

ON COMPLEX INFOMAX APPLIED TO FUNCTIONAL MRI DATA

V. Calhoun^{1,2}, T. Adali², G. Pearlson¹, J. Pekar^{3,4}

¹Johns Hopkins University Div. of Psychiatric Neuro-Imaging,
and ³Dept. of Radiology, Baltimore, MD 21287.

⁴FM Kirby Research Center for Functional Brain Imaging, Kennedy Krieger Institute, Baltimore, MD 21205.

²Dept. of CSEE, University of Maryland, Baltimore County, Baltimore, MD 21250.

ABSTRACT

Functional magnetic resonance imaging (fMRI) is a technique which produces complex data; however the vast majority of functional magnetic resonance imaging analyses utilize only magnitude images. In this paper, we derive a complex-valued independent component analysis (ICA) algorithm using the infomax approach which we then apply to fMRI analysis. Theoretical and empirical results demonstrate an improved sensitivity to functional changes when utilizing the complex data. Additionally, the complex infomax algorithm developed provides a powerful method for exploratory analysis of fMRI data.

1. INTRODUCTION

Since the early 1990's, fMRI has been utilized to observe small changes in blood oxygenation and flow coupled to neuronal activation [1]. Virtually all fMRI studies analyze only the magnitude images from the MRI scanner even though the data are acquired as complex images [2].

ICA has only recently been applied to fMRI data [3]. There is a need to study the properties of ICA as applied to fMRI data and to develop ICA methods to address specific issues involved in fMRI. We have recently proposed a generative model for fMRI data that can be used to optimize stages of ICA processing [4] and extended it for the analysis of multiple subjects using ICA [5]. In this paper, we extend the model to processing of complex images. Using a flexible analysis approach such as ICA is especially important since it is unclear how to model the complex fMRI data.

In the following, we describe complex fMRI data and the signal of interest, followed by a simple example showing the sensitivity improvement. Next we derive a complex ICA approach using the infomax principle. We then apply this method to complex fMRI data acquired from four subjects performing a simple visual paradigm and demonstrate an empirical sensitivity improvement.

2. THEORY

2.1. Complex Images

Commonly, reconstruction techniques in MRI acquire the data in complex frequency space, followed by an inverse discrete Fourier transform into the complex image space

[6]. The noise in the real and imaginary channels in MRI can be modeled as independent Gaussian [7].

In fMRI, the contrast-to-noise ratio (CNR) is of central importance, as the difference in signal intensity *between* two images is the signal of interest. We can demonstrate for a simple case that using the complex images results in an increased CNR.

Consider the complex data from an activated voxel with two conditions, *active* and *baseline*. The mean signal intensity in each condition can be written as $a = a_R + ja_I$, and $b = b_R + jb_I$, respectively. The contrast for a magnitude image approach can be written as:

$$c_{\text{mag}} = \|a\| - \|b\| = \sqrt{a_R^2 + a_I^2} - \sqrt{b_R^2 + b_I^2}, \quad (1)$$

and for a complex image approach as:

$$c_{\text{complex}} = a - b = (a_R - b_R) + i(a_I - b_I). \quad (2)$$

Assuming that the real and imaginary channels are independent zero-mean Gaussian with equal standard deviation, σ , we present the following two-sided hypothesis test for a given activation, t :

$$\text{if } \begin{cases} \|c_{\text{complex}/\text{mag}}\| > t & \text{choose } H_1 \\ \|c_{\text{complex}/\text{mag}}\| \leq t & \text{choose } H_0 \end{cases} \quad (3)$$

where H_0 denotes zero amplitude (no activation), and H_1 denotes a signal present (activation). The receiver operating characteristics (ROC) plot for CNR=1, 2, and 4 and $\sigma = 1$ is given in Figure 1.

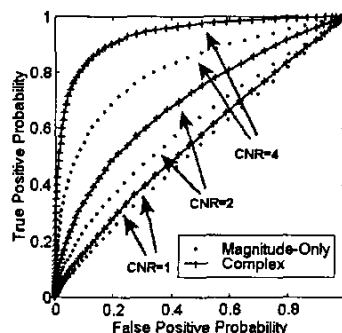


Figure 1: ROC for the magnitude and complex approaches for CNR = 1, 2, and 4. The complex approach performs better in all cases.

2.2. Complex Infomax

To perform ICA of complex-valued fMRI data by using the infomax approach, we formulate the problem as follows. In our notation, all variables (except indexes and dimensionality indicators) are assumed to be complex-valued unless subscripted by an R or I to denote the real or imaginary component of a complex variable (e.g. $u = u_R + ju_I$). The data, \mathbf{x}_i , at voxel i , is assumed to be a linear mixture of N statistically independent sources \mathbf{s}_i :

$$\mathbf{x}_i = \mathbf{A}\mathbf{s}_i \quad (4)$$

where \mathbf{A} is the $N \times N$ mixing matrix. The goal is to separate non-systematically overlapping (spatially independent) networks of activation. Our discussion will focus upon spatial independence although extension to temporal independence is straightforward [8]. The infomax algorithm proceeds by maximizing the entropy of the output of a single layer neural network. That is:

$$y_i = g(\mathbf{u}_i = \mathbf{W}\mathbf{x}_i) \quad (5)$$

where \mathbf{u}_i are the estimated sources and \mathbf{W} is the inverse of the mixing matrix. The update equation for infomax using the natural gradient can be written as [9]:

$$\Delta \mathbf{W} = \eta \left[\mathbf{I} + \frac{g''(\mathbf{u})}{g'(\mathbf{u})} \mathbf{u}^H \right] \mathbf{W} \quad (6)$$

where \mathbf{W} and \mathbf{I} are $N \times N$, \mathbf{u} is $N \times 1$, and

$$\frac{g''(\mathbf{u})}{g'(\mathbf{u})} = \begin{bmatrix} g''(u_1) & g''(u_N) \\ g'(u_1) & \dots & g'(u_N) \end{bmatrix} \quad (7)$$

and N is the number of sources. We define the nonlinearity as:

$$y = g(u_R + ju_I) \triangleq \text{sgm}(u_R) + j\text{jsgm}(u_I), \quad (8)$$

This is a complex-valued extension of the sigmoid function used in the original derivation [10].

$$\text{sgm}(u) = \frac{1}{1 + e^{-u}} \quad (9)$$

The function $g(\cdot)$ is bounded over the entire complex plane and obeys the properties of a joint cumulative distribution function (cdf), $F(u_R, u_I)$. It can be important to choose a nonlinearity which is able to "match" the input, that is, the nonlinearity can approximate the cdf of the sources. Additionally, when the nonlinearity has the form of a cdf, the infomax approach is equivalent to a maximum likelihood (ML) approach [11] thus providing access to the rich tools and properties of ML theory. Empirical evidence demonstrates that fMRI activations are mainly comprised of a set of unimodal distributions with positive kurtosis and Gaussian noise [4]. The sigmoid function is flexible enough to approximate either cdf hence the nonlinearity defined in (8) to model the joint distribution. Magnitude and phase plots of $g(\cdot)$ are depicted in the following figure.

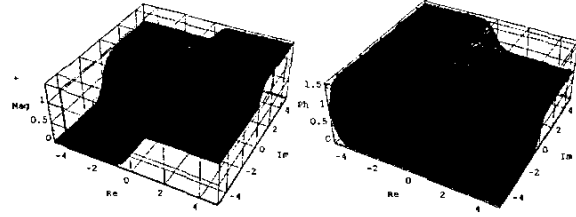


Figure 2: Magnitude (left) and Phase (right) of $g(\cdot)$

The infomax algorithm is derived by maximizing the entropy of the output from a single layer neural network. The entropy of a complex number is appropriately defined as the joint differential entropy of its real and imaginary parts [12]:

$$H(y) = H(y_R + jy_I) \triangleq H(y_R, y_I) \quad (10)$$

which can be written in terms of the output probability density function, $f_x(x)$, (which does not depend upon the weight w) and the Jacobian:

$$H(y) = E \left[\ln \left| \frac{\partial y}{\partial \mathbf{u}} \right| \right] - E[\ln f_x(x)] - \ln w \quad (11)$$

The fact that the complex nonlinearity y is bounded reveals, by a corollary to Liouville's theorem, that this function is non analytic. However we can use appropriate definitions of the first and second derivatives for this function as in [13] as follows:

$$y' = \frac{\partial y}{\partial u} \triangleq \frac{\partial y}{\partial u_R} + j \frac{\partial y}{\partial u_I} \quad (12)$$

$$y'' = \frac{\partial^2 y}{\partial u^2} = \frac{\partial^2 y}{\partial u_R^2} + j \frac{\partial^2 y}{\partial u_I^2} \quad (13)$$

The first derivative is then written as:

$$y' = y_R(1 - y_R) + jy_I(1 - y_I), \quad (14)$$

and the second derivative as:

$$y'' = y_R'(1 - 2y_R) + jy_I'(1 - 2y_I). \quad (15)$$

Replacing y_R' and y_I' with their definitions in terms of the complex variable y' and dividing by y' provides the following result:

$$\frac{y''}{y'} = (1 - y_R - y_I) - \frac{y''}{y'} (y_R - y_I). \quad (16)$$

In the implementation, a division can cause problems with stability. We can remove the division by noting that it can be written as a complex exponential in the following form:

$$\frac{y''}{y'} = (1 - y_R - y_I) - (y_R - y_I) e^{-j2\theta_{y'}}. \quad (17)$$

where $\theta_{y'} = \arg(y')$. We can now write the update equation for the complex nonlinearity defined in (8) as:

$$\Delta \mathbf{W} = \eta \left\{ \mathbf{I} + [h_1(u_1), \dots, h_N(u)] \mathbf{u}^H \right\} \mathbf{W} \quad (18)$$

where

$$h_i(u) = (1 - y_{iR} - y_{iI}) - (y_{iR} - y_{iI}) e^{-j2\theta_{y'}} \quad (19)$$

Note that if the data is real then the update equation reduces to the standard form:

$$\Delta W = \eta \{ \mathbf{I} + [\mathbf{I} - 2\mathbf{y}]\mathbf{u}^T \} W. \quad (20)$$

It is possible to use an analytic, non-bounded, function for the nonlinearity as well and to directly use the complex form of the update in equation (20). However, in this case, special attention has to be paid to the definition of a domain due to singularity and periodicity of the function as discussed in [14].

3. EXPERIMENTS AND METHODS

3.1. fMRI Paradigm & Subjects

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.

3.2. Pre-processing

Magnitude data were coregistered to an image averaged over all time points using SPM99 (Wellcome Department of Cognitive Neurology). These transformations were then applied to the real and imaginary images resulting in a motion-corrected set of complex images. The images were spatially smoothed with a 6x6x10 mm full-width at half maximum Gaussian kernel to reduce the amount of high-frequency spatial noise as well as to desensitize the images to errors in the motion correction. It is straightforward to demonstrate that spatial smoothing will not affect the (spatial) ICA estimate and thus is a reasonable preprocessing step [15].

3.3. ICA Estimation

The data matrix was of dimension $240 \times by \times M$ where M is the number of voxels inside the brain. This matrix was reduced [3] 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 2.2 to separate 15 sources. Time courses were reconstructed by multiplying the $15 \times by \times 15$ mixing matrix by the $240 \times by \times 15$ 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 [5].

4. RESULTS

We performed ten ICA estimations and calculated the average of the normalized sources. The thresholded spatial components from one subject are presented below. The complex-valued approach yielded a greater spatial extent of activation. Using the $|Z| > 2.5$ threshold, 46 voxels were in the magnitude-only maps but not in the complex maps and 193 voxels the reverse. A lower threshold of $|Z| > 1.5$ finds 48 voxels in the magnitude-only maps but not in the complex maps and 386 the reverse. There were thus considerably more locally connected voxels identified for the complex-valued approach.

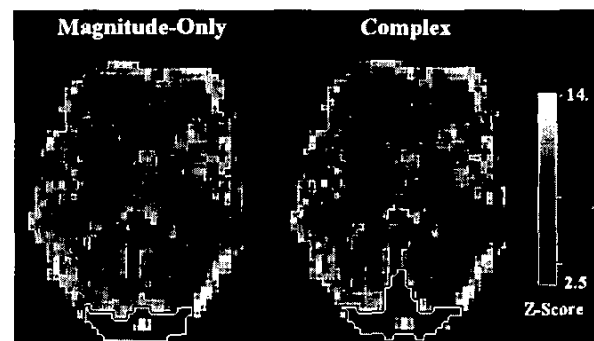


Figure 3: ICA Activation Maps: Supra threshold regions (outlined in white/black) overlaid onto anatomic image. The complex-valued approach results in a larger contiguously activated region in all subjects.

In the following figure we present the phase and amplitude time courses averaged over the regions that are detected by the complex approach. The relative average phase is calculated by subtracting from each voxel its average phase and averaging the supra-threshold voxels. Note that the phase demonstrates a periodicity similar to that of the amplitude curves. Approximately 63% of the detected voxels demonstrate the relationship depicted in Figure 4 whereas 37% of the voxels demonstrate the opposite relationship (see Figure 5).

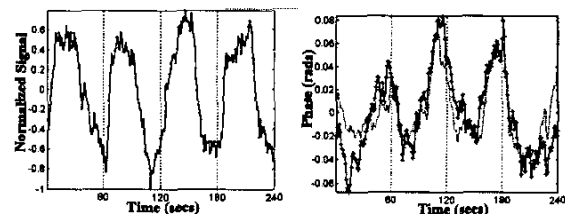


Figure 4: Magnitude (left) and Phase (right) Time Courses from One Subject: The phase time course for the complex-valued approach is plotted (dotted line), along with the relative average phase from the raw data (solid line), for those voxels that surpassed the threshold.

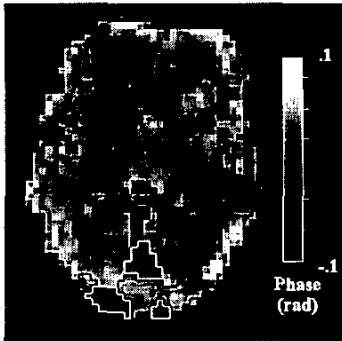


Figure 5: Phase Image from one subject: The results clearly indicates voxels in which the phase modulation is the same (white outline) or opposite the magnitude modulation (black outline).

5. DISCUSSION

In conclusion, we have developed an approach for analyzing complex-valued fMRI data using independent component analysis. We have demonstrated that utilizing the complex data provides a sensitivity increase in theory (see Figure 1) and on fMRI data as well. Our method for comparing the methods presented in the paper uses, as a criterion, the number of voxels that survive a magnitude threshold of the ICA images. The complex-valued ICA estimation detects an average of 25% more voxels than magnitude-only ICA at a threshold of $|Z| > 2.5$. Our data also demonstrate that some of the voxels have a phase modulation similar to or opposite that of the magnitude modulation. Using such flexible approaches for analyzing complex-valued fMRI data provides much needed gain in detection power when the contrast to noise ratio is small, without imposing strong modeling assumptions upon the data.

6. ACKNOWLEDGEMENTS

Data were acquired at the F.M. Kirby Research Center for Functional Brain Imaging, Kennedy Krieger Institute. Supported by the National Center for Research Resources, National Institutes of Health, under grant P41 RR15241.

7. REFERENCES

- [1] K.K.Kwong, J.W.Belliveau, *et al.*, "Dynamic Magnetic Resonance Imaging of Human Brain Activity During Primary Sensory Stimulation," *Proc. Natl. Acad. Sci.*, vol. 89, pp. 5675-5679, 1992.
- [2] P.A.Bandettini, A.Jesmanowicz, *et al.*, "Processing Strategies for Time-Course Data Sets in Functional MRI of the Human Brain," *Mag. Res. Med.*, vol. 30, pp. 161-173, 1993.
- [3] M.J.McKeown, S.Makeig, *et al.*, "Analysis of fMRI Data by Blind Separation Into Independent Spatial Components," *Hum. Brain Map.*, vol. 6, pp. 160-188, 1998.
- [4] V.Calhoun, T.Adali, and G.Pearlson, "Independent Components Analysis Applied to fMRI Data: A Generative Model for Validating Results," in *Proc. IEEE Workshop on Neural Networks for Signal Processing (NNSP)*, Falmouth, MA, 2001.
- [5] V.Calhoun, T.Adali, *et al.*, "A Method for Making Group Inferences From Functional MRI Data Using Independent Component Analysis," *Hum. Brain Map.*, vol. 14, pp. 140-151, 2001.
- [6] W.A.Edelstein, J.M.Hutchison, *et al.*, "Spin Warp NMR Imaging and Applications to Human Whole-Body Imaging," *Phys. Med. Biol.*, vol. 25, pp. 751-756, 1980.
- [7] A.Macovski, "Noise in MRI," *Mag. Res. Med.*, vol. 36, pp. 494-497, 1996.
- [8] V.D.Calhoun, T.Adali, *et al.*, "Spatial and Temporal Independent Component Analysis of Functional MRI Data Containing a Pair of Task-Related Waveforms," *Hum. Brain Map.*, vol. 13, pp. 43-53, 2001.
- [9] S.Amari, A.Cichocki, *et al.*, "A New Learning Algorithms for Blind Signal Separation," *Advances in Neural Information Processing Systems*, vol. 8, pp. 757-763, 1996.
- [10] 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.
- [11] B.A.Pearlmutter and L.C.Parra, "A Context-Sensitive Generalization of ICA," in *Proc. Int. Conf. on Neural Inf. Proc.*, Hong Kong, 1996.
- [12] F.D.Neuser, "Proper Complex Random Processes With Applications to Information Theory," *IEEE Trans. Inf. Th.*, vol. 39, pp. 1293-1302, 1993.
- [13] N.Benvenuto and F.Piazza, "On the Complex Backpropagation Algorithm," *IEEE Trans. Signal Proc.*, vol. 40, pp. 967-969, 1992.
- [14] T.Kim and T.Adali, "Complex Backpropagation Network Using Elementary Transcendental Functions," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, Salt Lake City, UT, 2000.
- [15] A.Hyvarinen and E.Oja, "Independent Component Analysis: Algorithms and Applications," *Neural Netw.*, vol. 13, pp. 411-430, 2000.