Adaptive Functional Thresholding for Sparse Covariance Function Estimation in High Dimensions

Qin Fang, Shaojun Guo & Xinghao Qiao

View supplementary material

Published online: 26 May 2023.

Submit your article to this journal

Article views: 271

View related articles

View Crossmark data

Taylor & Francis
Taylor & Francis Group

🔓 OPEN ACCESS

Check for updates

Adaptive Functional Thresholding for Sparse Covariance Function Estimation in High Dimensions

Qin Fang[a], Shaojun Guo[b], and Xinghao Qiao[a]

[a]Department of Statistics, London School of Economics and Political Science, London, UK; [b]Institute of Statistics and Big Data, Renmin University of China, Beijing, P.R. China

ABSTRACT

Covariance function estimation is a fundamental task in multivariate functional data analysis and arises in many applications. In this article, we consider estimating sparse covariance functions for high-dimensional functional data, where the number of random functions p is comparable to, or even larger than the sample size n. Aided by the Hilbert–Schmidt norm of functions, we introduce a new class of functional thresholding operators that combine functional versions of thresholding and shrinkage, and propose the adaptive functional thresholding estimator by incorporating the variance effects of individual entries of the sample covariance function into functional thresholding. To handle the practical scenario where curves are partially observed with errors, we also develop a nonparametric smoothing approach to obtain the smoothed adaptive functional thresholding estimator and its binned implementation to accelerate the computation. We investigate the theoretical properties of our proposals when p grows exponentially with n under both fully and partially observed functional scenarios. Finally, we demonstrate that the proposed adaptive functional thresholding estimators significantly outperform the competitors through extensive simulations and the functional connectivity analysis of two neuroimaging datasets. Supplementary materials for this article are available online.

1. Introduction

The covariance function estimation plays an important role in functional data analysis, while existing methods are restricted to data with a single or small number of random functions. Recent advances in technology have made multivariate or even high-dimensional functional datasets increasingly common in various applications: for example, time-course gene expression data in genomics (Storey et al. 2005), air pollution data in environmental studies (Kong et al. 2016) and different types of brain imaging data in neuroscience (Li and Solea 2018; Qiao, Guo, and James 2019). Under such scenarios, suppose we observe n independent samples $\mathbf{X}_i(\cdot) = \{X_{i1}(\cdot), \ldots, X_{ip}(\cdot)\}^{\mathrm{T}}$ for $i = 1, \ldots, n$ defined on a compact interval $\mathcal{U}$ with covariance function

$$\boldsymbol{\Sigma}(u, v) = \{\Sigma_{jk}(u, v)\}_{p \times p} = \mathrm{cov}\{\mathbf{X}_i(u), \mathbf{X}_i(v)\}, \quad u, v \in \mathcal{U}.$$

From a heuristic interpretation, we can simply treat each curve $X_{ij}(\cdot)$ as an infinitely long vector and replace the (j, k)th entry of $\boldsymbol{\Sigma}$ by $\Sigma_{jk}(\cdot, \cdot) = \mathrm{cov}\{X_{ij}(\cdot), X_{ik}(\cdot)\}$, the cross-covariance matrix of two infinitely long vectors. Then $\boldsymbol{\Sigma}$ can be understood as a block matrix with infinite sizes and its (j, k)th block being $\Sigma_{jk}(\cdot, \cdot)$. Besides being of interest in itself, an estimator of $\boldsymbol{\Sigma}$ is useful for many applications including, for example, multivariate functional principal components analysis (FPCA) (Happ and Greven 2018), multivariate functional linear regression (Chiou, Yang, and Chen 2016), functional factor model (Guo, Qiao, and

Wang 2022) and functional classification (Park, Ahn, and Jeon 2021). See Section 2.3 for details.

Our article focuses on estimating $\boldsymbol{\Sigma}$ under high-dimensional scaling, where p can be comparable to, or even larger than n. In this setting, the sample covariance function

$$\widehat{\boldsymbol{\Sigma}}(u, v) = \{\widehat{\Sigma}_{jk}(u, v)\}_{p \times p}$$
$$= \frac{1}{n-1} \sum_{i=1}^{n} \{\mathbf{X}_i(u) - \bar{\mathbf{X}}(u)\}\{\mathbf{X}_i(v) - \bar{\mathbf{X}}(v)\}^{\mathrm{T}}, \quad u, v \in \mathcal{U},$$

where $\bar{\mathbf{X}}(\cdot) = n^{-1} \sum_{i=1}^{n} \mathbf{X}_i(\cdot)$, performs poorly, and some lower-dimensional structural assumptions need to be imposed to estimate $\boldsymbol{\Sigma}$ consistently. In contrast to extensive work on estimating high-dimensional sparse covariance matrices (Bickel and Levina 2008; Rothman, Levina, and Zhu 2009; Cai and Liu 2011; Chen and Leng 2016; Avella-Medina et al. 2018; Wang et al. 2021), research on sparse covariance function estimation in high dimensions remains largely unaddressed in the literature.

In this article, we consider estimating sparse covariance functions via adaptive functional thresholding in the sense of shrinking some blocks $\widehat{\Sigma}_{jk}(\cdot, \cdot)$'s in an adaptive way. To achieve this, we introduce a new class of functional thresholding operators that combine functional versions of thresholding and shrinkage based on the Hilbert-Schmidt norm of functions, and develop an adaptive functional thresholding procedure on $\widehat{\boldsymbol{\Sigma}}(\cdot, \cdot)$ using

CONTACT Shaojun Guo ✉ sjguo@ruc.edu.cn 🏛 Institute of Statistics and Big Data, Renmin University of China, Beijing, P.R. China; Xinghao Qiao ✉ X.Qiao@lse.ac.uk 🏛 Department of Statistics, London School of Economics and Political Science, London, UK.

Supplementary materials for this article are available online. Please go to www.tandfonline.com/r/JASA.

entry-dependent functional thresholds that automatically adapt to the variability of blocks $\widehat{\Sigma}_{jk}(\cdot, \cdot)$'s. To provide theoretical guarantees of our method under high-dimensional scaling, it is essential to develop standardized concentration results taking into account the variability adjustment. Compared with adaptive thresholding for nonfunctional data (Cai and Liu 2011), the intrinsic infinite-dimensionality of each $X_{ij}(\cdot)$ leads to a substantial rise in the complexity of sparsity modeling and theoretical analysis, as one needs to rely on some functional norm of standardized $\widehat{\Sigma}_{jk}$'s, for example, the Hilbert–Schmidt norm, to enforce the functional sparsity in $\widehat{\Sigma}$ and tackle more technical challenges for standardized processes within an abstract Hilbert space. To handle the practical scenario where functions are partially observed with errors, it is desirable to apply nonparametric smoothers in conjunction with adaptive functional thresholding. This poses a computationally intensive task especially when p is large, thus calling for the development of fast implementation strategy.

There are many applications of the proposed sparse covariance function estimation method in neuroimaging analysis, where brain signals are measured over time at a large number of regions of interest (ROIs) for individuals. Examples include the brain-computer interface classification (Lotte et al. 2018) and the brain functional connectivity identification (Rogers et al. 2007). Traditional neuroimaging analysis models brain signals for each subject as multivariate random variables, where each ROI is represented by a random variable, and hence the covariance/correlation matrices of interest are estimated by treating the time-course data of each ROI as repeated observations. However, due to the nonstationary and dynamic features of signals (Chang and Glover 2010), the strategy of averaging over time fails to characterize the time-varying structure leading to the loss of information in the original space. To overcome these drawbacks, we follow recent proposals to model signals directly as multivariate random functions with each ROI represented by a random function (Li and Solea 2018; Qiao, Guo, and James 2019; Zapata, Oh, and Petersen 2022; Lee et al. in press). The identified functional sparsity pattern in our estimate of Σ can be used to recover the functional connectivity network among different ROIs, which is illustrated using examples of functional magnetic resonance imaging (fMRI) datasets in Section 6 and Section E.3 of the supplementary material.

Our article makes useful contributions at multiple fronts. On the method side, it generalizes the thresholding/sparsity concept in multivariate statistics to the functional setting and offers a novel adaptive functional thresholding proposal to handle the heteroscedastic problem of the sparse covariance function estimation motivated from neuroimaging analysis and many statistical applications, for example, those in Section 2.3 and Section C.2 of the supplementary material. It also provides an alternative way of identifying correlation-based functional connectivity with no need to specify the correlation function, the estimation of which poses challenges as the inverses of $\Sigma_{jj}(u, v)$'s are unbounded. In practice when functions are observed with errors at either a dense grid of points or a small subset of points, we also develop a unified local linear smoothing approach to obtain the smoothed adaptive functional thresholding estimator and its fast implementation via binning (Fan and Marron 1994) to speed

up the computation without sacrificing the estimation accuracy. On the theory side, we show that the proposed estimators enjoy the convergence and support recovery properties under both fully and partially observed functional scenarios when p grows exponentially fast relative to n. The proof relies on tools from empirical process theory due to the infinite-dimensional nature of functional data and some novel standardized concentration bounds in the Hilbert–Schmidt norm to deal with issues of high-dimensionality and variance adjustment. Our theoretical results and adopted techniques are general, and can be applied to other settings in high-dimensional functional data analysis.

The remainder of this article is organized as follows. Section 2 introduces a class of functional thresholding operators, based on which we propose the adaptive functional thresholding of the sample covariance function. We then discuss a couple of applications of the sparse covariance function estimation. Section 3 presents convergence and support recovery analysis of our proposed estimator. In Section 4, we develop a nonparametric smoothing approach and its binned implementation to deal with partially observed functional data, and then investigate its theoretical properties. In Sections 5 and 6, we demonstrate the uniform superiority of the adaptive functional thresholding estimators over the universal counterparts through an extensive set of simulation studies and the functional connectivity analysis of a neuroimaging dataset, respectively. All technical proofs are relegated to the supplementary material. We also provide the codes to reproduce the results for simulations and real data analysis in supplementary materials.

2. Methodology

2.1. Functional Thresholding

We begin by introducing some notation. Let $L_2(\mathcal{U})$ denotes a Hilbert space of square integrable functions defined on $\mathcal{U}$ and $\mathbb{S} = L_2(\mathcal{U}) \otimes L_2(\mathcal{U})$, where $\otimes$ is the Kronecker product. For any $Q \in \mathbb{S}$, we denote its Hilbert–Schmidt norm by $\|Q\|_{\mathcal{S}} = \{\int \int Q(u, v)^2 dudv\}^{1/2}$. With the aid of Hilbert–Schmidt norm, for any regularization parameter $\lambda \geq 0$, we first define a class of functional thresholding operators $s_\lambda : \mathbb{S} \to \mathbb{S}$ that satisfy the following conditions:

1. $\|s_\lambda(Z)\|_{\mathcal{S}} \leq c\|Y\|_{\mathcal{S}}$ for all Z and $Y \in \mathbb{S}$ that satisfy $\|Z - Y\|_{\mathcal{S}} \leq \lambda$ and some $c > 0$;
2. $\|s_\lambda(Z)\|_{\mathcal{S}} = 0$ for $\|Z\|_{\mathcal{S}} \leq \lambda$;
3. $\|s_\lambda(Z) - Z\|_{\mathcal{S}} \leq \lambda$ for all $Z \in \mathbb{S}$.

Our proposed functional thresholding operators can be viewed as the functional generalization of thresholding operators (Cai and Liu 2011). Instead of a simple pointwise extension of such thresholding operators under functional domain, we advocate a global thresholding rule based on the Hilbert–Schmidt norm of functions that encourages the functional sparsity, in the sense that $s_\lambda(Z)(u, v) = 0$, for all $u, v \in \mathcal{U}$, if $\|Z\|_{\mathcal{S}} \leq \lambda$ under condition (ii). Condition (iii) limits the amount of (global) functional shrinkage in the Hilbert–Schmidt norm to be no more than λ.

Conditions (i)–(iii) are satisfied by functional versions of some commonly adopted thresholding rules, which are intro-

duced as solutions to the following penalized quadratic loss problem with various penalties:

$$s_\lambda(Z) = \arg\min_{\theta \in \mathbb{S}} \left\{ \frac{1}{2} \|\theta - Z\|_{\mathcal{S}}^2 + p_\lambda(\theta) \right\} \qquad (1)$$

with $p_\lambda(\theta) = \tilde{p}_\lambda(\|\theta\|_{\mathcal{S}})$ being a penalty function of $\|\theta\|_{\mathcal{S}}$ to enforce the functional sparsity.

The soft functional thresholding rule results from solving (1) with an ℓ_1/ℓ_2 type of penalty, $p_\lambda(\theta) = \lambda\|\theta\|_{\mathcal{S}}$, and takes the form of $s_\lambda^{\text{s}}(Z) = Z(1 - \lambda/\|Z\|_{\mathcal{S}})_+$, where $(x)_+ = \max(x, 0)$ for $x \in \mathbb{R}$. This rule can be viewed as a functional generalization of the group lasso solution under the multivariate setting (Yuan and Lin 2006). To solve (1) with an ℓ_0/ℓ_2 type of penalty, $p_\lambda(\theta) = 2^{-1}\lambda^2 I(\|\theta\|_{\mathcal{S}} \neq 0)$, we obtain hard functional threhsolding rule as $ZI(\|Z\|_{\mathcal{S}} \geq \lambda)$, where $I(\cdot)$ is an indicator function. As a comparison, soft functional thresholding corresponds to the maximum amount of functional shrinkage allowed by condition (iii), whereas no shrinkage results from hard functional thresholding. Taking the compromise between soft and hard functional thresholding, we next propose functional versions of SCAD (Fan and Li 2001) and adaptive lasso (Zou 2006) thresholding rules. With a SCAD penalty (Fan and Li 2001) operating on $\|\cdot\|_{\mathcal{S}}$ instead of $|\cdot|$ for the univariate scalar case, SCAD functional thresholding $s_\lambda^{\text{SC}}(Z)$ is the same as soft functional thresholding if $\|Z\|_{\mathcal{S}} < 2\lambda$, and equals $Z\{(a-1) - a\lambda/\|Z\|_{\mathcal{S}}\}/(a-2)$ for $\|Z\|_{\mathcal{S}} \in [2\lambda, a\lambda]$ and Z if $\|Z\|_{\mathcal{S}} > a\lambda$, where $a > 2$. Analogously, adaptive lasso functional thresholding rule is $s_\lambda^{\text{AL}}(Z) = Z(1 - \lambda^{\eta+1}/\|Z\|_{\mathcal{S}}^{\eta+1})_+$ with $\eta \geq 0$.

Our proposed functional generalizations of soft, SCAD and adaptive lasso thresholding rules can be checked to satisfy conditions (i)–(iii), see Section B.1 of the supplementary material for details. To present a unified theoretical analysis, we focus on functional thresholding operators $s_\lambda(Z)$ satisfying conditions (i)–(iii). Note that, although the hard functional thresholding does not satisfy condition (i), theoretical results in Section 3 still hold for hard functional thresholding estimators under similar conditions with corresponding proofs differing slightly. For examples of functional data with some local spikes, one may possibly suggest supremum-norm-based class of functional thresholding operators. See the detailed discussion in Section C.1 of the supplementary material.

2.2. Estimation

We now discuss our estimation procedure based on $s_\lambda(Z)$. Note the variance of $\widehat{\Sigma}_{jk}(u, v)$ depends on the distribution of $\{X_{ij}(u), X_{ik}(v)\}$ through higher-order moments, which is intrinsically a heteroscedastic problem. Hence, it is more desirable to use entry-dependent functional thresholds that automatically takes into account the variability of blocks $\widehat{\Sigma}_{jk}(\cdot, \cdot)$'s to shrink some blocks to zero adaptively. To achieve this, define the variance factors $\Theta_{jk}(u, v) = \text{var}([X_{ij}(u) - \mathbb{E}\{X_{ij}(u)\}][X_{ik}(v) - \mathbb{E}\{X_{ik}(v)\}])$ with corresponding estimators

$$\widehat{\Theta}_{jk}(u, v) = \frac{1}{n}\sum_{i=1}^n \left[\{X_{ij}(u) - \bar{X}_j(u)\}\{X_{ik}(v) - \bar{X}_k(v)\} - \widehat{\Sigma}_{jk}(u, v)\right]^2,$$

$$j, k = 1, \ldots, p.$$

Then the adaptive functional thresholding estimator $\widehat{\mathbf{\Sigma}}_{\text{A}} = \{\widehat{\Sigma}_{jk}^{\text{A}}(\cdot, \cdot)\}_{p \times p}$ is defined by

$$\widehat{\Sigma}_{jk}^{\text{A}} = \widehat{\Theta}_{jk}^{1/2} \times s_\lambda\left(\frac{\widehat{\Sigma}_{jk}}{\widehat{\Theta}_{jk}^{1/2}}\right), \qquad (2)$$

which uses a single threshold level to functionally threshold standardized entries, $\widehat{\Sigma}_{jk}/\widehat{\Theta}_{jk}^{1/2}$ for all j, k, resulting in entry-dependent functional thresholds for $\widehat{\Sigma}_{jk}$'s. The selection of the optimal regularization parameter $\hat{\lambda}$ is discussed in Section 5.

An alternative approach to estimate $\mathbf{\Sigma}$ is the universal functional thresholding estimator

$$\widehat{\mathbf{\Sigma}}_{\text{U}} = \{\widehat{\Sigma}_{jk}^{\text{U}}(\cdot, \cdot)\}_{p \times p} \text{ with } \widehat{\Sigma}_{jk}^{\text{U}} = s_\lambda(\widehat{\Sigma}_{jk}),$$

where a universal threshold level is used for all entries. In a similar spirit to Rothman, Levina, and Zhu (2009), the consistency of $\widehat{\mathbf{\Sigma}}_{\text{U}}$ requires the assumption that marginal-covariance functions are uniformly bounded in nuclear norm, that is, $\max_j \|\Sigma_{jj}\|_{\mathcal{N}} \leq M$, where $\|\Sigma_{jj}\|_{\mathcal{N}} = \int_{\mathcal{U}} \Sigma_{jj}(u, u) du$. However, intuitively, such universal method does not perform well when nuclear norms vary over a wide range, or even fails when the uniform boundedness assumption is violated. Section 5 provides some empirical evidence to support this intuition.

2.3. Applications

Many statistical problems involving multivariate functional data $\{\mathbf{X}_i(\cdot)\}_{i=1}^n$ require estimating the covariance function $\mathbf{\Sigma}$. Under a high-dimensional regime, the functional sparsity assumption can be imposed on $\mathbf{\Sigma}$ to facilitate its consistent sparse estimates. Here we outline three applications of our proposals for the sparse covariance function estimation.

Our first application is *multivariate FPCA* serving as a natural dimension reduction approach for $\mathbf{X}_i(\cdot)$. With the aid of Karhunen-Loève expansion for multivariate functional data (Happ and Greven 2018), $\mathbf{X}_i(\cdot)$ admits the following expansion

$$\mathbf{X}_i(\cdot) = \mathbb{E}\{\mathbf{X}_i(\cdot)\} + \sum_{l=1}^\infty \xi_{il}\boldsymbol{\phi}_l(\cdot), \quad i = 1, \ldots, n, \qquad (3)$$

where the principal component scores $\xi_{il} = \sum_{j=1}^p \int [X_{ij}(u) - \mathbb{E}\{X_{ij}(u)\}]\phi_{lj}(u) du$ and eigenfunctions $\boldsymbol{\phi}_l(\cdot) = \{\phi_{l1}(\cdot), \ldots, \phi_{lp}(\cdot)\}^{\text{T}}$ are attainable by the eigenanalysis of $\mathbf{\Sigma}$. Under a large p scenario, we can adopt the proposed functional thresholding technique to obtain the sparse estimation of $\mathbf{\Sigma}$, which guarantees the consistencies of estimated eigenvalues/eigenfunctions pairs. In Section E.1 of the supplementary material, we follow the proposal of a normalized version of multivariate FPCA in Happ and Greven (2018) and use a simulated example to illustrate the superior sample performance of our functional thresholding approaches.

Our second application, *multivariate functional linear regression* (Chiou, Yang, and Chen 2016), takes the form of

$$Y_i = \beta_0 + \int_{\mathcal{U}} \mathbf{X}_i(u)^{\text{T}}\boldsymbol{\beta}(u) du + \epsilon_i, \quad i = 1, \ldots, n, \qquad (4)$$

where $\boldsymbol{\beta}(\cdot) = \{\beta_1(\cdot), \ldots, \beta_p(\cdot)\}^{\text{T}}$ is p-vector of functional coefficients to be estimated. The standard three-step procedure

involves performing (normalized) multivariate FPCA on $\mathbf{X}_i(\cdot)$'s based on $\widehat{\boldsymbol{\Sigma}}$, then estimating the basis coefficients vector of $\boldsymbol{\beta}(\cdot)$ and finally recovering the estimated functional coefficients, where details are presented in Section E.1 of the supplementary material and Chiou, Yang, and Chen (2016). When p is large, we can implement our functional thresholding proposals to obtain consistent estimators of $\boldsymbol{\Sigma}$ and hence $\boldsymbol{\beta}$. In Section E.1 of the supplementary material, we demonstrate via a simulated example the superiority of our adaptive-functional-thresholding-based estimator over its competitors.

Our third application considers another dimension reduction framework via *functional factor model* (Guo, Qiao, and Wang 2022) in the form of $\mathbf{X}_i(\cdot) = \mathbf{A}\mathbf{f}_i(\cdot) + \boldsymbol{\varepsilon}_i(\cdot)$, where the common components are driven by r functional factors $\mathbf{f}_i(\cdot) = \{f_{i1}(\cdot), \ldots, f_{ir}(\cdot)\}^{\mathrm{T}}$, the idiosyncratic components are $\boldsymbol{\varepsilon}_i(\cdot)$ and $\mathbf{A} \in \mathbb{R}^{p \times r}$ is the factor loading matrix. Denote the covariance functions of $\mathbf{X}_i(\cdot)$, $\mathbf{f}_i(\cdot)$ and $\boldsymbol{\varepsilon}_i(\cdot)$ by $\boldsymbol{\Sigma}_X$, $\boldsymbol{\Sigma}_f$ and $\boldsymbol{\Sigma}_\varepsilon$, respectively. Under the orthogonality of $\mathbf{A}$, $\int\int \boldsymbol{\Sigma}_X(u,v)\boldsymbol{\Sigma}_X(u,v)^{\mathrm{T}}\mathrm{d}u\mathrm{d}v$ can be decomposed as the sum of $\mathbf{A}\int\int \boldsymbol{\Sigma}_f(u,v)\boldsymbol{\Sigma}_f(u,v)^{\mathrm{T}}\mathrm{d}u\mathrm{d}v\mathbf{A}^{\mathrm{T}}$ and the remaining smaller order terms. Intuitively, with certain identifiable conditions, $\mathbf{A}$ can be recovered by carrying out an eigenanalysis of $\int\int \boldsymbol{\Sigma}_X(u,v)\boldsymbol{\Sigma}_X(u,v)^{\mathrm{T}}\mathrm{d}u\mathrm{d}v$. To provide a parsimonious model and enhance interpretability for near-zero loadings, we can impose subspace sparsity conditions (Vu and Lei 2013) on $\mathbf{A}$ that results in a functional sparse $\boldsymbol{\Sigma}_X$ and hence our functional thresholding estimators become applicable. See an application of our functional thresholding technique to improve the estimation quality when fitting sparse functional factor model in Guo, Qiao, and Wang (2022). See also Section C.2 of the supplementary material for other applications including functional graphical model estimation (Qiao, Guo, and James 2019) and multivariate functional classification.

3. Theoretical Properties

We begin with some notation. For a random variable W, define $\|W\|_\psi = \inf\{c > 0 : \mathbb{E}[\psi(|W|/c)] \leq 1\}$, where $\psi : [0,\infty) \to [0,\infty)$ is a nondecreasing, nonzero convex function with $\psi(0) = 0$ and the norm takes the value ∞ if no finite c exists for which $\mathbb{E}[\psi(|W|/c)] \leq 1$. Denote $\psi_k(x) = \exp(x^k) - 1$ for $k \geq 1$. Let the packing number $D(\epsilon, d)$ be the maximal number of points that can fit in the compact interval $\mathcal{U}$ while maintaining a distance greater than ϵ between all points with respect to the semimetric d. We refer to Chapter 8 of Kosorok (2008) for further explanations. For $\{X_{ij}(u) : u \in \mathcal{U}, i = 1, \ldots, n, j = 1, \ldots, p\}$, define the standardized processes by $Y_{ij}(u) = [X_{ij}(u) - \mathbb{E}\{X_{ij}(u)\}]/\sigma_j(u)^{1/2}$, where $\sigma_j(u) = \Sigma_{jj}(u,u)$.

To present the main theorems, we need the following regularity conditions.

Condition 1. (i) For each i and j, $Y_{ij}(\cdot)$ is a separable stochastic process with the semimetric $d_j(u,v) = \|Y_{1j}(u) - Y_{1j}(v)\|_{\psi_2}$ for $u, v \in \mathcal{U}$; (ii) For some $u_0 \in \mathcal{U}$, $\max_{1 \leq j \leq p} \|Y_{1j}(u_0)\|_{\psi_2}$ is bounded.

Condition 2. The packing numbers $D(\epsilon, d_j)$'s satisfy $\max_{1 \leq j \leq p} D(\epsilon, d_j) \leq C\epsilon^{-r}$ for some constants $C, r > 0$ and $\epsilon \in (0, 1]$.

Condition 3. There exists some constant $\tau > 0$ such that $\min_{j,k} \inf_{u,v \in \mathcal{U}} \mathrm{var}\{Y_{1j}(u)Y_{1k}(v)\} \geq \tau$.

Condition 4. The pair (n,p) satisfies $\log p/n^{1/4} \to 0$ as n and $p \to \infty$.

Conditions 1 and 2 are standard to characterize the modulus of continuity of sub-Gaussian processes $Y_{ij}(\cdot)$'s, see Chapter 8 of Kosorok (2008). These conditions also imply that there exist some positive constants C_0 and η such that $\mathbb{E}[\exp(t\|Y_{1j}\|^2)] \leq C_0$ for all $|t| \leq \eta$ and j with $\|Y_{1j}\| = \{\int_\mathcal{U} Y_{1j}(u)^2\mathrm{d}u\}^{1/2}$, which plays a crucial role in our proof when applying concentration inequalities within Hilbert space. Condition 3 restricts the variances of $Y_{ij}(u)Y_{ik}(v)$'s to be uniformly bounded away from zero so that they can be well estimated. It also facilitates the development of some standardized concentration results. This condition precludes the case of a Brownian motion $X_{ij}(\cdot)$ starting at 0 for some j. However, replacing $X_{ij}(\cdot)$ with a contaminated process $X_{ij}(\cdot) + \xi_{ij}$, where ξ_{ij}'s are independent from a normal distribution with zero mean and a small variance and are independent of $X_{ij}(\cdot)$'s, Condition 3 is fulfilled while the cross-covariance structure in $\boldsymbol{\Sigma}$ remains the same in the sense of $\mathrm{cov}\{X_{ij}(u) + \xi_{ij}, X_{ik}(v)\} = \mathrm{cov}\{X_{ij}(u), X_{ik}(v)\}$ for $k \neq j$ and $u, v \in \mathcal{U}$. Condition 4 allows the high-dimensional case, where p can diverge at some exponential rate as n increases.

We next establish the convergence rate of the adaptive functional thresholding estimator $\widehat{\boldsymbol{\Sigma}}_{\mathrm{A}}$ over a large class of "approximately sparse" covariance functions defined by

$$\mathcal{C}(q, s_0(p), \epsilon_0; \mathcal{U}) = \Big\{\boldsymbol{\Sigma}: \boldsymbol{\Sigma} \succeq 0, \max_{1 \leq j \leq p}\sum_{k=1}^p \|\sigma_j\|_\infty^{(1-q)/2}\|\sigma_k\|_\infty^{(1-q)/2}$$
$$\|\Sigma_{jk}\|_\mathcal{S}^q \leq s_0(p), \max_j \|\sigma_j^{-1}\|_\infty\|\sigma_j\|_\infty$$
$$\leq \epsilon_0^{-1} < \infty\Big\}$$

for some $0 \leq q < 1$, where $\|\sigma_j\|_\infty = \sup_{u \in \mathcal{U}} \sigma_j(u)$ and $\boldsymbol{\Sigma} \succeq 0$ means that $\boldsymbol{\Sigma} = \{\Sigma_{jk}(\cdot, \cdot)\}_{p \times p}$ is positive semidefinite, that is, $\sum_{j,k} \int\int \Sigma_{jk}(u,v)a_j(u)a_k(v)\mathrm{d}u\mathrm{d}v \geq 0$ for any $a_j(\cdot) \in L^2(\mathcal{U})$ and $j = 1, \ldots, p$. See Cai and Liu (2011) for a similar class of covariance matrices for nonfunctional data. Compared with the class

$$\mathcal{C}^*(q, s_0(p), M; \mathcal{U}) = \Big\{\boldsymbol{\Sigma} : \boldsymbol{\Sigma} \succeq 0, \max_j \|\sigma_j\|_\mathcal{N}$$
$$\leq M, \max_j \sum_{k=1}^p \|\Sigma_{jk}\|_\mathcal{S}^q$$
$$\leq s_0(p)\Big\},$$

over which the universal functional thresholding estimator $\widehat{\boldsymbol{\Sigma}}_{\mathrm{U}}$ can be shown to be consistent, the columns of a covariance function in $\mathcal{C}(q, s_0(p), \epsilon_0; \mathcal{U})$ are required to be within a weighted ℓ_q/ℓ_2 ball instead of a standard ℓ_q/ℓ_2 ball, where the weights are determined by $\|\sigma_j\|_\infty$'s. Unlike $\mathcal{C}^*(q, s_0(p), M; \mathcal{U})$, $\mathcal{C}(q, s_0(p), \epsilon_0; \mathcal{U})$ no longer requires the uniform boundedness assumption on $\|\sigma_j\|_\mathcal{N}$'s and allows $\max_j \|\sigma_j\|_\mathcal{N} \to \infty$. In the special case $q = 0$, $\mathcal{C}(q, s_0(p), \epsilon_0; \mathcal{U})$ corresponds to a class of truly sparse covariance functions. Notably, $s_0(p)$ can depend on p and be regarded implicitly as the restriction on functional sparsity.

Theorem 1. Suppose that Conditions 1–4 hold. Then there exists some constant $\delta > 0$ such that, uniformly on $\mathcal{C}(q, s_0(p), \epsilon_0; \mathcal{U})$, if $\lambda = \delta(\log p/n)^{1/2}$,

$$\|\widehat{\boldsymbol{\Sigma}}_A - \boldsymbol{\Sigma}\|_1 = \max_{1 \le k \le p} \sum_{j=1}^{p} \|\widehat{\Sigma}_{jk}^A - \Sigma_{jk}\|_{\mathcal{S}} = O_P\left\{s_0(p)\left(\frac{\log p}{n}\right)^{\frac{1-q}{2}}\right\}. \tag{5}$$

Theorem 1 presents the convergence result in the functional version of matrix ℓ_1 norm. The rate in (5) is consistent to those of sparse covariance matrix estimates in Rothman, Levina, and Zhu (2009) and Cai and Liu (2011).

We finally turn to investigate the support recovery consistency of $\widehat{\boldsymbol{\Sigma}}_A$ over the parameter space of truly sparse covariance functions defined by

$$\mathcal{C}_0(s_0(p); \mathcal{U}) = \left\{\boldsymbol{\Sigma} : \boldsymbol{\Sigma} \succeq 0, \max_{1 \le j \le p} \sum_{k=1}^{p} I(\|\Sigma_{jk}\|_{\mathcal{S}} \neq 0) \le s_0(p)\right\},$$

which assumes that $(\Sigma_{jk})_{p \times p}$ has at most $s_0(p)$ nonzero functional entries on each row. The following theorem shows that, with the choice of $\lambda = \delta(\log p/n)^{1/2}$ for some constant $\delta > 0$, $\widehat{\boldsymbol{\Sigma}}_A$ exactly recovers the support of $\boldsymbol{\Sigma}$, $\text{supp}(\boldsymbol{\Sigma}) = \{(j,k) : \|\Sigma_{jk}\|_{\mathcal{S}} \neq 0\}$, with probability approaching one.

Theorem 2. Suppose that Conditions 1–4 hold and $\|\Sigma_{jk}/\Theta_{jk}^{1/2}\|_{\mathcal{S}} > (2\delta + \gamma)(\log p/n)^{1/2}$ for all $(j,k) \in \text{supp}(\boldsymbol{\Sigma})$ and some $\gamma > 0$, where δ is stated in Theorem 1. Then we have that

$$\inf_{\boldsymbol{\Sigma} \in \mathcal{C}_0} P\{\text{supp}(\widehat{\boldsymbol{\Sigma}}_A) = \text{supp}(\boldsymbol{\Sigma})\} \to 1 \text{ as } n \to \infty.$$

Theorem 2 ensures that $\widehat{\boldsymbol{\Sigma}}_A$ achieves the exact recovery of functional sparsity structure in $\boldsymbol{\Sigma}$, that is, the graph support in functional connectivity analysis, with probability tending to 1. This theorem holds under the condition that the Hilbert-Schmidt norms of nonzero standardized functional entries exceed a certain threshold, which ensures that nonzero components are correctly retained. See an analogous minimum signal strength condition for sparse covariance matrices in Cai and Liu (2011).

4. Partially Observed Functional Data

In this section we consider a practical scenario where each $X_{ij}(\cdot)$ is partially observed, with errors, at random measurement locations $U_{ij1}, \ldots, U_{ijL_{ij}} \in \mathcal{U}$. Let Z_{ijl} be the observed value of $X_{ij}(U_{ijl})$. Then

$$Z_{ijl} = X_{ij}(U_{ijl}) + \varepsilon_{ijl}, \quad l = 1, \ldots, L_{ij}, \tag{6}$$

where ε_{ijl}'s are iid errors with $\mathbb{E}(\varepsilon_{ijl}) = 0$ and $\text{var}(\varepsilon_{ijl}) = \sigma^2$, independent of $X_{ij}(\cdot)$. For dense measurement designs all L_{ij}'s are larger than some order of n, while for sparse designs all L_{ij}'s are bounded (Zhang and Wang 2016; Qiao et al. 2020).

4.1. Estimation Procedure

Based on the observed data, $\{(U_{ijl}, Z_{ijl})\}_{1 \le i \le n, 1 \le j \le p, 1 \le l \le L_{ij}}$, we next present a unified estimation procedure that handles both densely and sparsely sampled functional data.

We first develop a nonparametric smoothing approach to estimate $\Sigma_{jk}(u, v)$'s. Without loss of generality, we assume that $\mathbf{X}_i(\cdot)$ has been centered to have mean zero. Denote $K_h(\cdot) = h^{-1}K(\cdot/h)$ for a univariate kernel function K with a bandwidth $h > 0$. A Local Linear Surface smoother (LLS) is employed to estimate cross-covariance functions $\Sigma_{jk}(u, v)$ $(j \neq k)$ by minimizing

$$\sum_{i=1}^{n} \sum_{l=1}^{L_{ij}} \sum_{m=1}^{L_{ik}} \left\{Z_{ijl}Z_{ikm} - \alpha_0 - \alpha_1(U_{ijl} - u) - \alpha_2(U_{ikm} - v)\right\}^2$$
$$K_{h_C}(U_{ijl} - u)K_{h_C}(U_{ikm} - v), \tag{7}$$

with respect to $(\alpha_0, \alpha_1, \alpha_2)$. Let the minimizer of (7) be $(\hat{\alpha}_0, \hat{\alpha}_1, \hat{\alpha}_2)$ and the resulting estimator is $\widetilde{\Sigma}_{jk}(u, v) = \hat{\alpha}_0$. To estimate marginal-covariance functions $\Sigma_{jj}(u, v)$'s, we observe that $\text{cov}(Z_{ijl}, Z_{ijm}) = \Sigma_{jj}(U_{ijl}, U_{ijm}) + \sigma^2 I(l = m)$, and hence apply a LLS to the off-diagonals of the raw covariances $(Z_{ijl}Z_{ijm})_{1 \le l \le m \le L_{ij}}$. We consider minimizing

$$\sum_{i=1}^{n} \sum_{1 \le l \neq m \le L_{ij}} \left\{Z_{ijl}Z_{ijm} - \beta_0 - \beta_1(U_{ijl} - u) - \beta_2(U_{ikm} - v)\right\}^2$$
$$K_{h_M}(U_{ijl} - u)K_{h_M}(U_{ikm} - v)$$

with respect to $(\beta_0, \beta_1, \beta_2)$, thus obtaining the estimate $\widetilde{\Sigma}_{jj}(u, v) = \hat{\beta}_0$. Note that we drop subscripts j, k of $h_{C,jk}$ and j of $h_{M,j}$ to simplify our notation in this section. However, we select different bandwidths $h_{C,jk}$ and $h_{M,j}$ across $j, k = 1, \ldots, p$ in our empirical studies.

To construct the corresponding adaptive functional thresholding estimator, a standard approach is to incorporate the variance effect of each $\widetilde{\Sigma}_{jk}(u, v)$ into functional thresholding. However, the estimation of $\text{var}\{\widetilde{\Sigma}_{jk}(u, v)\}$'s involves estimating multiple complicated fourth moment terms (Zhang and Wang 2016), which results in high computational burden especially for large p. Since our focus is on characterizing the main variability of $\widetilde{\Sigma}_{jk}(u, v)$ rather than estimating its variance precisely, we next develop a computationally simple yet effective approach to estimate the main terms in the asymptotic variance of $\widetilde{\Sigma}_{jk}(u, v)$. For $a, b = 0, 1, 2$, let

$$T_{ab,ijk}(u, v) = \sum_{l=1}^{L_{ij}} \sum_{m=1}^{L_{ik}} g_{ab}\{h_C, (u, v), (U_{ijl}, U_{ikm})\}Z_{ijl}Z_{ikm}, \tag{8}$$

where $g_{ab}\{h, (u, v), (U_{ijl}, U_{ikm})\} = K_h(U_{ijl} - u)K_h(U_{ikm} - v)(U_{ijl} - u)^a(U_{ikm} - v)^b$. According to Section D.1 of the supplementary material, minimizing (7) yields the resulting estimator

$$\widetilde{\Sigma}_{jk} = \sum_{i=1}^{n} (W_{1,jk}T_{00,ijk} + W_{2,jk}T_{10,ijk} + W_{3,jk}T_{01,ijk}), \tag{9}$$

where $W_{1,jk}, W_{2,jk}, W_{3,jk}$ can be represented via (S.12) in terms of

$$S_{ab,jk}(u, v) = \sum_{i=1}^{n} \sum_{l=1}^{L_{ij}} \sum_{m=1}^{L_{ik}} g_{ab}\{h_C, (u, v), (U_{ijl}, U_{ikm})\},$$
$$a, b = 0, 1, 2. \tag{10}$$

It is notable that the estimator $\widetilde{\Sigma}_{jk}$ in (9) is expressed as the sum of n independent terms. Ignoring the cross-covariances

among observations within the subject that are dominated by the corresponding variances, we propose a surrogate estimator for the asymptotic variance of $\widetilde{\Sigma}_{jk}$ by

$$\widetilde{\Psi}_{jk} = I_{jk} \sum_{i=1}^{n} \left(W_{1,jk} V_{00,ijk} + W_{2,jk} V_{10,ijk} + W_{3,jk} V_{01,ijk} \right)^2, \quad (11)$$

where

$$I_{jk} = \left(\sum_{i=1}^{n} L_{ij} L_{ik} \right)^2 \left\{ \sum_{i=1}^{n} \left(L_{ij} L_{ik} h_C^{-2} + L_{ij}^2 L_{ik} h_C^{-1} \right. \right.$$
$$\left. \left. + L_{ij} L_{ik}^2 h_C^{-1} + L_{ij}^2 L_{ik}^2 \right) \right\}^{-1}, \quad (12)$$

$$V_{ab,ijk}(u,v) = \sum_{l=1}^{L_{ij}} \sum_{m=1}^{L_{ik}} g_{ab} \left\{ h_C, (u,v), (U_{ijl}, U_{ikm}) \right\}$$
$$\left\{ Z_{ijl} Z_{ikm} - \widetilde{\Sigma}_{jk}(u,v) \right\}. \quad (13)$$

The rationale of multiplying the rate I_{jk} in (11) is to ensure that $\widetilde{\Psi}_{jk}(u,v)$ converges to some finite function when $n \to \infty$ and $h_C \to 0$ as justified in Section D.4 of the supplementary material. In particular, the rate I_{jk} can be simplified to $\sum_{i=1}^{n} L_{ij} L_{ik} h_C^2$ for the sparse or moderately dense case and to $(\sum_{i=1}^{n} L_{ij} L_{ik})^2 (\sum_{i=1}^{n} L_{ij}^2 L_{ik}^2)^{-1}$ for the very dense case. Note that I_{jk} is imposed in (11) mainly for the theoretical purpose and hence will not place a practical constraint on our method.

In a similar procedure as above, the estimated variance factor $\widetilde{\Psi}_{jj}$ of $\widetilde{\Sigma}_{jj}$ for each j can be obtained by operating on $\{Z_{ijl} Z_{ijm}\}_{1 \leq i \leq n, 1 \leq l \neq m \leq L_{ij}}$ instead of $\{Z_{ijl} Z_{ikm}\}_{1 \leq i \leq n, 1 \leq l \leq L_{ij}, 1 \leq m \leq L_{ik}}$ for $j \neq k$. Substituting $\widehat{\Theta}_{jk}$ in (2) by $\widetilde{\Psi}_{jk}$, we obtain the smoothed adaptive functional thresholding estimator

$$\widetilde{\boldsymbol{\Sigma}}_{\mathrm{A}} = (\widetilde{\Sigma}_{jk}^{\mathrm{A}})_{p \times p} \ \text{with} \ \widetilde{\Sigma}_{jk}^{\mathrm{A}} = \widetilde{\Psi}_{jk}^{1/2} \times s_\lambda \left(\frac{\widetilde{\Sigma}_{jk}}{\widetilde{\Psi}_{jk}^{1/2}} \right). \quad (14)$$

For comparison, we also define the smoothed universal functional thresholding estimator as $\widetilde{\boldsymbol{\Sigma}}_{\mathrm{U}} = (\widetilde{\Sigma}_{jk}^{\mathrm{U}})_{p \times p}$ with $\widetilde{\Sigma}_{jk}^{\mathrm{U}} = s_\lambda(\widetilde{\Sigma}_{jk})$.

A natural alternative to the proposed LLS-based smoothing procedure considers pre-smoothing each individual data. For densely sampled functional data, the observations $Z_{ij1}, \ldots, Z_{ijL_{ij}}$ for each i and j can be pre-smoothed through the local linear smoother to eliminate the contaminated noise, thus producing reconstructed random curves $\widehat{X}_{ij}(\cdot)$'s before subsequent analysis (Zhang and Chen 2007). See detailed implementation of pre-smoothing in Section D.2 of the supplementary material. For sparsely sampled functional data, such pre-smoothing step is not viable, while our smoothing proposal builds strength across functions by incorporating information from all the observations, and hence is still applicable. See also Section 5.3 for the numerical comparison between pre-smoothing and our smoothing approach under different measurement designs.

4.2. Theoretical Properties

In this section, we investigate the theoretical properties of $\widetilde{\boldsymbol{\Sigma}}_{\mathrm{A}}$ for partially observed functional data. We begin by introducing some notation. For two positive sequences $\{a_n\}$ and $\{b_n\}$, we write $a_n \lesssim b_n$ if there exits a positive constant c_0 such that $a_n/b_n \leq c_0$. We write $a_n \asymp b_n$ if and only if $a_n \lesssim b_n$ and $b_n \lesssim a_n$ hold simultaneously. Before presenting the theory, we impose the following regularity conditions.

Condition 5. (i) Let $\{U_{ijl} : i = 1, \ldots, n, j \in 1, \ldots, p, l = 1, \ldots, L_{ij}\}$ be iid copies of a random variable U with density $f_U(\cdot)$ defined on the compact set $\mathcal{U}$, with the L_{ij}'s fixed. There exist some constants m_f and M_f such that $0 < m_f \leq \inf_{\mathcal{U}} f_U(u) \leq \sup_{\mathcal{U}} f_U(u) \leq M_f < \infty$; (ii) X_{ij}, ε_{ijl} and U_{ijl} are independent for each i, j, l.

Condition 6. (i) Under the sparse measurement design, $L_{ij} \leq L_0 < \infty$ for all i, j and, under the dense design, $L_{ij} = L \to \infty$ as $n \to \infty$ with U_{ijl}'s independent of i; (ii) The bandwidth parameters $h_C \asymp h_M \asymp h \to 0$ as $n \to \infty$.

Condition 5 is standard in functional data analysis literature (Zhang and Wang 2016). Condition 6 (i) treats the number of measurement locations L_{ij} as bounded and diverging under sparse and dense measurement designs, respectively. To simplify notation, we assume that $L_{ij} = L$ for the dense case and h_C is of the same order as h_M in Condition 6 (ii).

Condition 7. There exists some constant $\gamma_1 \in (0, 1/2]$ such that

$$\max_{1 \leq j,k \leq p} \left\| \widetilde{\Sigma}_{jk} - \Sigma_{jk} \right\|_{\mathcal{S}} \lesssim \sqrt{\frac{\log p}{n^{2\gamma_1}}}$$
$$+ h^2 \ \text{with probability approaching one.} \quad (15)$$

Condition 8. There exist some positive constants c_1, $\gamma_2 \in (0, 1/2]$ and some deterministic functions $\Psi_{jk}(u,v)$'s with $\min_{j,k} \inf_{u,v \in \mathcal{U}} \Psi_{jk}(u,v) \geq c_1$ such that

$$\max_{1 \leq j,k \leq p} \sup_{u,v \in \mathcal{U}} \left| \widetilde{\Psi}_{jk}(u,v) - \Psi_{jk}(u,v) \right|$$
$$\lesssim \sqrt{\frac{\log p}{n^{2\gamma_2}}} + h^2 \ \text{with probability approaching one.} \quad (16)$$

Condition 9. The pair (n,p) satisfies $\log p / n^{\min(\gamma_1, \gamma_2)} \to 0$ and $\log p \geq c_2 n^{2\gamma_1} h^4$ for some positive constant c_2 as n and $p \to \infty$.

We follow Qiao et al. (2020) to impose Condition 7, in which the parameter γ_1 depends on h and possibly L under the dense design. This condition is satisfied if there exist some positive constants c_3, c_4, c_5 such that for each $j, k = 1, \ldots, p$ and $t \in (0, 1]$,

$$P(\|\widetilde{\Sigma}_{jk} - \Sigma_{jk}\|_{\mathcal{S}} \geq t + c_5 h^2) \leq c_4 \exp(-c_3 n^{2\gamma_1} t^2). \quad (17)$$

The presence of h^2 comes from the standard results for bias terms under the boundedness condition for the second-order partial derivatives of $\Sigma_{jk}(u,v)$ over $\mathcal{U}^2$ (Yao, Müller, and Wang 2005; Zhang and Wang 2016). This concentration result is fulfilled under different measurement schedules ranging from sparse to dense designs as γ_1 increases. For sparsely sampled functional data, Lemma 4 of Qiao et al. (2020) established L_2 concentration inequality for $\widetilde{\Sigma}_{jk}$ for $j = k$, which not only results in the same L_2 rate as that in the sparse case (Zhang and Wang 2016) but also ensures (17) with the choice of $\gamma_1 = 1/2 - a$

and $h \asymp n^{-a}$ for some positive constant $a < 1/2$. Following the same proof procedure, the same concentration inequality also applies for $j \neq k$ and hence Condition 7 is satisfied. This condition is also satisfied by densely sampled functional data, since it follows from Lemma 5 of Qiao et al. (2020) that (17) holds for $j = k$ and, with more efforts, also for $j \neq k$ by choosing $\gamma_1 = \min(1/2, 1/3 + b/6 - \epsilon'/2 - 2a/3)$ for some small constant $\epsilon' > 0$ when $h \asymp n^{-a}$ and $L \asymp n^b$ for some constants $a, b > 0$. As L grows sufficiently large, $\gamma_1 = 1/2$, thus leading to the same rate as that in the ultra-dense case (Zhang and Wang 2016). Condition 8 gives the uniform convergence rate for $\widehat{\Psi}_{jk}(u, v)$ in the same form as (15) but with different parameter γ_2. A denser measurement design corresponds to a larger value of γ_2 and a faster rate in (16). See the heuristic verification of Condition 8 in Section D.4 of the supplementary material. Condition 9 indicates that p can grow exponentially fast relative to n.

We next present the convergence rate of the smoothed adaptive functional thresholding estimator $\widetilde{\Sigma}_A$ over a class of "approximate sparse" covariance functions defined by

$$\widetilde{\mathcal{C}}(q, \tilde{s}_0(p), \epsilon_0; \mathcal{U}) = \left\{ \boldsymbol{\Sigma} : \boldsymbol{\Sigma} \succeq 0, \max_{1 \leq j \leq p} \sum_{k=1}^{p} \|\Psi_{jk}\|_\infty^{(1-q)/2} \|\Sigma_{jk}\|_{\mathcal{S}}^q \right.$$
$$\left. \leq \tilde{s}_0(p), \max_{j,k} \|\Psi_{jk}^{-1}\|_\infty \|\Psi_{jk}\|_\infty \leq \epsilon_0^{-1} < \infty \right\},$$

for some $0 \leq q < 1$.

Theorem 3. Suppose that Conditions 5–9 hold. Then there exists some constants $\tilde{\delta} > 0$ such that, uniformly on $\widetilde{\mathcal{C}}(q, \tilde{s}_0(p), \epsilon_0; \mathcal{U})$, if $\lambda = \tilde{\delta}(\log p / n^{2\gamma_1})^{1/2}$,

$$\|\widetilde{\Sigma}_A - \boldsymbol{\Sigma}\|_1 = \max_{1 \leq k \leq p} \sum_{j=1}^{p} \|\widetilde{\Sigma}_{jk}^A - \Sigma_{jk}\|_{\mathcal{S}} = O_P \left\{ \tilde{s}_0(p) \left(\frac{\log p}{n^{2\gamma_1}} \right)^{\frac{1-q}{2}} \right\}.$$
$$(18)$$

The convergence rate of $\widetilde{\Sigma}_A$ in (18) is governed by internal parameters (γ_1, q) and other dimensionality parameters. Larger values of γ_1 correspond to a more frequent measurement schedule with larger L and result in a faster rate. The convergence result implicitly reveals interesting phase transition phenomena depending on the relative order of L to n. As L grows fast enough, $\gamma_1 = 1/2$ and the rate is consistent to that for fully observed functional data in (5), presenting that the theory for very densely sampled functional data falls in the parametric paradigm. As L grows moderately fast, $\gamma_1 < 1/2$ and the rate is faster than that for sparsely sampled functional data but slower than the parametric rate.

We finally present Theorem 4 that guarantees the support recovery consistency of $\widetilde{\Sigma}_A$.

Theorem 4. Suppose that Conditions 5–9 hold and $\left\| \Sigma_{jk} / \Psi_{jk}^{1/2} \right\|_{\mathcal{S}} > (2\tilde{\delta} + \tilde{\gamma})(\log p / n^{2\gamma_1})^{1/2}$ for all $(j, k) \in \text{supp}(\boldsymbol{\Sigma})$ and some $\tilde{\gamma} > 0$, where $\tilde{\delta}$ is stated in Theorem 3, then

$$\inf_{\boldsymbol{\Sigma} \in \mathcal{C}_0} P\{\text{supp}(\widetilde{\Sigma}_A) = \text{supp}(\boldsymbol{\Sigma})\} \to 1 \text{ as } n \to \infty.$$

4.3. Fast Computation

Consider a common situation in practice, where, for each $i = 1, \ldots, n$, we observe the noisy versions of $X_{i1}(\cdot), \ldots, X_{ip}(\cdot)$ at the same set of points, $U_{i1}, \ldots, U_{iL_i} \in \mathcal{U}$, across $j = 1, \ldots, p$. Then the original model in (6) is simplified to

$$Z_{ijl} = X_{ij}(U_{il}) + \varepsilon_{ijl}, \ l = 1, \ldots, L_i, \quad (19)$$

under which the proposed estimation procedure in Section 4.1 can still be applied. Suppose that the estimated covariance function is evaluated at a grid of $R \times R$ locations, $\{(u_{r_1}, u_{r_2}) \in \mathcal{U}^2 : r_1, r_2 = 1, \ldots, R\}$. To serve the estimation of $p(p + 1)/2$ marginal- and cross-covariance functions and the corresponding variance factors, LLSs under the simplified model in (19) reduce the number of kernel evaluations from $O(\sum_{i=1}^{n} \sum_{j=1}^{p} L_{ij} R)$ to $O(\sum_{i=1}^{n} L_i R)$, which substantially accelerate the computation under a high-dimensional regime.

Apparently, such nonparametric smoothing approach is conceptually simple but suffers from high computational cost in kernel evaluations. To further reduce the computational burden, we consider fast implementations of LLSs by adopting a simple approximation technique, known as linear binning (Fan and Marron 1994), to the covariance function estimation. The key idea of the binning method is to greatly reduce the number of kernel evaluations through the fact that many of these evaluations are nearly the same. We start by dividing $\mathcal{U}$ into an equally spaced grid of R points, $u_1 < \cdots < u_R \in \mathcal{U}$, with binwidth $\Delta = u_2 - u_1$. Denote by $w_r(U_{il}) = \max(1 - \Delta^{-1}|U_{il} - u_r|, 0)$ the linear weight that U_{il} assigns to the grid point u_r for $r = 1, \ldots, R$. For the ith subject, we define its "binned weighted counts" and "binned weighted averages" as

$$\varpi_{r,i} = \sum_{l=1}^{L_i} w_r(U_{il}) \quad \text{and} \quad \mathcal{D}_{r,ij} = \sum_{l=1}^{L_i} w_r(U_{il}) Z_{ijl},$$

respectively. The binned implementation of smoothed adaptive functional thresholding can then be done using this modified dataset $\{(\varpi_{r,i}, \mathcal{D}_{r,ij})\}_{1 \leq i \leq n, 1 \leq j \leq p, 1 \leq r \leq R}$ and related kernel functions $g_{ab}\{h, (u, v), (u_{r_1}, u_{r_2})\}$ for $r_1, r_2 = 1, \ldots, R$. It is notable that, with the help of such binned implementation, the number of kernel evaluations required in the covariance function estimation is further reduced from $O(\sum_{i=1}^{n} L_i R)$ to $O(R)$, while only $O(\sum_{i=1}^{n} L_i)$ additional operations are involved for each j in the binning step (Fan and Marron 1994).

We next illustrate the binned implementation of LLS, denoted as BinLLS, using the example of smoothed estimates $\widetilde{\Sigma}_{jk}$ for $j \neq k$ in (9). Under Model (19), we drop subscripts j, k in $W_{1,jk}$, $W_{2,jk}$, $W_{3,jk}$, and $S_{ab,jk}$ due to the same set of points $\{U_{i1}, \ldots, U_{iL_i}\}$ across j, k. Denote the binned approximations of $T_{ab,ijk}$ and S_{ab} by $\check{T}_{ab,ijk}$ and $\check{S}_{ab}$, respectively. It follows from (8) and (10) that

$$\check{T}_{ab,ijk}(u, v) = \sum_{r_1=1}^{R} \sum_{r_2=1}^{R} g_{ab}\{h_C, (u, v), (u_{r_1}, u_{r_2})\} \mathcal{D}_{r_1,ij} \mathcal{D}_{r_2,ik},$$

$$\check{S}_{ab}(u, v) = \sum_{i=1}^{n} \sum_{r_1=1}^{R} \sum_{r_2=1}^{R} g_{ab}\{h_C, (u, v), (u_{r_1}, u_{r_2})\} \varpi_{r_1,i} \varpi_{r_2,i},$$

Table 1. The computational complexity analysis of LLS, BinLLS under Models (6), (19) when evaluating the corresponding smoothed covariance function estimates at a grid of $R \times R$ points.

Method	Model	Number of kernel evaluations	Number of operations (additions and multiplications)
LLS	(6)	$O(\sum_{i=1}^{n} \sum_{j=1}^{p} L_{ij}R)$	$O(R^2 \sum_{i=1}^{n} \sum_{j,k=1}^{p} L_{ij}L_{ik})$
LLS	(19)	$O(\sum_{i=1}^{n} L_i R)$	$O(p^2 R^2 \sum_{i=1}^{n} L_i^2)$
BinLLS	(19)	$O(R)$	$O(np^2R^2 + p^2R^4 + p\sum_{i=1}^{n} L_i)$

both of which together with (9) yield the binned approximation of $\widetilde{\Sigma}_{jk}$ as

$$\check{\Sigma}_{jk} = \sum_{i=1}^{n} \left(\check{W}_1 \check{T}_{00,ijk} + \check{W}_2 \check{T}_{10,ijk} + \check{W}_3 \check{T}_{01,ijk} \right),$$

where $\check{W}_1$, $\check{W}_2$, and $\check{W}_3$ are the binned approximations of $W_1, W_2,$ and W_3, computed by replacing the related S_{ab}'s in (S.12) of the supplementary material with the $\check{S}_{ab}$'s. It is worth noting that, for each pair (j, k), the above binned implementation reduces the number of operations (i.e., additions and multiplications) from $O(R^2 \sum_{i=1}^{n} L_i^2)$ to $O(nR^2 + R^4)$, since the kernel evaluations in $g_{ab}\{h_C, (u, v), (u_{r_1}, u_{r_2})\}$ no longer depend on individual observations. Table 1 presents the computational complexity analysis of LLS and BinLLS under Models (6) and (19). It reveals that the binned implementation dramatically improves the computational speed for both densely and sparsely sampled functional data, which is also supported by the empirical evidence in Section 5.3.

To aid the binned implementation of the smoothed adaptive functional thresholding estimator, we then derive the binned approximation of the variance factor $\widetilde{\Psi}_{jk}$, denoted by $\check{\Psi}_{jk}$. It follows from (13) that $V_{ab,ijk}$ can be approximated by

$$\check{V}_{ab,ijk}(u, v) = \sum_{r_1=1}^{R} \sum_{r_2=1}^{R} g_{ab}\left(h_C, (u, v), (u_{r_1}, u_{r_2}) \right)$$
$$\left\{ \mathcal{D}_{r_1,ij} \mathcal{D}_{r_2,ik} - \check{\Sigma}_{jk}(u, v) \varpi_{r_1,i} \varpi_{r_2,i} \right\}.$$

Substituting each term in (11) with its binned approximation, we obtain that

$$\check{\Psi}_{jk} = I_{jk} \sum_{i=1}^{n} \left(\check{W}_1 \check{V}_{00,ijk} + \check{W}_2 \check{V}_{10,ijk} + \check{W}_3 \check{V}_{01,ijk} \right)^2.$$

It is worth mentioning that, when $j = k$, the binned approximations of $\widetilde{\Sigma}_{jj}$ and $\widetilde{\Psi}_{jj}$ can be computed in a similar fashion except that the terms corresponding to $r_1 = r_2$ should be excluded from all double summations over $\{1, \ldots, R\}^2$. Finally, we obtain the binned adaptive functional thresholding estimator $\check{\Sigma}_{\mathrm{A}} = (\check{\Sigma}_{jk}^{\mathrm{A}})_{p \times p}$ with $\check{\Sigma}_{jk}^{\mathrm{A}} = \check{\Psi}_{jk}^{1/2} \times s_\lambda(\check{\Sigma}_{jk}/\check{\Psi}_{jk}^{1/2})$ and the corresponding universal thresholding estimator $\check{\Sigma}_{\mathrm{U}} = (\check{\Sigma}_{jk}^{\mathrm{U}})_{p \times p}$ with $\check{\Sigma}_{jk}^{\mathrm{U}} = s_\lambda(\check{\Sigma}_{jk})$.

5. Simulations

5.1. Setup

We conduct a number of simulations to compare adaptive functional thresholding estimators to universal functional thresholding estimators. Sections 5.2 and 5.3 consider scenarios where random functions are fully and partially observed, respectively.

In each scenario, to mimic the infinite-dimensionality of random curves, we generate functional variables by $X_{ij}(u) = \mathbf{s}(u)^{\mathrm{T}} \boldsymbol{\theta}_{ij}$ for $i = 1, \ldots, n, j = 1, \ldots, p$ and $u \in \mathcal{U} = [0, 1]$, where $\mathbf{s}(u)$ is a 50-dimensional Fourier basis function and $\boldsymbol{\theta}_i = (\boldsymbol{\theta}_{i1}^{\mathrm{T}}, \ldots, \boldsymbol{\theta}_{ip}^{\mathrm{T}})^{\mathrm{T}} \in \mathbb{R}^{50p}$ is generated from a mean zero multivariate Gaussian distribution with block covariance matrix $\boldsymbol{\Omega} \in \mathbb{R}^{50p \times 50p}$, whose (j, k)th block is $\boldsymbol{\Omega}_{jk} \in \mathbb{R}^{50 \times 50}$ for $j, k = 1, \ldots, p$. The functional sparsity pattern in $\boldsymbol{\Sigma} = \{\Sigma_{jk}(\cdot, \cdot)\}_{p \times p}$ with its (j, k)th entry $\Sigma_{jk}(u, v) = \mathbf{s}(u)^{\mathrm{T}} \boldsymbol{\Omega}_{jk} \mathbf{s}(v)$ can be characterized by the block sparsity structure in $\boldsymbol{\Omega}$. Define $\boldsymbol{\Omega}_{jk} = \omega_{jk}\mathbf{D}$ with $\mathbf{D} = \mathrm{diag}(1^{-2}, \ldots, 50^{-2})$ and hence $\mathrm{cov}(\theta_{ijk}, \theta_{ijk'}) \sim k^{-2}I(k = k')$ for $k, k' = 1, \ldots, 50$. Then we generate $\boldsymbol{\Omega}$ with different block sparsity patterns as follows.

- Model 1 (block banded). For $j, k = 1, \ldots, p/2$, $\omega_{jk} = (1 - |j - k|/10)_+$. For $j, k = p/2 + 1, \ldots, p$, $\omega_{jk} = 4I(j = k)$.
- Model 2 (block sparse without any special structure). For $j, k = p/2 + 1, \ldots, p$, $\omega_{jk} = 4I(j = k)$. For $j, k = 1, \ldots, p/2$, we generate $\boldsymbol{\omega} = (\omega_{jk})_{p/2 \times p/2} = \mathbf{B} + \delta' \mathbf{I}_{p/2}$, where elements of $\mathbf{B}$ are sampled independently from Uniform$[0.3, 0.8]$ with probability 0.2 or 0 with probability 0.8, and $\delta' = \max\{-\lambda_{\min}(\mathbf{B}), 0\} + 0.01$ to guarantee the positive definiteness of $\boldsymbol{\Omega}$.

We implement a cross-validation approach (Bickel and Levina 2008) for choosing the optimal thresholding parameter $\hat{\lambda}$ in $\widehat{\boldsymbol{\Sigma}}_{\mathrm{A}}$. Specifically, we randomly divide the sample $\{\mathbf{X}_i : i = 1, \ldots, n\}$ into two subsamples of size n_1 and n_2, where $n_1 = n(1 - 1/\log n)$ and $n_2 = n/\log n$ and repeat this N times. Let $\widehat{\boldsymbol{\Sigma}}_{\mathrm{A},1}^{(v)}(\lambda)$ and $\widehat{\boldsymbol{\Sigma}}_{\mathrm{S},2}^{(v)}$ be the adaptive functional thresholding estimator as a function of λ and the sample covariance function based on n_1 and n_2 observations, respectively, from the vth split. We select the optimal $\hat{\lambda}$ by minimizing

$$\widehat{\mathrm{err}}(\lambda) = N^{-1} \sum_{v=1}^{N} \|\widehat{\boldsymbol{\Sigma}}_{\mathrm{A},1}^{(v)}(\lambda) - \widehat{\boldsymbol{\Sigma}}_{\mathrm{S},2}^{(v)}\|_{\mathrm{F}}^2,$$

where $\| \cdot \|_{\mathrm{F}}$ denotes the functional version of Frobenius norm, that is, for any $\mathbf{Q} = \{Q_{jk}(\cdot, \cdot)\}_{p \times p}$ with each $Q_{jk} \in \mathbb{S}$, $\|\mathbf{Q}\|_{\mathrm{F}} = (\sum_{j,k} \|Q_{jk}\|_{\mathcal{S}}^2)^{1/2}$. The optimal thresholding parameters in $\widehat{\boldsymbol{\Sigma}}_{\mathrm{U}}, \widetilde{\boldsymbol{\Sigma}}_{\mathrm{A}}, \widetilde{\boldsymbol{\Sigma}}_{\mathrm{U}}, \check{\boldsymbol{\Sigma}}_{\mathrm{A}}, \check{\boldsymbol{\Sigma}}_{\mathrm{U}}$ can be selected in a similar fashion.

5.2. Fully Observed Functional Data

We compare the adaptive functional thresholding estimator $\widehat{\boldsymbol{\Sigma}}_{\mathrm{A}}$ to the universal functional thresholding estimator $\widehat{\boldsymbol{\Sigma}}_{\mathrm{U}}$ under hard, soft, SCAD (with $a = 3.7$) and adaptive lasso (with $\eta = 3$) functional thresholding rules, where the corresponding $\hat{\lambda}$'s are selected by the cross-validation with $N = 5$. We generate $n = 100$ observations for $p = 50, 100, 150$ and replicate

Table 2. The average (standard error) functional matrix losses over 100 simulation runs.

Model	Method	$p = 50$		$p = 100$		$p = 150$	
		$\widehat{\boldsymbol{\Sigma}}_A$	$\widehat{\boldsymbol{\Sigma}}_U$	$\widehat{\boldsymbol{\Sigma}}_A$	$\widehat{\boldsymbol{\Sigma}}_U$	$\widehat{\boldsymbol{\Sigma}}_A$	$\widehat{\boldsymbol{\Sigma}}_U$
		Functional Frobenius norm					
	Hard	5.67(0.03)	9.39(0.02)	9.48(0.04)	15.79(0.01)	14.00(0.05)	22.26(0.01)
	Soft	6.14(0.03)	8.55(0.04)	10.28(0.05)	15.00(0.05)	14.8(0.05)	21.89(0.04)
	SCAD	5.94(0.03)	8.59(0.04)	9.96(0.05)	15.02(0.05)	14.49(0.06)	21.91(0.04)
	Adap. lasso	5.44(0.03)	9.10(0.04)	8.99(0.04)	15.73(0.02)	13.02(0.05)	22.25(0.01)
	Sample	21.80(0.04)		43.51(0.06)		65.22(0.07)	
2		Functional matrix ℓ_1 norm					
	Hard	2.85(0.03)	4.74(0.01)	4.77(0.05)	7.11(0.01)	7.65(0.07)	10.31(0.01)
	Soft	3.31(0.03)	4.51(0.04)	5.37(0.04)	6.90(0.02)	8.21(0.05)	10.21(0.01)
	SCAD	3.22(0.03)	4.48(0.03)	5.29(0.04)	6.91(0.02)	8.14(0.05)	10.21(0.01)
	Adap. lasso	2.75(0.03)	4.66(0.02)	4.62(0.05)	7.08(0.01)	7.35(0.07)	10.30(0.01)
	Sample	28.06(0.12)		56.01(0.19)		84.13(0.23)	

Table 3. The average TPRs/ FPRs over 100 simulation runs.

Model	Method	$p = 50$		$p = 100$		$p = 150$	
		$\widehat{\boldsymbol{\Sigma}}_A$	$\widehat{\boldsymbol{\Sigma}}_U$	$\widehat{\boldsymbol{\Sigma}}_A$	$\widehat{\boldsymbol{\Sigma}}_U$	$\widehat{\boldsymbol{\Sigma}}_A$	$\widehat{\boldsymbol{\Sigma}}_U$
2	Hard	0.77/0.00	0.00/0.00	0.68/0.00	0.00/0.00	0.63/0.00	0.00/0.00
	Soft	0.99/0.06	0.50/0.07	0.97/0.04	0.30/0.04	0.96/0.04	0.11/0.02
	SCAD	0.99/0.06	0.47/0.06	0.98/0.05	0.29/0.04	0.97/0.05	0.10/0.01
	Adap. lasso	0.91/0.00	0.10/0.01	0.86/0.00	0.01/0.00	0.83/0.00	0.00/0.00

Table 4. The average (standard error) functional matrix losses and average CPU time for $p = 6$ over 100 simulation runs.

L_i	Method	Functional Frobenius norm	Functional matrix ℓ_1 norm	Elapsed time (sec)	Method	Functional Frobenius norm	Functional matrix ℓ_1 norm	Elapsed time (sec)
11	BinLLS	1.57(0.02)	1.72(0.03)	2.06	BinLLS-P	4.14(0.03)	4.36(0.04)	0.18
	LLS	1.62(0.02)	1.76(0.03)	50.52	LLS-P	4.23(0.04)	4.47(0.05)	0.22
21	BinLLS	1.28(0.02)	1.42(0.03)	2.07	BinLLS-P	2.66(0.02)	2.80(0.02)	0.19
	LLS	1.28(0.02)	1.42(0.03)	136.88	LLS-P	2.67(0.02)	2.82(0.03)	0.29
51	BinLLS	1.06(0.02)	1.20(0.03)	2.21	BinLLS-P	1.12(0.03)	1.26(0.03)	0.20
	LLS	1.04(0.02)	1.18(0.03)	967.75	LLS-P	1.12(0.03)	1.26(0.03)	0.39
101	BinLLS	1.00(0.02)	1.14(0.03)	2.23	BinLLS-P	0.99(0.02)	1.13(0.03)	0.21
	LLS	-	-	-	LLS-P	0.97(0.02)	1.11(0.03)	0.64

$\widehat{\boldsymbol{\Sigma}}_S$	Functional Frobenius norm	Functional matrix ℓ_1 norm	Elapsed time (sec)
	1.04(0.03)	1.20(0.03)	0.11

each simulation 100 times. We examine the performance of all competing approaches by estimation and support recovery accuracies. In terms of the estimation accuracy, Table 2 reports numerical summaries of losses measured by functional versions of Frobenius and matrix ℓ_1 norms. To assess the support recovery consistency, we present in Table 3 the average of true positive rates (TPRs) and false positive rates (FPRs), defined as TPR $= \#\{(j,k) : \|\widehat{\Sigma}_{jk}\|_{\mathcal{S}} \neq 0 \text{ and } \|\Sigma_{jk}\|_{\mathcal{S}} \neq 0\}/\#\{(j,k) : \|\Sigma_{jk}\|_{\mathcal{S}} \neq 0\}$ and FPR $= \#\{(j,k) : \|\widehat{\Sigma}_{jk}\|_{\mathcal{S}} \neq 0 \text{ and } \|\Sigma_{jk}\|_{\mathcal{S}} = 0\}/\#\{(j,k) : \|\Sigma_{jk}\|_{\mathcal{S}} = 0\}$. Since the results under Models 1 and 2 have similar trends, we only present the numerical results under Model 2 here to save space. See Tables 9 and 10 of the supplementary material for results under Model 1.

Several conclusions can be drawn from Tables 2–3 and 9–10 of the supplementary material. First, in all scenarios, $\widehat{\boldsymbol{\Sigma}}_A$ provides substantially improved accuracy over $\widehat{\boldsymbol{\Sigma}}_U$ regardless of the thresholding rule or the loss used. We also obtain the sample covariance function $\widehat{\boldsymbol{\Sigma}}_S$, the results of which deteriorate severely compared with $\widehat{\boldsymbol{\Sigma}}_A$ and $\widehat{\boldsymbol{\Sigma}}_U$. Second, for support recovery, again $\widehat{\boldsymbol{\Sigma}}_A$ uniformly outperforms $\widehat{\boldsymbol{\Sigma}}_U$, which fails to recover the functional sparsity pattern especially when p is large. Third,

the adaptive functional thresholding approach using the hard and the adaptive lasso functional thresholding rules tends to have lower losses and lower TPRs/FPRs than that using the soft and the SCAD functional thresholding rules.

5.3. Partially Observed Functional Data

In this section, we assess the finite-sample performance of LLS and BinLLS methods to handle partially observed functional data. We first generate random functions $X_{ij}(\cdot)$ for $i = 1, \ldots, n, j = 1, \ldots, p$ by the same procedure as in Section 5.1 with either nonsparse or sparse $\boldsymbol{\Sigma}$ depending on p. We then generate the observed values Z_{ijl} from (19), where the measurement locations U_{il} and errors ε_{ijl} are sampled independently from Uniform[0,1] and $\mathcal{N}(0, 0.5^2)$, respectively. We consider settings of $n = 100$ and $L_i = 11, 21, 51, 101$, changing from sparse to moderately dense to very dense measurement schedules. We use the Gaussian kernel with the optimal bandwidths proportional to $n^{-1/6}$, $(nL_i^2)^{-1/6}$ and $n^{-1/4}$, respectively, as suggested in Zhang and Wang (2016), so for the empirical work in this article we choose the proportionality

Table 5. The average (standard error) functional matrix losses for partially observed functional scenarios and $p = 50$ over 100 simulation runs.

Model	Method	$L_i = 11$ $\check{\boldsymbol{\Sigma}}_A$	$\check{\boldsymbol{\Sigma}}_U$	$L_i = 21$ $\check{\boldsymbol{\Sigma}}_A$	$\check{\boldsymbol{\Sigma}}_U$	$L_i = 51$ $\check{\boldsymbol{\Sigma}}_A$	$\check{\boldsymbol{\Sigma}}_U$	$L_i = 101$ $\check{\boldsymbol{\Sigma}}_A$	$\check{\boldsymbol{\Sigma}}_U$
					Functional Frobenius norm				
	Hard	8.12(0.03)	10.41(0.02)	6.85(0.04)	9.89(0.01)	6.06(0.04)	9.60(0.02)	5.75(0.04)	9.51(0.02)
	Soft	8.35(0.03)	10.37(0.01)	7.35(0.03)	9.60(0.03)	6.72(0.03)	8.86(0.04)	6.48(0.03)	8.56(0.04)
	SCAD	8.32(0.03)	10.37(0.01)	7.23(0.04)	9.60(0.03)	6.50(0.04)	8.89(0.04)	6.23(0.04)	8.61(0.04)
	Adap. lasso	7.83(0.03)	10.39(0.01)	6.69(0.04)	9.84(0.02)	5.97(0.04)	9.40(0.04)	5.71(0.04)	9.16(0.04)
2					Functional matrix ℓ_1 norm				
	Hard	3.82(0.04)	4.91(0.01)	3.36(0.04)	4.82(0.01)	3.00(0.05)	4.78(0.01)	2.85(0.05)	4.77(0.01)
	Soft	3.96(0.02)	4.88(0.01)	3.71(0.03)	4.72(0.02)	3.50(0.03)	4.55(0.03)	3.44(0.03)	4.47(0.03)
	SCAD	3.96(0.02)	4.88(0.01)	3.67(0.03)	4.72(0.02)	3.41(0.03)	4.55(0.02)	3.32(0.03)	4.48(0.02)
	Adap. lasso	3.65(0.04)	4.90(0.01)	3.28(0.04)	4.80(0.01)	2.96(0.04)	4.73(0.01)	2.88(0.04)	4.69(0.02)

Table 6. The average TPRs/ FPRs for partially observed functional scenarios and $p = 50$ over 100 simulation runs.

Model	Method	$L_i = 11$ $\check{\boldsymbol{\Sigma}}_A$	$\check{\boldsymbol{\Sigma}}_U$	$L_i = 21$ $\check{\boldsymbol{\Sigma}}_A$	$\check{\boldsymbol{\Sigma}}_U$	$L_i = 51$ $\check{\boldsymbol{\Sigma}}_A$	$\check{\boldsymbol{\Sigma}}_U$	$L_i = 101$ $\check{\boldsymbol{\Sigma}}_A$	$\check{\boldsymbol{\Sigma}}_U$
2	Hard	0.58/0.00	0.00/0.00	0.69/0.00	0.00/0.00	0.75/0.00	0.01/0.00	0.79/0.00	0.01/0.00
	Soft	0.95/0.04	0.03/0.01	0.97/0.05	0.22/0.03	0.99/0.06	0.48/0.06	0.99/0.06	0.58/0.07
	SCAD	0.95/0.04	0.03/0.01	0.97/0.06	0.22/0.03	0.99/0.07	0.46/0.06	0.99/0.07	0.54/0.06
	Adap. lasso	0.80/0.00	0.00/0.00	0.86/0.00	0.02/0.00	0.90/0.00	0.08/0.00	0.91/0.00	0.15/0.01

constants in the range $(0, 1]$, which gives good results in all settings we consider.

To compare BinLLS with LLS in terms of the computational speed and estimation accuracy, we first consider a low-dimensional example $p = 6$ with nonsparse $\boldsymbol{\Sigma}$ generated by modifying Model 1 with $\omega_{jk} = (1 - |j - k|/10)_+$ for $j, k = 1, \ldots, 6$. In addition to our proposed smoothing methods, we also implement local-linear-smoother-based pre-smoothing and its binned implementation, denoted as LLS-P and BinLLS-P, respectively. Table 4 reports numerical summaries of estimation errors evaluated at $R = 21$ equally spaced points in $[0, 1]$ and the corresponding CPU time on the processor Intel(R) Xeon(R) CPU E5-2690 v3 @ 2.60GHz. The results for the sample covariance function $\widehat{\boldsymbol{\Sigma}}_S$ based on fully observed $\mathbf{X}_1(\cdot), \ldots, \mathbf{X}_n(\cdot)$ are also provided as the baseline for comparison. Note that, LLS is too slow to implement for the case $L_i = 101$, so we do not report its result here.

A few trends are observable from Table 4. First, the binned implementations (BinLLS and BinLLS-P) attain similar or even lower estimation errors compared with their direct implementations (LLS and LLS-P) under all scenarios, while resulting in considerably faster computational speeds especially under dense designs. For example, BinLLS runs over 400 times faster than LLS when $L_i = 51$. Second, all methods provide higher estimation accuracies as L_i increases, and enjoy similar performance when functions are very densely observed, for example, $L_i = 51$ and 101, compared with the fully observed functional case. However, the performance of LLS-P and BinLLS-P deteriorates severely under sparse designs, for example, $L_i = 11$ and 21, since limited information is available from a small number of observations per subject. Among all competitors, we conclude that BinLLS is overall a unified approach that can handle both sparsely and densely sampled functional data well with increased computational efficiency and guaranteed estimation accuracy.

We next examine the performance of BinLLS-based adaptive and universal functional thresholding estimators in terms of estimation accuracy and support recovery consistency using the same performance measures as in Tables 2–3. Tables 5–6 and Tables 11–14 of the supplementary material report numerical results for settings of $p = 50$ and 100 satisfying Models 1 and 2 under different measurement schedules. We observe a few apparent patterns from Tables 5–6 and 11–14. First, $\check{\boldsymbol{\Sigma}}_A$ substantially outperforms $\check{\boldsymbol{\Sigma}}_U$ with significantly lower estimation errors in all settings. Second, $\check{\boldsymbol{\Sigma}}_A$ works consistently well in recovering the functional sparsity structures especially under the soft and SCAD functional thresholding rules, while $\check{\boldsymbol{\Sigma}}_U$ fails to identify such patterns. Third, the estimation and support recovery consistencies of $\check{\boldsymbol{\Sigma}}_A$ and $\check{\boldsymbol{\Sigma}}_U$ are improved as L_i increases. When curves are very densely observed, for example, $L_i = 101$, we observe that both estimators enjoy similar performance with $\widehat{\boldsymbol{\Sigma}}_A$ and $\widehat{\boldsymbol{\Sigma}}_U$ in Tables 2–3 and 9–10 of the supplementary material. Such observation provides empirical evidence to support our remark for Theorem 3 about the same convergence rate between very densely observed and fully observed functional scenarios.

6. Real Data

In this section, we aim to investigate the association between the brain functional connectivity and fluid intelligence (*gF*), the capacity to solve problems independently of acquired knowledge (Cattell 1987). The dataset contains subjects of resting-state fMRI scans and the corresponding *gF* scores, measured by the 24-item Raven's Progressive Matrices, from the Human Connectome Project (HCP). We follow many recent proposals based on HCP by modeling signals as multivariate random functions with each region of interest (ROI) representing one random function (Lee et al. in press; Miao, Zhang, and Wong in press; Zapata, Oh, and Petersen 2022). We focus our analysis on $n_{\text{low}} = 73$ subjects with intelligence scores $gF \leq 8$ and $n_{\text{high}} = 85$ subjects with $gF \geq 23$, and consider $p = 83$ ROIs of three generally acknowledged modules in neuroscience
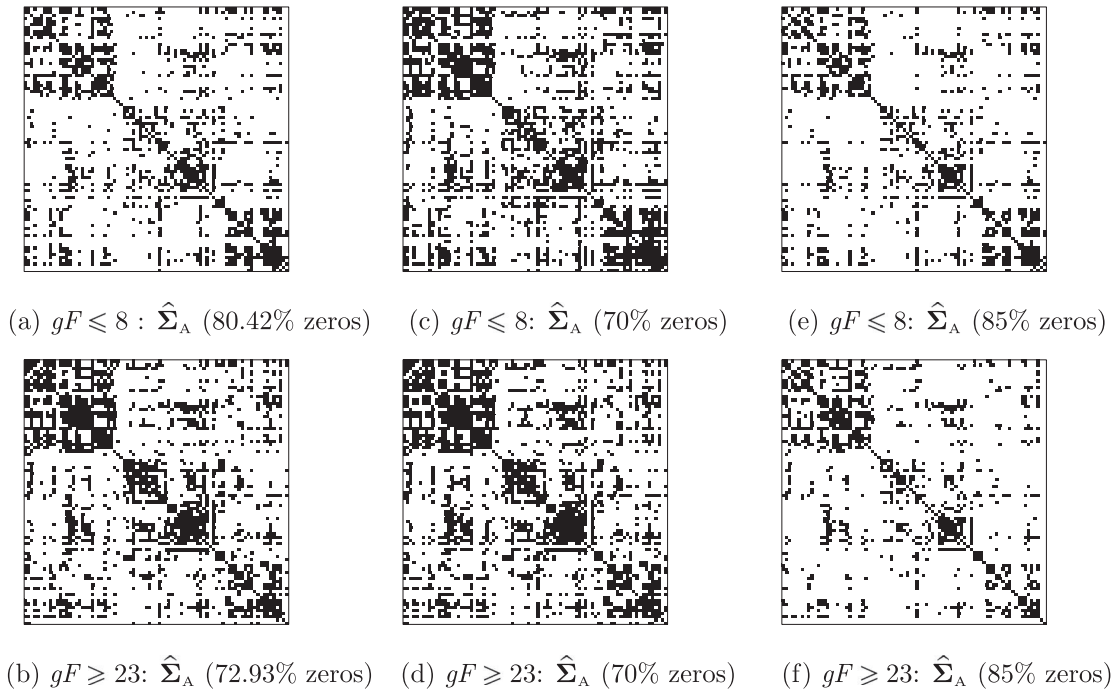
(a) $gF \leqslant 8 : \widehat{\boldsymbol{\Sigma}}_{\mathrm{A}}$ (80.42% zeros)   (c) $gF \leqslant 8 : \widehat{\boldsymbol{\Sigma}}_{\mathrm{A}}$ (70% zeros)   (e) $gF \leqslant 8 : \widehat{\boldsymbol{\Sigma}}_{\mathrm{A}}$ (85% zeros)

(b) $gF \geqslant 23 : \widehat{\boldsymbol{\Sigma}}_{\mathrm{A}}$ (72.93% zeros)   (d) $gF \geqslant 23 : \widehat{\boldsymbol{\Sigma}}_{\mathrm{A}}$ (70% zeros)   (f) $gF \geqslant 23 : \widehat{\boldsymbol{\Sigma}}_{\mathrm{A}}$ (85% zeros)

Figure 1. Estimated sparsity structures in $\widehat{\boldsymbol{\Sigma}}_{\mathrm{A}}$ using soft functional thresholding rule at fluid intelligence $gF \leq 8$ and $gF \geq 23$: (a)–(b) with the corresponding $\hat{\lambda}$ selected by 5-fold cross-validation; (c)–(f) with the estimated functional sparsity levels set at 70% and 85%.
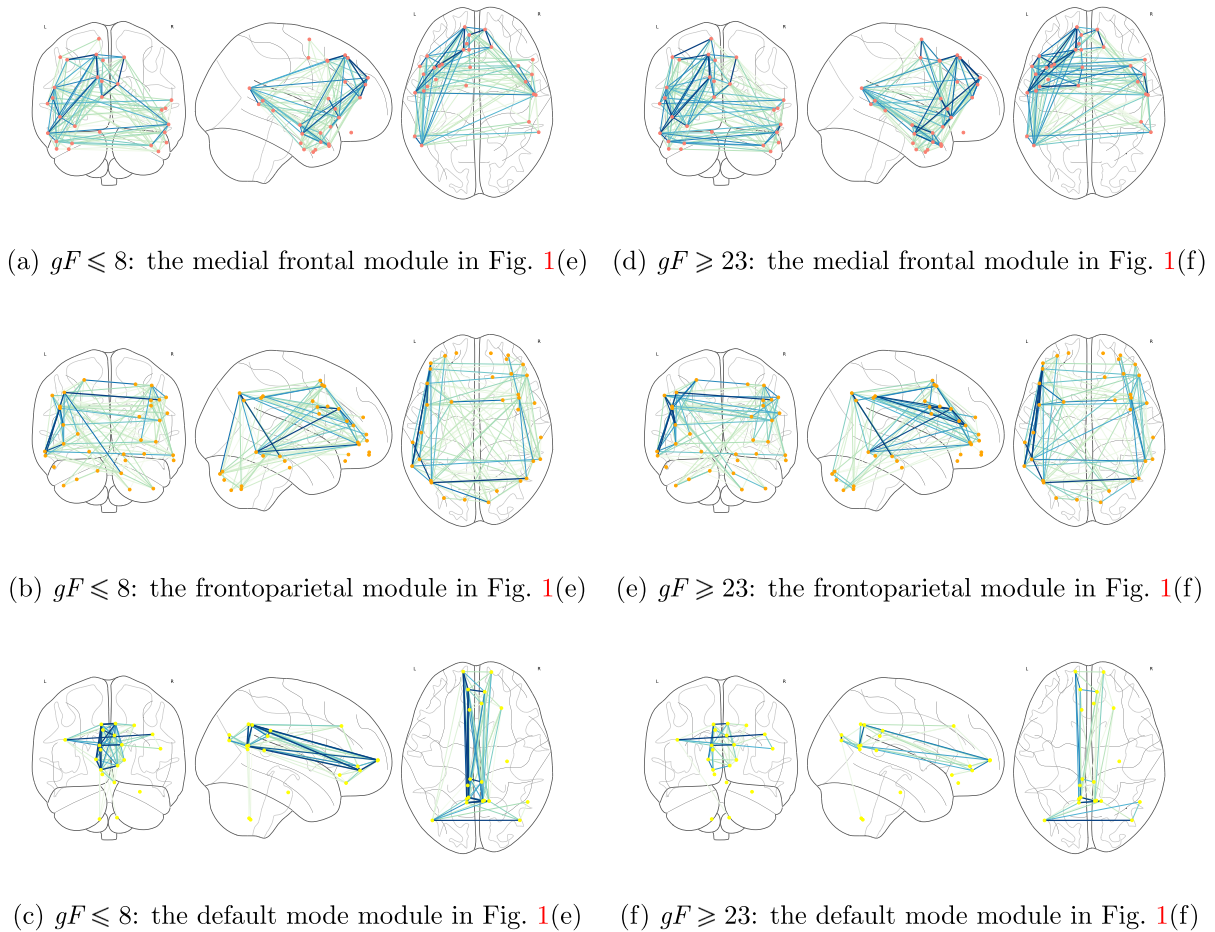

(a) $gF \leqslant 8$: the medial frontal module in Fig. 1(e)   (d) $gF \geqslant 23$: the medial frontal module in Fig. 1(f)

(b) $gF \leqslant 8$: the frontoparietal module in Fig. 1(e)   (e) $gF \geqslant 23$: the frontoparietal module in Fig. 1(f)

(c) $gF \leqslant 8$: the default mode module in Fig. 1(e)   (f) $gF \geqslant 23$: the default mode module in Fig. 1(f)

Figure 2. The connectivity strengths in Figure 1(e)–(f) at fluid intelligence $gF \leq 8$ and $gF \geq 23$. Salmon, orange and yellow nodes represent the ROIs in the medial frontal, frontoparietal and default mode modules, respectively. The edge color from cyan to blue corresponds to the value of $\|\widehat{\Sigma}_{jk}^{\mathrm{A}}\|_{\mathcal{S}} / (\|\widehat{\Sigma}_{jj}^{\mathrm{A}}\|_{\mathcal{S}} \|\widehat{\Sigma}_{kk}^{\mathrm{A}}\|_{\mathcal{S}})^{1/2}$ from small to large.

study (Finn et al. 2015): the medial frontal (29 ROIs), frontoparietal (34 ROIs) and default mode modules (20 ROIs). For each subject, the BOLD signals at each ROI are collected every 0.72 sec for a total of $L = 1200$ measurement locations (14.4 min). We first implement the ICA-FIX preprocessed pipeline (Glasser et al. 2013) and a standard band-pass filter at [0.01, 0.08] Hz to exclude frequency bands not implicated in resting state functional connectivity (Biswal et al. 1995). Figure 12 of the supplementary material displays exampified trajectories of pre-smoothed data. The adaptive functional thresholding method is then adopted to estimate the sparse covariance function and therefore the brain networks.

The sparsity structures in $\widehat{\mathbf{\Sigma}}_A$ for both groups are displayed in Figure 1. With $\widehat{\lambda}$ selected by the cross-validation, the network associated with $\widehat{\mathbf{\Sigma}}_A$ for subjects with $gF \geq 23$ is more densely connected than that with $gF \leq 8$, as evident from Figure 1(a)–(b). We further set the sparsity level to 70% and 85%, and present the corresponding sparsity patterns in Figure 1(c)–(f). The results clearly indicate the existence of three diagonal blocks under all sparsity levels, complying with the identification of the medial frontal, frontoparietal and default mode modules in Finn et al. (2015). We also implement the universal functional thresholding method. However, compared with $\widehat{\mathbf{\Sigma}}_A$, the results of $\widehat{\mathbf{\Sigma}}_U$ suffer from the heteroscedasticity, as demonstrated in Section 5 and Section E.3 of the supplementary material, and fail to detect any noticeable block structure, hence, we choose not to report them here. To explore the impact of gF on the functional connectivity, we compute the connectivity strength using the standardized form $\|\widehat{\Sigma}_{jk}^A\|_{\mathcal{S}}/(\|\widehat{\Sigma}_{jj}^A\|_{\mathcal{S}}\|\widehat{\Sigma}_{kk}^A\|_{\mathcal{S}})^{1/2}$ for $j, k = 1 \ldots, p$. Interestingly, we observe from Figure 2 that subjects with $gF \geq 23$ tend to have enhanced brain connectivity in the medial frontal and frontoparietal modules, while the connectivity strength in the default mode module declines. This agrees with existing neuroscience literature reporting a strong positive association between intelligence score and the medial frontal/frontoparietal functional connectivity in the resting state (Van Den Heuvel et al. 2009; Finn et al. 2015), and lends support to the conclusion that lower default mode module activity is associated with better cognitive performance (Anticevic et al. 2012). See also Section E.3 of the supplementary material, in which we illustrate our adaptive functional thresholding estimation using another ADHD dataset.

Supplementary Materials

The supplementary materials contain all the technical proofs, further methodological derivations and additional discussion and empirical results. We also provide the codes and datasets in Sections 5 and 6 in the supplementary materials.

Acknowledgments

We are grateful to the editor, the associate editor and two referees for their insightful comments and suggestions, which have led to significant improvement of our article.

Disclosure Statement

The authors report there are no competing interests to declare.

References

Anticevic, A., Cole, M. W., Murray, J. D., Corlett, P. R., Wang, X.-J., and Krystal, J. H. (2012), "The Role of Default Network Deactivation in Cognition and Disease," *Trends in Cognitive Sciences*, 16, 584–592. [12]

Avella-Medina, M., Battey, H. S., Fan, J., and Li, Q. (2018), "Robust estimation of High-Dimensional Covariance and Precision Matrices," *Biometrika*, 105, 271–284. [1]

Bickel, P. J., and Levina, E. (2008), "Covariance Regularization by Thresholding," *The Annals of Statistics*, 36, 2577–2604. [1,8]

Biswal, B., Zerrin Yetkin, F., Haughton, V. M., and Hyde, J. S. (1995), "Functional Connectivity in the Motor Cortex of Resting Human Brain Using Echo-Planar MRI," *Magnetic Resonance in Medicine*, 34, 537–541. [12]

Cai, T., and Liu, W. (2011), "Adaptive Thresholding for Sparse Covariance Matrix Estimation," *Journal of the American Statistical Association*, 106, 672–684. [1,2,4,5]

Cattell, R. B. (1987), *Intelligence: Its Structure, Growth and Action*, Amsterdam: Elsevier. [10]

Chang, C., and Glover, G. H. (2010), "Time–Frequency Dynamics of Resting-State Brain Connectivity Measured with fMRI," *Neuroimage*, 50, 81–98. [2]

Chen, Z., and Leng, C. (2016), "Dynamic Covariance Models," *Journal of the American Statistical Association*, 111, 1196–1207. [1]

Chiou, J.-M., Yang, Y.-F., and Chen, Y.-T. (2016), "Multivariate Functional Linear Regression and Prediction," *Journal of Multivariate Analysis*, 146, 301–312. [1,3,4]

Fan, J., and Li, R. (2001), "Variable Selection via Nonconcave Penalized Likelihood and its Oracle Properties," *Journal of the American Statistical Association*, 96, 1348–1360. [3]

Fan, J., and Marron, J. S. (1994), "Fast Implementations of Nonparametric Curve Estimators," *Journal of Computational and Graphical Statistics*, 3, 35–56. [2,7]

Finn, E. S., Shen, X., Scheinost, D., Rosenberg, M. D., Huang, J., Chun, M. M., Papademetris, X., and Constable, R. T. (2015), "Functional Connectome Fingerprinting: Identifying Individuals Using Patterns of Brain Connectivity," *Nature Neuroscience*, 18, 1664–1671. [12]

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, 105–124. [12]

Guo, S., Qiao, X., and Wang, Q. (2022), "Factor Modelling for High-Dimensional Functional Time Series," arXiv:2112.13651v2 . [1,4]

Happ, C., and Greven, S. (2018), "Multivariate Functional Principal Component Analysis for Data Observed on Different (dimensional) Domains," *Journal of the American Statistical Association*, 113, 649–659. [1,3]

Kong, D., Xue, K., Yao, F., and Zhang, H. H. (2016), "Partially Functional Linear Regression in High Dimensions," *Biometrika*, 103, 147–159. [1]

Kosorok, M. R. (2008), *Introduction to Empirical Processes and Semiparametric Inference*, Springer Series in Statistics, New York: Springer. [4]

Lee, K.-Y., Ji, D., Li, L., Constable, T., and Zhao, H. (in press), "Conditional Functional Graphical Models," *Journal of the American Statistical Association*, DOI: 10.1080/01621459.2021.1924178. [2,10]

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*, 113, 1637–1655. [1,2]

Lotte, F., Bougrain, L., Cichocki, A., Clerc, M., Congedo, M., Rakotomamonjy, A., and Yger, F. (2018), "A Review of Classification Algorithms for EEG-based Brain–Computer Interfaces: A 10 Year Update," *Journal of Neural Engineering*, 15, 031005. [2]

Miao, R., Zhang, X., and Wong, R. K. (in press), "A Wavelet-based Independence Test for Functional Data with an Application to MEG Functional Connectivity," *Journal of the American Statistical Association*, DOI: 10.1080/01621459.2021.2020126. [10]

Park, J., Ahn, J., and Jeon, Y. (2021), "Sparse Functional Linear Discriminant Analysis, *Biometrika*, 109, 209–226. [1]

Qiao, X., Guo, S., and James, G. (2019), "Functional Graphical Models," *Journal of the American Statistical Association*, 114, 211–222. [1,2,4]

Qiao, X., Qian, C., James, G. M., and Guo, S. (2020), "Doubly Functional Graphical Models in High Dimensions," *Biometrika*, 107, 415–431. [5,6,7]

Rogers, B. P., Morgan, V. L., Newton, A. T., and Gore, J. C. (2007), "Assessing Functional Connectivity in the Human Brain by fMRI," *Magnetic Resonance Imaging*, 25, 1347–1357. [2]

Rothman, A. J., Levina, E., and Zhu, J. (2009), "Generalized Thresholding of Large Covariance Matrices," *Journal of the American Statistical Association*, 104, 177–186. [1,3,5]

Storey, J. D., Xiao, W., Leek, J. T., Tompkins, R. G., and Davis, R. W. (2005), "Significance Analysis of Time Course Microarray Experiments," *Proceedings of the National Academy of Sciences*, 102, 12837–12842. [1]

Van Den Heuvel, M. P., Stam, C. J., Kahn, R. S., and Pol, H. E. H. (2009), "Efficiency of Functional Brain Networks and Intellectual Performance," *Journal of Neuroscience*, 29, 7619–7624. [12]

Vu, V. Q., and Lei, J. (2013), "Minimax Sparse Principal Subspace Estimation in High Dimensions," *The Annals of Statistics*, 41, 2905–2947. [4]

Wang, H., Peng, B., Li, D., and Leng, C. (2021), "Nonparametric Estimation of Large Covariance Matrices with Conditional Sparsity," *Journal of Econometrics*, 223, 53–72. [1]

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. [6]

Yuan, M., and Lin, Y. (2006), "Model Selection and Estimation in Regression with Grouped Variables," *Journal of the Royal Statistical Society*, Series B, 68, 49–67. [3]

Zapata, J., Oh, S. Y., and Petersen, A. (2022), "Partial Separability and Functional Graphical Models for Multivariate Gaussian Processes, *Biometrika*, 109, 665–681. [2,10]

Zhang, J.-T., and Chen, J. (2007), "Statistical Inferences for Functional Data," *The Annals of Statistics*, 35, 1052–1079. [6]

Zhang, X., and Wang, J.-L. (2016), "From Sparse to Dense Functional Data and Beyond," *The Annals of Statistics*, 44, 2281–2321. [5,6,7,9]

Zou, H. (2006), "The Adaptive Lasso and Its Oracle Properties," *Journal of the American Statistical Association*, 101, 1418–1429. [3]