Sparse Optimization in Exploring Brain Networks

Jitkomut Songsiri

Chula Engineering

Joint China-Cuba Laboratory for Frontier Research
in Translational Neurotechnology

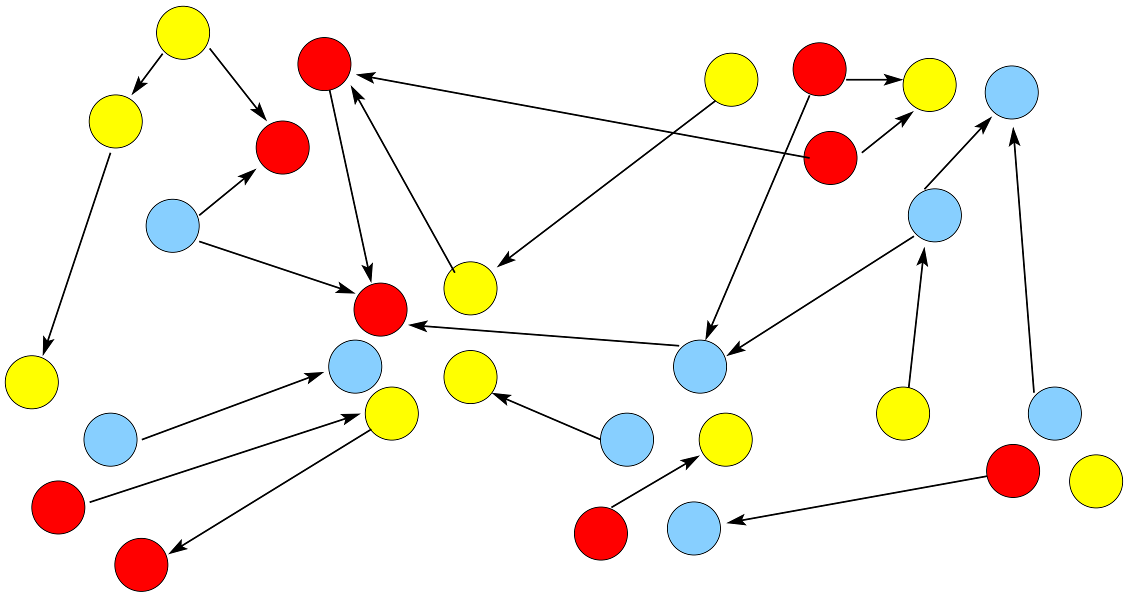
Chengdu, China

June 2-22, 2016

Outline

- **Brain connectivity**
- Granger graphical models
  - Learning a single Granger graph
  - Learning multiple Granger graphs having a common structure
  - Learning multiple Granger graphs having a partially common structure
  - Numerical examples
- Structural Equation Modeling

Brain connectivity



- a brain connectivity or a brain network is represented by a graph
- nodes represents voxels (or ROIs)
- a brain connectivity is explained by the graph topology
- the graph topology is described by a statistical dependence measure of interest

Dependence Measures

- data are treated as independent samples (no temporal consideration)
  - correlation (covariance matrix)
  - partial correlation (inverse of covariance matrix)
  - structural equation modeling (path coefficient matrix)

- data are treated as time series
  - cross coherence function (normalized correlation function)
  - partial coherence function (normalized inverse of correlation) – often done in frequency domain (inverse spectrum)
  - dynamical causal modeling (coupling matrices)
  - Granger causality (autoregressive coefficients)

Inference methods

estimation: covariance matrix, path coefficient of SEM, correlation function, spectrum, matrices in DCM, AR coefficients, etc.

inference methods can be roughly divided into:

- statistical tests (test if each of these measures is zero with a statistical significance)
- sparse estimation (estimation formulation promotes sparsity in these measures)

we pursue the latter approach by cooperating an ℓ_1-minimization in the formulation

goal: learning a zero pattern in an dependence measure

Outline

- Brain connectivity
- **Granger graphical models**
  - Learning a single Granger graph
  - Learning multiple Granger graphs having a common structure
  - Learning multiple Granger graphs having a partially common structure
  - Numerical examples
- Structural Equation Modeling

Granger causality (Granger 1969)

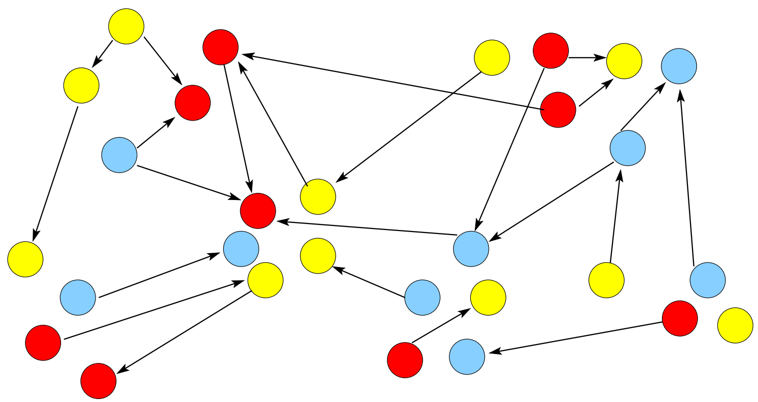
sparsity in coefficients A_k

$$(A_k)_{ij} = 0, \quad \text{for } k = 1, 2, \ldots, p$$

is the characterization of **Granger causality** in AR model:

$$y(t) = A_1 y(t-1) + A_2 y(t-2) + \cdots + A_p y(t-p) + \nu(t)$$

- y_i is not *Granger-caused* by y_j
- knowing y_j does not help to improve the prediction of y_i



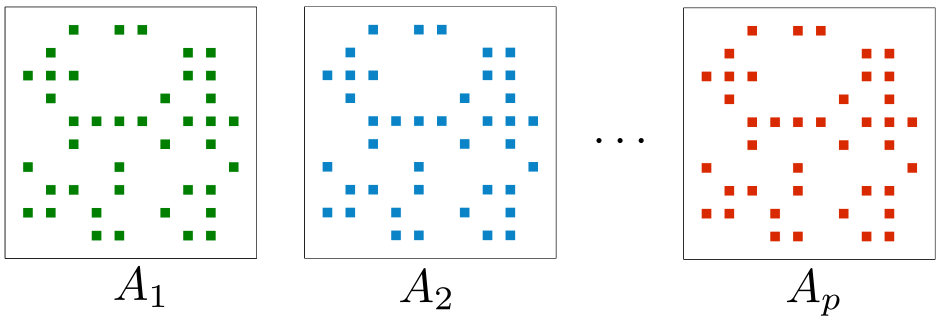
applications in neuroscience and system biology

(Salvador et al. 2005, Valdes-Sosa et al. 2005, Fujita et al. 2007, ...)

Learning a single Granger Graphical Model (J. Songsiri 2013)

Problem: find A_k's that minimize the mean-squared error and

- A_k's contain many zeros
- common zero locations in $A_1, A_2, \ldots, A_p$

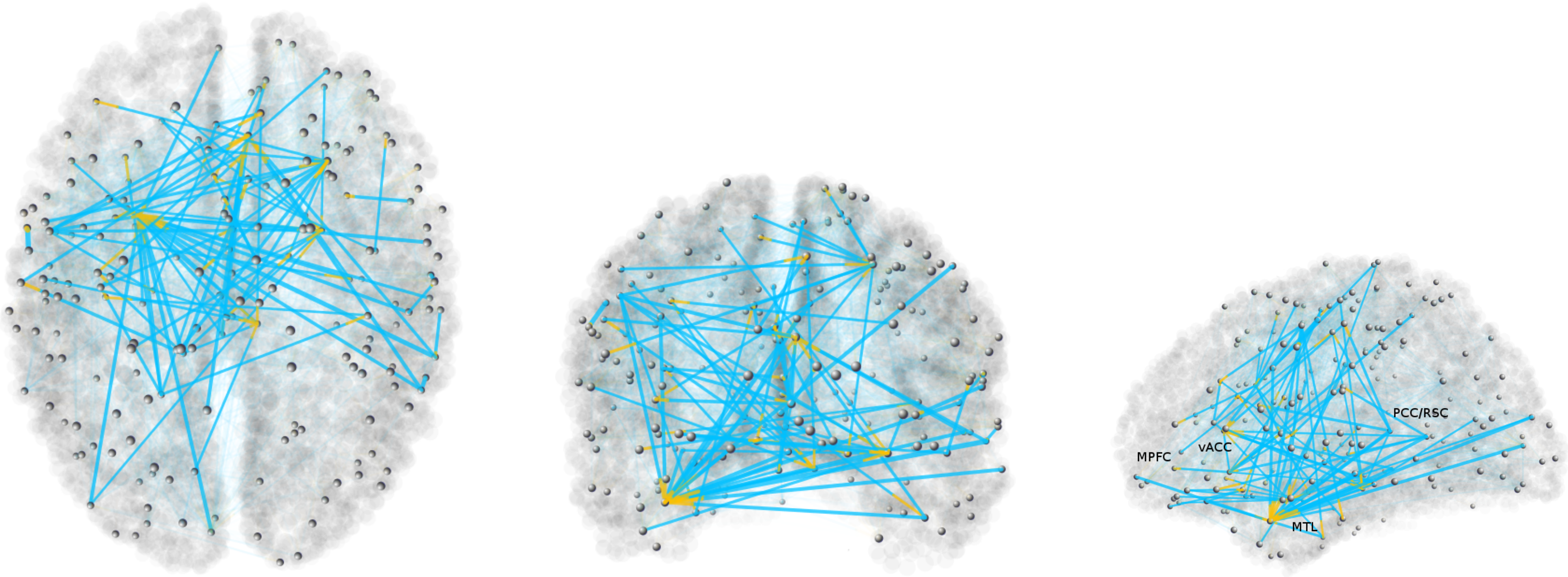


Formulation: least-squares with sum-of-ℓ_2-norm regularization

$$\min_A \ (1/2)\|Y - AH\|_2^2 + \lambda \sum_{i \neq j} \| \begin{bmatrix} (A_1)_{ij} & (A_2)_{ij} & \cdots & (A_p)_{ij} \end{bmatrix} \|_2$$

the problem falls into the framework of **Group Lasso**

Default Mode Network (A. Pongrattanakul, P. Lertkultanon and J. Songsiri 2013)

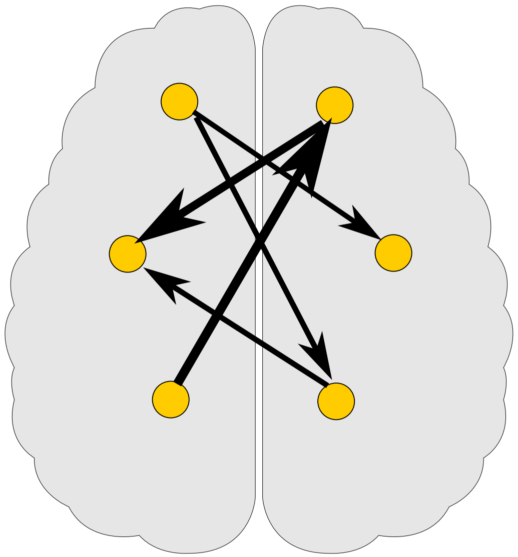


- many active nodes in vACC, MTLs and a few in MPFC and PCC/RSC
- strong connections between MTLs and PCC and vACC has a significant connectivity with PCC
- strong connections between left and right medial temporal lobes

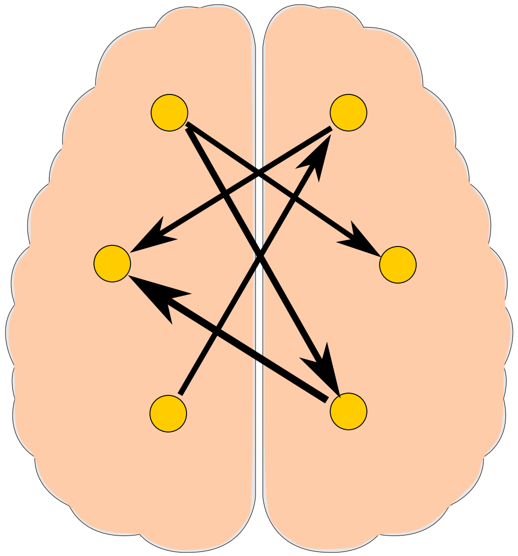
Outline

- Brain connectivity
- Granger graphical models
  - Learning a single Granger graph
  - **Learning multiple Granger graphs having a common structure**
  - Learning multiple Granger graphs having a partially common structure
  - Numerical examples
- Structural Equation Modeling

Application on learning a common brain structure



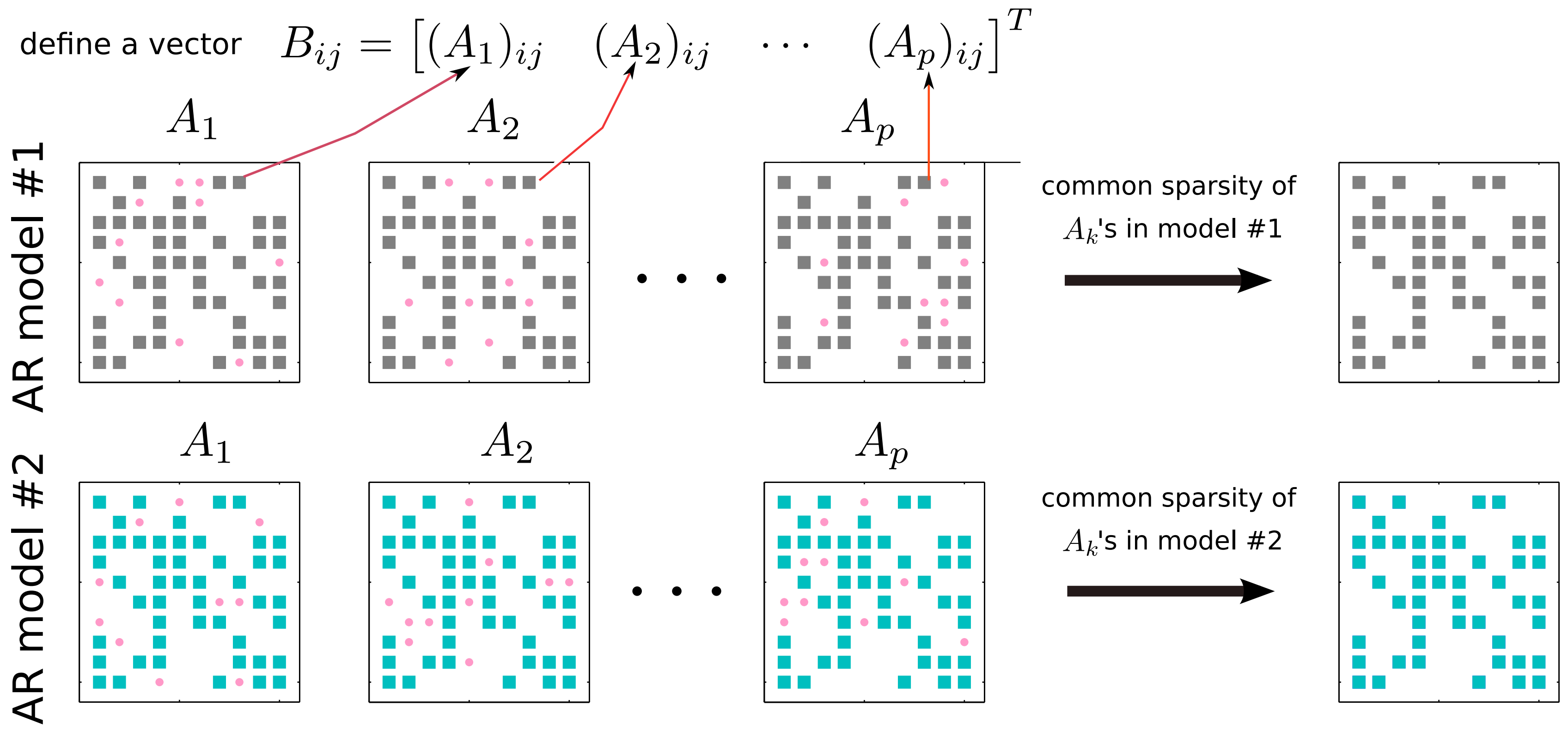
Brain of subject A



Brain of subject B

- brains of subjects under the same condition from a homogenous group are assumed to have a common brain connectivity
- the connection strengths of a pair of nodes from group subjects can be different

Common Granger causality of multiple AR models



- common sparsity of A_k's in each model defines its Granger causality structure
- our goal is to learn a **common** Granger causality structure among all models

Estimation Formulation

jointly estimate K AR models to have a **common** Granger-causality structure

reorder the variables

$$B_{ij}^{(k)} = \left[(A_1^{(k)})_{ij} \quad (A_2^{(k)})_{ij} \quad \cdots \quad (A_p^{(k)})_{ij} \right]^T, \quad C_{ij} = (B_{ij}^{(1)}, B_{ij}^{(2)}, \ldots, B_{ij}^{(K)})$$

optimization problem:

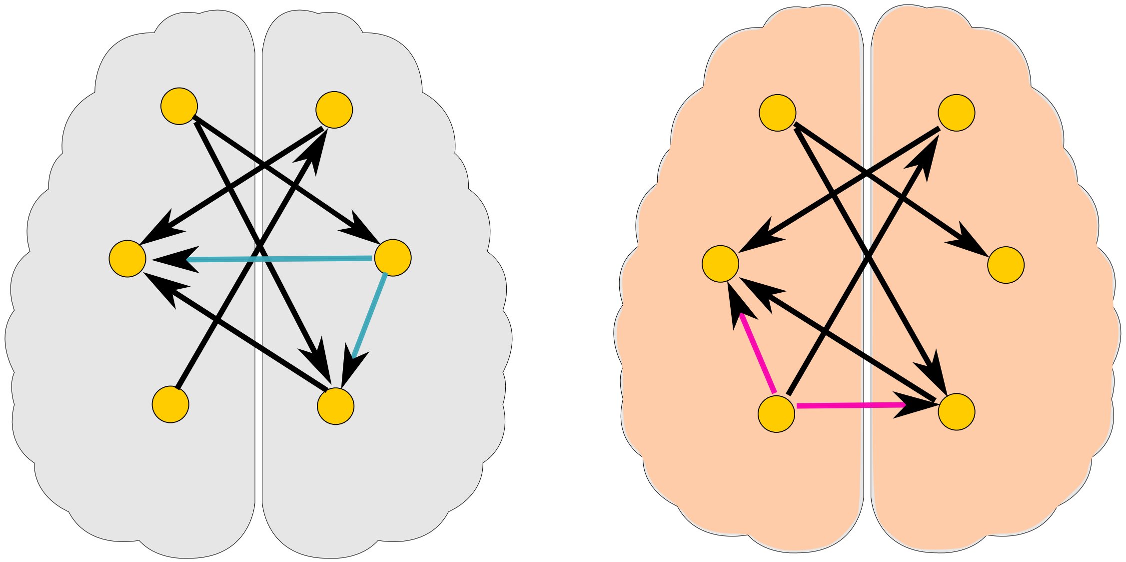
$$\underset{A^{(1)},\ldots,A^{(K)}}{\text{minimize}} \; (1/2) \sum_{k=1}^{K} \|Y^{(k)} - A^{(k)} H^{(k)}\|_F^2 + \lambda \sum_{i \neq j} \|C_{ij}\|_2$$

- the superscript $^{(k)}$ denotes the kth model
- 1st term: least-squares error of K models
- 2nd term: promote a common sparsity in all models

Outline

- Brain connectivity
- Granger graphical models
  - Learning a single Granger graph
  - Learning multiple Granger graphs having a common structure
  - **Learning multiple Granger graphs having a partially common structure**
  - Numerical examples
- Structural Equation Modeling

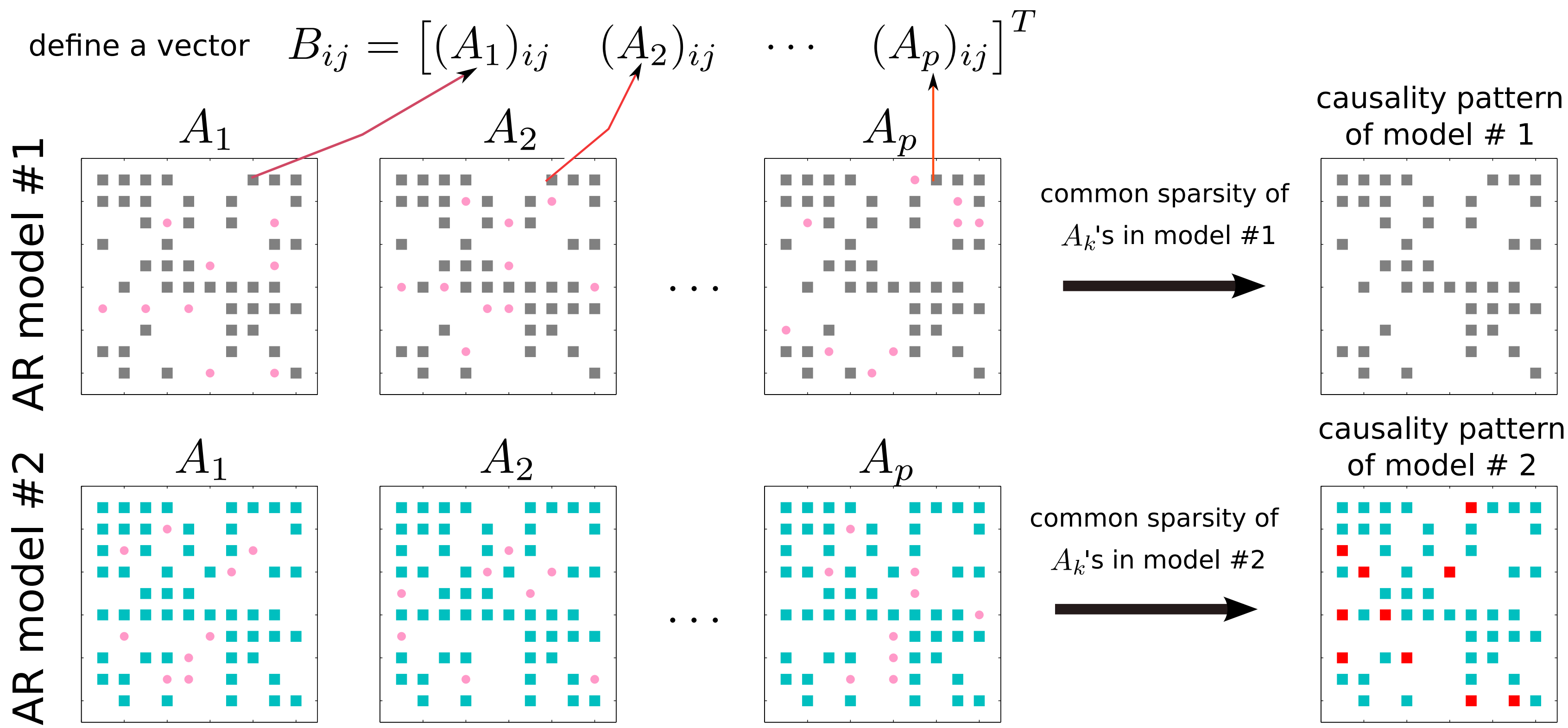
Application on classifying brain conditions



Brain under condition A   Brain under condition B

- brain under two conditions may share some **similar** connectivity patterns due to some normal functioning of the brain
- different conditions of the brain may lead to some **different** edges in the brain connectivity

Granger causality of multiple AR models



- common sparsity of A_k's in each model defines its Granger causality structure
- our goal is to learn similar Granger causality structures among all models

Estimation Formulation (J. Songsiri 2015)

jointly estimate K AR models to have **similar** Granger-causality structures

$$\underset{A^{(1)},\ldots,A^{(K)}}{\text{minimize}} \sum_{k=1}^{K} \frac{1}{2}\|Y^{(k)} - A^{(k)}H^{(k)}\|_2^2 + \lambda_1 \sum_{i\neq j}\sum_{k=1}^{K} \left\| B_{ij}^{(k)} \right\|_2 + \lambda_2 \sum_{i\neq j}\sum_{k=1}^{K-1} \left\| B_{ij}^{(k+1)} - B_{ij}^{(k)} \right\|_2$$

- the superscript $^{(k)}$ denotes the kth model
- $B_{ij}^{(k)} = \begin{bmatrix} (A_1^{(k)})_{ij} & (A_2^{(k)})_{ij} & \cdots & (A_p^{(k)})_{ij} \end{bmatrix}^T \in \mathbf{R}^p$
- 1st term: least-squares error of K models
- 2nd term: promote a sparsity in each model
- 3rd term: promote similarity in any two consecutive models
- a least-squares problem with sum-of-ℓ_2-norm regularization

Group Fused Lasso framework

the estimation problem can be regarded as a **Group Fused Lasso** problem

$$\underset{x}{\text{minimize}} \quad (1/2)\|Gx - b\|_2^2 + \lambda_1 \|\mathcal{P}x\|_{2,1} + \lambda_2 \|\mathcal{D}x\|_{2,1}$$

with variable $x \in \mathbf{R}^n$

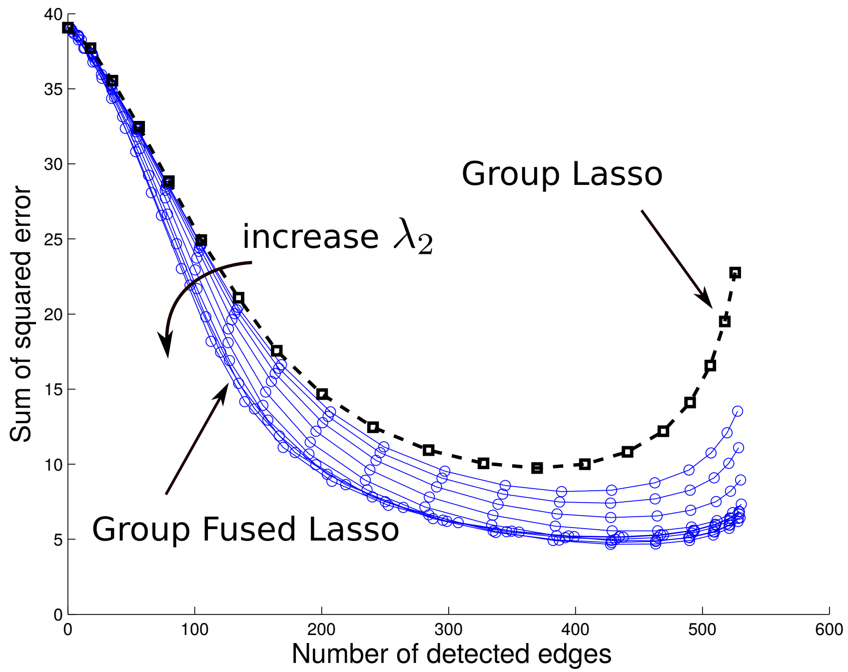
- $G \in \mathbf{R}^{m \times n}, b \in \mathbf{R}^m, \mathcal{P} \in \mathbf{R}^{s \times n}, \mathcal{D} \in \mathbf{R}^{r \times n}$ are problem parameters
- sum of 2-norm: $\|z\|_{2,1} = \sum_{k=1}^{L} \|z_k\|_2$
- $\mathcal{D}$ is a kronecker product of a projection and the forward difference matrix
- if $\lambda_2 = 0$ and $\lambda_1 > 0$, it reduces to a **group lasso** problem
- if $\lambda_2 > 0$ and $\lambda_1 = 0$, it is a class of **total variation regularized** problem

Outline

- Brain connectivity
- Granger graphical models
  - Learning a single Granger graph
  - Learning multiple Granger graphs having a common structure
  - Learning multiple Granger graphs having a partially common structure
  - **Numerical examples**
- Structural Equation Modeling

Numerical examples

generate 3 sparse AR models having similar Granger structures



Group Lasso:

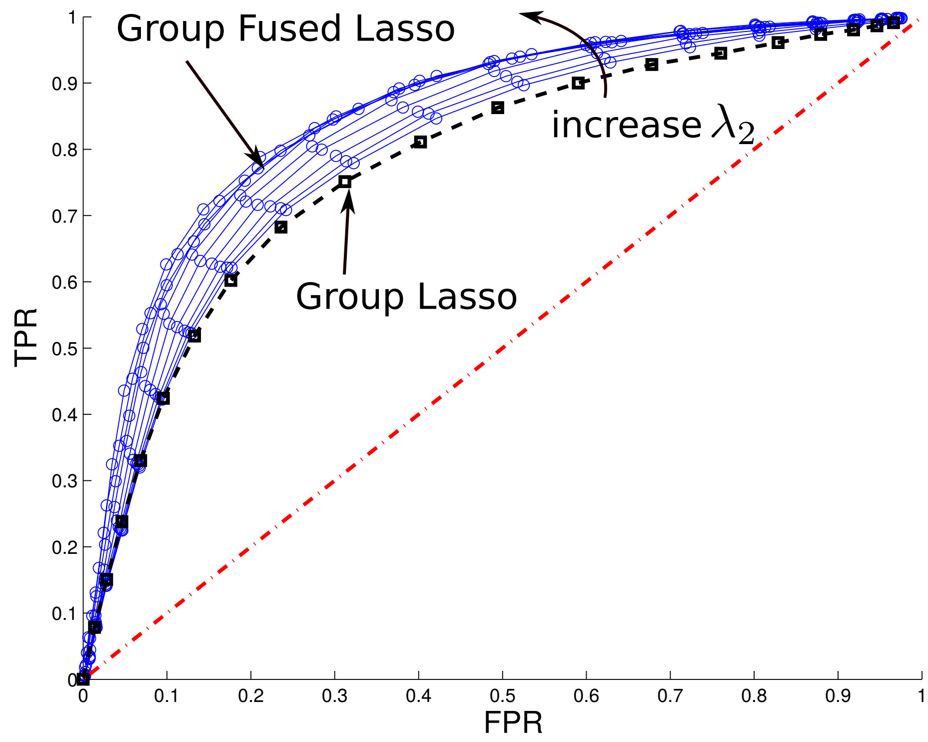
seperately estimate 3 models

Group Fused Lasso:

jointly estimate 3 models

- model errors are increasing as the estimated Granger network is too dense
- Group Fused Lasso yields a lower model error as λ_2 increases

ROC curves of Group Lasso VS Group Fused Lasso



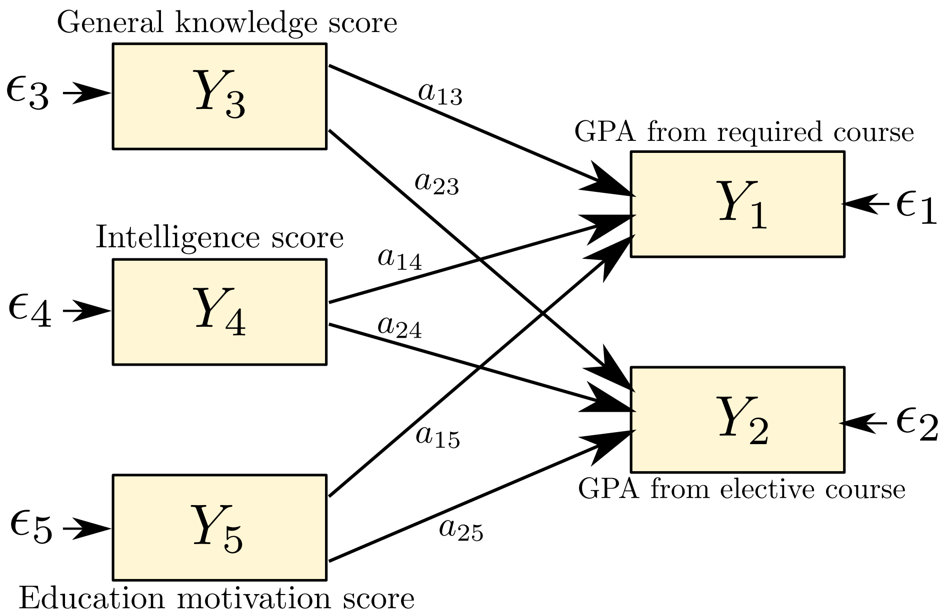
- at a fixed FPR (false positive rate), Group Fused Lasso yields a higher TPR (true positive rate) than Group Lasso
- obtain more accurate Granger structure as λ_2 increases

Outline

- Brain connectivity
- Granger graphical models
- **Structural Equation Modeling**
  - path analysis
  - confirmatory VS exploratory SEM
  - sparse SEM for exploratory SEM

Structural Equation Modeling (SEM)

path analysis is a special SEM that includes **only the observed variables**



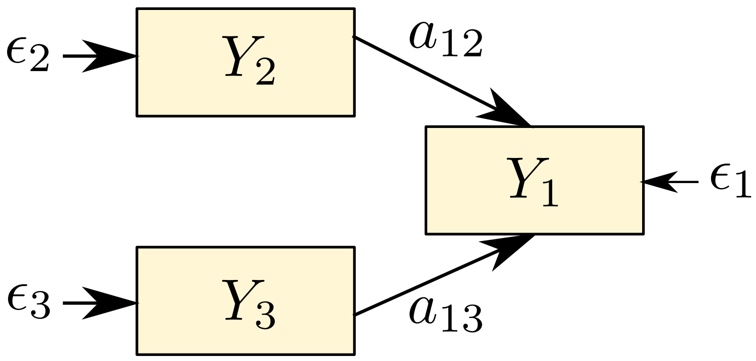
path analysis model

$$Y = AY + \epsilon$$

(multiple linear regression)

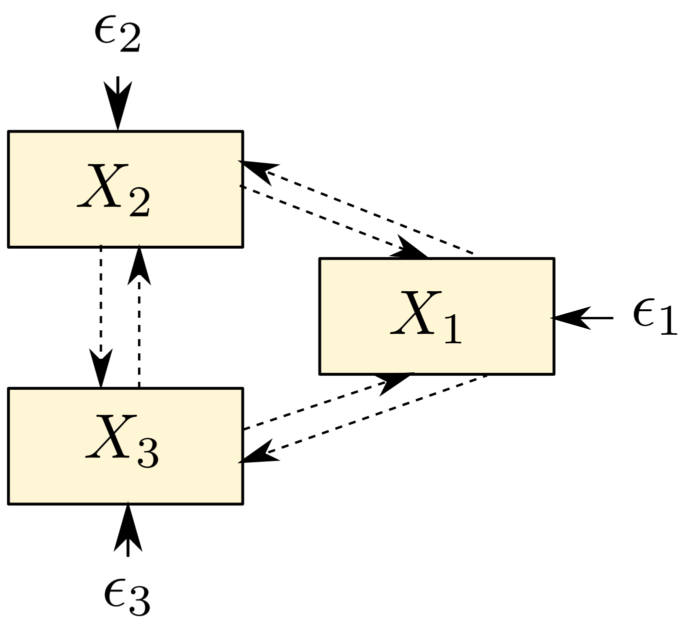
- ϵ is the model error, Y is the variable vector
- A is called the **path matrix** or path coefficient
- entries in the path matrix (a_{ij}) denotes a **causal relation** from Y_j to Y_i

Two important problems in path analysis



confirmatory SEM

- causal relationship is given
- zero pattern in A is given



exploratory SEM

- to explore a causal relationship among variables
- explore a zero pattern in A

Problem description

given samples of Y, we can compute the sample covariance S

the covariance of Y derived from $Y = AY + \epsilon$ is

$$\Sigma = (I - A)^{-1}\Psi(I - A)^{-T}, \qquad \text{where } \Psi = \mathbf{cov}\,(\epsilon)$$

goal: estimate Σ, Ψ and A so that Σ is close to S

in the sense that

$$d(S, \Sigma) = \log\det\Sigma + \mathbf{tr}(S\Sigma^{-1}) - \log\det S - n,$$

(Kullback-Leibler divergence function) is minimized

Sparse SEM (A. Pruttiakaravanich and J. Songsiri 2016)

a convex formulation for exploratory SEM

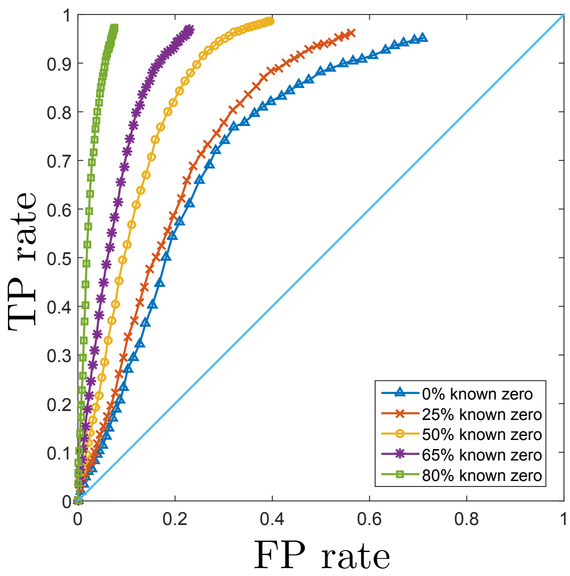
$$\begin{array}{ll} \underset{X,A,\Psi}{\text{minimize}} & -\log\det X + \mathbf{tr}(SX) + 2\gamma \sum_{(i,j)\notin I_A} |A_{ij}| \\ \text{subject to} & \begin{bmatrix} X & (I-A)^T \\ I-A & \Psi \end{bmatrix} \succeq 0, \\ & 0 \preceq \Psi \preceq \alpha I \\ & P(A) = 0 \end{array}$$

- α is a parameter representing a bound on covariance error
- $P(A)$ is a linear mapping giving the *prior* zero constraint in A, noted by the index set I_A
- $\sum_{(i,j)\notin I_A} |A_{ij}|$ is the ℓ_1-norm regularization to promote zeros in A
- if the optimal X has low rank, then Σ can be retrieved from $\Sigma = X^{-1}$

Effect of the percentage known zero in A

to see the effect of percentage known zero in A_{true}

- generate A_{true} with $n = 20$ and sparsity 10%.
- generate $S = (I - A_{\text{true}})^{-1}\Psi(I - A_{\text{true}})^{-T}$, $\Psi = 0.1I$
- solve sparse SEM by assuming that we know locations of zero in A_{true} about $0\%, 25\%, 50\%, 65\%, 80\%$



- knowing more correct zero structure in A_{true} provides the better accuracy of our learning causal structure method
- knowing 0% zero in A_{true} (underdetermined problem) provides poor estimation result since we may not obtain a true solution

Summary

- we have proposed a Group Fused Lasso formulation for estimating jointly multiple sparse AR models
- the formulation uses a sum of 2-norm penalty on the differences between consecutive AR models
- it finds applications in exploring a common structure of time series belonging to different/common classes
- we also proposed a convex formulation for learning causal pattern in SEM
- it finds applications in exploring causal relationships among static variables
- the problem is a type of ℓ_1-regularized estimation, can be solved by a convex solver
- the accuracy of learning the true network depends on the selection of the regularization parameter

(Selected) References

- J. Songsiri, "Learning multiple granger graphical models via group fused lasso," in *Proceedings of the 10th Asian Control Conference (ASCC)*, 2015.
- J. Songsiri, "Sparse autoregressive model estimation for learning Granger causality in time series, in *Proceedings of the 38th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2013, pp. 3198–3202.
- C. M. Alaìz, A. Barbero, and J. R. Dorronsoro, "Group fused lasso, *Artificial Neural Networks and Machine Learning*, pp. 66–73, 2013.
- A. Pruttiakaravanich and J. Songsiri, "A Convex Formulation for Path Analysis in Structural Equation Modeling," *To Appear in Proceeding of SICE conference 2016*.