Statistica Sinica **34** (2024), 1-22
doi:https://doi.org/10.5705/ss.202021.0393

NONLINEAR DIMENSION REDUCTION FOR FUNCTIONAL DATA WITH APPLICATION TO CLUSTERING

Ruoxu Tan*, Yiming Zang and Guosheng Yin

Tongji University, Université de Lorraine and Imperial College London

Abstract: Functional data often possess nonlinear structures, for example, phase variation, for which linear dimension-reduction techniques can be ineffective. We study nonlinear dimension reduction for functional data based on the assumption that the data lie on an unknown manifold contaminated with noise. We generalize a recently developed manifold learning method designed for high-dimensional data into our context, and derive asymptotic convergence results, taking noise into account. The results based on synthetic examples often produce more accurate geodesic distance estimations than those of the traditional functional Isomap method. We further develop a clustering strategy based on the manifold learning outcomes, and demonstrate that our method outperforms others if the data lie on a curved manifold. Two real-data examples are presented for illustration.

Key words and phrases: Geodesic distance, graph clustering, manifold learning, measurement error.

1. Introduction

Popular methods of dealing with high-/infinite-dimensional data reduce the dimension of the data, for example, using a principal component analysis or a linear discriminant analysis. More recently, nonlinear methods such as manifold learning have been developed to handle complex data patterns, particularly for high-dimensional data. Well-known methods include Isomap (Tenenbaum, De Silva and Langford. (2000)), local linear embedding (Roweis and Saul (2000)), Laplacian eigenmaps (Belkin and Niyogi (2003)), tangent space alignment (Zhang and Zha (2004)), and vector diffusion maps (Singer and Wu (2012)). These methods and their variants have been used successfully in many fields, for example, in imaging data analysis (Pless and Souvenir (2009)) when the pixels lie in a high-dimensional vector space, but are concentrated on a low-dimensional manifold.

Functional data are usually collected sequentially over time. Unlike high-dimensional data, functional data are intrinsically infinite dimensional, and thus the demand for dimension reduction is more pressing. Classical functional principal component analysis (FPCA) is a core technique of linear dimension

*Corresponding author.