A fixed-point iteration based constrained independent component analysis and its application in fMRI

Z. Wang[1]

[1]Department of Psychiatry, University of Pennsylvania, Philadelphia, PA, United States

Introduction. Independent component analysis (ICA) [1] is a useful tool for fMRI data analysis, but it suffers from the intrinsic order ambiguity, which makes it tricky to pick up the interested components across different runs and makes a problem of assessing the same or similar components across subjects. One solution is to incorporate prior information in the learning process as in the constrained ICA (cICA) [2,3]. However, the original cICA depends on a learning rate, which is not easy to be tuned. In this work, we developed a fast and learning rate free cICA algorithm and validated its performance for brain activation detection.

Materials and Methods. CICA can be described as an inequality constrained problem (ICP)[2,3]: maximize an object function $\vartheta(\mathbf{y})$, subject to $q(\mathbf{y}) \le 0$ and $h(\mathbf{y}) = 0$, where $q(\mathbf{y}) = (q(y_1), q(y_2), ..., q(y_M))^T, (M \le m)$ is the function controlling the closeness of the components to the priors, and a correlation based $q(\mathbf{y})$ was used in this work: $q(\mathbf{y}) = \xi^2 - E\{(y^T r)^2\}$, m is the number of components, $h(\mathbf{y}) = \mathbf{W}\mathbf{W}^T - \mathbf{I}$, and $\mathbf{W}$ is the weight to recover the latent source $\mathbf{s}$ from their mixed version $\mathbf{x}$ through $\mathbf{y}=\mathbf{W}\mathbf{x}$. A widely used $\vartheta(\mathbf{y})$ is the sum of negentropy $\sum_i^m J(y_i)$ and $J(.)$ can be approximated with $J(y_i) \approx c[E\{G(y_i) - G(v)\}]^2$ [1], where G is a nonquadratic function and is chosen to be $\log\cosh(.)$ in this work, $\mathcal{V}$ is a Gaussian variable with zero mean and unit variance. Converting $q(\mathbf{y}) \le 0$ into $q(\mathbf{y}) + \mathbf{z}^2 = 0$, this ICP can be solved using an augmented Lagrangian function: (1) $L(\mathbf{W}, \boldsymbol{\mu}, \mathbf{z}, \boldsymbol{\lambda}) = \vartheta(\mathbf{y}) - \boldsymbol{\mu}^T (q(\mathbf{y}) + \mathbf{z}^2) - 1/2\gamma \left\| q(\mathbf{y}) + \mathbf{z}^2 \right\|^2 - \boldsymbol{\lambda}^T h(\mathbf{y}) - 1/2\gamma \left\| h(\mathbf{y}) \right\|^2$, where $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$ are two sets of Lagrangian multipliers, and γ is a scalar penalty function, and $\|.\|$ is the Euclean norm. The optimal value of z^2 can be obtained through maximizing L wrt z^2. With the optima of z^2, $\boldsymbol{\mu}$ can be updated with: (2) $\boldsymbol{\mu}_{k+1} = \max\{0, \boldsymbol{\mu}_k + \gamma q(\mathbf{y})\}$. Neglecting the last item in Eq 1 and replacing z^2 by its optima, we can get: (3) $L(\mathbf{W}, \boldsymbol{\mu}, \boldsymbol{\lambda}) = \vartheta(\mathbf{y}) - 1/(2\gamma)(\max^2\{\boldsymbol{\mu} + \gamma q(\mathbf{y}), 0\} - \boldsymbol{\mu}^2) - \boldsymbol{\lambda}^T (\mathbf{W}\mathbf{W}^T - \mathbf{I})$. The optimal $\mathbf{W}$ can be then approached by setting the derivative of L wrt $\mathbf{W}$ to 0, resulting in the following equation: $2\lambda\mathbf{W} = \nabla_{\mathbf{W}}\vartheta - \nabla_{\mathbf{W}} L = E\{\mathbf{x}G'(\mathbf{W}\mathbf{x})\} - \boldsymbol{\mu}E\{\mathbf{x}q'(\mathbf{y})\}$, which can be further simplified as: (4) $\mathbf{W} = E\{\mathbf{x}G'(\mathbf{W}\mathbf{x})\} - \boldsymbol{\mu}E\{\mathbf{x}q'(\mathbf{y})\}$ since the scaling factor 2λ will be automatically dropped off if we force the weight to be normalized after each iteration using the fixed-point concept [1]: (5) $\mathbf{W} = \mathbf{W}/\|\mathbf{W}\|$. In summary, our proposed fixed-point cICA algorithm consists of Eqs. 4, 2, 5 and a deflation process: $\mathbf{W} = (\mathbf{W}\mathbf{W}^T)^{1/2}\mathbf{W}$ to prevent different weights from converging to the same optima.

The FastICA package [1] was modified to implement the algorithm in Matlab (Mathworks, Natick MA) scripts. All input data were centered and dewhitened using PCA decomposition. Four signals were generated with the normalized kurtoses of -0.2958, 4.3121, -1.2041, and -0.0422, respectively. The 1st source signal was a 1/f noise and its reference was a different 1/f signal. The references for the other 3 sources were different from the sources but their spectrums had the same major item as their corresponding sources. The mixed data were then separated 100 times using the original cICA and the proposed cICA. The aggregate performance index [2], signal-to-noise ratio (SNR), and separation time were collected. The proposed cICA was also applied to synthetic fMRI data that were generated as in [4] and previously published sensorimotor fMRI data [4].The boxcar function defining the activation sessions was used as the constraint for cICA to pick up the task related component. Univariate general linear model (GLM) provided in SPM5(http://www.fil.ion.ucl.ac.uk/spm) was used to generate the statistical parametric map of the activation-baseline contrast. ROC curves were collected from the cICA component map and the GLM t-map as described in [5]. FMRI data (n=17) preprocessing was performed using SPM5 based batch scripts [5], including realignment, coregistration, smoothing with an isotropic Gaussian filter (FWHM=6 mm), low-pass filtering with a Butterworth filter (cutoff frequency = 1/8 Hz), a high-pass Butterworth filtering (cutoff frequency = 1/128 Hz), spatially normalization to the MNI 152 standard brain. CICA was then applied using the design function as the constraint. GLM was also applied to collect the statistical parametric maps. A group analysis using the one-sample t-test was finally run on the individual cICA maps and GLM parametric maps.

cICA	SNR				PI	Time
Algorithms	S1	S2	S3	S4		(sec)
original	30.26	28.39	33.56	24.36	0.1022	0.2372
new	30.26	28.39	28.30	22.94	0.1287	0.5068

Results and discussions

As listed in the table on the right, the proposed cICA outperformed the original cICA in terms of higher SNR, lower PI, and shorter convergence time for separating the 1D synthetic data. These performance differences are due to the less computation burden of the proposed method as compared to the original one. The latter depends on a learning rate, which may not be easy to be tuned. In this work, the learning rate was initiated to be 1 and then gradually reduced until the learning process converged [1].

Fig. 1 shows that the proposed cICA demonstrated a much better sensitivity/specificity performance than standard univariate GLM for brain activation detection. This outperformance is due to the multivariate processing property and the source separation property of cICA. Treat each image as a single unit, cICA seeks the entire spatially coherent activation pattern as a whole rather than assessing them voxel by voxel separately. Since the spatially coherent activation patterns are more robust to noise interference than each single voxel if the noise is spatially independent which is a quite common assumption taken in the real MRI field, assessing the patterns as a whole should then gain an increased sensitivity as compared to assessing them separately at each voxel. Moreover, if there is just one Gaussian noise component and the desired component is not Gaussian distributed ICA could give a perfect detection to the spatially connected activation, since the spatial noise could be extracted as a separate component. Incorporating constraints in the learning procedures, cICA can do a better job by attracting the components to the desired location in the case of multiple local optima.

Fig. 2 shows the group level brain activation revealed by cICA and GLM from the sensorimotor fMRI. At the same significance level (t=4.67, P=0.05 with FDR correction), cICA yielded more extended activation clusters in visual cortex and motor cortex than GLM, while GLM revealed more focal activations in thalamus (Tha), left insula (lIns), and the supplementary motor area (SMA). The outperformance of cICA in visual and motor cortex is due to the same reasons stated above, while the underperformance in Tha, lIns and SMA could be caused by the less coherence of brain activities in these regions than in other regions underlying the overall activation pattern. Fig. 3 showed the aggregate correlation coefficients (CC) of the time series between lIns, Tha, right motor cortex, and visual cortex. We can see that fMRI time course in left insula has less aggregate correlation than other regions.

Reference [1] Hyvarinen et al, independent component analysis, Joh Wiley&Son, 2001, [2] Lu et al., ANIPS, 16:570-6,2000. [3] Lu et al., IEEE TNN, 2005,16:203-12. [4] Wang et al., NeuroImage, 46:608-615, 2009. [5] Wang et al., MRI, 2008, 26(2): 261-9.

Acknowledgement This work was supported by NIH grants NIH grants: R03DA023496, RR02305, R21DA026114, R01DA025906.

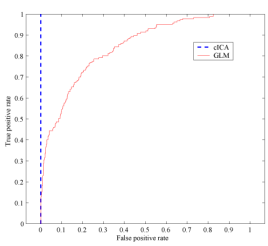

Fig 1. ROC curves of cICA and GLM for the synthetic brain activation detection.

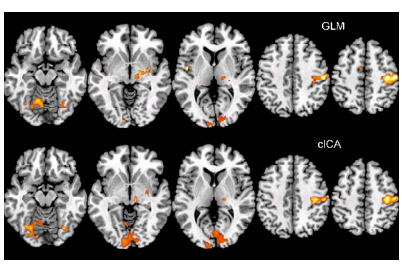

Fig 2. Group level statistical analysis results of the sensorimotor fMRI data. The color window is 4.67-8.

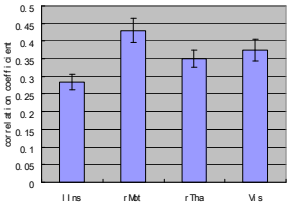

Fig 3. The mean and standard error of the aggregate CC of the 4 assessed ROIs for the sensorimotor fMRI data.