

STAT 591 Summary Report

Functional Data Analysis for Sparse Longitudinal Data
Fang Yao, Hans-Georg Müller & Jane-Ling Wang

Wangfei Wang
wwang75@uic.edu

April 19, 2020

1 SUMMARY OF CONTRIBUTIONS

Functional principal components (FPC) analysis can reduce random trajectories of random curves to a set of FPC scores, and therefore is popular in longitudinal data analysis. For a given sample of random trajectories, FPC analysis characterizes the dominant mode of variation around an overall mean trend function.

In longitudinal data analysis, there are extensive research of FPC analysis on repeated measures at a dense grid of regularly spaced time points. However, it is not uncommon that repeated measurements are infrequent, and these measurements are irregularly spaced per subject. In this situation, FPC analysis has limitations because of the sparse repeated measurements. The authors summarized a few available models that can be applied to irregular grid of time measurements. For example, kernel-based functional principal components analysis for repeated measurements with irregular time points was proposed by Staniswalis and Lee [refs]. Their method was further studied by other groups [refs]; however, when the measurement time points vary widely across individuals and the measurements per subject is very sparse (i.e., one or two measurements per subject), the FPC scores cannot be approximated by the usual integration method. Some other groups proposed that by using linear mixed models or reduced-rank mixed effects models, they could use B-splines to model the individual curves with random coefficients [refs]. But because of the complexity of the models, the asymptotic properties of the estimated components were not investigated.

Taken together, the authors proposed a simpler and more straightforward method to determine eigenfunctions, which they represent the trajectories directly using the Karhunen-Loève expansion. Their contributions include:

- They proposed a version of functional principal components (FPC) analysis, in which they framed the FPC scores as conditional expectations. And thus they coined this method “principal components analysis through conditional expectation (PACE)”.
- In the model, they took into account the additional measurement errors.
- They derived the asymptotic consistency properties.
- They derived the asymptotic distribution needed for obtaining point-wise confidence intervals for individual trajectories.

2 INNOVATION

- The proposed conditional model is designed for sparse and irregular longitudinal data.
- Under Gaussian assumptions, the authors showed that estimation of individual FPC scores are the best prediction; and under non-Gaussian assumption, they provide estimates for best linear prediction.
- One-curve-leave-out cross-validation was proposed to choose auxiliary parameters.
- Akaike information criterion (AIC) was used for faster computation to select eigenfunctions.

3 FUNCTIONAL PRINCIPAL COMPONENTS ANALYSIS FOR SPARSE DATA

Model with Measurement Errors

The authors modeled the sparse functional data as noisy sampled points from trajectories. These trajectories are assumed independent realizations of a smooth random function with unknown mean $\mathbb{E} X(t) = \mu(t)$ and covariance function $cov(X(s), X(t)) = G(s, t)$, where domain of $X(\cdot)$ is \mathcal{T} . It was assumed that G has an orthogonal expansion in terms of eigenfunction ϕ_k and eigenvalues λ_k : $G(s, t) = \sum_k \lambda_k \phi_k(s) \phi_k(t)$, $t, s \in \mathcal{T}$, where $\lambda_1 \geq \lambda_2 \geq \dots$. Assuming Y_{ij} is the j th observation of the random function $X(\cdot)$ made at a random time T_{ij} and let ϵ_{ij} be the measurement errors that are iid and are independent of random coefficients ξ_{ik} , where $i = 1, \dots, n; j = 1, \dots, N_i; k = 1, 2, \dots$, the authors constructed a model:

$$Y_{ij} = X_i(T_{ij}) + \epsilon_{ij} \quad (1)$$

$$= \mu(T_{ij}) + \sum_{k=1}^{\infty} \xi_{ik} \phi_k(T_{ij}) + \epsilon_{ij}, \quad T_{ij} \in \mathcal{T} \quad (2)$$

where $\mathbb{E} \epsilon_{ij} = 0$, $\text{var}(\epsilon_{ij}) = \sigma^2$

Estimation of the Model Components

Under the assumption that the mean, covariance and eigenfunctions are smooth, the authors first estimated the mean function μ based on the pooled data from all individuals. First, μ can be estimated by minimizing the following equation (3) respect to β_0 and β_1 , and obtained as $\hat{\mu}(t) = \hat{\beta}_0(t)$. Denote kernel functions $\kappa_1 : \mathbb{R} \rightarrow \mathbb{R}$ and $\kappa_2 : \mathbb{R}^2 \rightarrow \mathbb{R}$ that satisfy several conditions (omitted here; see appendix of the paper).

$$\sum_{i=1}^n \sum_{j=1}^{N_i} \kappa_1\left(\frac{T_{ij} - t}{h_\mu}\right) \{Y_{ij} - \beta_0 - \beta_1(t - T_{ij})\}^2 \quad (3)$$

$$cov(Y_{ij}, Y_{il}) = cov(X(T_{ij}), X(T_{il})) + \sigma^2 \delta_{ij} \quad (4)$$

where $\delta_{jl} = 1$ if $j = l$ and 0 otherwise. Denote “raw” covariances: $G_i(T_{ij}, T_{il}) = (Y_{ij} - \hat{\mu}(T_{ij}))(Y_{il} - \hat{\mu}(T_{il}))$. It can be shown that $\mathbb{E}[G_i(T_{ij}, T_{il}) | T_{ij}, T_{il}] \approx cov(X(T_{ij}), X(T_{il})) + \sigma^2 \delta_{jl}$, and thus only $G_i(T_{ij}, T_{il}), j \neq l$ should be included for the covariance surface smoothing step. One-curve-leave-out cross-validation is used to select smoothing parameter.

Denote $\hat{G}(s, t)$ be the estimate of $G(s, t)$. The local linear surface smoother for $G(s, t)$ can be estimated by minimizing the following equation (5) with respect to $\beta = (\beta_0, \beta_{11}, \beta_{12})$, yielding estimate $\hat{G}(s, t) = \hat{\beta}_0(s, t)$:

$$\sum_{i=1}^n \sum_{1 \leq j \neq l \leq N_i} \kappa_2\left(\frac{T_{ij} - s}{h_G}, \frac{T_{il} - t}{h_G}\right) \times \{G_i(T_{ij}, T_{il}) - f(\beta, (s, t), (T_{ij}, T_{il}))\}^2 \quad (5)$$

Functional Principal Components Analysis Through Conditional Expectation

Asymptotic Confidence Bands for Individual Trajectories

4 ASYMPTOTIC PROPERTIES

5 SIMULATION STUDIES

6 APPLICATIONS

Longitudinal CD4 Counts

Yeast Cell Cycle Gene Expression Profiles