Statistica Sinica **36** (2026), 1-22
doi:https://doi.org/10.5705/ss.202023.0147

INTRINSIC CORRELATION ANALYSIS FOR WASSERSTEIN FUNCTIONAL DATA

Hang Zhou, Zhenhua Lin and Fang Yao*

University of North Carolina at Chapel Hill,
National University of Singapore and Peking University

Abstract: We develop a framework of canonical correlation analysis for distribution-valued functional data within the geometry of Wasserstein spaces. Specifically, we formulate an intrinsic concept of correlation between random distributions, propose estimation methods based on functional principal component analysis and Tikhonov regularization, respectively, for the correlation and its corresponding weight functions, and establish the minimax convergence rates of the estimators. In order to overcome the challenge raised by nonlinearity of Wasserstein spaces, the key idea is to adopt tensor Hilbert spaces to distribution-valued functional data. The finite-sample performance of the proposed estimators is illustrated via simulation studies, and the practical merit is demonstrated via a study on the association of distributions of brain activities between brain regions.

Key words and phrases: Distribution-valued data, minimax rate, parallel transport, tensor Hilbert space.

1. Introduction

Thanks to rapid evolution of modern data collection technologies, functional data emerge ubiquitously and the challenges of analyzing such data lead to a major line of research. For instance, various methodologies for multivariate data have been successfully extended to functional data, including functional principal components analysis (FPCA) (Yao, Müller and Wang, 2005a; Hall and Hosseini-Nasab, 2006), linear regression (Yao, Müller and Wang, 2005b; Hall and Horowitz, 2007; Yuan and Cai, 2010; Dou, Pollard and Zhou, 2012), classification (Delaigle and Hall, 2012), and clustering (James and Sugar, 2003). For a comprehensive treatment on functional data, we recommend the monographs by Ramsay and Silverman (2006), Ferraty and Vieu (2006), Horváth and Kokoszka (2012), Hsing and Eubank (2015), and Kokoszka and Reimherr (2017). In addition, statistical analysis of functional data taking values in a nonlinear Riemannian manifold has gained increasing attention and been investigated by Dai and Müller (2018), Lin and Yao (2019), Dai, Lin and Müller (2021), and Shao, Lin and Yao (2022).

In addition to manifold-valued data, probability distributions are nowadays commonly seen in practice, for example, arising from studies on mortality rates

*Corresponding author. E-mail: fyao@math.pku.edu.cn