Nonparametric and high-dimensional graphical models for functional data

Eftychia Solea
CREST and ENSAI, Rennes, France
Joint work with Holger Dette
RUB, Bochum, Germany

June 24, 2021

Outline

Outline:

- Graphical models
- Graphical models for functional data
- Nonparametric and high-dimensional graphical models for functional data

Undirected graphical models

- Let $X = (X^1, \ldots, X^p)$ be a random vector.

- Let $\mathrm{G} = (\mathrm{V}, \mathrm{E})$ be an undirected graph, where $\mathrm{V} = \{1, \ldots, p\}$ is the set of nodes and $\mathrm{E} \subseteq \{(i,j) : \ i,j \in \mathrm{V}, \ i \neq j\}$ is the set of edges.

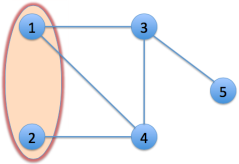
- Graphical models (GM) describe the conditional independencies (CI) among $X^1, \ldots, X^p$:

$$(i,j) \notin \mathrm{E} \quad \Longleftrightarrow \quad X^i \perp\!\!\!\perp X^j | X^{-(i,j)},$$

  where $X^{-(i,j)} = \{X^k : k \neq i,j\}$.

Graphical models (GM)

$$(1,2) \notin \mathrm{E} \iff X^1 \perp\!\!\!\perp X^2 | \{X^3, X^4, X^5\}.$$



- **In practice:** E is unknown

  $\Rightarrow$ Aim: **Estimate** E **based on a random sample from** X.

Gaussian graphical models (GGM)

- Let $X = (X^1, \ldots, X^p) \sim N(0, \Sigma)$. Let $\Theta = \Sigma^{-1}$ be the precision matrix. Then

$$X^i \perp\!\!\!\perp X^j | X^{-(i,j)} \iff \theta_{ij} = 0. \quad (1)$$

where θ_{ij} is the (i, j)th element of Θ.

- Estimating a GGM $\Leftrightarrow$ estimating the zero entries of Θ

$$(i, j) \notin \mathrm{E} \iff \theta_{ij} = 0. \quad (2)$$

  - Regression-based approach or neighborhood selection (Meinshausen and Buhlmann (2006), Peng et al. (2009)).
  - glasso/penalized maximum likelihood approach (Lasso, Yuan and Li (2007), SCAD and the adaptive lasso penalty (Lam and Fan, 2009), Dantzig selector (Cai et al., 2011) and hard-thresholding (Bickel and Levina, 2008)).

Gaussian graphical models

- Advantage of GGM: The equivalence

$$(i,j) \notin \mathsf{E} \quad \Longleftrightarrow \quad \theta_{ij} = 0,$$

that encodes conditional independence by the precision matrix.

- Disadvantage of GGM: The Gaussian assumption is very restrictive. For example,
  - The data are skewed.
  - There are nonlinear or heteroscedastic relations among the data.

$\Rightarrow$ We need more flexibility

Non-gaussian graphical models

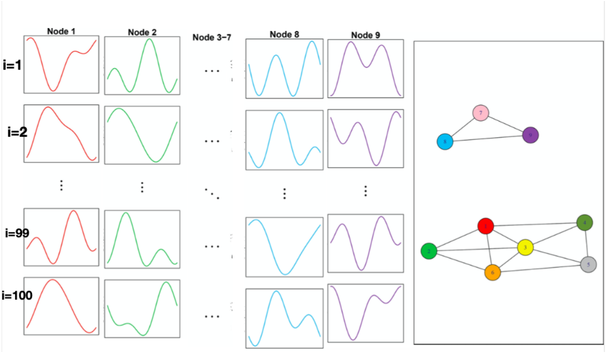
- Liu et al. (2009) and Xue and Zou (2012) relaxed the marginal Gaussian assumption using marginal copula transformations
- Voorman et al. (2004) imposed a generalized additive model between each node and its neighboring nodes.
- Li et al. (2014) and Lee et al. (2016a) developed a nonparametric GM by replacing conditional independence with additive conditional independence (ACI).
  - ACI satisfies the axioms of a semi-graphoid (Pearl, Geiger, and Verma, 1989), shared by conditional independence and the notion of separation
    $\rightarrow$ ACI can be used as an alternative criterion to construct a graph.
  - They defined several operators on additive Hilbert spaces that characterize ACI and applied hard-thresholding to determine the edges of the graph.

Graphical models for functional data

- Many applications, particularly in medical applications such as fMRI and EEG, produce multivariate functional data, where each sampling unit is modelled as a realization of a stochastic process varying over a time interval $T \subset \mathbb{R}$, $X(t) = (X^1(t), \ldots, X^p(t)), t \in T$.
- Problem: Construct functional graphical models, whose observations on the vertices are random functions.
- Goal: Represent statistical dependencies between random functions in the form of a network.

Toy example: Functional graphical models

- Left: Data, $n = 100$ observations of $X_i^j(t)$ for $j = 1, \ldots, 9$ nodes.
- Right: FGM of $p = 9$ nodes/functions.
- $(1, 9) \notin E \iff X^1 \perp\!\!\!\perp X^9 | X^{-(1,9)}$

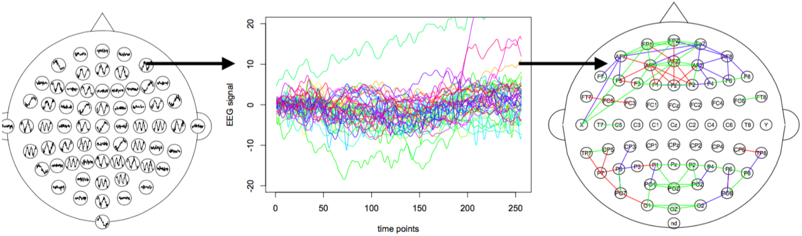


Picture taken from Qiao et al (2018). *Functional graphical models*.

Aim: Estimate E based on the data X_i^j, $j = 1, \ldots, p$, $i = 1, \ldots 100$.

Motivation - EEG data

- Data: 77 alcoholic subjects and 45 controls (UCI Mach. learn. repos.)
- Figure: 64 electrodes are placed on the subject's scalp.
- By the each electrode, EEG brain signals are recorded over time.



Aim: Characterise the dependence structure among the electrodes for the two groups of subjects.

Functional Gaussian graphical models (FGGM)

- Qiao et al. (2018) proposed a functional Gaussian graphical model (FGGM) where X is multivariate Gaussian stochastic process.
- They expand each X^i as a Karhunen-Loeve expansion.
- Extract the first m_n functional principal scores to form the Gaussian random vector, $\xi = (\xi^1, \ldots, \xi^p) \in \mathbb{R}^{pm_n}$ with the block precision matrix $\Theta_n = (\Theta_n^{ij})_{1 \le i,j \le p} = \Sigma_n^{-1} \in \mathbb{R}^{pm_n \times pm_n}$ (inverse covariance matrix).
- Then, under the Gaussian assumption

$$X^i \perp\!\!\!\perp X^j | X^{-(i,j)} \quad \Longleftrightarrow \quad \Theta_n^{ij} = 0. \tag{3}$$

- Hence, under the Gaussian assumption,

$$(i,j) \notin \mathrm{E}_n \quad \Longleftrightarrow \quad \Theta_n^{ij} = 0. \tag{4}$$

- Developed group-lasso penalized maximum likelihood estimation method which encourages blockwise sparsity of Θ_n.
- The FGGM is a generalization of the GGM (Yuan and Lin, (2006)) to the functional setting.

Functional Gaussian graphical models (FGGM)

- Qiao et al. (2018) showed that the FGGM can also be represented as a multivariate linear regression model with respect to the scores,

$$\xi_q^i = \sum_{j \neq i}^{p} \sum_{r=1}^{m_n} B_{qr}^{ij} \xi_r^j + \epsilon_q^i, \quad i \in \mathrm{V}, q = 1, \ldots, m_n, \tag{5}$$

such that $(\epsilon_q^i)_{1 \leq q \leq m_n}$ is uncorrelated with $(\xi_r^j)_{1 \leq r \leq m_n}, i \neq j$ if and only if

$$B_n^{ij} = -(\Theta_n^{ii})^{-1} \Theta_n^{ij}, \quad (i,j) \in \mathrm{V} \times \mathrm{V}, i \neq j, \tag{6}$$

where $B_n^{ij} = (B_{qr}^{ij})_{1 \leq q,r \leq m_n}$.

- Hence, under the Gaussian assumption the conditional relationships between nodes i and j are linear.

- Estimation of FGGM $\Leftrightarrow$ Estimation of the sparsity structure of B_n^{ij}.

$$(i,j) \notin \mathrm{E}_n \iff B_n^{ij} = 0 \iff \Theta_n^{ij} = 0.$$

Nonparametric functional graphical models

- Li and Solea (2018) developed a nonparametric FGM by extending ACI to its functional version, FACI, and proposing the functional additive precision operator (FAPO) to characterise FACI and also the graph.
- Solea and Li (2020) introduced the functional copula Gaussian distribution and they used it to develop a nonparametric FGM.

Nonparametric functional graphical models

Our objectives:

- Construct an alternative nonparametric graphical model for random functions.
- Remove the linearity assumption in FGGM by replacing the conditional linear relationships $B_{qr}^{ij}\xi_r^j$ among the scores with additive relationships $f_{qr}^{ij}(\xi_r^j)$.
- Develop concentration bounds for the resulting estimates at the high-dimensional setting.
- Construct brain networks based on EEG data.

Methodology

- For each $j = 1, \ldots, p$, let $X^j \in \mathcal{L}^2(T)$ such that $E\|X^i\|^2 < \infty$, where $\mathcal{L}^2(T)$ denotes the space of all square-integrable real-valued functions on T with the common inner product.

- Without loss of generality, we assume $\mu_{X^i}(t) = E(X^i(t)) = 0$ for all $t \in T$ and for all $i = 1, \ldots, p$.

- For each $(i, j) \in \mathrm{V} \times \mathrm{V}$, we define the autocovariance operator between the functions X^i and X^j as

$$\Sigma_{X^i X^j}(f)(t) = \int_T f(s)\sigma_{X^i X^j}(s, t)ds, \quad f \in \mathcal{L}^2(T),$$

  where $\sigma_{X^i X^j}(s, t) = \mathrm{cov}(X^i(s), X^j(t)) = E(X^i(s)X^j(t))$ is the cross-covariance function between X^i and X^j.

Karhunen-Loeve expansion

- Then each $X^j \in \mathcal{L}^2(T)$ can be represented by its Karhunen-Loève expansion

$$X^j = \sum_{r \in \mathbb{N}} \sqrt{\lambda_r^j} \xi_r^j \phi_r^j, \quad j = 1, \ldots, p.$$

- ξ_r^j are called the scores and they are uncorrelated random variables with $E(\xi_r^j) = 0$, $\mathrm{var}(\xi_r^j) = 1$,

- $\{(\lambda_r^j, \phi_r^j) : r = 1, 2, \ldots\}$ are eigenvalues and orthogonal eigenfunctions of $\Sigma_{X^j X^j}$.

- We assume the scores are independent and they take values in the closed and bounded interval e.g $[-1, 1]$.

Additive function-on-function model

Definition 1

A vector of random functions X follows the function-on-function additive model if for each pair $(i,j) \in \mathrm{V} \times \mathrm{V}$ there exists a sequence of smooth functions $f^{ij} = \{f^{ij}_{qr} : q, r \in \mathbb{N}\}$ defined on $\mathbb{R}$ with $E[f^{ij}_{qr}(\xi^j_r)] = 0, q, r \in \mathbb{N}$, such that

$$E[\xi^i_q | \{\xi^j_r, j \neq i\}] = \sum_{j\neq i}^{p} \sum_{r=1}^{\infty} f^{ij}_{qr}(\xi^j_r) \quad (7)$$

- Our model can be regarded as the nonparametric and additive version of the FGGM.
- Extends the model of Voorman et al. (2013) to the functional setting.

Additive functional graphical model

Definition 2

A vector of random functions X is said to follow an additive functional graphical model (AFGM) with respect to an undirected graph $\mathsf{G} = (\mathsf{V}, \mathsf{E})$ if and only if X is a function-on-function additive model of the form (7) and

$$(i, j) \notin \mathsf{E} \quad \Leftrightarrow \quad \mathsf{X}^{\mathsf{i}} \perp\!\!\!\perp \mathsf{X}^{\mathsf{j}} | \mathsf{X}^{-(\mathsf{i},\mathsf{j})}.$$

The definition implies

$$\mathsf{E} = \{(\mathsf{i}, \mathsf{j}) \in \mathsf{V} \times \mathsf{V} : \mathsf{i} \neq \mathsf{j}, \mathsf{f}^{\mathsf{ij}}_{\mathsf{qr}} \neq 0 \, \text{for some} \, \mathsf{q}, \mathsf{r} \in \mathbb{N}\}.$$

Additive functional graphical model

- Since each random function is infinite-dimensional, some type of regularisation is needed.
- We truncate the Karhunen-Loève expansion at a finite number of principal components m_n

$$E[\xi_q^i|\{\xi_r^j, j \neq i\}] = \sum_{j \neq i}^{p} \sum_{r=1}^{m_n} f_{qr}^{ij}(\xi_r^j),$$

where $m_n \to \infty$ as $n \to \infty$.

- Then, our goal is to estimate the truncated edge set

$$\mathrm{E}_n = \{(i,j) \in \mathrm{V} \times \mathrm{V} : i \neq j, f_{qr}^{ij} \neq 0 \text{ for some } q, r = 1, \ldots, m_n\}$$

Estimation

- Let $\hat{\xi}^i_{ur}, u = 1, \ldots, n, r = 1, \ldots, m_n, i \in \mathrm{V}$ be the estimated scores.

- Under some smoothness conditions, the additive functions f^{ij}_{qr} can be approximated by linear combinations of B-splines functions

$$f^{ij}_{qr}(x) \approx \sum_{k=1}^{k_n} h_k(x)\, \beta^{ij}_{qrk}, \quad q, r = 1, \ldots, m_n, \text{ where } k_n \to \infty.$$

- Then

$$f^{ij}_{qr} = 0 \quad \Leftrightarrow \quad \|\beta^{ij}_{qr}\|_2^2 = 0,$$

  where $\|\cdot\|_2$ denotes the Euclidean norm of $\beta^{ij}_{qr} = (\beta^{ij}_{qr1}, \ldots, \beta^{ij}_{qrk_n})^\mathsf{T} \in \mathbb{R}^{k_n}$, $q, r = 1, \ldots, m_n$.

- Let $B^{ij} = (\beta^{ij}_{qr})_{1 \le q \le m_n, 1 \le r \le m_n} \in \mathbb{R}^{k_n m_n \times m_n}$, then

$$(i, j) \notin \mathrm{E}_n \quad \Leftrightarrow \quad \|\mathrm{B}^{ij}\|_F = 0 \text{ for all } i \neq j,$$

  where $\|\cdot\|_F$ is the Frobenious norm.

- Inference of E_n $\Leftrightarrow$ Inference of sparsity structure of the spline coefficient matrix $B^i = (B^{ij}, i \neq j) \in \mathbb{R}^{(p-1)k_n m_n \times m_n}$.

Estimation procedure

- Estimate B^i by solving, separately for each $i \in \mathrm{V}$, a penalized additive regression of each node on all others (analog of neighborhood selection)

$$\widehat{PL}_i(B, \hat{\xi}) = \frac{1}{2n}\|\hat{\xi}^i - \tilde{H}_n(\hat{\xi}^{-i})B^i\|_F^2 + \lambda_n \textstyle\sum_{j\neq i}^p \|B^{ij}\|_F, \qquad (8)$$

  where $\tilde{H}_n(\hat{\xi}^{-i}) \in \mathbb{R}^{n\times(p-1)k_n m_n}$ design matrix of the center B-splines functions and $B^i = (B^{ij}, i \neq j) \in \mathbb{R}^{(p-1)k_n m_n \times m_n}$ coefficient regression matrix.

- Optimization is done by distance convex programming techniques.

- Given $\hat{B}_n^i$ as the solution of

$$\hat{B}_n^i = \operatorname{argmin}\{\widehat{PL}_i(B^i, \hat{\xi}) : B^i \in \mathbb{R}^{(p-1)k_n m_n \times m_n}\}.$$

  Estimate the set E_n by

$$\hat{\mathrm{E}}_n = \{(i,j) \in \mathrm{V}\times\mathrm{V} : i \neq j, \|\hat{B}_n^{ij}\|_F > 0 \text{ or } \|\hat{B}_n^{ji}\|_F > 0\}.$$

Algorithm

We summarize the algorithm below

1. Implement FPCA to obtain the estimated scores $\hat{\xi}^i_{ur}$ of each observation X^i_u. Transform the scores into the range $[-1, 1]$ using a monotone transformation. Choose m_n so that at least 90% of the total variation is explained.
2. For a given λ_n and for each $i \in \mathrm{V}$ solve the optimisation problem using, for example, distance convex programming techniques (e.g FISTA), to find a sparse estimate of B^i_n.
3. Declare that there is an edge between node i and node j if and only if either $\|\hat{B}^{ij}_n\|^2_F$ or $\|\hat{B}^{ji}_n\|^2_F$ are not zero.

Theoretical properties

- We develop model selection consistency of $\hat{\mathrm{E}}_n$ assuming
  - Random functions are fully observed for all t
  - (m_n, p_n, k_n) are allowed to grow as a function of n.
- The true population matrix $B^{*i}_{m_n} = (B^{*i1}_{m_n}, \ldots, B^{*ii-1}_{m_n}, B^{*ii+1}_{m_n}, \ldots, B^{*ip}_{m_n})$, with $B^{*ij}_{m_n} = \{\beta^{*ij}_{qrk} : 1 \le q, r \le m_n, k \in \mathbb{N}\}$ is defined by

$$B^{*i}_{m_n} = \operatorname{argmin}_{\beta^{ij}_{qrk}, 1 \le q, r \le m_n, k \in \mathbb{N}} \left\{ \sum_{q=1}^{m_n} E \left(\xi^i_q - \sum_{j \neq i}^{p} \sum_{r=1}^{m_n} \sum_{k=1}^{\infty} \tilde{h}_k(\xi^j_r)\, \beta^{ij}_{qrk} \right)^2 \right\},$$

where $\tilde{h}_k(\xi^j_r) = h_k(\xi^j_r) - E(h_k(\xi^j_r))$.

Theoretical properties

- Let $B_n^{*i} = (B_{m_n k_n}^{*i1}, \ldots, B_{m_n k_n}^{*ii-1}, B_{m_n k_n}^{*ii+1}, \ldots, B_{m_n k_n}^{*ip}) \in \mathbb{R}^{(p-1)k_n m_n \times m_n}$ denote the true truncated population matrix.

- The true truncated neighbourhood N_n^i of each node $i \in \mathsf{V}$ by

$$\mathsf{N}_n^i = \{j \in \mathsf{V} \setminus \{i\} : \|B_{m_n k_n}^{*ij}\|_F > 0\}.$$

- The true truncated edge set E_n

$$\mathsf{E}_n = \{(i,j) \in \mathsf{V} \times \mathsf{V} : i \neq j, i \in \mathsf{N}_n^j \text{ or } j \in \mathsf{N}_n^i\}.$$

Theoretical properties

- Let
$$f^{ij}_{qr}(\xi^j_r) = \sum_{k=1}^{\infty} \beta^{*ij}_{qrk} h_k(\xi^j_r) = \sum_{k=1}^{\infty} \beta^{*ij}_{qrk} \tilde{h}_k(\xi^j_r).$$
- We obtain from (7) the representation
$$\xi^i_q = \sum_{j \in \mathrm{N}^{\mathrm{i}}_{\mathrm{n}}} \sum_{r=1}^{m_n} f^{ij}_{qr}(\xi^j_r) + \epsilon^i_q, \quad q = 1, \ldots, m_n, i = 1, \ldots, p,$$
where ϵ_q are errors.
- The best approximation (in the least squares sense) of $E[\xi^i_q | \{\xi^j_r, j \neq i\}]$ is an additive function of the scores in the set of neighbours $\mathrm{N}^{\mathrm{i}}_{\mathrm{n}}$ of the node i only.

Theoretical properties

- We introduce the matrices

$$\Sigma^*_{\mathrm{N^i_n N^i_n}} = E\left(\tilde{\mathbf{H}}(\xi^{\mathrm{N^i_n}})\tilde{\mathbf{H}}(\xi^{\mathrm{N^i_n}})^\top\right) \in \mathbb{R}^{n^i k_n m_n \times n^i k_n m_n} \tag{9}$$

and

$$\Sigma^*_{\xi^j \mathrm{N^i_n}} = E\left(\tilde{H}(\xi^j)^\top \tilde{\mathbf{H}}(\xi^{\mathrm{N^i_n}})\right) \in \mathbb{R}^{k_n m_n \times n^i k_n m_n}, \tag{10}$$

where n^i is the cardinality of $\mathrm{N^i_n}$ and

$$\tilde{\mathbf{H}}(\xi^{\mathrm{N^i_n}})^\top = (\tilde{H}(\xi^j), j \in \mathrm{N^i_n}) \in \mathbb{R}^{\mathrm{n^i k_n m_n}}, \quad \tilde{\mathrm{H}}(\xi^{\mathrm{j}}) \quad = (\tilde{h}^\top(\xi^j_r))_{1 \le r \le m_n} \in \mathbb{R}^{k_n m_n}$$

Theoretical assumptions

Standard assumptions for the eigenvalues of covariance operators.

1. (i) There exist positive constants d_0, d_1 and d_2 such that

$$d_0 r^{-\beta} \leq \lambda_r^i \leq d_1 r^{-\beta}, \quad \lambda_r^i - \lambda_{r+1}^i \geq d_2^{-1} r^{-1-\beta} \quad \text{for } r \geq 1,$$

   and for some $\beta > 1$.

   (ii) The number of principal component scores m_n satisfies $m_n \asymp n^\alpha$ for some constant $\alpha \in [0, \frac{1}{2+3\beta})$,

   where $a_n \asymp b_n$ represents $A \leq \inf_n |\frac{a_n}{b_n}| \leq \sup_n |\frac{a_n}{b_n}| \leq B$, for $A > 0$ and $B > 0$.

Theoretical assumptions

The next two conditions refer to the smoothness of the functions f_{qr}^{ij}.

- Let l be a nonnegative integer, and let $\rho \in (0,1]$ be such that $d = l + \rho > 0.5$.
- Define $\mathscr{F}_{l,\rho}$, the Hölder space of functions $f : [-1,1] \to \mathbb{R}$ whose lth derivative exists and satisfies a Lipschitz condition of order ρ

$$\|f\|_\infty = \sup_{x\in[-1,1]}|f(x)| \le M \text{ for some } M > 0$$

2. $f_{qr}^{ij} \in \mathscr{F}_{l,\rho}$ and $E[f_{qr}^{ij}(\xi_{ur}^{j})] = 0$, for all $q, r = 1, \ldots, m_n$ and $(i,j) \in \mathrm{V} \times \mathrm{V}$.

3. The joint density function, say p^j, of the random vector $\xi^j = (\xi_1^j, \ldots, \xi_{m_n}^j)^\top$ is bounded away from zero and infinity on $[-1,1]^{m_n}$ for every $j = 1, \ldots, p$.

Theoretical assumptions

Assumptions to show model selection consistency for the lasso

4. Sub-gaussian tails There exists a constant $C > 0$ such that $P(|\epsilon_q^i| > x) \leq 2\exp(-Cx^2)$ for all $x \geq 0$ and $q = 1, \ldots, m_n, i \in \mathrm{V}$.

5. Sparsity $n^i = o(p)$ for all $i \in \mathrm{V}$, and there exists a constant $\theta > 0$ such that for all $i \in \mathrm{V}$

$$\sum_{j \in \mathrm{N}_n^i} \|B_{m_n k_n}^{*ij}\|_F < \theta.$$

6. Bounded eigenspectrum The minimum eigenvalue $\Lambda_{\min}(\Sigma^*_{\mathrm{N}_n^i \mathrm{N}_n^i})$ of the matrix $\Sigma^*_{\mathrm{N}^i \mathrm{N}^i}$ defined in (9) satisfies

$$\Lambda_{min}(\Sigma^*_{\mathrm{N}_n^i \mathrm{N}_n^i}) > C_{\min}. \qquad (11)$$

for some constant $C_{\min} > 0$.

7. Irrepresentable condition There exists a constant $0 < \eta \leq 1$ such that

$$\max_{j \notin \mathrm{N}_n^i} \|\Sigma^*_{\xi^j \mathrm{N}_n^i}(\Sigma^*_{\mathrm{N}_n^i \mathrm{N}_n^i})^{-1}\|_F \leq \frac{1-\eta}{\sqrt{n^i}}. \qquad (12)$$

Consistency of the $\mathrm{N}_\mathrm{n}^\mathrm{i}$

Theorem

If assumptions 1-7 are satisfied and the regularization parameter λ_n satisfies for all i

$$\frac{n^i m_n^{3/2}}{k_n^d \sum_{j\in \mathrm{N}_\mathrm{n}^\mathrm{i}} \|B^{*ij}_{m_n k_n}\|_F} \lesssim \lambda_n \lesssim (n^i)^{-3/2} (b_n^{*i})^3 \left(\sum_{j\in \mathrm{N}_\mathrm{n}^\mathrm{i}} \|B^{*ij}_{m_n k_n}\|_F\right)^{-2}, \tag{13}$$

where $b_n^{*i} = \min_{j\in \mathrm{N}_\mathrm{n}^\mathrm{i}} \|B^{*ij}_{m_n k_n}\|_F$. Then,

$$P\left(\hat{\mathrm{N}}_n^i \neq \mathrm{N}_n^i\right) \lesssim \exp\left(-C_1 \frac{n^{1-\alpha(2+3\beta)} \left(\lambda_n \sum_{j\in \mathrm{N}_\mathrm{n}^\mathrm{i}} \|B^{*ij}_{m_n k_n}\|_F\right)^2}{n^i m_n^2 k_n^4} + 2\log(p m_n k_n)\right),$$

where $C_1 > 0$.

Consistency of the E_n

Corollary

If the assumptions of Theorem 1 are satisfied, we have for a positive constant $C_1 > 0$

$$P(\hat{\mathrm{E}}_n \neq \mathrm{E}_n) \lesssim \exp\left(-C_1 \frac{n^{1-\alpha(2+3\beta)}(\lambda_n \min_{i=1}^{p} \sum_{j \in \mathrm{N}_n^i} \|B_{m_n k_n}^{*ij}\|_F)^2}{p m_n^2 k_n^4} + 2\log(p m_n k_n)\right).$$

Sketch of the proof

1. We show that if Assumptions (6) and (7) hold, then with high probability, the assumptions hold also for the corresponding **sample matrices**

$$\Sigma^{n}_{\mathrm{N^i_n N^i_n}} = \frac{1}{n}\tilde{\mathbf{H}}_{\mathbf{n}}(\xi^{\mathrm{N^i_n}})\tilde{\mathbf{H}}^{\mathsf{T}}_{\mathbf{n}}(\xi^{\mathrm{N^i_n}}),\ \Sigma^{n}_{\xi^j \mathrm{N^i_n}} = \frac{1}{n}\tilde{H}_n(\xi^j)^{\mathsf{T}}\tilde{\mathbf{H}}_{\mathbf{n}}(\xi^{\mathrm{N^i_n}})^{\mathsf{T}}$$

2. Then, we prove a conditional result of the Theorem, for the "fixed design" matrices using the technique of Bach (2008).

3. Additionally, the objective function to be minimised is based on the estimated scores
$\Rightarrow$ establish concentration bounds in the estimation of the sample design matrix $\Sigma^{n}_{\mathrm{N^i_n N^i_n}}$ using the estimated scores (rather than the true scores).

Simulation studies

- Compare numerically the performances of the AFGM estimator with 1) FGGM (Qiao et al, 2018) and 2) FAPO (Li and Solea, 2018).

- Given an edge set E of a directed acyclic graph, we generate functional data by the model

$$X_u^i(t_s) = \sum_{(i,j)\in \mathrm{E}} \sum_{q=1}^{5} \sum_{r=1}^{5} f_{qr}^{ij}(\xi_{ur}^j)\phi_q(t_s) + \epsilon_{us}^i, \quad u = 1, \ldots, n, s = 1, \ldots, 100$$

where $\phi_1^i(t), \ldots, \phi_5^i(t)$ are the first 5 functions of the orthonormal Fourier basis, and ϵ_{us}^i is an iid sample from $\mathcal{N}(0, \sigma^2)$.

- As a consequence the scores satisfy

$$\xi_{uq}^i = \sum_{(i,j)\in \mathrm{E}} \sum_{r=1}^{5} f_{qr}^{ij}(\xi_{ur}^j) + \tilde{\epsilon}_{uq}^i, \quad u = 1, \ldots, n, q = 1, \ldots, 5 \quad (14)$$

where the errors $\tilde{\epsilon}_{uq}^i$ form an iid sample a centred normal distribution.

Simulation studies

- For simplicity we assume $f_{qr}^{ij}(x) = f(x)$ for all $q, r = 1, \ldots, m_n$ and for all $(i, j) \in E$.
- In all examples, we center $f(\xi_{ur}^{j})$ to have 0 mean.
- We estimate each function X_u^i using 10 B-spline basis functions of order 4.
- We choose $m_n = 5$ functional principal components scores so that at least 90% of the total variation is explained.
- We approximate each f_{qr}^{ij} using B-splines of order 4 and take $k_n = 4 + \lceil \sqrt{n} \rceil$.

Simulation studies

We consider the following two nonlinear scenarios.

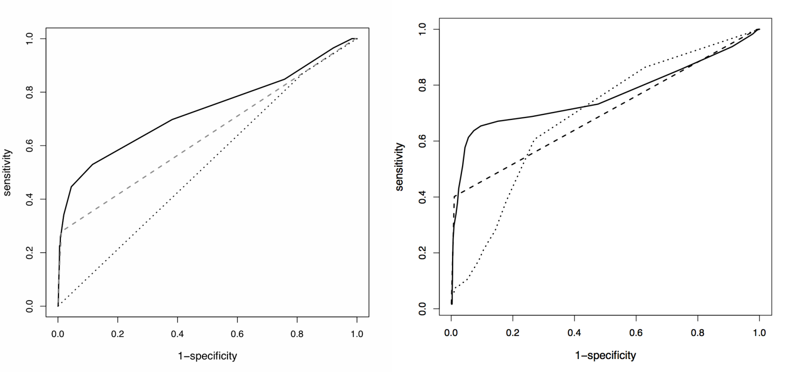
Model I: $f(x) = 1.4 + 3x - \frac{1}{2} + \sin(2\pi(x - \frac{1}{2})) + 8(x - \frac{1}{3})^2 - \frac{8}{9}$.

- For the choice of scores, we simulate ξ_{ur}^i independently from the uniform distribution $U[-1,1]$ for all $r = 1, \ldots, m_n, i \in \mathrm{V}, u = 1, \ldots, n$.
- The errors ϵ_{uq}^i simulated independently from $\mathcal{N}(0, 0.1)$.

Model II: $f(x) = -\sin(2x) + x^2 - 25/12 + x + \exp(-x) - 2/5 \cdot \sinh(5/2)$.

- ξ_{ur}^i were simulated independently from the uniform distribution $U[-2.5, 2.5]$ for all $r = 1, \ldots, m_n, i \in \mathrm{V}, u = 1, \ldots, n$.
- The errors ϵ_{uq}^i simulated independently from $\mathcal{N}(0, 1)$.

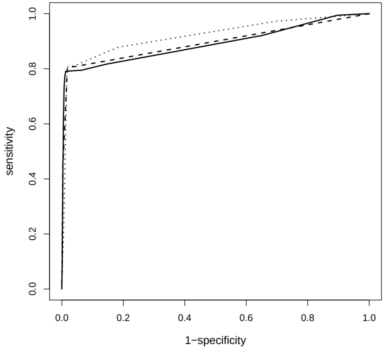
Nonlinear scenario



ROC curves ((AFGM (—), FAPO (— — —), FGGM ($\cdots$)) for Model I (left) and Model II (right) for $(p, n) = (100, 100)$.

- The areas under the ROC of the AFGM are larger than for the FGGM and FAPO, indicating the superior performance of the AFGM under a nonlinear, sparse and high-dimensional scenario.

Linear scenario

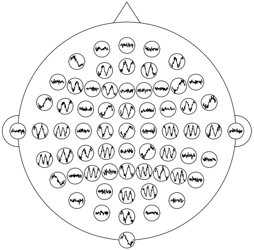


ROC curves ((AFGM (−), FAPO (− − −), FGGM (· · ·))

- The AFGM estimator is computed using the scale scores ξ^i_{ur}
- FAPO and the FGGM are computed using standard Gaussian scores
- There is some loss of efficiency by the nonparametric functional estimators, but the losses are quite modest.

Application to EEG data

- EEG data of 77 alcoholic subjects and 45 control subjects.
- For each subject EEG brain signals were recorded at 256 time points over a one second interval using 64 electrodes placed on the subject's scalp.



A schematic representation of the functional data collected by EEG from a subject.

Application to EEG data

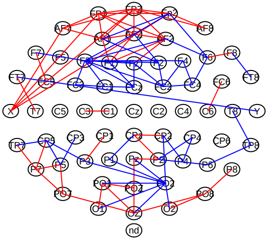
- Goal: Apply AFGM to identify differences in the brain network connectivity between the two groups of subjects.

- We take the tuning constant λ_n to be such that 5% of the $\binom{64}{2}$ pairs of vertices are retained as edges.

- We choose $k_n = 4 + \lceil \sqrt{n} \rceil$ B-spline functions of order 4, $m_n = 5$.

Application to EEG data

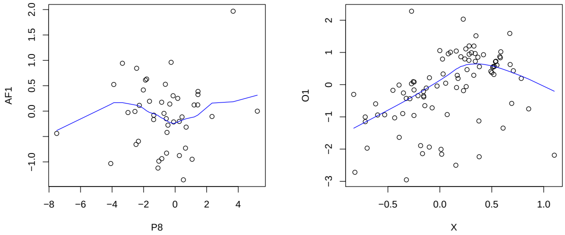
Differential brain network constructed by AFGM:



- red lines indicate the edges that are in the alcoholic network but not in the control network.
- blue lines indicate the edges that are in the control network but not in the alcoholic network.

Application to EEG data

- Pairwise scatterplots for the control group between channels AF1 and P8 (left) and channels O1 and X (right).



- Scatterplots show nonlinear relationships among the scores, violating the linearity assumption.

References

1. Bach, F. R. (2008), Consistency of the group lasso and multiple kernel learning, Journal of Machine Learning Research, 9, 11791225.
2. Lee, K.-Y., Li, B., and Zhao, H. (2016), ?On an additive partial correlation operator and nonparametric estimation of graphical models,? Biometrika, 103, 513?530.
3. Li, B and Solea, E, (2018), "A nonparametric graphical model for functional data with application to brain networks based on fMRI." Journal of the American Statistical Association.
4. Li, B and Solea, E, (2020), "Copula Gaussian graphical models for functional data." Journal of the American Statistical Association.
5. Li, B., Chun, H., and Zhao, H. (2014), ?On an Additive Semigraphoid Model for Statistical Networks With Application to Pathway Analysis,? Journal of the American Statistical Association, 109, 1188?1204
6. Liu, H., Lafferty, J., and Wasserman, L. (2009). The nonparanormal: Semiparametric estimation of high dimensional undirected graphs. Journal of Machine Learning Research, 10(Oct), 2295-2328.
7. Meinshausen, N., and Bhlmann, P. (2006). High-dimensional graphs and variable selection with the lasso. The annals of statistics, 34(3), 1436-1462.
8. Qiao, X., Guo, S., and James, G. M. (2018). Functional graphical models. Journal of the American Statistical Association, 1-12.
9. Voorman, A., Shojaie, A., Witten, D. (2013). Graph estimation with joint additive models. Biometrika, 101(1), 85-101.
10. Yuan, M., and Lin, Y. (2006). Model selection and estimation in regression with grouped variables. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 68(1), 49-67.
11. Xue, L., and Zou, H. (2012). Regularized rank-based estimation of high-dimensional nonparanormal graphical models. The

ACI definition

Let U, V and W be subvectors of $X = (X^1, \ldots, X^p)$. Let $\mathfrak{M}_U$, $\mathfrak{M}_V$, and $\mathfrak{M}_W$ be Hilbert spaces of additive functions of U, V and W. We say that U and V are additively conditionally independent given W iff

$$(I - P_{\mathfrak{M}_W})\mathfrak{M}_U \perp (I - P_{\mathfrak{M}_W})\mathfrak{M}_V,$$

where the orthogonality $\perp$ is in terms of the inner product in $L_2(P)$.

Lemma

Suppose that Assumption 11 holds. Then, there exists a constant $C_1 > 0$ *such that for any* $\delta > 0$,

$$P\left(\|\Sigma^{n}_{\mathrm{N}^i_n \mathrm{N}^i_n} - \Sigma^{*}_{\mathrm{N}^i_n \mathrm{N}^i_n}\|_F \geq \delta\right) \leq 2\exp\left(-C_1 \frac{n\delta^2}{(n^i m_n k_n)^2} + 2\log(n^i m_n k_n)\right). \tag{15}$$

$$P\left(\Lambda_{\min}(\Sigma^{n}_{\mathrm{N}^i_n \mathrm{N}^i_n}) \leq C_{\min} - \delta\right) \leq 2\exp\left(-C_1 \frac{n\delta^2}{(n^i m_n k_n)^2} + 2\log(n^i m_n k_n)\right). \tag{16}$$

The next Lemma guarantees that the sample matrices satisfy the irrepresentable condition in Assumption 12 with high probability.

Lemma

If Assumption 11 and 12 are satisfied for some $0 < \eta \leq 1$, then

$$P\Big(\max_{j \notin \mathrm{N_n^i}} \| \Sigma^n_{\xi^j \mathrm{N_n^i}} (\Sigma^n_{\mathrm{N_n^i N_n^i}})^{-1} \|_F \geq \frac{1 - \frac{\eta}{2}}{\sqrt{n^i}} \Big)$$

$$\lesssim \exp\Big(- C_1 \frac{n}{((n^i)^{5/4} m_n k_n)^2} + 2 \log(p m_n k_n) \Big),$$

where C_1 is a positive constant that depends only on $C_{\min}$ and η.

The next result provides tail bounds for all entries of the matrix $\hat{\Sigma}^n_{\mathrm{N}^i_n \mathrm{N}^i_n} - \Sigma^n_{\mathrm{N}^i_n \mathrm{N}^i_n}$.

Theorem

Suppose that Assumption (1) holds. Then, there exists a positive constants C_1 such that for any $\delta > 0$ satisfying $0 < \delta \le C_1$ and for all $(i,j) \in \mathrm{V} \times \mathrm{V}$, $i \neq j$, $r, q = 1, \dots, m_n$ and $k, \ell = 1, \dots, k_n$, we have

$$P\Big(\Big|\frac{1}{n}\sum_{u=1}^{n}\Big(\tilde{h}_{nk}(\hat{\xi}^i_{ur})\tilde{h}_{n\ell}(\hat{\xi}^j_{uq}) - \tilde{h}_{nk}(\xi^i_{ur})\tilde{h}_{n\ell}(\xi^j_{uq})\Big)\Big| \ge \delta\Big)$$
$$\lesssim \exp\left(-C_1 n^{1-\alpha(2+3\beta)} k_n^{-2} \delta^2\right).$$

Proposition

Suppose that Assumptions 2 and 3 are satisfied. Then, there exist functions $\tilde{f}^{ij}_{nqr} = \sum_{k=1}^{k_n} \beta^{ij}_{qrk} \tilde{h}_{nk}$ *and positive constants* c_1, C_1, *such that*

$$P\big(\Omega\big) \leq 2\exp\Big(- C_1 \frac{n k_n^{-2d}}{n^i m_n^2} + \log(n^i m_n^2)\Big),$$

where

$$\Omega = \Big\{ \max_{j \in \mathrm{N}^{\mathrm{i}}_{\mathrm{n}}} \max_{1 \leq q, r \leq m_n} \frac{1}{\sqrt{n}} \|\mathbf{f}^{\mathbf{ij}}_{\mathbf{qr}} - \tilde{\mathbf{f}}^{\mathbf{ij}}_{\mathbf{qr}}\|_2 \geq c_1 k_n^{-d} \Big\}\ , \tag{17}$$

and $\mathbf{f}^{\mathbf{ij}}_{\mathbf{qr}} = \big(f^{ij}_{qr}(\xi^j_{1r}), \ldots, f^{ij}_{qr}(\xi^j_{nr})\big)^\top$, $\tilde{\mathbf{f}}^{\mathbf{ij}}_{\mathbf{qr}} = \big(\tilde{f}^{ij}_{qr}(\xi^j_{1r}), \ldots, \tilde{f}^{ij}_{qr}(\xi^j_{nr})\big)^\top$.

The idea of the proof is to

1. first construct an estimator $\hat{B}_n^{\mathrm{N}_\mathrm{n}^\mathrm{i}}$ by minimizing the following restricted problem given the true support $\mathrm{N}_\mathrm{n}^\mathrm{i}$. That is,

$$\hat{B}_n^{\mathrm{N}_\mathrm{n}^\mathrm{i}} = \operatorname{argmin}\left\{\widehat{PL}_{\mathrm{N}_\mathrm{n}^\mathrm{i}}(B, \hat{\xi}) : B \in \mathbb{R}^{n^i k_n m_n \times m_n}\right\}, \tag{18}$$

   where

$$\widehat{PL}_{\mathrm{N}_\mathrm{n}^\mathrm{i}}(B, \hat{\xi}) = \frac{1}{2n}\|\hat{\xi}^i - \tilde{\mathbf{H}}_\mathbf{n}^\mathsf{T}(\hat{\xi}^{\mathrm{N}_\mathrm{n}^\mathrm{i}})B\|_F^2 + \frac{\lambda_n}{2}\left(\sum_{j\in\mathrm{N}_\mathrm{n}^\mathrm{i}}^p \|B^{ij}\|_F\right)^2, \tag{19}$$

   (note that $\widehat{PL}_{\mathrm{N}_\mathrm{n}^\mathrm{i}}(B, \hat{\xi})$ corresponds to the function $\widehat{PL}_i(B, \hat{\xi})$, where we put $B^{ij} = 0$ whenever $j \notin \mathrm{N}^\mathrm{i}$), and to show that the minimizer in (18) is "close" to the true matrix $B_n^{*\mathrm{N}_\mathrm{n}^\mathrm{i}}$. To achieve this we use similar arguments as in Bach (2008).

2. We show that $(\hat{B}_n^{\mathrm{N}_\mathrm{n}^\mathrm{i}}, \mathbf{0})$, with high probability, satisfies the second KKT-condition (20b), and thus, it is optimal for problem (18)

Lemma

(KKT conditions) A matrix $B^i = (B^{ij}, j \in \mathsf{V} \setminus \{\mathsf{i}\}) \in \mathbb{R}^{(\mathsf{p}-1)\mathsf{k_n m_n} \times \mathsf{m_n}}$ *with support* $\mathsf{N}^\mathsf{i}_\mathsf{n}$ *is optimal for problem* (8) *if and only if*

$$(\hat{\Sigma}^n_{\mathsf{N}^\mathsf{i}_\mathsf{n}\mathsf{N}^\mathsf{i}_\mathsf{n}} + \lambda_n \hat{D}_{\mathsf{N}^\mathsf{i}_\mathsf{n}}) B^{\mathsf{N}^\mathsf{i}_\mathsf{n}} - \hat{\Sigma}^n_{\mathsf{N}^\mathsf{i}_\mathsf{n}\xi^\mathsf{i}} = 0, \quad \textit{for all } j \in \mathsf{N}^\mathsf{i}_\mathsf{n}, \tag{20a}$$

$$\|\hat{\Sigma}^n_{\xi^j \mathsf{N}^\mathsf{i}_\mathsf{n}} B^{\mathsf{N}^\mathsf{i}_\mathsf{n}} - \hat{\Sigma}^n_{\xi^j \xi^i}\|_F \leq \lambda_n \textstyle\sum_{j \neq i}^p \|B^{ij}\|_F, \quad \textit{for all } j \notin \mathsf{N}^\mathsf{i}_\mathsf{n} \tag{20b}$$

where $\hat{\Sigma}^n_{\mathsf{N}^\mathsf{i}_\mathsf{n}\mathsf{N}^\mathsf{i}_\mathsf{n}}$, $B^{\mathsf{N}^\mathsf{i}_\mathsf{n}} = (B^{ij}, j \in \mathsf{N}^\mathsf{i}_\mathsf{n}) \in \mathbb{R}^{\mathsf{n}^\mathsf{i}\mathsf{k_n m_n} \times \mathsf{m_n}}$,
$B = (\beta^{ij}_{qrk} : 1 \leq q, r \leq m_n, 1 \leq k \leq k_n)$ *and*

$$\hat{D}_{\mathsf{N}^\mathsf{i}_\mathsf{n}} = \mathrm{diag}\big((\hat{D}_{\mathsf{N}^\mathsf{i}_\mathsf{n}})_{jj} : j \in \hat{\mathsf{N}}^\mathsf{i}_\mathsf{n}\big)\big)$$

is a block diagonal matrix with n^i *elements*
$(\hat{D}_{\mathsf{N}^\mathsf{i}_\mathsf{n}})_{jj} = \dfrac{\sum_{\ell \neq i}^p \|\hat{B}^{i\ell}\|_F}{\|\hat{B}^{ij}\|_F} I_{k_n m_n} \in \mathbb{R}^{k_n m_n \times k_n m_n}$.

Proposition

Suppose Assumptions of Proposition 1-5 are satisfied and that δ *satisfies*

$$\frac{2}{C_{\min}}\lambda_n(n^i)^{3/2}\left(\sum_{j\in \mathrm{N}_\mathrm{n}^\mathrm{i}}\|B^{*ij}_{m_nk_n}\|_F\right)^2 \le c_2 b_n^{*i}\delta \tag{21}$$

for some constant $c_2 > 0$*. Then,*

$$P\left(\|\hat{B}_n^{\mathrm{N}_\mathrm{n}^\mathrm{i}} - B_n^{*\mathrm{N}_\mathrm{n}^\mathrm{i}}\|_F \ge \delta\right) \lesssim \exp\left(-C_1\frac{n^{1-\alpha(2+3\beta)}(b_n^{*i})^2\delta^2}{(n^i)^4 m_n^2 k_n^4\left(\sum_{j\in \mathrm{N}_\mathrm{n}^\mathrm{i}}\|B^{*ij}_{m_nk_n}\|_F\right)^2} + 2\log(n^i m_n k_n)\right),$$

where $C_1 > 0$ *such that* $0 < \delta \le C_1$*.*

Proposition

The matrix $(\hat{B}_n^{\mathrm{N}_\mathrm{n}^\mathrm{i}}, \mathbf{0})$ *satisfies* (20b) *with high probability, in the sense that*

$$P(\max_{j \notin \mathrm{N}_\mathrm{n}^\mathrm{i}} \|\hat{\Sigma}^n_{\xi^j \mathrm{N}_\mathrm{n}^\mathrm{i}} \hat{B}_n^{\mathrm{N}_\mathrm{n}^\mathrm{i}} - \hat{\Sigma}^n_{\xi^j \xi^i}\|_F \geq \lambda_n \textstyle\sum_{j \neq i}^p \|\hat{B}_n^{ij}\|_F)$$

$$\lesssim \exp\Big(- C_1 \frac{n^{1-\alpha(2+3\beta)} (\lambda_n \sum_{j \in \mathrm{N}_\mathrm{n}^\mathrm{i}} \|B^{*ij}_{m_n k_n}\|_F)^2}{n^i m_n^2 k_n^4} + 2\log(n^i m_n k_n)\Big),$$

where C_1 *is a positive constant.*