Doubly Functional Graphical Models in High Dimensions

By Xinghao Qiao, Cheng Qian, and Gareth M. James

Presented by Javier Zapata

Preliminaries: Graphical Models

- $\boldsymbol{X} = (X_1, \dots, X_p)$ are p random variables with covariance matrix $\boldsymbol{\Sigma}$
- $\boldsymbol{\Theta}^{-1} = \boldsymbol{\Sigma}$ is the precision matrix of $\boldsymbol{X}$ where:

$$\boldsymbol{\Theta}_{\mathrm{ij}} = 0 \Leftrightarrow \mathrm{Cov}\left(X_i, X_j \middle| X_{-\{i,j\}}\right) = 0$$

- If $\boldsymbol{X}$ is multivariate normal:

$$\boldsymbol{\Theta}_{\mathrm{ij}} = 0 \Leftrightarrow X_i \perp X_j \mid X_{-\{i,j\}}$$

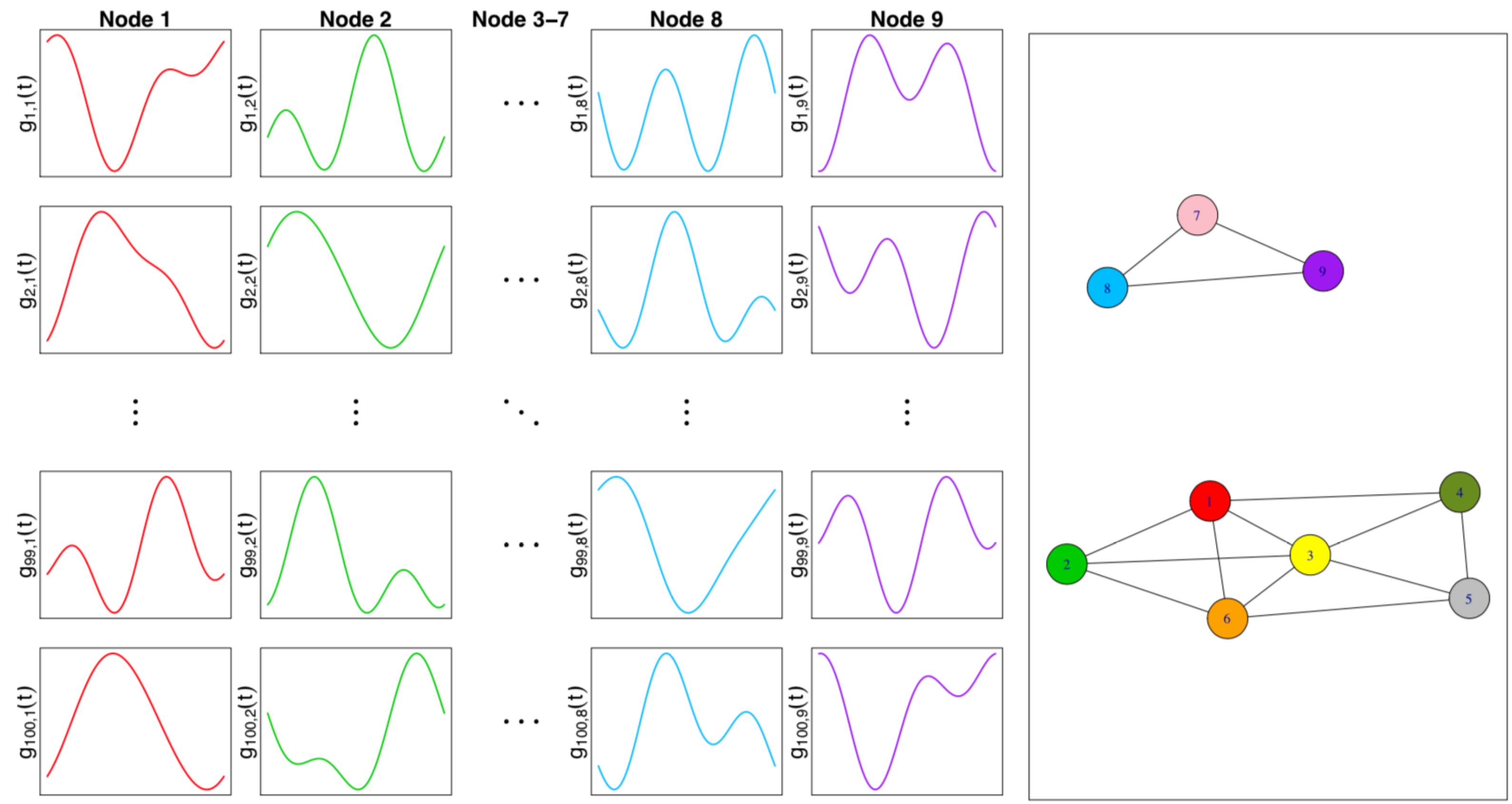
- This conditional dependence structure can be represented by an undirected graph $G = (V, E)$ where:
  - $V = \{1, \dots, p\}$ is the set of nodes
  - $E = \{(i, j): \boldsymbol{\Theta}_{\mathrm{ij}} \neq 0\}$ is the set of edges

Graphical Models for different data types

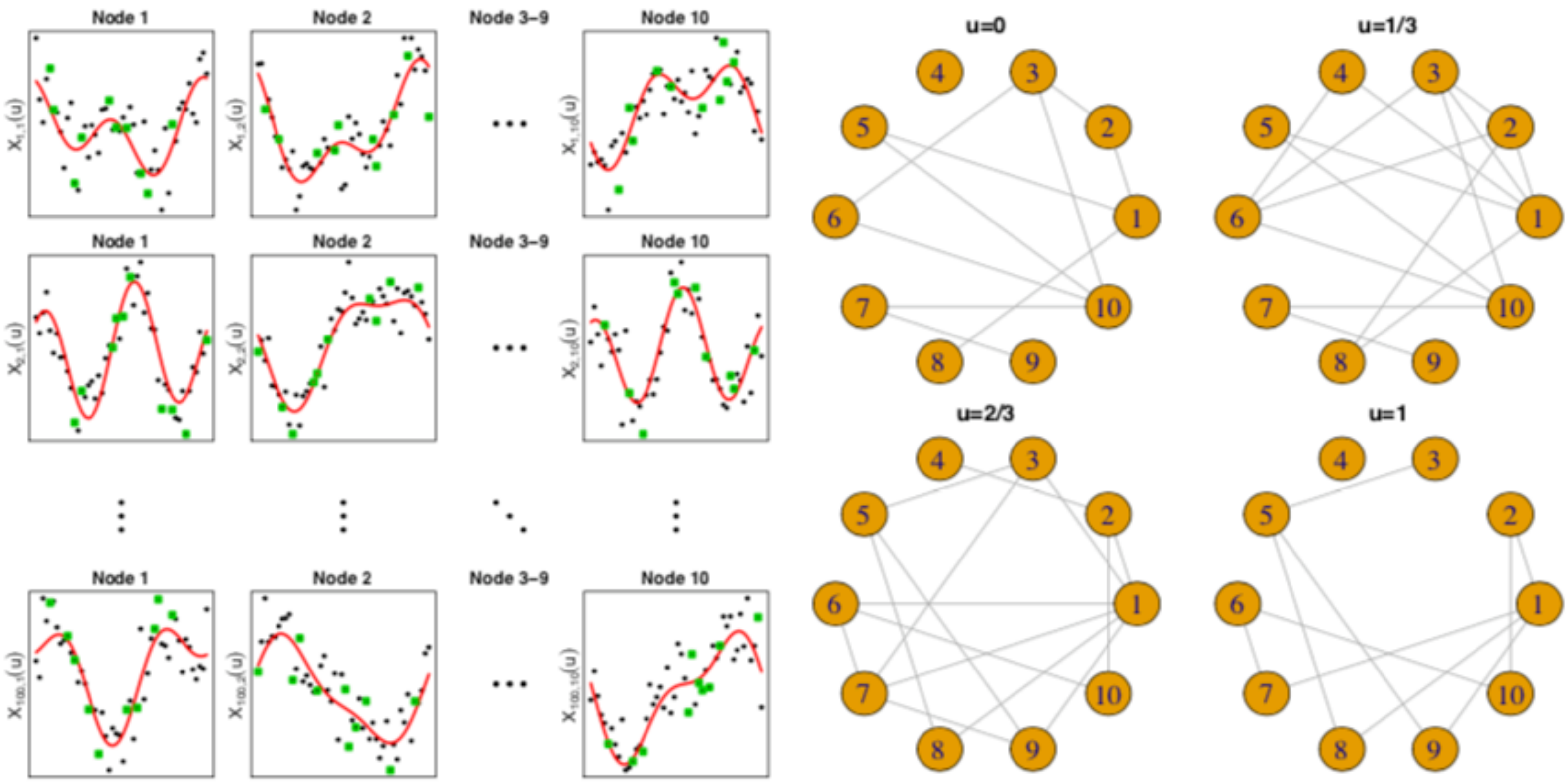
Table 1: Graphical models for different types of data and corresponding graph.

		Graphical Models	
		Static: $G = (V, E)$	Functional: $G(t) = (V, E(t))$
Data	Static: $X_1, \dots, X_p$	Gaussian graphical model	Dynamic graphical model
	Functional $X_1(t), \dots, X_p(t)$ for $t \in [0,1]$	Static functional graphical model	**Doubly functional graphical model**

Ex: Static Functional Graphical Model



Ex: Doubly Functional Graphical Model



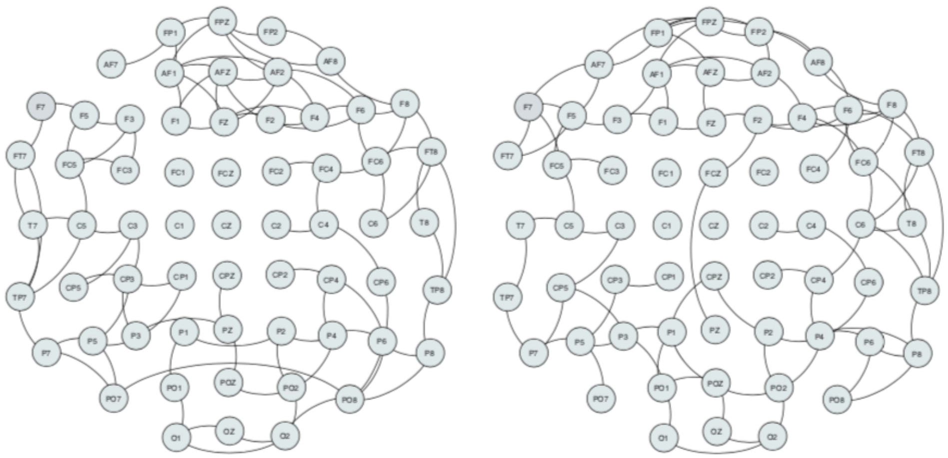
Application: EEG data

- The data consists of 77 alcoholic and 45 control subjects.
- Each subject, exposed to either a single stimulus or two stimuli, completed 120 trials. EEG signals were measured at 256 time points over a one second time interval at 64 electrodes/nodes.
- Hence:

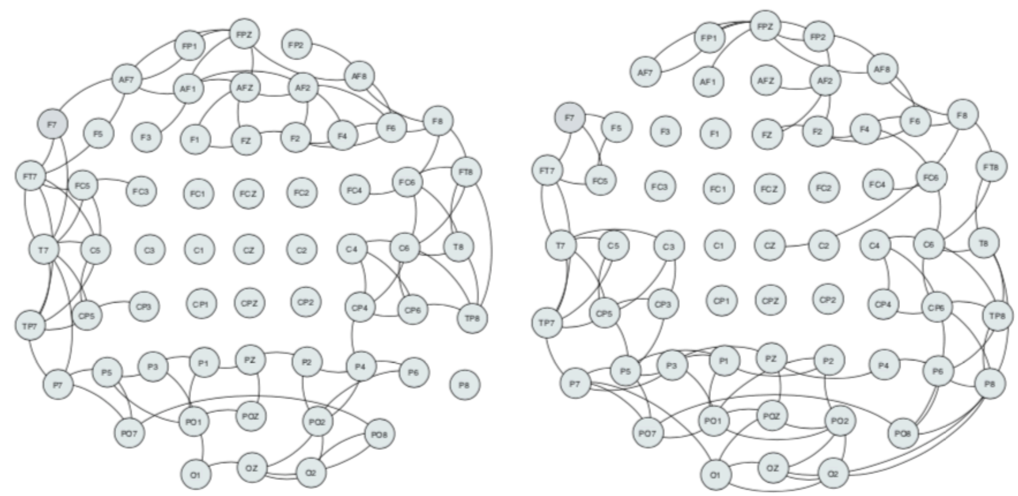
  $n_a = 77; n_c = 45$

  $\mathrm{p} = 64$

  $\mathrm{T_{ij}} = 256$ time points during $\mathcal{U}$=[0,1] (1 second)



$u = 0.2$



$u = 0.5$

Methodology

- $\boldsymbol{X}(u) = (X_1(u), \dots, X_p(u))$, $u \in \mathcal{U}$, denote a p-dimensional vector of Gaussian random functions, $X_j \in L_2(\mathcal{U})$, $\mathcal{U}$ compact subset of $\mathbb{R}$
- $\boldsymbol{C}(u, v) = \{C_{jk}(u, v)\}_{1 \leq j,k \leq p}$ with $C_{jk}(u, v) = Cov(X_j(u), X_k(v))$
- Hence: $\boldsymbol{X}(u) \sim \boldsymbol{N}(\boldsymbol{0}, \boldsymbol{\Sigma}(u)) = \boldsymbol{C}(u, u) \in \mathbb{R}^{p \times p}$, $\boldsymbol{\Theta}(u) = \boldsymbol{\Sigma}(u)^{-\mathbf{1}}$ with:
$$\boldsymbol{\Theta}_{\mathrm{ij}}(u) = 0 \Leftrightarrow \mathrm{Cov}\big(X_i(u), X_j(u) \big| \{X_l(u), l \neq j, k\}\big) = 0$$
- $G(u) = (V, E(u))$ an undirected functional graph for $u \in \mathcal{U}$ with:
$$E(u) = \{(j, k)\colon \boldsymbol{\Theta}_{\mathrm{jk}}(u) \neq 0, (j, k) \in V^2, j \neq k\}$$

When does $\mathbf{\Theta}(u) = \mathbf{\Sigma}(u)^{-1}$ exist?

- For M-dimensional $X_j(u)$ for $j = 1, \dots, p$.
- That is, X_j has KL decomposition $X_j(u) = \sum_{l=1}^{M} \xi_{jl}\, \phi_{jl}(u)$

with $\phi_{jl}, \dots, \phi_{jM}$ and $\omega_{jl}, \dots, \omega_{jM}$ the eigenfunctions and eigenvalues of X_j respectively, and principal component scores:

$$\xi_{jl} = \int_{\mathcal{U}} X_j(u)\phi_{jl}(u)du \sim N(0, \omega_{jl})$$

with ξ_{jl} independent of $\xi_{jl'}$ for $l \neq l'$

Three Steps Method:

1. Find the M-dimensional KL decomposition for $X_j(u)$:

$$X_j(u) = \sum_{l=1}^{M} \xi_{jl}\,\phi_{jl}(u)$$

2. Compute the functional covariance matrix at time $u \in \mathcal{U}$:

$$\Sigma_{jk,M}(u) = \sum_{l=1}^{M}\sum_{m=1}^{M} \mathrm{Cov}(\xi_{jl}, \xi_{km})\phi_{jl}(u)\phi_{km}(u).$$

3. Compute the precision matrix: $\boldsymbol{\Theta}_{\mathbf{M}}(u) = \boldsymbol{\Sigma}_{\boldsymbol{M}}(u)^{-\mathbf{1}}$

Estimation

- $\boldsymbol{X_i}(u) = \left(X_{i1}(u), \dots, X_{ip}(u)\right)^T$ for $i = 1, \dots, n$ (copies of $\boldsymbol{X}(u)$)
- $X_{ij}(u)$ observed without measurement errors at $U_{ijt} \in \mathcal{U}$ for $t = 1, \dots, T_{ij}$
- Y_{ijt} represent the observed value of $X_{ij}(U_{ijt})$:

$$Y_{ijt} = X_{ij}(U_{ijt}) + e_{ijt} = \sum_{l=1}^{\infty} \xi_{ijl}\phi_{jl}(U_{ijt}) + e_{ijt}$$

where the e_{ijt}'s are i.i.d. with $E(e_{ijt}) = 0$ and $\mathrm{Var}(e_{ijt}) = \sigma^2$, independent of X_{ij}, and the U_{ijt}'s are sampled from some specific density f_U.

Step 1. To perform functional principal components analysis based on realizations $\mathbf{Y}_{ij} = (Y_{ij1}, \ldots, Y_{ijT_{ij}})^{\mathrm{T}}, i = 1, \ldots, n$, for each $j \in V$, we first compute the estimator for $C_{jj}(u, v)$. Let $\boldsymbol{\Sigma}_{\mathbf{Y}_{ij}}$ be the covariance matrix for $\mathbf{Y}_{ij}$ with (t, t')-th element $\left(\boldsymbol{\Sigma}_{\mathbf{Y}_{ij}}\right)_{tt'} = \mathrm{Cov}(Y_{ijt}, Y_{ijt'}) = C_{jj}(U_{ijt}, U_{ijt'}) + \sigma^2 I(t = t')$. A local linear surface smoother is applied to the off-diagonals of the "raw covariances", $Y_{ijt}Y_{ijt'}, t \neq t'$. Denote $K_h(\cdot) = h^{-1}K(\cdot/h)$ for a univariate kernel function K with a positive bandwidth h. We consider minimizing

$$\sum_{i=1}^{n} w_{ij} \sum_{1 \leqslant t \neq t' \leqslant T_{ij}} \left\{Y_{ijt}Y_{ijt'} - \beta_0 - \beta_1(U_{ijt} - u) - \beta_2(U_{ijt'} - v)\right\}^2 K_{h_j}(U_{ijt} - u)K_{h_j}(U_{ijt'} - v), \quad (4)$$

with respect to $(\beta_0, \beta_1, \beta_2)$, where the weight w_{ij} is chosen for ith subject and the jth variable such that $\sum_{i=1}^{n} T_{ij}(T_{ij} - 1)w_{ij} = 1$.

We next perform eigen-decomposition on $\widehat{C}_{jj}(u,v)$ and obtain the estimated eigen-pairs $(\widehat{\omega}_{jl}, \widehat{\phi}_{jl}), l = 1, \ldots, M$. The estimated principal component scores are $\widehat{\xi}_{ijl} = \int_{\mathcal{U}} \widehat{X}_{ij}(u)\widehat{\phi}_{jl}(u)du$. However, this approach requires the estimated trajectories, $\widehat{X}_{ij}(u)$, which are unavailable, especially for sparse designs. Instead, we propose to use the best linear unbiased predictors $\widetilde{\xi}_{ijl} = \boldsymbol{\zeta}_{ijl}^{\mathrm{T}} \boldsymbol{\Sigma}_{\mathbf{Y}_{ij}}^{-1} \mathbf{Y}_{ij}$ (Rice and Wu, 2001), where $\boldsymbol{\zeta}_{ijl}$ is a T_{ij}-dimensional vector with t-th component

$$\zeta_{ijlt} = \mathrm{Cov}(\xi_{ijl}, Y_{ijt}) = E\Big\{\int X_{ij}(v)\phi_{jl}(v)dvX_{ij}(U_{ijt})\Big\} = \int C_{jj}(U_{ijt}, v)\phi_{jl}(v)dv.$$

From (Rice and Wu, 2001):

Conditioning on p and q, (2) is a classical linear mixed effects model, and the vector of observations on the ith subject can be expressed as

$$Y_i = X_i\beta + Z_i\gamma_i + \epsilon_i. \tag{4}$$

The covariance matrix of Y_i is $V_i = Z_i\Gamma Z_i^{\mathrm{T}} + \sigma^2 I$. We can thus use the methodology that has been developed for mixed effect models in this nonparametric context. Estimation of the parameters β, σ^2, and the covariance matrix Γ is accomplished by the EM algorithm (Laird and Ware, 1982). The BLUP estimate (Robinson, 1991) of the spline coefficients of the random effect for subject i is

$$\hat{\gamma}_i = \hat{\Gamma} Z_i^{\mathrm{T}} \left(Z_i\hat{\Gamma}Z_i^{\mathrm{T}} + \hat{\sigma}^2 I\right)^{-1} (Y_i - X_i\hat{\beta}). \tag{5}$$

Note, although we do not place any distributional assumptions on the errors, when e_{ijt} and ξ_{ijl} are jointly Gaussian, $\widetilde{\xi}_{ijl}$ reduces to the conditional expectation of ξ_{ijl} given $\mathbf{Y}_{ij}$ (Yao et al., 2005). We then obtain the estimator for $\widetilde{\xi}_{ijl}$ as

$$\widehat{\xi}_{ijl} = \widehat{\boldsymbol{\zeta}}_{ijl}^{\mathrm{T}} \widehat{\boldsymbol{\Sigma}}_{\mathbf{Y}_{ij}}^{-1} \mathbf{Y}_{ij}, \tag{5}$$

where $\widehat{\zeta}_{ijlt} = \int \widehat{C}_{jj}(U_{ijt}, v)\widehat{\phi}_{jl}(v)dv$, and $\left(\widehat{\boldsymbol{\Sigma}}_{\mathbf{Y}_{ij}}\right)_{tt'} = \widehat{C}_{jj}(U_{ijt}, U_{ijt'}) + \widehat{\sigma}^2 I(t = t')$. See Yao et al. (2005) for details on the estimate $\widehat{\sigma}^2$ of σ^2.

Step 2. Once the functional principal components analysis has been performed, we substitute the terms in (2) by their estimated values and thus obtain $\widehat{\boldsymbol{\Sigma}}(u)$ with its (j,k)-th entry given by $\widehat{\Sigma}_{jk}(u) = n^{-1}\sum_{i=1}^{n}\sum_{l=1}^{M}\sum_{m=1}^{M}\widehat{\xi}_{ijl}\widehat{\xi}_{ikm}\widehat{\phi}_{jl}(u)\widehat{\phi}_{km}(u)$.

Step 3. Finally, for a set of points $u \in \mathcal{U}$, we estimate $\Theta_{jk}(u)$. One of the advantages of our approach is that a variety of standard sparse precision matrix methods can be used to implement this step. Our empirical results suggest that the constrained ℓ_1-minimization (Cai et al., 2011) provides the most accurate results so we use that approach here. To be specific, we solve the following constrained optimization problem

$$\check{\boldsymbol{\Theta}}(u) = \underset{\boldsymbol{\Theta}\in\mathbb{R}^{p\times p}}{\arg\min}|\boldsymbol{\Theta}|_1 \quad \text{subject to } |\widehat{\boldsymbol{\Sigma}}(u)\boldsymbol{\Theta} - \mathbf{I}|_\infty \leqslant \lambda_n(u), \tag{6}$$

where $\mathbf{I} \in \mathbb{R}^{p\times p}$ is the identity matrix and $\lambda_n(u) \geqslant 0$ is a tuning parameter which controls the sparsity level of $\check{\boldsymbol{\Theta}}(u)$.

The convex problem (6) can be further decomposed into p separate optimization problems. For $j = 1, \ldots, p$, we solve

$$\widehat{\boldsymbol{\beta}}_j(u) = \underset{\boldsymbol{\beta} \in \mathbb{R}^p}{\arg\min} |\boldsymbol{\beta}|_1 \quad \text{subject to } |\widehat{\boldsymbol{\Sigma}}(u)\boldsymbol{\beta} - \mathbf{e}_j|_\infty \leqslant \lambda_n(u), \tag{7}$$

where $\mathbf{e}_j \in \mathbb{R}^p$ is the unit vector with j-th coordinate 1 and $\widehat{\boldsymbol{\beta}}_j(u)$ corresponds to the j-th column of $\breve{\boldsymbol{\Theta}}(u)$.

Our target estimator $\widehat{\boldsymbol{\Theta}}(u)$ is attained by the final step of symmetrizing $\breve{\boldsymbol{\Theta}}(u)$ whose (j, k) and (k, j)-th entries are obtained by

$$\widehat{\Theta}_{jk}(u) = \widehat{\Theta}_{kj}(u) = \breve{\Theta}_{jk}(u) I\big\{|\breve{\Theta}_{jk}(u)| \leqslant |\breve{\Theta}_{kj}(u)|\big\} + \breve{\Theta}_{kj}(u) I\big\{|\breve{\Theta}_{jk}(u)| > |\breve{\Theta}_{kj}(u)|\big\}. \tag{8}$$

This symmetrization procedure guarantees that our estimator $\widehat{\boldsymbol{\Theta}}(u)$ achieves the same elementwise ℓ_∞ estimation error rate as $\breve{\boldsymbol{\Theta}}(u)$.

Bibliography

- Qiao, Qian & James - Doubly Functional Graphical Models in High Dimensions, 2019

Cai, T., Liu, W. and Luo, X. (2011). A constrained l_1 minimization approach to sparse precision matrix estimation, *Journal of the American Statistical Association* **106**: 594–607.

Rice, J. and Wu, C. O. (2001). Nonparametric mixed effect models for unequally sampled noisy curves, *Biometrics* **57**: 253–259.