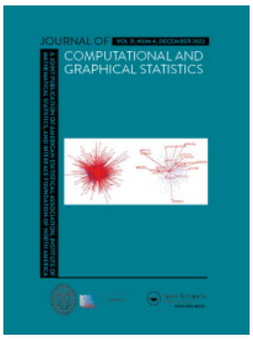

Journal of Computational and Graphical Statistics

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/ucgs20

Two-Dimensional Functional Principal Component Analysis for Image Feature Extraction

Haolun Shi, Yuping Yang, Liangliang Wang, Da Ma, Mirza Faisal Beg, Jian Pei & Jiguo Cao

To cite this article: Haolun Shi, Yuping Yang, Liangliang Wang, Da Ma, Mirza Faisal Beg, Jian Pei & Jiguo Cao (2022) Two-Dimensional Functional Principal Component Analysis for Image Feature Extraction, Journal of Computational and Graphical Statistics, 31:4, 1127-1140, DOI: 10.1080/10618600.2022.2035738

To link to this article: https://doi.org/10.1080/10618600.2022.2035738

JOURNAL OF COMPUTATIONAL AND GRAPHICAL STATISTICS
2022, VOL. 31, NO. 4, 1127–1140
https://doi.org/10.1080/10618600.2022.2035738

Taylor & Francis
Taylor & Francis Group

Two-Dimensional Functional Principal Component Analysis for Image Feature Extraction

Haolun Shi[a], Yuping Yang[a], Liangliang Wang[a], Da Ma[b,c], Mirza Faisal Beg[c], Jian Pei[d], and Jiguo Cao[a]

[a]Department of Statistics and Actuarial Science, Simon Fraser University, Burnaby, Canada; [b]School of Medicine, Wake Forest University, Winston-Salem, NC; [c]School of Engineering, Simon Fraser University, Burnaby, Canada; [d]School of Computing Science, Simon Fraser University, Burnaby, Canada

ABSTRACT

Methodologies for functional principal component analysis are well established in the one-dimensional setting. However, for two-dimensional surfaces, for example, images, conducting functional principal component analysis is complicated and challenging, because the conventional eigendecomposition approach would require the estimation of a four-dimensional covariance function, which may incur high cost in terms of time and machine memory. To circumvent such computational difficulties, we propose a novel two-dimensional functional principal component analysis for extracting functional principal components and achieving dimensionality reduction for images. Different from the conventional eigendecomposition approach, our proposed method is based on the direct estimation of the optimal two-dimensional functional principal components via tensor product B-spline, which opens up a new avenue for estimating functional principal components. We present theoretical results that prove the consistency of the proposed approach. Our method is illustrated by analyzing brain images of subjects with the Alzheimer's Disease and the handwritten digits images. The finite sample performance of our method is further assessed with some simulation studies. Supplementary materials for this article are available online.

ARTICLE HISTORY
Received March 2020
Revised October 2021

KEYWORDS
Eigendecomposition;
Functional data analysis;
Two-dimensional surfaces

1. Introduction

In functional data analysis, functional principal component analysis (FPCA) is widely used for dimensionality reduction. The goal of FPCA is to identify, decompose, and rank the major sources of variation in a sample of realizations from a stochastic process. FPCA seeks to decompose the underlying process into a linear combination of functional principal components (FPCs). The FPCs serve as a basis for the best approximation of the infinite-dimensional function (curve) of interest. Dimensionality reduction is achieved by selecting the top few FPCs that cumulatively explain the majority of variation.

The majority of existing works on FPCA focus on curves in the one-dimensional setting. For regularly and densely observed curve, several smoothing approaches for FPCA are proposed. Ramsay and Dalzell (1991) performed smoothing on the functional data in the first step, which is then followed by the regular FPCA procedure on the smoothed data. Rice and Silverman (1991) developed smoothed nonparametric estimates of the eigenfunctions and suggested a suitable way for determining the degree of smoothing. Pezzulli and Silverman (1993) introduced a roughness penalty term in the optimization criterion of FPCA. Silverman (1996) studied the theoretical properties of the smoothed FPCA. Lin et al. (2016) and Nie and Cao (2020) enhanced the interpretability of FPCA by estimating FPCs which are only nonzero in subregions. Sang et al. (2017) presented a parametric approach to estimate top FPCs to help users to interpret FPCs. Nie et al. (2018) proposed a supervised version of FPCA by considering the correlation of the functional predictor and response variable. Dong et al. (2018) analyzed GFR curves after kidney transplant by using FPCA.

In the case where the designs points are sparsely and irregularly spaced, several extensions of FPCA under one-dimensional setting have been proposed. James et al. (2000) considered FPCA for sparse and irregular functional data using a reduced rank mixed effects framework. Yao et al. (2005) developed an effective way to conduct FPCA where the FPC scores are computed based on conditional expectation. Hall et al. (2006) derived consistency results for the eigenfunction estimators under a sparse sampling scheme with measurement errors. Zipunnikov et al. (2011) developed a fast and scalable multilevel FPCA method for high-dimensional data. Di et al. (2014) proposed an FPCA method for sparse and multilevel functional data based on the multilevel structure in the covariance function. Goldsmith et al. (2015) adopted the multilevel FPCA in the generalized function-on-scalar regression model. Shi et al. (2021) suggested the informatively missing FPCA to analyze the longitudinal trajectories that are subject to informative missingness. Nie et al. (2022) proposed a method for conducting functional principal component analysis under a new regression framework. Other extensions include FPCA for structured data, FPCA on a variable domain, and for high-dimensional data (Shou et al. 2015; Xiao et al. 2016; Johns et al. 2019).

More recently, several methods under the Bayesian paradigm have been introduced to functional data analysis. Thompson and Rosen (2008) proposed a Bayesian nonparametric approach

CONTACT Jiguo Cao *jiguo_cao@sfu.ca* Department of Statistics and Actuarial Science, Simon Fraser University, Burnaby, V5A1S6, BC, Canada.
Supplementary materials for this article are available online. Please go to *www.tandfonline.com/r/JCGS*.

for modeling sparse functional data. Suarez and Ghosal (2017) developed a Bayesian FPCA where the covariance function is modeled using an approximate spectral decomposition. Jiang et al. (2021) extended the Bayesian sparse FPCA to the multivariate setting, which allows simultaneous estimation for multiple groups of trajectories.

FPCA in the one-dimensional setting is well studied in the literature. In the case of the two-dimensional surface, for example, images, conducting FPCA is a difficult and computationally challenging task. The usual approach that extends the one-dimensional FPCA would require the estimation of a four-dimensional covariance function, which may incur high cost in terms of time and machine memory. On the other hand, it is worth noting that concatenating columns/rows in the two-dimensional observed data matrix into a one-dimensional vector, based on which the FPCA is conducted, would lead to incorrect estimation as it ignores the crucial locational information on the two-dimensional surface. To circumvent these issues, we propose a two-dimensional functional principal component method for extracting the FPCs, extending the work of Nie et al. (2022). The method directly estimates the optimal empirical principal component functions and their related basis coefficients via tensor product B-spline. Theoretical studies reveal that the estimated empirical functional principal component converges to the true FPCs. The proposed method provides an alternative to the conventional approach of conducting eigen-decomposition on the sample covariance function to estimate the FPCs. It is also worth noting that while there have been abundant works that focus on the multivariate FPCA, for example, Happ and Greven (2018), our article attempts to tackle a different problem: rather than the observation having multiple dimensions, the domain of the stochastic process is multi-dimensional, which is a problem rarely explored in the existing literature.

Our approach has several highlights. The most important advantage of our approach is that it circumvents the difficulties related to the need of estimating the high-dimensional smoothed covariance function while still achieving the goal of effective dimensionality reduction. Moreover, our approach is based on a simple regression framework and is thus, robust to the irregularities in the distribution of the design points and can also naturally accommodate sparsely sampled data surfaces. The two-dimensional functional principal components obtained from our method have nice interpretability and may serve as a foundation for clustering and data recovery. The computing codes used in our simulation studies and the application are available at *https://github.com/haoluns/2DFPCA*.

The rest of the article is organized as follows. In Section 2, we present the proposed methodology of the two-dimensional functional principal component analysis method. In Section 3, we derive the theoretical properties of our approach, and in Section 4, we apply the proposed method on two datasets consisting of images of brain MRI scans and handwritten digits and provides interpretation to the results. In Section 5, simulation studies are conducted to assess the empirical accuracy in the estimation of the FPCs. Finally, Section 6 concludes the article with a discussion.

2. Methodology

We consider n independent realizations of a random two-dimensional stochastic process $X(s,t)$, where $(s,t) \in \mathcal{T}$. Each realization corresponds to a surface and each set of two-dimensional functional data is regarded as noisy samples from the surface. Let y_{ij} denote the jth observation of the random function $X_i(s,t)$, the ith realization of the stochastic process $X(s,t)$. For $i = 1,\ldots,n$ and $j = 1,\ldots,n_i$, where n_i is the number of sample points for the ith random function, the observed data point y_{ij} is modeled as

$$y_{ij} = X_i(s_{ij}, t_{ij}) + \epsilon_{ij},$$

where ϵ_{ij} is a random measurement error term with mean zero and variance σ^2. For simplicity, we assume that all the n_i are the same.

The two-dimensional random function $X_i(s,t)$ can be expressed in terms of the Mercer expansion of a series of orthonormal basis functions, that is,

$$X_i(s,t) = \mu(s,t) + \sum_{m=1}^{\infty} \alpha_{im}\psi_m(s,t),$$

where the basis functions should satisfy $||\psi_m||^2 = 1$ and $\langle \psi_m, \psi_l \rangle = 1$ if $m = l$, and 0 otherwise.

We propose a two-dimensional principal component analysis for estimating the first M functional principal components $\psi_m(s,t)$. The estimation procedure seeks to locate the minimizer of the objective function

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{n_i}\sum_{j=1}^{n_i}\left\{y_{ij} - \mu(s_{ij},t_{ij}) - \sum_{m=1}^{M}\alpha_{im}\psi_m(s_{ij},t_{ij})\right\}^2.$$

The M functional principal components are obtained in a sequential manner, that is, the mth FPC is approximated conditional on the estimated values of the first $m-1$ FPCs.

2.1. Estimating the First Functional Principal Component

The first FPC $\psi_1(s,t)$ is obtained by minimizing

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{n_i}\sum_{j=1}^{n_i}\left\{y_{ij} - \widehat{\mu}(s_{ij},t_{ij}) - \alpha_{i1}\psi_1(s_{ij},t_{ij})\right\}^2. \quad (1)$$

subject to $||\psi_1||^2 = 1$, where $\widehat{\mu}(s_{ij},t_{ij})$ is the estimated mean function by pooling all the data together and computing the sample mean at point (s_{ij},t_{ij}). For notational simplicity, we denote $y^*_{ij} = y_{ij} - \widehat{\mu}(s_{ij},t_{ij})$.

We express the two-dimensional surface $\psi_1(s,t)$ in terms of a basis expansion via the tensor product B-spline

$$\begin{aligned}\psi_1(s,t) &= \sum_{i=1}^{s_1}\sum_{j=1}^{s_2}\beta_{1,ij}b_i^{(1)}(s)b_j^{(2)}(t)\\ &= \mathbf{b}_1(s)^\top \mathbf{B}_1 \mathbf{b}_2(t), \quad (2)\end{aligned}$$

where $\mathbf{b}_1 = (b_1^{(1)},\ldots,b_{s_1}^{(1)})^\top$ and $\mathbf{b}_2 = (b_1^{(2)},\ldots,b_{s_2}^{(2)})^\top$ denote the spline basis functions for the first and second dimensions, respectively, s_1 and s_2 are the numbers of spline basis along each dimension, and $\mathbf{B}_1 = (\beta_{1,ij})$ is the matrix of spline coefficients.

For notational simplicity, we rearrange the $s_1 \times s_2$ matrix $\mathbf{B}_1$ into a $s_1 s_2 \times 1$ vector, and reorganize (2) as

$$\psi_1(s,t) = \boldsymbol{\beta}_1^\top \mathbf{b}(s,t),$$

where $\boldsymbol{\beta}_1 = \text{vec}(\mathbf{B}_1)$, and $\mathbf{b}(s,t) = \text{vec}(\mathbf{b}_1(s) \otimes \mathbf{b}_2(t))$. The $\text{vec}(\cdot)$ is the vectorization operator, which performs stacking of all the columns of a matrix.

Our goal is to estimate $\boldsymbol{\alpha}_1 = (\alpha_{11}, \ldots, \alpha_{n1})^\top$ and the coefficient vector $\boldsymbol{\beta}_1$ for the tensor product B-spline. The estimation is conducted iteratively, that is, conditional on the current estimate of $\boldsymbol{\alpha}_1$, the estimate of $\boldsymbol{\beta}_1$ is updated by minimizing the loss function in (1), and vice versa. The detailed algorithm is as follows.

1. Specify the spline basis functions $\mathbf{b}_1$ and $\mathbf{b}_2$, as well as their respective number of spline bases s_1 and s_2. Initialize $\boldsymbol{\beta}_1^{(0)}$ and hence, $\psi_1^{(0)}$ subject to $||\psi_1^{(0)}||^2 = 1$.
2. Let ℓ denote the current index of iteration. Conditional on the current estimate $\boldsymbol{\beta}_1^{(\ell)}$, the estimated functional principal component is

$$\psi_1^{(\ell)}(s,t) = \boldsymbol{\beta}_1^{(\ell)\top} \mathbf{b}(s,t).$$

   The least-square estimate of each element of $\boldsymbol{\alpha}_1$ can be obtained as

$$\alpha_{i1}^{(\ell)} = (\boldsymbol{\psi}_{i1}^\top \boldsymbol{\psi}_{i1})^{-1} \boldsymbol{\psi}_{i1}^\top \boldsymbol{y}_i, \qquad i = 1, \ldots, n,$$

   where $\boldsymbol{\psi}_{i1} = (\psi_1^{(\ell)}(s_{i1}, t_{i1}), \ldots, \psi_1^{(\ell)}(s_{in_i}, t_{in_i}))^\top$ and $\boldsymbol{y}_i = (y_{i1}^*, \ldots, y_{in_i}^*)^\top$.
3. Conditional on the current estimate $\boldsymbol{\alpha}_1^{(\ell)}$, update the estimate of $\boldsymbol{\beta}_1$ by minimizing (1) subject to $||\psi_1||^2 = 1$. The solution to such a constrained least-square problem is

$$\boldsymbol{\beta}_1^{(\ell+1)} = \frac{\widetilde{\boldsymbol{\beta}}_1^{(\ell+1)}}{||\widetilde{\psi}_1^{(\ell+1)}||},$$

   where $\widetilde{\boldsymbol{\beta}}_1^{(\ell+1)}$ is the unconstrained least-square estimate of $\boldsymbol{\beta}_1$ and $||\widetilde{\psi}_1^{(\ell+1)}||$ is the norm of the resulting functional principal component function.
4. Repeat Steps 2 and 3 until convergence.

2.2. Estimating Subsequent Functional Principal Components

The subsequent functional principal components are obtained in a sequential manner. Let J denote the current index of functional principal component. From the first $J-1$ estimation steps, we obtain the estimates $\widehat{\boldsymbol{\beta}}_m$ and the resulting $\widehat{\psi}_m$, where $m = 1, \ldots, J-1$. Given these estimated functional principal components, the Jth functional principal component ψ_J is obtained by minimizing

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{n_i}\sum_{j=1}^{n_i}\left\{y_{ij}^* - \sum_{m=1}^{J-1}\alpha_{im}\widehat{\psi}_m(s_{ij},t_{ij}) - \alpha_{iJ}\psi_J(s_{ij},t_{ij})\right\}^2, \quad (3)$$

subject to $||\psi_J||^2 = 1$ and $\langle \widehat{\psi}_m, \psi_J \rangle = 1$ if $m = J$, and 0 otherwise.

The detailed algorithm is as follows.

1. Initialize $\boldsymbol{\beta}_J^{(0)}$ and hence, $\psi_J^{(0)}$ subject to $||\psi_J^{(0)}||^2 = 1$.
2. Let ℓ denote the current index of iteration, and let $\boldsymbol{\alpha}_i = (\alpha_{i1}, \ldots, \alpha_{iJ})^\top$ denote the FPC scores for all the FPCs in the ith subject. Conditional on the current estimate $\boldsymbol{\beta}_J^{(\ell)}$ and all the estimated $\widehat{\boldsymbol{\beta}}_m$ from the previous steps, the estimated functional principal component functions are

$$\psi_m^{(\ell)}(s,t) = \widehat{\boldsymbol{\beta}}_m^\top \mathbf{b}(s,t), \quad m = 1, \ldots, J-1,$$
$$\psi_J^{(\ell)}(s,t) = \boldsymbol{\beta}_J^{(\ell)\top} \mathbf{b}(s,t).$$

   The least-square estimate of $\boldsymbol{\alpha}_i$ can be obtained as

$$\boldsymbol{\alpha}_i^{(\ell)} = (\boldsymbol{\psi}_i^\top \boldsymbol{\psi}_i)^{-1} \boldsymbol{\psi}_i^\top \boldsymbol{y}_i, \qquad i = 1, \ldots, n,$$

   where $\boldsymbol{\psi}_i = (\boldsymbol{\psi}_{i1}, \ldots, \boldsymbol{\psi}_{iJ})$, $\boldsymbol{\psi}_{im} = (\psi_m^{(\ell)}(s_{i1}, t_{i1}), \ldots, \psi_m^{(\ell)}(s_{in_i}, t_{in_i}))^\top$ and $\boldsymbol{y}_i = (y_{i1}^*, \ldots, y_{in_i}^*)^\top$.
3. Conditional on all the current estimates $\boldsymbol{\alpha}_i^{(\ell)}$, update the estimate of $\boldsymbol{\beta}_J$ by minimizing (3) subject to $||\psi_J||^2 = 1$, and $\langle \widehat{\psi}_m, \psi_J \rangle = 1$ if $m = J$, and 0 otherwise. Essentially, the estimate of $\boldsymbol{\beta}_J$ is the solution to a least-square problem with equality constraints (Lawson and Hanson 1974). To be specific,

$$\boldsymbol{\beta}_J^{(\ell+1)} = \arg\min_{\boldsymbol{\beta}_J} \sum_{i=1}^{n}\frac{1}{n_i}\sum_{j=1}^{n_i}\left\{y_{ij}^* - \sum_{m=1}^{J-1}\alpha_{im}^{(\ell)}\widehat{\boldsymbol{\beta}}_m^\top \mathbf{b}(s_{ij},t_{ij}) - \alpha_{ij}^{(\ell)}\boldsymbol{\beta}_J^\top \mathbf{b}(s_{ij},t_{ij})\right\}^2,$$

   subject to

$$\int\int_{\mathcal{T}}\left\{\widehat{\boldsymbol{\beta}}_m^\top \mathbf{b}(s,t)\right\}\left\{\boldsymbol{\beta}_J^\top \mathbf{b}(s,t)\right\}dsdt = 0, \quad m = 1, \ldots, J-1,$$

   and

$$\int\int_{\mathcal{T}}\left\{\boldsymbol{\beta}_J^\top \mathbf{b}(s,t)\right\}^2 dsdt = 1.$$

4. Repeat Steps 2 and 3 until convergence.

For each given value of M, the number of total FPCs, we select the optimal number of basis functions using the cross-validation procedure. To achieve the optimal smoothness control, we may loop through all the candidate values of the number of basis functions and select the optimal one with the smallest cross-validation criterion. To be specific, we divide the observed images into F folds, and perform the procedure on the $F-1$ folds while holding the remaining one out for testing. For simplicity, we assume the numbers of basis functions in the first and second dimensions are the same, that is, $L_1 = L_2 = L$. Under a candidate value of L, given the estimated mean function $\widehat{\mu}(s,t)$, FPCs $\widehat{\psi}_m(s,t)$, we may compute the FPC scores $\hat{\alpha}_{im}$ on the test dataset by regressing the observed $y_{ij} - \widehat{\mu}(s_{ij},t_{ij})$ against $\widehat{\psi}_m(s_{ij},t_{ij})$. The cross-validation score for such a fold can be calculated as

$$\text{CV}(L) = \sum_{i=1}^{n}\frac{1}{n_i}\sum_{j=1}^{n_i}\left\{y_{ij} - \widehat{\mu}(s_{ij},t_{ij}) - \sum_{m=1}^{M}\hat{\alpha}_{im}\widehat{\psi}_m(s_{ij},t_{ij})\right\},$$

Subsequently, we repeat the above procedure, leaving 1-fold for testing, and sum up the cross-validation score across all the folds as the final criterion.

Regarding the choice of the number of FPCs, we select the M by defining the relative proportion of variance explained by the top FPCs:

$$\sum_{m=1}^{M-1} \widehat{\text{var}}(\alpha_{im}) \Big/ \sum_{m=1}^{M} \widehat{\text{var}}(\alpha_{im}) = 1 - \widehat{\text{var}}(\alpha_{iM}) \Big/ \sum_{m=1}^{M} \widehat{\text{var}}(\alpha_{im}),$$

where $\widehat{\text{var}}(\alpha_{im})$ is the sample variance of the obtained estimates of α_{im} for $m = 1, \ldots, M$. Here, the sample variance of the FPC scores is computed based on an optimally chosen value of L for each given M. The cutoff value for this relative proportion of variance is chosen in a way similar to the one in the regular PCA procedure. It is either prespecified, for example, 90% or 95%, or determined adaptively based on the data, that is, at the elbow point where the observed $\widehat{\text{var}}(\alpha_{im})$ appears to decrease and reach a plateau.

2.3. Missing Data Restoration

When a certain data point on the new sampled data surface is missing, we may recover its value using the fitted FPCs. To be specific, after obtaining the basis coefficients $\widehat{\boldsymbol{\beta}}_m$ based on the training images, the estimate for the mth FPC is

$$\widehat{\psi}_m(s,t) = \widehat{\boldsymbol{\beta}}_m^{\top} \mathbf{b}(s,t), \qquad m = 1, \ldots, M.$$

Given a new set of possibly incomplete samples $\{(s_j, t_j, y_j);\ j = 1, \ldots, n_0\}$ for a damaged image surface $X(s,t)$, we may obtain the FPC scores $\widehat{\alpha}_m, m = 1, \ldots, M$, as the minimizer of the following loss function

$$\sum_{j=1}^{n_0} \left\{ y_j - \widehat{\mu}(s_j, t_j) - \sum_{m=1}^{M} \alpha_m \widehat{\psi}_m(s_j, t_j) \right\}^2,$$

which can be solved easily as a least-square fitting problem.

The recovered image surface is thus,

$$\widehat{X}(s,t) = \widehat{\mu}(s,t) + \sum_{m=1}^{M} \widehat{\alpha}_m \widehat{\psi}_m(s,t),$$

which can be used for recovering data on the damaged region.

3. Theoretical Properties

It can be shown that the minimizer $\widehat{\psi}_1$ of the loss functions in (1) converges to the first FPC, and the minimizer $\widehat{\psi}_J$ of (3) converges to the Jth FPC.

Denote $\mathbf{t} = (s,t)$ and let $X^*(\mathbf{t}) = X(\mathbf{t}) - \mu(\mathbf{t})$ be the underlying centered stochastic process. Let $\{\xi_m, \psi_m^0 : m = 1, 2, \ldots, \infty\}$ denote the Mercer expansion with kernel $K(\mathbf{t}, \mathbf{t}') = E\{X^*(\mathbf{t})X^*(\mathbf{t}')\}$. To be specific, there are random variables ξ_m and square integrable functions ψ_m^0 on $\mathcal{T}$ such that

$$X^*(\mathbf{t}) = \sum_{m=1}^{\infty} \xi_m \psi_m^0(\mathbf{t}).$$

Theorem 1. Under assumptions (A1)–(A3) in Appendix, supplementary materials, the minimizer $\widehat{\psi}_1$ of (1) converges to ψ_1^0 in $L^2(\mathcal{T})$ in probability as $n \to \infty$.

Theorem 2. Under assumptions (A1)–(A3) in Appendix, supplementary materials, the minimizer $\widehat{\psi}_J$ of (3), where $M = 2, \ldots, \infty$, converges to ψ_J^0 in $L^2(\mathcal{T})$ in probability as $n \to \infty$.

The two theorems point to a novel regression-based framework for conducting FPCA and estimating the FPCs, as opposed to the conventional methods that are based on decomposing the covariance functions. The proofs of the theorems primarily rely upon the asymptotics of the empirical processes (Pollard 1989), which are given in the Appendix, supplementary materials.

4. Applications

4.1. ADNI Brain MRI Scans

To demonstrate the scientific usefulness of the proposed two-dimensional FPCA in the field of medical imaging, we apply it to the T1-weighted structural brain MRI data obtained from the publicly available Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset (Jack et al. 2008). The dataset consists of subjects at different stages of clinically diagnosed Alzheimer's Disease, ranging from cognitive normal, to mild cognitive impairment, and ultimately Dementia of Alzheimer's Type (DAT). The MRI images are collected longitudinally from up to 13 time points across 120 months for participant with age ranging from 55 to 95 with mean ages at 75. To ensure the best representation of the diagnosis groups, we collected baseline data (i.e., the first time point acquired at the first time of visit) from the group of subjects ($n = 43$) who stay as cognitive normal across all the time points, as well as the group of patients ($n = 28$) who has been diagnosed as DAT across all the time points. We term the first group as stable normal control (sNC) and the second group stable Dementia of Alzheimer's Type (sDAT).

Before applying the two-dimensional FPCA on the brain MRI images, all volumetric images of the baseline scans from the subjects in the sDAT and sNC groups are transformed into the same sterotaxis space by affinely registering them to the same template using the image processing pipelines based on FreeSurfer version 5.3.0 (Ma et al. 2018). The middle axial slice of each volume are then extracted, upon which the two-dimensional FPCA is applied. The mean values of all the affinely co-registered images are first subtracted from the grayscale matrix of each matrix, and the FPCs are computed based on the residual.

We refer to the product of the FPC score and the FPC, that, $\alpha_{im}\psi_m(s,t)$, as "additional factor" (AF), and an image can thus, be viewed as the summation of the mean function and M additional factors plus the random noise. Figure 1 shows the plot of the averaged top five additional factors of the sDAT and sNC groups. The first AF component demonstrates an overall pattern of negative AF for the sDAT group (shown in red color), and positive AF for the sNC group (shown in blue color). The second AF component shows the contrast in the ventricle region (close to the posterior part of the brain, as indicated by the blue arrow), which again is negative for the sDAT group and positive for the

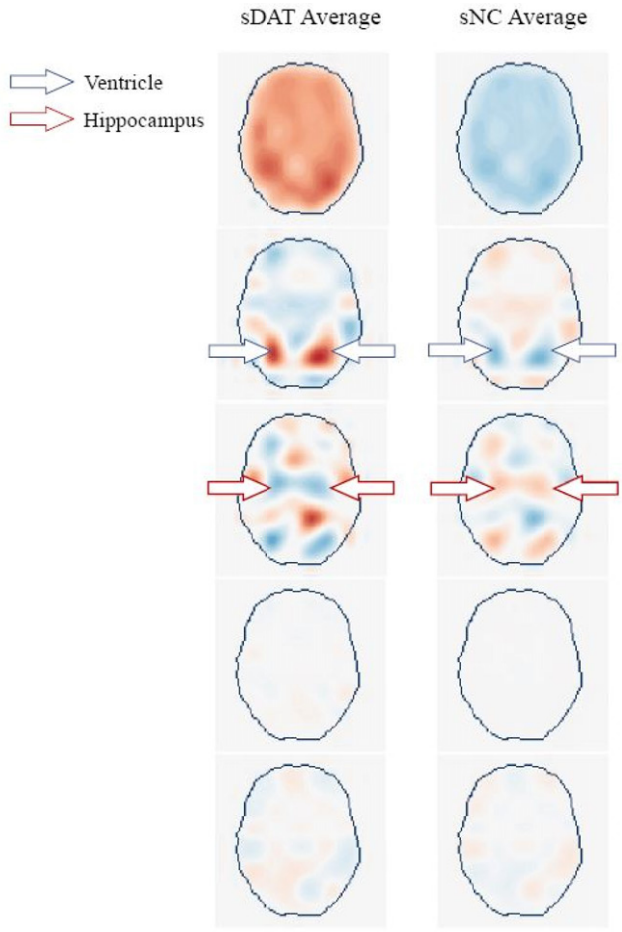

Figure 1. The averaged plots of the top five additional factors (AF), defined as functional principal component functions times the corresponding FPC scores, that is, $\alpha_{im}\psi_m(s,t)$, in the sDAT and sNC groups.

sNC group. The third AF component highlights the additional groupwise difference around the hippocampus region (around central part in the axial view of brain, as indicated by the red arrow), as well as a few cortical regions (at the surrounding part of the brain). The fourth and fifth AF components do not appear to have significant association with any particular anatomical regions in the brain.

Figure 2 further shows six examples of the original brain scans and their top functional principal components (3 from the sDAT group and 3 from the sNC group). The figures shows consistent patterns of the significant AF region across different subject within both the sDAT (3 columns in the left) and the sNC (3 columns in the right), which align well with the significant regions shown in the averaged images (Figure 1). Some additional variation appears among different subjects, which is more evident in the sDAT group (e.g., sharper AF shown in the third FPC of Image 3, and the fifth FPC of Image 1, which may reflect the nature of diversity in the diseased group (sDAT) in which different subjects at the different stage of the disease may present different level of pathologies in their local structures.

In terms of the proportion of variance explained, Figure 3 shows the values of variance, and the relative increase in the total variance explained for each FPC from the brain scans. The first five FPCs constitute the majority of explained variance, with the higher degree of FPC explaining smaller proportion of variance. As in the scree plot of regular PCA, the turning points or elbow occurs at the second FPC.

Moreover, under the framework of functional regression, it is feasible to regress the FPC scores against the outcomes. Such a model is known as the functional principal component regression (Reiss and Ogden 2007). It would be of interest to explore the feasibility of such a model for images. We use the first five FPC scores as the covariates to predict the status of sDAT and sNC under a generalized linear model. The results from this regression model are summarized in Table 1. We observe that the first three FPC scores are significant, with p-values being less than 0.10. In particular, the third FPC score is the most significant covariate, having the smallest p-value of 0.0028. As the second and third FPCs encompass the contrast in the ventricle and the hippocampus regions, the result is in line with the scientific finding that ventricle and hippocampus are the regions primarily affected by DAT. It is also worth noting that even though the elbow turning occurs at the second FPC in Figure 3, the truly useful predictors for DAT are the first three FPCs. The FPC that explains the largest proportion of variance might not necessarily be the one that has the highest predictive value. Moreover, the predictive performance of such a model based on the FPC scores is evaluated using a cross-validated approach, resulting in an AUC of 0.72 and the predictive accuracy of 0.78, respectively. While there are other models that directly targets the outcome and may potentially lead to even better predictive performance, for example, the scalar-on-image regression (Wang and Zhu 2017; Kang et al. 2018) and deep learning models (Gao et al. 2020), our results demonstrate the feasibility of functional principal component regression for images.

Moreover, we consider a more practical model setting with confounders and information from brain imaging studies. In addition to two confounder variables (age and gender), the model uses hippocampus structural volume derived from the analysis of brain MRI scan as the primary predictor in the model, which is constructed as follows. The original T1-weighted brain MRI images are first segmented into three tissue types: the gray matter, white matter, and cerebrospinal fluid using the FreeSurfer package version 5.3. The gray matter and cerebrospinal fluid were parcellated into 91 different anatomical regions of interest using FreeSurfer's parcellation pipelines. Among these regions, the hippocampus is chosen and its volume is extracted using a multi-atlas segmentation framework (Ma et al. 2018).

We sequentially add the FPC scores derived from the two-dimensional FPCA into the baseline model using a forward selection rule. The second and third FPC scores are selected into the model. The results from this regression model are summarized in Table 2. We observe that the hippocampus structural volume is highly significant, as it is derived directly from the brain image segmentation. The second and third FPC scores are also significant, indicating that the information from the FPCs may complement the primary predictor. In terms of the predictive accuracy, after incorporating the information of the FPCs, the 5-fold cross-validated AUC increases from 0.894 to 0.904, and the accuracy increases from 0.1127 to 0.1268, that is, an improvement in the predictive performance.

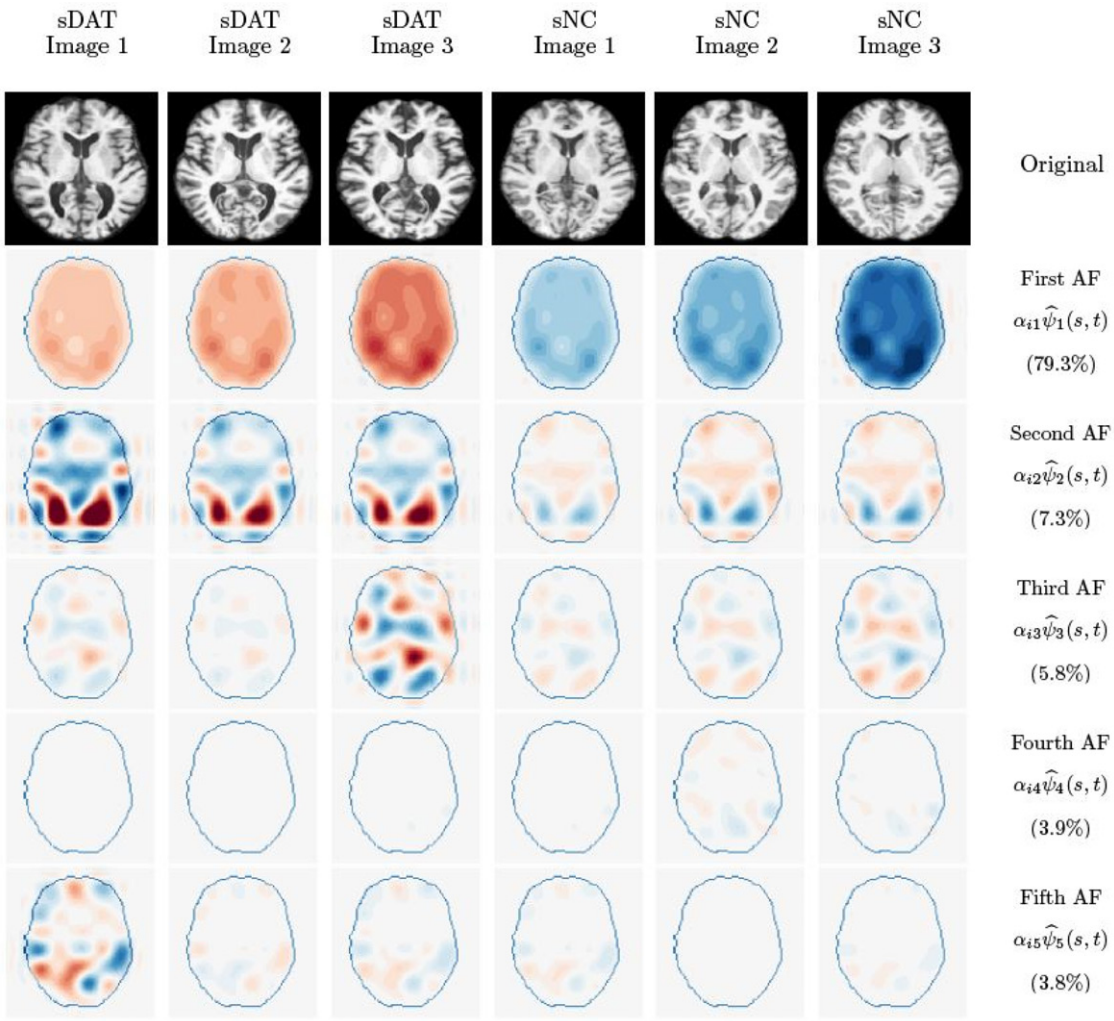

Figure 2. Six examples of brain MRI scans in the ADNI dataset (top row) and the plots of the top additional factors, defined as functional principal component functions times the corresponding FPC scores, that is, $\alpha_{im}\psi_m(s,t)$. The percentages in parenthesis are the proportion of variation explained.

4.2. MNIST Handwritten Digits

The MNIST dataset is a well-known database consisting of grayscale handwritten digits and is commonly used as a benchmark for empirical evaluation of image processing and machine learning (Yann et al. 2013). We apply the two-dimensional functional principal analysis on the MNIST dataset, which contains 70,000 images (of size 28×28 or $n_i = 784$ pixels) of handwritten digits. Each grayscale image is represented by a 28×28 matrix whose values represent the brightness of the pixels. We select images of digits "0" and "1" for our analysis.

Using our method, a two-dimensional FPCA is conducted on images of digit "0" only. For each image, we subtract the mean values of all the images and search for the FPCs on the residual. Figure 4 shows examples of the original images and their top functional principal components. It is worth noting that the functional principal components, as a representation of the deviation from the mean surface function, may take negative value, while a pixel in a grayscale image can only take integer values between 0 and 255. Therefore, for presentation purposes, the absolute value of the minimum negative value is added to the functional components. As a result, the brighter the pixel in the image of the FPC, the larger/positive the FPC score; whereas the darker pixel correlates with negative values of the FPC.

Interestingly, the obtained FPCs correspond to various handwriting style or calligraphic pattern of the digit "0". In general, the digit "0" can be written in a manner that is slightly skewed to the left (e.g., image 1, 3 and 5), or in a manner that is rounded and full (e.g., image 2 and 4). Correspondingly, the pixels in the FPCs that relate to the styles of handwriting are highlighted. For example, comparing the first FPC of image 1 against that of image 2, it is evident that the inner ring (skewed to the left) is very bright for image 1 and entirely dark for image 2, whereas the outer ring (rounded) is dark for image 1 and bright for image 2. Figure 5 shows similar results of the two-dimensional functional principal component analysis conducted on images of the digit "1" only. It is evident that the first FPC corresponds to the direction of skewness, for example, images 1 and 2 are left-skewed, whereas images 3 and 4 right-skewed. Moreover, image 5 has a different handwriting pattern from the rest of the images, and therefore, it has a highlighted component function at the seventh FPC.

Furthermore, a two-dimensional functional principal component analysis is performed on images of digits "0" and "1" combined. Compared with conducting separate functional principal component analysis on digit "0" or digit "1" only, where the focus is on the variation within the handwriting

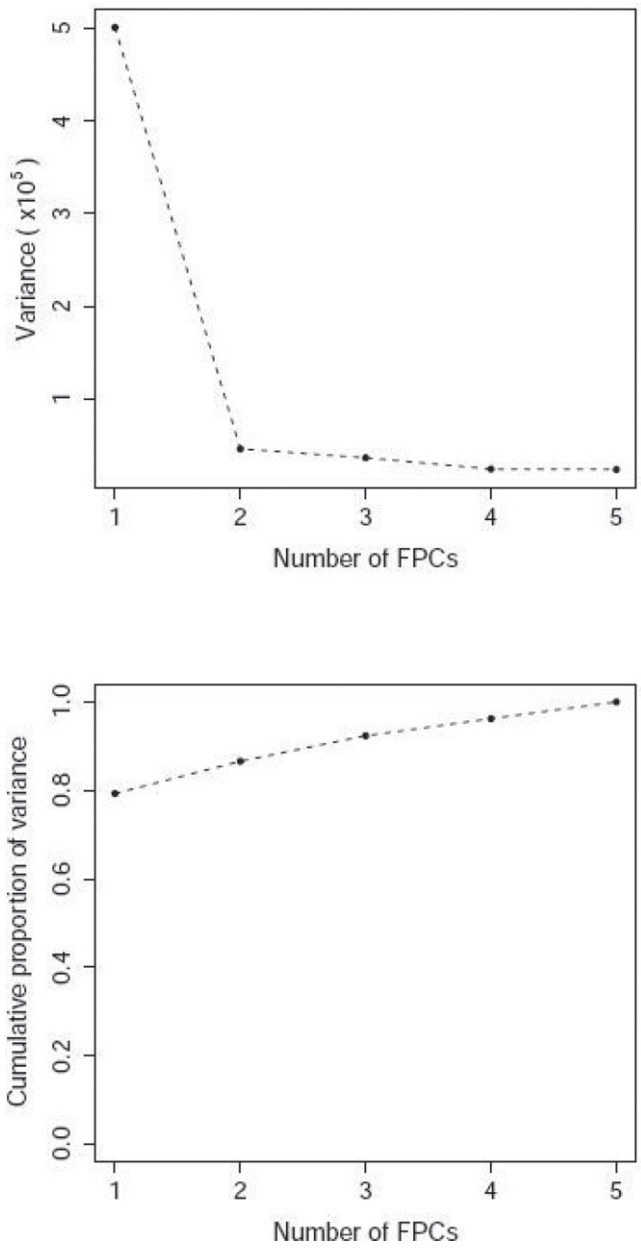

Figure 3. Variance and cumulative proportion of variance for each components in the analysis of brain MRI scans.

Table 1. The regression results from generalized linear model using FPC scores to predict the sDAT/sNC group in the ADNI brain scan dataset.

	Estimate	Std. Error	*p*-value
Intercept	−0.5116	0.2866	0.0743
FPC1	−0.0007	0.0004	0.0766
FPC2	0.0039	0.0016	0.0140
FPC3	0.0050	0.0017	0.0028
FPC4	0.0003	0.0017	0.8854
FPC5	−0.0015	0.0018	0.4018

patterns of the same digit, such a combined analysis naturally has the major variation resulting from the difference between "0" and "1". The mean of all the images is first subtracted from the pixel matrix in each image. The original images and the top functional components are shown in Figure 6. The first FPC shows a very evident pattern for distinguishing between digits

Table 2. The regression results from generalized linear model using hippocampus structural volume, gender, age and the second and third FPC scores to predict the sDAT/sNC group in the ADNI brain scan dataset.

	Estimate	Std. Error	*p*-value
Intercept	38.8381	12.8981	0.0026
FPC2	0.0060	0.0027	0.0309
FPC3	0.0043	0.0025	0.0835
Gender	1.5586	1.0069	0.1216
Age	−0.2791	0.1194	0.0194
Hippocampus	−26.6305	7.1847	0.0002

"0" and "1". For digit "0", the outer ring is brightened in the first FPC, whereas for digit "1", the inner skewed portion is highlighted.

In terms of the proportion of variance explained, Figure 7 shows the values of variance, and the relative increase in the total variance explained for each component in the FPCA conducted on digit "0" only, on digit "1" only, and on digits "0" and "1" combined. The first three FPCs constitute the majority of explained variance, and the usual pattern of diminishing increment as in the scree plot of regular PCA is observed. Moreover, as the digits "0" and "1" are distinctively different, the first FPC in the analysis of digits "0" and "1" combined explains a larger portion of the variances than that in the analysis of digit "0" only.

We experiment with the functionality of image completion or data restoration, that is, cropping a portion of the pixels from the image and then attempting to recover the full image using prediction based on the estimated FPCs. To be specific, we estimate the two-dimensional FPCs $\psi_m(s,t)$ from the training image data. Then we estimate the FPC scores for the damaged image, as outlined in Section 2.3. Based on the fitted FPC scores, we predict the value of a pixel in the cropped portion of the image. Figure 8 shows an example of image completion using FPCA. As a baseline for assessing how well the figure is reconstructed, the first panel in the plot shows the reconstructed figure under complete data. From the shape of the reconstructed plots, we observe that when the missing portion is not overly large, the reconstructed plots under missing data is similar to the one under complete data, indicating the effectiveness of reconstruction. While the proposed method performs relatively well in terms of image recovery, it may fail to render the correct completion when the cropped portion is overly large. It is also worth noting that the fitted images are slightly darker than the original ones. This might be attributed to the fact that the tensor product B-spline tends to fit a smoothed and continuous surface to the sample data points, whereas the images often have very extreme pixel values, that is, the empirical surface is not smooth.

Moreover, as the dimensions of the image are substantially reduced after extracting the top few FPCs, we compare the accuracy and speed of the classification of digits "0" and "1" using the FPC scores as predictors, versus the one using all the $28 \times 28 = 784$ grayscale pixels values as predictors. These FPCs and FPC scores are obtained by implementing FPCA on the combined set of digits "0" and "1". A boosting classifier is trained on the predictors using R package "ada". Both methods result in a prediction misclassification rate of around 0.1%. We have also experimented with extracting FPC scores separately from the set of digits "0" only and from the set of digits "1" only and use them for prediction, which results in a higher misclassification

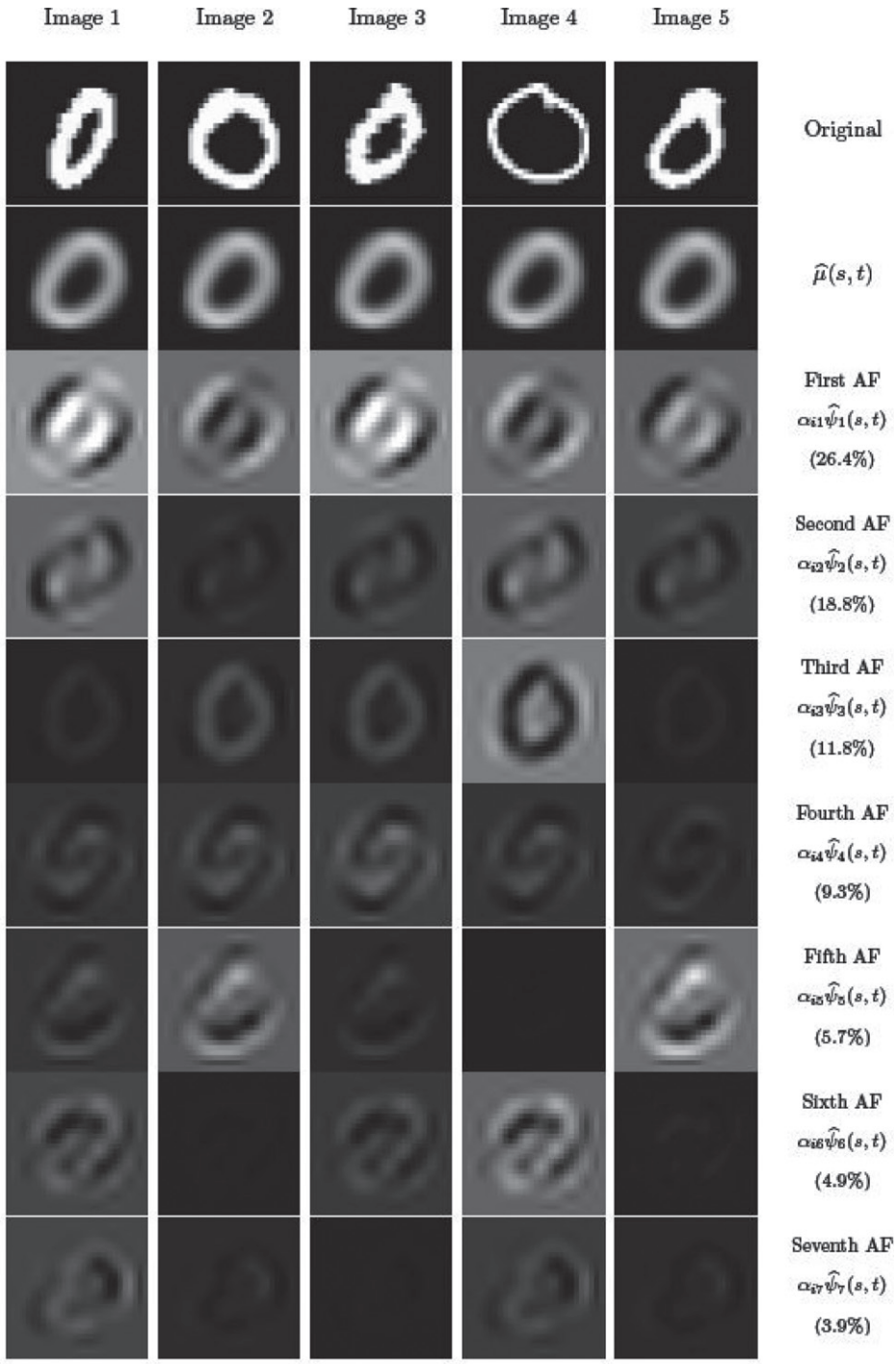

Figure 4. Five examples of digit "0" in the MNIST dataset (top row) and the plots of the top additional factors (AF), defined as functional principal component functions times the corresponding FPC scores, that is, $\alpha_{im}\psi_m(s,t)$. The percentages in parenthesis are the proportion of variation explained.

rate (0.4%). This indicates that conducting FPCA on a combined dataset is more suited for classification.

In terms of the computing time, the proposed method is substantially faster as it requires much fewer predictors. For a fair comparison, all the tuning parameters in the boosting classifiers are the same. On the task of producing classifications for 10,000 images, the method based on FPC scores has a total computing time (the sum of the time of computing the FPC score and the time of producing prediction from the boosting model) of 9.1 sec on a machine with Intel i5-5200 CPU and 12 Gigabytes of RAM, whereas the one based on all the 784 pixel values as predictors takes 38.3 sec; a running time reduction of more than 75% is achieved.

5. Simulation Studies

5.1. Simulation Setup

Simulation studies are conducted to assess the empirical performance of the proposed method. The goal is to evaluate the accuracy of extracting and recovering the FPCs based on our method. In particular, we are interested in how the estimation

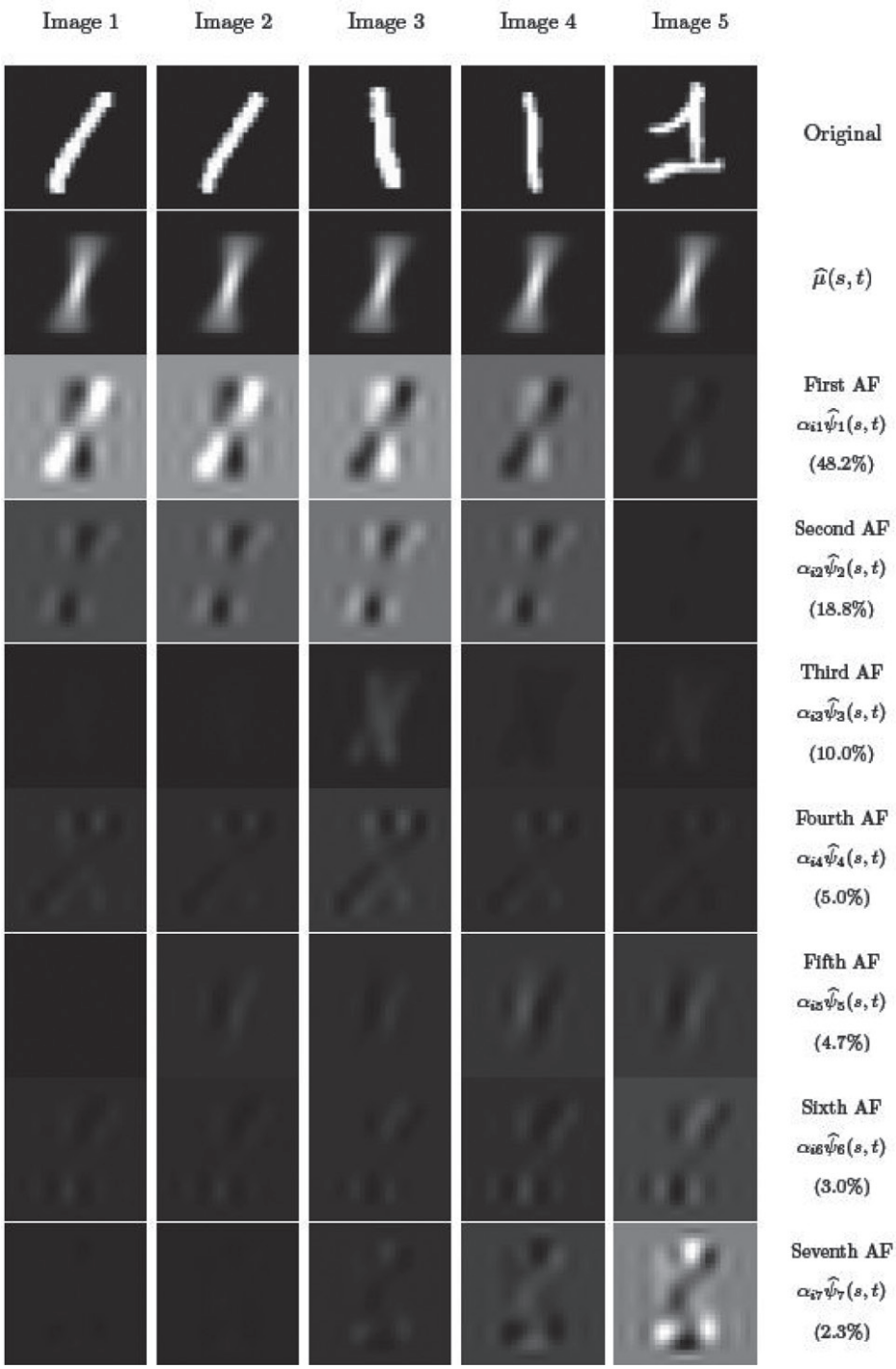

Figure 5. Five examples of digit"1" in the MNIST dataset (top row) and the plots the top additional factors (AF), defined as functional principal component functions times the corresponding FPC scores. The percentages in parenthesis are the proportion of variation explained.

error changes as sample size increases as it is expected that the estimation precision would improve under a larger sample size. Moreover, it is of interest to analyze the effect of the number of B-spline bases (s_1 or s_2) on the estimation accuracy.

We first generate the underlying true two-dimensional surfaces from a series of known FPCs. As the mean function in our method is used as a plugged-in value and the method is expected to perform well as long as the estimator for the mean function is consistent, our focus lies on the estimation of FPCs. We simulate the surfaces $X_i(s,t)$ whose mean function satisfies $\mu(s,t) = 0$, that is, the mean function remains 0. The underlying true two-dimensional surfaces are simulated as the sum product of three sets of FPCs and scores.

$$X_i(s,t) = \alpha_{i1}\psi_1(s,t) + \alpha_{i2}\psi_2(s,t) + \alpha_{i3}\psi_3(s,t), \quad i = 1,\ldots,n.$$

The underlying true FPCs $\psi_1(s,t)$ to $\psi_3(s,t)$ are the same as the first three functional principal component functions from the analysis on the images of digits "0" and "1" in the MNIST dataset, and thus, satisfy $||\psi_j||^2 = 1$, and $\langle\psi_j,\psi_k\rangle = 1$ if $j = k$, and 0 otherwise. These true underlying FPCs are constructed using a tensor product B-spline with $s_1 = s_2 = 12$ basis

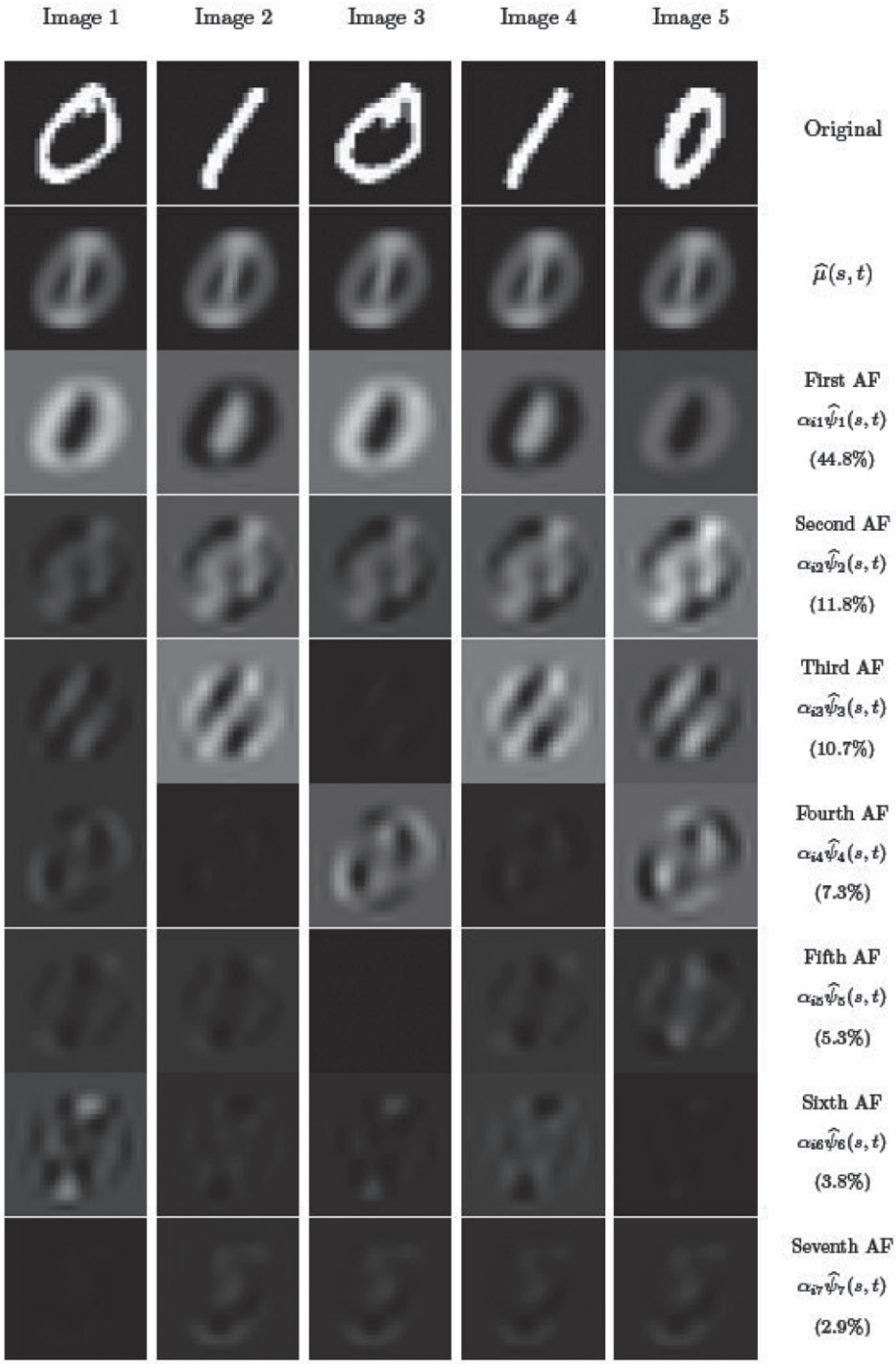

Figure 6. Five examples of digits "0" and "1" in the MNIST dataset (top row) and the plots of the top additional factors (AF), defined as functional principal component functions times the corresponding FPC scores. The percentages in parenthesis are the proportion of variation explained.

functions. The scores α_{i1} to α_{i3} are independently generated from both Gaussian and non-Gaussian distributions.

Having generated the true underlying surfaces, we simulate the observed data points from the surfaces subject to a random measurement error. As in the "MNIST" dataset, the observed data are simulated from a 28×28 grid as $y_{ij} = X_i(s_{ij}, t_{ij}) + \epsilon_{ij}$, where $i = 1, \ldots, 28$ and $j = 1, \ldots, 28$, and the error terms ϵ_{ij} are randomly drawn from a normal distribution $\mathrm{N}(0, 0.25^2)$.

We apply the two-dimensional functional principal component analysis on the simulated dataset and compare the extracted FPCs with the true ones. To measure how well the proposed estimation method approximates the true underlying FPCs, we use the mean squared error (MSE), defined as the mean of the squared differences between the estimated functional principal component functions $\widehat{\psi}_k$ and the true ones ψ_k on the grid,

$$\mathrm{MSE}(\widehat{\psi}_k) = \frac{1}{n_g^2} \sum_{i=1}^{n_g} \sum_{j=1}^{n_g} \left\{ \widehat{\psi}_k(s_{ij}, t_{ij}) - \psi_k(s_{ij}, t_{ij}) \right\}^2, \quad k = 1, 2, 3,$$

where $n_g = 28$ is the number of grid points on each dimension of the surface.

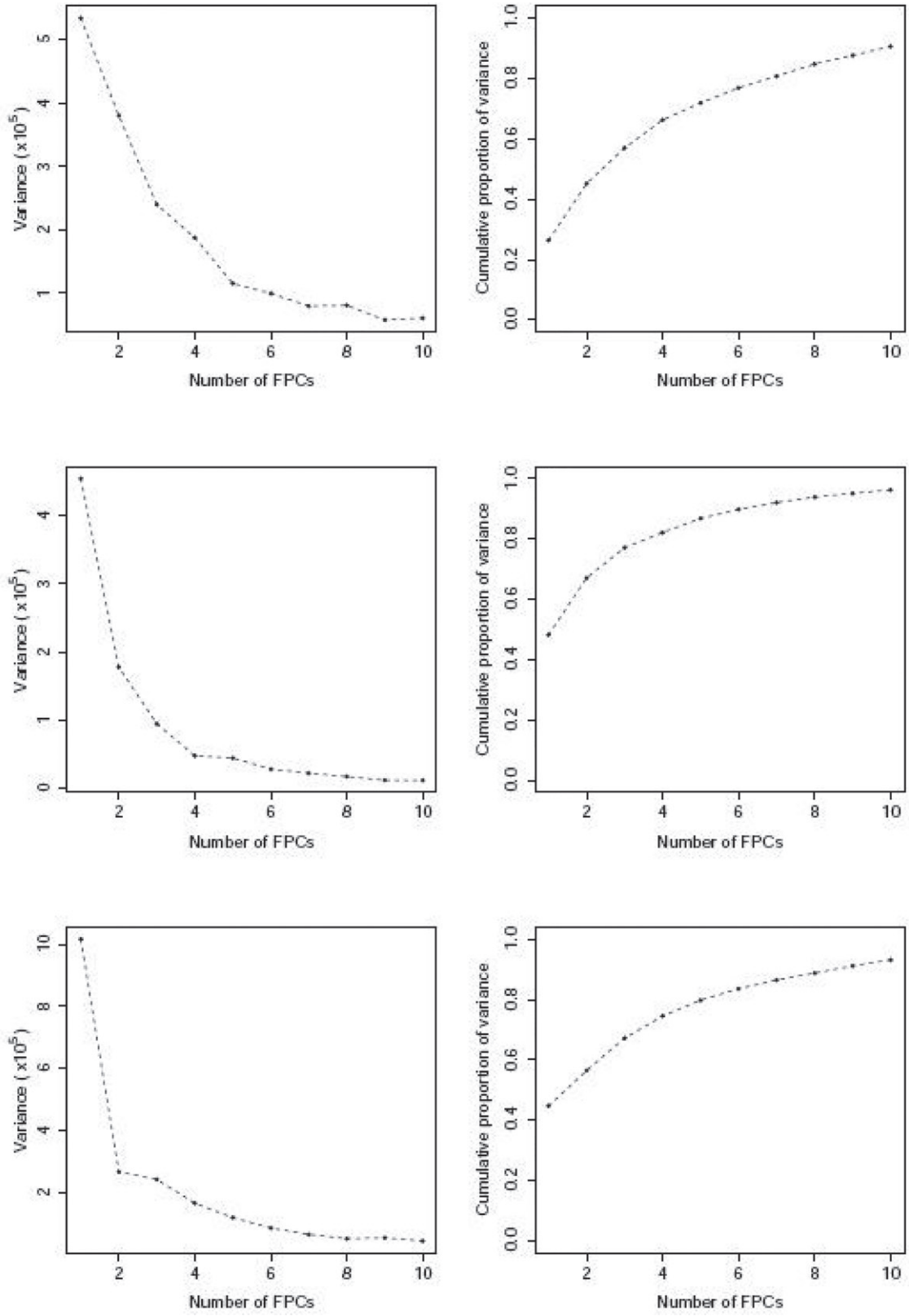

Figure 7. Variance and cumulative proportion of variance for each components in the analysis of digit "0" only (top row), of digit "1" only (middle row), and of digits "0" and "1" (bottom row) in the MNIST dataset.

Moreover, we compute the integrated mean squared error (IMSE), which is defined as

$$\text{IMSE}(\widehat{\psi}_k) = \int \int_{\mathcal{T}} \left\{\widehat{\psi}_k(s,t) - \psi_k(s,t)\right\}^2 dsdt, \qquad k = 1, 2, 3,$$

where $\mathcal{T} = [1, 28]$. The mean squared error and the integrated mean squared error are computed based on 100 data replications. Within each replication, a total of 100 sample data surfaces are generated.

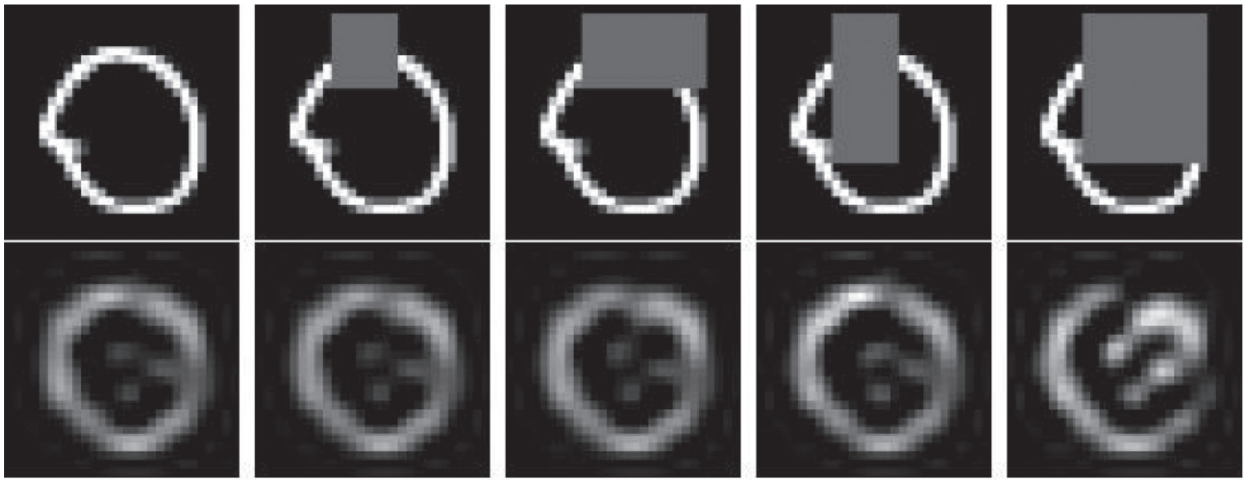

Figure 8. Examples of image completion of digit "0" in the MNIST dataset; the original images (cropped portion colored in gray) in the top row, completed images restored from FPCs in the bottom row.

5.2. Gaussian Scenario

Under the Gaussian scenario, we assume that the FPC scores follow a Gaussian distribution. The FPC scores α_{i1} to α_{i3} are randomly drawn from independent normal distributions with mean 0 and decreasing standard deviations of 100, 50, and 25, respectively,

$$\alpha_{i1} \sim \mathrm{N}(0, 100^2),$$
$$\alpha_{i2} \sim \mathrm{N}(0, 50^2),$$
$$\alpha_{i3} \sim \mathrm{N}(0, 25^2).$$

The decreasing sequence of standard deviations correspond well with the fact that the higher ranked FPC tends to explain a larger proportion of the total variation.

We experiment various values of sample size $n = 50, 100, 200, 400$ and number of spline basis functions $s_1 = s_2 = 8, 12, 16$, and compute the MSE and IMSE based on 100 simulation replications. Our goal is to understand the interplay between the sample size and the number of spline basis functions. Table 3 summarizes the results for various values of n and s_1 or s_2. It is evident that a larger sample size leads to a smaller MSE or IMSE. For example, the MSE of the first FPC when the number of basis functions is $s_1 = s_2 = 12$ decreases from 0.014 to 0.002 as the sample size increases from 50 to 400. However, the decrease is not as evident for the second and third FPCs, as the FPCs of higher order are harder to identify. Moreover, in general, the MSE and IMSE are larger for the second and third FPCs, as the higher-order FPCs contribute to less variation. The overall decreasing trend in the estimation error with respect to the sample size reflects the accuracy of the method.

The estimation performance may deteriorate when the number of basis functions s_1 or s_2 is too small. On the other hand, compared with the optimal scenario where $s_1, s_2 = 12$, a larger value of s_1 or s_2 results in very mild deterioration of the estimation accuracy. For example, under a sample size of 50, the MSE of the second FPC ψ_2 increases from 0.039 to 0.050 when the granularity of the tensor product B-spline becomes overly high ($s_1 = s_2 = 16$). However, when the tensor product B-spline is too blurry, that is, the number of basis functions being insufficiently low ($s_1 = s_2 = 8$), the MSE observes a much larger increase of nearly 10 folds, from 0.039 to 0.339. Moreover, the increase in sample size does not seem to ameliorate the loss in precision caused by the insufficient granularity of the tensor product B-spline; as the sample size increases from 50 to 200, the MSE with $s_1 = s_2 = 16$ is improved to a level similar to the one under the optimal case, whereas the MSE with $s_1 = s_2 = 8$ is still as high as 0.331.

Table 3. The MSE and IMSE of three estimated functional principal component functions for different values of the sample size n and the number of basis functions s_1 and s_2 under Gaussian distribution of FPC scores.

			MSE ($\times 10^{-3}$)			IMSE		
n	s_1	s_2	$\widehat{\psi}_1$	$\widehat{\psi}_2$	$\widehat{\psi}_3$	$\widehat{\psi}_1$	$\widehat{\psi}_2$	$\widehat{\psi}_3$
50	8	8	0.162	0.339	0.513	0.133	0.332	0.573
	12	12	0.014	0.039	0.048	0.012	0.034	0.039
	16	16	0.014	0.050	0.072	0.011	0.044	0.060
100	8	8	0.158	0.336	0.510	0.130	0.329	0.569
	12	12	0.008	0.012	0.008	0.006	0.010	0.006
	16	16	0.009	0.015	0.012	0.007	0.013	0.009
200	8	8	0.155	0.331	0.509	0.127	0.323	0.568
	12	12	0.004	0.031	0.040	0.003	0.026	0.033
	16	16	0.004	0.034	0.057	0.003	0.029	0.047
400	8	8	0.154	0.331	0.508	0.127	0.323	0.567
	12	12	0.002	0.003	0.002	0.001	0.002	0.001
	16	16	0.002	0.004	0.004	0.002	0.003	0.003

5.3. Non-Gaussian Scenario

Under the non-Gaussian scenario, the FPC scores α_{i1} to α_{i3} are independently drawn from shifted gamma distributions with mean 0 and standard deviations similar to the Gaussian case,

$$\alpha_{i1} \sim \mathrm{Gamma}(1, 0.01) - 100,$$
$$\alpha_{i2} \sim \mathrm{Gamma}(1, 0.025) - 40,$$
$$\alpha_{i3} \sim \mathrm{Gamma}(1, 0.05) - 20,$$

that is, a random draw would first be generated from a Gamma distribution and subsequently be deducted from it the mean of the Gamma distribution, such that the generated scores have an expected value of 0.

Table 4. The MSE and IMSE of three estimated functional principal components for different values of the sample size n and the number of basis functions s_1 and s_2 under non-Gaussian distribution of FPC scores.

			MSE ($\times 10^{-3}$)			IMSE		
n	s_1	s_2	$\widehat{\psi}_1$	$\widehat{\psi}_2$	$\widehat{\psi}_3$	$\widehat{\psi}_1$	$\widehat{\psi}_2$	$\widehat{\psi}_3$
50	8	8	0.156	0.335	0.518	0.128	0.328	0.575
	12	12	0.012	0.030	0.025	0.010	0.026	0.022
	16	16	0.008	0.031	0.033	0.007	0.030	0.032
100	8	8	0.155	0.332	0.512	0.127	0.323	0.570
	12	12	0.003	0.009	0.010	0.003	0.008	0.008
	16	16	0.006	0.012	0.012	0.004	0.010	0.009
200	8	8	0.154	0.330	0.508	0.126	0.320	0.568
	12	12	0.002	0.005	0.005	0.002	0.004	0.004
	16	16	0.002	0.006	0.006	0.002	0.004	0.005
400	8	8	0.153	0.330	0.508	0.126	0.320	0.567
	12	12	0.001	0.002	0.002	0.001	0.002	0.002
	16	16	0.001	0.003	0.004	0.001	0.003	0.003

Similar to the Gaussian case, we consider various values of $n = 50, 100, 200, 400$ and $s_1 = s_2 = 8, 12, 16$, and evaluate the precision accuracy using MSE and IMSE averaged from 100 simulation replications. Table 4 summarizes the results for various values of n and s_1 or s_2 under the non-Gaussian case. It is evident that a similar decreasing trend of the MSE and IMSE with respect to n can be observed. Moreover, compared with the Gaussian case, the estimation error is in general smaller; this is because the Gamma distribution has a heavier tail and thus, more extreme values of the FPC scores, which facilitate the contrast within the regression estimation.

6. Conclusions and Discussion

In this article, we have proposed a novel method for identifying major sources of variability in a two-dimensional surface process. The approach is distinctive from the conventional method of two-dimensional FPCA, as it circumvents the need for estimating the de-centered covariance function and then conducting eigen-decomposition on it. Besides, as our approach is regression-based, unlike the conventional two-dimensional FPCA, our method can handle sparse and irregularly sampled data.

Several extensions of our method can be considered as avenues for future research. A critical extension of our work is to develop a methodology for multi-level data surfaces, a typical example of which would be the three-level RGB color model for image processing. Furthermore, in the current example for medical image analysis, we have only demonstrated the application in 2D, that is, the middle sagittal view. This can be potentially be further extended to 3D version of the FPCA where volumetric analysis over the entire 3D brain images can be performed.

Supplementary Materials

Supplementary document: The supplementary document includes the detailed proofs for Theorems 1 and 2. (supplementarydoc.pdf)

R codes: We provide R codes for the simulation studies and the application on the MNIST Handwritten Digits; see the README file for more details. (Rcode_2DFPCA.zip, zip archive).

Acknowledgments

The authors thank to the editor, the associate editor, and two anonymous referees for many insightful comments. These comments are very helpful for us to improve our work.

Funding

This research is supported by the Strategic Partnership grant from the Natural Sciences and Engineering Research Council of Canada (NSERC).

ORCID

Liangliang Wang http://orcid.org/0000-0002-8509-7985
Jiguo Cao http://orcid.org/0000-0001-7417-6330

References

Di, C., Crainiceanu, C. M., and Jank, W. S. (2014), "Multilevel Sparse Functional Principal Component Analysis," *Stat*, 3, 126–143. [1127]

Dong, J., Wang, L., Gill, J., and Cao, J. (2018), "Functional Principal Component Analysis of GFR Curves after Kidney Transplant," *Statistical Methods in Medical Research*, 27, 3785–3796. [1127]

Gao, F., Yoon, H., Xu, Y., Goradia, D., Luo, J., Wu, T., Su, Y., and Alzheimer's Disease Neuroimaging Initiative. (2020). "AD-NET: Age-adjust Neural Network for Improved MCI to AD Conversion Prediction," *NeuroImage Clinical*, 27, 102290. [1131]

Goldsmith, J., Zipunnikov, V., and Schrack, J. (2015), "Generalized Multilevel Function-on-Scalar Regression and Principal Component Analysis," *Biometrics*, 71, 344–353. [1127]

Hall, P., Müller, H. G., and Wang, J. L. (2006), "Properties of Principal Component Methods for Functional and Longitudinal Data Analysis," *The Annals of Statistics*, 34, 1493–1517. [1127]

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. [1128]

Jiang, L., Elord, C., Kim, J. J., Swafford, A. D., Knight, R., and Thompson, W. K. (2021), "Bayesian Multivariate Sparse Functional Principal Components Analysis with Application to Longitudinal Microbiome Multi-Omics Data," available at *https://arxiv.org/pdf/2102.00067* [1128]

Jack, C. R., Bernstein, M. A., Fox, N. C., Thompson, P., Alexander, G., Harvey, D., and Weiner, M. W. (2008), "The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI Methods," *Journal of Magnetic Resonance Imaging*, 27, 685–691. [1130]

James, G. M., Hastie, T. J., and Sugar, C. A. (2000), "Principal Component Models for Sparse Functional Data," *Biometrika*, 87, 587–602. [1127]

Johns, J. T., Crainiceanu, C., Zipunnikov, V., and Gellar, J. (2019), "Variable-Domain Functional Principal Component Analysis," *Journal of Computational and Graphical Statistics*, 28, 993–1006. [1127]

Kang, J., Reich, B. J., Staicu, A. M. (2018), "Scalar-on-Image Regression via the Soft-Thresholded Gaussian Process," *Biometrika*, 105, 165–184. [1131]

Lawson, C. L., and Hanson, R. J. (1974), *Solving Least Squares Problems*, Englewood Cliffs: Prentice Hall. [1129]

Lin, Z., Wang, L., and Cao, J. (2016), "Interpretable Functional Principal Component Analysis," *Biometrics*, 72, 846–854. [1127]

Ma, D., Popuri, K., Bhalla, M., Sangha, O., Lu, D., Cao, J., Jacova, C., Wang, L., and Beg, M. F. (2018), "Quantitative Assessment of Field Strength, Total Intracranial Volume, Sex, and Age Effects on the Goodness of Harmonization for Volumetric Analysis on the ADNI Database," *Human Brain Mapping*, 40, 125–136.

Nie, Y., Wang, L., Liu, B., and Cao, J. (2018), "Supervised Functional Principal Component Analysis," *Statistics and Computing*, 28, 713–723. [1127]

Nie, Y. and Cao, J. (2020), "Sparse Functional Principal Component Analysis in a New Regression Framework," *Computational Statistics and Data Analysis*, 152, 107016. [1127]

Nie, Y., Yang, Y., Wang, L., and Cao, J. (2022), "Recovering the Underlying Trajectory from Sparse and Irregular Longitudinal Data," *Canadian Journal of Statistics*. https://doi.org/10.1002/cjs.11677. [1127,1128]

Pezzulli, S., and Silverman, B. (1993), "Some Properties of Smoothed Principal Components Analysis for Functional Data," *Computational Statistics*, 8, 1–16. [1127]

Pollard, D. (1989), "Asymptotics via Empirical Processes," *Statistical Science*, 4, 341–354. [1130]

Ramsay, J. O., and Dalzell, C. (1991), "Some Tools for Functional Data Analysis," *Journal of the Royal Statistical Society*, Series B, 53, 539–572. [1127]

Reiss, P. T, and Ogden, R. T. (2007), "Functional Principal Component Regression and Functional Partial Least Squares," *Journal of the American Statistical Association*, 102, 984–999.

Rice, J. A., and Silverman, B. W. (1991), "Estimating the Mean and Covariance Structure Nonparametrically When the Data are Curves," *Journal of the Royal Statistical Society*, Series B, 53, 23–243. [1127]

Sang, P., Wang, L., and Cao, J. (2017), "Parametric Functional Principal Component Analysis," *Biometrics*, 73, 802–810. [1127]

Shi, H., Dong, J., Wang, L., and Cao, J. (2021), "Functional Principal Component Analysis for Longitudinal Data with Informative Dropout," *Statistics in Medicine*, 40, 712–724. [1127]

Shou, H., Zipunnikov, V., Crainiceanu, C. M., and Greven, S. (2015), "Structured Functional Principal Component Analysis," *Biometrics*, 71, 247–257. [1127]

Silverman, B. W. (1996), "Smoothed Functional Principal Components Analysis by Choice of Norm," *The Annals of Statistics*, 24, 1–24. [1127]

Suarez, A. J., and Ghosal, S. (2017), "Bayesian Estimation of Principal Components for Functional Data," *Bayesian Analysis*, 12, 311–333. [1128]

Thompson, W., and Rosen, O. (2008), "A Bayesian Model for Sparse Functional Data," *Biometrics*, 64, 54–63. [1127]

Wang, X., and Zhu, H. (2017), "Generalized Scalar-on-Image Regression Models via Total Variation, "*Journal of the American Statistical Association*, 112, 1156–1168. [1131]

Xiao, L., Zipunnikov, V., Ruppert, D., and Crainiceanu, C. (2016), "Fast Covariance Estimation for High-Dimensional Functional Data," *Statistics and Computing*, 26, 409–421. [1127]

Yann, L., Cortes, C., and Burges, C. J. C. (2013), "MNIST Handwritten Digit Database." Available at http://yann.lecun.com/exdb/mnist. [1132]

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. [1127]

Zipunnikov, V., Caffo, B., Yousem, D. M., Davatzikos, C., Schwartz, B. S., and Crainiceanu, C. (2011), "Multilevel Functional Principal Component Analysis for High-Dimensional Data," *Journal of Computational and Graphical Statistics*, 20, 852–873. [1127]