Statistica Sinica **32** (2022), 1939-1960
doi:https://doi.org/10.5705/ss.202020.0445

SPARSE FUNCTIONAL PRINCIPAL COMPONENT ANALYSIS IN HIGH DIMENSIONS

Xiaoyu Hu and Fang Yao

Peking University

Abstract: Existing functional principal component analysis (FPCA) methods are restricted to data with a single or finite number of random functions (much smaller than the sample size n). In this work, we focus on high-dimensional functional processes where the number of random functions p is comparable to, or even much larger than n. Such data are ubiquitous in various fields, such as neuroimaging analysis, and cannot be modeled properly by existing methods. We propose a new algorithm, called sparse FPCA, that models principal eigenfunctions effectively under sensible sparsity regimes. The sparsity structure motivates a thresholding rule that is easy to compute by exploiting the relationship between univariate orthonormal basis expansions and the multivariate Karhunen–Loève representation. We investigate the theoretical properties of the resulting estimators, and illustrate the performance using simulated and real-data examples.

Key words and phrases: Basis expansion, multivariate Karhunen-Loève expansion, sparsity regime.

1. Introduction

Functional data are commonly encountered in modern statistics, and dimension reduction plays a key role, owing to the infinite dimensionality of such data. As an important tool for dimension reduction, functional principal component analysis (FPCA) is optimal in the sense that the integrated mean squared error is efficiently minimized, which has wide applications in functional regression, classification, and clustering (Rice and Silverman (1991); Yao, Müller and Wang (2005a,b); Müller and Stadtmüller (2005); Hall and Hosseini-Nasab (2006); Hall and Horowitz (2007); Horváth and Kokoszka (2012); Dai, Müller and Yao (2017); Wong, Li and Zhu (2019)). Despite progress being made in this field, existing methods often involve a single or finite number of random functions. In this study, we focus on modeling principal eigenfunctions of p random functions, where p is comparable to, or even much larger than the sample size n, that is, the number of subjects. Such high-dimensional functional data are becoming increasingly avail-

Corresponding author: Fang Yao, School of Mathematical Sciences, Center for Statistical Science, Peking University, P.R. China. E-mail: fyao@math.pku.edu.cn.