Principal Component Analysis of Functional Data based on Constant Numerical Characteristics

Meiling Chen[*,**] Huiwen Wang[*,**]

[*]School of Economics and Management, Beihang University, Beijing 100191, China
[**]Research Center of Complex Data Analysis, Beihang University, Beijing 100191, China
mlchen@foxmail.com,

Abstract. A new approach to principal component analysis (PCA) is proposed for functional data. In prevailing methods of functional principal component analysis (FPCA), the definition of a mean is in the form of a function. However, data centralisation based on this kind of mean actually obtains a residual function. The result of FPCA, given its matrix of residual functions, may thus fail to present the essential variation of the original data. Besides, applications in FPCA are mainly for types of one sample problems. Numerical characteristics of functional data are defined as real constants. Centralisation in terms of constant numerical characteristics implies the relocation of the entire matrix of functional variances in order to obtain original curves whose centres of gravity are settled on the origin. Furthermore, based on the covariance matrix obtained from constant numerical characteristics, functional principal components for multivariate sample problems are proposed. Conclusions are validated by simulation in a real situation.

1 Introduction

We propose an extension of principal component analysis (PCA) for classical data to functional data. Functional data analysis (FDA) was first introduced by Ramsay (1982). Later, Ramsay and Silverman (2005) provided a thorough treatment of FDA. In contrast to classical statistical methodology, FDA treats an entire sequence of measurements for an individual as a single functional entity rather than a set of discrete values. Treating the data as a function retains all of the information contained in the data.

We will focus on PCA of functional data. PCA for functional data (FPCA) is described in full by Ramsay and Silverman (2005). Each functional principal component (FPC) represents a particular movement pattern over the time interval considered and represents variation of all the observations around the mean. More recently, FPCA based on the Karhunen-Loeve decomposition has been successfully applied in many areas. Locantore et al. (1999) explored abnormalities in the curvature of the cornea in the human eye. Viviani et al. (2005) compared its functional and multivariate versions and discovered that the functional approach offers a rather better image of experimental manipulations underlying the data. These investigations have time as the covariate. Kneip and Utikal (2001) applied FPCA to describe a set of density curves where the covariate is income. These applications are for types of problems in one