Bayesian Analysis (2016) **11**, Number 2, pp. 447–475

Bayesian Registration of Functions and Curves

Wen Cheng*, Ian L. Dryden† and Xianzheng Huang‡

Abstract. Bayesian analysis of functions and curves is considered, where warping and other geometrical transformations are often required for meaningful comparisons. The functions and curves of interest are represented using the recently introduced square root velocity function, which enables a warping invariant elastic distance to be calculated in a straightforward manner. We distinguish between various spaces of interest: the original space, the ambient space after standardizing, and the quotient space after removing a group of transformations. Using Gaussian process models in the ambient space and Dirichlet priors for the warping functions, we explore Bayesian inference for curves and functions. Markov chain Monte Carlo algorithms are introduced for simulating from the posterior. We also compare ambient and quotient space estimators for mean shape, and explain their frequent similarity in many practical problems using a Laplace approximation. Simulation studies are carried out, as well as practical alignment of growth rate functions and shape classification of mouse vertebra outlines in evolutionary biology. We also compare the performance of our Bayesian method with some alternative approaches.

Keywords: ambient space, Dirichlet, Gaussian process, Quotient space, shape, warp.

1 Introduction

We consider statistical analysis of functions and curves where some form of registration or time warping is of interest. The two main applications that we focus on involve the alignment of growth rate curves and the classification of mouse vertebrae shape outlines in evolutionary biology. Both applications require methods which can take account of arbitrary reparameterizations of the functions or curves of interest. In order to help choose appropriate methods and models, we first describe three different spaces of interest: the original space, the ambient space and the quotient space. The choice of space in which to specify the statistical model is important, as it determines what type of mean estimation and subsequent statistical analyses are carried out.

Our main contribution is to introduce a Bayesian approach to the analysis of functions and curves, which is demonstrated to be effective in the two applications as well as in the comparison with some other existing methodologies. Inference is carried out using Markov chain Monte Carlo simulation, and prior beliefs about the amount of time warping or registration are included as part of the model.

We wish to consider applications where the functions or curves of interest may not be in alignment. For example, in the study of growth curves of children it makes sense

*The University of South Carolina, chengwen1985@gmail.com
†The University of Nottingham, ian.dryden@nottingham.ac.uk
‡The University of South Carolina, huang@stat.sc.edu

 DOI: 10.1214/15-BA957

to consider a time warping of the curves so that the curves match up in a biologically meaningful way. Children reach various stages of development such as puberty at different times, and so when comparing growth curves it is sensible to first align the curves in time and then compare the different heights and growth rates of the children using the time-warped curves (Ramsay and Li, 1998). The function registration problem has been considered by a large number of authors, including Kneip and Gasser (1992); Silverman (1995); Ramsay and Li (1998); Kneip et al. (2000) and Srivastava et al. (2011b), among many others. In earlier Bayesian approaches, Telesca and Inoue (2008) and Zhou et al. (2014) model the functions and warps by linear combinations of B-splines, and Claeskens et al. (2010) use another Bayesian method with multiresolution warping functions.

After alignment, quantities such as a population mean function and a population covariance function can then be estimated in the space of curves. In addition to the amplitude variability of the functions post registration, it is also of interest to analyze the variability in the registration transformations themselves, which is also known as phase variability. When analyzing curves in two or three dimensions we have additional potential invariances, such as translation, rotation and possibly scale invariance.

As a motivating example consider the functions in Figure 1, which are the growth rates (smoothed first derivative of height with respect to time) of two girls from the Berkeley growth study (Ramsay and Silverman, 2005). In order to compare the curves, it makes sense to align the main features, which here are two growth spurts given by the peaks. In the left hand plot of Figure 1, it can be seen that the scans are not well aligned, as the large peaks are not in the same positions in the x-axis. The goal of the alignment is to register the curves with a transformation of the x-axis so that peaks representing the growth spurts can be compared between individuals. After registration using the Bayesian methodology of this paper, it is clear in the middle plot of Figure 1 that the two main peaks have been lined up using the posterior mean warping function applied to the x-axis, and in the right hand plot the posterior mean warp and 95% pointwise posterior credibility intervals are displayed. Clearly, there is little uncertainty in the alignment in this case. In some applications much of the alignment can be accounted for locally by a translation of the x-axis, and so we develop a Bayesian method for alignment that places strong prior information on translations, if desired. A strong prior parameter $a = 50$ is used in Figure 1, which will be explained later in Section 4.

In this paper we first introduce the original, quotient and ambient spaces for representing functions and curves in Section 2. We focus on the square root velocity function and quotient space in Section 3. A Bayesian model in ambient space is described in Section 4, and inference is developed using Markov chain Monte Carlo simulation. Some properties of the methods are given in Section 5, including asymptotic properties and approximations. Various simulation studies and practical analysis of growth rate curves are given in Section 6, and we also consider a problem in shape analysis where it is of interest to classify mice vertebrae on the basis of the outline shape. We conclude with a brief discussion.

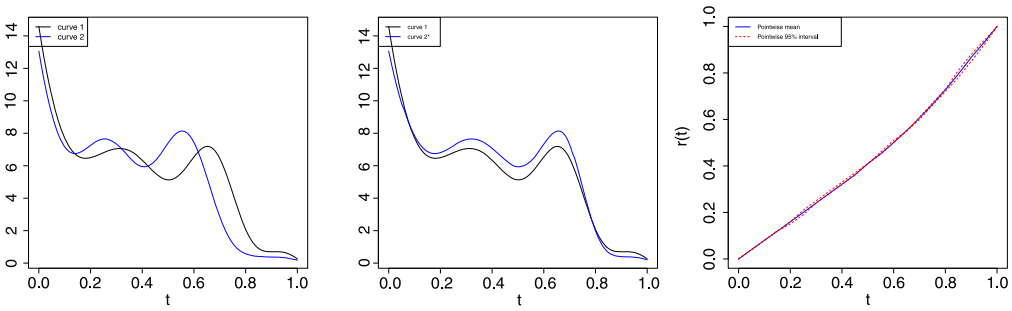

Figure 1: Growth rate curves from two girls in the Berkeley growth study (left), which have been aligned using the Bayesian procedure (center) using the estimated posterior mean warping function with 95% pointwise credibility intervals displayed (right). In each plot the x-axis represents time from 1 to 18 years, and has been rescaled to unit length.

2 The spaces of interest

2.1 Original, ambient and quotient spaces

Consider data of interest in the form of functions or curves

$$f_i(t) : [0,1] \to \mathbb{R}^m, \quad i = 1, \ldots, n.$$

In functional data analysis (Ramsay and Silverman, 2005), the function $f(t)$ is typically in $m = 1$ dimension. In statistical shape analysis (Klassen et al., 2003), the curve $f(t)$ is usually in $m = 2$ or $m = 3$ dimensions. In practice we cannot observe a complete continuous function but rather a finite set of discrete points $\{f(t_j) \in \mathbb{R}^m : j = 1, \ldots, k\}$, where the function is observed at times $t_j, j = 1, \ldots, k$.

In a general form of the registration problem, let us first consider the different spaces of interest. Each object f is located in the original space (e.g., a space of functions, a space of curves in $\mathbb{R}^m$, or a space of landmark coordinates). The original space is where we represent the raw objects under study.

It is very common to standardize the objects with a preliminary transformation, such as centering or rescaling so that the objects have unit norm, or perhaps taking a derivative with respect to time to be translation invariant. These initial transformations are simple in nature and carried out individually on each object, very much in the spirit of standardizing variables to have zero mean and unit variance in univariate statistics, or taking first differences in time series. The standardized object f^* is now represented in the ambient space S. Given that it is straightforward to transform to the ambient space, we will assume from now on that this initial standardization has been carried out.

Finally, we wish to investigate the equivalence class $[f]_Q \in Q$ which is obtained by removing transformations $\gamma \in G$ from the standardized f^*, where G is a group of

transformations and $Q = S/G$ is a quotient space. An important observation is that in order to compute distances in the quotient space, optimization over the transformation group G is required.

This notion of equivalence class and quotient space is precisely that introduced by Kendall (1984) for the representation of the shapes of k landmarks in $\mathbb{R}^m$, where $k > m$. The k landmarks are points located in m dimensions which represent the important features of the objects under study. In this situation the original space is the space of landmark coordinates $\mathbb{R}^{km} \setminus \{0\}$; the ambient space S is the pre-shape sphere $S^{(k-1)m-1}$ of landmark coordinates which are Helmertized (or centered) to remove location and scaled to have unit size; and the quotient space is Kendall's shape space Σ_m^k after quotienting out $SO(m) = \{R : R^T R = RR^T = I_m, \det(R) = 1\}$, where $SO(m)$ is the special orthogonal group of $m \times m$ rotation matrices. See Kendall et al. (1999) for a detailed description of the geometry of this space and Dryden and Mardia (1998) for statistical considerations.

In functional data analysis, the registration group is a transformation of the domain of the function, for example, a translation $\gamma(t) = t+c$, affine transformation $\gamma(t) = at+c$, or the full group of diffeomorphic transformations $\mathcal{D} = \{\gamma : [0,1] \to [0,1]\}$, such that γ is 1–1 and onto. The functions themselves lie in the original space, are then standardized to the ambient space, and then finally are decomposed such that the amplitude variability is represented in the quotient space and the phase variability is contained in the group of transformations G.

In curve analysis the registration of interest is the transformation of the domain, and in addition we may wish to register using the translation, rotation and scaling of the curve. In this case the curves lie in the original space, standardized versions lie in the ambient space, then the shapes of the curves are represented in the quotient space. The main spaces used in this paper are given in Table 1.

Original object	Ambient space	Distance	Quotient space distance
$X \in \mathbb{M}^{k\times m}$	$Z = \frac{HX}{\|HX\|} \in S^{(k-1)m}$	$\|Z_1 - Z_2\|$	$\inf_{\Gamma\in SO(m)}\|Z_1 - Z_2\Gamma\|$
$\{f(t) : t \in \mathbb{R}\}$	$q = \frac{\dot{f}}{\|\dot{f}\|^{1/2}} \in \mathbb{L}^2$	$\|q_1 - q_2\|_2$	$\inf_{\gamma\in\mathcal{D}}\|q_1 - q_2 \circ \gamma\|_2$
$\{f(t) : t \in \mathbb{R}^m\}$	$q = \frac{\dot{f}}{\|\dot{f}\|^{1/2}} \in \mathbb{L}^2$	$\|q_1 - q_2\|_2$	$\inf_{\gamma\in\mathcal{D},\Gamma\in SO(m)}\|q_1 - (q_2 \circ \gamma)\Gamma\|_2$

Table 1: Three examples of original objects, ambient spaces, ambient space distances and quotient space distances. Row 1: k landmarks in m dimensions, where H is a Helmert sub-matrix (Dryden and Mardia, 1998, p. 34) used for removing translation and Γ is an $m \times m$ rotation matrix; row 2: 1D functions, with warp $\gamma \in \mathcal{D}$ a re-parameterization of time; row 3: curves in mD with warp $\gamma \in \mathcal{D}$ a re-parameterization of arc-length and Γ is a rotation matrix in m-dimensions.

For our analysis of functions and curves, the original space and the ambient space S are standard classical spaces, such as $\mathbb{L}^2$, $\mathbb{L}^2 \times \cdots \times \mathbb{L}^2$ or S^{d-1}, where statistical models can be relatively easily formulated, and inference carried out. In terms of statistical modeling and inference, working with objects in the group G of transformations is more challenging, but can be undertaken. The geometry of the group is usually relatively

simple and well understood. However, the quotient space can be considerably more complicated in some situations. For example, the similarity shape space of a finite set of landmarks in three dimensions is very complicated, being a non-homogeneous space with singularities (Le and Kendall, 1993).

So, an important question is: In which space shall we define our statistical model, the original, ambient or quotient space? Since the transformation from the original to the ambient space is quite straightforward, the main issue is whether we should consider models in the ambient space or the quotient space. Ultimately the choice of model will depend on the goals of the study and what we are trying to make inference about.

Let us first consider two data objects X_1 and X_2, which could both be standardized functions, curves, landmarks or any other type of object in an ambient space S. How close are X_1 and X_2, ignoring arbitrary registrations $\gamma_1, \gamma_2 \in G$? Let $[X_1]_Q$ and $[X_2]_Q$ denote the amplitudes (or shapes) of X_1, X_2. A natural distance between the amplitude functions is in the quotient space:

$$d([X_1]_Q, [X_2]_Q) = \inf_{\gamma \in G} d(X_1, X_2 \circ \gamma),$$

where we must also have the isometric property

$$d([X_1 \circ \gamma^*]_Q, [X_2 \circ \gamma^*]_Q) = d([X_1]_Q, [X_2]_Q), \tag{1}$$

where an arbitrary common transformation γ^* can be applied to both objects and the quotient distance remains unchanged. This property is also known as a parallel orbit property, in that the orbits (transformations of an object by γ^*) are parallel, and it is also known as "right-invariance". This property is a necessity when thinking about practical statistical analyses which are invariant to transformations. If we apply an arbitrary transformation to our data then clearly all distances must remain invariant.

2.2 Statistical models and inference

Consider a distribution for a random object X, where it is the equivalence class up to transformations in $\gamma \in G$ that is of interest. We have several choices for specifying a distribution. We could model X in the ambient space with a population mean

$$\mu_A = \arg \inf_{\nu \in S} \int_S d(x, \nu)^2 h(x) dx, \tag{2}$$

where $h(x)$ is the probability density function (p.d.f.) of X. If $d(\cdot, \cdot)$ is the $\mathbb{L}^2$ or Euclidean norm then $\mu_A = E[X] = \int x h(x) dx$. The key location parameter of interest is then the amplitude (shape) of μ_A.

Statistical models in the ambient space are quite straightforward to specify because the ambient space is usually not complicated. For example, we specify a stochastic process/probability distribution for X, and then choose some coordinates in the quotient space, which we write as $U = [X]_Q$ together with registration parameters $\gamma \in G$. We can specify a probability distribution for X and transform from X to U (where

$U = X \circ \gamma^{-1} \in Q$ and $\gamma \in G$). Likelihood based inference about μ_A up to transformations γ is then carried out after marginalization, i.e., after integrating out the transformations γ from the distribution of X. This approach was used by Mardia and Dryden (1989); Dryden and Mardia (1991, 1992) in landmark shape analysis, for example.

Alternatively, we could model the equivalence class $U = [X]_Q$ directly in the quotient space with population Fréchet (1948) mean

$$\mu_Q = \arg \inf_{\mu \in Q} \int_Q d(u, \mu)^2 h(u) du, \tag{3}$$

where $h(u)$ is the p.d.f. of U, and $d(\cdot, \cdot)$ is an intrinsic distance in the space. An intrinsic distance is the length of the shortest geodesic path between two points, where the path remains in the space at all times (e.g., Huckemann et al., 2010). The minimized value of the expected squared distance is known as the Fréchet variance, and we assume that a global minimum is obtained. If instead only a local minimum has been found, we denote this as the Karcher mean (Karcher, 1977).

Also, we could consider extrinsic distance between two points, where a space is embedded in a higher dimensional Euclidean space. The extrinsic distance is taken as the Euclidean distance between the points in the embedding space. The population extrinsic mean

$$\mu_E = \arg \inf_{\mu \in Q} \int_Q d_E(u, \mu)^2 h(u) du, \tag{4}$$

where $d_E(\cdot, \cdot)$ is an extrinsic distance (e.g., Bhattacharya and Patrangenaru, 2003). Models can be specified in the quotient space itself and we can perform inference on μ_Q or μ_E. The method requires optimization over the γ parameters in order to compute the intrinsic distances in the shape spaces. This is the approach used in Procrustes analysis (Goodall, 1991) in landmark shape analysis.

Type of mean	Notation	Reference
Ambient space mean function	μ_A	Equation (2)
Quotient space/Fréchet/Karcher mean function	μ_Q	Equation (3)
Extrinsic mean function	μ_E	Equation (4)
Ambient space mean vector	$\mu_A([t])$	Section 5.1
Quotient space mean vector	$\mu_Q([t])$	Section 5.1

Table 2: Notation for the types of population means.

A summary of the notation for the different types of population means considered in the paper is given in Table 2.

In terms of practical guidance on which approach to use, statistical models in the original or ambient space are easier to specify and interpret, and so in many situations such models are more desirable to use. However, statistical inference can be difficult, with complicated marginal distributions obtained after integrating out the invariant transformations γ. On the other hand, models specified in the quotient space are easier to work with for inference; however, their interpretation in terms of the objects

under study is difficult. Ideally, then we would like to develop practical inference procedures for models in the ambient space, and our Bayesian inferential approach for ambient space models is the approach that we develop and recommend. The main disadvantage is that the computational time is longer than with quotient space based methods.

In the next section we shall describe some methods for computing distances and carrying out inference in quotient spaces for functions and curves. Then, in the following section we introduce our main approach to modeling using a Bayesian procedure in the ambient space.

3 Quotient space

3.1 SRVF and quotient space

Let f be a real valued differentiable curve function in the original space, $f(t) : [0,1] \to \mathbb{R}^m$. From Srivastava et al. (2011a) the Square Root Velocity Function (SRVF) of f is defined as $q : [0,1] \to \mathbb{R}^m$, where

$$q(t) = \frac{\dot{f}(t)}{\sqrt{\|\dot{f}(t)\|}},$$

and $\|f\|$ denotes the standard Euclidean norm. The q function is invariant under translation of the original function, and is in an ambient space. We consider situations when $m = 1$ for functions and $m = 2$ for planar shapes. In the one-dimensional functional case, the domain $t \in [0,1]$ often represents 'time' rescaled to unit length, whereas in two- and higher-dimensional cases, t represents the proportion of arc-length along the curve.

Let f be warped by a re-parameterization $\gamma \in G$, i.e., $f \circ \gamma$, where $\gamma \in G : [0,1] \to [0,1]$ is a strictly increasing differentiable warping function. The SRVF of $f \circ \gamma$ is then given as

$$q^*(t) = \sqrt{\dot{\gamma}(t)} q(\gamma(t)),$$

using the chain rule. An advantage of using the SRVF is that it can be used to compute the elastic metric

$$d(q_1, q_2) = d([q_1]_Q, [q_2]_Q) = \inf_{\gamma \in G} \|q_1 - \sqrt{\dot{\gamma}} q_2(\gamma)\|_2^2 = d_{\text{Elastic}}(f_1, f_2),$$

where $\|q\|_2 = \{\int_0^1 q(t)^2 dt\}^{1/2}$ denotes the $\mathbb{L}^2$ norm of q, and the elastic metric obeys the isometric property of (1):

$$d_{\text{Elastic}}(f_1 \circ \gamma, f_2 \circ \gamma) = d_{\text{Elastic}}(f_1, f_2),$$

where the distance is unchanged if both functions undergo a common reparameterization. For the $m = 1$ dimensional case the elastic metric is equivalent to the Fisher–Rao metric for measuring distances between probability density functions. If q_1 can be expressed as some warped version of q_2, i.e., they are in the same equivalence class, then

$d([q_1]_Q, [q_2]_Q) = 0$ in quotient space. Note that we sometimes wish to remove scale from the function or curve, and hence we can standardize so that

$$\int_0^1 q(t)^2 dt = 1. \tag{5}$$

In this case the ambient space would be the Hilbert sphere S^∞. In the $m \geq 2$ dimensional case, it is common to also require invariance under rotation of the original curve. Hence we may also wish to consider an elastic distance (Joshi et al., 2007; Srivastava et al., 2011a) defined in Q given as

$$d([q_1]_Q, [q_2]_Q) = \inf_{\gamma \in G, \Gamma \in SO(m)} \|q_1 - \sqrt{\dot{\gamma}} q_2(\gamma)\Gamma\|_2.$$

The $m = 2$ dimensional elastic metric for curves was first given by Younes (1998).

3.2 Quotient space inference

Inference can be carried out directly in the quotient space Q, and in this case the population mean is most naturally the Fréchet/Karcher mean μ_Q. Given a random sample $[q_1]_Q, \ldots, [q_n]_Q$, we obtain the sample Fréchet mean by optimizing over the warps for the 1D function case (Srivastava et al., 2011b):

$$\hat{\mu}_Q = \arg \inf_{\mu \in Q} \sum_{i=1}^n \inf_{\gamma_i \in G} \|\mu - \sqrt{\dot{\gamma}_i}(q_i \circ \gamma_i)\|_2^2.$$

In addition, for the $m \geq 2$ dimensional case (Srivastava et al., 2011a) we also need to optimize over the rotation matrices Γ_i where

$$\hat{\mu}_Q = \arg \inf_{\mu \in Q} \sum_{i=1}^n \inf_{\gamma_i \in G, \Gamma_i \in SO(m)} \|\mu - \sqrt{\dot{\gamma}_i}(q_i \circ \gamma_i)\Gamma_i\|_2^2.$$

This approach can be carried out using dynamic programming (Bradley et al., 1977, Chapter 11) for pairwise matching, then ordinary Procrustes matching (Dryden and Mardia, 1998, Chapter 5) for the rotation, and the sample mean is given by

$$\hat{\mu}_Q = \frac{1}{n} \sum_{i=1}^n \sqrt{\dot{\hat{\gamma}}_i}(q_i \circ \hat{\gamma}_i)\hat{\Gamma}_i.$$

Each of the parameters is then updated in an iterative algorithm until convergence.

4 A Bayesian ambient space model

4.1 The likelihood for functions

Our main approach is to consider a model in the ambient space, and then remove the unwanted transformations by marginalization. This Bayesian model was initially

described briefly by Cheng et al. (2014), and here we give a more complete description and extend the model to higher dimensions.

Since the q-function is a continuous function in the ambient space, naturally we consider a general stochastic process as the modeling framework for q, and we first consider the $m = 1$ dimensional case. We assume a zero mean Gaussian process for the difference of two 1D q functions, i.e., $\{q_1 - q_2^*|\gamma\} \sim GP$, where q_1 is untransformed and q_2^* is warped by a fixed reparameterization γ, i.e., $q_2^*(t) = \sqrt{\dot{\gamma}(t)}q_2(\gamma(t))$. The relative alignment function γ contains the parameters of interest.

If we use $q_1([t])$ and $q_2^*([t])$ to denote vectors evaluated at the same finite points on the domain of of $q_1(t)$ and $q_2^*(t)$, respectively, then the joint distribution of these finite differences $q_1([t]) - q_2^*([t])$ is a multivariate normal distribution based on the Gaussian process assumption, i.e.,

$$\{q_1([t]) - q_2^*([t])|\gamma\} \sim N_k(0_k, \Sigma_{k\times k}),$$

where here k is the number of points. We assume $\Sigma_{k\times k} = \frac{1}{2\kappa}I_{k\times k}$, where κ is a concentration parameter, although one may use more general covariance functions, such as the Gaussian or Matérn functions (Stein, 1999). Note that the assumption of independent differences implied in the Gaussian model above is at the level of q function, not on the original function itself. On the original function level, our model is analogous to a first order random walk model, which is commonly used in many applications. It encourages functions to have similar neighboring values, and hence the first derivatives are independent and identically distributed. One may also consider the second order random walk model, which encourages smoothness, e.g., similar neighboring slopes. The main advantage of our model is that it is a non-stationary model for the original function itself, allowing non-constant mean and also allowing the values on the function to be dependent. In Appendix E of the Supplementary Materials (Cheng et al., 2015), we demonstrate the performance of the proposed estimators obtained when this assumption of independent differences is violated, and we see that they are not very sensitive for pointwise estimation.

4.2 Prior distributions

The re-parameterization function $\gamma \in G$: $[0, 1] \to [0, 1]$ is a strictly increasing cumulative distribution function (c.d.f.), and this c.d.f. can be approximated by a set of equally spaced points along its domain $[0, 1]$ and linear interpolation. Let $\gamma([t])$ denote $\{\gamma(t_i), i = 0, 1, 2, \ldots, M\}$, the finite collection of $M + 1$ discretized points and $t_i = \frac{i}{M}$, then we have $\gamma(t_0) = \gamma(0) = 0$ and $\gamma(t_M) = \gamma(1) = 1$. We do not require equally spaced points, but we use this formulation for convenience. Further, if we let $p_i = \gamma(t_i) - \gamma(t_{i-1})$ for $i = 1, 2, \ldots, M$, we have $0 < p_i < 1$ and $\sum_{i=1}^{M} p_i = 1$. Denoting $\boldsymbol{p}_M = (p_1, p_2, \ldots, p_M)$ and treating $\boldsymbol{p}_M$ as a random vector, we can assign a Dirichlet prior to $\boldsymbol{p}_M|\gamma([t])$, i.e., $\pi(\boldsymbol{p}_M) \sim \text{Dirichlet}(a_1, \ldots, a_M)$. We take equal $a_i = a$ here, writing Dirichlet(a). This prior distribution is uniform when $a = 1$, and a larger value of a leads to a transformation more concentrated on $\dot{\gamma} = 1$ (i.e., translations). In the limit as $M \to \infty$, the warping function is a Dirichlet process. The choice of M is user specific, but it should be less

than the number of discrete points in the q functions, i.e., $M < k$. A possible way to choose M is to use the Deviance Information Criterion (Spiegelhalter et al., 2002), and further approaches to deal with the infinite dimensional nature of Dirichlet processes are finite truncations (Ishwaran and James, 2001), slice samplers (Walker, 2007; Kalli et al., 2011) and adaptive truncation (Griffin, 2014).

The prior distribution for the concentration parameter κ is taken as a Gamma(α, β) distribution, independent of γ. We use a fairly non-informative prior throughout with $\alpha = 1, \beta = 1{,}000$, and hence we have prior mean $E[\kappa] = \alpha\beta = 1{,}000$ and prior variance $\alpha\beta^2 = 1{,}000{,}000$.

4.3 Pairwise function comparison

Using Bayes Theorem, the posterior distribution for $\{\gamma([t]), \kappa\}$ given $(q_1([t]), q_2([t]))$ is

$$\pi(\gamma, \kappa | q_1, q_2) \propto \kappa^{p/2} e^{-\kappa \| q_1([t]) - \sqrt{\dot{\gamma}}(q_2([t]) \circ \gamma) \|^2} \pi(\gamma)\pi(\kappa).$$

In the above model, p represents the degrees of freedom in the model. If there is no unit scale length constraint (5) for q, then p would be calculated as follows: $p = km$, where k is the number of finite points taken from the q function, and m is the original function space dimension, i.e., $m = 1$ for functions and $m \geq 2$ for curves in higher dimensions. One degree of freedom will be lost in the constrained case (5) and thus $p = km - 1$.

We use a Markov chain Monte Carlo (MCMC) algorithm to simulate from the joint posterior distribution of $\{\gamma([t]), \kappa\}$. The concentration parameter κ is updated using a Gibbs sampler as the conditional posterior for κ given all other parameters is still Gamma distributed. For $\gamma([t])$ with $M + 1$ points, a shift in $\gamma([t])$ is proposed at each discrete point $(i = 1, \ldots, M - 1)$ and accepted/rejected according to a Metropolis–Hastings step. Note that $\gamma(t_0) = 0$ and $\gamma(t_M) = 1$ are both fixed and thus are not updated. The resulting Markov chain is irreducible and aperiodic, and hence dependent values from the posterior distribution can be simulated after a large number of iterations.

4.4 Multiple functions

If we are interested in multiple functions or curves, we can specify a mean process for q functions in the ambient space, i.e., $E(q_i^*) = \mu_A$, where $q_i^* = \sqrt{\dot{\gamma}_i} q_i(\gamma_i)$ is a warped version of q_i through some underlying fixed γ_i. Based on the Gaussian process assumption again, we have

$$\{q_i^*([t]) - \mu_A([t]) | \gamma_i([t]), \mu_A([t])\} \sim N(0_k, \Sigma_{k \times k})$$

for $i = 1, 2, \ldots, n$, where n is the number of q functions of interest. We take the prior distribution of μ_A to be a zero mean Gaussian process with large variance, independent of all other parameters. The joint posterior density for $(\mu_A, \gamma_1, \ldots, \gamma_n)$ is then

$$\pi(\mu_A, \gamma_1, \ldots, \gamma_n | q_1, \ldots, q_n) \propto \kappa^{np/2} e^{-\kappa \sum_{i=1}^n \| \mu_A([t]) - q_i^*([t]) \|^2} \pi(\mu_A)\pi(\gamma_1, \ldots, \gamma_n)\pi(\kappa).$$

To simulate from the posterior distribution, we again use an MCMC algorithm, consisting of pairwise MCMC updates from each curve to the current mean $\mu_A([t])$ and a Gibbs update for $\mu_A([t])$ itself.

Due to the simultaneous invariance of all the q functions to a common warping, $\mu_A([t])$ is only identifiable up to an equivalence class of warpings. Therefore, in order to compute the posterior mean estimate $\hat{\mu}_A([t])$, it is helpful to standardize in each MCMC iteration such that the Karcher mean of the warping functions from μ_A to each q_i is the identity function, i.e., $\hat{\dot{\gamma}} = 1$. The standardization is carried out by applying the inverse of Karcher mean of the warping functions from μ to the individual q_i's.

4.5 Curve warping

In the $m \geq 2$ dimensional case, we consider a Gaussian process for the difference of two vectorized q functions in a relative orientation Γ, i.e., $\{\text{vec}(q_1 - q_2^*)|\gamma, \Gamma\} \sim GP$, where $q_2^* = \sqrt{\dot{\gamma}} q_2(\gamma)\Gamma$. The matrix $\Gamma \in SO(m)$ is a rotation matrix with parameter vector θ. If we assign a prior for rotation parameters (Eulerian angles) θ corresponding to rotation matrix Γ, then the joint posterior distribution of $(\gamma([t]), \theta)$, given $(q_1([t]), q_2([t]))$ is

$$\pi(\gamma, \theta|q_1, q_2) \propto \kappa^{p/2} e^{-\kappa \|q_1([t]) - q_2^*([t])\|^2} \pi(\gamma)\pi(\theta)\pi(\kappa),$$

where γ, θ, κ are independent *a priori*. Throughout the paper we take Γ to have a uniform prior on the space of rotation matrices, e.g., see Czogiel et al. (2011).

For the multiple curves case, define $q_i^*(t) = \sqrt{\dot{\gamma}_i(t)} q_i(\gamma_i(t))\Gamma_i$ and $\mu_A = E(q_i^*)$ for fixed γ_i and Γ_i, and we assume

$$\text{vec}(\mu_A([t]) - q_i^*([t])) \sim N(0_{km}, \Sigma_{km \times km})$$

for fixed (γ_i, Γ_i), $i = 1, \ldots, n$. The joint posterior for $(\mu_A, \gamma_1, \ldots, \gamma_n, \Gamma_1, \ldots, \Gamma_n)$ is

$$\pi(\mu_A, \gamma_1, \ldots, \gamma_n, \Gamma_1, \ldots, \Gamma_n|q_1, \ldots, q_n) \propto$$

$$\kappa^{np/2} e^{-\kappa \sum_{i=1}^n \|\mu_A - q_i^*\|^2} \pi(\mu_A)\pi(\gamma_1, \ldots, \gamma_n)\pi(\Gamma_1, \ldots, \Gamma_n)\pi(\kappa),$$

with warps, rotations and κ independent *a priori*. Sampling from the posterior distribution is carried out through exactly the same procedure as when $m = 1$ but with an extra Metropolis–Hastings update for rotation angles.

4.6 Alternative Bayesian approaches

An earlier Bayesian method for curve registration is Bayesian hierarchical curve registration (BHCR) of Telesca and Inoue (2008) where B-splines were used to approximate both the mean functional curve $E[f(t)]$ and the alignment function $\gamma(t)$, with a monotonicity assumption. Rather than modeling coefficients of basis functions, our model uses discrete realizations of stochastic processes (Gaussian and Dirichlet), and we work with the derived function $q(t)$ and the alignment function $\gamma(t)$, with monotonicity built-in with a cumulative distribution function and linear interpolation. The key difference is that working with the q function enables the method to be invariant under simultaneous warping of all the curves, but this is not the case with the BHCR method. Both methods use MCMC simulation for posterior inference.

As an extension for further work we could also use a basis function approach, which would introduce more smoothness. In the case of high dimensional data, such as the mass spectrometry data of Koch et al. (2014) which has a large number of spikes, it is relatively expensive to approximate the original curve using B-splines. A recent Bayesian approach of Zhou et al. (2014) also uses B-splines for the warping function but with individual curves having different time domains. The same restrictions on the coefficients for monotonicity and a similar model to the BHCR approach is used, and an underlying common mean function applied to degradation signals of engineering components. A further Bayesian approach is introduced by Claeskens et al. (2010), where multi-resolution warping functions are considered. The warping functions are rather different, with multiresolution properties as in wavelets. MCMC methods are again used with both methods, but the key difference again is that the mean function $f(t)$ is used which is not invariant to simultaneous warping, unlike $q(t)$ that we use in our case.

5 Properties

5.1 Asymptotic properties

Let us write ϕ for the vector of all the parameters in $\{(\gamma_i, \Gamma_i), i = 1, \ldots, n.\}$, and consider μ_A to be represented by a piecewise linear function connecting a finite number k points given by km-vector $\mu_A([t])$. The marginal posterior density for ambient space inference is given by

$$\pi_A(\mu_A([t]), \kappa|X) = \int_\phi \pi(\mu_A([t]), \kappa, \phi|X)d\phi. \tag{6}$$

The posterior mode estimator of $(\mu_A([t]), \kappa)$ is written as $(\hat{\mu}_A([t]), \hat{\kappa})$ and is obtained by maximizing (6). If the prior distribution of $(\mu_A([t]), \kappa)$ is uniform then $(\hat{\mu}_A([t]), \hat{\kappa})$ is the maximum likelihood estimator. If the prior is absolutely continuous in a neighborhood of $\mu_A([t])$ with continuous positive density at $\mu_A([t])$ and the distribution satisfies certain regularity conditions (including differentiable in quadratic mean with non-singular Fisher information matrix $I_{\mu_A([t])}$), then consistency and asymptotic normality follow. Subject to the conditions of the Bernstein–von Mises theorem (van der Vaart, 1998, p. 141), we have

$$\sqrt{n}(\hat{\mu}_A([t]) - \mu_A([t])) \to N(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} I_\mu^{-1} \dot{\ell}_{\mu_A([t])}(X_i), I^{-1}_{\mu_A([t])})$$

in total variation norm as $n \to \infty$, where $\dot{\ell}_{\mu_A([t])}(X_i)$ is the derivative of the log-likelihood corresponding to observation i. If $\hat{\mu}_A$ is a piecewise linear function obtained from the vector $\hat{\mu}_A([t])$, because $\hat{\mu}_A([t])$ is consistent for $\mu_A([t])$ we can state that $\hat{\mu}_A \to \mu_A$ in probability as $n \to \infty$, and hence the ambient space mean is consistent. Allassonnière et al. (2007) and Allassonnière et al. (2010) give detailed discussion of consistency in ambient space models, in particular for deformable templates in image analysis.

The sample Fréchet mean vector $\hat{\mu}_Q([t])$ is consistent for the population Fréchet mean vector $\mu_Q([t])$ (Kendall, 1990; Le, 1991) provided the distribution has support within a regular geodesic ball, and if $\hat{\mu}_Q$ is a piecewise linear function obtained from the vector $\hat{\mu}_Q([t])$ and because $\hat{\mu}_Q([t])$ is consistent for $\mu_Q([t])$, we can state that $\hat{\mu}_Q \to \mu_Q$ in probability as $n \to \infty$.

The results in the subsection concern the asymptotic normality of a finite set of points on functions. Detailed consideration of similar results for the estimators of entire functions is a topic for further work.

5.2 Comparison of the quotient and ambient space methods

In general, the population Fréchet mean μ_Q in the quotient space and the ambient space mean μ_A do not have the same amplitude/shape, and hence the sample Fréchet mean can be inconsistent for the ambient space mean. Likewise, the sample ambient space mean can be inconsistent for the population Fréchet mean. It is most natural therefore to use the appropriate estimators given the choice of mean that is to be estimated. If we are interested in the amplitude/shape of the population ambient space mean then we use ambient space inference, while if we are interested in the population Fréchet mean then we use the sample Fréchet mean. As we see below there are situations where the sample ambient space and Fréchet estimators are very similar, and so our choice between them may be made on other grounds in this case, such as ease of computation.

When the prior distributions are uniform in the parameters an identical estimator to the sample Fréchet mean $\hat{\mu}_Q$ is obtained from the posterior mode in the Bayesian model of the previous section. If the priors are not uniform then the posterior mode is in fact a penalized quotient estimator, with the objective function

$$\hat{\mu}_{pen} = \arg \inf_{\mu \in Q} \sum_{i=1}^{n} \inf_{\gamma_i \in G, \Gamma_i \in SO(m)} \{-\log \pi(\mu, \kappa, \gamma_i, \Gamma_i | q_1, \ldots, q_n)\}$$

for the curve case.

Note that in many practical situations the ambient space estimator and penalized quotient space estimators are quite similar, although they do not have to be. One reason for the similarity in practice is due to a Laplace approximation, and the marginal posterior density (for ambient space inference) is given by

$$\pi_A(\mu, \kappa | X) = \int_{\phi} \pi(\mu, \kappa, \phi | X) d\phi. \tag{7}$$

whereas the penalized quotient space estimator is obtained by maximization of

$$\pi_Q(\mu, \kappa | X) \propto \sup_{\phi} \pi(\mu, \kappa, \phi | X). \tag{8}$$

where $X = \{q_1, \ldots, q_n\}$. Often we can consider $\pi_Q(\mu, \kappa | X)$ in (8) to be a good approximation to the marginal density (7) where the integral is approximated using Laplace's method:

$$\begin{aligned}\int_{\phi} \pi(\mu,\kappa,\phi|X)d\phi &= \int_{\phi} b(\phi)\exp\{-Ar(\phi)\}d\phi \\ &\approx b(\hat{\phi})\left(\frac{2\pi}{A}\right)^{p/2}|\Sigma_{\hat{\phi}}|^{1/2}\exp\{-Ar(\hat{\phi})\} \\ &\propto \sup_{\phi} \; b(\phi)\exp\{-Ar(\phi)\} \\ &\propto \pi_Q(\mu,\kappa|X),\end{aligned}$$

where the gradient of $r(\phi)$ is zero at $\hat{\phi}$, $\Sigma_{\hat{\phi}}$ is the inverse of the positive definite Hessian matrix of $r(\phi)$ at $\hat{\phi}$ and A is a constant. This approximation will be reasonable when the underlying distribution $(\phi|\mu,\kappa)$ is unimodal. Laplace's approximation is exact when $(\phi|\mu,\kappa)$ is multivariate Gaussian, i.e., $r(\phi)$ is a quadratic form in ϕ and $b(\phi)$ is constant. For further discussion on similar comparisons for unlabeled landmark shape analysis, see Kenobi and Dryden (2012). In cases where the Laplace approximation does not hold, e.g., under multimodality, then one should not rely on the quotient estimator as a reasonable estimate of the ambient space mean.

5.3 Multimodality

Multimodality of the posterior distribution can often be an issue with registration of functions and curves. Simulated tempering (Geyer and Thompson, 1995) is a powerful simulation technique designed to overcome problems in moving between local modes of the posterior. The key idea is to first jump from the "cold" temperature (target distribution), where it is difficult to move out of a local mode to a "hot" temperature where movement between modes is easier and then jump back to the "cold" temperature. Using this procedure, the MCMC algorithm can explore the sample space in a more efficient manner. Further details are given in Appendix F in the Supplementary Materials (Cheng et al., 2015).

6 Simulations and applications

6.1 1D data analysis

Simulation study

We consider now a simulation study to compare estimation properties of the quotient and ambient space estimators. The quotient space estimator $\hat{\mu}_Q$ is obtained by minimizing $\Sigma_{i=1}^{n}\|\mu - \sqrt{\dot{\gamma}_i}(q_i \circ \gamma_i)\|_2^2$ using dynamic programming while the ambient space estimator $\hat{\mu}_A$ is obtained using the pointwise mean of posterior samples from MCMC iterations after convergence. The detailed algorithm for dynamic programming is given in Appendix A of the Supplementary Materials; and the MCMC algorithm is described in greater detail in Appendix B of the Supplementary Materials (Cheng et al., 2015).

In a single Monte Carlo simulation repetition, a sample of q-functions in one dimension is generated through the model $q_i([t]) = \sqrt{\dot{\gamma}_i}\mu_A(\gamma_i([t])) + e_i([t])$, where $e_i \sim N(0_k, \Sigma_{k\times k}), \Sigma = \sigma^2 I_{k\times k}$ and $\gamma_i \sim \text{Dirichlet}(1)$ for $i = 1,\ldots,n$.

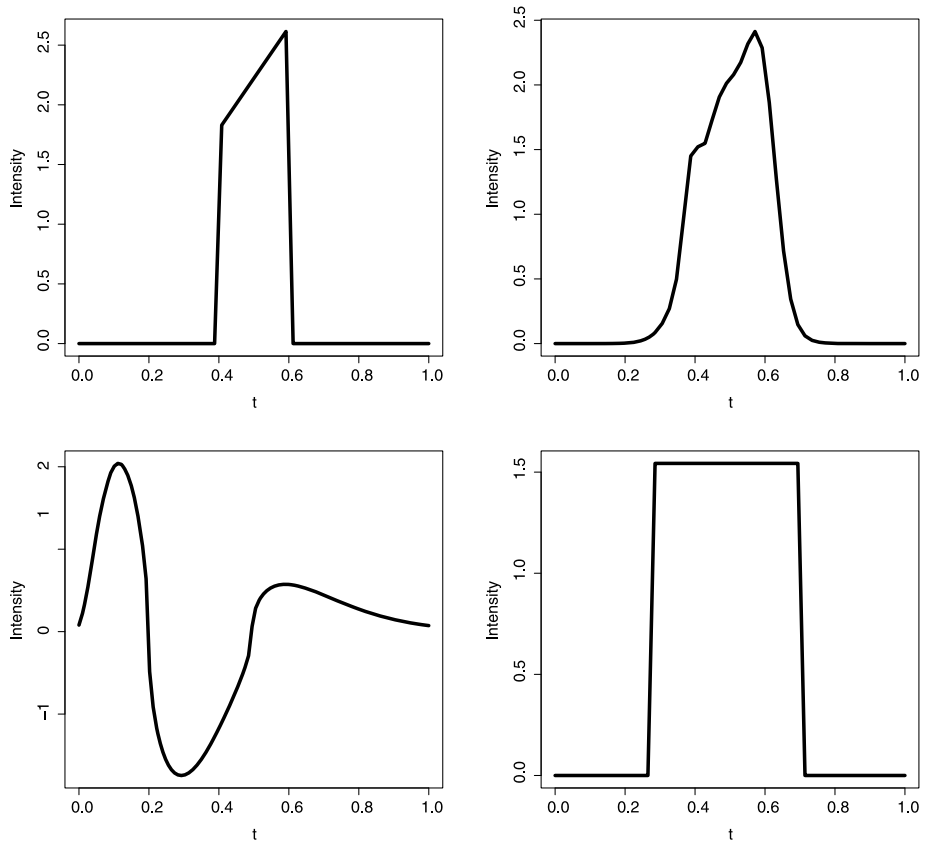

Figure 2: The true $\mu_A(t)$ functions used for simulation study. From left to right we denote the functions as Example I, II, III and IV, respectively.

Both $\hat{\mu}_Q$ and $\hat{\mu}_A$ are computed and their Fisher–Rao distances to the underlying true μ_A are calculated. Note that since the goal is to estimate μ_A in the ambient space, it is expected that $\hat{\mu}_A$ will be more appropriate than $\hat{\mu}_Q$. The MCMC algorithm for $\hat{\mu}_A$ is run for 50,000 iterations with a 25,000 iteration burn-in period. The prior for γ in the Bayesian model is taken as Dirichlet with $a = 1$, i.e., uniform. Given specific combinations of sample size $n \in \{5, 10, 20, 30, 50, 100, 200\}$ and error standard deviation $\sigma \in \{0.1, 0.3, 0.5, 1\}$, 100 Monte Carlo repetitions are run and the arithmetic means of squared Fisher–Rao distances from both estimators to μ_A are recorded.

Four examples for μ_A are considered, which are all scaled to have unit length and unit time. The functions μ_A in Examples I, II, III, IV given in Figure 2 are evaluated at k equal to $51, 51, 101, 51$ points, respectively, and the warping functions are parameterized using $M + 1$ points, where M is equal to $10, 10, 20, 10$, respectively. The underlying μ_A functions in Examples I and IV are piecewise linear, Example II is a mixture of three normal densities, and Example III is the derivative of the difference of two Gamma

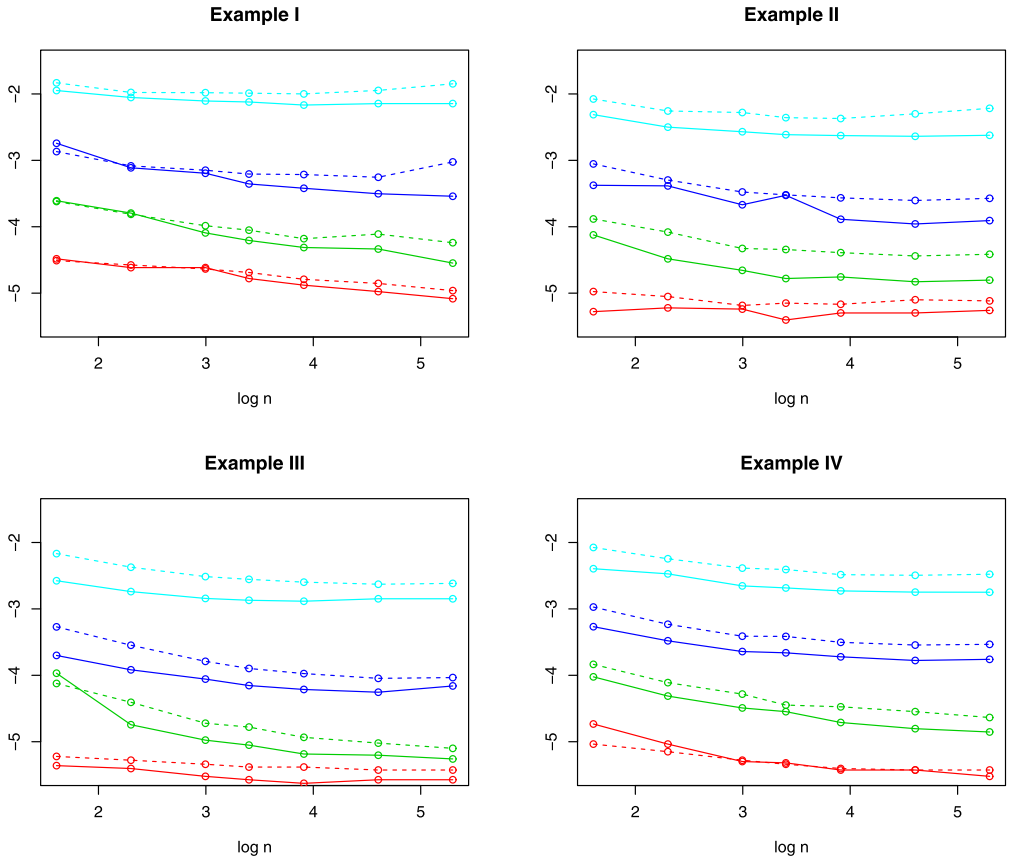

Figure 3: The logarithm of the mean square Fisher–Rao distance to the true mean μ_A versus logarithm of sample size n. The full line is the ambient space estimator and the dotted line is the quotient space estimator. The colors are red ($\sigma = 0.1$), green ($\sigma = 0.3$), blue ($\sigma = 0.5$) and cyan ($\sigma = 1$).

functions (in fact, it is the derivative of the canonical haemodynamic response function often used to model the blood oxygen level dependent signals in fMRI (Glover, 1999)). The corresponding distances from both estimators are given in Figure 3.

From Figure 3 we that see that when σ is smaller, the average squared distance between the estimate and true value is smaller, and as n increases in general the average squared distance becomes smaller. When σ is small (0.1), the performance of both estimators is almost equivalent. However, for larger σ in nearly all cases there is an advantage in using the ambient space estimator. One possible explanation could be over-warping of the quotient estimator to the noisy data due to the optimization over warpings, compared to the integration over warpings in the ambient space estimator. For large $\sigma \geq 0.5$ both procedures are clearly biased for these values of n, but it must

be borne in mind that the signal to noise ratio is very low in these cases and so the estimation is very challenging and the discrete implementation will have an important effect.

It is also worth pointing out that the quotient estimator we obtained here should be the same as that from Srivastava et al. (2011a) in theory although we use a quite different Dynamic Programming (DP) implementation to search the optimal warping function. To avoid the potential bias caused by different DP methods, we also perform a similar simulation study using their implementation for the quotient estimator for comparison. The observed pattern is almost the same, i.e., our ambient estimator has comparable performance to quotient counterpart for smaller σ while a better performance for larger σ.

Overall, from the above examples it does seem that there is an advantage in using the ambient space estimator as we expect, although this is at the cost of longer computational time. In Appendix C of the Supplementary Materials (Cheng et al., 2015), we report some computational times.

Growth rate curves

We consider the Berkeley growth study girls dataset (Tuddenham and Snyder, 1954; Ramsay and Silverman, 2005), where the smoothed growth rate curves of 54 girls over the period 1 to 18 years are displayed in Figure 4 (top left panel), with time rescaled to unit length, as well as the original q-functions (top right panel).

We apply our Bayesian methodology in order to align the full set of curves, and we use prior parameter $a = 50$ and $M = 20$ with 200,000 MCMC iterations and 100,000 burn-in. Convergence of the MCMC algorithm was monitored by trace plots in Figure D.1 of the Supplementary Materials (Cheng et al., 2015). In Figure 4 (bottom left panel), we display the registered q-functions using maximum a posteriori (MAP) estimates of the warping functions. In the same figure (bottom left panel), we plot the posterior mean and 95% credibility intervals for the mean q-function μ_A, calculated from the 100,000 MCMC iterations after burn-in.

Note that because μ_A is an equivalence class it can be helpful for interpretation to plot a particular member of the class, which is called an icon in shape analysis (Goodall, 1991). The mean and credibility intervals for the icon curves can be constructed using the inverse relationship (Srivastava et al., 2011a)

$$f(t) = \int_0^t q(s)|q(s)|ds, \tag{9}$$

and then a suitable scaling and translation is chosen for the icon on the scale of the original curves. We translate and scale so that the posterior mean icon curve matches the mean of the data curves at $t = 0$ and $t = 1$. Also, for each constructed icon curve after burn-in a Gaussian height shift is added with standard deviation given by the average of the standard errors of the sample mean curve, in order to plot the icons on the scale of the original data.

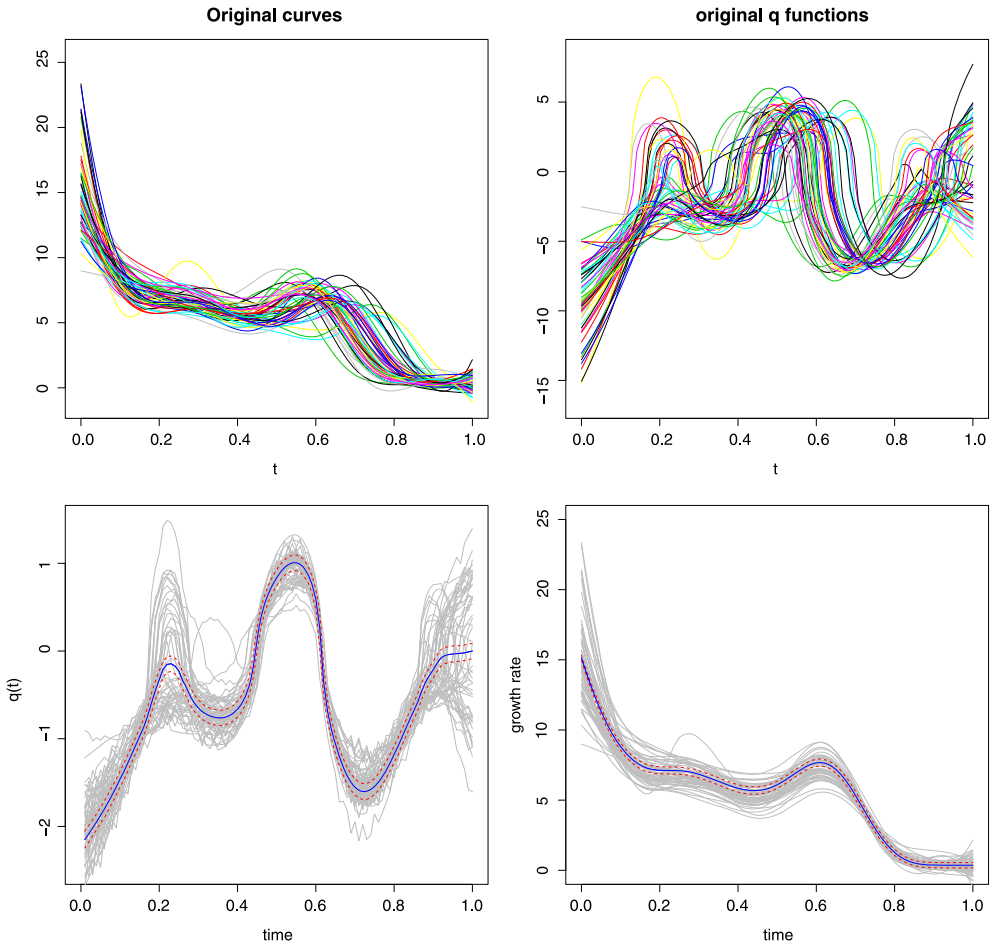

Figure 4: Growth rate curves from the girls in the Berkeley growth study (top left) and their unregistered q-functions (top right). In the bottom left figure, we display the aligned, scaled q-functions from our Bayesian procedure (using MAP estimated warps) together with the posterior mean of μ_A in blue and pointwise 95% credibility intervals as dashed red lines. In the bottom right figure, we display the registered curves and posterior mean (blue) and pointwise 95% credibility intervals (dashed red) as icons of μ_A. The prior parameter is $a = 50$ here. In each plot the x-axis represents time from 1 to 18 years, and has been rescaled to unit length.

The posterior mean and pointwise 97.5% and 2.5% quantiles of the 100,000 icons after burn-in are displayed, together with the registered curves, in the bottom right plot of Figure 4. There is good alignment of the curves, and we see clearly that there are two growth spurts in the posterior mean: the smaller mid-growth spurt around $t = 0.3$, and the strong pubertal growth spurt around $t = 0.6$. These two growth

spurts have been well-documented in the literature (e.g., Molinari et al., 1980; Gasser et al., 1985; Tang and Müller, 2008). Our method has also provided an indication of the uncertainty in the mean, which is relatively small due to the large sample size ($n = 54$). We consider a sensitivity analysis of the choice of prior parameter a in Appendix D of the Supplementary Materials (Cheng et al., 2015). Also, note that the posterior standard deviation of the mean curve is similar for different values of t. An extension of the model could be to have the covariance of the q-function given by a more general parameter matrix, e.g., with a Wishart prior, which could be considered in further work.

Comparison with other methods

We compare our method with some other existing methods (Ramsay and Silverman, 2005; Srivastava et al., 2011b) and we conduct an alignment comparison using four datasets: (1) Berkeley growth data – boys, (2) Berkeley growth data – girls, (3) Handwriting data, (4) Mass Spectrometry data. The first three datasets (1)–(3) are from Ramsay and Silverman (2005), and the mass spectrometry dataset is from Koch et al. (2014). The evaluation criteria used are:

- The synchronization (Sync) coefficient defined in James (2007),

$$\text{Sync} = \frac{1}{N}\sum_{i=1}^{N}\frac{||f_i^* - \frac{1}{N-1}\sum_{j\neq i} f_j^*||^2}{||f_i - \frac{1}{N-1}\sum_{j\neq i} f_j||^2}.$$

- The inverse of pairwise correlation (IPC),

$$\text{IPC} = \frac{\sum_{i\neq j} r(f_i, f_j)}{\sum_{i\neq j} r(f_i^*, f_j^*)},$$

where $r(\cdot,\cdot)$ is the pairwise Pearson's correlation between functions, and f denotes the original curve function while f^* denotes the aligned version. For both measures, smaller values indicates better alignment.

In the experimental setup, for Datasets 1 and 2, for the spline based method of Ramsay and Silverman (2005) we directly use the alignment result provided in the R package `fda` for males (Ramsay et al., 2013) and MATLAB code for females (Ramsay, 2013). For Datasets 3 and 4 we tune the model parameters in order to make the comparison as fair as possible. For the Square Root Velocity Function (SRVF) method of Srivastava et al. (2011b), we used the R package `fdasrvf` (Tucker, 2014) with its default setting for all four datasets. For our Bayesian version of SRVF (B-SRVF), we also use the default setting with the prior of warping function to be Dirichlet(1) and the iteration number for MCMC is fixed at 50,000 without further fine-tuning. For B-SRVF the posterior mean is used as the estimator in its standardized form (i.e., with the Karcher mean of the deformations equal to the identity).

From the table we can see that for Datasets 1 and 2, even though they are similar in nature (both are growth data), the relative performance of the methods is different.

	1D Alignment Comparison							
	Dataset 1		Dataset 2		Dataset 3		Dataset 4	
Methods	Sync	IPC	Sync	IPC	Sync	IPC	Sync	IPC
Spline	0.74	0.95	**0.54**	**0.95**	0.56	0.59	0.80	0.41
SRVF	0.67	0.93	0.68	**0.95**	**0.54**	**0.57**	**0.26**	0.18
B-SRVF	**0.64**	**0.90**	0.61	**0.95**	0.56	**0.57**	**0.26**	**0.17**

Table 3: 1D alignment result evaluation using four data sets with smallest values highlighted in bold.

Notice that for Dataset 1 the spline method performs the worst, while in Dataset 2 it performs the best. In this sense, there is some sensitivity in the relative performance of the methods to the datasets we use. In Dataset 4, where there are many spikes, we can see that our method does much better than the spline method, but it is comparable to the SRVF approach. Overall, our method generally has an advantage over the spline whereas it is comparable to SRVF in all four datasets. Although the performance in alignment between SRVF and B-SRVF is quite similar here, the advantage of the Bayesian approach is that assessment of the uncertainty is also provided via credibility intervals and prior information can be incorporated as desired.

6.2 2D data analysis

Mouse vertebrae

A two-dimensional application is the study of the shape of the second thoracic (T2) vertebrae in mice (Dryden and Mardia, 1998). Three groups of mouse vertebrae are available: 30 Control, 23 Large and 23 Small bones. The Large and Small group underwent genetic selection for large/small body weight, whereas the Control group consists of unselected mice. Each bone is represented by a curve consisting of 60 points which are determined through a semi-automatic procedure. Six landmarks are placed at points of high absolute curvature and then nine pseudo-landmarks are equally-spaced in-between each pair of landmarks. The main interests here include carrying out pairwise registration, obtaining mean shapes and credibility intervals, and carrying out classification based on the registered shapes. It is very common in many application areas to classify objects using shape information (Dryden and Mardia, 1998), and, for example, in studying the fossil record there is a need to classify bones from individuals into groups using size and/or shape as there is usually little or no other information available.

We start our analysis by performing a pairwise comparison from the ambient space model, and we use the MCMC algorithm for pairwise matching with 50,000 iterations. The q-functions are obtained by initial smoothing, and then normalized so that $\|q\|_2 = 1$. The registration is carried out using rotation through an angle θ about the origin, and a warping function γ. The original and registered pair (using a posterior mean) are shown in Figure 5 and the pointwise correspondence between the curves and a pointwise 95% credibility interval for $\gamma(t)$ are shown in Figure 6. The start point of the curve is fixed and is given by the left-most point on the curve in Figure 6 that has a red line connecting the two bones. The narrower regions in the credibility interval correspond well with high curvature regions in the shapes. We also applied the multiple curve registration, as

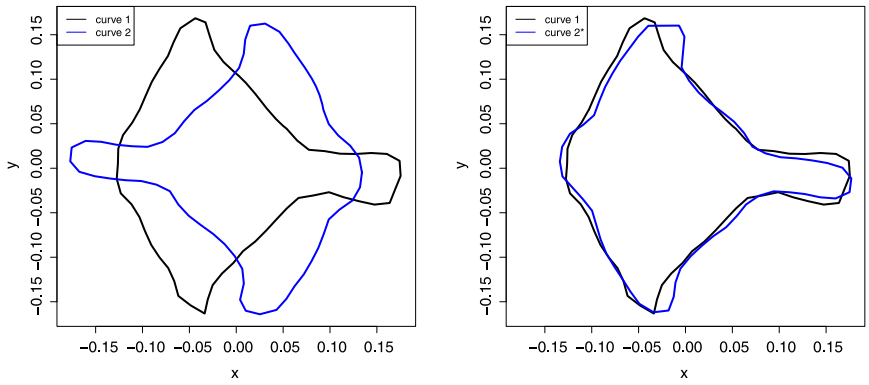

Figure 5: Unregistered curves on left and registration through $\hat{\gamma}(t)_A$ on right.

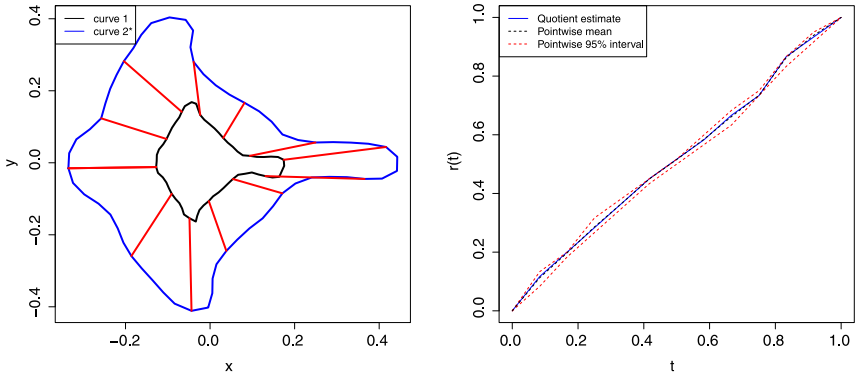

Figure 6: Correspondence based on $\hat{\gamma}(t)_A$ and 95% credibility interval for $\gamma(t)$. One of the bones is drawn artificially smaller in order to better illustrate the correspondence.

shown in Figure 7. Convergence of the MCMC schemes were monitored by trace plots, which can be seen in Appendix D, Figure D.2 of the Supplementary Materials (Cheng et al., 2015).

In order to investigate the differences between the new Bayesian method and the classic Procrustes analysis on the 60 landmarks we consider a classification study. For classification method A, the three group means are obtained through classical generalized Procrustes analysis (Goodall, 1991) using the `shapes` package in R (Dryden, 2014) and each test curve is assigned to the trained group which is closest in terms of Procrustes distance. The Procrustes distance is calculated by minimizing the Euclidean sum of squares between the landmark configurations using translation, rotating and scale. For method B, the three group means are obtained using the posterior mean from the Bayesian model and each test curve is classified based on the elastic distance to the mean (i.e., using amplitude variability). For method C, all training dataset curves are registered in one pooled group using generalized Procrustes analysis and the Procrustes registered curves are used as the training data. Each test curve is aligned to the mean by

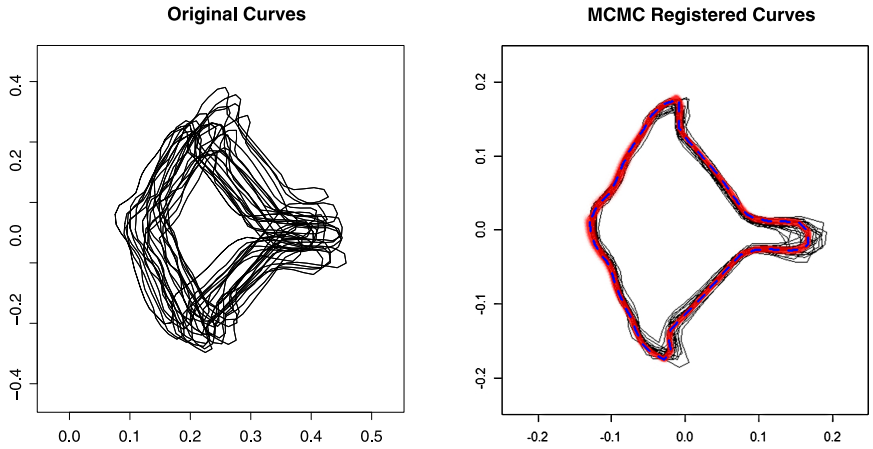

Figure 7: The original curves from Small group, without and with registration. The blue curve in the right panel is the estimated μ_A and red color shows the credible region given by 10,000 samples of mean.

ordinary Procrustes analysis. A Support Vector Machine (SVM) (Chang and Lin, 2001) is then trained on the registered training curves and applied to the registered test curves. For method D, all training dataset curves are registered through the Bayesian model and their warped, registered versions are used as the training data. Each test curve is aligned to the mean by pairwise registration using MCMC. An SVM is then trained on the MCMC registered training curves and applied to the registered test curves.

A total of 100 Monte Carlo repetitions are run for each exercise, where the training data and test data are sampled from each group without replacement. In a single Monte Carlo repetition, 16 curves from the Small group, 20 from the Control group and 16 from the Large group (about two-thirds of the original data) are randomly selected as the training data, while the remaining 24 curves are used as the test data. Method A gives an 80% correct classification rate for the test data, and method B gives 83% correct classification. In method C, the classification rate increases to 87% while method D has the highest classification rate of 92%. Under both A vs B and C vs D circumstances, we see some improvement in classification by using Bayesian alignment while we also notice an overall improvement in methods C and D compared to A and B by using SVM. The main reason for the improvement is that SVM is using hyperplanes to classify between distributions for each group, rather than shape distances which are isotropic in nature. Both method B and D demonstrate the advantage in using the Bayesian MCMC method for registration with warping, rather than just using the equally spaced pseudo-landmarks with no warping.

2D simulation and comparison

We now compare our method to generalized Procrustes analysis in the `shapes` package in R (Dryden, 2014) using the 60 points around the outline without warping. The Pro-

crustes alignment is carried out over translation, rotation and scale. We first begin with a simulation study, where independent isotropic Gaussian noise is added and uniform random rotations are also applied in order to generate the random curves. Four curves are selected at random from the T2 mouse vertebrae dataset and their q functions are used as the true μ process, and $n = 30$ curves are generated from the model. Multiple alignment is carried out and the Procrustes mean shape and the Bayesian SRVF standardized posterior mean are calculated. The comparisons are given in Table 4, where FR is the squared of Fisher–Rao distance while L2 is standard Euclidean norm and the measures are averaged over 50 Monte Carlo simulations.

		2D Estimation Comparison							
		Dataset 1		Dataset 2		Dataset 3		Dataset 4	
Methods		FR	L2	FR	L2	FR	L2	FR	L2
Procrustes	$\sigma = 0$	0.076	0.039	0.092	0.042	0.075	0.039	0.077	0.038
	$\sigma = 0.3$	0.168	0.078	0.151	0.072	0.169	0.074	0.154	0.072
	$\sigma = 0.5$	0.228	0.088	0.211	**0.085**	0.227	**0.088**	0.220	**0.085**
B-SRVF	$\sigma = 0$	**0.052**	**0.035**	**0.041**	**0.032**	**0.041**	**0.031**	**0.045**	**0.031**
	$\sigma = 0.3$	**0.086**	**0.074**	**0.075**	**0.070**	**0.077**	**0.071**	**0.079**	**0.070**
	$\sigma = 0.5$	**0.118**	**0.087**	**0.118**	0.089	**0.110**	**0.088**	**0.111**	0.086

Table 4: 2D estimation result evaluation using four functional curves.

From Table 4 we notice that Bayesian model is generally better than Procrustes method when a certain amount of warping effects truly exist in the dataset (as in this simulation study) and this is not surprising since the Procrustes method does not take warping effect into account. For datasets where alignment is not necessary, the performance of both methods is similar.

In the alignment comparison of real data sets, we use the T2 Small mouse vertebrae dataset plus three other 2D datasets which are used in Thakoor et al. (2007) (C. Subset of MPEG-7 CE Shape-1 Part-B, Datasets 4,5,6, respectively), available at:
http://visionlab.uta.edu/shape_data.htm
The alignment results are given in Table 5. It is clear again and not surprising that Bayesian model is an improvement on Procrustes in alignment, as warping is not taken into account in the Procrustes analysis.

	2D Alignment Comparison							
	T2 Small Mice		Thakoor 4		Thakoor 5		Thankoor 6	
Methods	Sync	IPC	Sync	IPC	Sync	IPC	Sync	IPC
Procrustes	0.064	0.972	0.218	0.989	0.144	0.873	0.160	0.523
B-SRVF	**0.043**	**0.967**	**0.073**	**0.974**	**0.069**	**0.849**	**0.136**	**0.467**

Table 5: 2D alignment result evaluation using 4 data sets with smallest values highlighted.

7 Discussion

In this paper we state the distinctions between three spaces of interest: the original, ambient and quotient space. We compare the ambient space estimator and quotient

space estimator in simulation studies as well as real data analysis and demonstrate a small improvement in performance in several examples. We also explain the similarity between the estimators in certain situations through a Laplace approximation.

An important component is that we incorporate prior information about the amount of warping, which is particularly useful in the growth curve application, where too much warping is not desirable. Naturally the choice of prior is important and will, of course, be problem specific; however, in the growth curve data it was clear the prior weighted towards translations was beneficial.

The methodology can be used for datasets with equally spaced points or non-equally spaced points on each curve or outline. However, there is a choice in whether to parameterize the warping function with equally spaced points or not. For our examples we did use equally spaced points in the warping function and this was reasonable in our applications. However, non-equally spaced points in the parameterization could also be used. One possible extension would be to estimate an initial warping function by equally spaced points and then identify the region where there exists a large change of slope in successive segments. More points can then be placed in those regions and a new MCMC run carried out. The choice of equal or non-equal points ought not to be crucial, except in more extreme cases.

Although some transformations such a translations (which are not 1–1 and onto) are not included in the family of deformations, a translation for most of the domain with non-linear kinks at each end will give practically the same warp and so our method retains flexibility without being restrictive.

Note that for matching between two functions we also can use multiple alignment, which also involves estimating the mean function, instead of the pairwise method. Although the multiple alignment method appears to be a less efficient approach due to the need for the mean function as parameters, it does have the property that the prior would be invariant under a common reparameterization of both curves.

Although we have focused on 1D and 2D applications the Bayesian methodology can be extended to higher dimensions, for example, analyzing the shape of 3D surface shapes using the square root normal fields (Jermyn et al., 2012).

Supplementary Material

Supplementary Materials: Bayesian Registration of Functions and Curves
(DOI: 10.1214/15-BA957SUPP; .pdf).

References

Allassonnière, S., Amit, Y., and Trouvé, A. (2007). "Towards a coherent statistical framework for dense deformable template estimation." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(1):3–29. MR2301497. doi: http://dx.doi.org/10.1111/j.1467-9868.2007.00574.x. 458

Allassonnière, S., Kuhn, E., and Trouvé, A. (2010). "Bayesian consistent estimation in deformable models using stochastic algorithms: applications to medical images." *Journal de la Société Française de Statistique*, 151(1):1–16. MR2652787. 458

Bhattacharya, R. and Patrangenaru, V. (2003). "Large sample theory of intrinsic and extrinsic sample means on manifolds. I." *The Annals of Statistics*, 31(1):1–29. MR1962498. doi: http://dx.doi.org/10.1214/aos/1046294456. 452

Bradley, S. P., Hax, A. C., and Magnanti, T. L. (1977). *Applied Mathematical Programming*. Addison-Wesley, Reading, MA. 454

Chang, C.-C. and Lin, C.-J. (2001). *LIBSVM: a library for support vector machines*. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm. 468

Cheng, W., Dryden, I. L., Hitchcock, D. B., and Le, H. (2014). "Analysis of proteomics data: Bayesian alignment of functions." *Electronic Journal of Statistics*, 8:1734–1741. MR3273588. doi: http://dx.doi.org/10.1214/14-EJS900C. 455

Cheng, W., Dryden, I. L., and Huang, X. (2015). "Supplementary materials: Bayesian registration of functions and curves." *Bayesian Analysis*. doi: http://dx.doi.org/10.1214/15-BA957SUPP. 455, 460, 463, 465, 467

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: Series B (Statistical Methodology)*, 72(5):673–694. MR2758241. doi: http://dx.doi.org/10.1111/j.1467-9868.2010.00752.x. 448, 458

Czogiel, I., Dryden, I. L., and Brignell, C. J. (2011). "Bayesian matching of unlabeled marked point sets using random fields, with an application to molecular alignment." *The Annals of Applied Statistics*, 5:2603–2629. MR2907128. doi: http://dx.doi.org/10.1214/11-AOAS486. 457

Dryden, I. L. (2014). *shapes: Statistical shape analysis*. R package version 1.1-10. 467, 468

Dryden, I. L. and Mardia, K. V. (1991). "General shape distributions in a plane." *Advances in Applied Probability*, 23:259–276. MR1104079. doi: http://dx.doi.org/10.2307/1427747. 452

Dryden, I. L. and Mardia, K. V. (1992). "Size and shape analysis of landmark data." *Biometrika*, 79:57–68. MR1158517. doi: http://dx.doi.org/10.1093/biomet/79.1.57. 452

Dryden, I. L. and Mardia, K. V. (1998). *Statistical Shape Analysis*. Wiley, Chichester. MR1646114. 450, 454, 466

Fréchet, M. (1948). "Les éléments aléatoires de nature quelconque dans un espace distancié." *Annales de l'Institut Henri Poincaré*, 10:215–310. MR0027464. 452

Gasser, T., Müller, H. G., Köhler, W., Prader, A., Largo, R., and Molinari, L. (1985). "An analysis of the mid-growth and adolescent spurts of height based on acceleration." *Annals of Human Biology*, 12:129–148. doi: http://dx.doi.org/10.1080/03014468500007631. 465

Geyer, C. J. and Thompson, E. A. (1995). "Annealing Markov chain Monte Carlo with applications to ancestral inference." *Journal of the American Statistical Association*, 90(431):909–920. doi: http://dx.doi.org/10.1080/01621459.1995.10476590. 460

Glover, G. (1999). "Deconvolution of impulse response in event-related BOLD fMRI". *NeuroImage*, 9(4):416–429. doi: http://dx.doi.org/10.1006/nimg.1998.0419. 462

Goodall, C. R. (1991). "Procrustes methods in the statistical analysis of shape (with discussion)." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 53:285–339. MR1108330. 452, 463, 467

Griffin, J. E. (2014). "An adaptive truncation method for inference in Bayesian nonparametric models." *Statistics and Computing*, to appear. doi: http://dx.doi.org/10.1007/s11222-014-9519-4. 456

Huckemann, S., Hotz, T., and Munk, A. (2010). "Intrinsic shape analysis: geodesic PCA for Riemannian manifolds modulo isometric Lie group actions." *Statistica Sinica*, 20(1):1–58. MR2640651. 452

Ishwaran, H. and James, L. F. (2001). "Gibbs sampling methods for stick-breaking priors." *Journal of the American Statistical Association*, 96(453):161–173. MR1952729. doi: http://dx.doi.org/10.1198/016214501750332758. 456

James, G. M. (2007). "Curve alignment by moments." *The Annals of Applied Statistics*, 1(2):480–501. MR2415744. doi: http://dx.doi.org/10.1214/07-AOAS127. 465

Jermyn, I. H., Kurtek, S., Klassen, E., and Srivastava, A. (2012). "Elastic shape matching of parameterized surfaces using square root normal fields." In: Fitzgibbon, A. W., Lazebnik, S., Perona, P., Sato, Y., and Schmid, C., editors, *Computer Vision – ECCV 2012 – 12th European Conference on Computer Vision, Florence, Italy, October 7–13, 2012, Proceedings, Part V*, volume 7576 of *Lecture Notes in Computer Science*, pages 804–817. Springer. 470

Joshi, S., Klassen, E., Srivastava, A., and Jermyn, I. (2007). "A novel representation for Riemannian analysis of elastic curves in $\mathbb{R}^n$." In: *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*, pages 1–7. 454

Kalli, M., Griffin, J., and Walker, S. (2011). "Slice sampling mixture models." *Statistics and Computing*, 21(1):93–105. MR2746606. doi: http://dx.doi.org/10.1007/s11222-009-9150-y. 456

Karcher, H. (1977). "Riemannian center of mass and mollifier smoothing." *Communications on Pure and Applied Mathematics*, 30(5):509–541. MR0442975. doi: http://dx.doi.org/10.1002/cpa.3160300502. 452

Kendall, D. G. (1984). "Shape manifolds, Procrustean metrics and complex projective spaces." *Bulletin of the London Mathematical Society*, 16:81–121. MR0737237. doi: http://dx.doi.org/10.1112/blms/16.2.81. 450

Kendall, D. G., Barden, D., Carne, T. K., and Le, H. (1999). *Shape and Shape Theory*. Wiley, Chichester. MR1891212. doi: http://dx.doi.org/10.1002/9780470317006. 450

Kendall, W. S. (1990). "The diffusion of Euclidean shape." In: Grimmett, G. R. and Welch, D. J. A., editors, *Disorder in Physical Systems*, pages 203–217, Oxford. Oxford University Press. MR1064562. 459

Kenobi, K. and Dryden, I. L. (2012). "Bayesian matching of unlabeled point sets using Procrustes and configuration models." *Bayesian Analysis*, 7(3):547–565. MR2981627. doi: http://dx.doi.org/10.1214/12-BA718. 460

Klassen, E., Srivastava, A., Mio, W., and Joshi, S. H. (2003). "Analysis of planar shapes using geodesic paths on shape spaces." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(3):372–383. doi: http://dx.doi.org/10.1109/TPAMI.2004.1262333. 449

Kneip, A. and Gasser, T. (1992). "Statistical tools to analyze data representing a sample of curves." *The Annals of Statistics*, 20(3):1266–1305. MR1186250. doi: http://dx.doi.org/10.1214/aos/1176348769. 448

Kneip, A., Li, X., MacGibbon, K. B., and Ramsay, J. O. (2000). "Curve registration by local regression." *Canadian Journal of Statistics*, 28(1):19–29. MR1789833. doi: http://dx.doi.org/10.2307/3315879. 448

Koch, I., Hoffmann, P., and Marron, J. S. (2014). "Proteomics profiles from mass spectrometry." *Electronic Journal of Statistics*, 8(2):1703–1713. MR3273585. doi: http://dx.doi.org/10.1214/14-EJS900. 458, 465

Le, H.-L. (1991). "A stochastic calculus approach to the shape distribution induced by a complex normal model." *Mathematical Proceedings of the Cambridge Philosophical Society*, 109:221–228. MR1075133. doi: http://dx.doi.org/10.1017/S0305004100069681. 459

Le, H.-L. and Kendall, D. G. (1993). "The Riemannian structure of Euclidean shape spaces: a novel environment for statistics." *The Annals of Statistics*, 21:1225–1271. MR1241266. doi: http://dx.doi.org/10.1214/aos/1176349259. 451

Mardia, K. V. and Dryden, I. L. (1989). "The statistical analysis of shape data." *Biometrika*, 76:271–282. MR1016017. doi: http://dx.doi.org/10.1093/biomet/76.2.271. 452

Molinari, L., Largo, R. H., and A., P. (1980). "Analysis of the growth spurt at age seven (mid-growth spurt)." *Helvetica Paediatrica Acta*, 35:325–334. 465

Ramsay, J. (2013). "Functional data analysis software." Technical report, McGill University. http://www.psych.mcgill.ca/misc/fda/software.html. 465

Ramsay, J. O. and Li, X. (1998). "Curve registration." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 60(2):351–363. MR1616045. doi: http://dx.doi.org/10.1111/1467-9868.00129. 448

Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis, Second Edition*. Springer, New York. MR2168993. doi: http://dx.doi.org/10.1002/0470013192.bsa239. 448, 449, 463, 465

Ramsay, J. O., Wickham, H., Graves, S., and Hooker, G. (2013). *fda: Functional Data Analysis*. R package version 2.4.0. 465

Silverman, B. W. (1995). "Incorporating parametric effects into functional principal components analysis." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 57(4):673–689. MR1354074. 448

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and van der Linde, A. (2002). "Bayesian measures of model complexity and fit." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4):583–639. MR1979380. doi: http://dx.doi.org/10.1111/1467-9868.00353. 456

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. doi: http://dx.doi.org/10.1109/TPAMI.2010.184. 453, 454, 463

Srivastava, A., Wu, W., Kurtek, S., Klassen, E., and Marron, J. S. (2011b). "Registration of functional data using the Fisher–Rao metric." Technical report, Florida State University. arXiv:1103.3817v2 [math.ST]. 448, 454, 465

Stein, M. L. (1999). *Interpolation of Spatial Data: Some Theory for Kriging*. Springer, New York. MR1697409. doi: http://dx.doi.org/10.1007/978-1-4612-1494-6. 455

Tang, R. and Müller, H.-G. (2008). "Pairwise curve synchronization for functional data." *Biometrika*, 95(4):875–889. MR2461217. doi: http://dx.doi.org/10.1093/biomet/asn047. 465

Telesca, D. and Inoue, L. Y. T. (2008). "Bayesian hierarchical curve registration." *Journal of the American Statistical Association*, 103(481):328–339. MR2420237. doi: http://dx.doi.org/10.1198/016214507000001139. 448, 457

Thakoor, N., Gao, J., and Jung, S. (2007). "Hidden Markov model-based weighted likelihood discriminant for 2-d shape classification." *IEEE Transactions on Image Processing*, 16(11):2707–2719. MR2472413. doi: http://dx.doi.org/10.1109/TIP.2007.908076. 469

Tucker, J. D. (2014). *fda: Functional Data Analysis*. R package version 1.4.2. 465

Tuddenham, R. D. and Snyder, M. M. (1954). "Physical growth of California boys and girls from birth to age 18." *University of California Publications in Child Development*, 1:183–364. 463

van der Vaart, A. W. (1998). *Asymptotic Statistics*, volume 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge. MR1652247. doi: http://dx.doi.org/10.1017/CBO9780511802256. 458

Walker, S. G. (2007). "Sampling the Dirichlet mixture model with slices." *Communications in Statistics – Simulation and Computation*, 36(1):45–54. MR2370888. doi: http://dx.doi.org/10.1080/03610910601096262. 456

Younes, L. (1998). "Computable elastic distances between shapes." *SIAM Journal on Applied Mathematics*, 58(2):565–586 (electronic). MR1617630. doi: http://dx.doi.org/10.1137/S0036139995287685. 454

Zhou, R. R., Serban, N., Gebraeel, N., and Müller, H.-G. (2014). "A functional time warping approach to modeling and monitoring truncated degradation signals." *Technometrics*, 56(1):67–77. MR3176573. doi: http://dx.doi.org/10.1080/00401706.2013.805661. 448, 458

Acknowledgments

We thank David Hitchcock and Huiling Le for their comments, and acknowledge the support of a Royal Society Wolfson Research Merit Award, EPSRC grant EP/K022547/1 and the Mathematical Biosciences Institute.