

# Toolbox ‘BSBM’

This manual aims to show detailed guidelines on using our toolbox, including the usages of primary functions, our settings of initial values, and a control panel for reproducing our simulation results.

---

BSBM\_ContPnl

*An example of toolbox ‘BSBM’*

---

## Description:

The Control Panel of Our Proposed Bayesian Method with a SBM-motivated Prior (BSBM)

This control panel serves as an example of analyzing data using our toolbox. It is written as a function which can be easily modified to use on computing clusters. Users can use this control panel to reproduce the results of our simulation study (Section 3 in our paper) and modify this file to analyze their data without difficulty.

## Details:

This control panel shows a step-by-step procedure to analyze the simulated data we used in Section 3.1. The two simulated datasets we used in Section 3.2 can be analyzed by only changing the file names in the control panel. All the simulated data are provided under the subfolder *Simulated\_data*:

- Simulation1.mat: The simulated data we used in Section 3.1;
- Simulation2\_1.mat: The simulated data we used in Section 3.2, with 2714 time points;
- Simulation2\_2.mat: The simulated data we used in Section 3.2, with 1000 time points.

Each data contains two variables:

- **Y**: The simulated data, which is a real  $d \times T$  matrix.  $d$  is the number of channels and  $T$  is the number of time points. For example, in `Simulation1.mat`, **Y** is a  $50 \times 1000$  matrix;
- **A\_true**: The true network structure, which is a binary  $d \times d$  matrix. “**A\_true**( $i,j$ ) = 1” indicates that there is a directional connection from channel  $j$  to channel  $i$ . For example, in `Simulation1.mat`, **A\_true** is a  $50 \times 50$  matrix.

The systems and the error structures we used to generate these simulated datasets are described in Section 3.1 and Section 3.2 of our paper in detail.

The control panel also includes the calculation of the clustering probabilities and the network edge probabilities. The analyses based on the clustering probabilities and the network edge probabilities, including generating figures, are described in Section 2.4, Section 3, and Section 4 of our paper in detail.

---

BSBM\_EM

*EM Algorithm for Setting Starting Values and Hyperparameter of the Partially Collapsed Gibbs Sampler*

---

### Description:

This is the primary function of the EM algorithm, which is used to set starting values and hyperparameter for the MCMC algorithm.

### Usage:

```
[Gam_all, A_all, C_all, R_all, mu_all, m_all, B_all]
= BSBM_EM(Y, mite, l0, u0, e0, xi_sq, xi_sq2, A0, C0, mu0, R0, m0, B0, Gam0, b_sq)
```

### Input:

|             |   |
|-------------|---|
| <b>Y</b>    | The observed data <b>Y</b> , a real $d \times T$ matrix, where $d$ is the number of channels and $T$ is the number of time points |
| <b>mite</b> | The max number of iterations  |

|                                      |   |
|--------------------------------------|---|
| <code>l0,u0</code>                   | Hyperparamters in $\mathbf{B}$ 's prior   |
| <code>e0</code>                      | Hyperparameter in $\mathbf{R}$ 's prior   |
| <code>xi_sq</code>                   | Hyperparameter in $\mathbf{A}$ 's prior   |
| <code>xi_sq2</code>                  | Hyperparameter in $\mathbf{C}$ 's prior and $\boldsymbol{\mu}$ 's prior   |
| <code>A0,C0,mu0,R0,m0,B0,Gam0</code> | The initial values of $\mathbf{A}, \mathbf{C}, \boldsymbol{\mu}, \mathbf{R}, \mathbf{m}, \mathbf{B}, \boldsymbol{\Gamma}$ |
| <code>b_sq</code>                    | Fixed at 1 in current version of BSBM   |

### Output:

`Gam_all, A_all, C_all, R_all, mu_all, m_all, B_all` record the updated values of  $\boldsymbol{\Gamma}, \mathbf{A}, \mathbf{C}, \mathbf{R}, \boldsymbol{\mu}, \mathbf{m}, \mathbf{B}$  at each iteration.

### Details:

The outputs of BSBM\_EM contain the values of parameters at every iteration. We use the updated values at the final iteration to set the starting values and hyperparameter of the MCMC algorithm.

Other functions under the subfolder *EM* (BSBM\_Kalman, BSBM\_EM\_update\_A\_Gam, BSBM\_EM\_update\_B, BSBM\_EM\_update\_C, BSBM\_EM\_update\_m, BSBM\_EM\_update\_mu, BSBM\_EM\_update\_R) are the updating steps in the EM algorithm.

---

BSBM\_MCMC

*MCMC Simulation Algorithm for Posterior Inference*

---

### Description:

This is the primary function of the MCMC algorithm, which is used for posterior inference of our proposed Bayesian hierarchical model.

### Usage:

[A\_all, C\_all, X, V, Q, R\_all, mu\_all, P\_all, m\_all, B\_all, Gam\_all]  
= BSBM\_MCMC(Y, mcmc\_mite, l0, u0, e0, xi\_sq, xi\_sq2, X0,  
A0, C0, mu0, Q0,R0, V0, P0, m0, B0, Gam0)

### Input:

|                            |   |
|----------------------------|---|
| Y                          | The observed data $\mathbf{Y}$ , a real $d \times T$ matrix, where $d$ is the number of channels and $T$ is the number of time points |
| mcmc_mite                  | The max number of iterations  |
| l0,u0                      | Hyperparamters in $\mathbf{B}$ 's prior   |
| e0                         | Hyperparameter in $\mathbf{R}$ 's prior   |
| xi_sq                      | Hyperparameter in $\mathbf{A}$ 's prior   |
| xi_sq2                     | Hyperparameter in $\mathbf{C}$ 's prior and $\boldsymbol{\mu}$ 's prior   |
| X0                         | The initial values of the underlying neuronal states $\mathbf{X}$   |
| A0,C0,mu0,R0,P0,m0,B0,Gam0 | The initial values of $\mathbf{A}, \mathbf{C}, \boldsymbol{\mu}, \mathbf{R}, \mathbf{P}, \mathbf{m}, \mathbf{B}, \boldsymbol{\Gamma}$ |
| Q0,V0                      | Fixed at the identity matrices in current version of BSBM   |

### Output:

Gam\_all, A\_all, C\_all, R\_all, mu\_all, P\_all, m\_all, B\_all record the simulated values of  $\boldsymbol{\Gamma}, \mathbf{A}, \mathbf{C}, \mathbf{R}, \boldsymbol{\mu}, \mathbf{P}, \mathbf{m}, \mathbf{B}$  at each iteration. X record the simulated values of  $\mathbf{X}$  at the last iteration. V, Q are fixed at the identity matrices in the current version of BSBM.

### Details:

The outputs of BSBM\_MCMC contain the simulated markov chains of all parameters. We use the simulated values of  $\mathbf{m}$  and  $\boldsymbol{\Gamma}$  to calculate the clustering probabilities and the network edge probabilities, which we defined in Secion 2.4 in our paper.

Other functions under the subfolder *MCMC* (BSBM\_MCMC\_Update\_A, BSBM\_MCMC\_update\_B, BSBM\_MCMC\_update\_C, BSBM\_MCMC\_update\_Gam,

BSBM\_MCMC\_update\_m, BSBM\_MCMC\_update\_mu, BSBM\_MCMC\_update\_P, BSBM\_MCMC\_update\_Q, BSBM\_MCMC\_update\_R, BSBM\_MCMC\_update\_X) are the simulation steps in the MCMC simulation algorithm.

**Reference:**

Huazhang Li, Yaotian Wang, Guofen Yan, Ying Sun, Seiji Tanabe, Chang-Chia Liu, Mark S. Quigg, Tingting Zhang (2020). A Bayesian State-Space Approach to Mapping Directional Brain Networks, *Journal of the American Statistical Association*. Under Review.