
Functional data analysis for multivariate distributions through Wasserstein slicing

Han Chen

Department of Statistics
University of California, Davis
Davis, CA 95616
nahchen@ucdavis.edu

Hans-Georg Müller

Department of Statistics
University of California, Davis
Davis, CA 95616
hgmueeller@ucdavis.edu

Abstract

The modeling of samples of distributions is a major challenge since distributions do not form a vector space. While various approaches exist for univariate distributions, including transformations to a Hilbert space, far less is known about the multivariate case. We utilize a transformation approach to map multivariate distributions to a Hilbert space via a Wasserstein slicing method that is invertible. This approach combines functional data analysis tools, such as functional principal component analysis and modes of variation, with the facility to map back to interpretable distributions. We also provide convergence guarantees for the Hilbert space representations under a broad class of such transforms. The method is illustrated using joint systolic and diastolic blood pressure data.

1 Introduction

Data that consist of multivariate distributions are becoming increasingly available. Instances of such data encompass anthropometric data distributions [24], stock price data distributions [19] and distributions of temperatures in climate modeling. Analyzing distributional data presents distinctive challenges, as they do not form a vector space, rendering conventional methods designed for vector and functional data inapplicable.

For data that correspond to samples of univariate distributions, invertible global transformations such as the log quantile density transform and log hazard transform have been employed to map distributions to a Hilbert space L^2 [35]. While various maps were proposed to embed metric spaces into a Hilbert space, starting with Schoenberg’s kernel embedding [43, 44] and also including the centered log-ratio transformation [23, 30, 24], these maps are generally not 1:1 and therefore not invertible. However, invertibility is crucial as functional principal component analysis or regression in Hilbert space involves elements that can be located anywhere in the Hilbert space. To get back to the original distribution space, which is a must for interpretability and visualization, therefore requires invertibility of the map.

Another example of such a non-invertible embedding for the Wasserstein space of univariate distributions is to use Riemannian type log maps to map densities to tangent bundles. The resulting tangent spaces, are Hilbert spaces, where the development of principal component analysis [5] and regression models [12] is straightforward, however the reverse exp map is not defined on the entire Hilbert space but only on a convex cone and this poses a major problem, as only ad hoc solutions such as projecting elements outside the invertibility cone on the cone are available. Statistical regression models targeting univariate distributions as responses and Euclidean predictors can be implemented with Fréchet regression [36], where also other distribution-to-distribution regression models with viable implementations have been proposed [16, 49].

While these approaches have been developed for univariate distributions, multivariate distributional data have been less investigated; while kernel embeddings are still available, the invertibility problem remains. For regression with multivariate distributions, [34] introduce a regression model but its scope is limited to the case of multivariate Gaussian distributions, while [17] provides some theoretical results for a more general settings for an optimal transport map regression approach in analogy to the case of univariate distributions [16]. Another approach utilizes a Bayes Hilbert space over bivariate domains to model a time series of copula probability density functions [19], further developed for bivariate distributional data through spline representations for bivariate densities in the Bayes Hilbert space [24]. The Bayes Hilbert space proposals do not offer theoretical support and also do not address the estimation of densities from available random samples that they generate. Wasserstein geodesic principal component analysis represents variability among probability measures along geodesics in Wasserstein space, offering closed-form solutions for the case of Gaussian distributions and neural-network approximations for more general scenarios. However, it lacks an explicit inverse map and does not guarantee invertibility for non-Gaussian distributions [45].

Our goal in this paper is to extend the univariate log quantile density and log hazard transformation method [35] to the case of multivariate distributions. Then functional data analysis methodology, including functional principal component analysis (FPCA), functional modes of variation and functional regression can be applied in the Hilbert space into which the distributions are mapped. The densities corresponding to the Hilbert space representations are then obtained through a regularized inverse map [9]. Theoretical results regarding the convergence of these representations in density space are derived under regularity conditions, drawing strength from known results for FPCA and addressing the complexities introduced by the transformations. Given that multivariate distributions are often not fully observed and need to be recovered from random samples generated by the underlying distributions, we also consider the challenges introduced by the estimation step.

2 Preliminaries

2.1 Univariate Density Transformation

Let $\mathcal{G} = \{g \in \mathcal{C}^0(\mathcal{Z}) : g > 0, \int_0^1 g(z)dz = 1\}$ be the space of strictly positive densities on a common compact interval $\mathcal{Z}$. Without loss of generality, assume the support for the univariate density function is $\mathcal{Z} = [0, 1]$. [35] propose a family of transformations $\psi : \mathcal{G} \rightarrow L^2(\mathcal{S})$ that satisfy assumptions (L1)-(L3) that are listed in Appendix A. These transformations map continuous univariate densities to a new space $L^2(\mathcal{S})$, where $\mathcal{S}$ is a compact interval.

Here we focus on the log quantile density (LQD) transformation, while the log hazard transformation is another prominent example for this class of transformations and is reviewed in Appendix D. Choosing $\mathcal{S} = [0, 1]$, the LQD transformation $\psi_Q : \mathcal{G} \rightarrow L^2(\mathcal{S})$ is given by

$$\psi_Q(g)(s) = -\log(g(G^{-1}(s))), \quad s \in \mathcal{S}, \quad (1)$$

where $G(z) = \int_0^z g(u)du$ is the cumulative distribution function and G^{-1} is the corresponding quantile function supported on $\mathcal{S}$. The inverse of a continuous function φ on $\mathcal{S}$ can be written as $\exp\{-\varphi(G_\varphi(z))\}$, where $G_\varphi^{-1}(s) = \int_0^s e^{\varphi(u)}du$. Since $G_\varphi^{-1}(1)$ is not fixed, a boundary adjustment guarantees that the inverse will be a density on $\mathcal{Z} = [0, 1]$ [37],

$$\psi_Q^{-1}(\varphi)(z) = \alpha_\varphi \exp\{-\varphi(G_\varphi(z))\}, \quad G_\varphi^{-1}(s) = \alpha_\varphi^{-1} \int_0^s e^{\varphi(u)}du,$$

where $\alpha_\varphi = \int_0^1 e^{\varphi(u)}du$. The LQD transformation is an element of a family of a larger class of univariate density transformations; however, similar transformations are currently not available for multivariate density functions. We fill this gap by introducing a generalized multivariate transformation in Section 3.2, which will be illustrated with examples based on multivariate extensions of the LQD transformation.

2.2 Density Slicing Transformation

To connect multivariate and univariate distributions, we use the Radon transform $\mathcal{R}$ [40], which is an integral transform that takes an integrable p -dimensional function and maps it to an infinite set of

integrals over hyperplanes in $\mathbb{R}^p$. Let θ be a unit vector in $\Theta = \{z \in \mathbb{R}^p, \|z\|_2 = 1\}$ and z an element in $\mathbb{R}$. We denote affine hyperplanes by $l_{z,\theta} = \{x \in \mathbb{R}^p : \langle x, \theta \rangle = z\}$ and select an orthonormal basis $\{\theta, e_1, \dots, e_{p-1}\}$ in $\mathbb{R}^p$ such that $\langle \theta, e_j \rangle = 0$ and $\langle e_j, e_l \rangle = \delta_{jl}$, for $j, l = 1, \dots, p-1$. We then define the p -dimensional Radon transform $\mathcal{R}$ through the integral over $l_{z,\theta}$ as follows,

$$\mathcal{R}(f)(\theta, z) = \int_{\mathbb{R}^{p-1}} f\left(z\theta + \sum_{j=1}^{p-1} s_j e_j\right) ds_1 \cdots ds_{p-1}, \quad \text{for } \theta \in \Theta \text{ and } z \in \mathbb{R}, \quad (2)$$

where f is a p -dimensional multivariate density function supported on D .

The inverse Radon transform has undergone thorough investigation, both theoretically and numerically [1, 22, 31]. The commonly used method to reconstruct the original multivariate distribution from its Radon transform ϕ is through the filtered back-projection given by

$$\mathcal{R}^{-1}(\phi)(x) = \frac{1}{2(2\pi)^p} \int_{\Theta} \int_{\mathbb{R}} \mathcal{J}(\phi(\theta, \cdot))(v) e^{ir\langle \theta, x \rangle} |v|^{p-1} dv d\theta, \quad (3)$$

where the Fourier transform is $\mathcal{J}(g)(v) = \int_{\mathbb{R}} g(u) e^{-iuv} du$ for all $v \in \mathbb{R}$. There are cases where Radon transforms are bona fide probability densities, while the inverse is not, such as the Wigner quasi-probability distribution in quantum state tomography [2]. Moreover, the inverse map $\mathcal{R}^{-1}$ is not continuous, and small reconstruction errors in $\mathcal{R}(f)$ can be magnified [15]. To address this, a regularized inverse [15, 25, 9] using a filter function in the Fourier domain is commonly applied,

$$\check{\mathcal{R}}_{\tau}^{-1}(\phi)(x) = \frac{1}{2(2\pi)^p} \int_{\Theta} \int_{\mathbb{R}} \mathcal{J}(\phi(\theta, \cdot))(v) e^{ir\langle \theta, x \rangle} |v|^{p-1} \mathbf{1}_{|v| \leq \tau} dv d\theta. \quad (4)$$

Note that the regularization parameter τ balances approximation accuracy and the continuity of the inverse map [9]. As this regularized inverse is not guaranteed to be a multivariate density function, normalization is applied to map the resulting L^1 function into the multivariate density space $\mathcal{F}$ via

$$\mathcal{R}_{\tau}^{-1}(\phi)(x) = \begin{cases} |\check{\mathcal{R}}_{\tau}^{-1}(\phi)(x)| / \int_D |\check{\mathcal{R}}_{\tau}^{-1}(\phi)(x)| dx & \text{if } \int_D |\check{\mathcal{R}}_{\tau}^{-1}(\phi)(x)| dx > 0, \\ 1/|D| & \text{otherwise.} \end{cases} \quad (5)$$

Here $|D|$ represents the Lebesgue measure of D which is the domain of the multivariate density function f . Note that the Radon transform and its inverse are part of a generalized slicing transform family, for which an explicit invertible embedding between multivariate and sliced univariate densities was introduced in [9], where further details can be found.

3 Multivariate Density Transformation

3.1 Assumptions for Multivariate Densities

By convention, $\mathcal{C}^0$ denotes the set encompassing all continuous functions while $\mathcal{C}^1$ denotes the set consisting of all continuously differentiable functions. Let $\|\cdot\|_2$ be the L^2 metric and $\|\cdot\|_{\infty}$ be the uniform metric for measurable functions. Throughout, $C_0, C_1, \dots$ represent various constants, and their dependency on pertinent quantities ϱ will be denoted by writing $C_0(\varrho), C_1(\varrho), \dots$.

We assume that the multivariate distributions that we consider have density functions and that their common support is $D \subset \mathbb{R}^p$, where

(D1) The support set $D \subset \mathbb{R}^p$ is compact and convex.

Denote the space of strictly positive multivariate density functions on $\mathbb{R}^p$ with compact support D by

$$\mathcal{F} = \left\{ f(x) \in L^1(\mathbb{R}^p) : f(x) > 0, \int_{\mathbb{R}^p} f(x) dx = 1, \text{support}(f) = D, f \text{ satisfies (F1)} \right\}.$$

(F1) For all $f \in \mathcal{F}$, there exists a constant $M_0 > 0$ such that $\max\{\|f\|_{\infty}, \|1/f\|_{\infty}\} \leq M_0$ and for a $\kappa \geq p+1$, f is continuously differentiable of order κ on $D \subset \mathbb{R}^p$ and has uniformly bounded partial derivatives.

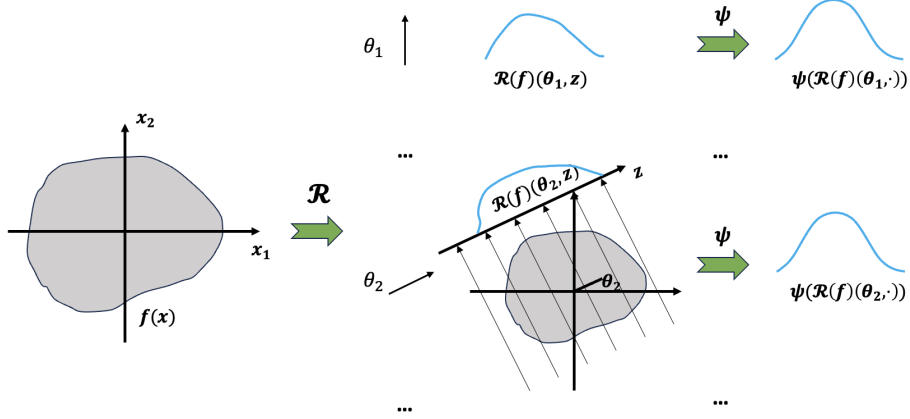


Figure 1: Scheme for the Radon Log Quantile Density (RLQD) transformation for the bivariate case: Slicing the bivariate density function along each projection (represented by an angle) through the Radon transform, followed by the slice-wise log quantile density transform.

3.2 Multivariate Density Transformation

We introduce multivariate density transformations $\Psi : \mathcal{F} \rightarrow L^2(\Theta \times \mathcal{S})$ from multivariate density functions to a Hilbert space as follows. Let ψ, ψ^{-1} be the generalized univariate density transformation and its inverse reviewed in Section 2.1, then

$$\Psi(f)(\theta, s) = \psi(\mathcal{R}(f)(\theta, \cdot))(s), \quad \text{for } (\theta, s) \in \Theta \times \mathcal{S}, f \in \mathcal{F}. \quad (6)$$

By using the regularized inverse of the Radon transform, the regularized inverse can be defined for any continuous function ζ as

$$\Psi_{\tau}^{-1}(\zeta)(x) = \mathcal{R}_{\tau}^{-1}((\psi^{-1}(\zeta(\theta, \cdot)))(z))(x). \quad (7)$$

A concrete example is the Radon log quantile density (RLQD) transformation Ψ_Q ,

$$\Psi_Q(f)(\theta, s) = \psi_Q(\mathcal{R}(f)(\theta, \cdot))(s) = -\log(\mathcal{R}(f)(\theta, R_{\theta}^{-1}(s))), \quad \text{for } (\theta, s) \in \Theta \times \mathcal{S}, f \in \mathcal{F}, \quad (8)$$

where $R_{\theta}(z) = \int_0^z \mathcal{R}(f)(\theta, u) du$ is the cumulative distribution function along z when fixing the first argument θ and $R_{\theta}^{-1}(s)$ is the corresponding sliced quantile function supported on $[0, 1]$. The inverse map can be defined for any continuous function ζ as

$$\Psi_{Q, \tau}^{-1}(\zeta)(x) = \mathcal{R}_{\tau}^{-1}(\psi_Q^{-1}(\zeta(\theta, \cdot)))(x). \quad (9)$$

The steps defining the RLQD transformation are visualized in Fig. 1.

For the theoretical analysis of these transformations we require certain smoothness and regularity conditions. We show that when the univariate density transformation ψ satisfies assumptions (L1)-(L3) in the Appendix, the corresponding multivariate density transformation Ψ according to (6) satisfies the four smoothness properties (T1)-(T4) in Appendix A.

4 Representations for Transformed Distributions

4.1 Metric for Multivariate Distributions

Let $\mathcal{B}(D)$ be the Borel σ -algebra on D . We focus on random probability distributions on $(D, \mathcal{B}(D))$ and each probability distribution is characterized by a density function with respect to Lebesgue measure. While classical metrics such as L^1, L^2, L^∞ or total variation metric [14] are commonly used, recent work has emphasized the Fisher-Rao [13, 50] and Wasserstein metrics [18, 32, 26] for samples of distributional data. The 2-Wasserstein distance between two densities is defined as

$$d_W^2(f_1, f_2) = \inf_{\substack{\mu \in \Pi(f_1, f_2), \\ (Z_1, Z_2) \sim \mu}} E(\|Z_1 - Z_2\|^2), \quad \text{for all } f_1, f_2 \in \mathcal{F}.$$

Here Z_1 and Z_2 are random variables on $\mathbb{R}^p$, and $\Pi(f_1, f_2)$ is the space of joint probability measures on $D \times D$ with marginal densities f_1 and f_2 . For one-dimensional distributions, the 2-Wasserstein metric corresponds to the L^2 distance between their quantile functions,

$$d_W^2(g_1, g_2) = \int_0^1 (G_1^{-1}(s) - G_2^{-1}(s))^2 ds, \quad \text{for all } g_1, g_2 \in \mathcal{G},$$

where G_1^{-1}, G_2^{-1} are the corresponding quantile functions of g_1, g_2 . Obtaining the Wasserstein distance for the multivariate distribution involves intricate algorithms and does not have an analytic solution [39, 38]. An alternative approach is the sliced Wasserstein distance [7]. It offers a computationally more efficient solution compared to the conventional Wasserstein distance and is defined as

$$d_{SW}^2(f_1, f_2) = \int_{\theta \in \Theta} d_W^2(\mathcal{R}(f_1)(\theta, \cdot), \mathcal{R}(f_2)(\theta, \cdot)) d\theta, \quad \text{for all } f_1, f_2 \in \mathcal{F}, \quad (10)$$

where the Radon transform $\mathcal{R}$ is introduced in (2).

4.2 Transformed Distribution Representations

Suppose the random densities are generated as $f \sim \mathfrak{F}$ where $\mathfrak{F}$ is a probability measure on $\mathcal{F}$. Applying the transformation in Section 3 yields a random element in the Hilbert space $X := \Psi(f) \in L^2(\Theta \times \mathcal{S})$, which corresponds to the realization of a stochastic process. This enables functional data analysis [47] and linear representations in the Hilbert space L^2 , with results mapped back to the density space via the regularized inverse Ψ_τ^{-1} that was introduced in (9).

Denote the mean function of the process X by $\nu(\theta, s) = E(X(\theta, s))$ and the covariance function by $H(\theta_1, s_1, \theta_2, s_2) = \text{Cov}(X(\theta_1, s_1), X(\theta_2, s_2))$, where the eigenfunctions of the covariance operator associated with the kernel H are denoted by $\{\gamma_k(\theta, s)\}_{k=1}^\infty$ and their corresponding eigenvalues by $\{\lambda_k\}_{k=1}^\infty$. The transformed process can be represented through the Karhunen-Loève expansion [27]

$$X(\theta, s) = \nu(\theta, s) + \sum_{k=1}^\infty \chi_k \gamma_k(\theta, s), \quad (\theta, s) \in \Theta \times \mathcal{S}, \quad (11)$$

with the principal components $\chi_k = \int_{\Theta \times \mathcal{S}} (X(\theta, s) - \nu(\theta, s)) \gamma_k(\theta, s) d\theta ds$. This expansion exhibits an optimality characteristic where the initial K terms constitute the K -dimensional representation of $X(s, t)$ that minimizes the unexplained variance. Empirical estimators of $\hat{\nu}, \hat{H}, \hat{\gamma}_k, \hat{\lambda}_k, \hat{\chi}_k$ can be obtained based on the sample of transformed processes [4, 20, 3, 47]. For extensions to multivariate functional data, including product and marginal FPCA, see [11, 10].

Truncating the Karhunen-Loève expansion to the first K components yields an L^2 -optimal approximation; with judicious choice of K it captures a pre-specified fraction of the variation,

$$X_K(\theta, s) = \nu(\theta, s) + \sum_{k=1}^K \chi_k \gamma_k(\theta, s), \quad (\theta, s) \in \Theta \times \mathcal{S}. \quad (12)$$

To interpret the contribution of each principal component in the original density space, we extend the notion of univariate transformation modes of variation [35] to the multivariate setting. Specifically, we compute modes of variation in the transformed L^2 space and map them back to the density space via the regularized inverse. The k -th transformation mode of variation is defined as a family of functions indexed by $\alpha \in \mathbb{R}$,

$$m_k(x, \alpha, \tau) = \Psi_\tau^{-1} \left(\nu + \alpha \sqrt{\lambda_k} \gamma_k \right) (x). \quad (13)$$

The data application in Section 8 demonstrates that transformation modes of variation play an important role in understanding the effect of each principal component.

5 Estimation

In practice, random densities $\{f_i\}_{i=1}^n$ are often not fully observed and one has only have observations generated by each of these densities. In this more realistic case, a preliminary density estimation step

is needed, which is described in Appendix F. If the random densities f_i are estimated by a kernel density estimator $\hat{f}_i$ through (31), the corresponding sample mean and variance are based on the new transformed sample $\{\tilde{X}_i\}_{i=1}^n = \{\Psi(\hat{f}_i)\}_{i=1}^n$,

$$\hat{\nu}(\theta, s) = \frac{1}{n} \sum_{i=1}^n \tilde{X}_i(\theta, s), \quad \hat{H}(\theta_1, s_1, \theta_2, s_2) = \frac{1}{n} \sum_{i=1}^n \tilde{X}_i(\theta_1, s_1) \tilde{X}_i(\theta_2, s_2) - \hat{\nu}(\theta_1, s_1) \hat{\nu}(\theta_2, s_2). \quad (14)$$

Another application of Mercer's theorem implies

$$\hat{H}(\theta_1, s_1, \theta_2, s_2) = \sum_{k=1}^{\infty} \hat{\lambda}_k \hat{\gamma}_k(\theta_1, s_1) \hat{\gamma}_k(\theta_2, s_2), \quad (15)$$

where $\hat{\lambda}_k$ and $\hat{\gamma}_k$ are estimated eigenvalues and eigenfunctions of the covariance operator associated with the covariance function $\hat{H}$. The principal components χ_{ik} are estimated as

$$\hat{\chi}_{ik} = \int_{\Theta} \int_{\mathcal{S}} (\tilde{X}_i(\theta, s) - \hat{\nu}(\theta, s)) \hat{\gamma}_k(\theta, s) d\theta ds \quad (16)$$

and the plugin estimator for the transformation modes of variation (13) is

$$\hat{m}_k(x, \alpha, \tau) = \Psi_{\tau}^{-1} \left(\hat{\nu} + \alpha \sqrt{\hat{\lambda}_k} \hat{\gamma}_k \right) (x). \quad (17)$$

Selection of tuning parameter K and τ are further discussed in Appendix E.

6 Asymptotic Convergence

In this section we establish asymptotic properties of the empirical estimators proposed in Section 5. A key feature in the convergence rate is the spacing between eigenvalues, where we define

$$\sigma_k = \min_{1 \leq j \leq k} (\lambda_j - \lambda_{j+1}). \quad (18)$$

for any integer k . For the i^{th} distribution, we assume there are N_i independent observations available to estimate the density function $\hat{f}_i$ in (31), with details in the Appendix. The lower bound $N(n) = \min_{1 \leq i \leq n} N_i$ is relevant for the overall asymptotic convergence rate. For the following result, we require that the sample size per distribution is large enough so that the effect of density estimation can be neglected when deriving the asymptotic rates. Specifically, we require

(E1) $\lim_{n \rightarrow \infty} N(n)/n^r \geq c$ for some $r > 1 + p/4$, where $c > 0$ is a constant and p denotes the dimension of the domain of the densities.

Convergence analysis of the representations is presented in Appendix G. For the following, we require an additional assumption (S1) which imposes restrictions on the eigencomponents. It reflects typical assumptions of exponentially or polynomially declining eigenvalues.

(S1) The sequence $K = K(n)$ of approximating K -representations is such that $\sigma_K^{-1} \lambda_K^{-1} n^{-1/2} = O(1)$. With C_4 as in (T5) in Appendix A, we use the following bound to derive the convergence of transformation modes of variation, where for a fixed constant $\alpha_0 > 0$,

$$\begin{aligned} S(\tau, K) &= \max_{1 \leq k \leq K} \sup_{|\alpha| \leq \alpha_0} C_4 \left(\tau, \|X_{k,\alpha}\|_{\infty}, \|\partial X_{k,\alpha}(\theta, s)/\partial s\|_{\infty}, d_{\infty}(X_{k,\alpha}, \hat{X}_{k,\alpha}) \right), \\ &= O \left(\tau^p \max_{1 \leq k \leq K} \left\{ e^{6\|\gamma_k\|_{\infty}} \|\partial \gamma_k(\theta, s)/\partial s\|_{\infty} e^{2d_{\infty}(X_{k,\alpha}, \hat{X}_{k,\alpha})} \right\} \right), \end{aligned} \quad (19)$$

with $X_{k,\alpha} = \nu + \alpha \sqrt{\lambda_k} \gamma_k$ and $\hat{X}_{k,\alpha} = \hat{\nu} + \alpha \sqrt{\hat{\lambda}_k} \hat{\gamma}_k$. Under (S1), we have $d_{\infty}(X_{k,\alpha}, \hat{X}_{k,\alpha}) = O_p(1)$. When K is fixed, $\|X_{k,\alpha}\|_{\infty}$ and $\|\partial X_{k,\alpha}(\theta, s)/\partial s\|_{\infty}$ can be uniformly bounded, implying $S(\tau, K) = O_p(\tau^p)$. When $K = K(n) \rightarrow \infty$, the convergence of $S(\tau, K)$ requires boundedness of $\|X_{k,\alpha}\|_{\infty}$ and $\|\partial X_{k,\alpha}(\theta, s)/\partial s\|_{\infty}$, which is implied by the boundedness of $\|\gamma_k\|_{\infty}$ and $\|\partial \gamma_k(\theta, s)/\partial s\|_{\infty}$. The following result provides the rate of convergence of the estimated transformation modes of variation in (17) to the true modes in (13), uniformly over the parameter $|\alpha| \leq \alpha_0$ and for any constant $\alpha_0 > 0$.

Theorem 1. Under assumptions (D1), (F1), (T1)-(T4), (E1), (K1)-(K2), for fixed K and α_0 , $m_k(x, \alpha, \tau)$ and $\hat{m}_k(x, \alpha, \tau)$ as per (13) and (17),

$$\max_{1 \leq k \leq K} \sup_{|\alpha| \leq \alpha_0} d_{SW}(m_k(x, \alpha, \tau), \hat{m}_k(x, \alpha, \tau)) = O_p(\tau^p n^{-1/2}).$$

If we assume $K = K(n) \rightarrow \infty$ satisfies (S1), then

$$\begin{aligned} & \max_{1 \leq k \leq K(n)} \sup_{|\alpha| \leq \alpha_0} d_{SW}(m_k(x, \alpha, \tau), \hat{m}_k(x, \alpha, \tau)) \\ &= O_p(\tau^p \max_{1 \leq k \leq K} \{e^{6\|\gamma_k\|_\infty} \|\partial \gamma_k(\theta, s)/\partial s\|_\infty\} \sigma_K^{-1} n^{-1/2}), \end{aligned}$$

where τ, γ_k, σ_K are as in equations (4), (11) and (18).

Corollary 1. When assuming exponential decay for the eigenvalues, i.e., $\lambda_k = ce^{-\theta k}$ for some constants $c, \theta > 0$, with $K(n) = \lfloor \frac{1}{4\theta} \log(n) \rfloor \rightarrow \infty$, Theorem 1 provides that

$$\begin{aligned} & \max_{1 \leq k \leq K(n)} \sup_{|\alpha| \leq \alpha_0} d_{SW}(m_k(x, \alpha, \tau), \hat{m}_k(x, \alpha, \tau)) = \\ & O_p(\tau^p \max_{1 \leq k \leq K} \{e^{6\|\gamma_k\|_\infty} \|\partial \gamma_k(\theta, s)/\partial s\|_\infty\} n^{-1/4}). \end{aligned}$$

We also present the convergence rate of the truncated density estimator in 25. See G for additional asymptotic convergence results in the Appendix.

Theorem 2. Under assumptions (D1), (F1), (T1)-(T5), (E1), (K1)-(K2), assume exponentially decaying eigenvalues $\lambda_k = ce^{-\theta k}$ for some constants c and θ . With $K(n) = \lfloor \frac{1}{3\theta} \log(n) \rfloor \rightarrow \infty$,

$$\max_{1 \leq i \leq n} d_{SW}(f_i(\cdot), \hat{f}_i(\cdot, K, \tau)) = O_p(\tau^{-(\kappa-p)} + \tau^p n^{-1/6}),$$

and choosing $\tau \sim n^{1/6k}$ leads to

$$\max_{1 \leq i \leq n} d_{SW}(f_i(\cdot), \hat{f}_i(\cdot, K, \tau)) = O_p(n^{-(\kappa-p)/6k}).$$

7 Experiments

We illustrate our methods by comparing the performance of three approaches: (1) Directly applying FPCA and modes of variation to estimated densities viewed as elements of L^2 , ignoring density constraints [48]. The resulting functions are truncated to their positive part and renormalized to ensure valid densities; (2) Performing PCA in the tangent space (TPCA) of the Bures–Wasserstein manifold of covariance operators at a reference density and maps the resulting components back to the density space via the exponential map [29] and (3) the proposed RLQD method, which takes the constraints into account and where K -component approximations as well as modes of variation are guaranteed to reside in the density space and additional adjustments are not necessary,

We conducted Monte Carlo experiments under two simulation settings: (A) random bivariate normals and (B) mixtures of two bivariate normals. Details of the data generation process appear in Table 1. In both settings, we generate 50 random densities, each from latent parameters, and sample 200 observations per density. As evaluation criterion we use the fraction of variance explained $\hat{V}_K/\hat{V}_\infty$ as defined in Section E, where a higher fraction of variance explained for the same choice of K reflects better representations. Boxplots depicting the fraction of variance explained with respect to the sliced Wasserstein distance for FPCA and RLQD approaches are in Figure 4. It is clear that in these settings the proposed RLQD method performs the best throughout in comparison with both the standard FPCA approach and the tangent FPCA approach. The advantage of RLQD becomes particularly evident beyond the first two modes, where FPCA’s explained variance plateaus and TPCA yields diminishing returns. In contrast, RLQD continues to capture meaningful higher-order variation, demonstrating stronger structural representation and efficiency with fewer retained components.

We further visualize the modes of variation for the first two principal components under Setting A in Figures 2 and 3. Under the Gaussian assumption, both the proposed RLQD transformation and the classic FPCA exhibit similar patterns of variation. Specifically, the first mode primarily captures the overall intensity or concentration of the density around the mean, whereas the second mode reflects a rotational change in the orientation of the joint distribution.

Table 1: Simulation settings for bivariate normal and mixture distributions.

Setting	Random components	Resulting density
A	$\mu_i \sim \mathcal{U}[-0.2, 0.2]$ $\rho_i \sim \mathcal{U}[0.4, 0.8]$ $\log(s_{1i}) \sim \mathcal{U}[-0.3, 0.3]$ $\log(s_{2i}) \sim \mathcal{U}[-0.3, 0.3]$ $\Sigma_i = \begin{pmatrix} s_{1i}^2 & s_{1i}s_{2i}\rho_i \\ s_{1i}s_{2i}\rho_i & s_{2i}^2 \end{pmatrix}$	$\{X_{ij}\}_{j=1}^{200} \sim \mathcal{N}(\mu_i, \Sigma_i)$ truncated on $[-2, 2]^2$ $i = 1, \dots, 50$
B	$\mu_{1,i}, \mu_{2,i} \sim \mathcal{U}[-0.2, 0.2]$ $\rho_{1,i}, \rho_{2,i} \sim \mathcal{U}[0.4, 0.8]$ $\log(s_{1,1i}), \log(s_{2,1i}) \sim \mathcal{U}[-0.3, 0.3]$ $\log(s_{1,2i}), \log(s_{2,2i}) \sim \mathcal{U}[-0.3, 0.3]$ $\Sigma_{1,i} = \begin{pmatrix} s_{1,1i}^2 & s_{1,1i}s_{1,2i}\rho_i \\ s_{1,1i}s_{1,2i}\rho_i & s_{1,2i}^2 \end{pmatrix}$ $\Sigma_{2,i} = \begin{pmatrix} s_{2,1i}^2 & s_{2,1i}s_{2,2i}r_i \\ s_{2,1i}s_{2,2i}r_i & s_{2,2i}^2 \end{pmatrix}$	$\{X_{ij}\}_{j=1}^{200} \sim \frac{1}{2}\mathcal{N}(\mu_{1,i}, \Sigma_{1,i})$ $+ \frac{1}{2}\mathcal{N}(\mu_{2,i}, \Sigma_{2,i})$, truncated on $[-2, 2]^2$ $i = 1, \dots, 50$

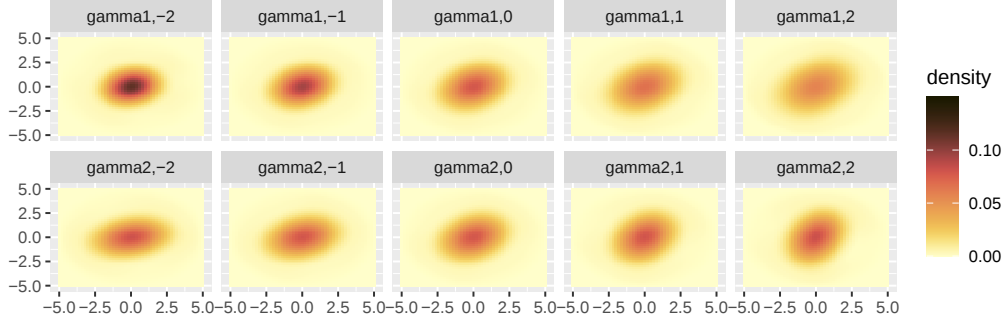


Figure 2: Proposed RLQD transformation modes of variation as per 13 for Setting A

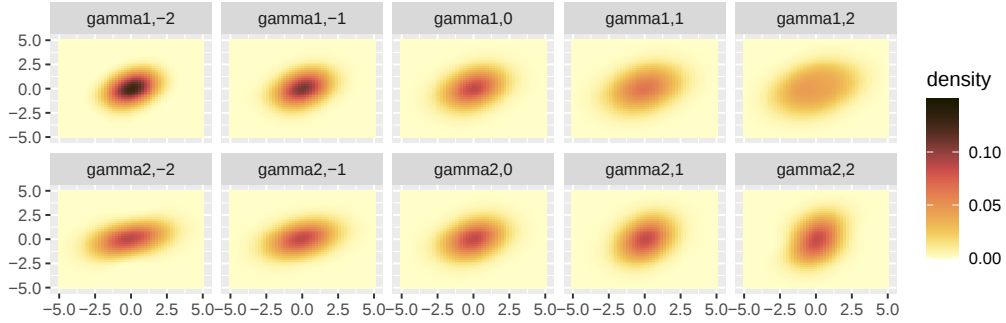


Figure 3: Classic FPCA Modes of variation for Setting A

8 Data Illustrations

We analyze blood pressure data from the Baltimore Longitudinal Study of Aging (BLSA) <https://www.blsa.nih.gov/>. The dataset comprises 16,715 bivariate measurements (SBP, DBP) for approximately 2,800 individuals, with the number of visits per individual ranging from 1 to 26 and ages spanning from 17 to 75. For each age group, we estimate joint two-dimensional densities using a kernel density estimator. The estimator is applied over 51 equidistant grid points in each direction,

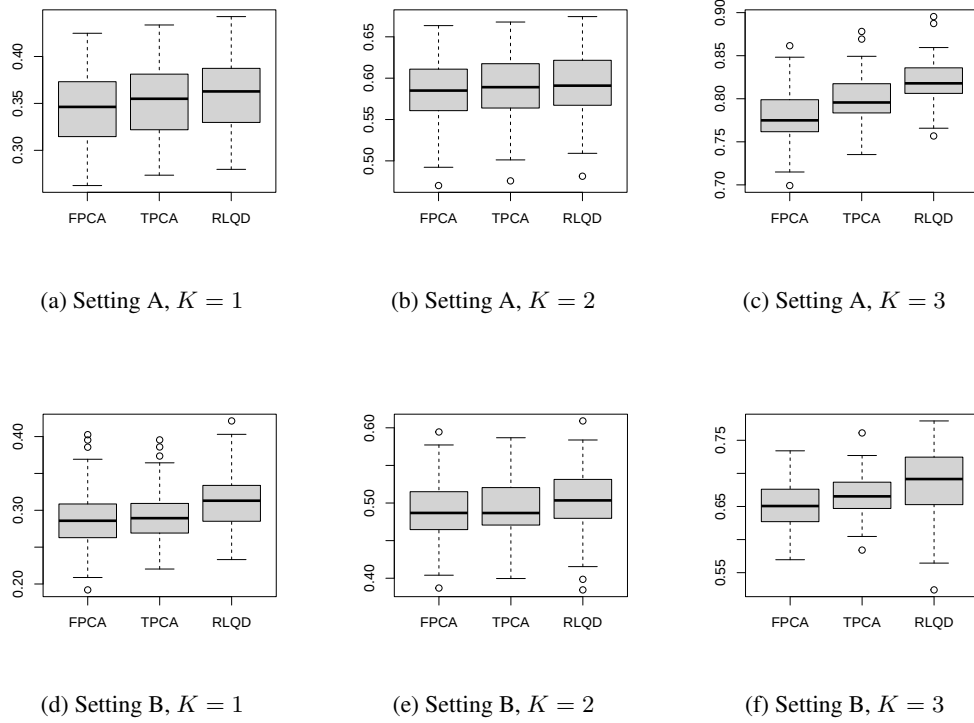


Figure 4: Boxplots show the fraction of variance explained (via sliced Wasserstein distance) across 100 simulations. Each row represents a different setting, with varying component numbers K . Larger values indicate that a higher fraction of variation is explained and are preferred. The proposed RLQD approach is compared with unconstrained FPCA and tangent PCA for densities.

covering the domains $[50, 205]$ for SBP and $[40, 125]$ for DBP with Gaussian kernel and bandwidths $h_{SBP} = 24$ and $h_{DBP} = 15$. The fitted densities (Figure 7 in the Appendix B) show a clear positive correlation between SBP and DBP, along with increasing variability at older ages, highlighting the need for low-dimensional representations that capture these distributional shifts.

Using the RLQD approach with $K = 2$ components, we achieve strong dimension reduction while explaining 87% of the variance in sliced Wasserstein distance as shown in Figure 8 of the Appendix. The corresponding transformation modes of variation in Figure 5 provide insight into principal patterns of change. The predominant mode of variation in the top row reflects a horizontal mean shift coupled with increasing variance, primarily in systolic blood pressure. This suggests that the primary source of variation is a shift in the mean and increased variance of systolic blood pressure. The second mode of variation highlights changes in the variation of diastolic blood pressure and in the covariance structure between diastolic and systolic blood pressures. The first two FPCA components explain 70% and 11%. Under a standard threshold of 85% cumulative variance, RLQD requires only 2 components for effective dimension reduction, whereas FPCA requires at least 3. The eigenfunctions and modes of variation produced by FPCA typically resemble Gaussian-shaped perturbations around the mean as presented in Figure 9 of the Appendix, which can be restrictive when modeling data with inherent skewness or non-Gaussian features. In contrast, RLQD captures more flexible, non-Gaussian modes that better reflect the underlying distributions. For example, the blood pressure data exhibit clear asymmetry, with a noticeable skew toward lower pressure levels. This is statistically confirmed by Mardia’s test [28] for multivariate normality: both skewness (3063.39) and kurtosis (42.75) statistics are highly significant (p -values < 0.001), indicating a strong departure from multivariate normality. These findings demonstrate that RLQD not only provides more informative components with fewer dimensions but also produces representations that are more faithful to the structural properties of complex real-world distributions.

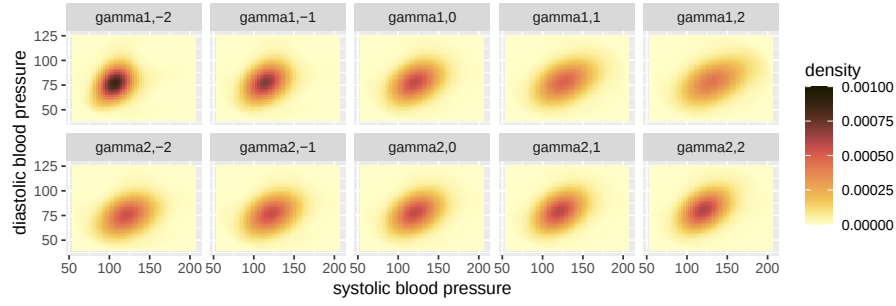


Figure 5: Transformation modes of variation as per (13) for bivariate blood pressure distributions. The first row illustrates the action of the first mode of variation m_1 and the second row that of the second mode of variation m_2 as the values of α are varied from -2 to 2 as per (17).

One can also use the K -component representations—here with $K = 2$, yielding bivariate score vectors—to characterize evolving blood pressure patterns. Local linear smoothing of the scores against age is shown in Figure 6, using a Gaussian kernel with bandwidth 6, selected via leave-one-out cross-validation. The first score, χ_1 , associated with γ_1 , shows an increasing trend with age, suggesting rising blood pressure variation and increased diastolic pressure in older individuals. The second score, χ_2 , linked to γ_2 , increases until midlife then declines, indicating that diastolic pressure peaks in midlife. An additional case study on temperature data in the United States is provided in Appendix H.

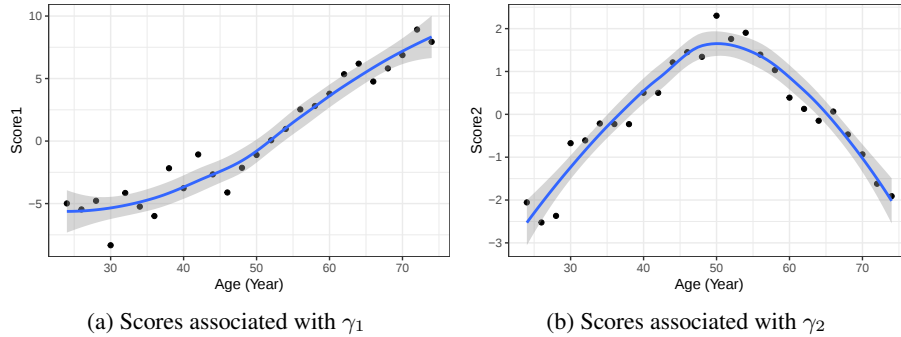


Figure 6: Blood pressure data score smoothing. Local linear smoothing of the score representations along age. Left: Smoothed first scores χ_1 versus age. Right: Smoothed second scores χ_2 against age.

9 Conclusion and Limitations

This work presents a transformation-based framework for the functional analysis of multivariate distributions. By combining Radon slicing with univariate log quantile density transforms, we map multivariate densities into a Hilbert space where classical FDA tools such as FPCA and modes of variation become applicable. The regularized inverse transform ensures that the structure of the original density space is preserved. While not the focus of this paper, we note that this framework facilitates distributional regression once vectorized representations are obtained. Limitations include the dependence on accurate density estimation from finite samples that is subject to the curse of dimensionality and the need for a regularization parameter.

Acknowledgement and Disclosure of Funding

We would like to thank the reviewers for their constructive feedback. This research was partially supported by NSF grant DMS-2310450.

References

- [1] Habti Abeida, Qilin Zhang, Jian Li, and Nadjim Merabtine. Iterative sparse asymptotic minimum variance based approaches for array processing. *IEEE Transactions on Signal Processing*, 61(4):933–944, 2012.
- [2] Luis M Artiles, Richard D Gill, and MI Guț ă. An invitation to quantum tomography. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 67(1):109–134, 2005.
- [3] Juan Lucas Bali, Graciela Boente, David E Tyler, and Jane-Ling Wang. Robust functional principal components: A projection-pursuit approach. *The Annals of Statistics*, 39:2852–2882, 2011.
- [4] Philippe Besse and James O. Ramsay. Principal components analysis of sampled functions. *Psychometrika*, 51(2):285–311, 1986.
- [5] Jérémie Bigot, Raúl Gouet, Thierry Klein, and Alfredo López. Geodesic PCA in the Wasserstein space by convex PCA. *Annales de l’Institut Henri Poincaré B: Probability and Statistics*, 53:1–26, 2017.
- [6] Nicolas Bonneel and Hanspeter Pfister. Sliced Wasserstein barycenter of multiple densities. Technical report, Harvard Technical Report. TR-02-13. <ftp://ftp.deas.harvard.edu/techreports/tr-02-13.pdf>, 2013.
- [7] Nicolas Bonneel, Julien Rabin, Gabriel Peyré, and Hanspeter Pfister. Sliced and Radon Wasserstein barycenters of measures. *Journal of Mathematical Imaging and Vision*, 51(1):22–45, 2015.
- [8] Denis Bosq. *Linear Processes in Function Spaces: Theory and Applications*. Springer-Verlag, New York, 2000.
- [9] Han Chen and Hans-Georg Müller. Sliced Wasserstein regression. *arXiv preprint arXiv:2306.10601*, 2023.
- [10] Han Chen, Hans-Georg Müller, Vasilis G Rodovitis, Nikos T Papadopoulos, and James R Carey. Daily activity profiles over the lifespan of female medflies as biomarkers of aging and longevity. *Aging Cell*, page e14080, 2024.
- [11] Kehui Chen, Pedro Delicado, and Hans-Georg Müller. Modeling function-valued stochastic processes, with applications to fertility dynamics. *Journal of the Royal Statistical Society, Series B (Theory and Methodology)*, 79:177–196, 2017.
- [12] Yaqing Chen, Zhenhua Lin, and Hans-Georg Müller. Wasserstein regression. *Journal of the American Statistical Association*, 118:869–882, 2023.
- [13] Xiongtao Dai. Statistical inference on the Hilbert sphere with application to random densities. *Electronic Journal of Statistics*, 16(1):700–736, 2022.
- [14] L. Devroye and L. Györfi. *Nonparametric Density Estimation: the L1 view*. Wiley, 1985.
- [15] Charles L Epstein. *Introduction to the Mathematics of Medical Imaging*. SIAM, 2007.
- [16] Laya Ghodrati and Victor M Panaretos. Distribution-on-distribution regression via optimal transport maps. *Biometrika*, 109(4):957–974, 2022.
- [17] Laya Ghodrati and Victor M Panaretos. Transportation of measure regression in higher dimensions. *arXiv preprint arXiv:2305.17503*, 2023.
- [18] Alison L Gibbs and Francis Edward Su. On choosing and bounding probability metrics. *International Statistical Review*, 70(3):419–435, 2002.
- [19] Dominique Guégan and Matteo Iacopini. Nonparametric forecasting of multivariate probability density functions. *arXiv preprint arXiv:1803.06823*, 2018.
- [20] Peter Hall and Mohammad Hosseini-Nasab. On properties of functional principal components analysis. *Journal of the Royal Statistical Society: Series B*, 68(1):109–126, 2006.

- [21] Sigurdur Helgason and S Helgason. *The Radon Transform*, volume 2. Springer, 1999.
- [22] Gabor T Herman. *Fundamentals of Computerized Tomography: Image Reconstruction from Projections*. Springer Science & Business Media, 2009.
- [23] K Hron, A Menafoglio, M Templ, K Hruzova, and P Filzmoser. Simplicial principal component analysis for density functions in Bayes spaces. *Computational Statistics and Data Analysis*, 94:330–350, 2016.
- [24] Karel Hron, Jitka Machalová, and Alessandra Menafoglio. Bivariate densities in Bayes spaces: Orthogonal decomposition and spline representation. *Statistical Papers*, 64:1629–1667, 2023.
- [25] Avinash C Kak and Malcolm Slaney. *Principles of Computerized Tomographic Imaging*. SIAM, 2001.
- [26] Leonid V Kantorovich. On the translocation of masses. *Journal of Mathematical Sciences*, 133(4):1381–1382, 2006.
- [27] Kari Karhunen. Zur Spektraltheorie stochastischer Prozesse. *Ann. Acad. Sci. Fennicae, Ser. A*, 34, 1946.
- [28] Kanti V Mardia. 9 tests of univariate and multivariate normality. *Handbook of Statistics*, 1:279–320, 1980.
- [29] Valentina Masarotto, Guido Masarotto, Maintainer Valentina Masarotto, and Panaretos Masarotto. Package ‘fdwasserstein’. *R package version*, 2024.
- [30] Alessandra Menafoglio, Marco Grasso, Piercesare Secchi, and Bianca Maria Colosimo. Profile monitoring of probability density functions via simplicial functional pca with application to image data. *Technometrics*, 60(4):497–510, 2018.
- [31] Russell Manning Mersereau and Alan V Oppenheim. Digital reconstruction of multidimensional signals from their projections. *Proceedings of the IEEE*, 62(10):1319–1338, 1974.
- [32] Gaspard Monge. Mémoire sur la théorie des déblais et des remblais. *Histoire de l’Académie Royale des Sciences de Paris*, 1781.
- [33] Hans-Georg Müller and Fang Yao. Functional additive models. *Journal of the American Statistical Association*, 103(484):1534–1544, 2008.
- [34] Ryo Okano and Masaaki Imaizumi. Distribution-on-distribution regression with Wasserstein metric: Multivariate Gaussian case. *arXiv preprint arXiv:2307.06137*, 2023.
- [35] Alexander Petersen and Hans-Georg Müller. Functional data analysis for density functions by transformation to a Hilbert space. *The Annals of Statistics*, 44(1):183–218, 2016.
- [36] Alexander Petersen and Hans-Georg Müller. Fréchet regression for random objects with Euclidean predictors. *The Annals of Statistics*, 47:691–719, 2019.
- [37] Alexander Petersen, Chao Zhang, and Piotr Kokoszka. Modeling probability density functions as data objects. *Econometrics and Statistics*, 21:159–178, 2022.
- [38] Gabriel Peyré and Marco Cuturi. Computational optimal transport: With applications to data science. *Foundations and Trends in Machine Learning*, 11(5-6):355–607, 2019.
- [39] Julien Rabin, Gabriel Peyré, Julie Delon, and Marc Bernot. Wasserstein barycenter and its application to texture mixing. In *Scale Space and Variational Methods in Computer Vision: Third International Conference, SSVM 2011, Ein-Gedi, Israel, May 29–June 2, 2011, Revised Selected Papers 3*, pages 435–446. Springer, 2012.
- [40] J. Radon. Über die Bestimmung von Funktionen durch ihre Integralwerte längs gewisser Mannigfaltigkeiten. *Akad. Wiss.*, 69:262–277, 1917.
- [41] GR Ramesh, N Srinivasa, and K Rajgopal. An algorithm for computing the discrete radon transform with some applications. In *Fourth IEEE Region 10 International Conference TENCON*, pages 78–81. IEEE, 1989.

- [42] Alessandro Rinaldo and Larry Wasserman. Generalized density clustering. *The Annals of Statistics*, 38(5):2678–2722, 2010.
- [43] Isaac J Schoenberg. On certain metric spaces arising from Euclidean spaces by a change of metric and their imbedding in Hilbert space. *Annals of Mathematics*, pages 787–793, 1937.
- [44] Isaac J Schoenberg. Metric spaces and positive definite functions. *Transactions of the American Mathematical Society*, 44(3):522–536, 1938.
- [45] Nina Vesseron, Elsa Cazelles, Alice Le Brigant, and Thierry Klein. On the Wasserstein geodesic principal component analysis of probability measures. *arXiv preprint arXiv:2506.04480*, 2025.
- [46] Martin J Wainwright. *High-dimensional Statistics: A Non-asymptotic Viewpoint*, volume 48. Cambridge University Press, 2019.
- [47] Jane-Ling Wang, Jeng-Min Chiou, and Hans-Georg Müller. Functional data analysis. *Annual Review of Statistics and Its Application*, 3:257–295, 2016.
- [48] Yidong Zhou, Han Chen, Su I Iao, Poorbita Kundu, Hang Zhou, Satarupa Bhattacharjee, Cody Carroll, Yaqing Chen, Xiongtao Dai, Jianing Fan, Alvaro Gajardo, Pantelis Z Hadji-pantelis, Kyunghee Han, Hao Ji, Changbo Zhu, Paromita Dubey, Shu-Chin Lin, Hans-Georg Müller, and Jane-Ling Wang. fdapace: Functional data analysis and empirical dynamics. URL <https://CRAN.R-project.org/package=fdapace>. *R package version 0.6.0*, 2024.
- [49] Changbo Zhu and Hans-Georg Müller. Autoregressive optimal transport models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 85(3):1012–1033, 2023.
- [50] Changbo Zhu and Hans-Georg Müller. Spherical autoregressive models, with application to distributional and compositional time series. *Journal of Econometrics*, 2023.

NeurIPS Paper Checklist

1. Claims

Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

Answer: [\[Yes\]](#)

Justification: The abstract and introduction accurately summarize the paper's key contributions and are consistent with the scope and content presented in the main text.

Guidelines:

- The answer NA means that the abstract and introduction do not include the claims made in the paper.
- The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
- The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
- It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. Limitations

Question: Does the paper discuss the limitations of the work performed by the authors?

Answer: [\[Yes\]](#)

Justification: Please refer to Section 9 for limitations

Guidelines:

- The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
- The authors are encouraged to create a separate "Limitations" section in their paper.
- The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
- The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
- The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
- The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
- If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. Theory assumptions and proofs

Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [\[Yes\]](#)

Justification: We present three main theorems: one in Section 6 of the main paper and two in Appendix G. All are accompanied by detailed proofs in Section J.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. Experimental result reproducibility

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [\[Yes\]](#)

Justification: Please refer to Section 5 of the main paper and Section E of the Appendix for the choice of tuning parameters.

Guidelines:

- The answer NA means that the paper does not include experiments.
- If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general, releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
 - (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
 - (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
 - (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
 - (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. Open access to data and code

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [Yes]

Justification: Code is provided along with the submission.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (<https://nips.cc/public/guides/CodeSubmissionPolicy>) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so “No” is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (<https://nips.cc/public/guides/CodeSubmissionPolicy>) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. Experimental setting/details

Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

Answer: [Yes]

Justification: Please refer to Section 5 of the main paper and Section E of the Supplement for the estimation details. Code is also provided along with the submission.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

7. Experiment statistical significance

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [Yes]

Justification: The error bars is provided in Section 7 of the main paper.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).

- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. Experiments compute resources

Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

Answer: [Yes]

Justification: Please refer to Section I we use and execution time.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
- The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
- The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. Code of ethics

Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics <https://neurips.cc/public/EthicsGuidelines>?

Answer: [Yes]

Justification: The research conducted in the paper conforms with the NeurIPS Code of Ethics.

Guidelines:

- The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
- If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. Broader impacts

Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: [NA]

Justification: The paper presents foundational research that is not directly linked to specific applications, and therefore a discussion of societal impacts is not applicable.

Guidelines:

- The answer NA means that there is no societal impact of the work performed.
- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.

- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. Safeguards

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

Answer: [NA]

Justification: The paper does not involve the release of models or datasets with a high risk of misuse. It focuses on theoretical and methodological contributions rather than deployable systems.

Guidelines:

- The answer NA means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. Licenses for existing assets

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [Yes]

Justification: All external assets used in the paper, including datasets and code, are properly cited with references to the original sources. The licenses and terms of use for these assets are explicitly stated, and their usage complies with the specified terms. Version information and URLs are included where applicable.

Guidelines:

- The answer NA means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.

- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, paperswithcode.com/datasets has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [NA]

Justification: The paper does not release new assets.

Guidelines:

- The answer NA means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [NA]

Justification: This paper does not involve crowdsourcing nor research with human subjects.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [NA]

Justification: The paper does not involve crowdsourcing nor research with human subjects.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.

- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

Answer: [NA]

Justification: The core methods and contributions of this research do not involve the use of large language models (LLMs) in any important, original, or non-standard way.

Guidelines:

- The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy (<https://neurips.cc/Conferences/2025/LLM>) for what should or should not be described.

A Additional Assumptions

The univariate transformation ψ is required to satisfy the following assumptions (L1)-(L3), where we denote the L^2 and uniform metrics as d_2 and d_∞ , respectively.

(L1) Let $g_1, g_2 \in \mathcal{G}$ with g_1 differentiable and $\|g'_1\|_\infty < \infty$. Set

$$\tilde{D}_0 \geq \max(\|g_1\|_\infty, \|1/g_1\|_\infty, \|g_2\|_\infty, \|1/g_2\|_\infty, \|g'_1\|_\infty).$$

There exists a const $\tilde{C}_0$ depending only on $\tilde{D}_0$ such that

$$d_2(\psi(g_1), \psi(g_2)) \leq \tilde{C}_0 d_2(g_1, g_2), \quad d_\infty(\psi(g_1), \psi(g_2)) \leq \tilde{C}_0 d_\infty(g_1, g_2).$$

(L2) Assume $g \in \mathcal{G}$ is differentiable with $\|g'\|_\infty < \infty$ and $\tilde{D}_1$ is a constant bounded below by $\max(\|g\|_\infty, \|1/g\|_\infty, \|g'\|_\infty)$. Then $\psi(g)$ is differentiable and there exists a constant $\tilde{C}_1 > 0$ depending only on $\tilde{D}_1$ such that

$$\|\psi(g)\|_\infty \leq \tilde{C}_1, \quad \|\psi(g)'\|_\infty \leq \tilde{C}_1.$$

(L3) Assume φ_1, φ_2 are continuous functions, where φ_1 is differentiable on $\mathcal{S}$ with $\|\varphi'_1\|_\infty < \infty$. There exist constants $c, \tilde{C}_2, \tilde{C}_3 > 0$ where $\tilde{C}_2(\|\varphi_1\|_\infty, \|\varphi'_1\|_\infty) \leq ce^{6\|\varphi_1\|_\infty} \|\partial\varphi_1(\theta, s)/\partial s\|_\infty$ and $\tilde{C}_3(d_\infty(\varphi_1, \varphi_2)) \leq ce^{2d_\infty(\varphi_1, \varphi_2)}$, such that

$$d_2(\psi^{-1}(\varphi_1), \psi^{-1}(\varphi_2)) \leq \tilde{C}_2 \tilde{C}_3 d_2(\varphi_1, \varphi_2),$$

where $\tilde{C}_2$ and $\tilde{C}_3$ are increasing in their respective arguments.

The multivariate transformation Ψ is required to satisfy the following assumptions (T1)-(T5).

(T1) There exists a constant C_1 such that

$$d_2(\Psi(f_1), \Psi(f_2)) \leq C_1 d_2(f_1, f_2), \quad d_\infty(\Psi(f_1), \Psi(f_2)) \leq C_1 d_\infty(f_1, f_2), \quad \text{for all } f_1, f_2 \in \mathcal{F}.$$

(T2) For all $f \in \mathcal{F}$, $\Psi(f)(\theta, s)$ is continuous and has a continuous partial derivative with respect to s with a constant C_2 such that

$$\|\Psi(f)\|_\infty \leq C_2, \quad \|\partial\Psi(f)(\theta, s)/\partial s\|_\infty \leq C_2, \quad |\Psi(f)(\theta_1, s) - \Psi(f)(\theta_2, s)| \leq C_2 \|\theta_1 - \theta_2\|_2.$$

(T3) An inverse transformation Ψ^{-1} exists such that $\Psi^{-1} \circ \Psi(f) = f$ for all $f \in \mathcal{F}$.

(T4) There exists a sequence of approximating inverses $\Psi_\tau^{-1}, \tau \rightarrow \infty$, and a constant $C_3 = C_3(\tau) \leq c\tau^{-(\kappa-p)}$ for a constant c , such that

$$d_\infty(\Psi_\tau^{-1} \circ \Psi(f), f) \leq C_3, \quad \text{for all } f \in \mathcal{F},$$

where C_3 is decreasing to 0 as $\tau \rightarrow \infty$.

(T5) Let $\zeta_1, \zeta_2 : \Theta \times \mathcal{S} \rightarrow \mathbb{R}$ be continuous functions, where ζ_1 has a continuous partial derivative $\partial\zeta_1(\theta, s)/\partial s$. There exist constants $c, C_4 > 0$ with

$$C_4 = C_4(\tau, \|\zeta_1\|_\infty, \|\partial\zeta_1(\theta, s)/\partial s\|_\infty, d_\infty(\zeta_1, \zeta_2)) \leq c\tau^p \|\partial\zeta_1(\theta, s)/\partial s\|_\infty e^{6\|\zeta_1\|_\infty} e^{2d_\infty(\zeta_1, \zeta_2)},$$

such that

$$d_\infty(\Psi_\tau^{-1}(\zeta_1), \Psi_\tau^{-1}(\zeta_2)) \leq C_4 d_2(\zeta_1, \zeta_2),$$

where C_4 is increasing in each of its four arguments.

Assumptions (T1) and (T2) ensure the continuity and the boundedness of the forward transformation, while assumption (T3) guarantees invertibility on the image space $\Psi(\mathcal{F})$. Note that $\Psi(\mathcal{F})$ is not guaranteed to cover the entire C^0 space and the inverse transformation Ψ^{-1} is only defined on the image space $\Psi(\mathcal{F})$. A sequence of approximating inverses Ψ_τ^{-1} is provided in assumption (T4) and (T5) and is used to map L^2 elements back to $\mathcal{F}$. Note that the regularization parameter τ controls the approximation accuracy and the roles of ζ_1 and ζ_2 in assumption (T5) could be reversed.

B Additional Figures

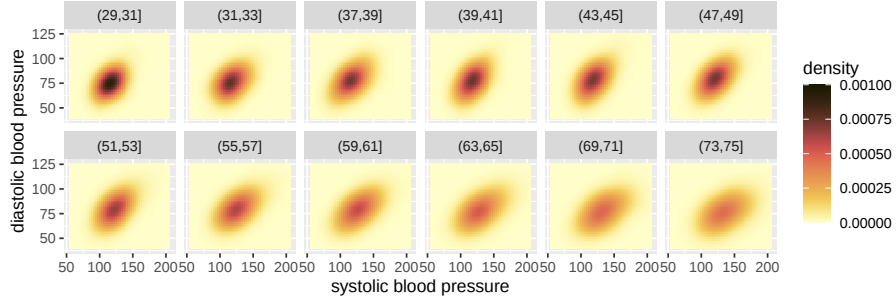


Figure 7: Fitted kernel density surfaces for bivariate blood pressure data for selected age groups.

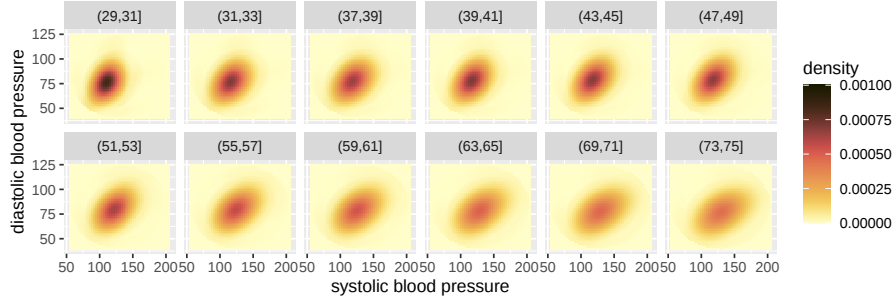


Figure 8: Reconstructed two-dimensional blood pressure distributions using the proposed RLQD approach, with sliced Wasserstein variance explained of 0.87—0.73 from the first component and 0.14 from the second. The two-dimensional representations capture most of the variation in the original density space and enable substantial dimension reduction.

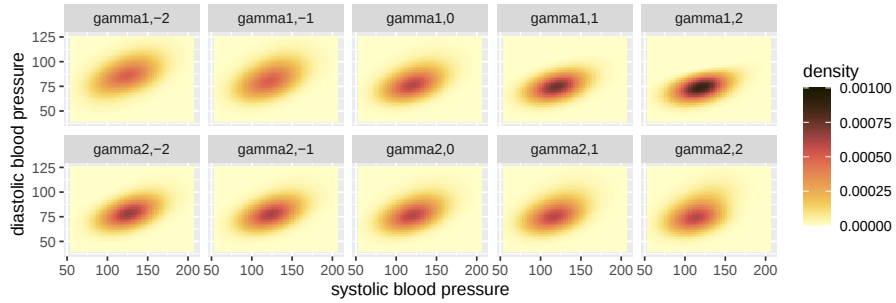


Figure 9: Modes of variation of the first two components by using standard FPCA approach

C Reconstruction of the L2 process

Proposition 1. Assume (D1), (F1) and (T2)-(T5). Let $f \in \mathcal{F}$ and ζ be a continuous function on $L^2(\Theta \times \mathcal{S})$. Then

$$d_{SW}(f, \Psi_\tau^{-1} \circ \zeta) = O(C_3 + C_4 d_2(\Psi(f), \zeta)), \quad \text{as } \tau \rightarrow \infty.$$

Here, C_3 , C_4 , and Ψ are defined as per assumptions (T4)-(T5) and in equation (6).

This bound indicates that by approximating the transformed density within the transformed L^2 space, one can obtain a consistent reconstruction of the multivariate density through the regularized inverse map Ψ_τ^{-1} introduced in (9), where the rate of convergence can be further specified as in the Proposition.

D Radon Log Hazard and Log Quantile Density Transformations

We review here an additional specific univariate density transformation, the log hazard transformation [35]. We adopt the assumption that has been made in Section 2, namely that the support for a univariate density function is $\mathcal{Z} = [0, 1]$. As hazard functions diverge at the right endpoint 1 of the support of the distribution, we consider quotient spaces by equating densities that are equal on a subdomain $\mathcal{S} = [0, 1_v]$, where $1_v = 1 - v$ for some $0 < v < 1$. The log hazard transformation $\psi_H : \mathcal{G} \rightarrow L^2(\mathcal{S})$ is expressed as follows:

$$\psi_H(g)(s) = \log \left\{ \frac{g(s)}{1 - G(s)} \right\}, \quad s \in \mathcal{S}, \quad (20)$$

where $G(s) = \int_0^s g(z)dz$ is the cumulative distribution function. The inverse of a continuous function φ on $\mathcal{S}$ is

$$\psi_H^{-1}(\varphi)(z) = \begin{cases} \exp \left\{ \varphi(z) - \int_0^z e^{\varphi(u)} du \right\} & \text{for } z \in [0, 1_v], \\ v^{-1} \exp \left\{ - \int_0^{1_v} e^{\varphi(u)} du \right\} & \text{for } z \in [1_v, 1]. \end{cases}$$

Employing the Radon transform $\mathcal{R}$ as defined in (2), the Radon log hazard transformation (RLH) Ψ_H is

$$\Psi_H(f)(\theta, s) = \psi_H(\mathcal{R}(f)(\theta, \cdot))(s) = \log \left\{ \frac{\mathcal{R}(f)(\theta, s)}{1 - R_\theta(s)} \right\}, \quad (21)$$

where $R_\theta(s) = \int_{-\infty}^s \mathcal{R}(f)(\theta, u)du$ is the cumulative distribution function along $\mathcal{R}(f)(\theta, \cdot)$, fixing the first argument at θ .

Corollary 2. For C_3, C_4 as in assumption (T4)-(T5), when using the RLH transform Ψ_H as per (21), it holds that $C_3(\tau) \sim \tau^{-(\kappa-p)}$ and $C_4(\tau, \|\zeta_1\|_\infty, \|\partial\zeta_1(\theta, s)/\partial s\|_\infty, d_\infty(\zeta_1, \zeta_2)) \sim \tau^p e^{2\|\zeta_1\|_\infty} e^{d_\infty(\zeta_1, \zeta_2)}$ as $\tau \rightarrow \infty$.

Similarly, for Radon log quantile density (RLQD) transformation, we have

Proposition 2. If (D1), (F1) and (L1)-(L3) hold, assumptions (T1)-(T5) are satisfied for the multivariate density transformation Ψ in (6). When using the RLQD transformation Ψ_Q as per (8), we have $C_3(\tau) \sim \tau^{-(\kappa-p)}$ and $C_4(\tau, \|\zeta_1\|_\infty, \|\partial\zeta_1(\theta, s)/\partial s\|_\infty, d_\infty(\zeta_1, \zeta_2)) \sim \tau^p \|\partial\zeta_1(\theta, s)/\partial s\|_\infty e^{6\|\zeta_1\|_\infty} e^{2d_\infty(\zeta_1, \zeta_2)}$.

The following corollary follows directly from Corollary 2 and Theorem 1.

Corollary 3. When assuming exponentially decaying eigenvalues, $\lambda_k = ce^{-\theta k}$ for some constants $c, \theta > 0$, with $K(n) = \lfloor \frac{1}{4\theta} \log(n) \rfloor \rightarrow \infty$, it holds that

$$\max_{1 \leq k \leq K(n)} \sup_{|\alpha| \leq \alpha_0} d_{SW}(m_k(x, \alpha, \tau), \hat{m}_k(x, \alpha, \tau)) = O_p \left(\tau^p \max_{1 \leq k \leq K} \left\{ e^{2\|\gamma_k\|_\infty} \right\} n^{-1/4} \right).$$

Furthermore, if the eigenfunctions correspond to the trigonometric basis, then

$$\max_{1 \leq k \leq K(n)} \sup_{|\alpha| \leq \alpha_0} d_{SW}(m_k(x, \alpha, \tau), \hat{m}_k(x, \alpha, \tau)) = O_p \left(\tau^p n^{-1/4} \right).$$

E Estimation and Selection of Tuning Parameter

We discuss the empirical estimation in Section 5. We now present the population version to provide a complete formulation and to facilitate the proofs.

Suppose we have a sample of independent random densities $\{f_i\}_{i=1}^n \sim \mathfrak{F}$. Writing $X_i = \Psi(f_i)$, empirical sample mean and variance are

$$\tilde{\nu}(\theta, s) = \frac{1}{n} \sum_{i=1}^n X_i(\theta, s), \quad \tilde{H}(\theta_1, s_1, \theta_2, s_2) = \frac{1}{n} \sum_{i=1}^n X_i(\theta_1, s_1) X_i(\theta_2, s_2) - \tilde{\nu}(\theta_1, s_1) \tilde{\nu}(\theta_2, s_2). \quad (22)$$

Mercer's theorem implies that the spectral decomposition of $\tilde{H}$ leads to

$$\tilde{H}(\theta_1, s_1, \theta_2, s_2) = \sum_{k=1}^{\infty} \tilde{\lambda}_k \tilde{\gamma}_k(\theta_1, s_1) \tilde{\gamma}_k(\theta_2, s_2), \quad (23)$$

where $\tilde{\lambda}_k$ and $\tilde{\gamma}_k$ are empirical eigenvalues and eigenfunctions of the covariance operator with the covariance kernel $\tilde{H}$.

In terms of the selection of tuning parameter, By selecting the first K components in the expansion (11) and applying the regularized inverse transformation, the truncated representations can be obtained through

$$f_i(x, K, \tau) = \Psi_{\tau}^{-1} \left(\nu + \sum_{k=1}^K \chi_{ik} \gamma_k \right) (x), \quad (24)$$

with estimated version

$$\hat{f}_i(x, K, \tau) = \Psi_{\tau}^{-1} \left(\hat{\nu} + \sum_{k=1}^K \hat{\chi}_{ik} \hat{\gamma}_k \right) (x). \quad (25)$$

A simple and often well-working approach for the selection of the tuning parameter K can be based on the fraction of variance explained in the density space. To implement this method, we need to first define the sliced Wasserstein Fréchet mean with respect to the metric d_{SW} , which is

$$f_{\oplus} = \operatorname{arginf}_{q \in \mathcal{F}} E(d_{SW}(f, q)^2), \quad (26)$$

with empirical estimates $\hat{f}_{\oplus} = \operatorname{argmin}_{q \in \mathcal{F}} n^{-1} \sum_{i=1}^n d_{SW}^2(\hat{f}_i, q)$. We note that $f_{\oplus}$ and $\hat{f}_{\oplus}$ might be sets with more than one element. Further results regarding the sliced Wasserstein distance and sliced Wasserstein barycenter can be found in [38, 6, 7]. To quantify the variance explained in the density space, we use the Fréchet variance

$$V_{\infty} = E(d_{SW}(f, f_{\oplus})^2), \quad (27)$$

which is unique irrespective of whether the Fréchet mean is unique or not. Its empirical estimate is $\hat{V}_{\infty} = \frac{1}{n} \sum_{i=1}^n d_{SW}(\hat{f}_i, \hat{f}_{\oplus})^2$. The variance explained by the K component representation is

$$V_K = V_{\infty} - E(d_{SW}(f, f_{K, \tau})^2), \quad (28)$$

where $f_{K, \tau} = \Psi_{\tau}^{-1}(\nu + \sum_{k=1}^K \chi_k \gamma_k)$ is the truncated representation of the density process. This variance explained is estimated by $\hat{V}_K = \hat{V}_{\infty} - \frac{1}{n} \sum_{i=1}^n d_{SW}(\hat{f}_i(\cdot), \hat{f}_i(\cdot, K, \tau))^2$. Control of the tuning parameter τ is implemented by adjusting the sampling rate in the density domain, which determines the Nyquist frequency—the highest frequency representable without aliasing. This Nyquist frequency directly sets the cutoff for the regularized inverse Radon transform, thus controlling the level of smoothing in the reconstructed density [25, 41].

The fraction of variance explained is the ratio V_K/V_{∞} , with estimated ratio $\hat{V}_K/\hat{V}_{\infty}$. The choice of K is then the smallest K that explains a fraction of variance above a given threshold ϑ with $0 < \vartheta < 1$, formally

$$K^* = \operatorname{argmin}_K \left\{ K : \frac{V_K}{V_{\infty}} > \vartheta \right\}, \quad (29)$$

with plug-in estimator

$$\hat{K}^* = \operatorname{argmin}_K \left\{ K : \frac{\hat{V}_K}{\hat{V}_{\infty}} > \vartheta \right\}. \quad (30)$$

Here the default choices of ϑ are $\vartheta = .95$ or $\vartheta = .9$.

F Random Distribution Estimation

We assume that the multivariate distributions we are dealing with are well-defined and have density functions on fixed domains, which might be unknown but encompass a fixed known domain D . The domain D is as in assumption (D1) and is determined either by prior knowledge or selected by the user in practical scenarios. Our focus is on distributions truncated to the common domain D and we presume that these truncated distributions are well-defined with density functions on D . Similar to the preliminary density estimation step as in Section S.3 of [9], we assume that there exists $\epsilon > 0$ such that $D_\epsilon = \bigcup_{z \in D} B(z, \epsilon)$ belongs to the domain of all underlying multivariate distributions, where $B(z, \epsilon)$ is a ball with radius ϵ centered at z . This assumption guarantees that boundary effects can be ignored in the density estimation step. The continuous differentiability assumption (F1) is extended to (F1') below. Additional assumptions (P1), (K1), (K2) are also listed below.

(F1') There exists a constant $M_0 > 0$ and an integer $\kappa \geq p + 1$ such that for all $f \in \mathcal{F}$, the density function f is the density of a truncated distribution on the domain D , where the original distribution has a domain D_f with $D_f \supset D_\epsilon$ and a density f_{D_f} that is defined on D_f . It holds that $\max\{\|f_{D_f}\|_\infty, \|1/f_{D_f}\|_\infty\} \leq M_0$ on D_ϵ and that f_{D_f} is continuously differentiable of order k on D_ϵ , with uniformly bounded partial derivatives. The target distributions with densities f are the truncated versions of the f_{D_f} on D , so that for all $z \in D$ one has $f(z) = f_{D_f}(z)/\int_D f_{D_f}(u)du$.

(P1) $N(n) = \min_{1 \leq i \leq n} N_i \geq cn^r$, where N_i is the number of random observations for the i -th distribution, and the constants satisfy $r > 1 + p/4$ and $c > 0$.

(K1) The kernel function $\mathcal{K}$ as per (31) is a probability density function that has compact support and is a symmetric, bounded and k times continuously differentiable function (without loss of generality, the support is assumed to be contained in the unit cube of $\mathbb{R}^p$).

(K2) For some $A > 0$ and $\omega > 0$, the class of functions $\mathcal{I}_b = \{\mathcal{K}(\frac{z-\cdot}{b}), z \in \mathbb{R}^p, b > 0\}$ satisfies

$$\sup_{\mathcal{P}} M(\mathcal{I}_b, L_2(\mathcal{P}), \varepsilon \|F\|_{L_2(\mathcal{P})}) \leq \left(\frac{A}{\varepsilon}\right)^\omega,$$

where $M(T, d, \varepsilon)$ denotes the ε -covering number of the metric space (T, d) , F is the envelope function of $\mathcal{I}_b$ and the supremum is taken over the set of all probability measures on $\mathbb{R}^p$. The quantities A and ω correspond to the Vapnik-Cervonenkis (VC) characteristic of $\mathcal{K}$.

Note that assumption (P1) ensures that the L^2 convergence of the multivariate density is faster than the parametric rate $n^{-1/2}$ while assumptions (K1)-(K2) are standard to derive the L^2 and uniform convergence of the multivariate density estimator. Let μ_{D_f} be a random probability distribution with density function f_{D_f} that satisfies (F1'), from which random observations $W_1, \dots, W_N$ are independently sampled. A standard kernel estimator $\check{f}$ is then applied to estimate f_{D_f} , and its truncated version $\hat{f}$ is employed to estimate the density f , which is the distribution μ_{D_f} truncated to the domain D .

$$\check{f}(x) = \frac{1}{Nh^p} \sum_{j=1}^N \mathcal{K}\left(\frac{x - W_j}{h}\right), \quad \hat{f}(x) = \check{f}(x) / \int_D \check{f}(u)du, \quad x \in D \subset \mathbb{R}^p. \quad (31)$$

The following proposition is a direct consequence of Propositions S2 and S3 in [9].

Proposition 3. Assume (D1), (F1') and (K1). Choosing $h \sim N^{-\frac{1}{p+4}}$, the kernel density estimator $\hat{f}$ in (31) satisfies

$$\sup_{f \in \mathcal{F}} E \left[d_2(\hat{f}, f)^2 \middle| f \right] = O\left(N^{-4/(p+4)}\right). \quad (32)$$

If we assume further (K2), for any $\delta > 0$,

$$\sup_{f \in \mathcal{F}} P\left(d_\infty(\hat{f}, f) > N^{-\frac{2}{p+4} + \delta} \middle| f\right) = o(1) \quad \text{as } N \rightarrow \infty. \quad (33)$$

Proposition 3 indicates the convergence rate of the proposed density estimator based on the observation sample size N from the random distribution. The following proposition provides an enhanced form of uniform convergence for the multivariate kernel density estimator under assumption (P1), which is essential for establishing uniform convergence of the representation in the L^2 space.

Proposition 4. Assume (D1), (F1'), (K1), (K2) and (P1). For any $R > 0, \delta > 0$,

$$n \sup_{f \in \mathcal{F}} P \left(d_\infty \left(\hat{f}, f \right) > RN^{-\frac{2}{p+4} + \delta} \right) = o(1) \quad \text{as } n \rightarrow \infty. \quad (34)$$

G Additional Asymptotic Convergence Analysis

Convergence analysis of the representations of the process is presented as follows.

Theorem 3. Under assumptions (F1), (T1)-(T2), (E1) and (K1)-(K2), and with $\hat{\nu}$, $\hat{H}$, $\hat{\lambda}_k$, $\hat{\gamma}_k$ as per (14) and (15), for the k -th component of the eigen-expansion,

$$\begin{aligned} d_2(\nu, \hat{\nu}) &= O_p \left(n^{-1/2} \right), \quad d_\infty(\nu, \hat{\nu}) = O_p \left(\left(\frac{\log(n)}{n} \right)^{1/2} \right), \\ d_2(H, \hat{H}) &= O_p \left(n^{-1/2} \right), \quad d_\infty(H, \hat{H}) = O_p \left(\left(\frac{\log(n)}{n} \right)^{1/2} \right), \\ |\lambda_k - \hat{\lambda}_k| &= O_p \left(n^{-1/2} \right), \\ d_2(\gamma_k, \hat{\gamma}_k) &= \sigma_k^{-1} O_p \left(n^{-1/2} \right), \\ d_\infty(\gamma_k, \hat{\gamma}_k) &= O_p \left(\lambda_k^{-1} \sigma_k^{-1} n^{-1/2} \right). \end{aligned}$$

Furthermore, if the eigenfunctions correspond to the trigonometric basis, then $\|\gamma_k\|_\infty = O(1)$ and $\|\partial \gamma_k(\theta, s)/\partial s\|_\infty = O(k)$, so that,

$$\max_{1 \leq k \leq K(n)} \sup_{|\alpha| \leq \alpha_0} d_{SW} (m_k(x, \alpha, \tau), \hat{m}_k(x, \alpha, \tau)) = O_p(\tau^p \log(n) n^{-1/4}).$$

The following additional assumptions impose additional restrictions on the eigencomponents. They are satisfied for eigenvalue sequences with exponential decay and are needed to show convergence of the K -representations to the target densities in Theorem 4. Both conditions involve a sequence of included components $K = K(n)$ with $K \rightarrow \infty$ as $n \rightarrow \infty$.

$$(S2) \left(\sum_{k=1}^K \lambda_k^{-1} \sigma_k^{-1} \right) n^{-1/2} = O(1).$$

$$(S3) \left(\sum_{k=1}^K \lambda_k^{-1} \right) \left(\sum_{k>K} \lambda_k \right) = O(1).$$

The following bound is essential,

$$R(\tau) = \max_{1 \leq k \leq K} \max_{1 \leq i \leq n} C_4 \left(\tau, \|X_i\|_\infty, \|\partial X_i(\theta, s)/\partial s\|_\infty, d_\infty(X_i, \hat{X}_{ik}) \right) \quad (35)$$

$$= O \left(\tau^p \max_{1 \leq k \leq K, 1 \leq i \leq n} e^{2d_\infty(X_i, \hat{X}_{ik})} \right). \quad (36)$$

Here $\|X_i\|_\infty$ and $\|\partial X_i(\theta, s)/\partial s\|_\infty$ are bounded, as per (T2). The convergence of $d_\infty(X_i, \hat{X}_{ik})$ can be decomposed into two parts, the convergence of $d_\infty(X_{ik}, \hat{X}_{ik})$ and that of $d_\infty(X_i, X_{ik})$. Both are $O_p(1)$, using assumptions (S2), respectively, (S3).

The convergence rate of the truncated density estimator in (25) is as follows.

Theorem 4. Under assumptions (D1), (F1), (T1)-(T5), (E1), (K1)-(K2), if $K(n) \rightarrow \infty$ satisfies conditions (S2)-(S3), then

$$\max_{1 \leq i \leq n} d_{SW} \left(f_i(\cdot), \hat{f}_i(\cdot, K, \tau) \right) = O_p \left(\tau^{-(\kappa-p)} + \tau^p \left(n^{-1/2} \sum_{k=1}^K \sigma_k^{-1} + \left(\sum_{k>K} \lambda_k \right)^{1/2} \right) \right).$$

Under exponentially decaying eigenvalues, one can derive convergence rates that depend on the dimensionality of the domain and the growth behavior of the regularization parameter τ . These rates are obtained for specific choices of K and τ , as stated in Theorem 2, which follows as a corollary of Theorem 4.

H Temperature Data Case Study

Our second data illustration involves analyzing maximum and minimum temperature data obtained from the National Centers for Environmental Information, with raw data conveniently accessible at <https://www.ncdc.noaa.gov/>. This extensive dataset comprehensively covers weather observations spanning the entirety of the United States. We focus on climate stations positioned at major international airports within the 44 largest U.S. cities, since the quality of the data from airport weather stations is superior compared to other general weather stations due to the obvious relevance of weather conditions for the operation of aircraft. The temporal scope of our investigation spans from January 1, 2020, to December 31, 2020, capturing a comprehensive spectrum of climatic conditions throughout the year 2020. In preliminary data processing, we converted the min/max daily temperature into bivariate vectors (min,range), where range=max-min temperature for each day. This simple linear transformation yields bivariate data that are equivalent to the original data but avoids the somewhat awkward triangular support set of the original (min/max) data.

Utilizing kernel density estimation, as illustrated in Figure 10, with a Gaussian density kernel and bandwidth set at 20 for the maximum temperature and 10 for the temperature range, we obtain estimated densities as depicted in Figure 11. Here we utilize five-dimensional representations. Dimension reduction is achieved by employing only five components to represent the infinite-dimensional space while explaining a fraction of variance with respect to the sliced Wasserstein distance that exceeds 0.85.

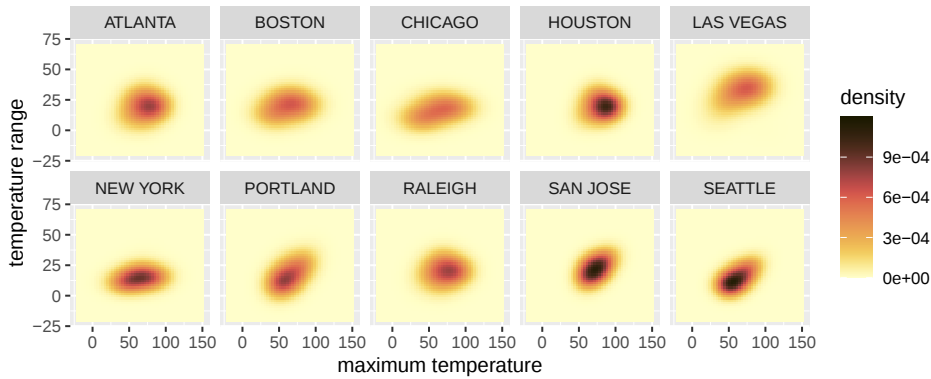


Figure 10: Observed smoothed densities for a few randomly selected airport weather stations.

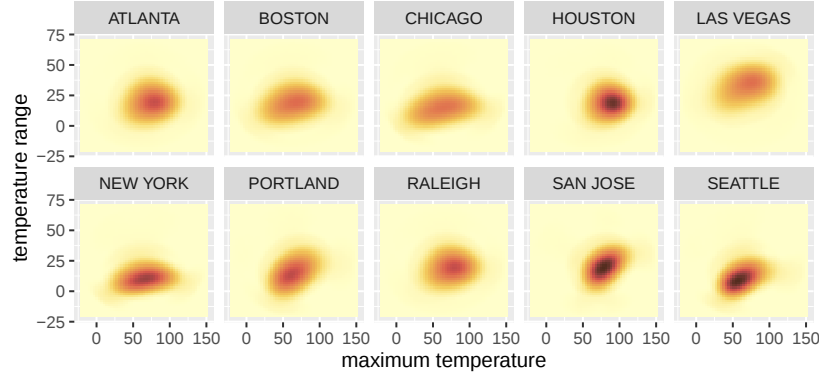


Figure 11: Fitted density surfaces for a few airports with sliced Wasserstein fraction of variance explained at level 0.85 with five components, where the first component contributes 0.41 and the second 0.18.

The first two transformation modes of variation are shown in Figure 12 while scores related to eigenfunctions are presented in Figure 13 and 14. The first model of variation reflects both the intensity of variation in the underlying distribution and the average level of temperature range. A lower score of the first eigenfunction indicates reduced variation and a lower temperature range, while a higher score corresponds to amplified variation, particularly in maximum temperature, and an overall higher temperature range. The second mode of variation reflects the shape of the joint distribution. For smaller scores it reflects a flatter joint distribution, where the temperature range exhibits a lower overall variation and a weaker correlation with the maximum temperature, while higher scores are associated with a more positive correlation between the temperature range and maximum temperature and is accompanied by increased overall variation in the temperature range.

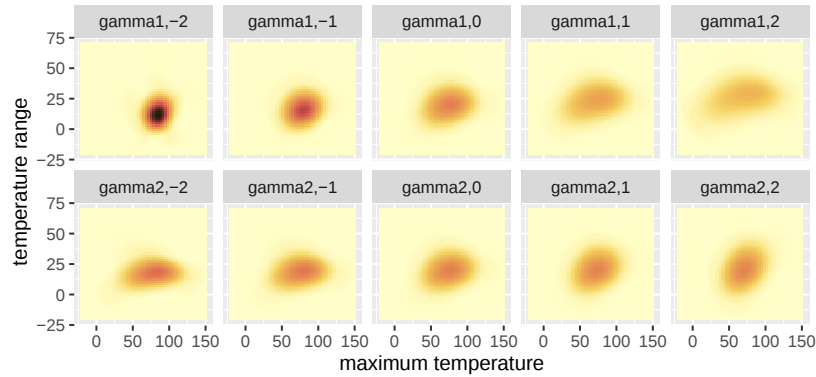


Figure 12: Transformation modes of variation as per (13) for bivariate temperature distributions. The first row corresponds to the first mode of variation and the second row to the second mode of variation, for α varying from -2 to 2 as per (17).

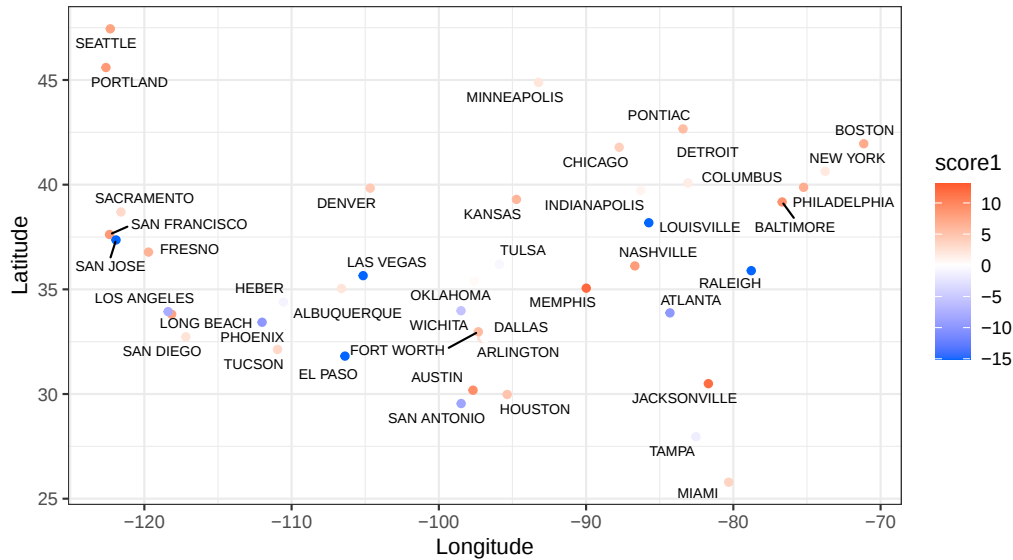


Figure 13: Visualization of the scores for the first eigenfunction.

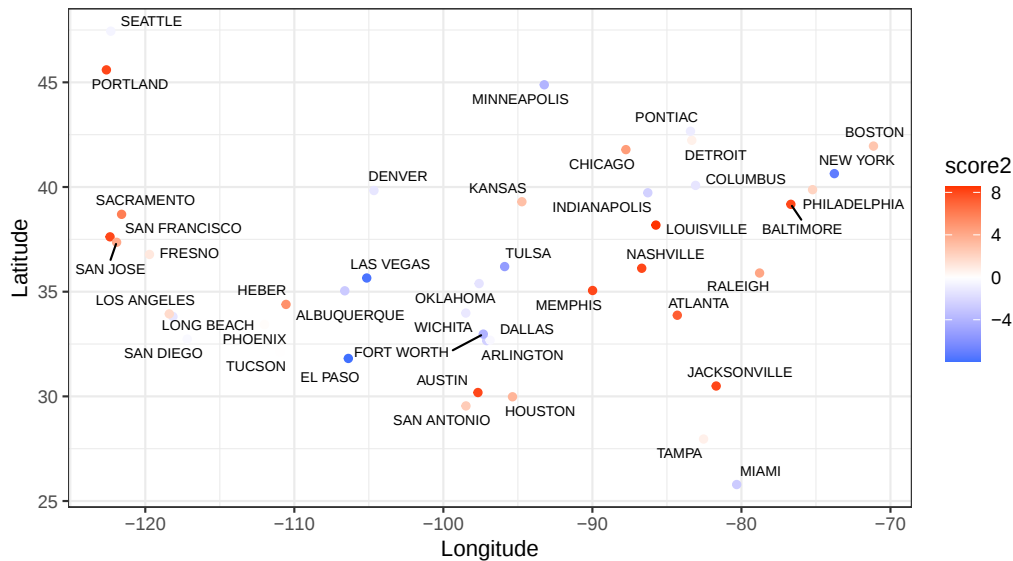


Figure 14: Visualization of the scores corresponding to the second eigenfunction.

Take San Jose and Portland. Both have positive scores for the second eigenfunction, indicating a higher correlation between temperature range and maximum temperature. Nevertheless, San Jose's variance is smaller, attributable to its negative score for the first eigenfunction, in contrast to Portland's positive score. On the other side of the spectrum, New York exhibits a negative score for the second eigenfunction, resulting in a flattened shape of the joint distribution. This pattern reflects a weak association between the temperature range and the maximum temperature, accompanied by a lower overall variation in the temperature range. San Jose and Portland's increased correlation between temperature range and maximum temperature likely is due to regional factors, such as maritime influence due to their proximity to the Pacific Ocean, resulting in temperature amplitudes that are relatively smaller in winter compared to summer. New York's weakened association and flatter distribution is also likely linked to its specific geography with relatively little variation in the temperature range.

I Computation Resources

All data analyses were performed on a MacBook Pro with an Apple M1 chip and 16GB RAM, running R version 4.3.1 (2023-06-16) under macOS Big Sur. The full analysis for the BLSA dataset and the temperature dataset (presented in the Appendix) each took approximately 5 minutes to complete.

J Proofs

J.1 Proof of Proposition 3

Proof of Proposition 3 follows directly from the proofs of Propositions S2 and S3 in [9].

J.2 Proof of Proposition 4

Denote the density f_{D_f} on the domain D as $\tilde{f} = f_{D_f} \mathbf{1}_D$. We first show that for any $R > 0, \delta > 0$,

$$n \sup_{f \in \mathcal{F}} P \left(d_\infty \left(\check{f}, \tilde{f} \right) > RN^{-\frac{2}{p+4} + \delta} \middle| f \right) = o(1) \quad \text{as } N \rightarrow \infty.$$

By the triangle inequality,

$$d_\infty \left(\check{f}, \tilde{f} \right) \leq d_\infty \left(\check{f}, E(\check{f}) \right) + d_\infty \left(E(\check{f}), \tilde{f} \right).$$

From equation (S.19) in [9] we find

$$\sup_{x \in D} \sup_{f \in \mathcal{F}} \left| \tilde{f}(x) - E(\check{f}(x)) \right| = O(h^2).$$

By Proposition 9 in [42], there exist constants L_1 and L_2 , which depend only on the VC characteristics of $\mathcal{K}$ in (K2), such that for any $R > 0$ and sequences $a_N, L_h = O(h^2)$,

$$\begin{aligned} P \left(d_\infty(\check{f}, \tilde{f}) > Ra_N \middle| f \right) &\leq P \left(d_\infty(\check{f}, E(\check{f})) > Ra_N \middle| f \right) + P \left(d_\infty(E(\check{f}), \tilde{f}) > Ra_N \middle| f \right) \\ &\leq L_1 \exp\{-L_2 N h^p R^2 a_N^2\} + I\{L_h > Ra_N\}, \end{aligned}$$

where I is the indicator function. Let $a_N = N^{-2/(p+4) + \delta}$ for any $\delta > 0$, and $h \sim N^{-\frac{1}{p+4}}$. Thus

$$\sup_{f \in \mathcal{F}} P \left(d_\infty \left(\check{f}, \tilde{f} \right) > Ra_N \middle| f \right) \leq L_1 \exp\{-L_2 R^2 N^{2\delta}\}.$$

Under assumption (P1), $N(n) \sim n^r$, for a constant $r > 1 + p/4$, therefore

$$n \sup_{f \in \mathcal{F}} P \left(d_\infty \left(\check{f}, \tilde{f} \right) > Ra_N \middle| f \right) \leq n L_1 \exp\{-L_2 R^2 n^{2\delta r}\} = o(1) \quad \text{as } n \rightarrow \infty.$$

Next, we show the convergence of the truncated density $\hat{f}$, where the target is $f(x) = \tilde{f}(x)/\int_D \tilde{f}(u)du, x \in D$. Note that when N is large enough we have

$$\int_D \tilde{f}(u)du \geq \frac{|D|}{M_0}, \quad \int_D \check{f}(u)du \geq \frac{|D|}{2M_0},$$

whence

$$\begin{aligned} d_\infty \left(\hat{f}, f \right) &\leq d_\infty \left(\frac{\check{f}(x)}{\int_D \check{f}(u)du}, \frac{\tilde{f}(x)}{\int_D \tilde{f}(u)du} \right) + d_\infty \left(\frac{\tilde{f}(x)}{\int_D \tilde{f}(u)du}, \frac{\tilde{f}(x)}{\int_D \tilde{f}(u)du} \right) \\ &\leq \frac{2M_0}{|D|} d_\infty(\check{f}, \tilde{f}) + \frac{2M_0^3}{|D|} d_\infty(\check{f}, \tilde{f}). \end{aligned}$$

It follows that for any $R > 0$ and $\delta > 0$,

$$n \sup_{f \in \mathcal{F}} P \left(d_\infty \left(\hat{f}, f \right) > RN^{-\frac{2}{p+4} + \delta} \middle| f \right) = o(1) \quad \text{as } n \rightarrow \infty.$$

J.3 Proof of Proposition 2

Lemma 1. *Under (D1) and (F1), there exists a constant A such that,*

$$\sup_{f \in \mathcal{F}} |\mathcal{R}(f)(\theta_1, s) - \mathcal{R}(f)(\theta_2, s)| \leq A \|\theta_1 - \theta_2\|_2, \quad \text{for all } \theta_1, \theta_2 \in \Theta.$$

Proof of Lemma 1. We first show the result for $p = 2$. Write $\theta(\alpha) = (\cos \alpha, \sin \alpha)^T \in \mathbb{S}^1$ where $\alpha \in [0, 2\pi]$ and $x = (x_1, x_2)^T \in \mathbb{R}^2$. Let Π_Δ be the two-dimensional rotation operator that employs the angle Δ and $f_{\Pi_\Delta} = f(\Pi_\Delta \circ x)$. Note that for all z ,

$$\mathcal{R}(f_{\Pi_\Delta})(\theta(\alpha), z) = \mathcal{R}(f)(\theta(\alpha + \Delta), z).$$

Calculating the derivative with respect to α ,

$$\mathcal{R}(f)(\theta(\alpha + \Delta), z) - \mathcal{R}(f)(\theta(\alpha), z) = \mathcal{R}(f_{\Pi_\Delta} - f)(\theta(\alpha), z). \quad (37)$$

Note that the two-dimensional rotation operator can be represented as a rotation matrix,

$$\begin{aligned} \Pi_\Delta \circ x &= \begin{pmatrix} \cos \Delta & -\sin \Delta \\ \sin \Delta & \cos \Delta \end{pmatrix} \circ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \\ &= \begin{pmatrix} x_1 \cos \Delta - x_2 \sin \Delta \\ x_1 \sin \Delta + x_2 \cos \Delta \end{pmatrix} \end{aligned} \quad (38)$$

Then,

$$\begin{aligned} f_{\Pi_\Delta}(x_1, x_2) - f(x_1, x_2) &= f(x_1 \cos \Delta - x_2 \sin \Delta, x_1 \sin \Delta + x_2 \cos \Delta) - f(x_1, x_2) \\ &= \frac{\partial f}{\partial x_1}(x_1^*, x_2^*)(x_1(\cos \Delta - 1) - x_2 \sin \Delta) \\ &\quad + \frac{\partial f}{\partial x_2}(x_1^*, x_2^*)(x_1 \sin \Delta + x_2(\cos \Delta - 1)), \end{aligned}$$

where $x_1^* = (1 - c_0)x_1 + c_0(x_1 \cos \Delta - x_2 \sin \Delta)$ and $x_2^* = (1 - c_0)x_2 + c_0(x_1 \sin \Delta + x_2 \cos \Delta)$ for a $c_0 \in (0, 1)$. Thus,

$$\lim_{\Delta \rightarrow 0} \frac{f_{\Pi_\Delta}(x_1, x_2) - f(x_1, x_2)}{\Delta} = -x_2 \frac{\partial f}{\partial x_1}(x_1, x_2) + x_1 \frac{\partial f}{\partial x_2}(x_1, x_2). \quad (39)$$

Combining (37) and (39), we have

$$\begin{aligned} \frac{\partial \mathcal{R}(f(x_1, x_2))}{\partial \alpha}(\theta(\alpha), z) &= \lim_{\Delta \rightarrow 0} \frac{\mathcal{R}(f(x_1, x_2))(\theta(\alpha + \Delta), z) - \mathcal{R}(f(x_1, x_2))(\theta(\alpha), z)}{\Delta} \\ &= \lim_{\Delta \rightarrow 0} \frac{\mathcal{R}(f_{\Pi_\Delta}(x_1, x_2) - f(x_1, x_2))(\theta(\alpha), z)}{\Delta} \\ &= \mathcal{R} \left[\left(x_1 \frac{\partial f}{\partial x_2} - x_2 \frac{\partial f}{\partial x_1} \right) (x_1, x_2) \right] (\theta(\alpha), z). \end{aligned}$$

Under (D1) and (F1), there exists a constant A_0 such that

$$\left| \frac{\partial \mathcal{R}(f)}{\partial \alpha}(\theta(\alpha), z) \right| \leq A_0.$$

Let $\theta_1 = \theta(\alpha_1)$ and $\theta_2 = \theta(\alpha_2) \in \mathbb{S}^2$, $\alpha_1, \alpha_2 \in [0, 2\pi]$, where we note that $\|\theta_1 - \theta_2\|_2 = |2 \sin((\alpha_1 - \alpha_2)/2)| \gtrsim |\alpha_1 - \alpha_2|$ when $|\alpha_1 - \alpha_2| \rightarrow 0$. Thus, for any $\theta_1, \theta_2 \in \mathbb{S}^1$, $z \in \mathbb{R}$, $f \in \mathcal{F}$,

$$|\mathcal{R}(f)(\theta_1, z) - \mathcal{R}(f)(\theta_2, z)| \leq A_0 |\alpha_1 - \alpha_2| \lesssim 2A_0 \|\theta_1 - \theta_2\|_2, \quad \text{as } \|\theta_1 - \theta_2\|_2 \rightarrow 0.$$

Next, we provide the proof for $p = 3$.

Let $\alpha \in [0, 2\pi]$, $\beta \in [0, \pi]$, $\theta(\alpha, \beta) = (\sin \beta \cos \alpha, \sin \beta \sin \alpha, \cos \beta)^T \in \mathbb{S}^2$ and $x = (x_1, x_2, x_3)^T \in \mathbb{R}^3$. Noting that

$$\begin{aligned} &\frac{\partial \mathcal{R}(f(x_1, x_2, x_3))}{\partial \alpha}(\theta(\alpha, \beta), z) \\ &= -(\sin \alpha \sin \beta) \mathcal{R} \left[\frac{\partial f}{\partial x_1}(x_1, x_2, x_3) \right] (\theta(\alpha, \beta), z) + (\cos \alpha \sin \beta) \mathcal{R} \left[\frac{\partial f}{\partial x_2}(x_1, x_2, x_3) \right] (\theta(\alpha, \beta), z) \end{aligned}$$

$$\begin{aligned}
& \frac{\partial R(f(x_1, x_2, x_3))}{\partial \beta}(\theta(\alpha, \beta), z) \\
&= -(\sin \alpha \cos \beta) \mathcal{R} \left[\frac{\partial f}{\partial x_1}(x_1, x_2, x_3) \right] (\theta(\alpha, \beta), z) + (\cos \alpha \cos \beta) \mathcal{R} \left[\frac{\partial f}{\partial x_2}(x_1, x_2, x_3) \right] (\theta(\alpha, \beta), z) \\
&\quad - (\sin \beta) \mathcal{R} \left[\frac{\partial f}{\partial x_3}(x_1, x_2, x_3) \right] (\theta(\alpha, \beta), z),
\end{aligned}$$

under (D1) and (F1), there exists a constant A_0 such that

$$\left| \frac{\partial \mathcal{R}(f)}{\partial \alpha}(\theta(\alpha, \beta), z) \right| \leq 2A_0 |\sin \beta|, \quad \left| \frac{\partial \mathcal{R}(f)}{\partial \beta}(\theta(\alpha, \beta), z) \right| \leq 3A_0.$$

Let $\theta_1 = \theta(\alpha_1, \beta_1)$ and $\theta_2 = \theta(\alpha_2, \beta_2) \in \mathbb{S}^3$, $\alpha_1, \alpha_2 \in [0, 2\pi]$ and $\beta_1, \beta_2 \in [0, \pi]$. Since $\|\theta_1 - \theta_2\|_2 = \sqrt{2 - 2(\sin \beta_1 \sin \beta_2 \cos(\alpha_2 - \alpha_1) + \cos \beta_1 \cos \beta_2)} \gtrsim |\sin \beta_1| |\alpha_2 - \alpha_1| + |\beta_2 - \beta_1|$ when $\|\theta_2 - \theta_1\|_2 \rightarrow 0$, for any $\theta_1, \theta_2 \in \mathbb{S}^2$, $z \in \mathbb{R}$, $f \in \mathcal{F}$, as $\|\theta_1 - \theta_2\|_2 \rightarrow 0$,

$$\begin{aligned}
|\mathcal{R}(f)(\theta_1, z) - \mathcal{R}(f)(\theta_2, z)| &\leq 2A_0 |\sin \beta_1| |\alpha_2 - \alpha_1| + 3A_0 |\beta_2 - \beta_1| \\
&\lesssim \|\theta_2 - \theta_1\|_2.
\end{aligned}$$

Following these arguments, it is clear how the proof can be extended to the multivariate case when $p > 3$, and we omit the details. $\square$

The analytical tools in Section 2.2 provide the following property for the sliced density function, which implies that the sliced univariate density function adheres to the domain assumption (A1) in [35]. This observation is basic for our subsequent exploration of multivariate density transformation.

Lemma 2. *Under assumptions (D1) and (F1), $\mathcal{R}(f)(\theta, \cdot)$, $f \in \mathcal{F}$, is κ -differentiable, $\kappa \geq p + 1$, and there exists a constant C_0 such that*

$$\sup_{f \in \mathcal{F}} \|\mathcal{R}(f)\|_\infty < C_0, \quad \sup_{f \in \mathcal{F}} \|1/\mathcal{R}(f)\|_\infty < C_0, \quad \sup_{f \in \mathcal{F}} \left\| \frac{\partial^\kappa \mathcal{R}(f)}{\partial z}(\theta, z) \right\|_\infty < C_0.$$

Proof of Lemma 2. From assumption (F1) which states that the density f is uniformly bounded above and below, it follows that

$$\|\mathcal{R}(f)\|_\infty \leq M_0 |D|, \quad \|1/\mathcal{R}(f)\|_\infty \leq M_0 |D|,$$

where $|D|$ represents the Lebesgue measure of the domain set D . From (S1) in [9] and assumption (F1), there exists an constant $C_0 \geq M_0 |D|$ such that

$$\left\| \frac{\partial^\kappa \mathcal{R}(f)}{\partial z}(\theta, z) \right\|_\infty \leq C_0.$$

Note that the constant C_0 is chosen independently of the function f , which completes the proof of the proposition. $\square$

Proof of Proposition 2. Denote the slice-wise density transformation $\tilde{\psi}$ as

$$\tilde{\psi}(h)(\theta, s) = \psi(h(\theta, \cdot))(s), \quad h \in \mathcal{R}(\mathcal{F})$$

and the slice-wise inverse transformation $\tilde{\psi}^{-1}$ as

$$\tilde{\psi}^{-1}(\zeta)(\theta, z) = \psi^{-1}(\zeta(\theta, \cdot))(z), \quad \zeta \in \Psi(\mathcal{F}). \quad (40)$$

The multivariate density transformation in (6) is $\Psi(f)(\theta, s) = \tilde{\psi} \circ \mathcal{R}(f)(\theta, s)$. From assumption (D1) and Proposition 2 in [9], for any $f_1, f_2 \in \mathcal{F}$, there exists a constant A_1 such that

$$\begin{aligned}
d_2(\mathcal{R}(f_1), \mathcal{R}(f_2)) &\leq A_1 d_2(f_1, f_2), \\
d_\infty(\mathcal{R}(f_1), \mathcal{R}(f_2)) &\leq A_1 d_\infty(f_1, f_2).
\end{aligned}$$

Furthermore, from Proposition 2, there exists a constant A_2 such that

$$\|\mathcal{R}(f)\|_\infty \leq A_2, \quad \|1/\mathcal{R}(f)\|_\infty \leq A_2, \quad \left\| \frac{\partial \mathcal{R}(f)(\theta, z)}{\partial z} \right\|_\infty \leq A_2, \quad \text{for all } f \in \mathcal{F}.$$

From (L1), for any $f_1, f_2 \in \mathcal{F}$, there exists a constant A_3 such that

$$\begin{aligned} d_2(\tilde{\psi} \circ \mathcal{R}(f_1), \tilde{\psi} \circ \mathcal{R}(f_2)) &= \left(\int_{\theta \in \Theta} \int_{u \in \mathcal{S}} \left(\tilde{\psi} \circ \mathcal{R}(f_1)(\theta, u) - \tilde{\psi} \circ \mathcal{R}(f_2)(\theta, u) \right)^2 \right)^{1/2} \\ &\leq \left(\int_{\theta \in \Theta} A_3^2 d_2^2(\mathcal{R}(f_1)(\theta, \cdot), \mathcal{R}(f_2)(\theta, \cdot)) \right)^{1/2} \\ &= A_3 d_2(\mathcal{R}(f_1), \mathcal{R}(f_2)). \end{aligned}$$

Similarly,

$$d_\infty(\tilde{\psi} \circ \mathcal{R}(f_1), \tilde{\psi} \circ \mathcal{R}(f_2)) \leq A_3 d_\infty(\mathcal{R}(f_1), \mathcal{R}(f_2)),$$

which proves the existence of $C_1 = A_1 A_3$ in (T1). Furthermore from (L2), there exists a constant A_4 such that

$$\sup_{f \in \mathcal{F}} \left\| \left(\tilde{\psi} \circ \mathcal{R} \right) (f) \right\|_\infty \leq A_4, \quad \sup_{f \in \mathcal{F}} \left\| \frac{\partial \left(\tilde{\psi} \circ \mathcal{R} \right) (f)}{\partial z} (\theta, z) \right\|_\infty \leq A_4.$$

From Lemma 1 and (L1), it follows that there also exists a constant A_5 such that for any $f \in \mathcal{F}$,

$$\begin{aligned} \sup_s \left| \left(\tilde{\psi} \circ \mathcal{R} \right) (f)(\theta_1, s) - \left(\tilde{\psi} \circ \mathcal{R} \right) (f)(\theta_2, s) \right| &\leq \tilde{C}_0 \sup_z |\mathcal{R}(f)(\theta_1, z) - \mathcal{R}(f)(\theta_2, z)| \\ &\leq \tilde{C}_0 A_5 \|\theta_1 - \theta_2\|_2. \end{aligned}$$

This shows the existence of C_2 in (T2). Assumption (T3) is supported by the inverse transformation formula (40) and the injectivity of the Radon transform on rapidly decreasing functions, as established in Theorem 2.4 of [21]. The regularized inverse (7) is $\Psi_\tau^{-1} = \mathcal{R}_\tau^{-1} \circ \tilde{\psi}$. From Corollary 1 of [9], there exists a constant depending on $C_3(\tau) = O(\tau^{-(\kappa-p)})$ as $\tau \rightarrow \infty$ such that

$$\begin{aligned} d_\infty(\Psi_\tau^{-1} \circ \Psi(f), f) &= d_\infty(\mathcal{R}_\tau^{-1} \circ \tilde{\psi}^{-1} \circ \tilde{\psi} \circ \mathcal{R}(f), f) \\ &= d_\infty(\mathcal{R}_\tau^{-1} \circ \mathcal{R}(f), f) \\ &\leq C_3(\tau), \end{aligned}$$

which shows the existence of C_3 in (T4). Furthermore, there exists a constant $A_6(\tau) = O(\tau^p)$ as $\tau \rightarrow \infty$, such that

$$\begin{aligned} d_\infty(\Psi_\tau^{-1}(\zeta_1), \Psi_\tau^{-1}(\zeta_2)) &= d_\infty(\mathcal{R}_\tau^{-1} \circ \tilde{\psi}^{-1}(\zeta_1), \mathcal{R}_\tau^{-1} \circ \tilde{\psi}^{-1}(\zeta_2)) \\ &\leq A_6(\tau) d_2(\tilde{\psi}^{-1}(\zeta_1), \tilde{\psi}^{-1}(\zeta_2)). \end{aligned} \tag{41}$$

From assumption (L3), there exist constants $A_7(\|\zeta_1\|_\infty, \|\partial \zeta_1(\theta, s)/\partial s\|_\infty)$ and $A_8(d_\infty(\zeta_1, \zeta_2))$,

$$\begin{aligned} d_2(\tilde{\psi}^{-1}(\zeta_1), \tilde{\psi}^{-1}(\zeta_2)) &= \left(\int_{\theta \in \mathcal{S}^{p-1}} \int_z (\psi^{-1}(\zeta_1(\theta, \cdot))(z) - \psi^{-1}(\zeta_2(\theta, \cdot))(z))^2 \right)^{1/2} \\ &\leq \left(\int_{\theta \in \mathcal{S}^{p-1}} A_7^2 A_8^2 d_2(\zeta_1(\theta, \cdot), \zeta_2(\theta, \cdot))^2 \right)^{1/2} \\ &= A_7 A_8 d_2(\zeta_1, \zeta_2). \end{aligned} \tag{42}$$

Combining (41) and (42), we have

$$d_\infty(\Psi_\tau^{-1}(\zeta_1), \Psi_\tau^{-1}(\zeta_2)) \leq A_6(\tau) A_7(\|\zeta_1\|_\infty, \|\partial \zeta_1(\theta, s)/\partial s\|_\infty) A_8(d_\infty(\zeta_1, \zeta_2)) d_2(\zeta_1, \zeta_2),$$

which proves the existence of $C_4 = A_6 A_7 A_8$ in (T5). From Proposition 2 of [35], when using the LQD transformation ψ_Q ,

$$A_7 = O\left(\|\partial\zeta_1(\theta, s)/\partial s\|_\infty e^{6\|\zeta_1\|_\infty}\right), \quad A_8 = O\left(e^{2d_\infty(\zeta_1, \zeta_2)}\right).$$

Thus, when using the RLQD transformation Ψ_Q , we have

$$C_4(\tau, \|\zeta_1\|_\infty, \|\partial\zeta_1(\theta, s)/\partial s\|_\infty, d_\infty(\zeta_1, \zeta_2)) \sim \tau^p \|\partial\zeta_1(\theta, s)/\partial s\|_\infty e^{6\|\zeta_1\|_\infty} e^{2d_\infty(\zeta_1, \zeta_2)}.$$

For the LH transformation ψ_H ,

$$A_7 = O\left(e^{2\|\zeta_1\|_\infty}\right), \quad A_8 = O\left(e^{d_\infty(\zeta_1, \zeta_2)}\right).$$

Thus, when using the RLH transformation Ψ_H , we have

$$C_4(\tau, \|\zeta_1\|_\infty, \|\partial\zeta_1(\theta, s)/\partial s\|_\infty, d_\infty(\zeta_1, \zeta_2)) \sim \tau^p e^{2\|\zeta_1\|_\infty} e^{d_\infty(\zeta_1, \zeta_2)}.$$

□

J.4 Proof of Proposition 1

Proof. From assumption (T2),

$$\|\Psi(f)\|_\infty \leq C_2, \quad \|\partial s \Psi(f)(\theta, s)/\partial s\|_\infty \leq C_2.$$

From the fact that $f \in \mathcal{F}$ and ζ is a continuous function on $L^2(\Theta \times \mathcal{S})$,

$$d_\infty(\Psi(f), \zeta) \leq \|\Psi(f)\|_\infty + \|\zeta\|_\infty < \infty.$$

Thus

$$C_4(\tau, \|\Psi(f)\|_\infty, \|\partial \Psi(f)(\theta, s)/\partial s\|_\infty, d_\infty(\Psi(f), \zeta)) < \infty.$$

From assumption (T5),

$$d_\infty(\Psi_\tau^{-1}(\Psi(f)), \Psi_\tau^{-1}(\zeta)) \leq C_4 d_2(\Psi(f), \zeta),$$

and from assumption (T4),

$$d_\infty(f, \Psi_\tau^{-1}(\Psi(f))) \leq C_3.$$

Then

$$d_\infty(f, \Psi_\tau^{-1}(\zeta)) \leq C_3 + C_4 d_2(\Psi(f), \zeta).$$

From Assumption (D1), we have

$$d_2(f, \Psi_\tau^{-1}(\zeta)) \leq C_3 + C_4 d_2(\Psi(f), \zeta).$$

From Lemma S3 in [9] and assumption (D1), it follows that

$$d_{SW}(f, \Psi_\tau^{-1}(\zeta)) = O(C_3 + C_4 d_2(\Psi(f), \zeta)).$$

□

J.5 Proof of Theorem 3

Lemma 3. Assume (T2) and let $\tilde{\nu}$ as per (22), then

$$d_\infty(\tilde{\nu}, \nu) = O_p\left((\log n/n)^{1/2}\right). \quad (43)$$

Proof of Lemma 3. As $\mathcal{S} \subset \mathbb{R}$ is a compact interval, one can find a grid $\mathcal{T}_\mathcal{S} = \{s_1, \dots, s_a\}$ with $|\mathcal{T}_\mathcal{S}| = O(n)$ such that for any $s \in \mathcal{S}$, there exists a point $s_m \in \mathcal{T}_\mathcal{S}$ satisfying $|s - s_m| < 1/n$. Similarly, one can find a set $\mathcal{T}_\Theta = \{\theta_1, \dots, \theta_b\}$ such that for any $\theta \in \Theta$, there exists a point $\theta_m \in \mathcal{T}_\Theta$ satisfying $\|\theta - \theta_m\|_2 < 1/n$. The size of the set $|\mathcal{T}_\Theta|$ is determined by the covering number of Θ . From Lemma 5.7 and Example 5.8 in [46], one can choose $\mathcal{T}_\Theta$ such that its size satisfies

$|\mathcal{T}_\Theta| = O(n^p)$. Forming the product of $\mathcal{T}_\mathcal{S}$ and $\mathcal{T}_\Theta$ leads to the grid $\mathcal{T} = \mathcal{T}_\Theta \times \mathcal{T}_\mathcal{S}$ on $\Theta \times \mathcal{S}$ with $|\mathcal{T}| = O(n^{p+1})$.

From the smoothness assumption in (T2),

$$\sup_{\theta \in \Theta, s \in \mathcal{S}} |\tilde{\nu}(\theta, s) - \nu(\theta, s)| = \sup_{\theta, s \in \mathcal{T}} |\tilde{\nu}(\theta, s) - \nu(\theta, s)| + O(1/n).$$

Under (T2), $\|X_i(\theta, s)\|_\infty \leq C_2$. Applying Bernstein inequality for a fixed $(\theta, s) \in \mathcal{T}$, for any $R > 0$,

$$\begin{aligned} P\left(|\tilde{\nu}(\theta, s) - \nu(\theta, s)| \geq R \left(\frac{\log(n)}{n}\right)^{1/2}\right) &= P\left(\left|\frac{1}{n} \sum_{i=1}^n X_i(\theta, s) - \nu(\theta, s)\right| \geq R \left(\frac{\log(n)}{n}\right)^{1/2}\right) \\ &\leq 2 \exp\left(-\frac{3(n \log n)R^2}{6nC_2^2 + 2C_2R(n \log n)^{1/2}}\right). \end{aligned}$$

Then

$$\begin{aligned} P\left(\sup_{\theta, s \in \mathcal{T}} |\tilde{\nu}(\theta, s) - \nu(\theta, s)| \geq R \left(\frac{\log(n)}{n}\right)^{1/2}\right) &\leq 2n^{p+1} \exp\left(-\frac{3(n \log n)R^2}{6nC_2^2 + 2C_2R(n \log n)^{1/2}}\right) \\ &= 2n^{p+1-R^*}, \end{aligned} \quad (44)$$

where $R^* = \frac{3R^2}{6C_2^2 + 2C_2R(\log n/n)^{1/2}}$. For a fixed dimension p , choose a large enough R such that $R^* > p + 1$. Then

$$\sup_{\theta \in \Theta, s \in \mathcal{S}} |\tilde{\nu}(\theta, s) - \nu(\theta, s)| = O_p\left(\left(\frac{\log n}{n}\right)^{1/2}\right). \quad (45)$$

□

Proof of Theorem 3. Assumption (T2) implies that $E\|X\|_2^2 < \infty$. Denote by $\tilde{\nu}, \tilde{G}$ the population mean function and covariance function as per (22). Using Theorems 3.9 and 4.2 in [8], it follows that

$$d_2(\nu, \tilde{\nu}) = O_p(n^{-1/2}), \quad d_2(G, \tilde{G}) = O_p(n^{-1/2}).$$

From Proposition 3,

$$\sup_{f \in \mathcal{F}} E\left[d_2(f, \check{f})^2 | f\right] = O\left(N^{-\frac{4}{p+4}}\right).$$

The convergence rate of $d_2(\tilde{\nu}, \hat{\nu})$ is

$$\begin{aligned} P(\{d_2(\tilde{\nu}, \hat{\nu}) > RN^{-\frac{2}{p+4}}\}) &\leq P\left(\frac{1}{n} \sum_{i=1}^n d_2(X_i, \check{X}_i) > RN^{-\frac{2}{p+4}}\right) \\ &\leq \frac{\frac{1}{n} \sum_{i=1}^n E(d_2(X_i, \check{X}_i))}{RN^{-\frac{2}{p+4}}} \\ &\leq \frac{\frac{1}{n} \sum_{i=1}^n \sqrt{E(d_2(X_i, \check{X}_i)^2)}}{RN^{-\frac{2}{p+4}}} \\ &\leq \frac{\frac{1}{n} \sum_{i=1}^n C_1 \sqrt{E(d_2(f_i, \check{f}_i)^2)}}{RN^{-\frac{2}{p+4}}} \\ &= O\left(\frac{1}{R}\right), \end{aligned}$$

which indicates that $d_2(\tilde{\nu}, \hat{\nu}) = O_p(N^{-\frac{2}{p+4}})$, then $d_2(\nu, \hat{\nu}) = O_p(n^{-1/2} + N^{-\frac{2}{p+4}})$. Similarly, we have $d_2(H, \hat{H}) = O_p(n^{-1/2} + N^{-\frac{2}{p+4}})$. In terms of uniform convergence, Lemma 3 provides the uniform convergence for $\tilde{\nu}$ and similarly for $\tilde{H}$,

$$d_\infty(\tilde{\nu}, \nu) = O_p\left(\left(\frac{\log n}{n}\right)^{1/2}\right), \quad d_\infty(\tilde{H}, H) = O_p\left(\left(\frac{\log n}{n}\right)^{1/2}\right).$$

From Proposition 3, for any $R > 0, \delta > 0$,

$$P\left(d_\infty(\tilde{\nu}, \hat{\nu}) > RN^{-\frac{2}{p+4}+\delta}\right) \leq n \max_{1 \leq i \leq n} P(d_\infty(f_i, \check{f}) > C_1^{-1} RN^{-\frac{2}{p+4}+\delta}) \\ = o(1).$$

Thus $d_\infty(\hat{\nu}, \nu) = O_p\left(\left(\frac{\log n}{n}\right)^{1/2} + N^{-\frac{2}{p+4}+\delta}\right)$ and similarly,

$d_\infty(\hat{H}, H) = O_p\left(\left(\frac{\log n}{n}\right)^{1/2} + N^{-\frac{2}{p+4}+\delta}\right)$. Based on Lemma (4.2) and (4.3) of [8], it follows that

$$|\lambda_k - \hat{\lambda}_k| = O_p(n^{-1/2} + N^{-\frac{2}{p+4}}), \\ d_2(\gamma_k, \hat{\gamma}_k) = \sigma_k^{-1} O_p(n^{-1/2} + N^{-\frac{2}{p+4}}).$$

In terms of the uniform convergence rate of γ_k , employing the techniques in Lemma 1 of [33] and the proof of Corollary 1 of [47],

$$d_\infty(\hat{\lambda}_k \gamma_k, \hat{\lambda}_k \hat{\gamma}_k) \leq d_\infty(\lambda_k \gamma_k, \hat{\lambda}_k \hat{\gamma}_k) + d_\infty(\lambda_k \gamma_k, \hat{\lambda}_k \gamma_k) \\ \lesssim d_\infty(H, \hat{H}) + d_2(\gamma_k, \hat{\gamma}_k) \|H\|_\infty + |\lambda_k - \hat{\lambda}_k| \|\gamma_k\|_\infty.$$

Thus,

$$d_\infty(\gamma_k, \hat{\gamma}_k) = O_p\left(\lambda_k^{-1} \sigma_k^{-1} n^{-1/2}\right).$$

□

J.6 Proof of Theorem 1

Lemma 4. Under (D1), (F1) and (T2), there exists a constant C such that

$$\|\nu\|_\infty \leq C, \quad \|\partial \nu(\theta, s)/\partial s\|_\infty \leq C, \quad (46)$$

$$\|\gamma_k\|_\infty \leq C \lambda_k^{-1}, \quad \|\partial \gamma_k(\theta, s)/\partial s\|_\infty \leq C \lambda_k^{-1}. \quad (47)$$

Proof of Lemma 4. The derivation of (46) follows directly from (T2). Note that

$$\gamma_k(\theta, s) = \lambda_k^{-1} \int_{\Theta \times \mathcal{S}} H(\theta, s, \alpha, t) \gamma_k(\alpha, t) d\alpha dt.$$

(T2) and the fact that $\|\gamma_k\|_2 = 1$ then imply (47). □

Proof of Theorem 1. Assume K is fixed. Then from Lemma 4,

$$\|X_{k,\alpha}\|_\infty \leq \|\nu\|_\infty + \alpha_0 \sqrt{\lambda_1} \|\gamma_k\|_\infty \leq C \left(1 + \alpha_0 \sqrt{\lambda_1} \lambda_K^{-1}\right). \quad (48)$$

Similarly,

$$\|\partial X_{k,\alpha}(\theta, s)/\partial s\|_\infty \leq C \left(1 + \alpha_0 \sqrt{\lambda_1} \lambda_K^{-1}\right). \quad (49)$$

Employing Theorem 3,

$$d_\infty(X_{k,\alpha}, \hat{X}_{k,\alpha}) \leq d_\infty(\nu, \hat{\nu}) + \alpha_0 \left(\sqrt{\hat{\lambda}_1} d_\infty(\gamma_k, \hat{\gamma}_k) + \|\gamma_k\|_\infty \left|\sqrt{\lambda_k} - \sqrt{\hat{\lambda}_k}\right|\right) \\ = O_p(\sigma_K^{-1} \lambda_K^{-1} n^{-\frac{1}{2}}).$$

Here we use the fact that

$$\left|\sqrt{\lambda_k} - \sqrt{\hat{\lambda}_k}\right| = \frac{|\lambda_k - \hat{\lambda}_k|}{\sqrt{\lambda_k} + \sqrt{\hat{\lambda}_k}} \sim \lambda_k^{-\frac{1}{2}} |\lambda_k - \hat{\lambda}_k|, \quad \text{as } n \rightarrow \infty.$$

For the L^2 convergence,

$$\begin{aligned} d_2(X_{k,\alpha}, \hat{X}_{k,\alpha}) &\leq d_2(\nu, \hat{\nu}) + \alpha_0 \left(\sqrt{\lambda_1} d_2(\gamma_k, \hat{\gamma}_k) + \left| \sqrt{\lambda_k} - \sqrt{\hat{\lambda}_k} \right| \right) \\ &= O_p(\sigma_K^{-1} n^{-1/2}). \end{aligned}$$

From assumption (T4),

$$\begin{aligned} &d_\infty(m_k(\alpha, \Psi, \tau), \hat{m}_k(\alpha, \Psi, \tau)) \\ &\leq C_4(\tau, \|X_{k,\alpha}\|_\infty, \|\partial X_{k,\alpha}(\theta, s)/\partial s\|_\infty, d_\infty(X_{k,\alpha}, \hat{X}_{k,\alpha})) d_2(X_{k,\alpha}, \hat{X}_{k,\alpha}). \end{aligned}$$

Case 1: When K is fixed,

$$\max\{\|X_{k,\alpha}\|_\infty, \|\partial X_{k,\alpha}(\theta, s)/\partial s\|_\infty, d_\infty(X_{k,\alpha}, \hat{X}_{k,\alpha})\} = O_p(1).$$

Writing $S(\tau, K)$ as per (19), by (T5) it follows that $S(\tau, K) = O(\tau^p)$. Thus

$$d_\infty(m_k(\alpha, \Psi, \tau), \hat{m}_k(\alpha, \Psi, \tau)) = O_p(\tau^p n^{-\frac{1}{2}}).$$

From Lemma S3 in [9] and assumption (D1),

$$d_{SW}(m_k(\alpha, \Psi, \tau), \hat{m}_k(\alpha, \Psi, \tau)) = O_p(\tau^p n^{-\frac{1}{2}}).$$

Case 2: When $K = K(n) \rightarrow \infty$,

$$d_\infty(X_{k,\alpha}, \hat{X}_{k,\alpha}) = O_p(1).$$

From assumption (T5), (48) and (49),

$$\begin{aligned} &d_\infty(m_k(\alpha, \Psi, \tau), \hat{m}_k(\alpha, \Psi, \tau)) \\ &\leq C_4 \left(\tau, \|X_{k,\alpha}\|_\infty, \|\partial X_{k,\alpha}(\theta, s)/\partial s\|_\infty, d_\infty(X_{k,\alpha}, \hat{X}_{k,\alpha}) \right) \sigma_K^{-1} n^{-1/2} \\ &= O_p \left(\tau^p \max_{1 \leq k \leq K} \left\{ e^{5\|\gamma_k\|_\infty} \|\partial \gamma_k(\theta, s)/\partial s\|_\infty \right\} \sigma_K^{-1} n^{-1/2} \right). \end{aligned}$$

From Lemma S3 in [9] and assumption (D1),

$$d_{SW}(m_k(\alpha, \Psi, \tau), \hat{m}_k(\alpha, \Psi, \tau)) = O_p \left(\tau^p \max_{1 \leq k \leq K} \left\{ e^{5\|\gamma_k\|_\infty} \|\partial \gamma_k(\theta, s)/\partial s\|_\infty \right\} \sigma_K^{-1} n^{-1/2} \right).$$

□

J.7 Proof of Theorem 4

Lemma 5. Note that under assumption (S3),

$$d_\infty(X, X_K) = O_p(1).$$

Proof of Lemma 5. Recall that $X_K(\theta, s) = \nu(\theta, s) + \sum_{k=1}^K \chi_k \gamma_k(\theta, s)$ and let $Z_K = X - X_K$. For the k -th principal component $\chi_k = \int_{\Theta \times \mathcal{S}} X(\theta, s) \gamma_k(\theta, s) d\theta ds \leq C_2$ from the assumption (T2). We first show that the process Z_K is Lipschitz continuous,

$$|Z_K(\theta_1, s_1) - Z_K(\theta_2, s_2)| \leq L_K (\|\theta_1 - \theta_2\|_2 + |s_1 - s_2|),$$

where $L_k = O(\sum_{k=1}^K \lambda_k^{-1})$. From assumption (T2),

$$\begin{aligned} |X(\theta_1, s_1) - X(\theta_2, s_2)| &\leq |X(\theta_1, s_1) - X(\theta_2, s_2)| + |X(\theta_1, s_1) - X(\theta_2, s_2)| \\ &\leq C_2 (\|\theta_1 - \theta_2\|_2 + |s_1 - s_2|). \end{aligned}$$

Similarly,

$$\begin{aligned} |\nu(\theta_1, s_1) - \nu(\theta_2, s_2)| &\leq C_2 (\|\theta_1 - \theta_2\|_2 + |s_1 - s_2|), \\ |H(\theta_1, s_1, \alpha, t) - H(\theta_2, s_2, \alpha, t)| &\leq 2C_2 (\|\theta_1 - \theta_2\|_2 + |s_1 - s_2|). \end{aligned}$$

Expanding the eigenfunction $\gamma_k(\theta, s) = \lambda_k^{-1} \int_{\Theta \times \mathcal{S}} H(\theta, s, \alpha, t) \gamma_k(\alpha, t) d\alpha dt$,

$$\begin{aligned} |\gamma_k(\theta_1, s_1) - \gamma_k(\theta_2, s_2)| &\leq \lambda_k^{-1} \int_{\Theta \times \mathcal{S}} (H(\theta_1, s_1, \alpha, t) - H(\theta_2, s_2, \alpha, t)) \gamma_k(\alpha, t) d\alpha dt \\ &\leq 2C_2 \lambda_k^{-1} (\|\theta_1 - \theta_2\|_2 + |s_1 - s_2|). \end{aligned}$$

Then

$$\begin{aligned} |X_K(\theta_1, s_1) - X_K(\theta_2, s_2)| &\leq |\nu(\theta_1, s_1) - \nu(\theta_2, s_2)| + \sum_{k=1}^K |\chi_k| |\gamma_k(\theta_1, s_1) - \gamma_k(\theta_2, s_2)| \\ &\leq C_2 \left(1 + 2C_2 \sum_{k=1}^K \lambda_k^{-1} \right) (\|\theta_1 - \theta_2\|_2 + |s_1 - s_2|), \end{aligned}$$

whence

$$\begin{aligned} |Z_K(\theta_1, s_1) - Z_K(\theta_2, s_2)| &\leq C_2 \left(2 + 2C_2 \sum_{k=1}^K \lambda_k^{-1} \right) (\|\theta_1 - \theta_2\|_2 + |s_1 - s_2|) \\ &\leq L_K (\|\theta_1 - \theta_2\|_2 + |s_1 - s_2|), \end{aligned}$$

where $L_K = O(\sum_{k=1}^K \lambda_k^{-1})$. Similar to Lemma 1 in [35], it follows that

$$\|Z_K\|_\infty \leq 2 \max \left(|\Theta|^{-1/2} |\mathcal{S}|^{-1/2} \|Z_K\|_2, L_K^{1/3} \|Z_K\|_2^{2/3} \right). \quad (50)$$

Regarding the L^2 norm of Z_K ,

$$\begin{aligned} E(\|Z_K\|_2^2) &= \int_{\Theta \times \mathcal{S}} E(X - X_K)^2 d\theta ds \\ &= \int_{\Theta \times \mathcal{S}} E \left(\sum_{k>K} \chi_k \gamma_k \right)^2 d\theta ds \\ &= \sum_{k>K} \lambda_k. \end{aligned}$$

Then

$$\|Z_K\|_2^2 = \|X - X_K\|_2^2 = O_p \left(\sum_{k>K} \lambda_k \right). \quad (51)$$

Combining (50) and assumption (S3),

$$d_\infty(X, X_K) = O_p \left(\left(\sum_{k=1}^K \lambda_k^{-1} \sum_{k>K} \lambda_k \right)^{1/3} \right) = O_p(1).$$

□

Proof of Theorem 4. Writing $X_{iK} = \nu + \sum_{k=1}^K \chi_{ik} \gamma_k$ and $\hat{X}_{iK} = \hat{\nu} + \sum_{k=1}^K \hat{\chi}_{ik} \hat{\gamma}_k$, note that $\chi_{ik} = \int_{\Theta \times \mathcal{S}} X_i(\theta, s) \gamma_k(\theta, s) d\theta ds$ and $\hat{\chi}_{ik} = \int_{\Theta \times \mathcal{S}} \check{X}_i(\theta, s) \hat{\gamma}_k(\theta, s) d\theta ds$. By the triangle inequality and Theorem 3,

$$\begin{aligned} |\chi_{ik} - \hat{\chi}_{ik}| &= \left| \int_{\Theta \times \mathcal{S}} (X_i(\theta, s) \gamma_k(\theta, s) - \check{X}_i(\theta, s) \hat{\gamma}_k(\theta, s)) d\theta ds \right| \\ &\leq \left| \int_{\Theta \times \mathcal{S}} X_i(\theta, s) (\gamma_k(\theta, s) - \hat{\gamma}_k(\theta, s)) d\theta ds \right| + \left| \int_{\Theta \times \mathcal{S}} (X_i(\theta, s) - \check{X}_i(\theta, s)) \hat{\gamma}_k(\theta, s) d\theta ds \right| \\ &\lesssim d_2(\gamma_k, \hat{\gamma}_k) + d_2(X_i, \check{X}_i) \\ &= O_p(\sigma_k^{-1} n^{-1/2}). \end{aligned}$$

Here the last inequality follows from Proposition 3 and assumption (E1). Assumption (S3) and Lemma 5 imply $d_\infty(X_{iK}, X_i) = O_p(1)$. The triangle inequality and results in Theorem 3 and Lemma 4 imply

$$\begin{aligned} d_\infty(X_i, \hat{X}_{iK}) &\leq d_\infty(\nu, \hat{\nu}) + \sum_{k=1}^K d_\infty(\chi_{ik}\gamma_k, \hat{\chi}_{ik}\hat{\gamma}_k) + d_\infty(X_{iK}, X_i) \\ &\leq d_\infty(\nu, \hat{\nu}) + \sum_{k=1}^K (|\chi_{ik} - \hat{\chi}_{ik}| \|\gamma_k\|_\infty + \hat{\chi}_{ik} d_\infty(\gamma_k, \hat{\gamma}_k)) + d_\infty(X_{iK}, X_i) \\ &= O_p\left(\left(\sum_{k=1}^K \sigma_k^{-1} \lambda_k^{-1}\right) n^{-1/2}\right). \end{aligned}$$

From (S2) and assumption (T2),

$$\max_{1 \leq i \leq n} \left\{ \|X_i\|_\infty, \|\partial X_i(\theta, s)/\partial s\|_\infty, d_\infty(X_i, \hat{X}_{iK}) \right\} = O_p(1).$$

From assumption (T2) and (T5),

$$\begin{aligned} R(\tau) &= \max_{1 \leq k \leq K} \max_{1 \leq i \leq n} C_4 \left(\tau, \|X_i\|_\infty, \|\partial X_i(\theta, s)/\partial s\|_\infty, d_\infty(X_i, \hat{X}_{iK}) \right) \\ &= O_p(\tau^p). \end{aligned}$$

Regarding the L^2 convergence,

$$\begin{aligned} d_2(X_i, \hat{X}_{iK}) &\leq d_2(\nu, \hat{\nu}) + \sum_{k=1}^K d_2(\chi_{ik}\gamma_k, \hat{\chi}_{ik}\hat{\gamma}_k) + d_2(X_{iK}, X_i) \\ &\leq d_2(\nu, \hat{\nu}) + \sum_{k=1}^K (|\chi_{ik} - \hat{\chi}_{ik}| \|\gamma_k\|_2 + \hat{\chi}_{ik} d_2(\gamma_k, \hat{\gamma}_k)) + d_2(X_{iK}, X_i) \\ &= O_p\left(\left(\sum_{k=1}^K \sigma_k^{-1}\right) n^{-1/2} + \left(\sum_{k>K} \lambda_k\right)^{1/2}\right). \end{aligned}$$

Here we use (51) from Lemma 5 and $d_2(X_{iK}, X_i) = O_p\left(\left(\sum_{k>K} \lambda_k\right)^{1/2}\right)$. Using assumptions (T4) and (T5),

$$\begin{aligned} d_\infty(f_i(\cdot), \hat{f}_i(\cdot, K, \Psi, \tau)) &= d_\infty(\Psi^{-1}(X_i), \Psi_\tau^{-1}(\hat{X}_{iK})) \\ &\leq d_\infty(\Psi^{-1}(X_i), \Psi_\tau^{-1}(X_i)) + d_\infty(\Psi_\tau^{-1}(X_i), \Psi_\tau^{-1}(\hat{X}_{iK})) \\ &= O_p\left(C_3(\tau) + R(\tau) \left(\left(\sum_{k=1}^K \sigma_k^{-1}\right) n^{-1/2} + \left(\sum_{k>K} \lambda_k\right)^{1/2}\right)\right) \\ &= O_p\left(\tau^{-(\kappa-p)} + \tau^p \left(\left(\sum_{k=1}^K \sigma_k^{-1}\right) n^{-1/2} + \left(\sum_{k>K} \lambda_k\right)^{1/2}\right)\right). \end{aligned}$$

From Lemma S3 in [9] and assumption (D1),

$$d_{SW}(f_i(\cdot), \hat{f}_i(\cdot, K, \Psi, \tau)) = O_p\left(\tau^{-(\kappa-p)} + \tau^p \left(\left(\sum_{k=1}^K \sigma_k^{-1}\right) n^{-1/2} + \left(\sum_{k>K} \lambda_k\right)^{1/2}\right)\right).$$

With eigenvalues $\lambda_k = ce^{-\theta k}$ for some constants c and θ , it follows that

$$\sum_{k=1}^K \sigma_k^{-1} \sim \sum_{k=1}^K e^{\theta k} \sim e^{\theta K}.$$

Similarly,

$$\sum_{k>K} \lambda_k = \sum_{k>K} e^{-\theta k} \sim e^{-\theta K}.$$

Choosing $K(n) = \lfloor \frac{1}{3\theta} \log(n) \rfloor$,

$$d_{SW} \left(f_i(\cdot), \hat{f}_i(\cdot, K, \Psi, \tau) \right) = O_p \left(\tau^{-(\kappa-p)} + \tau^p n^{-1/6} \right),$$

and choosing $\tau \sim n^{1/6k}$,

$$d_{SW} \left(f_i(\cdot), \hat{f}_i(\cdot, K, \Psi, \tau) \right) = O_p \left(\tau^{-(\kappa-p)/6k} \right).$$

□