The Annals of Statistics
2022, Vol. 50, No. 3, 1573–1592
https://doi.org/10.1214/21-AOS2163

UNIFORM CONVERGENCE OF LOCAL FRÉCHET REGRESSION WITH APPLICATIONS TO LOCATING EXTREMA AND TIME WARPING FOR METRIC SPACE VALUED TRAJECTORIES

BY YAQING CHEN[a] AND HANS-GEORG MÜLLER[b]

Department of Statistics, University of California, Davis
[a]*yaqchen@ucdavis.edu,* [b]*hgmueller@ucdavis.edu*

Local Fréchet regression is a nonparametric regression method for metric space valued responses and Euclidean predictors, which can be utilized to obtain estimates of smooth trajectories taking values in general metric spaces from noisy metric space valued random objects. We derive uniform rates of convergence, which so far have eluded theoretical analysis of this method, for both fixed and random target trajectories, where we utilize tools from empirical processes. These results are shown to be widely applicable in metric space valued data analysis. In addition to simulations, we provide two pertinent examples where these results are important: The consistent estimation of the location of properly defined extrema in metric space valued trajectories, which we illustrate with the problem of locating the age of minimum brain connectivity as obtained from fMRI data; and time warping for metric space valued trajectories, illustrated with yearly age-at-death distributions for different countries.

1. Introduction. Non-Euclidean data, or random object data taking values in metric spaces have become increasingly common in modern data analysis and data science while there is a lack of principled and statistically justified methodology. Since such data are metric space valued, they generally do not lie in a vector space, which means that many classical notions of statistics such as the definition of sample or population mean as an average or expected value do not apply anymore and need to be replaced by barycenters or Fréchet means (Fréchet (1948)), the mathematical and statistical properties of which have been studied for various metric spaces. These include finite-dimensional Riemannian manifolds, the space of symmetric positive definite matrices, Kendall's shape space or the Wasserstein space of distributions (Agueh and Carlier (2011), Bhattacharya and Patrangenaru (2003), Bhattacharya and Patrangenaru (2005), Dryden, Koloydenko and Zhou (2009), Huckemann (2012), Le Gouic and Loubes (2017), among others); the latter is not a Riemannian manifold (Ambrosio, Gigli and Savaré (2004)).

Another important topic is to study the relationship of such random objects with other variables, where regression analysis comes into play. Nonparametric (local) regression techniques have been used for a long time for smoothing and interpolation of Euclidean responses. While Nadaraya–Watson type methods have been proposed when data lie in finite-dimensional Riemannian manifolds (Davis et al. (2007), Pelletier (2006), Steinke and Hein (2009), Steinke, Hein and Schölkopf (2010), Yuan et al. (2012)), and also generic metric spaces (Hein (2009)), local Fréchet regression (Petersen and Müller (2019)), can be viewed as a generalization of local linear regression for metric space valued responses. While pointwise asymptotic results for the corresponding estimates have been previously derived, uniform convergence is much more challenging. Here, we derive uniform rates of convergence

Received August 2020; revised December 2021.
MSC2020 subject classifications. Primary 62G05, 62G20; secondary 62G08.
Key words and phrases. Random objects, metric space valued random processes, rates of convergence, smoothing, fMRI, mortality distributions.

for local Fréchet regression estimates of the fixed conditional Fréchet mean trajectory using tools from empirical process theory; see Theorem 1. We then extend this result to the case where local Fréchet regression is applied to recover metric space valued random processes from discrete noisy observations; see Theorem 2. While these results are of interest in their own right, our derivations are motivated by important applications of uniform convergence. These include the estimation of the location of suitably defined extrema in metric space valued functions as well as time warping for metric space valued functional data.

Estimation of modes or maximum locations has been well studied for regression functions in nonparametric regression for real-valued data (e.g., Belitser, Ghosal and van Zanten (2012), Devroye (1978), Müller (1989)) and densities of probability distributions (e.g., Balabdaoui, Rufibach and Wellner (2009), Parzen (1962), Vieu (1996)). For object data in metric spaces, the location of extrema with regard to functionals of interest can be obtained based on the estimation of the complete conditional Fréchet mean trajectory through local Fréchet regression, where the consistency of the derived estimates of the location of an extremum is guaranteed by the uniform convergence of local Fréchet regression under regularity conditions.

For real-valued functional data, a random function may be considered to reflect two types of random variation: amplitude variation and phase (or time) variation. Confounding these two types of variation may compromise conventional statistical methods (Kneip and Gasser (1992)). This issue has been addressed by introducing time warping, also referred to as curve synchronization, registration or alignment. The prototypical method is dynamic time warping (DTW) (Sakoe and Chiba (1978)) and various statistical approaches have been developed over the years for real-valued functional data (Gasser and Kneip (1995), Gervini and Gasser (2004), James (2007), Kneip and Gasser (1992), Ramsay and Li (1998), Wang and Gasser (1997), among others); see Marron et al. (2015) for a review. Beyond classical functional data in $\mathcal{L}^2$ Hilbert space, time synchronization has been investigated in engineering for non-Euclidean semimetric spaces, also referred to as dissimilarity spaces (Faragó, Linder and Lugosi (1993)), where the DTW method and its variants have been adopted with applications including human motion recognition and video classification (Gong and Medioni (2011), Trigeorgis et al. (2018), Vu, Carey and Mahadevan (2012), among others). We note that no theoretical results were provided in these works. To our knowledge, no comprehensive studies exist of time warping for samples of metric space valued trajectories that include an investigation of statistical properties or asymptotic behavior.

Statistical methods devised for real-valued functional data, or random elements of a Hilbert space, are usually not applicable to functional data taking values in a metric space (Huckemann (2015)). Even extending functional data analysis methods without time warping to metric space valued trajectories is challenging (Dubey and Müller (2020)), due to the fact that in general metric spaces one cannot make use of an algebraic structure. In this paper, we tackle the even more challenging task to extend pairwise warping for real-valued curves (Tang and Müller (2008)) to metric space valued functional data. Since random processes are usually not fully observed and only discrete and noisy measurements are available, local Fréchet regression needs to be used to obtain complete subject-specific trajectories. The uniform convergence result in Theorem 2 for local Fréchet regression estimates of random processes is crucial to derive the uniform consistency of estimates of the pairwise warping functions, which form the backbone of the proposed warping method; the uniform consistency provides the major justification for this approach.

The remainder of the paper is organized as follows. A key result on the uniform rate of convergence for local Fréchet regression is presented in Section 2, followed by a study of the case where the target of the local Fréchet regression is a random process rather than a fixed trajectory in Section 3. We present two applications, where the estimation of the location of

extrema is based on Theorem 2 and presented in Section 4. A second key application is the time synchronization for metric space valued functional data in Section 5, which is based on Theorem 2. The proposed methods are shown to lead to consistent estimation of time warping functions in Theorem 3 and Corollary 3. We then demonstrate the estimation of extrema locations with functional magnetic resonance imaging (fMRI) data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database in Section 6.1, where we find the time of minimum brain connectivity, quantified by the Fiedler values of the brain network. The time warping for metric space valued functional data is illustrated with yearly age-at-death distribution data for different countries from the Human Mortality Database in Section 6.2. We also report the results of various simulation studies for the proposed warping method for distribution-valued functional data in Section S.5 in the Supplementary Material (Chen and Müller (2022)).

2. Uniform rates of convergence for local Fréchet regression. Let $(\mathcal{M}, d_{\mathcal{M}})$ be a totally bounded separable metric space, and $\mathcal{T} = [0, \tau]$ be a closed interval in $\mathbb{R}$. This will be assumed throughout. Consider a random pair (U, V) with a joint distribution on the product space $\mathcal{T} \times \mathcal{M}$, where U is a real-valued predictor and V is a metric space valued response. Suppose $\{(U_j, V_j)\}_{j=1}^m$ are i.i.d. realizations of (U, V). For any $t \in \mathcal{T}$, the conditional Fréchet mean of V given $U = t$ is defined by

$$\nu(t) = \operatorname*{argmin}_{z \in \mathcal{M}} L(z, t), \qquad L(z, t) = \mathbb{E}\big[d_{\mathcal{M}}^2(V, z) \mid U = t\big]. \tag{2.1}$$

We consider local Fréchet means (Petersen and Müller (2019))

$$\widetilde{\nu}_b(t) = \operatorname*{argmin}_{z \in \mathcal{M}} \widetilde{L}_b(z, t), \qquad \widetilde{L}_b(z, t) = \mathbb{E}\big[w(U, t, b) d_{\mathcal{M}}^2(V, z)\big], \tag{2.2}$$

where $w(s, t, b) = K_b(s-t)[\rho_{2,b}(t) - \rho_{1,b}(t)(s-t)]/\sigma_b^2(t)$, $\rho_{l,b}(t) = \mathbb{E}[K_b(U-t)(U-t)^l]$, for $l = 0, 1, 2$, $\sigma_b^2(t) = \rho_{0,b}(t)\rho_{2,b}(t) - \rho_{1,b}(t)^2$, $K_b(\cdot) = K(\cdot/b)/b$, K is a smoothing kernel, and $b = b(m) > 0$ is a bandwidth sequence. Local Fréchet regression estimates of $\nu(t)$ are given by

$$\widehat{\nu}_m(t) = \operatorname*{argmin}_{z \in \mathcal{M}} \widehat{L}_m(z, t), \qquad \widehat{L}_m(z, t) = m^{-1} \sum_{j=1}^m \widehat{w}(U_j, t, b) d_{\mathcal{M}}^2(V_j, z), \tag{2.3}$$

where $\widehat{w}(s, t, b) = K_b(s-t)[\widehat{\rho}_{2,m}(t) - \widehat{\rho}_{1,m}(t)(s-t)]/\widehat{\sigma}_m^2(t)$, $\widehat{\rho}_{l,m}(t) = m^{-1} \sum_{j=1}^m [K_b(U_j - t)(U_j - t)^l]$, $l = 0, 1, 2$, $\widehat{\sigma}_m^2(t) = \widehat{\rho}_{0,m}(t)\widehat{\rho}_{2,m}(t) - \widehat{\rho}_{1,m}(t)^2$.

Let $\mathcal{T}^\circ = (0, \tau)$ be the interior of $\mathcal{T}$. We require the following assumptions to obtain uniform rates of convergence over $t \in \mathcal{T}$ for local Fréchet regression estimators in (2.3).

(K0) The kernel K is a probability density function, symmetric around zero and uniformly continuous on $\mathbb{R}$. Defining $K_{kl} = \int_{\mathbb{R}} K(x)^k x^l \, \mathrm{d}x < \infty$, for $k, l \in \mathbb{N}$, K_{14} and K_{26} are finite. The derivative K' exists and is bounded on the support of K, that is, $\sup_{K(x)>0} |K'(x)| < \infty$; additionally, $\int_{\mathbb{R}} x^2 |K'(x)| \sqrt{|x \log |x||} \, \mathrm{d}x < \infty$.

(R0) The marginal density f_U of U and the conditional densities $f_{U|V}(\cdot, z)$ of U given $V = z$ exist and are continuous on $\mathcal{T}$ and twice continuously differentiable on $\mathcal{T}^\circ$, the latter for all $z \in \mathcal{M}$. The marginal density f_U is bounded away from zero on $\mathcal{T}$, $\inf_{t \in \mathcal{T}} f_U(t) > 0$. The second-order derivative f_U'' is bounded, $\sup_{t \in \mathcal{T}^\circ} |f_U''(t)| < \infty$, and the second-order partial derivatives $(\partial^2 f_{U|V}/\partial t^2)(\cdot, z)$ are uniformly bounded, $\sup_{t \in \mathcal{T}^\circ, z \in \mathcal{M}} |(\partial^2 f_{U|V}/\partial t^2)(t, z)| < \infty$. Additionally, for any open set $E \subset \mathcal{M}$, $P(V \in E \mid U = t)$ is continuous as a function of t; for any $t \in \mathcal{T}$, $L(z, t)$ is equicontinuous, that is,

$$\limsup_{s \to t} \sup_{z \in \mathcal{M}} \big|L(z, s) - L(z, t)\big| = 0. \tag{2.4}$$

(R1) For all $t \in \mathcal{T}$, the minimizers $\nu(t)$, $\widetilde{\nu}_b(t)$ and $\widehat{\nu}_m(t)$ exist and are unique, the last P-almost surely. In addition, for any $\epsilon > 0$,

$$\inf_{t\in\mathcal{T}} \inf_{d_{\mathcal{M}}(\nu(t),z)>\epsilon} \big(L(z,t) - L\big(\nu(t),t\big)\big) > 0,$$

$$\liminf_{b\to 0} \inf_{t\in\mathcal{T}} \inf_{d_{\mathcal{M}}(\widetilde{\nu}_b(t),z)>\epsilon} \Big(\widetilde{L}_b(z,t) - \widetilde{L}_b\big(\widetilde{\nu}_b(t),t\big)\Big) > 0,$$

and there exists $c = c(\epsilon) > 0$ such that

$$P\left(\inf_{t\in\mathcal{T}} \inf_{d_{\mathcal{M}}(\widehat{\nu}_m(t),z)>\epsilon} \Big(\widehat{L}_m(z,t) - \widehat{L}_m\big(\widehat{\nu}_m(t),t\big)\Big) \geq c\right) \to 1.$$

(R2) Let $B_r(\nu(t)) \subset \mathcal{M}$ be a ball of radius r centered at $\nu(t)$ and $N(\epsilon, B_r(\nu(t)), d_{\mathcal{M}})$ be its covering number using balls of radius ϵ. Then

$$\int_0^1 \sup_{t\in\mathcal{T}} \sqrt{1 + \log N\big(r\epsilon, B_r\big(\nu(t)\big), d_{\mathcal{M}}\big)}\, \mathrm{d}\epsilon = O(1), \quad \text{as } r \to 0+.$$

(R3) There exists $r_1, r_2 > 0$, $c_1, c_2 > 0$ and $\beta_1, \beta_2 > 1$ such that

$$\inf_{t\in\mathcal{T}} \inf_{d_{\mathcal{M}}(z,\nu(t))<r_1} \big[L(z,t) - L\big(\nu(t),t\big) - c_1 d_{\mathcal{M}}\big(z,\nu(t)\big)^{\beta_1}\big] \geq 0,$$

$$\liminf_{b\to 0} \inf_{t\in\mathcal{T}} \inf_{d_{\mathcal{M}}(z,\widetilde{\nu}_b(t))<r_2} \big[\widetilde{L}_b(z,t) - \widetilde{L}_b\big(\widetilde{\nu}_b(t),t\big) - c_2 d_{\mathcal{M}}\big(z,\widetilde{\nu}_b(t)\big)^{\beta_2}\big] \geq 0.$$

Similar but weaker assumptions have been made by Petersen and Müller (2019) for pointwise rates of convergence for local Fréchet regression estimators. Assumption (K0) is needed to apply results of Silverman (1978) and Mack and Silverman (1982), and (R0) is a standard distributional assumption for local nonparametric regression. These assumptions guarantee the asymptotic uniform equicontinuity of $\widetilde{L}_b$ and control the behavior of $(\widetilde{L}_b - L)$ around $\nu(t)$ uniformly over $t \in \mathcal{T}$, whence we obtain the uniform rate for the bias part $d_{\mathcal{M}}(\nu(t), \widetilde{\nu}_b(t))$ and the uniform consistency of the stochastic part $d_{\mathcal{M}}(\widetilde{\nu}_b(t), \widehat{\nu}_m(t))$ for the local Fréchet regression estimators. In particular, (2.4) guarantees the $d_{\mathcal{M}}$-continuity of $\nu(t)$ in conjunction with (R1). Assumption (R1) is commonly used to establish the uniform consistency of M-estimators (van der Vaart and Wellner (1996)). It ensures the uniform convergence of $\widetilde{L}_b(\cdot,t)$ to $L(\cdot,t)$ and the weak convergence of the empirical process $\widehat{L}_m(\cdot,t)$ to $\widetilde{L}_b(\cdot,t)$, which, in conjunction with the assumption that the metric space $\mathcal{M}$ is totally bounded, implies the pointwise convergence of the minimizers for any given $t \in \mathcal{T}$; it also ensures that the (asymptotic) uniform equicontinuity of $\widetilde{L}_b$ and $\widehat{L}_m$ implies the (asymptotic) uniform equicontinuity of $\widetilde{\nu}_b(\cdot)$ and $\widehat{\nu}_m(\cdot)$, whence the uniform convergence of the minimizers follows as the time domain $\mathcal{T}$ is compact. Assumptions (R2) and (R3) are adapted from empirical process theory to control the differences $(\widehat{L}_m - \widetilde{L}_b)$ and $(\widetilde{L}_b - L)$ near the minimizers $\widetilde{\nu}_b(t)$ and $\nu(t)$, respectively, which is necessary to obtain the convergence rates for the bias and stochastic parts.

In the following, we discuss assumptions (R1)–(R3) in the context of some specific metric spaces.

EXAMPLE 1. Let $\mathcal{M}$ be the set of probability distributions on a closed interval of $\mathbb{R}$ with finite second moments, endowed with the $\mathcal{L}^2$-Wasserstein distance d_W; specifically, for any two distributions $z_1, z_2 \in \mathcal{M}$,

$$d_W(z_1,z_2) = \left(\int_0^1 \big(Q_{z_1}(x) - Q_{z_2}(x)\big)^2\, \mathrm{d}x\right)^{1/2} = d_{\mathcal{L}^2}(Q_{z_1}, Q_{z_2}),$$

where Q_z is the quantile function for any given distribution $z \in \mathcal{M}$. The Wasserstein space $(\mathcal{M}, d_W)$ satisfies (R1)–(R3) with $\beta_1 = \beta_2 = 2$.

EXAMPLE 2. Let $\mathcal{M}$ be the space of r-dimensional correlation matrices, that is, symmetric positive semidefinite matrices in $\mathbb{R}^{r\times r}$ with diagonal elements all equal to 1, endowed with the Frobenius metric d_F. The space $(\mathcal{M}, d_F)$ satisfies (R1)–(R3) with $\beta_1 = \beta_2 = 2$.

For Examples 1–2, we note that since the Wasserstein space and the space of correlation matrices are Hadamard spaces (the former as per Kloeckner (2010)), there exists a unique minimizer of $L(\cdot, t)$, for any $t \in \mathcal{T}$ (Sturm (2003)). Examples 1–2 follow from similar arguments as those in the proofs of Propositions 1–2 of Petersen and Müller (2019); we omit the details.

We then obtain uniform rates of convergence over $t \in \mathcal{T}$ for local Fréchet regression estimators as follows. Proofs and auxiliary results are in the Supplementary Material (Chen and Müller (2022)).

THEOREM 1. *Under* (K0), (R0)–(R3) *and if* $b \to 0$, $mb^2(-\log b)^{-1} \to \infty$, *as* $m \to \infty$, *for any* $\varepsilon > 0$, *it holds for* $\nu(t)$, $\widetilde{\nu}_b(t)$ *and* $\widehat{\nu}_m(t)$ *as per* (2.1)–(2.3), *respectively, that*

$$\sup_{t\in\mathcal{T}} d_{\mathcal{M}}\big(\nu(t), \widetilde{\nu}_b(t)\big) = O\big(b^{2/(\beta_1-1)}\big), \tag{2.5}$$

$$\begin{aligned}&\sup_{t\in\mathcal{T}} d_{\mathcal{M}}\big(\widetilde{\nu}_b(t), \widehat{\nu}_m(t)\big)\\ &\quad = O_P\big(\max\big\{\big(mb^2\big)^{-1/[2(\beta_2-1)+\varepsilon]}, \big(mb^2(-\log b)^{-1}\big)^{-1/[2(\beta_2-1)]}\big\}\big).\end{aligned} \tag{2.6}$$

Furthermore, with $b \sim m^{-(\beta_1-1)/(2\beta_1+4\beta_2-6+2\varepsilon)}$, *it holds that*

$$\sup_{t\in\mathcal{T}} d_{\mathcal{M}}\big(\nu(t), \widehat{\nu}_m(t)\big) = O_P\big(m^{-1/(\beta_1+2\beta_2-3+\varepsilon)}\big). \tag{2.7}$$

Theorem 1 is a novel and relevant result for local Fréchet regression in its own right; we expect it to be a useful and widely applicable tool for the study of metric space valued data. We note that although $\mathcal{T}$ is a closed interval, boundary effects do not pose a problem, similar to the situation for local polynomial regression with real-valued responses (e.g., Fan and Gijbels (1996)). For the bias part in (2.5), we obtain the same rate as for pointwise results. For the stochastic part, the proof is substantially more involved. When $\beta_1 = \beta_2 = 2$ as in Examples 1–2, the uniform convergence rate is found to be arbitrarily close to $O_P(m^{-1/3})$.

3. Recovering metric space valued random processes from discrete noisy measurements. We consider a metric space valued random process $Y\colon \mathcal{T} \to \mathcal{M}$ that is assumed to be $d_{\mathcal{M}}$-continuous over $\mathcal{T}$. In practice, the process Y is usually not fully observed; instead one observes noisy measurements at discrete time points. Since a metric space in general is not a vector space, and hence does not afford additive operations, it is not obvious how to express the deviation of noisy observations from the underlying process Y. To address this issue, we introduce a random perturbation map $\mathcal{P}\colon \mathcal{M} \to \mathcal{M}$ such that

$$z' = \operatorname*{argmin}_{z\in\mathcal{M}} \mathbb{E}\big[d^2_{\mathcal{M}}\big(\mathcal{P}(z'), z\big)\big], \quad \text{for all } z' \in \mathcal{M}. \tag{3.1}$$

Consider a random pair (T, Z) following a joint distribution on $\mathcal{T} \times \mathcal{M}$, where T is the time of observation and Z is a noisy observation of the process Y at a random time T, given by

$$Z = \mathcal{P}\big(Y(T)\big). \tag{3.2}$$

Then the conditional Fréchet mean of the observed object Z given the process Y and time T is the process evaluated at that time, that is,

$$Y(T) = \operatorname*{argmin}_{z\in\mathcal{M}} \mathbb{E}\big[d^2_{\mathcal{M}}(Z, z) \mid Y, T\big]. \tag{3.3}$$

Furthermore, we assume:

(P1) The time of observation T and the random perturbation map $\mathcal{P}$ are independent of the random process Y.

The analogue of assumption (P1) in Euclidean regression is the standard assumption of independence between additive noise and underlying process.

Suppose that available noisy observations of the process Y are $\{(T_j, Z_j)\}_{j=1}^m$, where $Z_j = \mathcal{P}_j(Y(T_j))$, and $\{(T_j, \mathcal{P}_j)\}_{j=1}^m$ are independent realizations of $(T, \mathcal{P})$. Hence, $\{(T_j, Z_j)\}_{j=1}^m$ are conditionally independent realizations of (T, Z) given the process Y. Local Fréchet regression can be utilized to estimate the process trajectories Y via (2.3), with trajectory estimates

$$\widehat{Y}(t) = \operatorname*{argmin}_{z\in\mathcal{M}} \frac{1}{m}\sum_{j=1}^m \widehat{v}(T_j, t, b) d^2_{\mathcal{M}}(Z_j, z), \quad \text{for all } t \in \mathcal{T}. \tag{3.4}$$

Here, $\widehat{v}(s,t,b) = K_b(s-t)[\widehat{\varrho}_{2,m}(t) - \widehat{\varrho}_{1,m}(t)(s-t)]/\widehat{\varsigma}^2_m(t)$, $b = b(m) > 0$ is a bandwidth sequence, $\widehat{\varrho}_{l,m}(t) = m^{-1}\sum_{j=1}^m K_b(T_j - t)(T_j - t)^l$, $l = 0, 1, 2$, $\widehat{\varsigma}^2_m(t) = \widehat{\varrho}_{0,m}(t)\widehat{\varrho}_{2,m}(t) - \widehat{\varrho}_{1,m}(t)^2$, $K_b(\cdot) = K(\cdot/b)/b$, and K is a kernel function.

We note that while $\widehat{v}_m$ in (2.3) is a local Fréchet regression estimate of the fixed trajectory v as per (2.1), the target of the local Fréchet regression implemented as per (3.4) is the random process Y. We next extend the results in Section 2 for local Fréchet regression with fixed targets to the case of such random targets, and obtain the uniform convergence rates for $\widehat{Y}$ over $\mathcal{T}$.

Let $(\Omega, \mathscr{F}, P)$ be the probability space on which the observed data (T_j, Z_j) are defined, where Ω is the sample space, $\mathscr{F}$ is the σ-algebra of events, and $P\colon \mathscr{F} \to [0,1]$ is the probability measure. As the random mechanisms that generate the data as per (P1) are independent, the probability space $(\Omega, \mathscr{F}, P)$ is a product space of two probability spaces, $(\Omega_1, \mathscr{F}_1, P_{\Omega_1})$, where the metric space valued process Y is defined, and $(\Omega_2, \mathscr{F}_2, P_{\Omega_2})$, where the observed times T_j and the random perturbation map $\mathcal{P}_j$ associated with the noisy observations Z_j are defined. Fixing an element $\omega_1 \in \Omega_1$ corresponds to a realization of the metric space valued process Y. Given a fixed $\omega_1 \in \Omega_1$, the observed pairs $\{(T_j, Z_j)\}_{j=1}^m$ are independent in $(\Omega_2, \mathscr{F}_2, P_{\Omega_2})$ and T and T_j do not depend on ω_1. We use Y_{ω_1}, Z_{ω_1}, $Z_{\omega_1 j}$, T and T_j to represent the corresponding quantities given $\omega_1 \in \Omega_1$ in what follows, and also $\mathbb{E}_{\Omega_2}$ for the expectation (integral) with respect to P_{Ω_2}. For any fixed $\omega_1 \in \Omega_1$, $\{(T_j, Z_{\omega_1 j})\}_{j=1}^m$ are i.i.d. realizations of (T, Z_{ω_1}). For any $t \in \mathcal{T}$, as per (3.3),

$$Y_{\omega_1}(t) = \operatorname*{argmin}_{z\in\mathcal{M}} M_{\omega_1}(z,t), \qquad M_{\omega_1}(z,t) = \mathbb{E}_{\Omega_2}\big[d^2_{\mathcal{M}}(Z_{\omega_1}, z) \mid T = t\big]. \tag{3.5}$$

The localized Fréchet mean (Petersen and Müller (2019)) is

$$\widetilde{Y}_{\omega_1,b}(t) = \operatorname*{argmin}_{z\in\mathcal{M}} \widetilde{M}_{\omega_1,b}(z,t), \qquad \widetilde{M}_{\omega_1,b}(z,t) = \mathbb{E}_{\Omega_2}\big[v(T,t,b) d^2_{\mathcal{M}}(Z_{\omega_1}, z)\big]. \tag{3.6}$$

Here, $v(s,t,b) = K_b(s-t)[\varrho_{2,b}(t) - \varrho_{1,b}(t)(s-t)]/\varsigma^2_b(t)$, where $\varrho_{l,b}(t) = \mathbb{E}_{\Omega_2}[K_b(T-t)(T-t)^l]$, for $l = 0, 1, 2$ and $\varsigma^2_b(t) = \varrho_{0,b}(t)\varrho_{2,b}(t) - \varrho_{1,b}(t)^2$. The local Fréchet regression estimates $\widehat{Y}(t)$ in (3.4) can be expressed as

$$\begin{aligned} \widehat{Y}_{\omega_1,m}(t) &= \operatorname*{argmin}_{z\in\mathcal{M}} \widehat{M}_{\omega_1,m}(z,t), \\ \widehat{M}_{\omega_1,m}(z,t) &= m^{-1}\sum_{j=1}^m \widehat{v}(T_j,t,b) d^2_{\mathcal{M}}(Z_{\omega_1 j}, z). \end{aligned} \tag{3.7}$$

Let $\mathcal{T}^\circ = (0, \tau)$ be the interior of the time domain $\mathcal{T}$. Considering an arbitrarily fixed $\omega_1 \in \Omega_1$, for local Fréchet regression as described in (3.5)–(3.7), assumptions (R0)–(R3) can be

adapted to obtain uniform rates of convergence of local Fréchet regression estimates $\widehat{Y}_{\omega_1,m}(t)$ over $t \in \mathcal{T}$. To obtain uniform rates of convergence of the local Fréchet regression estimate $\widehat{Y}$ of the random process Y over $\mathcal{T}$, we need to deal with different $\omega_1 \in \Omega_1$ simultaneously, for which we require the following stronger variants of assumptions (R0)–(R3).

(U0) The marginal density f_T of T and the conditional densities $f_{T|Z_{\omega_1}}(\cdot, z)$ of T given $Z_{\omega_1} = z$ exist and are continuous on $\mathcal{T}$ and twice continuously differentiable on $\mathcal{T}^\circ$, the latter for all $z \in \mathcal{M}$ and $\omega_1 \in \Omega_1$. The marginal density f_T is bounded away from zero on $\mathcal{T}$, $\inf_{t\in\mathcal{T}} f_T(t) > 0$. The second-order derivative f_T'' is bounded, $\sup_{t\in\mathcal{T}^\circ} |f_T''(t)| < \infty$. The second-order partial derivatives $(\partial^2 f_{T|Z_{\omega_1}}/\partial t^2)(\cdot, z)$ are uniformly bounded, $\sup_{\omega_1\in\Omega_1, t\in\mathcal{T}^\circ, z\in\mathcal{M}} |(\partial^2 f_{T|Z_{\omega_1}}/\partial t^2)(t, z)| < \infty$. Additionally, for any open set $E \subset \mathcal{M}$, $P_{\Omega_2}(Z_{\omega_1} \in E \mid T = t)$ is continuous as a function of t for all $\omega_1 \in \Omega_1$.

(U1) For all $\omega_1 \in \Omega_1$ and $t \in \mathcal{T}$, the minimizers $Y_{\omega_1}(t)$, $\widetilde{Y}_{\omega_1,b}(t)$ and $\widehat{Y}_{\omega_1,m}(t)$ exist and are unique, the last P_{Ω_2}-almost surely. Additionally, for any $\epsilon > 0$,

$$\inf_{\omega_1\in\Omega_1, t\in\mathcal{T}} \inf_{\substack{z\in\mathcal{M} \text{ s.t.} \\ d_{\mathcal{M}}(Y_{\omega_1}(t),z)>\epsilon}} \big(M_{\omega_1}(z,t) - M_{\omega_1}\big(Y_{\omega_1}(t), t\big)\big) > 0,$$

$$\liminf_{b\to 0} \inf_{\omega_1\in\Omega_1, t\in\mathcal{T}} \inf_{\substack{z\in\mathcal{M} \text{ s.t.} \\ d_{\mathcal{M}}(\widetilde{Y}_{\omega_1,b}(t),z)>\epsilon}} \Big(\widetilde{M}_{\omega_1,b}(z,t) - \widetilde{M}_{\omega_1,b}\big(\widetilde{Y}_{\omega_1,b}(t), t\big)\Big) > 0.$$

(U2) Let $B_r(Y_{\omega_1}(t)) \subset \mathcal{M}$ be a ball of radius r centered at $Y_{\omega_1}(t)$ and $N(\epsilon, B_r(Y_{\omega_1}(t)), d_{\mathcal{M}})$ be its covering number using balls of radius ϵ. Then

$$\sup_{r>0} \sup_{\omega_1\in\Omega_1} \int_0^1 \sup_{t\in\mathcal{T}} \sqrt{1 + \log N\big(r\epsilon, B_r\big(Y_{\omega_1}(t)\big), d_{\mathcal{M}}\big)}\, \mathrm{d}\epsilon < \infty.$$

(U3) There exist $c_1, c_2 > 0$ and $\beta_1, \beta_2 > 1$ such that for any $r_1, r_2 > 0$,

$$\inf_{\omega_1\in\Omega_1, t\in\mathcal{T}} \inf_{\substack{z\in\mathcal{M} \text{ s.t.} \\ d_{\mathcal{M}}(z,Y_{\omega_1}(t))<r_1}} \big[M_{\omega_1}(z,t) - M_{\omega_1}\big(Y_{\omega_1}(t), t\big) - c_1 d_{\mathcal{M}}\big(z, Y_{\omega_1}(t)\big)^{\beta_1}\big] \geq 0,$$

$$\liminf_{b\to 0} \inf_{\substack{\omega_1\in\Omega_1, \\ t\in\mathcal{T}}} \inf_{\substack{z\in\mathcal{M} \text{ s.t.} \\ d_{\mathcal{M}}(z,\widetilde{Y}_{\omega_1,b}(t))<r_2}} \big[\widetilde{M}_{\omega_1,b}(z,t) - \widetilde{M}_{\omega_1,b}\big(\widetilde{Y}_{\omega_1,b}(t), t\big) - c_2 d_{\mathcal{M}}\big(z, \widetilde{Y}_{\omega_1,b}(t)\big)^{\beta_2}\big] \geq 0.$$

We note that the assumption of equicontinuity of L as per (2.4) to guarantee the $d_{\mathcal{M}}$-continuity of ν is not needed in this case, since the process Y is assumed to be $d_{\mathcal{M}}$-continuous. We then obtain the uniform convergence rates for $\widehat{Y}$ over $\mathcal{T}$ as follows.

THEOREM 2. *Under* (P1), (K0) *and* (U0)–(U3), *for any* $\varepsilon > 0$,

$$\sup_{t\in\mathcal{T}} d_{\mathcal{M}}\big(Y(t), \widetilde{Y}(t)\big) = O\big(b^{2/(\beta_1-1)}\big);$$

$$\sup_{t\in\mathcal{T}} d_{\mathcal{M}}\big(\widetilde{Y}(t), \widehat{Y}(t)\big) = O_P\big(\max\big\{(mb^2)^{-1/[2(\beta_2-1)+\varepsilon]}, \big(mb^2(-\log b)^{-1}\big)^{-1/[2(\beta_2-1)]}\big\}\big). \tag{3.8}$$

Furthermore, if $b \sim m^{-(\beta_1-1)/(2\beta_1+4\beta_2-6+2\varepsilon)}$,

$$\sup_{t\in\mathcal{T}} d_{\mathcal{M}}\big(Y(t), \widehat{Y}(t)\big) = O_P\big(m^{-1/(\beta_1+2\beta_2-3+\varepsilon)}\big). \tag{3.9}$$

We note that Examples 1 and 2 indeed satisfy (U1)–(U3) with $\beta_1 = \beta_2 = 2$, where the uniform convergence rate in (3.9) can be arbitrarily close to $O_P(m^{-1/3})$. We also note that

the uniform convergence results for local Fréchet regression with fixed and random targets in Theorems 1 and 2, respectively, can be extended to the case of multivariate predictors at the expense of more tedious algebra similarly to multivariate nonparametric regression with scalar responses (Ruppert and Wand (1994)).

4. Estimation of the location of extrema. In this section, we consider the problem of estimating the locations of extrema (maxima and/or minima) of the conditional Fréchet mean trajectory $\nu(t)$ as per (2.1) with regard to some functional of interest that is defined on the metric space of random objects and will depend on the specific nature of these objects, corresponding to a map from the object space to the real line. Without loss of generality, we focus on the case of extrema that are minima and note that the case of zero crossings can be handled analogously.

Consider a functional $\Gamma\colon \mathcal{M} \to \mathbb{R}$ that quantifies a property of interest of the objects situated in the metric space $\mathcal{M}$. Then the conditional Fréchet mean $\nu(t)$ of V given $U = t$ as per (2.1) induces a real function $\Lambda(t)$ that reflects the dependence of the functional on the covariate t,

$$\Lambda(t) = \Gamma\big(\nu(t)\big), \quad \text{for all } t \in \mathcal{T}.$$

Our goal is to find the location where $\Lambda(\cdot)$ is minimized,

$$t_{\min} = \operatorname*{argmin}_{t\in\mathcal{T}} \Lambda(t). \tag{4.1}$$

An estimate of the minimizer $t_{\min}$ of $\Lambda(\cdot)$ is obtained by replacing $\nu(t)$ with its local Fréchet regression estimate, that is,

$$\widehat{t}_{\min} = \operatorname*{argmin}_{t\in\mathcal{T}} \widehat{\Lambda}(t), \quad \text{with } \widehat{\Lambda}(t) = \Gamma\big(\widehat{\nu}_m(t)\big). \tag{4.2}$$

In addition, we assume:

(D1) There exists $C_1 > 0$ and $\alpha_1 > 1$ such that for all $z_1, z_2 \in \mathcal{M}$, $|\Gamma(z_1) - \Gamma(z_2)| \le C_1 d_{\mathcal{M}}(z_1, z_2)^{\alpha_1}$.

(D2) The minimizer $t_{\min}$ exists and is unique. Additionally, for any $\epsilon > 0$, $\inf_{|t-t_{\min}|>\epsilon}[\Lambda(t) - \Lambda(t_{\min})] > 0$.

(D3) There exists $r, C_2 > 0$ and $\alpha_2 > 1$ such that $\inf_{|t-t_{\min}|<r}[\Lambda(t) - \Lambda(t_{\min}) - C_2|t - t_{\min}|^{\alpha_2}] \ge 0$.

Assumptions (D1) and (D2) guarantee the consistency of the minimizer estimate $\widehat{t}_{\min}$, and hence can be used to obtain the corresponding convergence rate in conjunction with (D3). An example scenario where (D1) holds with $\alpha_1 = 1$ will be given in Section 6.1. For (D2) and (D3), a sufficient condition is, for instance, that $\Lambda(\cdot)$ is twice continuously differentiable on $\mathcal{T}$ with unique minimizer $t_{\min}$ and $\Lambda''(t_{\min}) > 0$; specifically, $\alpha_2 = 2$ in (D3).

Applying Theorem 1, we obtain the following result of the minimum location estimate $\widehat{t}_{\min}$ based on local Fréchet regression.

COROLLARY 1. *Under* (K0), (R0)–(R3) *and* (D1)–(D3), *for any* $\varepsilon > 0$ *and for* $b \sim m^{-(\beta_1-1)/(2\beta_1+4\beta_2-6+2\varepsilon)}$, *it holds for the estimate* $\widehat{t}_{\min}$ *in* (4.2) *of the minimizer* $t_{\min}$ *in* (4.1) *that*

$$|\widehat{t}_{\min} - t_{\min}| = O_P\big(m^{-\alpha_1/[\alpha_2(\beta_1+2\beta_2-3+\varepsilon)]}\big). \tag{4.3}$$

We will illustrate this approach with an application to the study of brain connectivity utilizing fMRI data in Section 6.1. More generally, for an aggregation statistic determined by a functional Γ^* such that $|\Gamma^*(\widehat{\nu}_m) - \Gamma^*(\nu)| \le C_1^* \sup_{t\in\mathcal{T}} d_{\mathcal{M}}(\widehat{\nu}_m(t), \nu(t))^{\alpha_1^*}$, for some $C_1^* > 0$

and $\alpha_1^* > 1$, where $\Gamma^*(\nu), \Gamma^*(\widehat{\nu}_m) \in \mathbb{R}$, analogous rates of convergence as in (4.3) can be obtained for $|\Gamma^*(\widehat{\nu}_m) - \Gamma^*(\nu)|$. Examples where such results are useful include the estimation of zero crossings or more general level crossings and the estimation of intervals where $\Gamma^*(\nu)$ exceeds a certain level.

5. Time warping for metric space valued functional data.

5.1. *Global warping.* We consider the time warping problem for metric space valued random trajectories. With $\mathcal{T} = [0, \tau]$ being the time domain, consider a set of warping functions $\mathcal{W} = \{g : \mathcal{T} \to \mathcal{T} \mid g(0) = 0, g(\tau) = \tau, g \text{ is continuous and strictly increasing on } \mathcal{T}\}$. Note that for each function $g \in \mathcal{W}$, $g(\cdot)/\tau$ is a strictly increasing cdf on $\mathcal{T}$. Suppose $\mu : \mathcal{T} \to \mathcal{M}$ is a fixed metric space valued trajectory, and $h \in \mathcal{W}$ is a random (global) warping function such that $\mathbb{E}[h(t)] = t$, for all $t \in \mathcal{T}$. We consider the following model for the metric space valued random process $Y : \mathcal{T} \to \mathcal{M}$ in Section 3:

$$Y(t) = \mu\big(h^{-1}(t)\big), \quad \text{for all } t \in \mathcal{T}, \tag{5.1}$$

where μ is referred to as the mean trajectory, and the stochastic fluctuations of the random warping function h around the identity function id determines the phase variation of the process Y. Considering a random pair (T, Z) consisting of time of observation T and process Z, which is observed with a perturbation that is determined by the map $\mathcal{P}$ satisfying (3.1), suppose $\{(h_i, Y_i, T_i, \mathcal{P}_i, Z_i)\}_{i=1}^n$ is a set of n independent realizations of the quintuple $(h, Y, T, \mathcal{P}, Z)$, where as per (5.1) the metric space valued processes Y_i are

$$Y_i(t) = \mu\big(h_i^{-1}(t)\big), \quad \text{for all } t \in \mathcal{T}, \tag{5.2}$$

and the observed objects are $Z_i = \mathcal{P}_i(Y_i(T_i))$, as per (3.2).

Furthermore, we make the following assumptions regarding the fixed mean trajectory μ and random warping function $h \in \mathcal{W}$.

(W1) The trajectory μ is $d_{\mathcal{M}}$-continuous, that is, $\lim_{\Delta \to 0} d_{\mathcal{M}}(\mu(t + \Delta), \mu(t)) = 0$, for any $t \in \mathcal{T}$.

(W2) Defining a bivariate function $d_\mu : \mathcal{T}^2 \to \mathbb{R}$ as $d_\mu(s, t) = d_{\mathcal{M}}(\mu(s), \mu(t))$, d_μ is twice continuously differentiable with $\inf_{s=t \in \mathcal{T}} |(\partial d_\mu/\partial s)(s, t)| > 0$. For any $t_1, t_2 \in \mathcal{T}$ with $t_1 < t_2$, $\int_{t_1}^{t_2} [(\partial d_\mu/\partial s)(s, t)]^2 \, \mathrm{d}s > 0$, for all $t \in \mathcal{T}$.

(W3) The difference quotients of the global warping function h are bounded from above and below, that is, there exist constants $c, C \in (0, +\infty)$ with $c < C$ and $cC \leq 1$ such that $c \leq [h(s) - h(t)]/(s - t) \leq C$, for all $s, t \in \mathcal{T}$ with $s < t$.

Assumption (W1) implies the $d_{\mathcal{M}}$-continuity of the random process Y in conjunction with the continuity of the warping function h; (W2) excludes the possibility that any part of the trajectory μ could be flat. This is necessary to ensure the uniqueness of the warping functions, and will be used to establish the uniform convergence of the proposed estimates for the pairwise warping functions; see Section 5.4. Assumption (W3) guarantees there are no plateaus or steep increases in the global warping function and its inverse.

5.2. *Pairwise warping.* For any $i, i' \in \{1, \ldots, n\}$ such that $i \neq i'$, the random pairwise warping function $g_{i'i} : \mathcal{T} \to \mathcal{T}$ is a temporal transformation from $Y_{i'}$ toward Y_i defined by

$$g_{i'i}(t) = h_{i'}\big(h_i^{-1}(t)\big), \quad \text{for all } t \in \mathcal{T}.$$

We note that $g_{i'i} \in \mathcal{W}$. Moreover, we assume that (warping) functions in $\mathcal{W}$ can be parameterized by linear splines (as in Tang and Müller (2008)). Let $t_k = k\tau/(p + 1)$, for $k = 1, \ldots, p$, be p equidistant knots in $\mathcal{T}$, with $t_0 = 0$ and $t_{p+1} = \tau$. For any function $g \in \mathcal{W}$, defining a

coefficient vector $\theta_g = [g(t_1), \ldots, g(t_{p+1})]^\top$, the piecewise linear formulation of g can be expressed as

$$g(t) = \theta_g^\top A(t), \quad \text{for all } t \in \mathcal{T}, \tag{5.3}$$

where $A(t) = [A_1(t), \ldots, A_{p+1}(t)]^\top$, $A_k(t) = A_k^{(1)}(t) - A_{k+1}^{(2)}(t)$, $A_k^{(1)}(t) = (t - t_{k-1})/(t_k - t_{k-1}) \cdot \mathbf{1}_{[t_{k-1}, t_k)}$, $A_k^{(2)}(t) = (t - t_k)/(t_k - t_{k-1}) \cdot \mathbf{1}_{[t_{k-1}, t_k)}$, for $k = 1, \ldots, p+1$ and $A_{p+2}^{(2)} = 0$. Due to the definition of the warping function space $\mathcal{W}$, the parameter space Θ of the spline coefficient vector θ_g is

$$\Theta = \{\theta \in \mathbb{R}^{p+1} : 0 < \theta_1 < \cdots < \theta_{p+1} = \tau\}. \tag{5.4}$$

The corresponding family of warping functions is $\mathcal{W} = \mathcal{W}_{\mathrm{LS}} = \{g \in \mathcal{W} : g = \theta_g^\top A$ with $\theta_g \in \Theta\}$. We assume that the pairwise warping function $g_{i'i}$ can be represented by (5.3), that is,

$$g_{i'i}(\cdot) = \theta_{g_{i'i}}^\top A(\cdot), \quad \text{with } \theta_{g_{i'i}} \in \Theta. \tag{5.5}$$

5.3. *Samples and estimation.* For each $i = 1, \ldots, n$, suppose available observations for the process Y_i are $\{(T_{ij}, Z_{ij})\}_{j=1}^{m_i}$, where $Z_{ij} = \mathcal{P}_{ij}(Y_i(T_{ij}))$, and $\{(T_{ij}, \mathcal{P}_{ij})\}_{j=1}^{m_i}$ are m_i independent realizations of $(T_i, \mathcal{P}_i)$. To estimate the warping functions h_i, a first step is to estimate the processes Y_i by local Fréchet regression. Specifically, as per (3.4), the estimated trajectories are

$$\widehat{Y}_i(t) = \operatorname*{argmin}_{z \in \mathcal{M}} \frac{1}{m_i} \sum_{j=1}^{m_i} \widehat{v}(T_{ij}, t, b_i) d_{\mathcal{M}}^2(Z_{ij}, z), \quad \text{for all } t \in \mathcal{T}, \tag{5.6}$$

where $\widehat{v}$ is as defined after (3.4) and $b_i = b_i(m_i) > 0$ are bandwidth sequences.

Our next step is to obtain an estimator for the pairwise warping functions $g_{i'i}$ as per (5.5), for any distinct $i', i \in \{1, \ldots, n\}$. This is equivalent to estimating the corresponding spline coefficients $\theta_{g_{i'i}} \in \Theta$, which can be obtained by minimizing the integral of the squared distance between $\widehat{Y}_{i'}$ with time shifted toward $\widehat{Y}_i$ and $\widehat{Y}_i$ over the time domain $\mathcal{T}$, with a regularization penalty on the magnitude of warping. Specifically, an estimator for $\theta_{g_{i'i}}$ is

$$\begin{aligned} \widehat{\theta}_{g_{i'i}} &= \operatorname*{argmin}_{\theta \in \Theta} \mathscr{C}_{\widehat{Y},\lambda}(\theta; \widehat{Y}_{i'}, \widehat{Y}_i), \\ \text{with} \quad \mathscr{C}_{\widehat{Y},\lambda}(\theta; \widehat{Y}_{i'}, \widehat{Y}_i) &= \int_{\mathcal{T}} \big[d_{\mathcal{M}}^2\big(\widehat{Y}_{i'}(\theta^\top A(t)), \widehat{Y}_i(t)\big) + \lambda\big(\theta^\top A(t) - t\big)^2 \big] \mathrm{d}t, \end{aligned} \tag{5.7}$$

whence we obtain an estimator $\widehat{g}_{i'i}$ of the pairwise warping functions

$$\widehat{g}_{i'i}(t) = \widehat{\theta}_{g_{i'i}}^\top A(t), \quad \text{for all } t \in \mathcal{T}. \tag{5.8}$$

By the assumption $\mathbb{E}[h(t)] = t$, we have $\mathbb{E}[g_{i'i}(t) \mid h_i] = \mathbb{E}[h_{i'}(h_i^{-1}(t)) \mid h_i] = h_i^{-1}(t)$, for all $t \in \mathcal{T}$, which justifies estimating the inverse global warping functions h_i^{-1} by

$$\widehat{h}_i^{-1}(t) = n^{-1} \sum_{i'=1}^{n} \widehat{g}_{i'i}(t), \quad \text{for all } t \in \mathcal{T}. \tag{5.9}$$

Hence, estimators $\widehat{h}_i$ for the global warping functions h_i can be obtained by inversion, with estimated aligned trajectories given by $\widehat{Y}_i(\widehat{h}_i(t))$, for $t \in \mathcal{T}$.

5.4. *Asymptotic results for time warping.* In order to obtain the convergence rate for the proposed estimates $\widehat{h}_i$ for the warping functions as per (5.9) based on discrete and noisy observations $\{(T_{ij}, Z_{ij})\}_{j=1}^{m_i}$, an initial step is to derive bounds for the difference between the actual metric space valued processes Y_i and their estimates $\widehat{Y}_i$ as per (5.6), obtained by local Fréchet regression. Specifically, a uniform rate of convergence over the time domain $\mathcal{T}$, beyond the pointwise results shown by Petersen and Müller (2019), is needed. Furthermore, the targets of the local Fréchet regression implemented here are random processes Y_i rather than fixed trajectories as per (2.1). Thus, Theorem 2, where the targets are random processes, needs to be invoked. Subsequently, we derive the rate of convergence for the estimates for warping functions and time synchronized processes.

For any distinct $i', i = 1, \ldots, n$, define functions $\mathscr{C}_\mu(\cdot; h_{i'}, h_i) \colon \mathbb{R}^{p+1} \to \mathbb{R}$,

$$\mathscr{C}_\mu(\theta; h_{i'}, h_i) = \int_{\mathcal{T}} d_{\mathcal{M}}^2\left(\mu\big(h_{i'}^{-1}[\theta^\top A(t)]\big), \mu\big(h_i^{-1}(t)\big)\right) \mathrm{d}t, \quad \theta \in \mathbb{R}^{p+1}. \tag{5.10}$$

We show in Lemma S.2 in the Supplementary Material (Chen and Müller (2022)) that for any distinct $i, i' = 1, \ldots, n$, the coefficient vector $\theta_{g_{i'i}}$ corresponding to the pairwise warping functions $g_{i'i}$ is the unique minimizer of $\mathscr{C}_\mu(\theta; h_{i'}, h_i)$ under certain constraints.

In order to deal with the estimation of n trajectories simultaneously, we make the following assumption on the bandwidths b_i and numbers of discrete observations per trajectory m_i.

(W4) There exist sequences $m = m(n)$ and $b = b(n)$ such that (1) $\inf_{1\le i\le n} m_i \ge m$; (2) $0 < C_1 < \inf_{1\le i\le n} b_i/b \le \sup_{1\le i\le n} b_i/b < C_2 < \infty$, for some constants C_1 and C_2; and (3) $m \to \infty$, $b \to 0$ and $mb^2(-\log b)^{-1} \to \infty$, as $n \to \infty$.

We then derive an asymptotic bound for the discrepancy between the two objective functions $\mathscr{C}_\mu$ and $\mathscr{C}_{\widehat{Y},\lambda}$, whence we obtain the convergence rates for the estimates of the coefficient vector $\theta_{g_{i'i}}$ and the corresponding pairwise warping function in conjunction with Theorem 2 as follows.

THEOREM 3. *Under* (P1), (W1)–(W4), (K0) *and* (U0)–(U3), *for any* $\varepsilon > 0$, *if* $b_i \sim m_i^{-(\beta_1-1)/(2\beta_1+4\beta_2-6+2\varepsilon)}$ *for all* $i = 1, \ldots, n$, *and if* $\lambda \to 0$, *as* $n \to \infty$, *then for any distinct* i' *and* i, *it holds for the constrained minimizer* $\widehat{\theta}_{g_{i'i}}$ *in* (5.7) *that*

$$\|\widehat{\theta}_{g_{i'i}} - \theta_{g_{i'i}}\| = O\big(\lambda^{1/2}\big) + O_P\big(m^{-1/[2(\beta_1+2\beta_2-3+\varepsilon)]}\big), \tag{5.11}$$

where m *is defined in* (W4). *Furthermore, for the corresponding estimate of the pairwise warping function* $\widehat{g}_{i'i}$ *in* (5.8) *it holds that*

$$\sup_{t\in\mathcal{T}} |\widehat{g}_{i'i}(t) - g_{i'i}(t)| = O\big(\lambda^{1/2}\big) + O_P\big(m^{-1/[2(\beta_1+2\beta_2-3+\varepsilon)]}\big). \tag{5.12}$$

We next obtain asymptotic results for local Fréchet regression estimates $\widehat{Y}_i$ across trajectories $i = 1, \ldots, n$ in Corollary 2, which is used in conjunction with Theorem 3 to obtain the convergence rates for the estimates of the warping functions $\widehat{h}_i$, and the aligned trajectories $\widehat{Y}_i(\widehat{h}_i(\cdot))$ in Corollary 3.

COROLLARY 2. *Under* (P1), (W1), (W4), (K0) *and* (U0)–(U3), *for any* $\varepsilon > 0$ *and* $\varepsilon' \in (0, 1)$, *if* $b_i \sim m_i^{-(\beta_1-1)(1-\varepsilon')/(2\beta_1+4\beta_2-6+2\varepsilon)}$ *and* $\limsup_{n\to\infty} nm^{-\varepsilon'(\beta_2-1)/[2(\beta_2-1+\varepsilon/2)]} < \infty$, *it holds that*

$$\sup_{t\in\mathcal{T}} d_{\mathcal{M}}\big(Y_i(t), \widehat{Y}_i(t)\big) = O_P\big(m^{-(1-\varepsilon')/(\beta_1+2\beta_2-3+\varepsilon)}\big), \tag{5.13}$$

for all $i = 1, \ldots, n$;

$$\sup_{t\in\mathcal{T}} n^{-1}\sum_{i=1}^{n} d_{\mathcal{M}}\big(Y_i(t), \widehat{Y}_i(t)\big)^{\alpha} = O_P\big(m^{-\alpha(1-\varepsilon')/(\beta_1+2\beta_2-3+\varepsilon)}\big), \tag{5.14}$$

for any given $\alpha \in (0, 1]$.

COROLLARY 3. *Under* (P1), (W1)–(W4), (K0) *and* (U0)–(U3), *for any* $\varepsilon > 0$ *and for any* $\varepsilon' \in (0, 1)$, *if* $b_i \sim m_i^{-(\beta_1-1)(1-\varepsilon')/(2\beta_1+4\beta_2-6+2\varepsilon)}$ *for all* $i = 1, \ldots, n$ *and* $\limsup_{n\to\infty} nm^{-\varepsilon'(\beta_2-1)/[2(\beta_2-1+\varepsilon/2)]} < \infty$, *and if* $\lambda \to 0$, *as* $n \to \infty$, *it holds for the estimated warping functions* $\widehat{h}_i$ *that*

$$\sup_{t\in\mathcal{T}} |\widehat{h}_i(t) - h_i(t)| = O\big(\lambda^{1/2}\big) + O_P\big(m^{-(1-\varepsilon')/[2(\beta_1+2\beta_2-3+\varepsilon)]}\big) + O_P\big(n^{-1/2}\big). \tag{5.15}$$

Furthermore, if the mean trajectory μ *is Lipschitz* $d_{\mathcal{M}}$*-continuous, that is, there exists* $C_\mu > 0$ *such that* $d_{\mathcal{M}}(\mu(t_1), \mu(t_2)) \le C_\mu |t_1 - t_2|$ *for all* $t_1, t_2 \in \mathcal{T}$, *then it holds for the estimates of the aligned trajectories* $\widehat{Y}_i(\widehat{h}_i(\cdot))$ *that*

$$\begin{aligned}&\sup_{t\in\mathcal{T}} d_{\mathcal{M}}\big(\widehat{Y}_i\big(\widehat{h}_i(t)\big), Y_i\big(h_i(t)\big)\big)\\ &\quad = O\big(\lambda^{1/2}\big) + O_P\big(m^{-(1-\varepsilon')/[2(\beta_1+2\beta_2-3+\varepsilon)]}\big) + O_P\big(n^{-1/2}\big).\end{aligned} \tag{5.16}$$

Defining $\gamma = (1-\varepsilon')^{-1} - 1$, $\varepsilon' \in (0, 1)$ entails $\gamma > 0$. To discuss some more specific rates, under the assumptions of Corollary 3, the minimum number of observations per trajectory m should be bounded below by a multiple of $n^{2[1-(\gamma+1)^{-1}]^{-1}[1+\varepsilon/(2(\beta_2-1))]}$, which implies that the rates in the second terms on the right-hand sides of (5.15)–(5.16) are bounded above by a multiple of $n^{-\gamma^{-1}[(\beta_2-1+\varepsilon/2)/(\beta_2-1)](\beta_1+2\beta_2-3+\varepsilon)^{-1}}$, where the latter can be arbitrarily close to $n^{-1/[\gamma(\beta_1+2\beta_2-3)]}$. Consider $\lambda = O(n^{-1})$. Then, if $\gamma \in (0, 2(\beta_1+2\beta_2-3)^{-1}]$, the estimates for the warping functions h_i and mean trajectory μ as per (5.15)–(5.16) converge with a rate of $n^{-1/2}$. Otherwise, if $\gamma > 2(\beta_1+2\beta_2-3)^{-1}$, the rates in (5.15)–(5.16) can be arbitrarily close to $n^{-1/[\gamma(\beta_1+2\beta_2-3)]}$. Taking $\beta_1 = \beta_2 = 2$ as in Examples 1–2, the estimates $\widehat{h}_i$ and $\widehat{Y}_i(\widehat{h}_i(\cdot))$ achieve the root-n rate when $\gamma = 2(\beta_1+2\beta_2-3)^{-1} = 2/3$ and $m \gtrsim n^{5(1+\varepsilon/2)}$. When $\gamma > 2/3$ and $m \gtrsim n^{(2+\varepsilon)/(1+(\gamma+1)^{-1})}$, the rate becomes approximately $n^{-1/(3\gamma)}$.

6. Data illustrations.

6.1. *Age of minimum connectivity in brain networks*: *fMRI data*. Much work has been done in recent years to investigate how normal aging affects functional connectivity in human brains, which reflects spatial integration of brain activity based on resting-state functional magnetic resonance imaging (rs-fMRI) (Dennis and Thompson (2014), Ferreira and Busatto (2013), Zonneveld et al. (2019)). Fluctuations in regional brain activity are recorded by blood oxygen-level dependent (BOLD) signals while subjects relax. This leads to voxel-specific time series of activation strength. Patterns of subject-specific functional connectivity are frequently analyzed invoking a spatial parcellation of the brain into a set of predefined regions (Bullmore and Sporns (2009)). Connectivity between pairwise brain regions in the parcellation is then usually quantified by what is referred to in the field as temporal Pearson correlation of the fMRI time series of the corresponding regions in neuroimaging. When considering r distinct brain regions, the temporal Pearson correlations then yield correlation matrices in $\mathbb{R}^{r\times r}$, where each row and column represent one brain region, and one such matrix is obtained for each of m subjects, where in the ADNI data that we analyze for each subject one fMRI scan is available.

To study the relationship between age and functional connectivity, it is then natural to apply local Fréchet regression for the case where the random objects that form the responses are situated in the space of correlation matrices and age is a scalar predictor. Based on the correlation matrices, networks of connectivity across regions are constructed by standard procedures in neuroimaging (Rubinov and Sporns (2010)); see also Phillips et al. (2015) and Petersen et al. (2016). The resulting networks can then be converted to graph Laplacians, for which the second smallest eigenvalue is known as the Fiedler value, also referred to as algebraic connectivity (Fiedler (1973)). The Fiedler value is a measure of the global connectivity of a graph that indicates how well connected a network is (Cai et al. (2019), de Haan et al. (2012), Phillips et al. (2015)). Based on the results obtained from local Fréchet regression, we can then express the Fiedler value as a function of age of a subject and identify the age at which the resting human brain attains the minimum level of connectivity. This is of interest to understand the aging brain and as brain connectivity has been reported to mostly decrease during aging while also increases have been reported (Ferreira and Busatto (2013)).

We investigated the dependence of brain connectivity on age for elderly cognitively normal people using the resting-state fMRI data obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (http://adni.loni.usc.edu). The data used in our analysis consist of fMRI scans from $m = 402$ clinically normal elderly subjects at ages ranging from 55.6 to 95.4 years old, where one randomly selected scan is taken for subjects for whom multiple scans are available.

Our analysis focused on the interregional connectivity of $r = 10$ hubs (Buckner et al. (2009), Table 3). Specifically, we considered spherical seed regions of diameter 8 mm centered at the seed voxels of these hubs. Preprocessing of the BOLD signals was implemented by adopting the standard procedures of head motion correction, slice-timing correction, coregistration, normalization and spatial smoothing. Subsequently, average signals of voxels within each seed region were extracted, whence linear detrending and bandpass filtering were performed to account for signal drift and global cerebral spinal fluid and white matter signals, including only frequencies between 0.01 and 0.1 Hz, respectively. These steps were performed in MATLAB using the Statistical Parametric Mapping (SPM12, http://www.fil.ion.ucl.ac.uk/spm) and Resting-State fMRI Data Analysis Toolkit V1.8 (REST1.8, http://restfmri.net/forum/?q=rest).

Let $\{y_{jks}\}_{s=1}^{S}$ be the signal time series of seed region k of subject j excluding the first four time points, which were discarded to eliminate nonequilibrium effects of magnetization, for $k = 1, \dots, r$ and $j = 1, \dots, m$. For subject j, the correlation matrix calculated for analyzing connectivity in fMRI is

$$\mathbf{R}_j = (R_{j,kl})_{1\le k,l\le r}, \qquad R_{j,kl} = \frac{\sum_{s=1}^{S}(y_{jks} - \bar{y}_{jk})(y_{jls} - \bar{y}_{jl})}{[\sum_{s=1}^{S}(y_{jks} - \bar{y}_{jk})^2 \sum_{s=1}^{S}(y_{jls} - \bar{y}_{jl})^2]^{1/2}}. \tag{6.1}$$

In the local Fréchet regression, we used age-at-scan as predictor, and the correlation matrices $\mathbf{R}_j$ as response, taken to be elements in the space of correlation matrices of dimension r equipped with the Frobenius metric, $(\mathcal{M}, d_F)$, as in Example 2.

For any correlation matrix $\mathbf{R} \in \mathcal{M}$, the Fiedler value is the second smallest eigenvalue of the corresponding graph Laplacian matrix

$$L(\mathbf{R}) = D(\mathbf{R}) - A(\mathbf{R}).$$

Here, $A(\mathbf{R}) = (\mathbf{R} - \mathbf{I}_r)_+$ is the adjacency matrix obtained by applying a threshold and setting the diagonal elements to zero, and $D(\mathbf{R}) = \operatorname{diag}\{A(\mathbf{R})\mathbf{1}_r\}$ is the (node) degree matrix, where $\mathbf{I}_r = \operatorname{diag}\{\mathbf{1}_r\}$, $\mathbf{1}_r = (1, \dots, 1)^\top \in \mathbb{R}^r$ and $\mathbf{B}_+ = (\max\{B_{kl}, 0\})_{1\le k,l\le r}$, for any $\mathbf{B} \in \mathbb{R}^{r\times r}$. Then the Fiedler value corresponding to $\mathbf{R}$ is given by a map $\Gamma\colon \mathcal{M} \to \mathbb{R}$,

$$\Gamma(\mathbf{R}) = \lambda_{r-1}\big(L(\mathbf{R})\big),$$

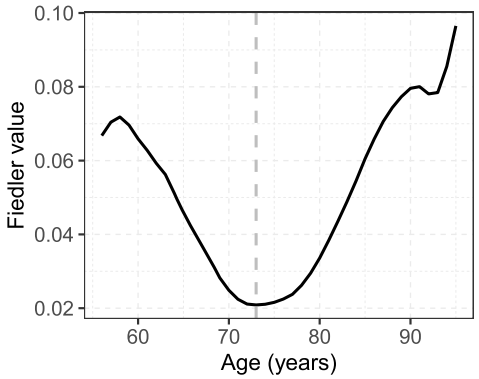

FIG. 1. *Fiedler values as a function of age, corresponding to the local Fréchet regression estimate of the correlation matrix valued conditional Fréchet mean trajectory as per* (6.2), *with the minimum attained at* 73 *years of age marked by a dashed line.*

that yields the $(r-1)$th largest, that is, second smallest eigenvalue of $L(\mathbf{R})$, for any $\mathbf{R} \in \mathcal{M}$. Note that $d_F(L(\mathbf{R}_1), L(\mathbf{R}_2))^2 \le 3d_F(\mathbf{R}_1, \mathbf{R}_2)^2$. In view of the Hoffman–Wielandt inequality (Hoffman and Wielandt (1953)), Γ satisfies (D1) with $C_1 = \sqrt{3}$ and $\alpha_1 = 1$. Applying local Fréchet regression with bandwidth $b = 13.26$, chosen by leave-one-out cross validation, the Fiedler values for the local Fréchet regression estimates $\widehat{\nu}_m(t)$ as per (2.3) of the conditional mean correlation matrix at age t are

$$\widehat{\Lambda}(t) = \Gamma\big(\widehat{\nu}_m(t)\big) = \lambda_{r-1}\big(L\big(\widehat{\nu}_m(t)\big)\big), \quad \text{for } t \in \mathcal{T}. \tag{6.2}$$

Figure 1 displays the trajectory $\widehat{\Lambda}$ of age-varying Fiedler values obtained for the local Fréchet regression estimate of the correlation matrix valued conditional Fréchet mean trajectory according to (6.2), based on the correlation matrices obtained from fMRI scans as per (6.1) for $m = 402$ normal subjects in the ADNI data. A convex pattern can be seen around the minimum of $\widehat{\Lambda}$, which is attained at 73 years of age. While some studies have found that functional connectivity decreases during normal aging processes before 80 years of age (Ferreira and Busatto (2013), Mevel et al. (2013)), we observe for these data that the decrease is reversed for older ages.

We mention here that a limitation of the current approach is that our methods and theoretical justifications are for point estimates only. At this time, we do not have theoretically justified tools for uncertainty quantification such as confidence intervals for the location of the extremum point. The construction of such intervals and other tools for uncertainty quantification will be left for future research.

6.2. *Time warping for distributional trajectories*: *Human mortality data.* There has been perpetual interest in understanding human longevity. One particular goal is to gain an understanding of how the distribution of age-at-death evolves over time. To study this evolution, human mortality data for different countries are available from the Human Mortality Database (http://www.mortality.org/). We consider the calendar time period from 1983 to 2013, for which the mortality data for 28 countries are available throughout the period. It is known that the mortality distributions generally shift to higher ages during this time interval, which reflects increasing longevity. It is then of interest to ascertain which countries move faster and which move slower toward increased longevity, quantified by the rightward shift of the densities of age-at-death.

To address this question, we apply the proposed time warping method in the metric space of probability distributions with the Wasserstein metric, that is, the Wasserstein space

as per Example 1. Before going into the details, we note that we are dealing with time warping for Wasserstein space valued functional data, which is distinct from time warping of density functions, where in this context the latter can be considered a special case of real-valued functional data; the warping of age-at-death density functions at a fixed calendar year would be a different topic. To our knowledge, this is the first time a warping model for metric-space valued random processes is being proposed. For now, due to the absence of algebraic operations, the proposed warping model is limited to an adaptation of the pairwise warping paradigm. In contrast, for real-valued functional data various other competitive time-warping methods are available (Marron et al. (2015)) and it would be an interesting future research topic to extend other warping methods from the real-valued to the metric space-valued scenario and to compare the performance of such extensions.

In 1983, all countries start out with their warping functions taking values at the initial calendar year 1983, and in 2013 they all assume the value at the ending year 2013, so that the warping effect is considered between these two endpoints. A warping function with values below those of the identity function indicates that the country to which it belongs is on an accelerating course towards enhanced longevity, while countries with warping functions assuming values above those of the identity are on a delayed course.

Comparing the estimated warping functions across countries, we found that for males the enhancement in longevity of Japanese from 1983 to 2007 and for Icelanders from 2008 to 2013 accelerates the fastest among all of the 28 countries that we consider between 1983 and 2013, while males have the most delayed increased longevity for Lithuania throughout the period (Figure 2). For females, the movement toward increased longevity is found to be fastest for Japanese women and slowest for Latvian women. The relative delay in increasing longevity for Lithuania and Latvia, former Soviet republics, is likely due to the aftermath of the breakup of the Soviet Union.

The original and aligned trajectories along with the estimated warping functions for two selected groups of countries are demonstrated in Figures 3 and 4. The former group includes representative countries for which the pattern of the estimated warping functions are similar between females and males, while the latter group consists of representative countries for which the estimated warping functions are mismatched between females and males.

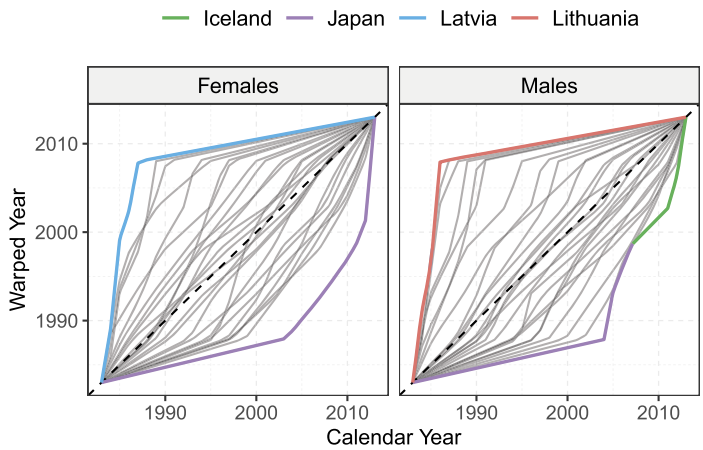

FIG. 2. *Estimated warping functions $\widehat{h}_i$ as per (5.9) for the mortality distribution trajectories for females (left) and males (right) for each country in the sample (grey solid curves), where cross-sectional minimum and maximum warping functions are identified and highlighted in different colors. The black dashed lines represent identity functions.*

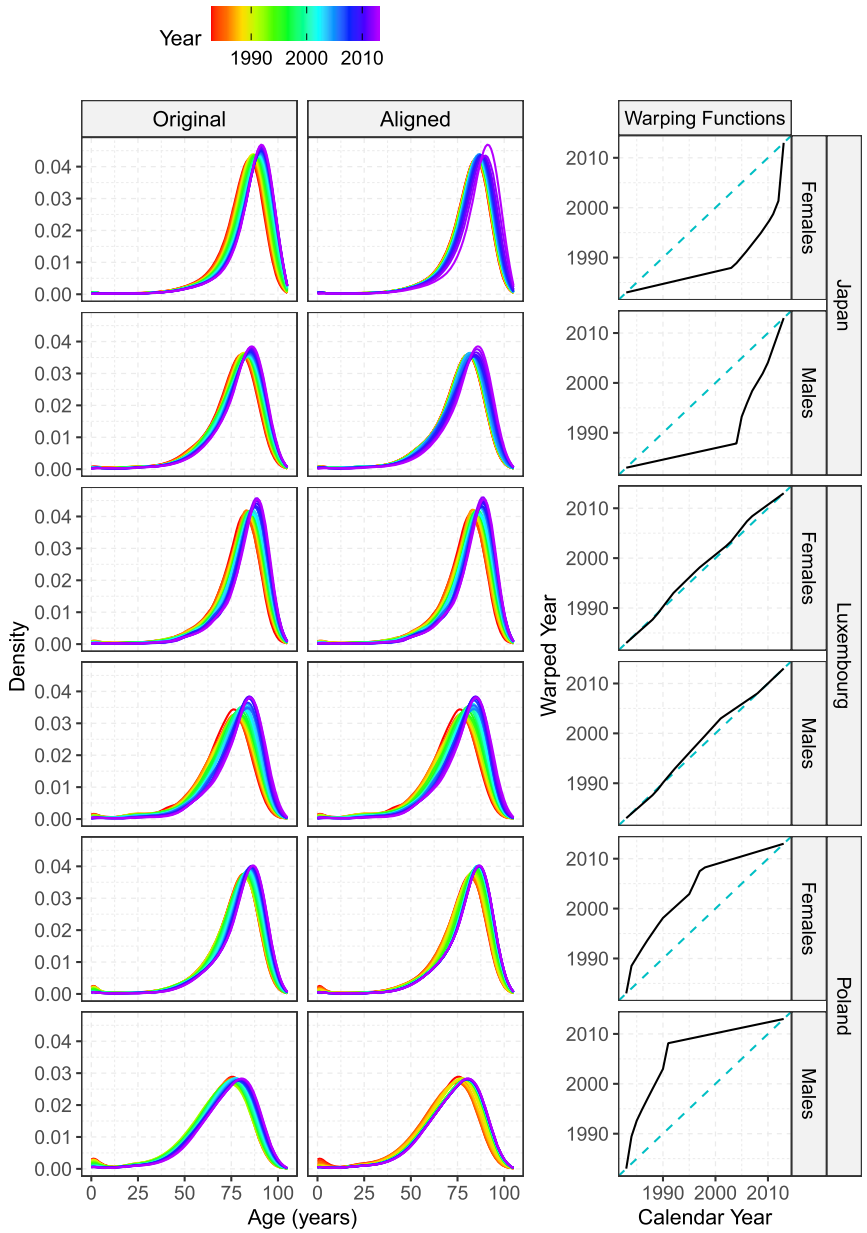

FIG. 3. *Density functions corresponding to the original* (*left*) *and aligned* (*middle*) *trajectories*, $\widehat{Y}_i(\cdot)$ *and* $\widehat{Y}_i(\widehat{h}_i(\cdot))$, *and the estimated warping functions* $\widehat{h}_i$ (*right*) *during* 1983–2013 *for three countries*, *Japan* (*top*), *Luxembourg* (*middle*) *and Poland* (*bottom*), *for which the estimated warping functions for females and males are similar, where* $\widehat{Y}_i$ *and* $\widehat{h}_i$ *are as per* (5.6) *and* (5.9). *The blue dashed lines on the right panels represent identity functions.*

Among the countries shown in Figure 3 with similar warping patterns between males and females, Luxembourg's warping functions are close to the identity and, therefore, its longevity increase represents the average increase across all countries, for both males and females. For Japan, both male and female longevity are strongly accelerated compared to the other countries considered, in contrast to the situation for Poland, where the increase in longevity for both males and females is much delayed relative to the average. In addition, countries shown in Figure 4 exhibit an interesting gender heterogeneity. Both France and Israel show average longevity increase patterns for one gender, namely males in France and

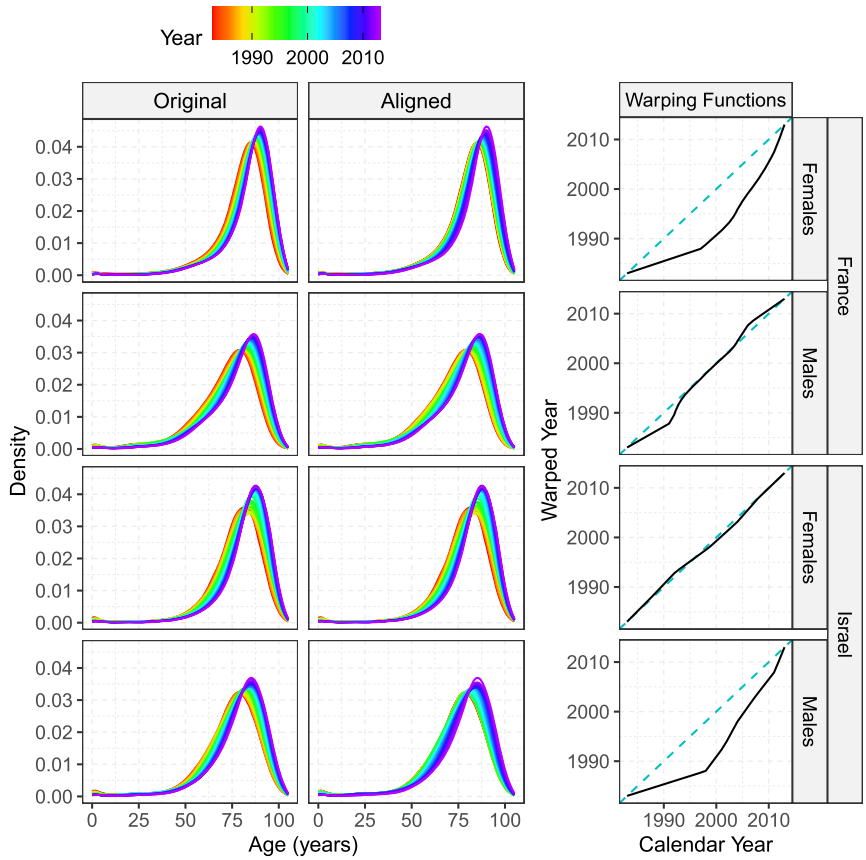

FIG. 4. *Density functions corresponding to the original* (*left*) *and aligned* (*middle*) *trajectories*, $\widehat{Y}_i(\cdot)$ *and* $\widehat{Y}_i(\widehat{h}_i(\cdot))$, *and the estimated warping functions* $\widehat{h}_i$ (*right*) *during* 1983–2013 *for two countries*, *France* (*top*) *and Israel* (*bottom*), *for which the estimated warping functions for females and males differ, where* $\widehat{Y}_i$ *and* $\widehat{h}_i$ *are as per* (5.6) *and* (5.9). *The blue dashed lines on the right panels represent identity functions*.

females in Israel, but not for the other gender, as females in France and males in Israel exhibit accelerated increase in longevity.

Acknowledgments. We wish to thank two anonymous referees, an Associate Editor, and the Editor for their helpful and constructive comments which led to numerous improvements in the paper. Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database. As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators and databases can be found at http://adni.loni.usc.edu

Funding. Research supported in part by NSF Grants DMS-1712864 and DMS-2014626. Funding sources for ADNI are as listed at http://adni.loni.usc.edu.

SUPPLEMENTARY MATERIAL

Proofs and simulation results (DOI: 10.1214/21-AOS2163SUPP; .pdf). The supplement includes proofs and auxiliary lemmas as well as simulation results.

REFERENCES

AGUEH, M. and CARLIER, G. (2011). Barycenters in the Wasserstein space. *SIAM J. Math. Anal.* **43** 904–924. MR2801182 https://doi.org/10.1137/100805741

AMBROSIO, L., GIGLI, N. and SAVARÉ, G. (2004). Gradient flows with metric and differentiable structures, and applications to the Wasserstein space. *Atti Accad. Naz. Lincei Cl. Sci. Fis. Mat. Natur. Rend. Lincei* (9) *Mat. Appl.* **15** 327–343. MR2148889

BALABDAOUI, F., RUFIBACH, K. and WELLNER, J. A. (2009). Limit distribution theory for maximum likelihood estimation of a log-concave density. *Ann. Statist.* **37** 1299–1331. MR2509075 https://doi.org/10.1214/08-AOS609

BELITSER, E., GHOSAL, S. and VAN ZANTEN, H. (2012). Optimal two-stage procedures for estimating location and size of the maximum of a multivariate regression function. *Ann. Statist.* **40** 2850–2876. MR3097962 https://doi.org/10.1214/12-AOS1053

BHATTACHARYA, R. and PATRANGENARU, V. (2003). Large sample theory of intrinsic and extrinsic sample means on manifolds. I. *Ann. Statist.* **31** 1–29. MR1962498 https://doi.org/10.1214/aos/1046294456

BHATTACHARYA, R. and PATRANGENARU, V. (2005). Large sample theory of intrinsic and extrinsic sample means on manifolds. II. *Ann. Statist.* **33** 1225–1259. MR2195634 https://doi.org/10.1214/009053605000000093

BUCKNER, R. L., SEPULCRE, J., TALUKDAR, T., KRIENEN, F. M., LIU, H., HEDDEN, T., ANDREWS-HANNA, J. R., SPERLING, R. A. and JOHNSON, K. A. (2009). Cortical hubs revealed by intrinsic functional connectivity: Mapping, assessment of stability, and relation to Alzheimer's disease. *J. Neurosci.* **29** 1860–1873.

BULLMORE, E. and SPORNS, O. (2009). Complex brain networks: Graph theoretical analysis of structural and functional systems. *Nat. Rev. Neurosci.* **10** 186–198.

CAI, J., LIU, A., MI, T., GARG, S., TRAPPE, W., MCKEOWN, M. J. and WANG, Z. J. (2019). Dynamic graph theoretical analysis of functional connectivity in Parkinson's disease: The importance of Fiedler value. *IEEE J. Biomed. Health Inform.* **23** 1720–1729.

CHEN, Y. and MÜLLER, H.-G. (2022). Supplement to "Uniform convergence of local Fréchet regression with applications to locating extrema and time warping for metric space valued trajectories." https://doi.org/10.1214/21-AOS2163SUPP

DAVIS, B. C., FLETCHER, P. T., BULLITT, E. and JOSHI, S. (2007). Population shape regression from random design data. In 2007 *IEEE* 11*th International Conference on Computer Vision*.

DE HAAN, W., VAN DER FLIER, W. M., WANG, H., VAN MIEGHEM, P. F., SCHELTENS, P. and STAM, C. J. (2012). Disruption of functional brain networks in Alzheimer's disease: What can we learn from graph spectral analysis of resting-state magnetoencephalography? *Brain Connect.* **2** 45–55.

DENNIS, E. L. and THOMPSON, P. M. (2014). Functional brain connectivity using fMRI in aging and Alzheimer's disease. *Neuropsychol. Rev.* **24** 49–62. https://doi.org/10.1007/s11065-014-9249-6

DEVROYE, L. P. (1978). The uniform convergence of nearest neighbor regression function estimators and their application in optimization. *IEEE Trans. Inf. Theory* **24** 142–151. MR0518942 https://doi.org/10.1109/tit.1978.1055865

DRYDEN, I. L., KOLOYDENKO, A. and ZHOU, D. (2009). Non-Euclidean statistics for covariance matrices, with applications to diffusion tensor imaging. *Ann. Appl. Stat.* **3** 1102–1123. MR2750388 https://doi.org/10.1214/09-AOAS249

DUBEY, P. and MÜLLER, H.-G. (2020). Functional models for time-varying random objects. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **82** 275–327. MR4084166

FAN, J. and GIJBELS, I. (1996). *Local Polynomial Modelling and Its Applications. Monographs on Statistics and Applied Probability* **66**. CRC Press, London. MR1383587

FARAGÓ, A., LINDER, T. and LUGOSI, G. (1993). Fast nearest-neighbor search in dissimilarity spaces. *IEEE Trans. Pattern Anal. Mach. Intell.* **15** 957–962.

FERREIRA, L. K. and BUSATTO, G. F. (2013). Resting-state functional connectivity in normal brain aging. *Neurosci. Biobehav. Rev.* **37** 384–400.

FIEDLER, M. (1973). Algebraic connectivity of graphs. *Czechoslovak Math. J.* **23** 298–305. MR0318007

FRÉCHET, M. (1948). Les éléments aléatoires de nature quelconque dans un espace distancié. *Ann. Inst. Henri Poincaré* **10** 215–310. MR0027464

GASSER, T. and KNEIP, A. (1995). Searching for structure in curve samples. *J. Amer. Statist. Assoc.* **90** 1179–1188.

GERVINI, D. and GASSER, T. (2004). Self-modelling warping functions. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **66** 959–971. MR2102475 https://doi.org/10.1111/j.1467-9868.2004.B5582.x

GONG, D. and MEDIONI, G. (2011). Dynamic manifold warping for view invariant action recognition. In *Proceedings of International Conference on Computer Vision* 571–578.

HEIN, M. (2009). Robust nonparametric regression with metric-space valued output. In *Advances in Neural Information Processing Systems* 718–726.

HOFFMAN, A. J. and WIELANDT, H. W. (1953). The variation of the spectrum of a normal matrix. *Duke Math. J.* **20** 37–39. MR0052379

HUCKEMANN, S. F. (2012). On the meaning of mean shape: Manifold stability, locus and the two sample test. *Ann. Inst. Statist. Math.* **64** 1227–1259. MR2981621 https://doi.org/10.1007/s10463-012-0352-2

HUCKEMANN, S. F. (2015). (Semi-)Intrinsic statistical analysis on non-Euclidean spaces. In *Advances in Complex Data Modeling and Computational Methods in Statistics* 103–118. Springer, New York.

JAMES, G. M. (2007). Curve alignment by moments. *Ann. Appl. Stat.* **1** 480–501. MR2415744 https://doi.org/10.1214/07-AOAS127

KLOECKNER, B. (2010). A geometric study of Wasserstein spaces: Euclidean spaces. *Ann. Sc. Norm. Super. Pisa Cl. Sci.* (5) **9** 297–323. MR2731158

KNEIP, A. and GASSER, T. (1992). Statistical tools to analyze data representing a sample of curves. *Ann. Statist.* **20** 1266–1305. MR1186250 https://doi.org/10.1214/aos/1176348769

LE GOUIC, T. and LOUBES, J.-M. (2017). Existence and consistency of Wasserstein barycenters. *Probab. Theory Related Fields* **168** 901–917. MR3663634 https://doi.org/10.1007/s00440-016-0727-z

MACK, Y. P. and SILVERMAN, B. W. (1982). Weak and strong uniform consistency of kernel regression estimates. *Z. Wahrsch. Verw. Gebiete* **61** 405–415. MR0679685 https://doi.org/10.1007/BF00539840

MARRON, J. S., RAMSAY, J. O., SANGALLI, L. M. and SRIVASTAVA, A. (2015). Functional data analysis of amplitude and phase variation. *Statist. Sci.* **30** 468–484. MR3432837 https://doi.org/10.1214/15-STS524

MEVEL, K., LANDEAU, B., FOUQUET, M., LA JOIE, R., VILLAIN, N., MÉZENGE, F., PERROTIN, A., EUSTACHE, F., DESGRANGES, B. and CHÉTELAT, G. (2013). Age effect on the default mode network, inner thoughts, and cognitive abilities. *Neurobiol. Aging* **34** 1292–1301.

MÜLLER, H.-G. (1989). Adaptive nonparametric peak estimation. *Ann. Statist.* **17** 1053–1069. MR1015137 https://doi.org/10.1214/aos/1176347255

PARZEN, E. (1962). On estimation of a probability density function and mode. *Ann. Math. Stat.* **33** 1065–1076. MR0143282 https://doi.org/10.1214/aoms/1177704472

PELLETIER, B. (2006). Non-parametric regression estimation on closed Riemannian manifolds. *J. Nonparametr. Stat.* **18** 57–67. MR2214065 https://doi.org/10.1080/10485250500504828

PETERSEN, A. and MÜLLER, H.-G. (2019). Fréchet regression for random objects with Euclidean predictors. *Ann. Statist.* **47** 691–719. MR3909947 https://doi.org/10.1214/17-AOS1624

PETERSEN, A., ZHAO, J., CARMICHAEL, O. and MÜLLER, H.-G. (2016). Quantifying individual brain connectivity with functional principal component analysis for networks. *Brain Connect.* **6** 540–547.

PHILLIPS, D. J., MCGLAUGHLIN, A., RUTH, D., JAGER, L. R., SOLDAN, A. and ADNI (2015). Graph theoretic analysis of structural connectivity across the spectrum of Alzheimer's disease: The importance of graph creation methods. *NeuroImage Clin.* **7** 377–390.

RAMSAY, J. O. and LI, X. (1998). Curve registration. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **60** 351–363. MR1616045 https://doi.org/10.1111/1467-9868.00129

RUBINOV, M. and SPORNS, O. (2010). Complex network measures of brain connectivity: Uses and interpretations. *NeuroImage* **52** 1059–1069.

RUPPERT, D. and WAND, M. P. (1994). Multivariate locally weighted least squares regression. *Ann. Statist.* **22** 1346–1370. MR1311979 https://doi.org/10.1214/aos/1176325632

SAKOE, H. and CHIBA, S. (1978). Dynamic programming algorithm optimization for spoken word recognition. *IEEE Trans. Acoust. Speech Signal Process.* **26** 43–49.

SILVERMAN, B. W. (1978). Weak and strong uniform consistency of the kernel estimate of a density and its derivatives. *Ann. Statist.* **6** 177–184. MR0471166

STEINKE, F. and HEIN, M. (2009). Non-parametric regression between manifolds. In *Advances in Neural Information Processing Systems* 1561–1568.

STEINKE, F., HEIN, M. and SCHÖLKOPF, B. (2010). Nonparametric regression between general Riemannian manifolds. *SIAM J. Imaging Sci.* **3** 527–563. MR2736019 https://doi.org/10.1137/080744189

STURM, K.-T. (2003). Probability measures on metric spaces of nonpositive curvature. In *Heat Kernels and Analysis on Manifolds, Graphs, and Metric Spaces* (*Paris*, 2002). *Contemp. Math.* **338** 357–390. Amer. Math. Soc., Providence, RI. MR2039961 https://doi.org/10.1090/conm/338/06080

TANG, R. and MÜLLER, H.-G. (2008). Pairwise curve synchronization for functional data. *Biometrika* **95** 875–889. MR2461217 https://doi.org/10.1093/biomet/asn047

TRIGEORGIS, G., NICOLAOU, M. A., SCHULLER, B. W. and ZAFEIRIOU, S. (2018). Deep canonical time warping for simultaneous alignment and representation learning of sequences. *IEEE Trans. Pattern Anal. Mach. Intell.* **40** 1128–1138. https://doi.org/10.1109/TPAMI.2017.2710047

VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes. Springer Series in Statistics*. Springer, New York. MR1385671 https://doi.org/10.1007/978-1-4757-2545-2

VIEU, P. (1996). A note on density mode estimation. *Statist. Probab. Lett.* **26** 297–307. MR1393913 https://doi.org/10.1016/0167-7152(95)00024-0

VU, H. T., CAREY, C. and MAHADEVAN, S. (2012). Manifold warping: Manifold alignment over time. In *Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence* 1155–1161.

WANG, K. and GASSER, T. (1997). Alignment of curves by dynamic time warping. *Ann. Statist.* **25** 1251–1276. MR1447750 https://doi.org/10.1214/aos/1069362747

YUAN, Y., ZHU, H., LIN, W. and MARRON, J. S. (2012). Local polynomial regression for symmetric positive definite matrices. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **74** 697–719. MR2965956 https://doi.org/10.1111/j.1467-9868.2011.01022.x

ZONNEVELD, H. I., PRUIM, R. H. R., BOS, D., VROOMAN, H. A., MUETZEL, R. L., HOFMAN, A., ROMBOUTS, S. A. R. B., VAN DER LUGT, A., NIESSEN, W. J. IKRAM, M. A. and VERNOOIJAB, M. W. (2019). Patterns of functional connectivity in an aging population: The Rotterdam study. *NeuroImage* **189** 432–444.