Sparse Functional Principal Component Analysis in a New Regression Framework

Yunlong Nie[a], Jiguo Cao[a,*]

[a] *Department of Statistics and Actuarial Science,*
Simon Fraser University

Abstract

The functional principal component analysis is widely used to explore major sources of variation in a sample of random curves. These major sources of variation are represented by functional principal components (FPCs). The FPCs from the conventional FPCA method are often nonzero in the whole domain, and are hard to interpret in practice. The main focus is to estimate functional principal components (FPCs), which are only nonzero in subregions and are called sparse FPCs. These sparse FPCs not only represent the major variance resources but also can be used to identify the subregions where those major variations exist. The current methods obtain sparse FPCs by adding a penalty term on the length of nonzero regions of FPCs in the conventional eigendecomposition framework. However, these methods become an NP-hard optimization problem. To overcome this issue, a novel regression framework is proposed to estimate FPCs and the corresponding optimization is not NP-hard. The FPCs estimated with the proposed sparse FPCA method is shown to be equivalent to the FPCs using the conventional FPCA method when the sparsity parameter is zero. Simulation studies illustrate that the proposed sparse FPCA method can provide more accurate estimates for FPCs than other available methods when those FPCs are only nonzero in subregions. The proposed method is demonstrated by exploring the major variations among the acceleration rate curves of 107 diesel trucks, where the nonzero regions of the estimated sparse FPCs are found well separated.

Keywords: Dimension Reduction, Eigendecomposition, Empirical Basis Approximation, Functional Data Analysis

[*] Corresponding Author. Postal Address: 8888 University Dr, Burnaby, BC, Canada, V5A1S6. Tel:(+1)778-782-7600; Fax: (+1)778-782-4368; Email: jiguo_cao@sfu.ca

[1] An R package "sparseFPCA" is developed to implement the proposed method. The computing scripts for the simulation study can be downloaded at https://github.com/caojiguo/sparseFPCA.

1. Introduction

Functional principal component analysis (FPCA) is a crucial dimension reduction tool in functional data analysis. FPCA explores major sources of variability in a sample of random curves by finding functional principal components (FPCs) that maximize the curve variation. Consequently, the top few FPCs explain most of the variability in the random curves. Besides, each random curve can be approximated by a linear combination of the top few FPCs. Therefore, the infinite-dimensional curves are projected to a low-dimensional space defined by the top FPCs. This powerful dimensional reduction feature also promotes the popularity of FPCA.

The theoretical properties of FPCA have been carefully studied at length. For example, (Dauxois et al., 1982) first studied the asymptotic properties of PCA estimators for the infinite-dimensional data from a linear operator viewpoint. Following this point of view, Mas (2002); Bosq (2000) utilized functional analysis to study FPCA theoretically. On the other hand, Hall and Horowitz (2007); Hall et al. (2006); Yao et al. (2005) studied FPCA from the kernel perspective. Sang et al. (2017) proposed a parametric approach for estimating FPCs to enhance their interpretability for users. Nie et al. (2018) propose a supervised version of FPCA by considering the correlation of the functional predictor and response variable. In addition, FPCA has been widely and successfully applied in many applications such as functional linear regression (Yao et al., 2005), classification and clustering of functional data (Ramsay and Silverman (2005); Yao et al. (2005); Müller (2005); Müller and Stadtmüller (2005); Peng and Müller (2008); Dong et al. (2018)). All these applications assume the functional data are densely and regularly observed. When it comes to sparse and irregularly observed data, (Yao et al., 2005) proposed to estimate the FPC score using conditional expectation, which allows recovering the individual trajectory by borrowing information across all the subjects. The smooth version of functional principal component analysis is carefully studied by Rice and Silverman (1991); Pezzulli (1993); Silverman (1996), and Yao et al. (2005). There are mainly three methods to achieve smoothness. The first method smooths the functional data in the first step and conducts the regular FPCA on the sample covariance function. The second method smooths the covariance function first and then eigendecomposes the resulting smoothed covariance function to estimate the smoothed FPCs. The last method directly adds a roughness penalty in the optimization criterion for estimating the FPCs.

The conventional FPCA aims to estimate FPCs that maximize the curve variation. These FPCs represent the source or direction of maximum variations among curves, and the curves are projected to the low-dimensional space defined by these FPCs. Therefore, it is essential to interpret them. However, these FPCs are usually nonzero in the whole observed domain, and users often find it hard to interpret these FPCs. On the other hand, if the estimated FPC is only nonzero in a subregion of the entire domain, we can easily use them to identify the subregions from which the major variation of the curves exhibits. In this paper, our goal is to propose a method to estimate the sparse functional principal components, which are only nonzero in a subregion and, at the same time, account for an almost maximum amount of variation within the curves.

Several methods have been proposed to enhance the interpretability of functional principal components. The first method is the interpretable functional principal components analysis (iFPCA) proposed by Lin et al. (2016). This method adds an ℓ_0-penalty on the length of the nonzero region of FPCs and obtains FPCs, which are only nonzero in subregions. However, the optimization in their framework is an NP-hard problem because of the use of the ℓ_0-penalty. A greedy backward elimination algorithm is proposed to solve this optimization problem approximately. The second method is called a localized functional principal components analysis (LFPCA) method proposed by Chen and Lei (2015). This method adds an ℓ_1 penalty to the original eigendecomposition problem of smoothed FPCs, which is also not a convex optimization problem. They approximate this non-convex problem through a Deflated Fantope Localization method and propose a novel estimation procedure in a sequential manner. In addition, Di et al. (2014) considered the functional principal component analysis on sparsely sampled multilevel functional data. The sparsity in their work refers to the situations when the functional data are not fully observed rather than the sparsity of the FPCs. Li et al. (2016) studied the problem when the low-rank structure of the functional data was related to multivariate supervision data. The resulting supervised FPCs incorporate the information carried within the response data. In comparison, our work needs no supervision information and assumes the underlying FPCs are sparse on their own.

This paper has three major contributions. Firstly, we propose a new regression-type framework for the sparse functional principal component analysis. The estimated sparse FPCs can not only account for a reasonable variation within the functional data but also be sparse on the whole domain. We also show that the FPCs estimated with our proposed sparse FPCA method is equivalent to the FPCs with the conventional FPCA method when the sparsity parameter is zero. Secondly, our approach is not an NP-hard optimization problem, and the computation is very efficient. Lastly, our method estimates the top sparse FPCs simultaneously rather than sequentially estimating each FPC. Sequentially estimating the FPCs often leads to a quadratic optimization problem with multiple linear constraints. The numerical complexity increases as the rank of FPCs increases. Besides, the sequential manner does not allow parallel computing because the Kth FPC can only be estimated after obtaining the first $K-1$ FPCs. In our regression framework, the regression step of our algorithm only involves individual FPC such that it can be solved in a parallel way. An R package "sparseFPCA" is developed to implement our proposed sparse FPC (SFPCA) method. The computing scripts for our simulation study can be downloaded at `https://github.com/caojiguo/sparseFPCA`.

The rest of the paper is organized as follows. In Section 2, we introduce our SFPCA method and show its connection with the conventional FPCA. Details of our approach and the computation algorithm are described in Section 3, followed by theoretical results in Section 4. In Section 5, we apply our proposed method in a real-data application to explore major sources of variation among the acceleration rates of 107 diesel trucks. In Section 6, two carefully-designed simulations are conducted to evaluate the finite sample performance of our proposed method in comparison with other alternative methods in different settings. Section 7 provides concluding remarks.

2. Sparse Functional Principal Component Analysis

Consider a stochastic process X, which is square-integrable on the compact domain $\mathcal{T}$. In other words, $X \in L^2(\mathcal{T})$ almost surely, where $L^2(\mathcal{T})$ is the Hilbert space of square-integrable functions on $\mathcal{T}$. We denote the inner product between two functions $f, g \in L^2(\mathcal{T})$ as $\langle f, g \rangle = \int_{\mathcal{T}} f(t)g(t)dt$ with the corresponding norm as $||f|| = \sqrt{\langle f, f \rangle}$. If $\langle f, g \rangle = 0$, we also use the notation $f \perp g$. Let $x_i, i = 1, \ldots, n$, be the observed functional data for the stochastic process X. The rest of the paper assumes that the functional data are fully observed. When the functional data is not fully observed, we recommend to estimate the functional data first and then use the proposed method to estimate the sparse eigenfunctions based on the estimated functional data. When the data are densely observed, the smoothing spline method can be used to estimate the curves from the noisy data. When the data are only sparsely observed, the PACE method is recommended to estimate the curves. Without loss of generality, we assume that $E(X) = 0$ in the rest of this paper. In practice, users can always center the observed functional data first to remove this assumption.

We propose to estimate the first J leading unnormalized sparse FPCs $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_J)^T$ by minimizing the following criterion:

$$\frac{1}{n} \sum_{i=1}^{n} \left|\left| x_i - \sum_{j=1}^{J} \alpha_j \langle \beta_j, x_i \rangle \right|\right|^2 + \text{PEN}(\boldsymbol{\beta}) + \tau \sum_{j=1}^{J} \int \beta_j^2(t)dt, \qquad (1)$$

with respect to β_j and the ancillary parameter α_j, $j = 1, \ldots, J$, with the constraints $||\alpha_j||^2 = 1$ and $\langle \alpha_\ell, \alpha_j \rangle = 0$ when $\ell \neq j$. Introducing ancillary parameter α_j helps to remove the orthogonal constraint on β_j. Here, $\text{PEN}(\boldsymbol{\beta})$ in (1) consists of a sparsity penalty denoted by $\text{SPEN}_\lambda(\boldsymbol{\beta})$ and a roughness penalty term on $\boldsymbol{\beta}$ denoted by $\text{RPEN}_\gamma(\boldsymbol{\beta})$:

$$\text{PEN}(\boldsymbol{\beta}) = \text{SPEN}_\lambda(\boldsymbol{\beta}) + \text{RPEN}_\gamma(\boldsymbol{\beta}). \qquad (2)$$

The sparsity term penalizes the length of nonzero regions of each β and the roughness part prevents the estimated FPCs being too ‘wiggle’. When the penalty term $\text{PEN}(\boldsymbol{\beta})$ becomes zero, we will show in Section 2.2 that minimizing the first two terms in (1) leads to the conventional FPCs. In other words, the resulting FPC will be the optimal basis functions in explaining or recovering observed functional data $\{x_i(t), i = 1, \ldots, n\}$. In fact, when PEN($\boldsymbol{\beta}$) becomes zero, $\alpha_j(t)$ is equal to the conventional j-th FPC, and $\alpha_j = \beta_j/||\beta_j||$. Hence, $\langle \beta_j, x_i \rangle/||\beta_j||$ will be the corresponding conventional FPC score for x_i. On the other hand, with the sparsity penalty, our criterion (1) not only considers the resulting FPC's ability to explain the maximum variation among the functional data, but also takes the sparsity of the FPCs into account. The term $\sum_{j=1}^{J} \int \beta_j^2(t)dt$ forces β_j to stay in the space spanned by $\{x_1, \ldots, x_n\}$, or in some sense, ensures the identifiability of $\beta_j(t)$. After obtaining the estimate for the first J leading unnormalized sparse FPCs $\{\hat{\beta}_j, j = 1, \ldots, J\}$, we normalize each $\hat{\beta}_j$ to obtain the normalized sparse FPCs $\hat{\xi}_j = \widehat{\beta}_j/||\widehat{\beta}_j||$.

2.1. Sparsity Penalty

The sparse penalty term SPEN$_\lambda(\boldsymbol{\beta})$ in (2) penalizes the length of nonzero regions of $\boldsymbol{\beta}$. The functional SCAD method proposed by Lin et al. (2017) is a functional generalization of the SCAD method (Fan and Li, 2001). The functional SCAD method is used in Lin et al. (2017) to find a locally sparse estimator for the coefficient function in functional linear regression models. The nice shrinkage property of functional SCAD allows the proposed estimator to locate null subregions of the coefficient function without over shrinking nonzero values of the coefficient functions.

In this article, we employ the functional SCAD penalty to achieve a locally sparse estimator of FPCs by defining:

$$\mathrm{SPEN}_\lambda(\boldsymbol{\beta}) = \sum_{j=1}^{J} \int p_{\lambda_j}(|\beta_j(t)|)dt,$$

in which $p_\lambda(\cdot)$ is the SCAD function defined in Fan and Li (2001):

$$p_\lambda(u) = \begin{cases} \lambda u & \text{if } 0 \le u \le \lambda, \\ -\frac{u^2 - 2a\lambda u + \lambda^2}{2(a-1)} & \text{if } \lambda < u < a\lambda, \\ \frac{(a+1)\lambda^2}{2} & \text{if } u \ge a\lambda, \end{cases}$$

where a is 3.7, as suggested by Fan and Li (2001), and λ is the tuning parameter. A large value of the tuning parameter λ_j will penalize the nonzero region of the corresponding β_j, hence leading to a sparse estimation. On the other hand, when the sparse parameter, λ_j, is zero, the resulting β_j reduces to the conventional functional principal components. Before showing the details of estimating the sparse FPCs given the tuning parameters, we first show that the FPCs estimated with our proposed sparse FPCA method is equivalent to the FPCs with the conventional FPCA method when the penalty term $\mathrm{PEN}(\boldsymbol{\beta})$ in (1) is zero.

2.2. Connection to the Conventional FPCA

The conventional FPCA method estimates the top FPCs with the eigendecomposition method. We can show that the j-th FPC $\phi_j(t)$ is the j-th eigenfunction of the covariance function $G(s,t) = \mathrm{E}(X(s)X(t))$, and satisfies the following eigenequation:

$$\int G(s,t)\phi_j(s)ds = \lambda_j \phi_j(t), \tag{3}$$

where λ_j is the corresponding eigenvalue and $\lambda_1 \geq \lambda_2 \geq \ldots \geq 0$. The conventional method estimates the FPCs by solving the above eigenequation (3), in which the covariance function $G(s,t)$ is replaced by the empirical covariance function $g(s,t) = \frac{1}{n}\sum_{i=1}^{n} x_i(s)x_i(t)$. The FPC score s_{ij} can be calculated as $s_{ij} = \langle x_i, \phi_j \rangle$. The FPC score s_{ij} has mean 0 and variance λ_j. One widely-used strategy to determine the number of FPCs is to choose a value such that the first J leading FPCs account for more than 90% of the total variation:

$$J = \inf\left\{k : \frac{\sum_{j=1}^{k}\lambda_j}{\sum_{j=1}^{\infty}\lambda_j} \geq 90\%\right\}.$$

Another conventional way to understand FPCs is through the Karhunen-Loève(KL) expansion (Fukunaga and Koontz, 1970). More specifically, according to the KL expansion, $x_i(t)$ can be expressed as

$$x_i(t) = \sum_{j=1}^{\infty} s_{ij}\phi_j(t), \quad i = 1, \ldots, n, \tag{4}$$

in which $\langle \phi_i, \phi_j \rangle = \delta_{ij}$, and δ_{ij} is the Kronecker's delta. A major advantage of FPCA is that by projecting each $x_i(t)$ onto orthogonal FPCs with uncorrelated scores, it allows us to approximate each $x_i(t)$ using the first J leading FPCs:

$$x_i(t) \approx \sum_{j=1}^{J} s_{ij}\phi_j(t), \quad i = 1, \dots, n. \tag{5}$$

In fact, there are many other basis functions on which $x_i(t)$ can be projected. However, the FPCs obtained from eigendecomposing the empirical covariance function are the optimal basis functions in the sense that they minimize the squared approximation errors (see Tran (2008)). Formally speaking, for any fixed $K \in \mathbb{N}$, the first J FPCs, $\{\phi_j, j = 1, \dots, J\}$, satisfy

$$\{\phi_j, j = 1, \dots, J\} = \underset{\langle \phi_\ell, \phi_j \rangle = \delta_{\ell j}}{\arg\min} \sum_{i=1}^{n} \left|\left| x_i(t) - \sum_{j=1}^{J} \langle x_i, \phi_j \rangle \phi_j \right|\right|^2.$$

This 'best-approximation' point of view inspires us to estimate the FPCs by searching for the optimal basis functions to approximate $x_i(t), i = 1, \dots, n$.

We will show that the first empirical leading FPC is the solution of a least square optimization with some constraints. Formally,

Proposition 1. *For any $\tau > 0$, let*

$$(\hat{\beta}, \hat{\alpha}) = \arg\min \frac{1}{n} \sum_{i=1}^{n} \left|\left| x_i - \alpha \langle \beta, x_i \rangle \right|\right|^2 + \tau \int \beta^2(t) dt, \tag{6}$$

subject to $||\alpha||^2 = 1$, then $\hat{\alpha} = \hat{\phi}_1$ and $\hat{\beta} = c\hat{\phi}_1$, where $\hat{\phi}_1$ is the first empirical eigenfunctions of the sample covariance function $g(s,t) = \frac{1}{n}\sum_{i=1}^{n} x_i(s)x_i(t)$ and c is a constant scale factor.

The ridge type penalty term essentially ensures the identifiablity of the primary parameter (i.e., β_j). The reason can be explained by the following example. We assume that the observed functional data $x(t)$ can be expressed by $K \leq n$ eigenfunctions:

$$x_i(t) = \sum_{k=1}^{K} \alpha_{ik}\phi_k(t).$$

Then there always exists a fixed function $\eta(t)$ such as $\eta(t) \perp span(\phi_1, \dots, \phi_K)$. Now if the ridge penalty term is removed in equation (6), the objective function in Proposition 1 becomes

$$Q(\beta) = \frac{1}{n} \sum_{i=1}^{n} \left|\left| x_i - \alpha \langle \beta, x_i \rangle \right|\right|^2,$$

subject to $||\alpha||^2 = 1$. However, this objective function doesn't result in a single minimizer. For example, given $\alpha(t)$, for any $\beta(t) = \sum_{k=1}^{K} b_k\phi_k(t)$, there always exists a corresponding $\beta'(t) = \beta(t)+\eta(t)$ such that these two objective functions are equal. On the other hand, with $\tau > 0$, the new objective function becomes

$$Q_\tau(\beta) = \frac{1}{n}\sum_{i=1}^{n} \left|\left|x_i - \alpha\langle\beta, x_i\rangle\right|\right|^2 + \tau \int \beta^2(t)dt.$$

It guarantees that, for any $\eta(t)$ such that $\eta \perp span(\phi_1, \ldots, \phi_K)$ and for any $\beta(t) = \sum_{k=1}^{K} b_k\phi_k(t)$, if $\beta'(t) = \beta(t) + \eta(t)$, then $Q_\tau(\beta') > Q_\tau(\beta)$. Therefore, it guarantees that the minimizor of $Q_\tau(\beta)$ is identifiable.

Similarly, for the first J leading FPCs, let $\hat{\beta}_1, \ldots, \hat{\beta}_J$ be the solution of minimizing

$$\frac{1}{n}\sum_{i=1}^{n} \left|\left|x_i - \sum_{j=1}^{J} \alpha_j\langle\beta_j, x_i\rangle\right|\right|^2 + \tau\sum_{j=1}^{J}\int \beta_j^2(t)dt,$$

with respect to $\{\beta_j, j = 1, \ldots, J\}$ and ancillary parameter $\{\alpha_j, j = 1, \ldots, J\}$ with the constaints $\langle\alpha_i, \alpha_j\rangle = \delta_{ij}$. Then $\hat{\beta}_j = c_j\hat{\phi}_j$, where $\hat{\phi}_j$ is the jth FPC and c_j is a scale factor. Formally,

Proposition 2. *Let $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_J)$ and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_J)$. For any $\tau > 0$, let*

$$(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}}) = \arg\min \frac{1}{n}\sum_{i=1}^{n} ||x_i - \sum_{j=1}^{J} \alpha_j\langle\beta_j, x_i\rangle||^2 + \tau\sum_{j=1}^{J}\int \beta_j^2(t)dt,$$

subject to $\langle\alpha_i, \alpha_j\rangle = \delta_{ij}$ and δ_{ij} is the Kronecker delta, then $\hat{\beta}_j = c_j\hat{\phi}_j, j = 1, \ldots, J$, where $\hat{\phi}_j$ is the j-th empirical eigenfunctions of the sample covariance function $g(s,t) = \frac{1}{n}\sum_{i=1}^{n} x_i(s)x_i(t)$ and c_j is a scale factor.

Proposition 2 shows that when the penalty term PEN($\boldsymbol{\beta}$) in (1) is zero, the corresponding estimated sparse FPC $\hat{\xi}_j$ is equivalent to the conventional FPC $\hat{\phi}_j$. Therefore, the FPCs estimated with our proposed sparse FPCA method is equivalent to the FPCs with the conventional FPCA method when penalty term is zero. The detailed proofs for these two propositions are in the supplementary file.

3. Estimation Method

We propose to estimate the first J unnormalized sparse FPCs, $\beta_1, \ldots, \beta_J$, in an iterative optimization method. More specifically, within each iteration, the first step is to find the optimal β_j that minimizes the criterion (1) given the current estimate of α_j and the second step is to search for a new α_j which further minimizes the proposed criterion conditional on the optimal β_j from the first step. This procedure is repeated until it converges. In the rest of this section, we first give the details of these two steps. Then we discuss the tuning parameter selection and the adjusted variance explained in the end.

3.1. Estimate β_j for Given α_j

Given the jth α_j, the corresponding β_j is obtained by minimizing

$$Q_{\tau,\lambda,\gamma}(\beta_j) = \frac{1}{n}\sum_{i=1}^{n} \left|\left| \langle x_i, \alpha_j\rangle - \langle \beta_j, x_i\rangle \right|\right|^2 \\ + \tau \int \beta_j^2(t)dt + \int p_\lambda(|\beta_j(t)|)dt + \gamma \int \left[\frac{d^2\beta_j(t)}{dt^2}\right]^2 dt,$$

in which the last term $\gamma \int \left[\frac{d^2\beta_j(t)}{dt^2}\right]^2 dt$ represents the roughness penalty RPEN_γ on β_j. A larger value of γ will prevent the estimated SFPC from being too 'wiggly'. Without any parametric assumption on β_j, we first represent β_j as a linear combination of basis functions

$$\beta_j(t) = \sum_{m=1}^{M} b_{jm}\psi_m(t) = \boldsymbol{\psi}(t)^T \mathbf{b}_j, \tag{7}$$

where $\boldsymbol{\psi}(t) = (\psi_1(t), \psi_2(t), \ldots, \psi_M(t))^T$ denotes the vector of B-spline basis functions, $\mathbf{b}_j$ is the corresponding vector of basis coefficients, and M denotes the number of basis functions.

For simplicity, we recast each part in $Q_{\tau,\lambda,\gamma}(\beta_j)$ into a matrix form. Let $\mathbf{a}_j = (a_{1j}, a_{2j}, \ldots, a_{nj})^T$ with $a_{ij} = \int x_i(t)\alpha_j(t)dt$, then the first term in the loss function can be expressed as

$$\sum_{i=1}^{n} ||\langle x_i, \alpha_j\rangle - \langle \beta_j, x_i\rangle||^2 = (\mathbf{a}_j - \mathbf{Z}\mathbf{b}_j)^T(\mathbf{a}_j - \mathbf{Z}\mathbf{b}_j). \tag{8}$$

Here $\mathbf{Z}$ is an $n \times M$ matrix with entries $z_{ij} = \int x_i(t)\psi_j(t)dt$ for $1 \le i \le n$ and $1 \le j \le M$. The second term in the loss function can be expressed as

$$\tau \int \beta^2(t)dt = \tau \mathbf{b}_j^T \mathbf{\Psi} \mathbf{b}_j, \tag{9}$$

in which $\mathbf{\Psi}$ denotes an $M \times M$ matrix with entries $\mathbf{\Psi}_{ij} = \int \psi_i(t)\psi_j(t)dt$ for $1 \le i, j \le M$. The roughness penalty term in the loss function can be expressed as

$$\gamma \int \left[\frac{d^2\beta_j(t)}{dt^2}\right]^2 dt = \gamma \mathbf{b}_j^T \mathbf{R} \mathbf{b}_j, \tag{10}$$

in which $\mathbf{R}$ denotes an $M \times M$ matrix with the (i,j)-th entry $\int \psi_i''(t)\psi_j''(t)dt$ for $1 \le i, j \le M$.

The sparsity penalty term in the loss function, as shown in Lin et al. (2017), can be approximated as

$$\int p_\lambda(|\beta_j(t)|)dt \approx \frac{T}{M-d} \sum_{\ell=1}^{M-d} p_\lambda\left(\sqrt{\frac{M-d}{T} \int_{t_{\ell-1}}^{t_\ell} \beta_j^2(t)dt}\right),$$

in which $t_0, t_1, \ldots, t_{M-d}$ denote the sequence of the knots of B-spline basis functions $\boldsymbol{\psi}(t)$, and d denotes the order of the basis functions. We choose the B-spline basis because it enjoys the local compact support property. Note that when M is large, each B-spline basis function is only nonzero for no more than d consecutive subintervals. In other words, if consecutive coefficients of the B-spline basis functions are zero, the corresponding estimated $\beta(t)$ would become strictly zero in the subregions.

We further define

$$||\beta_{[\ell]}(t)||_2^2 \overset{def}{=} \int_{t_{\ell-1}}^{t_\ell} \beta_j^2(t)dt = \mathbf{b}_j^T \mathbf{\Psi}_j \mathbf{b}_j,$$

in which $\mathbf{\Psi}_j$ denotes an $M \times M$ matrix with the (p,q)-entry as $\int_{t_{j-1}}^{t_j} \psi_p(t)\psi_q(t)dt$ when $j \le p, q \le j + d$ and zero elsewhere. Using the local quadratic approximation (LQA) method proposed in Fan and Li (2001), given some initial estimate $\mathbf{b}_j^{(0)}$, we can derive that

$$\int p_\lambda(|\beta_j(t)|)dt \approx \frac{T}{M-d}\left[\mathbf{b}_j^T \boldsymbol{W}^{(0)} \mathbf{b}_j + G(\mathbf{b}_j^{(0)})\right], \tag{11}$$

where

$$\boldsymbol{W}^{(0)} = \frac{1}{2}\sum_{\ell=1}^{M-d}\left(\frac{p'_\lambda(||\beta_{[\ell]}(t)||_2\sqrt{M-d/T})}{||\beta_{[\ell]}(t)||_2\sqrt{T/M-d}}\boldsymbol{\Psi}_j\right),$$

and

$$G(\boldsymbol{\beta}^{(0)}) \equiv \sum_{\ell=1}^{M} p_\lambda\left(\frac{||\beta_{[\ell]}^{(0)}||_2}{\sqrt{T/M-d}}\right) - \frac{1}{2}\sum_{\ell=1}^{M} p'_\lambda\left(\frac{||\beta_{[\ell]}^{(0)}||_2}{\sqrt{T/M-d}}\right)\frac{||\beta_{[\ell]}^{(0)}||_2}{\sqrt{T/M-d}}.$$

Putting (8),(9),(10) and (11) together,

$$\begin{aligned} Q_{\tau,\lambda,\gamma}(\beta_j) =& \frac{1}{n}(\mathbf{a}-\mathbf{Z}\boldsymbol{\beta})^T(\mathbf{a}-\mathbf{Z}\boldsymbol{\beta}) + \tau\boldsymbol{\beta}^T\boldsymbol{\Psi}\boldsymbol{\beta} + \gamma\boldsymbol{\beta}^T\mathbf{R}\boldsymbol{\beta} + \\ & \frac{T}{M-d}\boldsymbol{\beta}^T\boldsymbol{W}^{(0)}\boldsymbol{\beta} + \frac{T}{M-d}G(\boldsymbol{\beta}^{(0)}). \end{aligned}$$

By minimizing $Q_{\tau,\lambda,\gamma}(\beta_j)$, we obtain the estimate for the basis coefficients

$$\widehat{\boldsymbol{\beta}} = \left(\frac{1}{n}\mathbf{Z}^T\mathbf{Z} + \tau\boldsymbol{\Psi} + \gamma\mathbf{R} + \frac{T}{M-d}\mathbf{W}^{(0)}\right)^{-1}\mathbf{Z}^T\mathbf{a}.$$

Then we plug the estimate $\widehat{\boldsymbol{\beta}}$, into (7) to obtain the estimates for $\beta_j(t)$:

$$\widehat{\beta}(t) = \boldsymbol{\phi}(t)^T\widehat{\boldsymbol{\beta}}.$$

3.2. Estimate α_j for Given β_j

Let $\boldsymbol{\alpha}(t) = (\alpha_1, \ldots, \alpha_J)^T$, $\boldsymbol{\beta}(t) = (\beta_1, \ldots, \beta_J)^T$, and $u_{ij} = \langle \beta_j, x_i \rangle$. We obtain the estimate for $\boldsymbol{\alpha}(t)$ by minimizing

$$\begin{aligned} Q_{\boldsymbol{\beta}_j}(\alpha_1) &= \sum_{i=1}^{n}\left|\left|x_i(t) - \sum_{j=1}^{J}\alpha_j(t)u_{ij}\right|\right|^2 \\ &= \sum_{i=1}^{n}||x_i||^2 - \int 2\sum_{i=1}^{n}x_i(t)\left(\sum_{j=1}^{J}\alpha_j(t)u_{ij}\right)dt + \int\sum_{i=1}^{n}\left(\sum_{j=1}^{J}\alpha_j(t)u_{ij}\right)^2 dt. \end{aligned}$$

First, we can see that the first term is equivalent to the sum of the norm of each observed $x_i(t)$, which does not depend on the value of $\boldsymbol{\alpha}(t)$. Second, the last term can be recast into:

$$\int\sum_{i=1}^{n}\left(\sum_{j=1}^{J}\alpha_j(t)u_{ij}\right)^2 dt = \int\sum_{i=1}^{n}\sum_{j=1}^{J}\alpha_j^2(t)u_{ij}^2 + \sum_{i=1}^{n}\sum_{l<k}2\alpha_l(t)u_{il}\alpha_k(t)u_{ik}dt = \sum_{i=1}^{n}u_{ij}^2,$$

due to the fact that $\langle \alpha_i, \alpha_j \rangle = \delta_{ij}$. Thus, the last term does not depend on the value of $\boldsymbol{\alpha}(t)$, either. Therefore, minimizing $Q(\boldsymbol{\alpha})$ is equivalent to minimizing the second term. We can further recast the second term into the following form:

$$\begin{aligned}
\int \sum_{i=1}^{n} x_i(t) \Big(\sum_{j=1}^{J} \alpha_j(t) u_{ij} \Big) dt &= \int \bigg(\sum_{j=1}^{J} \alpha_j(t) \int \sum_{i=1}^{n} (x_i(t) x_i(s)) \beta_j(s) ds \bigg) dt \\
&= n \int \bigg(\sum_{j=1}^{J} \alpha_j(t) \int g(t,s) \beta_j(s) ds \bigg) dt \\
&= n \int \bigg(\sum_{j=1}^{J} \alpha_j(t) \sum_{k=1}^{K} \lambda_k \langle \beta_j, \phi_k \rangle \phi_k(t) \bigg) dt,
\end{aligned}$$

where $\phi_k, k = 1, \ldots, K$, denotes the empirical eigenfunctions obtained from decomposing the sample covariance function $g(t,s)$ as discussed in Equation (3) and λ_k is the corresponding eigenvalues. The last step in the above equation uses the fact that $g(t,s) = \sum_{k=1}^{K} \{\lambda_k \phi_k(t) \phi_k(s)\}$, where K is the number of nonzero eigenvalues of the sample covariance function, and $K \leq n$.

To simplify the notation, let $\xi_j(t) = \sum_{k=1}^{K} \lambda_k \langle \beta_j, \phi_k \rangle \phi_k(t)$, so that the above equation becomes

$$\int \sum_{i=1}^{n} x_i(t) \Big(\sum_{j=1}^{J} \alpha_j(t) u_{ij} \Big) dt = n \int \bigg(\sum_{j=1}^{J} \alpha_j(t) \xi_j(t) \bigg) dt. \tag{12}$$

Let $h_l, l = 1, \ldots, J$, denote the eigenfunctions of $\frac{1}{J} \sum_{j=1}^{J} \xi_j(s) \xi_j(t)$. To maximize (12) with respect to α_j, we first express $\{\xi_j, j = 1, \ldots, J\}$ using $h_l, l = 1, \ldots, J$:

$$\xi_j = \sum_{l=1}^{J} \rho_{jl} h_l, \tag{13}$$

in which $\rho_{jl} = \langle \xi_j, h_l \rangle$. Then we plug equation (13) back into equation (12):

$$\begin{aligned}
\int \sum_{i=1}^{n} x_i(t) \Big(\sum_{j=1}^{J} \alpha_j(t) u_{ij} \Big) dt &= n \int \bigg(\sum_{j=1}^{J} \alpha_j(t) \sum_{l=1}^{J} \rho_{jl} h_l(t) \bigg) dt \\
&= n \sum_{j=1}^{J} \sum_{l=1}^{J} \rho_{jl} \int \Big(\alpha_j(t) h_l(t) \Big) dt.
\end{aligned}$$

By the Cauchy-Swartz inequality, the above equation is maximized when

$$\int \alpha_j(t) h_l(t) dt \propto \rho_{jl}.$$

Note that $h_l, l = 1, \ldots, J$, can be viewed as the orthogonal basis functions that α_j is projected onto. Therefore, the solution should be given as

$$\hat{\alpha}_j = \frac{\sum_{l=1}^{J} \rho_{jl} h_l}{\sum_l \rho_{jl}^2}, j = 1, \ldots, J.$$

Putting into a matrix form, we have

$$\widehat{\boldsymbol{\alpha}}(t) = \mathbf{P}\mathbf{h}(t),$$

in which $\mathbf{P}$ is a $J \times J$ matrix with (j,l)-th element being $\frac{\rho_{jl}}{\sum_l \rho_{jl}^2}$. To check whether the resulting $\widehat{\boldsymbol{\alpha}}(t)$ satisfies the orthonormal condition, we can see that the coefficients matrix

$$\mathbf{P} = \begin{bmatrix} \frac{\rho_{11}}{\sum_l \rho_{1l}^2} & \cdots & \frac{\rho_{1J}}{\sum_l \rho_{Jl}^2} \\ \vdots & \cdots & \vdots \\ \frac{\rho_{J1}}{\sum_l \rho_{1l}^2} & \cdots & \frac{\rho_{JJ}}{\sum_l \rho_{Jl}^2} \end{bmatrix},$$

is an orthogonal matrix because $\mathbf{P}^T\mathbf{P} = \mathbf{I}$. Therefore, the resulting $\hat{\alpha}_j$ satisfy $\langle \hat{\alpha}_i, \hat{\alpha}_j \rangle = \delta_{ij}$.

3.3. Detailed Algorithms

Below we summarize the proposed estimation method step by step:

Step I: Initialize $\alpha_j^{(0)} = \hat{\phi}_j, j = 1, \ldots, J$, where $\hat{\phi}_j$ is the estimated FPC using the conventional FPCA method, which satisfy $||\hat{\phi}_j||^2 = 1$ and $\langle \hat{\phi}_i, \hat{\phi}_j \rangle = 0, i \neq j$;

Step II: Given $\alpha_j^{(i)}$, obtain the corresponding $\beta_j^{(i)}$ by minimizing

$$\begin{aligned} Q_{\tau,\lambda,\gamma}(\beta) = \frac{1}{n}(&\sum_{i=1}^{n} || \int x_i(t)\alpha^{(i)}(t)dt - \langle \beta, x_i \rangle||^2 \\ &+ \tau \int \beta^2(t)dt + \gamma \int \beta''^2(t)dt + \int p_\lambda(|\beta(t)|)dt. \end{aligned}$$

Due to the fact that $\langle \alpha_k^{(i)}, \alpha_l^{(j)} \rangle = 0$ for $k \neq l$, we can obtain each $\beta_j^{(i)}$ separately. The details of this step is discussed in Section 3.1.

Step III: Given $\beta_1^{(i)}, \ldots, \beta_J^{(i)}$, obtain the corresponding $\alpha_1^{(i+1)}, \ldots, \alpha_J^{(i+1)}$ by minimizing

$$Q_{\boldsymbol{\beta}_j}(\boldsymbol{\alpha}) = \frac{1}{n}\sum_{i=1}^{n} ||x_i - \sum_{j=1}^{J} \alpha_j \int \beta_j^{(i)}(t) x_i(t)||^2 + \text{Constant}.$$

The 'Constant' term represents $\sum_{j=1}^{J} \tau \int {\beta_j^{(i)}}^2(t)dt$ and the remaining constant terms when $\beta^{(i)}$ is given. Unlike Step II, we can obtain $\alpha_1^{(i+1)}, \ldots, \alpha_J^{(i+1)}$ simultaneously. The details of this step is provided in Section 3.2.

Step IV: Repeat Step II to Step III until they converge.

3.4. Choosing Tuning Parameters

There are three tuning parameters: the ridge-type parameter τ, the sparsity parameter λ, and the smoothing parameter γ. The ridge-type parameter τ is only required to be positive by Proposition 2. Our numerical study in the supplementary file suggests that the estimated sparse FPCs are almost the same as the value of τ varies. The sparsity parameter λ controls the sparsity of the estimated sparse FPCs $\widehat{\boldsymbol{\beta}}(t)$. Note that the more compact the resulting sparse FPCs are, the less variation they can explain for the original functional data in comparison to the conventional FPCs or equivalently the errors of approximating the original functional data become larger. We recommend choosing a value of λ that balances between sparsity and the errors of approximating the original functional data. In addition, the smoothing parameter γ prevents the resulting $\beta(t)$ from being too 'wiggly'. Again, a smoother $\boldsymbol{\beta}(t)$ explains less variation of the functional data and we recommend to choose a value that balances these two aspects. In practice, one can choose both the smoothing parameter and the sparsity parameter by cross-validation. More specifically, one can split the data into the training and test datasets. Then we can use the proposed sparse FPCA method to obtain the estimated sparse FPCs. Next, we can use the resulting sparse FPCs to recover the trajectory in the test set and compare it with what is observed to obtain the cross-validation errors. However, this two-dimensional cross-validation method might be computationally intensive in practice. We alternatively propose a two-step procedure to choose the tuning parameter. Our computational experience suggests this two-step procedure yields reasonable results.

In the first step, the smoothing parameter γ is chosen using cross-validation as described in Ramsay and Silverman (2005). In the second step, we introduce the following AIC criterion to select the sparsity parameter λ:

$$\mathrm{AIC} = n \log\left(\frac{\sum_{i=1}^{n} ||x_i - \hat{x}_i||^2}{nT}\right) + 2df(\lambda), \tag{14}$$

where x_i and $\hat{x}_i = \sum_{k=1}^{K} \langle x_i, \hat{\phi}_k \rangle \hat{\phi}_k$ represent the ith sample curve and its corresponding estimate using the estimated FPCs, respectively. The degree of freedom is the number of nonzero B-spline coefficients in the estimated FPCs under different values of the sparsity parameter λ. The goal is to balance the errors of approximating functional data and the length of the support regions of the estimated FPCs. We will demonstrate this procedure in both real data application and simulation studies.

3.5. Adjusted Total Variance Explained

Due to the fact that the sparse functional principal component scores are not necessarily uncorrelated, the variance explained by the jth SFPC is not simply the variance of the corresponding SFPC scores and need to take the correlation between SFPCs into account. Here we propose a new approach to compute the total variance explained by the jth SFPC. Let $\phi_j(t)$ and $\mathbf{s}_j$ denote the jth SFPC and the corresponding score vector. We regress $\mathbf{s}_j$ on $\mathbf{s}_1, \ldots, \mathbf{s}_{j-1}$ and denote the resulting residuals as $\mathbf{r}_j$. Then the adjusted variance explained by $\phi_j(t)$ is $||\mathbf{r}_j||^2$.

4. Application

Our proposed method is demonstrated by analyzing a real dataset relating to particulate matter (PM) emissions from diesel trucks (Clark et al., 2007). In the experiment, trucks are driven through a pre-determined driving cycle and PM at the exhaust pipe is measured every second via a particulate matter counter. Hall and Hooker (2016) analyzed this dataset to predict PM using the acceleration rate with a functional linear model. Figure 1 displays the acceleration rate curves for 107 diesel trucks. We will demonstrate our proposed sparse FPCA method by analyzing the major variations among these acceleration curves.

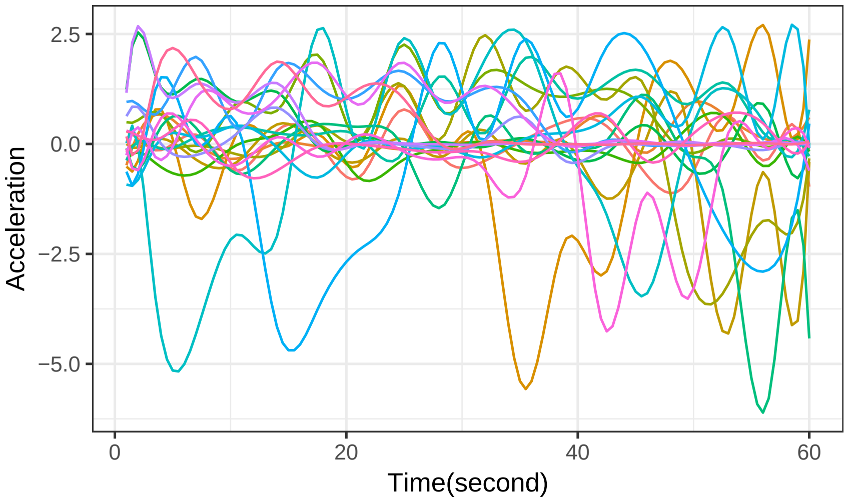

Figure 1: The observed acceleration rates for 20 randomly selected diesel trucks out of all 107 diesel trucks. Each curve respects one truck's observations.

We first applied the conventional FPCA method (Ramsay and Silverman, 2005) to analyze the major variations among these acceleration curves. The top four estimated FPCs are shown in Figure 2. They account for 25.7%, 24.6%, 17.4% and 15.3% of the total variation among the acceleration curves, respectively. In total, the first four FPCs explain 83.0% of the total variation. As expected, these estimated FPCs are nonzero on the entire time domain.

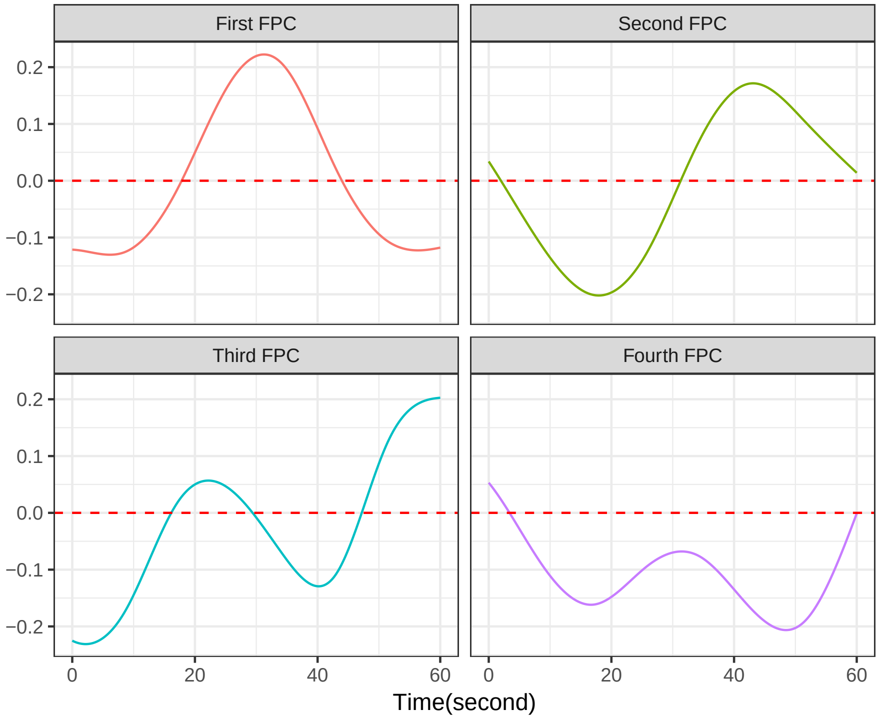

Figure 2: The estimated first four leading functional principal components using conventional FPCA for analyzing the acceleration curves. They account for 25.7%, 24.6%, 17.4% and 15.3% of the total variation among the acceleration curves, respectively.

We then apply our proposed sparse FPCA method to analyze the major variations among these acceleration curves. For comparison, we also estimate the first four sparse FPCs. We select the sparsity parameter $\lambda = 40$ based on AIC. The corresponding estimated sparse FPCs are shown in Figure 3. They account for 20%, 19%, 17% and 14% of the total variation among the acceleration curves, respectively.

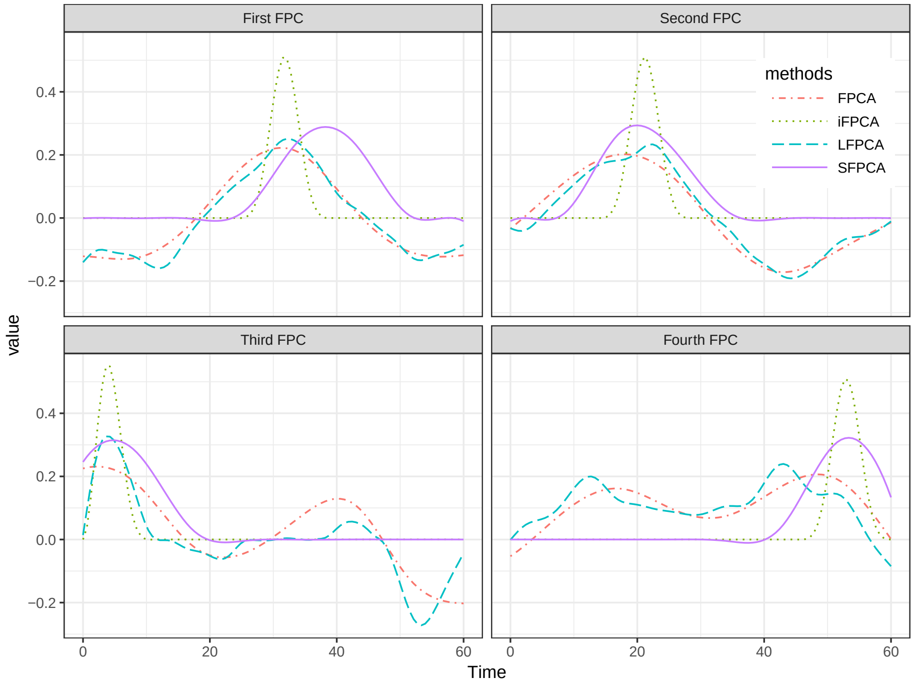

Figure 3: Estimated functional principal components using our proposed sparse FPCA (SFPCA) method (solid line) and three other methods, including the conventional FPCA method (dotted-dash line) (Ramsay and Silverman, 2005), the interpretable functional principal component analysis (iFPCA) (dotted line) proposed by Lin et al. (2016) and the localized functional principal component analysis (LFPCA) (dashed line) proposed by Chen and Lei (2015).

Figure 3 shows that the estimated sparse FPCs tend to be nonzero at different intervals along with the whole time domain. For instance, the first sparse FPC is nonzero roughly in [25,45], which indicates that the major variation among the acceleration curves comes from this interval. Compared to the conventional FPCs, which are nonzero everywhere, the estimated SFPCs are nonzero at different intervals. More specifically, the second sparse FPC is nonzero between $[10, 25]$, the third sparse FPC is nonzero between $[0, 20]$, and the last sparse FPC is nonzero between $[40, 60]$. This observation suggests the major variation within the acceleration curves can be separated into different subintervals rather than mixing in the entire domain.

For the other existing methods, we can see that the estimated FPCs using the iFPCA method are the most sparse. However, all four FPCs estimated with iFPCA have the non-sparse domains aligned with those FPCs estimated with SPCA. For example, both methods' first FPC are non-sparse in $[30, 40]$. LFPCA seems very similar to the conventional FPCA method for the first, second and fourth FPCs and only results in a sparse estimate for the third FPC.

In addition, we utilize the resulting sparse FPCs shown in Figure 3 to study the relationship between the PM of the acceleration curve using the following functional linear model:

$$Y_i = \int_0^T \beta(t) X_i(t) dt + \epsilon_i,$$

in which Y_i and $X_i(t)$ represents the log PMs level and the acceleration curve for the ith observation. Note that we can approximate both $X_i(t)$ and unknown $\beta(t)$ using the estimated sparse FPCs, that is, $X_i(t) = \sum_k s_{ik}\xi_k(t)$ and $\beta(t) = \sum_k \beta_k \xi_k(t)$, where $s_{ik} = \int \xi(t) X_i(t) dt$ denotes the kth score for the ith observation. Then the functional linear model above becomes:

$$\begin{aligned} Y_i &= \int_0^T \beta(t) X_i(t) dt + \epsilon_i \\ &= \int_0^T \left[\sum_k \beta_k \xi_k(t)\right]\left[\sum_k s_{ik}\xi_k(t)\right] dt + \epsilon_i \\ &= \sum_k \beta_k s_{ik} + \epsilon_i. \end{aligned} \tag{15}$$

Therefore we can simply regress Y_i on those scores s_{ik} to estimate β_k. Table 1 shows the estimated $\hat{\beta}_k$. It can be seen clearly that the PM is only significantly related to the third sparse FPC, as $\hat{\beta}_3$ is the only significant coefficient estimated. Because the third sparse FPC is only nonzero in $[0, 20]$ as shown in Figure 3, it suggests that the PM level is significantly affected by the acceleration curve between 0 and 20 seconds. This conclusion coincides with the analysis in Hall and Hooker (2016).

Table 1: Estimated coefficients for the functional linear model (15) with their standard errors (SEs), t statistics, and p values.

	Estimate	SEs	t Statistics	P-value
β_1	-0.40	0.27	-1.51	1.35e-01
β_2	-0.15	0.27	-0.56	5.78e-01
β_3	1.87	0.29	6.39	4.69e-09
β_4	0.42	0.28	1.47	1.44e-01

Furthermore, we conduct a 5-fold cross-validation to access the prediction accuracy using the estimated sparse FPCs. In each repetition, we select one fold of observations as the test set and leave the rest as the training set, then we estimate the proposed sparse FPCs using the training set only. After that, we obtain the predicted log PM level using the estimated sparse FPCs in the test set and compare the predicted values with the true values. Table 4 summarizes the mean squared prediction errors in the 5-fold cross-validation process. We also apply the same procedure on three other existing methods, including the conventional FPCA method (Ramsay and Silverman, 2005), the interpretable functional principal component analysis (iFPCA) proposed by Lin et al. (2016) and the localized functional principal component analysis (LFPCA) proposed by Chen and Lei (2015). The mean squared prediction errors using our proposed method is 4.4% and 6.5% lower than those using the conventional FPCA and LPCA methods, respectively. Our method has the mean squared prediction errors 1.7% larger than the iFPCA method.

Table 2: The mean squared prediction errors using the sparse FPCA (SFPCA), iFPCA, LFPCA and conventional FPCA methods in a 5-fold cross-validation.

Replicates	FPCA	SFPCA	iFPCA	LFPCA
1	98.79	95.55	84.03	98.63
2	44.20	39.48	58.55	47.06
3	67.03	62.89	63.81	66.41
4	69.32	48.21	47.83	71.31
5	50.46	69.10	55.68	53.77
Average	65.96	63.05	61.98	67.44

5. Simulation Study

We conduct two simulation studies to evaluate our proposed sparse FPCA method by comparing it with three available methods, including the conventional FPCA method (Ramsay and Silverman, 2005), the interpretable functional principal component analysis (iFPCA) proposed by Lin et al. (2016) and the localized functional principal component analysis (LFPCA) proposed by Chen and Lei (2015).

More specifically, the true underlying functional curves are generated using

$$X_i(t) = s_{i1}\xi_1(t) + s_{i2}\xi_2(t) + s_{i3}\xi_3(t) + s_{i4}\xi_4(t),$$

$t \in [1, 60]$. We consider two scenarios depending on whether the true FPCs are sparse or not. In the sparse scenario, the true FPCs, $\xi_k(t), k = 1, 2, 3, 4$, are set as the estimated FPCs using our proposed sparse FPCA method from the real data, shown as solid lines in Figure 3. In the non-sparse scenario, the true FPCs, $\xi_k(t), k = 1, 2, 3, 4$, are set as the estimated FPCs using the conventional FPCA method from the real data, shown as dotted-dash lines in Figure 3. In both cases, the FPC scores $\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3, \mathbf{s}_4$ are generated from multivariate normal distribution with mean zero and the variance-covariance matrix $\Sigma = \text{diag}(30, 20, 10, 3)$. The observed trajectories are generated by $Y_{ij} = X_i(t_j) + \epsilon_{ij}$ for $j = 1, \ldots, 60$, where t_j is the j-th observed point equally spaced in $[1, 60]$ and $\epsilon_{ij} \overset{i.i.d}{\sim} N(0, 1)$. In each simulation run, we generate N sample curves and then apply the four methods to estimate the true FPCs. For the iFPCA method, following the authors' recommendation, the smoothing parameter is chosen by CV and the sparsity parameter is also selected using CV after the smoothing parameter is determined. For the LFPCA method, we also use the recommended CV method to determine the tuning parameters. For the proposed sparse FPCA method, we use AIC as our criterion to choose the tuning parameters as described in Section 3.4.

We compare the performance of the four methods using the integrated error (IE) defined as follows:

$$\mathrm{IE}(\hat{\xi}_k) = \int (\hat{\xi}_k(t) - \xi_k(t))^2 dt, \tag{16}$$

in which $\xi_k(t)$ and $\hat{\xi}_k(t)$ represent the k-th true FPC and the corresponding estimated FPC, respectively. Each simulation is repeated 100 times in order to compute the average IEs and the corresponding standard deviations.

5.1. Sparse Setting

Table 5.1 shows the mean and standard deviations of the integrated errors (IEs) using different methods for each of the true underlying FPCs with different sample sizes when the true FPCs are spars. The sparse FPCA method yields the lowest IEs in comparison with all three alternative methods under different sample sizes. We also notice that the average IE and the corresponding standard deviations increase with the rank of estimated FPC. Besides, as the sample size N increases from 100 to 1000, the average IE decreases. Among the alternative methods, the performance of the LFPCA method is quite comparable with the sparse FPCA method, especially when the sample size is large, and the rank of FPC is low. The iFPCA method gives quite large IEs when the sample size is 100 and produces reasonable estimates for the first FPC when the sample size increases above 500. However, it is unable to yield reasonable estimates for the lower rank FPCs. For instance, the average IE remains quite large for ξ_2 even when the sample size is 1000. Lastly, the conventional FPCA method is not able to estimate the underlying true FPCs well, and its performance does not improve as the sample size goes up. This is because the conventional FPCA method only aims at maximizing the explained variations and is not able to recover the sparsity structure of the underlying true FPCs.

Table 3: The means and standard deviations of integrated errors (IE) using the sparse FPCA (SFPCA), iFPCA, LFPCA and the conventional FPCA with different sample size $N = 100, 500, 1000$ when the true FPCs are sparse.

N	Method	IE($\hat{\xi}_1$)	IE($\hat{\xi}_2$)	IE($\hat{\xi}_3$)	IE($\hat{\xi}_4$)
100	FPCA	0.233(0.040)	0.41(0.076)	0.25(0.063)	0.148(0.018)
100	iFPCA	0.506(0.023)	0.509(0.033)	0.823(0.052)	1.109(0.128)
100	LFPCA	0.007(0.018)	0.061(0.085)	0.126(0.138)	0.124(0.055)
100	SFPCA	0.002(0.001)	0.004(0.003)	0.009(0.007)	0.106(0.067)
500	FPCA	0.225(0.019)	0.406(0.033)	0.251(0.026)	0.143(0.007)
500	iFPCA	0.040(0.199)	0.489(0.717)	1.073(0.208)	1.414(0.184)
500	LFPCA	0.014(0.025)	0.055(0.048)	0.085(0.053)	0.090(0.016)
500	SFPCA	7e-04(4e-04)	0.002(9e-04)	0.007(0.003)	0.080(0.024)
1000	FPCA	0.222(0.012)	0.399(0.02)	0.246(0.016)	0.142(0.005)
1000	iFPCA	0.024(0.016)	0.325(0.572)	1.125(0.146)	1.427(0.146)
1000	LFPCA	0.018(0.024)	0.062(0.046)	0.088(0.039)	0.078(0.018)
1000	SFPCA	5e-04(2e-04)	0.001(7e-04)	0.008(0.002)	0.081(0.019)

We plot those estimated FPCs from one simulation replicate in Figure 4 using sparse FPCA, LFPCA and iFPCA. We can see that the estimated FPCs from the sparse FPCA method are closest to the true FPCs. More specifically, iFPCA and LFPCA perform similarly as the sparse FPCA method in estimating the first two FPCs, but they become worse than the sparse FPCA method when estimating the 3rd and 4th FPCs.

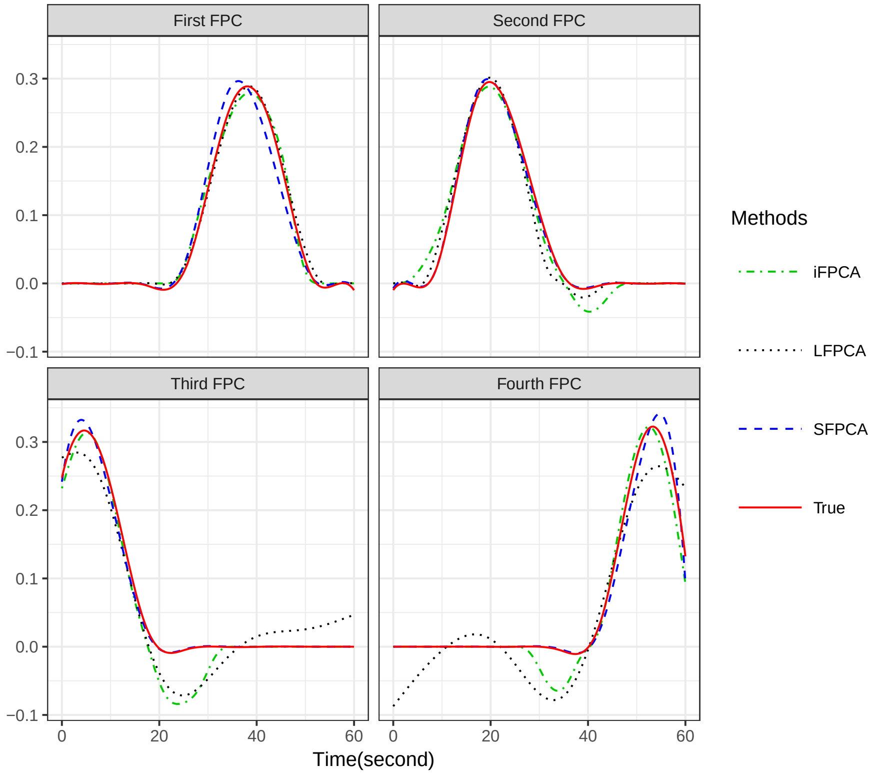

Figure 4: The estimated FPCs using three methods including iFPCA (dotted-dashed line), LFPCA (dotted line), and sparse FPCA (dashed line) in comparison with the true FPCs (solid line) in one simulation replicate when the true FPCs are sparse.

5.2. NonSparse Setting

Table 4 shows the mean and standard deviations of the integrated errors (IEs) using different methods for each of the true underlying FPCs with different sample sizes when the true FPCs are not sparse. In the nonsparse setting, the conventional FPCA method performs best, as expected, because the true FPCs are the estimated FPCs from the real data using the conventional FPCA method and the true FPCs are nonsparse. Both the proposed sparse FPCA method and LFPCA are close to the conventional FPCA method in terms of average integrated errors for all FPCs.

Table 4: The means and standard deviations of integrated errors (IE) using the sparse FPCA (SFPCA), iFPCA, LFPCA and the conventional FPCA with different sample size $N = 100, 500, 1000$ when the true FPCs are not sparse.

N	Method	IE($\hat{\xi}_1$)	IE($\hat{\xi}_2$)	IE($\hat{\xi}_3$)	IE($\hat{\xi}_4$)
100	FPCA	0.121(0.194)	0.137(0.198)	0.290(0.328)	0.279(0.321)
100	iFPCA	0.873(0.318)	1.513(0.268)	1.669(0.243)	1.057(0.214)
100	LFPCA	0.122(0.200)	0.130(0.196)	0.314(0.278)	0.267(0.273)
100	SFPCA	0.143(0.194)	0.163(0.204)	0.343(0.336)	0.434(0.344)
500	FPCA	0.022(0.018)	0.027(0.020)	0.142(0.209)	0.133(0.205)
500	iFPCA	0.334(0.127)	0.524(0.307)	1.128(0.314)	1.049(0.459)
500	LFPCA	0.021(0.018)	0.025(0.020)	0.114(0.092)	0.100(0.086)
500	SFPCA	0.042(0.022)	0.051(0.025)	0.183(0.209)	0.236(0.209)
1000	FPCA	0.018(0.015)	0.021(0.016)	0.086(0.047)	0.081(0.047)
1000	iFPCA	0.241(0.071)	0.305(0.186)	1.111(0.226)	1.273(0.438)
1000	LFPCA	0.013(0.010)	0.016(0.012)	0.068(0.036)	0.060(0.038)
1000	SFPCA	0.037(0.021)	0.049(0.023)	0.125(0.051)	0.165(0.068)

Figure 5 shows the estimated FPCs from one simulation replicate using the conventional FPCA, the sparse FPCA, LFPCA and iFPCA. It shows that iFPCA tends to over shrink the FPCs' nonzero domains. Therefore, the FPCs estimated with iFPCA are quite far from the true FPCs in comparison with the other three methods.

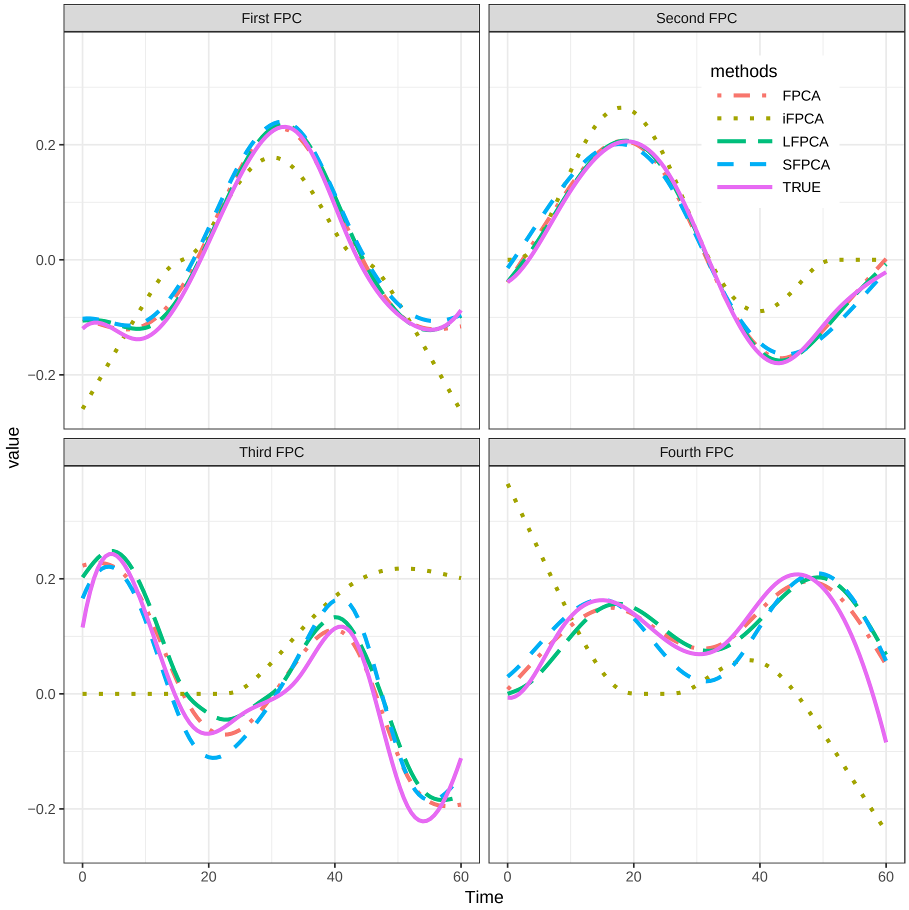

Figure 5: The estimated FPCs using four methods including sparse FPCA (dashed line), iFPCA (dotted line), LFPCA (long dashed line), and the conventional FPCA (dotted-dash line) in comparison with the true FPCs (solid line) in one simulation replicate when the true FPCs are not sparse.

6. Conclusions

In this paper, we focused on the problem of estimating FPCs with compact support. The conventional FPCA method estimate FPCs by maximizing the variation among the functional data. But these estimated FPCs are nonzero in the entire domain. Hence they are often hard to interpret.

Conventional FPCA methods estimate FPCs by eigendecomposing the sample covariance function. However, when we need to add the regulation penalty to the FPCs, this eigendecomposition method always leads to an NP-hard problem. We proposed a new regression framework to estimate the sparse FPCs by minimizing the errors of approximating functional data. One major advantage of our framework is that the optimization problem is not NP-hard when adding a penalty term to regulate the FPCs. We also showed that the FPCs estimated with our proposed sparse FPCA method is equivalent to the FPCs with the conventional FPCA method when the sparsity parameter is zero.

Our sparse FPCA method was applied to explore the major variations among the acceleration rate curves of 107 diesel trucks. We found that the nonzero regions of the estimated sparse FPCs are well separated, which shows that the major variation within the acceleration curves can be separated into different subintervals rather than mixing with each other in the entire domain. We also compare our proposed sparse FPCA method with the conventional FPCA method (Ramsay and Silverman, 2005), the interpretable functional principal component analysis (Lin et al., 2016) and the localized functional principal component analysis (Chen and Lei, 2015) using a simulation study. The simulation study shows that the sparse FPCA method obtains more accurate estimates of FPCs in comparison with the alternative three methods when the true FPCs have compact support regions. When the true FPCs are nonsparse, the proposed sparse FPCA method can also estimate the FPCs as well as the conventional FPCA method. On the other hand, our proposed sparse FPCA method does not guarantee orthogonality of the estimated FPCs, while the localized functional principal component analysis (Chen and Lei, 2015) enforces the orthogonality of the estimated FPCs.

Acknowledgments

The authors are very grateful for the constructive comments from the Editor, an Associate Editor, and two reviewers. These comments are very helpful for the authors to improve their work. This research was supported by the

Postgraduate Scholarship-Doctorial (PGS-D) to Y. Nie from the Natural Sciences and Engineering Research Council of Canada (NSERC) and the NSERC Discovery grant (RGPIN-2018-06008) of J. Cao.

Supplementary Materials

The supplementary document contains the detailed proofs for the two propositions in Section 2 and some additional simulation studies. An R package "sparseFPCA" is developed to implement the proposed SFPC method. The computing scripts for the simulation study can be downloaded at `https://github.com/caojiguo/sparseFPCA`.

References

J. Dauxois, A. Pousse, Y. Romain, Asymptotic theory for the principal component analysis of a vector random function: some applications to statistical inference, Journal of Multivariate Analysis 12 (1982) 136–154.

A. Mas, Weak convergence for the covariance operators of a hilbertian linear process, Stochastic Processes and their Applications 99 (2002) 117–135.

D. Bosq, Linear Processes in Function Spaces: Theory and Applications, Springer-Verlag, New York, 2000.

P. Hall, J. L. Horowitz, Methodology and convergence rates for functional linear regression, The Annals of Statistics 35 (2007) 70–91.

P. Hall, H.-G. Müller, J.-L. Wang, Properties of principal component methods for functional and longitudinal data analysis, The Annals of Statistics 34 (2006) 1493–1517.

F. Yao, H.-G. Müller, J.-L. Wang, Functional data analysis for sparse longitudinal data, Journal of the American Statistical Association 100 (2005) 577–590.

P. Sang, L. Wang, J. Cao, Parametric functional principal component analysis, Biometrics 73 (2017) 802–810.

Y. Nie, L. Wang, B. Liu, J. Cao, Supervised functional principal component analysis, Statistics and Computing 28 (2018) 713–723.

F. Yao, H.-G. Müller, J.-L. Wang, Functional linear regression analysis for longitudinal data, The Annals of Statistics 33 (2005) 2873–2903.

J. Ramsay, B. Silverman, Functional data analysis, second ed., Springer-Verlag, New York, 2005.

H.-G. Müller, Functional modelling and classification of longitudinal data, Scandinavian Journal of Statistics 32 (2005) 223–240.

H.-G. Müller, U. Stadtmüller, Generalized functional linear models, The Annals of Statistics 33 (2005) 774–805.

J. Peng, H.-G. Müller, Distance-based clustering of sparsely observed stochastic processes, with applications to online auctions, The Annals of Applied Statistics 2 (2008) 1056–1077.

J. Dong, L. Wang, J. Gill, J. Cao, Functional principal component analysis of gfr curves after kidney transplant, Statistical Methods in Medical Research 27 (2018) 3785–3796.

J. Rice, B. Silverman, Estimating the mean and covariance structure nonparametrically when the data are curves, Journal of the Royal Statistical Society. Series B. 53 (1991) 233–243.

S. Pezzulli, Some properties of smoothed principal components analysis for functional data, Computational Statistics 8 (1993) 1–16.

B. W. Silverman, Smoothed functional principal components analysis by choice of norm, The Annals of Statistics 24 (1996) 1–24.

Z. Lin, L. Wang, J. Cao, Interpretable functional principal component analysis, Biometrics 72 (2016) 846–854.

K. Chen, J. Lei, Localized functional principal component analysis, Journal of the American Statistical Association 110 (2015) 1266–1275.

C. Di, C. M. Crainiceanu, W. S. Jank, Multilevel sparse functional principal component analysis, Stat 3 (2014) 126–143.

G. Li, H. Shen, J. Z. Huang, Supervised sparse and functional principal component analysis, Journal of Computational and Graphical Statistics 25 (2016) 859–878.

Z. Lin, J. Cao, L. Wang, H. Wang, Locally sparse estimator for functional linear regression models, Journal of Computational and Graphical Statistics 26 (2017) 306–318.

J. Fan, R. Li, Variable selection via nonconcave penalized likelihood and its oracle properties, Journal of the American Statistical Association 96 (2001) 1348–1360.

K. Fukunaga, W. L. Koontz, Representation of random processes using the finite karhunen-loeve expansion, Information and Control 16 (1970) 85–101.

N. M. Tran, An introduction to theoretical properties of functional principal component analysis, Ph.D. thesis, Honours thesis, The University of Melbourne., 2008.

N. Clark, M. Gautam, W. Wayne, D. Lyons, G. Thompson, B. Zielinska, Heavy-duty vehicle chassis dynamometer testing for emissions inventory, air quality modeling, source apportionment and air toxics emissions inventory, Coordinating Research Council, incorporated (2007).

P. Hall, G. Hooker, Truncated linear models for functional data, Journal of the Royal Statistical Society: Series B 78 (2016) 637–653.