{\mathbf{s}}^1, \ldots, \widehat{\mathbf{s}}^K)$$

where $\widehat{\mathbf{s}}^k$ is the estimate of the k-th subspace. This is a valid approach which has seen some use; however, in general, this approach is a combinatorial optimization problem [10], since one does not know which of the estimated sources should be grouped together [11, 12]. One must then either test all possible groupings, which grow very quickly with K and rapidly become intractable, or solve a discrete optimization problem by following, e.g., a greedy approach. Nevertheless, this approach is very prone to local minima.

Furthermore, the aforementioned approach involves the computation of the entropy of random vectors of dimension N_k, for $k = 1, \ldots, K$. This computation is non-trivial for $N_k \geq 2$ [13], further increasing the computational complexity of this approach. Nevertheless, this approach has been tackled, e.g., by estimating the entropy of multi-dimensional components using minimum spanning trees [12], or using variational Bayes approaches [11].

In this paper we divide ISA into three successive parts, and consider only the first one, similarly to what other groups have done [14–16]. Thus, the goal is to solve an easier problem than the full ISA problem; we call this easier problem "partial ISA". This partial problem is not combinatorial.

2.4. Partial ISA and full ISA. The ISA procedure can be split into three parts. The first part could be called inter-subspace separation, and is the primary focus of this study. The goal of this first part is to obtain a demixing matrix $\mathbf{W}_{inter}$, such that the gain matrix, $\mathbf{W}_{inter}\mathbf{M}$, is a permutation of a block diagonal matrix with blocks corresponding to the subspaces.

For example, suppose that there are three subspaces ($K = 3$), the first of which has three components ($N_1 = 3$) while the second and third subspaces have two components ($N_2 = N_3 = 2$). In this case, the goal is to find a matrix $\mathbf{W}_{inter}$ of the form $\mathbf{W}_{inter} \equiv \mathbf{P}\mathbf{B}_{inter}$ where **P** is a permutation matrix and $\mathbf{B}_{inter}$ is such that

$$\mathbf{B}_{inter}\mathbf{M} = \begin{bmatrix} \mathbf{U}^1 & \mathbf{0}_{3\times 2} & \mathbf{0}_{3\times 2} \\ \mathbf{0}_{2\times 3} & \mathbf{U}^2 & \mathbf{0}_{2\times 2} \\ \mathbf{0}_{2\times 3} & \mathbf{0}_{2\times 2} & \mathbf{U}^3 \end{bmatrix}.$$

Here, $\mathbf{0}_{m\times n}$ is the m-by-n zero matrix, $\mathbf{U}^1$ is a 3-by-3 invertible matrix, and $\mathbf{U}^2$ and $\mathbf{U}^3$ are 2-by-2 invertible matrices. In this case, each entry of the random vector $\mathbf{x}_{inter} \equiv \mathbf{W}_{inter}\mathbf{y}$ is a linear combination of sources from one subspace only. The second part (which is not studied in this paper) is called permutation compensation or subspace detection. The goal is to group the entries of the random vector $\mathbf{x}_{inter}$ so that the first N_1 entries of $\mathbf{x}_{inter}$ are linear combinations of sources from the first subspace, the next N_2 entries are linear combinations of sources from the second subspace, and so on. This can be achieved by multiplying $\mathbf{x}_{inter}$ by a suitable permutation matrix, **Q**. In general, this step involves the use of a measure of dependency or independency, and therefore will depend on the specific type of dependencies between sources in each subspace.

After the subspace detection is completed, one can define

$$\mathbf{y}^1 \equiv \mathbf{Q}_{(1:N_1 \, , \, :)}\mathbf{x}_{inter} \quad (1)$$

$$\mathbf{y}^2 \equiv \mathbf{Q}_{(N_1+1:N_1+N_2 \, , \, :)}\mathbf{x}_{inter} \quad (2)$$

$$\vdots \quad (3)$$

$$\mathbf{y}^k \equiv \mathbf{Q}_{(N-N_k+1:N \, , \, :)}\mathbf{x}_{inter}, \quad (4)$$

where $\mathbf{Q}_{(a:b \, , \, :)}$ is a matrix composed of the rows a to b of matrix $\mathbf{Q}$.

The third and last part (which is also not studied here) is called intra-subspace separation, and involves finding square matrices $\mathbf{W}^k$ (of size N_k by N_k) such that $\mathbf{s}^k = \mathbf{W}^k\mathbf{y}^k$. There are K such matrices to be found, and each can be estimated separately once the inter-subspace separation and the subspace detection have been performed.

Unfortunately, the available literature is not quite clear on what the term "Independent Subspace Analysis" means. Some authors (cf., [17, 18]) define "ISA" as the task of performing all three steps, while others (cf., [15, 16]) define the same term as solving the first or the first two steps only. To prevent confusion, throughout this paper we will define "Full ISA" as the task of performing all three steps, and "Partial ISA" as the task of performing the first step alone.

ISA (full and partial) has seen increasing interest from the scientific community in recent years. It is known by different names, including "Independent Subspace Analysis" [12, 14, 15, 19], "Subspace Independent Component Analysis" [16], among others. Relevant theoretical results have been published about this topic, such as sufficient conditions on the distribution of the sources for full ISA to be achievable through maximization of kurtosis [18] or minimization of mutual information [17]. Under these conditions, then, simple ICA algorithms which maximize kurtosis (such as some variants of FastICA) or minimize mutual information (such as Infomax) can be safely used to perform the full ISA task, even though the assumption of independence of the sources is violated. On the other hand, dedicated algorithms for partial ISA have also been proposed; see, e.g., [15] and [16]. Techniques for subspace detection have also been presented recently [14].

2.5. Phase synchrony. Motivated by studies of the human brain, where synchrony plays an important role, we assume here that the ISA subspaces are composed of synchronous sources. The external measurements such as EEG or MEG are the result of superpositions of brain regions' activities, and due to the low frequencies involved, the mixing process can be assumed to be linear and instantaneous.

We can thus safely assume that the sources follow the ISA model and, in addition, they are assumed to have perfect phase synchrony within each subspace, as measured by the Phase Locking Factor (PLF) [9]; in other words, within each subspace all pairwise PLFs are 1. Such signals follow

a specific model: it can be shown that the sources of each subspace k are of the form [25]

$$s_i^k(t) \equiv a_i^k(t)\cos(\alpha_i^k(t)), \quad \text{where} \quad \alpha_i^k(t) \equiv \phi_i^k + \psi^k(t), \quad (5)$$

where $a_i^k(t)$ are non-negative amplitudes, and $\alpha_i^k(t)$ is called the phase of source s_i^k. Informally, in each subspace the phase difference between any two sources is constant, *i.e.*, $\alpha_i^k(t) - \alpha_j^k(t) = \phi_i^k - \phi_j^k = \text{const.}$ [9, 25]. For this reason, $\psi^k(t)$ is sometimes called the common oscillation of subspace k (although it does not have to be an oscillatory signal), and ϕ_i^k is called the phase lag of source s_i^k.

3. Algorithms

3.1. ICA algorithms. We now give a brief overview of the specific ICA algorithms used in this study. Note that ICA algorithms have not been designed to be used for ISA tasks, therefore it is not necessarily expected that they perform ideally. The Infomax algorithm [5] can be derived as the maximum likelihood estimator (MLE) of $\mathbf{W}$, if the observations $\mathbf{y}$ follow the ICA model $\mathbf{y} = \mathbf{Ms}$, with $\mathbf{s}$ having independent components. The optimization of the likelihood function is usually done with gradient steps. In our experiments, we use the implementation of Infomax in the MISEP package[2] [20] which not only estimates the demixing matrix $\mathbf{W}$ but is also adaptive in the probability density functions of the sources, thus avoiding the usual parametrization between sub- and super-gaussian densities [5].

The FastICA algorithm [5] can also be derived as an MLE, just like Infomax; however, the optimization is done using fixed-point iterations instead of gradient steps. FastICA is actually a family of algorithms: one can separate the sources using as contrast function the kurtosis of the estimated sources or their entropy, among other possibilities. We use the authors' implementation of FastICA[3]; the symmetric version with the cubic nonlinearity is used.

The Second Order Blind Identification (SOBI) algorithm and the Temporal Decorrelation SEParation (TDSEP) algorithms [21] are somewhat different from the previous two. They are based on the following principle: if the estimated sources $\mathbf{x}$ are independent from time-lagged versions of each other, then the correlation matrix $\mathbf{C}(\tau) \equiv \mathbb{E}[\mathbf{x}(t)\mathbf{x}(t-\tau)^\mathsf{T}]$, where $\mathbb{E}$ is the expectation value operator, should be diagonal for any value of the time lag τ. SOBI and TDSEP choose a set of M lags $\tau_1, \dots, \tau_M$ and find $\mathbf{W}$ such that the correlation matrices $\mathbf{C}(\tau_1), \dots, \mathbf{C}(\tau_M)$ are as diagonal as possible. If the data does not follow the ICA model exactly due to, e.g., noise, it is in general impossible to exactly diagonalize these matrices, and algorithms for approximate joint diagonalization must be employed.[4] For more details of these algorithms, see [5, 6].

3.2. ISA algorithms. We now briefly discuss the ISA algorithms used in this study. FastISA [15] assumes that the di-

---

[2]Available from http://www.lx.it.pt/lbalmeida/ica/mitoolbox.html

[3]Available from http://research.ics.tkk.fi/ica/fastica/

[4]We use the TDSEP implementation in [21].

mensions of the subspaces are known. It is an algorithm which searches for a matrix $\mathbf{W}$ such that the norms of the projections of $\mathbf{x}$ onto each of the K subspaces are independent. This can be viewed as FastICA applied to a scalar function (the norm) of the vector variables $\mathbf{x}^k$. The goal of FastISA is to perform Partial ISA only.

In [16], it was argued that the kurtosis of a scalar random variable can be generalized to random vectors. The authors prove that, if the dimensions of all subspaces are equal and known (*i.e.*, if $N_1 = \ldots = N_K$ and K is known), then one of the stationary points of $\text{vkurt}(\mathbf{x}^1) + \ldots + \text{vkurt}(\mathbf{x}^K)$, where vkurt is the vector kurtosis function in [16], corresponds to a solution of Partial ISA.

The third algorithm we use is sJADE, a generalization of JADE [5] based on joint block diagonalization [14]. Unlike the previous two algorithms, sJADE does not assume that the subspace sizes are known.

While these three algorithms only perform Partial ISA, algorithms for subsequent subspace detection have been proposed recently as well [14].

4. Experimental results

4.1. MEG-like data. The choice of the data to use in this study is non-trivial. On one hand, knowledge of the true sources in real EEG and MEG requires simultaneous data from outside the scalp (EEG or MEG, which correspond to the mixed signals $\mathbf{y}$) and from inside the scalp (intra-cranial recordings, which are more directly related to the sources $\mathbf{s}$). Aquisition of intra-cranial recordings is, obviously, a very invasive procedure which is seldom available. When they are not available, results can only be assessed qualitatively by human experts, who can tell whether the extracted sources are meaningful or not. On the other hand, synthetic data is too simple to assess the usefulness of these algorithms in real data.

In order to retain the advantages of using simulated and real data, we use MEG-like data, generated from actual MEG recordings. These data were originally used in [22]; they consist of data from 122 MEG channels, each with 17730 time samples. The sampling frequency is 297 Hz, and the data had already been subjected to a low-pass filtering with cutoff at 90 Hz. We begin by selecting 4 of the 122 channels randomly. Those 4 channels are bandpass filtered in the band [23, 24] Hz. The resulting filtered data is then downsampled with a factor of $\lfloor \frac{0.25}{B} \rfloor$, where $B = 14/297$ is the passband width in relative frequency units. This downsampling is performed to make the size of the data manageable by some of the more computationally intensive algorithms. We then compute the Hilbert Transform [23] of those 4 channels and use it to obtain the corresponding analytic signals, $\hat{u}_i(t)$, $i = 1, \ldots, 4$. Note that $\hat{u}_i(t)$ are complex-valued signals.

Since the 4 MEG channels are not independent, we split the observation period into two parts, one with $t = 1, \ldots, \frac{17730}{2}$, and one with $t = \frac{17730}{2} + 1, \ldots, 17730$. The first half is used to generate data for the first subspace, whereas the second half is used for the second subspace.

The common oscillation of the first subspace is taken as $\psi^1(t) \equiv e^{\text{i angle}(\hat{u}_1(t))}$. The phase lags ϕ_1^1 and ϕ_2^1 are drawn from a Uniform$(0,2\pi)$ distribution. The amplitudes are simply taken as the amplitudes of the first and second analytic signals: $a_{1,2}^1(t) \equiv |\hat{u}_{1,2}(t)|$. Recall that, for the first subspace, only the first half of the observation period is used.

The second subspace is generated in a similar manner, with the following differences: the common oscillation is now generated from the third MEG channel, $\psi^2(t) \equiv e^{\text{i angle}(\hat{u}_3(t))}$; the amplitudes are now taken as the amplitudes of the third and fourth analytic signals, $a_{1,2}^2(t) \equiv |\hat{u}_{3,4}(t)|$; and, as mentioned earlier, the second half of the observation period is used instead of the first half.

Furthermore, to ensure that the mixing process is physiologically plausible, we use the EEGIFT [24] package with its example dataset and default options to estimate 20 independent components on a dataset of 64 EEG recordings. We select a random 4-by-4 submatrix from the 20-by-64 mixing matrix estimated by EEGIFT and use it to mix the MEG-like sources described in the previous paragraphs.

The above process generates a set of 4 sources, in which sources 1 and 2 form a subspace and sources 3 and 4 form another subspace. An example of such sources is shown in Fig. 1. We call this a "2+2" structure. We used an analogous procedure to generate two other datasets: one in a "3+3" structure, *i.e.*, a set of 6 sources where sources 1, 2, and 3 form one subspace, as do sources 4, 5, and 6; and one in a "2+2+2" structure, *i.e.*, a set of 6 sources where sources 1 and 2 form one subspace, sources 3 and 4 form another subspace, and sources 5 and 6 form yet another subspace.
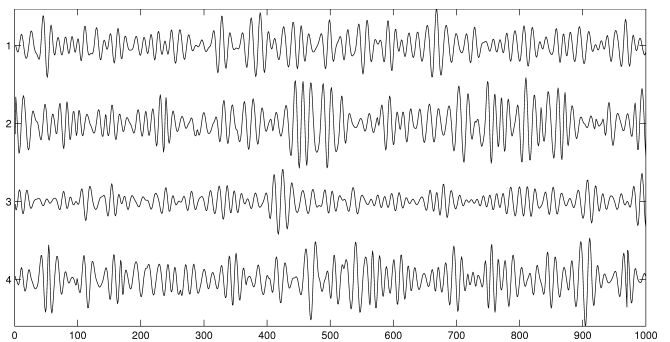

Fig. 1. Excerpt of one dataset with 2+2 structure. The figure shows the first 1000 time points of the four sources

By choosing randomly which MEG channels are used to generate the sources and which sub-matrix from the 20-by-64 mixing matrix estimated by EEGIFT is used to mix the sources, we can generate different sets of data. We generate 500 datasets of each of the structures mentioned in the previous paragraph and run the six algorithms mentioned above on them.

4.2. Results. The results of these experiments are summarized in Figs. 2, 3, and 4. We use the Subspace Amari Performance Index (SAPI; see, e.g., [18]) to measure the quality of the inter-subspace separation. This quantity is zero if and

only if the gain matrix $\mathbf{WM}$ is a permutation of a block diagonal matrix, with block sizes corresponding to the subspace dimensions. Otherwise, it is positive, with greater values of the SAPI corresponding to worse separations.
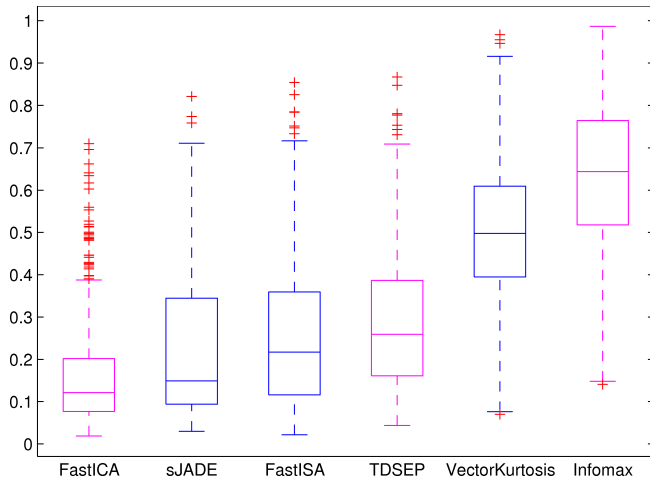

Fig. 2. [2+2] Values of the subspace Amari Performance Index (SAPI) for the six algorithms applied to the 500 datasets with 2+2 structure. The box plots represent the following quantiles: 0 (bottom whisker), 0.25 (bottom edge of the box), 0.5 (solid line within the box), 0.75 (top edge of the box) and 1 (top whisker). If some values of the SAPI would make the whiskers longer than 1.5 times the height of the corresponding box, they are considered outliers, are represented as individual "+" symbols. They are not considered in the box plots. Magenta boxes correspond to ICA algorithms whereas blue boxes correspond to ISA algorithms
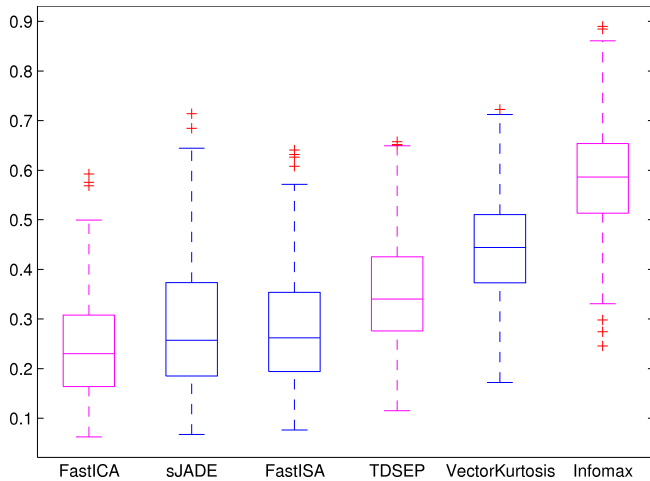

Fig. 3. [2+2+2] Values of the subspace Amari Performance Index (SAPI) for the six algorithms on the 500 datasets with 2+2+2 structure. The box plots represent similar quantities as Fig. 2

The main conclusion is very interesting, albeit unexpected: the best overall algorithm for this type of data appears to be FastICA, an algorithm which was not taylored for tackling subspaces. The performance depends on the random elements (phase lags, choice of MEG channels, choice of submatrix), but the typical SAPI value for FastICA is of the order of 0.1 to 0.2 for the 2+2 and 3+3 cases, which correspond to good

separations. The 2+2+2 case worsens the results on all algorithms, but FastICA's results of 0.18 to 0.3 are still acceptable.
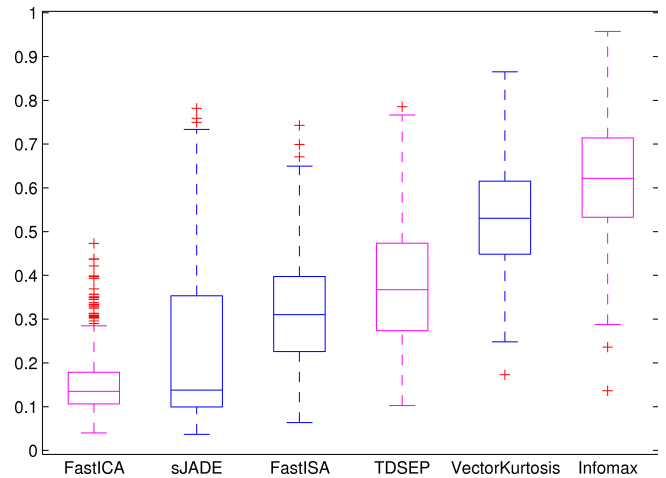

Fig. 4. [3+3] Values of the subspace Amari Performance Index (SAPI) for the six algorithms on the 500 datasets with 3+3 structure. The box plots represent similar quantities as Fig. 2

5. Discussion

The algorithms used here require some parameter choices. We found that the results shown here are relatively robust to such choices: for example, FastICA can be used with several non-linearities, all of which yield similar results. It can also be used in a symmetric or deflation mode, and again, both yield similar results. The choice of lags in TDSEP is the only somewhat critical parameter; however, since we downsample the data according to the sampling frequency and the filter bandwidth, we simply used all lags from 0 to 8. The reader is referred to the references given in Sec. 3 for more information on the effect of these parameters and on how they should be chosen.

The results shown here allow an important conclusion for the goal of performing full ISA when each subspace is composed of synchronous sources. FastICA yields good results despite being an ICA algorithm and not an ISA algorithm. Algorithms for the intra-subspace separation already exist for this type of sources [9, 25]; thus, the full ISA task can, in theory, be satisfactorily performed if one devises a proper subspace detection procedure. Note, however, that this may depend strongly on the type of dependency in each subspace, since other authors have found rather different results on other types of data [14].

In practice, some questions still need to be addressed. The most important of these is the performance of these methods on actual EEG/MEG data. The MEG-like data used here allows a quantitative measurement of the quality of the separation; such measurement is not available on real EEG or MEG. On the other hand, natural EEG/MEG data contain disturbing properties (such as artifacts) which were simplified in the present data generation.

Another relevant issue is the performance when the number of sources and measurements is large (more than 100 as in EEG/MEG): in this case, one can compress the data onto a

smaller dimension using, e.g., Principal Component Analysis [5], but the effective number of sources will still likely be larger than 6. Unfortunately, due to time limitations, we could not run simulations for larger numbers of sources in time for this submission. Finally, the choice of the frequency band to analyze is not innocuous, since different frequency bands contain information about different brain phenomena [26].

6. Conclusions

We have studied the performance of several algorithms for the separation of subspaces of synchronous sources, using MEG-like data to assess the quality of the separation. We have found that FastICA yields the best results on this type of data, despite the fact that it is an ICA algorithm and not an ISA algorithm. Nevertheless, sJADE (in general) or FastISA (if the subspace dimensions are known in advance) also seem to be able to deal reasonably well with the problem of partial ISA.

REFERENCES

[1] A. Pikovsky, M. Rosenblum, and J. Kurths, *Synchronization: a Universal Concept in Nonlinear Sciences, Cambridge Nonlinear Science Series*, Cambridge University Press, Cambridge, 2001.

[2] P.J. Uhlhaas and W. Singer, "Neural synchrony in brain disorders: relevance for cognitive dysfunctions and pathophysiology", *Neuron* 52, 155–168 (2006).

[3] P.L. Nunez, R. Srinivasan, A.F. Westdorp, R.S. Wijesinghe, D.M. Tucker, R.B. Silberstein, and P.J. Cadusch, "EEG coherency I: statistics, reference electrode, volume conduction, laplacians, cortical imaging, and interpretation at multiple scales", *Electroencephalography and Clinical Neurophysiology* 103, 499–515 (1997).

[4] R. Vigário, J. Särelä, V. Jousmäki, M. Hämäläinen, and E. Oja, "Independent component approach to the analysis of EEG and MEG recordings", *IEEE Trans. on Biom. Eng.* 47 (5), 589–593 (2000).

[5] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. John Wiley & Sons, London, 2001.

[6] A. Cichocki and S. Amari, *Adaptive Blind Signal and Image Processing – Learning Algorithms and Applications*, John Wiley & Sons, London, 2002.

[7] P. Hoyer, "Non-negative matrix factorization with sparseness constraints", *J. Machine Learning Research* 5, 1457–1469 (2004).

[8] D. Lee and H. Seung, "Algorithms for non-negative matrix factorization", *Advances in Neural Information Processing Systems* 13, 556–562 (2001).

[9] M. Almeida, J.-H. Schleimer, J. Bioucas-Dias, and R. Vigário, "Source separation and clustering of phase-locked subspaces", *IEEE Trans. on Neural Networks* 22 (9), 1419–1434 (2011).

[10] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, Cambridge, 2004.

[11] E. Alhoniemi, A. Honkela, K. Lagus, and J. Seppä, "Compact modeling of data using independent variable group analysis", *IEEE Trans. on Neural Networks* 18, 1762–1776 (2007).

[12] B. Póczos and A. Lörincz, "Independent subspace analysis using geodesic spanning trees", *Proc. Int. Conf. on Machine Learning (ICML)* 1, CD-ROM (2005).

[13] J. Beirlant, E. Dudewicz, L. Gyorfi, and E. van der Meulen, "Nonparametric entropy estimation: an overview", *Int. J. Mathematical and Statistical Sciences* 6, 17–39 (1997).

[14] H.W. Gutch, J. Krumsiek, and F.J. Theis, "An ISA algorithm with unknown group sizes identifies meaningful clusters in metabolomics data", *Proc. Eur. Signal Processing Conf. (EUSIPCO)* 1, CD-ROM (2011).

[15] A. Hyvärinen and U. Köster, "FastISA: A fast fixed-point algorithm for independent subspace analysis", *Proc. Eur. Symposium on Artificial Neural Networks (ESANN)* 1, CD-ROM (2006).

[16] A. Sharma and K.K. Paliwal, "Subspace independent component analysis using vector kurtosis", *Pattern Recognition* 39, 2227–2232 (2006).

[17] J.A. Palmer and S. Makeig, "Blind separation of dependent sources and subspaces by minimum mutual information", in *Technical Report*, University of California, San Diego, 2010.

[18] Z. Szabó, B. Póczos, and A. Lorincz, "Undercomplete blind subspace deconvolution", *J. Machine Learning Research* 8, 1063–1095 (2007).

[19] F.J. Theis, "Towards a general independent subspace analysis", *Advances in Neural Information Processing Systems (NIPS)* 1, CD-ROM (2007).

[20] L.B. Almeida, "MISEP – linear and nonlinear ICA based on mutual information", *J. Machine Learning Research* 4, 1297–1318 (2004).

[21] A. Ziehe and K.-R. Müller, "TDSEP – an efficient algorithm for blind separation using time structure", *Int. Conf. on Artificial Neural Networks* 1, 675–680 (1998).

[22] R. Vigário, V. Jousmäki, M. Hämäläinen, R. Hari, and E. Oja, "Independent component analysis for identification of artifacts in magnetoencephalographic recordings", *Advances in NIPS* 10, 229–235 (1997).

[23] A.V. Oppenheim, R.W. Schafer, and J.R. Buck, *Discrete-Time Signal Processing*, Prentice-Hall International Editions, London, 1999.

[24] T. Eichele, S. Rachakonda, B. Brakedal, R. Eikeland, and V. D. Calhoun, "EEGIFT: Group independent component analysis for event-related EEG data", *Computational Intelligence and Neuroscience* 2011, 1–9 (2011).

[25] M. Almeida, R. Vigário, and J. Bioucas-Dias, "Estimation of the common oscillation for phase locked matrix factorization", *Proc. Int. Conf. on Pattern Recognition Applications and Methods (ICPRAM)* 1, CD-ROM (2012).

[26] P.L. Nunez and R. Srinivasan, *Electric Fields of the Brain: the Neurophysics of EEG*, Oxford University Press, Oxford, 2006.