Sparse System Identification for Discovering Brain Connectivity from fMRI time series

Arnan Pongrattanakul, Puttichai Lertkultanon

Jitkomut Songsiri

Department of Electrical Engineering

Chulalongkorn University

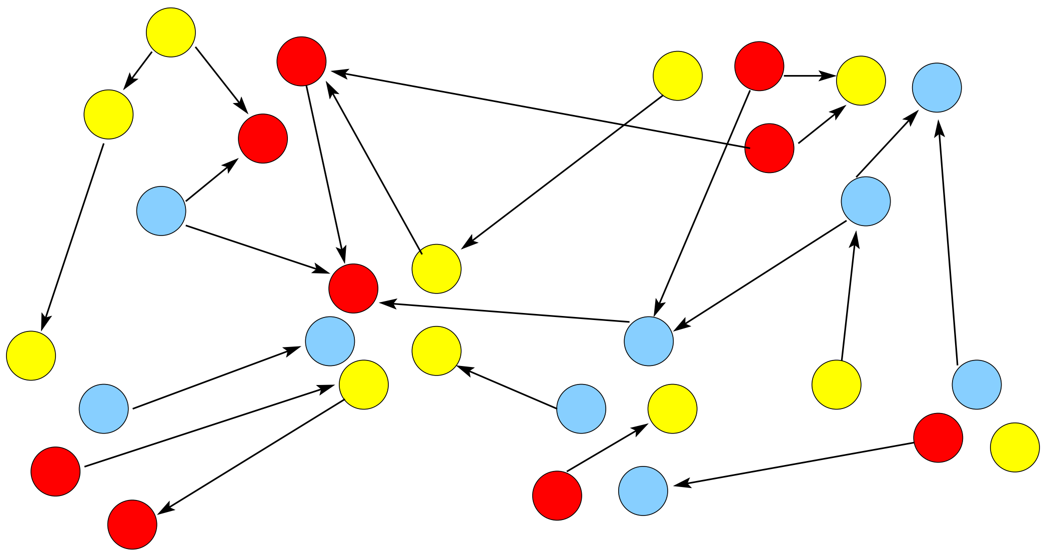
Joint meeting at Universiti Sains Malaysia

April 23-26, 2013

Outline

- **Granger Graphical Models**
- Sparse multivariate autoregressive models
- Numerical examples

Graphical Models



a graphical model consists of

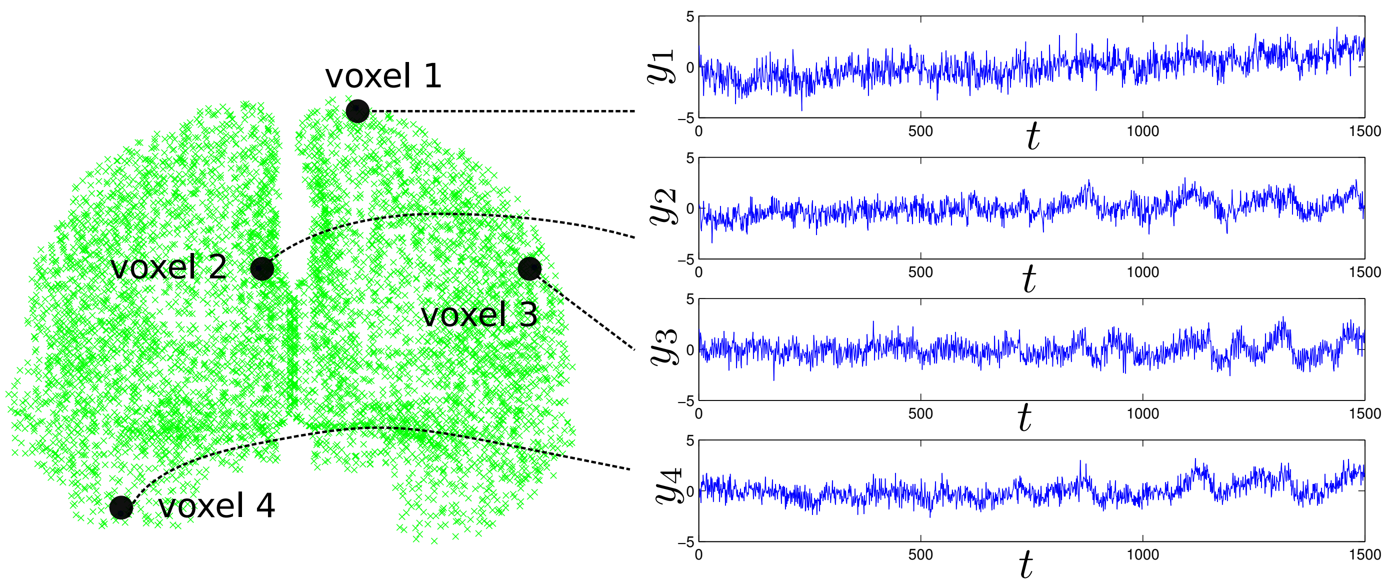
- **nodes:** represent variables of interest
  here the ith node is fMRI time series at the ith voxel
- **edges:** explain relationships between variables

Autoregressive Models

explain a multivariate time series by a vector AR process of order p

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

$y(t) \in \mathbf{R}^n$, $A_k \in \mathbf{R}^{n \times n}$, $k = 1, 2, \ldots, p$, $\nu(t)$ is noise



$n = 6004$ (number of voxels)
y_k represents the time series from the kth voxel

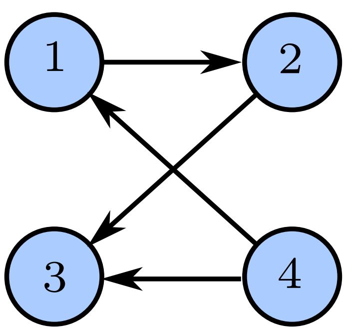
Granger Graphical Models (Granger 1969)

sparsity in coefficients A_k

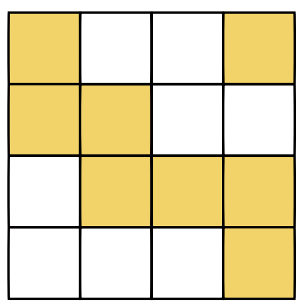
$$(A_k)_{ij} = 0, \quad \text{for } k = 1, 2, \ldots, p$$

is the characterization of **Granger causality** of AR models

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



granger graphical model



zero patterns in A_k

for example, 4-dimensional AR

y_2 is Granger caused by y_1

y_4 is NOT Granger caused by y_2

Outline

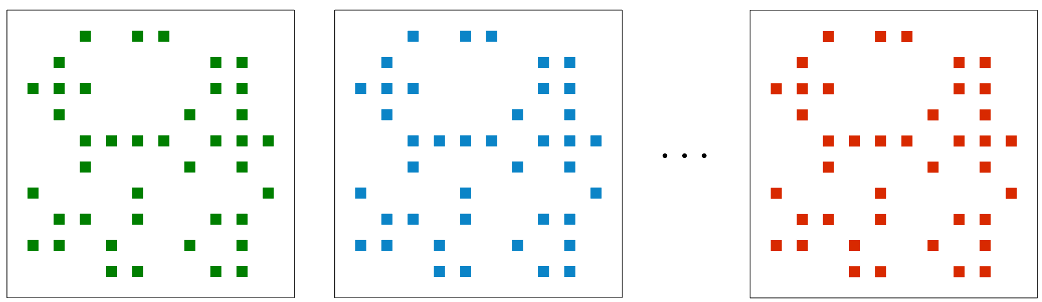
- Granger Graphical Models
- **Sparse multivariate autoregressive models**
- Numerical examples

Sparse Autoregressive (AR) Models

Problem: find A_k's that minimize the sum-square error

$$\sum_{t=p+1}^{N} \|y(t) - \sum_{k=1}^{p} A_k y(t-k)\|_2^2$$

- A_k's contain many zeros (to infer Granger causality among variables)
- $A_1, A_2, \ldots, A_p$ have a common zero pattern



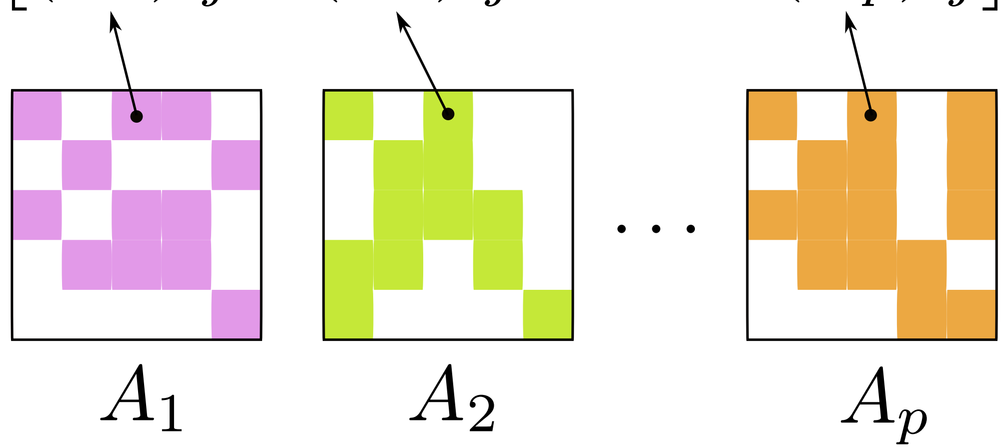
A_1 A_2 A_p

this formulation finds many applications in neuroscience and system biology (Salvador et al. 2005, Valdes-Sosa et al. 2005, Fujita et al. 2007, ...)

Group sparsity

stack the (i, j) entries of all A_k's in vector $B_{ij} \in \mathbf{R}^p$

$$\left[(A_1)_{ij} \quad (A_2)_{ij} \quad \cdots \quad (A_p)_{ij}\right] \triangleq B_{ij}$$



$$\|B_{ij}\|_2 = 0 \quad \Longrightarrow \quad (A_1)_{ij} = (A_2)_{ij} = \cdots = (A_p)_{ij} = 0$$

obtain a group sparsity in A_k's if we can enforce

$$\|B_{ij}\|_2 = 0, \quad \text{or} \quad \|\left[(A_1)_{ij} \quad (A_2)_{ij} \quad \cdots \quad (A_p)_{ij}\right]\|_2 = 0$$

for some (i, j)

given the measurements $y(1), y(2), \ldots, y(N)$

$$\begin{array}{ll} \text{minimize} & \sum_{t=p+1}^{N} \|y(t) - \sum_{k=1}^{p} A_k y(t-k)\|^2 \\ \text{subject to} & (A_1)_{ij} = (A_2)_{ij} = \cdots = (A_p)_{ij} = 0, \quad (i,j) \notin \mathcal{V} \end{array}$$

with variables $A_k \in \mathbf{R}^{n\times n}$ for $k = 1, 2, \ldots, p$

- $\mathcal{V}$ is the index set of a given Granger causality constraints
- the equality constraints can be eliminated, resulting in a reduced least-squares
- the solution is then analytically obtained

given the measurements $y(1), y(2), \ldots, y(N)$

$$\text{minimize} \quad \sum_{t=p+1}^{N} \|y(t) - \sum_{k=1}^{p} A_k y(t-k)\|^2 + \lambda \sum_{i \neq j} \left\| \begin{bmatrix} (A_1)_{ij} & (A_2)_{ij} & \cdots & (A_p)_{ij} \end{bmatrix} \right\|_2$$

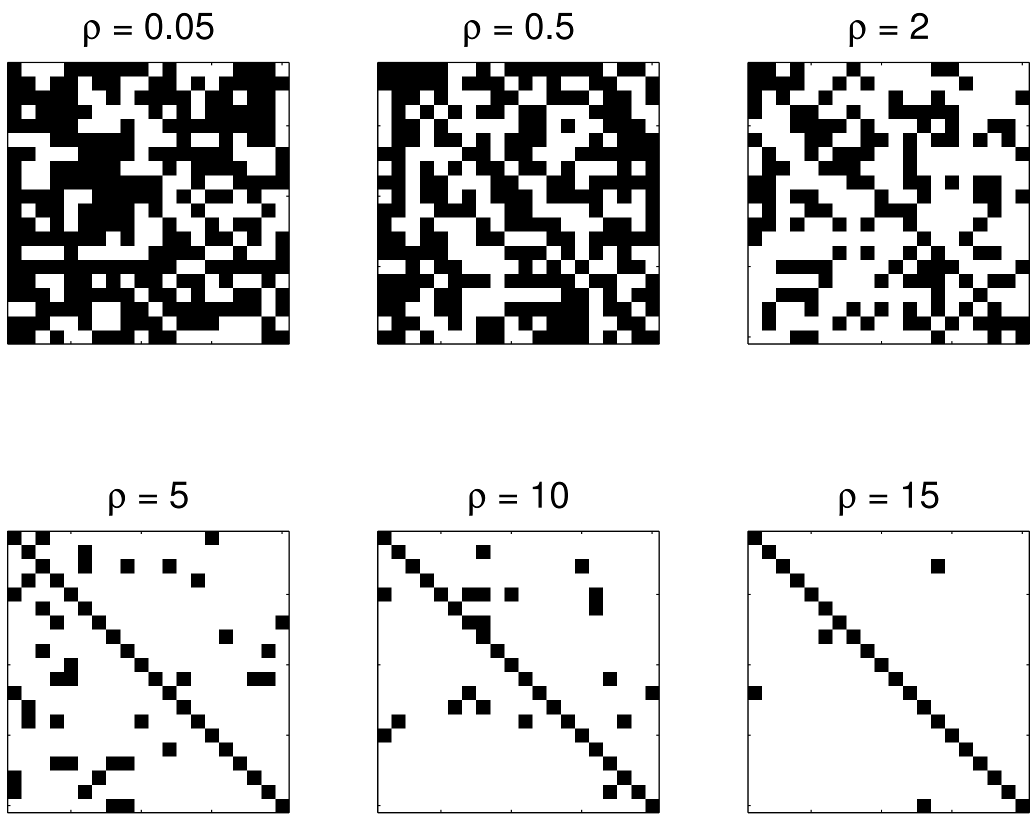
with variables $A_k \in \mathbf{R}^{n \times n}$ for $k = 1, 2, \ldots, p$

- regarded as an ℓ_1-regularized least-squares problem
- summation over (i, j) plays a role of ℓ_1-type norm
- using the ℓ_2 norm of p-tuple of $(A_k)_{ij}$ yields a *group sparsity*
- λ is called a regularization parameter ($\lambda > 0$)

a heuristic convex approach to obtain sparse AR coefficients

Example: 20-dimensional AR of order 3

a common zero pattern of a solution A_1, A_2 and A_3



as ρ increases, A_k's contain more zeros

Numerical method

the estimation problem can be expressed as

$$\text{minimize} \quad f(x) + \rho\|x\|_1$$

- uncontrained convex problem
- *nonsmooth* problem; make it challenging to solve in *large scale*
- suitable for ADMM algorithm; simple and fast in practice

  (S. Boyd, et al. *Distributed optimization and statistical learning via the alternating direction method of multipliers*, 2010)
- many approaches on choosing ρ have been proposed; BIC, cross validation, etc.

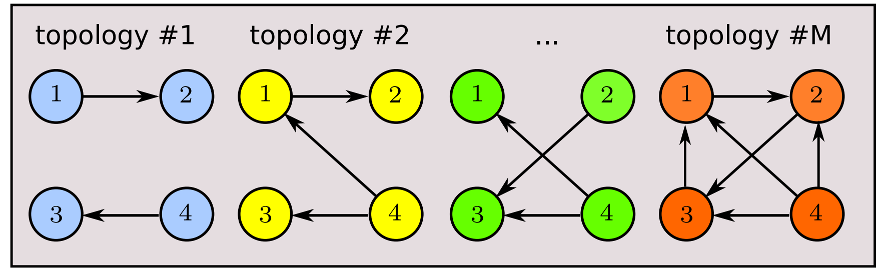
Model Selection

sparse AR estimation

minimize $(1/2)\|Y - AH\|_2^2 + \lambda g(A)$

where $g(A) = \sum_{i \neq j} \|[(A_1)_{ij}\ (A_2)_{ij}\ \cdots\ (A_p)_{ij}]\|_2$

vary λ



Granger constraints

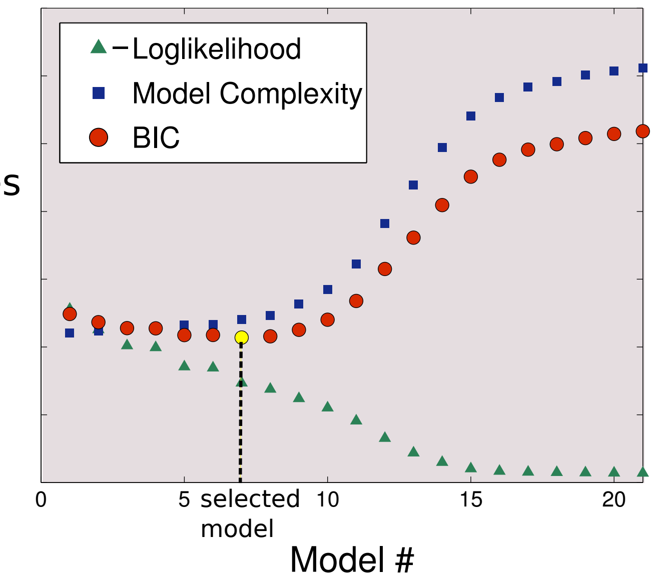
constrained AR estimation

minimize $(1/2)\|Y - AH\|_2^2$

subject to $(A_1)_{ij} = (A_2)_{ij} = \cdots = (A_p)_{ij} = 0$

Model Candidates

BIC = -2 Loglikelihood + Model Complexity



select the model that minimizes Bayes information criterion (BIC)

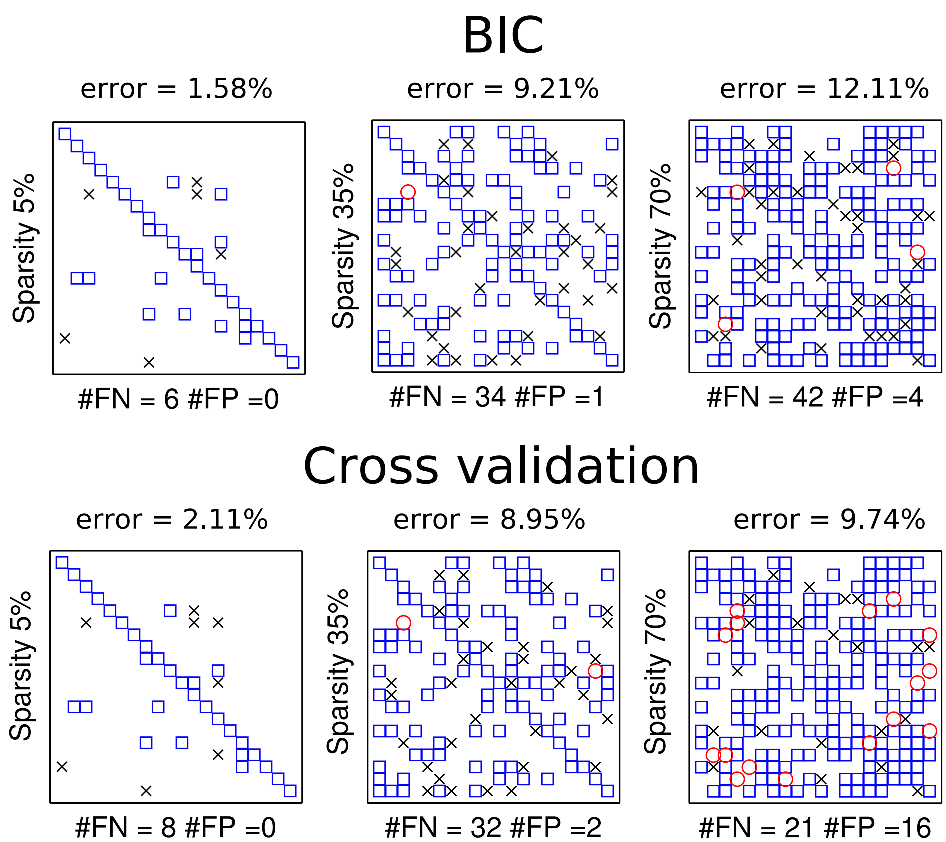
$$\text{BIC} = -2 \cdot \text{Loglikelihood} + d \log N$$

Outline

- Granger Graphical Models
- Sparse multivariate autoregressive models
- **Numerical examples**

Sparse AR estimation

generate 1000 time points from a sparse AR process with $n = 20$ and $p = 3$

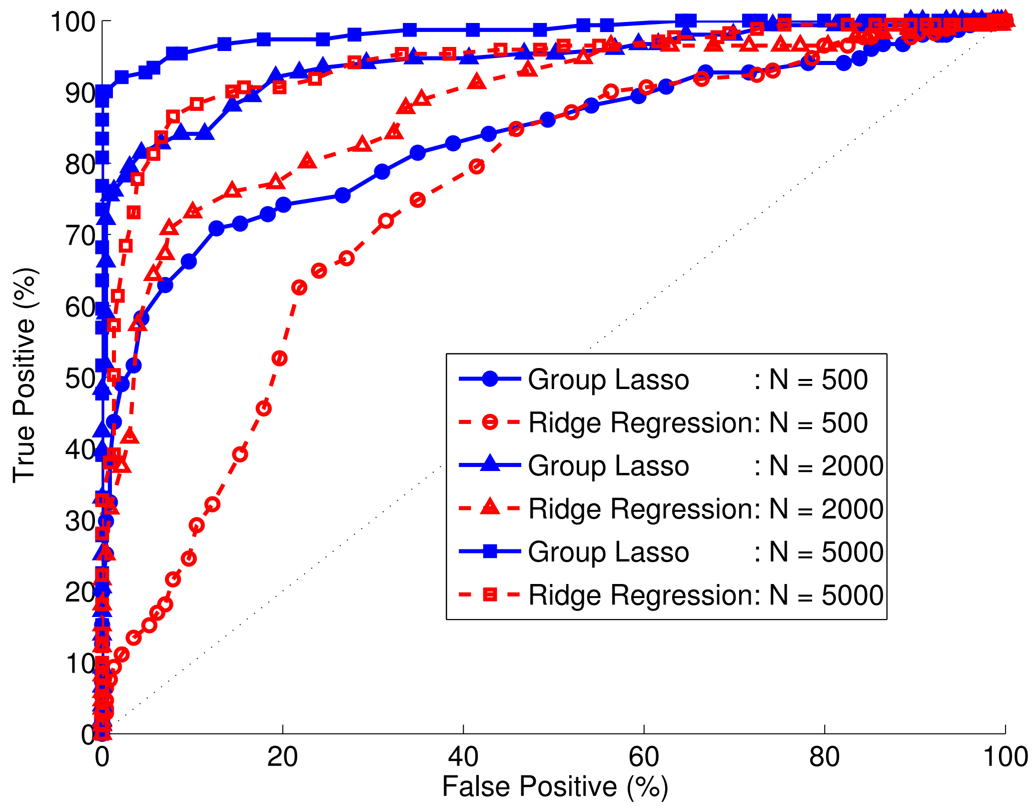


- **blue** squares are the correctly identified nonzero entries
- **red** circles are misclassified entries as nonzeros
- **black** crosses are misclassified entries as zeros

BIC yields a smaller error than cross validation if the true model is sparse

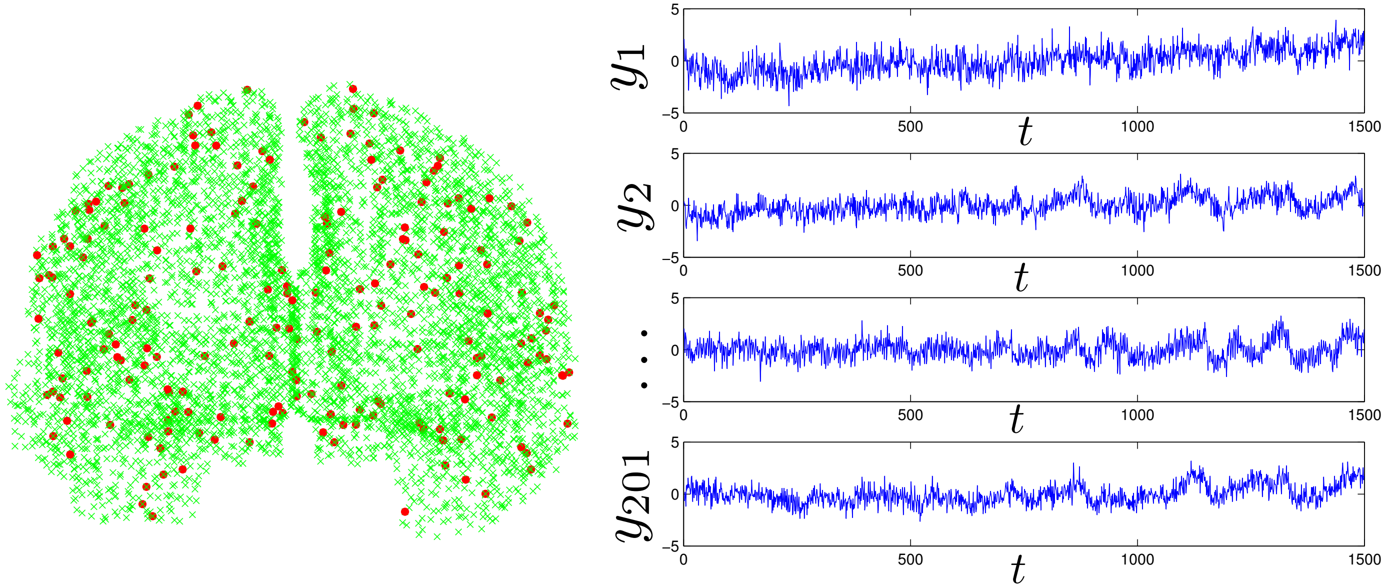
Receiver Operating Characteristic (ROC) curves

- Group sparse AR estimation: vary ρ
- Ridge regression: vary threshold value applied on the estimates



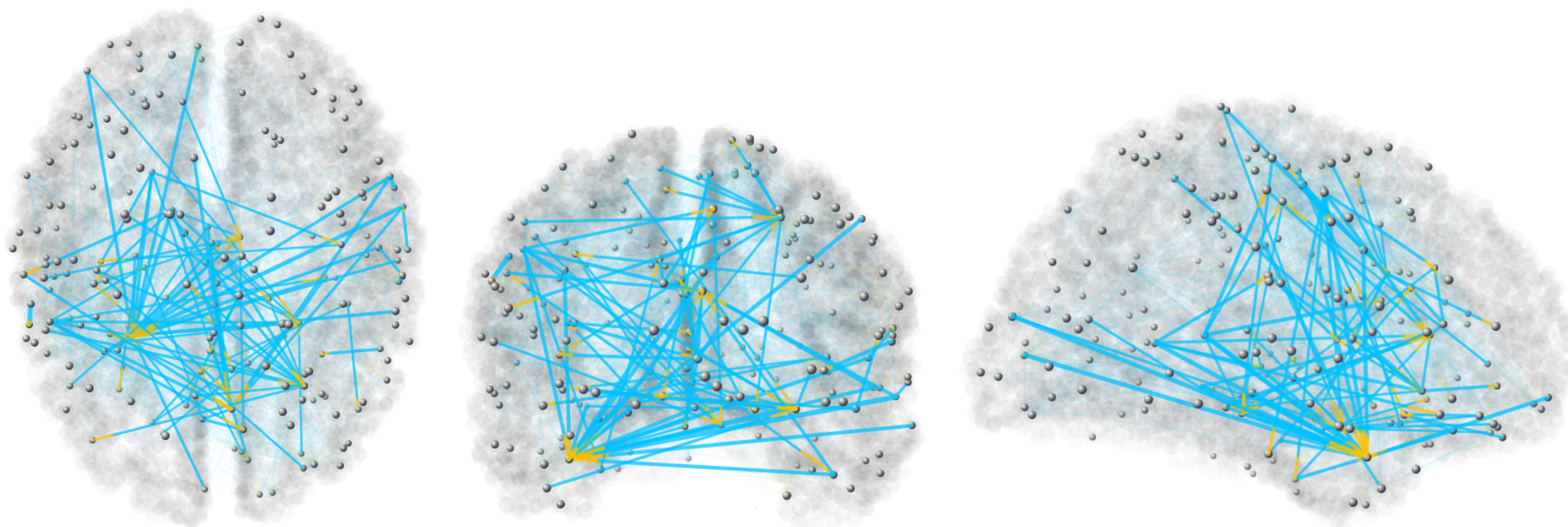
ROC of group sparse AR estimation lies above that of ridge regression

fMRI time series (Feinberg et. al. 2010)



- the data were obtained while a subject was in the resting state
- BOLD signals recorded at 6004 voxels with 1499 time samples
- reduce the number of voxels to 201 (red dots)

Granger Graphical Models for fMRI time series



- BIC selects the AR model of order 1 and the graph density is 7%
- the link width is proportional to $\|B_{ij}\|_2$
- orange color painted at the link end towards node j represents that the node j is Granger-caused by other nodes.

Summary

- graphical models are useful for explaining relationships in complex systems
- a problem of learning graph topologies can be formulated as a sparse identification problem
- to obtain a sparse model, we add an ℓ_1-type regularization to the estimation problem
- the resulting problem is unconstrained convex but nondifferentiable
- solving the problem in large scale is done by ADMM algorithm (shown to be efficiently fast in many applications)