A CONSTRAINED COEFFICIENT ICA ALGORITHM FOR GROUP DIFFERENCE ENHANCEMENT

Jing Sui[*1], *Jingyu Liu*[1,2], *Lei Wu*[2], *Andrew Michael*[1], *Lai Xu*[2], *Tulay Adali*[3], *Vince D. Calhoun*[1,2]

[1]The Mind Research Network, 1101 Yale Blvd, NE, Albuquerque, NM, 87131
[2]Dept. of Electrical and Computer Engineering, University of New Mexico, Albuquerque, NM, 87131
[3]Dept of CSEE, University of Maryland, Baltimore County, Baltimore, MD, 21250

ABSTRACT

Independent component analysis (ICA) is a statistical and computational technique for revealing hidden factors that underlie sets of signals. We propose an improved ICA framework for group data analysis by adding an adaptive constraint to the mixing coefficients, namely, constrained coefficients ICA (CCICA). The method is dedicated to identification and increasing the accuracy of components that show significant group differences reflected in the mixing coefficients. Performance of CCICA is assessed by simulations under different signal to noise ratios. An application to multitask functional magnetic resonance imaging analysis is conducted to illustrate the advantages of CCICA. It is shown that CCICA provides stable results and can estimate both the components and the mixing coefficients with a relatively high accuracy compared to Infomax, hence is a promising tool for the identification of biomarkers from brain imaging data.

Index Terms— Independent component analysis, functional magnetic resonance imaging, mixing coefficients, Infomax, joint ICA

1. INTRODUCTION

Independent component analysis (ICA) is a theoretically rich, extendable approach that uses higher order statistical information and can reveal hidden associations between variables whose explicit relationship is not well understood [1]. ICA has been increasingly applied to analysis of multimedia streams and brain imaging data with success.

In the joint ICA application of multitask/multimodal (MTMM) fMRI data analysis [2,3], the MTMM data of all subjects are first preprocessed to produce an activation map. Then the activation images are flattened to vectors and stacked side by side with e.g. another task to obtain the observed data (subjects by voxels). Each task shares the same mixing matrix, and the components showing group difference are identified by a 2 sample *t*-test on the mixing coefficients to test whether their means are different. We have found that the component exhibiting the largest group difference don't always have the smallest *p* value for the 2 sample *t*-test, especially when the known sources added to each group have some variation.

Motivated by the constrained ICA framework [4,5], we incorporate hypothesis testing knowledge as additional constraint into a general ICA cost function $G(Y)$ which is usually maximized in order to extract ICs, where Y is a nonlinear function of the observed data. We expect to increase the average difference of the mixing coefficients while simultaneously maximizing $G(Y)$. Thus, for a fixed number of group members, a higher statistic T value of a two-sample *t*-test, and hence a lower *p* value will be attained. The CCICA method can then be formulated by expanding the cost function and choosing an appropriate optimization strategy.

In this manuscript, CCICA is introduced to Infomax ICA, and the joint CCICA is demonstrated with a two task fMRI data set. Note that since CCICA is designed for group analysis, it can be used for either single modality data or MTMM data. The use of MTMM data enables us to capture patient versus controls differences by using cross-information between brain networks.

2. METHOD

2.1. Introduction to Infomax

Infomax is a self-organizing learning algorithm which maximizes the information transferred in a network of nonlinear units [7]. The ICA framework is defined as equation (1), where the observed data X consists of measurements such as MRI images or speech signals. The components, S, contain independent sources such as brain activation networks or multiple speakers' recordings as in

$$X = A \cdot S \tag{1}$$

where X is assumed to be a linear mixture of components S, and A is the mixing matrix containing the loading parameters. The aim of ICA is to find the unmixing matrix $W=A^{-1}$ so that WX is as close as possible to the true source S. Using natural gradient learning rules, the Infomax algorithm attempts to find the W matrix by maximizing the

entropy of the nonlinear output Y defined as

$$H(Y) = -E[\ln P(Y)] \tag{2}$$

$$Y = \frac{1}{1+e^{-U}}, U = W \cdot X \tag{3}$$

where $P(Y)$ is the probability density function of Y, E is the expectation operator and H the differential entropy .

2.2. CCICA model

To provide for group inferences, the ICA model can be extended as

$$\begin{bmatrix} X_h \\ X_p \end{bmatrix} = \begin{bmatrix} A_h \\ A_p \end{bmatrix} \cdot S \tag{4}$$

where the suffixes h and p denote healthy controls and patients, A_{h}, A_{p} are corresponding mixing coefficients matrix of which each column represents loading parameters for one shared component, their inverse matrixes W_{h}, W_{p} are key parts used to back-reconstruct sources for each group. The goal is to find the components that significantly differentiate the groups as measured by computing a two sample t-test of the mixing coefficients.

We start by considering a fixed number of subjects, including n_h healthy controls and n_p patients. It is assumed that distributions of $A_{\mathrm{h,i}}$, $A_{\mathrm{p,i}}$ have the same variance(If this assumption is violated, the unequal variance t-test can be used alternatively). The T statistic to test whether their means are different can be calculated as equation (5):

$$T_i = \frac{\overline{A_{h,i}} - \overline{A_{p,i}}}{\sqrt{(n_h-1)s_h^2 + (n_p-1)s_p^2}} \times \sqrt{\frac{n_h \cdot n_p (n_h+n_p-2)}{n_h+n_p}} \sim t(n_h+n_p-2) \tag{5}$$

where the symbol "$\overline{\ }$" is the average function, vector $A_{h,i}$ $A_{p,i}$ (the i^{th} column of matrix A_h and A_p), denote the loading parameters of each group related with the i^{th} component, s^2 is the unbiased estimator of the vector's variance, n_h+n_p-2 is the number of degrees of freedom for two-tailed significance testing. The larger the absolute value of T_{i}, with the higher probability that the mixing coefficients' average are significantly different for two groups.

Therefore the cost function is constructed such that in addition to the given objective (output entropy in the case of Infomax) for achieving independence, the squared statistic T_i between $A_{h,i}$ and $A_{p,i}$ is selected as a constraint for joint maximization of C such that

$$C = Max\{H(Y) + \lambda \cdot \sum T_i^2\} \tag{6}$$

where $H(Y)$ is entropy of the nonlinear output, λ is the weight of the T^2 term, the suffix i represents the column index of the constrained component(s) that may differentiate two group significantly. We next discuss an adaptive approach to determine which components are constrained.

2.3. Dynamic CCICA optimization

The goal of ICA is to estimate the unmixing matrix W that maximizes the independence among the components. Most solutions depend on minimizing or maximizing a multivariate cost function. For the optimization process of CCICA, we emphasize two points: 1) dynamically constrained components and 2) adaptive constraint strength.

First, the specific components constrained are not fixed, instead, they are allowed to vary during the optimization process. In practice, we first run Infomax for several iterations without a constraint and when the change ΔW stabilizes, the corresponding T values are calculated and sorted for all components, one or more columns of matrix A (depending on the component number used) with the highest T^2 values are constrained in this iteration. The fewer components we constrain, the more they fluctuate between iterations initially, but as W converges, the constrained components become fixed at every iteration. This flexibility provides an ability to update the components selectively.

Second, we adaptively adjust the constraint strength of the T^2 term in the cost function C. Specifically, the learning rates are allowed to continuously change during the optimization. In most cases, the entropy term and the T_i^2 term converge at different rates. To compensate for the differences, we first maximize entropy term using the natural gradient [7], then we optimize the $\sum T_i^2$ term using a steepest ascent algorithm with a step length (learning rate) estimated from selected components at each iteration. Note that the entropy maximization should dominate the W update process, if the learning rate is too large, the constraint T^2 may infringe upon the independence criterion and decrease the entropy, so the tendency of ΔW during the last l (i.e., l=10) iterations is used as a supervision to determine the learning rate. If the cost function is maximized, ΔW should decrease on the whole and eventually go to zero, otherwise the slope of the last l ΔW s could increase. If this condition occurs, the learning rate is scaled down, ensuring that the maximum entropy attained.

2.4. Performance evaluation metrics

CCICA is designed to reliably obtain desired ICs with high accuracy, which simultaneously can differentiate two groups as much as possible. Three metrics are selected to evaluate performance of CCICA: the smallest p value that determines the probability of whether two groups have significantly different mean, accuracy of the recovered ICs and accuracy of the mixing coefficients. The estimated components should be the same as the "true" sources; the correlation between the true source and the extracted component under different signal to noise ratios (SNRs) are used to indicate the component accuracy. Finally, we compute the correlation of constrained columns of the mixing matrix A with the real loading parameters to the mixing coefficients.

3. SIMULATED SIGNALS TESTING

We demonstrate performance of CCICA compared with standard Infomax using synthetic data created according to the joint ICA framework in which the observed data of two tasks share the same mixing matrix. Infomax when used with the sigmoid nonlinearity performs well with super-Gaussian signals, so four zero-mean super-Gaussian signals each including two tasks in length of 1000 samples were randomly mixed to obtain 16 mixtures. The mixing matrix A is vertical combination of two 8×4 matrices A_h and A_p with uniform distribution in range of [0–1]. The average of the fourth column of A_h is set to be larger than the corresponding one of A_p, so that mixing coefficients of this component shows significant group difference in the mean.

Sources are allowed to exhibit some variation between the two groups. Specifically, the shapes of true sources for two groups are similar, as figure 1(a) shows, but with different frequency and width of the signal periods.

All components are constrained here since there are only four ICs. As expected, CCICA converged to produce the output signals identical to the desired sources under different SNRs. Figure 1(b) displays the back-reconstructed sources U_h, U_p by CCICA at low SNR, where $U_h = W_h X_h$, $U_p = W_p X_p$.

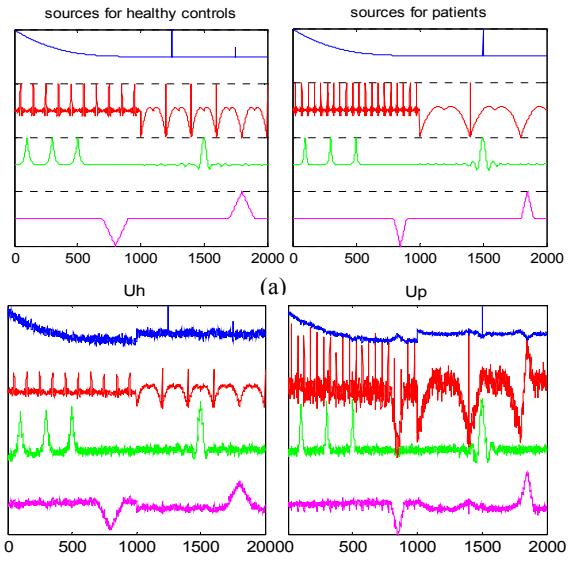

Figure 1. (a) Real sources for each group (b) back-reconstructed ICs derived from CCICA at SNR=12.5

Accuracy of all recovered ICs and the mixing coefficients are estimated and listed in Table 1. Note that for the 4th component, both algorithms showed it had significant group different mixing parameters and CCICA showed the smallest p value. For the 2nd component, correlations of either sources or coefficients attained from CCICA with the ground truth are increased remarkably in both SNR conditions relative to Infomax, confirming its effectiveness.

		CCICA				Infomax			
IC		1	2	3	4	1	2	3	4
Low SNR = 12.5	*Correlation of ICA loading with true mixing coefficients*								
	A	0.996	**0.898**	0.991	0.934	0.996	**0.819**	0.992	0.912
	A_h	0.973	0.972	0.992	0.974	0.981	0.999	0.993	0.969
	A_p	0.998	**0.912**	0.994	0.906	0.998	**0.254**	0.995	0.867
	Correlation of extracted IC with true source								
	U_h	0.790	**0.910**	0.944	0.841	0.790	**0.887**	0.943	0.892
	U_p	0.811	**0.711**	0.965	**0.912**	0.831	**0.635**	0.965	**0.625**
	p value of 2 sample t-test between A_h and A_p								
	p	0.359	0.952	0.279	**0.003**	0.415	0.774	0.328	**0.012**
High SNR = 31.9	*Correlation of ICA loading with true mixing coefficients*								
	A	0.997	**0.856**	0.992	0.927	0.996	**0.801**	0.992	0.901
	A_h	0.980	0.997	0.994	0.993	0.983	0.999	0.994	0.956
	A_p	0.999	**0.779**	0.995	0.877	0.998	**0.387**	0.995	0.858
	Correlation of extracted IC with true source								
	U_h	0.845	**0.926**	0.975	0.908	0.845	**0.889**	0.975	0.914
	U_p	0.835	**0.699**	0.997	**0.861**	0.845	**0.650**	0.997	**0.568**
	p value of 2 sample t-test between A_h and A_p								
	p	0.343	0.905	0.252	**0.003**	0.402	0.725	0.313	**0.013**

Table 1 Simulation results under different SNR

4. APPLICATION TO MULTITASK FMRI DATA

An approach to examine the "coupled" activation across multiple tasks has been proposed and applied to fMRI data [2]. We found that when the artificially added sources are same for two groups, Infomax works well by extracting them directly with the smallest p value. But when working with real data which is noisy, Infomax does not always get optimum results.

As figure 2 (a) shows, we created hybrid data by superimposing variant sources on actual SPM contrast images generated from the auditory oddball and Sternberg tasks. For healthy controls, it is a 21×21 half-cycle sinusoid pattern, and for patients, the pattern is contra-rotated by 45°. The mixing coefficients are random numbers drawn from a uniform distribution. The ranges used were [1.5–0.5] and [0–1] for patients and controls respectively.

CCICA and Infomax are then used to perform joint ICA on the mixed data to extract maximally spatially independent maps for each task that are together by a shared loading parameter under the contrast to noise ratio (CNR) of 5. CNR is computed as the maximum value of the known source divided by the standard deviation of the fMRI data at the same voxel.

The first three components extracted from the two algorithms show a significant difference between groups. Figure 2(b) illustrates the three components sorted by two sample t-test p values. It is obvious that the component ranked first in CCICA with p<4.03e-5 is the IC that we expected, with known sources showing at the right position of both features. However, the corresponding IC in Infomax is only sorted third with p<0.1199, hence not showing a significant group difference. Note that the shape of the activations is almost identical to the sources we added to the

patients, which may be because patients have higher mean of loading parameters than controls when mixed. Also, we calculate the J divergence between the back-reconstructed source distributions for controls and patients. The desired IC from two ICA algorithms both show the highest divergence value among all the unmixed ICs, which confirmed that it is the right source that shows the largest group difference.

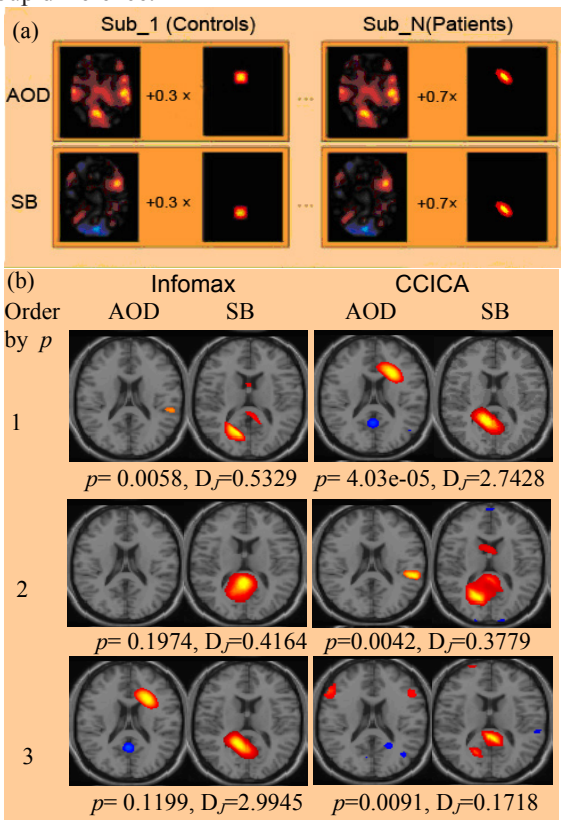

Figure 2 (a) generation of hybrid data (b) the first three ICs sorted by p value derived from two ICA algorithms

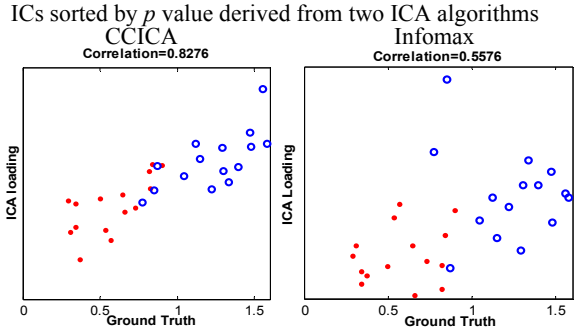

Figure 3 Correlation of the loading parameters of the desired IC extracted from two algorithms with the ground truth. (patients are coded in blue circles, controls in red points)

The correlation between loading parameters of the desired ICs and the true mixing coefficients are calculated to determine the mixing coefficient accuracy. As expected, the first component of CCICA and the third component of Infomax again show the highest correlation. Figure 3 plots the loading parameter versus the ground truth (patients are coded in blue circle, controls in red points). In both cases, the controls showed a lower mean than the patients did, and CCICA outperformed Infomax with a higher unmixing accuracy (0.8276 vs. 0.5576). In addition, the IC with the largest two sample t-test result is consistent with the divergence sort order.

The presented results support the claim that, by incorporating an additional constraint with ICA, CCICA can better identify the IC showing largest group difference compared to standard Infomax which shows less consistent results.

5. CONCLUSIONS

We present a novel algorithm, CCICA, to extract independent components that show significant group differences in the mixing coefficients. The problem is formulated using the constrained ICA framework, and prior knowledge of group membership is incorporated into conventional entropy maximization algorithms. Applications to both simulated signals and real multitask fMRI data demonstrated several advantages of CCICA. Compared to Infomax, CCICA can identify components and mixing coefficients with a relatively high accuracy at different SNR values.

Note that CCICA can be applied to either single modality or to MTMM data. It has great potential for the precise group analysis of brain imaging data and specifically for the identification of features which may serve as potential biomarkers.

The authors would like to thank the Mind Research Network for providing the fMRI data. This research was supported by NIH grants No. 1R01EB006841 and No. 1R01EB005846.

REFERENCES

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

[2] V.D. Calhoun, et al., "A method for multitask fMRI data fusion applied to schizophrenia," *Hum Brain Mapp*, 27(7): pp. 598-610, 2006.

[3] V.D. Calhoun, et al., "Method for multimodal analysis of independent source differences in schizophrenia: combining gray matter structural and auditory oddball functional data ," *Hum Brain Mapp*, 27(1): pp. 47-62, 2006

[4] W. Lu and J.C. Rajapakse, "Approach and applications of constrained ICA", *IEEE Trans Neural Netw*, 16(1): pp.203-212, 2005.

[5] J. Liu, V.D. Calhoun**.** "Parallel independent component analysis for multimodal analysis: application to fMRI and EEG data," *Proc. of IEEE ISBI conference, Arlington, USA.* pp.1028-1231, 2007.

[6] V.D. Calhoun, T. Adali, J. Liu. "A feature-based approach to combine multimodal brain imaging data". *Proc. of the 28th IEEE EMBS Conference*, *New York City*, USA. pp.3672-3675, 2006.

[7] J.F. Cardoso, "Infomax and maximum likelihood for blind source separation", *IEEE Signal Processing Letters*, 4(4): pp. 112-114, 1997.