

Copula Gaussian graphical models for functional data

Eftychia Solea and Bing Li

University of Cyprus and Pennsylvania State University

Abstract

We introduce a statistical graphical model for multivariate functional data, which are common in medical applications such as EEG and fMRI. Recently published functional graphical models rely on the multivariate Gaussian process assumption, but we relax it by introducing the Functional Copula Gaussian Graphical Model (FCGGM). This model removes the marginal Gaussian assumption but retains the simplicity of the Gaussian dependence structure, which is particularly attractive for large data. We develop four estimators for the FCGGM and establish the consistency and the convergence rates of one of them. We compare our FCGGM with the existing functional Gaussian graphical model by **simulations**, and apply our method to an EEG data set to construct brain networks.

Keywords: Conditional independence; Covariance operator; EEG data; Karhunen-Loeve expansion; Non-Gaussian random functions; Precision operator; Rank transformation.

1 Introduction

Functional graphical models were recently developed by Zhu et al. (2016), Qiao et al. (2019a), and Li and Solea (2018) to construct networks with function-valued observations. This type of data arises frequently in medical applications such as EEG and fMRI. See, for example, Lazar et al. (2002), Cheng and Herskovits (2007), and Li, Kim, and Altman (2010). The functional graphical model is a continuation of the recent research on graphical models for scalar-valued observations (Meinshausen and Bühlmann, 2006; Yuan and Lin, 2007), which can be traced back to Darroch et al. (1980) and Lauritzen et al. (1984). Other important early references include Wer-muth and Lauritzen (1983, 1990), Dawid and Lauritzen (1993), Whittaker (1990), and Lauritzen (1996).

The functional Gaussian graphical model (FGGM) of Qiao et al. (2019a) is developed under the multivariate Gaussian process assumption; it applied group lasso to the coefficients of the Karhunen Loeve expansions (Bosq, 2000). Also under the Gaussian assumption, Zhu et al. (2016) introduced a Bayesian approach by imposing an invert Wishart prior distribution on the covariance matrix of the vector-valued functional data and integrating out the covariance operator. Li and Solea (2018) introduced a nonparametric functional graphical model based on the *additive conditional independence* introduced by Li et al. (2014). Furthermore, a dynamic functional graphical model was developed recently by Qiao et al. (2019b) under the Gaussian assumption.

In this paper we extend the functional Gaussian graphical model to the functional copula Gaussian graphical model. To explain the ideas of our extension, we first give an overview the recent developments of the copula Gaussian graphical model in the classical setting. A special character of the multivariate Gaussian distribution is that conditional independence is completely specified by the second moments, so that the

estimation of the edge set can be reduced to sparse estimation of the precision matrix. However, the Gaussian assumption is very restrictive: skewness and kurtosis are but two of many ways that it can be violated. To relax the Gaussian assumption while retaining its simple conditional independence structure, Liu, Lafferty, and Wasserman (2009), Liu et al. (2012), and Xue and Zou (2012) proposed several versions of copula Gaussian graphical model. The copula model assumes that the random variables can be marginally transformed to multivariate Gaussian, and leads to substantial gain in accuracy under marginal violation of the Gaussian assumption. Inspired by the above developments we propose a copula Gaussian model for a vector of random functions, leading to the Functional Copula Gaussian Graphical Model (FCGGM). In doing so we encounter two challenges. One is that a random function has no marginal variables to apply the copula transformations to, and we solve this by applying them to the coefficients of the Karhunen-Loeve expansions. The other is that the copula transformations are not applied to observed data, but instead to estimated quantities, which means the standard asymptotic tools for the copula model cannot be applied, and special techniques need to be developed.

The significance of a copula Gaussian model for functional data goes far beyond the current setting: we expect it to have wide applications in functional data analysis, such as variable selection, variable screening, and functional time series analysis. Furthermore, since many dimension reduction methods require the predictors to have an elliptical distribution, the proposed copula model opens up wide possibilities for developing simple and efficient functional dimension reduction methods. See Li and Song (2017) for a recent development in functional sufficient dimension reduction.

In Sections 2 and 3, we introduce the copula Gaussian random function and the FCGGM. In Sections 4 and 5, we develop estimation methods and establish their consistency and convergence rates. In Sections 6 and 7, we compare FCGGM with FGGM by **simulations** and apply the former to an EEG data. Some concluding re-

marks are made in Section 8. All the proofs are given in the Supplementary Material.

2 Copula Gaussian random functions

In this section we give a rigorous definition of the copula Gaussian random element, and lay out some basic concepts and notations that will be used in the rest of the paper.

The extension of the copula model to functional data is not as straightforward as it might seem, not least because, unlike in the multivariate case, here we do not have natural “marginals” on which to impose the copula assumption. At first glance it might seem plausible to impose the copula Gaussian assumption on the observations $X(t)$ themselves. However, it is theoretically cumbersome to do so because t varies over an uncountable set. Another possibility is to impose the copula assumption on the linear coefficients of a pre-assigned orthonormal basis, such as the Fourier series. But this seems arbitrary because there are infinitely many orthonormal bases. Our idea is to impose the copula assumption on the coefficients in the Karhunen-Loeve expansion, which is independent of the choices of basis at the population level.

For two generic Hilbert spaces, say \mathcal{S} and \mathcal{T} , let $\mathcal{B}(\mathcal{S}, \mathcal{T})$ denote the class of all bounded linear operators from \mathcal{S} to \mathcal{T} . For $A \in \mathcal{B}(\mathcal{S}, \mathcal{T})$, let $\ker(A)$ denote the kernel of A , that is, $\{\phi \in \mathcal{S} : A(\phi) = 0\}$, let $\text{ran}(A)$ denote the range of A ; that is, $\{A\phi : \phi \in \mathcal{S}\}$, and let $\overline{\text{ran}}(A)$ denote closure of $\text{ran}(A)$. For $s \in \mathcal{S}$ and $t \in \mathcal{T}$, their tensor product $t \otimes s$ is the operator from \mathcal{S} to \mathcal{T} that maps an $h \in \mathcal{S}$ to $\langle s, h \rangle_{\mathcal{S}} t \in \mathcal{T}$. Let \mathbb{N} be the set of positive integers $\{1, 2, \dots\}$.

Let (Ω, \mathcal{F}, P) be a probability space. A random element U in \mathcal{S} is a mapping from Ω to \mathcal{S} measurable with respect to the Borel σ -field generated by the open sets in \mathcal{S} ; a random element $A \in \mathcal{B}(\mathcal{S}, \mathcal{T})$ is a mapping from Ω to $\mathcal{B}(\mathcal{S}, \mathcal{T})$ measurable with respect to the Borel σ -field generated by the open sets in $\mathcal{B}(\mathcal{S}, \mathcal{T})$. We make

the following assumption.

Assumption 1. $E\|U\|_{\mathcal{S}}^2 < \infty$.

This assumption implies $E\|U\|_{\mathcal{S}} < \infty$, under which the linear functional $\mathcal{S} \rightarrow \mathbb{R}$, $s \mapsto E\langle s, U \rangle_{\mathcal{S}}$ is bounded and its Riesz representation is defined as the mean of U , written as $\mu_U = E(U)$. Under Assumption 1, the linear operator

$$E[(U - E(U)) \otimes (U - E(U))]$$

is also a well defined trace-class operator, and is called the covariance operator of U , written as $\Sigma_{UU} = \text{var}(U)$. See, for example, Bosq (2000). In the rest of the paper we take $\mathcal{S} = L_2(T)$. Thus, U is a mapping from Ω to $L_2(T)$ measurable with respect to the Borel σ -field in $L_2(T)$. This amounts to assuming, for each $\omega \in \Omega$, the sample path $U(\omega)$ is a square-integrable function with respect to the Lebesgue measure on T . More rigorously, U should be defined as a random element in the quotient space $L_2(T)/\sim$, where \sim is the equivalence relation defined by almost sure equality with respect to the Lebesgue measure on T . This would avoid the problem that a member of $L_2(T)$ cannot be uniquely identified by the $L_2(T)$ -norm. See, for example, Li and Babu (2019, page 226) for a detailed discussion on this point. With some extra conditions that can identify an equivalence class in $L_2(T)/\sim$ with a single function (e.g. continuous sample path), we can make explicit connections between the mean element μ_U and the covariance operator Σ_{UU} with pointwise means and covariances of the stochastic process $\{U(t) : t \in T\}$: $\mu_U(t)$ is, in fact, $E[U(t)]$, and Σ_{UU} is the linear operator

$$\Sigma_{UU}(f) = \int_T f(s) \sigma_{UU}(s, t) dt,$$

where $\sigma_{UU}(s, t) = \text{cov}[U(s), U(t)]$.

We now give a formal definition of a Gaussian random element in \mathcal{S} .

Definition 1. *A random element U of \mathcal{S} is Gaussian if, for any $t \in \mathcal{S}$,*

$$E[\exp(\iota \langle t, U \rangle_{\mathcal{S}})] = \exp[\iota \langle t, \mu_U \rangle_{\mathcal{S}} - (1/2) \langle t, \Sigma_{UU} t \rangle_{\mathcal{S}}], \quad (2.1)$$

where $\iota = \sqrt{-1}$.

Equivalently, a random element in \mathcal{S} is Gaussian if and only if, for each $f \in \mathcal{S}$, $\langle f, U \rangle_{\mathcal{S}}$ is a Gaussian random variable. Since Σ_{UU} is a trace-class operator, $U - \mu_U$ can be expressed as the Karhunen-Loeve expansion (Bosq, 2000, Theorem 1.5):

$$U - \mu_U = \sum_{r \in \mathbb{N}} \lambda_r^{1/2} \xi_r \phi_r. \quad (2.2)$$

where $\lambda_1 \geq \lambda_2 \dots$ and ϕ_1, ϕ_2, \dots are the eigenvalues and eigenfunctions of Σ_{UU} , and ξ_1, ξ_2, \dots are i.i.d. $N(0, 1)$ variables.

We first consider univariate random functions; that is, consider the special case where \mathcal{S} is a Hilbert space of functions defined on an interval $T \subseteq \mathbb{R}$ and taking values in \mathbb{R} . We call such a space a Hilbert space of \mathbb{R} -valued functions.

Definition 2. *Let U be a random element in a Hilbert space \mathcal{S} of \mathbb{R} -valued functions, and suppose it has Karhunen-Loeve expansion (2.2). We say that U follows a copula Gaussian distribution if there exists a sequence of monotone increasing functions c_1, c_2, \dots defined on \mathbb{R} with $E[c_r(\xi_r)] = 0$, $\text{var}[c_r(\xi_r)] = 1$ such that*

$$V = \sum_{r \in \mathbb{N}} \lambda_r^{1/2} c_r(\xi_r) \phi_r$$

is a Gaussian element in \mathcal{S} with the right-hand side as its Karhunen-Loeve expansion.

Note that, since $N(0, 1)$ is symmetric, $c_r(\xi_r) \sim N(0, 1)$ if and only if $-c_r(\xi_r) \sim N(0, 1)$. So the assumption “monotone increasing function” can be replaced by

“monotone function”. We restrict ourselves to monotone increasing function — which does not lose generality — to avoid complication in the theoretical development. The next proposition gives a sufficient condition for a random element in \mathcal{S} to have a copula Gaussian distribution.

Proposition 1. *If U has Karhunen-Loeve expansion (2.2) where ξ_1, ξ_2, \dots are continuous and independent random variables, then U has a copula Gaussian distribution.*

To see what is involved in the univariate functional copula assumption in Definition 2, recall that, in the classical setting, any continuous random variable U can be transformed to Gaussian by the transformation $\Phi^{-1} \circ F_U$, where F_U is the cumulative distribution function of U . In the functional setting, in addition to requiring the scores ξ_r to be continuous, we also require them to be independent. Since by definition ξ_r are already uncorrelated, the real extra assumption in the univariate functional copula model is to strengthen the uncorrelation of ξ_r to independence of ξ_r .

The assumption in Proposition 1 that ξ_r are independent is not unduely strong — it defines a much larger class of random elements in \mathcal{S} than the class of Gaussian random elements. Indeed, if U is a Gaussian random element in \mathcal{S} having a Karhunen-Loeve expansion $\sum_{r=1}^{\infty} \lambda_r^{1/2} \xi_r \phi_r$, then any series of the form $\sum_{r=1}^{\infty} \lambda_r^{1/2} f_r(\xi_r) \phi_r$, where f_r are injections such that $E[f_r(\xi_r)] = 0$ and $\text{var}[f_r(\xi_r)] = 1$, is the Karhunen-Loeve expansion of some non-Gaussian random element in \mathcal{S} . Thus, the functional copula model does significantly expand the applicability of the Gaussian model.

Another piece of intuition in this construction is that we are not interested in the dependence structure in U itself: by imposing copula transformation on the $\langle \phi_r, U \rangle_{\mathcal{S}}$ we in effect leave the dependence structure of U intact.

Let

$$\rho_r = F_r(\xi_r), \quad \eta_r = \Phi^{-1}(\rho_r).$$

We refer to ρ_r as the ranked scores and η_r as the normal scores of the random element U . The next corollary gives a specific form of the Gaussian copula functions.

Corollary 1. *If U is an \mathbb{R} -valued copula Gaussian random function for some sequence $C = \{c_r : r \in \mathbb{N}\}$ of monotone increasing functions, then $c_r = \Phi^{-1} \circ F_r$, where F_r is the c.d.f. of ξ_r .*

We denote the sequence $\{\Phi^{-1} \circ F_r : r \in \mathbb{N}\}$ by C_U , and denote the transformed random element $\sum_{r \in \mathbb{N}} \lambda_r^{1/2} \Phi^{-1} \circ F_r(\xi_r)$ by $C_U(U)$.

We now turn to vector-valued random functions. For each $i = 1, \dots, p$, let \mathcal{H}_i be a separable Hilbert space of \mathbb{R} -valued functions on T with inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}_i}$. Let $\oplus_{i=1}^p \mathcal{H}_i$ be the direct sum of $\mathcal{H}_1, \dots, \mathcal{H}_p$. That is, $\oplus_{i=1}^p \mathcal{H}_i$ is the Cartesian product $\mathcal{H}_1 \times \dots \times \mathcal{H}_p$ with its inner product defined by

$$\langle f, g \rangle_{\oplus \mathcal{H}} = \sum_{i=1}^p \langle f_i, g_i \rangle_{\mathcal{H}_i}$$

where f and g are members of $\oplus_{i=1}^p \mathcal{H}_i$ and f_i and g_i are the i th components of f and g , respectively. Let $X = (X^1, \dots, X^p)$ be a random element in $\oplus_{i=1}^p \mathcal{H}_i$. For each $(i, j) \in \mathbb{V} \times \mathbb{V}$, define the covariance operator between X^j and X^i as

$$\Sigma_{X^j X^i} = \text{cov}(X^j, X^i) = E[(X^i - \mu_{X^i}) \otimes (X^j - \mu_{X^j})].$$

This operator has a one-to-one correspondence with the kernel function $\sigma_{X^i, X^j}(s, t) = \text{cov}[X^i(s), X^j(t)]$. Note that, $\Sigma_{X^j X^i} \in \mathcal{B}(\mathcal{H}_i, \mathcal{H}_j)$. We say that X is a Gaussian random element of $\oplus_{i=1}^p \mathcal{H}_i$ if, for each $(t_1, \dots, t_p) \in \oplus_{i=1}^p \mathcal{H}_i$,

$$E \left[\exp \left(\iota \sum_{i \in \mathbb{V}} \langle t_i, X^i \rangle_{\mathcal{H}_i} \right) \right] = \exp \left[\iota \sum_{i \in \mathbb{V}} \langle t_i, \mu_{X^i} \rangle_{\mathcal{H}_i} - \sum_{i, j \in \mathbb{V}} \langle t_i, \Sigma_{X^i X^j} t_j \rangle_{\mathcal{H}_i} / 2 \right].$$

Define Σ_{XX} to be the following operator

$$\oplus_{i=1}^p \mathcal{H}_i \rightarrow \oplus_{i=1}^p \mathcal{H}_i, \quad (t_1, \dots, t_p) \mapsto (\sum_{i \in \mathbf{V}} \Sigma_{X^1 X^i} t_i, \dots, \sum_{i \in \mathbf{V}} \Sigma_{X^p X^i} t_i).$$

Intuitively, Σ_{XX} can be interpreted as the $p \times p$ matrix whose (i, j) th entry is $\Sigma_{X^i X^j}$; that is,

$$\Sigma_{XX} = \begin{pmatrix} \Sigma_{X^1 X^1} & \cdots & \Sigma_{X^1 X^p} \\ \vdots & \ddots & \vdots \\ \Sigma_{X^p X^1} & \cdots & \Sigma_{X^p X^p} \end{pmatrix}.$$

In general, suppose, for each $(i, j) \in \mathbf{V} \times \mathbf{V}$, A_{ij} is an operator in $\mathcal{B}(\mathcal{H}_j, \mathcal{H}_i)$. We define the matrix of operators $A = \{A_{ij}\}_{i,j=1}^p$ as the mapping

$$A : \oplus_{i=1}^p \mathcal{H}_i \rightarrow \oplus_{i=1}^p \mathcal{H}_i, \quad (t_1, \dots, t_p) \mapsto (\sum_{\ell=1}^p A_{1\ell} t_\ell, \dots, \sum_{\ell=1}^p A_{p\ell} t_\ell).$$

The class of all such operators are denoted by $\times_{i,j \in \mathbf{V}} \mathcal{B}(\mathcal{H}_i, \mathcal{H}_j)$. Notice that

$$\times_{i,j \in \mathbf{V}} \mathcal{B}(\mathcal{H}_i, \mathcal{H}_j) = \mathcal{B}(\oplus_{i=1}^p \mathcal{H}_i, \oplus_{i=1}^p \mathcal{H}_i).$$

Using this convention of matrix of operators we can write a Gaussian random element X equivalently as

$$E[\exp(\iota \langle t, X \rangle_{\oplus \mathcal{H}_i})] = \exp[\iota \langle t, \mu_X \rangle_{\oplus \mathcal{H}_i} - \langle t, \Sigma_{XX} t \rangle_{\oplus \mathcal{H}_i} / 2],$$

which has the same form as (2.1). We now define the copula Gaussian random element in $\oplus_{i=1}^p \mathcal{H}_i$.

Definition 3. Suppose $X = (X^1, \dots, X^p)$ is a random element in $\oplus_{i=1}^p \mathcal{H}_i$, where

each X^i is a copula Gaussian random element in \mathcal{H}_i . We say that X is a copula Gaussian random element in $\oplus_{i=1}^p \mathcal{H}_i$ iff the \mathbb{R}^p -valued random function

$$(C_{X^1}(X^1), \dots, C_{X^p}(X^p))$$

is a Gaussian random element in $\oplus_{i=1}^p \mathcal{H}_i$.

As we mentioned earlier the direct reason for us to impose the copula transformations on the Karhunen-Loeve coefficients is that we do not have natural marginals to impose it on. Here, we would like to further argue that, even if the Hilbert space is finite-dimensional, where X^i does have “natural” coordinate, it is still reasonable to assign copula transformations to the Karhunen-Loeve coefficients (or the principal components). For example, suppose X^i lies in a space spanned by the polynomials $1, t, t^2, \dots, t^k$, then there seems no more reason to believe that all the non-Gaussian feature happens along the coefficients of these polynomials than along the principal components. In other words, imposing copula transformations on principal components may well be a viable alternative even in the finite-dimensional setting.

3 Functional copula Gaussian graphical models

As in Qiao, Guo, and James (2019a), we use conditional independence to define graphical models for random functions.

Definition 4. A random element in $\oplus_{i=1}^p \mathcal{H}_i$ is said to follow a functional graphical model with respect to an undirected graph $G = (V, E)$ iff

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

Furthermore, if X is a copula Gaussian random element in $\oplus_{i=1}^p \mathcal{H}_i$ then we say X

follows a functional copula Gaussian graphical model, and write this statement as $X \sim \text{FCGGM}(\mathbf{G})$.

In the classical setting, the most attractive feature of the Gaussian graphical model is its ability to encode conditional independence as the zero entries of the precision matrix, so that estimating an undirected graph reduces to sparse estimation of the precision matrix. More specifically, suppose $Z = (Z^1, \dots, Z^p)$ is a multivariate Gaussian random vector with covariance matrix Σ , and let $\Theta = \Sigma^{-1}$ be its precision matrix. Then

$$Z^i \perp\!\!\!\perp Z^j | Z^{-\{i,j\}} \quad \text{iff} \quad \theta_{ij} = 0, \quad (3.1)$$

where θ_{ij} is the (i, j) th element of Θ . The copula Gaussian graphical models of Liu, Lafferty, and Wasserman (2009) and Xue and Zou (2012) inherit this property but allows the marginal distributions of Z^i to be non-Gaussian.

Our motivation for introducing the functional copula Gaussian model is to inherit the Gaussian-like conditional independence structure without having to assume X to be a Gaussian random function. Specifically, the following scheme seems plausible. Suppose X is a copula Gaussian random element in $\oplus_{i=1}^p \mathcal{H}_i$ and let $Z = (C_{X^1}(X^1), \dots, C_{X^p}(X^p))$. Then Z is a Gaussian random element in $\oplus_{i=1}^p \mathcal{H}_i$. Let $\Theta_{ZZ} = \Sigma_{ZZ}^{-1}$. Then

$$X^i \perp\!\!\!\perp X^j | X^{-\{i,j\}} \Leftrightarrow Z^i \perp\!\!\!\perp Z^j | Z^{-\{i,j\}} \Leftrightarrow \Theta_{Z^i Z^j} = 0,$$

where the first equivalence holds because there is a one-to-one correspondence between X^i and Z^i , and the second holds because Z is a Gaussian random element. However, there is additional complication here: since Σ_{XX} is a trace-class operator, its inverse is an unbounded operator, which makes it problematic to use Σ_{XX}^{-1} directly as the

object of estimation. To avoid this difficulty, we introduce the correlation operator of a random element in $\oplus_{i=1}^p \mathcal{H}_i$.

Let \mathcal{R}_i denote the subspace $\overline{\text{ran}}(\Sigma_{Z^i Z^i})$, \mathcal{R}_j the subspace $\overline{\text{ran}}(\Sigma_{Z^j Z^j})$, and $P_{\mathcal{R}_i}$ and $P_{\mathcal{R}_j}$ the projections on to \mathcal{R}_i and \mathcal{R}_j , respectively. By Baker (1973), there is a unique operator $R_{Z^i Z^j} \in \mathcal{B}(\mathcal{H}_j, \mathcal{H}_i)$ such that $\|R_{Z^i Z^j}\| \leq 1$, $R_{Z^i Z^j} = P_{\mathcal{R}_i} R_{Z^i Z^j} P_{\mathcal{R}_j}$, and

$$\Sigma_{Z^i Z^i}^{1/2} R_{Z^i Z^j} \Sigma_{Z^j Z^j}^{1/2} = \Sigma_{Z^i Z^j}, \quad (3.2)$$

where $\Sigma_{Z^i Z^i}^{1/2}$ is the square-root of the operator $\Sigma_{Z^i Z^i}$. That is,

$$\Sigma_{Z^i Z^i}^{1/2} = \sum_{r=1}^{\infty} (\lambda_r^i)^{1/2} (\phi_r^i \otimes \phi_r^i),$$

where $\{(\lambda_r^i, \phi_r^i) : r \in \mathbb{N}\}$ are the eigenvalues and eigenfunctions of $\Sigma_{X^i X^i}$. The operator $R_{Z^i Z^j} : \mathcal{H}_j \rightarrow \mathcal{H}_i$ is called the correlation operator between Z^i and Z^j . Note that, when $i = j$, the identity mapping $I_{\mathcal{H}_i}$ in \mathcal{H}_i satisfies the above relation. Hence $R_{Z^i Z^i} = I_{\mathcal{H}_i}$, which is consistent with the classical definition of correlation. Using this operator we define the correlation operator of a random element in $\oplus_{i=1}^p \mathcal{H}_i$ as follows.

Definition 5. Suppose $Z = (Z^1, \dots, Z^p)$ is random element in $\oplus_{i=1}^p \mathcal{H}_i$ with $E\|Z\|^2 < \infty$. Let $R_{Z^i Z^j}$ be the correlation operator between Z^i and Z^j . We call the operator $R_{ZZ} = \{R_{Z^i Z^j}\}_{i,j \in \mathcal{V}} \in \times_{i,j \in \mathcal{V}} \mathcal{B}(\mathcal{H}_i, \mathcal{H}_j)$ the additive correlation operator of Z .

It turns out that, unlike additive covariance operator Σ_{ZZ} , the additive correlation operator R_{ZZ} has a bounded inverse under very reasonable assumptions.

Assumption 2. For $i \neq j$, $R_{Z^i Z^j}$ are compact; R_{ZZ} is invertible.

This assumption was also made in Bach (2008). The assumption is quite mild: as argued in Li (2018), the compact assumption of $R_{Z^i Z^j}$ imposes a type of collective smoothness between the functions in \mathcal{H}_i and functions in \mathcal{H}_j . Also note that R_{ZZ}

is invertible if and only if Σ_{ZZ} invertible, which holds if and only if $\ker(\Sigma_{ZZ}) = \{0\}$. Since a function belongs to $\ker(\Sigma_{ZZ})$ iff $\text{var}(\langle f, X \rangle_{\mathcal{H}}) = 0$, the invertibility simply means that X does not reside in a proper subspace of \mathcal{H} . The next proposition shows that under Assumption 2 R_{ZZ} has a bounded inverse.

Proposition 2. *Under Assumption 2, $R_{ZZ} \geq cI$ for some $c > 0$, and consequently R_{ZZ}^{-1} is a bounded operator.*

We call R_{ZZ}^{-1} the additive precision operator of Z , and denote it by Θ_{ZZ} . Note that both R_{ZZ} and Θ_{ZZ} are members of $\times_{i,j \in \mathcal{V}} \mathcal{B}(\mathcal{H}_i, \mathcal{H}_j)$. We denote the (i, j) th entry of Θ_{ZZ} by $\Theta_{Z^i Z^j}$.

Theorem 1. *Suppose Z is an \mathbb{R}^p -valued Gaussian random function in \mathcal{H} , with Θ_{ZZ} as its precision operator. Suppose, furthermore, that, for each $i \neq j$, the operators $\Sigma_{Z^i Z^i}^{-1} \Sigma_{Z^i Z^{-(i,j)}}$ and $\Sigma_{Z^j Z^j}^{-1} \Sigma_{Z^j Z^{-(i,j)}}$ are bounded linear operators. Then*

$$Z^i \perp\!\!\!\perp Z^j | Z^{-(i,j)} \quad \text{iff} \quad \Theta_{Z^i Z^j} = 0.$$

Note that, even though $\Sigma_{Z^i Z^i}^{-1}$ is an unbounded operator, it is reasonable to assume $\Sigma_{Z^i Z^i}^{-1} \Sigma_{Z^i Z^{-(i,j)}}$ to be a bounded operator, which can again be interpreted as a type of “collective smoothness” between functions in $\mathcal{H}_{-(i,j)}$ and \mathcal{H}_i . **For further discussion on this point, see Li (2018).** Because there is a one-to-one relation between X^i and Z^i , the above proposition leads immediately to the following corollary.

Corollary 2. *Suppose X is an \mathbb{R}^p -valued copula Gaussian random function. Let $Z = (C_{X^1}(X^1), \dots, C_{X^p}(X^p))$. Then*

$$X^i \perp\!\!\!\perp X^j | X^{-(i,j)} \quad \text{iff} \quad \Theta_{Z^i Z^j} = 0.$$

In our case the operator $R_{Z^i Z^j}$ takes a rather simple form, as given in the next theorem.

Theorem 2. Suppose X is a copula Gaussian random element in $\oplus_{i=1}^p \mathcal{H}_i$ and let Z be the random element $(C_{X^1}(X^1), \dots, C_{X^p}(X^p))$. Suppose, for each $i, j = 1, \dots, p$,

$$\sum_{r \in \mathbb{N}} (\lambda_r^i)^{1/2} < \infty, \quad \sum_{(r,s) \in \mathbb{N} \times \mathbb{N}} |\text{corr}(\eta_r^i, \eta_s^j)| < \infty.$$

Then

$$R_{Z^i Z^j} = \sum_{r \in \mathbb{N}} \sum_{t \in \mathbb{N}} \text{corr}(\eta_r^i, \eta_t^j) (\phi_r^i \otimes \phi_t^j) \quad (3.3)$$

where $\eta_r^i = \Phi^{-1} \circ F_r^i(\xi_r^i)$, $\xi_r^i = \langle X^i, \phi_r^i \rangle_{\mathcal{H}_i} / (\lambda_r^i)^{1/2}$.

The assumption $\sum_{r \in \mathbb{N}} (\lambda_r^i)^{1/2} < \infty$ simply means that $\Sigma_{X^i X^i}^{1/2}$ is a trace-class operator, which is needed in the proof of this theorem. There is an alternative way of representing $R_{Z^i Z^j}$. According to Kendall (1948) and Kruskal (1958), if (U, V) is a bivariate Gaussian random vector, then

$$\text{corr}(U, V) = 2 \sin \left[\frac{\pi}{6} \text{corr}(F_U(U), F_V(V)) \right],$$

where F_U and F_V are the cumulative distribution functions of U and V . From this relation we can easily derive the following alternative expression of $R_{Z^i Z^j}$.

Corollary 3. Under the conditions of Theorem 2 we have

$$R_{Z^i Z^j} = 2 \sum_{r \in \mathbb{N}} \sum_{t \in \mathbb{N}} \sin[(\pi/6) \text{corr}(\rho_r^i, \rho_t^j)] (\phi_r^i \otimes \phi_t^j) \quad (3.4)$$

where $\rho_r^i = F_r^i(\langle X^i, \phi_r^i \rangle_{\mathcal{H}_i} / (\lambda_r^i)^{1/2})$.

4 Estimation

Theorem 2 and Corollary 3 suggest two different ways of estimating $R_{Z^i Z^j}$: the first is to estimate the correlations between the normal scores η_r^i and η_t^j ; the second is to estimate the Kendall's transformations of correlations between the ranked scores ρ_r^i and ρ_t^j . Both of these approaches will be developed in this section, and we refer to them as the normal score (NS) method and ranked score (RS) method. We also employ two types of sparse penalties: thresholding and the group lasso. The development of the first three subsections is at the operator level, which conveys the main ideas of estimation without getting too much involved in the details of coordinate mapping. The last subsection gives the outline of the algorithms using coordinate mapping, with full details developed in the Supplementary Material. We begin with the Karhunen-Loeve expansion at the sample level.

4.1 Empirical Karhunen-Loeve expansion

Let X_1, \dots, X_n be an independent sample from the random element X . For each $u = 1, \dots, n$, X_u is the vector (X_u^1, \dots, X_u^p) , where, for each $i = 1, \dots, p$, X_u^i is a \mathbb{R} -valued random function defined on the interval T . We use E_n to denote the sample average operator. That is, for a sample of random elements, say W_1, \dots, W_n , $E_n(W)$ means $n^{-1} \sum_{u=1}^n W_u$.

We estimate the mean element $\mu_{X^i} = E(X^i)$ by $\hat{\mu}_{X^i} = E_n(X^i)$. For each $i = 1, \dots, p$, we construct the operator

$$\hat{\Sigma}_{X^i X^i} = E_n[(X^i - \hat{\mu}_{X^i}) \otimes (X^i - \hat{\mu}_{X^i})], \quad (4.1)$$

which is well defined on \mathcal{H}_i . Let $\{(\hat{\lambda}_r^i, \hat{\phi}_r^i) : r \in \mathbb{N}\}$ be the sequence of eigenvalue-eigenfunction pairs of $\hat{\Sigma}_{X^i X^i}$ with $\hat{\lambda}_1^i \geq \hat{\lambda}_2^i \geq \dots$. Then, for each $u = 1, \dots, n$, $X_u^i - \hat{\mu}_{X^i}$

has the following empirical Karhunen-Loeve expansion:

$$X_u^i - \hat{\mu}_{X^i} = \sum_{r \in \mathbb{N}} \langle X_u^i, \hat{\phi}_r^i \rangle_{\mathcal{H}_i} \hat{\phi}_r^i \equiv \sum_{r \in \mathbb{N}} (\hat{\lambda}_r^i)^{1/2} \hat{\xi}_{ur}^i \hat{\phi}_r^i,$$

where $\hat{\xi}_{ur}^i = \langle X_u^i, \hat{\phi}_r^i \rangle_{\mathcal{H}_i} / (\hat{\lambda}_r^i)^{1/2}$. To filter out noise we truncate the sum at some m_n , which goes to ∞ with n at a rate specified in Section 7. Thus we now have the truncated empirical Karhunen Loeve expansion:

$$X_u^i - \hat{\mu}_{X^i} \approx \sum_{r=1}^{m_n} (\hat{\lambda}_r^i)^{1/2} \hat{\xi}_{ur}^i \hat{\phi}_r^i.$$

The set of estimated $\hat{\xi}_{ur}^i, \hat{\phi}_r^i$,

$$\{(\hat{\xi}_{ur}^i, \hat{\phi}_r^i) : u = 1, \dots, n, r = 1, \dots, m_n, i \in \mathbb{V}\},$$

will be the basis for all the estimation procedures to follow.

4.2 Normal-Score estimators

We first derive the sample-level approximation of the principal copula transformations $C_X = (C_{X^1}, \dots, C_{X^p})$ where $C_{X^i} = \{\Phi^{-1} \circ F_r^i : r \in \mathbb{N}\}$, F_r^i being the c.d.f. of ξ_r^i . As in Liu, Lafferty, and Wasserman (2009), we use Windsorized empirical distribution based on the sample $\hat{\xi}_{1r}^i, \dots, \hat{\xi}_{nr}^i$. Let \hat{F}_r^i be the empirical distribution of $\hat{\xi}_{1r}^i, \dots, \hat{\xi}_{nr}^i$:

$$\hat{F}_r^i(y) = E_n[I(\hat{\xi}_r^i \leq y)].$$

The Winsorized empirical distribution \tilde{F}_r^i is

$$\tilde{F}_r^i(y) = \begin{cases} \delta_n & \text{if } \hat{F}_r^i(y) < \delta_n \\ \hat{F}_r^i(y) & \text{if } \delta_n \leq \hat{F}_r^i(y) \leq 1 - \delta_n \\ 1 - \delta_n & \text{if } \hat{F}_r^i(y) > 1 - \delta_n \end{cases}$$

where $\delta_n > 0$ and $\lim_{n \rightarrow \infty} \delta_n = 0$. Following Liu et al. (2009), we choose δ_n to be $[4n^{1/4} \sqrt{\pi \log(n)}]^{-1}$.

The empirical principal copula transformations are then given by

$$\hat{C}_{X^i} = \{\Phi^{-1} \circ \tilde{F}_r^i : r = 1, \dots, m_n\}.$$

Let

$$\hat{\eta}_{ur}^i = \Phi^{-1} \circ \tilde{F}_r^i(\hat{\xi}_{ur}^i), \quad u = 1, \dots, n, \quad r = 1, \dots, m_n, \quad i = 1, \dots, p$$

be the empirical normal scores. We estimate the correlation operator $R_{Z^i Z^j}$ by

$$\hat{R}_{Z^i Z^j} = \sum_{r,t=1}^{m_n} \text{corr}_n(\hat{\eta}_r^i, \hat{\eta}_t^j)(\hat{\phi}_r^i \otimes \hat{\phi}_t^j) \quad (4.2)$$

where corr_n is the empirical correlation defined by replacing all the expectations in corr with its empirical counterpart E_n . We then form the operator \hat{R}_{ZZ} as the matrix of operators whose (i, j) th entry is $\hat{R}_{Z^i Z^j}$. Using this operator we develop sparse estimators of Θ_{ZZ} either by thresholding or group lasso, which are detailed below separately.

Thresholding Let $\hat{\Theta}_{ZZ} = \hat{R}_{ZZ}^\dagger$ be the Moore-Penrose inverse of the operator \hat{R}_{ZZ} (Hsing and Eubank, 2015, Definition 3.5.7). Let ϵ_n be a positive sequence that goes

to 0 as $n \rightarrow \infty$, which will be specified in the asymptotic section. We estimate \mathbf{E} by

$$\hat{\mathbf{E}}(\epsilon_n) = \{(i, j) \in \mathbf{V} \times \mathbf{V} : \|\hat{\Theta}_{ZZ^{ij}}\| > \epsilon_n\} \quad (4.3)$$

where $\|\cdot\|$ is the operator norm or the Hilbert Schmidt norm of an operator.

Group lasso Inspired by Qiao, Guo, and James (2019a), we define

$$L_n : \times_{i,j \in \mathbf{V}} \mathcal{B}(\mathcal{H}_i, \mathcal{H}_j) \rightarrow \mathbb{R},$$

$$\Theta \mapsto -\log \det(\Theta) + \text{trace}(\Theta \hat{R}_{ZZ}) + \lambda_n \sum_{i \neq j} \|\Theta_{ij}\|_{\text{HS}}.$$

Here, $\det(\Theta)$ means the product of all the nonzero eigenvalues of Θ and Θ_{ij} denotes for the sub-operator of Θ in $\mathcal{B}(\mathcal{H}_i, \mathcal{H}_j)$. The precision operator Θ_{ZZ} is then estimated by maximizing $L_n(\Theta)$ among all positive semidefinite operators in $\times_{i,j \in \mathbf{V}} \mathcal{B}(\mathcal{H}_i, \mathcal{H}_j)$. This is essentially a group lasso procedure because the parameters in a suboperator Θ_{ij} are shrunk to 0 together.

Since, at the sample level, $\otimes_{i=1}^p \mathcal{H}_i$ is a finite-dimensional space, the operator $\Theta \hat{R}_{ZZ}$ can be represented by coordinate mapping as a matrix with respect to a basis in $\oplus_{i=1}^p \mathcal{H}_i$. The trace of the operator $\Theta \hat{R}_{ZZ}$ is simply the trace of its matrix representation. This trace is a well defined function of an operator because it does not depend on the basis with respect to which the operator is represented. The function $-\log \det(\Theta) + \text{trace}(\Theta \hat{R}_{ZZ})$ need not be regarded as the Gaussian likelihood with its parameter ranging over the class of linear operators in $\times_{i,j \in \mathbf{V}} \mathcal{B}(\mathcal{H}_i, \mathcal{H}_j)$; it is simply a reasonable objective function to minimize, because, without the sparse penalty, it is minimized at $\Theta = \hat{R}_{ZZ}^{-1}$. It can also be regarded as the natural generalization of the objective function for the glasso to the operator level.

4.3 Estimators based on ranked scores

Here, we replace (4.2) by an estimator $R_{Z^i Z^j}$ using Corollary 3:

$$\hat{R}_{Z^i Z^j} = 2 \sum_{r,t=1}^{m_n} \sin[(\pi/6) \text{corr}_n(\hat{\rho}_r^i, \hat{\rho}_t^j)] (\hat{\phi}_r^i \otimes \hat{\phi}_t^j), \quad (4.4)$$

where $\hat{\rho}_{ur}^i$ are the empirical ranked scores

$$\hat{\rho}_{ur}^i = \hat{F}_r^i(\hat{\xi}_{ur}^i) : \quad u = 1, \dots, n, \quad r = 1, \dots, m_n, \quad i = 1, \dots, p.$$

Note that we use the usual empirical likelihood \hat{F}_r^i rather than the Winsorized version. The thresholding and the group lasso procedures based on the ranked scores are exactly the same as those based on the normal scores, except that the estimator (4.2) is replaced by the estimator (4.4).

Up to this point we have described our estimators at the population level, under the assumption that the random functions X_1, \dots, X_n are observed in their entirety. In the next section we describe how to implement them numerically using coordinate representation in finite-dimensional Hilbert spaces. The normal score methods based on truncation and group lasso will be abbreviated by NS-T and NS-L; the ranked score methods based on truncation and group lasso will be abbreviated by RS-T and RS-L.

4.4 Algorithms

In the previous subsections we described the estimation procedures at the operator level. To make them executable algorithms we need to represent operators as matrices through coordinating mapping. To save space we leave the full details of the development of the coordinate mapping in the Supplementary Material, while only present the final results here.

Algorithm for Karhunen-Loeve expansion

1. For $i = 1, \dots, p$, choose a family of functions $\mathcal{H}_i = \text{span}\{h_1^i, \dots, h_{k_n}^i\}$ on T ; compute the Gram matrix K_i and its centered version G_i ;
2. Reset X_u^i to its centered version $X_u^i - E_n(X^i)$ and compute the coordinates $[X_u^i]_{\mathcal{B}_i}$ relative to the basis \mathcal{B}_i of \mathcal{H}_i ;
3. Perform spectral decomposition on $G_i^{1/2} E_n([X^i][X^i]^\top) G_i^{1/2}$ to obtain the eigen pairs $\{(\hat{\lambda}_r^i, v_r^i) : r = 1, \dots, k_n\}$. Compute $\hat{\xi}_{ur}^i = (\hat{\lambda}_r^i)^{-1/2} [X_u^i]_{\mathcal{B}_i}^\top G_i^{1/2} v_r^i$.

Algorithm for NS-T, RS-T, NS-L, RS-L

In the following, for NS-T, follow $1 \rightarrow 2 \rightarrow 3$; for RS-T, follow $1' \rightarrow 2' \rightarrow 3'$; for NS-L, follow $1 \rightarrow 2 \rightarrow 3''$; for NS-R, follow $1' \rightarrow 2' \rightarrow 3'''$.

1. For NS-T and NS-L, compute $\hat{\eta}_{ur}^i$ according to $\hat{\eta}_{ur}^i = \Phi^{-1} \circ \tilde{F}_r^i(\hat{\xi}_{ur}^i)$;
- 1'. For RS-T and RS-L, compute $\hat{\rho}_{ur}^i$ according to $\hat{\rho}_{ur}^i = \hat{F}_r^i(\hat{\xi}_{ur}^i)$ (note that the un-truncated empirical distribution is used);
2. For NS-T and NS-L, compute $\Omega^{(N)} = \{\Omega_{ij}^{(N)}\}_{i,j=1}^p$, where

$$\Omega_{ij}^{(N)} = \sum_{r,t=1}^{m_n} \text{corr}_n(\hat{\eta}_r^i, \hat{\eta}_t^j) G_i^{\dagger 1/2} v_r^i (v_t^j)^\top G_j^{\dagger 1/2}.$$

- 2'. For RS-T and RS-L, compute $\Omega^{(R)} = \{\Omega_{ij}^{(R)}\}_{i,j=1}^p$, where

$$\Omega_{ij}^{(R)} = 2 \sum_{r,t=1}^{m_n} \sin\{(\pi/6) \text{corr}_n(\hat{\rho}_r^i, \hat{\rho}_t^j)\} G_i^{\dagger 1/2} v_r^i (v_t^j)^\top G_j^{\dagger 1/2}.$$

3. For NS-T, compute $\Lambda^{(N)} = (\Omega^{(N)})^\dagger$, and let $\Lambda_{ij}^{(N)}$ be the (i, j) th block of $\Lambda^{(N)}$. Choose a threshold ϵ_n , and estimate the edge set by $\|G_j^{\dagger 1/2} \Lambda_{ji}^{(N)} G_i^{\dagger 1/2}\|_F > \epsilon_n$.

3'. For RS-T, perform step 3 with $\Omega^{(N)}, \Omega_{ij}^{(N)}, \Lambda^{(N)}$ replaced by $\Omega^{(R)}, \Omega_{ij}^{(R)}, \Lambda^{(R)}$.

3''. For NS-L, use group lasso (Qiao et al., 2019a) to minimize the objective function

$$M_n^{(N)}(A) = -\log \det(A) + \text{trace}\{A(\oplus_{i \in \mathbf{V}} G_i^{1/2})\Omega^{(N)}(\oplus_{i \in \mathbf{V}} G_i^{1/2})\} + \lambda_n \sum_{i \neq j} \|A_{ij}\|_{\mathbb{F}}$$

to obtain $\hat{A}^{(N)} = \{\hat{A}_{ij}^{(N)}\}_{i,j=1}^p$ and estimate the edge set by $\{(i, j) \in \mathbf{V} \times \mathbf{V} : i \neq j, \hat{A}_{ij} \neq 0\}$.

3'''. For NS-R, follow step 3'' with $\Omega^{(N)}, \hat{A}^{(N)}, \hat{A}_{ij}^{(N)}$, replaced by $\Omega^{(N)}, \hat{A}^{(N)}, \hat{A}_{ij}^{(N)}$.

5 Consistency and convergence rate

In this section we develop the consistency and convergence rate of one of the proposed estimators. Due to space limit, we focus on the RS-T procedure with the operator norm. There are several challenges for developing an asymptotic theory for our copula functional model. First, a difference between our proposed functional copula model and the conventional copula model is that our copula transformations are applied to the estimated scores $\hat{\xi}_{ur}^i$ rather than the true scores ξ_{ur}^i , whereas the conventional copula models are applied directly to observed data. Accounting for this approximation is a major undertaking: since rank transformation is not continuous, standard methods are not applicable, and new techniques must to be developed. The second challenge is that in various places we need to find uniform bounds for eigenvalues and eigenfunctions, and a variety of quantities derived from them.

If a_n and b_n are positive sequences, we write $a_n \prec b_n$ (or $b_n \succ a_n$) if $a_n/b_n \rightarrow 0$ as $n \rightarrow \infty$. We write $a_n \asymp b_n$ if

$$0 < \liminf_{n \rightarrow \infty} (b_n/a_n) \leq \limsup_{n \rightarrow \infty} (b_n/a_n) < \infty.$$

If U_n is a sequence of random variables that is bounded in probability; that is, for any $\epsilon > 0$, there is a $K > 0$ such that $\limsup_{n \rightarrow \infty} P(|U_n| > K) < \epsilon$, then we write $U_n = O_P(1)$. If $U_n/a_n = O_P(1)$, then we write $U_n = O_P(a_n)$. If $U_n = O_P(a_n)$ and $a_n \prec b_n$, then we write $U_n \stackrel{P}{\prec} b_n$ or $b_n \stackrel{P}{\succ} U_n$.

We first establish the consistency of $\hat{R}_{Z^i Z^j}$. For any integer m , let

$$\omega_m = \min\{\lambda_r^i - \lambda_{r+1}^i : r = 1, \dots, m, i = 1, \dots, p\},$$

and let $c_{rt}^{ij} = 2 \sin[(\pi/6)\text{corr}(\rho_r^i, \rho_t^j)]$. We impose the additional condition $\sum_{r,t=1}^{\infty} |c_{rt}^{ij}| < \infty$, which requires c_{rt}^{ij} to decay sufficiently fast. Note that, if $R_{Z^i Z^j}$ is a Hilbert Schmidt operator, then $\sum_{r,t=1}^{\infty} (c_{rt}^{ij})^2 < \infty$. So our condition is stronger than this, but is in the same spirit. As we will further discuss below, requiring c_{rt}^{ij} to decay fast suggests a certain type of smoothness.

Theorem 3. *Suppose, for $i \neq j$,*

1. $\sum_{r,t=1}^{\infty} |c_{rt}^{ij}| < \infty$,
2. $E\|X^k\|_{\mathcal{H}_i}^4 < \infty$, for $k = 1, \dots, p$,
3. *for some sequence $\{m_n : n \in \mathbb{N}\}$ and some $0 < \alpha < 1/2$,*

$$1 \succ \omega_{m_n} \succ n^{2\alpha/3-1/3}, \quad m_n^2 \prec \omega_{m_n}^{3/2} n^{1/2-\alpha}.$$

Then $\|\hat{R}_{Z^i Z^j} - R_{Z^i Z^j}\|_{\text{OP}} \xrightarrow{P} 0$.

In the theorem, the condition $\omega_{m_n} \succ n^{2\alpha/3-1/3}$ is to ensure that m_n can be chosen to go to ∞ . Note that it also implies $\omega_{m_n} \succ n^{-1/2}$, which is used in the proof. **The condition $\sum_{r,t} |c_{rt}^{ij}| < \infty$ is needed to ensure that the operator norm of the difference between $R_{Z^i Z^j}$ and its truncated version goes to 0, which control the bias of the estimator $\hat{R}_{Z^i Z^j}$.** Next, we develop the convergence rates of $\hat{R}_{Z^i Z^j}$. To accomplish this

we need to strengthen three conditions. The first is the rate at which c_{rt}^{ij} decays as $r \rightarrow \infty$ and $t \rightarrow \infty$. This rate of decay characterizes the degree of smoothness in the relation between Z^i and Z^j : c_{rt}^{ij} decaying fast means that most of their correlations are concentrated on the low-frequency components of Z^i and Z^j . The second is the rate of decay of λ_r^i , which characterizes the smoothness of X^i itself. The third is the tail probability of the random variable $\|X^i\|_{\mathcal{H}_i}$, as reflected in the existence of its higher-order moments or its moment generating function. In the following, let $S(m_n)$ be the tail index set

$$\{m_n + 1, m_n + 2, \dots\} \times \{m_n + 1, m_n + 2, \dots\}.$$

Theorem 4. *Suppose, for $i \neq j$, there exist $\beta > 0$ and $s > 4$ such that*

1. $\sum_{r,t=1}^{\infty} |c_{rt}^{ij}| < \infty$;
2. $[\sum_{(r,t) \in S(m_n)} (c_{rt}^{ij})^2]^{1/2} = O(m_n^{-\beta}) \quad \text{as } n \rightarrow \infty$;
3. $E\|X^i\|_{\mathcal{H}_i}^s < \infty$;
4. *for some sequence $\{m_n : n \in \mathbb{N}\}$ and some $0 < \alpha < 1/s$,*

$$1 \succ \omega_{m_n} \succ n^{2\alpha/3-1/3}, \quad m_n^2 \prec \omega_{m_n}^{3/2} n^{1/2-\alpha}.$$

Then, $\|\hat{R}_{Z^i Z^j} - R_{Z^i Z^j}\|_{OP} = O_P(m_n^2 n^{-1/2} + m_n^2 \omega_{m_n}^{-3/2} n^{-1/2+\alpha} + m_n^{-\beta})$.

A condition somewhat similar to the above condition 2 is also employed in Li and Song (2017) and Li and Solea (2018) in the context of nonlinear sufficient dimension reduction and nonparametric graphical models for functional data. See also Li (2018) for further discussions on this point.

Some discussion of the roles played by different constants in the convergence rate is in order. As can be seen from the proofs of Theorems 3~5 in the Supplementary

Material, the estimation error $\|\hat{R}_{Z^i Z^j} - R_{Z^i Z^j}\|_{\text{OP}}$ is bounded from above by

$$\|\hat{R}_{Z^i Z^j} - R_{Z^i Z^j}^{(m_n)}\|_{\text{OP}} + \|R_{Z^i Z^j}^{(m_n)} - R_{Z^i Z^j}\|_{\text{OP}},$$

where $R_{Z^i Z^j}^{(m_n)}$ is the first m_n terms in the expansion of $R_{Z^i Z^j}$. Roughly, the second term above represents the bias of the estimate; the first term the variance. The positive constant α controls the tail of the random variable $\|X^i\|_{\mathcal{H}_i}$: the smaller α is, the thinner the tail. A thinner tail helps to reduce the variance term. The integer m_n is the length of the truncated Karhunen-Loeve expansion of X^i , and a larger m_n reduces the bias term. The number ω_{m_n} is the overall eigenvalue gap of the first $m_n + 1$ eigenvalues of $\Sigma_{Z^i Z^i}$. A larger eigenvalue gap also helps to reduce the variance term. With these tendencies in mind, the condition $m_n^2 \prec \omega_{m_n}^{3/2} n^{1/2-\alpha}$ in Theorem 4 means (a) if the tail of $\|X^i\|_{\mathcal{H}_i}$ is thin, then the variance term is small, and we can afford to choose a larger m_n to reduce the bias and (b) similarly, if the eigenvalue gap is large, then the variance term is small, and we can choose a larger m_n .

If we ignore the term $m_n^{-\beta}$, then the convergence rate is faster when m_n is smaller. This is the “parametric part” of the rate. However, as m_n becomes small, $m_n^{-\beta}$ increases. This is the nonparametric part of the rate. If the smoothness index β is large, then $m_n^{-\beta}$ is small even if m_n increases slowly with n . As will be seen in Example 1, the rate of $m_n^2 \omega_{m_n}^{-3/2}$ is determined by how fast λ_r^i decays as $r \rightarrow \infty$. The faster it decays, the slower rate (to ∞) of $m_n^2 \omega_{m_n}^{-3/2}$ can be tolerated. Thus, if c_{rs}^{ij} and λ_r^i are allowed to decay arbitrarily fast, the rate in Theorem 4 can get arbitrarily close to $n^{-1/2+1/s}$.

We can further improve the convergence rate by strengthening the moment assumption on $\|X^i\|_{\mathcal{H}_i}$ to existence of its moment generating function. In that case, the convergence rate can get arbitrarily close to $n^{-1/2} \log(n)$.

Theorem 5. *Suppose*

1. as $n \rightarrow \infty$, $[\sum_{(r,t) \in S(m_n)} (c_{rt}^{ij})^2]^{1/2} = O(m_n^{-\beta})$;
2. for each $i = 1, \dots, p$, the moment generating function of $\|X^i\|_{\mathcal{H}_i}$ is finite in a neighborhood of 0;
3. for some sequence $\{m_n : n \in \mathbb{N}\}$, and some $\alpha > 1$,

$$1 \succ \omega_{m_n} \succ n^{-1/2}, \quad m_n^2 \prec \omega_{m_n}^{-3/2} n^{-1/2} (\log(n))^\alpha.$$

Then, $\|\hat{R}_{Z^i Z^j} - R_{Z^i Z^j}\|_{\text{OP}} = O_P(m_n^2 n^{-1/2} + m_n^2 \omega_{m_n}^{-3/2} (\log(n))^\alpha + m_n^{-\beta})$.

As mentioned earlier, the rate in Theorem 5 can be made arbitrarily close to $n^{-1/2} \log(n)$ if c_{rt}^{ij} and λ_r^i decay sufficiently fast. To provide intuition regarding how $m_n^2 \omega_{m_n}^{-3/2}$ is related to the decaying rate of λ_r^i , we give an example below using $\lambda_r^i \propto r^{-a}$, $a > 0$ as a prototype.

Example 1. Because $\Sigma_{X^i X^i}$ is a trace-class operator, we have $a > 1$. Then, for any integer m , $\omega_m = \lambda_m - \lambda_{m+1} = m^{-a} - (m+1)^{-a}$. By elementary calculations, we can show that this is of the order $O(m^{-a-1})$. So if we want to choose m_n so that $\omega_n = n^{-1/2+b}$ for some $b > 0$, then we need $m_n^{-(a+1)} = n^{-1/2+b}$, which is satisfied if $m_n = n^{\frac{1/2-b}{a+1}}$. Hence

$$m_n^2 \omega_{m_n}^{-3/2} = n^{\frac{1-2b}{a+1}} n^{(-1/2+b)(-3/2)} = n^{\frac{1-2b}{a+1} + \frac{3}{4} - \frac{3b}{2}}.$$

Because $b < 1/2$, we have $\frac{1-2b}{a+1} + \frac{3}{4} - \frac{3b}{2} > 0$. So $m_n^2 \omega_{m_n}^{-3/2} \rightarrow \infty$. But if a is large and b is chosen to be small, the increasing rate can be arbitrarily slow, so that the rate in Theorem 5 can be arbitrarily close to $n^{-1/2} \log(n)$ except the term $m_n^{-\beta}$, which itself can be arbitrarily small if β is large. \square

Finally, we establish the consistency and convergence rates of $\hat{\Theta}_{ZZ}$, as well as the consistency of the graph estimator $\hat{E}(\epsilon_n)$ defined in (4.3). Because p is a constant,

$\|\hat{R}_{ZZ} - R_{ZZ}\|_{\text{OP}}$ is also consistent under the conditions of Theorem 3, and has the same convergence rates in Theorems 4 and 5, under the respective conditions. To derive the consistency and convergence rate of $\hat{\Theta}_{ZZ}$ we need the following lemma from Li and Solea (2018), which can be verified by straightforward calculation. The next theorem shows that $\|\hat{\Theta}_{ZZ} - \Theta_{ZZ}\|_{\text{OP}}$ and $\|\hat{R}_{ZZ} - R_{ZZ}\|_{\text{OP}}$ have the same convergence rate under mild conditions.

Theorem 6. *If $\|\hat{R}_{ZZ} - R_{ZZ}\|_{\text{OP}} \xrightarrow{P} 0$ and $R_{ZZ} \geq cI$ for some $c > 0$, then, for any positive sequence $a_n \rightarrow 0$,*

$$\|\hat{R}_{ZZ} - R_{ZZ}\|_{\text{OP}} = O_P(a_n) \Rightarrow \|\hat{\Theta}_{ZZ} - \Theta_{ZZ}\|_{\text{OP}} = O_P(a_n).$$

We say that an estimator $\hat{\mathbf{E}}$ of the true edge set \mathbf{E} is consistent if the probability of the event $\hat{\mathbf{E}} = \mathbf{E}$ tends to 1 as $n \rightarrow \infty$. The next corollary establishes the consistency of $\hat{\mathbf{E}}(\epsilon_n)$ as defined in (4.3), as well as the rate of the threshold ϵ_n .

Corollary 4. *Let γ_n denote the convergence rates in Theorems 4. If $1 \succ \epsilon_n \succ \gamma_n$, then, under the conditions in Theorem 4, $P(\hat{\mathbf{E}}(\epsilon_n) = \mathbf{E}) \rightarrow 1$. The same can be said if Theorem 4 is replaced by Theorem 5.*

6 Simulation Studies

In this section we compare numerically the performances of our FCGGM estimators, including NS-T, RS-T, NS-L, RS-L, with two versions of the FGGM estimator, one based on the group Lasso (FGGM-L) as proposed by Qiao, Guo, and James (2019a), the other based on thresholding the Frobenius norm of the (i, j) th block of the estimated precision matrix of the functional principal components (FGGM-T, this version was not contained in Qiao, Guo, and James (2019a)). We consider two scenarios: one in which the random elements on the vertices are copula Gaussian random functions,

and one in which they are Gaussian random functions.

To simulate copula Gaussian random functions, we first draw n independent Gaussian random functions using five Fourier basis functions, as in Qiao, Guo, and James (2019a):

$$X_u^i(t) = \sum_{r=1}^m \xi_{ur}^i v_r(t), \quad u = 1, \dots, n, \quad i = 1, \dots, p, \quad (6.1)$$

where $m = 5$, and $\{v_r, r = 1, 2, 3, 4, 5\}$ are the first 5 functions in the Fourier basis

$$1, \sqrt{2} \sin(2\pi t), \sqrt{2} \cos(2\pi t), \sqrt{2} \sin(4\pi t), \sqrt{2} \cos(4\pi t),$$

and, for each u , $(\xi_{u1}^1, \dots, \xi_{um}^1, \dots, \xi_{u1}^p, \dots, \xi_{um}^p)^\top$ is multivariate Gaussian with mean 0 and block precision matrix $\Lambda \in \mathbb{R}^{pm \times pm}$. We consider choices of Λ :

$$\begin{aligned} \text{(a)} \quad & \Lambda_{j,j} = I_m, \quad \Lambda_{j,j+1} = 0.4I_m; \\ \text{(b)} \quad & \Lambda_{j,j} = I_m, \quad \Lambda_{1,j} = 0.2I_m. \end{aligned} \quad (6.2)$$

Then we transform ξ_{ur}^i to $c_{ri}(\xi_{ur}^i)$ where, for simplicity, we choose $c_{ri} = c_r$ to be the same across $i = 1, \dots, p$, which are taken to be the following functions

$$c_1(x) = x^3, \quad c_2(x) = e^x, \quad c_3(x) = \frac{e^x}{1 + e^x}, \quad c_4(x) = (1 + x)^5, \quad c_5(x) = x. \quad (6.3)$$

Each function $X_u^i(t)$ is sampled at 10 equally spaced time points t_1, \dots, t_{10} , where $t_1 = 0$ and $t_{10} = 1$. In the simulations, the network size and the sample size are taken to be $(p, n) = (10, 100), (10, 200)$. The simulation sample size is $n_{\text{sim}} = 100$.

To approximate each of the X_u^i based on its sampled points

$$\{(t_a, X_u^i(t_a)) : a = 1, \dots, 10\},$$

we use cubic spline functions with 3 interior nodes equally spaced within $[0, 1]$. That is, we employ 4 piecewise polynomials that are connected smoothly so that they are continuous and have continuous first two derivatives. With 16 parameters for the 4 cubic polynomials and 9 constraints for smoothness, we have 7 free parameters left to describe these functions. Equivalently, each \mathcal{H}_i is spanned by $k_n = 7$ linearly independent functions $\{h_1, \dots, h_7\}$. Each X_u^j is then approximated as the linear combination of these 7 functions:

$$X_u^j = \sum_{k=1}^{k_n} [X_u^j]_k h_k^j, \quad j = 1, \dots, p,$$

where the linear coefficient vector $[X_u^j] \in \mathbb{R}^7$ is determined by least squares. For simplicity we choose $\mathcal{H}_1, \dots, \mathcal{H}_p$ to be the space spanned by these 7 spline functions. We retain the first 3 ($m_n = 3$) functional principal components for both FCGGM and FGGM.

In all the simulations, the truncation parameter m_n is chosen so that the first m_n eigenvalues in the Karhunen-Loeve expansion explains 90% of the total variation in the functional PCA.

6.1 Case 1: Non-Gaussian data

We first consider two models, Model I and Model II, where the functional Gaussian assumption does not hold. Both of the models are generated by

$$X_u^i(t) = \sum_{r=1}^m c_r(\xi_{ur}^i) v_r(t), \quad u = 1, \dots, n, \quad i = 1, \dots, p,$$

where c_r are as defined in (6.3). The precision matrix Λ for Model I is specified by the first line in (6.2); that for Model II is specified by the second line in (6.2). Figure 1 presents the ROC curves of the Models I and II, averaged across the $n_{\text{sim}} = 100$

simulation runs.

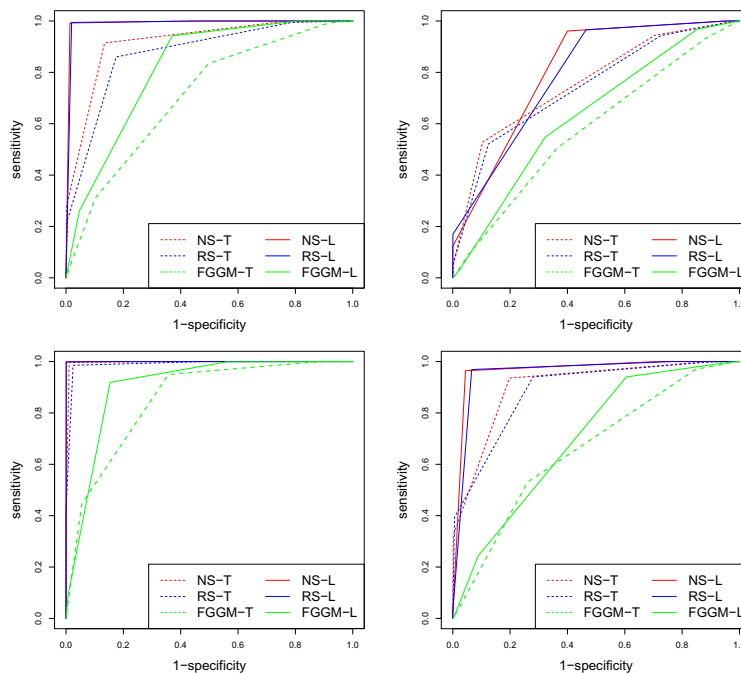


Figure 1. ROC curves for Models I-II (first, second columns), and for $n = 100$ (first row), and $n = 200$ (second row).

In Table 1, we report the means and standard deviations (in parentheses) of the associated area-under-curve values (AUC). As expected, our FCGGM estimators perform much better than the FGGM estimators in this case. Also, it can be seen in Figure 1 that the group-lasso based procedures NS-L, RS-L and FGGM are more efficient than thresholding.

In Table 2 we repeat the above calculation for $p = 100$, where the edge sets of Model I and Model II remain the same pattern in high dimension. Since the computation of AUC for FGGM-L, NS-L, and RS-L is quite time consuming for larger p (the group Lasso has to be performed repeatedly for each sparse penalty constant λ_n), we only calculated the results for FGGM-T, NS-T, and RS0-T. The table indicates that the same pattern of comparison upholds in high dimension.

Table 1. Means and standard errors (in parentheses) for AUC for models I-II.

n	Models	Methods					
		NS-T	RS-T	FGGM-T	NS-L	RS-L	FGGM-L
100	I	0.93	0.90	0.71	0.99	0.98	0.82
		(0.05)	(0.06)	(0.08)	(0.01)	(0.02)	(0.06)
	II	0.79	0.77	0.57	0.81	0.79	0.62
		(0.08)	(0.09)	(0.15)	(0.05)	(0.05)	(0.1)
200	I	0.99	0.98	0.85	0.99	0.99	0.90
		(0.01)	(0.01)	(0.07)	(0.002)	(0.003)	(0.04)
	II	0.93	0.91	0.65	0.96	0.95	0.70
		(0.04)	(0.04)	(0.09)	(0.04)	(0.03)	(0.08)

Table 2. Means and standard errors (in parentheses) for AUC for models I-II.

n	Models	Methods		
		NS-T	RS-T	FGGM-T
100	I	0.96 (0.01)	0.96 (0.01)	0.85 (0.02)
	II	0.78 (0.11)	0.79 (0.11)	0.43 (0.15)
200	I	0.99 (0.00)	0.99 (0.003)	0.92 (0.01)
	II	0.80 (0.13)	0.83 (0.13)	0.45 (0.15)

6.2 Case 2: Gaussian data

Next, we consider two models, Model III and Model IV, where the functional Gaussian assumption holds, to see how much information might be lost by employing a functional copula Gaussian model under the Gaussian assumption. Both models are generated by (6.1), with Model III corresponding to the precision matrix specified by

the first line of (6.2), and Model IV the second line. Figure 2 presents the averaged ROC curves across the $n_{\text{sim}} = 100$ simulated samples. Table 3 reports the means and standard deviations of AUC. Overall, although there is some loss of efficiency by the functional copula estimators, the losses are quite modest. In Table 4 we repeat the above calculation for $p = 100$. Again the same pattern of comparison upholds in high dimension.

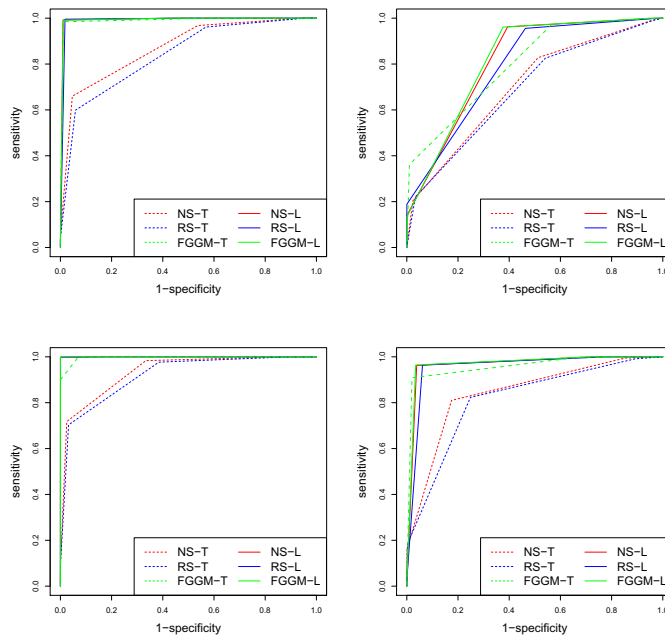


Figure 2. ROC curves for Models III and IV (first, second columns),
and for $n = 100$ (first row), and $n = 200$ (second row).

In the above simulation studies of Case 1 and Case 2, the results are inevitably affected by the choice of the number of knots in the splines. To examine the sensitivity of this choice, we conducted further simulations with number of knots equal to 5, 6, 8. Overall, the performances of the estimators (as measured by the areas under the ROC curves) are relatively stable, although NS-T and RS-T seem to perform better for larger number of knots. Due to the limited space, we present corresponding ROC curves in Section 15 in the Supplementary Material.

Table 3. Means and standard errors of AUC for Models III, IV.

n	Models	Methods					
		NS-T	RS-T	FGGM-T	NS-L	RS-L	FGGM-L
100	I	0.93	0.90	0.99	0.99	0.98	1
		(0.05)	(0.05)	(0.02)	(0.01)	(0.01)	(0.01)
	II	0.81	0.82	0.84	0.80	0.79	0.82
		(0.1)	(0.1)	(0.07)	(0.05)	(0.05)	(0.05)
200	I	0.99	0.99	1	1	0.99	1
		(0.006)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)
	II	0.93	0.92	0.96	0.97	0.96	0.97
		(0.05)	(0.05)	(0.04)	(0.03)	(0.03)	(0.02)

Table 4. Means and standard errors for AUC for models III-IV.

n	Models	Methods		
		NS-T	RS-T	FGGM-T
100	III	0.84 (0.03)	0.84 (0.03)	0.99 (0.00)
	IV	0.62 (0.12)	0.62 (0.12)	0.87 (0.02)
200	III	0.87 (0.02)	0.86 (0.02)	0.99 (0.00)
	IV	0.64 (0.10)	0.64 (0.10)	0.92 (0.02)

It is interesting to observe from Tables 2 and 4 that, even at $p = 100$, the AUC values are still relatively high. On the surface, at a total dimension of $pm_n = 300$, the thresholding method shouldn't perform this well. But our experiences often indicate that the observations on functional data actually help rather than hamper estimation. It seems as if the observations on functional data shouldn't be simply counted as increase of dimension. This is an important theoretical question for functional data analysis that deserved further careful investigation.

7 Application to EEG data

In this section we apply FCGGM (versions NS-L and RS-L) and the group lasso-based FGGM to the EEG data set used in Li, Kim, and Altman (2010) and Qiao, Guo, and James (2019a). We also applied the functional additive precision operator (FAPO) method introduced recently by Li and Solea (2018) this data set. The EEG study involved two groups of subjects: an alcoholic group of 77 subjects and a control group of 45 subjects. Each subject was exposed to a stimulus while brain activities were recorded from the 64 electrodes placed on the subject’s scalp, over a one-second period in which 256 time points were sampled. See Zhang et al. (1995) and Ingber (1997) for more backgrounds of this data. Our goal is to construct brain networks of the 64 nodes for the two groups, based on the functional data collected from the electrodes on each subject.

We choose $m_n = 6$ for all three methods. To construct \mathcal{H}_i , we use spline functions with 20 equally-spaced interior nodes, which means the dimension of \mathcal{H}_i is $k_n = (20+1) \times 4 - 20 \times 3 = 24$. Thus the dimension of Ω is $(64 \times 24) \times (64 \times 24) = 1536 \times 1536$. We take the penalizing constant λ_n for both NS-L and RS-L to be such that 3% of the $\binom{64}{2}$ pairs of vertices are retained as edges. Similarly, the penalty constant in group Lasso for FGGM is tuned so that roughly 3% of the pairs of nodes are edges.

The choice of 3% of the $\binom{64}{2}$ edges is to avoid the network looking too crowded, while showing the most outstanding connections. A more systematic method for determining the number of edges, for example via a significance test, needs to be developed. This is, however, beyond the scope of the current paper and will be left for future research.

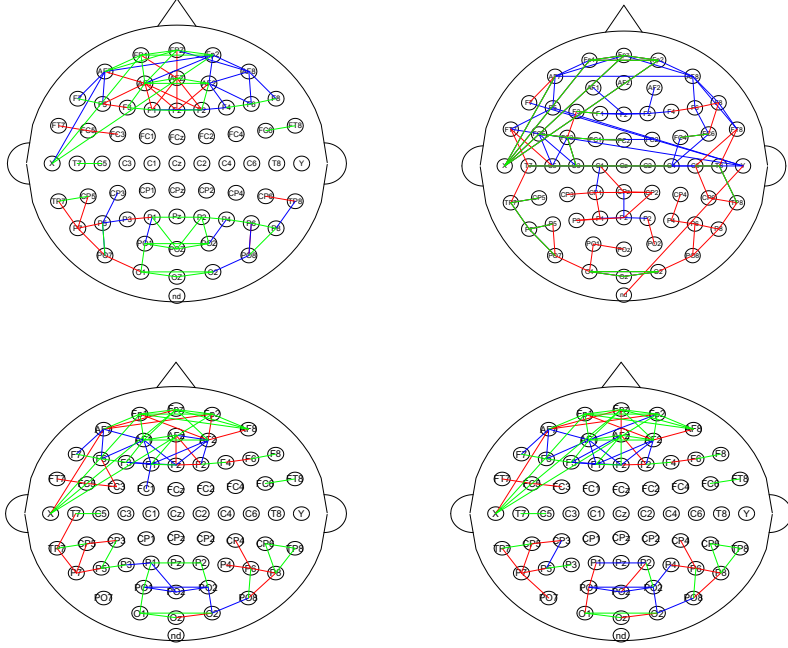


Figure 3. Brain networks for the alcoholic and non-alcoholic groups constructed by FGGM-L (upper-left), FAPO (upper-right), NS-L (lower-left), and RS-L (lower-right). The green lines indicate the edges shared by the alcoholic and non-alcoholic networks; the red lines indicate the edges in the alcoholic network but not in the non-alcoholic network; the blue lines indicate the edges in the non-alcoholic network but not in the alcoholic network.

Figure 3 shows the networks constructed by the four methods, where the green lines indicate the edges shared by the alcoholic and non-alcoholic networks, the red lines indicate the edges that are in the alcoholic network but not in the non-alcoholic network, and the blue lines indicate the edges that are in the non-alcoholic network but not in the alcoholic network. We see that the networks produced by FGGM, NS-L, and RS-L are quite similar. Moreover, the networks in the frontal lobe produced by all four methods are also similar to a degree.

We also investigated the degree to which the Gaussian assumption, as required by the FGGM, is violated in this data, which might be one of the contributing factors of the difference between FGGM and the copula-based NS-L and RS-L. Figure 4 shows the histograms of the first coefficients in the Karhunen-Loeve expansions for the random functions from three channels: channel Fp1 for the alcoholic group, channel Fz for the alcoholic group, and channel Fz for the non-alcoholic group. These histograms display strong skewness, violating the Gaussian assumption.

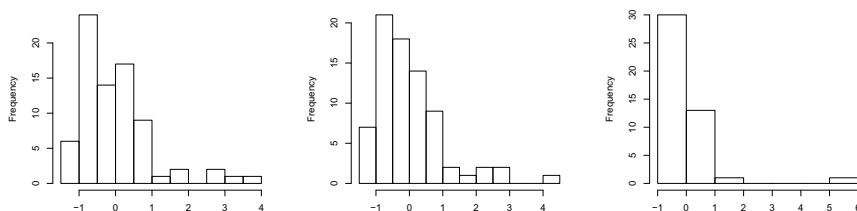


Figure 4. Histograms for the first coefficients in the Karhunen-Loeve expansions for channel Fp1 for the alcoholic group (left), channel Fz for the alcoholic group (middle), and channel Fz for the non-alcoholic group (right).

8 Discussion

In this paper we put forward the idea of the functional copula Gaussian model, and use it to develop a flexible non-Gaussian functional graphical model. The crux of this idea is to apply copula transformations to the coefficients in the Karhunen-Loeve expansion of a random function, which, at the sample level, amounts to first taking the ranks of these coefficients and then transform them by the Gaussian quantile function. The advantage of the functional copula approach is that it retains dynamics within a random function but makes the conditional dependence among random

functions in the same observation unit behave like Gaussian conditional dependence. This not only simplifies the computation but also avoids any high-dimensional kernels that can hamper estimation accuracy. We have established the consistency and convergence rates of this approach, and in the process introduced novel techniques for the asymptotic analysis for the functional copula models.

The functional copula model leads to many theoretical and computational problems that cannot all be tackled within the scope of the current paper. We now outline six directions of research that need further development. First, the asymptotic developments here are focussed on the case where the dimension p is fixed when the sample size n tends to infinity. It is plausible that some or all of these results can be extended to the case where p tends to infinity with n , perhaps along the lines of Liu et al. (2012) and Xue and Zou (2012). Second, the asymptotic developments in this paper are based on the assumption that the random function X^i is observed in its entirety, ignoring the fact that in practice they can only be observed on a finite set of time points. Third, we have yet to develop the asymptotic distribution of the proposed functional copula estimators. Fourth, in the multivariate and high-dimensional setting, Gu et al. (2015) further developed statistical inference procedures for the copula Gaussian graphical model, including a test procedure for the presence of a single edge, and a confidence subgraph. We expect that the techniques employed there can be adopted the current functional graphical model for statistical inference. Fifth, in this paper we have chosen the truncation constant m_n and the dimension k_n empirically, for example, in the simulation we set $k_n = 7$ and then select $m_n = 3$ such that more than 90% of the total variation can be explained. A more systematic tuning constant selection procedure needs to be developed, for example, by cross-validation. Finally, as a referee pointed out, the current paper is based on truncated Karhunen-Loeve expansions of each marginal random function X^i , but it would be more efficient to perform a multivariate Karhunen-Loeve expansion (Chiou et al.,

2014) on (X^1, \dots, X^p) and then apply the copula transformations to this multivariate sequence.

It is true that even the Gaussian copula model is still a restrictive assumption, which is not satisfied by many stochastic functions. Nevertheless, the class of copula Gaussian random functions is a much larger than the class of Gaussian random functions. Essentially, by using the functional copula model we have enlarged the family of applicable models from

$$\left\{ \sum_{r=1}^{\infty} \lambda_r^{1/2} \xi_r \phi_r : \sum_{r=1}^{\infty} \lambda_r < \infty, \xi_r \text{'s are i.i.d. } N(0, 1) \right\}$$

to

$$\left\{ \sum_{r=1}^{\infty} \lambda_r^{1/2} h_r(\xi_r) \phi_r : \sum_{r=1}^{\infty} \lambda_r < \infty, \xi_r \text{'s are i.i.d. } N(0, 1), \right. \\ \left. h_r \text{'s are increasing functions with } E h_r(\xi_r) = 0, \text{ var}[h_r(\xi_r)] = 1 \right\}.$$

While it can be argued — validly — that this family is still not large enough, the same criticism also applies to the classical copula Gaussian graphical models, which have been quite successful in various applications in spite of its limitation.

Finally, although we have focussed on functional graphical models, the idea of functional copula model can have far wider implications. Functional data have become increasingly common in modern data analysis, and many estimation and testing procedures have been developed, as can be found in Ramsay and Silverman (2005), Yao, Müller, and Wang (2005), Ferraty and Vieu (2006), Horváth and Kokoszka (2012), and Hsing and Eubank (2015), among many others. It is our hope that the functional copula model as well as the related asymptotic theory presented in this paper can open an avenue for further developing many of the above methods.

Acknowledgements

We thank two referees and an Associate Editor for their many insightful and constructive comments and suggestions, which helped us greatly in improving this work. Bing Li's research is supported in part by NSF grant DMS-1713078.

References

- Bach, F. R. (2008), "Consistency of the group lasso and multiple kernel learning," *The Journal of Machine Learning Research*, 9, 1179–1225.
- Baker, C. R. (1973), "Joint Measures and Cross-Covariance Operators," *Transactions of The American Mathematical Society*, 186, 273–289.
- Bosq, D. (2000), *Linear Process in Function Spaces: Theory and Application. Lecture Notes in Statistics 149*, Springer.
- Cheng, R. and Herskovits, E. H. (2007), "Graphical-model-based multivariate analysis of functional magnetic resonance data," *NeuroImage*, 35(2), 635–647.
- Chiou, J.-M., Chen, Y.-T., and Yang, Y.-F. (2014), "Multivariate functional principal component analysis: A normalization approach," *Statistica Sinica*, 24, 1571–1596.
- Darroch, J. N., Lauritzen, S. L., and Speed, T. P. (1980), "Markov fields and log linear models for contingency tables," *The Annals of Statistics*, 8, 522–539.
- Dawid, A. P. and Lauritzen, S. L. (1993), "Hyper Markov laws in the statistical analysis of decomposable graphical models," *The Annals of Statistics*, 21, 1272–1317.
- Ferraty, F. and Vieu, P. (2006), *Nonparametric functional data analysis: theory and practice*, Springer Science & Business Media.
- Gu, Q., Cao, Y., Ning, Y., and Liu, H. (2015), "Local and Global Inference for High Dimensional Nonparanormal Graphical Models," Available at <https://arxiv.org/abs/1502.02347>.
- Horváth, L. and Kokoszka, P. (2012), *Inference for functional data with applications*, vol. 200, Springer Science & Business Media.

- Hsing, T. and Eubank, R. (2015), *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*, Wiley.
- Ingber, L. (1997), “Statistical mechanics of neocortical interactions: Canonical momenta indicators of electroencephalography,” *Physical Review E*, 55, 4578.
- Kendall, M. G. (1948), *Rank correlation methods*, Charles Griffin and Co. Ltd. London.
- Kruskal, W. H. (1958), “Ordinal measures of association,” *Journal of the American Statistical Association*, 53, 814–861.
- Lauritzen, S. L. (1996), *Graphical Models*, Oxford: Clarendon Press.
- Lauritzen, S. L., Speed, T. P., and Vijayan, K. (1984), “Decomposable graphs and hypergraphs,” *Journal of the Australian Mathematical Society, Series A*, 36, 12–29.
- Lazar, N. A., Lura, B., Sweeney, J. A., and Eddy, W. F. (2002), “Combining brains: a survey of methods for statistical pooling of information,” *NeuroImage*, 16(2), 538–550.
- Li, B. (2018), “Linear operator-based statistical analysis: a useful paradigm for big data,” *The Canadian Journal of Statistics*, 46, 79–103.
- Li, B. and Babu, J. (2019), *A Graduate Course on Statistical Inference*, Springer.
- 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.
- Li, B., Kim, M. K., and Altman, N. (2010), “On dimension folding of matrix-or array-valued statistical objects,” *The Annals of Statistics*, 1094–1121.
- 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*. Early view online at <https://amstat.tandfonline.com/toc/uasa20/0/ja>, 113, 1637–1655.
- Li, B. and Song, J. (2017), “Nonlinear sufficient dimension reduction for functional data,” *Annals of Statistics*, 45, 1059–1095.
- Liu, H., Han, F., Yuan, M., Lafferty, J., and Wasserman, L. (2012), “The nonparametric skeptic,” *arXiv preprint arXiv:1206.6488*.

- Liu, H., Lafferty, J., and Wasserman, L. (2009), “The nonparanormal: Semiparametric estimation of high dimensional undirected graphs,” *The Journal of Machine Learning Research*, 10, 2295–2328.
- Meinshausen, N. and Bühlmann, P. (2006), “High-dimensional graphs with the lasso,” *The Annals of Statistics*, 34, 1436–1462.
- Qiao, X., Guo, S., and James, G. M. (2019a), “Functional Graphical Models,” *Journal of the American Statistical Association*, 114, 211–222.
- Qiao, X., Qian, C., James, G. M., and Guo, S. (2019b), “Doubly functional graphical models in high dimensions,” *Biometrika*, to appear.
- Ramsay, J. O. and Silverman, B. W. (2005), *Functional Data Analysis*, Springer.
- Wermuth, N. and Lauritzen, S. L. (1983), “Graphical and recursive models for contingency tables,” *Biometrika*, 70, 537–552.
- (1990), “On substantive research hypotheses, conditional independence graphs and graphical chain models (with discussion),” *Journal of the Royal Statistical Society, Series B*, 52, 21–72.
- Whittaker, J. (1990), *Graphical Models in Applied Multivariate Statistics*, Wiley, Chichester.
- Xue, L. and Zou, H. (2012), “Regularized rank-based estimation of high-dimensional nonparanormal graphical models,” *The Annals of Statistics*, 40, 2541–2571.
- Yao, F., Müller, H.-G., and Wang, J.-L. (2005), “Functional data analysis for sparse longitudinal data,” *Journal of the American Statistical Association*, 100, 577–590.
- Yuan, M. and Lin, Y. (2007), “Model selection and estimation in Gaussian graphical model,” *Biometrika*, 94, 19–35.
- Zhang, X. L., Begleiter, H., Porjesz, B., Wang, W., and Litke, A. (1995), “Event related potentials during object recognition tasks,” *Brain Research Bulletin*, 38, 531–538.
- Zhu, H., Strawn, N., and Dunson, D. B. (2016), “Bayesian Graphical Models for Multivariate Functional Data,” *Journal of Machine Learning Research*, 17, 1–7.