Bayesian Registration of Functions with a Gaussian Process Prior

Yi Lu, Radu Herbei and Sebastian Kurtek
Department of Statistics
The Ohio State University
Columbus, OH

January 13, 2017

Abstract

We present a Bayesian framework for registration of real-valued functional data. At the core of our approach is a series of transformations of the data and functional parameters, developed under a differential geometric framework. We aim to avoid discretization of functional objects for as long as possible, thus minimizing the potential pitfalls associated with high-dimensional Bayesian inference. Approximate draws from the posterior distribution are obtained using a novel Markov chain Monte Carlo (MCMC) algorithm which is well suited for estimation of functions. We illustrate our approach via pairwise and multiple functional data registration, using both simulated and real data sets.

Keywords: Function registration; time warping; Markov chain Monte Carlo.

1. INTRODUCTION

Registration of functional data refers to the process of matching two (pairwise registration) or more (multiple registration) functions, usually done as a pre-processing step, followed by other statistical procedures. Function registration is referred to by many names in various communities, including *matching, amplitude-phase separation, reparameterization*, etc. It has a wide range of applications. For example, in image registration, each function is defined on a unit square or cube domain, and one is interested in matching pixel or voxel values across images. In shape registration, the functions are parameterized Euclidean curves, and one would like to match common geometric features of the curves. In this work, we consider the classical case of univariate functions defined on a subinterval of the real line; registration of such functional data aims to match common features (i.e., peaks and valleys) for improved subsequent statistical inference.

As a motivating example, we consider the growth rate curves from the Berkeley growth study (Tuddenham and Snyder 1954) available in the R package *fda* (Ramsay, Wickham, Graves and Hooker 2014). Figure 1 provides an example of growth rate curves for two girls. The two functions have very similar structure (number of peaks and valleys) but the timing of these features is different as seen in the left panel. One can use function registration to align the timing of features across the two curves such that they are better aligned; this is shown in the middle panel. This is achieved using a special function called a warping function, which is displayed in the right panel for this example. The resulting alignment improves comparison of functions by accounting for the confounding factor of when important features occur on the two curves. The Berkeley growth study has been used as a motivating example in other function registration papers including Ramsay and Li (1998), Telesca and Inoue (2008) and Cheng et al. (2016).

Other application examples that require registration of real-valued functions include handwriting analysis (Kneip et al. 2000; Ramsay 2000), environmetrics such as the study of monthly temperature curves (Ramsay and Silverman 2005), time course microarray and proteomics data analysis (Telesca and Inoue 2008; Koch et al. 2014), spike train analysis in neuroscience (Wu et al. 2014), biomechanical experiments (Ramsay and Li 1998; Ramsay, Gribble and Kurtek 2014), medical imaging (Sangalli et al. 2014), and many others. A textbook explanation of the functional data registration problem, as well as descriptions of many

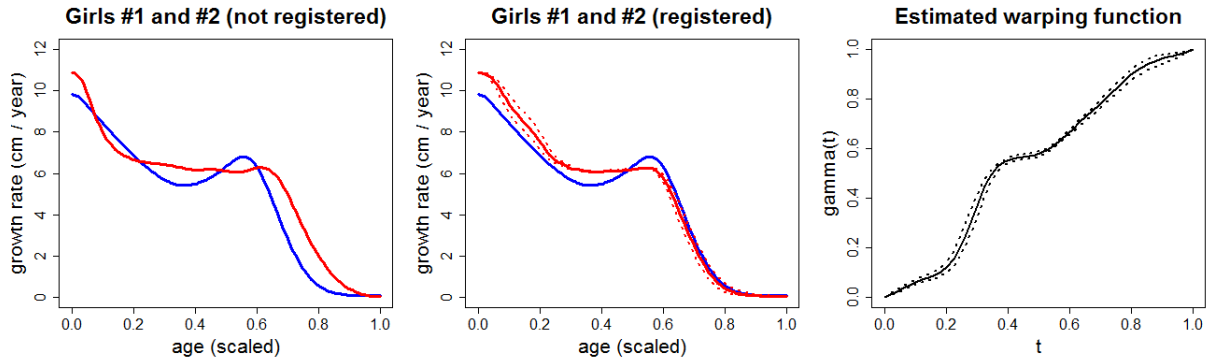

Figure 1: *Growth rate curves of two girls from the Berkeley growth study dataset. Curves before and after registration are shown in the left and middle panels, respectively. The right panel displays the estimated warping function γ (posterior mean). The dotted lines in the middle and right panel represent* 95% *credible intervals.*

of the aforementioned examples, can be found in Ramsay and Silverman (2002, 2005). The area of application we focus on in this work is concerned with analyzing biomechanical data in the form of gait cycles during a walking sequence of test subjects. Five common methods to perform function registration for gait cycle data are compared in Helwig et al. (2011). We discuss their drawbacks in detail in Section 5.3.

The main underlying goal in function registration is to separate two types of variation present in functional data: amplitude or y-axis variation and phase or x-axis variation. Amplitude refers to the magnitude of functions at different points of interest on the domain, and is invariant to the timing of these features. Phase variation refers to the difference in timing of the features across functions (Ramsay and Li 1998), and can be induced via domain warping. Function registration methods can be categorized by whether or not user-specified points of interest, called landmarks, are required. For landmark-based registration, users identify a few marker events usually corresponding to some salient or mathematical features (e.g., local maxima and minima in temperature functions). Registration is then performed such that those events occur at the same time (Kneip et al. 2000). Examples of such methods include Sadeghi et al. (2000) and Piecewise Linear Length Normalization Helwig et al. (2011). The main disadvantage of these approaches is that landmarks often must be specified manually, which can be extremely time consuming, especially for large datasets.

Other methods adopt automatic marker-less registration and align the full target functions. Many of these methods are variational in nature and include Ramsay and Li (1998), Kneip et al. (2000), Gervini and Gasser (2004), Liu and Müller (2004), James (2007), Tang and Müller (2008), Kneip and Ramsay (2008), Srivastava et al. (2011b), Kurtek et al. (2011), and Raket et al. (2014). The proposed method belongs to the second category and does not require user-specified markers.

Most classical methods (all mentioned in the previous paragraph) are non-Bayesian in nature. Bayesian registration of functions is a relatively new concept, see Telesca and Inoue (2008), Kurtek (2015), and Cheng et al. (2016) for the most recent advances. Main benefits of a Bayesian approach include a comprehensive exploration of the warping function parameter space, principled uncertainty analysis via posterior inference (e.g., the discovery of multiple plausible alignments), and the ability to incorporate prior information into the model (e.g., marker locations). Another emerging trend in all functional data analysis is to build models and algorithms that avoid discretization of functions until the last possible step (Stuart 2010; Cotter et al. 2013). We aim to contribute to the field of functional data registration by constructing a Bayesian function registration model that avoids discretization of the infinite-dimensional parameter space (made precise later) until the last step. We utilize a new Z-mixture preconditioned Crank-Nicolson (pCN) algorithm, modified from the pCN algorithm presented in Cotter et al. (2013), which allows efficient sampling from the posterior distribution on the space of warping functions. The proposed method is defined for both pairwise and multiple function registration.

The rest of this paper is organized as follows. We formally set-up the registration problem in Section 2, in which we introduce the different function spaces of interest. In Section 3, we derive a Bayesian model for pairwise registration and present a Metropolis within Gibbs algorithm to sample from the posterior distribution. Section 4 extends this model to the case of multiple function registration. Section 5 evaluates the proposed approach using simulated and real data examples. We focus on an application to a comprehensive gait cycle dataset, where the need for function registration is clear. We close with a discussion as well as possible directions for future work in Section 6.

2. NOTATION AND GENERAL SET-UP

We first consider the case of pairwise registration where we observe two real-valued functions, denoted by f_1 and f_2, defined on a common interval $[a, b]$; without loss of generality, we restrict our analysis to the interval $[0, 1]$. We seek a mapping $\gamma : [0, 1] \mapsto [0, 1]$ such that f_1 and $f_2 \circ \gamma$ are *optimally matched*, where the optimality criterion is defined later. The function γ is called a *warping function* and is an orientation preserving diffeomorphism of $[0, 1]$. In this setting, we are *aligning* f_2 *to* f_1. One could also consider aligning f_1 to f_2, and, in general, the two problems are not equivalent. We return to this issue in Section 2.1. The role of γ is to warp the domain of the observed function f_2 by composition, $f_2 \circ \gamma$. The functional values of f_2 are retained, but f_2 travels to those values at a different speed (e.g., f_2 and $f_2 \circ \gamma$ will both reach the same local maxima, but $f_2 \circ \gamma$ may reach them faster or slower, depending on γ). In this framework, we regard f_1 and f_2 as data, belonging to the observation space $\mathcal{F} = \{f : [0, 1] \mapsto \mathbb{R} \mid f \text{ is absolutely continuous}\}$. The warping function γ is viewed as a parameter to be estimated, belonging to the Lie group $\Gamma = \{\gamma : [0, 1] \mapsto [0, 1] \mid \gamma(0) = 0, \gamma(1) = 1,\ 0 < \gamma' < \infty\}$ where the identity element is $\gamma_{id}(t) = t$ and inverse element γ^{-1} is the usual inverse. This is also known as the *reparameterization* or *warping group*.

Function registration has traditionally been treated as an optimization problem, commonly referred to as *Dynamic Time Warping (DTW)* (see Wang et al. 1997, Keogh and Pazzani 2001, Senin 2008, and Clifford et al. 2009). In this setting, the estimate of γ is

$$\hat{\gamma} = \arg\inf_{\gamma \in \Gamma} \Big(d(f_1, f_2 \circ \gamma) + \lambda \mathtt{R}(\gamma) \Big),$$

where λ is a tuning parameter, $\mathtt{R}(\gamma)$ is a regularization term usually involving the first derivative of γ, and d is the $\mathbb{L}^2$–metric on $\mathcal{F}$ (we maintain this notation throughout the manuscript).

Our aim in this work is to describe a Bayesian model for estimating the warping function γ given the data f_1 and f_2. The ingredients we require are (i) a likelihood that measures the misregistration between two functions, (ii) a prior distribution on the function space Γ, and (iii) a Markov chain Monte Carlo (MCMC) algorithm for exploring the posterior distribution. Before describing these three components we require several transformations, both for the data f_1 and f_2, and for the parameter γ. These transformations are motivated

by the geometry of the function spaces described above, and are described in the following two subsections.

Discretization. At the computer implementation stage, some level of discretization is always necessary. Throughout this paper, we assume that the functions to be registered (f_1 and f_2) are observed on a given, finite grid of size N. Following ideas from Stuart (2010) and Cotter et al. (2013), we describe a sampling algorithm which is scale-free and thus unaffected by the need to also discretize the function γ. We use $h([t])$ to denote the values of a general function h (defined on a continuous interval) at a discrete set of time points $[t] = \{t_1, \ldots, t_N\}$. Similarly, $\int_0^{[t]} h(s)ds$ denotes the N-dimensional vector $\left\{\int_0^{t_1} h(s)ds, \ldots, \int_0^{t_N} h(s)ds\right\}$.

2.1 Observation Space Transformation

For $f_1, f_2 \in \mathcal{F}$ and $\gamma \in \Gamma$, one can easily verify that, in general, $d(f_1, f_2) \neq d(f_1 \circ \gamma, f_2 \circ \gamma)$. It follows that

$$\inf_{\gamma_1,\gamma_2\in\Gamma} d(f_1 \circ \gamma_1, f_2 \circ \gamma_2) \neq \inf_{\gamma_1\in\Gamma} d(f_1 \circ \gamma_1, f_2) \neq \inf_{\gamma_2\in\Gamma} d(f_1, f_2 \circ \gamma_2) \ .$$

This leads to an undesirable registration asymmetry since it is not always clear if one should align f_1 to f_2 or f_2 to f_1. One could consider a double optimization approach, i.e., search for a γ_1 and γ_2 to minimize $d(f_1 \circ \gamma_1, f_2 \circ \gamma_2)$. However, in this case one can find γ_1 and γ_2 such that $f_1 \circ \gamma_1$ and $f_2 \circ \gamma_2$ are arbitrarily close, by "pinching" them to have different values only on an arbitrarily small interval (termed the *pinching effect* in Marron et al. 2015). From a geometric perspective, let $[f] = \{f \circ \gamma \ \mid \ \gamma \in \Gamma\}$ be the equivalence class of f induced by Γ. Under the $\mathbb{L}^2$–metric $d(\cdot, \cdot)$, two equivalence classes $[f_1]$ and $[f_2]$ are "not parallel" (i.e., $d(f_1, f_2) \neq d(f_1 \circ \gamma, f_2 \circ \gamma)$) and thus can get arbitrarily close to each other (i.e., the pinching effect). This indicates that d is not an appropriate distance to measure how close f_1 and $f_2 \circ \gamma$ are to each other. In order to address this, we suggest transforming f_1 and f_2 to their *square-root velocity functions* (SRVFs).

The SRVF framework was originally developed for the purpose of shape analysis in Srivastava et al. (2011a) and later extended to the case of univariate functions in (Srivastava et al. 2011b). A function $f \in \mathcal{F}$ is mapped to its SRVF $Q(f) \in \mathbb{L}^2([0, 1])$ defined as

$$Q(f) : [0, 1] \to \mathbb{R}, \quad Q(f)(t) = \text{sign}(f'(t))\sqrt{|f'(t)|} \ , \tag{1}$$

where f' denotes the derivative of f. To simplify notation, let $q \equiv Q(f)$. Note that Q is a bijective mapping up to a translation. Given any $q \in \mathbb{L}^2([0,1])$, the corresponding f is determined uniquely up to a constant $f(0)$, by the inverse mapping

$$Q^{-1}(q)(t) = f(0) + \int_0^t q(s)|q(s)|\, ds \ .$$

For $f \in \mathcal{F}$, the SRVF $q = Q(f)$ is square integrable. Similarly, $f \circ \gamma \in \mathcal{F}$ is mapped to

$$Q(f \circ \gamma)(t) = Q(f)\left(\gamma(t)\right)\sqrt{|\gamma'(t)|} \equiv (q, \gamma) \ . \tag{2}$$

One can easily verify that $d(q_1, q_2) = d((q_1, \gamma), (q_2, \gamma))$, and thus $[q_1]$ and $[q_2]$ are "parallel" under the $\mathbb{L}^2$–metric, where $[q] = \{(q, \gamma) \ \mid \ \gamma \in \Gamma\}$ is the equivalence class of q induced by Γ. The collection of equivalence classes $[q]$ is called a quotient space, denoted by $\mathbb{L}^2([0,1])/\Gamma$. It is referred to as the amplitude space, since each $[q]$ corresponds to a collection of functions in $\mathcal{F}$ with the same amplitude. Thus, instead of searching for a warping function γ such that f_1 and $f_2 \circ \gamma$ are close in the $\mathbb{L}^2$–sense, we aim for $q_1 = Q(f_1)$ and $(q_2, \gamma) = Q(f_2 \circ \gamma)$ to be close under the $\mathbb{L}^2$–distance. This avoids the aforementioned issues of pinching and asymmetry of registration solutions.

2.2 Parameter Space Transformation

There are two traditional approaches to selecting a prior distribution over a functional space. If one is interested in specifying a certain level of smoothness, a Gaussian Process (GP) prior is more suitable, whereas basis function representations are typically used when one is interested in capturing certain details. In each case, there is a vast literature, covering various applications: see, for example, Telesca and Inoue (2008), Claeskens et al. (2010), and Kurtek (2015). In this application, we require that γ satisfies rather restrictive conditions, and thus neither of the two solutions above are applicable directly (note that Γ is not a linear space). In addition, we aim to exploit the Riemannian-geometric structure of Γ when performing posterior inference. To that end, another transformation is required.

For $\gamma \in \Gamma$, let $\psi = Q(\gamma) = \sqrt{\gamma'}$ be the corresponding SRVF (note that $\text{sign}(\gamma') = 1$). The resulting SRVF space, denoted by $\Psi^+ = \{Q(\gamma) \mid \gamma \in \Gamma\}$, is called the space of *square-root densities* (SRD). This special case of the SRVF representation was first proposed by Bhattachayya (1943) to simplify the geometry of the space of probability density functions,

and has been discussed more recently in Srivastava et al. (2007), Kurtek (2015), and Kurtek and Bharath (2015). Since $\gamma(0) = 0$, this mapping is a bijection and a unique γ can be reconstructed from a given $\psi \in \Psi^+$ as $Q^{-1}(\psi)(t) = \int_0^t \psi^2(s)ds$. We also note two properties of the resulting SRD functions $\psi \in \Psi^+$: (1) it is positive, and (2) it has unit $\mathbb{L}^2$–norm since

$$\|\psi\|^2 = \int_0^1 \psi(t)^2 dt = \int_0^1 \gamma'(t)dt = \gamma(1) - \gamma(0) = 1.$$

Thus, the SRD space $\Psi^+ = \{\psi : [0,1] \mapsto \mathbb{R}^+ \ | \ \|\psi\|^2 = 1\}$ is the positive orthant of the unit sphere in the Hilbert space $\mathbb{L}^2([0,1])$ denoted by $\Psi = \{\psi : [0,1] \mapsto \mathbb{R} \ | \ \|\psi\|^2 = 1\}$. On the space Ψ, we use the arclength distance, which is equivalent to using the Fisher-Rao metric on the space Γ (see Srivastava et al. 2011b):

$$d_{FR}(\gamma_1, \gamma_2) = d(\psi_1, \psi_2) = \cos^{-1}(\langle \psi_1, \psi_2 \rangle) = \cos^{-1}\left(\int_0^1 \psi_1(t)\psi_2(t)dt \right). \tag{3}$$

In addition, since we have defined a distance on Γ, we can also define the Karcher mean function $\bar{\gamma}_n$ (Karcher 1977) for a set of warping functions $\{\gamma_1, \ldots, \gamma_n\}$ via

$$\bar{\gamma}_n = \text{argmin}_{\gamma \in \Gamma} \sum_{i=1}^n d_{FR}(\gamma, \gamma_i)^2. \tag{4}$$

A gradient-based algorithm to calculate this Karcher mean is provided in Srivastava et al. (2011b). The Karcher mean provides a way to define, and to conveniently calculate, the posterior mean function of a collection of warping functions.

Although the SRD representation has simplified the complicated geometry of Γ, the resulting space Ψ is still nonlinear (for instance, it is clearly not closed under scalar multiplication). To simplify this representation further, we map the unit sphere onto a tangent space (also called "unwrapping"). The identity warping function $\gamma_{id} \in \Gamma$ maps to a constant function $1 \in \Psi$, and the tangent space of Ψ at this point is defined as

$$T_1(\Psi) = \left\{ g : [0,1] \mapsto \mathbb{R} \ | \ \langle g, 1 \rangle = \int_0^1 g(t)dt = 0 \right\}. \tag{5}$$

The tangent space $T_1(\Psi)$ is a linear space (closed under point-wise addition and scalar multiplication with the constant function 0, which is identified with the identity element). Our goal is to represent warping functions in this tangent space, and to assign a Gaussian process prior distribution on this space. One way to connect Ψ and $T_1(\Psi)$ is via the exponential

map and its inverse, which are defined as

$$\exp_1 : T_1(\Psi) \mapsto \Psi \qquad \exp_1(g) = \cos(\|g\|) + \frac{\sin(\|g\|)}{\|g\|} g, \quad g \in T_1(\Psi), \tag{6}$$

$$\exp_1^{-1} : \Psi \mapsto T_1(\Psi) \qquad \exp_1^{-1}(\psi) = \frac{\theta}{\sin(\theta)} \left(\psi - \cos(\theta)\right), \quad \theta = d(1, \psi), \quad \psi \in \Psi. \tag{7}$$

Additional details on the tangent space and its geometry are given in Srivastava et al. (2007) and Kurtek et al. (2012).

Our approach is to use $T_1(\Psi)$ as the parameter space for inference (recall that when we map $T_1(\Psi)$ back to Ψ, we need to truncate it to Ψ^+). The transformations mentioned in Sections 2.1 and 2.2 are summarized in the following diagram:

$$\begin{array}{ccccc} f \in \mathcal{F} & \underset{\mathrm{Q}^{-1}}{\overset{\mathrm{Q}}{\rightleftarrows}} & q \in \mathbb{L}^2, & & \\ \gamma \in \Gamma & \underset{\mathrm{Q}^{-1}}{\overset{\mathrm{Q}}{\rightleftarrows}} & \psi \in \Psi^+ & \underset{\exp_1}{\overset{\exp_1^{-1}}{\rightleftarrows}} & g \in T_1(\Psi). \end{array}$$

3. PAIRWISE REGISTRATION

3.1 Bayesian Model Specification

Suppose we want to register two functions, f_1 and f_2, each observed on a finite grid $[t]$ of length N. As mentioned before, we use the SRVF representation detailed in Section 2. We model the difference $Q(f_1)([t]) - Q(f_2 \circ \gamma)([t])$, which is a vector of length N, by a zero-mean multivariate Gaussian distribution. Using the transformations defined in Section 2.2, we reparameterize the likelihood using g instead of γ and work with

$$Q(f_2 \circ \gamma)(t) = Q(f_2) \left(\int_0^t \exp_1^2(g)(s) ds \right) \exp_1(g)(t). \tag{8}$$

We restrict the prior distribution on g to a subset $A \subset T_1(\Psi)$ defined as

$$g \in A \subset T_1(\Psi) \iff \exp_1(g) > 0. \tag{9}$$

The proposed Bayesian model is fully specified below.

Model 1.

$$Q(f_1)([t]) - Q(f_2) \left(\int_0^{[t]} \exp_1^2(g)(s) ds \right) \exp_1(g)([t]) \;\Big|\; g, \sigma_1^2 \sim N(0_N, \sigma_1^2 I_N),$$

$$g \sim \mathit{Gaussian}(0, \mathcal{C}_g; I_A),$$

$$\sigma_1^2 \sim IG(shape = a, scale = b),$$

where Gaussian$(\cdot, \cdot; I_A)$ *refers to the Gaussian process restricted to the domain A (Equation 9),* $\mathcal{C}_g$ *is a pre-specified covariance operator,* $IG(\cdot, \cdot)$ *refers to the inverse gamma distribution, and* a *and* b *are known constants. A brief discussion on incorporating prior knowledge onto this model is given in the Supplementary Material.*

Note that the prior and posterior distributions of (g, σ_1^2) are probability measures on the product space $T_1(\Psi) \times \mathbb{R}^+$. The prior measure is the product measure $\mu_0 \equiv Gaussian(0, \mathcal{C}_g; I_A) \times IG(a, b)$. The posterior measure, denoted by μ, is absolutely continuous with respect to the prior measure, with the Radon-Nikodym derivative given by the Bayes' formula

$$\frac{d\mu}{d\mu_0}(g, \sigma_1^2) \propto L(g, \sigma_1^2 \mid f_1, f_2). \tag{10}$$

Here, $L(\cdot, \cdot \mid f_1, f_2)$ is the likelihood function given by

$$L(g, \sigma_1^2 \mid f_1, f_2) \propto \left(\frac{1}{\sigma_1^2}\right)^{N/2} \exp\left\{-\frac{1}{2\sigma_1^2}\texttt{SSE}(g)\right\}, \tag{11}$$

where the sum of squared errors (SSE) is defined as

$$\texttt{SSE}(g) = \sum_{i=1}^{N} \left(Q(f_1)(t_i) - Q(f_2)\left(\int_0^{t_i} \exp_1^2(g)(s)ds\right)\exp_1(g)(t_i)\right)^2. \tag{12}$$

The posterior distribution given by (10) is explored using a Metropolis within Gibbs (MwG) algorithm. At each iteration, the component σ_1^2 is updated via a draw from the conditional distribution

$$\sigma_1^2 \mid g, f_1, f_2 \sim IG\left(shape = \frac{N}{2} + a, scale = \frac{1}{2}\texttt{SSE}(g) + b\right). \tag{13}$$

The functional component g is updated using a Metropolis-Hastings (MH) step, following Cotter et al. (2013). The proposal is a linear combination of the current state g and a new draw ξ from the prior distribution $Gaussian(0, \mathcal{C}_g; I_A)$. We use a *Z-mixture pCN algorithm*, which is a modification of the pCN (*preconditioned Crank-Nicolson*) algorithm derived in Cotter et al. (2013) (the original pCN algorithm is a special case where $Z = 1$). The proposal g' at each iteration is of the form

$$g' \sim \sum_{z=1}^{Z} p_z \, Gaussian(g\sqrt{1-\beta_z^2}, \beta_z^2 \mathcal{C}_g; I_A), \tag{14}$$

where g is the current state, $\{\beta_z \in (0,1), z = 1, \ldots, Z\}$ are fixed tuning parameters, and $\{p_z \in (0,1), \sum p_z = 1\}$ are fixed mixture probabilities. In other words, we first draw β_z with probability p_z and set $g' = g\sqrt{(1-\beta_z^2)} + \beta_z \xi$, where ξ is a draw from the prior.

There are two advantages for using the Z-mixture pCN proposal. First, the MH acceptance ratio simplifies to $\rho = 1 \wedge L(g', \sigma_1^2 \mid f_1, f_2)/L(g, \sigma_1^2 \mid f_1, f_2)$, which is independent of the dimension of the discretized g. The pCN algorithm leads to significant speed-ups when the parameter is a function evaluated on a fine grid of size d_u, and is robust to increasing d_u. Alternatively, one can consider a standard random walk proposal of the form $\texttt{prop}^{(t)} = \texttt{curr}^{(t)} + \beta\xi$, where ξ is a draw from the prior. With the random walk proposal, calculation of the acceptance ratio can be done on a grid, but will become arbitrarily slow as the grid size increases. For more information on this issue, we refer the reader to Cotter et al. (2013).

The second advantage of using the Z-mixture proposal, compared to the pCN proposal with a *fixed* β, is that it allows proposals that are sometimes close to the current state (for β_z values that are close to 0), and sometimes far from the current state (for β_z values that are close to 1). This provides a flexible way to control the convergence and mixing rate of the MCMC algorithm. Small values of β result in better discoveries of small, local features while large values of β lead to faster explorations of all regions of the posterior distribution.

3.2 Sampling from the prior measure

The MCMC algorithm in the previous section requires sampling $\xi \sim \mathit{Gaussian}(0, \mathcal{C}_g; I_A)$. This is done via the Karhunen-Loève expansion. The covariance operator $\mathcal{C}_g$ is specified via its eigenpairs $(b_i, \lambda_i^2,\ i \geq 1)$, where $\{b_i(\cdot),\ i \geq 1\}$ forms an orthonormal basis for the function space $T_1(\Psi)$ and $\sum \lambda_i^2 < \infty$. We sample independent variates $\xi_i \sim N(0, \lambda_i^2)$ and set

$$\xi = \sum_{i=1}^{M} \xi_i b_i$$

where the constant $M \in \{1, 2, \ldots\}$ is selected apriori. A simple accept-reject step is required to ensure that $g' \in A$. In our simulations, we find that setting $M \geq 20$ leads to satisfactory results (see Section 5.1). For the remainder of the paper, the eigenfunctions are taken to be the Fourier functions

$$b_i(t) \in \left\{\sqrt{2}\sin(2i\pi t),\ \sqrt{2}\cos(2i\pi t)\right\}, \quad i \geq 1$$

which form an orthonormal basis for the parameter space $T_1(\Psi)$. The eigenvalues are set to $\lambda_i^2 = \sigma_g^2/i^2$, where σ_g^2 is assumed known. A detailed description of the algorithm to perform pairwise registration is provided in the Supplementary Material.

4. MULTIPLE FUNCTION REGISTRATION

We can generalize pairwise registration to the case of multiple registration for $c > 2$ functions. We observe $f_1, \ldots, f_c$ and register them simultaneously to some *unknown template function*, f^*. Each observed function f_i is associated with a warping function γ_i, $i = 1, \ldots, c$. As in the pairwise case, we first transform each observed function f_i to q_i and use q^* to denote the template function on the SRVF space. We also represent the warping functions using $g_1, \ldots, g_c \in T_1(\Psi)$. Each warping function g_i is modeled independently using a Gaussian process prior with mean function 0 and the same covariance operator $\mathcal{C}_g$. The template q^* is also assigned a Gaussian process prior, with mean function 0 and covariance operator $\mathcal{C}_q$. We want to *jointly estimate* the template q^* and the individual warping functions that align $q_1, \ldots, q_c$ to q^*. In contrast, many existing methods either specify a *known* template function (e.g., the cross-sectional mean before registration or one of the sample curves) and only update it once the f_i's are registered (Kneip et al. 2000), or use a Procrustes fitting process that alternates between estimating the warpings and estimating the template (Silverman 1995, Ramsay and Li 1998).

Given the parameters $g_1, \ldots, g_c$ and q^*, we model the differences $q^* - (q_i, \gamma_i)$ with zero-mean normal distributions as in the pairwise case. We assume that these distributions are mutually independent and have the same covariance structure. The full Bayesian model for multiple function registration is given below.

Model 2.

$$Q(f_1)\left(\int_0^{[t]} \exp_1^2(g_1)(s)ds\right)\exp_1(g_1)([t]) - q^*([t]) \;\mid\; q^*, g_1, \sigma_1 \sim N(0_N, \sigma_1^2 I_N),$$

$$\vdots$$

$$Q(f_c)\left(\int_0^{[t]} \exp_1^2(g_c)(s)ds\right)\exp_1(g_c)([t]) - q^*([t]) \;\mid\; q^*, g_c, \sigma_1 \sim N(0_N, \sigma_1^2 I_N),$$

$$g_1, \ldots, g_c \sim \textit{(indep) Gaussian}(0, \mathcal{C}_g; I_A),$$

$$q^* \sim Gaussian(0, \mathcal{C}_q),$$

$$\sigma_1^2 \sim IG(shape = a, scale = b).$$

Bayesian multiple function registration is also studied in Cheng et al. (2016). The main difference is that Cheng et al. (2016) put a Dirichlet process prior directly on Γ instead of using a Gaussian process prior on the tangent space $T_1(\Psi)$.

As in the pairwise case, sampling from the posterior distribution of the parameters $\{g_1, \ldots, g_c; q^*; \sigma_1^2\}$ is achieved using a MwG algorithm. At each update, we sample from the full conditional distribution of each parameter. Specifically, we use a MH algorithm with the Z-mixture pCN proposal to draw iteratively from the full conditional distributions of $g_1, \ldots, g_c$ as well as q^*. The acceptance ratios simplify to the likelihood ratios in the same way as in the pairwise case. The full conditional distribution of σ_1^2 is given by

$$\sigma_1^2 \mid \{q_i\}, \{g_i\}, q^* \sim IG\left(shape = \frac{1}{2}cN + a, scale = \frac{1}{2}\sum_{i=1}^{c} \texttt{SSE}(g_i, q^*) + b\right), \tag{15}$$

where

$$\texttt{SSE}(g_i, q^*) = \sum_{j=1}^{N} \left(q_i\left(\int_0^{t_j} \exp_1^2(g_i)(s)ds\right)\exp_1(g_i)(t_j) - q^*(t_j)\right)^2. \tag{16}$$

As noted in Cheng et al. (2016), since function amplitudes are invariant to a common warping, the template function q^* is only identifiable up to an equivalence class of warpings. To avoid this issue, we "center" $\gamma_1, \ldots, \gamma_c$ in each iteration such that their Karcher mean is the identity warping. This standardization is carried out by applying the inverse of the Karcher mean function, $\bar{\gamma}^{-1}$, to each of the γ_i's and to the template q^*. This procedure is justifiable because it does not change the likelihood (shown in Appendix A). A detailed description of the multiple function registration algorithm is provided in the Supplementary Material.

5. SIMULATIONS AND REAL DATA APPLICATIONS

5.1 Pairwise Bayesian Registration Simulation Studies

At the implementation stage, we need to pre-specify the following two quantities: (1) number of basis functions M_g used to estimate g (in pairwise and multiple function registration), and (2) the number of basis functions M_{q^*} used to estimate the template q^* (in multiple

function registration). To test the performance of our algorithm with the main interest in its robustness to the choice of M_g, we carry out simulations for the pairwise registration case. We randomly generate true warping functions (denoted by γ_{true}) and see how well we can estimate them using the proposed method with different pre-specified values of M_g. We set one of the observed functions $f_2(t) = \sin(4\pi t^2)$ (Ramsay and Li 1998) and let $f_1 = f_2 \circ \gamma_{true}$. We then register f_2 to f_1 using the proposed method with six different M_g values: $\{5, 10, 20, 30, 40, 50\}$. Other constants in the prior distribution are specified as $\sigma_g = 4$, $a = 0.1$ and $b = 0.1$. The chain is run for 10^6 iterations with a burn-in period of 0.5×10^6.

We simulate registration problems using the following three methods: (1) randomly generate $g \in T_1(\Psi)$ using the first pair of cos and sin functions in the Fourier basis and transform it back to γ_{true} ("*fourier1*"), (2) randomly generate $g \in T_1(\Psi)$ using the first 20 pairs of cos and sin functions in the Fourier basis and transform it back to γ_{true} ("*fourier20*"), and (3) generate a random γ_{true} directly by smoothing a randomly generated step function without the use of any basis functions ("*random*"). For each method and value of M_g, we generate ten random γ_{true} functions and estimate them using the proposed method via the posterior Karcher mean denoted by γ_{est}. We then calculate the FR distance (Equation 3) between each pair of γ_{true} and γ_{est}. The results are plotted in Figure 2. Each panel in Figure 2 corresponds to a method for randomly generating γ_{true} and shows the results of 60 independent registration problems (six values of M_g with ten replicates for each value). For comparison, we also solve each registration problem using the Dynamic Programming (DP) algorithm (using the R function `optimum.reparam` in package *fdasrvf*, Tucker 2016). This optimization-based method is described in detail in Srivastava et al. (2011b). We plot the average of the FR distances between γ_{true} and γ_{DP} as a horizontal blue line in Figure 2.

The results show that our method performs better than DP. For our method, the largest FR distance between γ_{true} and γ_{est} is less than 0.4. When γ_{true} is generated without using basis functions, a choice of $M_g \geq 20$ yields sufficiently good estimates (with the largest FR distance around 0.1). When γ_{true} is generated using the Fourier basis, a conservative setting of M_g to a larger number of basis elements than the true number used to generate γ_{true} yields FR distances below 0.1. When M_g is underestimated (e.g., we use the first 20 pairs of the Fourier basis to generate γ_{true} but set M_g equal to 5 or 10), the FR distances are larger, as expected. Two examples in the simulation set `fourier20` are shown in detail in

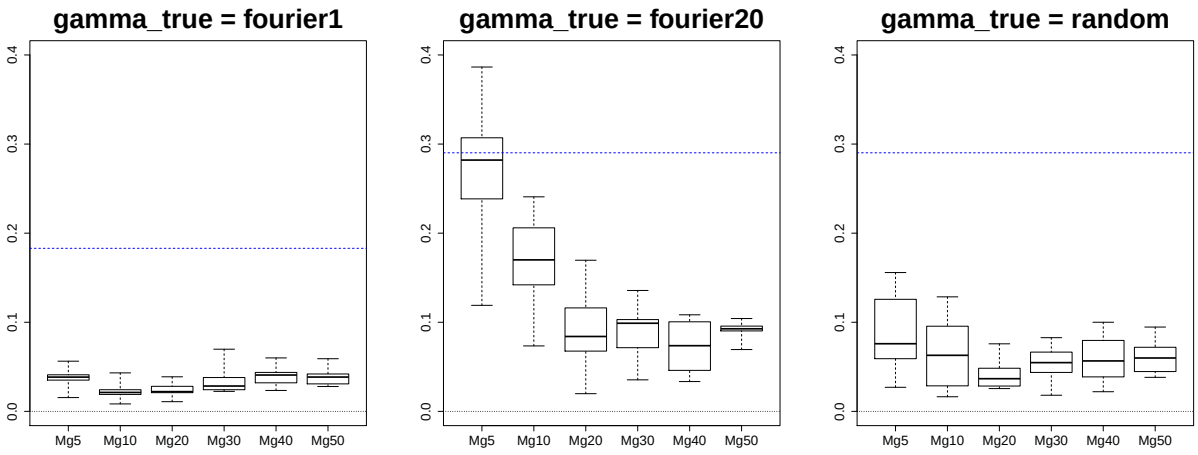

Figure 2: *FR distances between a randomly generated true warping function and its posterior mean estimate for different values of* $M_g \in \{5, 10, 20, 30, 40, 50\}$. *The three panels correspond to the cases when the true warping functions are generated using the first pair of the Fourier basis functions (left), the first* 20 *pairs of the Fourier basis functions (middle), and without the use of basis functions (right). Each boxplot is constructed using FR distances corresponding to ten independent registration problems. The blue dotted lines show the average FR distance when the warping functions are estimated using Dynamic Programming.*

Figure 3. Since γ_{true} is generated using the first 20 pairs of the Fourier basis, it exhibits many local features. The estimate obtained by setting $M_g = 5$ can only capture the general shape of γ_{true} (top panel). As a result, γ_{est} appears smoother than γ_{true}. This issue does not exist when we set $M_g = 50$ as shown in the example in the bottom panel. In addition, we have performed preliminary model assessment based on the Deviance Information Criterion (DIC), which shows that a choice of M_g between 20 and 30 is reasonable.

5.2 Berkeley Growth Data Registration

In the Berkeley growth study (Tuddenham and Snyder 1954), the heights of 54 girls were recorded from ages 1 to 18, each at 31 different time points. We are interested in comparing the girls' growth rates, which are given by the first derivatives of the growth curves. The main motivation for registering growth rate curves is that each individual reaches different phases of height growth at different times. We can better compare the girls' growth rates at a certain phase, such as the obvious pubertal spurt, after we register all of the curves.

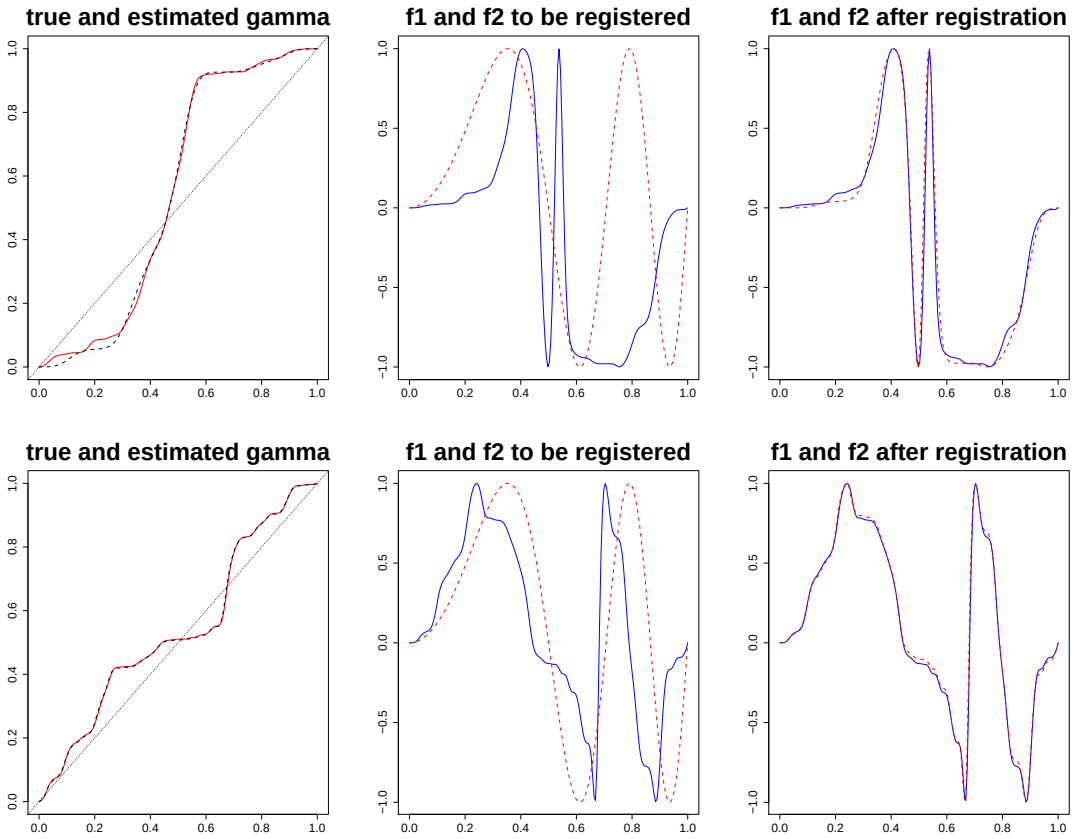

Figure 3: *Two different pairwise registration simulation examples when* γ_{true} *is generated using the first* 20 *pairs of the Fourier basis. Top: The left panel shows* γ_{true} *(red solid curve) and* γ_{est} *(black dashed curve) obtained using* $M_g = 5$*. The middle panel shows* f_1 *(red) and* f_2 *(blue). The right panel shows* f_1 *(red) and* $f_2 \circ \gamma_{est}$ *(blue). Bottom: Same as top but with* $M_g = 50$*.*

We use the Z-mixture pCN algorithm to register the 54 girls' growth rate curves simultaneously. The first 20 pairs of Fourier basis functions are used to estimate the functions g and the template function q^* (i.e., $M_g = 20$ and $M_{q^*} = 20$). Other constants for the prior specification are $\sigma_g = 4$, $\sigma_{q^*} = 0.8q_{95}$ (q_{95} is defined as the range of the middle 95% values of all q_i), $\alpha_{\sigma_1} = 0.1$ and $\beta_{\sigma_1} = 0.1$. The MCMC chain is run for 10^5 iterations with a burn-in period of 0.2×10^5. The results are shown in Figure 4. The top panel of this figure shows the original 54 observed functions $f_1, \ldots, f_{54}$, their registered versions (with the template and 95% pointwise credible interval in red), the original SRVFs $q_1, \ldots, q_{54}$, and the SRVFs of the registered functions (with the template and 95% pointwise credible interval in red),

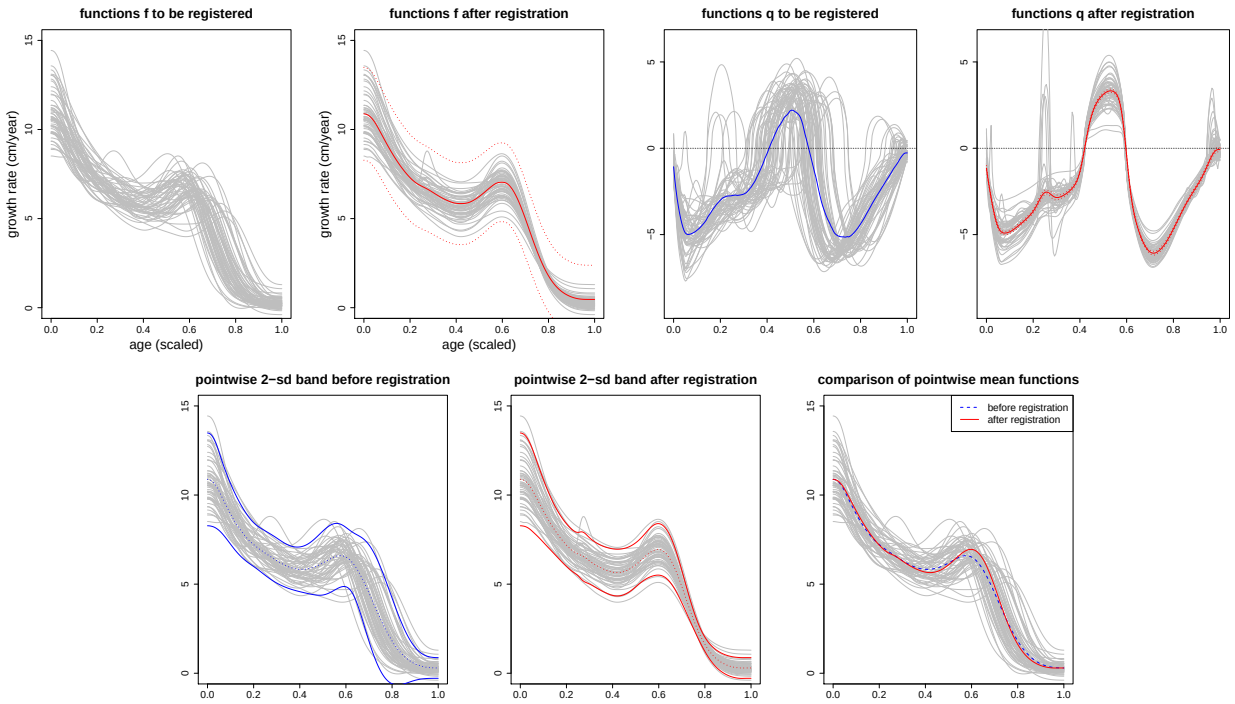

Figure 4: *Registration results for the growth rate curves of* 54 *girls in the Berkeley growth study. Top row left to right: (1) The original* 54 *observed growth rate functions, (2) the registered growth rate functions with the template (red solid line) and* 95% *credible band overlayed (red dotted lines), (3) the original SRVFs with the initial value of* q^* *(blue), and (4) the SRVFs of the registered growth rate functions with the final template (red solid line) and* 95% *credible band overlayed (red dotted lines). Bottom row left to right: (1) Cross-sectional mean (dotted) and pointwise* ± 2*-standard deviation band for the original* 54 *observed growth rate functions, (2) same as (1) but for the registered functions, and (3) comparison of the pointwise cross-sectional means.*

respectively. To plot an estimate of the template after registration f^*, we set $f^*(0)$ to be the average of $f_1(0), \ldots, f_{54}(0)$ (recall that we can only recover f^* from q^* up to a translation). For the credible interval of f^*, we transfer back the upper (lower) bound of q^* and add (subtract) 2 times the standard deviation of $\{f_i(0)\}$ to form the upper (lower) bound of f^*. The bottom panel shows the cross-sectional mean with a two standard deviation band prior to registration (left) and after registration (middle). The two mean functions are directly compared in the bottom right. After registration, we see that the pointwise mean function

exhibits a higher peak and a lower valley, corresponding to the growth spurt phase during puberty and a slowing-down phase before puberty. The pointwise standard deviation decreases significantly in some regions of the domain. For instance, during the post-puberty phase, growth rates exhibit much less amplitude variation across individuals, a feature not shown prior to registration. Thus, the mean function and the ± 2-standard deviation band are better representatives of the overall shape of the individual growth rate functions after registration.

5.3 Gait Cycle Data

As a commonly studied area of mechanical analysis of human movement, gait cycle analysis considers functional data that are collected as participants walk over a period of time, possibly under different conditions. Variables of interest include velocity and angles (kinematics) as well as muscle forces and moments (kinetics) (van den Bogert et al. 2013). Some variables, such as joint angles and foot floor reaction forces, are directly measurable whereas other variables, such as moments of force, need to be computed from some model (Kadaba et al. 1989). Measurements are usually taken by placing makers on test subjects (detailed description of data acquisition can be found in most of the papers cited here), and data collected vary in magnitude as well as in timings of within gait cycle events (Helwig et al. 2011). Kadaba et al. (1989) notes that data can differ in timing simply due to misalignment of markers when measuring the same subject on a different day.

Many tasks encountered in gait analysis involve comparing and averaging functional data. We explain three major tasks using the outline given in Duhamel et al. (2004). First, for one individual, we want to know if all of the gait curves are consistent and, if so, calculate one average curve as a summary statistic for that individual. Second, for two different populations, we want to compare their gait properties by comparing, for instance, the population mean curves. The two populations under investigation can be two independent populations (e.g., "pathological" versus "normal") or one population under different conditions (e.g., braced versus non-braced conditions in Shorter et al. 2008). Third, combining the above two tasks, we also want to compare the mean gait curve of one subject to that of a population so that we can classify a new subject to a given population.

Many existing methods for statistical analysis of gait cycle data compare gait curves by

superimposing those curves for a simple visual assessment, and by calculating a numeric value as a measurement of how consistent (or "reliable") those curves are. Examples of such numeric summaries include coefficient of variation (Winter 1984), adjusted coefficient of multiple determination (Kadaba et al. 1989), and intra-class correlation coefficient (Duhamel et al. 2004). When summarizing multiple curves of one person or of one population, the mean curve and a confidence band are usually calculated in a pointwise fashion (Duhamel et al. 2004). Surveys of traditionally used statistical methods can be found in Chau (2001) and Duhamel et al. (2004). A discussion of how to better use those methods in clinical problems can be found in Simon (2004). As a result of the need to compare and average functions, function registration arises naturally as an initial step in the analysis. It allows one to compare phase variation and amplitude variation separately. It is also necessary if we want to carry out statistical analysis in a point-by-point fashion as is commonly done in this community. As one of the earliest papers to use curve registration on gait data, Sadeghi et al. (2000) shows that curve registration reduces inter-subject variability. We hope to contribute to the current literature by applying the proposed Bayesian registration technique to gait cycle data.

Moore et al. (2015) provides a list of available gait cycle datasets. Most of them, however, only give individual mean curves or group mean curves, possibly due to the large size of the unaveraged raw data. Data used here are part of the online supplementary material of van den Bogert et al. (2013). Gait cycle data of 12 healthy individuals (11 males and 1 female, 28.3 ± 3.9 years of age) are reported, which include 44 kinematic variables and 300 kinetic variables. We use one of the variables, *Right Knee Flexion (v39)*, as an illustrative example. This variable is investigated by many gait cycle studies including Duhamel et al. (2004) and Helwig et al. (2011). Note that the time axis of each gait cycle is linearly scaled such that they all have the same length. Data are then expressed as a function of percentage of one gait cycle. This standardization (called Linear Length Normalization in Helwig et al. 2011) is used frequently as a preprocessing step in gait analysis since it makes the data easier to manipulate and to interpret. Each standardized gait curve is a real-valued function defined on $[0, 1]$, which is well-suited for the function registration set-up in this paper.

We try different values for prior specification; specifically, we use $M_g \in \{5, 20\}$, $\sigma_g \in \{0.5, 4\}$, $M_{q^*} \in \{5, 20\}$, $\sigma_{q^*} = q_{95}$, $\alpha_{\sigma_1} = 0.1$, and $\beta_{\sigma_1} = 0.1$. The MCMC chain is run for

σ_g, M_g, M_q^*	$0.5, 5, 5$	$4, 5, 5$	$0.5, 5, 20$	$4, 5, 20$	$0.5, 20, 5$	$4, 20, 5$	$0.5, 20, 20$	$4, 20, 20$
Sync	0.6072	0.5996	0.5987	0.5896	0.5963	0.5923	0.5863	**0.5853**
IPC	0.9828	0.9826	0.9825	0.9823	0.9825	0.9823	**0.9822**	**0.9822**

Table 1: *Sync and IPC criteria measuring the quality of alignment of the* 12 *Right Knee Flexion functions under different prior distribution choices.*

5×10^5 updates with a burn-in period of 10^5. To compare estimation results for different prior configurations, we use the following two criteria (Cheng et al. 2016).

- The synchronization (Sync) coefficient is given by:

$$\text{Sync} = \frac{1}{c} \sum_{i=1}^{c} \frac{\|\tilde{f}_i - \frac{1}{(c-1)} \sum_{j \neq i} \tilde{f}_j\|^2}{\|f_i - \frac{1}{(c-1)} \sum_{j \neq i} f_j\|^2}; \tag{17}$$

- The inverse of pairwise correlation (IPC) is given by:

$$\text{IPC} = \frac{\sum_{i \neq j} r(f_i, f_j)}{\sum_{i \neq j} r(\tilde{f}_i, \tilde{f}_j)}, \tag{18}$$

where $r(\cdot, \cdot)$ is the pairwise Pearson's correlation between functions, c is the number of functions to be registered (here, $c = 12$), f denotes the original unregistered functions, and $\tilde{f}$ denotes the registered functions. For both Sync and IPC, a smaller value indicates better alignment. A summary of the results is given in Table 1. This comparison shows no significant difference between different prior configurations, although one might note that larger M_g and M_{q^*} values yield slightly better alignment results as expected. Registration results for one of the configurations is plotted in Figure 5. For this dataset, the cross-sectional mean curves do not show much difference after registration (although it is clear that the functions are better matched overall after registration), but the ± 2-standard deviation band becomes narrower in some regions of the domain, and more accurately reflects the overall shape of the observed gait cycle functions.

6. DISCUSSION

In this paper, we presented a Bayesian model to perform pairwise and multiple function registration. We transform the observation space to the SRVF space such that the $\mathbb{L}^2$–norm satisfies desirable properties for the registration problem. We transform the parameter space

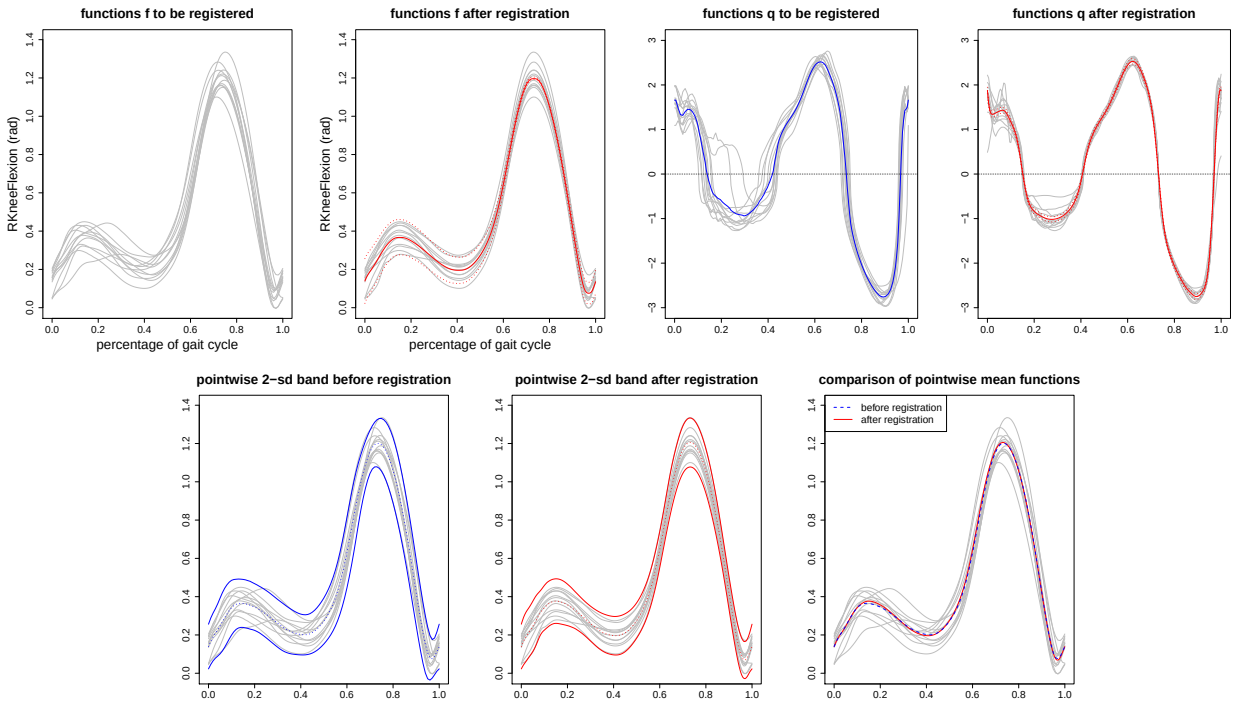

Figure 5: *Registration results for Right Knee Flexion functions of* 12 *individuals. Top row left to right: (1) The original* 12 *observed Right Knee Flexion functions, (2) the registered functions with the template (red solid line) and* 95% *credible band overlayed (red dotted lines), (3) the original SRVFs with the initial value of* q^* *(blue), and (4) the SRVFs of the registered Right Knee Flexion functions with the final template (red solid line) and* 95% *credible band overlayed (red dotted lines). Bottom row left to right: (1) Cross-sectional mean (dotted) and pointwise* ± 2*-standard deviation band for the original* 12 *observed Right Knee Flexion functions, (2) same as (1) but for the registered functions, and (3) comparison of the pointwise cross-sectional means.*

to a linear space, $T_1(\Psi)$, which allows us to assign a Gaussian process prior on the warping functions, as opposed to the Dirichlet process prior in Cheng et al. (2016). Our model also has the added benefit of being robust to different discretization choices for the warping function. We perform inference by drawing from the posterior distribution using a Metropolis within Gibbs algorithm. We define a novel Z-mixture pCN proposal, which is a generalization of the pCN proposal in Cotter et al. (2013), to make the Metropolis step more efficient. A simulation study for pairwise registration shows that our method accurately recovers the

true warping function. Applying our method to growth rate curves and gait cycle curves shows that multiple function registration is a useful data analysis step, especially when we want to analyze the phase and amplitude variations separately.

In our simulations, we find that a grid of 200 time points is sufficiently fine so that approximation errors caused by evaluating the transformations in Section 2 (first order derivatives and integrals) are negligible. Also, for the centering step, we can calculate $\gamma^{shift} = \gamma \circ \bar{\gamma}^{-1}$ or equivalently calculate $\psi^{shift} \equiv \phi(\gamma^{shift}) = \psi \circ \bar{\gamma}^{-1}\sqrt{(\bar{\gamma}^{-1})'}$. When both are calculated on a finite grid, we note that calculating ψ^{shift} leads to a smaller approximation error. More details on the grid size and approximation error can be found in the Supplementary Material. Note that, as discussed earlier, the warping functions are represented via known basis functions, so they are actually known at any given point on the domain. A grid is simply needed for storing the warping functions on the computer.

There are a few directions for future work that are of interest. The algorithms presented in this paper require pre-specifications of *fixed* M_g and M_{q^*}, the number of basis functions used to approximate g and the template q^*. Alternatively, we can select prior distributions for M_g and M_{q^*}. In this random truncation case, the prior distributions of g and q^* are no longer Gaussian (instead, they are called *sieve priors* in Cotter et al. 2013). When representing a non-periodic template q^* on $[0, 1]$ using the Fourier basis (as in the Berkeley growth study example), setting the period to equal two works well in practice. Alternatively, one can use other non-periodic bases such as splines or wavelets. Since the proposed algorithm involves calculating the sum of Fourier basis functions on a grid, further efficiency can be gained by incorporating the Fast Fourier Transform (FFT) in this step.

APPENDIX A: DERIVATIONS

We show that applying $\tilde{\gamma} \in \Gamma$ to $(\gamma_1, \ldots, \gamma_c; q^*)$ does not change the likelihood. In the centering step for multiple function registration in Section 4, this fact is used with $\tilde{\gamma} = \bar{\gamma}^{-1}$. We first show that `SSE` remains the same after applying $\tilde{\gamma}$. Recall that $\texttt{SSE}(\gamma, q^*) \equiv \|(q, \gamma) - q^*\|^2$, where q is the observed function that we want to register using γ, and $\|\cdot\|$ refers to the $\mathbb{L}^2$–norm. We want to show that

$$\|(q, \gamma) - q^*\|^2 = \|(q, \gamma \circ \tilde{\gamma}) - (q^*, \tilde{\gamma})\|^2. \tag{19}$$

One can easily verify that $(q, \gamma \circ \tilde{\gamma}) = ((q, \gamma), \tilde{\gamma})$. Now, Equation 19 is true since it is equivalent to

$$\|(q, \gamma) - q^*\|^2 = \|((q, \gamma), \tilde{\gamma}) - (q^*, \tilde{\gamma})\|^2, \tag{20}$$

which is true due to the action of Γ on $\mathbb{L}^2([0, 1])$ being an isometry. It then follows directly that the likelihood

$$L(g_1, \ldots, g_c, q^*, \sigma_1^2) \propto \exp\left\{\sum_{i=1}^{c}\left(-N\log(\sigma_1) - \frac{1}{2\sigma_1^2}\texttt{SSE}(g_i, q^*)\right)\right\} \tag{21}$$

remains the same after applying any $\tilde{\gamma} \in \Gamma$ to $(\gamma_1, \ldots, \gamma_c; q^*)$, where c is the number of functions to be registered and N is the dimension of the grid on which the functions were observed.

APPENDIX B: SUPPLEMENTARY MATERIALS

Supplementary Material: The Supplementary Material includes: 1) a detailed descriptions of the MCMC algorithms used for pairwise and multiple function registration, 2) a discussion on the implementation details (estimation error and the choice of grid size), 3) additional plots, 4) discussion on choices of hyperparameters in the Gaussian process prior for warping functions. (supplementary.pdf, pdf file)

Code: R code and examples to reproduce results in the paper are available in the zip file. Details can be found in the readme.txt file included. (code.zip, zip file)

REFERENCES

Bhattachayya, A. (1943), "On a Measure of Divergence Between Two Statistical Population Defined by their Population Distributions," *Bulletin Calcutta Mathematical Society*, 35, 99–109.

Chau, T. (2001), "A Review of Analytical Techniques for Gait Data. Part 1: Fuzzy, Statistical and Fractal Methods," *Gait & Posture*, 13(1), 49–66.

Cheng, W., Dryden, I. L., and Huang, X. (2016), "Bayesian Registration of Functions and Curves," *Bayesian Anal.*, 11(2), 447–475.

Claeskens, G., Silverman, B. W., and Slaets, L. (2010), "A Multiresolution Approach to Time Warping Achieved by a Bayesian Prior–posterior Transfer Fitting Strategy," *Journal of the Royal Statistical Society*, 72(5), 673–694.

Clifford, D., Stone, G., Montoliu, I., Rezzi, S., Martin, F.-P., Guy, P., Bruce, S., and Kochhar, S. (2009), "Alignment Using Variable Penalty Dynamic Time Warping," *Analytical Chemistry*, 81(3), 1000–1007.

Cotter, S. L., Roberts, G. O., Stuart, A. M., White, D. et al. (2013), "MCMC Methods for Functions: Modifying Old Algorithms to Make them Faster," *Statistical Science*, 28(3), 424–446.

Duhamel, A., Bourriez, J. L., Devos, P., Krystkowiak, P., Destee, A., Derambure, P., and Defebvre, L. (2004), "Statistical Tools for Clinical Gait Analysis," *Gait & posture*, 20(2), 204–212.

Gervini, D., and Gasser, T. (2004), "Self-modeling Warping Functions," *Journal of the Royal Statistical Society*, 66, 959–971.

Helwig, N. E., Hong, S., Hsiao-Wecksler, E. T., and Polk, J. D. (2011), "Methods to Temporally Align Gait Cycle Data," *Journal of Biomechanics*, 44(3), 561–566.

James, G. (2007), "Curve Alignment by Moments," *Annals of Applied Statistics*, 1(2), 480–501.

Kadaba, M. P., Ramakrishnan, H. K., Wootten, M. E., Gainey, J., Gorton, G., and Cochran, G. V. B. (1989), "Repeatability of Kinematic, Kinetic, and Electromyographic Data in Normal Adult Gait," *Journal of Orthopaedic Research*, 7(6), 849–860.

Karcher, H. (1977), "Riemannian Center of Mass and Mollifier Smoothing," *Communications on Pure and Applied Mathematics*, 30(5), 509–541.

Keogh, E. J., and Pazzani, M. J. (2001), "Derivative Dynamic Time Warping," *SIAM Proceedings of the International Conference on Data Mining*, pp. 1–11.

Kneip, A., Li, X., MacGibbon, K. B., and Ramsay, J. O. (2000), "Curve Registration by Local Regression," *The Canadian Journal of Statistics*, pp. 19–29.

Kneip, A., and Ramsay, J. O. (2008), "Combining Registration and Fitting for Functional Models," *Journal of the American Statistical Association*, 103(483), 1155–1165.

Koch, I., Hoffmann, P., and Marron, J. S. (2014), "Proteomics Profiles from Mass Spectrometry," *Electronic Journal of Statistics*, 8(2), 1703–1713.

Kurtek, S. (2015), "A Geometric Approach to Pairwise Bayesian Alignment of Functional Data Using Importance Sampling," *arXiv preprint*, arXiv:1505.06954v2.

Kurtek, S., and Bharath, K. (2015), "Bayesian Sensitivity Analysis with the Fisher–Rao Metric," *Biometrika*, 102(3), 601–616.

Kurtek, S., Srivastava, A., Klassen, E., and Ding, Z. (2012), "Statistical Modeling of Curves Using Shapes and Related Features," *Journal of the American Statistical Association*, 107(499), 1152–1165.

Kurtek, S., Srivastava, A., and Wu, W. (2011), "Signal Estimation under Random Time-warpings and Nonlinear Signal Alignment," *Neural Information Processing Systems (NIPS)*, pp. 675–683.

Liu, X., and Müller, H. G. (2004), "Functional Convex Averaging and Synchronization for Time-warped Random Curves," *Journal of the American Statistical Association*, 99, 687–699.

Marron, J. S., Ramsay, J. O., Sangalli, L. M., and Srivastava, A. (2015), "Functional Data Analysis of Amplitude and Phase Variation," *Statistical Science*, 30(4), 468–484.

Moore, J. K., Hnat, S. K., and van den Bogert, A. J. (2015), "An Elaborate Dataset on Human Gait and the Effect of Mechanical Perturbations," *PeerJ*, 3, e918.

Raket, L. L., Sommer, S., and Markussen, B. (2014), "A Nonlinear Mixed-effects Model for Simultaneous Smoothing and Registration of Functional Data," *Pattern Recognition Letters*, 38, 1–7.

Ramsay, J. O. (2000), "Functional Components of Variation in Handwriting," *Journal of the American Statistical Association*, 95(449), 9–15.

Ramsay, J. O., Gribble, P., and Kurtek, S. (2014), "Description and Processing of Functional Data Arising from Juggling Trajectories," *Electronic Journal of Statistics*, 8(2), 1811–1816.

Ramsay, J. O., and Li, X. (1998), "Curve Registration," *Journal of the Royal Statistical Society. Series B*, 60, 351–363.

Ramsay, J. O., and Silverman, B. W. (2002), *Applied Functional Data Analysis: Methods and Case Studies*, 1 edn, New York: Springer-Verlag.

Ramsay, J. O., and Silverman, B. W. (2005), *Functional Data Analysis*, 2 edn, New York: Springer-Verlag.

Ramsay, J. O., Wickham, H., Graves, S., and Hooker, G. (2014), *Functional Data Analysis*, R package version 2.4.4.
URL: *https://cran.r-project.org/web/packages/fda/fda.pdf*

Sadeghi, H., Allard, P., Shafie, K., Mathieu, P. A., Sadeghi, S., Prince, F., and Ramsay, J. O. (2000), "Reduction of Gait Data Variability Using Curve Registration," *Gait & posture*, 12(3), 257–264.

Sangalli, L. M., Secchi, P., and Vantini, S. (2014), "AneuRisk65: A Dataset of Three-dimensional Cerebral Vascular Geometries," *Electronic Journal of Statistics*, 8(2), 1879–1890.

Senin, P. (2008), *Dynamic Time Warping Algorithm Review*, Information and Computer Science Department, University of Hawaii at Manoa Honolulu, USA.

Shorter, K. A., Polk, J. D., Rosengren, K. S., and Hsiao-Wecksler, E. T. (2008), "A New Approach to Detecting Asymmetries in Gait," *Clinical Biomechanics*, 23(4), 459–467.

Silverman, B. W. (1995), "Incorporating Parametric Effects into Functional Principal Components Analysis," *Journal of the Royal Statistical Society*, pp. 673–689.

Simon, S. R. (2004), "Quantification of Human Motion: Gait Analysis-Benefits and Limitations to its Application to Clinical Problems," *Journal of Biomechanics*, 37(12), 1869–1880.

Srivastava, A., Jermyn, I., and Joshi, S. (2007), "Riemannian Analysis of Probability Density Functions with Applications in Vision," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8.

Srivastava, A., Klassen, E., Joshi, S. H., and Jermyn, I. H. (2011a), "Shape Analysis of Elastic Curves in Euclidean Spaces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(7), 1415–1428.

Srivastava, A., Wu, W., Kurtek, S., Klassen, E., and Marron, J. S. (2011b), "Registration of Functional Data Using Fisher-Rao Metric," *arXiv preprint*, arXiv:1103.3817.

Stuart, A. M. (2010), "Inverse Problems: a Bayesian Perspective," *Acta Numerica*, 19, 451–559.

Tang, R., and Müller, H. G. (2008), "Pairwise Curve Synchronization for Functional Data," *Biometrika*, 95(4), 875–889.

Telesca, D., and Inoue, L. Y. T. (2008), "Bayesian Hierarchical Curve Registration," *Journal of the American Statistical Association*, 103(481), 328–339.

Tucker, J. D. (2016), *Elastic Functional Data Analysis*, R package version 1.6.0.
URL: *https://cran.r-project.org/web/packages/fdasrvf/fdasrvf.pdf*

Tuddenham, R. D., and Snyder, M. M. (1954), "Physical Growth of California Boys and Girls from Birth to Eighteen Years," *Publications in Child Development at University of California, Berkeley*, 1(2), 183.

van den Bogert, A. J., Geijtenbeek, T., Even-Zohar, O., Steenbrink, F., and Hardin, E. C. (2013), "A Real-time System for Biomechanical Analysis of Human Movement and Muscle Function," *Medical & biological engineering & computing*, 51(10), 1069–1077.

Wang, K., Gasser, T. et al. (1997), "Alignment of Curves by Dynamic Time Warping," *The Annals of Statistics*, 25(3), 1251–1276.

Winter, D. A. (1984), "Kinematic and Kinetic Patterns in Human Gait: Variability and Compensating Effects," *Human Movement Science*, 3(1), 51–76.

Wu, W., Hatsopoulos, N. G., and Srivastava, A. (2014), "Introduction to Neural Spike Train Data for Phase-Amplitude Analysis," *Electronic Journal of Statistics*, 8(2), 1759–1768.