Published in final edited form as:

IEEE Trans Biomed Eng. 2011 October ; 58(10): 2794–2803. doi:10.1109/TBME.2011.2159841.

Application of Independent Component Analysis with Adaptive Density Model to Complex-valued fMRI Data

Hualiang Li[Member, IEEE], Nicolle M. Correa, Pedro A. Rodriguez, Vince D. Calhoun[Senior Member, IEEE], and Tülay Adalı[Fellow, IEEE]

Hualiang Li, Nicolle M. Correa, Pedro A. Rodriguez and Tülay Adalı are with the Department of Computer Science and Electrical Engineering, University of Maryland, Baltimore County, MD. Vince D. Calhoun is with the Department of ECE, University of New Mexico, Albuquerque, NM and the Mind Research Network, Albuquerque, NM

Abstract

Independent component analysis (ICA) has proven quite useful for the analysis of functional resonance magnetic imaging (fMRI) data, especially when the underlying nature of the data is hard to model. It is especially attractive for the analysis of fMRI data in its native complex form since very little is known about the nature of phase, which is typically discarded in most analyses. In this paper, we show that a complex ICA approach using a flexible nonlinearity that adapts to the source density is the more desirable one for performing ICA of complex fMRI data compared to the those that use fixed nonlinearity. We show that by adaptively matching the underlying fMRI density model, the analysis performance can be improved in terms of both the estimation of the spatial maps and the task-related time courses. We also define a procedure for analysis and visualization of complex-valued fMRI results, which includes the construction of bivariate t-maps for multiple subjects and a complex-valued ICASSO [1] scheme for evaluating the consistency of ICA algorithms.

Index Terms

Functional magnetic resonance imaging; Independent component analysis; Complex-valued signal processing; Adaptive density model

I. Introduction

Independent component analysis (ICA) has been shown to be very useful for the analysis of fMRI data [2]–[4]. By assuming that the observed data (fMRI images) are linear mixtures of unknown independent sources (activation maps), ICA algorithms can recover both the unknown linear mixing coefficients and the unknown underlying sources subject to some scaling and permutation ambiguity. ICA can minimize the constraints imposed on the temporal—or the spatial—dimension of the fMRI data, and hence provides valuable new insights, especially when studying paradigms for which reliable models of brain activity are not available. Virtually all fMRI studies analyze only the magnitude images from the MRI scanner. The information contained in the phase images is ignored in such analyses. However, it has been shown that there can be activation-dependent differences in the phase images, especially for voxels with larger venous blood fractions [5]. It has also been shown that the phase of the MR signal can provide excellent contrast for identifying in vivo anatomy with very high resolution [6]. Recent studies suggest that employing phase information in fMRI time series analysis can improve the statistical power of activation maps [7] or improve the localization of BOLD signal changes in gray matter by excluding larger vessels based on the phase signal changes they exhibit in response to activation [8].

There is growing evidence that using both the magnitude and phase in the analysis provides significant advantages, especially in terms of increased sensitivity and specificity [9]-[11]. More important, fMRI data are typically noisy and the problem is especially compounded in the case of complex data because of the high noise in the phase of the data [12]-[14]. Hence even small improvement in complex fMRI data analysis is significant and useful.

ICA using nonlinear functions, which forms by far the majority of ICA algorithms, is based on the maximum likelihood (ML) or the maximization of negentropy (MN) principle. The two criteria are equivalent when the demixing matrix is constrained to be unitary [15], the case for the algorithms that use maximization of negentropy such as the popular FastICA [16] algorithm. Both of these classes of algorithms achieve the large sample optimality properties when the form of the nonlinear function matches the form of the pdf of the sources [17], [18]. For ML estimation, the nonlinear function is also called score function, which is the derivative of log function of the source density. Hence, fixing the nonlinearity limits the achievable performance. However, the complex ICA algorithms implemented in these aforementioned references all use a fixed nonlinearity. The two early extensions of two popular ICA algorithms, the complex fastICA algorithm [19] and circular Infomax algorithm [20], assume a fixed nonlinearity to match the underlying density model, though practical and more widely used for complex ICA, will not be optimal for most cases. Since very little is known about the density model of the estimated fMRI sources within the ICA framework, and in particular of the complex case, it would be desirable to use an adaptive algorithm such that the density model is estimated for each source adaptively. Another limitation of complex fastICA algorithm and circular Infomax algorithm is their circularity assumption on the sources. A complex random variable is called circular when its pdf is invariant subject to a phase rotation. Since different fMRI sources might be circular or noncircular, it would be desirable to apply complex ICA algorithms to fMRI data without making the circularity assumption on the sources.

In this paper, we will show that adaptivity or density matching is important when noise is high. We study the performance of these two classes of algorithms, those that are based on ML and MN, and within each class, consider one commonly used solution with fixed nonlinearity and another recently proposed approach that adapts to the form of the source nonlinearity during the adaptation. The two adaptive algorithms we use in the comparison are adaptive complex maximization of non-Gaussianity (A-CMN) algorithm introduced in [21]-[23] based on the MN principle and the ICA by complex entropy bound minimization (complex ICA-EBM) algorithm presented in [24] based on the ML principle. Both algorithms assume a flexible density model and specifically, A-CMN assumes a generalized Gaussian density model and complex ICA-EBM assumes a maximum entropy density model. For both algorithms, there is no strict circularity assumption made on the sources and the parameters of the model are estimated explicitly or implicitly during the source separation.

In section IV, we study the performance of the two adaptive complex ICA algorithms along with the two popular solutions using fixed nonlinearities to an fMRI data set collected during a simple finger-tapping motor task. We show that, adaptively matching the source density model, leads to improved performance. Since the two algorithm groups, MN and ML, have fundamental differences and the most important difference is the unitary constraint on the demixing matrix for the MN-based ICA algorithm, in order to conduct a fair comparison, we have divided the four algorithms into two groups. One is the MN group that includes A-CMN and complex fastICA. The other is ML group that includes ICA-EBM and the circular Infomax algorithm. We conduct the comparisons within each group and demonstrate the advantages of adaptive density matching for the problem. In addition, we address a number of issues in the analysis and visualization of ICA results of complex fMRI

data and demonstrate the results using actual fMRI data for multiple subjects performing a motor task. Some of the preliminary results have been published in a short conference paper [25].

In the next section, we first give a brief review of complex statistics and ICA, and then introduce the A-CMN algorithm and complex ICA-EBM algorithm in Section II-B. In Section III, we explain the pre- and post-processing for the complex fMRI data and provide comparison of the performances of four competitive complex ICA algorithms when applied to fMRI analysis in section IV. The conclusions are given in the last section.

II. Theory

A. Complex ICA of fMRI data

The probability density function (pdf) of a complex random variable $X = X_r + jX_i$ is defined through its joint density function $p_{\mathbf{x}}(\mathbf{x}) \triangleq p_{X_r, X_i}(x_r, x_i)$, where the subscript r and i denote the real and imaginary parts of the variable. A complex random variable X is called circular in the strict-sense when X and $Xe^{j\theta}$ have the same pdf. A zero-mean random variable is called *second-order circular* (or proper) when its pseudo-covariance is zero, i.e., $E\{X^2\} = 0$.

In the ICA analysis of fMRI data, we assume independence of spatial brain activations (for spatial ICA), and write the complex ICA model as

$$\mathbf{x} = \mathbf{A}\mathbf{s}$$

where $\mathbf{x}$ is the random vector of observation, $\mathbf{A}$ is the mixing matrix and $\mathbf{s}$ is the independent source vector. For the volume image data of v voxels and n time points, we can write $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_v]$ and $\mathbf{S} = [\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_v]$ where $\mathbf{X} \in \mathbb{C}^{n \times v}$ is formed using the fMRI data such that the k th row is formed by flattening the volume image data of v voxels into a row. The rows of $\mathbf{X}$ are indexed as a function of time from one to n and the k th row vector of $\mathbf{S}$ is the k th spatial map. The assumed k th mixing column vector of $\mathbf{A}$ represents the time course for the k th spatial map.

Previous statistical analysis on fMRI data show that the real part $\mathbf{x}_r$ and imaginary part $\mathbf{x}_i$ of fMRI data are Gaussian distributed. This results agrees with the ICA model since the mixtures of independent random variables tend to be Gaussian distributed by the central limit theorem. Given Gaussian distributed $\mathbf{x}_r$ and imaginary part $\mathbf{x}_i$, it is straightforward to show that the magnitude of $\mathbf{x}$ is Rician distributed [26]. It has also been shown that the distribution of phase is bimodal with two peaks at $\psi/2$ and $\psi/2 - \pi$ where ψ is the phase of the circularity coefficients [27]. When ICA is performed on fMRI data, non-Gaussian distributions are assumed on the sources $\mathbf{s}$. For a given fMRI data set $\mathbf{x}$, ICA estimates the unknown $\mathbf{A}$ and $\mathbf{s}$ simultaneously. The task of the ICA algorithm is to determine a weight matrix $\mathbf{W}$ such that $\mathbf{y} = \mathbf{W}\mathbf{x} = \mathbf{P}\mathbf{D}\mathbf{s}$, where $\mathbf{P}$, a permutation matrix, represents the permutation ambiguity and $\mathbf{D}$, a diagonal matrix, represents the complex-valued scaling ambiguity of ICA. As we can observe, $\mathbf{x} = \mathbf{A}\mathbf{s} = \mathbf{A}'\mathbf{s}'$ where $\mathbf{A}' = \mathbf{A}\mathbf{D}^{-1}$ and $\mathbf{s}' = \mathbf{D}\mathbf{s}$. Therefore there are infinitely many solutions of $\mathbf{A}$ and $\mathbf{s}$ since $\mathbf{D}$ is an arbitrary diagonal phase rotation matrix. That is, for general ICA problem, it is impossible to recover the original scale of the sources, which in the complex case includes a magnitude and a phase term.

Besides using a fixed nonlinearity, many complex ICA algorithms also assume that the sources are circular, such as the complex fastICA algorithm and the circular Infomax algorithm. However, the preprocessing of the phase data is likely to lead the fMRI data to be noncircular [28]. In Figure 1, we show an example of the scatter plot of fMRI data averaging

over the model time course which clearly demonstrates the noncircular nature of fMRI data. Therefore, it is desirable to relax the assumption of circularity so that the estimated source distributions are not constrained. In addition, as we have discussed in the introduction, it is desirable to match the source distributions during the ICA estimation process. In the next section, we introduce such adaptive ICA algorithms.

B. Adaptive complex ICA algorithms

A-CMN algorithm—The adaptable complex maximization of non-Gaussianity (A-CMN) algorithm [21], [23] uses a complex generalized Gaussian distribution (CGGD) model for the underlying sources. The form of the CGGD density model is given as

$$p_{\mathbf{s}}(\bar{\mathbf{s}}) = \frac{\beta(c)}{\sqrt{|\mathbf{C}|}} e^{-[\eta(c)\bar{\mathbf{s}}^H \mathbf{C}^{-1} \bar{\mathbf{s}}]^c}$$

where $\bar{\mathbf{s}} = \begin{bmatrix} \mathbf{s} \\ \mathbf{s}^* \end{bmatrix}$ is the augmented source vector, $\mathbf{C}$ is the covariance matrix of $\bar{\mathbf{s}}$,

$\eta(c) = \frac{\Gamma(2/c)}{2\Gamma(1/c)}$, $\beta(c) = \frac{c\Gamma(2/c)}{\pi\Gamma(1/c)^2}$ and c is the shape parameter. Using this CGGD model, both sub-Gaussian ($c > 1$) and super-Gaussian ($0 < c < 1$) sources can be modeled through the flexible shape parameter and we can model both circular and noncircular sources through the estimation of the covariance matrix. The A-CMN cost function is defined as

$$J(\mathbf{y}) = E\{(\mathbf{y}\mathbf{y}^*)^c\}$$

where $\mathbf{y}$ is the estimated source in a deflationary mode. The shape parameter c and covariance matrix $\mathbf{C}$ are estimated on-line using a maximum likelihood estimator. This on-line adaptation, modifies the cost function to match each source distribution and hence improves the overall performance. When the covariance matrix $\mathbf{C}$ is fixed to be an identity matrix and only the shape parameter c is estimated, we obtain a circular version of the A-CMN algorithm.

The other algorithm in the MN group is complex fastICA algorithm. Complex fastICA algorithm is closely related to the A-CMN algorithm and also maximizes the non-Gaussianity. The cost function used by complex fastICA algorithm is defined as $J(\mathbf{y}) = E\{G(|\mathbf{y}|^2)\}$ where G is some fixed nonlinearity. A typical choice of G is $G(x) = \log(x + a)$ where a is some small constant.

Complex ICA-EBM algorithm—The complex ICA-EBM algorithm uses the minimization of the mutual information principle, which is equivalent to ML, to perform source separation. The cost function can be written as [24]

$$I(\mathbf{y}) = \sum_{k=1}^n H(\mathbf{y}_k) - 2 \log|\det(\mathbf{W})| - H(\mathbf{x})$$

where $H(\mathbf{y}_k)$ is the entropy of the k th spatial map. Instead of estimating $H(\mathbf{y}_k)$ directly, complex ICA-EBM estimate the tightest bound of the entropy by assuming that the density of the sources is either weighted linear combinations or elliptical distribution. Complex ICA-EBM can obtain a reliable estimate of the bound of entropy by solving for the maximum entropy distribution that maximize the entropy under certain constraints. The

associated maximum entropy distribution includes many bivariate distributions, such as Gaussian, uniform, (double) exponential, Student t , and GGD. Let us define two density forms as

$$\begin{aligned} p(u) &= A \exp[-au^2 - bu - cG_m(u)], \\ q(|u|) &= A \exp[-a|u|^2 - cG(|u|)] \end{aligned}$$

where G_m is one of a set of pre-determined measure functions and the parameters A , a , b , and c are solved by using some normalization constraint, the associated maximum entropy distribution of s could be $Ap(s_r)p(s_i)$ or $Aq(|s|)$ for two different entropy bounds. Four function forms of G , including unbounded fourth order symmetric and bounded second-order asymmetric, are considered for the first entropy bound where the density is weighted linear combinations. Two function forms of G , including fourth order symmetric and first order asymmetric, are considered for the second entropy bound where the sources are elliptical distributions. Among all the entropy estimates, only the minimum one is used as the final estimate of the entropy.

The other algorithm in the ML group is circular Infomax algorithm. Circular Infomax algorithm uses the ML principle, which is equivalent to the principle of minimization of mutual information used by complex ICA-EBM. As shown in [20], the nonlinear score function of circular Infomax used to associate the super-Gaussian densities takes the fixed form

$$\psi(y) = \text{sign}(y) \frac{1 - \exp(-|y|)}{1 + \exp(-|y|)}$$

where only the magnitude of the data has been considered and all the phase information has been discarded. Therefore, the density model assumed by complex ICA-EBM is more flexible than circular Infomax.

It should be noted that the MN group, *i.e.*, the A-CMN algorithm and complex fastICA algorithm, both impose the unitary constraint on W . In contrast, the ML group, *i.e.*, the ICA-EBM and circular Infomax algorithm, do not impose such a constraint. As expected, constraining the set of solutions, *i.e.*, constraining the demixing matrix affects the performance, and the ML group will have better performance than those in the MN group when all other variables are the same as shown in [15] with simulated data. However, constraining the demixing matrix to be unitary allows for the implementation of a deflationary mode [17] which provides flexibility in density matching. In addition, in deflationary mode, one can only extract sources of interest and can easily include constraints in this class of algorithms. Hence both classes of algorithms have certain advantages.

III. Materials and methods

A. fMRI data

The dataset used in this paper was from 16 subjects performing a finger-tapping motor task while receiving auditory instructions. IRB-approved informed consent at the University of New Mexico was obtained from all the participants. The paradigm has a block design with alternating periods of 30 seconds on (finger tapping) and 30 seconds off (rest). The experiments were performed on a 3T Siemens TIM TRIO system with a 12-channel radio frequency (RF) coil. The fMRI experiment used a standard Siemens gradient-echo EPI sequence modified to store real and imaginary data separately. The following parameters

were used: Field-of-View (FOV)= 24cm, Slice Thickness= 3.5mm, Slice Gap= 1mm, Number of slices= 32, Matrix size= 64×64 , TE= 29ms, TR= 2s, flip angle 70 deg. We collected 15 whole head fMRI images during each 'ON' or 'OFF' period. Data collection was prefaced by a 12s rest period that was collected to allow for T1 effects to stabilize. The preprocessing of the data was similar to the procedures used in [29]. The magnitude and phase images were written out as 4D nifti files using a custom reconstruction program on the scanner. Preprocessing of the data was done using the SPM5 software package[30]. Magnitude data was co-registered to compensate for movements in the fMRI time series images. Images were then spatially normalized into the standard Montreal Neurological Institute space. Following spatial normalization, the data (real and imaginary images) were slightly sub-sampled, resulting in $53 \times 63 \times 46$ voxels. Motion correction and spatial normalization parameters were computed from the magnitude data and then applied to the phase data. Then the real and imaginary images were both spatially smoothed with a $10 \times 10 \times 10 \text{ mm}^3$ full width at half-maximum Gaussian kernel. It is important to note that, in the following analysis, the ICA estimations were performed in a complex form, not in a magnitude form and a phase form separately. The magnitude images and phase images shown in the following sections are just for displaying purpose.

B. Denoising

First, we applied the quality map phase de-noising (QPMd) method presented in [31] to generate a mask to minimize the effects of noise in the phase of the fMRI data. The binary mask generated by QPMd identifies the good quality voxels in each fMRI slice for every fMRI data volume obtained at every time point. Then the fMRI data were multiplied by the mask and smoothed for the real part and imaginary part separately. The original dimension of the fMRI data for each subject is 153594×165 where the number of voxels in all the slices is 153594, and the number of time points is 165. After masking and smoothing, the dimension decreases to approximately 50000×165 . Then we performed 10 ICA estimations for each subject and calculated the averages across 16 subjects in order to provide a fair comparison.

C. Data dimension reduction

Since the original data dimension 165 is relatively high for the complex ICA algorithms to process, data dimension reduction is required for further ICA processing. A common approach of data dimension reduction used for feature extraction is principal component analysis (PCA) [17], which is linear. For a real-valued $n \times 1$ random vector $\mathbf{x}$, PCA can be used to reduce the dimension of $\mathbf{x}$ from n to p by using $\mathbf{z} = \mathbf{M}^T \mathbf{x}$ where the matrix $\mathbf{M}$ is composed of p eigenvectors of the covariance matrix of $\mathbf{x}$ corresponding to p largest eigenvalues. Here we consider the extension of PCA to the complex domain by considering the covariance matrix $\mathbf{C}$ of $\mathbf{x}$ where $\mathbf{C} = E\{\mathbf{x}\mathbf{x}^H\}$.

To determine the optimal order p , the most common approach has been using information theoretic criterion (ITC) as in [32] where the two underlying assumptions are that the data is assumed to be multivariate Gaussian distributed and the noise within the data is white and uncorrelated with the signal of interest. Then any ITC such as Akaike's information criterion (AIC), Kullback-Leibler information criterion (KIC) and the minimum description length (MDL) can be used for the complex-valued fMRI data [11], [33]. In the experiments, we have tried different orders and found that 26 was a reasonable number that provided a good tradeoff between preserving much of the information in the data while reducing the size of the results, thus making the analysis and interpretation less intensive. As observed in the experiments, the estimation results were not sensitive to small changes in the order.

Therefore, we applied complex PCA to reduce the dimension of time points from 165 to 26 for each subject, where the PCA weighting matrix $\mathbf{M}$ is composed of the first 26 significant eigenvectors of the covariance matrix of the data. Next we applied the four complex ICA algorithms, A-CMN, complex fastICA, complex ICA-EBM and circular Infomax to the data set to find a demixing matrix $\mathbf{W}$ and the estimated spatial map $\hat{\mathbf{s}} = \mathbf{W}\mathbf{z}$ for each algorithm. The resultant mixing matrix $\mathbf{B}$ estimated by ICA algorithms was used to back-reconstruct the time courses for each subject, which is $(\mathbf{M}^T)^+\mathbf{B}$ with the superscript $+$ denoting the pseudoinverse.

D. Consistency analysis using complex ICASSO

For each subject, we performed 10 ICA estimations for each algorithm with different initial conditions. Since the ICA estimates for iterative algorithms change for different runs, the issue of consistency of the estimates needs to be addressed. For the real-valued case, a toolbox, ICASSO has been developed to evaluate similarities among multiple estimates of the fastICA algorithm [1]. ICASSO collects the components estimated from all runs and then matches components across runs by clustering components based on the absolute value of the correlation between squared source estimates. We modified the toolbox and made it available for the complex case as well. The similarity measure used in real-valued ICASSO for clustering is based on the correlation matrix $\mathbf{R}$ where $\mathbf{R} = E\{\mathbf{ss}^T\} = \mathbf{WCW}^T$. For the complex case, since the source spatial maps are complex valued, we can consider the correlation matrix between the magnitude spatial map or the phase spatial map. Since the resultant magnitude spatial map was more informative in the experiments, we used the correlations between the magnitude of the estimates to perform the clustering. The clustering of the estimated components is expected to yield information on the reliability of the estimation.

If the ICA algorithms give consistent results over different runs, the output of ICASSO will have 26 compact clusters for each subject and each cluster will have 10 members. More importantly, complex ICASSO is used for determining the centrotpe of 10 runs that has the highest correlation with other members within one cluster. It should be noted that this centrotpe is not the average of all the members within one cluster. For each subject, there are 26 centrotpe estimates for each algorithm.

E. Bivariate thresholding with Mahalanobis distance metric for single subject

For real-valued ICA on the fMRI data, after the ICA estimation, the images were converted to Z-score and thresholded at $|Z| > a$ where a is a predetermined positive constant. It is generally understood that the ICA Z-scores are used for descriptive purposes and do not have statistical interpretation similar to t-test results [2]. However, this Z-score value represents the distance between the value of the original pixel and the zero mean in units of the standard deviation of the image.

For the complex case, we can have three different thresholding schemes for each complex-valued voxel: thresholding only the magnitude of the data, thresholding only the phase of the data, and thresholding both the magnitude and phase of the data jointly. The results of the experiments showed that the last one, bivariate thresholding, provided the best performance. We define a Mahalanobis Z-score using the Mahalanobis distance metric for the bivariate thresholding. The metric is given by

$$D_M(s) = \bar{s}^T \mathbf{C}_s^{-1} \bar{s} \quad (1)$$

where $\bar{s} = \begin{bmatrix} s_r \\ s_i \end{bmatrix}$ and $s = s_r + js_i$ is a complex-valued voxel in the fMRI spatial map $\mathbf{s}$ where we assume the spatial map is centralized and $\mathbf{C}_{\bar{s}}$ is the 2-by-2 covariance matrix of $\bar{s}$. For a 1-by- v estimated complex-valued spatial map $\mathbf{s}$ or equivalently a 2-by- w real-valued spatial map, the 2×2 covariance matrix $\mathbf{C}_{\bar{s}}$ is calculated using all the voxels and hence there is one Mahalanobis Z-score for each voxel. In the experiments, after the ICA stage, the images were converted to Mahalanobis Z-score and thresholded at $D_M(\mathbf{s}) > a$ where a is a predetermined positive constant. Mahalanobis Z-score is preferable to the univariate Z-score because it takes the bivariate nature and, more importantly, the noncircular nature of the complex data into account. This Mahalanobis Z-score represents the distance between the value of each voxel and the zero mean in units of the standard deviation of the image. It is computed for a single subject and not across multiple subjects. After the Mahalanobis thresholding, we were left with three set of images: the Mahalanobis image (*i.e.*, the Mahalanobis Z-scores), the real image and imaginary image (or the magnitude image and phase image). Of course the Mahalanobis image can be used on its own [34]. However, in this paper, we are using the Mahalanobis metric as a mask to determine the thresholded magnitude images and phase images.

F. Construction of bivariate t-map for multiple subjects

For a group of subjects, we can define a bivariate t-map using the Hotelling T^2 -Test. The Hotelling T^2 -Test is defined as [35]

$$T^2(\mathbf{s}) = N \langle \bar{\mathbf{s}} \rangle^T \mathbf{C}_{\bar{\mathbf{s}}}^{-1} \langle \bar{\mathbf{s}} \rangle \quad (2)$$

where N is the number of realizations of random vector $\mathbf{s}$. For Hotelling T^2 -Test, $\langle \bar{\mathbf{s}} \rangle$ in (2) represents the sample mean vector of a set of realizations from multivariate Gaussian distributed $\mathbf{s}$. For a group of subjects, we can calculate the mean image $\langle \bar{\mathbf{s}} \rangle$ of 16 subjects, where $\langle \bar{\mathbf{s}} \rangle$ represents the sample mean vector of $16 \times 2 \times 1$ vectors for each voxel. The covariance matrix $\mathbf{C}_{\bar{\mathbf{s}}}$ is also defined with respect to the $16 \times 2 \times 1$ vectors, where each 2×1 vector is treated as a realization of a 2×1 Gaussian random vector. Therefore we can construct an image of size v using (2) and the value of each pixel represents a Hotelling T^2 value, where we have assumed that the $16 \times 2 \times 1$ vectors for each voxel are the realizations of a multivariate Gaussian distribution. Such an image might be called a group bivariate t-map. It should be noted that 16 may be too small as a sample size and the multivariate Gaussian distribution across different subjects is also assumed for simplicity.

G. Receiver operating characteristic (ROC) analysis

A receiver operating characteristic curve is a graphical plot of true positive rate (TP) vs. false positive rate (FP) for a binary classifier system as its discrimination threshold is varied. Similar to the ROC analysis in [36], here we define that the probability of true positives (TP) represents the normalized number of voxels that survive the threshold and that fall inside the motor cortex mask. Similarly, the probability of false positive (FP) is defined as the normalized number of voxels that pass the various thresholds and that fall outside the motor cortex mask. Each quantity is normalized with respect to the number of all the voxels in the mask or the number of voxels outside the mask. Comparison of the ROC curves for different algorithms is done by comparing their respective area under the curve (AUC). Better performance, as measured by the larger AUC, indicates overall higher sensitivity and specificity at the various thresholds used in the ROC curves. Higher sensitivity means that a higher number of voxels are identified as active. Similarly, an increase in specificity indicates that a large number of the additional identified voxels are located inside the motor cortex area mask. However we noted that such ROC curves should not be interpreted as the

traditional ones since we don't have the knowledge of the density function of the estimated components with respect to the true or null hypothesis.

IV. Results

To demonstrate the performance of the algorithms, two components were considered for this fMRI data set: a task-related motor component and a non-task-related temporal component. The motor component is defined as the right motor activation area responding to the task, while the temporal component is a commonly identified resting network occurring in the bilateral temporal lobe. These two components were consistently estimated by all four algorithms. To compare their performance, we consider the following measures:

1. the number of activated voxels for both components;
2. the maximum activation value for both components;
3. the correlation between the estimated time course and a model time course for the motor (task-related) component, where the model time course was generated by convolving a temporal model of the on-off task with a canonical hemodynamic response function.

In addition to the above measures, we also show some typical results of complex ICASSO and the bivariate t-map, which show significant activations across 16 subjects for each algorithm.

A. Comparison of numbers of activated voxels and maximum activation values in the region of interest

To count the number of activated voxels for an activation map, we used WFU Pickatlas [37] to create the mask by selecting different areas of the brain. We created two masks for the right motor component and the bilateral temporal component. These masks included neuronal areas that were expected to be activated during the task as well as those in the temporal lobe. The areas included in each of the two masks are as follows: 1) right motor task-related: Brodmann areas 1, 2, 3, 4 and 6. 2) bilateral temporal lobe: Brodmann areas 20, 21, 22, 37, 38 and 42.

After the identification of the right motor and the temporal component, the resultant spatial maps were thresholded at $D_M(s) > 2$. Then we counted the number of voxels in the estimated components that overlapped with the corresponding masks for each algorithm and each subject. The average number of activated voxels for the right motor component with all the subjects are displayed in Table I. As shown in the table, the number of activated voxels estimated by the adaptive ICA algorithms were larger than the number estimated by non-adaptive ones within each group. The number of activated voxels was 728 for A-CMN and 622 for complex fastICA for the MN group. The number of activated voxels was around 761 for complex ICA-EBM and 667 for circular Infomax for the ML group. Therefore, an average of 14 to 22 percent more activated voxels were estimated for the right motor component when using adaptive ICA algorithms. This increase in activated voxels for the adaptive ICA algorithms was also found to be insensitive to the threshold used.

The results for the temporal component are also displayed in Table I. As shown in the table, the number of activated voxels estimated by the adaptive ICA algorithms was larger than the number estimated by non-adaptive ones. However, the advantage of using adaptive nonlinearity was not as obvious as it was for the motor component. For the MN group, the number of activated voxels was 1855 for A-CMN and 1782 for complex fastICA. For the ML group, the number of activated voxels was around 1912 for Complex ICA-EBM and

1829 for circular Infomax. Therefore, an average of 4 to 7 percent more voxels contributing to the temporal component have been estimated by the adaptive ICA algorithms.

The maximum $D_M(s)$ values are displayed in Table II. For this case, the A-CMN algorithm achieved the best performance with the circular Infomax algorithm closely behind. The complex fastICA algorithm obtained the lowest activation value for both of the two components. It was noted that complex ICA-EBM always had a lower maximum $D_M(s)$ than the circular Infomax algorithm. However, all the $D_M(s)$ values were quite close when we consider the standard deviation.

B. Comparison of correlation values between the estimated time courses and model time course

The correlations between the estimated time course and the model time course are displayed in Table III. As shown in the table, the phase timecourses estimated by the adaptive ICA algorithms showed a higher correlation with the model time course. We had an average of 16 percent higher estimation on the phase time course for the motor component. An example of time courses estimated by A-CMN algorithm along with the complete spatial map is shown in Figure 2(a), where both the magnitude time course and phase time course have correlation values as high as 0.83 with the model time course. As a comparison, one sample estimate by complex fastICA algorithm is shown in Figure 2(b). The two figures together are representative examples which demonstrate that the A-CMN algorithm results in timecourses that are more highly correlated with the model versus complex fastICA.

C. Results of consistency analysis

For the discussion of consistency of all the four algorithms, one typical result is given in Figure 3 where the result is from complex ICA-EBM for one subject. In the figure, each estimate is plotted as a point on the display, and a convex hull (the grey lines) bounds the estimates belonging to the same cluster. The points are connected with lines whose thickness/color represent the similarities between them. Red lines indicate that there are high correlation values between the estimates. The ICASSO results of other three algorithms were all similar to Figure 3 and there was no significant difference between the consistency results of adaptive ICA algorithms and non-adaptive ones. In the figure, there are 26 clusters. For the ideal case, *i.e.*, the results of 10 runs are identical, we will expect that there are only 26 dots in the figure. As observed from the figure, cluster one is quite compact, which indicates high consistency of performance of ICA-EBM algorithm for estimating this component. Actually this cluster represents the right motor component. Cluster three is also compact and it represents the temporal component. Other clusters showing large variation, such as cluster 26, represent noise within the data.

D. Results of bivariate t-maps and difference t-maps

As observed from previous results, there were some variations among different subjects. Using the Hotelling T^2 statistic defined in (2), we can construct a bivariate t-map for each algorithm to check for voxels that are significant active across the 16 subjects. However, the fMRI images estimated by the ICA algorithms have phase rotation ambiguity and we cannot simply apply the T^2 -test. The phase ambiguity, as explained in Section II-A, is due to the ICA model $\mathbf{x} = \mathbf{A}\mathbf{s}$. As we can observe, $\mathbf{x} = \mathbf{A}\mathbf{s} = \mathbf{A}'\mathbf{s}'$ where $\mathbf{A}' = \mathbf{A}\mathbf{D}^{-1}$, $\mathbf{s}' = \mathbf{D}\mathbf{s}$ and $\mathbf{D}$ represents a diagonal matrix whose entries are complex, and hence include a magnitude and phase part. Therefore there are infinitely many solutions of $\mathbf{A}$ and $\mathbf{s}$ since $\mathbf{D}$ is an arbitrary diagonal phase rotation matrix. That is, for general ICA problem, it is impossible to recover the original scale of the sources, which in the complex case includes a magnitude and a phase term.

Therefore, we need to perform the phase rotation on each estimate to ensure that the largest magnitude of the estimate is on the real axis. Here we have used the prior knowledge of fMRI data where the real part of the signal of interest is dominant over the imaginary part [38]. The value for each voxel in the bivariate t-map tells us how the voxel values are distributed across different subjects. High T^2 -values in such figures might be regarded as an index indicating that the voxel values are probably high for all the subjects at that specific pixel and with probably lower variations across different subjects. We observed that the bivariate t-maps looked visually similar for all the four algorithms. However, in terms of hypothesis testing, the results were not the same. For hypothesis testing, we are testing $H_0 : \mu_s = 0$ versus $H_1 : \mu_s \neq 0$ where μ_s is the mean spatial map of 16 subjects. We accept H_0 if $T^2 < T^2_{\alpha, d, N-1}$ where α is the probability of Type I error (accept H_1 when H_0 is in fact true), $d = 2$ is the dimension of random vector, and $N - 1 = 15$ is the degrees of freedom. For instance, $T^2_{0.05, 2, 15} = 8.01$. In Table IV, we show the number of voxels in the mean spatial map that have $T^2 > T^2_{0.05, 2, 15}$. As observed from Table IV, the A-CMN algorithm detected larger number of voxels than complex fastICA and the ICA-EBM algorithm detected larger number of voxels than circular Infomax.

We can also compute difference t-maps for the comparison of algorithms. Given any two sets of estimated spatial maps with voxels $\mathbf{X}_{ijk}$ and $\mathbf{Y}_{ijk}$, where $\mathbf{X}_{ijk}$ denotes the k th voxel in the j th component of the i th subject, each voxel of 16 difference images are calculated as $\mathbf{D}_{ijk} = \mathbf{X}_{ijk} - \mathbf{Y}_{ijk}$. The difference images of 16 subjects were calculated first, then a T^2 -test was performed. The resulting difference t-maps of A-CMN and complex fastICA is shown in Figure 4 (a) and the resulting difference t-maps of ICA-EBM and circular Infomax is shown in Figure 4 (b). Since the bivariate t-maps measure the relative difference in the activation values across subjects, we observed that A-CMN algorithm showed significantly higher activation within the motor area for the MN group. Similarly, ICA-EBM showed significantly higher activation within the motor area for the ML group as well.

E. Results of ROC analysis

In Figure 5, we show the results of ROC curves for the two classes of algorithms. As observed from the figure, within each class, the adaptive ICA algorithms had larger areas under the curve (AUC), which implied higher sensitivity and specificity. As shown in the figure, for a given low false positive rate, the results of adaptive ICA algorithms had achieved a higher true positive rate. We also observed that the difference between the AUC of A-CMN and complex fastICA was larger than the difference between the AUC of ICA-EBM and circular Infomax. The difference between two ROC curves within each group is small because the numbers were normalized by the number of voxels in the mask, which is a large quantity.

Hence, based on all the performance indices, we note the importance of adaptively matching the underlying fMRI source model for improving the performance of the complex ICA of fMRI data.

V. Discussion

ICA is one of the most fruitful data-driven methods for the study of fMRI data and the use of phase information in addition to the typically used magnitude promises to provide new insights for the analysis. In this paper, we showed that, by adaptively matching the density model and relaxing the circularity assumption, more activated voxels in the task-related region were detected and the correlation between the estimated time course and model time course was improved especially for the phase. In particular, we demonstrated that the

adaptive ICA algorithms, A-CMN and ICA-EBM, provided a more flexible way to extract the intrinsic features of the noisy fMRI data. In addition, we addressed important issues when performing the ICA of fMRI data in its native, complex form, such as denoising, thresholding of the results for visualization, and construction of bivariate t-map for multiple subjects.

Complex ICA-EBM algorithm achieves the best performance in the simulation studies [24] and A-CMN algorithm provides competitive performance and faster convergence property [21]. The Matlab code for the algorithms can be found at <http://mlsp.umbc.edu/resources.html>.

The results of using adaptive ICA algorithms for complex fMRI data show that this is a promising direction and there are some interesting future directions worth investigating:

1. in the preprocessing, we have used complex PCA to reduce the dimension of the data where only the covariance matrix is used. Other dimension reduction techniques, such as noncircular PCA and kernel PCA, may better represent the original complex valued fMRI data in a lower dimension subspace.
2. although the two adaptive ICA algorithms provide better performance than non-adaptive ones, it should be noted that the computation cost has increased. For one subject and one run, it took about three to seven minutes to run the complex fastICA algorithm. The computational cost of A-CMN algorithm or circular Infomax algorithm was similar to that of complex fastICA. For ICA-EBM, it took about ten to 20 minutes. Therefore, it is desirable to investigate how to make adaptive ICA algorithms more efficient especially for high-dimensional data.
3. in future work, we plan to extend our approach beyond single-subject ICA to group ICA.

References

1. Himberg, J.; Hyvarinen, A. Proc NNSP. Toulouse, France: 2003. ICASSO: software for investigating the reliability of ICA estimates by clustering and visualization.
2. McKeown MJ, Makeig S, Brown GG, Jung T, Kindermann SS, Bell AJ, Sejnowski TJ. Analysis of fMRI data by blind separation into independent spatial components. *Human Brain Mapping*. 1998; 6:160–188. [PubMed: 9673671]
3. Calhoun, VD.; Adali, T.; Hansen, LK.; Larsen, J.; Pekar, JJ. Proc ICA. Nara, Japan: Ica of functional MRI data: an overview.
4. McKeown MJ, Hansen LK, Sejnowski TJ. Independent component analysis of functional MRI: what is signal and what is noise? *Curr Opin Neurobiol*. 2003; 13:620–629. [PubMed: 14630228]
5. Hoogenraad FGC, Reichenbach JR, Haacke EM, Lai S, Kuppusamy K, Sprenger M. In vivo measurement of changes in venous blood-oxygenation with high resolution functional MRI at 0.95 Tesla by measuring changes in susceptibility and velocity. *Magn Reson Med*. 1998; 39:97–107. [PubMed: 9438443]
6. Duyn JH, van Gelderen P, Li TQ, de Zwart JA, Koretsky AP, Fukunaga M. High-field MRI of brain cortical substructure based on signal phase. *Proc Natl Acad Sci USA*. 2007; 104(28):11 796–801.
7. Logan BR, Rowe DB. A complex way to compute fMRI activation. *NeuroImage*. 2004; 23:1078–92. [PubMed: 15528108]
8. Menon R. Postacquisition suppression of large-vessel BOLD signals in high-resolution fMRI. *Magn Reson Med*. 2002; 47:1–9. [PubMed: 11754436]
9. Calhoun VD, Adali T, Pearlson GD, van Zijl PC, Pekar JJ. Independent component analysis of fMRI data in the complex domain. *Magn Reson Med*. 2002; 48:180–192. [PubMed: 12111945]

10. Calhoun, VD.; Adali, T.; Li, YO. Proc IEEE Int Symp Bio Imag (ISBI). Arlington, VA: 2004. Independent component analysis of complex-valued functional MRI data by complex nonlinearities.
11. Xiong, W.; Li, YO.; Li, H.; Adali, T. Proc ICASSP. Las Vegas, Nevada: April. 2008 On ICA of complex-valued fMRI: Advantages and order selection.
12. Ghiglia, DC.; Pritt, MD. Two-dimensional phase unwrapping. New York: Wiley; 1998.
13. Herráez MA, Gdeisat MA, Burton DR, Lalor MJ. Robust, fast, and effective two dimensional automatic phase unwrapping algorithm based on image decomposition. Applied Optics. 2002; 41:7445–7455. [PubMed: 12502302]
14. Rodriguez P, Correa N, Eichele T, Calhoun V, Adali T. Quality map thresholding for de-noising of complex-valued fMRI data and its application to ICA of fMRI. J Sign Process Syst. 2010:1–12.
15. Adali T, Li H, Novey M, Cardoso J. Complex ICA using nonlinear functions. IEEE Trans Signal Processing. 2008; 56:4536–4544.
16. Hyvärinen A. Fast and robust fixed-point algorithms for independent component analysis. IEEE Trans Neural Networks. 1999; 10(3):626–634.
17. Hyvärinen, A.; Karhunen, J.; Oja, E. Independent Component Analysis. Wiley; 2001.
18. Pham D, Garat P. Blind separation of mixtures of independent sources through a quasi maximum likelihood approach. IEEE Trans Signal Processing. 1997; 45(7):1712–1725.
19. Bingham E, Hyvarinen A. A fast fixed-point algorithm for independent component analysis of complex valued signals. Int J Neural Systems. 2000; 10:1–8.
20. Anemüller J, Sejnowski TJ, Makeig S. Complex independent component analysis of frequency-domain electroencephalographic data. Neural Networks. 2003; 16:1311–1323. [PubMed: 14622887]
21. Novey, M.; Adali, T. Proc MLSP. Maynooth, Ireland: Sept. 2006 Adaptable nonlinearity for complex maximization of nongaussianity and a fixed-point algorithm.
22. Novey M, Adali T. On extending the complex FastICA algorithm to noncircular sources. IEEE Trans Signal Processing. 2008; 56(5):2148–2154.
23. Novey M, Adali T, Roy A. A complex generalized Gaussian distribution- characterization, generation, and estimation. IEEE Trans Signal Processing. 2010; 58(3):1427–1433.
24. Li X-L, Adali T. Complex independent component analysis by entropy bound minimization. IEEE Trans Circuits and Systems I. 2010; 57:1417–1430.
25. Li, H.; Adali, T.; Correa, N.; Rodriguez, P.; Calhoun, V. Proc ICASSP. Dallas: March. 2010 Flexible complex ICA of fMRI data.
26. Gudbjartsson H, Patz S. The Rician distribution of noisy MRI data. Magn Reson Med. 1995; 34:910–914. [PubMed: 8598820]
27. Schreier, P.; Scharf, L. Statistical Signal Processing of Complex-Valued Data: The Theory of Improper and Non-circular signals. Cambridge, UK: University Press; 2010.
28. Adali T, Calhoun VD. Complex ICA of brain imaging data. IEEE Signal Processing Magazine. 2007; 24:136–139.
29. Arja S, Feng Z, Chen Z, Caprihan A, Kiehl KA, Adali T, Calhoun VD. Changes in fMRI magnitude data and phase data observed in block-design and event-related tasks. NeuroImage. 2010; 49:3149–60. [PubMed: 19900561]
30. SPM. <http://www.fil.ion.ucl.ac.uk/spm>
31. Rodriguez P, Correa N, Adali T, Calhoun V. Quality map thresholding for de-noising of complex-valued fMRI data and its application to ICA of fMRI. Proc MLSP. Sep.2009
32. Wax M, Kailath T. Detection of signals by information theoretic criteria. IEEE Trans Acoust, Speech, and Signal Processing. 1985; 33:387–392.
33. Li Y-O, Adali T, Calhoun VD. Estimating the number of independent components for fMRI data. Human Brain Mapping. 2007; 28:1251–66. [PubMed: 17274023]
34. Rodriguez P, Adali T, Li H, Correa N, Calhoun V. Phase correction and denoising for ICA of complex fMRI data. Proc ICASSP. March.2010
35. Rencher, AC. Methods of Multivariate Analysis. Wiley-Interscience; 2002.

36. Skudlarski P, Constable RT, Gore JC. ROC analysis of statistical methods used in functional MRI: Individual subjects. *NeuroImage*. 1999; 9:311–329. [PubMed: 10075901]
37. Maldjian J, Laurienti P, Kraft R, Burdette J. An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data set. *NeuroImage*. 2003; 19:1233–1239. [PubMed: 12880848]
38. Hardy E, Hoferer J, Mertens D, Kasper G. Automated phase correction via maximization of the real signal. *Magnetic Resonance Imaging*. 2009; 27:393–400. [PubMed: 18760554]

Fig. 1.

An example of the scatter plot of fMRI data averaging over the model time course which clearly demonstrates the noncircular nature of fMRI data; (left) The model time course; (right) Averaged fMRI signal weighted by the model time course.

Fig. 2.

An example to show that the A-CMN algorithm results in time courses that are more highly correlated with the model versus complex fastICA; (a) The magnitude time course and phase time course for the right motor component estimated by A-CMN (b) The magnitude time course and phase time course for the right motor component estimated by complex fastICA algorithm; Both figures are from single subject and are thresholded at $D_M(s) > 4$; The dotted yellow line represents the model time course and the solid purple lines are the estimated time courses.

Fig. 3.

An example to show the typical consistency results of complex ICA-EBM algorithm for one subject and 10 runs; The ICASSO results of other three algorithms are all similar to this figure and there is no significant difference among them; There are 26 estimated components for each run and each component is plotted as a point; Convex hull (the grey lines) bounds the estimates belonging to the same cluster and the thickness/color of lines represent the similarities between the estimates.

Fig. 4. Results of difference bivariate t-maps to show that the adaptive ICA algorithms have significantly higher activation within the motor area for each class; (a) Difference bivariate t-map for the motor component of A-CMN and complex fastICA; (b) Difference bivariate t-map for the motor component of ICA-EBM and circular Infomax; The results are thresholded at $\alpha = 0.19$ ($t > 4$).

Fig. 5.

Results to show that the adaptive ICA algorithms had achieved a higher true positive rate than the non-adaptive ones for a given false positive rate, which implies higher sensitivity and specificity; The probability of true positives (TP) in the generated ROC curves indicates the number of voxels that survive the threshold and that fall inside the motor cortex mask; The probability of false positive (FP) indicates the number of voxels that pass the various thresholds and that fall outside the motor cortex mask; The difference between two ROC curves is small because the numbers were normalized by the number of voxels in the mask, which is a large quantity.

TABLE I

Results to show that the number of activated voxels estimated by the adaptive ICA algorithms are larger than the number estimated by non-adaptive ones within each group; The number in each cell (mean $\pm$ std) represents the number of voxels in activation area of components that overlap with the corresponding mask; The mean and standard deviation are evaluated across 16 subjects.

	MN-ICA		ML-ICA	
Right Motor	A-CMN	CfastICA	ICA-EBM	C-Infomax
No of voxels	728 $\pm$ 275	622 $\pm$ 310	761 $\pm$ 226	667 $\pm$ 238
Temporal	A-CMN	CFastICA	ICA-EBM	C-Infomax
No of voxels	1855 $\pm$ 352	1782 $\pm$ 369	1912 $\pm$ 393	1829 $\pm$ 440

TABLE II

Results to show that A-CMN algorithm achieves the best performance in terms of maximum $D_M(s)$ of the estimated components and complex ica-ebm has a slightly lower maximum $D_M(s)$ than the circular infomax algorithm; The number in each cell (mean $\pm$ std) represents the maximum $D_M(s)$; The mean and standard deviation are evaluated across 16 subjects.

	MN-ICA		ML-ICA	
Right Motor	A-CMN	CfastICA	ICA-EBM	C-Infomax
Maximum $D_M(s)$	12 $\pm$ 2.4	11 $\pm$ 4	11.7 $\pm$ 2.5	12 $\pm$ 2.4
Temporal	A-CMN	CfastICA	ICA-EBM	C-Infomax
Maximum $D_M(s)$	11 $\pm$ 2.1	9 $\pm$ 2.6	9.5 $\pm$ 1.8	10 $\pm$ 2.3

TABLE III

Results to show that the phase time courses estimated by the adaptive ICA algorithms has higher correlation with the model time course than the non-adaptive ones within each group; The number in each cell (mean $\pm$ std) represents the correlation of the estimated time courses with model time courses; The mean and standard deviation are evaluated across 16 subjects.

	MN-ICA		ML-ICA	
Right Motor	A-CMN	CfastICA	ICA-EBM	C-Infomax
Magnitude	0.68 $\pm$ 0.24	0.64 $\pm$ 0.25	0.68 $\pm$ 0.2	0.68 $\pm$ 0.26
Right Motor	A-CMN	CfastICA	ICA-EBM	C-Infomax
Phase	0.69 $\pm$ 0.19	0.56 $\pm$ 0.24	0.66 $\pm$ 0.2	0.6 $\pm$ 0.27

TABLE IV

Results to show that the A-CMN algorithm detects larger number of voxels in the bivariate t-maps than complex fastICA and the ICA-EBM algorithm detects larger number of voxels than circular Infomax; The number of voxels are thresholded at

$$T^2 > T_{0.05,2,15}^2$$

	MN-ICA		ML-ICA	
Right Motor	A-CMN	CfastICA	ICA-EBM	C-Infomax
Number of voxels	2655	2184	2951	2761