Biometrika (2022), **109**, 3, *pp.* 665–681 doi: 10.1093/biomet/asab046
Advance Access publication 19 December 2021

Partial separability and functional graphical models for multivariate Gaussian processes

By J. ZAPATA, S. Y. OH and A. PETERSEN

Department of Statistics and Applied Probability, University of California Santa Barbara, Santa Barbara, California 93106, U.S.A.

zapata@pstat.ucsb.edu syoh@pstat.ucsb.edu petersen@stat.byu.edu

Summary

The covariance structure of multivariate functional data can be highly complex, especially if the multivariate dimension is large, making extensions of statistical methods for standard multivariate data to the functional data setting challenging. For example, Gaussian graphical models have recently been extended to the setting of multivariate functional data by applying multivariate methods to the coefficients of truncated basis expansions. However, compared with multivariate data, a key difficulty is that the covariance operator is compact and thus not invertible. This paper addresses the general problem of covariance modelling for multivariate functional data, and functional Gaussian graphical models in particular. As a first step, a new notion of separability for the covariance operator of multivariate functional data is proposed, termed partial separability, leading to a novel Karhunen–Loève-type expansion for such data. Next, the partial separability structure is shown to be particularly useful in providing a well-defined functional Gaussian graphical model that can be identified with a sequence of finite-dimensional graphical models, each of identical fixed dimension. This motivates a simple and efficient estimation procedure through application of the joint graphical lasso. Empirical performance of the proposed method for graphical model estimation is assessed through simulation and analysis of functional brain connectivity during a motor task.

Some key words: Functional brain connectivity; Functional data; Inverse covariance; Separability.

1. Introduction

The analysis of functional data continues to be an important field for statistical development given the abundance of data collected over time via sensors or other tracking equipment. Frequently, such time-dependent signals are vector-valued, resulting in multivariate functional data. Prominent examples include longitudinal behavioural tracking (Chiou & Müller, 2014), blood protein levels (Dubin & Müller, 2005), traffic measurements (Chiou et al., 2014, 2016) and neuroimaging data (Petersen & Müller, 2016; Happ & Greven, 2018), for which dimensionality reduction and regression have been the primary methods investigated. As for standard multivariate data, the nature of dependencies between component functions of multivariate functional data is an important question requiring careful consideration.

Dependencies between functional magnetic resonance imaging, fMRI, signals for a large number of regions across the brain during a motor task experiment consititute the motivating example for this paper. Since fMRI signals are collected simultaneously, it is natural to model

them as a multivariate process $\{X(t) \in \mathbb{R}^p : t \in \mathcal{T}\}$, where $\mathcal{T} \subset \mathbb{R}$ is a time interval over which the scans are taken (Qiao et al., 2019). The dual multivariate and functional aspects of the data make the covariance structure of X complex, particularly if the multivariate dimension p is large. This leads to difficulties in extending highly useful multivariate analysis techniques, such as graphical models, to multivariate functional data without further structural assumptions. For example, in the analogous setting of spatiotemporal data, it is common to impose further structure on the covariance, usually assuming that the spatial and temporal effects can be separated in some way. However, similar notions for multivariate functional data have not yet been considered.

As for ordinary multivariate data, the conditional independence properties of X are perhaps of greater interest than the marginal covariance, leading to the consideration of inverse covariance operators and graphical models for functional data. If X is Gaussian, each component function X_j corresponds to a node in the functional Gaussian graphical model, which is a single network of p nodes. This is inherently different from the estimation of time-dependent graphical models (Zhou et al., 2010; Kolar & Xing, 2011; Qiu et al., 2016; Qiao et al., 2020), in which the graph is dynamic and has nodes corresponding to scalar random variables. In this paper, the graph is considered to be static, while each node represents an infinite-dimensional functional object. This is an important distinction, as covariance operators for functional data are compact and thus not invertible in the usual sense, so that presence or absence of edges cannot in general be identified immediately with zeros in any precision operator. In the past few years there has been some investigation into this problem. Zhu et al. (2016) developed a Bayesian framework for graphical models on product function spaces, including the extension of Markov laws and appropriate prior distributions. Qiao et al. (2019) implemented a truncation approach, where each function is represented by the coefficients of a truncated basis expansion using functional principal component analysis, and a finite-dimensional graphical model is estimated by a modified graphical lasso criterion. Li & Solea (2018) developed a non-Gaussian variant, where conditional independence is replaced by a notion of so-called additive conditional independence.

The methodology proposed in this paper is within the setting of multivariate Gaussian processes as in Qiao et al. (2019), and exploits a notion of separability for multivariate functional data to develop efficient estimation of suitable inverse covariance objects. There are at least three novel contributions of this methodology to the fields of functional data analysis and Gaussian graphical models. First, a structure termed partial separability is defined for the covariance operator of multivariate functional data, yielding a novel Karhunen–Loève-type representation. The second contribution is to show that when the process is indeed partially separable, the functional graphical model is well-defined and can be identified with a sequence of finite-dimensional graphical models. In particular, the assumption of partial separability overcomes the problem of noninvertibility of the covariance operator when X is infinite-dimensional, in contrast with Zhu et al. (2016) and Qiao et al. (2019), which assume that the functional data are concentrated on finite-dimensional subspaces. Third, an intuitive estimation procedure is developed based on simultaneous estimation of multiple graphical models. Furthermore, theoretical properties are derived under the regime of fully observed functional data. Empirical performance of the proposed method is then compared with that of Qiao et al. (2019) through simulations involving dense and noisily observed functional data, including a setting where partial separability is violated. Finally, the method is applied to the study of brain connectivity, also known as functional connectivity in the neuroscience literature, using data from the Human Connectome Project corresponding to a motor task experiment. Through these practical examples, our proposed method is shown to yield improved efficiency in estimation and computation. An `R` package (R Development Core Team, 2022) `fgm` implementing the proposed methods is freely available from the CRAN repository.

2. Preliminaries

2.1. *Gaussian graphical models*

Consider a p-variate random variable $\theta = (\theta_1, \ldots, \theta_p)^{\mathrm{T}}$, where $p > 2$. For any distinct indices $j, k = 1, \ldots, p$ let $\theta_{-(j,k)} \in \mathbb{R}^{p-2}$ denote the subvector of θ obtained by removing its jth and kth entries. A graphical model (Lauritzen, 1996) for θ is an undirected graph $G = (V, E)$, where $V = \{1, \ldots, p\}$ is the node set and $E \subset V \times V \setminus \{(j,j) : j \in V\}$ is the edge set. The edges in E encode the presence or absence of conditional independencies among the distinct components of θ by excluding (j, k) from E if and only if $\theta_j \perp\!\!\!\perp \theta_k \mid \theta_{-(j,k)}$. In the case where $\theta \sim N_p(0, \Sigma)$, the corresponding Gaussian graphical model is intimately connected to the positive-definite covariance matrix Σ through its inverse $\Omega = \Sigma^{-1}$, known as the precision matrix of θ. Specifically, the edge set E can be readily obtained from Ω by the relation $(j, k) \in E$ if and only if $\Omega_{jk} \neq 0$ (Lauritzen, 1996). This identification of edges in E with the nonzero off-diagonal entries of Ω arises from the simple fact that the latter are proportional to the conditional covariance between components. Thus, the zero/nonzero structure of Ω serves as an adjacency matrix of the graph G, allowing the employment of a vast number of statistical tools for sparse inverse covariance estimation to recover a sparse graph structure from data.

2.2. *Functional Gaussian graphical models*

We first introduce some notation. Let $L^2[0,1]$ denote the space of square-integrable measurable functions on $[0,1]$ endowed with the standard inner product $\langle g_1, g_2 \rangle = \int_0^1 g_1(t) g_2(t)\, \mathrm{d}t$ and associated norm $\|\cdot\|$. Let $(L^2[0,1])^p$ be its p-fold Cartesian product or direct sum, endowed with the inner product $\langle f_1, f_2 \rangle_p = \sum_{j=1}^p \langle f_{1j}, f_{2j} \rangle$ and associated norm $\|\cdot\|_p$. For a generic compact covariance operator $\mathcal{A}$ defined on an arbitrary Hilbert space, let $\lambda_j^{\mathcal{A}}$ denote its jth-largest eigenvalue. Suppose $f \in (L^2[0,1])^p$, $g \in L^2[0,1]$, $a \in \mathbb{R}^p$, Δ is a $p \times p$ matrix and $\mathcal{B} : L^2[0,1] \to L^2[0,1]$ is a linear operator. Then $ag \in (L^2[0,1])^p$ takes values $\{g(x)\}a \in \mathbb{R}^p$, $\Delta f \in (L^2[0,1])^p$ takes values $\Delta\{f(x)\} \in \mathbb{R}^p$, $\mathcal{B}(f) = \{\mathcal{B}(f_1), \ldots, \mathcal{B}(f_p)\} \in (L^2[0,1])^p$ and $(\Delta\mathcal{B})(f) = \mathcal{B}(\Delta f)$. The tensor products $g \otimes g$ and $f \otimes_p f$ signify the operators $(g \otimes g)(\cdot) = \langle g, \cdot \rangle g$ and $(f \otimes_p f)(\cdot) = \langle f, \cdot \rangle_p f$ on $L^2[0,1]$ and $(L^2[0,1])^p$, respectively.

In this paper, multivariate functional data constitute a random sample from a multivariate process $\{X(t) \in \mathbb{R}^p : t \in [0,1]\}$, which is assumed, for the moment, to be zero-mean such that $X \in (L^2[0,1])^p$ almost surely and $E(\|X\|_p^2) < \infty$. If X is also assumed to be Gaussian, then its distribution is uniquely characterized by its covariance operator $\mathcal{G}$, the infinite-dimensional counterpart of the covariance matrix for standard multivariate data. In fact, one can think of it as a matrix of operators $\mathcal{G} = \{\mathcal{G}_{jk} : j, k \in V\}$, where each entry $\mathcal{G}_{jk}$ is a linear, trace-class integral operator on $L^2[0,1]$ (Hsing & Eubank, 2015) with kernel $G_{jk}(s,t) = \mathrm{cov}\{X_j(s), X_k(t)\}$. That is, for any $g \in L^2[0,1]$, $\mathcal{G}_{jk}(g)(\cdot) = \int_0^1 G_{jk}(\cdot, t) g(t)\, \mathrm{d}t$. Then $\mathcal{G}$ is an integral operator on $(L^2[0,1])^p$ with $\{\mathcal{G}(f)\}_j = \sum_{k=1}^p \mathcal{G}_{jk}(f_k)$ for $f \in (L^2[0,1])^p$ and $j \in V$.

A functional Gaussian graphical model for X is a graph $G = (V, E)$ that encodes the conditional independence structure among its components. As in the finite-dimensional case, the edge set can be recovered from the conditional covariance functions

$$C_{jk}(s,t) = \mathrm{cov}\{X_j(s), X_k(t) \mid X_{-(j,k)}\} \quad (j, k \in V, j \neq k) \tag{1}$$

through the relation $(j, k) \in E$ if and only if $C_{jk}(s,t) = 0$ for all $s, t \in [0,1]$. However, unlike in the finite-dimensional case, the covariance operator $\mathcal{G}$ is compact and thus not invertible, with the consequence that the connection between conditional independence and an inverse covariance

operator is lost, as the latter does not exist. This is a known issue for infinite-dimensional functional data, for instance in linear regression models with functional predictors; see Wang et al. (2016) and references therein. Hence, a common approach is to regularize the problem by first performing dimensionality reduction, most often through a truncated basis expansion of the functional data. Specifically, one chooses an orthonormal functional basis $\{\phi_{jl}\}_{l=1}^{\infty}$ of $L^2[0,1]$ for each j and expresses each component of X as

$$X_j(t) = \sum_{l=1}^{\infty} \xi_{jl}\phi_{jl}(t), \quad \xi_{jl} = \int_0^1 X_j(t)\phi_{jl}(t)\,\mathrm{d}t. \tag{2}$$

These expansions are then truncated at a finite number of basis functions to perform estimation, and the basis size is allowed to diverge with the sample size to obtain asymptotic properties.

Previous work related to functional Gaussian graphical models includes Zhu et al. (2016) and Qiao et al. (2019). The former rigorously considered the notion of conditional independence for functional data and proposed a family of priors for the covariance operator $\mathcal{G}$. The latter truncated (2) at L terms using the functional principal component basis (Hsing & Eubank, 2015) and set $\xi_j = (\xi_{j1}, \ldots, \xi_{jL})^{\mathrm{T}}$ $(j \in V)$. Qiao et al. (2019) defined a $pL \times pL$ covariance matrix Γ blockwise for the concatenated vector $(\xi_1{}^{\mathrm{T}}, \ldots, \xi_p{}^{\mathrm{T}})^{\mathrm{T}}$ as $\Gamma = (\Gamma_{jk})_{j,k=1}^{p}$ with $(\Gamma_{jk})_{lm} = \mathrm{cov}(\xi_{jl}, \xi_{km})$ $(l, m = 1, \ldots, L)$. Then, a functional graphical lasso algorithm was developed to estimate Γ^{-1} with sparse off-diagonal blocks in order to estimate the edge set.

The method of Qiao et al. (2019) represents an intuitive approach to functional graphical model estimation, but it encounters some difficulties, which we seek to address in this paper. From a theoretical point of view, even when p is fixed, consistent estimation of the graphical model requires that one permits L to diverge, so that the number of covariance parameters needing to be estimated is $(pL)^2$. Additionally, the identification of zero off-diagonal blocks in Γ^{-1} was shown to be linked to the true functional graphical model only under the strict assumption that each X_j take values in a finite-dimensional space almost surely. In many practical applications, the dimension p can be high, the number of basis functions L may need to be large in order to retain a suitable representation of the observed data, or both of these may occur simultaneously. It is thus desirable to introduce structure on $\mathcal{G}$ so as to provide a parsimonious basis expansion for multivariate functional data that is amenable to graphical model estimation.

3. PARTIAL SEPARABILITY

3.1. *A parsimonious basis for multivariate functional data*

Functional principal component analysis is a commonly used tool for functional data, with one of its most useful features being the parsimonious reduction of each univariate component X_j to a countable sequence of uncorrelated random variables through the Karhunen–Loève expansion (Hsing & Eubank, 2015), taking the form of (2) when the basis is chosen as the eigenbasis of $\mathcal{G}_{jj}$. Chiou et al. (2014) extended this approach to multivariate functional data via the spectral decomposition $\mathcal{G} = \sum_{m=1}^{\infty} \lambda_m^{\mathcal{G}} \rho_m \otimes_p \rho_m$, leading to the multivariate Karhunen–Loève expansion $X(t) = \sum_{l=1}^{\infty} \langle X, \rho_l \rangle_p \rho_l(t)$, where $\{\rho_l\}_{l=1}^{\infty}$ is an orthonormal basis of $(L^2[0,1])^p$. While this decomposition is indeed parsimonious, the multivariate aspect of the data is lost since the random coefficients $\langle X, \rho_l \rangle_p$ are scalar. As a consequence, one cannot readily apply tools from finite-dimensional Gaussian graphical model estimation to these coefficients. We begin by proposing a novel structural assumption on the eigenfunctions of $\mathcal{G}$, termed partial separability, and will then demonstrate its advantages for defining and estimating functional Gaussian graphical models.

DEFINITION 1. *A covariance operator $\mathcal{G}$ on $(L^2[0,1])^p$ is partially separable if there exist orthonormal bases $\{e_{lj}\}_{j=1}^p$ $(l \in \mathbb{N})$ of $\mathbb{R}^p$ and $\{\varphi_l\}_{l=1}^\infty$ of $L^2[0,1]$ such that the eigenfunctions of $\mathcal{G}$ take the form $e_{lj}\varphi_l$ $(l \in \mathbb{N}, j \in V)$.*

We first draw a connection to separability of covariance operators as they appear in spatiotemporal analyses, after which the implications of partial separability will be further explored. Dependent functional data arise naturally in the context of a spatiotemporal random field that is sampled at p discrete spatial locations. In many instances, it is assumed that the covariance of X is separable, meaning that there exist a $p \times p$ covariance matrix Δ and a covariance operator $\mathcal{B}$ on $L^2[0,1]$ such that $\mathcal{G} = \Delta\mathcal{B}$ (Gneiting et al., 2006; Genton, 2007; Aston et al., 2017). Letting $\{e_j\}_{j=1}^p$ and $\{\varphi_l\}_{l=1}^\infty$ be the orthonormal eigenbases of Δ and $\mathcal{B}$, respectively, it is clear that $e_j\varphi_l$ are the eigenfunctions of $\mathcal{G}$. Hence, a separable covariance operator $\mathcal{G}$ satisfies the conditions of Definition 1. It should be noted that the property of $\mathcal{G}$ having eigenfunctions of the form $e_j\varphi_l$ has also been referred to as weak separability (Lynch & Chen, 2018), and is a consequence and not a characterization of separability. The connections between these three separability notions are summarized in the following result, whose proof is simple and therefore omitted.

PROPOSITION 1. *Suppose $\mathcal{G}$ is partially separable according to Definition 1. Then $\mathcal{G}$ is also weakly separable if and only if the bases $\{e_{lj}\}_{j=1}^p$ do not depend on l. If $\mathcal{G}$ is weakly separable, then it is also separable if and only if the eigenvalues take the form $\langle \mathcal{G}(e_j\varphi_l), e_j\varphi_l\rangle_p = c_j d_l$ for positive sequences $\{c_j\}_{j=1}^p$ and $\{d_l\}_{l=1}^\infty$.*

The next result gives several characterizations of partial separability. The proof of this theorem and all remaining theoretical results can be found in the Supplementary Material.

THEOREM 1. *Let $\{\varphi_l\}_{l=1}^\infty$ be an orthonormal basis of $L^2[0,1]$. The following are equivalent.*

(i) *$\mathcal{G}$ is partially separable with $L^2[0,1]$ basis $\{\varphi_l\}_{l=1}^\infty$.*
(ii) *There exists a sequence of $p \times p$ covariance matrices $\{\Sigma_l\}_{l=1}^\infty$ such that*

$$\mathcal{G} = \sum_{l=1}^{\infty} \Sigma_l \varphi_l \otimes \varphi_l.$$

(iii) *The covariance operator of each X_j can be written as $\mathcal{G}_{jj} = \sum_{l=1}^\infty \sigma_{ljj}\varphi_l \otimes \varphi_l$, with $\sigma_{ljj} > 0$ and $\sum_{l=1}^\infty \sigma_{ljj} < \infty$ and with $\operatorname{cov}(\langle X_j, \varphi_l\rangle, \langle X_k, \varphi_{l'}\rangle) = 0$ for $j, k \in V$ and $l \neq l'$.*
(iv) *The expansion*

$$X = \sum_{l=1}^{\infty} \theta_l \varphi_l, \quad \theta_l = (\langle X_1, \varphi_l\rangle, \ldots, \langle X_p, \varphi_l\rangle)^{\mathrm{T}}, \tag{3}$$

holds almost surely in $(L^2[0,1])^p$, where the θ_l are mutually uncorrelated random vectors.

As will be seen in § 3.2, the matrices Σ_l in statement (ii) of Theorem 1 contain all the necessary information to form the functional graphical model when X is Gaussian and $\mathcal{G}$ is partially separable. For clarity, when $\mathcal{G}$ is partially separable, the expansion in (ii) is assumed to be ordered according to decreasing value of $\operatorname{tr}(\Sigma_l)$. Statement (iii) reveals that the $\mathcal{G}_{jj}$ share common eigenfunctions and are thus simultaneously diagonalizable, with projections of any features onto

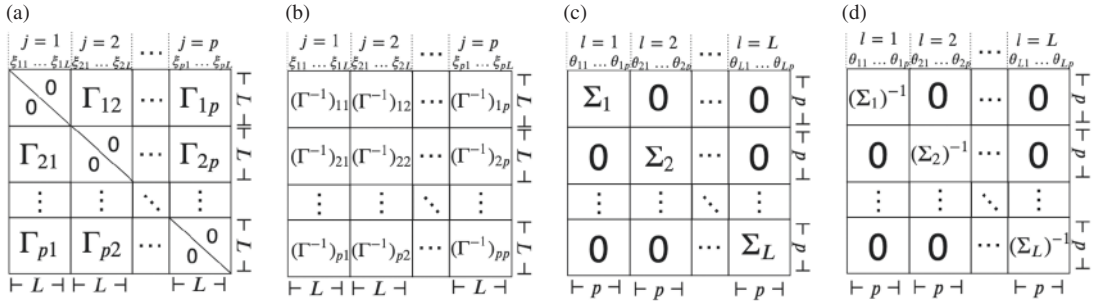

Fig. 1. Covariance structures of $\mathbb{R}^{Lp}$-valued random coefficients from different L-truncated Karhunen–Loève-type expansions: (a) and (b) show covariance and precision matrices, respectively, of functional principal component coefficients $(\xi_1^{\mathrm{T}}, \ldots, \xi_p^{\mathrm{T}})^{\mathrm{T}}$ in (2) as in Qiao et al. (2019); (c) and (d) show block-diagonal covariance and precision matrices, respectively, of coefficients $(\theta_1^{\mathrm{T}}, \ldots, \theta_L^{\mathrm{T}})^{\mathrm{T}}$ under partial separability in (3).

different eigenfunctions being uncorrelated. This is related to the concept known as coregionalization (Banerjee & Johnson, 2006), but is more general in that it allows different eigenvalues for each $\mathcal{G}_{jj}$. Consequently, one obtains the vector Karhunen–Loève-type expansion in (3). If one truncates (3) at L components, the covariance matrix of the concatenated vector $(\theta_1{}^{\mathrm{T}}, \ldots, \theta_L{}^{\mathrm{T}})^{\mathrm{T}}$ is block-diagonal, with the $p \times p$ matrices $\Sigma_l = \mathrm{var}(\theta_l)$ constituting the diagonal blocks. Figure 1 visualizes this covariance structure in comparison with that of Qiao et al. (2019), along with comparisons of the inverse covariance structure. The latter comparison is the more striking and relevant one, since the model of Qiao et al. (2019) possesses a potentially full inverse covariance structure, whereas that under partial separability remains block-diagonal. As a consequence, the model of Qiao et al. (2019) has $O(L^2p^2)$ free parameters, while the corresponding model under partial separability has only $O(Lp^2)$ free parameters.

Lastly, we establish optimality and uniqueness properties for the basis $\{\varphi_l\}_{l=1}^{\infty}$ of a partially separable $\mathcal{G}$. A key object is the trace-class covariance operator

$$\mathcal{H} = \frac{1}{p}\sum_{j=1}^{p} \mathcal{G}_{jj}. \tag{4}$$

Let $\lambda_l = \lambda_l^{\mathcal{H}}$ $(l \in \mathbb{N})$ denote the eigenvalues of $\mathcal{H}$, in nonincreasing order.

THEOREM 2. *Suppose the eigenvalues of $\mathcal{H}$ in (4) have multiplicity* 1.

(i) *For any $L \in \mathbb{N}$ and for any orthonormal set $\{\tilde{\varphi}_l\}_{l=1}^{L}$ in $L^2[0,1]$, $\sum_{l=1}^{L}\sum_{j=1}^{p} \mathrm{var}(\langle X_j, \tilde{\varphi}_l\rangle) \leqslant \sum_{l=1}^{L} \lambda_l$, with equality if and only if $\{\tilde{\varphi}\}_{l=1}^{L}$ span the first L eigenspaces of $\mathcal{H}$.*

(ii) *If $\mathcal{G}$ is partially separable with $L^2[0,1]$ basis $\{\varphi_l\}_{l=1}^{\infty}$, then*

$$\mathcal{H} = \sum_{l=1}^{\infty} \lambda_l \varphi_l \otimes \varphi_l, \quad \lambda_l = \frac{1}{p}\,\mathrm{tr}(\Sigma_l).$$

Statement (i) says that, independent of partial separability, the eigenbasis of $\mathcal{H}$ is optimal in terms of retaining the greatest amount of total variability in vectors of the form $(\langle X_1, \tilde{\varphi}_l\rangle, \ldots, \langle X_p, \tilde{\varphi}_l\rangle)^{\mathrm{T}}$, subject to orthogonality constraints. Statement (ii) indicates that if $\mathcal{G}$ is partially separable, the unique basis of $L^2[0,1]$ that makes Definition 1 hold corresponds to this optimal basis.

3.2. *Consequences for functional Gaussian graphical models*

Assume that $\mathcal{G}$ is partially separable according to Definition 1, so that the partially separable Karhunen–Loève expansion in (3) holds. If we further assume that X is Gaussian, then the $\theta_l \sim N(0, \Sigma_l)$ $(l \in \mathbb{N})$ are independent, where Σ_l is positive definite for each l. These facts follow from Theorem 1. Recall that in order to define a coherent functional Gaussian graphical model, the conditional covariance functions C_{jk} in (1) between component functions X_j and X_k need to be well-defined. The expansion in (3) facilitates a simple connection between the C_{jk} and the inverse covariance matrices $\Omega_l = \Sigma_l^{-1}$, as follows. Let $\Sigma_l = (\sigma_{ljk})_{j,k=1}^p$. For any fixed $j, k \in V$ define the partial covariance between θ_{lj} and θ_{lk} as

$$\tilde{\sigma}_{ljk} = \sigma_{ljk} - \operatorname{cov}\{\theta_{lj}, \theta_{l,-(j,k)}\} \operatorname{var}\{\theta_{l,-(j,k)}\}^{-1} \operatorname{cov}\{\theta_{l,-(j,k)}, \theta_{lk}\}.$$

It is well known that these partial covariances are directly related to the precision matrix $\Omega_l = (\omega_{ljk})_{j,k=1}^p$ by $\tilde{\sigma}_{ljk} = -\omega_{ljk}/(\omega_{ljj}\omega_{lkk} - \omega_{ljk}^2)$, so that $\tilde{\sigma}_{ljk} = 0$ if and only if $\omega_{ljk} = 0$. The next result establishes that the conditional covariance functions C_{jk} can be expanded in the partial separability basis $\{\varphi_l\}_{l=1}^\infty$ with coefficients $\tilde{\sigma}_{ljk}$.

THEOREM 3. *If $\mathcal{G}$ is partially separable, then the cross-covariance kernel between X_j and X_k, conditional on the multivariate subprocess $\{X_{-(j,k)}(u) : u \in [0,1]\}$, is*

$$C_{jk}(s,t) = \sum_{l=1}^{\infty} \tilde{\sigma}_{ljk} \varphi_l(s) \varphi_l(t) \quad (j, k \in V, j \neq k; s, t \in [0,1]).$$

Now, the conditional independence graph for the multivariate Gaussian process can be defined by $(j,k) \notin E$ if and only if $C_{jk}(s,t) \equiv 0$. Owing to the above result, the edge set E is connected to the sequence of edge sets $\{E_l\}_{l=1}^\infty$ for which $(j,k) \notin E_l$ if and only if $\tilde{\sigma}_{ljk} = \omega_{ljk} = 0$, corresponding to the sequence of Gaussian graphical models (V, E_l) for each θ_l.

COROLLARY 1. *In the setting of Theorem 3, the functional graph edge set E is related to the sequence of edge sets E_l by $E = \bigcup_{l=1}^\infty E_l$.*

We have thus established that under partial separability, the problem of functional graphical model estimation can be simplified to estimation of a sequence of decoupled graphical models. When partial separability fails, the edge sets remain meaningful. Recall from Theorem 2 that the eigenbasis of $\mathcal{H}$ is optimal in a sense that is independent of partial separability, so that the vectors $\theta_l = (\langle X_1, \varphi_l \rangle, \ldots, \langle X_p, \varphi_l \rangle)^{\mathrm{T}}$ are still the coefficients of X in an optimal expansion. Although one loses a direct connection between the E_l and the edge set of the functional graph, each E_l remains the edge set of the Gaussian graphical model for the coefficient vector θ_l in this optimal expansion. Moreover, the equivalence $E = \bigcup_{l=1}^\infty E_l$ may hold independent of partial separability. For instance, the Supplementary Material gives sufficient conditions, based on a Markov-type property and an edge coherence assumption, under which the equivalence holds.

4. GRAPH ESTIMATION AND THEORY

4.1. *Joint graphical lasso estimator*

Consider a p-variate process X, with means $\mu_j(t) = E\{X_j(t)\}$ and covariance operator $\mathcal{G}$. Let $\{\varphi_l\}_{l=1}^\infty$ be an orthonormal eigenbasis of $\mathcal{H}$ in (4), and set $\theta_{lj} = \langle X_j, \varphi_l \rangle$ and $\Sigma_l = \operatorname{var}(\theta_l)$.

The targets are the edge sets E_l, where $(j, k) \in E_l$ if and only if $(\Sigma_l^{-1})_{jk} \neq 0$, as motivated by the developments in § 3.2. Specifically, when X is Gaussian and $\mathcal{G}$ is partially separable, the conditional independence graph of X has edge set $E = \bigcup_{l=1}^{\infty} E_l$. When partial separability fails, these targets still provide useful information about the conditional independencies of X when projected onto the eigenbasis of $\mathcal{H}$, which is optimal in the sense of Theorem 2. Furthermore, when X is not Gaussian, rather than representing conditional independence, the E_l represent the sparsity structure of the partial correlations of θ_l, which may still be of interest. By Theorem 1, $\mathrm{tr}(\Sigma_l) = \lambda_l \downarrow 0$ as $l \to \infty$. A practical consideration is that this makes estimators of Σ_l progressively more unstable to work with as l increases. To avoid this difficulty, we work with $\Xi_l = R_l^{-1}$, where R_l is the correlation matrix corresponding to Σ_l; Ξ_l and Ω_l share the same edge information as entries in these two matrices are either zero or nonzero simultaneously.

We first define the estimation procedure with targets Ξ_l from a random sample $X_1, \ldots, X_n$, each distributed as X. We remark that X is not required to be Gaussian, nor $\mathcal{G}$ partially separable, in developing the theoretical properties of the estimators, which also allow the dimension p to diverge with n. In order to make these methods applicable to any functional dataset, it is assumed that preliminary mean and covariance estimates $\hat{\mu}_j$ and $\hat{\mathcal{G}}_{jk}$ $(j, k = 1, \ldots, p)$ have been computed for each component. As an example, if the X_i are fully observed, cross-sectional estimates

$$\hat{\mu}_j = \frac{1}{n}\sum_{i=1}^{n} X_{ij}, \quad \hat{\mathcal{G}}_{jk} = \frac{1}{n}\sum_{i=1}^{n}(X_{ij} - \hat{\mu}_j) \otimes (X_{ik} - \hat{\mu}_k) \tag{5}$$

can be used. For practical observational designs, smoothing can be applied to the pooled data to estimate these quantities (Yao et al., 2005; Yang et al., 2011). Given such preliminary estimates, the estimate of $\mathcal{H}$ is $\hat{\mathcal{H}} = p^{-1}\sum_{j=1}^{p} \hat{\mathcal{G}}_{jj}$, leading to empirical eigenfunctions $\hat{\varphi}_l$, which are uniquely defined only up to a sign and for $1 \leqslant l \leqslant np = \mathrm{rank}(\mathcal{H})$. These quantities produce estimates of $\sigma_{ljk} = \langle \mathcal{G}_{jk}(\varphi_l), \varphi_l \rangle$ by plug-in as

$$s_{ljk} = (S_l)_{jk} = \langle \hat{\mathcal{G}}_{jk}(\hat{\varphi}_l), \hat{\varphi}_l \rangle. \tag{6}$$

A group graphical lasso approach (Danaher et al., 2014) will be used to estimate the Ξ_l. Let $(\hat{R}_l)_{jk} = \hat{r}_{ljk} = s_{ljk}/(s_{ljj}s_{lkk})^{1/2}$ be the estimated correlations. The estimates target the first $L \leqslant np$ inverse correlation matrices Ξ_l by

$$(\hat{\Xi}_1, \ldots, \hat{\Xi}_L) = \underset{\Upsilon_l \succ 0,\, \Upsilon_l = \Upsilon_l^{\mathrm{T}}}{\arg\min} \sum_{l=1}^{L} \left\{ \mathrm{tr}(\hat{R}_l \Upsilon_l) - \log(|\Upsilon_l|) \right\} + P(\Upsilon_1, \ldots, \Upsilon_L). \tag{7}$$

In the Gaussian case, these are penalized likelihood estimators with penalty

$$P(\Upsilon_1, \ldots, \Upsilon_L) = \gamma \left\{ \alpha \sum_{l=1}^{L} \sum_{j \neq k} |\upsilon_{ljk}| + (1 - \alpha) \sum_{j \neq k} \left(\sum_{l=1}^{L} \upsilon_{ljk}^2 \right)^{1/2} \right\}, \quad (\Upsilon_l)_{jk} = \upsilon_{ljk}. \tag{8}$$

The parameter $\gamma > 0$ controls the overall penalty level, while $\alpha \in [0, 1]$ distributes the penalty between the two penalty terms. Then the estimated edge set is $(j, k) \in \hat{E}_l$ if and only if $\hat{\Xi}_{ljk} \neq 0$. The joint graphical lasso was chosen to borrow structural information across multiple bases instead of multiple classes, as was done in Danaher et al. (2014). If $\alpha = 1$, the first penalty will encourage sparsity in each $\hat{\Xi}_l$ and the corresponding edge set $\hat{E}_l$, but the overall estimate

$\hat{E} = \bigcup_{l=1}^{L} \hat{E}_l$ may not be sparse. While consistent graph recovery is still possible with $\alpha = 1$, as demonstrated below in Theorem 5, the influence of the second penalty term when $\alpha < 1$ ensures that the overall graph estimate is sparse, enhancing interpretation.

In practice, the tuning parameters γ and α can be chosen with cross-validation to minimize (7) for out-of-sample data. Specifically, the procedure would select γ and α that minimize the average of (7) evaluated over each fold, where $\Upsilon_1, \ldots, \Upsilon_L$ are computed with the training set and $\hat{R}_l$ are from the validation set. Another practically useful, and less computationally intensive, approach is to choose these parameters to yield a desired sparsity level of the estimated graph (Qiao et al., 2019). This latter approach is implemented in the data example of § 6.

4.2. *Asymptotic properties*

The goal of the current subsection is to provide lower bounds on the sample size n so that with high probability, $\hat{E}_l = E_l$ $(l = 1, \ldots, L)$. The approach follows that of Ravikumar et al. (2011), adapting the results to the case of functional graphical model estimation in which multiple graphs are estimated simultaneously. For simplicity, and to facilitate comparisons with the asymptotic properties of Qiao et al. (2019), the results are derived in the setting of fully observed functional data, so that $\hat{\mu}$ and $\hat{\mathcal{G}}_{jk}$ are as in (5). As a preliminary result, we first derive a concentration inequality for the estimated covariances s_{ljk} in (6), requiring the following mild assumptions.

Assumption 1. The eigenvalues λ_l of $\mathcal{H}$ have multiplicity 1 and are thus strictly decreasing.

Assumption 2. There exists $\varsigma^2 > 0$ such that $E\{\exp(s\theta_{lj})\} \leqslant \exp(s^2\varsigma^2\sigma_{ljj}/2)$ for all $l \in \mathbb{N}$, $j \in V$ and $s \in \mathbb{R}$; that is, the standardized scores $\theta_{lj}/\sigma_{ljj}^{1/2}$ are sub-Gaussian random variables with parameter ς^2. Furthermore, there exists M independent of p such that $\sup_{j \in V} \sum_{l=1}^{\infty} \sigma_{ljj} < M < \infty$.

Assumption 2 can be relaxed to accommodate eigenvalues with multiplicity greater than 1, at the cost of an increased notational burden. The eigenvalue spacings play a key role through the quantities $\tau_1 = 2\sqrt{2}(\lambda_1 - \lambda_2)^{-1}$ and $\tau_l = 2\sqrt{2}\max\left\{(\lambda_{l-1} - \lambda_l)^{-1}, (\lambda_l - \lambda_{l+1})^{-1}\right\}$, for $l \geqslant 2$. Assumption 2 clearly holds in the Gaussian case, and can be relaxed to accommodate different parameters ς_l^2 for each l, though for simplicity these are assumed to be uniform.

THEOREM 4. *Suppose that Assumptions 1 and 2 hold. Then there exist constants $C_1, C_2, C_3 > 0$ such that for any $0 < \delta \leqslant C_3$ and for all $l \leqslant np$ and $j, k \in V$,*

$$\mathrm{pr}(|s_{ljk} - \sigma_{ljk}| \geqslant \delta) \leqslant C_2 \exp(-C_1 \tau_l^{-2} n\delta^2). \tag{9}$$

Concentration inequalities such as (9) are generally required in penalized estimation problems where the dimension diverges to infinity. For the current problem, even if the dimension p of the process remains fixed, the dimension still diverges since one requires L to diverge with n. Furthermore, in contrast to standard multivariate scenarios, the bound in Theorem 4 contains the additional factor τ_l^{-2}. Since $\lambda_l \downarrow 0$, τ_l diverges to infinity with l, so that (9) reflects the increased difficulty of estimating covariances corresponding to eigenfunctions with smaller eigenvalue gaps.

Remark 1. A similar result to Theorem 4 was obtained by Qiao et al. (2019) under a specific eigenvalue decay rate and truncation parameter scheme. Imposing similar assumptions on the

eigenvalues of $\mathcal{H}$, we have $\tau_l = O(l^{1+\beta_1})$ for some $\beta_1 > 1$, so that for any $0 < \beta_2 < 1/(4\beta_1)$ and $L = n^{\beta_2}$, (9) implies

$$\max_{l=1,\ldots,L} \max_{j,k\in V} \mathrm{pr}(|s_{ljk} - \sigma_{ljk}| \geqslant \delta) \leqslant C_2 \exp(-C_1 n^{1-2\beta_2(1+\beta_1)}\delta^2),$$

matching the rate of Qiao et al. (2019). In addition to establishing the concentration inequality for a general eigenvalue decay rate, our proof is greatly simplified by using the inequality

$$|s_{ljk} - \sigma_{ljk}| \leqslant 2\tau_l \|\mathcal{G}_{jk}\|_{\mathrm{HS}} \|\hat{\mathcal{H}} - \mathcal{H}\|_{\mathrm{HS}} + \|\hat{\mathcal{G}}_{jk} - \mathcal{G}_{jk}\|_{\mathrm{HS}}, \tag{10}$$

where $\|\cdot\|_{\mathrm{HS}}$ is the Hilbert–Schmidt operator norm.

Remark 2. The bound in (10) utilizes a basic eigenfunction inequality found, for example, in Bosq (2000, Lemma 4.3); see also Bhatia et al. (1983). However, using expansions instead of geometric inequalities, Jirak (2016) and other authors cited therein established stronger results for differences between true and estimated eigenfunctions in the form of limiting distributions and moment bounds. Therefore, it is likely that the bound in (9) is suboptimal, although improvements along the lines of Jirak (2016) would require further challenging work to establish the required exponential tail bounds.

As the objective (7) utilizes the correlations $\hat{r}_{ljk}$, the following corollary is needed.

COROLLARY 2. *Under the assumptions of Theorem 4, there exist constants $D_1, D_2, D_3 > 0$ such that for any $0 < \delta \leqslant D_3$ and for all $l \leqslant np$ and $j, k \in V$,*

$$\mathrm{pr}(|\hat{r}_{ljk} - r_{ljk}| \geqslant \delta) \leqslant D_2 \exp(-D_1 n m_l^2 \delta^2), \quad m_l = \tau_l^{-1}\pi_l, \quad \pi_l = \min_{j\in V} \sigma_{ljj}.$$

To establish consistency of $\hat{E}_l$, we introduce some additional notation. Let $\Psi_l = R_l \tilde{\otimes} R_l$, where $\tilde{\otimes}$ is the Kronecker product, and let $\bar{E}_l = E_l \cup \{(1,1), \ldots, (p,p)\}$. For $B \subset V \times V$, let $\Psi_{l,BB}$ denote the submatrix of Ψ_l indexed by sets of pairs $(j, k) \in B$, where $\Psi_{l,(j,k),(j',k')} = R_{ljj'} R_{lkk'}$. For a $p \times p$ matrix Δ, let $\|\Delta\|_\infty = \max_{j=1,\ldots,p} \sum_{k=1}^p |\Delta_{jk}|$. The following assumption corresponds to the irrepresentability or neighbourhood stability condition often seen in sparse matrix and regression estimation (Meinshausen & Bühlmann, 2006; Ravikumar et al., 2011).

Assumption 3. For $l = 1, \ldots, L$, there exists $\eta_l \in (0, 1]$ such that

$$\left\| \Psi_{l,\bar{E}_l^c \bar{E}_l} \left(\Psi_{l,\bar{E}_l \bar{E}_l} \right)^{-1} \right\|_\infty \leqslant 1 - \eta_l.$$

For fixed l, Assumption 3 was employed by Ravikumar et al. (2011) as a sufficient condition for model selection consistency. As Theorem 5 below implies simultaneous consistency of the first L edge sets, we require the assumption for each l.

Set $\kappa_{R_l} = \|R_l\|_\infty$ and $\kappa_{\Psi_l} = \|(\Psi_{l,\bar{E}_l\bar{E}_l})^{-1}\|_\infty$, let $y_l = \max_{j\in V} |\{k \in V : \Xi_{ljk} \neq 0\}|$ be the maximum degree of the graph (V, E_l), and define $\xi_{\min,l} = \min\{|\Xi_{ljk}| : \Xi_{ljk} \neq 0\}$. Finally, when Assumptions 1–3 hold, for any $\alpha > 1 - \min_{l=1,\ldots,L} \eta_l$ define $\eta_l' = \alpha + \eta_l - 1 > 0$ and

$\epsilon_L = \min_{1 \leqslant l \leqslant L} \eta_l' m_l$. Then, with D_3 as in Corollary 2, set

$$\mathfrak{a}_L = D_3 \min_{l=1,\ldots,L} m_l,$$

$$\mathfrak{b}_L = \min_{l=1,\ldots,L} \left\{ 6 y_l m_l \max(\kappa_{\Psi_l}^2 \kappa_{R_l}^3, \kappa_{\Psi_l} \kappa_{R_l})(m_l^{-1} + 8\epsilon_L^{-1})^2 \right\}^{-1},$$

$$\mathfrak{c}_L = \min_{l=1,\ldots,L} \xi_{\min,l} \left\{ 4\kappa_{\Psi_l}(m_l^{-1} + 8\epsilon_L^{-1}) \right\}^{-1}.$$

Tracking $\mathfrak{a}_L$ and $\mathfrak{b}_L$, including the maximal degrees y_l, allows one to obtain uniform consistency of the estimates $\hat{\Xi}_l$ in (7), and to conclude that $\hat{E}_l \subset E_l$ with high probability; see the Supplementary Material. The quantity $\mathfrak{c}_L$ involves the weakest nonzero signal $\xi_{\min,l}$ of each graph, with weaker signals requiring larger n for recovery. The quantities $\mathfrak{b}_L$ and $\mathfrak{c}_L$ decrease with α, so that smaller values of α also require larger n.

THEOREM 5. *Suppose Assumptions 1–3 hold,* $(1 - \min_{l=1,\ldots,L} \eta_l) < \alpha \leqslant 1$, $L \leqslant np$ *and that for some* $\varrho > 2$ *one has* $\gamma = 8\epsilon_L^{-1}\{(D_1 n)^{-1} \log(D_2 L^{\varrho-1} p^{\varrho})\}^{1/2}$, *where* D_1 *and* D_2 *are constants from Corollary 2. If the sample size n satisfies*

$$n \min(\mathfrak{a}_L, \mathfrak{b}_L, \mathfrak{c}_L)^2 > D_1^{-1} \{\log(D_2) + (\varrho - 1)\log(n) + (2\varrho - 1)\log(p)\}, \tag{11}$$

then with probability at least $1 - (Lp)^{2-\varrho}$, $\hat{E}_l = E_l$ *for all* $l = 1, \ldots, L$.

Remark 3. If L grows sufficiently slowly with respect to n, (11) becomes $n \min(\mathfrak{a}_L, \mathfrak{b}_L, \mathfrak{c}_L)^2 \gtrsim \varrho \log(p)$ for large n, where $\gtrsim$ denotes inequality up to a multiplicative constant. Hence, the conclusion of Theorem 5 holds for large n when $\log(p) = o(n)$ by suitably limiting the growth of L.

Remark 4. To understand how the graph properties affect the lower bound, assume that κ_{Ψ_l}, κ_{R_l} and η_l do not depend on l, n or p, and that $\min_{1 \leqslant l \leqslant n} m_l \gtrsim n^{-d}$ for $0 < d < 1/4$. Then (11) becomes

$$n \gtrsim \left[\left\{ \left(\max_{1 \leqslant l \leqslant L} \xi_{\min,l}^{-2} \right) + \left(\max_{1 \leqslant l \leqslant L} y_l^2 \right) \right\} \varrho \log(p) \right]^{1-4d}$$

asymptotically. In particular, if L remains bounded so that $d = 0$, the above bound is consistent with that of Ravikumar et al. (2011), where the maxima over l reflect the need to satisfy the bound for the edge set E_l that is most difficult to estimate.

If the conditional independence graph of X is $E = \bigcup_{l=1}^{\infty} E_l$, as is the case when X is Gaussian and $\mathcal{G}$ partially separable, Theorem 5 can lead to edge selection consistency in the functional graphical model. For a given n, there exists a finite L_p^* such that $E = \bigcup_{l=1}^{L_p^*} E_l$, where L_p^* can only diverge with n if p does so. The following corollary is then immediate.

COROLLARY 3. *Under the assumptions of Theorem 5, suppose* $\log(p) = o(n)$, $L \to \infty$ *such that* $\min(\mathfrak{a}_L, \mathfrak{b}_L, \mathfrak{c}_L)[n/\log\{\max(n,p)\}]^{1/2} \to \infty$ *and that for large n one has* $L \geqslant L_p^*$. *Then for large n,* $E = \bigcup_{l=1}^{L} \hat{E}_l$ *with probability at least* $1 - (Lp)^{2-\varrho}$.

5. NUMERICAL EXPERIMENTS

5.1. *Simulation settings*

The simulations in this section compare the proposed method for partially separable functional Gaussian graphical models with that of Qiao et al. (2019). Throughout this section we denote these methods by psFGGM and FGGM, respectively. Other potentially competing non-functional-based approaches are not included, since they are clearly outperformed by the latter (Qiao et al., 2019). An initial conditional independence graph $G = (V, E)$ is generated from a power-law distribution with parameter $\pi = \text{pr}\{(j, k) \in E\}$. Then, for a fixed M, a sequence of edge sets $E_1, \ldots, E_M$ is generated so that $E = \bigcup_{l=1}^{M} E_l$. A set of common edges to all edge sets is computed for a given proportion of common edges $\tau \in [0, 1]$. Next, $p \times p$ precision matrices Ω_l are generated for each E_l based on the algorithm of Peng et al. (2009). A detailed description of this step is included in the Supplementary Material.

Random vectors $\theta_i \in \mathbb{R}^{Mp}$ are then generated from a zero-mean multivariate normal distribution with covariance matrix Σ, yielding discrete and noisy functional data

$$Y_{ijk} = X_{ij}(t_k) + \varepsilon_{ijk}, \quad \varepsilon_{ijk} \sim N(0, \sigma_\varepsilon^2) \quad (i = 1, \ldots, n;\ j = 1, \ldots, p;\ k = 1, \ldots, M).$$

Here, $\sigma_\epsilon^2 = 0.05 \sum_{l=1}^{M} \text{tr}(\Sigma_l)/p$ and $X_{ij}(t_k) = \sum_{l=1}^{M} \theta_{ilj}\varphi_l(t_k)$ according to the partially separable Karhunen–Loève expansion in (3). Fourier basis functions $\varphi_1, \ldots, \varphi_M$ evaluated on an equally spaced time grid $t_1, \ldots, t_T$, with $t_1 = 0$ and $t_T = 1$, were used to generate the data. In all settings, 100 simulations were conducted. To mimic a real-data example from § 6, we set $T = 30$, $M = 20$ and $\pi = 5\%$ for a sparse graph.

We consider two models for Σ, corresponding to partially separable and non-partially separable X. In the first, the covariance Σ_{ps} is formed as a block-diagonal matrix with $p \times p$ diagonal blocks $\Sigma_l = a_l \Omega_l^{-1}$. The decaying factors $a_l = 3l^{-1.8}$ guarantee that $\text{tr}(\Sigma_l)$ decreases monotonically in l. In the second model, Σ_{ps} is modified to violate partial separability. Specifically, a block-banded precision matrix Ω is computed with $p \times p$ blocks $\Omega_{l,l} = \Omega_l$ and $\Omega_{l+1,l} = \Omega_{l,l+1} = 0.5(\Omega_l^* + \Omega_{l+1}^*)$, where $\Omega_l^* = \Omega_l - \text{diag}(\Omega_l)$. Then, the non-partially separable covariance is computed as $\Sigma_{\text{non-ps}} = \text{diag}(\Sigma_{\text{ps}})^{1/2}\Omega^{-1}\text{diag}(\Sigma_{\text{ps}})^{1/2}$.

5.2. *Comparison of results*

Presented here are comparisons between the proposed method and that of Qiao et al. (2019), implemented using code provided by the authors. Additional comparisons obtained by thresholding correlations are reported in the Supplementary Material. Although this alternative method does not estimate a sparse inverse covariance structure, its graph recovery is competitive with that of the proposed method in some settings. As performance metrics, the true and false positive rates of correctly identifying edges in graph G are computed over a range of γ values and a coarse grid of five evenly spaced points $\alpha \in [0, 1]$. The value of α maximizing the area under the receiver operating characteristic curve is considered for the comparison. In all cases, we set $\pi = 0.05$ and $\tau = 0$. The two methods are compared using L principal components that explain at least 90% of the variance. For all simulations and both methods, this threshold results in a choice of $L = 5$ or $L = 6$ components. For higher variance-explained thresholds, however, we see a sharp contrast. While the proposed method consistently converges to a solution, the method of Qiao et al. (2019) does not, due to increasing numerical instability. The reason for the instability is the need to estimate $L = 5$ or 6 times more parameters than the proposed method. The proposed method can thus accommodate larger L, thereby incorporating more information from the data. In the figures

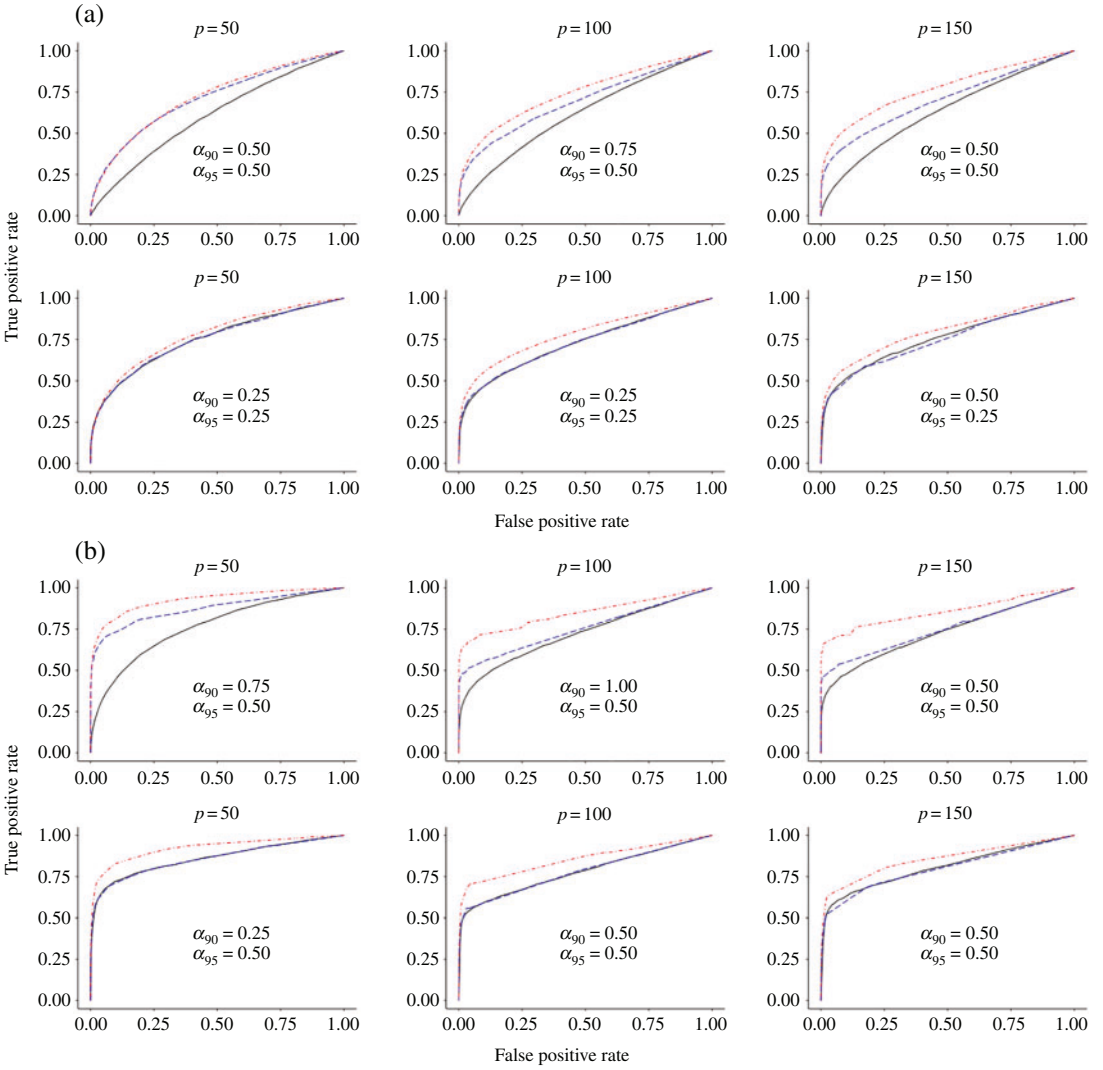

Fig. 2. Mean receiver operating characteristic curves for the proposed method, denoted by psFGGM, and that of Qiao et al. (2019), denoted by FGGM: (a) high-dimensional case of $n = p/2$; (b) large-sample case of $n = 1.5p$. In both (a) and (b), Σ_{ps} was used for the upper row and $\Sigma_{\text{non-ps}}$ for the lower row, with $p = 50, 100, 150$. Curves are coded as psFGGM (– – –) and FGGM (——) at 90% of variance and psFGGM (—·—) at 95% of variance explained. For psFGGM, the values of α used to compute the curve values are printed in each panel.

and tables, additional results are reported for the proposed method when L is increased to explain at least 95% of the variance.

Figure 2(a) shows average true versus false positive rate curves for the high-dimensional case of $n = p/2$. The smoothed curves are computed using the `R` package `supsmu`, which implements SuperSmoother (Friedman, 1984), a variable-bandwidth smoother that uses cross-validation to find the best bandwidth. Table 1 shows the mean and standard deviation of area-under-the-curve estimates for various settings. When partial separability holds, i.e., $\Sigma = \Sigma_{\text{ps}}$, the proposed method exhibits uniformly higher true positive rates across the full range of false positive rates. Even when partial separability is violated, i.e., $\Sigma = \Sigma_{\text{non-ps}}$, the two methods perform comparably. More importantly, and in all cases, the proposed method is able to leverage a 95% level of variance explained, owing to the numerical stability mentioned above. Figure 2(b) and Table 1

Table 1. *Values of mean area under the receiver operating characteristic curve (with standard errors in parentheses) for Fig. 2*

			$\Sigma = \Sigma_{\text{ps}}$			$\Sigma = \Sigma_{\text{non-ps}}$		
n			$p = 50$	$p = 100$	$p = 150$	$p = 50$	$p = 100$	$p = 150$
$p/2$	AUC	$\text{FGGM}_{90\%}$	0.60 (0.03)	0.62 (0.02)	0.63 (0.01)	0.75 (0.03)	0.72 (0.02)	0.75 (0.02)
		$\text{psFGGM}_{90\%}$	0.71 (0.04)	0.69 (0.02)	0.70 (0.01)	0.75 (0.03)	0.73 (0.02)	0.74 (0.03)
		$\text{psFGGM}_{95\%}$	0.72 (0.04)	0.74 (0.02)	0.77 (0.02)	0.77 (0.03)	0.78 (0.02)	0.79 (0.02)
	AUC15	$\text{FGGM}_{90\%}$	0.15 (0.04)	0.18 (0.02)	0.20 (0.01)	0.39 (0.04)	0.40 (0.02)	0.45 (0.03)
		$\text{psFGGM}_{90\%}$	0.30 (0.05)	0.35 (0.02)	0.37 (0.02)	0.39 (0.04)	0.42 (0.03)	0.44 (0.04)
		$\text{psFGGM}_{95\%}$	0.29 (0.05)	0.40 (0.03)	0.46 (0.03)	0.41 (0.05)	0.48 (0.03)	0.51 (0.03)
$1.5p$	AUC	$\text{FGGM}_{90\%}$	0.76 (0.02)	0.72 (0.02)	0.73 (0.01)	0.86 (0.02)	0.78 (0.02)	0.80 (0.03)
		$\text{psFGGM}_{90\%}$	0.87 (0.03)	0.75 (0.02)	0.75 (0.01)	0.85 (0.02)	0.78 (0.02)	0.79 (0.03)
		$\text{psFGGM}_{95\%}$	0.92 (0.02)	0.84 (0.02)	0.85 (0.02)	0.92 (0.03)	0.85 (0.02)	0.85 (0.02)
	AUC15	$\text{FGGM}_{90\%}$	0.37 (0.04)	0.41 (0.02)	0.44 (0.02)	0.66 (0.03)	0.55 (0.03)	0.57 (0.04)
		$\text{psFGGM}_{90\%}$	0.69 (0.04)	0.52 (0.02)	0.52 (0.02)	0.65 (0.04)	0.56 (0.04)	0.55 (0.05)
		$\text{psFGGM}_{95\%}$	0.75 (0.04)	0.68 (0.03)	0.69 (0.03)	0.76 (0.06)	0.68 (0.03)	0.64 (0.03)

AUC, area under the receiver operating characteristic curve; AUC15, AUC computed for false positive rates in the interval $[0, 0.15]$, normalized to have maximum area 1.

summarize results for the large-sample case $n = 1.5p$, with similar conclusions. Comparisons under additional simulation settings can be found in the Supplementary Material.

6. APPLICATION TO FUNCTIONAL BRAIN CONNECTIVITY

In this section, the proposed method is used to reconstruct the connectivity structure of the brain using fMRI data from the Human Connectome Project. We analyse the ICA-FIX pre-processed data variant that controls for spatial distortions and alignments across both subjects and modalities (Glasser et al., 2013). In particular, we use the 1200 Subjects 3T MR imaging data available at `https://db.humanconnectome.org`, which consists of fMRI scans of individuals performing basic body movements. During each scan, a three-second visual cue signals the subject to move a specific body part, which is then recorded for 12 seconds at a temporal resolution of 0.72 seconds. For this work, we consider only the data from left- and right-hand finger movements.

The left- and right-hand tasks data for $n = 1054$ subjects with complete metadata were pre-processed by averaging the blood oxygen level-dependent signals over $p = 360$ regions of interest, ROIs (Glasser et al., 2016). After removing cool-down and ramp-up observations, $T = 16$ time-points of pure movement tasks remained. Motivated by Theorem 1(iii), diagnostics were performed to assess the plausibility of the partial separability assumption, with no indications to the contrary; see the Supplementary Material for more details. Penalty parameters $\gamma = 0.91$ and $\alpha = 0.95$ were used to estimate very sparse graphs in both tasks.

Figure 3 compares activation patterns from left- and right-hand task datasets. Panels (a) and (b) show the recovered ROI graph on a flat brain map, and only those ROIs with positive degree of connectivity are coloured. Panels (c) and (d) show connected ROIs that are unique to each task, whereas (e) shows only those that are common to both tasks. In this map, one can see that almost all of the visual cortex ROIs in the occipital lobe are shared by both maps. This is expected as both tasks require individuals to watch visual cues. Furthermore, the primary sensory cortex, corresponding to touch and motor sensory inputs, and the intraparietal sulcus, corresponding to

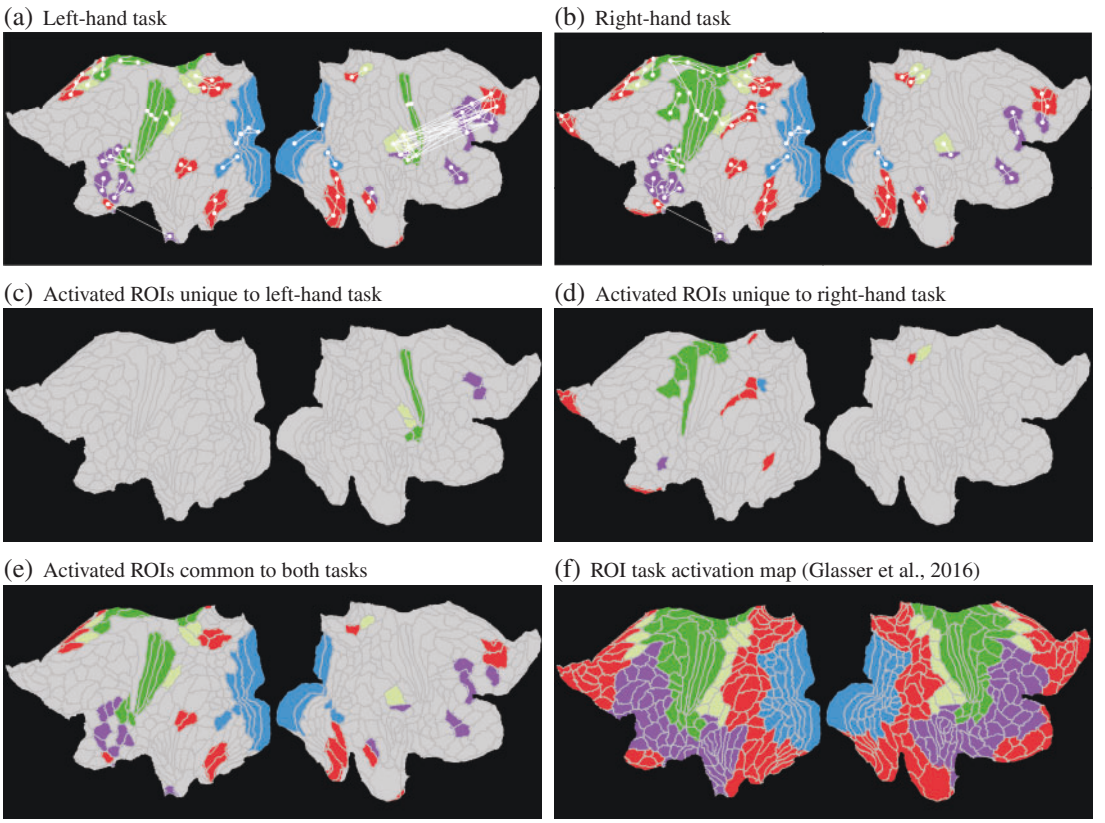

Fig. 3. psFGGM-estimated functionally connected cortical ROIs for the left- and right-hand motor tasks. Each panel shows a flat brain map of the left and right hemispheres, in that order. ROIs having a positive degree of connectivity in each estimated graph are coloured based on their functionality (Glasser et al., 2016): visual (blue), motor (green), mixed motor (light green), mixed other (red) and other (purple).

perceptual motor coordination, are activated during both left- and right-hand tasks. On the other hand, the main difference between these motor tasks lies in the motor cortex near the central sulcus. In Fig. 3 panels (c) and (d) the functional maps for the left- and right-hand tasks present particular motor-related cortical areas in the right and left hemispheres, respectively. These results are in line with the motor task activation maps obtained by Barch et al. (2013).

7. Discussion

Partial separability for multivariate functional data is a novel structural assumption with further potential applications beyond graphical models. For example, it is well known that the functional linear model is simplified by univariate functional principal component analysis, which parses out the problem into a sequence of simple linear regressions; see Wang et al. (2016) and references therein. The partially separable Karhunen–Loève expansion in (3) demonstrates a similar potential, namely to break down a problem involving multivariate functional data into a sequence of standard multivariate problems. This potential was shown in the present paper by decomposing a functional graphical model into a sequence of standard multivariate graphical models.

Motivated by the brain connectivity example, we have presented partial separability for multivariate processes with components X_j defined on the same domain. However, this restriction is not

necessary for defining partial separability. If the X_j are elements of $L^2(\mathcal{T}_j)$ $(j = 1, \ldots, p)$, a more general definition of partial separability would be the existence of orthonormal bases $\{\varphi_{jl}\}_{l=1}^{\infty}$ of $L^2(\mathcal{T}_j)$ $(j = 1, \ldots, p)$, such that the vectors $\theta_l = (\theta_{l1}, \ldots, \theta_{lp})^{\mathrm{T}}$, where $\theta_{lj} = \int_{\mathcal{T}_j} X_j(t)\varphi_{jl}(t)\,\mathrm{d}t$, are mutually uncorrelated across l. Such a generalization is highly desirable, as many multivariate functional datasets consist of functions on different domains. In fact, the above notion is applicable even when the domains $\mathcal{T}_j$ are of different dimensions (Happ & Greven, 2018), and even to a complex manifold such as the surface of the brain.

The proposed method for functional graphical model estimation is equally applicable to dense or sparse functional data, observed with or without noise. However, rates of convergence will inevitably suffer, as observations become more sparse or are contaminated with higher levels of noise. The results in Theorem 4 of this paper or Theorem 1 of Qiao et al. (2019) have been derived in the setting of fully observed functional data, so future work will include similar derivations under more general observation schemes.

ACKNOWLEDGEMENT

The authors thank Scott Grafton for an interesting discussion of the neuroscience data analysis results, as well as two referees and an associate editor for their helpful suggestions. Oh is also affiliated with the Lawrence Berkeley Lab, Berkeley, California, and Petersen with the Department of Statistics at Brigham Young University, Utah.

SUPPLEMENTARY MATERIAL

Supplementary Material includes proofs of all theoretical results, further details on simulated data generation, additional simulation results, and illustrations of diagnostic procedures for assessing partial separability in the fMRI data of § 6.

REFERENCES

ASTON, J. A., PIGOLI, D. & TAVAKOLI, S. (2017). Tests for separability in nonparametric covariance operators of random surfaces. *Ann. Statist.* **45**, 1431–61.

BANERJEE, S. & JOHNSON, G. A. (2006). Coregionalized single- and multiresolution spatially varying growth curve modeling with application to weed growth. *Biometrics* **62**, 864–76.

BARCH, D. M., BURGESS, G. C., HARMS, M. P., PETERSEN, S. E., SCHLAGGAR, B. L., CORBETTA, M., GLASSER, M. F., CURTISS, S., DIXIT, S., FELDT, C. et al. (2013). Function in the human connectome: Task-fMRI and individual differences in behavior. *NeuroImage* **80**, 169–89.

BHATIA, R., DAVIS, C. & MCINTOSH, A. (1983). Perturbation of spectral subspaces and solution of linear operator equations. *Lin. Algeb. Applic.* **52**, 45–67.

BOSQ, D. (2000). *Linear Processes in Function Spaces: Theory and Applications*. New York: Springer.

CHIOU, J.-M., CHEN, Y.-T. & YANG, Y.-F. (2014). Multivariate functional principal component analysis: A normalization approach. *Statist. Sinica* **24**, 1571–96.

CHIOU, J.-M. & MÜLLER, H.-G. (2014). Linear manifold modelling of multivariate functional data. *J. R. Statist. Soc.* B **76**, 605–26.

CHIOU, J.-M., YANG, Y.-F. & CHEN, Y.-T. (2016). Multivariate functional linear regression and prediction. *J. Mult. Anal.* **146**, 301–12.

DANAHER, P., WANG, P. & WITTEN, D. M. (2014). The joint graphical lasso for inverse covariance estimation across multiple classes. *J. R. Statist. Soc.* B **76**, 373–97.

DUBIN, J. A. & MÜLLER, H.-G. (2005). Dynamical correlation for multivariate longitudinal data. *J. Am. Statist. Assoc.* **100**, 872–81.

FRIEDMAN, J. H. (1984). A variable span scatterplot smoother. Technical Report No. 5, Laboratory for Computational Statistics, Stanford University.

Genton, M. G. (2007). Separable approximations of space-time covariance matrices. *Environmetrics* **18**, 681–95.
Glasser, M. F., Coalson, T. S., Robinson, E. C., Hacker, C. D., Harwell, J., Yacoub, E., Ugurbil, K., Andersson, J., Beckmann, C. F., Jenkinson, M. et al. (2016). A multi-modal parcellation of human cerebral cortex. *Nature* **536**, 171–8.
Glasser, M. F., Sotiropoulos, S. N., Wilson, J. A., Coalson, T. S., Fischl, B., Andersson, J. L., Xu, J., Jbabdi, S., Webster, M., Polimeni, J. R. et al. (2013). The minimal preprocessing pipelines for the Human Connectome Project. *NeuroImage* **80**, 105–24.
Gneiting, T., Genton, M. G. & Guttorp, P. (2006). *Geostatistical Space-Time Models, Stationarity, Separability, and Full Symmetry*, vol. 107 of *Monographs on Statistics and Applied Probability*. Boca Raton, Florida: Chapman & Hall/CRC.
Happ, C. & Greven, S. (2018). Multivariate functional principal component analysis for data observed on different (dimensional) domains. *J. Am. Statist. Assoc.* **113**, 649–59.
Hsing, T. & Eubank, R. (2015). *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*. Chichester: John Wiley & Sons.
Jirak, M. (2016). Optimal eigen expansions and uniform bounds. *Prob. Theory Rel. Fields* **166**, 753–99.
Kolar, M. & Xing, E. (2011). On time varying undirected graphs. *Proc. Mach. Learn. Res.* **15**, 407–15. Proc. 14th Int. Conf. Artificial Intelligence and Statistics.
Lauritzen, S. L. (1996). *Graphical Models*, vol. 17 of *Oxford Statistical Science Series*. Oxford: Clarendon Press.
Li, B. & Solea, E. (2018). A nonparametric graphical model for functional data with application to brain networks based on fMRI. *J. Am. Statist. Assoc.* **113**, 1637–55.
Lynch, B. & Chen, K. (2018). A test of weak separability for multi-way functional data, with application to brain connectivity studies. *Biometrika* **105**, 815–31.
Meinshausen, N. & Bühlmann, P. (2006). High-dimensional graphs and variable selection with the lasso. *Ann. Statist.* **34**, 1436–62.
Peng, J., Wang, P., Zhou, N. & Zhu, J. (2009). Partial correlation estimation by joint sparse regression models. *J. Am. Statist. Assoc.* **104**, 735–46.
Petersen, A. & Müller, H.-G. (2016). Fréchet integration and adaptive metric selection for interpretable covariances of multivariate functional data. *Biometrika* **103**, 103–20.
Qiao, X., Guo, S. & James, G. M. (2019). Functional graphical models. *J. Am. Statist. Assoc.* **114**, 211–22.
Qiao, X., Qian, C., James, G. M. & Guo, S. (2020). Doubly functional graphical models in high dimensions. *Biometrika* **107**, 415–31.
Qiu, H., Han, F., Liu, H. & Caffo, B. (2016). Joint estimation of multiple graphical models from high dimensional time series. *J. R. Statist. Soc.* B **78**, 487–504.
R Development Core Team (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. http://www.R-project.org.
Ravikumar, P., Wainwright, M. J., Raskutti, G. & Yu, B. (2011). High-dimensional covariance estimation by minimizing ℓ_1-penalized log-determinant divergence. *Electron. J. Statist.* **5**, 935–80.
Wang, J.-L., Chiou, J.-M. & Müller, H.-G. (2016). Functional data analysis. *Annu. Rev. Statist. Appl.* **3**, 257–95.
Yang, W., Müller, H.-G. & Stadtmüller, U. (2011). Functional singular component analysis. *J. R. Statist. Soc.* B **73**, 303–24.
Yao, F., Müller, H.-G. & Wang, J.-L. (2005). Functional data analysis for sparse longitudinal data. *J. Am. Statist. Assoc.* **100**, 577–90.
Zhou, S., Lafferty, J. & Wasserman, L. (2010). Time varying undirected graphs. *Mach. Learn.* **80**, 295–319.
Zhu, H., Strawn, N. & Dunson, D. B. (2016). Bayesian graphical models for multivariate functional data. *J. Mach. Learn. Res.* **17**, 1–27.

[*Received on* 25 *March* 2020. *Editorial decision on* 5 *August* 2021]