Direct Estimation of Differential Functional Graphical Models

Boxin Zhao, Y. Samuel Wang, Mladen Kolar

Department of Statistics & Booth School of Business, The University of Chicago

Abstract

Background: In many applications, it is more natural to consider data as random function vectors rather than multivariate scalars; e.g., electroencephalography (EEG) data are more appropriately treated as functions of time. As in the multivariate scalar case, undirected graphs can be used to encode conditional independence relationships between the random functions.

What we do: We directly estimate the difference of structures between two functional graphs without estimating the individual graphs. The direct estimation approach allows better computational efficiency and statistically consistent results in a high-dimensional setting even when individual graphs are dense.

Algorithm contribution: We propose an efficient proximal gradient-descent algorithm to estimate the differential graph.

Theoretical result: We show that our method recovers the differential graph structure with high probability even in high-dimensional setting with dense individual graphs as long as the underlying differential structure is sparse.

Experiments: We demonstrate the benefits of our method by simulations and an analysis of real EEG data. We also provide a setting where the functional data approach is not ideal.

Problem Setting

Functional differential graphical model:
Let $X(t) = (X_1(t), \dots, X_p(t))$, and $Y(t) = (Y_1(t), \dots, Y_p(t))$, be two p-dimensional *multivariate Gaussian processes* with mean zero and common domain $\mathcal{T}$ from two different populations; assume that $\mathcal{T}$ is a closed subset of the real line. The conditional cross-covariance function for $X(t)$ is defined as:

$$C^X_{jl}(s,t) \; = \; \mathrm{Cov}\left(X_j(s), X_l(t) \;\middle|\; \{X_k(\cdot)\}_{k \neq j,l}\right),$$

and $C^\Delta_{jl}(s,t) = \; C^X_{jl}(s,t) - C^Y_{jl}(s,t)$. We define the differential graph to be $G_\Delta = \{V, E_\Delta\}$, where

$$E_\Delta \; = \; \{(j,l) \in V^2 : \; j \neq l \;\text{ and }\; \|C^\Delta_{jl}\|_{HS} \neq 0\}.$$

Data: $X_i(t) = (X_{i1}(t), \dots, X_{ip}(t)) \sim X(t)$, i.i.d., $i = 1, \dots, n_X$; and $Y_i(t) = (Y_{i1}(t), \dots, Y_{ip}(t)) \sim Y(t)$, i.i.d., $i = 1, \dots, n_Y$.

Object: Return estimate $\widehat{E}_\Delta$ based on data.

Our Method

Functional PCA: Estimate the covariance function $K^X_{jj}(t,s) = \mathrm{Cov}(X_{ij}(t), X_{ij}(s))$ with $\bar{X}_j(t) = n_X^{-1} \sum_{i=1}^{n_X} X_{ij}(t)$:

$$\widehat{K}^X_{jj}(s,t) = \frac{1}{n_X} \sum_{i=1}^{n_X} (X_{ij}(s) - \bar{X}_j(s))(X_{ij}(t) - \bar{X}_j(t)).$$

For each $j \in [p]$, an eigen-decomposition of $\widehat{K}^X_{jj}(s,t)$ gives estimated eigenfunctions $\widehat{\phi}^X_{jk}$ and scores $\widehat{a}^X_{ij} = \{\widehat{a}^X_{ijk}\}_{k \in [M]}$ with

$$X_{ij} \approx \textstyle\sum_{k=1}^M \widehat{\phi}^X_{jk} \widehat{a}^X_{ijk}.$$

FuDGE (Functional Differential Graph Estimation):
Suppose $\widehat{a}^{X,M}_i = \{\widehat{a}^X_{i1}, \dots, \widehat{a}^X_{ip}\}$ has sample covariance S^X and population precision matrix

$$\Theta^M_X = \begin{bmatrix} \mathrm{Cov}(\widehat{a}^T_{i1}\widehat{a}_{i1}) & \cdots & \mathrm{Cov}(\widehat{a}^T_{i1}\widehat{a}_{ip}) \\ \cdots & \ddots & \cdots \\ \mathrm{Cov}(\widehat{a}^T_{ip}\widehat{a}_{i1}) & \cdots & \mathrm{Cov}(\widehat{a}^T_{ip}\widehat{a}_{ip}) \end{bmatrix}^{-1} \quad \text{with } \mathrm{Cov}(\widehat{a}^T_{ij}\widehat{a}_{il}) \in \mathbb{R}^{M \times M}.$$

Roughly speaking, if the (j,l) block of Θ^M_X only contains 0, then $X_{ij} \perp X_{il} \mid \{X_{ik}\}_{k \neq l}$. Thus, we estimate $\widehat{\Delta}^M \approx \Theta^M_X - \Theta^M_Y$ with block sparsity by solving

$$\widehat{\Delta}^M \in \underset{\Delta \in \mathbb{R}^{pM \times pM}}{\operatorname{argmin}} \; L(\Delta) + \lambda_n \sum_{\{i,j\} \in V^2} \|\Delta_{ij}\|_F,$$

$$\widehat{E}_\Delta = \{(j,l) \in V^2 : \|\widehat{\Delta}^M_{jl}\|_F > \epsilon_n \text{ or } \|\widehat{\Delta}^M_{lj}\|_F > \epsilon_n\}.$$

Algorithm 1 FuDGE: Proximal gradient-descent

Input S^X, S^Y, λ_n, ϵ_n; **Output** $\widehat{\Delta}$
Initialize $\Delta^{(0)} = 0_{pM}$
repeat
  $A = \Delta - \eta \nabla L(\Delta) = \Delta - \eta \left(S_X^{(M)} \Delta S_Y^{(M)} - (S_Y^{(M)} - S_X^{(M)}) \right)$
  for $1 \leq i, j \leq p$ **do**
    $\Delta_{jl} \leftarrow \left(\frac{\|A_{jl}\|_F - \lambda_n \eta}{\|A_{jl}\|_F} \right)_+ \cdot A_{jl}$
  end for
until Converge

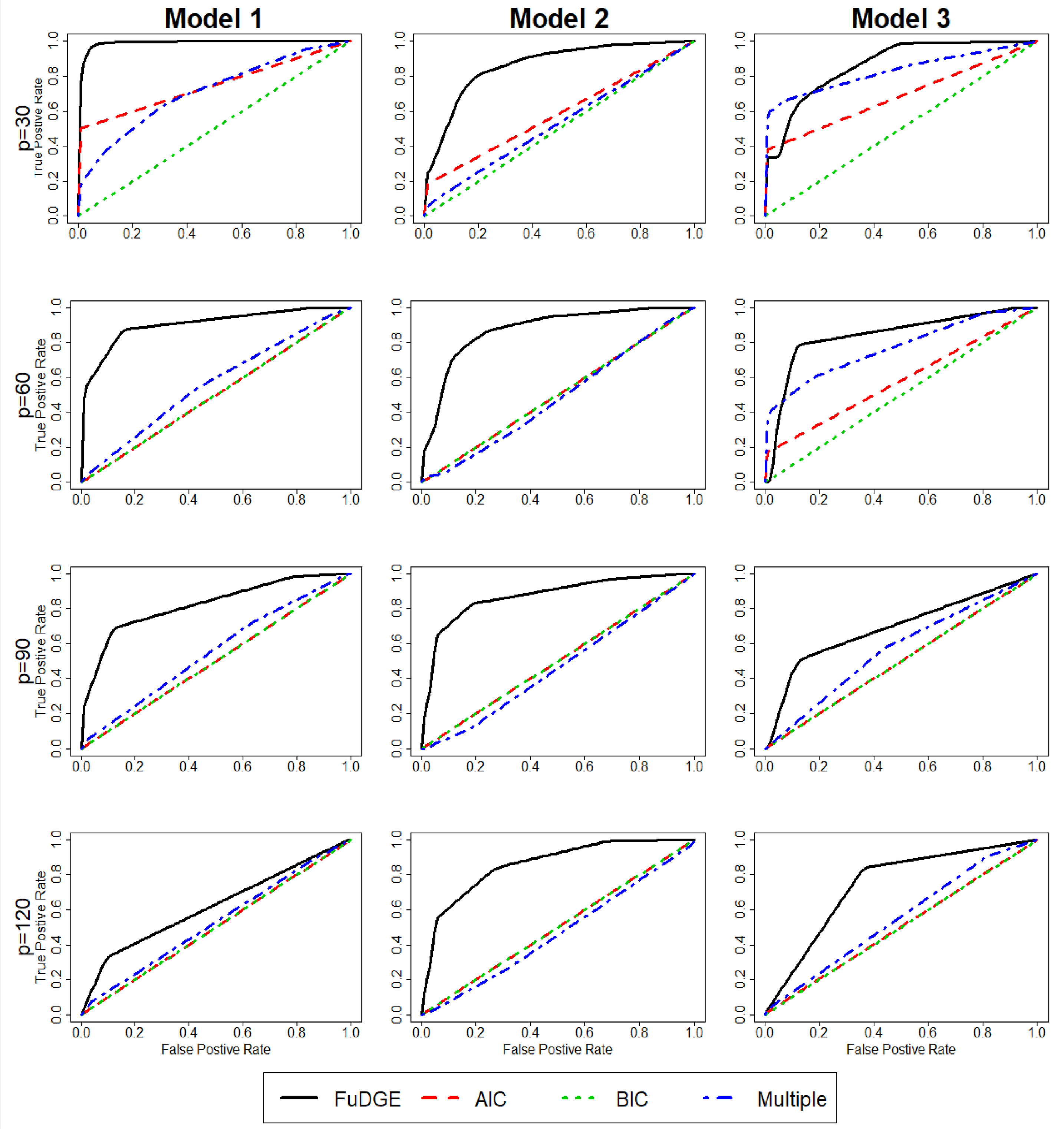
Theoretical Guarantees

Under mild conditions, for n and M large enough and appropriately selected ϵ_n, we have $P\left(\widehat{E}_\Delta = E_\Delta\right) \geq 1 - 2c_2/n^2$.

For fixed M, the sample complexity is $O(s \log(p))$, where s is the number of differential edges. Thus, the method is consistent even in the high-dimensional case when the differential graph is sparse.

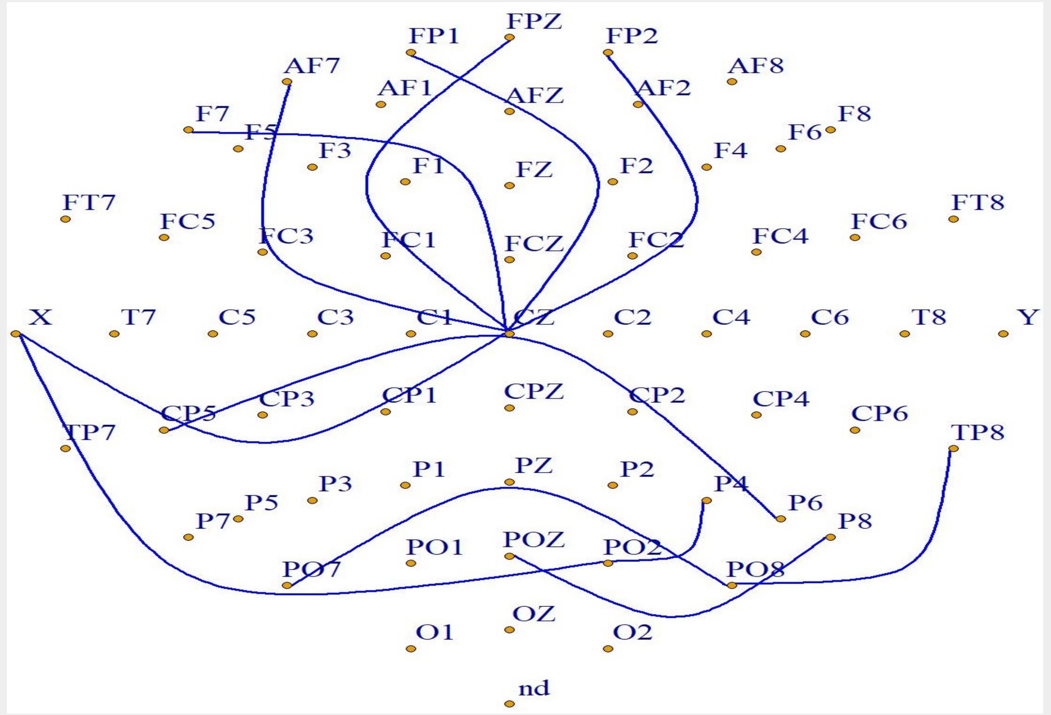
Simulation Results

Detailed simulation settings and more simulation results are given in the paper.



EEG Data Application

We use FuDGE to analyze electroencephalogram (EEG) data obtained from an alcoholism study, which included 122 total subjects; 77 in an alcoholic group and 45 in the control group. The EEG data was measured by placing $p = 64$ electrodes on various locations on the subject's scalp and measuring voltage values across time. The estimated differential graph for the two groups is below.



boxinz@uchicago.edu, swang24@uchicago.edu, mkolar@chicagobooth.edu