COMPLEX ICA FOR FMRI ANALYSIS: PERFORMANCE OF SEVERAL APPROACHES

V. Calhoun[1,2] and T. Adali[3]

[1]Neuropsychiatry Research Center, Institute of Living, Hartford, CT 06106.
[2]Yale University School of Medicine, New Haven, CT 06520.
[3]Dept. of CSEE, University of Maryland, Baltimore County, Baltimore, MD 21250.

ABSTRACT

Independent component analysis (ICA) for separating complex-valued sources is needed for convolutive source-separation in the frequency domain, or for performing source separation on complex-valued data, such as functional magnetic resonance imaging data. Functional magnetic resonance imaging (fMRI) is a technique that produces complex-valued data; however the vast majority of fMRI analyses utilize only magnitude images. We compare the performance of the complex infomax algorithm that uses an analytic (and hence unbounded) nonlinearity with the traditional complex infomax approaches that employ bounded (and hence non-analytic) nonlinearities as well as with a cumulant-based approach. We compare the performances of these algorithms for processing both simulated and real fMRI data and show that the complex infomax using analytic nonlinearity has the ability to separate both sub- and super-Gaussian sources with a hyperbolic tangent nonlinearity. The complex infomax algorithm that uses analytic nonlinearity thus provides a potentially powerful method for exploratory analysis of fMRI data.

1. INTRODUCTION

Independent component analysis (ICA) has been successfully used for blind source separation (BSS), and to separate mixtures in a variety of applications. In the cocktail party problem, if the mixing is convolutive, then it is common to work in the frequency domain, thus requiring an algorithm for complex-valued data. Likewise, it is useful for analyzing complex-valued data, such as the functional magnetic resonance (fMRI) data. Virtually all fMRI studies analyze only the magnitude images from the MRI scanner even though the data are acquired as complex images. When performing an analysis of fMRI data using the complex-valued images, results demonstrate an increased ability to isolate the task-related functional changes [1,2] illustrating the importance of performing source separation directly on the acquired data that is complex-valued.

In the following, we first review a number of approaches for performing complex ICA, the JADE algorithm [3], two infomax approaches that use bounded nonlinearities [3], and then an infomax approach that uses an analytic and thus unbounded nonlinearity for infomax [4]. We demonstrate their application to fMRI analysis and discuss their performance differences, in particular when processing sub- and super-Gaussian sources.

2. BACKGROUND

We briefly review the four different approaches for performing complex ICA, three based on infomax and the fourth-order cumulant-based joint approximate diagonalization of eigenmatrices (JADE) algorithm.

2.1. Infomax

We use bold for matrix and vector quantities and underline to differentiate complex-valued variables and functions from those that are real-valued. In the complex ICA problem, the data, $\underline{\mathbf{x}}_i$, at sample i, is assumed to be a linear mixture of N statistically independent, complex-valued sources $\underline{\mathbf{s}}_i$:

$$\underline{\mathbf{x}}_i = \underline{\mathbf{A}}\underline{\mathbf{s}}_i \quad i = 1, \ldots, M \tag{1}$$

where $\underline{\mathbf{A}} \in \mathbb{C}^{N\times N}$, $\underline{\mathbf{x}}_i, \underline{\mathbf{s}}_i \in \mathbb{C}^N$, and M is the number of samples. In this derivation, we assume that the number of sources is equal to the number of sensors. The infomax algorithm proceeds by maximizing the entropy of the output of a single layer neural network. That is:

$$\underline{\mathbf{y}}_i = \underline{g}(\underline{\mathbf{u}}_i) \quad i = 1, \ldots, M \tag{2}$$

where $\underline{\mathbf{W}} \in \mathbb{C}^{N\times N}$ is the inverse of the mixing matrix, and $\underline{\mathbf{y}}_i$, $\underline{\mathbf{u}}_i = \underline{\mathbf{W}}\underline{\mathbf{x}}_i \in \mathbb{C}^N$ are the output and estimated sources, respectively. The entropy of a complex number is appropriately defined as the joint differential entropy of its real and imaginary parts [5]:

$$H(\underline{y}) = H(y_{\mathrm{Re}} + \mathrm{j}y_{\mathrm{Im}}) \triangleq H(y_{\mathrm{Re}}, y_{\mathrm{Im}}), \tag{3}$$

which can be written, in terms of the output probability density function (pdf), $p(\underline{\mathbf{x}})$, (which does not depend upon the weights, $\underline{\mathbf{W}}$) and the Jacobian:

$$H(\underline{\mathbf{y}}) = \sum_{i=1}^{N} E\left[\ln\left|\frac{\partial \underline{y}_i}{\partial \underline{u}_i}\right|\right] - E\left[\ln p(\underline{\mathbf{x}})\right] - \ln\left|\underline{\mathbf{W}}\right|. \tag{4}$$

The weight update equation for infomax using the natural gradient [6] can be written for the complex case as:

$$\Delta\underline{\mathbf{W}} \propto \frac{\partial H(\underline{\mathbf{y}})}{\partial \underline{\mathbf{W}}}\underline{\mathbf{W}}^H\underline{\mathbf{W}} = h\left[\mathbf{I} + \underline{j}(\underline{\mathbf{u}})\underline{\mathbf{u}}^H\right]\underline{\mathbf{W}} \tag{5}$$

where

$$\underline{j}(\underline{\mathbf{u}}) \triangleq \frac{\underline{g}''(\underline{\mathbf{u}})}{\underline{g}'(\underline{\mathbf{u}})} = \left[\frac{\underline{g}''(\underline{u}_1)}{\underline{g}'(\underline{u}_1)}, \ldots, \frac{\underline{g}''(\underline{u}_N)}{\underline{g}'(\underline{u}_N)}\right]. \tag{6}$$

Note that it is *entropy* that is being maximized, which is equivalent to maximization of the mutual information between the input and the output of the neural network. This works well provided the nonlinearity can at least roughly approximate the source distributions [7].

2.1.1. Split-Complex Infomax 1 (SCI-1)

Complex infomax using a non-analytic nonlinearity is introduced in [8]. In this case, the nonlinearity is defined as:

$$\underline{y} = \underline{g}(u_{\mathrm{Re}} + \mathrm{j}u_{\mathrm{Im}}) \triangleq \tanh(u_{\mathrm{Re}}) + \mathrm{j}\tanh(u_{\mathrm{Im}}). \qquad (7)$$

Following the terminology of [9], we call this implementation of complex infomax, the *split* complex infomax (SCI-1), since the real and imaginary values have been split into separate channels and define this nonlinearity as $\mathrm{stanh}(\cdot)$. This is introduced as a complex-valued extension of the $\tanh(\cdot)$ function used in the original derivation of infomax [7]. A split-complex function similar to (7) has also been used for deriving complex back-propagation for training the multi-layer perceptron network, and has been the traditional form of a *complex* neural network. The importance of using this function for back-propagation is that $\underline{g}(\cdot)$ is bounded over the entire complex plane. A motivation for using this function for complex infomax can be given as the need to use a function whose magnitude obeys the properties of a joint cumulative distribution function (cdf), $F(u_{\mathrm{R}}, u_{\mathrm{I}})$. Such an argument, however, ignores the fact that the nonlinearity is a complex-valued function and thus also contains phase information.

2.1.2. Split-Complex Infomax 2 (SCI-2)

In [8], where a shifted and scaled version of (7) is used for complex infomax, the update equations are directly written as a function of the nonlinearity, however the form given is a special case that only holds for the real-valued hyperbolic tangent. The function $\underline{j}(\underline{\mathbf{u}})$ is directly defined as the nonlinearity; a definition that is only true when $\tanh(\cdot)$ is used as the nonlinearity. The update equations involve the first and second order derivatives of the nonlinearity, thus it is important to derive these equations for the desired nonlinearity. The infomax update using this split nonlinearity is given in [1] where each element of $\underline{j}(\underline{\mathbf{u}}) \in \mathbb{C}^N$ is given by:

$$\underline{j}_i(\underline{u}) = -(y_{\mathrm{iRe}} + y_{\mathrm{iIm}}) - (y_{\mathrm{iRe}} - y_{\mathrm{iIm}})e^{-\mathrm{j}2q_{\underline{y}_i}} \qquad (8)$$

We refer to this approach as SCI-2.

2.1.3. Fully-Complex Infomax (FCI)

Instead of using bounded but non-analytic functions as the nonlinearity, we can use an *analytic* nonlinearity for the complex infomax algorithm. Since in this case, the real and imaginary values are not split into separate channels, we refer to this algorithm as the *fully-complex* infomax (FCI). Because boundedness and analyticity are conflicting requirements in the complex plane (stated by Liouville's theorem), split processing has been the traditional way of defining complex neural networks such as MLP networks. Since learning on the MLP requires backpropagation of error through the layers of neurons, the boundedness of the activation functions has been noted as being essential for stability. However, recent work has shown that, for the MLP, a complex counterpart of the universal approximation theorem can be shown with activation functions that are entire (analytic for all values of z) but bounded only almost everywhere [9]. In the case of infomax, since the network is single layer, and the output of the first layer is not used beyond the optimization stage, the problem is simpler. We can proceed by noting that the singular points have measure zero and hence will not be realized in practice and convergence thus will not be affected by their presence.

It is thus possible to use an analytic function for the nonlinearity in infomax as well and to directly use the complex form of the update in equation (5) where $\underline{j}(\underline{\mathbf{u}}) = 2\tanh(\underline{\mathbf{u}})$. The real and imaginary parts of the $\tanh(\cdot)$ function are plotted in Figure 1. One of the appealing properties of the infomax approach has been its straightforward implementation and low computational complexity. With $\tanh(\cdot)$ as the nonlinearity, these properties are preserved for complex infomax. This is not necessarily the case for its split-complex implementations.

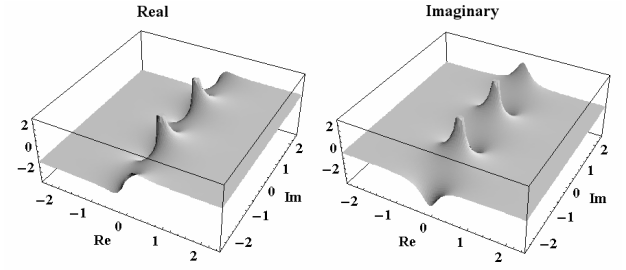

Figure 1: Plots of $g(\cdot)$ for Fully-Complex Infomax

2.2. Cumulant-based Approach

The JADE algorithm [10] uses fourth-order cumulants to achieve statistical independence. A computationally efficient technique is used for the blind estimation of directional vectors based on joint diagonalization of fourth-order cumulant matrices. It has been successfully applied to blind-beamforming and other areas.

3. APPLICATION TO FMRI ANALYSIS

3.1. Complex-Valued fMRI Images

Commonly, reconstruction techniques in MRI acquire the data in the complex k-space, and the data are mapped to a complex image space by an inverse discrete Fourier transform. The complex data in the magnitude/phase domain can be considered useful in terms of MR physics. For example, voxels with high blood volume fractions tend to exhibit the most detectable changes in phase and thus one can use phase to exclude large veins from further analysis. A different approach is to consider that the phase information can be used to improve the ability to identify a given source (by using both phase and magnitude behavior to "separate" or classify the source). For this approach, the complex data is operated upon in the imaginary/real domain, although following detection may be converted to magnitude/phase images and time courses for interpretation.

3.2. Simulated fMRI Data

One application of interest for complex infomax is the analysis of fMRI data where the data are naturally acquired as a spatio-temporal complex data set. The goal is to separate non-systematically overlapping (spatially independent) networks of activation. A source can be

interpreted as an image indicating areas that are related to one another functionally.

Two complex sources with two complex mixing vectors are generated. The source(s) consisted of $\sqrt{M}\times\sqrt{M}$ images (shaped into an $M\times 2$ matrix, $\mathbf{S}$), where M is the number of spatial samples, and the (random) mixing matrix is 2×2. We generate a data set with two sources, one having a slightly super-Gaussian (random) spatial distribution, modeling a "non-interesting" source, and one having between two and sixteen "activation" regions. The activations are generated by calculating the outer product of two $\sqrt{M}\times 1$ sinusoidal functions with periods of 2, 3, or 4. To make the activations "focused" (and their distributions super-Gaussian) this image was the raised to the power of 3,5,7, or 9 (odd powers were used to preserve "negative" activations). Two sets of the sources, $\mathbf{s}_{1i}$ and $\mathbf{s}_{2i}$ were generated to create for each column of $\mathbf{S}_i$:

$$\underline{\mathbf{S}}_i = \mathbf{s}_{1i}\sin(q) + \mathrm{j}\mathbf{s}_{2i}\cos(q) \qquad (9)$$

where q is uniformly distributed on $[-p, p]$ (different for each column), and $\underline{\mathbf{S}}_i$ represents the i^{th} column of $\underline{\mathbf{S}}$. An example of the activations is presented in Figure 2. Our simulated data set also included an additive complex Gaussian noise contribution.

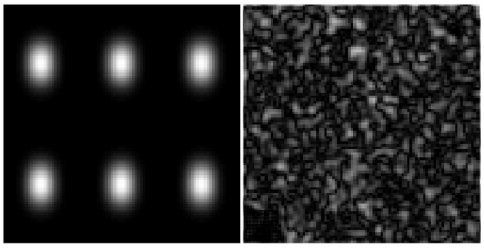

Figure 2: Sample of Simulated Sources

3.3. Real fMRI Data

BOLD fMRI data were acquired in a Philips 1.5T Gyroscan NT PT-6000 scanner. Nine 5 mm, single-shot gradient-echo echo-planar slices were acquired (repeat time=1s, echo time=39ms, flip angle=90°) over a four-minute period. The paradigm consisted of an 8 Hz reversing checkerboard turned off and on with a period of 60 seconds. The visual stimuli were provided on a rear-projection screen subtending 25 degrees of visual field via an LCD projector. Following review and approval by the Johns Hopkins University Institutional Review Board, four subjects gave informed consent.

3.4. ICA Estimation

The data matrix was of dimension $240\times M$ where M is the number of voxels inside the brain, hence performing spatial ICA. This matrix was reduced from 240 time points to 15 time points using principal component analysis (PCA). Following PCA, independent component estimation was performed using the algorithm described in section 2 to separate 15 sources. Time courses were reconstructed by multiplying the 15×15 mixing matrix by the 240×15 reducing matrix from the PCA stage. The source of interest was selected by correlating the magnitude of the time courses with the experimental paradigm (after convolution with a standard empirical hemodynamic response function). This source image was then converted to a Z-score image and thresholded at $|Z| > 2.5$. In this work, the ICA for each subject was estimated separately although one ICA estimation for all subjects is also possible [11].

4. RESULTS

4.1. Convergence

A summary of our results is presented in Table 1, where convergence is measured in both number of iterations and number of seconds. Convergence occurred when the weight change was less than 10e-6. Our algorithm was implemented in Matlab™ 6.5, running on a dual-processor AMD 1800+ machine with Windows XP™ as the operating system.

M = 3600	**Iterations**	**Time**
Split CI-1	45.6±2.1	1.1±0.1s
Split CI-2	247.4±13.5	9.74±0.55s
Fully CI	58.0±2.97	1.3±0.1s
JADE	--	0.02±0.01s

Table 1: Convergence Results from 100 Experiments

The JADE algorithm was by far the fastest, as it is highly optimized. The SCI-1 and FC approaches performed comparably whereas the SCI-2 took considerably longer to converge.

4.2. Approximation

It is important to examine the properties of each complex approach in terms of its ability to separate the final sources. The infomax algorithm works as the nonlinearity "matches" the distribution of the sources up to a scaling factor [7]. In general, the best performance (in terms of both convergence and approximation properties) should be obtained when the assumed distributions are selected to match the true distributions closely.

For infomax algorithms, it is important to have a nonlinearity that is capable of approximating the pdf of the sources. We show that a fully-complex activation function has reduced computational complexity compared with a split-complex approach, and, for a real-valued input the presence of cross terms in the Jacobian enables the analytic nonlinearity to approximate a more general class of input distributions. Contrary to previous assumptions [8], for infomax, any complex activation function that is unbounded (such as the hyperbolic tangent we use) provides convergence problems, we demonstrate that the opposite is the case, and the unbounded fully-complex activation function improves the shape of the performance surface, and, as a result, the algorithm converges faster.

We now measure, for simulated sources, the performance using the Kullback-Leibler (KL) divergence between the "true" and estimated distributions, and the normalized correlation between the "true" and estimated sources (given as r in the table). We also calculate the bias and variance of the estimated sources. The results, averaged over 100 experiments, along with their standard errors, are presented in Table 2.

M = 3600	r	KL	Bias	Var
Split-1	0.513±0.015	**1.172±0.067**	0.360±0.024	0.494±0.549
Split-2	0.460±0.014	1.363±0.068	0.420±0.023	0.575±0.052
FC	**0.518±0.016**	1.185±0.066	**0.087±0.012**	**0.341±0.025**
Jade	0.455±0.015	1.252±0.070	0.418±0.024	0.572±0.054

Table 2: Approximation Results from 100 Experiments

The fully-complex and the first split infomax approaches appeared to perform the best, with FCI slightly outperforming SCI-1 on all but the KL measure on the average (note that the difference is insignificant given the standard deviance). In the experiments above, we used a contrast-to-noise level of 1.0 (typical for fMRI data). If the additive noise is doubled, then FCI performs slightly better than SCI-1 (and the other two algorithms) on all four measures, suggesting robustness to noise.

4.3. Sub-Gaussianity

We are also interested in determining how well various complex ICA methods can separate sources that are sub-Gaussian. For temporal ICA of fMRI data, the expected signals are often bimodal in nature, and thus sub-Gaussian. In the next simulation we have a mixture of two sub-Gaussian sources (generated from uniform distributions). The results are presented in Table 3. In both cases, the fully-complex approach significantly outperforms the other three approaches.

Sub-Gaussian Sources				
M = 3600	r	KL	Bias	Var
Split CI-1	0.534±0.026	0.698±0.051	0.384±0.025	0.557±0.059
Split CI-2	0.598±0.020	0.614±0.036	0.137±0.0091	0.035±0.005
Fully CI	**0.980±0.008**	**0.047±0.095**	**0.087±0.012**	**0.008±0.037**
JADE	0.563±0.020	1.004±0.065	0.382±0.025	0.555±0.025

Table 3: Results for Sub-Gaussian Sources (N=100)

In [12], it is shown that the sigmoid nonlinearity is specialized for a super-Gaussian distribution, but a bimodal-distribution is necessary to appropriately separate a sub-Gaussian distribution. If we examine the fully-complex nonlinearity (see Figure 1), we see that it has the properties of a real-valued sigmoid nonlinearity along the real axis in that it increases smoothly from –1 to 1. However as one moves closer to the imaginary axis the profile begins to resemble the cdf of the bimodal density suggested in [12] for modeling sub-Gaussian sources. Thus it appears that the fully-complex nonlinearity has properties important for modeling both sub- and super-Gaussian distributions. This property potentially makes the complex $\tanh(\cdot)$ nonlinearity a very general tool for source separation with complex-valued data.

4.4. fMRI Data

We performed ten ICA estimations and calculated the average of the normalized sources. All complex-valued algorithms yielded a greater spatial extent of activation. Figure 3 demonstrates the empirical performance increase gained when using the complex-valued fMRI data.

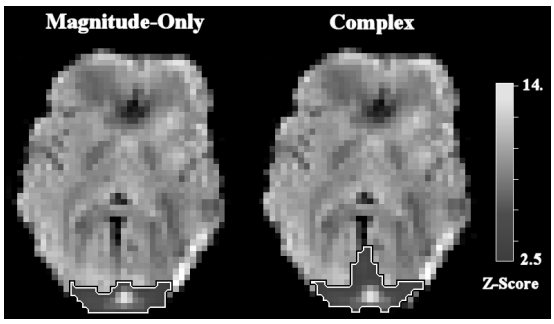

Figure 3: ICA Activation Maps: Supra threshold regions (outlined in white/black) overlaid onto anatomic image.

We compared performance of the various complex-valued ICA approaches again for this case, *i.e.*, using real fMRI data, and noted results similar to simulated data results given in 4.1 for convergence and increased sensitivity to detected activations with all four algorithms.

5. DISCUSSION

Using simulations and application to complex-valued functional MRI data, we have presented results comparing the convergence and approximation properties of several complex ICA approaches. Three infomax approaches (two split-complex and one fully-complex approach) and the JADE algorithm are compared. In general the fully-complex and the first split-complex infomax both perform comparably on super-Gaussian sources. However the fully-complex infomax outperforms all other approaches when the mixture involved sub-Gaussian sources. In this paper, we have presented only one choice for fully-complex nonlinearities. It is worth investigating the properties of infomax with other complex nonlinear functions such as the elementary transcendental functions proposed for MLP in [13] and study the performance of other complex ICA approaches such as the FastICA algorithm.

6. REFERENCES

[1] V.D. Calhoun, T. Adali, G.D. Pearlson, and J.J. Pekar, "On Complex Infomax Applied to Complex FMRI Data," in *Proc. ICASSP*, Orlando, FL, 2002.

[2] V.D. Calhoun, T. Adali, G.D. Pearlson, P.C. van Zijl, and J.J. Pekar, "Independent Component Analysis of FMRI Data in the Complex Domain," *Magn Reson. Med.*, vol. 48, pp. 180-192, 2002.

[3] P.A. Bandettini, A. Jesmanowicz, E.C. Wong, and J.S. Hyde, "Processing Strategies for Time-Course Data Sets in Functional MRI of the Human Brain," *Magn. Res. Med.*, vol. 30, pp. 161-173, 1993.

[4] V.D. Calhoun and T. Adali, "Complex Infomax: Convergence and Approximation of Infomax With Complex Nonlinearities," in *Proc. IEEE Workshop on Neural Networks for Signal Processing (NNSP)*, Switzerland, 2002.

[5] F.D. Neeser, "Proper Complex Random Processes With Applications to Information Theory," *IEEE Trans. Inf. Th.*, vol. 39, pp. 1293-1302, 1993.

[6] S. Amari, A. Cichocki, and H.H. Yang, "A New Learning Algorithm for Blind Signal Separation," *Advances in Neural Information Processing Systems*, vol. 8, pp. 757-763, 1996.

[7] A.J. Bell and T.J. Sejnowski, "An Information Maximisation Approach to Blind Separation and Blind Deconvolution," *Neural Comput.*, vol. 7, pp. 1129-1159, 1995.

[8] P. Smaragdis, "Blind Separation of Convolved Mixtures in the Frequency Domain," *Neurocomputing*, vol. 22, pp. 21-34, 1998.

[9] T. Kim and T. Adali, "Fully Complex Backpropagation for Constant Envelope Signal Processing," in *Proc. IEEE Workshop on Neural Networks for Signal Processing (NNSP)*, Sydney, Australia, pp. 231-240, 2000.

[10] J.F. Cardoso and A. Souloumiac, "Blind Beamforming for Non Gaussian Signals," *IEE-Proceeding-F*, vol. 140, pp. 362-370, 1993.

[11] V.D. Calhoun, T. Adali, G.D. Pearlson, and J.J. Pekar, "A Method for Making Group Inferences From Functional MRI Data Using Independent Component Analysis," *Hum. Brain Map.*, vol. 14, pp. 140-151, 2001.

[12] T.W. Lee, M. Girolami, and T.J. Sejnowski, "Independent Component Analysis Using an Extended Infomax Algorithm for Mixed Subgaussian and Supergaussian Sources," *Neural Comput.*, vol. 11, pp. 417-441, 1999.

[13] T. Kim and T. Adali, "Universal Approximation of Fully Complex Feed-Forward Neural Networks," in *Proc. ICASSP*, Orlando, FL, 2002.