# Learning Multiple Granger Graphical Models via Group Fused Lasso

**Jitkomut Songsiri**

Department of Electrical Engineering

Chulalongkorn University

Asian Control Conference (ASCC), May 31- June 3, 2015

Kota Kinabalu, Malaysia

# Outline

- **Granger graphical models**
- Learning multiple Granger graphical models
- Algorithm
- Numerical examples

# Granger causality (Granger 1969)

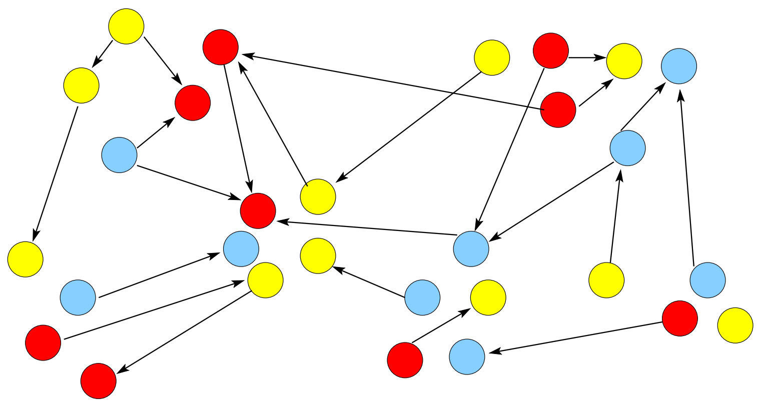
sparsity in coefficients $A_k$

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

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 $y_i$



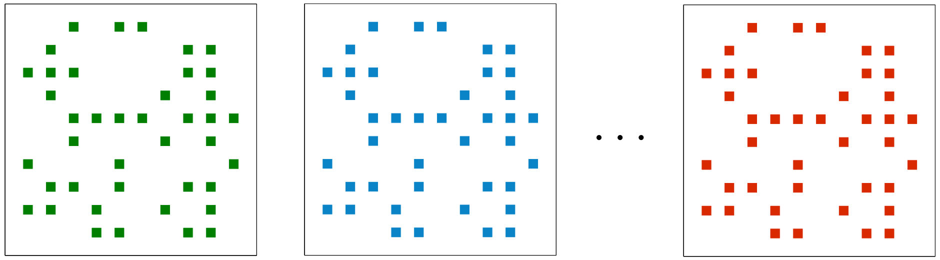
applications in neuroscience and system biology

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

## Learning Granger Graphical Models (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 $A_1, A_2, \ldots, A_p$



$A_1$ $A_2$ $\cdots$ $A_p$

**Formulation:** least-squares with sum-of-$\ell_2$-norm regularization

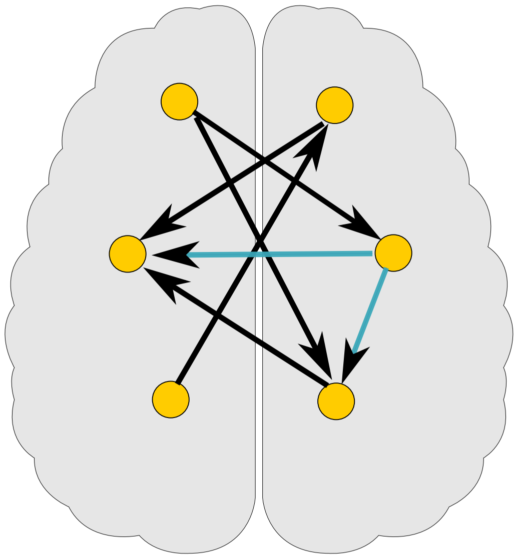
$$\min_A \; (1/2)\|Y - AH\|_2^2 + \lambda \sum_{i \neq j} \| \left[ (A_1)_{ij} \quad (A_2)_{ij} \quad \cdots \quad (A_p)_{ij} \right] \|_2$$

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

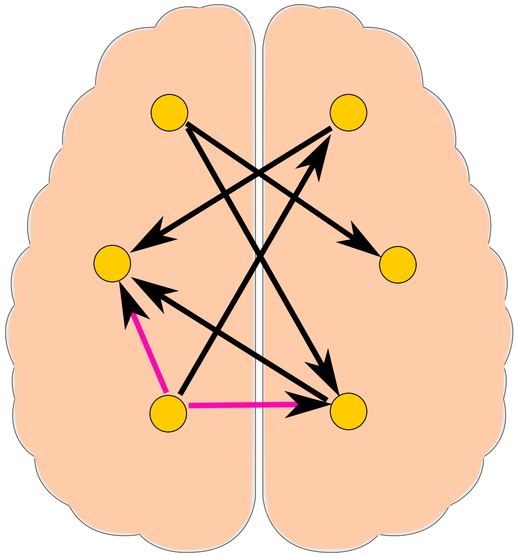
# Outline

- Granger graphical models
- **Learning multiple Granger graphical models**
- Algorithm
- Numerical examples

# Application on classifying brain conditions



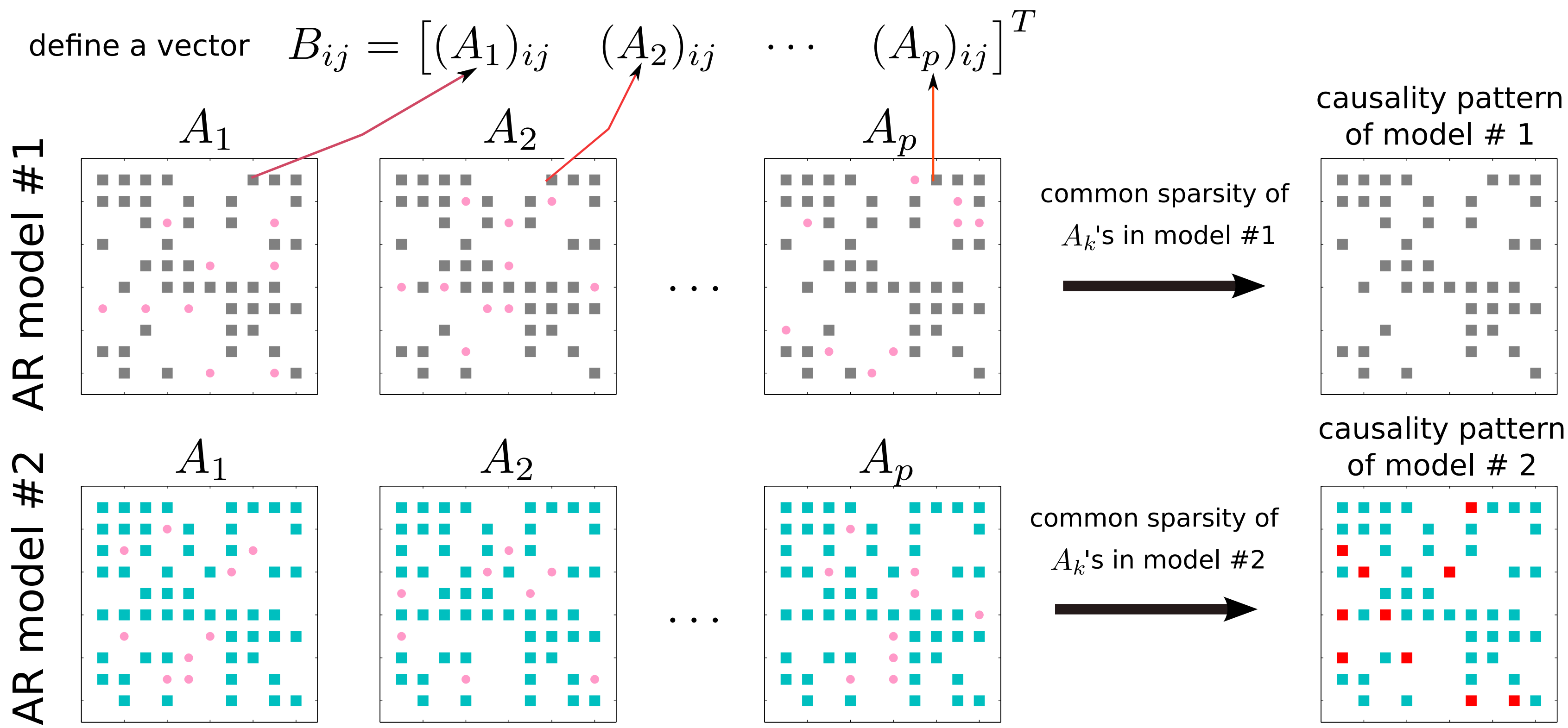
Brain under condition A



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

# Formulation for learning multiple graphical models

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

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

- the superscript $^{(k)}$ denotes the $k$th model
- $B_{ij}^{(k)} = \left[ (A_1^{(k)})_{ij} \quad (A_2^{(k)})_{ij} \quad \cdots \quad (A_p^{(k)})_{ij} \right]^T \in \mathbf{R}^p$
- 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-$\ell_2$-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

- Granger graphical models
- Learning multiple Granger graphical models
- **Algorithm**
- Numerical examples

## Splitting technique

by splitting the cost objective into three terms

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

and define the following functions

$$f(x) = (1/2)\|Gx - b\|_2^2, \quad g(x) = \lambda_1 \|x\|_{2,1}, \quad h(x) = \lambda_2 \|x\|_{2,1}$$

arranged into ADMM (Alternating Direction Multiplier Method) format as

$$\begin{array}{ll} \text{minimize} & f(x_1) + g(x_2) + h(x_3) \\ \text{subject to} & \begin{bmatrix} \mathcal{P} \\ \mathcal{D} \end{bmatrix} x_1 = \begin{bmatrix} x_2 \\ x_3 \end{bmatrix} \end{array}$$

with variables $x_1 \in \mathbf{R}^n, x_2 \in \mathbf{R}^s$ and $x_3 \in \mathbf{R}^r$

see the detail of the algorithm in Parikh and Boyd 2014, Proximal algorithms

# Computational cost in ADMM

the iteration update in ADMM involves

- basic matrix algebraic operations: addition, multiplication
- solving linear equations with positive definite matrix (using Cholesky)
- computing the proximal operator of $f(x) = \|x\|_{2,1} = \sum_{k=1}^{L} \|x_k\|_2$

$$(\mathbf{prox}_{\gamma f}(x))_k = \max\left\{1 - \frac{\gamma}{\|x_k\|_2}, 0\right\} x_k,$$

for $k = 1, 2, \dots, L$ known as **block solft thresholding operator**

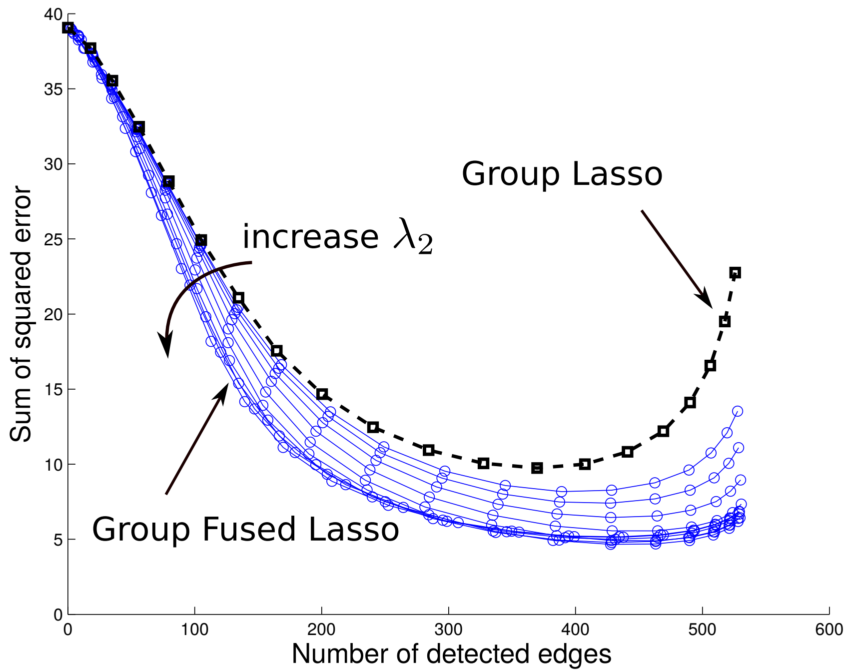
the algorithm applied in this problem is computationally cheap

# Outline

- Granger graphical models
- Learning multiple Granger graphical models
- Algorithm
- **Numerical examples**

# Numerical examples

generate $3$ sparse AR models having similar Granger structures




**Group Lasso:**

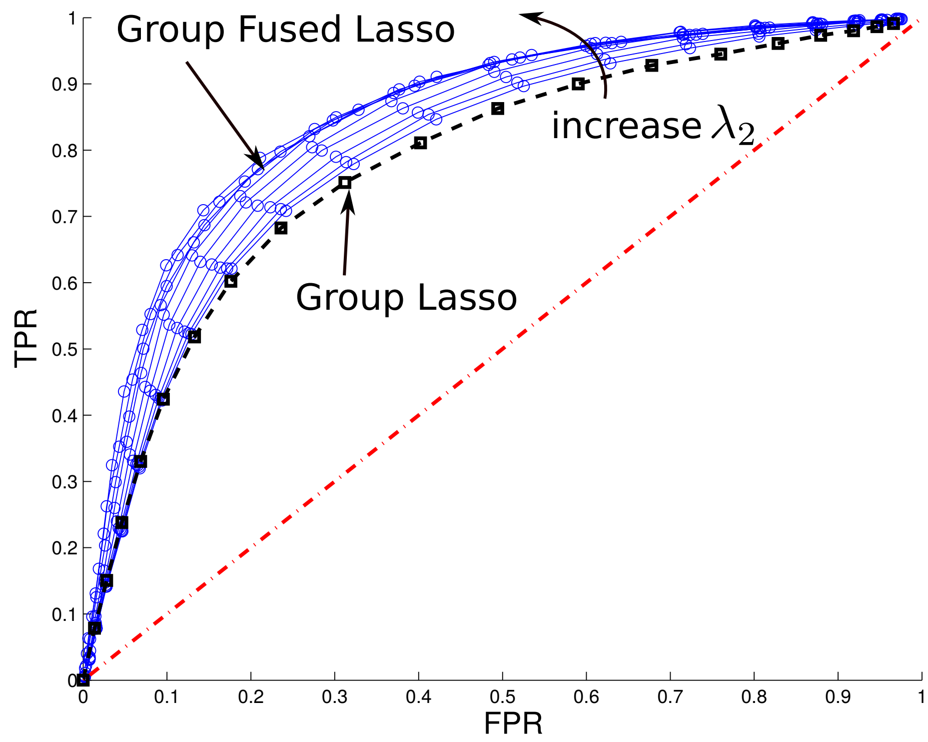
*seperately* estimate $3$ models

**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 $\lambda_2$ increases

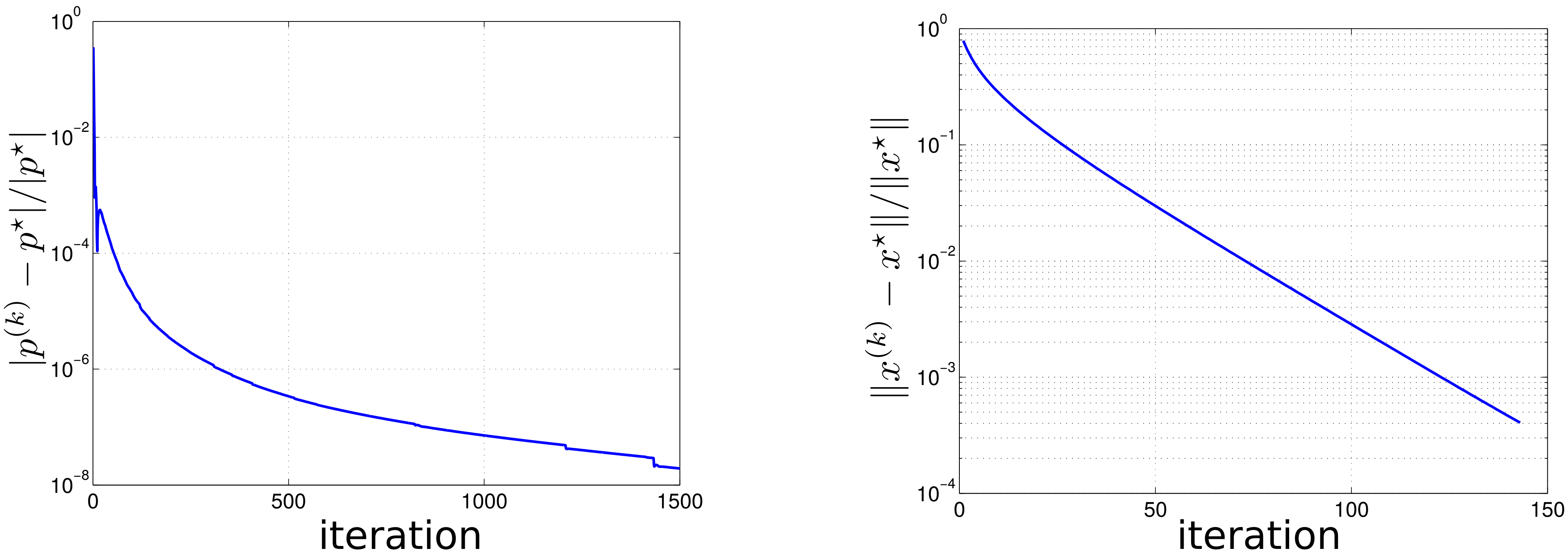
# ROC curves of Group Lasso VS Group Fused Lasso




- 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 $\lambda_2$ increases

## Performance of ADMM algorithm

solved the problem with $30,000$ variables by ADMM in 300-400 seconds




- **Left:** relative error of the primal objective
- **Right:** relative error of the 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 classes
- ADMM algorithm as an proximal method is shown to be efficient to solve the problem in large scale

# (Selected) References


- 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.

- N. Parikh and S. Boyd, “Proximal algorithms, *Foundations and Trends in Optimization*, vol. 1, no. 3, pp. 127–239, 2014. [Online]. Available: http://dx.doi.org/10.1561/2400000003

- D. OConnor and L. Vandenberghe, “Primal-dual decomposition by operator splitting and applications to image deblurring, *SIAM Journal on Imaging Sciences*, vol. 7, no. 3, pp. 1724–1754, 2014.