

Distribution on Warp Maps for Alignment of Open and Closed Curves

Karthik Bharath¹ and Sebastian Kurtek²

¹School of Mathematical Sciences, University of Nottingham, UK

²Department of Statistics, The Ohio State University, USA

Abstract

Alignment of curve data is an integral part of their statistical analysis, and can be achieved using model- or optimization-based approaches. The parameter space is usually the set of monotone, continuous warp maps of a domain. Infinite-dimensional nature of the parameter space encourages sampling based approaches, which require a distribution on the set of warp maps. Moreover, the distribution should also enable sampling in the presence of important landmark information on the curves which constrain the warp maps. For alignment of closed and open curves in $\mathbb{R}^d, d = 1, 2, 3$, possibly with landmark information, we provide a constructive, point-process based definition of a distribution on the set of warp maps of $[0, 1]$ and the unit circle \mathbb{S} that is (1) simple to sample from, and (2) possesses the desiderata for decomposition of the alignment problem with landmark constraints into multiple unconstrained ones. For warp maps on $[0, 1]$, the distribution is related to the Dirichlet process. We demonstrate its utility by using it as a prior distribution on warp maps in a Bayesian model for alignment of two univariate curves, and as a proposal distribution in a stochastic algorithm that optimizes a suitable alignment functional for higher-dimensional curves. Several examples from simulated and real datasets are provided.

Keywords: Stochastic curve registration; Functional data; Point processes; Simulated Annealing.

1 Introduction

In contrast to standard multivariate analysis, the concept of phase variation is a unique feature of functional data. For functional data obtained as parametric curves representing geometric objects in high-resolution images, establishing correspondence between points on the curves is an important task. Failure to isolate and quantify variation due to phase or lack of correspondence between points can be detrimental when computing descriptive summaries, and for subsequent inferential tasks. This process of isolation and quantification is referred to as registration or alignment.

There exist several approaches to alignment. One popular approach is continuous monotone alignment, which refers to the alignment of two curves $g_i : D \rightarrow \mathbb{R}^d$, $d \geq 1$, $i = 1, 2$, where D is a compact domain, by estimating a homeomorphic self map $\gamma : D \rightarrow D$, known as a warp map, that best matches $g_2 \circ \gamma$ to g_1 (or vice versa). A variational formulation quantifies the matching through a cost or energy functional, and alignment is defined as the determination of an optimal warp map γ^* from a class W that minimizes the cost functional. Alternatively, a statistical model-based formulation of the alignment task seeks to estimate the warp map γ based on a likelihood function defined using some discrepancy measure between g_1 and $g_2 \circ \gamma$, conditional on γ .

1.1 Motivation and related work

The focus of this work is on the construction of a probability distribution on W and development of a simple sampling scheme, which would enable quantification of uncertainty on the optimal alignment of open and closed curves using *any* of the available continuous monotone methods. The need for a distribution enabling stochastic approaches to alignment arises in two contexts: (1) since the parameter space of warp maps is a non-linear function space, deterministic algorithms for solving the variational problem can get stuck in local minima, and (2) in a Bayesian model a prior distribution on W is required. The construction and the desirable features of the distribution are motivated by the following common modeling scenarios that arise in the analysis of biomedical and biological datasets.

Boundaries of data objects such as organs in high-resolution medical images are often represented as parametrized open curves (connected curves that begin and end at different

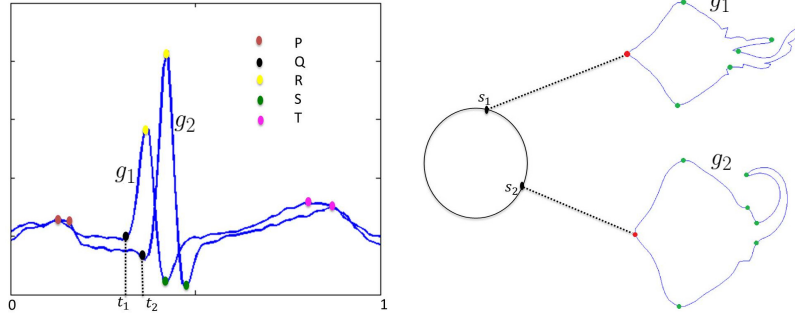


Figure 1: Left: Alignment of two ECG cycles with corresponding PQRST complex landmarks. Right: Alignment of the shapes of two stingrays, represented as simple closed curves with landmark features (red and green points).

points) or simple closed curves (connected curves that do not cross themselves, and begin and end at the same point). In certain cases, explicit physiological information can constrain the warp maps. In Figure 1, the left panel shows two open curves (univariate functions) $g_i : [0, 1] \rightarrow \mathbb{R}$, $i = 1, 2$ representing two electrocardiogram (ECG) cycles with corresponding PQRST complexes marked as fixed landmarks (maxima and minima marked on each cycle). For example, the point Q on g_1 at t_1 is to be registered to the corresponding Q point on g_2 attained at t_2 (so do the feature points P, R, S and T). This introduces constraints on any warp map $\gamma : [0, 1] \rightarrow [0, 1]$: $\gamma(t_1) = t_2$ to ensure the matching of the Q point across the two functions (additional constraints are used to ensure the matching of the P, R, S and T points as well); between the landmarks, γ is unconstrained. The example in the right panel shows the outlines of two stingrays, represented as embeddings of the circle of unit circumference \mathbb{S} in \mathbb{R}^2 , or equivalently as planar closed curves $g_i : \mathbb{S} \rightarrow \mathbb{R}^2$, $i = 1, 2$. The snout of the stingrays (marked in red) as well as other landmark points (marked in green) are to be matched. This again imposes constraints on a warp map $\gamma : \mathbb{S} \rightarrow \mathbb{S}$ requiring exact matching of landmarks with unconstrained matching in-between.

There are two complementary requirements of the alignment procedure in the above scenarios: (1) uncertainty around the ‘best’ alignment needs to be captured, and (2) constraints due to landmark points need to be automatically incorporated into the alignment procedure. Operationally, this demands a probability distribution on W that satisfies the following.

- (i) An alignment task with m constraints can be decomposed into $m+1$ unconstrained ones by performing *independent* alignment of the curves across $m+1$ subsets of the domain

D. This requires the initial constrained alignment problem to satisfy the desirable *subset invariance* condition: if $\gamma^* : D \rightarrow D$ is the optimal warp map and if $\gamma_B^* : B \rightarrow B$, $B \subset D$ is the optimal map when alignment is performed only on a subset B induced by landmarks, then $\gamma^* = \gamma_B^*$ on $B \subset D$ (Trouvé and Younes, 2000). In the presence of landmark-induced constraints, a probability distribution π on W is said to satisfy subset invariance if its restriction π_B to warp maps of $B \subset D$ is a suitably re-scaled version of π , and is independent of $\pi_{D \setminus B}$.

- (ii) For efficient exploration of W , the distribution should be flexible enough to be centred at any warp map of choice. For example, in the Figure 1 examples, as a first step, one can align the two curves *only* at the constraint points with a piecewise linear (PL) warp map. Alignment of the remaining regions can be carried out by sampling in the neighborhood of the PL warp map within a stochastic algorithm, or by employing a prior distribution centred at the PL warp map in a Bayesian model.

Literature on alignment has mostly focused on functional data defined on a closed interval $[a, b]$: for frequentist approaches see Kneip and Gasser (1992); Gervini and Gasser (2004); Gasser and Kneip (1995); Zhou et al. (2014) and Tang and Müller (2008); for Bayesian approaches see Telesca and Inoue (2008); Cheng et al. (2016); Claeskens et al. (2010); Lu et al. (2017) and Kurtek (2017). Alignment methods for closed curves with $D = \mathbb{S}$ are conspicuous by their absence within statistics literature; a notable exception is the differential-geometric approach of Srivastava et al. (2011). For a good account of variational strategies for alignment of closed and open curves, we refer the reader to Srivastava and Klassen (2016).

Landmark-constrained alignment, under a geometric, non-stochastic framework, was only recently studied (Strait et al., 2017; Bauer et al., 2017). In those methods, it is not possible to capture or model uncertainty in the optimal alignment. As far as we know, stochastic approaches to curve alignment in the presence of landmarks have not been considered before.

1.2 Contributions

In a Bayesian model for alignment, Cheng et al. (2016) define a finite-dimensional Dirichlet distribution on discretized warp maps as a prior distribution; this results in a finite-dimensional specification of the prior instead of a functional one. We first prove that their

construction results in a limiting degenerate distribution on W as dimensionality increases (Theorem 1). We further establish degeneracy of the limiting distribution for a class \mathcal{C} of finite-dimensional distributions that generalize the Dirichlet (Corollary 1).

Employing point process methods, we propose a modification of their finite-dimensional specification that results in a constructive definition of a non-degenerate distribution on the set of warp maps of $[0, 1]$ and \mathbb{S} that satisfies requirements (i) and (ii) (Theorem 2). We show that the distribution is a canonical one in the sense that it remains the limiting distribution for all finite-dimensional specifications from a class \mathcal{C} . The distribution is related to the Dirichlet process (Ferguson, 1973) on the set of probability measures on $[0, 1]$ (Remark 2). Our approach provides an explicit link between the self-similarity- and Markov-type properties of the Dirichlet process and requirements (i) and (ii) (Proposition 1). The distribution can be parametrized by a concentration parameter $\theta > 0$ whose value determines how close the probability mass is distributed around the chosen average warp map. The warp maps sampled from the distribution are discontinuous with probability one. *We show that this is unavoidable if one insists on subset invariance in requirement (i).* The constructive definition, along with the concentration parameter θ , identifies the finite-dimensional projections (coordinates) of the distribution with a Dirichlet distributed random vector, with parameters that depend intimately on the discretization and the choice of the average warp map. This leads to a simple algorithm for sampling PL warp maps.

Finally, we propose a novel stochastic algorithm based on the proposed distribution for alignment of open and closed curves in $\mathbb{R}^k, k = 1, 2, 3$, possibly with landmark constraints. We elucidate on the importance of requirements (i) and (ii) on the distribution through several simulation and real-data examples. In addition, we employ the proposed distribution in a Bayesian model for alignment, similar to the one used by Cheng et al. (2016).

The rest of this paper is organized as follows. Section 2 examines the algorithm of Cheng et al. (2016) and details its shortcomings. Section 3 modifies the preceding construction, and proposes a theoretically-justified approach for warp maps of $[0, 1]$; the properties of the constructed distribution are studied in Section 4. Section 5 extends the construction mechanism to a corresponding distribution on warp maps of \mathbb{S} . Section 6 presents sample warp maps under different settings, and results from different analyses of real open and closed curve data, possibly with landmark constraints, under a Bayesian model and using a novel

stochastic algorithm (Section 6.2.2). Section 7 discusses extensions of the proposed methods. The Supplementary Material contains proofs of all results, an alternative construction of a distribution on warp maps of \mathbb{S} , and detailed descriptions of the datasets used in this work.

2 Construction using fixed partitions and issues

We first consider the construction of a distribution on the set of warp maps of a closed subinterval of the real line, which, without loss of generality, can be assumed to be $[0, 1]$. The possibility of landmarks on the observed curve data implies that the class of smooth warp maps is inappropriate. Instead, consider the class given by $W_I := \{\gamma : [0, 1] \rightarrow [0, 1], \text{ increasing, continuous, } \gamma(0) = 0, \gamma(1) = 1\}$. In a recent paper on Bayesian alignment of curves, Cheng et al. (2016) proposed a simple method to obtain a continuous random warp map in W_I . The following is a summary of their algorithm.

Algorithm 1. Fixed partition-based sampling of warp maps.

1. Choose a deterministic set of ordered points $0 =: t_0 < t_1 < t_2 < \dots < t_{n-1} < t_n := 1$ that induces a partition \mathcal{T}_n of $[0, 1]$.
2. Sample an n -dimensional Dirichlet distributed random vector with all parameters set to the same value $\alpha > 0$.
3. Construct a warp map on $[0, 1]$ by linear interpolation of the increments.

The resulting warp maps are continuous and are elements of W_I . The idea behind this approach is based on the fact that for $\gamma \in W_I$, its increments $p_i := \gamma(t_i) - \gamma(t_{i-1})$ are positive and satisfy $\sum_{i=1}^n p_i = 1$. Hence, p_i can be identified with coordinates of the simplex $\Delta_{n-1} := \{x = (x_1, \dots, x_n) \in \mathbb{R}^n : x_i \geq 0, \sum_{i=1}^n x_i = 1\}$. The parameter space of warp maps generated in this fashion is essentially finite-dimensional, since a warp map is fully determined by its values at the ordered set of points t_i . Cheng et al. (2016) state (without proof) that as $n \rightarrow \infty$, this results in a Dirichlet process.

For a fixed partition \mathcal{T}_n , Algorithm 1 recommends simulating a Dirichlet random vector with all n parameters equal to $\alpha > 0$, or equivalently $\alpha(1, 1, \dots, 1)$. The parameter vector is hence independent of the size n of partition. The natural way to incorporate

information from the partition \mathcal{T}_n is to assume that the parameter vector is of the form $\alpha^*(\frac{1}{n}, \dots, \frac{1}{n})$ where $\alpha^* = n\alpha$. This can be generalized to an arbitrary deterministic partition $\mathcal{T} = \{B_1, \dots, B_k, k \geq 1\}$ of $[0, 1]$ by simulating a $|\mathcal{T}|$ -dimensional Dirichlet random vector with parameters $\alpha(\mu(B_1), \dots, \mu(B_k))$, where $\alpha > 0$, μ is a finite measure on $[0, 1]$, and $|T|$ denotes the cardinality of the set T . An equi-spaced or uniform partition arises by setting $\mathcal{T} = \mathcal{T}_n$ with $t_i - t_{i-1} = 1/n, i = 1, \dots, n$, and taking μ as the Lebesgue measure.

From a practical perspective, Algorithm 1 is appealing since the partition \mathcal{T}_n can be chosen in various ways; the increments can be sampled from distributions different from the Dirichlet on Δ_{n-1} , and linear interpolation results in continuous maps. The two pertinent questions are: (1) What is the corresponding distribution on W_I as $n \rightarrow \infty$? and (2) can the sampled Dirichlet random vector be identified with the finite-dimensional distributions of a stochastic process? The lack of partition information in the parameter vector has a significant implication on the answers to questions (1) and (2).

The algorithm constructs a warp map by linear interpolation of the increments obtained from a fixed partition \mathcal{T}_n . The natural setting for its examination is the space $C([0, 1])$ of real-valued, continuous functions on $[0, 1]$, with the linearly interpolated process based on the partial sum of the increments:

$$Y_n(t) := \sum_{i=1}^{\lfloor nt \rfloor} p_i + (nt - \lfloor nt \rfloor)p_{\lfloor nt \rfloor + 1}, \quad t \in [0, 1]. \quad (1)$$

Clearly, $Y_n(0) = 0$, $Y_n(1) = 1$, and Y_n is continuous, increasing in $(0, 1)$ and an element of W_I . The case when $\alpha = 1$ and (p_1, \dots, p_n) is uniform on Δ_{n-1} (for a given partition of \mathcal{T}_n) is particularly instructive as it captures the key features of the algorithm.

Theorem 1. *Suppose (p_1, \dots, p_n) based on a fixed equi-spaced partition \mathcal{T}_n is uniformly distributed on Δ_{n-1} , and $\gamma_{id} : [0, 1] \rightarrow [0, 1]$ with $\gamma_{id}(t) = t$. In $C([0, 1])$, equipped with the uniform topology, Y_n converges in probability to the identity warp map γ_{id} . The process $\sqrt{n}(Y_n(t) - \gamma_{id}(t))$ converges in distribution to a standard Brownian Bridge process.*

Theorem 1 states that for large n , when $\alpha = 1$, the sampling algorithm proposed by Cheng et al. (2016) results in a distribution on W_I that is *degenerate at the identity warp map*. The conclusion also remains true with γ_{id} replaced by another deterministic warp map for a fixed

non-equi-spaced partition \mathcal{T}_n (see Supplementary Material). Moreover, the fluctuations away from the identity warp can be captured by the behavior of a standard Brownian Bridge. This result suggests that the resulting distribution on the class W_I , in the limit, is governed only by the value of the Dirichlet scalar concentration parameter α , and concentrates on the identity warp map. See Section 6.1 for numerical illustrations of such degenerate behavior.

In fact, such an uninteresting distribution on W_I resulting from Algorithm 1 is not restricted to the case where the increments are uniformly distributed on Δ_{n-1} ; *this phenomenon is applicable to a rather large class of distributions on Δ_{n-1} based on spacings of random variables*. Suppose x_1, \dots, x_n are independent from an absolutely continuous distribution function F on $[0, 1]$, with density f and quantile function Q on $(0, 1)$. Extend the definition of Q to $[0, 1]$ by setting $0 =: Q(0) = \lim_{t \rightarrow 0} Q(t)$ and $1 =: Q(1) = \lim_{t \rightarrow 1} Q(t)$. Set $x_{0:n} := 0$, $x_{n:n} := 1$, let $0 < x_{1:n} < x_{2:n} < \dots < x_{n-1:n} < 1$ a.s. denote the corresponding order statistics, and let $p_i = x_{i:n} - x_{i-1:n}$ for $i = 1, \dots, n$ be the spacings. Since $\sum_{i=1}^n p_i = 1$, (p_1, \dots, p_n) is a random vector on Δ_{n-1} . When f is the uniform density on $[0, 1]$, (p_1, \dots, p_n) is a Dirichlet distributed random vector with each parameter equalling one. This class of distributions is hence a natural extension of the one used in Algorithm 1.

Corollary 1. *Suppose $(p_1, \dots, p_n) \in \Delta_{n-1}$ based on any fixed equi-spaced partition \mathcal{T}_n is the vector of spacings of i.i.d. random variables with a twice differentiable distribution function F and quantile function Q . If f is the corresponding probability density, assume that $\inf_{0 \leq x \leq 1} f(Q(x)) > 0$ and $\sup_{0 \leq x \leq 1} |f'(Q(x))| < \infty$. Then, Y_n converges in probability to Q in $C([0, 1])$ equipped with the uniform topology.*

The conditions on F in Corollary 1 are not too restrictive and are satisfied by several densities with support on $[0, 1]$. For example, one can easily check that if F is the distribution function of a non-central Beta (Hodges, 1955) with both shape parameters equal to one, and a non-centrality parameter $\kappa > 0$, then F is twice differentiable with bounded second derivative, and $\inf_{0 \leq x \leq 1} f(x)$ is $e^{-\kappa/2}(\kappa/2 + 1) > 0$, which is attained in the limit at zero. As with Theorem 1 the conclusion of Corollary 1 holds for a non-equi-spaced partition as well (See Supplementary Material).

3 Improved construction via random partitions

While the simplicity of Algorithm 1 is attractive, the degeneracy of the resulting distribution on W_I (as $n \rightarrow \infty$) is disconcerting. In this section, we offer a simple modification of the previous approach that salvages the situation. The sampling method in Algorithm 1 depends on the choice of the partition \mathcal{T}_n of $[0, 1]$. We demonstrate that choosing a random partition $\mathcal{T}_n(H)$ based on order statistics of an i.i.d. sample from distribution H on $[0, 1]$, in conjunction with a point process representation, results in a limit process with sample paths that lie in W_I centred at a desired warp map. The motivation for using partitions induced by order statistics is the fact that conditional on n , $t_{i:n}$ have the same distributions as the order statistics of an i.i.d. sample from $h(t)/\int_0^1 h(u)du$, where h is the intensity of a non-homogeneous Poisson process on $[0, 1]$. Next, we provide a new algorithm for sampling warp maps in W_I , and study the theoretical properties of the associated distribution.

Algorithm 2. Random partition-based sampling of warp maps on $[0, 1]$.

1. Choose order statistics $0 =: t_{0:n} < t_{1:n} < t_{2:n} < \dots < t_{n-1:n} < t_{n:n} := 1$ of a random sample from distribution H on $[0, 1]$.
2. Sample an n -dimensional Dirichlet distributed random vector (p_1, \dots, p_n) with parameters set to $(t_{1:n} - t_{0:n}, \dots, t_{n:n} - t_{n-1:n})$.
3. Construct a warp map on $[0, 1]$ by linear interpolation of the increments.

Algorithm 2 is easy to implement, and extends Algorithm 1 to one based on random partitions (and a subsequent change in Dirichlet parameters).

3.1 Theoretical support for Algorithm 2

As $n \rightarrow \infty$, the limiting distribution associated with Algorithm 2 can be identified in two ways. The first approach is to start with finite-dimensional distributions at chosen time points and posit the existence of a process with the chosen finite-dimensional projections based on Kolomogorov's consistency theorem (see [von Renesse and Sturm \(2009\)](#)). *With this approach, it is then difficult to centre the distribution at a desired warp map.* The approach

we adopt in this paper is to constructively define a distribution that arises as a limit (as $n \rightarrow \infty$) based on a point process formulation using transformed increments.

We first review the Gamma subordinator process. A process $\mathcal{G}(t)$, $t \in [0, 1]$ is a Gamma subordinator taking values in \mathbb{R}_+ if, for $0 \leq s < t \leq 1$, $\mathcal{G}(t) - \mathcal{G}(s)$ is Gamma distributed with shape parameter $t - s$ and scale parameter equal to one. It is a Levy process with Levy measure $\lambda(dy) = y^{-1}e^{-y}dy$ and sample paths that are discrete with probability one, and thereby allows for a point process representation: $\mathcal{G}(t) := \sum_{\nu_x \leq t} \nu_y$, $t \in [0, 1]$, where $\nu = (\nu_x, \nu_y) \in [0, 1] \times \mathbb{R}_+$ is a Poisson point process with intensity measure $dx \times \lambda(dy)$. For a distribution function (not necessarily one corresponding to a probability measure) H on $[0, 1]$, such that $\lim_{x \rightarrow 1} H(x) = c < \infty$, consider the time-changed Gamma process $\mathcal{G}(H(t))$ whose increments $\mathcal{G}(H(t)) - \mathcal{G}(H(s))$ are Gamma distributed with shape parameter $H(t) - H(s)$ and scale parameter equal to one. Then, the normalized Gamma process $t \mapsto \mathcal{G}(H(t))/\mathcal{G}(c)$ is the Dirichlet process $\mathcal{D}(H(t))$ with base measure or parameter H , taking values in $[0, 1]$. Its sample paths are hence random functions mapping $[0, 1]$ to itself. We denote the laws of \mathcal{G} and \mathcal{D} by \mathbb{G} and \mathbb{D} , and $\mathcal{G} \circ H$ and $\mathcal{D} \circ H$ by $\mathbb{G} \circ H$ and $\mathbb{D} \circ H$, respectively.

Using the representation of a pure jump Levy process by [Ferguson and Klass \(1972\)](#), we examine the existence of a limit process with sample paths in W_I with ‘finite-dimensional’ Dirichlet distributions with appropriate parameters. The following result is formulated for the general class of distributions on Δ_{n-1} induced by spacings of i.i.d. random variables on $[0, 1]$ with density f and distribution function F , of which the Dirichlet with all parameters set to the same value is a special case. Consider a random partition $\mathcal{T}_n(H)$ of $[0, 1]$ based on order statistics $0 =: t_{0:n} < t_{1:n} < \dots < t_{n-1:n} < t_{n:n} := 1$ from an i.i.d sample $\{t_i, i = 1, \dots, n-1\}$ with absolutely continuous distribution function H on $[0, 1]$. Independent of t_i , consider $(p_1, \dots, p_n) \in \Delta_{n-1}$ obtained as spacings of an i.i.d sequence x_i from a density f chosen as described in [Corollary 1](#). Let F and Q be the corresponding distribution and quantile functions of x_i , respectively. Set $v_1 = p_1$ and $v_i = p_1 + \dots + p_i$, $i = 2, \dots, n$, and consider the transformed random variables $z_{i,n} = n f(Q(\zeta_{i,n})) v_i$ where $0 \leq \zeta_{i,n} \leq 1$ is a deterministic sequence such that $\max_{1 \leq i \leq n} |\frac{i}{n} - \zeta_{i,n}| = O(1/n)$. Define $\lambda^{-1}(x) := \inf\{t \in \mathbb{R} : \lambda(t) \geq x\}$.

Theorem 2. *Let $\mathcal{P}_n := \sum_{i=1}^n \delta_{\{t_i, \lambda^{-1}(z_{i,n})\}}$ be a sequence of point processes, where $\delta_{\{x,y\}}$ is the point measure at $(x, y) \in [0, 1] \times \mathbb{R}_+$. Assume f to be continuous and positive on $(0, 1)$ with $\lim_{x \downarrow 0} f(x) > 0$ and finite.*

- (1) Suppose p_i for each $i = 1, \dots, n$ possesses a unimodal density. The sequence \mathcal{P}_n converges in the vague topology to the Poisson point process \mathcal{P} with intensity measure $H(dt) \times \lambda(dy)$, where $\lambda(dy) = y^{-1}e^{-y}dy$.
- (2) If \mathcal{P}_n converges in the vague topology to \mathcal{P} , then the sequence of processes $G_n(t) := \sum_i \lambda^{-1}(z_{i,n}) \mathbb{I}_{t_i \leq t}$, $t \in [0, 1]$ converges weakly to the time-changed Gamma process $\mathcal{G} \circ H$ in the Skorohod J_1 topology. However, the linearly interpolated version of G_n converges to $\mathcal{G} \circ H$ in the Skorohod M_1 topology.

Evidently, the limit process \mathcal{P} is the Gamma process $\mathcal{G} \circ H$, and \mathcal{P}_n , when normalized, converges weakly to a limit, which we refer to as the *Dirichlet process* $\mathcal{D} \circ H$ with dH as the base measure, where dH is the Lebesgue-Stieltjes measure corresponding to the distribution function H . The probability measure $\mathbb{D} \circ H$ on the class W_I is constructed *only* using the increments $p_i = \gamma(t_{i:n}) - \gamma(t_{i-1:n})$ of warp maps on a random partition $\mathcal{T}_n(H)$ based on t_i i.i.d. from H . Thus, Theorem 2 ensures that the finite-dimensional Dirichlet distributions can be identified with the finite-dimensional distributions of $\mathcal{D} \circ H$, and form a consistent family in the Kolmogorov sense (von Renesse and Sturm, 2009). It can be interpreted as follows: conditioned on a partition $\mathcal{T}_n(H) = \{0 =: t_{0:n} < t_{1:n} < \dots < t_{n-1:n} < t_{n:n} := 1\}$,

$$\begin{aligned} \mathbb{D} \circ H(\gamma(t_{1:n}) \in dx_1, \dots, \gamma(t_{n-1:n}) \in dx_{n-1}) &= \\ &= \frac{\Gamma(1)}{\prod_{i=1}^n \Gamma(t_{i:n} - t_{i-1:n})} \prod_{i=1}^n (x_i - x_{i-1})^{(t_{i:n} - t_{i-1:n})} dx_1 \dots dx_{n-1}, \end{aligned} \quad (2)$$

where $t_{n:n} = x_n = 1$ and $t_{0:n} = x_0 = 0$. A few remarks are in order at this stage.

Remark 1. The Ferguson-Klass representation is based on the transformation of a homogeneous Poisson random measure on $\mathbb{R}_+ \times \mathbb{R}_+$ under the inverse of the tail of a Levy measure g on \mathbb{R}_+ . Theorem 2 uses this representation with Levy measure λ to (1) retain the simplicity of Algorithm 1, (2) ensure that the distribution on W_I satisfies subset invariance, and (3) ensure that the distribution on W_I can be centred at a desired warp map. Proposition 1 in the next section clarifies (2) in view of Lukac's characterization of the Gamma distribution; in other words, the subset invariance requirement fixes the Levy measure to be the Gamma tail measure $\lambda(dx) = x^{-1}e^{-x}dx$. For a *fixed* partition, the measure $\mathbb{D} \circ H$ with $H(t) = t$ was constructed by von Renesse and Sturm (2009) (Proposition 3.4) using Kolmogorov's exten-

sion theorem; as mentioned earlier, such an approach cannot be used to obtain a distribution that can be centred at any desired warp map.

Remark 2. Let H be the uniform distribution function. The definition of the probability measure $\mathbb{D} \circ H = \mathbb{D}$ on W_I does not identify the increments of a γ with the set of probability measures on a finite set. This is in contrast to Ferguson's Dirichlet process, say $\bar{\mathcal{D}}$, which is constructed on the set $\mathbb{P}([0, 1])$ of probability measures on $[0, 1]$, topologized by weak convergence and indexed by Borel sets of $\mathbb{P}([0, 1])$. Suppose W_I is identified with distribution functions on $[0, 1]$. Consider the homeomorphism $h : W_I \rightarrow \mathbb{P}([0, 1])$ that assigns to each $\gamma \in W_I$ its Stieltjes measure $d\gamma$. Equip W_I now with the image of the weak topology on $\mathbb{P}([0, 1])$ under the map $h^{-1} : \mathbb{P}([0, 1]) \rightarrow W_I$. Then, the law of $\bar{\mathcal{D}}$ can be viewed as the push-forward $h_*\mathbb{D}$ of \mathbb{D} . On the other hand, if $\bar{h} : \mathbb{P}([0, 1]) \rightarrow W_I$ such that \bar{h} assigns to each $\mu \in \mathbb{P}([0, 1])$ a function $\gamma(t) := \sup\{u \in [0, 1] : \mu[0, u] \leq t\}$, then W_I is identified with quantile functions on $[0, 1]$. Interestingly, the pullback of \mathbb{D} on W_I is topologically very different from the law of Ferguson's process (see p.1131 of [von Renesse and Sturm \(2009\)](#)). Importantly, when considering warp maps of \mathbb{S} , the relationships to distribution or quantile functions of measures on \mathbb{S} are unavailable.

Remark 3. Part (2) of Theorem 2 is striking: linear interpolation of the sample paths of G_n does not affect the limit process. This has an important implication for Algorithm 2 using random partitions, while not giving up continuity of obtained warp maps. The weaker Skorohod's M_1 topology is used since the linear interpolation of G_n implies that we seek convergence to a limit jump process with unmatched jumps in the converging sequence of processes; this cannot be achieved under the usual J_1 topology.

Remark 4. Part (1) of Theorem 2 states that the limit process is unchanged if the vector of increments of the warp map γ is assumed to have a distribution on the simplex Δ_{n-1} obtained as the spacings of i.i.d. random variables from an arbitrary density on $[0, 1]$. This, in a certain sense, makes the Dirichlet process a natural choice as a probability measure on W_I . The assumption of unimodality of the densities of p_i is not critical and can be relaxed at some technical cost.

4 Properties of the proposed distribution

Next, we study the theoretical properties of the constructed distribution $\mathbb{D} \circ H$ on warp maps of $[0, 1]$. We also consider the case when landmark constraints need to be enforced during the registration process.

4.1 Automatic regularization

The limit Gamma process $\mathcal{G} \circ H$ (and hence $\mathcal{D} \circ H$) in Theorem 2 is a pure jump process with a.s. discrete paths, which leads to a pure jump warp map γ . Nonetheless, for a fixed $u \in [0, 1]$, the function $u \mapsto \gamma(u)$ is continuous at u $\mathbb{D} \circ H$ -almost surely since $\mathcal{G} \circ H$ is a Levy process, and hence stochastically continuous: $|\mathcal{G}(H(t+u)) - \mathcal{G}(H(u))| \xrightarrow{P} 0$ as $t \rightarrow 0$ since H is absolutely continuous. We gather from Part (2) of Theorem 2 that linear interpolation does not affect the discrete nature of the limit Gamma process. For the alignment problem, the number and frequency of large jumps are pertinent since Algorithm 2 is based on choosing a fine partition (large n); the occurrence and likelihood of warp maps that contain regions of ‘large warping’ are particularly important. Moreover, this also sheds light on how likely we are to sample warp maps from $\mathbb{D} \circ H$ that deviate considerably from their average H .

Based on a vector of increments (p_1, \dots, p_n) uniformly distributed on Δ_{n-1} , we consider the ‘large jumps’ process defined as the real-valued partial sum process $Y_n(t) := \sum_{i=1}^{\lfloor nt \rfloor} [\xi_i - E(\xi_i \mathbb{I}_{\xi_i \leq 1})]$, $0 \leq t \leq 1$ taking values in $D([0, 1])$, where $\xi_{i,n} := np_i - \log n$ is a triangular array sequence. The definition of $\xi_{i,n}$ is motivated by the fact that the largest jump amongst the p_i is of size $O_P(\log n/n)$ when (p_1, \dots, p_n) is uniform on Δ_{n-1} (Devroye, 1981). Jumps of smaller order can also be studied using intermediate spacings amongst the p_i under a different normalizing transformation (see Nagaraja et al. (2015) for details).

Theorem 3. *Let $Y(t)$, $t \in [0, 1]$ be a real-valued Levy jump process with Levy measure $\nu(dy) = e^{-y}dy$. The sequence Y_n converges weakly to Y in $D([0, 1])$ equipped with the Skorohod J_1 topology.*

Theorem 3 is a functional limit theorem for the process Y_n , and describes the probabilistic behavior of Algorithm 2 in generation of warp maps which contain regions of high warping. The centering term in Theorem 3 cannot be dispensed with, although other equivalent terms

can be chosen. The arrivals of the large increments are governed by a finite activity Levy process since $\int_0^\infty e^{-u} du < \infty$, which implies that large jumps occur infrequently; Y is hence a compound Poisson process. This ensures that under Algorithm 2, once an H that generates the random partition and sets the average warp map is chosen, *we are not likely to sample warp maps from $\mathbb{D} \circ H$ that contain far more large jumps relative to those in the average warp*. Thus, Algorithm 2 offers automatic regularization toward the average warp map.

4.2 Landmark constraints and a global concentration parameter

In practice, in the presence of landmarks, one decomposes the unconstrained registration problem into multiple sub-problems corresponding to intervals formed due to the landmark constraints. For example, in the case of $m = 2$ landmarks, suppose that the landmark locations on the domain of two open curves g_1 and g_2 are at $t_i, t_k \in [0, 1]$, and at $t_j, t_l \in [0, 1]$, respectively, with $1 \leq i < k \leq n - 1$ and $1 \leq j < l \leq n - 1$. This induces three intervals of interest on each curve: $\{[0, t_i], [t_i, t_k], [t_k, 1]\}$ for g_1 and $\{[0, t_j], [t_j, t_l], [t_l, 1]\}$ for g_2 . The points t_i and t_k can now be exactly matched to t_j and t_l using a PL warp map, resulting in two new points t_i^* and t_k^* . This leads to three classes of warp maps: $W_1 = \{\gamma : [0, t_i^*] \rightarrow [0, t_i^*]\}$, $W_2 = \{\gamma : [t_i^*, t_k^*] \rightarrow [t_i^*, t_k^*]\}$, and $W_3 = \{\gamma : [t_k^*, 1] \rightarrow [t_k^*, 1]\}$. The original registration problem on $[0, 1]$ has thus been decomposed into three similar ones: (1) match $g_1|_{[0, t_i^*]}$ and $g_2|_{[0, t_i^*]}$, (2) match $g_1|_{[t_i^*, t_k^*]}$ and $g_2|_{[t_i^*, t_k^*]}$, and (3) match $g_1|_{[t_k^*, 1]}$ and $g_2|_{[t_k^*, 1]}$, where $g|_{[a, b]}$ is the restriction of g to the subinterval $[a, b]$. Note that $g_1|_{[0, t_i^*]}(t_i^*) = g_2|_{[0, t_i^*]}(t_i^*) = g_1|_{[t_i^*, t_k^*]}(t_i^*) = g_2|_{[t_i^*, t_k^*]}(t_i^*)$, and $g_1|_{[t_i^*, t_k^*]}(t_k^*) = g_2|_{[t_i^*, t_k^*]}(t_k^*) = g_1|_{[t_k^*, 1]}(t_k^*) = g_2|_{[t_k^*, 1]}(t_k^*)$; such relationships are true for all warp maps in the classes W_1, W_2 and W_3 .

The above decomposition requires a property of warp maps called subset invariance. The corresponding requirement on a probability measure on W_I is the following. Consider the map $S([a, b]) : W_I \rightarrow W_I$,

$$S([a, b])(\gamma(u)) = \frac{\gamma((1 - u)a + ub) - \gamma(a)}{\gamma(b) - \gamma(a)}, \quad u \in [0, 1], \quad 0 \leq a < b \leq 1. \quad (3)$$

A parametrized probability measure \mathbb{P}_θ , $\theta \in \Theta$ on W_I is said to satisfy subset invariance if its push-forward under the map $S([a, b])$, $S([a, b])_\# \mathbb{P}_\theta$ is $\mathbb{P}_{\theta(b-a)}$. To facilitate subset invariance, and to ensure that the three sub-problems borrow strength, we introduce a concentration

parameter $\theta > 0$, which allows the distribution to interpolate between the indicator (step) function and the average warp map. Under the notation employed in the preceding section, consider the parametrized Levy measure $\lambda_\theta(y) = \theta \int_y^\infty e^{-t} t^{-1} dt$. The limit process \mathcal{P} in Theorem 2 is then a Gamma process with Levy measure λ_θ , such that, for any Borel set A and any $0 \leq s < t \leq 1$,

$$P(\mathcal{G}(\theta t) - \mathcal{G}(\theta s) \in A) = \int_A \frac{1}{\Gamma(\theta(t-s))} y^{\theta(t-s)-1} e^{-y} dy.$$

Then $\mathcal{D}_\theta(t) = \mathcal{G}(\theta t)/\mathcal{G}(\theta)$, $t \in [0, 1]$ is the corresponding Dirichlet process.

Proposition 1. *The distribution $\mathbb{D}_\theta \circ H$ satisfies the following properties (von Renesse and Sturm, 2009).*

- (1) Concentration around the mean: For the partition based on H , $E_\theta(\gamma_t) = H(t)$ and $\text{Var}_\theta(\gamma_t) = \frac{1}{1+\theta} H(t)(1 - H(t))$, $t \in [0, 1]$, $\theta > 0$.
- (2) Subset invariance: For the map $S([a, b])$ in Equation (3), the push-forward measure $S([a, b])_\# \mathbb{D}_\theta \circ H$ equals $\mathbb{D}_{\theta(H(b)-H(a))} \circ H$.
- (3) Markov-type property: For $\theta > 0$, the push-forward measures $S([a, b])_\# \mathbb{D}_\theta \circ H$ and $S([0, 1] \setminus [a, b])_\# \mathbb{D}_\theta \circ H$ depend on each other only at the endpoints a and b .
- (4) As $\theta \rightarrow 0$, $\mathbb{D}_\theta \circ H$ converges to a uniform distribution on the subset $\{\gamma : [0, 1] \rightarrow [0, 1] : \gamma(t) = \mathbb{I}_{[0, H(t)]}\}$ of W_I .
- (5) As $\theta \rightarrow \infty$, $\mathbb{D}_\theta \circ H$ converges to the point mass distribution δ_H at H .
- (6) Let $\bar{\gamma} : [0, 1] \rightarrow [0, 1]$ be continuous and increasing. For every $\theta > 0$, $\bar{\gamma}_*^{-1} \mathbb{D}_\theta \circ H$ is absolutely continuous with respect to $\mathbb{D}_\theta \circ H$.

The proof of (1) follows from direct computation using Equation (4). Proofs of (2) and (3) are easily obtained from the representation of the Dirichlet process as a normalized Gamma process, $\mathcal{D}_\theta(t) = \mathcal{G}(\theta t)/\mathcal{G}(\theta)$, and the fact that $\mathcal{D}_\theta(t)$ is independent of $\mathcal{G}(\theta)$ for every $t \in [0, 1]$ and $\theta > 0$. The independence and self-similarity properties are hence based on Lukac's characterization of the Gamma distribution: if X_1, \dots, X_n are independent Gamma random variables, then $Y = \sum_{i=1}^n X_i$ and the vector $(X_1/Y, \dots, X_n/Y)$ are independent. For proofs

of (3) and (4), we refer the reader to the proof of Proposition 3.14 by [von Renesse and Sturm \(2009\)](#). Properties (1), (4) and (5) ensure that the distribution on W_I is centred at H . The parameter θ behaves like a concentration parameter: varying θ moves mass away from or toward H , and offers a rich class of probability models for W_I . Property (6) is a crucial distributional property: W_I is closed under composition, and hence any distribution on W_I should be quasi-invariant with respect to composition. Its proof can be found in Theorem 4.3 of [von Renesse and Sturm \(2009\)](#).

With the introduction of $\theta > 0$, the distribution $\mathbb{D}_\theta \circ H$ has finite-dimensional projections interpreted in the following manner: conditioned on a partition $\mathcal{T}_n(H) = \{0 =: t_{0:n} < t_{1:n} < \dots < t_{n-1:n} < t_{n:n} := 1\}$,

$$\begin{aligned} \mathbb{D}_\theta \circ H(\gamma(t_{1:n}) \in dx_1, \dots, \gamma(t_{n-1:n}) \in dx_{n-1}) &= \\ &= \frac{\Gamma(\theta)}{\prod_{i=1}^n \Gamma(\theta(t_{i:n} - t_{i-1:n}))} \prod_{i=1}^n (x_i - x_{i-1})^{\theta(t_{i:n} - t_{i-1:n})} dx_1 \dots dx_{n-1}. \end{aligned} \quad (4)$$

Remark 5. Part (2) of Proposition 1 identifies $\mathbb{D}_\theta \circ H$ with the subset invariance property. *In other words, if subset invariance is a requisite property for a distribution on warp maps, with or without landmarks, it is not possible to construct one that does not concentrate on discontinuous warp maps.* This stands in contrast to the usual smoothness assumptions associated with warp maps ([Ramsay and Silverman, 2005](#); [Claeskens et al., 2010](#)), but is rarely an issue in practice. Intuitively, decomposition of alignment based on landmarks can be linked with an independent increments property of the stochastic process. Furthermore, under the popular square-root velocity transform that we employ in the applications for alignment (see Section 6.2 for details), [Lahiri et al. \(2015\)](#) proved that the optimal warp map is PL when at least one of the two curves to be aligned is PL (see Theorem 6). The lack of smoothness of the sample paths of $\mathcal{D} \circ H$ is thus not unrealistic.

5 Extension to distributions on warp maps of \mathbb{S}

Our aim in this section is to construct a probability measure on the set $W_{\mathbb{S}} := \{\gamma : \mathbb{S} \rightarrow \mathbb{S} : \text{continuous, orientation preserving}\}$ of warp maps of \mathbb{S} . From a practical perspective, we wish to develop an easy-to-implement sampling method, similar to Algorithm 2, for generating

random warp maps of \mathbb{S} based on a suitable discretization. Our approach ‘unwraps’ \mathbb{S} at a specific point c and proceeds to identify $W_{\mathbb{S}}$ with the product space $W_I \times \mathbb{S}$ through the identification of \mathbb{S} with $[0, 1]$. This amounts to using the probability measure $\mathbb{D}_{\theta} \circ H$ on W_I along with one on \mathbb{S} , based on viewing the circle of unit circumference \mathbb{S} as the quotient group $\mathbb{S} = \mathbb{R}/2\pi\mathbb{Z}$ with the addition operation inherited from \mathbb{R} . We thus move from $[0, 1]$ to the circle with unit length \mathbb{S} by identifying the endpoints of the interval.

Through the identification of \mathbb{S} with $[0, 1]$, every continuous mapping $\beta : \mathbb{R} \rightarrow \mathbb{R}$ induces a continuous mapping of \mathbb{S} onto itself such that $\beta(t+j) = \beta(t) + j$ for all $t \in \mathbb{R}$, where j is an integer (β is unique up to addition of an integer and $\beta(t) - t$ is periodic with period j). If β is monotone increasing and $j = +1$, we say that the induced map on \mathbb{S} is orientation-preserving (based on a choice of clockwise or anti-clockwise orientation). Specifically, consider the class $W_{\mathbb{R}} := \{\beta : \mathbb{R} \rightarrow \mathbb{R} : \beta(t+1) = \beta(t) + 1, \text{ continuous and non-decreasing}\}$. Each member β of $W_{\mathbb{R}}$ induces a warp map $\tilde{\beta} : \mathbb{S} \rightarrow \mathbb{S}$ with $\tilde{\beta}(e^{2\pi it}) = e^{2\pi i\beta(t)}$, where β is referred to as the lift of $\tilde{\beta}$. This β satisfies $\beta(t+1) = \beta(t) + 1$ for all $t \in [0, 1]$, and consequently we have, for $t \in [0, 1]$, $\beta(t) = \gamma(t) + c$, where γ is a warp map of $[0, 1]$ and $c \in (0, 1]$ (through the identification of $[0, 1]$ with $\mathbb{R}/2\pi\mathbb{Z}$). This procedure can be viewed as one that produces a warp map of \mathbb{S} by ‘unwrapping’ \mathbb{S} at a chosen point c and generating a warp map of $[0, 1]$.

The random version of this corresponds to choosing a c according to a non-atomic probability measure μ on $[0, 1]$, independent of $\mathbb{D}_{\theta} \circ H$, resulting in a product probability measure $\mu \times \mathbb{D}_{\theta} \circ H$ on $(0, 1] \times W_I$. The procedure outlined above induces a bijection between the set $W_{\mathbb{S}}$ and $(0, 1] \times W_I$, and results in the following algorithm.

Algorithm 3. Random partition-based sampling of warp maps on \mathbb{S} .

1. Choose c from μ on $[0, 1]$.
2. Sample γ from $\mathbb{D}_{\theta} \circ H$ using Algorithm 2.
3. Set $\gamma_s(t) := (\gamma(t) + c) \bmod 1$.

This method was used in the work of [Graf et al. \(1986\)](#) while constructing random homeomorphisms of \mathbb{S} . The map γ_s is a warp map of $[0, 1]$ with a single point of discontinuity $t_c \in [0, 1]$ at which $\gamma(t_c) + c = 1$, thereby ensuring that $\gamma_s(t_c) = 0$. The point t_c is unique

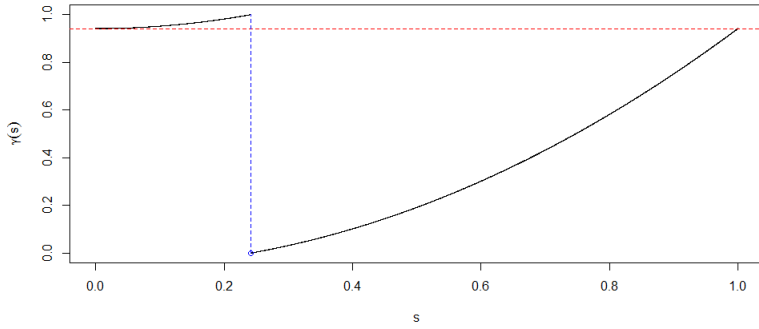


Figure 2: A warp map γ_s based on $\gamma(t) = t^2$ and $c = 0.94$ with $t_c = 0.24$.

to γ_s . Figure 2 offers an illustration of this approach for $\gamma(t) = t^2$ with $c = 0.94$, leading to $t_c = 0.24$. Proposition 2 formalizes this for a random γ_s generated in this fashion.

Proposition 2. *Conditional on c from μ , for each γ_s from $\mathbb{D}_\theta \circ H$, the following hold with probability one:*

- (1) $\gamma_s(0) = c$.
- (2) A unique t_c exists in the interior of $[0, 1]$ such that $\lim_{t \uparrow t_c} \gamma_s(t) = 0$ and $\lim_{t \downarrow t_c} \gamma_s(t) = 1$.

Thus, starting with probability measures $\mathbb{D}_\theta \circ H$ on W_I and μ on $(0, 1]$, the sampling scheme induces the product probability measure $\mu \times \mathbb{D}_\theta \circ H$ on the set of warp maps of \mathbb{S} . The corresponding measure, independent of the unwrapping point c , can be obtained by integrating the product probability measure with respect to μ . Note that the bijection ensures that (trivially) the resulting measure on warp maps of \mathbb{S} is necessarily absolutely continuous with respect to the probability measure on W_I used in the construction. The Supplementary Material contains an ‘intrinsic’ construction that circumvents the need to unwrap \mathbb{S} . But, the construction provided above is easier to implement in practice, and is thus the only one used in subsequent applications.

6 Numerical illustrations

The degeneracy phenomenon in Algorithm 1 can arise due to two reasons: (1) for large values of the concentration parameter α regardless of the size n and manner of construction of the

deterministic partition \mathcal{T}_n ; (2) moderate values for α but large n regardless of the manner of construction of the deterministic partition \mathcal{T}_n . High values of the concentration parameter θ in Algorithm 2 also results in a distribution $\mathbb{D}_\theta \circ H$ highly concentrated around H . We first illustrate this behavior in a simulation exercise. We then study the utility of Algorithms 2 and 3 in alignment problems in the context of real data analysis based on finite-dimensional distributions of the process $\mathcal{D}_\theta \circ H$ with law $\mathbb{D}_\theta \circ H$ in view of Theorem 2. Note that the theorems merely offer theoretical support for the two algorithms; in particular, there is no need to transform the increments with the inverse Levy measure.

6.1 Simulation examples

Figure 3 demonstrates the degeneracy issue arising from the use of Algorithm 1 discussed in Theorem 1. For deterministic partitions \mathcal{T}_n , as $n \rightarrow \infty$, samples from Algorithm 1 eventually concentrate around a deterministic warp map determined by the construction of \mathcal{T}_n . The top row of Figure 3 illustrates this behavior when \mathcal{T}_n is constructed using values of a Beta(2,1) distribution function; for increasing n , samples concentrate around the Beta(2,1) quantile function (see the Supplementary Material for a version of Theorem 1 for non-equi-spaced partitions). The bottom row illustrates the same for a uniform partition \mathcal{T}_n wherein the samples concentrate around the identity warp map. We now demonstrate the flexibility of $\mathbb{D}_\theta \circ H$ in modeling warp maps in W_I . For this purpose, we have chosen two partitions leading to two different choices of H : the uniform and the Beta(5,1). This results in $\mathbb{D}_\theta \circ H$ with average warp maps corresponding to uniform and Beta(5,1) distribution functions, respectively. Then, we simulated 300 warp maps for each case under the following settings: $n = 5, 20, 80$ and $\theta = 0.1, 10, 100$. The simulated warp maps under the uniform partition are shown in Figure 4(a), while panel (b) shows the warp maps sampled based on the Beta(5,1) partition. In both cases, we see that the proposed distribution is very flexible, exhibiting a variety of possible shapes of warp maps under different combinations of n and θ . A partition created with $n = 5$ results in few large jumps while a partition created with $n = 80$ generates warp maps with many small jumps. As θ is increased from 0.1 to 100, we notice the sample tightening around the warp map corresponding to the average map induced by the partition.

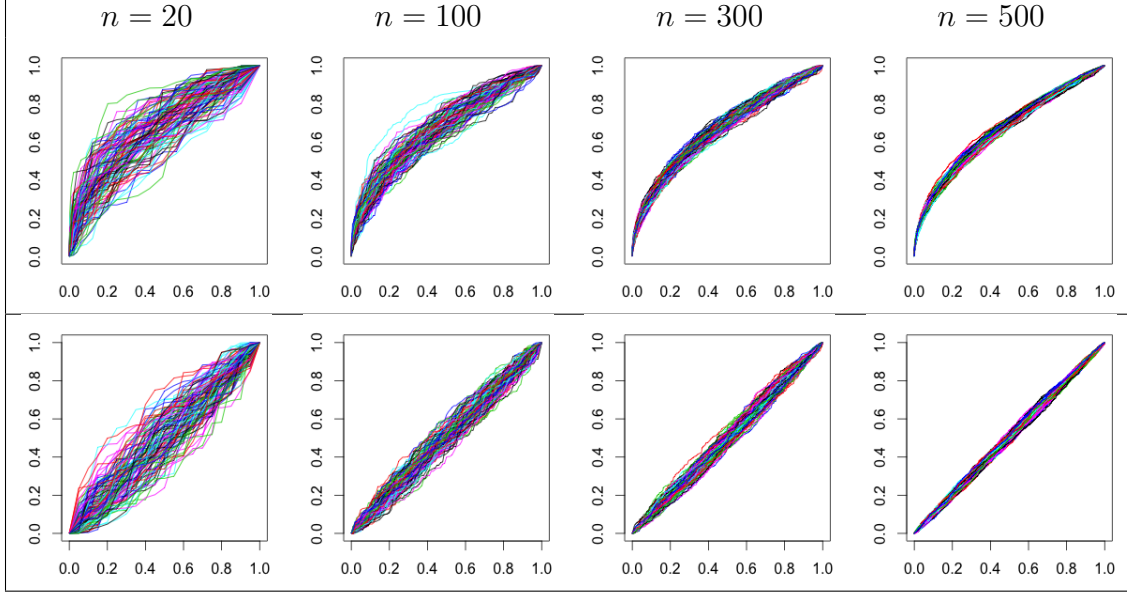


Figure 3: Sample warp maps from Algorithm 1 based on different deterministic partitions \mathcal{T}_n and sizes n with $\alpha = 1.2$. Top: \mathcal{T}_n is a non-equi-spaced partition based on a Beta(2,1). Bottom: Equi-spaced partition based on a U[0,1].

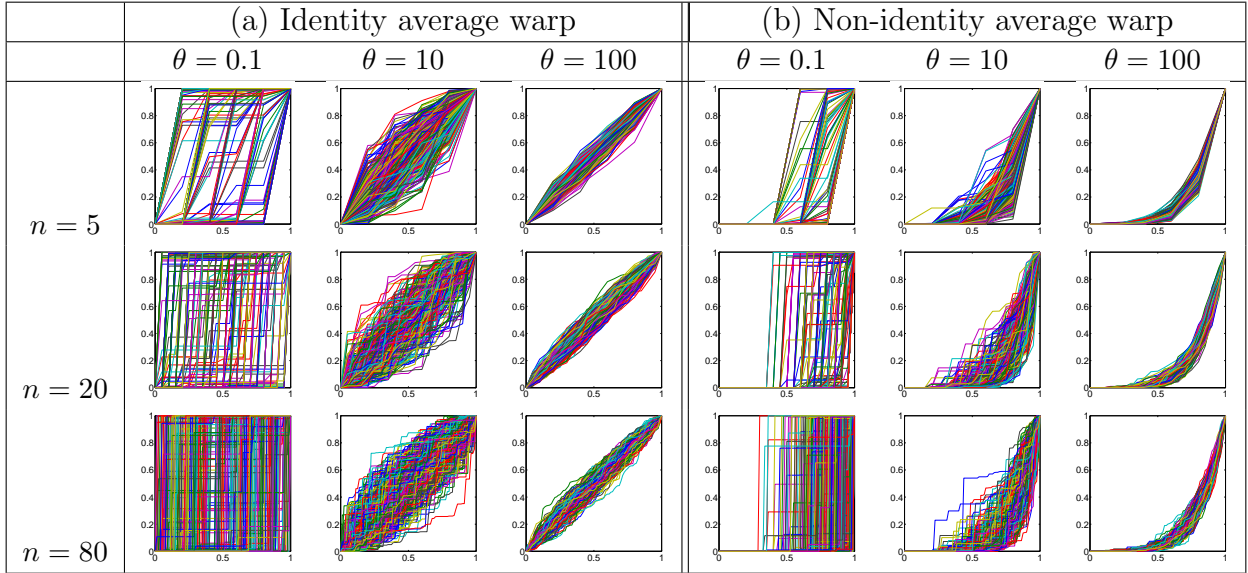


Figure 4: Sample warp maps from $\mathbb{D}_\theta \circ H$ based on Algorithm 2 centred at (a) $H(t) = t$ and (b) H corresponding to a Beta(5,1), under different choices of the partition size n and the concentration parameter θ .

6.2 Illustrations on real data

For detailed descriptions of the datasets used throughout this section, please refer to the Supplementary Material. We illustrate the utility of the proposed distribution and associated sampling scheme in two pairwise alignment tasks: (1) Bayesian alignment of univariate functions with and without landmark constraints, and (2) unconstrained alignment of univariate functions and higher-dimensional open and closed curves using a novel Simulated Annealing-based algorithm (see [Robert and Casella \(2005\)](#) for details). While the distribution $\mathbb{D}_\theta \circ H$ for sampling warp maps can be used with any alignment method, we use the framework based on the square-root velocity function (SRVF) representation of curves in \mathbb{R}^d , $d \geq 1$: $f \mapsto q := \dot{f}(|\dot{f}|)^{-1/2}$, where \dot{f} is the derivative of f and $|\cdot|$ is the Euclidean norm in \mathbb{R}^d . We use this representation due to its many nice properties for the registration problem (see [Srivastava et al. \(2011\)](#); [Kurtek et al. \(2012\)](#); [Srivastava et al. \(2011\)](#); [Lahiri et al. \(2015\)](#)). Under this representation, warping of a function $f \mapsto f \circ \gamma$ is given by $q \mapsto (q \circ \gamma)\sqrt{\dot{\gamma}}$.

6.2.1 Bayesian alignment of curves

Suppose we have two functions $g_i : [0, 1] \rightarrow \mathbb{R}$, $i = 1, 2$. A Bayesian registration model can be defined using their SRVF representations. The alignment problem then centres around an \mathbb{R} -valued, square-integrable, separable stochastic process $X(t) := q_1(t) - q_2(t)$ for $t \in [0, 1]$ with law \mathbb{P} and density $p = \frac{d\mathbb{P}}{d\mu}$ with respect to a σ -finite measure μ on $\mathbb{L}^2([0, 1])$. For a fixed $\gamma \in W_I$, assume that the law \mathbb{P}_γ of the process $X_\gamma(t) = q_1(t) - q_2(\gamma(t))\sqrt{\dot{\gamma}(t)}$ is absolutely continuous with respect to \mathbb{P} with density p_γ (see Theorem 6.4.5 in [Bogachev \(1998\)](#) for sufficient conditions). Suppose that the SRVFs, discretized at points $[t] = \{t_{0:n}, \dots, t_{n:n}\}$, are represented as vectors $q_i([t]) := (q_i(t_{0:n}), \dots, q_i(t_{n:n}))$, $i = 1, 2$. Then, $X_\gamma([t]) := q_1([t]) - q_2(\gamma([t]))\sqrt{\dot{\gamma}([t])} \sim p_\gamma$ prescribes a likelihood through the finite-dimensional projections of X_γ . The PL discretization of the functions, as well as the warp map γ , are theoretically supported by the work of [Lahiri et al. \(2015\)](#). Thus, in this context, the finite dimensional restriction of the probability measure $\mathbb{D}_\theta \circ H$ is well-suited for defining priors on warp maps.

The Bayesian model employed here is very similar to the one presented in [Cheng et al. \(2016\)](#) and [Kurtek \(2017\)](#). In short, the likelihood is a zero-mean multivariate Gaussian distribution with a diagonal covariance matrix. We select a conjugate, vague Gamma prior for the likelihood precision and analytically integrate it out of the posterior. While [Cheng et al.](#)

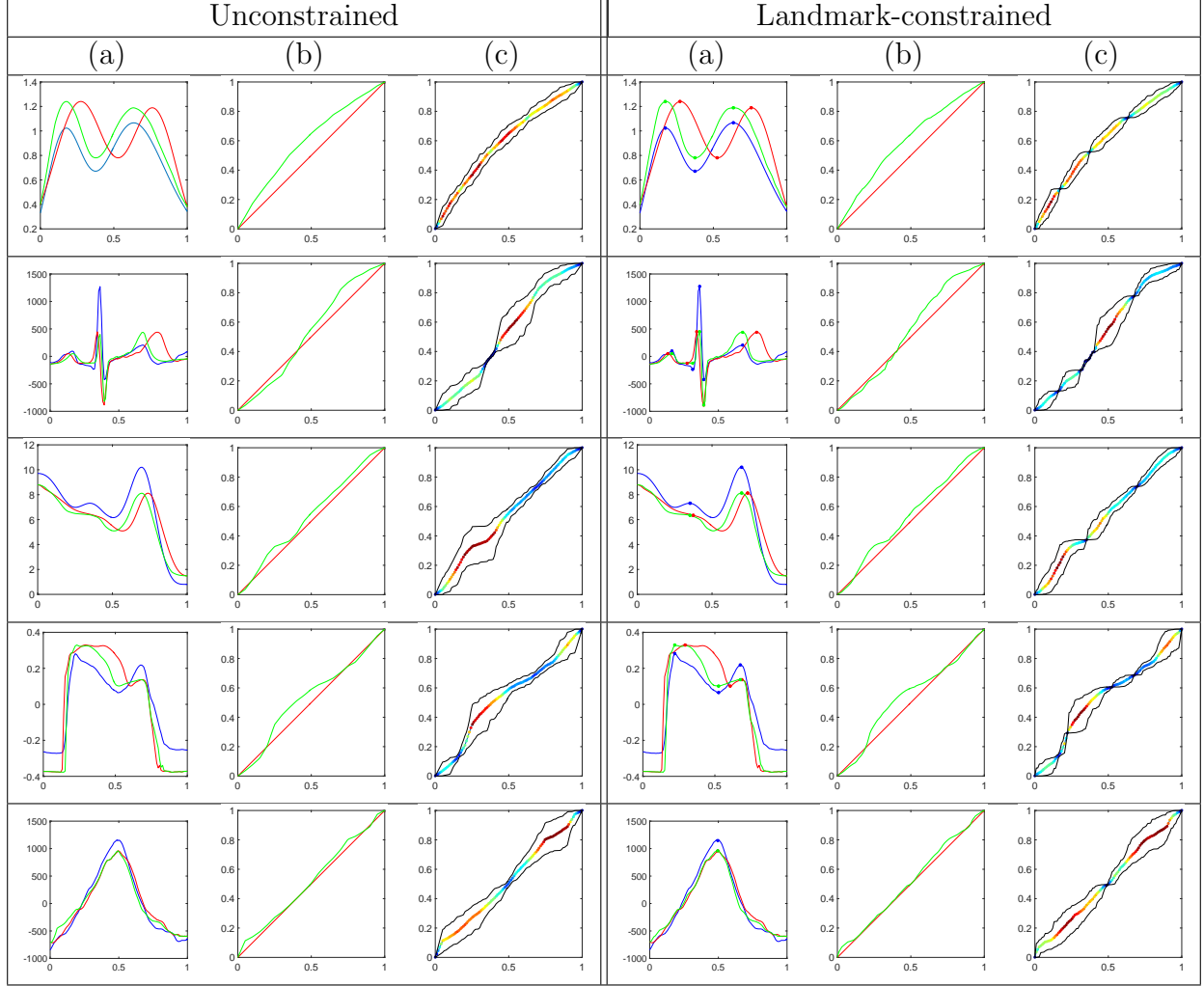


Figure 5: Bayesian alignment of functions. From top to bottom: (1) simulated data, (2) PQRST complexes, (3) growth rate functions, (4) gait pressure cycles, (5) respiration cycles. (a) Two functions before alignment (blue and red), and red function after alignment (green); landmarks are marked in the constrained version. (b) Posterior mean (green) and identity (red) warp maps. (c) Posterior mean warp map with a pointwise 95% credible interval (colors correspond to width of interval: blue=less uncertainty, red=more uncertainty).

(2016) sampled from the posterior distribution using Markov chain Monte Carlo (MCMC), we use a simple sampling importance resampling (SIR) algorithm, with the importance function set to the prior distribution as in Kurtsek (2017). This may not be the best approach to sample from the posterior, but it provides a fast approximation and seems to work well in the settings we considered. We assess the utility of the proposed prior distribution on W_I in two settings: (1) unconstrained, and (2) landmark-constrained function alignment. Note that the presented model can also be used for alignment of d -dimensional open and closed

curves where $d > 1$ with minor adjustments.

In order to sample from the distribution $\mathbb{D}_\theta \circ H$ on W_I via the easy-to-implement Algorithm 2, we require three specifications: (1) choice of partition that determines the average warp map H , (2) n , which controls the size of the partition, and (3) θ , which controls the spread around the average warp. We set $H(t) = t$, the uniform partition, which ensures regularization toward identity warping. We resample all functions with 100 points and choose $n = 20$, $\theta = 10$. This gives flexibility in the prior to explore extreme warpings (small θ) while also ensuring that the resulting warp maps are fairly regular (smaller partition prevents many small jumps). *In the case of landmark-constrained alignment, owing to the properties of $\mathbb{D}_\theta \circ H$ in Proposition 1, we re-scale the θ and n proportionally to the length of each function segment, and consider each sub-problem independently.*

Figure 5 presents results of unconstrained and landmark-constrained alignment for one simulated example and four real datasets. In the case of landmark-constrained alignment, the landmarks were selected either based on semantic features of the signals (e.g., PQIRST points in a complex) or mathematical features (e.g., peaks and valleys). In panels (b) and (c), we show the posterior mean warp map (cross-sectional average of the posterior sample), and a pointwise 95% credible interval. In all cases, the registration results are visually very good. Comparing columns (b) and (c) for the unconstrained and landmark-constrained cases, we observe intuitive differences in the posterior mean warp maps and corresponding credible intervals. The key observation is that in each of the datasets, when using additional information provided by the landmarks, the distribution $\mathbb{D}_\theta \circ H$ allows us to decompose the alignment problem into unconstrained sub-problems by enabling subset invariance; the constraints on the warp maps, and the properties of $\mathbb{D}_\theta \circ H$, ensure that there is almost zero uncertainty in regions close to the landmarks.

6.2.2 Curve alignment with Simulated Annealing

Based on the distributions $\mathbb{D}_\theta \circ H$ and $\mathbb{D}_\theta^s \circ H$, we present a novel stochastic algorithm for unconstrained alignment of three different types of functions: (1) univariate functions: $g : [0, 1] \rightarrow \mathbb{R}$, (2) shapes of 3D open curves: $g : [0, 1] \rightarrow \mathbb{R}^3$, and (3) shapes of planar closed curves: $g : \mathbb{S} \rightarrow \mathbb{R}^2$. A crucial step of the algorithm is based on the ability to propose warp maps in the neighborhood of any other warp map. The distributions \mathbb{D}^θ and \mathbb{D}_s^θ are well-suited

for this purpose: (1) we can centre the distributions at any warp map, and (2) we can control the size of the neighborhood via the parameter θ . We only consider the unconstrained case; the extension to landmark-constrained alignment, as demonstrated in the previous section, is achieved through subset invariance properties of the distributions.

As with the Bayesian model, we base our stochastic alignment algorithm on the SRVF representation of curves. The energy functional that we seek to optimize is $E(\gamma) = \|q_1 - (q_2 \circ \gamma)\sqrt{\tilde{\gamma}}\|^2$, for γ in W_I or W_S . When the domain of q_1 and q_2 is $[0, 1]$, a solution to this optimization problem can be obtained using a Dynamic Programming (DP) algorithm (Robinson, 2012). The resulting solution is deterministic and depends on the fineness of the discretization and the size of the neighborhood search. In the case of closed curves, one has to either resort to a gradient descent algorithm (Srivastava et al., 2011), which has the obvious limitation of getting stuck in a local solution, or a DP approach with an additional seed (the point at which \mathbb{S} is unwrapped to $[0, 1]$) search, which only gives an approximate solution. The energy functional E is a natural choice for curve registration, because (1) the \mathbb{L}^2 distance in the energy corresponds to an elastic metric on the space of curves, and (2) this elastic metric is preserved under identical warping (isometry); see Srivastava and Klassen (2016) for details.

The Simulated Annealing alignment algorithm for functional data (i.e., $g : [0, 1] \rightarrow \mathbb{R}$) using $\mathbb{D}_\theta \circ H$ on W_I is given as Algorithm 4. The extension to alignment of open and closed curves for the purpose of shape analysis is commented upon below.

Algorithm 4. Alignment of functions via Simulated Annealing.

Inputs: $g_i : [0, 1] \rightarrow \mathbb{R}$, $i = 1, 2$ (SRVFs $q_i : [0, 1] \rightarrow \mathbb{R}^d$, $i = 1, 2$).

Outputs: Optimal warp map $\gamma^* : [0, 1] \rightarrow [0, 1]$.

Initialize: $n = 20$, $\theta = 100$, $T = 10$, $\gamma_0 = \gamma_{id}$, $E_0 = \|q_1 - q_2\|^2$ and $j = 0$.

1. Generate a random $\tilde{\gamma}$ from $\mathbb{D}_\theta \circ H$ with H set to the warp map γ_j . Set $\gamma_p = 0.9\tilde{\gamma} + 0.1\gamma_{id}$.
2. Compute $E(\gamma_p) = \|q_1 - (q_2 \circ \gamma_p)\sqrt{\tilde{\gamma}_p}\|^2$.
3. Accept $\gamma_{j+1} = \gamma_p$, and set $E(\gamma_{j+1}) = E(\gamma_p)$, with probability $\min\left\{1, e^{\frac{E(\gamma_j) - E(\gamma_p)}{T}}\right\}$.
Otherwise, let $\gamma_{j+1} = \gamma_j$ and $E(\gamma_{j+1}) = E(\gamma_j)$.
4. Set $j = j + 1$ and update the temperature to $T = T/c$ (we suggest $c = 1.0001$).

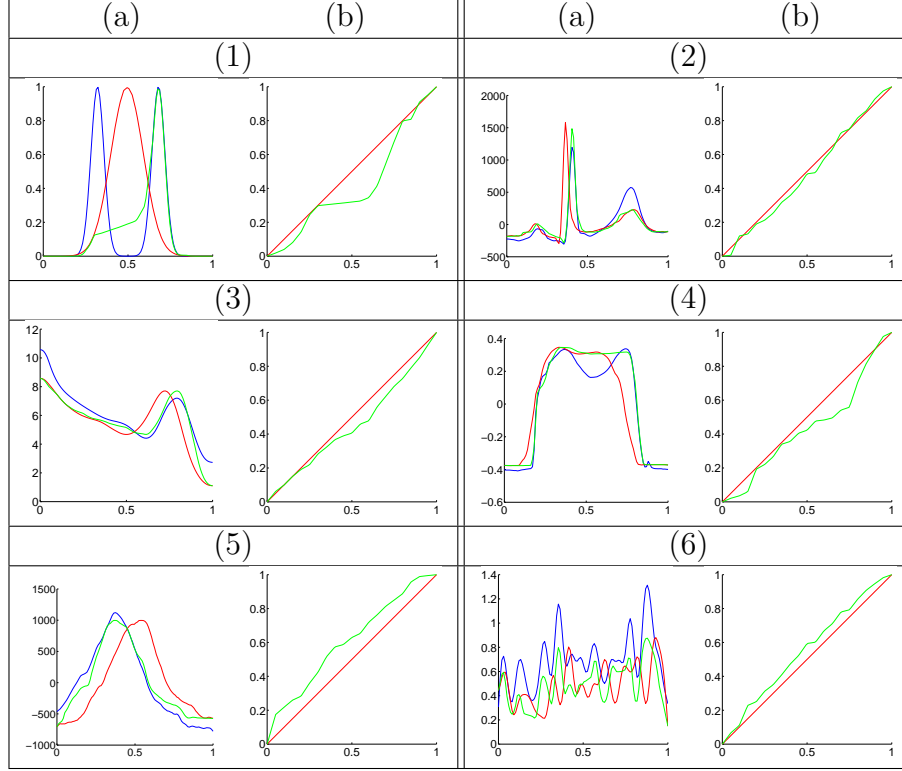


Figure 6: Simulated Annealing-based alignment on six examples: (1) simulated data, (2) PQRST complexes, (3) growth rate functions, (4) gait pressure cycles, (5) respiration cycles, (6) tangential acceleration of signature curves. (a) Two functions before alignment (blue and red), and red function after alignment (green). (b) Optimal (green) and identity (red) warp maps.

The algorithm is fairly robust to the choices of n, θ, T and c . Numerical illustrations investigating robustness are presented in the Supplementary Material. In the first step of the algorithm, we propose a candidate warp map that is a linear combination of a random warping sampled from the distribution $\mathbb{D}_\theta \circ H$ centred at the previously accepted warp map and the identity warping; while not necessary, this ensures extra regularization toward identity warping. The key here is that efficient sampling from $\mathbb{D}_\theta \circ H$ centred at an arbitrary warp map is easily enabled through Algorithm 2. Furthermore, the relative values of the concentration parameter θ (neighborhood size) and the T control the dynamics of the algorithm.

Next, we briefly comment on how Algorithm 4 is extended to the case of alignment of open and closed curves for the purpose of shape analysis. In the case of shapes, translation, scale and rotation variations are nuisances and have to be removed in addition to alignment via warp maps. Translation is removed automatically through the SRVF representation; scale variation is removed by normalizing all curves to unit length. Rotation variability is

	(a)	(b)	(c)		(a)	(b)	(c)
(1)	2.74	1.43	1.37	(2)	92.06	17.58	21.77
(3)	3.07	0.71	0.82	(4)	1.36	0.53	0.52
(5)	57.57	11.09	15.43	(6)	3.73	0.76	1.26

Table 1: Distances (a) before alignment, (b) after DP alignment, and (c) after Simulated Annealing alignment, for the six examples in Figure 6.

accounted for by adding a Procrustes step at each iteration of the algorithm (see [Dryden and Mardia \(1998\)](#) for details on Procrustes alignment). In the case of closed curves, we must additionally propose the seed point, which is used to unwrap \mathbb{S} to $[0, 1]$ and eventually sample from $\mathbb{D}_\theta \circ H$; this is accomplished via a random proposal, in the neighborhood of the current seed point, from the von-Mises distribution on \mathbb{S} .

We begin with six examples that consider univariate functional data. The alignment results are presented in Figure 6. In all cases, the computed warp map provides a nice alignment of features across functions. For example, in panel (2), we consider alignment of two PQIRST complexes without imposing landmark constraints on the warp maps. Originally, the PQIRST peaks and valleys are not well aligned; this is especially evident in the case of the R peaks (highest peak in each function). The proposed method is able to align the peaks very well. This is also the case in the more complex example (6) that considers alignment of two signature tangential acceleration functions. These functions contain many peaks and valleys that are not in correspondence before alignment. The proposed method is able to effectively align all of the peaks and valleys via a suitable warp map. Table 1 provides a numerical evaluation of our approach. For each of the six examples, we compare three different distances between the functions: (a) distance before alignment, (b) distance after alignment using DP, and (c) distance after alignment using Simulated Annealing. The proposed method provides comparable performance to DP.

Next, we present several results of registering shapes of 3D open curves. In this case, we use two datasets that were previously considered in [Kurtek et al. \(2012\)](#): (1) simulated spirals, and (2) fibers extracted from diffusion tensor magnetic resonance images (DT-MRIs). The results are presented in Figure 7. For each example, we show the optimal warp map, the evolution of the energy $E(\gamma)$ as a function of the number of iterations, and the geodesic path (shortest distance deformation under the \mathbb{L}^2 metric on SRVF representations) between the two shapes before and after Simulated Annealing-based alignment. For the simulated spirals,

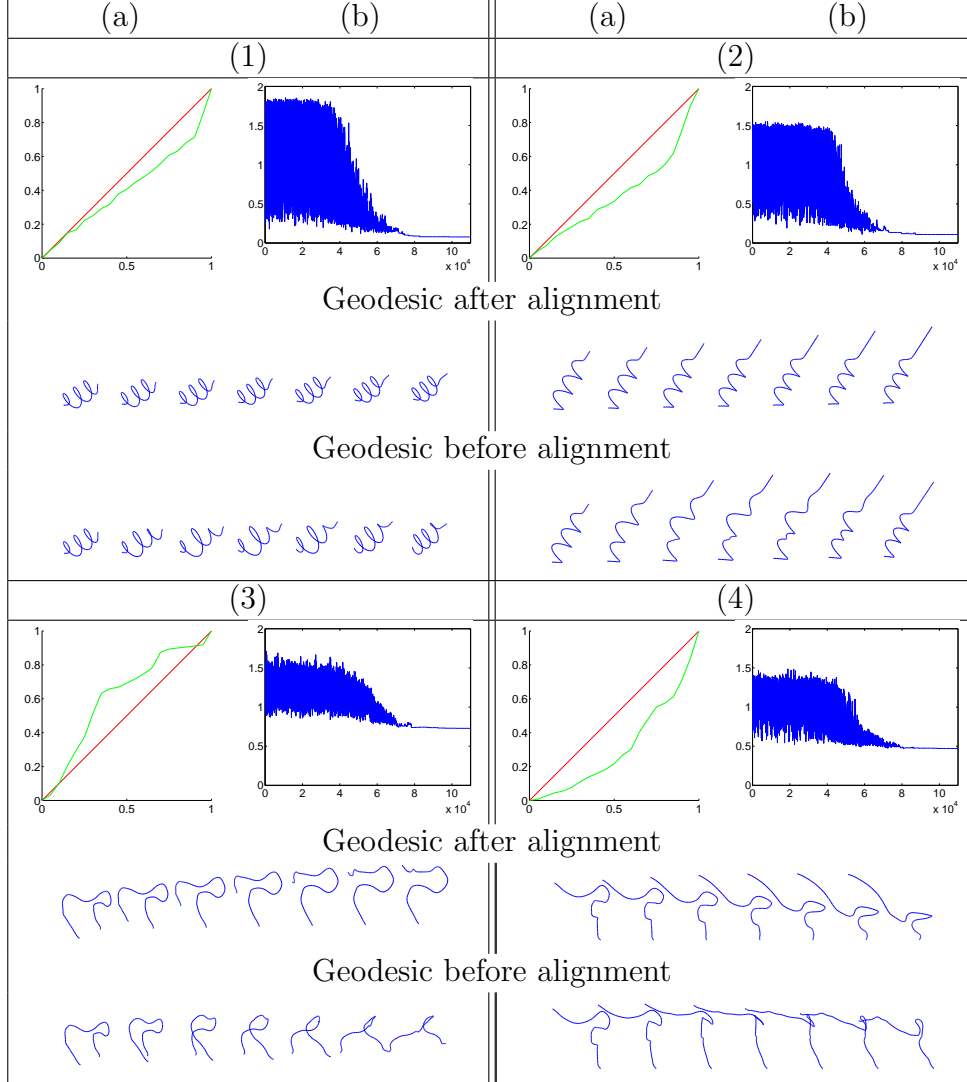


Figure 7: Results of Simulated Annealing-based alignment on four different examples of open curves. (a) Optimal (red) and identity (green) warp maps. (b) Evolution of $E(\gamma)$.

the additional alignment via warp maps results in a much more natural geodesic deformation between them, where the shapes of the individual spirals are better preserved. This is also the case for the DT-MRI fibers, albeit not as clear. The top portion of Table 2 provides a quantitative comparison of DP-based alignment and the proposed method. In three out of the four given examples, the proposed method results in a significantly shorter distance between the considered shapes (the maximum distance on this shape space is $\pi/2$).

We close this section with four examples of registering closed curves from the MPEG-7 dataset using Simulated Annealing; these curves represent fairly complex shapes including

	(a)	(b)	(c)		(a)	(b)	(c)
Open Curves							
(1)	1.20	0.39	0.28	(2)	1.24	0.25	0.33
(3)	1.17	0.97	0.88	(4)	1.11	0.92	0.70
Closed Curves							
(1)	0.57	0.42	0.26	(2)	1.03	0.68	0.54
(3)	0.63	0.42	0.45	(4)	0.77	0.50	0.52

Table 2: Distances (a) before alignment, (b) after DP alignment, and (c) after Simulated Annealing alignment, for the examples shown in Figures 7 and 8.

a cup with a handle and a stingray. Recall that in the case of closed curves, we must optimize over the seed placement (point at which \mathbb{S} is unwrapped to $[0, 1]$) on the curve in addition to the warp map. The results are presented in Figure 8. We provide the same displays as in the open curve examples. As previously, the geodesic paths after alignment represent more natural deformations between the shapes than those before alignment. This is especially evident in the cup example. We present our quantitative assessment in the bottom portion of Table 2. Here, we compare to a DP approach with an additional seed search. The proposed method performs significantly better on examples (1) and (2), and gives comparable performance on examples (3) and (4).

7 Discussion

The class of warp maps of $[0, 1]$ can be identified with the set of distribution or quantile functions on $[0, 1]$. Sample paths of Levy subordinators normalized to obtain normalized random measures (Dirichlet process is a special case), and variants thereof (Hjort, 1990; Regazzini et al., 2003; Nieto-Barajas et al., 2004), can be used to define distributions on warp maps along with numerous corresponding sampling schemes (see Griffin (2016) and references therein). However, warp maps of \mathbb{S} cannot be identified with distribution functions, and an ‘intrinsic’ distribution based on arc lengths (as described in the Supplementary Material) is not easily obtainable as laws of normalized random measures.

We have given two methods for *pairwise* matching of curves: (1) a Bayesian registration model, and (2) a stochastic search algorithm via Simulated Annealing. In many applications, it may be of interest to match multiple curves simultaneously, also termed *multiple* registra-

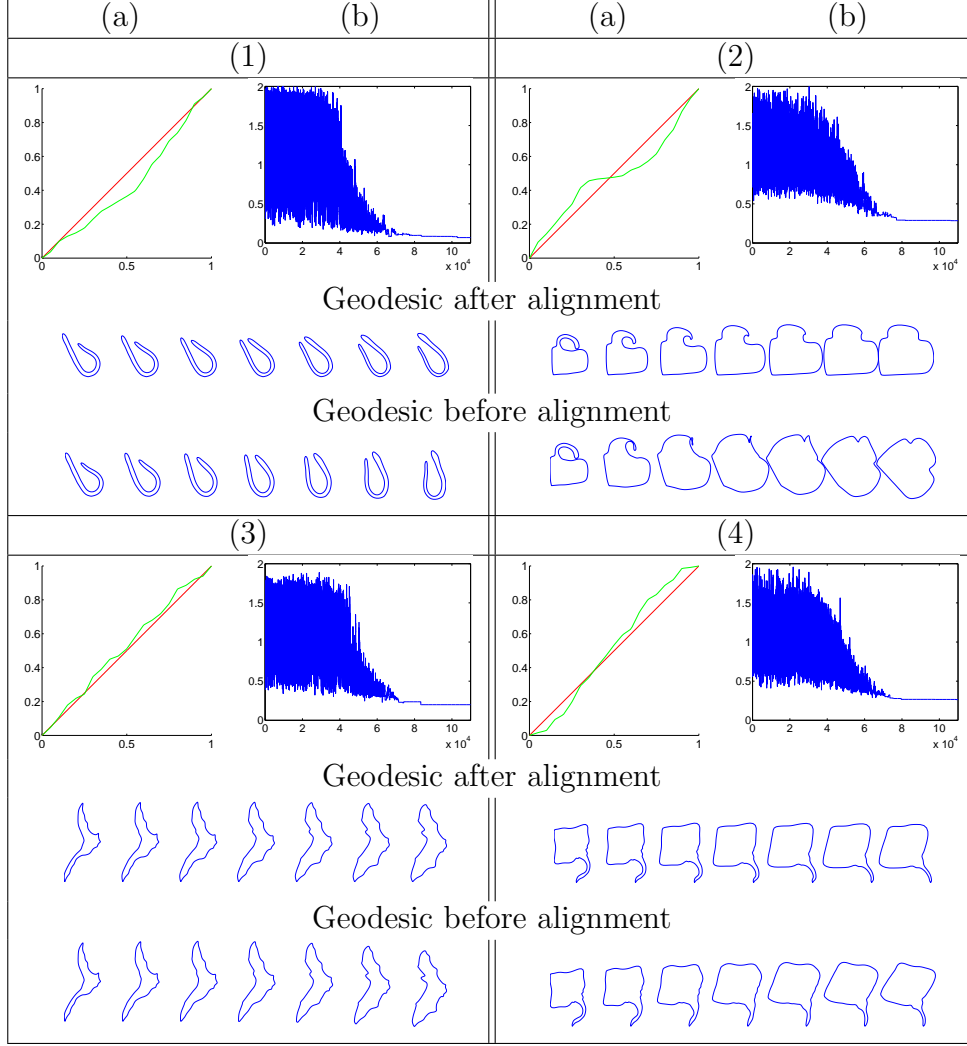


Figure 8: Results of Simulated Annealing-based alignment on four different examples of closed curves. (a) Optimal (green) and identity (red) warp maps. (b) Evolution of $E(\gamma)$.

tion. This is usually accomplished by joint estimation of a template curve and an additional matching step. The proposed Simulated Annealing-based alignment can be easily incorporated into a multiple registration algorithm by replacing the commonly used DP approach. Multiple alignment via a formal Bayesian model requires a prior on the template curve in addition to the warp maps. Nonetheless, our approach can be readily built into existing Bayesian multiple registration models such as the one presented in [Cheng et al. \(2016\)](#).

Although unexplored in this paper, the proposed distribution on warp maps is well-suited for curve registration with landmarks observed with uncertainty in their placement. The desirable properties allow us to develop priors centred at piecewise linear warp maps that match

the landmarks exactly. This allows for incorporating prior information into the problem by regularizing the warp maps toward a landmark induced warping.

Acknowledgements: We thank Ian Dryden, Huiling Le and Eric Klassen for helpful discussions. We also thank two anonymous reviewers for their suggestions. This research was partially supported by NSF DMS 1613054 and NIH R01 CA214955 (to KB and SK). SK was also partially supported by NSF CCF 1740761.

References

- Bauer, M., M. Eslitzbichler, and M. Grasmair (2017). Landmark-guided elastic shape analysis of human character motions. *Inverse Problems and Imaging* 11(4), 601–621. 4
- Bogachev, V. I. (1998). *Gaussian Measures*. American Mathematical Society. 21
- Cheng, W., I. L. Dryden, and X. Huang (2016). Bayesian registration of functions and curves. *Bayesian Analysis* 11(2), 447–475. 4, 5, 6, 7, 21, 22, 29
- Claeskens, G., B. W. Silverman, and L. Slaets (2010). A multiresolution approach to time warping achieved by a Bayesian prior-posterior transfer fitting strategy. *Journal of the Royal Statistical Society, Series B* 72(5), 673–694. 4, 16
- Devroye, L. (1981). Laws of iterated logarithm for order statistics of uniform spacings. *Annals of Probability* 9(1), 860–867. 13
- Dryden, I. L. and K. V. Mardia (1998). *Statistical Shape Analysis*. John Wiley & Sons, Chichester. 26
- Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *Annals of Statistics* 1(2), 209–230. 5
- Ferguson, T. S. and M. J. Klass (1972). A representation of independent increments processes without Gaussian components. *Annals of Mathematical Statistics* 43(2), 1634–1643. 10
- Gasser, T. and A. Kneip (1995). Searching for structure in curve samples. *Journal of the American Statistical Association* 90, 1179–1188. 4

- Gervini, D. and T. Gasser (2004). Self-modelling warping functions. *Journal of the Royal Statistical Society, Series B* 66(2), 959–971. 4
- Graf, S., R. D. Mauldin, and S. C. Williams (1986). Random homeomorphisms. *Advances in Mathematics* 60, 239–359. 17
- Griffin, J. E. (2016). An adaptive truncation method for inference in Bayesian nonparametric models. *Statistics and Computing* 26(1), 423–441. 28
- Hjort, N. L. (1990). Nonparametric Bayes estimators based on Beta processes in models for life history data. *Annals of Statistics* 18(2), 1259–1294. 28
- Hodges, J. L. (1955). On the noncentral beta-distribution. *Annals of Mathematical Statistics* 26(5), 648–653. 8
- Kneip, A. and T. Gasser (1992). Statistical tools to analyze data representing a sample of curves. *Annals of Statistics* 20, 1266–1305. 4
- Kurtek, S. (2017). A geometric approach to pairwise Bayesian alignment of functional data using importance sampling. *Electronic Journal of Statistics* 11(1), 502–531. 4, 21, 22
- Kurtek, S., A. Srivastava, E. Klassen, and Z. Ding (2012). Statistical modeling of curves using shapes and related features. *Journal of the American Statistical Association* 107(499), 1152–1165. 21, 26
- Lahiri, S., D. Robinson, and E. Klassen (2015). Precise matching of PL curves in \mathbb{R}^n in the Square Root Velocity framework. *Geometry, Imaging and Computing* 2, 133–186. 16, 21
- Lu, Y., R. Herbei, and S. Kurtek (2017). Bayesian registration of functions with a Gaussian process prior. *Journal of Computational and Graphical Statistics* 26(4), 894–904. 4
- Nagaraja, H. N., K. Bharath, and F. Zhang (2015). Spacings around an order statistic. *Annals of Institute of Statistical Mathematics* 67, 515–540. 13
- Nieto-Barajas, L. E., I. Prünester, and S. G. Walker (2004). Normalized random measures driven by increasing additive processes. *Annals of Statistics* 32(2), 2343–2360. 28

- Ramsay, J. O. and B. W. Silverman (2005). *Functional Data Analysis*. 2nd edition, Springer, New York. 16
- Regazzini, E., A. Lijoi, and I. Prünester (2003). Distributional results for means of normalized random measures with independent increments. *Annals of Statistics* 31(2), 560–585. 28
- Robert, C. P. and G. Casella (2005). *Monte Carlo Statistical Methods*. Springer-Verlag, New York. 21
- Robinson, D. T. (2012). *Functional Data Analysis and Partial Shape Matching in the Square Root Velocity Framework*. Ph. D. thesis, Florida State University. 24
- Srivastava, A., E. Klassen, S. H. Joshi, and I. H. Jermyn (2011). Shape analysis of elastic curves in Euclidean spaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33, 1415–1428. 21, 24
- Srivastava, A. and E. P. Klassen (2016). *Functional and Shape Data Analysis*. Springer-Verlag, New York. 4, 24
- Srivastava, A., W. Wu, S. Kurtek, E. Klassen, and J. S. Marron (2011). Registration of functional data using the Fisher–Rao metric. *arXiv:1103.3817v2*. 4, 21
- Strait, J., S. Kurtek, E. Bartha, and S. M. MacEachern (2017). Landmark-constrained elastic shape analysis of planar curves. *Journal of the American Statistical Association* 112(518), 521–533. 4
- Tang, R. and H.-G. Müller (2008). Pairwise curve synchronization for functional data. *Biometrika* 95, 875–889. 4
- Telesca, D. and L. Y. T. Inoue (2008). Bayesian hierarchical curve registration. *Journal of the American Statistical Association* 103, 328–339. 4
- Trouné, A. and L. Younes (2000). On a class of diffeomorphic matching problems in one dimension. *SIAM Journal on Control and Optimization* 39(4), 1112–1135. 4
- von Renesse, M.-K. and K.-T. Sturm (2009). Entropic measure and Wasserstein diffusion. *Annals of Probability* 37(2), 1114–1191. 9, 11, 12, 15, 16

Zhou, R. R., N. Serban, N. Gebraeel, and H.-G. Müller (2014). A functional time warping approach to modeling and monitoring truncated degradation signals. *Technometrics* 56, 67–77. 4