A hierarchical group ICA model for assessing covariate effects on brain functional networks

Ying Guo

Joint work with Ran Shi

Department of Biostatistics and Bioinformatics

Emory University

06/03/2013
Introduction of independent component analysis (ICA) for fMRI data
A hierarchical group PICA model for modeling covariate effects on brain functional networks
Estimation methods for the proposed group ICA model
Simulation studies
A fMRI data example
Summary
Source Separation in fMRI studies

Goal: Decompose observed fMRI data as a linear combination of spatio-temporal processes of underlying source signals.
Single-subject ICA

Classic noise-free ICA

\[ Y = AS, \]  \hspace{1cm} (1)

- \( Y(T \times V) \) is the observed fMRI data.
- \( A(T \times q) \) is the mixing matrix. Each column represents the time series associated with an IC.
- \( S(q \times V) \) is statistically independent spatial source signals in its rows.

The distribution of the spatial source signal is non-Gaussian.

Probabilistic ICA (PICA):

\[ Y = AS + E \]

\( E(T \times V) \) noise term representing residual variability in the data that are not explained by the extracted ICs.
Schematic plot for Single-subject ICA

\[ Y = AV \]

Voxels

Statistically independent

\[ f(s_1, \ldots, s_q) = f(s_1) \times \cdots \times f(s_q) \]
Existing group ICA methods

- Temporal-concatenation group ICA (TC-GICA)
  - GIFT (Calhoun et al., 2001)
  - Tensor PICA model (Beckmann and Smith, 2005)
  - A general group PICA model (Guo and Pagnoni, 2008; Guo 2011)

Assume common IC spatial maps across subjects in group ICA decomposition.

- A hierarchical group PICA (Guo and Tang, 2013)
  Incorporate subject-specific random effects in IC spatial maps in decomposing multi-subject fMRI data.
The brain functional networks may differ depending on subjects’ clinical and demographical characteristics. For example, Greicius MD et al. (2007): increased default-mode network functional connectivity in subjects with major depression.

Existing approaches for assessing covariate effects in ICA

- Single-subject ICA approach (e.g. Greicius MD et al., 2007)
  - Run single-subject ICA on each subject
  - select IC in each subject that most closely matched to the functional network of interest
  - group analysis based on subjects’ best-matched IC images

- Existing group ICA:
  - Estimate subject-specific spatial maps: back-construction, dual-regression, or model-based estimation
  - group analysis based on subject-specific IC maps
A hierarchical group PICA regression model

The first-level model

\[ Y_i(v) = A_i s_i(v) + \gamma_i^{(1)}(v), \]

The second-level model

\[ s_i(v) = s(v) + \beta(v)' x_i + \gamma_i^{(2)}(v), \]

- **\( Y_i(v) \) \((T \times 1)\):** the observed fMRI data from subject \( i \) at voxel \( v \)
- **\( s_i(v) \) \((q \times 1)\):** spatial source signals for subject \( i \) at voxel \( v \)
- **\( A_i \) \((T \times q)\):** mixing matrix for subject \( i \)
- **\( \gamma_i^{(1)}(v) \):** residual variability in the observed data \( \sim N(0, \Sigma^{(1)}) \).
- **\( s(v) \) \((q \times 1)\):** statistically independent population spatial source signals,
- **\( x_i \) \((p \times 1)\):** \( p \) covariates of subject \( i \)
- **\( \beta(v) \) \((p \times q)\):** regression parameters of the \( p \) covariates for the \( q \) IC
- **\( \gamma_i^{(2)}(v) \):** subject random variability, \( \sim N(0, \Sigma^{(2)}) \).
Source Distribution Model

Based on the characteristics of fMRI signals, we model
\[ s(v) = [s_1(v), \ldots, s_q(v)]' \] with a mixture of Gaussian distributions,
\[ f(s_\ell(v); \varphi_\ell) = \sum_{j=1}^{m} \pi_{\ell j} g(s_\ell(v); \mu_{\ell j}, \sigma_{\ell j}^2) \]

where \( g(\cdot) \) is the Gaussian pdf.

Define \( z_\ell(v) \) as latent states variable for the \( \ell \)th IC at voxel \( v \) with
\[ p(z_\ell(v) = j) = \pi_{\ell j}. \]

Given \( z_\ell(v) \), we have
\[ p(s_\ell(v)|z_\ell(v) = j) = g(s_\ell(v); \mu_{\ell j}, \sigma_{\ell j}^2), \]

Let \( z(v) \) as \( q \times 1 \) latent states vector for the \( q \) ICs at voxel \( v \).
\[ z(v) = [z_1(v), \ldots, z_q(v)]', \]
Complete Log-likelihood Function

\[
\log \ell(\theta \mid Y, SI, S, Z) = \sum_{v=1}^{V} \left\{ \sum_{i=1}^{N} \left[ \log g \left( Y_i(v); A_i s_i(v), \Sigma^{(1)} \right) + \log g \left( s_i(v); s(v) + \beta(v)' x_i, \Sigma^{(2)} \right) \right] + \log g \left( s(v); \mu_{z(v)}, \Sigma_z^{(3)} \right) + \sum_{l=1}^{q} \log \pi_{lz(v)} \right\}
\]

where \( Y = \{Y_i(v)\} \), \( SI = \{s_i(v)\} \), \( S = \{s(v)\} \) and \( Z = \{z(v)\} \), \( \theta = \{\{A_i\}, \Sigma^{(1)}, \Sigma^{(2)}, \varphi\} \) where \( \varphi \) include parameters in the source distribution model.
An EM algorithm

- E-Step: we find the conditional expectation for the complete data log-likelihood,

\[
E_{S,SI,Z\mid Y, \hat{\theta}^{(k)}}[\ell(\theta \mid Y, S, SI, Z)]
\]

\[
= Q_1(\theta \mid \hat{\theta}^{(k)}) + Q_2(\theta \mid \hat{\theta}^{(k)}) + Q_2(\theta \mid \hat{\theta}^{(k)}) + Q_4(\theta \mid \hat{\theta}^{(k)}),
\]
The EM algorithm, Cont.

\[ Q_1(\theta | \hat{\theta}^{(k)}) = \sum_{v=1}^{V} \sum_{i=1}^{N} E_{S,S_i,Z|Y,\theta^{(k)}} \left[ \log g \left( Y_i(v); A_i s_i(v), \Sigma^{(1)} \right) \right] \]

\[ Q_2(\theta | \hat{\theta}^{(k)}) = \sum_{v=1}^{V} \sum_{i=1}^{N} E_{S,S_i,Z|Y,\theta^{(k)}} \left[ \log g \left( s_i(v); s(v) + \beta(v)'x_i, \Sigma^{(2)} \right) \right] \]

\[ Q_3(\theta | \hat{\theta}^{(k)}) = \sum_{v=1}^{V} E_{S,S_i,Z|Y,\theta^{(k)}} \left[ \log g \left( s(v); \mu_{z(v)}, \Sigma_{z(v)}^{(3)} \right) \right] \]

\[ Q_4(\theta | \hat{\theta}^{(k)}) = \sum_{v=1}^{V} \sum_{l=1}^{q} E_{S,S_i,Z|Y,\theta^{(k)}} \left[ \log \pi_{lz(v)} \right] \]
Derive the conditional distributions needed for $Q(\theta | \hat{\theta}^{(k)})$
$p(s_i|y, \theta), \ p(s|y, \theta), \ p(s,s_i|y, \theta)$ and $p(z|y, \theta))$,

- We first derive $p(s_i|z, y, \theta), \ p(s|z, y, \theta), \ p(s,s_i|z, y, \theta)$.
  Conditioning on $z(v)$, rewrite the source distribution model as,

$$s(v) = G_{Z(v)}\mu + \gamma^{(3)}(v), \quad \text{Third-level Model}$$

where $\mu = [\mu_{11}, \ldots, \mu_{qm}], \ G_{Z(v)}$ is $q \times mq$ binary indicator matrix, and $\gamma^{(3)}(v) \sim \text{MVN}(0, \Sigma^{(3)}_{Z(v)})$.

- We then derive $p(z|y, \theta)$

- Obtain $p(s_i|y, \theta), \ p(s|y, \theta), \text{ and } p(s,s_i|y, \theta)$. 
The EM algorithm, Cont.

- **M-Step**

*Parameter updates for the first-level model:*

\[
\hat{A}_i^{(k+1)} = \left\{ \sum_{v=1}^{V} Y_i(v) E \left( s_i(v) | Y(v); \hat{\theta}^{(k)} \right) \right\}
\]

\[
\left\{ \sum_{v=1}^{V} E \left( s_i(v) s_i(v)' | Y(v); \hat{\theta}^{(k)} \right) \right\}^{-1}
\]

\[
\hat{\Sigma}^{(1)(k+1)} = \frac{1}{NV} \sum_{v=1}^{V} \sum_{i=1}^{N} \left\{ Y_i(v) Y_i(v)' - 2 Y_i(v) E \left( s_i(v) | Y(v); \hat{\theta}^{(k)} \right) \hat{A}_i^{(k+1)} \right.
\]

\[
+ \hat{A}_i^{(k+1)} E \left( s_i(v) s_i(v)' | Y(v); \hat{\theta}^{(k)} \right) \hat{A}_i^{(k+1)'} \right\}
\]
For the second-level model, the covariate effects is updated by

\[
\hat{\beta}(v)^{(k+1)} = \left( \sum_{i=1}^{N} x_i x'_i \right)^{-1} \sum_{i=1}^{N} x_i \left[ E \left( s_i(v)'|Y(v); \hat{\theta}^{(k)} \right) - E \left( s(v)'|Y(v); \hat{\theta}^{(k)} \right) \right],
\]

while \( \Sigma^{(2)} = \text{diag}(\nu_1^2, \nu_2^2, \ldots, \nu_q^2) \) is updated by

\[
\hat{\nu}_l^{2(k+1)} = \frac{1}{NV} \sum_{v} \sum_{i=1}^{N} \left\{ E \left( s_{il}(v)^2|Y(v); \hat{\theta}^{(k)} \right) + E \left( s_l(v)^2|Y(v); \hat{\theta}^{(k)} \right) - 2E \left( s_{il}(v)s_l(v)|Y(v); \hat{\theta}^{(k)} \right) + \hat{\beta}_l(v)^{(k+1)'x_i x'_i \hat{\beta}_l(v)^{(k+1)}}\right. \\
\left. + 2 \left( E \left( s_l(v)|Y(v); \hat{\theta}^{(k)} \right) - E \left( s_{il}(v)|Y(v); \hat{\theta}^{(k)} \right) \right) x'_i \hat{\beta}_l(v)^{(k+1)} \right\},
\]

(2)
Parameter updates for the source distribution model:

$$\hat{\pi}_{lj}^{(k+1)} = \frac{1}{V} \sum_{v=1}^{V} p(z(v)_l = j|Y(v); \hat{\theta}^{(k)})$$

$$\hat{\mu}_{lj}^{(k+1)} = \frac{\sum_{v=1}^{V} p(z(v)_l = j|Y(v); \hat{\theta}^{(k)})E\left(s(v)_l|z(v)_l = j, Y(v); \hat{\theta}^{(k)}\right)}{V\hat{\pi}_{lj}^{(k+1)}}$$

$$\hat{\sigma}_{lj}^{2(k+1)} = \frac{\sum_{v=1}^{V} p(z(v)_l = j|Y(v); \hat{\theta}^{(k)})E\left(s(v)_l^2|z(v)_l = j, Y(v); \hat{\theta}^{(k)}\right)}{V\hat{\pi}_{lj}^{(k+1)}} - (\hat{\mu}_{lj}^{(k+1)})^2.$$
Steps for the model estimation

Step 1. Start with initial parameter values
\[ \hat{\theta}^{(0)} = (\hat{A}_i^{(0)}, \{\hat{\Sigma}^{(1,0)}\}, \{\hat{\Sigma}^{(2,0)}\}, \hat{\phi}^{(0)}, \beta^{(0)}) \].

Step 2. E-step: compute the conditional expectation function \( Q(\theta|\hat{\theta}^{(k)}) \).

Step 3. M-step: updating parameter estimates
\[ \hat{\theta}^{(k+1)} = \arg\max \limits_{\theta} Q(\theta|\hat{\theta}^{(k)}) \]
using the explicit solutions.

Step 4. Iterate between steps 2-3 until convergence.

Step 5. Estimate IC spatial maps based on \( \hat{\theta} \) for specified covariates \( x^* \).
\[ \hat{s}^* (v) = E(s(v)|Y, \hat{\theta}) + \hat{\beta}(v)'x^* \]
Prior to ICA, dimension reduction and whitening is performed,

\[ \tilde{Y}_i(v) = (\Lambda_{i,q} - \tilde{\sigma}^2_{i,q} I_q)^{-1/2} U_{i,q} Y_i(v), \]  

(3)

\( U_{i,q} \) and \( \Lambda_{i,q} \): the first \( q \) eigenvectors and eigenvalues from SVD of \( [Y_i(1), \ldots, Y_i(V)]_{T \times V} \).

The first-level model:

\[ \tilde{Y}_i(v) = \tilde{A}_i s_i(v) + \tilde{\gamma}^{(1)}_i(v), \]

where

- \( \tilde{Y}_i(v) = H_i Y_i(v) \), \( \tilde{A}_i(v) = H_i A_i(v) \), \( \tilde{\gamma}^{(1)}_i(v) = H_i \gamma^{(1)}_i(v) \)

- \( H_i = (\Lambda_{i,q} - \tilde{\sigma}^2_{i,q} I_q)^{-1/2} U_{i,q} \): transformation matrix
Simulation Studies

We compare the performance of our method (Covariate ICA) vs. an existing method (Dual regression) under the following simulation settings

- simulate fMRI data from $q = 3$ source signals and consider sample sizes of $n = 10, 20, 40$ subjects
- for each IC: 3D population spatial maps $(25 \times 25 \times 4)$ and time course of length $T = 200$
- simulate two covariates $\mathbf{X} = (X_1, X_2), X_1 \sim U(-1, 1), X_2 \sim \text{Bernoulli}(0.5)$
- multi-subject fMRI data with low and high between-subject variability in spatial distributions of source signals
- simulate subject-specific time courses for each source signal
Simulation Study: Group Level IC maps

Group IC (True)  Group IC (Cov.ICA.)  Group IC (Dual.Reg.)
Simulation Study: Covariate Effect Maps for $X_1$

- $\beta(v)$ (True)
- $\beta(v)$ (Cov.ICA.)
- $\beta(v)$ (Dual.Reg.)
Simulation Study: Covariate Effect Maps for $X_2$

$\beta(v)$(True)  

$\beta(v)$(Cov.ICA.)  

$\beta(v)$(Dual.Reg.)
Simulation Studies

<table>
<thead>
<tr>
<th>Btw-subj Var</th>
<th>Population-level spatial maps Corr(SD)</th>
<th>Covariate Effects MSE(SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N=10</td>
<td>0.982(0.003)</td>
<td>0.953(0.017)</td>
</tr>
<tr>
<td>N=20</td>
<td>0.990(0.002)</td>
<td>0.955(0.002)</td>
</tr>
<tr>
<td>N=40</td>
<td>0.992(0.002)</td>
<td>0.952(0.003)</td>
</tr>
</tbody>
</table>

\(MSE(\hat{\beta})\) is defined as \(\frac{1}{V} E(\sum_{v=1}^{V} \|\hat{\beta}(v) - \beta(v)\|_F)\)
An Approximated EM algorithm

- The proposed exact EM algorithm:
  - Pros: provide explicit solutions for E-step and M-step
  - Cons: this exact estimation becomes computational expensive when the number of ICs increases.

- We propose an approximation EM algorithm for faster computation. It is derived based on the fact that fMRI source signals are generally sparse and well-separated in distributed patterns.
Simulation Studies: the Exact EM vs. the Approximated EM

<table>
<thead>
<tr>
<th>Num. of</th>
<th>Population-level spatial maps</th>
<th>Covariate Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>IC</td>
<td>Corr(SD)</td>
<td>MSE(SD)</td>
</tr>
<tr>
<td></td>
<td>Exact EM</td>
<td>Approx EM</td>
</tr>
<tr>
<td>q=3</td>
<td>0.981(0.003)</td>
<td>0.981(0.001)</td>
</tr>
<tr>
<td>q=6</td>
<td>0.980(0.006)</td>
<td>0.980(0.006)</td>
</tr>
<tr>
<td>q=9</td>
<td>0.974(0.019)</td>
<td>0.970(0.018)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Num. of</th>
<th>Computation Time (min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>IC</td>
<td>Exact EM</td>
</tr>
<tr>
<td>q=3</td>
<td>9.91</td>
</tr>
<tr>
<td>q=6</td>
<td>71.05</td>
</tr>
<tr>
<td>q=9</td>
<td>523.45</td>
</tr>
</tbody>
</table>
A Data Example

A Zen meditation study

- 24 subjects: Zen meditation group (n=12) and control group (n=12).
- The two groups were matched for gender, age, and education level.
- A series of 520 functional MRI images were acquired with a 3.0 Tesla Siemens Magnetom Trio scanner, each containing $53 \times 63 \times 46$ voxels.
- Word or nonword items presented visually in random order; press a button with the left hand (index finger = ”yes”, middle finger = ”no”); focus on breathing in other time points.
- We fit the proposed ICA regression model with a covariate indicating group membership: $x = 1$ for meditator and $x = 0$ for controls.
**Figure:** Model-based estimation of subgroup spatial maps for functional networks

- **Task-related network**
  - Control: Temporal corr. with task time series: 0.78
  - Meditator: Temporal corr. with task time series: 0.86

- **Default mode network**
  - -6 19 30
  - Temporal corr. with task time series: 0.86
We propose a hierarchical group PICA regression model

- The proposed model provides a formal statistical framework for modeling covariate effects in group ICA.
- Our model can provide model-based estimation of spatial distributed patterns of brain functional networks based on subjects’ demographic and clinical characteristics.
- We propose an exact EM algorithm and an approximation EM for model estimation.
- Simulation results show that our proposed model has better performance than the existing method.
Acknowledgements

- Giuseppe Pagnoni, PhD. Department of Biomedical Sciences and Technologies, University of Modena and Reggio Emilia, Modena, Italy
- Emory University URC grant and NIH grant R01-MH079251.
References


