Spatially Constrained ICA algorithms with applications in EEG processing

Maarten De Vos, Lieven De Lathauwer and Sabine Van Huffel
ESAT-SISTA, K.U. Leuven
Kasteelpark Arenberg 10, B-3001 Leuven-Heverlee, Belgium

March 4, 2009

Abstract

TO DO

1 Introduction

Independent Component Analysis (ICA) or Blind Source Separation (BSS) [7, 15, 9] is a technique that extracts statistically independent components from a set of measured signals. Nowadays, it has become a standard technique in neuroscience in order to decompose electroencephalogram (EEG) and functional Magnetic Resonance Imaging (fMRI) recordings into individual brain activation components. Standard ICA has two limitations, which form the starting point for this paper. First, the extracted components are not ordered. When someone is only looking for the presence of one component, all the extracted components have to be visually inspected. Usually, complex features have then to be computed in order to recognise the signal of interest. Second, in some applications, there is prior knowledge about the source signals being sought. ICA methods are blind box techniques that do not exploit this available prior knowledge.

In order to make this BSS methods more intelligent, several attempts have been made to incorporate the available signal characteristics and develop 'semi-blind' source separation techniques. If one has the available information on the statistical distributions, one could incorporate prior knowledge in a Bayesian framework, see e.g. [20, 12]. This has the disadvantage of having to specify all priors for the distribution of the model parameters. Furthermore, no closed form equations can be derived. If the source distributions are not fully known, but it is known that some components have large second- or higher-order statistics, a so-called 'robust BSS approach' has been introduced in [19]. When only kurtosis signs are known, a BSS algorithm is proposed in [23]. When information about the shape of one or more source signals are assumed, this prior knowledge can be used as a constraint in the ICA optimisation algorithm, leading to the introduction of temporally constrained ICA [10]. Developing on this idea, in e.g. [18, 11, 3] other optimisation algorithms are proposed to extract sources resembling a reference signal.
Besides trying to use information about the source signals, another way to constrain ICA is to impose constraints on the columns of the mixing vector. This is useful when the spatial distribution of one (or more) source(s) is (approximately) known. In [8], a first algorithm was proposed. At every iteration step, the BSS criterion was optimised, and afterwards the spatial constraint was enforced. The paper also contains a thorough discussion on the difference between hard and soft constraints and on the trade-off between imposing prior knowledge constraints and statistically independence constraints. It should be clear for the reader that in some extreme situations prior knowledge can be conflicting with mathematically independence assumptions. In [17], it is mentioned (without giving details) that spatial constraints can be incorporated with Lagrange multipliers to solve an unconstrained problem.
In this paper, we propose new algorithms to incorporate prior knowledge into the popular ICA

implementations Jade [6] and Sobi [2]. The algorithms assume that there is information about one or more mixing vectors. The novelty in this paper is threefold. First, we derive detailed closed-form formulas based on numerical algebra. Second, we show how the prior knowledge should be taken into account during the so-called pre-whitening step. Third, we evaluate the improvement in performance by incorporating prior knowledge in a simulation study.
This paper is organized as follows. First (§1.1), the necessary background on ICA algorithms, and the specific Jade and Sobi algorithms, is repeated. Second (§2), we develop our ideas about Spatially Constrained ICA (scICA). Two variants are discussed. A first algorithm computes the ICA decomposition when a mixing vector is exactly known. A second algorithm shows how the constrained decomposition can be computed when there is only limited confidence in the mixing vector. Low-level formulas will be given to weigh the prior knowledge assumption against the statistically independence assumption. We conclude this paper with validating the methods in a simulation study. Both methods are also illustrated on real-life electroencephalogram (EEG) data. The algorithm with fixed constraints can be useful for investigating brain activity at a specified location, e.g. the irritative zone in epilepsy. Incorporating weighted prior knowledge will be incorporated in an accurate eye artifact removal technique in neuroscience.

1.1 Algorithmic details on Independent Component Analysis

Assume the basic linear statistical model

$$\mathbf{Y} = \mathbf{M} \cdot \mathbf{S} \quad (+N) \tag{1}$$

where $\mathbf{Y} \in \mathbb{R}^{I \times T}$ is called the observation matrix, $\mathbf{S} \in \mathbb{R}^{J \times T}$ the source matrix and $\mathbf{N} \in \mathbb{R}^{I \times T}$ the additive noise. $\mathbf{M} \in \mathbb{R}^{I \times J}$ is the mixing matrix. I is the number of sensors, J is the number of sources, and T the number of samples in each observation.
The goal of Independent Component Analysis is to estimate the mixing matrix $\mathbf{M}$, and/or the source matrix $\mathbf{X}$, given only realizations of the observation matrix $\mathbf{Y}$. In this work, we restrict our discussion to the case that there are more sensors than sources, i.e. $J \leqslant I$. For strategies to solve the underdetermined ICA problem, when there are more sources then sensors, we refer to [13] and references herein.
Blind identification of $\mathbf{M}$ in (1) is only possible when other assumptions about the sources are made. The most commonly used assumption is that the sources are mutually statistically independent, as well as independent from the noise components. A computationally efficient algorithm, called Joint Approximate Diagonalization of Eigenmatrices (JADE) is described in [6]. An overview of other algorithms is given in [9, 15]. Another assumption that can be made is that the sources are mutually uncorrelated but individually autocorrelated. Then, an algorithm based on second order statistics can be used to estimate the underlying sources, Second Order Blind Identification (Sobi) [2].
In the following we start with a note on pre-whitening, because we propose alternatives in §2.1.1 and §2.2.1. We also shortly review the idea of JADE and SOBI as starting point for §2.1.2 (and §2.2.2) and simultaneous matrix diagonalization as a starting point for §2.2.2.

1.2 Pre-whitening

From an algebraic point of view, saying that realizations of X are uncorrelated means that the covariance matrix of X, $\mathbf{C}_X$, is a diagonal matrix. Saying that realizations of X are independent also implies that all the higher order cumulants of X, $\mathcal{C}_x^{(N)}$, are diagonal. With the linear model from equation (1), we can also write:

$$\mathbf{C}_Y = \mathbf{C}_X \times_1 \mathbf{M} \times_2 \mathbf{M} + \mathbf{C}_n \tag{2}$$

$$\mathcal{C}_y^{(N)} = \mathcal{C}_x^{(N)} \times_1 \mathbf{M} \times_2 \mathbf{M} \ldots \times_N \mathbf{M} + \mathcal{C}_n^{(N)} \tag{3}$$

in which $\mathbf{C}_X$ and $\mathcal{C}_x^{(N)}$ are diagonal and $\mathcal{C}_n^{(N)}$ drops if the noise is gaussian. Different kinds of ICA algorithms can be distinguished, depending on how equations (2) and (3) are combined. One

approach is to solve the ICA problem with only higher-order statistics. For algebraic techniques, we refer to [4, 14]. However, most algorithms start with a so-called pre-whitening step. During this pre-whitening step, as much information as possible is extracted from (2). In this stage, the mixing matrix can be identified up to an orthogonal rotation matrix. Afterwards, (3) has to be solved in some approximative manner.
During the pre-whitening step, the observation matrix $\mathbf{Y}$ is transformed into another stochastic variable with unit variance. In formulas, pre-whitening of the data amounts to the computation of the covariance matrix $\mathbf{R}_Y$ of the signal:

$$\begin{aligned}\mathbf{R}_Y &= E\{\mathbf{Y}\cdot\mathbf{Y})^T\} && (4)\\ &= E\{(\mathbf{M}\cdot\mathbf{X})(\mathbf{M}\cdot\mathbf{X})^T\} && (5)\\ &= \mathbf{M}\cdot\mathbf{R}_X\cdot\mathbf{M}^T && (6)\\ &= (\mathbf{US})\cdot(\mathbf{US})^T && (7)\end{aligned}$$

In the last step, we assume without loss of generality that the source signals have unit variance, and we replace $\mathbf{M}$ by its singular value decomposition $\mathbf{USV}^T$.The whitening matrix is then given by the pseudo-inverse of this square root of the covariance matrix., $\mathbf{US}$. In many applications of ICA (e.g. EEG, fMRI, ...), the number of observations is typically higher than the number of sources. In the pre-whitening step, the dimensionality of the given ICA problem is reduced: when $R < I$, a projection of Y onto the signal subspace is carried out. From these equations, it can be seen that the mixing matrix $\mathbf{M} = \mathbf{USV}^T$ can only be estimated from 7 up to an orthogonal matrix $\mathbf{V}$.
Although many ICA algorithms start with a pre-whitening step, this step has the disadvantage that the calculations are directly affected by additive Gaussian noise. Recall that (3) is blind to additive Gaussian noise. Errors introduced in the pre-whitening step cannot be compensated in the higher-order step [5, 16].

1.3 Algoritmes

1.3.1 JADE

The assumption of mutually statistically independent sources means algebraically that all the higher-order cumulants of X are diagonal tensors. From (3) it can be seen that the mixing matrix can be estimated as the matrix that diagonalizes the higher-order cumulant. This mixing matrix will be an orthogonal matrix when the observation matrix is first pre-whitened. A very efficient algorithm "to diagonalize the cumulant" is described in [6]. When the cumulant slices are rearranged into a matrix $\mathcal{C}_y^{(4)} \in \mathbb{R}^{J^2\times J^2}$, it was shown that the mixing matrix can be estimated as the orthogonal matrix that *simultaneously* diagonalizes different matrices that form a basis of the range of this matrix. The best choice for this set of matrices are the P dominant eigenmatrices $\mathbf{E}_p$ of $(\mathcal{C}_y^4)$. So in order to estimate the mixing matrix, the P dominant eigenmatrices $\mathbf{E}_p$ of $(\mathcal{C}_y^4)$ have to be jointly (approximately) diagonalized. For the computation of an orthogonal simultaneous matrix diagonalization, one can resort to techniques developed in e.g. [6, 21]. We review in §1.4 the technique based on Jacobi angles.

1.3.2 SOBI

When the sources are individually correlated in time, but mutually uncorrelated, an ICA algorithm based on second order statistics can be derived [1].
Mathematically, this means that not only the covariance matrix of the sources is diagonal, but also that the covariance matrices at different time lags $t+\tau$ are diagonal. Under these assumptions, the mixing matrix $\mathbf{M}$ is the matrix that jointly diagonalizes all the covariance matrices:

$$\begin{aligned}\mathbf{R}_y(\tau) &= E\{\mathbf{Y}(t)\mathbf{Y}(t+\tau)^T\} && (8)\\ &= \mathbf{M}\cdot\mathbf{R}_x(\tau)\cdot\mathbf{M}^T \qquad \forall\tau\neq 0 && (9)\end{aligned}$$

Again, after pre-whitening of the observation matrix, optimal Jacobi angles can be derived for this orthogonal simultaneous diagonalization.

1.4 Jacobi angles for simultaneous diagonalization

A Jacobi rotation is defined as a rank-2 correction at positions i and j to the identity of the form

$$\mathbf{J}(i,j) = \begin{pmatrix} 1 & \dots & 0 & \dots & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & & \vdots & & \vdots \\ 0 & \dots & c & \dots & s & \dots & 0 \\ \vdots & & \vdots & \ddots & \vdots & & \vdots \\ 0 & \dots & -s & \dots & c & \dots & 0 \\ \vdots & & \vdots & & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & \dots & 0 & \dots & 1 \end{pmatrix} \tag{10}$$

with $c = cos(\theta), s = sin(\theta)$ In [6], it is shown that the optimal Jacobi angles for simultaneous diagonalization are derived from an eigenvalue problem. It is important to emphasize that this technique can be used for both the orthogonal diagonalization of cumulant slices and the correlation matrices. The goal is to maximize the diagonal elements of $\mathbf{J}^T\mathbf{R}_p(\tau)\mathbf{J}$ with $\mathbf{J}$ a Jacobi rotation. As can be seen from (10), only the elements at positions i and j are changed. At every iteration, we can write the corresponding contrast function as

$$f = \sum_{p}^{\tau} \tilde{u}_{iip}^2 + \tilde{u}_{jjp}^2 \tag{11}$$

where
$$\tilde{u}_{iip} = (c \quad -s) \begin{pmatrix} u_{iip} & u_{ijp} \\ u_{jip} & u_{jjp} \end{pmatrix} \begin{pmatrix} c \\ -s \end{pmatrix} \tag{12}$$

and
$$\tilde{u}_{jjp} = (s \quad c) \begin{pmatrix} u_{iip} & u_{ijp} \\ u_{jip} & u_{jjp} \end{pmatrix} \begin{pmatrix} s \\ c \end{pmatrix}. \tag{13}$$

u_{ijp} denotes the $(i,j)^{th}$ entry of the pth correlation matrix, $i \leqslant i < j \leqslant J$. It has been shown that this contrast can also be written as:

$$f = W^T \cdot \mathbf{G} \cdot W \tag{14}$$

when we define (15)

$$W = (\cos(2\theta) \qquad \sin(2\theta))^T, \tag{16}$$

$$\mathbf{G} = \sum_{p} Z_p Z_p^T \tag{17}$$

and
$$Z_p = \frac{1}{2} \begin{pmatrix} u_{iip} - u_{jjp} \\ u_{ijp} + u_{jip} \end{pmatrix} \tag{18}$$

(14) is an eigenvalue problem. This means that the rotation angle θ for the optimal Jacobi rotation at one iteration is given by the dominant eigenvector W of this eigenvalue problem.

2 Imposing spatial constraints

There are two ways to incorporate prior knowledge about a mixing vector during ICA computation. Either we assume that we exactly know a mixing vector (§2.1), i.e., that we know the location of one source with respect to the sensors. Then the extracted mixing matrix should have this vector as one of its columns. Or we have a good idea about a mixing vector (§2.2), and we can weight the degree of confidence in this location with respect to the independence constraint. In this case,

the use of prior knowledge can help the ICA algorithm to converge towards the desired solution. For both situations, we explain first how we assure that the mixing vector is fully (§2.1.1) or partially (§2.2.1) contained by the signal subspace. This step is necessary for reducing the dimensionality of our problem. In a next step (§2.1.2 and §2.2.2), we estimate the full mixing matrix.

2.1 Fixed constraint

We assume the observations are I-dimensional, that there are J sources and that we know exactly one mixing vector M_1. The extension to more mixing vectors is straightforward.

2.1.1 Pre-whitening step

We have to define the dominant signal subspace $\mathbf{U}$ of dimension J that include M_1. In other words, we have to find an orthonormal matrix $\mathbf{U}$ that maximizes the norm of the projection of our data on $\mathbf{U}$ constrained to M_1 is contained in span($\mathbf{U}$). This can be done as follows:

1. Normalise M_1 to unit length. The result will be U_1.
2. Project the observation matrix $\mathbf{Y}$ on the orthogonal complement of U_1 in order to obtain $\mathbf{Y2}$.
3. Define U_2 to U_J as the $J-1$ dominant eigenvectors from $\mathbf{Y2}$.

It can be seen that by this procedure, M_1 is constrained to be in the column space of $\mathbf{U}$. In other words, M_1 lies in the estimated signal subspace. In order to reduce the dimensionality, we project our datamatrix on this subspace:

$$\mathbf{Y}_s = \mathbf{U}^T \cdot \mathbf{Y} \tag{19}$$

.

2.1.2 Identifying the mixing matrix

After the pre-whitening step, we know the column space of the mixing matrix.

$$\begin{aligned}\mathbf{C}_2^{Y_s} &= E\{\mathbf{Y}_s \cdot \mathbf{Y}_s^T\} \\ &= \mathbf{M} \cdot \mathbf{C}_2^X \cdot \mathbf{M}^T \\ &= (\mathbf{VS}) \cdot (\mathbf{VS})^T\end{aligned}$$

and we are dealing with a square ICA problem. The full mixing matrix can be expressed as:

$$\mathbf{M} = \mathbf{VSQ} \tag{20}$$

, where $\mathbf{V}$ and $\mathbf{S}$ are known. In order to estimate the full mixing matrix,

1. We notice that we can compute from this equation the first column Q_1 of $\mathbf{Q}$, because we know M_1.
2. In order to identify the other columns of $\mathbf{Q}$, we project $\mathbf{Y}_s$ on the orthogonal complement of Q_1.
3. The other columns of $\mathbf{J}$ can be estimated by computing an ICA decomposition of dimensions $J-1$ with Sobi or Jade.

2.2 Weighted constraint

We give a second algorithm that can be used when one has a good idea about a mixing vector M_1.

2.2.1 Pre-whitening step

Similar to §2.1.1, we have to define a signal subspace $\mathbf{U}$ of dimension J. This time, we are only up to a certain level confident that the subspace fully includes M_1. We have to find a trade-off by defining an orthonormal matrix $\mathbf{U}$ between maximizing the norm of the projection of our data on $\mathbf{U}$ and the constraint that M_1 is contained in span($\mathbf{U}$). Following procedure fulfills our objectives:

1. Compute the covariance matrix of the observation matrix:

$$\begin{aligned}\mathbf{C}_2^Y &= E\{\mathbf{Y}\cdot\mathbf{Y}^T\}\\ &= \mathbf{M}\cdot\mathbf{C}_2^X\cdot\mathbf{M}^T\\ &= (\mathbf{PS})\cdot(\mathbf{PS})^T\end{aligned}$$

2. Keep the J-1 eigenvectors $[P_1,\ldots,P_{J-1}]$ corresponding to the J-1 largest eigenvalues.

3. Compute the component of M_1 orthogonal to this subspace and normalize this vector to unit length in order to obtain M_1^o.

4. The Jth component U_J is then defined by the weighted average between M_1^o and P_J. The weighting factor should be defined by the confidence in accuracy of M_1 and the Jth singular value.

5. $\mathbf{U}_J$ is normalized to unit length in order to obtain $\mathbf{U}$.

In order to reduce the dimensionality, we project our datamatrix on this subspace:

$$\mathbf{Y}_s = \mathbf{U}^T\cdot\mathbf{Y} \tag{21}$$

.

2.2.2 Identifying the mixing matrix

Recall that Jade and Sobi rely on Jacobi iterations. We mentioned in §1.4 that a Jacobi iteration, working on rows i and j, is computed by an eigenvalue problem.
In this section, we show how we modify the computation of the optimal Jacobi rotation to take into account that the first mixing vector should be similar to M_1. When $i \neq 1$, nothing changes and equation (14) defines the optimal rotation. When $i = 1$, we modify the contrast function by adding the weighted squared projection of M_1 on the updated mixing vector. This updated mixing vector will be called $\widetilde{M}_1$ and is the first column of the product of all the previous Jacobi matrices. This additional term ensures that $\widetilde{M}_1$ will become more similar to M_1. We will derive below that this modified contrast function can be written as the modified eigen value problem:

$$f = W^T\cdot\mathbf{G}\cdot W + w\cdot K^T W\,(+C). \tag{22}$$

w is a predefined weight reflecting the confidence in the known mixing vector and
$W = (\cos(2\theta)\sin(2\theta))^T$, as in (14) .
The extra term $w*|M_1^T(\widetilde{M}_{new})_1|^2$ has to be written as a function of W. Starting from the update of the mixing matrix from $\widetilde{\mathbf{M}}_{old}$ to $\widetilde{\mathbf{M}}_{new}$ during an iteration with Jacobi matrix (working on i=1 and j):

$$\widetilde{\mathbf{M}}_{new} = \widetilde{\mathbf{M}}_{old}\cdot\mathbf{J}^T, \tag{23}$$

the first column of the updated mixing matrix $\widetilde{\mathbf{M}}_{new}$ can be written as:

$$(\widetilde{M}_{new})_1 = s\cdot(\widetilde{M}_{old})_j + c\cdot(\widetilde{M}_{old})_1 \tag{24}$$

Hence,

$$w * |M_1^T(\widetilde{M}_{new})_1|^2 = w * c \sum_k m_k(\tilde{m}_{old})_{k1} \left(c \sum_l m_l(\tilde{m}_{old})_{l1} + s \sum_l m_l(\tilde{m}_{old})_{lj}\right) + w * s \sum_k m_k(\tilde{m}_{old})_{kj} \left(s \sum_l m_l(\tilde{m}_{old})_{lj} + c \sum_l m_l(\tilde{m}_{old})_{l1}\right) \quad (25)$$

With defining

$$e_1 = w \sum_k m_k(\tilde{m}_{old})_{k1} \left(\sum_l m_l(\tilde{m}_{old})_{l1}\right),$$

$$e_2 = w \sum_k m_k(\tilde{m}_{old})_{kj} \left(\sum_l m_l(\tilde{m}_{old})_{lj}\right) \quad (26)$$

and

$$e_3 = w \sum_k m_k \left((\tilde{m}_{old})_{k1} \sum_l m_l(\tilde{m}_{old})_{lj} + (\tilde{m}_{old})_{kj} \sum_l m_l(\tilde{m}_{old})_{l1}\right), \quad (27)$$

we can write:

$$w * |M_1^T(\widetilde{M}_{new})_1|^2 = e_1 \cos^2(\theta) + e_2 \sin^2(\theta) + e_3 \cos(\theta)\sin(\theta). \quad (28)$$

With following substitution

$$\cos(2\theta) = \cos^2(\theta) - \sin^2(\theta), \quad (29)$$

$$\sin(2\theta) = 2\cos(\theta)\sin(\theta), \quad (30)$$

and

$$1 = \cos^2(\theta) + \sin^2(\theta,) \quad (31)$$

equation (22) can be rewritten as a homogeneous equation in cos and sin:

$$f = g_{11}(\cos^4(\theta) + \sin^4(\theta) - 2\cos^2(\theta)\sin^2(\theta)) + (2g_{12} + 2g_{21})(\sin(\theta)\cos^3(\theta) - \sin^3(\theta)\cos(\theta)) + 4g_{22}\sin^2(\theta)cos^2(\theta) + e_1(\cos^4(\theta) + \cos^2(\theta)\sin^2(\theta)) + e_2(\sin^4(\theta) + \cos^2(\theta)\sin^2(\theta)) + e_3(\cos^3(\theta)\sin(\theta) + \cos(\theta)\sin^3(\theta)).$$

In order to compute the θ that maximises this contrast function, we set the derivative of this function with respect to θ to 0:

$$cos^4(\theta)(2g_{21} + 2g_{12} + e_3) + cos^3(\theta)sin(\theta)(8g_{22} - 8g_{11} - 2e_1 + 2e_2) + cos^2(\theta)sin^2(\theta)(-12g_{12} - 12g_{21}) + cos(\theta)sin^3(\theta)(8g_{11} + 2e_2 - 2e_1 - 8g_{22}) + sin^4(\theta)(-e_3 + 2g_{12} + 2g_{21}) = 0$$

Dividing this homogeneous equation by $\cos(\theta)$ gives a fourth-order polynomial in $\tan(\theta)$. Selecting the root that maximises (22) gives the optimal Jacobi rotation θ.

3 Simulation study

In this section, we illustrate the performance of the proposed algorithms by means of numerical experiments and compare it to standard ICA.
Independent sources $\mathbf{S}$ are generated from binary source distributions (1 or -1), with an equal probability of both values. 600 samples of each source were considered.
Rank-R observation matrices $\tilde{\mathbf{Y}}$ are generated in the following way:

$$\tilde{\mathbf{Y}} = \frac{\mathbf{Y}}{||\mathbf{Y}||_F} + \sigma_N \frac{\mathbf{N}}{||\mathbf{N}||_F}, \quad (32)$$

in which $\mathbf{Y} = \mathbf{MS}$ exactly satisfies a linear mixture model with $R = 4$ independent sources and $\mathbf{N}$ represents gaussian noise.
We conduct Monte Carlo simulations consisting of 250 runs. We evaluate the performance of the different algorithms by means of the normalized Frobenius norm of the difference between the estimated and the real sources:

$$error_S = \frac{||\mathbf{S} - \hat{\mathbf{S}}||_F}{||\mathbf{S}||_F}. \quad (33)$$

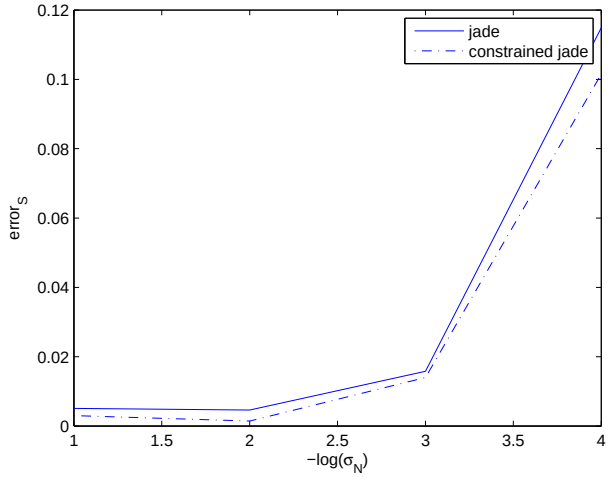

Figure 1: The mean value of $error_S$ as a function of the noise level σ_N for unconstrained jade (solid) and fixing 2 mixing vecto

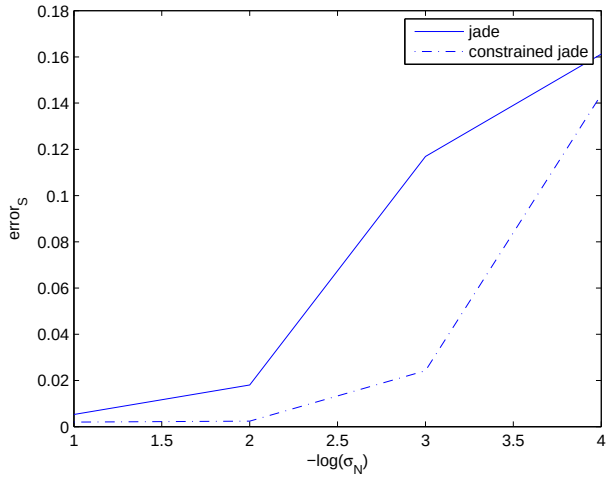

Figure 2: The mean value of $error_C$ as a function of the noise level σ_N for unconstrained jade (solid) and fixing 2 mixing vectors (dash-dot). Some sources have 10 x larger variance than others.

We performed 3 simulations. In the first (figure 1), we gave all sources equal contributions in the mixture and we assumed 2 mixing vectors to be known. The figure shows that having perfect knowledge of 2 mixing vectors improves slightly the estimation of the sources. In a second simulation (figure 2), we also considered that 2 sources were exactly known. Here, we gave 2 sources a 10 x stronger contribution in the mixture. It can be seen that knowing the source subspace improves source estimation a lot. At noise level 0.1, some sources are with the standard ICA not well estimated. By including prior knowledge, these sources can also be detected. In a last simulation, we used the algorithm where we weigh the confidence in the known mixing vectors. The choice of the specific weight is an application dependent issue, and outside the scope of this article. It should be clear for the reader that the algorithm with a low weight approaches unconstrained ICA, while imposing a high weight approaches our algorithm with fixed mixing vector. The advantage for practical use over the algorithm with fixed constrains is that no perfect knowledge of a mixing vector is required. The third figure shows the result when we allow 2% variations on the prior knowledge, an put the weight to 95 %.
These simulations show that a higher accuracy can certainly be obtained when prior knowledge is incorporated during the estimation of the independent component. Furthermore, as we mentioned in the introduction, the components obtained from standard ICA are not ordered and have to be inspected afterwards. ICA with prior knowledge puts the signal of interest in the first component(s), so postprocessing is not needed anymore.

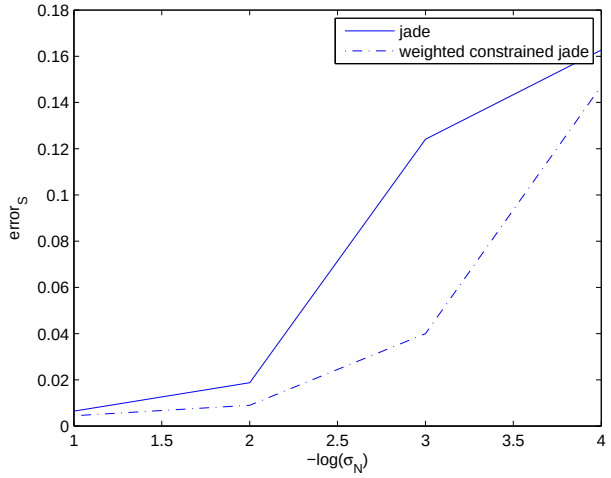

Figure 3: The mean value of $error_C$ as a function of the noise level σ_N for unconstrained jade (solid) and adding weighted prior knowledge for 2 mixing vectors (dash-dot). Some sources have 10 x larger variance than others.

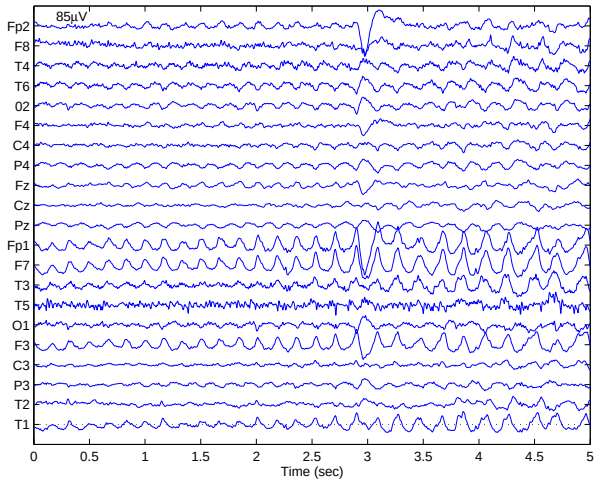

Figure 4: The original EEG.

4 Applications in EEG signal processing

We briefly illustrate two main applications of the newly proposed method.

4.1 Eye-artefact removal

We illustrate the use of weighted constrained ICA with the removal of eye-artifacts. The spatial distribution of eye blinks can be supposed to be known from other ICA decompositions. There is little variation between artifacts from different decompositions and/or different people, so the algorithm with weighted prior knowledge can be used. Figure 4 shows the original EEG dataset containing an eye blink. Figure 5 shows the EEG when the artefact is removed in an automatic way by imposing during the computation of ICA that the first component is related to eye-blinks, and afterwards subtracting this component. In [22], we showed that standard ICA was not able to remove the artifact.

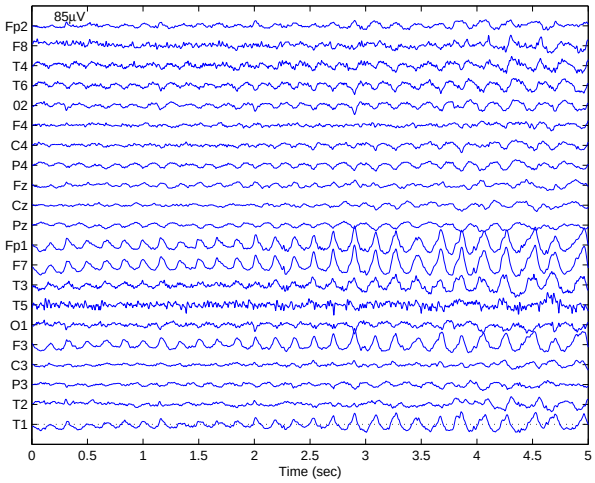

Figure 5: The EEG after eye artifact removal.

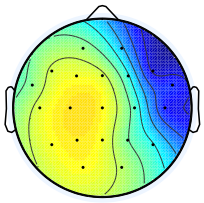

Figure 6: The mixing vector of the place in the head we want to monitor.

4.2 Spatial channel monitoring

In epilepsy research, one is interested in delineating the epileptogenic zone, i.e., the zone that is responsible for causing epileptic seizures. An other zone is the irritative zone, which is defined as the region in which interictal transients are generated. It can be very useful to monitor the signal generated in this irritative zone that is independent from the rest of the brain. When the spatial position of your irritative region is known, it can be considered to largely overlap with the epileptogenic zone. In other words, an imaginary electrode can be created at the position of interest in order to monitor the signal at this location. We have already the spatial distribution from the eye artifacts from this patient. We derive also from a spike the location of the irritative zone (figure 6). This can be done by computing a singular value decomposition of the spike, and keeping the first singular vector as prior knowledge. We decompose a seizure from that patient (figure 7) with an ICA decomposition with fixed prior knowledge. Figure 8 then shows the electrical signal corresponding to the region of observation. It is clear from comparison between the 2 figures that it is much easier to detect the seizure on the signle channel as the rhythmical peaks are prominent. In this way, instead of implementing seizure detection methods, which investigate multiple channels, only the activity of one known region has to be monitored.

5 Conclusion

TO DO

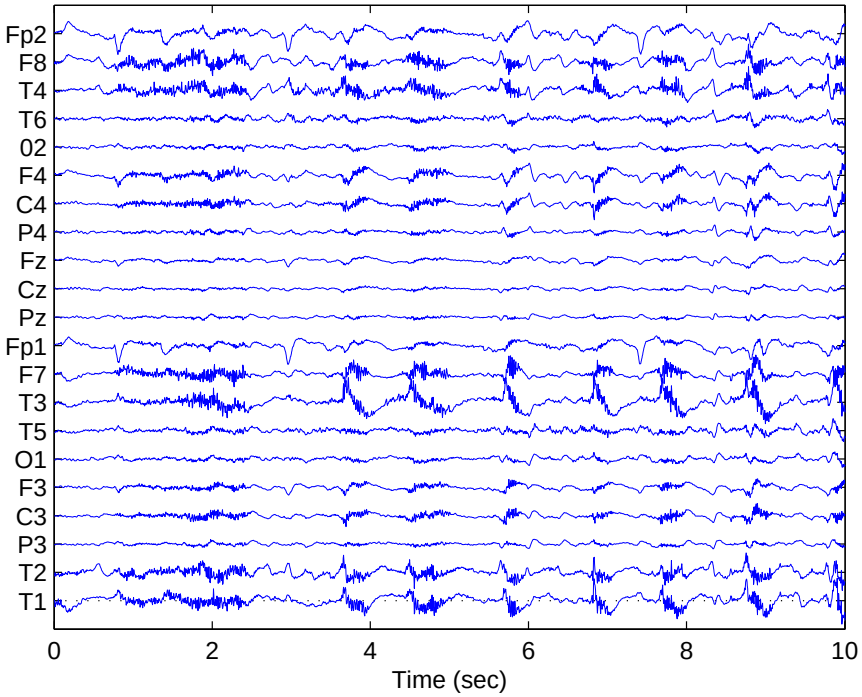

Figure 7: An epileptic seizure.

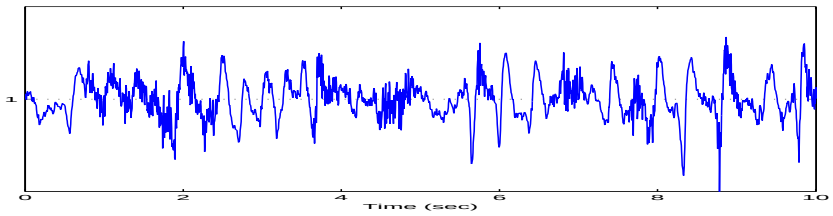

Figure 8: The epileptic signal generated at the location of interest.

References

[1] A. Belouchrani, K. Abed-Meraim, J.-F. Cardoso, and E. Moulines. A blind source separation technique using second order statistics. *IEEE Trans. Signal Process.*, 45:434–444, 1997.

[2] A. Belouchrani, K. Abed-Meraim, and Y. Hua. Jacobi-like algorithms for joint block diagonalization: Applications to source localization. In *Proc. IEEE Int Work on Intelligent Sig Proc and Comm Sys (ISPACS)*, Melbourne, Australia, 1998.

[3] V. D. Calhoun, T. Adali, M. Stevens, K. A. Kiehl, and J. J. Pekar. Semi-blind ICA of FMRI: A method for utilizing hypothesis-derived time courses in a spatial ICA analysis. *NeuroImage*, 35, 2005.

[4] J.-F. Cardoso. Iterative techniques for blind source separation using only fourth-order cumulants. In *Proc. EUSIPCO '92*, pages 739–742, Brussels, Belgium, 1992.

[5] J.-F. Cardoso. On the performance of orthogonal source separation algorithms. In *Proc. EUSIPCO '94*, pages 776–779, Edinburgh, 1998.

[6] J.-F. Cardoso and A. Souloumiac. Blind beamforming for non-gaussian signals. *IEE Proc. F*, 140:362–370, 1994.

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

[8] C.W. Hesse and James C.J. On semi-blind source separation using spatial constraints with applications in EEG analysis. *IEEE Trans. Biomed. Eng.*, 2006.

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

[10] Ch. J. James and O. J. Gibson. Temporally constrained ICA: An application to artifact rejection in electromagnetic brain signal analysis. *IEEE Trans. Biomed. Eng.*, 50:1108–1116, 2003.

[11] M. Jing and S. Sanei. A new constrained topographic independent component analysis for separation of epileptic seizure signals. *Computational Intelligence and Neuroscience*, id 21315, 2007.

[12] K.H. Knuth. Informed source separation: A Bayesian tutorial. In *Proc. 13th EUSIPCO '05*, Turkey, 2005.

[13] L. De Lathauwer, J. Castaing, and J.-F. Cardoso. Fourth-order cumulant based underdetermined independent component analysis. *IEEE Trans. Signal. Proc.*

[14] L. De Lathauwer, B. De Moor, and J. Vandewalle. Independent component analysis based on higher-order-only ica. In *Proc. of 8th IEEE SP workshop on Statistical Signal and Array Processing*, pages 356–359, Corfu, Greece, 1996.

[15] L. De Lathauwer, B. De Moor, and J. Vandewalle. An introduction to independent component analysis. *J Chemometrics*, 14:123–149, 2000.

[16] L. De Lathauwer, B. De Moor, and J. Vandewalle. A prewhitening-induced bound on the identification error in independent component analysis. *IEEE Trans. on Circuits and Systems*, 52:546–554, 2005.

[17] M.A. Latif, S. Sanei, J. Chambers, and L. Shoker. Localization of abnormal EEG sources using blind source separation partially constrained by the locations of known sources. *IEEE Signal Processing Letters*, 13:117–120, 2006.

[18] W. Lu and J.C. Rajapakse. Approach and applications of constrained ICA. *IEEE transactions on neural networks*, 16:203–212, 2005.

[19] M. Sahmoudi and K. Abed-Meraim. Robust blind separation algorithms for heavy-tailed sources. In *Proc. of ISSPIT '04, Processing and Information Technology*, Rome, Italy, 2005.

[20] H. Snoussi and A.M. Djafari. Bayesian unsupervised learning for source separation with mixture of gaussians prior. *J. VLSI Signal Process. Syst.*, 37(2-3):263–279, 2004. ISSN 0922-5773.

[21] A.-J. Van Der Veen and A. Paulraj. An analytical constant modulus algorithm. *IEEE Trans. Signal. Proc.*, 44:1136–1155, 1996.

[22] M. De Vos, L. De Lathauwer, A. Vergult, W. De Clercq, W. Van Paesschen, and S. Van Huffel. Spatially constrained independent component analysis for real-time eye artifact removal from the electroencephalogram. In *Proc. of the first Annual Symposium of the IEEE/EMBS Benelux Chapter*, pages 159–163, Brussels, Belgium, 2006.

[23] V. Zarzoso, R. Phlypo, and P. Comon. A contrast for independent component analysis with priors on the source kurtosis signs. *IEEE Signal Processing Letters*, 2008.