



A stochastic process representation for time warping functions

Yijia Ma^{*}, Xinyu Zhou, Wei Wu

Department of Statistics, Florida State University, 117 N. Woodward Ave., Tallahassee, FL 32306-4330, United States of America

ARTICLE INFO

Keywords:

Functional boxplot
Hilbert space
Isometric isomorphism
Stochastic process
Time warping functions
Penalized function registration

ABSTRACT

Time warping function provides a mathematical representation to measure phase variability in functional data. Recent studies have developed various approaches to estimate optimal warping between functions. However, a principled, linear, generative representation on time warping functions is still under-explored. This is highly challenging because the warping functions are non-linear in the conventional \mathbb{L}^2 space. To address this problem, a new linear warping space is defined and a stochastic process representation is proposed to characterize time warping functions. The key is to define an inner-product structure on the time warping space, followed by a transformation which maps the warping functions into a sub-space of the \mathbb{L}^2 space. With certain constraints on the warping functions, this transformation is an isometric isomorphism. In the transformed space, the \mathbb{L}^2 basis in the Hilbert space is adopted for representation, which can be easily utilized to generate time warping functions by using different types of stochastic process. The effectiveness of this representation is demonstrated through its use as a new penalty in the penalized function registration, accompanied by an efficient gradient method to minimize the cost function. The new penalized method is illustrated through simulations that properly characterize nonuniform and correlated constraints in the time domain. Furthermore, this representation is utilized to develop a boxplot for warping functions, which can estimate templates and identify warping outliers. Finally, this representation is applied to a Covid-19 dataset to construct boxplots and identify states with outlying growth patterns.

1. Introduction

Temporal phase variability has been a central topic in the field of functional data analysis. In function registration or alignment, the goal is often to separate phase and amplitude variabilities, where the phase variation is represented using a time warping function. In many studies, it is critically important to find the aligned functions, because warping is considered as a nuisance variable in the measurement process, and its variability needs to be removed (Ramsay and Silverman, 2006). However, in other cases, phase is considered an essential and critical feature in the data (Marron et al., 2015). For either purpose, one needs to estimate optimal time warpings to align functional observations properly. A common space of warping functions is defined as $\Gamma = \{\gamma : [0, 1] \rightarrow [0, 1] | \gamma(0) = 0, \gamma(1) = 1, 0 < \dot{\gamma}(t) < \infty\}$ ($\dot{\gamma}$ denotes the derivative of γ) (Srivastava et al., 2011), which is a nonlinear space under the conventional \mathbb{L}^2 metric. Over the past two-to-three decades, various approaches have been developed for robust and efficient estimations. Early approaches formulated a least-square problem by representing warping function with a linear combination of B-spline basis functions, and the warping can be obtained by estimating the corresponding coefficients (Ramsay and Li, 1998; Gervini and Gasser, 2004; James, 2007; Eilers, 2004). Recent approaches conducted registration by minimizing the Fisher-Rao metric

^{*} Corresponding author.

E-mail address: ym19f@fsu.edu (Y. Ma).

<https://doi.org/10.1016/j.csda.2024.107941>

Received 27 September 2022; Received in revised form 13 February 2024; Accepted 14 February 2024

Available online 20 February 2024

0167-9473/Â© 2024 Elsevier B.V. All rights reserved.

(Srivastava et al., 2011; Wu and Srivastava, 2014), or resampling with Bayesian registration (Cheng et al., 2014, 2016; Kurtek, 2017; Lu et al., 2017). In particular, there have been attempts to analyze the phase variations for functional regression (Hadjipantelis et al., 2014; Gervini, 2015), classification (Tucker et al., 2013) and functional PCA (Lee and Jung, 2016; Happ et al., 2019).

We point out that most of these studies focused on estimating optimal warping function for alignment, but not building a probabilistic model on it. Studies in Bayesian registration have examined statistical models on time warping, whereas the methods mainly focus on the simplified Dirichlet distribution (Cheng et al., 2016) or Gaussian process in the tangent space of the nonlinear Hilbert unit sphere (Kurtek, 2017; Lu et al., 2017). Because of this non-linearity, the modeling of time warping is still very challenging in the field. To explore the benefits in a linear, inner product space, we will at first define linear and inner-product operations in the warping functions. Our goal is to build an isometric isomorphism to transform the warping space to an \mathbb{L}^2 sub-space and then adopt stochastic process such as Gaussian process to represent the transformed functions.

To achieve this goal, we at first use the fact that a warping function in Γ can be represented as a probability density function by simply taking its derivative. In addition, it is well known that there exists an isometric isomorphism between the density space and a sub-space in \mathbb{L}^2 via the Centered Logratio transformation in the notion of Bayes Hilbert Spaces (Egozcue et al., 2006). Based on these results, we propose to represent the time warping function in three steps: 1) transform warping function to density function, 2) transform density function to \mathbb{L}^2 sub-space, and 3) develop a stochastic process model in the \mathbb{L}^2 sub-space. A similar idea was explored in (Happ et al., 2019), where the focus was on the functional principal component analysis (fPCA) and joint-modeling phase and amplitude components. Our proposed representation adopts this framework, whereas our goal is to build a principled stochastic process on time warping functions.

Our representation provides a linear, inner-product system specifically on the time warping functions. It has the following four apparent features: Firstly, it has a principled theoretical foundation presented in an explicitly generative form (Egozcue et al., 2006), allowing for the straightforward sampling of various time warping functions using pre-specified or data-driven (fPCA) basis elements (Happ et al., 2019). Secondly, the proposed representation can be utilized as a new penalty in the penalized function registration. Thirdly, the proposed representation can be employed to construct functional bagplots for warping functions, facilitating the identification of outliers. Fourthly, we can conduct other statistical analyses such as ANOVA or regressions on the original data by building one-to-one mapping from the time warping space to a conventional \mathbb{L}^2 space.

The rest of this manuscript is organized as follows: In Section 2, we at first define a space that contains only warping functions with bounded derivative and show it is isometrically isomorphic to an inner-product space by applying the Centered Logratio transformation to the derivative of the warping function. We then extend the inner-product space to a Hilbert space so that we can build a stochastic-process-based representation with bounded Hilbert basis functions. In addition, we describe how to estimate a model from observations. In Section 3, we present a new approach for penalized function registration with a new penalty by using our representation, and illustrate the method with simulations. Based on the new representation, in Section 4, we propose to construct a new boxplot for warping functions that can identify the shape outliers and then compare with the-state-of-art methods. A real-world application is given in Section 5. Finally, we summarize our work in Section 6. All mathematical details, as well as extra data analyses and illustrations, are given in the appendices. The programming scripts are deposited at a public repository (<https://github.com/YMa2019/warping>).

2. Methods

2.1. Warping functions with bounded derivatives

Time warping functions have been studied extensively in the literature, and a common space for all warpings in a finite domain $[0, 1]$ is defined as

$$\Gamma = \{ \gamma : [0, 1] \rightarrow [0, 1] | \gamma(0) = 0, \gamma(1) = 1, 0 < \dot{\gamma}(t) < \infty \}. \tag{1}$$

A simple function $\gamma(t) = t^\alpha$, $\alpha > 0$, is often used as a time warping example. See Fig. 1(a) for a few example curves. This paper aims to provide a stochastic-process-based representation on those warping functions. To make the mathematical representation feasible, we need to provide certain basic assumptions on the process. For example, processes with well-defined mean and covariance are referred to as second-order processes (Hsing and Eubank, 2015). Therefore, we often assume the process is second-order (such as a Gaussian process) so that the standard covariance-based methods can be adopted. However, $\dot{\gamma}$ in Equation (1) is simply positive without any other constraints, which makes it challenging to develop an appropriate model.

One simple and effective solution is to provide a lower and an upper bound on the derivative function, and this can lead to finite integrations such as the \mathbb{L}^p norms. That is, we can study time warping functions in the following domain:

$$\Gamma_1 = \{ \gamma : [0, 1] \rightarrow [0, 1] | \gamma(0) = 0, \gamma(1) = 1, 0 < m_\gamma < \dot{\gamma}(t) < M_\gamma < \infty \}. \tag{2}$$

In this new domain the two bounds m_γ and M_γ vary with respect to the function γ . One typical example in Γ_1 is $\gamma(t) = \frac{e^{at} - 1}{e^a - 1}$, with $a \neq 0$. A few example curves of this type of warping is given in Fig. 1(b). We point out that $\gamma(t) = t^\alpha$ is not in Γ_1 . Note that the derivative of warping function $\dot{\gamma}$ is essentially a probability density function on $[0, 1]$, and Γ_1 is a group with function composition. Motivated by the Centered Logratio (CLR) transformation between a density space and the \mathbb{L}^2 space (Leonard, 1978; Egozcue et al., 2006; Menafoglio and Hron, 2014; Menafoglio et al., 2016), we aim to build an isometric isomorphism to transform Γ_1 to a proper \mathbb{L}^2 subspace. At first, we need to build an inner-product structure on Γ_1 .

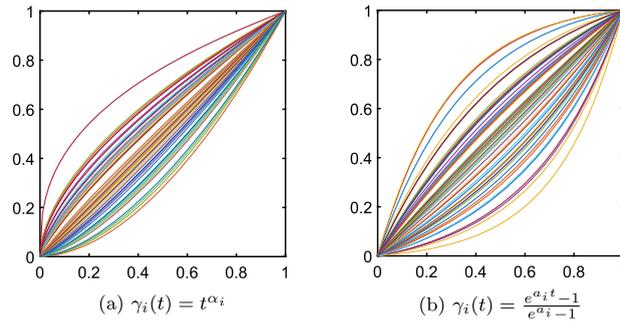


Fig. 1. Simulations of 50 warping functions using two typical methods, respectively. (a) $\gamma_i(t) = t^{\alpha_i}$ with $\alpha_i \sim \Gamma(5, 0.2)$, i.e., Gamma distribution with mean 1 and variance 0.2. (b) $\gamma_i(t) = \frac{e^{\alpha_i t} - 1}{e^{\alpha_i} - 1}$ with $\alpha_i \sim N(0, 4)$, i.e. normal distribution with mean 0 and variance 4.

It is apparent that Γ_1 has constraints and is not even a vector space under the conventional \mathbb{L}^2 metric. In this paper, we propose to define perturbation, power, and inner-product operations to make Γ_1 an inner-product space. For $f, g \in \Gamma_1$ and $\alpha \in \mathbb{R}$, the perturbation with operator $\oplus_\Gamma : \Gamma_1 \times \Gamma_1 \rightarrow \Gamma_1$ is given by

$$[f \oplus_\Gamma g](t) = \frac{\int_0^t \dot{f}(s)\dot{g}(s) ds}{\int_0^1 \dot{f}(\tau)\dot{g}(\tau) d\tau}.$$

The power operation with operator $\odot_\Gamma : \mathbb{R} \times \Gamma_1 \rightarrow \Gamma_1$ is given by:

$$[\alpha \odot_\Gamma f](t) = \frac{\int_0^t \dot{f}^\alpha(s) ds}{\int_0^1 \dot{f}^\alpha(\tau) d\tau}.$$

In addition, the inner product is defined as the functional $\langle \cdot, \cdot \rangle_\Gamma : \Gamma_1 \times \Gamma_1 \rightarrow \mathbb{R}$ in the following form:

$$\langle f, g \rangle_\Gamma = \int_0^1 \log(\dot{f}(t)) \log(\dot{g}(t)) dt - \int_0^1 \log(\dot{f}(s)) ds \int_0^1 \log(\dot{g}(t)) dt. \tag{3}$$

With the inner-product given above, the associated norm and metric distance can be easily defined. An illustrative example of the concepts of warping norm and inner-product is presented in Appendix A. Based on the CLR transformation result in (Egozcue et al., 2006), we select a \mathbb{L}^2 subspace under the conventional \mathbb{L}^2 norm in the following form:

$$H(0, 1) = \left\{ h \in \mathbb{L}^2([0, 1]) \mid \int_0^1 h(t) dt = 0, -\infty < m_h < h(t) < M_h < \infty \right\}. \tag{4}$$

It is easy to see that $H(0, 1)$ is a subspace of $\mathbb{L}^2([0, 1])$. The main result between Γ_1 and $H(0, 1)$ is given in the following theorem:

Theorem 2.1. Given the mapping $\psi_B : \Gamma_1 \rightarrow H(0, 1)$:

$$h(t) = \psi_B(\gamma)(t) = \log(\dot{\gamma}(t)) - \int_0^1 \log(\dot{\gamma}(s)) ds, \tag{5}$$

the space $H(0, 1)$ and Γ_1 are isometric isomorphism (under the linear and inner-product operations). In particular, the inverse mapping $\psi_B^{-1} : H(0, 1) \rightarrow \Gamma_1$ is given by:

$$\gamma(t) = \psi_B^{-1}(h)(t) = \frac{\int_0^t \exp(h(s)) ds}{\int_0^1 \exp(h(\tau)) d\tau}.$$

2.2. Extension to a Hilbert space

Hilbert space, a.k.a. complete inner-product space, is a natural extension of finite Euclidean spaces to the infinite-dimensional case. Because the space is complete, all limiting operations are closed, and techniques in calculus can be directly used. In this paper, we focus on using a stochastic process to provide a linear framework for time warping functions, where a key step is to transform the warping functions into a space with the conventional \mathbb{L}^2 metric, and then the orthonormal basis representations can be fully exploited.

However, we can see that the space $H(0, 1)$ defined in Equation (4) contains only bounded functions, and is therefore not a Hilbert space. In this section, we aim to extend it to a Hilbert space in the following form:

$$E(0, 1) = \left\{ h \in \mathbb{L}^2([0, 1]) \mid \int_0^1 h(t) dt = 0 \right\}. \tag{6}$$

It is easy to verify that $E(0, 1)$ is indeed the smallest Hilbert space containing the space $H(0, 1)$. Basically, we just remove the lower bound m_γ and upper bound M_γ in Equation (4).

As $E(0, 1)$ is also a subspace of $\mathbb{L}^2([0, 1])$, we can at first find a complete orthonormal system for $\mathbb{L}^2([0, 1])$:

$$\left\{ \phi_0(t) = 1, \phi_{2j-1}(t) = \sqrt{2} \sin(2j\pi t), \phi_{2j}(t) = \sqrt{2} \cos(2j\pi t), j \geq 1, t \in [0, 1] \right\}.$$

In Equation (6), the only constraint is that $\int_0^1 \phi(t) dt = 0$. Therefore, by removing the constant term $\phi_0(t) = 1$, we obtain the complete orthonormal system in the space $E(0, 1)$ as follows:

$$B = \left\{ \phi_{2j-1}(t) = \sqrt{2} \sin(2j\pi t), \phi_{2j}(t) = \sqrt{2} \cos(2j\pi t), j \geq 1, t \in [0, 1] \right\}. \tag{7}$$

We also extend the warping space Γ_1 in Equation (2) to the following form:

$$\Gamma_2 = \left\{ \gamma : [0, 1] \rightarrow [0, 1] \mid \dot{\gamma} \in \mathcal{F} \right\}, \tag{8}$$

where the \mathcal{F} space is an extended probability density function space in (Egozcue et al., 2006) and given as

$$\mathcal{F} = \left\{ f : [0, 1] \rightarrow \mathbb{R} \mid f > 0, \log f \in \mathbb{L}^2([0, 1]) \right\}. \tag{9}$$

It was shown in (Van den Boogaart et al., 2014) that under the CLR transformation from \mathcal{F} to $E(0, 1) : clr(f) = \log(f) - \int_0^1 \log(f(s)) ds$, \mathcal{F} and $E(0, 1)$ are isometric isomorphism and the inverse transformation is given by

$$clr^{-1}(h) = \begin{cases} \frac{\exp(h)}{\int_0^1 \exp(h(s)) ds}, & \text{if } \int_0^1 \exp(h(s)) ds < \infty \\ \exp(h), & \text{o.w.} \end{cases}.$$

Thus, if $\{\phi_j\}_{j \geq 1}$ is a set of bounded complete orthonormal basis functions in $E(0, 1)$ (e.g. Set B in Equation (7)), then $\{\psi_j\}_{j \geq 1}$, with $\psi_j = \frac{\exp[\phi_j]}{\int_0^1 \exp[\phi_j(s)] ds}$, is also a complete orthonormal basis for \mathcal{F} .

It is easy to see that Γ_2 contains Γ_1 . However, we point out that a derivative operation is not a bijective mapping between Γ_2 and \mathcal{F} . For any $\gamma \in \Gamma_2$, we have $\dot{\gamma} \in \mathcal{F}$. However, for any $f \in \mathcal{F}$, if $\int_0^1 f(t) dt = \infty$, then there will not be a $\gamma \in \Gamma_2$ such that $f = \dot{\gamma}$. Moreover, the linear operations in \mathcal{F} cannot be directly used in Γ_2 . This is obvious because the product of two density functions may not have a finite integration value on $[0, 1]$.

2.3. Stochastic process representation for time warpings

In Section 2.2, we have extended the bounded inner-product space $H(0, 1)$ to a Hilbert space $E(0, 1)$. We also show that using the CLR transformation, the density function space \mathcal{F} is isometrically isomorphic to the $E(0, 1)$ space. This implies we can build a systematical representation on space $E(0, 1)$, and then use the inverse CLR transformation to project it back into space \mathcal{F} .

2.3.1. Characterize time warping via second-order stochastic process

In this section, we will develop a new procedure to characterize time warping functions in Γ_1 . By the isometric isomorphism, we only need to represent functions in $H(0, 1)$. This representation process is based on a second-order stochastic process in the Hilbert space $E(0, 1)$ (smallest extension of $H(0, 1)$).

In the Karhunen-Loève expansion (Hsing and Eubank, 2015), any mean-square continuous stochastic process can be represented as an infinite linear combination of a set orthonormal basis in $\mathbb{L}^2([0, 1])$, where the basis functions are eigenfunctions for the covariance kernel $K(s, t)$. According to Mercer's Theorem (Hsing and Eubank, 2015), the kernel function must meet the requirements of being continuous, symmetric, and non-negative definite. Therefore, our focus is on developing a suitable kernel function that satisfies these requirements to properly utilize the second-order process. In principle, any complete orthonormal basis can be used. For example, we may adopt the bounded Fourier set B in Equation (7) to characterize bounded function in $h(0, 1)$. The construction of the kernel is given in the following proposition, where the detailed proof is given in Appendix B.

Proposition 2.2. For any non-negative sequence $\{\mu_i\}_{i=1}^\infty$ such that $\sum_{i=1}^\infty \mu_i < \infty$, let

$$K(s, t) = \sum_{i=1}^\infty \mu_i \phi_i(s) \phi_i(t), \quad \text{for all } s, t \in [0, 1],$$

where $\{\phi_i\}_{i=1}^\infty$, given in Equation (7), is the complete orthonormal system in $E(0, 1)$. Then 1) K converges absolutely and uniformly, and 2) K is a continuous, symmetric, non-negative definite function.

Remark. By Proposition 2.2, we can easily construct a kernel function using convergent sequence $\{\mu_i\}$ and basis $\{\phi_i\}$. Note that $\{\mu_i\}$ do not necessarily follow the decreasing order. By the uniqueness of eigenvalues $\{\lambda_i\}$ in Mercer’s theorem (Hsing and Eubank, 2015), $\{\lambda_i\}$ are in fact the ordered sequence (from large to small) of $\{\mu_i\}$. In practical use, common choices of $\{\mu_i\}$ are $\mu_i = \frac{1}{i^s}$, or $\frac{1}{i(\log(i))^s}$, with $s \geq 2$.

Based on the above result, we can simulate a random process in $E(0, 1)$ as follows: Given the orthonormal basis $B = \{\phi_i(t)\}$ and non-negative sequence $\{\mu_i\}_{i=1}^\infty$ with convergent sum, we can generate a mean-centered second-order process X in the following form:

$$X(t) = \sum_{i=1}^\infty G_i \phi_i(t), \tag{10}$$

where G_i are uncorrelated random variables with mean 0 and variance μ_i . Note that there is no constraint for the type of distribution on G_i , which fully characterizes the randomness in $X(t)$. We can choose any distribution to explore all possible variabilities. In particular, to generate a Gaussian process, we only need to set $G_i \sim N(0, \mu_i)$.

In practice, we can only simulate a second-order stochastic process $X(t)$ in Equation (10) with finite m terms in the sum. That is, $\mu_i = 0$ and $G_i = 0$ when $i > m$. In this case, it is always true that $\sum_{i=1}^\infty \mu_i < \infty$ and the corresponding covariance kernel K is well-defined. This truncated version can be written as:

$$X_m(t) = \sum_{i=1}^m G_i \phi_i(t). \tag{11}$$

As ϕ_i are all bounded functions, $X_m(t)$ is also bounded and therefore in $H(0, 1)$. By the isometric isomorphism between $H(0, 1)$ and Γ_1 , we can transform $X_m(t)$ to build a warping function. In summary, the generative procedure for time warping function in Γ_1 is given in Algorithm 1.

Algorithm 1 Generative procedure for warping function in Γ_1 .

Require: Basis functions $\{\phi_i\}_{i=1}^m$ in Equation (7).

Generate coefficient sequence G_i with any probability distribution with mean 0 and variance $\mu_i, i = 1, \dots, m$.

$$X_m(t) = \sum_{i=1}^m G_i \phi_i(t).$$

$$\gamma_m(t) = \frac{\int_0^1 \exp(X_m(s)) ds}{\int_0^1 \exp(X_m(\tau)) d\tau}.$$

Output γ_m .

Remark. If we allow $m = \infty$ in Algorithm 1, then we will need to add two conditions to simulate warping function: 1) $\sum_{i=1}^\infty \mu_i < \infty$, and 2) $\int_0^1 \exp(X_m(s)) ds < \infty$. In this case, the simulated warping function may not be in Γ_1 (i.e., bounded), but will be in Γ_2 as defined in Equation (8).

2.3.2. Illustration

We now illustrate Algorithm 1 with $m = 20$, where the coefficients $\{G_i\}_{i=1}^{20}$ are from each of the following 5 different distributions:

- (a) $G_i \sim N(\mu_i, \sigma_i^2)$, where $\mu_i = 0$, and $\sigma_i = \frac{1}{i}$, i.e., normal distribution with mean 0 and variance $\frac{1}{i^2}$.
- (b) $G_i \sim La(\mu_i, b_i)$, where $\mu_i = 0$, $b_i = \frac{1}{\sqrt{2i}}$, i.e., Laplacian distribution with mean 0 and variance $\frac{1}{i^2}$.
- (c) $G_i \sim U(a_i, b_i)$, where $a_i = -\frac{\sqrt{3}}{i}$, $b_i = \frac{\sqrt{3}}{i}$, i.e., uniform distribution with mean 0 and variance $\frac{1}{i^2}$.
- (d) $G_i \sim N(\mu_i, \sigma_i^2)$, where $\mu_i = 0$, and $\sigma_i = \frac{1}{2i}$, i.e., normal distribution with mean 0 and variance $\frac{1}{(2i)^2}$.
- (e) $G_i \sim N(\mu_i, \sigma_i^2)$, where $\mu_i = 0$, and $\sigma_i = \frac{1}{5i}$, i.e., normal distribution with mean 0 and variance $\frac{1}{(5i)^2}$.

In each of these 5 cases, we generate 10 stochastic processes. The results are shown in Fig. 2. It is easy to see that the simulated warping functions have more variabilities than the previous example $\gamma(t) = \frac{e^{at}-1}{ea-1}$ in Fig. 1(b). The first three columns show time warpings and their corresponding functions in $H(0, 1)$ from one Gaussian process (Column (a)) and two non-Gaussian processes (Columns (b) and (c)). The warping functions exhibit different types of variabilities, whereas the degrees of warping look similar as they share the same variances for the coefficients G_i . As a comparison, we also show two other Gaussian processes ((Columns (d) and (e)) with smaller variances. It is obvious that when the variance gets smaller, the corresponding warping functions stay closer to the identity warping function $\gamma_{id}(t) = t$.

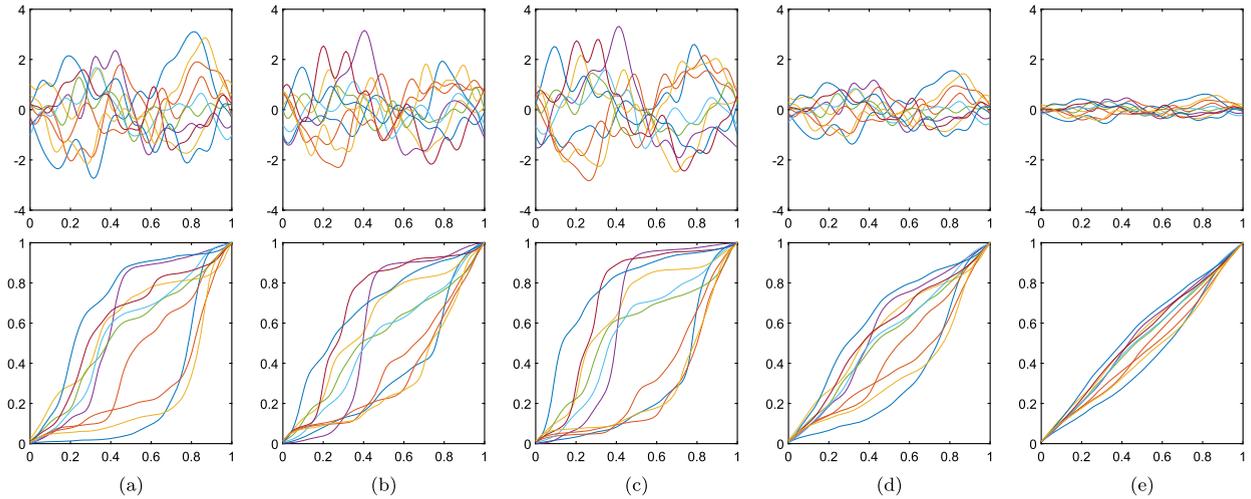


Fig. 2. Simulation examples using Algorithm 1, with 10 stochastic processes in $H(0, 1)$ in the top row and the corresponding 10 time warping functions in Γ_1 in the bottom row. Column (a): $G_i \sim N(0, (1/i)^2)$, Column (b) $G_i \sim La(0, 1/(\sqrt{2}i))$, Column (c) $G_i \sim U(-\sqrt{3}/i, \sqrt{3}/i)$, Column (d) $G_i \sim N(0, (1/(2i))^2)$, Column (e) $G_i \sim N(0, (1/(5i))^2)$.

2.4. Estimation of basis functions via principal component analysis

In Section 2.3, we have developed a stochastic process representation to represent the time warping functions. This representation is based on a complete orthonormal system in the \mathbb{L}^2 space. In practice, we may look for an alternative basis with the given observations. In this section, we will explore modeling with the functional principal component analysis (fPCA) method.

2.4.1. Modeling and resampling via fPCA

fPCA is a basis representation method in the \mathbb{L}^2 space. When time warping observations in Γ_1 are given, we may transform them (stated in Theorem 2.1) into the $H(0, 1)$ space and then estimate orthonormal basis via the fPCA method. This fPCA method has been exploited in (Happ et al., 2019), where the goal was dimension reduction on functional data. In this paper, we will extend the idea to model and resample warping functions. The detailed process is given in Algorithm 2.

Algorithm 2 Modeling and resampling with fPCA.

Require: N observed warping functions γ_n in Γ_1 .

Transform the warping functions γ_n in to $H(0, 1)$: $X_n(t) = \log(\gamma_n(t)) - \int_0^1 \log(\gamma_n(s)) ds$.

Calculate the mean $\hat{\mu}(t) = \frac{1}{N} \sum_{n=1}^N X_n(t)$ and the covariance $\hat{K}(s, t) = \frac{1}{N-1} \sum_{n=1}^N (X_n(s) - \hat{\mu}(s))(X_n(t) - \hat{\mu}(t))$.

Apply spectral decomposition to \hat{K} to get the eigen sequence $\{(\lambda_i, e_i)\}_{i=1}^N$.

Find cutoff threshold δ , let $m = \max\{i | \lambda_i > \delta\}$.

for $k = 1 : m$ **do**

 Calculate the coefficients $Z_{nk} = \int_0^1 (X_n(t) - \hat{\mu}(t))e_k(t) dt, n = 1, 2, \dots, N$.

 Use the sample $\{Z_{nk}\}_{n=1}^N$ to estimate their distribution D_k .

end for

for $r = 1 : R$ **do** (resample warping functions)

 Simulate coefficient G_k using the estimated distribution D_k .

$X_r(t) = \hat{\mu}(t) + \sum_{k=1}^m G_k e_k(t)$.

$\gamma_r(t) = \frac{\int_0^t \exp(X_r(s)) ds}{\int_0^1 \exp(X_r(\tau)) d\tau}$.

end for

Output $\{\gamma_r(t)\}_{r=1}^R$.

Remark. We simplify the resampling process in Algorithm 2 by assuming the coefficients $\{G_i\}_{i=1}^N$ are independent and then generate samples independently. This is true if the process is a Gaussian process. However, in the framework of Karhunen-Loève expansion, they are, in general, only uncorrelated. In practical use, we may need to model the coefficients simultaneously for a more appropriate resampling.

Model evaluation is an essential aspect of any data analysis process. The choice of evaluation technique depends on the type of model under consideration. For regression models, goodness-of-fit tests such as the one described by Maier et al. (2021) can be used to assess accuracy. In this paper, we choose to perform a direct visual inspection by doing resampling from the estimated model. This technique allows us to intuitively inspect the resampled processes and evaluate the model's performance.

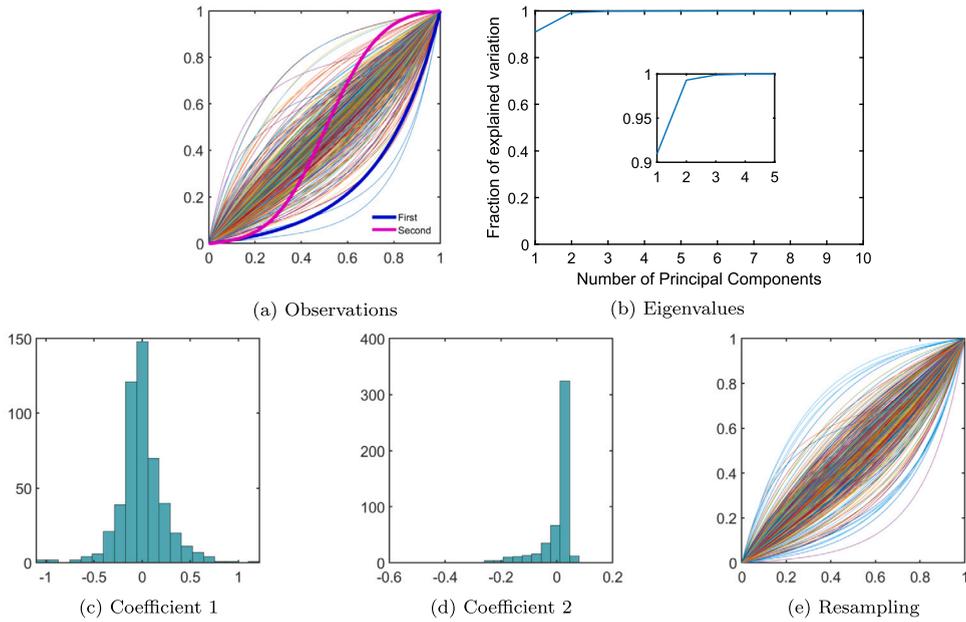


Fig. 3. Result on Simulation 2. (a) The curves represent 500 simulated time warping functions, and the bold blue and magenta curves represent the first and second eigenfunctions, respectively. (b) Fraction of variance explained by the first n principal components. The inset plot provides a magnified view of the first five components. (c) Histogram of the first principal component. (d) Histogram of the second principal component. (e) 500 resampled functions with the estimated model. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

2.4.2. Resampling examples

We will use one example to illustrate Algorithm 2. In this example, 500 warping functions are given by: $\gamma_i(t) = \alpha_{1i} \frac{e^{a_i t} - 1}{e^{a_i} - 1} + \alpha_{2i} \frac{e^{b_i t} - 1}{e^{b_i} - 1} + (1 - \alpha_{1i} - \alpha_{2i}) \frac{d_i \left(\frac{e^{-c_i t} - 1}{e^{-c_i} - 1} \right) - 1}{e^{d_i} - 1}$, where $a_i, d_i \sim \text{Exp}(\frac{1}{3})$ (i.e., exponential distribution with mean $\frac{1}{3}$), $b_i \sim \chi^2(3)$ (i.e., Chi-square distribution with mean 3), and $c_i \sim \Gamma(0.5, 2)$ (i.e., gamma distribution with mean 1). In addition, we have $x_i, y_i \sim U(0, 1)$, and set $\alpha_{1i} = x_i, \alpha_{2i} = \max(y_i - x_i, 0)$. The principal component analysis and resampling result are shown in Fig. 3. At first, the 500 warping functions are shown in Panel (a). The top 10 eigenvalues are shown in Panel (b). We can see that the first two principal components explain over 99% of the total variance, and thus the analysis is conducted only on the these two components. To visualize the variability, we superimpose the first two eigenfunctions in Panel (a). The distributions of the first two principal components are shown in Panels (c) and (d), respectively. There is no simple parametric form to describe the distributions, and we choose to adopt the conventional Gaussian kernel method to estimate distribution functions. Based on the estimated two distributions, we can use Algorithm 2 to resample 500 warping functions, and the result is shown in Panel (e). It can be easily seen that the resampled curves also look very similar to the original time warping functions in Panel (a), which indicates the effectiveness of the fPCA modeling procedure.

2.4.3. Resampling on the Berkeley growth data

In this section, we demonstrate the practical application of our resampling process using the well-known Berkeley Growth curve data (available at the link: <https://rdrr.io/cran/fda/man/growth.html>), where the heights of 39 boys and 54 girls were recorded at thirty-one time points from age 1 to age 18 (Ramsay and Silverman, 2006). As each growth curve is increasing in the age interval $[1, 18]$, it is also a warping function, albeit with a different domain and range. To examine the variability of time warping, we linearly transform the growth functions into standard warping functions from $[0, 1]$ onto $[0, 1]$. The original recording time points are not evenly spaced, and we adopt a smoothing procedure using cubic splines. The smoothed curves in both male and female groups are shown in Column (a) of Fig. 4. We will use the proposed method in Section 2.4 to model and resample these observations.

Fig. 4 also shows the principal component analysis and resampling result. The eigenvalues are shown in Column (b). We can see that the first twenty principal components explain almost 100% of the total variance in both groups. To visualize the variability contributed by the first three eigenfunctions, we plot curves representing the effects of these three eigenfunctions as perturbation from the mean in Columns (c), (d) and (e), respectively. Moreover, we adopt the conventional Gaussian kernel method to estimate the first twenty coefficient distributions. Based on these estimated results, we apply Algorithm 2 to resample the same number of warping functions as in the original dataset. The resampling results, shown in Column (e), closely resemble the original curves in Column (a) for both groups, demonstrating the practical effectiveness of the fPCA modeling procedure.

Remark. By utilizing the CLR-transformation, we can transform the warping space Γ_1 into a subspace of \mathbb{L}^2 , thereby enabling conventional statistical inference methods on time warping functions within the subspace. For example, the proposed representation

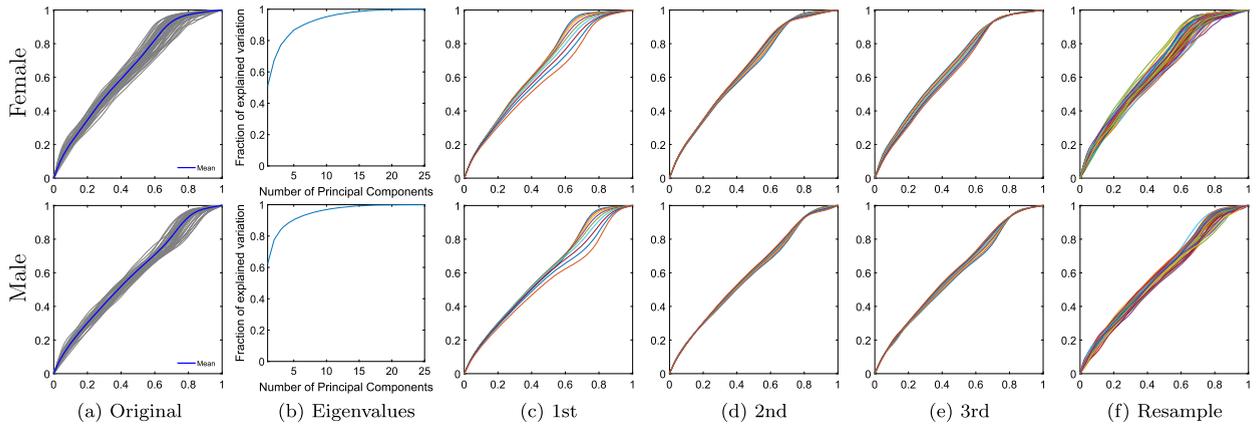


Fig. 4. Results on Berkeley growth data for the female group (1st row) and male group (2nd row). (a) Grey curves represent original growth functions in the linearly-transformed space from $[0, 1]$ to $[0, 1]$. The blue curve represents the mean function. (b) Fraction of variance explained by the first n principal components. (c) Function curves in the form of $\hat{\mu} + c\hat{\lambda}_1^{1/2}\hat{f}_1$ to visualize the effect of the first eigenfunction as perturbation from the mean, where $\hat{\mu}$ is the estimated mean function, c ranges from -2 to 2 with a 0.5 step size, \hat{f}_1 is the estimated first eigenfunction, and $\hat{\lambda}_1$ is the estimated first eigenvalue. (d) and (e), same as (c) except for the second and third eigenfunctions and eigenvalues, respectively. (f) resampled functions with the same sample size as the original ones.

can be used to conduct functional ANOVA (fANOVA) to test the difference between the means of two groups of warping functions. In addition, the CLR-transformed warping functions can be utilized as predictors for logistic regression. To exemplify these applications, we present an illustrative example in Appendix C, with the same Berkeley growth data.

3. Penalized function registration

In this section, we will utilize the proposed representation on time warping to provide a new approach for penalized function registration. Conventional penalties make the warping close to identity $\gamma_{id}(t) = t$, via second order (Ramsay and Silverman, 2006) or first order constraints (Srivastava and Klassen, 2016). Other types of penalties are given in the form of a prior term in a Bayesian framework. Bayesian registration is a relatively new paradigm that incorporates the prior information of warping function to conduct function registration (Cheng et al., 2014, 2016; Lu et al., 2017; Kurtek, 2017; Tucker et al., 2021; Matuk et al., 2021). Majority of these approaches are based on the SRVF (Square Root Velocity Function) transformation and explore appropriate representation of the warping functions on the corresponding tangent space, where a Gaussian process was used to model the inverse exponential transformed warping function. However, it was pointed out that this model is restricted to a bounded region of the positive orthant of the tangent space and linear operations on this region may get out of it and result in undesirable nonincreasing warping functions (Happ et al., 2019).

Our registration is based on the penalized framework in Srivastava and Klassen (2016), whereas we propose a new penalty on the CLR transformed warping space to control the degree of phase variation. The optimal warping is estimated by minimizing the penalized form with a gradient method instead of MCMC simulations in Bayesian approaches (Cheng et al., 2016; Kurtek, 2017). Unlike previous isotropic covariance representation (Cheng et al., 2016), our full covariance can characterize nonuniform temporal variance as well as correlated relationship in the time domain. We emphasize that our covariance is well defined in the CLR transformed \mathbb{L}^2 subspace so that we can utilize a second order stochastic process (e.g. a Gaussian process) to characterize the penalty term.

3.1. New penalty on time warping

Let f be an absolutely continuous function on the interval $[0, 1]$. Its SRVF is defined as $q : [0, 1] \rightarrow \mathbb{R}$, $q(t) = \dot{f}(t)/\sqrt{|\dot{f}(t)|}$ (Srivastava et al., 2011). For $\gamma \in \Gamma_1$, the SRVF of $f \circ \gamma$ is given by: $(q, \gamma) = \sqrt{\dot{\gamma}(t)}q(\gamma(t))$. For two function f_1, f_2 with corresponding SRVF functions q_1, q_2 , the unpenalized optimal time warping between f_1 and f_2 is given in the following form:

$$\gamma_{opt} = \arg \min_{\gamma \in \Gamma_1} \|q_1 - (q_2, \gamma)\|^2.$$

In Cheng et al. (2016), a Dirichlet prior is assigned to model the warping γ . Here, we propose to use a Gaussian process to model the transformed warping functions in the \mathbb{L}^2 space, i.e., $\log(\dot{\gamma}(t)) - \int_0^1 \log(\dot{\gamma}(s))ds \sim GP(0, c^{-1})$, where $c(s, t) : [0, 1] \times [0, 1] \rightarrow \mathbb{R}$ denotes the inverse of the covariance operator (i.e., we treat the covariance as a linear operator from \mathbb{L}^2 to \mathbb{L}^2 and assume it is invertible). Motivated by the quadratic exponent term in a Gaussian density function, the penalized function registration can be given in the following form:

$$\|q_1 - (q_2, \gamma)\|^2 + \lambda \int_0^1 \int_0^1 \left(\log(\dot{\gamma}(s)) - \int_0^1 \log(\dot{\gamma}(u)) du \right) c(s, t) \left(\log(\dot{\gamma}(t)) - \int_0^1 \log(\dot{\gamma}(u)) du \right) ds dt,$$

where $\lambda > 0$ is the penalty coefficient. The optimal time warping is obtained by minimizing the above form. As $\|q_1 - (q_2, \gamma)\|^2 = \|q_1\|^2 + \|q_2\|^2 - \int_0^1 2q_1(t)(q_2, \gamma)(t)dt$, we can get the loss function $J(\gamma)$ as follows:

$$J(\gamma) = \int_0^1 -2q_1(t)q_2(\gamma(t))\sqrt{\dot{\gamma}(t)} dt + \lambda \int_0^1 \int_0^1 \left(\log(\dot{\gamma}(s)) - \int_0^1 \log(\dot{\gamma}(u)) du \right) c(s, t) \left(\log(\dot{\gamma}(t)) - \int_0^1 \log(\dot{\gamma}(u)) du \right) ds dt. \tag{12}$$

Remark. The above loss function $J(\gamma)$ is for any symmetric positive definite kernel $c(s, t)$. There are two important special cases: isotropic and diagonal.

1. If the covariance is isotropic, we can set $c(s, t) = a\delta(s - t)$ in Equation (12), where $a > 0$ is a scale parameter. In this case, the penalty term is a scaled, squared \mathbb{L}^2 norm $a\|\log(\dot{\gamma}) - \int_0^1 \log(\dot{\gamma}(u)) du\|^2$.
2. If the covariance is diagonal, we can set $c(s, t) = r(t)\delta(s - t)$ in Equation (12), where $r(t) > 0$ is a weight function to denote the inverse variability at each time t . In this case, the penalty term is a weighted, squared \mathbb{L}^2 norm $\|\sqrt{r}(\log(\dot{\gamma}) - \int_0^1 \log(\dot{\gamma}(u)) du)\|^2$.

Regarding the selection of covariance, it should align with specific applications. In general, when one aims to apply a uniform penalty over all time points, it is advisable to utilize isotropic covariance. On the other hand, if the intention is to impose varying penalties at different time points, choosing a diagonal one is recommended. In situations where prior knowledge about co-variability is available, and there is a desire to incorporate it into the function alignment, the full case can be employed.

3.2. Optimization and the alignment algorithm

When there is no penalty term, the loss function is an integration with respect to the warping function and a dynamic programming procedure can be applied to get the optimal warping function, albeit on a discrete grid (Srivastava et al., 2011). However, with the penalty, the dynamic programming cannot be used because the loss function in Equation (12) can no longer be written under one integration. To deal with this problem, we propose to conduct the optimization via a gradient-based method. Note that the time warping function is in a non-vector space with conventional \mathbb{L}^2 metric, and the gradient on warping cannot be used for optimization. Analogous to the CLR transformation, we let $\phi(t) = \log(\dot{\gamma}(t)) \in \mathbb{L}^2([0, 1])$, and then we can get the new loss function of ϕ in the following form:

$$J(\phi) = \int_0^1 -2q_1(t)q_2\left(\int_0^t \exp(\phi(s)) ds\right)\sqrt{\exp(\phi(t))} dt + \lambda \int_0^1 \int_0^1 \left(\phi(s) - \int_0^1 \phi(u) du \right) c(s, t) \left(\phi(t) - \int_0^1 \phi(u) du \right) ds dt. \tag{13}$$

Note that we still have one constraint on $\phi(t)$, i.e., $\int_0^1 \exp(\phi(t)) dt = 1$. So when we apply the gradient descend, we will need to conduct this normalization to update time warping function in each iteration. Using the variational method, we can calculate the gradient of the loss function as follows (see details in Appendix D):

$$\begin{aligned} \frac{\partial J}{\partial \phi}(t) &= -2 \exp(\phi(t)) \int_t^1 q_1(\mu)q_2\left(\int_0^\mu \exp(\phi(s)) ds\right)\sqrt{\exp(\phi(\mu))} d\mu \\ &\quad - q_1(t)q_2\left(\int_0^t \exp(\phi(s)) ds\right)\sqrt{\exp(\phi(t))} + \lambda \left(\int_0^1 c(t, s)\phi(s) ds + \int_0^1 \phi(s)c(s, t) ds \right. \\ &\quad \left. - \int_0^1 \int_0^1 \phi(s)c(s, u) ds du - \int_0^1 \phi(u) du \int_0^1 c(t, s) ds - \int_0^1 \int_0^1 c(s, u)\phi(u) ds du \right) \end{aligned} \tag{14}$$

$$- \int_0^1 c(s, t) ds \int_0^1 \phi(u) du + 2 \int_0^1 \phi(u) du \int_0^1 \int_0^1 c(s, v) ds dv \Big).$$

The gradients on the two special cases (isotropic covariance and diagonal covariance) are also given in Appendix D, where the calculations are more efficient because of the simplified structures on the covariance.

Based on the gradient function in Equation (14), we can apply the gradient descent method. We emphasize that this method has linear computational order w.r.t. the number of discrete points, and is highly efficient in practical calculation. In contrast, the well-known dynamic programming is in the quadratic order and can be time-consuming when the number of discrete points is large. In summary, the overall alignment process is given in the following algorithm:

Algorithm 3 Alignment with Penalized Registration.

Require: Two real valued functions f_1, f_2 on interval $[0, 1]$, initial warping γ_0 , learning rate ϵ , threshold δ , tuning parameter λ .

Calculate the SRVF functions $q_1(t), q_2(t)$ of $f_1(t), f_2(t)$, respectively.

Let $\phi(t) = \log(\gamma_0(t))$, and estimate the loss function $J(\phi)$ using Equation (13).

Calculate derivative $\frac{\partial J}{\partial \phi}(t)$ of the loss function using Equation (14).

while $\|\frac{\partial J}{\partial \phi}\| > \delta$ **do**

$\phi(t) \leftarrow \phi(t) - \epsilon \frac{\partial J}{\partial \phi}$.

$\phi(t) \leftarrow \phi(t) - \log \left(\int_0^1 \exp(\phi(s)) ds \right)$.

Recalculate the loss function with the new $\phi(t)$.

end while

Let ϕ_{new} be the last ϕ in the while loop. Then the optimal warping is: $\gamma_{new}(t) = \int_0^t \exp(\phi_{new}(s)) ds$.

Output $\gamma_{new}(t)$.

Remark. 1) In addition to normalizing ϕ at each iteration, we can also use the Lagrange multiplier technique to solve the optimization with constraint $\int_0^1 \exp(\phi(t)) dt = 1$. It is found that this method provides similar optimization performance as that in Algorithm 3 and is therefore omitted in this paper. **2)** Without the penalty (or prior) term, i.e., by setting the $\lambda = 0$, the registration procedure is symmetric. This is a main result in the SRVF-based registration method (Srivastava et al., 2011). However, when the penalty term is added, the registration is no longer symmetric. **3)** For the selection of the penalty coefficient λ , typically, a larger value will put more constraints on the warping and push it towards the identity warping function, while a smaller value will allow more flexibility to warping. An illustrative example can be found in Appendix E.1. The optimal λ value may vary depending on the specific objectives of practical applications.

3.3. Alignment illustration

We will now illustrate the penalized alignment with diagonal covariance kernel which describes nonuniform and uncorrelated constraints in the time domain. We also include other examples which focus on isotropic and full covariances in Appendix E. To the best of our knowledge, such studies have not been well explored in function registration.

We at first simulate two multimodal functions, i.e., $f_1(t) = 6 \cdot 0.8^{20t} \cdot \cos(10\pi t - \frac{\pi}{4})$ and $g(t) = 5 \cdot 0.8^{20t} \cdot \sin(10\pi t), t \in [0, 1]$.

Then we generate a warped version of $g(t)$ by defining $f_2(t) = g(\gamma(t))$ with warping function $\gamma(t) = \frac{e^{2t}-1}{e^2-1}$. The functions $f_1(t)$ and $f_2(t)$ are shown as blue and green solid curves in Fig. 5(a), respectively. Moreover, we set $c(s, t) = r(t)\delta(s - t)$, where

$$r(t) = \begin{cases} 0.025(t + 0.1) & \text{if } 0 \leq t \leq 0.6 \\ 250t & \text{if } 0.6 < t \leq 1 \end{cases}.$$

$r(t)$ is positive and piecewise linear on $[0, 1]$. Its function value is close to 0 on $[0, 0.6]$, and much larger in magnitude on $(0.6, 1]$, which indicates nonuniform penalty in the time domain.

Our goal is find optimal warping function γ^* to minimize the loss function in Equation (12). The registration results are shown in Fig. 5. When $\lambda = 0$, and 10, the optimal warping functions are calculated using Algorithm 3 and shown in Fig. 5(b). We can see that when there is no penalty (i.e., $\lambda = 0$), the optimal warping can align f_2 to f_1 very well. The aligned functions $f_2(\gamma^*(t))$ are also shown in Fig. 5(a). We can see $f_2(\gamma^*(t))$ is right on the top of $f_1(t)$ when $\lambda = 0$. When there is a penalty, the optimal warping at the first part in the domain overlap the optimal warping when $\lambda = 0$, but the latter part gets closer to the identity warping $\gamma_{id}(t) = t$ (optimal warping when $\lambda = \infty$). Indeed, $f_2(\gamma^*(t))$ is right on the top of $f_1(t)$ when $\lambda = 10$ for $t \in [0, 0.3]$ and start to be lagged compare to the $f_1(t)$ from $t = 0.3$ when $\lambda = 10$. These results clearly demonstrate the effectiveness of the new penalty term for the nonuniform constraint in the time domain in the alignment process.

4. Boxplot of warping functions and outlier detection

In this section, we will utilize the proposed representation on time warping to construct functional boxplot for time warping functions. Boxplot is a standardized way of displaying the distribution of data and one of very few statistical graph methods that show outliers. Functional data are infinite dimensional, which brings challenge on theory as well as computation for a proper boxplot. The challenge is even higher for time warping functions due to their nonlinear structure in \mathbb{L}^2 . To the best of our knowledge, boxplot representation for time warping is still an under-explored area.

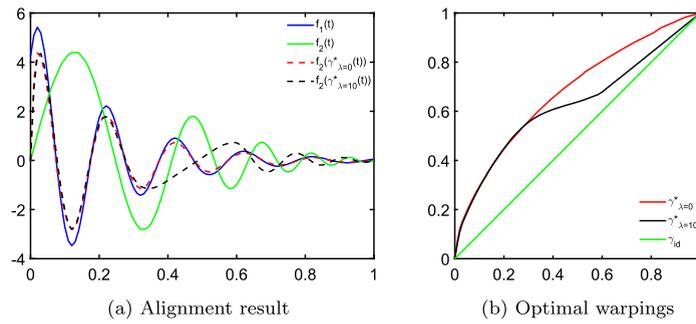


Fig. 5. Penalized registration illustration with diagonal covariance kernel. (a) Original functions and alignment functions. The blue and green solid curves are the two given functions f_1 and f_2 , the red and black dotted curves are the aligned f_2 using Algorithm 3 with λ equal to 0 and 10, respectively. (b) Optimal warping functions in the alignment. The red and black curve are the optimal warping functions from Algorithm 3 with λ equal to 0 and 10, respectively. The green curve is the identity warping function.

Sun and Genton (2011) first generated functional boxplot using the notion of band depth. We can apply this method directly to the warping functions. Alternatively, to better address the conventional \mathbb{L}^2 property, we can construct the boxplot by applying this functional boxplot to the CLR-transformed warping functions. However, either method constructs the boxplot in a point-wise manner which ignores the overall structure of the warping functions and may lose essential information such as shape and smoothness of the functions. Xie et al. (2017) proposed a metric-based method for constructing warping boxplot. This method is based on the square-root-velocity-function (SRVF) method and tangent space. It was pointed out that there is no one-to-one mapping between the tangent space and warping space, which may result in un-desired (e.g. non-increasing) warping functions in the estimation process.

To deal with these problems, we propose to construct the warping boxplot by adopting the idea of functional bagplot (Hyndman and Shang, 2010), which we will describe in this section. The functional bagplot uses only the first two principal components to provide a bivariate version of boxplot, which may cause severe information loss. In contrast, to preserve more useful and complex information, we will use the first p dominant principal components (e.g., with a 95% cutoff on cumulative variance). Specifically, we first find the principal component vector of the CLR-transformed warping functions, and then rank the vector according to the Tukey’s halfspace location depth (Tukey, 1975). After that, we build the inner and outer regions, which contain approximately 50% and 99% of the warping data with respect to the largest Tukey’s depth, respectively. Finally, we will display the region in the original warping space, which is the warping boxplot we built. The detailed procedure is given below.

4.1. Functional boxplot

The estimation of the principal components has been thoroughly described in Section 2.4. To generate the inner region in the boxplot, we order the vectors of the estimated principal components according to decreasing depth values. Let $\gamma_{[n]}$ denote the sample curve associated with the n th largest Tukey depth value. We select the first 50% of data points, i.e., $\gamma_{n=[1, \dots, \lfloor N/2 \rfloor]}$, and use these data points to construct a convex hull; this is the inner region which contains at least 50% of the total number of functional observations. Then we find the corresponding original warping functions, and define the first quartile warping as $\gamma_{Q_1}(t) = \min\{\gamma_{[1]}(t), \dots, \gamma_{[N/2]}(t)\}$, and the third quartile warping as $\gamma_{Q_3}(t) = \max\{\gamma_{[1]}(t), \dots, \gamma_{[N/2]}(t)\}$.

The outer region (or fence) of the boxplot is the convex hull of the points contained within the region obtained by inflating the inner region (relative to the Tukey median) by a factor ρ . Hyndman and Shang (2010) used a value of $\rho = 2.58$, as that will allow the outer region to contain 99% of the observations when only two principal components are used, which follows bivariate normal distributions. Here we stick with this strategy and set inflation factor ρ to make it contain 99% of the observation if the coefficient vectors follow a multivariate Gaussian distribution, i.e., $\rho = \sqrt{\frac{\chi^2_{p,0.99}}{\chi^2_{p,0.5}}}$, where p is the dimension of coefficient vectors (the number of principal components we choose in the fPCA step). We define γ_{LF} and γ_{UF} , i.e., the lower and upper fences, that give the minimum or maximum value of all the warping functions stay in this inflated region along their domain. Furthermore, the warping median is defined as the function that corresponds to the Tukey’s depth median, and the corresponding functions whose principal component coordinates stay outside the outer region are identified as outliers. In summary, the generative steps of construction of the boxplot for time warping function in Γ_1 is given in Algorithm 4.

Remark. 1) The influence of the inflation factor ρ on the construction of the boxplot is demonstrated in Appendix F. It is apparent that by elevating the value of ρ , we can observe a reduction in the detection of outliers. The choice of ρ should ideally be informed by prior information about the percentage of contamination in the data. **2)** We in fact use the convex hulls built in $H(0, 1)$ space to detect outliers, i.e., any warping outside the convex hull will be identified as outliers. The boxplot in the warping space is only for visualization purpose. That is, in this boxplot, any curves outside the phase outlier cutoffs will be identified as outliers. However, we point out that the curves inside the boxplot may also be outliers.

Algorithm 4 Construct boxplot for warping functions.

Require: N observed warping functions γ_n in Γ_1 .

Transform the warping functions γ_n in to $H(0, 1)$: $X_n(t) = \log(\dot{\gamma}_n(t)) - \int_0^1 \log(\dot{\gamma}_n(s)) ds$.

Calculate the mean $\hat{\mu}(t) = \frac{1}{N} \sum_{n=1}^N X_n(t)$ and the covariance $\hat{K}(s, t) = \frac{1}{N-1} \sum_{n=1}^N (X_n(s) - \hat{\mu}(s))(X_n(t) - \hat{\mu}(t))$.

Apply spectral decomposition to \hat{K} to get the eigen sequence $\{(\lambda_i, e_i)\}_{i=1}^N$.

Find cutoff threshold δ , let $m = \max\{i | \lambda_i > \delta\}$.

for $k = 1 : m$ **do**

 Calculate the coefficients $Z_{nk} = \int_0^1 (X_n(t) - \hat{\mu}(t))e_k(t)dt, n = 1, 2, \dots, N$.

end for

Calculate the Tukey's depth $D(n)$ for $Z_n = (Z_{n1}, \dots, Z_{nm}), n = 1, 2, \dots, N$.

Construct convex hull 1 using 50% of Z_n with highest depth $D(n)$.

Construct convex hull 2 using data points staying in the region by inflating the convex hull 1 by a factor ρ w.r.t. the Tukey's depth median.

1st quartile: $\gamma_{Q_1}(t) = \min_{1 \leq n \leq N} \{\gamma_n(t) : Z_n \in \text{Convex Hull 1 (50\%)}\}$;

3rd quartile: $\gamma_{Q_3}(t) = \max_{1 \leq n \leq N} \{\gamma_n(t) : Z_n \in \text{Convex Hull 1 (50\%)}\}$;

Lower fence: $\gamma_{LF}(t) = \min_{1 \leq n \leq N} \{\gamma_n(t) : Z_n \in \text{Convex Hull 2 (99\%)}\}$;

Upper fence: $\gamma_{UF}(t) = \max_{1 \leq n \leq N} \{\gamma_n(t) : Z_n \in \text{Convex Hull 2 (99\%)}\}$;

Flag any points outside convex hull 2 as outliers: $\{\gamma_n(t) : Z_n \notin \text{Convex Hull 2}\}$.

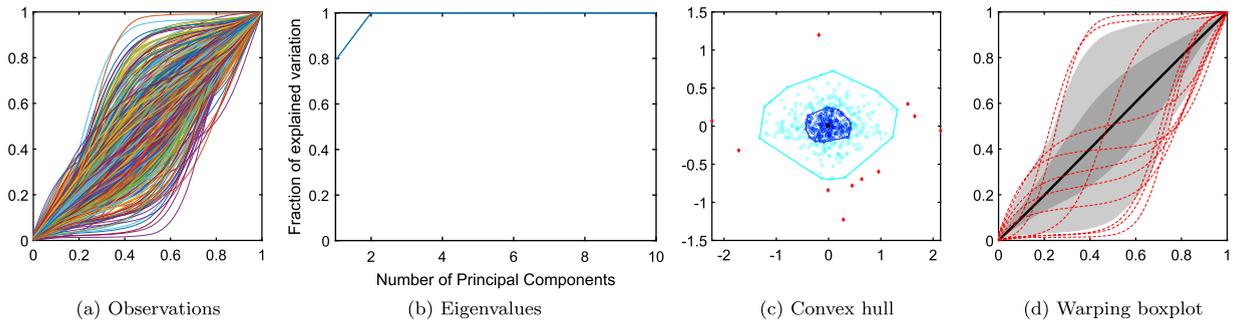


Fig. 6. Result on Illustration. (a) The curves show simulated 500 time warping functions. (b) Fraction of variance explained by the first n principal components. (c) 2 layer convex hulls (d) the warping boxplot generated by Algorithm 4, the red dashed curves are the outlier candidates.

We also emphasize that we can apply our proposed representation to perform functional interpolation for time warping function, which is a process of determining the unknown functions that lie in between the known functions. A brief description is given in Appendix G.

4.2. Illustration

We use one example to thoroughly illustrate the proposed boxplot for warping functions in Algorithm 4. In this example, 500 warping functions $\gamma_i, i = 1, \dots, 500$ on $[0, 1]$ are generated in the following form:

$$\gamma_i(t) = \psi_B^{-1}(G_{1,i}\phi_1(t) + G_{2,i}\phi_2(t)), \quad t \in [0, 1],$$

where the coefficients $G_{k,i} \sim La(0, \frac{1}{\sqrt{2k}})$, i.e., Laplacian distribution with mean 0 and variance $\frac{1}{4k^2}, k = 1, 2, i = 1, \dots, 500$, and the orthonormal basis functions $\phi_1, \phi_2 \in H(0, 1)$ are given as $\phi_1(t) = \sqrt{2} \sin(2\pi t)$ and $\phi_2(t) = \sqrt{2} \cos(2\pi t)$.

These 500 warping functions are shown in Panel (a) of Fig. 6. The top 10 eigenvalues in the covariance function are shown in Panel (b). We can see that the first two principal components explain over 99% of the total variance. Therefore, we construct the boxplot only use the first two principal components and set inflation factor ρ to 2.58.

In Panel (c), we show the 50% and 99% convex hulls. Specifically, the region surrounded by blue lines represents the convex hull 1 generated using Algorithm 4, which is analogous to the bag in Hyndman and Shang (2010), and the one surrounded by the cyan lines represents the convex hull 2. The dots represent the bivariate principal components of the CLR-transformed warping functions. Tukey's depth median is shown as the black dot, and the red dots outside the convex hulls are the identified outliers. In Panel (d), we show the corresponding boxplot in warping space. The dark and light grey regions correspond to the blue and cyan convex hulls in Panel (c), which contains all the warpings in these two convex hulls, respectively. The bold black curve and the red dashed curves are the warping median and the identified warping outliers, which correspond to the black and red points in Panel (c), respectively.

This proposed boxplot apparently provides a reasonable way to characterize the main body; the gray region contains most of the warping functions. In addition, this method can appropriately identify the outliers in the data. 11 of the 500 warpings are identified as outliers, accounting for about 2.2% of the entire function set. We can further divide these outliers into three categories: 1) warpings that increase too slowly at the beginning, corresponding to the relatively flat red dashed curves in the first half, 2) warpings that increase too quickly at the beginning, corresponding to the red dashed curves in which the latter half is very flat, and 3) warpings that increase relatively fast at the beginning, very fast at the end, but very slow in the middle.

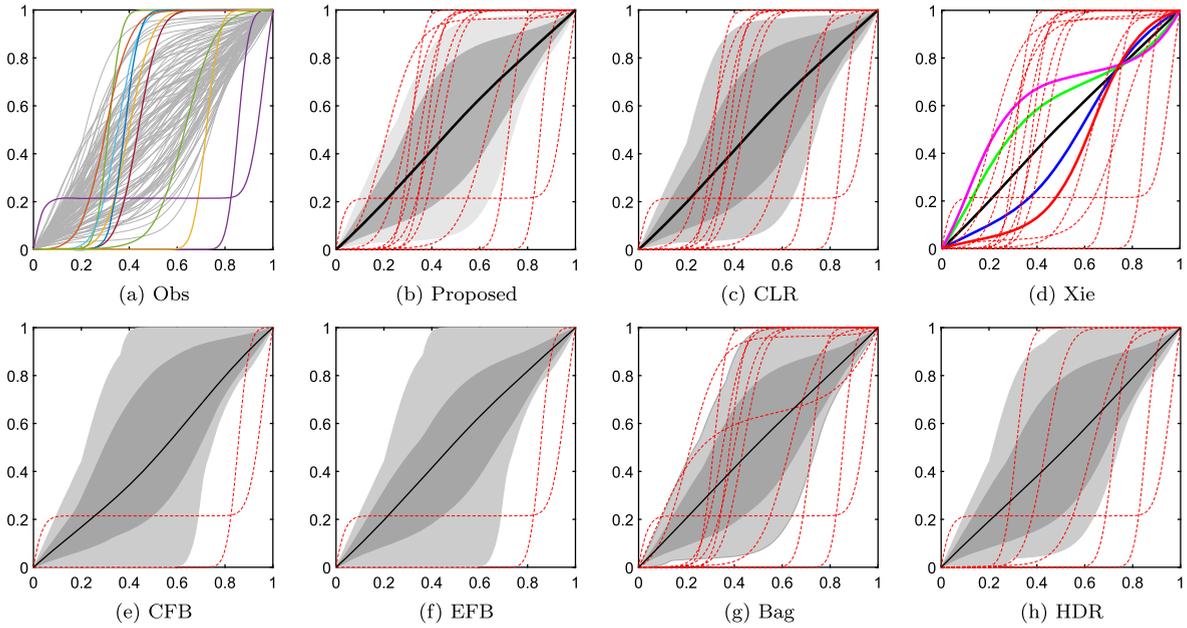


Fig. 7. Result in Case 1. (a) 110 simulated observations, where the gray and colorful curves represent 100 main body and 10 outlier warping functions, respectively. (b)(c)(d) Boxplots generated by the proposed method, functional boxplot to CLR-transformed function, and the Xie method, respectively. The thin red lines in each plot are the identified outliers. (e) Conventional functional boxplot (CFB). (f) Enhanced functional boxplot (EFB). (g) Functional bagplot (Bag). (h) Functional HDR plot.

4.3. Outlier detection and comparison with other methods

In section 4.2, we have illustrated the proposed method using one simulation, which provides all algorithmic details in the functional boxplot construction. In this section, we will use another simulation to examine the performance of this boxplot for outlier detection. We will also compare its performance with competing methods.

In this simulation, we will generate two groups of warpings, where the first group includes 100 warpings as the main body, and the second group includes 10 warpings, treated as outliers. At first, we will test if our method can detect these outliers or not. Then we will compare our method with other functional plot methods, including the Xie method (Xie et al., 2017), functional boxplot (Sun and Genton, 2011), enhanced functional boxplot (Sun and Genton, 2011), functional bagplot (Hyndman and Shang, 2010), and functional Highest density region (HDR) boxplot (Hyndman and Shang, 2010).

Simulation: In this example, 110 warping functions are generated in the following form:

$$\gamma_i(t) = \psi_B^{-1} \left(\sum_{j=1}^m G_{j,i} \phi_j(t) \right), \quad t \in [0, 1].$$

The first group of 100 warpings are generated by letting $m = 2$ and coefficients $G_{j,i} \sim N(0, \frac{1}{2i})$, i.e., Gaussian distribution with mean 0 and variance $\frac{1}{4i^2}, j = 1, 2, i = 1, \dots, 100$, and the orthonormal basis functions ϕ_j is the first 2 basis functions in Equation (7).

We consider 3 different cases for the 10 warping functions in the second group, which represent different types of functional outliers. They are generated as follows:

- Case 1: $m = 30$ and coefficients $\{G_{ji}\} \sim N(0, \frac{10}{i})$, i.e., Gaussian distribution with mean 0 and variance $\frac{100}{i^2}, j = 1, \dots, 30, i = 1, \dots, 10$.
- Case 2: $m = 30$ and coefficients $\{G_i\} \sim La(0, \frac{1}{\sqrt{2i}})$, i.e., Laplacian distribution with mean 0 and variance $\frac{1}{i^2}, j = 1, \dots, 30, i = 1, \dots, 10$.
- Case 3: $m = 10$ and coefficients $\{G_i\} \sim N(0, \frac{i}{40})$, i.e., Gaussian distribution with mean 0 and variance $\frac{i^2}{1600}, j = 1, \dots, 10, i = 1, \dots, 10$.

We show the results for the three cases in Figs. 7, 8 and 9, respectively.

Results in Case 1 (Magnitude): In this case, coefficients $\{G_i\}$ of larger variances are used to generate the warping outliers compared to the main body. The results are shown in Fig. 7. In Panel (a), we show 110 simulated warping curves. The 100 warping functions in the first group, i.e., the main body, are shown in grey. In contrast, the ten outlier warpings are shown as colored curves. As we can see, all ten outlier curves are partially located outside the central region formed by the first set of warpings. We call these outliers magnitude outliers to properly indicate their outlying pattern. The boxplot constructed by the proposed method is shown in

Table 1
Outliers Detection Performance: the two numbers in each cell represent sensitivity and specificity, respectively.

Method	Proposed	CLR	Xie	CFB	EFB	Bag	HDR
Case 1:							
Magnitude	$(\frac{10}{10}, \frac{99}{100})$	$(\frac{10}{10}, \frac{100}{100})$	$(\frac{10}{10}, \frac{97}{100})$	$(\frac{2}{10}, \frac{100}{100})$	$(\frac{2}{10}, \frac{100}{100})$	$(\frac{10}{10}, \frac{99}{100})$	$(\frac{6}{10}, \frac{100}{100})$
Case 2:							
Magnitude & Shape	$(\frac{10}{10}, \frac{100}{100})$	$(\frac{3}{10}, \frac{100}{100})$	$(\frac{1}{10}, \frac{97}{100})$	$(\frac{0}{10}, \frac{100}{100})$	$(\frac{0}{10}, \frac{100}{100})$	$(\frac{1}{10}, \frac{97}{100})$	$(\frac{2}{10}, \frac{96}{100})$
Case 3:							
Shape	$(\frac{10}{10}, \frac{100}{100})$	$(\frac{0}{10}, \frac{100}{100})$	$(\frac{0}{10}, \frac{96}{100})$	$(\frac{0}{10}, \frac{100}{100})$	$(\frac{0}{10}, \frac{100}{100})$	$(\frac{0}{10}, \frac{96}{100})$	$(\frac{0}{10}, \frac{94}{100})$

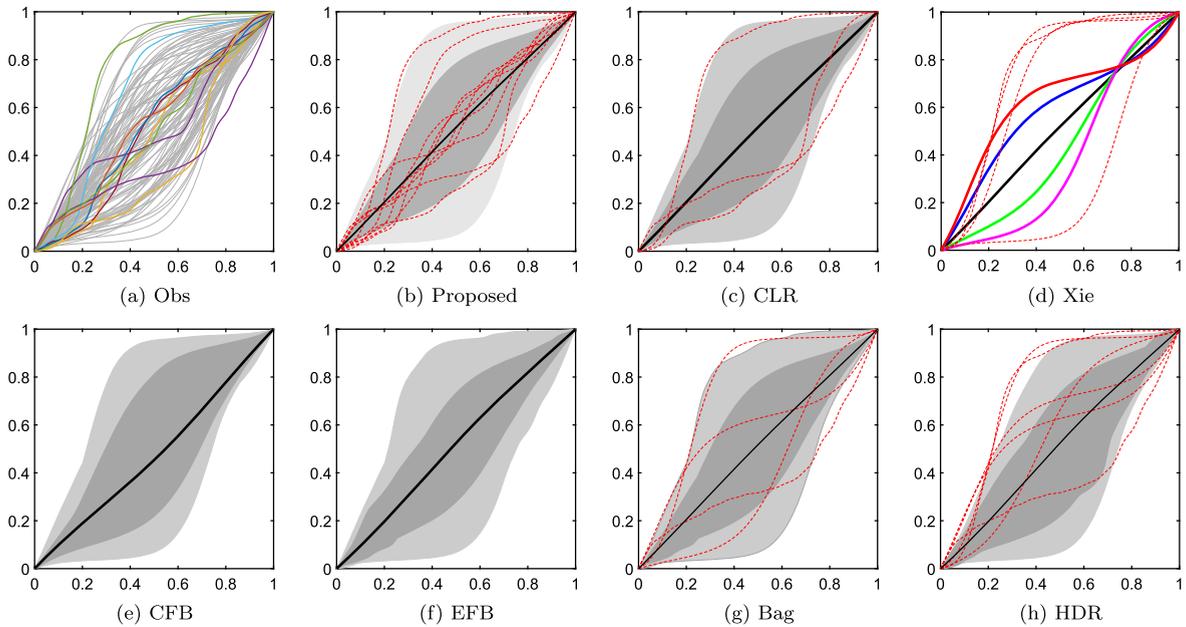


Fig. 8. Same as that in Fig. 7 except for result in Case 2.

Panel (b), where we mark outliers candidates as thin red lines. We use the first two principal components in this case and set the inflation ratio ρ to 2.58. It can be seen that we correctly identified all 10 outlier warpings. However, we mistakenly classified one main body warping as outlier.

In Panel (c), we show the boxplot generated by applying our proposed functional boxplot on CLR-transformed warping functions. We also marked outliers candidates as thin red lines. This method also correctly identified the 10 magnitude warping outliers. In Panel (d), we show the boxplot generated by the Xie Method. The bold black curve represents the warping median curve. The blue, green, bold red, and magenta curves represent two quartiles $\gamma_{Q_1}, \gamma_{Q_3}$, and two outlier cutoffs γ_{LF}, γ_{UF} , respectively. Moreover, the thin dotted red curves are the identified outlier candidates. This method also correctly identified the 10 outliers, but misclassified 3 main body ones as outliers. In Panels (e) and (f), we show the boxplots generated using functional boxplot with Band depth and functional boxplot with modified band depth, and the identified outliers are shown in red. It is easy to see that both methods correctly identified only two of the 10 outliers. In Panels (g) and (h), we show the functional bagplot and functional highest density region boxplot, and the identified outliers are also shown as dotted red curves. Functional bagplot gets all the outliers identified but misclassified one normal one, whereas the functional HDR plot only identifies six outliers.

For the convenience of comparison across different methods, the detailed outlier detection results for all methods in this case are shown in the 1st row of Table 1.

Result in Case 2 (Magnitude & Shape): In this case, the warpings in outlier group are generated using a Laplacian distribution. We also set the variance of the coefficients smaller than the coefficients in Case 1 so that the magnitude is no longer the dominant factor. The result in Case 2 is shown in Fig. 8. We display the simulated warping functions in Panel (a). We can see that all outlier curves have higher curvature than those in the main body. Furthermore, some of them are partially located outside the central region formed by the first set of warpings, so we say these outliers are not only outlying in shape but also in magnitude. The boxplot constructed by the proposed method is shown in Panel (b). We use the first three principal components and set the inflation ratio ρ to 2.99. It can be seen that all 10 outliers are correctly identified. We show the boxplots generated by all the other six methods

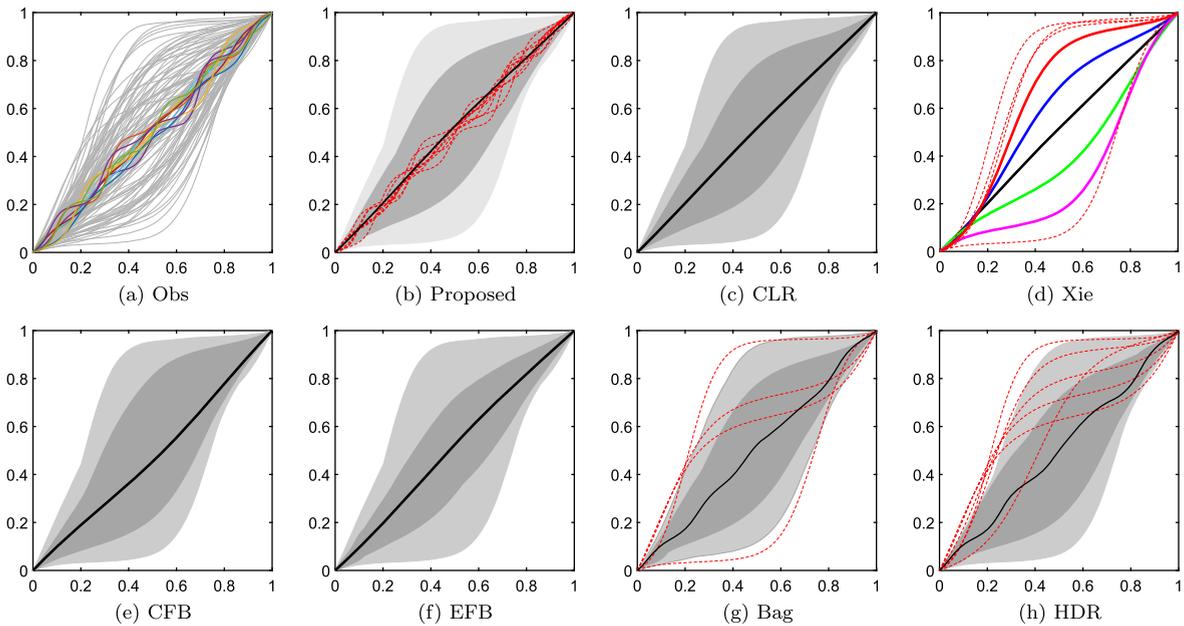


Fig. 9. Same as that in Fig. 7 except for result in Case 3.

in Panels (b) to (h), respectively. The detailed outlier detection results in all methods in this case are summarized in the 2nd row of Table 1.

Result in Case 3 (Shape): In this case, we use 10 basis functions and a relatively smaller variance of normally-distributed coefficients to generate the outlier warpings. All 110 functions are shown in Panel (a) of Fig. 9. We see that the outlier warpings stay inside the main body and are very close to the identity, but are different from the main body with respect to their high-curvature shape. The boxplots generated by all methods are shown in Panels (b) to (h), respectively. The outlier detection results of all 7 methods are shown in the 3rd row of Table 1. We use the first three principal components and set the inflation ratio ρ to 2.99 for our proposed method. It can be seen that only our proposed method can correctly identify these simulated outliers.

To sum up the results in these three cases, we can see that only our proposed method can achieve high values in both sensitivity and specificity. Compared with our method, the competing methods have no problem identifying outliers differing on magnitude, but they often fail to identify outliers differing on shape. We also notice that all methods can have high values on specificity, which indicates they all perform well in terms of true negative.

5. Real data application

In this section, we apply our method to the CDC COVID data (available at <https://covid.cdc.gov>) to test our proposed method in real-world data outlier detection. We select the cumulative COVID cases from 51 states, 3 territories, and New York City, respectively, from 03-11-2020 to 07-12-2022. Since the growth curve (i.e. cumulative cases over time) in each region is increasing during this period, it can be treated as a warping function, albeit with a different domain and range. To examine the variability of time warping, we linearly transform the growth functions into standard warping functions from $[0, 1]$ onto $[0, 1]$. The original curves are shown in Panel (a) of Fig. 10. For better visualization, we change the x-axis into the original data range, i.e., March-2020 to July-2022. We will use the proposed method in Section 4 to construct the boxplot and identify locations with different growth patterns compared to the other states.

In this example, the first three, six, and eight principal components explained more than 75%, 90%, and 95% of the total variance, respectively. However, we have only 55 data points. Due to the sparsity of the data, we cannot construct the Convex hull in high dimensions very effectively. Therefore, we only use the first three principal components to construct the boxplot and set the inflation factor to 2.99. Fig. 10 (b) shows the boxplot constructed by our proposed method. To visualize the difference between the identified outliers and the main body, we superimpose the 10 identified outliers on the boxplot. They are Alaska, Guam, Hawaii, Maine, Montana, New York City, Puerto Rico, Virgin Islands, Vermont, and Wyoming. In Panel (c), we showed the map of the United States and highlighted locations identified as outliers in red.

We will now provide interpretation on these identified outliers. At first, Guam, Puerto Rico, Virgin Islands are united states territories, and Alaska and Hawaii are not on the main continent, so it makes sense that their total case growth curves differ from other states. Phylogenetic analysis and sentinel surveillance suggest that the introduction of COVID-19 into NYC from travelers started during early to mid-February 2020. CDC pointed out COVID-19 spreads when an infected person breathes out droplets and very small particles that contain the virus, and people who are closer than 6 feet from the infected person are most likely to get infected. According to demographic statistics, New York City has the highest population density of any major city in the United

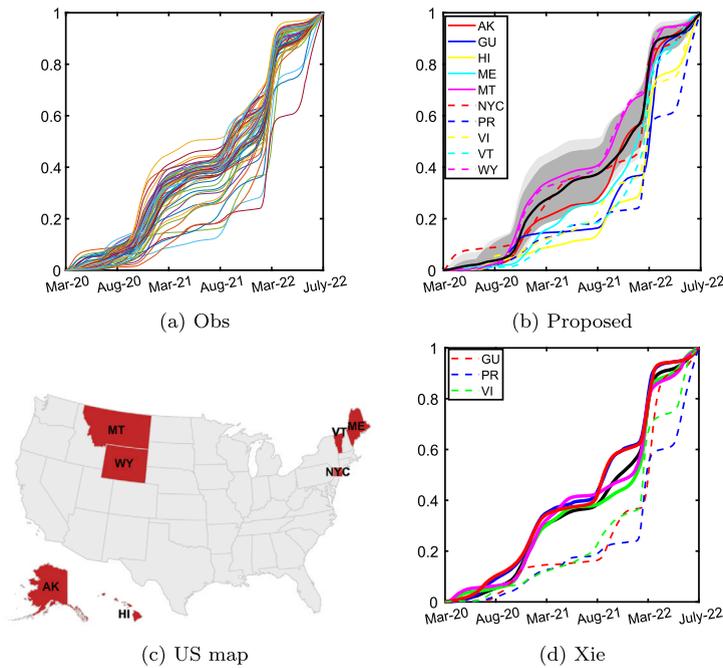


Fig. 10. Result on COVID-19 data. (a) The curves represent the original total case functions, linearly-transformed from $[0, 1]$ to $[0, 1]$. For better visualize the time range of the real data, the x-axis is displayed from March 2020 to July 2022. (b) The proposed boxplot and outlier candidates generated using Algorithm 4. (c) The US state map, where the red ones are the identified outlier states. (d) The boxplot and outlier candidates generated using the Xie method.

States, with over 27,000 people per square mile. COVID-19 in New York City had the largest following outbreak due to the highest population density. The growth curve, shown by the red dashed line in Fig. 10(b), was the sharpest at the beginning, so it is reasonable to identify New York City as an outlier candidate. On the contrary, Alaska, Wyoming, and Montana are the three states with the lowest population densities, which are 1.28, 6, and 7.42 per square mile, respectively. The first cases were found relatively late, and the growth curves were more flat at the beginning than elsewhere. As seen from Fig. 10 (b), Maine and Montana are identified as outliers because of the slow growth at the beginning.

As a comparison, we show the boxplot generated using the Xie Method in Panel (d), where the four boundaries are overlapped with each other and are difficult to interpret. This method detects only three outliers: Guam, Puerto Rico, and the Virgin Islands. This result does not seem reasonable because Vermont, and Hawaii have similar growth curves with the identified three curves, but are not identified.

Another aspect of interest is the rate of increase in COVID-19 cases. To analyze this, we take the derivatives of the case increase functions and perform functional alignment to obtain warping functions. By constructing boxplots on the estimated functions, we can gain insights into the variability of the rate of increase across different locations. Due to the space limitation, we have included this analysis as an alternative study in Appendix H.

6. Summary

In this paper, we have proposed a new representation for time warping functions as a linear inner-product space, which is an apparent advantage over the previous nonlinear approximation methods. The critical element of this process is a derivative operation of the warping function and then a centered logratio transformation to map the warping functions into a conventional \mathbb{L}^2 space. We have also defined two warping spaces to make the transformation mathematically precise. The first one, bounded warping space Γ_1 , is isometrically isomorphic to the space of bounded, centered \mathbb{L}^2 functions. We extended this bounded \mathbb{L}^2 to a Hilbert space, mapping it to a more general warping space Γ_2 . These two warping spaces provide sufficient representation for practical use. We then stated several statistical inferences under this new representation, including using fPCA to construct a model for functional warping observations, applying our new representation in penalized function registration to provide time-variant and temporally correlated constraints in function alignment, and constructing boxplot for time warping functions. Finally, we illustrate the method in a real-world dataset and obtain reasonable result.

We point out that Γ_2 is not a vector space, which limits its usefulness in the modeling process. We will aim to extend the warping space to a Hilbert space in the future. If this can be done, the warping will be fully described by a stochastic process in \mathbb{L}^2 space. In the context of penalized function registration, we have used a Gaussian process for warping. To capture the intricacies of practical data, we will explore non-Gaussian processes with broader applicability in the future. The determination of the optimal regularization coefficient λ , multiple penalized registration, and comparing different registration methods are all challenging problems of critical importance that require further investigation. In addition, a symmetric penalized form and Bayesian alignment represents an

intriguing and promising topic of research that deserves further exploration. Our current work focused on constructing the boxplot for time warping functions. We can further apply this into the general functional boxplot and outlier detection, i.e., we can decompose functional data into amplitude and phase components, and then construct individual boxplot for each of these two components to better visualize variability in functional data.

Appendix A. Norm and inner-product in Γ_1

The norm of the identity warping function $\gamma(t) = t$ is 0, representing the origin in the warping space. As the warping’s curvature increases, the norm also increases, indicating that it deviates further from the identity warping (or the origin). Additionally, an increase in the inner-product value signifies a decrease in the angle between two warping functions, indicating that they are more similar or closer to each other. The following simple example is used to illustrate this concept:

The method $\gamma_i(t) = \frac{e^{a_i t} - 1}{e^{a_i} - 1}$ is used to simulate six warping functions. The values of a_i are set to be -5, -3, -1, 1, 3, and 5 for $i = 1, \dots, 6, respectively$. The warping functions are shown in Fig. A.11. The inner-product matrix for these six warping functions can be calculated using Equation (3), and is as follows:

$$M_{inp} = \begin{matrix} & \begin{matrix} \gamma_1 & \gamma_2 & \gamma_3 & \gamma_{id} & \gamma_4 & \gamma_5 & \gamma_6 \end{matrix} \\ \begin{matrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \\ \gamma_{id} \\ \gamma_4 \\ \gamma_5 \\ \gamma_6 \end{matrix} & \begin{pmatrix} 2.08 & 1.25 & 0.42 & 0 & -0.42 & -1.25 & -2.08 \\ 1.25 & 0.75 & 0.25 & 0 & -0.25 & -0.75 & -1.25 \\ 0.42 & 0.25 & 0.08 & 0 & -0.08 & -0.25 & -0.41 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -0.42 & -0.25 & -0.08 & 0 & 0.08 & 0.25 & 0.41 \\ -1.25 & -0.75 & -0.25 & 0 & 0.25 & 0.75 & 1.25 \\ -2.08 & -1.25 & -0.42 & 0 & 0.42 & 1.25 & 2.08 \end{pmatrix} \end{matrix}.$$

The diagonal elements of the matrix M_{inp} represent the norm of these warping functions. The off-diagonal elements represent the inner product between the corresponding row and column labels. The warping functions exhibit the greatest curvature in γ_1 and γ_6 , which also have the highest norms. Conversely, γ_3 and γ_4 exhibit the smallest norms, except for the identity warping. In summary, we can deduce that as the warping approaches the identity function, its norm decreases. Moreover, we can observe that the inner product of γ_1 and γ_2 is greater than the inner product of γ_1 and γ_3 , and the inner product of γ_1 and γ_4 , among others. This indicates that when two warping functions are closer to each other, their inner product is larger, implying that the angle between them is smaller.

Appendix B. Proof of Proposition 2.2

For any $n \in \mathbb{N}$,

$$\sum_{i=1}^n \left| \mu_i \phi_i(s) \phi_i(t) \right| \leq \sum_{i=1}^n \left| \mu_i \sqrt{2} \sqrt{2} \right| = 2 \sum_{i=1}^n \mu_i \leq 2 \sum_{i=1}^{\infty} \mu_i < \infty.$$

Then, $\forall s, t \in [0, 1]$

$$\sum_{i=1}^{\infty} \left| \mu_i \phi_i(s) \phi_i(t) \right| = \lim_{n \rightarrow \infty} \sum_{i=1}^n \left| \mu_i \phi_i(s) \phi_i(t) \right| \leq \lim_{n \rightarrow \infty} 2 \sum_{i=1}^n \mu_i = 2 \sum_{i=1}^{\infty} \mu_i < \infty.$$

Thus, $K(s, t) = \sum_{i=1}^{\infty} \mu_i \phi_i(s) \phi_i(t)$ converges absolutely.

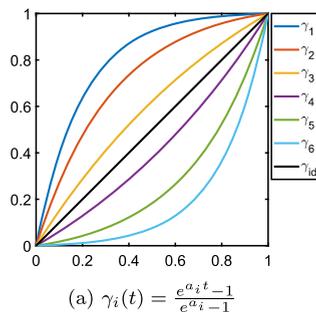


Fig. A.11. Six warping functions were simulated using the standard method $\gamma_i(t) = \frac{e^{a_i t} - 1}{e^{a_i} - 1}$, where $a_i = -5, -3, -1, 1, 3, 5$ for i ranging from 1 to 6, with the bold black line representing the identity warping.

For any $s, t \in [0, 1]$, we have:

$$\left| K(s, t) - \sum_{i=1}^n \mu_i \phi_i(s) \phi_i(t) \right| \leq \sum_{i=n+1}^{\infty} \left| \mu_i \phi_i(s) \phi_i(t) \right| \leq 2 \sum_{i=n+1}^{\infty} \mu_i.$$

As $\sum_{i=1}^{\infty} \mu_i < \infty$, we can get:

$$\lim_{n \rightarrow \infty} \left| K(s, t) - \sum_{i=1}^n \mu_i \phi_i(s) \phi_i(t) \right| = 0.$$

Thus, $\sum_{i=1}^{\infty} \mu_i \phi_i(s) \phi_i(t)$ converges uniformly.

We will then prove that K is symmetric, non-negative definite, and continuous:

- **Symmetry:** It is easy to see that $K(s, t) = \sum_{i=1}^{\infty} \mu_i \phi_i(s) \phi_i(t) = \sum_{i=1}^{\infty} \mu_i \phi_i(t) \phi_i(s) = K(t, s), \quad \forall s, t \in [0, 1]$.
- **Non-negative definiteness:** $\forall f \in L^2([0, 1])$, we have:

$$\begin{aligned} \int_0^1 \int_0^1 f(s) K(s, t) f(t) ds dt &= \int_0^1 \int_0^1 f(s) \sum_{i=1}^{\infty} \mu_i \phi_i(s) \phi_i(t) f(t) ds dt \\ &= \sum_{i=1}^{\infty} \mu_i \int_0^1 \int_0^1 f(s) \phi_i(s) \phi_i(t) f(t) ds dt \\ &= \sum_{i=1}^{\infty} \mu_i \left(\int_0^1 f(s) \phi_i(s) ds \right)^2 \geq 0. \end{aligned}$$

- **Continuity:** Define $K_n(s, t) := \sum_{i=1}^n \mu_i \phi_i(s) \phi_i(t)$. As K is uniformly convergent, for any $\epsilon > 0$ there exists $n_\epsilon \in \mathbb{N}$ such that for any $s, t \in [0, 1]$:

$$\left| K(s, t) - K_{n_\epsilon}(s, t) \right| < \epsilon/3.$$

Because the basis function ϕ_i is uniformly continuous, there exists $\delta > 0$ such that for any $s, s', t, t' \in [0, 1]$:

$$\left| K_{n_\epsilon}(s, t) - K_{n_\epsilon}(s', t') \right| = \left| \sum_{i=1}^{n_\epsilon} \mu_i \phi_i(s) \phi_i(t) - \sum_{i=1}^{n_\epsilon} \mu_i \phi_i(s') \phi_i(t') \right| < \epsilon/3,$$

whenever $|s - s'| < \delta$ and $|t - t'| < \delta$. Hence,

$$\begin{aligned} \left| K(s, t) - K(s', t') \right| &\leq \left| K(s, t) - K_{n_\epsilon}(s, t) \right| + \left| K_{n_\epsilon}(s, t) - K_{n_\epsilon}(s', t') \right| + \\ &\quad \left| K_{n_\epsilon}(s', t') - K(s', t') \right| < \epsilon. \end{aligned}$$

Appendix C. Real data application on Berkeley growth data

C.1. fANOVA

We can apply the proposed framework to test if there is any significant difference between the mean growth curves of male and female groups. We first use the CLR-transformation in Equation (5) to convert growth warping curves to functions in the \mathbb{L}^2 space, this becomes a classical two-sample problem for functional data, and we adopt the functional ANOVA method for comparison (Zhang, 2013). Without any Gaussian process assumption on the transformed data, we can use a bootstrap approach with 10000 replicates for comparison. The functional ANOVA has two types of bootstrap test statistics: \mathbb{L}^2 -norm-based test statistic and the F -type test statistic. It is found that the corresponding test statistics are 572.9428 and 49.7896, respectively, and the associated p-values for both statistics are less than 10^{-4} . This indicates a significant difference between the mean growth curves of females and males.

C.2. Classification with logistic regression

We can also use the CLR-transformed warping functions as a predictor to classify whether the growth curve is male or female. Table C.2 shows the classification confusion matrix calculated using the first two eigenfunctions. It turns out that 33 out of 39 male growth curves and 47 out of 54 female growth curves were correctly classified by the model. Table C.3 presents the corresponding classification performance with various criteria such as precision, sensitivity, specificity, accuracy, and F-measure. We can see that all these criteria have high values at around 0.85. This desirable performance indicates that the logistic regression using transformed warping functions as an explanatory variable is an appropriate classification method.

Table C.2
Confusion Matrix.

Classification Result	True gender		Total
	Male	Female	
Male	33	7	40
Female	6	47	53
Total	39	54	93

Table C.3
Classification Performance.

TP	FP	FN	TN	precision	sensitivity	specificity	accuracy	F-Measure
33	7	6	47	0.83	0.85	0.87	0.86	0.84
47	6	7	33	0.89	0.87	0.85	0.86	0.88

Appendix D. Calculation of the gradient of $J(\phi)$

To get $\nabla J(\phi)$, we define $\tilde{\phi} = \phi + \epsilon g$, where $\epsilon \in \mathbb{R}$, $g \in L^2(0, 1)$, then,

$$\begin{aligned}
 J(\tilde{\phi}) &= \int_0^1 -2q_1(t)q_2 \left(\int_0^t \exp((\phi + \epsilon g)(s)) ds \right) \sqrt{\exp((\phi + \epsilon g)(t))} dt \\
 &+ \lambda \int_0^1 \int_0^1 \left(\phi(s) - \int_0^1 \phi(u) du \right) c(s, t) \left(\phi(t) - \int_0^1 \phi(u) du \right) ds dt \\
 &= J_1(\tilde{\phi}) + \lambda J_2(\tilde{\phi}).
 \end{aligned}$$

The directional derivative of J in the direction g is given by, $D_g J(\phi) = \langle \nabla J(\phi), g \rangle = \left. \frac{dJ(\tilde{\phi})}{d\epsilon} \right|_{\epsilon=0}$, we calculate it by two parts, the first part:

$$\begin{aligned}
 \frac{dJ_1(\tilde{\phi})}{d\epsilon} &= \int_0^1 -2q_1(t)q_2 \left(\int_0^t \exp((\phi + \epsilon g)(s)) ds \right) \\
 &\int_0^t \exp((\phi + \epsilon g)(u))g(u) du \sqrt{\exp((\phi + \epsilon g)(t))} dt \\
 &- \int_0^1 q_1(t)q_2 \left(\int_0^t \exp((\phi + \epsilon g)(s)) ds \right) \sqrt{\exp((\phi + \epsilon g)(t))} g(t) dt.
 \end{aligned}$$

The second part is:

$$\begin{aligned}
 \frac{dJ_2(\tilde{\phi})}{d\epsilon} &= \int_0^1 \int_0^1 (\phi + \epsilon g)(s)c(s, t)(\phi + \epsilon g)(t) ds dt \\
 &- \int_0^1 (\phi + \epsilon g)(u) du \int_0^1 \int_0^1 (\phi + \epsilon g)(s)c(s, t) ds dt \\
 &- \int_0^1 (\phi + \epsilon g)(u) du \int_0^1 \int_0^1 c(s, t)(\phi + \epsilon g)(t) ds dt \\
 &+ \left(\int_0^1 (\phi + \epsilon g)(u) du \right)^2 \int_0^1 \int_0^1 c(s, t) ds dt.
 \end{aligned}$$

Let $\epsilon = 0$:

$$\begin{aligned}
 D_g J_1(\phi) &= \int_0^1 -2q_1(t)\dot{q}_2 \left(\int_0^t \exp(\phi(s)) ds \right) \int_0^t \exp(\phi(u))g(u) du \sqrt{\exp(\phi(t))} dt \\
 &\quad - \int_0^1 q_1(t)q_2 \left(\int_0^t \exp(\phi(s)) ds \right) \sqrt{\exp(\phi(t))}g(t) dt \\
 &= -2 \int_0^1 \int_0^t q_1(t)\dot{q}_2 \left(\int_0^t \exp(\phi(s)) ds \right) \exp(\phi(u))g(u) \sqrt{\exp(\phi(t))} du dt \\
 &\quad - \int_0^1 q_1(t)q_2 \left(\int_0^t \exp(\phi(s)) ds \right) \sqrt{\exp(\phi(t))}g(t) dt \\
 &= -2 \int_0^1 \int_u^1 q_1(t)\dot{q}_2 \left(\int_0^t \exp(\phi(s)) ds \right) \sqrt{\exp(\phi(t))} dt \exp(\phi(u))g(u) du \\
 &\quad - \int_0^1 q_1(t)q_2 \left(\int_0^t \exp(\phi(s)) ds \right) \sqrt{\exp(\phi(t))}g(t) dt \\
 &= -2 \int_0^1 \int_t^1 q_1(u)\dot{q}_2 \left(\int_0^u \exp(\phi(s)) ds \right) \sqrt{\exp(\phi(u))} du \exp(\phi(t))g(t) dt \\
 &\quad - \int_0^1 q_1(t)q_2 \left(\int_0^t \exp(\phi(s)) ds \right) \sqrt{\exp(\phi(t))}g(t) dt \\
 &= \left\langle g, -2 \exp(\phi(t)) \int_t^1 q_1(u)\dot{q}_2 \left(\int_0^u \exp(\phi(s)) ds \right) \sqrt{\exp(\phi(u))} du \right. \\
 &\quad \left. - q_1(t)q_2 \left(\int_0^t \exp(\phi(s)) ds \right) \sqrt{\exp(\phi(t))} \right\rangle.
 \end{aligned}$$

And,

$$\begin{aligned}
 D_g J_2(\phi) &= \int_0^1 \int_0^1 g(s)c(s,t)\phi(t) ds dt + \int_0^1 \int_0^1 \phi(s)c(s,t)g(t) ds dt \\
 &\quad - \int_0^1 g(u) du \int_0^1 \int_0^1 \phi(s)c(s,t) ds dt - \int_0^1 \phi(u) du \int_0^1 \int_0^1 g(s)c(s,t) ds dt \\
 &\quad - \int_0^1 g(u) du \int_0^1 \int_0^1 c(s,t)\phi(t) ds dt - \int_0^1 \phi(u) du \int_0^1 \int_0^1 c(s,t)g(t) ds dt \\
 &\quad + 2 \int_0^1 \phi(u) du \int_0^1 g(v) dv \int_0^1 \int_0^1 c(s,t) ds dt \\
 &= \langle g, \int_0^1 c(t,s)\phi(s) ds \rangle + \langle g, \int_0^1 \phi(s)c(s,t) ds \rangle - \langle g, \int_0^1 \int_0^1 \phi(s)c(s,u) ds du \rangle \\
 &\quad - \langle g, \int_0^1 \phi(u) du \int_0^1 c(t,s) ds \rangle - \langle g, \int_0^1 \int_0^1 c(s,u)\phi(u) ds du \rangle
 \end{aligned}$$

$$-\langle g, \int_0^1 c(s, t) ds \int_0^1 \phi(u) du \rangle + 2\langle g, \int_0^1 \phi(u) du \int_0^1 \int_0^1 c(s, v) ds dv \rangle.$$

Thus, the gradient is given by

$$\begin{aligned} \nabla J(\phi) = & -2 \exp(\phi(t)) \int_t^1 q_1(u) \dot{q}_2 \left(\int_0^u \exp(\phi(s)) ds \right) \sqrt{\exp(\phi(u))} du \\ & - q_1(t) q_2 \left(\int_0^t \exp(\phi(s)) ds \right) \sqrt{\exp(\phi(t))} + \lambda \left(\int_0^1 c(t, s) \phi(s) ds + \int_0^1 \phi(s) c(s, t) ds \right. \\ & - \int_0^1 \int_0^1 \phi(s) c(s, u) ds du - \int_0^1 \phi(u) du \int_0^1 c(t, s) ds - \int_0^1 \int_0^1 c(s, u) \phi(u) ds du \\ & \left. - \int_0^1 c(s, t) ds \int_0^1 \phi(u) du + 2 \int_0^1 \phi(u) du \int_0^1 \int_0^1 c(s, v) ds dv \right). \end{aligned}$$

In particular, we show two special cases on the covariance structure.

1. h is a diagonal covariance: By setting $c(s, t) = r(t)\delta(t - s)$, we can derive the gradient as:

$$\begin{aligned} \nabla J(\phi) = & -2 \exp(\phi(t)) \int_t^1 q_1(\mu) \dot{q}_2 \left(\int_0^\mu \exp(\phi(s)) ds \right) \sqrt{\exp(\phi(\mu))} d\mu \\ & - q_1(t) q_2 \left(\int_0^t \exp(\phi(s)) ds \right) \sqrt{\exp(\phi(t))} + 2\lambda \left(r(t)\phi(t) \right. \\ & \left. + \int_0^1 \phi(u) du \int_0^1 r(s) ds - r(t) \int_0^1 \phi(s) ds - \int_0^1 \phi(s)r(s) ds \right). \end{aligned}$$

2. h is an isotropic covariance: By setting $c(s, t) = a\delta(t - s)$, we can derive the gradient as:

$$\begin{aligned} \nabla J(\phi) = & -2 \exp(\phi(t)) \int_t^1 q_1(\mu) \dot{q}_2 \left(\int_0^\mu \exp(\phi(s)) ds \right) \sqrt{\exp(\phi(\mu))} d\mu \\ & - q_1(t) q_2 \left(\int_0^t \exp(\phi(s)) ds \right) \sqrt{\exp(\phi(t))} + 2a\lambda \left(\phi(t) - \int_0^1 \phi(s) ds \right). \end{aligned}$$

Appendix E. Penalized registration illustrations

E.1. Isotropic covariance

We here use one example to illustrate the penalized alignment with isotropic covariance kernel. We at first simulate one bimodal function, i.e., $f(t) = z_1 e^{-(t-0.22)^2/2} + z_2 e^{-(t-0.78)^2/2}$, where $z_1, z_2 \sim U(0.75, 1.25), t \in [0, 1]$. Then we obtain two functions $f_i(t) = f(\gamma_i(t))$ with warping functions $\gamma_i(t) = \frac{e^{a_i t} - 1}{e^{a_i} - 1}, i = 1, 2$, where $a_1 = -0.5$ and $a_2 = 2$. In addition, we scale $f_2(t)$ up by 1.1 for better visualization. The functions $f_1(t)$ and $f_2(t)$ are shown as blue and green solid curves, respectively, in Fig. E.12(a). Our goal is find optimal warping function γ^* to minimize the loss function in Equation (12). When $\lambda = 0, 40$, and 80 , the optimal warping functions are calculated using Algorithm 3 and shown in Fig. E.12(b). We can see that when $\lambda = 0$, the optimal warping is very close to the one estimated using dynamic programming; the difference is only about numerical errors. When λ gets larger, the optimal warping is closer to the identity warping $\gamma_{id}(t) = t$ (optimal warping when $\lambda = \infty$). The aligned functions $f_2(\gamma^*(t))$ are also shown in Fig. E.12(a). We can see $f_2(\gamma^*(t))$ is right on the top of $f_1(t)$ when $\lambda = 0$, and only slightly shift from $f_2(t)$ when $\lambda = 80$. These results clearly demonstrate the effectiveness of the penalty term and penalty coefficient in the alignment process.

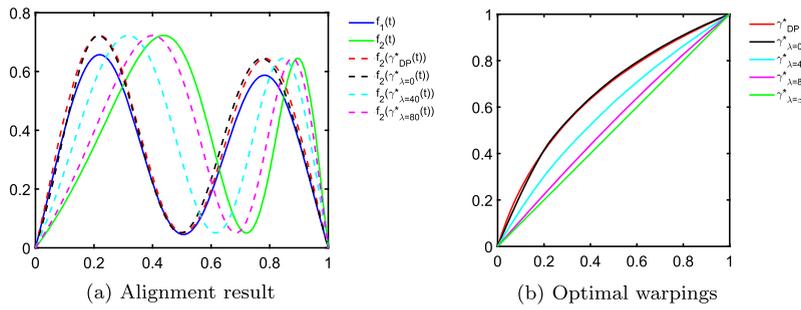


Fig. E.12. Penalized registration illustration with isotropic covariance kernel. (a) Original functions and alignment functions. The blue and green solid curves are the two given functions f_1 and f_2 , respectively, the red dotted curve is the aligned f_2 using dynamic programming, and the black, cyan, and magenta dotted curves are the aligned f_2 using Algorithm 3 with λ equal to 0, 40, 80, respectively. (b) Optimal warping functions in the alignment. The red curve is the optimal warping function from dynamic programming. The black, cyan, magenta, and green curves are the optimal warping functions from Algorithm 3 with λ equal to 0, 40, 80, and ∞ , respectively.

E.2. Full covariance

We have shown two examples to illustrate penalty on time warping using the diagonal terms on the covariance kernel, which essentially describes the variability at each time point. Now we use another example to illustrate the penalized alignment with general non-diagonal covariance which takes into account co-variability between two different time points. In general, $c(s, t)$ in Equation (13) can be any symmetric, positive definite and invertible kernel. To simplify the illustration, we here assume $c(s, t)$ has the following block form:

$$c(s, t) = \begin{cases} a \delta(s - t), & \text{if } s, t \in [0, \frac{1}{2}] \times [0, \frac{1}{2}] \\ b \delta(s - t), & \text{if } s, t \in [\frac{1}{2}, 1] \times [\frac{1}{2}, 1] \\ z \delta(s - t - \frac{1}{2}), & \text{if } s, t \in [\frac{1}{2}, 1] \times [0, \frac{1}{2}] \\ z \delta(s - t + \frac{1}{2}), & \text{if } s, t \in [0, \frac{1}{2}] \times [\frac{1}{2}, 1] \end{cases} \tag{E.1}$$

where $a > 0, b > 0$, and $ab > z^2$. It is easy to verify that $c(s, t)$ is symmetric and positive definite.

Let $\zeta(s) = \phi(s) - \int_0^1 \phi(u) du$, the penalty term in the loss function in Equation (13) with the given $c(s, t)$ can be rewritten as:

$$\begin{aligned} & \int_0^1 \int_0^1 \left(\phi(s) - \int_0^1 \phi(u) du \right) c(s, t) \left(\phi(t) - \int_0^1 \phi(u) du \right) ds dt \\ &= \int_0^{\frac{1}{2}} \left(\sqrt{a} \zeta(s) ds + \sqrt{b} \zeta(s + \frac{1}{2}) \right)^2 ds + 2(z - \sqrt{ab}) \int_0^{\frac{1}{2}} \zeta(s) \zeta(s + \frac{1}{2}) ds. \end{aligned}$$

By fixing the diagonal coefficients a and b , we focus on the penalty with respect to the off-diagonal coefficient z . It is easy to see that 1) if $\int_0^{\frac{1}{2}} \zeta(s) \zeta(s + \frac{1}{2}) ds < 0$, then the penalty is a decreasing function of z , and 2) if $\int_0^{\frac{1}{2}} \zeta(s) \zeta(s + \frac{1}{2}) ds > 0$, then the penalty is an increasing function of z . We now use two simulations to illustrate these two cases, respectively. In each case, we let $a = b = 1$ and set $z = 0.9, 0.2, -0.2, -0.9$ to see how the co-variates will affect the function alignment.

Case 1. $[\int_0^{\frac{1}{2}} \zeta(s) \zeta(s + \frac{1}{2}) ds < 0]$: We first simulate one bimodal function $f_1(t) = 2 \sin(4\pi t), t \in [0, 1]$, and then warp f_1 to get f_2 as follows:

$$f_2(t) = \begin{cases} f_1(0.5\gamma_1(2t)), & \text{if } 0 \leq t \leq 0.5 \\ f_1(0.5\gamma_2(2t - 1) + 0.5), & \text{if } 0.5 < t \leq 1 \end{cases}$$

where $\gamma_i(t) = \frac{e^{a_i t} - 1}{e^{a_i} - 1}, i = 1, 2$, with $a_1 = -5, a_2 = 5$. The functions $f_1(t)$ and $f_2(t)$ are shown as yellow and green solid curves, respectively, in Fig. E.13(a). The optimal warping functions $\gamma^*(t)$ are shown in Fig. E.13(b). We can see that when there is no penalty, the optimal warping can align f_2 to f_1 very well. The corresponding $f_2(\gamma^*(t))$ in Fig. E.13(a) stays right on the top of $f_1(t)$ when there is no penalty. When there is a penalty, same as the previous two simulation examples, the optimal warping also gets closer to the identity warping $\gamma_{id}(t) = t$. In addition, it can be seen that as z becomes smaller, the optimal warping gets closer to the identity warping, and the corresponding $f_2(\gamma^*(t))$ moves further away from f_1 . We point out that

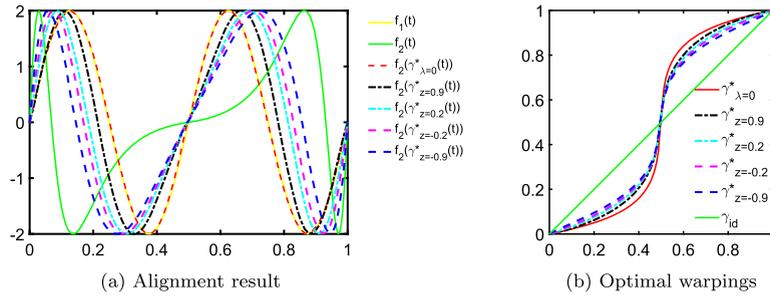


Fig. E.13. Penalized registration illustration with full covariance for Case 1. (a) Original functions and alignment functions. The yellow and green solid curves are the two given functions f_1 and f_2 , respectively, and the red dotted curve is the aligned f_2 without penalty, i.e., $\lambda = 0$. The black, cyan, magenta, and blue dotted curve are the aligned f_2 with penalty term with same $\lambda = 10^5$ and z in Equation (E.1) equal to 0.9, 0.2, -0.2, and -0.9, respectively. (b) Optimal warping functions in the alignment. The red curve is the optimal warping functions with λ equal to 0. The black, cyan, magenta, and blue dotted curve are the optimal warping functions with penalty term with z equal to 0.9, 0.2, -0.2, and -0.9, respectively.

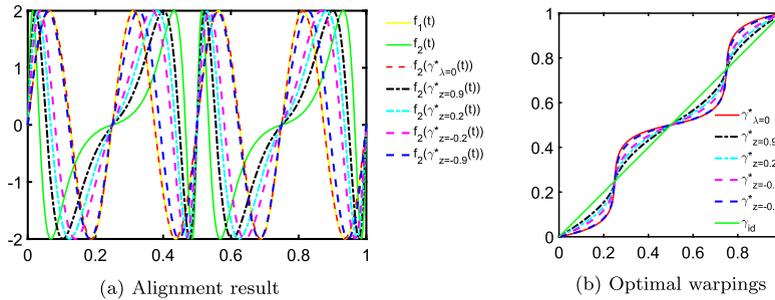


Fig. E.14. Same as Fig. E.13 except for illustration with full covariance for Case 2.

because $2(z - \sqrt{ab}) \int_0^{\frac{1}{2}} \zeta(s)\zeta(s + \frac{1}{2}) ds$ is always non-negative, the off-diagonal terms add more penalty to the warping than the diagonal terms only, given in the integration $\int_0^{\frac{1}{2}} (\sqrt{a}\zeta(s) ds + \sqrt{b}\zeta(s + \frac{1}{2}))^2 ds$.

Case 2. $[\int_0^{\frac{1}{2}} \zeta(s)\zeta(s + \frac{1}{2}) ds \geq 0]$: We first simulate one multi-modal function $f_1(t) = 2 \sin(8\pi t), t \in [0, 1]$, and then warp f_1 to get f_2 as follows:

$$f_2(t) = \begin{cases} f_1(0.25\gamma_1(4t)), & \text{if } t \in [0, 0.25) \\ f_1(0.25\gamma_2(4t - 1) + 0.25), & \text{if } t \in [0.25, 0.5) \\ f_1(0.25\gamma_1(4t - 2) + 0.5), & \text{if } t \in [0.5, 0.75) \\ f_1(0.25\gamma_2(4t - 3) + 0.75), & \text{if } t \in [0.75, 1] \end{cases}$$

where $\gamma_i(t) = \frac{e^{a_i t} - 1}{e^{a_i} - 1}, i = 1, 2$, with $a_1 = -5, a_2 = 5$. The functions $f_1(t)$ and $f_2(t)$ are shown as yellow and green solid curves, respectively, in Fig. E.14(a). The optimal warping functions $\gamma^*(t)$ are shown in Fig. E.14(b). Same as in case 1, when there is no penalty, the optimal warping can align f_2 to f_1 very well. When there is a penalty, same as in the previous two simulation examples, the optimal warping also gets closer to the identity warping $\gamma_{id}(t) = t$. However, in contrast to the result in Case 1, as z becomes larger, the optimal warping gets closer to the identity warping $\gamma_{id}(t) = t$, and the corresponding $f_2(\gamma^*(t))$ moves further away from f_1 . Because now $2(z - \sqrt{ab}) \int_0^{\frac{1}{2}} \zeta(s)\zeta(s + \frac{1}{2}) ds$ is always non-positive, the off-diagonal terms reduce penalty from the diagonal terms. This further demonstrates of the effect of the off-diagonal terms in penalized registration.

E.3. Full covariance vs. diagonal

We have presented two examples to elucidate the impact of non-diagonal terms in the covariance kernel during penalized function alignment. It can be seen that these non-diagonal terms can either diminish or amplify the influence of diagonal terms. We now introduce a slightly more complex example that further illustrates how we can manipulate non-diagonal terms to alter the penalty across time points and compare the alignment result with a diagonal covariance.

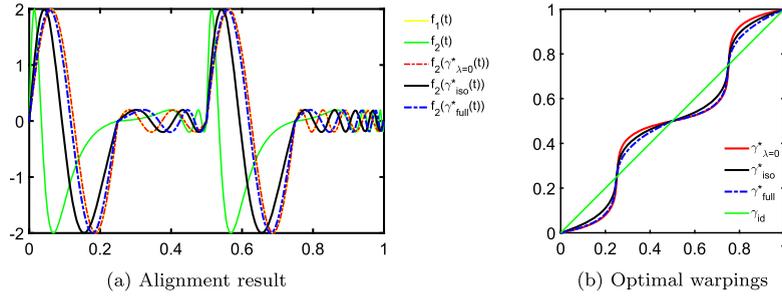


Fig. E.15. Illustration of penalized registration comparing full covariance and isotropic cases. (a) Original functions and alignment functions. The yellow and green solid curves are the two given functions f_1 and f_2 , respectively, and the red dotted curve is the aligned f_2 without penalty, i.e., $\lambda = 0$. The black and blue dotted curve are the aligned f_2 with penalty term using isotropic and full covariance, respectively. (b) Optimal warping functions in the alignment. The red curve is the optimal warping functions without penalty. The black and blue dotted curve are the optimal warping functions with isotropic and full covariance, respectively.

We at first simulate function $f_1(t) = \begin{cases} 2 \sin(8\pi t), & \text{if } t \in [0, 0.25) \cup [0.5, 0.75) \\ 0.2 \sin(16\pi t), & \text{if } t \in [0.25, 0.5) \\ 0.2 \sin(32\pi t), & \text{if } t \in [0.75, 1] \end{cases}$, and then warp f_1 to get f_2 :

$$f_2(t) = \begin{cases} f_1(0.25\gamma_1(4t)), & \text{if } t \in [0, 0.25) \\ f_1(0.25\gamma_2(4t - 1) + 0.25), & \text{if } t \in [0.25, 0.5) \\ f_1(0.25\gamma_1(4t - 2) + 0.5), & \text{if } t \in [0.5, 0.75) \\ f_1(0.25\gamma_2(4t - 3) + 0.75), & \text{if } t \in [0.75, 1] \end{cases}$$

where $\gamma_i(t) = \frac{e^{a_i t} - 1}{e^{a_i} - 1}$, $i = 1, 2$, with $a_1 = -5, a_2 = 5$. The functions $f_1(t)$ and $f_2(t)$ are shown as yellow and green solid curves, respectively, in Fig. E.15(a).

In this example, we assume $c(s, t)$ has the following block form:

$$c(s, t) = \begin{cases} a \delta(s - t), & \text{if } s, t \in A_1 \\ -b \delta(s - t + \frac{1}{2}), & \text{if } s, t \in [0, \frac{1}{4}) \times [\frac{1}{2}, \frac{3}{4}) \\ -b \delta(s - t - \frac{1}{2}), & \text{if } s, t \in [\frac{1}{2}, \frac{3}{4}) \times [0, \frac{1}{4}) \\ b \delta(s - t + \frac{1}{2}), & \text{if } s, t \in [\frac{1}{4}, \frac{1}{2}) \times [\frac{3}{4}, 1] \\ b \delta(s - t - \frac{1}{2}), & \text{if } s, t \in [\frac{3}{4}, 1] \times [\frac{1}{4}, \frac{1}{2}) \\ 0, & \text{o.w.} \end{cases} \tag{E.2}$$

where $A_1 = [0, \frac{1}{4}) \times [0, \frac{1}{4}) \cup [\frac{1}{4}, \frac{1}{2}) \times [\frac{1}{4}, \frac{1}{2}) \cup [\frac{1}{2}, \frac{3}{4}) \times [\frac{1}{2}, \frac{3}{4}) \cup [\frac{3}{4}, 1] \times [\frac{3}{4}, 1]$. By assigning values of $a = 1, b = 0.9$, we achieve the optimal warping, as illustrated in Fig. E.15. Fig. E.15(b) showcases the optimal warping functions $\gamma_{\lambda=0}^*$. Notably, in the absence of any penalty, the optimal warping effectively aligns f_2 with f_1 , as evident from the overlap of $f_2(\gamma_{\lambda=0}^*)$ with $f_1(t)$ in Fig. E.15(a).

However, with the introduction of a penalty, the behavior varies between the diagonal (isotropic in this example) and full covariance scenarios. In the isotropic case, where the covariance is defined as $a\delta(s - t), (s, t) \in [0, 1] \times [0, 1]$, the optimal warping approaches the identity warping $\gamma_{id}(t) = t$ across different points, as illustrated by the black curve in Fig. E.15(b). Nevertheless, in the full covariance case, the optimal warping still closely resembles the penalty-free scenario in the sub-domain $[0, \frac{1}{4}) \times [\frac{1}{2}, \frac{3}{4})$. In contrast to the isotropic case, where the penalty on time s and time t is $a\phi(s)^2 + a\phi(t)^2$, here the penalty becomes $a\phi(s)^2 + a\phi(t)^2 - 2b\phi(s)\phi(t)$, resulting in a diminished impact of the penalty introduced by diagonal terms. On the other hand, in the sub-domain $[\frac{1}{4}, \frac{1}{2}) \times [\frac{3}{4}, 1]$, the penalty becomes $a\phi(s)^2 + a\phi(t)^2 + 2b\phi(s)\phi(t)$. This amplifies the influence of the penalty, causing the optimal warping to be more constrained by the penalty. Consequently, the flexibility of the optimal warping here is limited in the sub-domain $[\frac{1}{4}, \frac{1}{2}) \times [\frac{3}{4}, 1]$.

Appendix F. Functional boxplot

We aim to investigate the impact of the inflation factor ρ on outlier detection outcomes. In this context, we employ the example provided in Section 4.2 to demonstrate these effects. We have chosen values for ρ of 1.28, 1.82, 2.08, and 2.9 to correspond to the inclusion of 68%, 90%, 95%, and 99.7% of the observations, assuming that the coefficient vectors adhere to a multivariate Gaussian distribution. The findings are presented in Fig. F.16. It becomes apparent that by increasing the value of ρ , we are able to detect an decreasing number of outliers.

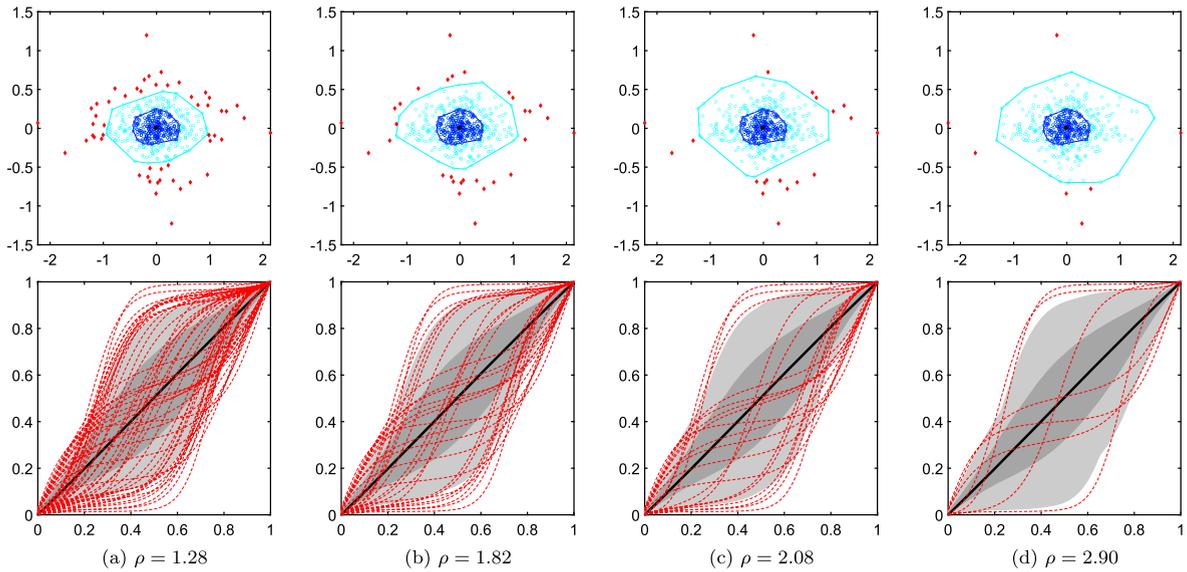


Fig. F.16. Results Demonstrating the Impact of ρ in Outlier Detection, with 2 layer convex hulls in the top row and the warping boxplot generated by Algorithm 4 in the bottom row, the red dashed curves are the outliers candidates.

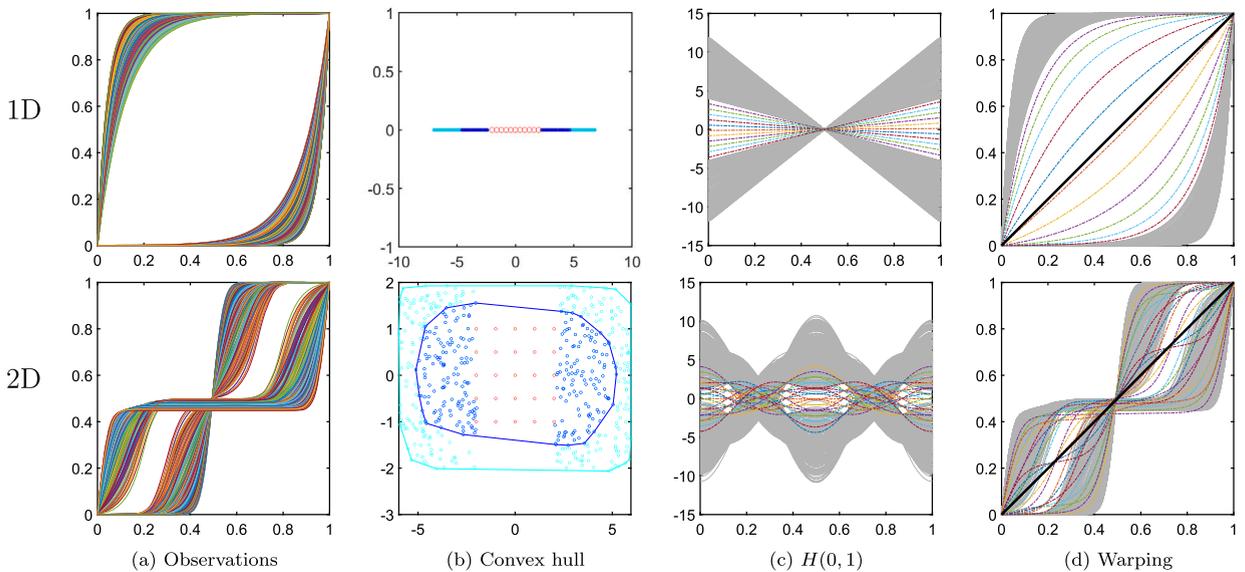


Fig. G.17. Results on Interpolation with 1D case in top row and 2D case in the bottom row. (a) the curves show simulated 500 time warping functions, (b) 2 layer convex hull and the red dots are uniformly random points (c) The corresponding curves in $H(0,1)$ space. (d) The corresponding warping of the red dots in (c). The bold black curve is the corresponding Tukey's depth median.

Appendix G. Functional interpolation

Functional interpolation is a process of determining the unknown functions that lie in between the known functions, and it can be easily achieved in the process of constructing the boxplot using fPCA method. In Fig. G.17 (a), the set of warping observations can be divided into two groups in both case. Specifically, in 1D case (1st row), one group of warping is convex function and the other is concave, and in 2D case (2nd row), one group of warping is arcsin-ish function, the other arctangent-ish. This can be easily seen in the fPCA results shown in Panel (b). The blue and cyan dots represent warping observations in Panel (a). And it is obvious that these points can be grouped into two clusters. It is difficult for us to get the functional interpolation directly in the warping space. Whereas by uniformly sampling (red) points in the gap between the two groups, we can easily plot out the corresponding clr-transformed warpings and the warpings in $H(0, 1)$ and Γ_1 spaces, respectively.

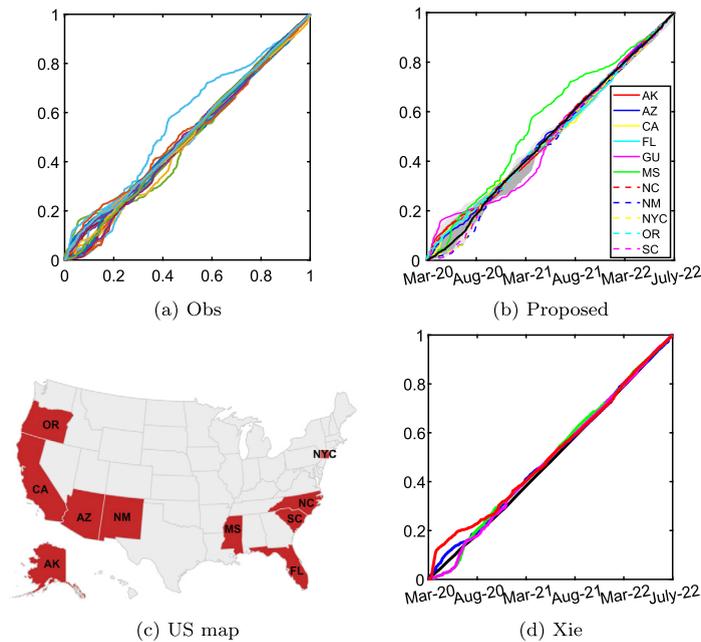


Fig. H.18. Result on COVID-19 data. (a) The curves represent the estimated warping function for the derivative functions of the covid data, the x-axis is displayed from March 2020 to July 2022. (b) The proposed boxplot and outlier candidates generated using Algorithm 4. (c) The US state map, where the red ones are the identified outlier states. (d) The boxplot and outlier candidates generated using the Xie method.

Appendix H. Boxplot for the rate of increase in the COVID data

Apart from linearly transforming the COVID case growth functions into warping functions, we can get the warping functions by performing the Fisher-Rao functional alignment (Srivastava et al., 2011) to the derivative functions of the growth function. The estimated curves are shown in Panel (a) of Fig. H.18. We will use the proposed method in Section 4 to construct the boxplot and identify locations with different rate of increase patterns compared to the other states.

In this example, the first three, seven, and nine principal components explained more than 54%, 83%, and 95% of the total variance, respectively. However, we have only 55 data points. Due to the sparsity of the data, we cannot construct the Convex hull in high dimensions very effectively. Therefore, we only use the first three principal components to construct the boxplot and set the inflation factor to 2.99. Fig. H.18 (b) shows the boxplot constructed by our proposed method. To visualize the difference between the identified outliers and the main body, we superimpose the 11 identified outliers on the boxplot. They are Alaska, Arizona, California, Florida, Puerto Rico, Mississippi, North Carolina, New Mexico, New York City, Oregon, and South Carolina. In Panel (c), we showed the map of the United States and highlighted locations identified as outliers in red.

As a comparison, we show the boxplot generated using the Xie Method in Panel (d), where the four boundaries are overlapped with each other and are difficult to interpret. This method detects no outliers.

References

- Van den Boogaart, K.G., Egozcue, J.J., Pawlowsky-Glahn, V., 2014. Bayes Hilbert spaces. *Aust. N. Z. J. Stat.* 56, 171–194.
- Cheng, W., Dryden, I.L., Hitchcock, D.B., Le, H., 2014. Analysis of proteomics data: Bayesian alignment of functions. *Electron. J. Stat.* 8, 1734–1741.
- Cheng, W., Dryden, I.L., Huang, X., 2016. Bayesian registration of functions and curves. *Bayesian Anal.* 11, 447–475.
- Egozcue, J.J., Díaz-Barrero, J.L., Pawlowsky-Glahn, V., 2006. Hilbert space of probability density functions based on Aitchison geometry. *Acta Math. Sin.* 22, 1175–1182.
- Eilers, P.H., 2004. Parametric time warping. *Anal. Chem.* 76, 404–411.
- Gervini, D., 2015. Warped functional regression. *Biometrika* 102, 1–14.
- Gervini, D., Gasser, T., 2004. Self-modelling warping functions. *J. R. Stat. Soc., Ser. B, Stat. Methodol.* 66, 959–971.
- Hadjipantelis, P.Z., Aston, J.A., Müller, H.G., Moriarty, J., 2014. Analysis of spike train data: a multivariate mixed effects model for phase and amplitude. *Electron. J. Stat.* 8, 1797–1807.
- Happ, C., Scheipl, F., Gabriel, A.A., Greven, S., 2019. A general framework for multivariate functional principal component analysis of amplitude and phase variation. *Stat* 8, e220.
- Hsing, T., Eubank, R., 2015. *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*, vol. 997. John Wiley & Sons.
- Hyndman, R.J., Shang, H.L., 2010. Rainbow plots, bagplots, and boxplots for functional data. *J. Comput. Graph. Stat.* 19, 29–45.
- James, G.M., 2007. Curve alignment by moments. *Ann. Appl. Stat.* 1, 480–501.
- Kurtek, S., 2017. A geometric approach to pairwise Bayesian alignment of functional data using importance sampling. *Electron. J. Stat.* 11, 502–531.
- Lee, S., Jung, S., 2016. Combined analysis of amplitude and phase variations in functional data. *ArXiv preprint arXiv:1603.01775*.
- Leonard, T., 1978. Density estimation, stochastic processes and prior information. *J. R. Stat. Soc., Ser. B, Methodol.* 40, 113–132.
- Lu, Y., Herbei, R., Kurtek, S., 2017. Bayesian registration of functions with a Gaussian process prior. *J. Comput. Graph. Stat.* 26, 894–904.

- Maier, E.M., Stöcker, A., Fitzenberger, B., Greven, S., 2021. Additive density-on-scalar regression in Bayes Hilbert spaces with an application to gender economics. ArXiv preprint arXiv:2110.11771.
- Marron, J.S., Ramsay, J.O., Sangalli, L.M., Srivastava, A., 2015. Functional data analysis of amplitude and phase variation. *Stat. Sci.*, 468–484.
- Matuk, J., Bharath, K., Chkrebti, O., Kurtek, S., 2021. Bayesian framework for simultaneous registration and estimation of noisy, sparse, and fragmented functional data. *J. Am. Stat. Assoc.*, 1–17.
- Menafoglio, A., Guadagnini, A., Secchi, P., 2016. Stochastic simulation of soil particle-size curves in heterogeneous aquifer systems through a Bayes space approach. *Water Resour. Res.* 52, 5708–5726.
- Menafoglio, A., Hron, K., 2014. Bayes spaces. In: *Wiley StatsRef: Statistics Reference Online*, pp. 1–5.
- Ramsay, J., Silverman, B., 2006. *Functional Data Analysis*. Springer, New York.
- Ramsay, J.O., Li, X., 1998. Curve registration. *J. R. Stat. Soc., Ser. B, Stat. Methodol.* 60, 351–363.
- Srivastava, A., Klassen, E.P., 2016. *Functional and Shape Data Analysis*, vol. 1. Springer.
- Srivastava, A., Wu, W., Kurtek, S., Klassen, E., Marron, J.S., 2011. Registration of functional data using Fisher-Rao metric. ArXiv preprint arXiv:1103.3817.
- Sun, Y., Genton, M.G., 2011. Functional boxplots. *J. Comput. Graph. Stat.* 20, 316–334.
- Tucker, J.D., Shand, L., Chowdhary, K., 2021. Multimodal Bayesian registration of noisy functions using Hamiltonian Monte Carlo. *Comput. Stat. Data Anal.* 163, 107298.
- Tucker, J.D., Wu, W., Srivastava, A., 2013. Generative models for functional data using phase and amplitude separation. *Comput. Stat. Data Anal.* 61, 50–66.
- Tukey, J.W., 1975. Mathematics and the picturing of data. In: *Proceedings of the International Congress of Mathematicians. Vancouver, 1975*, pp. 523–531.
- Wu, W., Srivastava, A., 2014. Analysis of spike train data: alignment and comparisons using the extended Fisher-Rao metric. *Electron. J. Stat.* 8, 1776–1785.
- Xie, W., Kurtek, S., Bharath, K., Sun, Y., 2017. A geometric approach to visualization of variability in functional data. *J. Am. Stat. Assoc.* 112, 979–993.
- Zhang, J.T., 2013. *Analysis of Variance for Functional Data*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. CRC Press.