# Intrinsic Correlation Analysis for
# Wasserstein Functional Data

Hang Zhou[1], Zhenhua Lin[2] and Fang Yao[3]

[1]*Department of Statistics, University of California, Davis, U.S.A.*

[2]*Department of Statistics and Applied Probability*

*National University of Singapore, Singapore*

[3]*School of Mathematical Sciences, Center for Statistical Science,*

*Peking University, Beijing, China*

*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 et al., 2005a; Hall and Hosseini-Nasab, 2006), linear regression (Yao et al., 2005b; Hall and Horowitz, 2007; Yuan and Cai, 2010; Dou et al., 2012), classification (Delaigle and Hall, 2012) and clustering (James and Sugar, 2003). For a comprehensive treatment on functional data, we recommend the monographs 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 et al. (2021) and Shao et al. (2022).

In addition to manifold-valued data, probability distributions are nowadays commonly seen in practice, for example, arising from studies on mortality rates (Lin and Müller, 2021), economics/housing (Chen et al., 2023), healthcare (Lin et al., 2023), fingerprints and metagenomics (Sommerfeld and Munk, 2018), flow

cytometry (Freulon et al., 2023), wearable device (Zhang et al., 2022) and COVID-19 cases and deaths (Gajardo and Müller, 2023). The space of probability distributions defined in a common domain, referred to as Wasserstein space, is clearly not a linear space as a linear combination of two probability measures may not be a probability measure. In order to tackle the nonlinear structure of the Wasserstein space, Petersen and Müller (2016) proposed a log quantile density (LQD) transformation to turn probability density functions to unconstrained functions. Dai (2022) adopted a square root transformation to map density functions into the positive orthant of a unit Hilbert sphere $\mathcal{S}^{\infty}$. However, none of these consider the more natural geometry that is compatible with optimal transport on the Wasserstein space. Since the Wasserstein space comes with a formal Riemannian structure (Ambrosio et al., 2008), it is natural to transform probability distributions via Riemannian logarithmic maps that have been well utilized (Lin and Yao, 2019; Shao et al., 2022). For instance, based on this idea, Bigot et al. (2017) proposed a geodesic principal component analysis for data sampled from a Wasserstein space, Petersen and Müller (2019) studied Wasserstein covariance for multiple random densities, and Chen et al. (2023) developed a class of regression models on Wasserstein space.

In this paper, we push further the frontier of statistical analysis into Wasserstein functional data that refer to functions taking values in a Wasserstein space.

Such data, for example, could naturally arise from functional magnetic resonance imaging (fMRI) studies, where the distribution of brain signals in a region is longitudinally available for a period; see Figure 1 for an illustration. For statistical analysis of such data, in addition to the challenging issue of infinite dimensionality shared by the ordinary functional data analysis, a major obstacle comes from the nonlinear nature of the Wasserstein space that creates difficulties especially in modeling the covariance structure. Such nonlinearity is also presented in the Riemannian functional data analysis and is addressed by the intrinsic device of tensor Hilbert spaces proposed in Lin and Yao (2019). However, although Wasserstein spaces have a geometric construction that is similar to the Riemannian structure, they are not Riemannian manifolds. On one hand, when considering the measures defined on the real line, the Wasserstein space can be regarded as the convex closed subset of $\mathcal{L}^2(0,1)$ formed by equivalence classes of quantile functions. Therefore, the Fréchet mean can be expressed by the quantile functions and some regularity conditions on the Fréchet functional are no longer needed to ensure the existence and uniqueness of the empirical and population Fréchet mean. In addition, the McCann's interpolation defines a constant-speed geodesic on the Wasserstein space and the flatness property facilitates the theoretical analysis and asymptotic behavior of our proposed estimators. On the other hand, since the tangent space at each point of Wasserstein space is an infinite-dimensional linear space, it is non-trivial
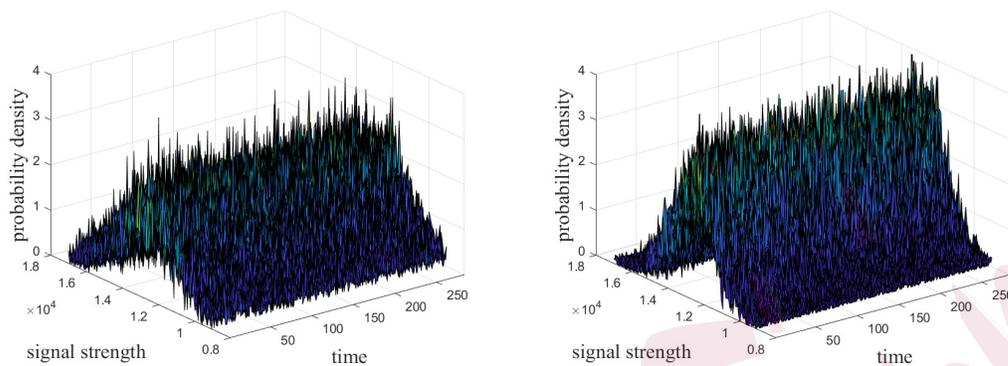
Figure 1: The observed longitudinal densities of signal strength in Caudate nucleus (left) and Putamen area (right) from resting-state fMRI examinations.

to extend the framework in Lin and Yao (2019) developed for (finite-dimensional) Riemannian manifolds to Wasserstein spaces. Given the aforementioned formal Riemannian structure of the Wasserstein space, we propose to circumvent the challenge by extending the device to Wasserstein functional data.

Specifically, we investigate correlation analysis for Wasserstein space valued functional data. Canonical correlation is one of the key tools for statistical analysis, and has been extensively studied for multivariate data and Euclidean functional data (He et al., 2003; Eubank and Hsing, 2008; Yang et al., 2011; Lian, 2014; Zhou and Chen, 2020), but is yet to be explored for Wasserstein functional data. Our main contribution is to formulate an intrinsic concept of correlation between random distributions, to propose an FPCA-based estimator and a Tikhonov reg-

ularized estimator, respectively, for the correlation and its corresponding weight functions, to establish the minimax convergence rates of the estimators, and to extend the framework of tensor Hilbert spaces to Wasserstein functional data. In addition, our arguments for the minimax rate can be straightforwardly extended to the setting of Euclidean functional data, while existing works seem to lack of rigorous arguments; see Remark 1 in supplementary for details.

The rest of the paper is organized as follows. We develop the foundational framework for Wasserstein space valued functional data in Section 2. Intrinsic Wasserstein correlation analysis is presented in Section 3. The simulation studies are offered in Section 4 and an application to fMRI dataset can be found in Section 5. The proofs of the main and ancillary results are collected together in the supplementary material.

## 2. Wasserstein Functional Data

In this section, we shall give a synopsis of the functional data valued in Wasserstein space. We first introduce the Wasserstein metric and geometry on a density class, which has similar structures to a Riemannian manifold. Based on the concept of tensor Hilbert space, we propose a new framework for functional data on Wasserstein space and discuss its properties. Then the random elements perspective in tensor Hilbert space are investigated, including an estimation procedure for the

mean surface and covariance operator. Finally, the asymptotic properties of the

proposed estimators are studied.

## 2.1    Wasserstein Metric and Geometry

Let $\mathcal{W}_2(\mathbb{R})$ be the collection of probability measures on the real line $\mathbb{R}$ with finite

second-order moments, that is,

$$\mathcal{W}_2(\mathbb{R}) = \left\{ \mu \in \mathcal{P}(\mathbb{R}) : \int_{\mathbb{R}} |x|^2 \mathrm{d}\mu(x) < \infty \right\},$$

where $\mathcal{P}(\mathbb{R})$ is the set of probability measures on $\mathbb{R}$. For $\mu \in \mathcal{W}_2(\mathbb{R})$ and a $\mu$-

measurable map $T : \mathbb{R} \to \mathbb{R}$, the push-forward measure of $\mu$ through $T$ is defined

by $(T\#\mu)(A) = \mu\{x \in \mathbb{R} \mid T(x) \in A\}$, for $A \in \mathcal{B}(\mathbb{R})$, where $\mathcal{B}(\mathbb{R})$ denotes the Borel

space of $\mathbb{R}$. In the sequel, we use $F_\mu$ and $F_\mu^{-1}$ to denote the distribution function

and the right-continuous quantile function of $\mu$, respectively. If $\mu$ is absolutely

continuous to the Lebesgue measure, its density function is denoted by $f_\mu$. The

Wasserstein distance between two measures $\mu, \nu \in \mathcal{W}_2(\mathbb{R})$ is defined by

$$d(\mu, \nu) := \inf \left\{ \int_{\mathbb{R}^2} |x_1 - x_2|^2 \mathrm{d}\gamma(x_1, x_2) : \gamma \in \Gamma(\mu, \nu) \right\}, \tag{2.1}$$

where $\Gamma(\mu, \nu)$ is the class of joint probability measures with marginal measures $\mu$

and $\nu$. The $\mathcal{W}_2(\mathbb{R})$ space endowed with the Wasserstein distance, denoted by the

Wasserstein space $(\mathcal{W}_2, d)$, is a separable and complete metric space (Ambrosio

et al., 2008).

The minimization problem (2.1) is known as the Kantorovich's formulation

(Kantorovich, 2006) that is a relaxation of the Monge problem

$$\inf\left\{\int_{\mathbb{R}}\{T(u)-u\}^2\mathrm{d}\mu(u),\ \text{such that }T\#\mu=\nu\right\}. \tag{2.2}$$

Unlike the Kantorovich's formulation, the Monge problem (2.2) can be ill-posed

when $\mu$ is a Dirac mass and $\nu$ has no atom, under which there is no transport map

$T$ such that $T\#\mu=\nu$. If $\mu$ is absolutely continuous to Lebesgue measure, which

implies $\mu$ has no atom and $F_\mu$ is continuous, the Monge problem (2.2) is equivalent

to (2.1) and has a unique solution $T=F_\nu^{-1}\circ F_\mu:=T_\mu^\nu$ (Major, 1978; Cuesta and

Matran, 1989; Brenier, 1991; Gangbo and McCann, 1996; Petersen and Müller,

2019). This solution, called the optimal transport map from $\mu$ to $\nu$, also induces

a geodesic between $\mu$ and $\nu$; here, a curve $\eta(t):I\to\mathcal{W}_2(\mathbb{R})$ parameterized by

an interval $I\subset\mathbb{R}$ is a geodesic if $d\{\eta(t),\eta(t+\epsilon)\}=a\epsilon$ for some constant $a>0$,

all $t\in I$ and sufficiently small $\epsilon>0$. Specifically, the McCann's interpolation

(McCann, 1997)

$$\mu_t=\left\{\mathbf{id}+t\left(T_\mu^\nu-\mathbf{id}\right)\right\}\#\mu:[0,1]\to\mathcal{W}_2(\mathbb{R})$$

is a geodesic connecting $\mu$ to $\nu$ (Ambrosio et al., 2008), where $\mathbf{id}$ denotes the

identity map.

Although $(\mathcal{W}_2,d)$ is not a Riemannian manifold, it can be endowed with a

formal Riemannian structure, in which geometric concepts essential to statistical

analysis, such as tangent spaces, Riemannian exponential maps and logarithmic
maps can be defined (Bigot et al., 2017; Chen et al., 2023). This motivates us to
extend the framework of Lin and Yao (2019) to the Wasserstein space, as follows.
Define

$$\mathcal{L}^2(\mu; \mathbb{R}) := \left\{ T : \mathbb{R} \to \mathbb{R} \quad \mu\text{-measurable} : \int_{\mathbb{R}} |T(x)|^2 \mathrm{d}\mu(x) < +\infty \right\},$$

which is a separable Hilbert space for any probability measure $\mu$, where the inner
product is given by $\langle T_1, T_2 \rangle_\mu = \int_{\mathbb{R}} T_1(u) T_2(u) \mathrm{d}\mu(u)$ for $T_1, T_2 \in \mathcal{L}^2(\mu; \mathbb{R})$. The
tangent space at $\mu$ is defined as the closure of

$$\mathrm{Tan}_\mu^\circ = \left\{ \tau \left( F_\nu^{-1} \circ F_\mu - \mathbf{id} \right) : \tau > 0, \nu \in \mathcal{W}_2(\mathbb{R}) \right\}$$

within $\mathcal{L}^2(\mu; \mathbb{R})$, i.e., $\mathrm{Tan}_\mu = \overline{\mathrm{Tan}_\mu^\circ}^{\mathcal{L}^2(\mu;\mathbb{R})}$. The term $F_\nu^{-1} \circ F_\mu - \mathbf{id}$ can be interpreted
as a "direction", just as in the Euclidean case where $x + t(y - x)$ has direction $y - x$.
It follows from the definition that $\mathrm{Tan}_\mu$ is a complete and separable subspace of
$\mathcal{L}^2(\mu, \mathbb{R})$. $\mathrm{Tan}_\mu$ is also a linear space (Chapter 2.3, Panaretos and Zemel, 2020),
and thus a separable Hilbert space endowed with the inner product $\langle \cdot, \cdot \rangle_\mu$. The
exponential map $\mathrm{Exp}_\mu : \mathrm{Tan}_\mu \to \mathcal{W}_2(\mathbb{R})$ and the corresponding logarithmic map
$\mathrm{Log}_\mu : \mathcal{W}_2(\mathbb{R}) \to \mathrm{Tan}_\mu$ at $\mu$ are defined by

$$\mathrm{Exp}_\mu(\mathrm{T}) = (\mathrm{T} + \mathbf{id}) \# \mu \quad \text{and} \quad \mathrm{Log}_\mu(\nu) = \mathrm{F}_\nu^{-1} \circ \mathrm{F}_\mu - \mathbf{id}, \tag{2.3}$$

respectively, for $\mathrm{T} \in \mathrm{Tan}_\mu$ and $\mu, \nu \in \mathcal{W}_2(\mathbb{R})$. Unlike ordinary Riemannian expo-
nential maps, the exponential map at $\mu$ defined above may not be a local homeo-

morphism between a neighborhood of $\mu$ and a neighborhood of the origin of $\mathrm{Tan}_{\mu}$, which shows that $\mathcal{W}_2(\mathbb{R})$ is not a genuine Riemannian manifold. Nevertheless, Theorem 2.2 in Bigot et al. (2017) shows that the exponential map $\mathrm{Exp}_{\mu}$ restricted to the image of the logarithmic map $\mathrm{Log}_{\mu}$ is an isometric homeomorphism with $\mathrm{Log}_{\mu}$ being its inverse, and this is sufficient for our statistical analysis.

As in most Wasserstein data the observed distributions have density functions supported in a compact domain $\mathcal{S}$, in this paper we restrict our attention to $\mathcal{W}_2(\mathcal{S})$ that contains all probability measures supported in $\mathcal{S}$.

## 2.2    Distribution-valued Functional Data

Let $X(t)$ be a random process indexed by $t \in \mathcal{T}$ and taking values in $\mathcal{W}_2(\mathcal{S})$, where $\mathcal{T}$ is a compact subset of $\mathbb{R}$, that is, for each $t \in \mathcal{T}$, $X(t)$ is a random measure on $\mathcal{S}$; in the statistical analysis, the random process $X$ serves as a prototype of the observed distribution-valued functional data. To quantify the first-order behaviors of $X$, as in Dai and Müller (2018); Lin and Yao (2019), we utilize the concept of the Fréchet mean (Fréchet, 1948; Agueh and Carlier, 2011), which in our context is defined by

$$\mu(t) = \underset{p \in \mathcal{W}_2(\mathcal{S})}{\arg\min} F(p, t) \text{ with } F(p, t) = \mathbf{E}d^2\{X(t), p\} \quad, t \in \mathcal{T}.$$

According to Lemma S2 of Lin et al. (2023), the Fréchet mean function exists and is unique for each $t \in \mathcal{T}$.

To characterize the second-order structure of $X$, we exploit the formal Riemannian structure of $\mathcal{W}_2(\mathcal{S})$ to consider the process $\mathrm{Log}_{\mu(t)}X(t)$ indexed by $t$ and taking values in the vector space $\mathrm{Tan}_{\mu(t)}$ for each $t \in \mathcal{T}$. Like Lin and Yao (2019), we treat the process $\mathrm{Log}_{\mu(\cdot)}X(\cdot)$, that is also denoted by $\mathrm{Log}_\mu X$ in the sequel for simplicity, as a vector field along the mean curve $\mu$, and then further view it as a random element in the space of vector fields along $\mu$,

$$\mathscr{T}(\mu) := \left\{ Z : Z(\cdot) \in \mathrm{Tan}_{\mu(\cdot)}, \int \langle Z(t), Z(t) \rangle_{\mu(t)} \, \mathrm{d}t < \infty \right\}.$$

For Riemannian functional data, i.e., data of functions taking value in a (finite-dimensional) Riemannian manifold, the space $\mathscr{T}(\mu)$, termed a tensor Hilbert space and shown by Lin and Yao (2019), is a separable Hilbert space. Now we extend this result to the Wasserstein space $\mathcal{W}_2(\mathcal{S})$ by endowing $\mathscr{T}(\mu)$ with the inner product $\langle\!\langle Z_1, Z_2 \rangle\!\rangle_\mu := \int \langle Z_1(t), Z_2(t) \rangle_{\mu(t)} \mathrm{d}t$ and the induced norm $\|\!|\cdot|\!\|_\mu$, where we recall that $\langle \cdot, \cdot \rangle_{\mu(t)}$ denotes the inner product in $\mathrm{Tan}_{\mu(t)}$.

**Proposition 1.** $\mathscr{T}(\mu)$ *is a separable Hilbert space with the inner product* $\langle\!\langle \cdot, \cdot \rangle\!\rangle_\mu$.

Since $\mathcal{S}$ is compact, then $\mathcal{W}_2(\mathcal{S})$ is compact and thus $\mathbf{E}\|\!|\mathrm{Log}_\mu X|\!\|_\mu^2 < \infty$, according to Theorem 7.4.2 in Hsing and Eubank (2015), $\mathrm{Log}_\mu X$ can be viewed as a random element in $\mathscr{T}(\mu)$. The auto-covariance operator $\mathbf{C} : \mathscr{T}(\mu) \to \mathscr{T}(\mu)$ for $X$ can be defined by

$$\langle\!\langle \mathbf{C}U, V \rangle\!\rangle_\mu := \mathbf{E}\left( \langle\!\langle \mathrm{Log}_\mu X, U \rangle\!\rangle_\mu \langle\!\langle \mathrm{Log}_\mu X, V \rangle\!\rangle_\mu \right), \quad \text{for } U, V \in \mathscr{T}(\mu).$$

This operator is a nonnegative–definite trace–class operator with the following eigendecomposition (Theorem 7.2.6, Hsing and Eubank, 2015)

$$\mathbf{C} = \sum_{k=1}^{\infty} \lambda_k \Phi_k \otimes \Phi_k$$

with eigenvalues $\lambda_1 > \lambda_2 > \cdots > 0$, that are assumed of multiplicity 1 without loss of generality, and orthonormal eigenelements $\Phi_k$ that form a complete orthonormal system for $\mathscr{T}(\mu)$. In addition, the process $\mathrm{Log}_\mu X$ admits the following Karhunen–Loève expansion

$$\mathrm{Log}_\mu X = \sum_{k=1}^{\infty} \xi_k \Phi_k$$

with $\xi_k := \langle\!\langle \mathrm{Log}_\mu X, \Phi_k \rangle\!\rangle_\mu$ being uncorrelated and centered random variables.

Given a sample of independently and identically distributed (i.i.d.) copies $X_1, \cdots, X_n$ of $X$, the Fréchet mean function $\mu$ is estimated by its sample version

$$\hat{\mu}(t) = \underset{p \in \mathcal{W}_2}{\arg\min}\, F_n(p, t) \text{ with } F_n(p, t) = \frac{1}{n} \sum_{i=1}^{n} d^2\{X_i(t), p\}, \quad p \in \mathcal{W}_2,\, t \in \mathcal{T}.$$

According to Chen et al. (2023) and Lemma S2 of Lin et al. (2023),

$$F_{\mu(t)}^{-1} = \mathbf{E} F_{X(t)}^{-1} \quad \text{and} \quad F_{\hat{\mu}(t)}^{-1} = \frac{1}{n} \sum_{i=1}^{n} F_{X_i(t)}^{-1} \text{ for each } t \in \mathcal{T}. \tag{2.4}$$

Similarly, the auto-covariance operator is estimated by its sample version

$$\hat{\mathbf{C}} = \frac{1}{n} \sum_{i=1}^{n} (\mathrm{Log}_{\hat{\mu}} X_i) \otimes (\mathrm{Log}_{\hat{\mu}} X_i),$$

which admits the eigendecompoxtion $\hat{\mathbf{C}} = \sum_{k=1}^{\infty} \hat{\lambda}_k \hat{\Phi}_k \otimes \hat{\Phi}_k$ for $\hat{\lambda}_1 > \hat{\lambda}_2 > \cdots \geq 0$ with the estimated eigenvalues $\hat{\lambda}_k$ and eigenfunctions $\hat{\Phi}_k$.

To assess the estimation quality, for the Fréchet mean function, one may use the integrated squared error $\int d^2\{\hat{\mu}(t), \mu(t)\}\mathrm{d}t$. It turns out to be challenging to quantify the discrepancy between $\hat{\mathbf{C}}$ and $\mathbf{C}$, as when $\hat{\mu}$ and $\mu$ are not identical, the spaces $\mathscr{T}(\hat{\mu})$ and $\mathscr{T}(\mu)$ are distinct Hilbert spaces. For Riemannian functional data, Lin and Yao (2019) addressed this problem by the parallel transport induced by the Levi–Civita connection that is intrinsic to the Riemannian manifold under consideration. Fortunately, as shown in Ambrosio et al. (2008), the Wasserstein space also has also a similar geometric structure that is defined via the Benamou–Brenier formula (Benamou and Brenier, 2000), which can be adopted to $\mathcal{W}_2(\mathcal{S})$ as follows.

First, we begin with the parallel transport of tangent vectors at an element of $\mathcal{W}_2(\mathcal{S})$ to another element. To this end, let $\nu$ and $\nu'$ be elements in $\mathcal{W}_2(\mathcal{S})$, a parallel transport operator can be defined between the entire Hilbert spaces $\mathcal{L}^2(\nu; \mathcal{S})$ and $\mathcal{L}^2(\nu'; \mathcal{S})$ (Chen et al., 2023), that is, $\mathrm{P}_\nu^{\nu'} u := u \circ F_\nu^{-1} \circ F_{\nu'}$ for $u \in \mathcal{L}^2(\nu; \mathcal{S})$, where $F_\nu^{-1}$ and $F_{\nu'}$ are the quantile function of $\nu$ and distribution function of $\nu'$. Assuming that $\nu$ is atomless, the parallel transport $\mathcal{P}_\nu^{\nu'}$ from tangent space $\mathrm{Tan}_\nu$ to $\mathrm{Tan}_{\nu'}$ is defined by $\mathrm{P}_\nu^{\nu'}$ restricted to $\mathrm{Tan}_\nu$, i.e., $\mathcal{P}_\nu^{\nu'} = \mathrm{P}_\nu^{\nu'}|_{\mathrm{Tan}_\nu}$.

To extend the concept of parallel transport defined in Chen et al. (2023) to tensor Hilbert spaces, let $\mu(\cdot), \mu'(\cdot), \nu(\cdot)$ and $\nu'(\cdot)$ be measurable curves on $\mathcal{W}_2(\mathcal{S})$. For $U \in \mathscr{T}(\mu)$, the parallel transport of $U$ from $\mathscr{T}(\mu)$ to $\mathscr{T}(\mu')$ is defined by

$(\mathcal{P}_\mu^{\mu'} U)(\cdot) = \mathcal{P}_{\mu(\cdot)}^{\mu'(\cdot)} \circ U(\cdot)$. Let $\mathfrak{B}(\mu, \nu)$ denotes the set of all bounded linear operators

on $\mathscr{T}(\mu)$ mapping to $\mathscr{T}(\nu)$, which is a Banach space with the norm $|||\mathbf{A}|||_{\mathfrak{B}(\mu,\nu)} =$

$\sup_{U \in \mathscr{T}(\mu), \|U\|_\mu = 1} \|\mathbf{A}U\|_\nu$ (Hsing and Eubank, 2015, Theorem 3.1.3). The operator

$\mathcal{P}_\mu^\nu$ also gives rise to a mapping $\mathcal{P}_{\mathfrak{B}(\mu,\nu)}^{\mathfrak{B}(\mu',\nu')}$ from $\mathfrak{B}(\mu, \nu)$ to $\mathfrak{B}(\mu', \nu')$, defined by

$\left\{ \mathcal{P}_{\mathfrak{B}(\mu,\nu)}^{\mathfrak{B}(\mu',\nu')} \mathbf{A} \right\} V = \mathcal{P}_\nu^{\nu'} \mathbf{A}(\mathcal{P}_{\mu'}^\mu V)$ for $\mathbf{A} \in \mathfrak{B}(\mu, \nu)$ and $V \in \mathscr{T}(\mu')$.

**Proposition 2.** *Let $\mu, \mu', \nu, \nu'$ be measurable curves on $\mathcal{W}_2(\mathcal{S})$ and assume for each*

*$t \in \mathcal{T}$, $\mu(t)$, $\mu'(t)$, $\nu(t)$, $\nu'(t)$ are atomless; $U, U' \in \mathscr{T}(\mu)$, $V \in \mathscr{T}(\nu)$, $\mathbf{A} \in \mathfrak{B}(\mu, \nu)$,*

*and $\mathbf{B} \in \mathfrak{B}(\mu', \nu')$.*

(a) *$\mathcal{P}_{\mu(t)}^{\nu(t)}$ is a unitary transportation form $\mathrm{Tan}_{\mu(t)}$ to $\mathrm{Tan}_{\nu(t)}$ and the adjoint operator of $\mathcal{P}_{\mu(t)}^{\nu(t)}$ is $\mathcal{P}_{\nu(t)}^{\mu(t)}$.*

(b) *$\left\langle \mathcal{P}_{\mu(t)}^{\nu(t)} u, v \right\rangle_{\nu(t)} = \left\langle u, \mathcal{P}_{\nu(t)}^{\mu(t)} v \right\rangle_{\mu(t)}$ and $\left\| \mathcal{P}_{\mu(t)}^{\nu(t)} u - v \right\|_{\nu(t)} = \left\| u - \mathcal{P}_{\nu(t)}^{\mu(t)} v \right\|_{\mu(t)}$ for $u \in \mathrm{Tan}_{\mu(t)}$ and $v \in \mathrm{Tan}_{\nu(t)}$.*

(c) *$\langle\!\langle U, U' \rangle\!\rangle_\mu = \langle\!\langle \mathcal{P}_\mu^\nu U, \mathcal{P}_\mu^\nu U' \rangle\!\rangle_{\nu}$, $\langle\!\langle \mathcal{P}_\mu^\nu U, V \rangle\!\rangle_\nu = \langle\!\langle U, \mathcal{P}_\nu^\mu V \rangle\!\rangle_\mu$ and $\left\| \mathcal{P}_\mu^\nu U - V \right\|_\nu = \|\!\|\!| U - \mathcal{P}_\nu^\mu V |\!\|\!\|_\mu$.*

(d) *$\mathcal{P}_\nu^{\nu'}(\mathbf{A}U) = \left\{ \mathcal{P}_{\mathfrak{B}(\mu,\nu)}^{\mathfrak{B}(\mu',\nu')} \mathbf{A} \right\} (\mathcal{P}_\mu^{\mu'} U)$.*

(e) *$\left\|\!\left\|\!\left\| \mathcal{P}_{\mathfrak{B}(\mu,\nu)}^{\mathfrak{B}(\mu',\nu')} \mathbf{A} - \mathbf{B} \right\|\!\right\|\!\right\|_{\mathfrak{B}(\mu',\nu')} = \left\|\!\left\|\!\left\| \mathbf{A} - \mathcal{P}_{\mathfrak{B}(\mu',\nu')}^{\mathfrak{B}(\mu,\nu)} \mathbf{B} \right\|\!\right\|\!\right\|_{\mathfrak{B}(\mu,\nu)}$.*

(f) *$\mathcal{P}_{\mathfrak{B}(\mu,\nu)}^{\mathfrak{B}(\mu',\nu')} \sum_k c_k \Phi_{\mu,k} \otimes \Phi_{\nu,k} = \sum_k c_k \left( \mathcal{P}_\mu^{\mu'} \Phi_{\mu,k} \right) \otimes \left( \mathcal{P}_\nu^{\nu'} \Phi_{\nu,k} \right)$, where $c_k$ are scalar constants, $\Phi_{\mu,k} \in \mathscr{T}(\mu)$, and $\Phi_{\nu,k} \in \mathscr{T}(\nu)$.*

In the above we exploit the geometry of the Wasserstein space to develop the parallel transport of elements and linear operators on tensor Hilbert spaces. This contrasts with the mechanism adopted in Lin and Yao (2019) for Riemannian manifolds in which the Levi–Civita connection exists and can be leveraged. Now we are ready to quantify the discrepancy between objects of the same kind in $\mathscr{T}(\mu)$ and $\mathscr{T}(\hat{\mu})$ by utilizing the above parallel transport.

**Theorem 1.** *Assume for each $t \in \mathcal{T}$, $\mu(t) \in \mathcal{W}_2$ is atomless.*

(a) *$\sqrt{n}\mathrm{Log}_\mu \hat{\mu}$ converges in distribution to a Gaussian measure on the tensor Hilbert space $\mathscr{T}(\mu)$.*

(b) *$\sup_{t \in \mathcal{T}} d^2\{\mu(t), \hat{\mu}(t)\} = O_p(n^{-1})$ and $\int_{\mathcal{T}} d^2\{\mu(t), \hat{\mu}(t)\}\mathrm{d}t = O_p(n^{-1})$.*

(c) *$\left|\left|\left|\mathcal{P}^{\mathfrak{B}(\mu,\mu)}_{\mathfrak{B}(\hat{\mu},\hat{\mu})}\hat{\mathbf{C}} - \mathbf{C}\right|\right|\right|^2_{\mathfrak{B}(\mu,\mu)} = O_p(n^{-1})$ and $\sup_{k \geqslant 1}|\hat{\lambda}_k - \lambda_k|^2 = O_p(n^{-1})$.*

(d) *Let $\eta_j = (1/2)\inf_{k \neq j}|\lambda_k - \lambda_j|$ and $\Delta = \mathcal{P}^{\mathfrak{B}(\mu,\mu)}_{\mathfrak{B}(\hat{\mu},\hat{\mu})}\hat{\mathbf{C}} - \mathbf{C}$. If $\lambda_j \sim j^{-a}$ and $\eta_j \sim j^{-(a+1)}$, then for all $j$ such that $|||\Delta|||_{\mathfrak{B}(\mu,\mu)} < \eta_j/2$ and a constant $C$ that does not depend on $j$ and $n$, we have*

$$\mathbf{E}\left\|\mathcal{P}^\mu_{\hat{\mu}}\hat{\Phi}_j - \Phi_j\right\|^2_{\mathscr{T}(\mu)} \leqslant C\frac{j^2}{n}.$$

Part (a), (b) and (c) in Theorem 1 show that the convergence rates for the mean and covariance estimators are root-$n$, which is consistent with the classic results for fully observed Euclidean (Hall and Hosseini-Nasab, 2006; Hall and Horowitz,

2007) and Riemannian manifold functional data (Lin and Yao, 2019). In view of the properties for $\mathcal{W}_2(\mathcal{S})$, some regularity conditions for the Fréchet functional are no longer needed to ensure the existence of the population and empirical Fréchet mean. Furthermore, due to the flatness of the Wasserstein space, the high order terms in the Taylor expansion of the logarithm process are vanished, which facilitates the theoretical derivation. The last statement of Theorem 1 extends the classic result in eigenfunctions for fully observed Euclidean functional data, which is essential in most FPCA-based methods, especially related to regression problems (Hall and Hosseini-Nasab, 2006; Hall and Horowitz, 2007; Dou et al., 2012). We stress that the component number $j$ is not fixed and could diverge slowly with $n$, and this rate is optimal in the minimax sense (Wahl, 2022).

## 3. Intrinsic Wasserstein Correlation Analysis

With the preparation of the groundwork for Wasserstein functional data, we are ready to discuss the correlation analysis between two sets of Wasserstein functional data.

### 3.1 Wasserstein Correlation

Let $X$ and $Y$ be two $\mathcal{W}_2(\mathcal{S})$-valued random processes with mean functions $\mu_X, \mu_Y$ and auto-covariance operators $\mathbf{C}_X, \mathbf{C}_Y$, respectively. The cross-covariance operator

$\mathbf{C}_{XY} : \mathscr{T}(\mu_Y) \longmapsto \mathscr{T}(\mu_X)$ for $X$ and $Y$ is defined as

$$\langle\!\langle \mathbf{C}_{XY}V, U \rangle\!\rangle_{\mu_X} := \mathbf{E}\left( \langle\!\langle \mathrm{Log}_{\mu_Y} Y, V \rangle\!\rangle_{\mu_Y} \langle\!\langle \mathrm{Log}_{\mu_X} X, U \rangle\!\rangle_{\mu_X} \right) \text{ for } V \in \mathscr{T}(\mu_Y), U \in \mathscr{T}(\mu_X),$$

and $\mathbf{C}_{YX}$ is defined analogously. We then define the intrinsic Wasserstein correlation between $X$ and $Y$ as

$$\rho = \max_{\substack{U \in \mathscr{T}(\mu_X): \langle U, \mathbf{C}_X U \rangle_{\mu_X} = 1 \\ V \in \mathscr{T}(\mu_Y): \langle V, \mathbf{C}_Y V \rangle_{\mu_Y} = 1}} \langle\!\langle U, \mathbf{C}_{XY} V \rangle\!\rangle_{\mu_X}, \tag{3.1}$$

which generalizes the canonical correlation for classic multivariate data and Euclidean functional data (He et al., 2003; Lian, 2014). Unlike the latter two types of data, functions valued in Wasserstein space are nonlinear, and such nonlinearity is overcome by the device of tensor Hilbert space introduced in Section 2. Note that the maximization problem (3.1) is equivalent to finding $U$ in $\mathscr{T}(\mu_X)$ and $V$ in $\mathscr{T}(\mu_Y)$ to maximize the correlation between $\langle\!\langle U, \mathrm{Log}_{\mu_X} X \rangle\!\rangle_{\mu_X}$ and $\langle\!\langle V, \mathrm{Log}_{\mu_Y} Y \rangle\!\rangle_{\mu_Y}$. By similar arguments as Theorem 10.1.2 of Hsing and Eubank (2015), the $k$th correlation can be defined by optimizing (3.1) within the subspace that is orthogonal to the first $(k-1)$th weight functions.

In the case of multivariate data, the solution to canonical correlation analysis is reduced to singular value decomposition of $\mathbf{C}_X^{-1/2} \mathbf{C}_{XY} \mathbf{C}_Y^{-1/2}$, which could not be applied to functional data because both $\mathbf{C}_X$ and $\mathbf{C}_Y$ are infinite-dimensional compact operators and thus have an invertibility issue. By Theorem 7.2.10 in Hsing and Eubank (2015), there exists an operator $\mathbf{R}_{XY} \in \mathfrak{B}(\mu_Y, \mu_X)$ with $|||\mathbf{R}_{XY}|||_{\mathfrak{B}(\mu_Y, \mu_X)} \leqslant$

1 such that $\mathbf{C}_{XY} = \mathbf{C}_X^{1/2} \mathbf{R}_{XY} \mathbf{C}_Y^{1/2}$. This suggests that the operator $\mathbf{C}_X^{-1/2} \mathbf{C}_{XY} \mathbf{C}_Y^{-1/2}$ is definable on the range of $\mathbf{C}_Y^{1/2}$. To formulate this idea and link it to the optimization problem (3.1), we first recall the Karhunen–Loève expansions for $X$ and $Y$, given by

$$\mathrm{Log}_{\mu_X} X = \sum_{k=1}^{\infty} \xi_k \Phi_{X,k}, \quad \mathrm{Log}_{\mu_Y} Y = \sum_{k=1}^{\infty} \eta_k \Phi_{Y,k},$$

where $\{\Phi_{X,k}\}_{k=1}^{\infty}$ and $\{\Phi_{Y,k}\}_{k=1}^{\infty}$ are respectively the eigenbases of $\mathbf{C}_X$ and $\mathbf{C}_Y$, and $\xi_j, \eta_j$ are principal component scores respectively with variance $\lambda_{X,j}$ and $\lambda_{Y,j}$. It is then seen that $\mathbf{C}_{YX}$ and $\mathbf{C}_{XY}$ can be expressed as

$$\mathbf{C}_{YX} = \sum_{j_1=1}^{\infty} \sum_{j_2=1}^{\infty} \gamma_{j_1 j_2} \Phi_{X,j_1} \otimes \Phi_{Y,j_2}, \quad \mathbf{C}_{XY} = \sum_{j_1=1}^{\infty} \sum_{j_2=1}^{\infty} \gamma_{j_1 j_2} \Phi_{Y,j_2} \otimes \Phi_{X,j_1}$$

with $\gamma_{j_1 j_2} = \mathbf{E}\{\xi_{j_1} \eta_{j_2}\}$. Now we impose the following assumption on the interplay among $\gamma_{j_1 j_2}$, $\lambda_{X,j_1}$ and $\lambda_{Y,j_2}$.

**Assumption (B.0).** $\sum_{j_1,j_2}^{\infty} \frac{\gamma_{j_1 j_2}^2}{\lambda_{X,j_1}^2 \lambda_{Y,j_2}} < \infty$ and $\sum_{j_1,j_2}^{\infty} \frac{\gamma_{j_1 j_2}^2}{\lambda_{X,j_1} \lambda_{Y,j_2}^2} < \infty$.

The above assumption is the same as Condition 4.5 in He et al. (2003) and requires that cross-covariance operator of $X$ and $Y$ be aligned with the eigenfunctions of $\mathrm{Log}_{\mu_X} X$ and $\mathrm{Log}_{\mu_Y} Y$; such a requirement is commonly adopted in FPCA-based functional regression models (Hall and Horowitz, 2007; Dou et al., 2012). Under this assumption, the following proposition, inspiring estimators that are proposed in the next section, asserts the boundedness of the operator $\mathbf{C}_X^{-1/2} \mathbf{C}_{XY} \mathbf{C}_Y^{-1/2}$ and provides a solution to the maximization problem (3.1).

**Proposition 3.** *Under Assumption* B.0, $\mathbf{C}_X^{-1/2}\mathbf{C}_{XY}\mathbf{C}_Y^{-1/2}$ *and* $\mathbf{C}_X^{-1}\mathbf{C}_{XY}\mathbf{C}_Y^{-1/2}$ *are Hilbert Schmidt operators defined on* $\mathscr{T}(\mu_Y)$. *The maximum in* (3.1) *is achieved for the weight functions* $U \in \mathscr{T}(\mu_X)$ *and* $V \in \mathscr{T}(\mu_Y)$ *with maximum* $\rho = \sqrt{\alpha}$, *where* $(\alpha, U)$ *is the first eigenpair of* $\mathbf{C}_X^{-1}\mathbf{C}_{XY}\mathbf{C}_Y^{-1}\mathbf{C}_{YX}$, *and* $V = \mathbf{C}_Y^{-1}\mathbf{C}_{YX}U/\|\mathbf{C}_Y^{-1/2}\mathbf{C}_{YX}U\|_{\mu_Y}$.

## 3.2 Estimation and Theoretical Properties

Given a random sample of functions $\{(X_i, Y_i)\}_{i=1}^n$ of $(X, Y)$, the mean functions for $X$ and $Y$ are estimated by $\hat{\mu}_X$ and $\hat{\mu}_Y$ that are respectively represented by

$$F_{\hat{\mu}_X(t)}^{-1} = \frac{1}{n}\sum_{i=1}^n F_{X_i(t)}^{-1} \text{ and } F_{\hat{\mu}_Y(t)}^{-1} = \frac{1}{n}\sum_{i=1}^n F_{Y_i(t)}^{-1} \text{ for each } t \in \mathcal{T}.$$

The estimators for autocovariance operators of $X$ and $Y$ are

$$\hat{\mathbf{C}}_X = \frac{1}{n}\sum_{i=1}^n (\mathrm{Log}_{\hat{\mu}_X} X_i) \otimes (\mathrm{Log}_{\hat{\mu}_X} X_i) \text{ and } \hat{\mathbf{C}}_Y = \frac{1}{n}\sum_{i=1}^n (\mathrm{Log}_{\hat{\mu}_Y} Y_i) \otimes (\mathrm{Log}_{\hat{\mu}_Y} Y_i),$$

respectively. In addition, they admit the following decompositions,

$$\hat{\mathbf{C}}_X = \sum_{k=1}^\infty \hat{\lambda}_{X,k}\hat{\Phi}_{X,k} \otimes \hat{\Phi}_{X,k} \text{ and } \hat{\mathbf{C}}_Y = \sum_{k=1}^\infty \hat{\lambda}_{Y,k}\hat{\Phi}_{Y,k} \otimes \hat{\Phi}_{Y,k},$$

where $(\hat{\lambda}_{X,k}\hat{\Phi}_{X,k})$ and $(\hat{\lambda}_{Y,k}, \hat{\Phi}_{Y,k})$ serve as estimators for $(\lambda_{X,k}, \Phi_{X,k})$ and $(\lambda_{Y,k}, \Phi_{Y,k})$, respectively. Similarly, the cross covariance operators between $X$ and $Y$ are estimated by

$$\hat{\mathbf{C}}_{YX} = \frac{1}{n}\sum_{i=1}^n (\mathrm{Log}_{\hat{\mu}_X} X_i) \otimes (\mathrm{Log}_{\hat{\mu}_Y} Y_i) \text{ and } \hat{\mathbf{C}}_{XY} = \frac{1}{n}\sum_{i=1}^n (\mathrm{Log}_{\hat{\mu}_Y} Y_i) \otimes (\mathrm{Log}_{\hat{\mu}_X} X_i).$$

Based on the above estimators, the estimator of $\mathbf{C}_X^{-1}\mathbf{C}_{XY}\mathbf{C}_Y^{-1}\mathbf{C}_{YX}$ is denoted by $\hat{\mathbf{C}}_{X,k_X}^{-1}\hat{\mathbf{C}}_{XY}\ \hat{\mathbf{C}}_{Y,k_Y}^{-1}\hat{\mathbf{C}}_{YX}$, where $\hat{\mathbf{C}}_{X,k_X}^{-1} = \sum_{j=1}^{k_X}\hat{\lambda}_{X,j}^{-1}\hat{\Phi}_{X,j}$, $\hat{\mathbf{C}}_{Y,k_Y}^{-1} = \sum_{j=1}^{k_Y}\hat{\lambda}_{Y,j}^{-1}\hat{\Phi}_{Y,j}$, and $k_X, k_Y \in \mathbb{N}^+$ are two tuning parameters. The truncation of $\hat{\mathbf{C}}_{X,k_X}^{-1}$ and $\hat{\mathbf{C}}_{X,k_X}^{-1}$ at respectively finite levels $k_X$ and $k_Y$ serves as a way of regularization that is needed to address the invertibility issue of infinite-dimensional compact operators (Yao et al., 2005b; Hall and Horowitz, 2007; Dou et al., 2012). Then, the estimator of $U$, denoted by $\hat{U}$, is the eigenfunction of $\hat{\mathbf{C}}_{X,k_X}^{-1}\ \hat{\mathbf{C}}_{XY}\hat{\mathbf{C}}_{Y,k_Y}^{-1}\ \hat{\mathbf{C}}_{YX}$ associated with its largest eigenvalue $\hat{\alpha}$, and the estimators for $V, \rho$ are defined by

$$\hat{V} = \hat{\mathbf{C}}_{Y,k_Y}^{-1}\hat{\mathbf{C}}_{YX}\hat{U}/\|\hat{\mathbf{C}}_{Y,k_Y}^{-1/2}\hat{\mathbf{C}}_{YX}\hat{U}\|_{\hat{\mu}_Y} \text{ and } \hat{\rho} = \sqrt{\hat{\alpha}},$$

respectively. In the subsequent analyses, we focus on the first correlation pairs. The $k$th canonical correlation with its associated weight functions can be obtained from the $k$th eigenfunction of $\hat{\mathbf{C}}_{X,k_X}^{-1}\ \hat{\mathbf{C}}_{XY}\hat{\mathbf{C}}_{Y,k_Y}^{-1}\ \hat{\mathbf{C}}_{YX}$ and our theoretical results hold for all fixed $k$.

Alternatively, we may utilize Tikhonov regularization (Hall and Horowitz, 2007) to estimate $\mathbf{C}_X^{-1}\mathbf{C}_{XY}\ \mathbf{C}_Y^{-1}\mathbf{C}_{YX}$ by $(\hat{\mathbf{C}}_X + \epsilon_X\hat{\mathbf{id}}_X)^{-1}\hat{\mathbf{C}}_{XY}(\hat{\mathbf{C}}_Y + \epsilon_Y\hat{\mathbf{id}}_Y)^{-1}\hat{\mathbf{C}}_{YX}$, where $\epsilon_X, \epsilon_Y$ are positive tuning parameters and $\hat{\mathbf{id}}_X, \hat{\mathbf{id}}_Y$ are the identity operators, respectively, on $\mathscr{T}(\hat{\mu}_X)$ and $\mathscr{T}(\hat{\mu}_Y)$. Then $U$ is estimated by the eigenfunction $\tilde{U}$ of $(\hat{\mathbf{C}}_X + \epsilon_X\hat{\mathbf{id}}_X)^{-1}\hat{\mathbf{C}}_{XY}(\hat{\mathbf{C}}_Y + \epsilon_Y\hat{\mathbf{id}}_Y)^{-1}\hat{\mathbf{C}}_{YX}$ associated with the largest eigenvalue $\tilde{\alpha}$, $V$ is estimated by $\tilde{V} = (\hat{\mathbf{C}}_Y + \epsilon_Y\hat{\mathbf{id}}_Y)^{-1}\hat{\mathbf{C}}_{YX}\tilde{U}/\|(\hat{\mathbf{C}}_Y + \epsilon_Y\hat{\mathbf{id}}_Y)^{-1/2}\hat{\mathbf{C}}_{YX}\tilde{U}\|_{\hat{\mu}_Y}$, and $\rho$ is estimated by $\tilde{\rho} = \sqrt{\tilde{\alpha}}$.

To study the theoretical properties of the estimators $\hat{U}$ and $\hat{V}$, we require the following assumption to utilize the parallel transportation operators defined in Section 2.

**Assumption (A.1).** For each $t \in \mathcal{T}$, $\mu_X(t)$ and $\mu_Y(t)$ are atomless.

In addition, we assume the eigenspace for the largest eigenvalue of $\mathbf{C}_X^{-1}\mathbf{C}_{XY}\mathbf{C}_Y^{-1}\mathbf{C}_{YX}$ has multiplicity one to ensure the uniqueness of the solution to (3.1). We also make the following assumptions, where in the sequel $C$ denotes some positive constant which may vary cross places.

**Assumption (A.2).** $\mathbf{E}\xi_j^4 \leqslant C\lambda_{X,j}^2$ and $\mathbf{E}\eta_j^4 \leqslant C\eta_{Y,j}^2$.

**Assumption (B.1).** $Cj^{-a_X} \geqslant \lambda_{X,j} \geqslant \lambda_{X,j+1} + C^{-1}j^{-a_X-1}$, and $Cj^{-a_Y} \geqslant \lambda_{Y,j} \geqslant \lambda_{Y,j+1} + C^{-1}j^{-a_Y-1}$ for some $a_X, a_Y > 1$ and each $j \geqslant 1$.

**Assumption (B.2).** $\sum_{j_2} \gamma_{j_1j_2}^2 \leqslant Cj_1^{-2a_X-2b_X}, \sum_{j_1} \gamma_{j_1j_2}^2 \leqslant Cj_2^{-2a_Y-2b_Y}, k_X \asymp n^{1/(a_X+2b_X)}$ and $k_Y \asymp n^{1/(a_Y+2b_Y)}$, for some constants $b_X > a_X/2+1, b_Y > a_Y/2+1$.

Assumption A.2 is common in the FPCA literature (Hall and Horowitz, 2007). Assumptions B.1 and B.2 define a class of random processes $X$ and $Y$ for which we are able to establish the minimax rate for the proposed estimators; similar conditions have been adopted in the literature of functional linear regression (Hall and Horowitz, 2007; Dou et al., 2012). The following theorem presents an upper bound on the convergence rate of estimated weight functions $\hat{U}$ and $\hat{V}$.

**Theorem 2.** *Under assumptions* A.1*,* A.2*,* B.1 *and* B.2*, we have*

$$\|\mathcal{P}_{\hat{\mu}_X}^{\mu_X}\hat{U} - U\|_{\mu_X}^2 = O_p\left(\max\left\{n^{-(2b_X-1)/(a_X+2b_X)}, n^{-(2b_Y-1)/(a_Y+2b_Y)}\right\}\right);$$

$$\|\mathcal{P}_{\hat{\mu}_Y}^{\mu_Y}\hat{V} - V\|_{\mu_Y}^2 = O_p\left(\max\left\{n^{-(2b_X-1)/(a_X+2b_X)}, n^{-(2b_Y-1)/(a_Y+2b_Y)}\right\}\right);$$

$$|\hat{\rho} - \rho|^2 = O_p\left(\max\left\{n^{-(2b_X-1)/(a_X+2b_X)}, n^{-(2b_Y-1)/(a_Y+2b_Y)}\right\}\right).$$

The convergence rate in Theorem 2 is in accordance to the rate in classic functional regression problems (Hall and Horowitz, 2007; Yuan and Cai, 2010; Dou et al., 2012), as well as the rate for non-functional Wasserstein regression (Chen et al., 2023). This is not surprising, since canonical correlation analysis is intimately related to two regression problems, in our context, one in which $\text{Log}_{\mu_X}X$ is regressed on $\text{Log}_{\mu_Y}Y$ and the other in which $\text{Log}_{\mu_Y}Y$ is regressed on $\text{Log}_{\mu_X}X$. This is slightly different from the attained rate of the functional linear regression involving the Riemannian manifold (Lin and Yao, 2019, 2021), as the nonlinear structure does not affect the convergence rate due to the flatness of the geodesic in Wasserstein spaces.

To study the asymptotic properties of the Tikhonov estimators $\tilde{U}$ and $\tilde{V}$, we require the following assumptions.

**Assumption (B.1$^*$ ).** $\lambda_{X,j} \leqslant Cj^{-a_X}$ and $\lambda_{Y,j} \leqslant Cj^{-a_Y}$ for some $a_X, a_Y > 1$ and each $j \geqslant 1$.

**Assumption (B.2$^*$ ).** $\sum_{j_2} \gamma_{j_1 j_2}^2 \leqslant Cj_1^{-2a_X-2b_X}$, $\sum_{j_1} \gamma_{j_1 j_2}^2 \leqslant Cj_2^{-2a_Y-2b_Y}$,

$\epsilon_X \asymp n^{-a_X/(a_X+2b_X)}$, $\epsilon_Y \asymp n^{-a_Y/(a_Y+2b_Y)}$ for some constants $b_X > a_X - 1/2, b_Y > a_Y - 1/2,$.

The following theorem provides an upper bound on the convergence rate of both $\tilde{U}$ and $\tilde{V}$.

**Theorem 3.** *Under assumptions* A.1, A.2, B.1* *and* B.2*, *we have*

$$\|\mathcal{P}_{\hat{\mu}_X}^{\mu_X}\tilde{U} - U\|_{\mu_X}^2 = O_p\left(\max\left\{n^{-(2b_X-1)/(a_X+2b_X)}, n^{-(2b_Y-1)/(a_Y+2b_Y)}\right\}\right);$$

$$\|\mathcal{P}_{\hat{\mu}_Y}^{\mu_Y}\tilde{V} - V\|_{\mu_Y}^2 = O_p\left(\max\left\{n^{-(2b_X-1)/(a_X+2b_X)}, n^{-(2b_Y-1)/(a_Y+2b_Y)}\right\}\right);$$

$$|\tilde{\rho} - \rho|^2 = O_p\left(\max\left\{n^{-(2b_X-1)/(a_X+2b_X)}, n^{-(2b_Y-1)/(a_Y+2b_Y)}\right\}\right).$$

Now we show that the bounds in Theorems 2 and 3 are tight. To this end, for two mean surfaces $\mu_X$ and $\mu_Y$, recall that $\text{Log}_{\mu_X}X$ and $\text{Log}_{\mu_Y}Y$ admit the expansions $\text{Log}_{\mu_X}X = \sum_j \xi_j \Phi_{X,j}$ and $\text{Log}_{\mu_Y}Y = \sum_j \eta_j \Phi_{Y,j}$ for the orthogonal bases $\{\Phi_{X,j}\}_{j=1}^\infty \subset \mathscr{T}(\mu_X)$ and $\{\Phi_{Y,j}\}_{j=1}^\infty \subset \mathscr{T}(\mu_Y)$. Let $P_{XY}$ be the distribution of $(\text{Log}_{\mu_X}X, \text{Log}_{\mu_Y}Y)$ and define the family

$$\mathcal{F}(C,a,b) := \left\{P_{XY} : \sum_{i,j=1}^\infty \frac{\gamma_{ij}^2}{\lambda_{X,j_1}^2\lambda_{Y,j_2}} \leqslant C, \sum_{i,j=1}^\infty \frac{\gamma_{ij}^2}{\lambda_{X,j_1}\lambda_{Y,j_2}^2} \leqslant C, C^{-1}j^{-a} \leqslant \lambda_{X,j} \leqslant Cj^{-a}, \right.$$
$$\left. C^{-1}j^{-a} \leqslant \lambda_{Y,j} \leqslant Cj^{-a}, \sum_j \gamma_{ij}^2 \leqslant Ci^{-2a-2b}, \sum_i \gamma_{ij}^2 \leqslant Cj^{-2a-2b} \right\}.$$

The following result establishes a lower bound on the convergence rate of an estimator of $(U, V)$. The bound matches the upper bound in Theorems 2 and 3 and

thus implies the optimality of the estimators $(\hat{U}, \hat{V})$ and $(\tilde{U}, \tilde{V})$ in the minimax sense.

**Proposition 4.** *Suppose* $(X_1, Y_1), \ldots (X_n, Y_n)$ *form a random sample of* $(X, Y)$ *with* $(\text{Log}_{\mu_X} X, \text{Log}_{\mu_Y} Y)$ *following the distribution* $P_{XY} \in \mathcal{F}(C, a, b, \mu)$*, and that* $(U, V)$ *maximizes* $(3.1)$*. Then*

$$\lim_{c \to 0} \liminf_{n \to \infty} \inf_{(\check{U}, \check{V})} \sup_{P_{XY} \in \mathcal{F}(C, a, b)} P_{XY} \left( \|\check{U} - U\|_{\mu_X}^2 + \|\check{V} - V\|_{\mu_Y}^2 \geqslant C n^{-\frac{2b-1}{a+2b}} \right) = 1,$$

*where* $(\check{U}, \check{V})$ *denotes an estimator of* $(U, V)$ *based on the data* $(X_1, Y_1), \ldots (X_n, Y_n)$.

## 4. Simulation Studies

To illustrate the numerical behavior of the proposed methods, we set $\mu_X(t)$ to the Beta distribution with parameters $(2 + t, 3 - (t^2 + t)/2)$ and $\mu_Y(t)$ to the Beta distribution with parameters $(3 - t, 2 + (t^2 + t)/2)$ for each $t \in [0, 1]$ as the mean surfaces. We consider the set of orthonormal functions

$$\phi_j(x) = \sqrt{2} \sin(\pi j x), \quad \text{for } x \in [0, 1], \text{ and } j \in \mathbb{N}_+.$$

Then $\Phi_{X,j}(x, t) = \phi_j \circ F_{X,t}(x)$ and $\Phi_{Y,j}(x, t) = \phi_j \circ F_{Y,t}(x)$, $j = 1, 2, \ldots$, form orthonormal bases for $\mathscr{T}(\mu_X)$ and $\mathscr{T}(\mu_Y)$, where $F_{X,t}$ and $F_{Y,t}$ denote the distribution function of $\mu_X(t)$ and $\mu_Y(t)$, respectively.

Write $\text{Log}_{\mu_X} X_i(t) = \sum_{j=1}^{\infty} \xi_{ij} \Phi_{X,j} = \sum_{j=1}^{\infty} \xi_{ij} \phi_j \circ F_{X,t}(x)$, where $\xi_{ij}$ are uncorrelated random variables with zero mean such that $\sum_{j=1}^{\infty} \xi_{ij}^2 < \infty$ almost surely.

To guarantee $\sum_{j=1}^{\infty} \xi_{ij}\phi_j \circ F_{X,t}(x) \in \mathrm{Log}_{\mu_X} \mathcal{W}_2^{ac}([0,1])$, where $\mathcal{W}_2^{ac}([0,1])$ denotes

the set of absolutely continuous measures on $[0,1]$, it suffices to require

$$\sum_{j=1}^{\infty} \xi_{ij}\phi_j'\{F_{X,t}(x)\}f_{X,t} + 1 \geqslant 0 \text{ for all } x \in [0,1] \text{ and } t \in \mathcal{T}, \qquad (4.1)$$

where $f_{X,t}$ is the density function of $F_{X,t}$. Condition (4.1) is satisfied, e.g., when

$$\xi_{ij} \leqslant \frac{v_j}{\sup_{x\in[0,1]} |\phi_j'(x)| \sup_{t,x\in[0,1]} f_{X,t}(x) \sum_{j=1}^{\infty} v_j},$$

where $\{v_j\}_{j=1}^{\infty}$ is a non-negative sequence of constants such that $\sum_{j=1}^{\infty} v_j < \infty$,

e.g., $v_j = a^{-j}$ for a given $a > 1$.

Taking $K = 20$, $v_j = 2^{-j}$, $V_j := \sup_{x\in[0,1]} |\phi_j'(x)| = \sqrt{2\pi}j$ and $M = 1.78 >$

$\sup_{x,t\in[0,1]} f_{X,t}(x) = \sup_{x,t\in[0,1]} f_{Y,t}(x)$, we set $X_i(t) = \mathrm{Exp}_{\mu_X}(\sum_{k=1}^{K} \xi_{ik}\Phi_{X,k})$ and

$Y_i(t) = \mathrm{Exp}_{\mu_Y}(\sum_{k=1}^{K} \eta_{ik}\Phi_{Y,k})$. We first consider the following two types of scores

that are widely used in functional data analysis.

- Case 1 (Truncated normal): We sample $\xi_{ik} \sim v_k(V_kM)^{-1}\theta_{ik}$ with $\theta_{ik} \sim$

  $\mathrm{TN}_{[-1,1]}(0,1)$ independently for $i = 1,2,\cdots,n$ and $k = 1,2,\cdots,K$, and

  $\eta_{ik} \sim v_k(V_kM)^{-1}\vartheta_{ik}$ with $\vartheta_{ik} \sim \mathrm{TN}_{[-1,1]}(0,1)$, except that $\eta_{i2} = 0.5(\xi_{i1} +$

  $\xi_{i2}) + \sigma\mathbf{E}(\xi_{i1}^2 + \xi_{i2}^2)\vartheta_{i2}$, where $\mathrm{TN}_{[-1,1]}(0,1)$ denotes the Gaussian distribution

  $N(0,1)$ truncated on $[-1,1]$, and $\sigma$ is a constant representing the noise level.

- Case 2 (Uniform): We sample $\xi_{ik} \sim \mathrm{Unif}[-v_k(V_kM)^{-1}, v_k(V_kM)^{-1}]$ indepen-

  dently for $i = 1,\cdots,n$ and $k = 1,\cdots,K$, and $\eta_{ik} \sim \mathrm{Unif}[-v_k(V_kM)^{-1}, v_k(V_kM)^{-1}]$,

  except that $\eta_{i2} = 0.5(\xi_{i1} + \xi_{i2}) + \sigma\mathbf{E}(\xi_{i1}^2 + \xi_{i2}^2)\vartheta_{i2}$ with $\vartheta_{i2} \sim \mathrm{Unif}[-1,1]$.
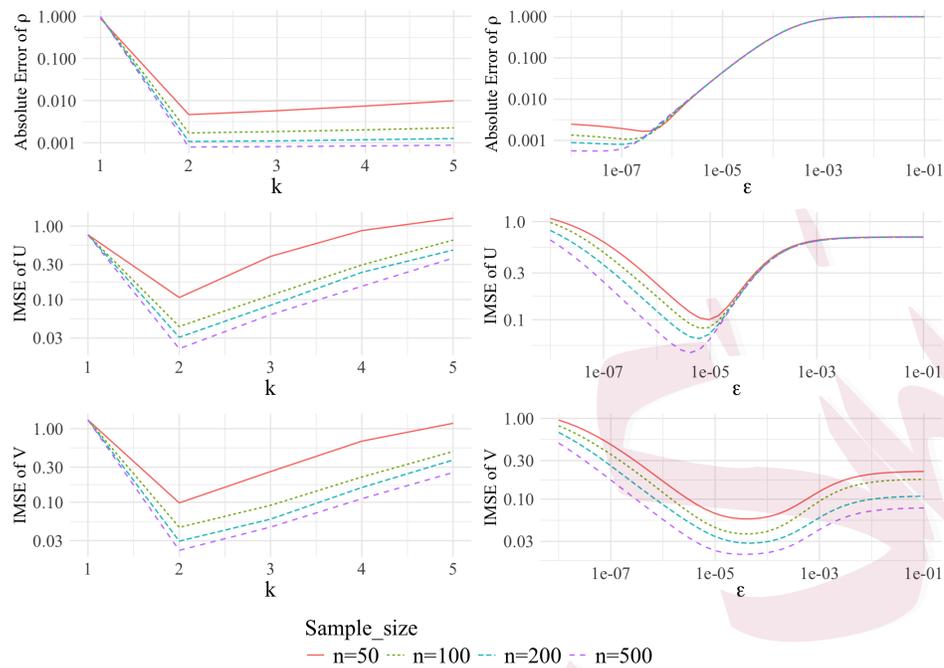
Figure 2: The absolute error $|\hat{\rho} - \rho|$ ($|\tilde{\rho} - \rho|$, respectively) and IMSE for $\hat{U}$ and $\hat{V}$ ($\tilde{U}$ and $\tilde{V}$, respectively)on different tuning parameters for the FPCA (left column) and Tikhonov (right column) methods by the average of 200 Monte Carlo replicates with noise level $\sigma = 0.05$ in Case 1.

In this construction we have $U \propto \Phi_{X,1} + \Phi_{X,2}$, $V \propto \Phi_{Y,2}$ and $\rho^2 = 0.5^2/(0.5^2 + \sigma^2)$. We first inspect how the estimation error responds to the tuning parameters for FPCA and Tikhonov methods, where we set the same tuning parameters for $X$ and $Y$ for simplicity, i.e., $k_X = k_Y = k$ and $\epsilon_X = \epsilon_Y = \epsilon$, respectively. Four sample sizes from 50 to 500 are considered, and each simulation is repeated 200 times independently. We use the absolute error $|\hat{\rho} - \rho|$ to quantify the estimation

error of $\hat{\rho}$ and integrated mean squared error $\text{IMSE}(\hat{U}) = (\mathbb{E}\|\hat{U} - U\|_\mu^2)^{1/2}$ of $\hat{U}$; estimation errors for $\hat{V}$, $\tilde{\rho}$, $\tilde{U}$ and $\tilde{V}$ are quantified analogously. From the results in Figure 2 we see that the estimators converge rapidly as the sample size $n$ increases. In addition, we observe that the optimal tuning parameters for $\rho$, $U$ and $V$ are rather different for Tikhonov estimators, e.g., the $\epsilon$ that minimizes $|\hat{\rho} - \rho|$ may not be the optimal choice for estimating the weight functions $U$ and $V$. By contrast, the optimal truncation parameter $k$ for $\rho, U$ and $V$ seems to have little difference in FPCA method.

In practice, we propose to choose the tuning parameters via a $\kappa$-fold cross-validation (CV) approach with $\kappa = 5$. Specifically, we split the dataset into $\kappa$ partitions of roughly even size. Taking the FPCA estimation for example, let $\hat{U}_k^{-l}$ and $\hat{V}_k^{-l}$ be the FPCA estimators of the weight functions with the tuning (truncation) parameter $k$ obtained without using data from the $l$-th partition. The cross-validation score of $k$ is defined based on the squared Pearson correlation (Leurgans et al., 1993)

$$\text{CV}(k) = \left[ \frac{n \left( \sum_{i=1}^n x_{i,k} y_{i,k} \right) - \left( \sum_{i=1}^n x_{i,k} \right) \left( \sum_{i=1}^n y_{i,k} \right)}{\sqrt{\left\{ n \sum_{i=1}^n x_{i,k}^2 - \left( \sum_{i=1}^n x_{i,k} \right)^2 \right\} \left\{ n \sum_{i=1}^n y_{i,k}^2 - \left( \sum_{i=1}^n y_{i,k} \right)^2 \right\}}} \right]^2 \tag{4.2}$$

with

$$x_{i,k} = \sum_{l=1}^\kappa \langle \text{Log}_{\hat{\mu}_X} X_i, \hat{U}_k^{-l} \rangle \mathbf{1}_{i \in I_l} \text{ and } y_{i,k} = \sum_{k=1}^\kappa \langle \text{Log}_{\hat{\mu}_Y} Y_i, \hat{V}_k^{-l} \rangle \mathbf{1}_{i \in I_l},$$

Table 1: The absolute error $|\hat{\rho} - \rho|$ ($|\tilde{\rho} - \rho|$, respectively) for different noise levels with tuning parameters chosen by 5-fold CV, where the values are multiplied by $10^3$ for visualization.

|  | | FPCA | | | | Tikhonov | | | |
|---|---|---|---|---|---|---|---|---|---|
|  | $\sigma$ | n=50 | n=100 | n=200 | n=500 | n=50 | n=100 | n=200 | n=500 |
| Case 1 | 0.05 | 1.892 | 1.097 | .8242 | .5604 | 38.41 | 12.61 | 6.853 | 4.703 |
|  | 0.1 | 7.381 | 4.311 | 3.232 | 2.172 | 44.89 | 21.65 | 14.28 | 9.289 |
|  | 0.2 | 26.17 | 15.55 | 11.58 | 7.449 | 74.89 | 35.73 | 25.25 | 17.18 |
|  | 0.3 | 48.67 | 28.72 | 21.63 | 13.47 | 111.8 | 49.53 | 36.76 | 22.42 |
|  | 0.5 | 84.16 | 49.89 | 37.40 | 21.87 | 159.7 | 74.89 | 51.28 | 27.18 |
| Case 2 | 0.05 | 1.822 | 1.049 | .7877 | .5194 | 28.80 | 12.77 | 6.753 | 4.683 |
|  | 0.1 | 7.125 | 4.095 | 3.092 | 2.008 | 44.63 | 21.65 | 13.99 | 9.218 |
|  | 0.2 | 25.17 | 14.88 | 11.06 | 6.967 | 74.31 | 35.71 | 24.67 | 17.01 |
|  | 0.3 | 47.61 | 28.18 | 20.96 | 12.75 | 101.7 | 49.74 | 32.47 | 21.95 |
|  | 0.5 | 82.50 | 49.17 | 36.46 | 21.10 | 137.9 | 75.76 | 47.43 | 26.49 |

where $I_l$ denotes the set of indices of data in the $l$-th partition. Then we choose the value of $k$ that maximizes $\mathrm{CV}(k)$. Selection for the tuning parameters associated with the Tikhonov estimators $\tilde{U}$ and $\tilde{V}$ can be performed analogously.

Tables 1–2 report the absolute error $|\hat{\rho} - \rho|$ ($|\tilde{\rho} - \rho|$, respectively) and the IMSE for estimating $U$, where the tuning parameters are selected by the 5-fold CV (4.2). The supplementary material provides the IMSE for estimating $V$. From Table 1, we see that the correlation $\rho$ is estimated well by both methods and converges rapidly as the sample size increases. When inspecting the estimation error for $\rho$ in Figure 2 and Table 1 closely, one may find that Figure 2 shows similar results between two methods ($|\hat{\rho} - \rho|$ v.s. $|\tilde{\rho} - \rho|$) over the illustrated ranges of tuning parameters, while Table 1 shows that $|\hat{\rho} - \rho|$ by the FPCA method is clearly lower than that by the Tikhonov method when the tuning parameters are chosen by the 5-fold CV. This might be explained by Figure 3, which presents the selected tuning parameters in both methods. For the FPCA method, most of the selected truncation numbers are concentrated in 2, that is, exactly the optimal tuning parameter. Figure 2 also indicates that the optimal range of $\epsilon$ is around $(10^{-8}, 10^{-6})$ (the optimal choice of $\epsilon$ is decreasing as the sample size grows, which is consistent with our theoretical results). However, as shown in Figure 3, a substantial proportion of the selected values do not fall into this range even as the sample size increases, which deteriorates the performance of the Tikhonov method. In addition, for the

Table 2: IMSE of $\hat{U}$ ($\tilde{U}$, respectively) for different noise levels with tuning parameters chosen by 5-fold CV.

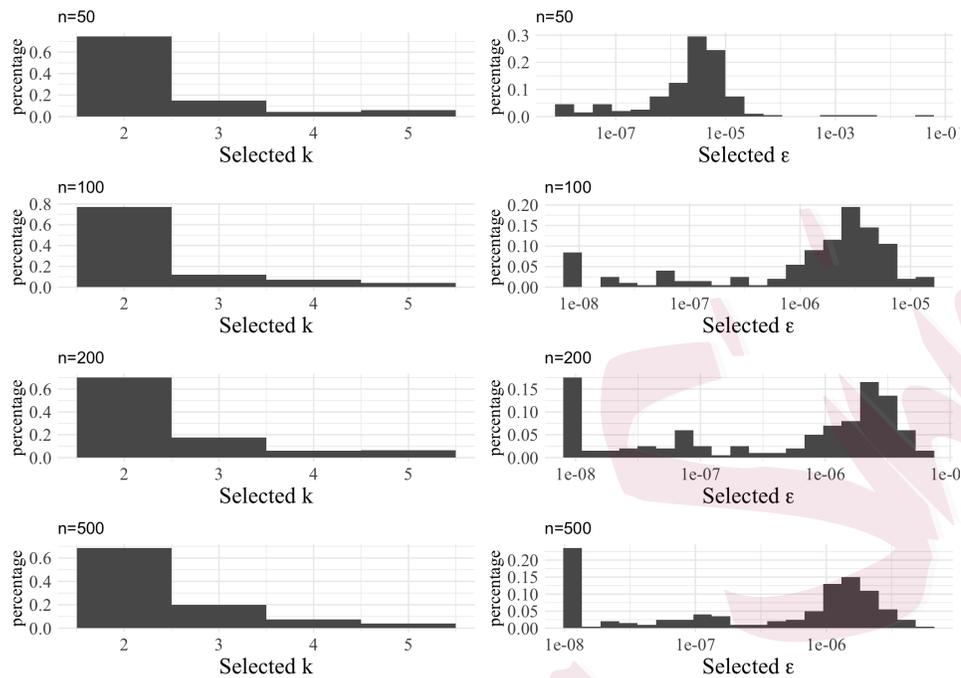|  |  | FPCA | | | | Tikhonov | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | $\sigma$ | n=50 | n=100 | n=200 | n=500 | n=50 | n=100 | n=200 | n=500 |
| Case 1 | 0.05 | .1281 | .0900 | .0850 | .0481 | .2536 | .2580 | .3033 | .2498 |
|  | 0.1 | .1935 | .1446 | .1377 | .0863 | .2846 | .2519 | .2977 | .2888 |
|  | 0.2 | .2979 | .2194 | .2303 | .1417 | .3538 | .3154 | .3156 | .2481 |
|  | 0.3 | .3806 | .2880 | .2747 | .1873 | .4318 | .3783 | .3328 | .2551 |
|  | 0.5 | .5237 | .3816 | .3429 | .2808 | .5322 | .4876 | .3912 | .3067 |
| Case 2 | 0.05 | .1312 | .0870 | .0847 | .0492 | .2546 | .2537 | .3006 | .2577 |
|  | 0.1 | .1994 | .1419 | .1447 | .0828 | .2928 | .2494 | .3037 | .2966 |
|  | 0.2 | .2938 | .2124 | .2278 | .1460 | .3562 | .3131 | .3199 | .2497 |
|  | 0.3 | .3789 | .2811 | .2744 | .1911 | .4279 | .3723 | .3359 | .2582 |
|  | 0.5 | .5178 | .3867 | .3457 | .2764 | .5241 | .4803 | .3853 | .3030 |

Figure 3: Histograms of selected tuning parameters in FPCA (left column) and Tikhonov (right column) methods by 5-fold CV with noise level $\sigma = 0.05$ in Case 1.

Tikhonov method, the CV procedure (4.2) aiming for $\rho$ seems not optimal for the weight functions $U$ and $V$, as observed from Figure 2. This may partially explain why the Tikhonov method yields larger estimation errors for the weight functions than the FPCA method, as exhibited in Tables 2. Additional figures and results can be found in supplementary material.

In theory, we only require the decay rates of eigenvalues of the covariance operator, which correspond to the decay rates of the variances of the functional

principal scores. Besides the normal and uniform scores, we also consider the following setting with log-normal scores to see whether our proposed method can adapt to the heavy-tailed distribution. Details on the settings and convergence rates for the weight functions can be found in the supplementary material. The results exhibit a pattern similar to that in Figure 2, demonstrating the robustness of our proposed method.

## 5. Application

We apply the proposed Wasserstein correlation analysis to study functional connectivities between different brain regions by using data from the HCP 1200 subjects release (Van Essen et al., 2013) that is available at `https://db.humanconnectome.org/app/template/Index.vm`. Specifically, we focus on the correlation of longitudinally measured distributions of the signal strength between two specific areas in the brain.

The dataset we use consists of $n = 209$ subjects who are healthy young adults and have been scanned for both a resting-state fMRI (rsfMRI) image and a task-evoked fMRI (tfMRI) image related to fine motor skills. The rsfMRI data were acquired in four runs of approximately 15 minutes each, two runs in one session and two in the other session. During data acquisition, subjects were instructed to keep their eyes open and fixed on a projected bright cross-hair on a dark background that

Table 3:   Top five largest correlations obtained by FPCA, Tikhonov and traditional FDA methods for the fMRI data.

| $\rho$ | | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| FPCA | rfMRI | .6966 | .3505 | .2970 | .2593 | .2474 |
| | tfMRI | .7476 | .4448 | .4108 | .3093 | .2809 |
| Tikhonov | rfMRI | .7250 | .3246 | .2510 | .2160 | .1840 |
| | tfMRI | .7457 | .3450 | .2960 | .2135 | .1300 |
| Traditional FDA | rfMRI | .6967 | .4314 | .0773 | .0032 | .0001 |
| | tfMRI | .6874 | .6676 | .3827 | .2907 | .0860 |

was presented in a darkened room. In the tfMRI data acquisition, subjects were presented with visual cues that asked them to either tap their left or right fingers, squeeze their left or right toes, or move their tongues to map motor areas. During the experiment, the brain of each subject was scanned and the neural activities were recorded at 284 equal-spaced time points. More details of the experiment and data acquisition can be found in the reference manual of WU-Minn HCP 1200 Subjects Data Release.

We consider the Putamen and Caudate nucleus areas, which are known to be related to motor skills in medical literature. At each time $t$, $X_i^{rest}(t)$ and $X_i^{motor}(t)$

denote the distributions of the signal strength in the left Caudate nucleus of the $i$-th

subject for rsfMRI and tfMRI, respectively, and $Y_i^{rest}(t)$ and $Y_i^{motor}(t)$ represent the

distributions of the signal strength in the Putamen area. The proposed Wasserstein

correlation analysis is applied to investigate the correlation between $X_i^{rest}(t)$ and

$Y_i^{rest}(t)$, as well as the correlation between $X_i^{motor}(t)$ and $Y_i^{motor}(t)$. The tuning

parameters are selected by 5-fold CV (4.2).

From Table 3 that reports the top five correlations, we observe that FPCA

and Tikhonov methods yield similar patterns that the correlation between these

two areas in the tfMRI images is larger relative to its rsfMRI counterpart, while

difference seems more pronounced by FPCA method. For comparison, we ran the

traditional functional canonical correlation analysis by taking the average of the

signals within each brain region for every observation time. As shown in Table 3,

the traditional functional canonical correlation analysis fails to capture the change

in correlations from resting-state to motor fMRI. From Figure 4, we see that,

compared with rsfMRI, fluctuation of weight functions along time in tfMRI is

more intensive, which suggests that the association between the two brain regions

is more dynamic during a motor task. In summary, via the proposed method

we find that the correlations for distributions of the signal strength between the

Putamen area and Caudate nucleus increase during the motor task in contrast to

the resting state, moreover, the weight functions in tfMRI express a more active
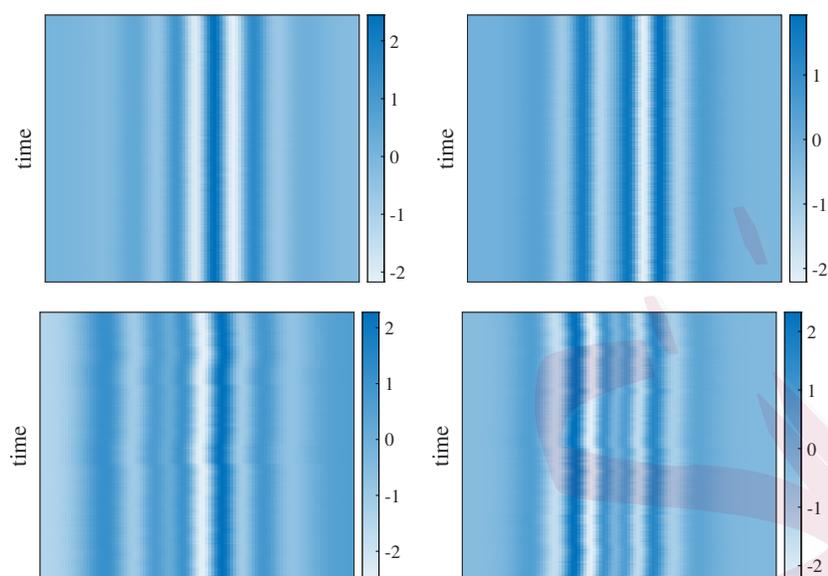
Figure 4: Heat-map for the estimated weight functions $\hat{U}$ (left) and $\hat{V}$ (right) for the rsfMRI data (top) and tfMRI data (bottom).

pattern than those in rsfMRI.

## Supplementary Material

The supplementary materials include detailed technical proofs and additional simulation results.

## Acknowledgments

## References

Agueh, M. and G. Carlier (2011). Barycenters in the wasserstein space. SIAM J. Math. Anal. 43, 904–924.

Ambrosio, L., N. Gigli, and G. Savaré (2008). Gradient Flows: in Metric Spaces and in the Space of Probability Measures. Springer Science & Business Media.

Benamou, J.-D. and Y. Brenier (2000). A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. Numer. Math. 84(3), 375–393.

Bigot, J., R. Gouet, T. Klein, A. López, et al. (2017). Geodesic PCA in the wasserstein space by convex PCA. Ann. inst. Henri Poincare (B) Probab. Stat. 53(1), 1–26.

Brenier, Y. (1991). Polar factorization and monotone rearrangement of vector-valued functions. Commun. Pure Appl. Math. 44(4), 375–417.

Chen, Y., Z. Lin, and H.-G. Müller (2023). Wasserstein regression. J. Amer. Statist. Assoc. 118(542), 869–882.

Cuesta, J. A. and C. Matran (1989). Notes on the Wasserstein Metric in Hilbert Spaces. Ann. Probab. 17(3), 1264 – 1276.

Dai, X. (2022). Statistical inference on the hilbert sphere with application to random densities. Electron.

J. Stat. <u>16</u>(1), 700–736.

Dai, X., Z. Lin, and H.-G. Müller (2021). Modeling sparse longitudinal data on Riemannian manifolds. <u>Biometrics</u> <u>77</u>(4), 1328–1341.

Dai, X. and H.-G. Müller (2018). Principal component analysis for functional data on Riemannian manifolds and spheres. <u>Ann. Stat.</u> <u>46</u>(6B), 3334 – 3361.

Delaigle, A. and P. Hall (2012). Achieving near perfect classification for functional data. <u>J. R. Stat. Soc., B: Stat. Methodol.</u> <u>74</u>(2), 267–286.

Dou, W. W., D. Pollard, H. H. Zhou, et al. (2012). Estimation in functional regression for general exponential families. <u>Ann. Stat.</u> <u>40</u>(5), 2421–2451.

Eubank, R. and T. Hsing (2008). Canonical correlation for stochastic processes. <u>Stoch. Process. their Appl.</u> <u>118</u>(9), 1634–1661.

Ferraty, F. and P. Vieu (2006). <u>Nonparametric Functional Data Analysis: Theory and Practice</u>. Springer Science & Business Media.

Fréchet, M. (1948). Les éléments aléatoires de nature quelconque dans un espace distancié. In <u>Ann. inst. Henri Poincare</u>, Volume 10, pp. 215–310.

Freulon, P., J. Bigot, and B. P. Hejblum (2023). Cytopt: Optimal transport with domain adaptation for interpreting flow cytometry data. <u>Ann. Appl. Stat.</u> <u>17</u>(2), 1086–1104.

Gajardo, Á. and H.-G. Müller (2023). Point process models for covid-19 cases and deaths. <u>J. Appl. Stat.</u> <u>50</u>(11-12), 2294–2309.

## REFERENCES

Gangbo, W. and R. J. McCann (1996). The geometry of optimal transportation. Acta Math. 177(2), 113–161.

Hall, P. and J. L. Horowitz (2007). Methodology and convergence rates for functional linear regression. Ann. Stat. 35(1), 70–91.

Hall, P. and M. Hosseini-Nasab (2006). On properties of functional principal components analysis. J. R. Stat. Soc., B: Stat. Methodol. 68(1), 109–126.

He, G., H.-G. Müller, and J.-L. Wang (2003). Functional canonical analysis for square integrable stochastic processes. J. Multivar. Anal. 85(1), 54–77.

Horváth, L. and P. Kokoszka (2012). Inference for Functional Data with Applications. Springer Science & Business Media.

Hsing, T. and R. Eubank (2015). Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators. John Wiley & Sons.

James, G. M. and C. A. Sugar (2003). Clustering for sparsely sampled functional data. J. Amer. Statist. Assoc. 98(462), 397–408.

Kantorovich, L. V. (2006). On a problem of monge. J. Math. Sci.(NY) 133, 1383.

Kokoszka, P. and M. Reimherr (2017). Introduction to Functional Data Analysis. CRC press.

Leurgans, S. E., R. A. Moyeed, and B. W. Silverman (1993). Canonical correlation analysis when the data are curves. J. R. Stat. Soc., B: Stat. Methodol. 55(3), 725–740.

Lian, H. (2014). Some asymptotic properties for functional canonical correlation analysis. J. Stat. Plan.

Inference $\underline{153}$, 1–10.

Lin, Z., D. Kong, and L. Wang (2023). Causal inference on distribution functions. J. R. Stat. Soc., B: Stat. Methodol. $\underline{85}$(2), 378–398.

Lin, Z. and H.-G. Müller (2021). Total variation regularized fréchet regression for metric-space valued data. Ann. Stat. $\underline{49}$(6), 3510–3533.

Lin, Z. and F. Yao (2019). Intrinsic Riemannian functional data analysis. Ann. Stat. $\underline{47}$(6), 3533 – 3577.

Lin, Z. and F. Yao (2021). Functional regression on the manifold with contamination. Biometrika $\underline{108}$(1), 167–181.

Major, P. (1978). On the invariance principle for sums of independent identically distributed random variables. J. Multivar. Anal. $\underline{8}$(4), 487–517.

McCann, R. J. (1997). A convexity principle for interacting gases. Advances in Mathematics $\underline{128}$(1), 153–179.

Panaretos, V. M. and Y. Zemel (2020). An Invitation to Statistics in Wasserstein Space. Springer Nature.

Petersen, A. and H.-G. Müller (2016). Functional data analysis for density functions by transformation to a Hilbert space. Ann. Stat. $\underline{44}$(1), 183–218.

Petersen, A. and H.-G. Müller (2019). Wasserstein covariance for multiple random densities. Biometrika $\underline{106}$(2), 339–351.

Ramsay, J. and B. Silverman (2006). Functional Data Analysis. Springer Science & Business Media.

Shao, L., Z. Lin, and F. Yao (2022). Intrinsic riemannian functional data analysis for sparse longitudinal

observations. Ann. Stat. 50(3), 1696–1721.

Sommerfeld, M. and A. Munk (2018). Inference for empirical wasserstein distances on finite spaces. J. R. Stat. Soc., B: Stat. Methodol. 80(1), 219–238.

Van Essen, D. C., S. M. Smith, D. M. Barch, T. E. Behrens, E. Yacoub, K. Ugurbil, W.-M. H. Consortium, et al. (2013). The WU-Minn human connectome project: an overview. Neuroimage 80, 62–79.

Wahl, M. (2022). Information inequalities for the estimation of principal components. Ann. Inst. Henri Poincare (B) Probab. Stat. 58(3), 1565–1589.

Yang, W., H.-G. Müller, and U. Stadtmüller (2011). Functional singular component analysis. J. R. Stat. Soc., B: Stat. Methodol. 73(3), 303–324.

Yao, F., H.-G. Müller, and J.-L. Wang (2005a). Functional data analysis for sparse longitudinal data. J. Amer. Statist. Assoc. 100(470), 577–590.

Yao, F., H.-G. Müller, and J.-L. Wang (2005b). Functional linear regression analysis for longitudinal data. Ann. Stat. 33(6), 2873–2903.

Yuan, M. and T. T. Cai (2010). A reproducing kernel Hilbert space approach to functional linear regression. Ann. Stat. 38(6), 3412–3444.

Zhang, J., K. R. Merikangas, H. Li, and H. Shou (2022). Two-sample tests for multivariate repeated measurements of histogram objects with applications to wearable device data. Ann. Appl. Stat. 16(4), 2396–2416.

Zhou, Y. and D.-R. Chen (2020). The optimal rate of canonical correlation analysis for stochastic

processes. J. Stat. Plan. Inference <u>207</u>, 276–287.

Hang Zhou

Department of Statistics, University of California, Davis, U.S.A.

E-mail: hgzhou@ucdavis.edu

Zhenhua Lin

Department of Statistics and Applied Probability, National University of Singapore, Singapore

E-mail: linz@nus.edu.sg

Fang Yao

School of Mathematical Sciences, Center for Statistical Science, Peking University, Beijing, China

E-mail: fyao@math.pku.edu.cn