Journal of the American Statistical Association

ISSN: 0162-1459 (Print) 1537-274X (Online) Journal homepage: https://www.tandfonline.com/loi/uasa20

Automatic Detection and Uncertainty Quantification of Landmarks on Elastic Curves

Justin Strait, Oksana Chkrebtii & Sebastian Kurtek

To cite this article: Justin Strait, Oksana Chkrebtii & Sebastian Kurtek (2019): Automatic Detection and Uncertainty Quantification of Landmarks on Elastic Curves, Journal of the American Statistical Association, DOI: 10.1080/01621459.2018.1527224

To link to this article: https://doi.org/10.1080/01621459.2018.1527224

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION
2019, VOL. 00, NO. 0, 1–16: Applications and Case Studies
https://doi.org/10.1080/01621459.2019.1527224

Automatic Detection and Uncertainty Quantification of Landmarks on Elastic Curves

Justin Strait[a], Oksana Chkrebtii[b], and Sebastian Kurtek[b]

[a]Department of Statistics, University of Georgia, Athens, GA; [b]Department of Statistics, The Ohio State University, Columbus, OH

ABSTRACT

A population quantity of interest in statistical shape analysis is the location of landmarks, which are points that aid in reconstructing and representing shapes of objects. We provide an automated, model-based approach to inferring landmarks given a sample of shape data. The model is formulated based on a linear reconstruction of the shape, passing through the specified points, and a Bayesian inferential approach is described for estimating unknown landmark locations. The question of how many landmarks to select is addressed in two different ways: (1) by defining a criterion-based approach and (2) joint estimation of the number of landmarks along with their locations. Efficient methods for posterior sampling are also discussed. We motivate our approach using several simulated examples, as well as data obtained from applications in computer vision, biology, and medical imaging. Supplementary materials for this article, including a standardized description of the materials available for reproducing the work, are available as an online supplement.

ARTICLE HISTORY
Received September 2017
Revised September 2018

KEYWORDS
Elastic metric; Landmarks; Linear reconstruction; Markov chain Monte Carlo; Shape analysis.

1. Introduction

Shape analysis is an emerging field within statistics due to the necessity of making inference on the shapes of objects. Shape is an important physical property of objects and emerges in various application areas including medical imaging, pattern recognition, computer vision, biometrics, biology, bioinformatics, and many others. In general, statistical shape analysis refers to a set of tools, which can be used for alignment, comparison, averaging, summarization of variability, statistical inference, and other tasks performed on shape spaces. These tools can resemble standard statistical methods developed for multivariate numerical data; however, developing these methods require extra care for several reasons. First, there is no consensus on the choice of shape representation, which determines how these objects are treated mathematically. Many representations have been developed; two of the most common classes are landmark-based and function-based, and are described later in this section. Second, most shape representation spaces are quotients of nonlinear manifolds (where performing operations like adding shapes is not straightforward as in a linear space). This is due to the most common definition of shape as an inherited property of an object, which remains unchanged under some transformations (most commonly rotation, scaling, and translation). Quotient spaces are required to deem two shapes equivalent when they only differ by this set of transformations. Third, in the case of functional representations of shape, the underlying shape spaces are infinite-dimensional. Thus, any statistical analysis on these spaces requires tools from functional data analysis.

Initially, the statistical shape analysis community represented an object's shape using a finite point set comprised of so-called *landmarks*. These ideas were first introduced by Kendall (1984), who defined shape as a property of an object which remains unchanged under rigid motion and scaling. The landmark points represent important mathematical (e.g., curvature) or salient anatomical features of the objects and are in correspondence across a population of shapes—this means that, for instance, if a landmark is placed at the tip of a human's nose, then this particular point should be matched with the nose of another human that may be compared to it. In this framework, the entire object is represented by a low-dimensional landmark configuration matrix which is based on the coordinates of the landmarks. After some adjustments to account for the desired shape invariances, standard multivariate analyses can be performed on these shape representation spaces (details provided in Dryden and Mardia 2016; Small 1996; Dryden and Mardia 1992; Bookstein 1986). If landmarks can be located on objects of interest, then this approach provides a low-dimensional shape representation for which many statistical tools are readily available.

As computing technology improved, researchers developed infinite-dimensional, functional representations of shape based on parameterized curves of the objects' outlines. These representations allow one to model the full structure of the object of interest, but also lead to some additional challenges. Most notably, statistical shape analysis of parameterized curves often should be invariant to reparameterization of the curves (in addition to rigid motion and scaling). In other words, any statistical analyses should be the same regardless of the rate at which the curve is traversed. To overcome this challenge, *elastic statistical shape analysis* (Younes 1998; Michor et al. 2007; Srivastava et al. 2011) was introduced as a parameterization-invariant way to compare and model curves. This is accomplished by

CONTACT Justin Strait *justin.strait@uga.edu* Department of Statistics, University of Georgia, 310 Herty Dr., Athens, GA 30602.
Color versions of one or more of the figures in the article can be found online at *www.tandfonline.com/r/JASA*.
Supplementary materials for this article are available online. Please go to *www.tandfonline.com/r/JASA*.
These materials were reviewed for reproducibility.

matching corresponding geometric features across shapes (for instance, the tails of animals being compared), which provides improved results over arc-length parameterization methods (Zahn and Roskies 1972; Klassen et al. 2004). Srivastava et al. (2011) introduced a novel representation of shape called the square-root velocity function (SRVF), which greatly simplifies statistical analysis under the elastic shape analysis paradigm.

More recently, Strait et al. (2017) extended this work to allow hard landmark constraints in the SRVF representation (known as landmark-constrained elastic shape analysis). This new development provides tools for statistical modeling of the full parameterized curve representation of an object's boundary while at the same time respecting given landmark constraints, that is, by enforcing exact landmark matching. These methods are useful for comparison of shapes where the entire object is treated as a function, but special points are "forced" to match (for instance, if a particular shape feature is visible on one object and known but not visible on the other object). Other examples of landmark-constrained elastic shape analysis include the works of Bauer, Eslitzbichler, and Grasmair (2017) and Liu, Srivastava, and Zhang (2010), where landmarks are treated as soft constraints, that is, landmarks are used to compute optimal deformations and distances between shapes, but are not necessarily matched exactly.

This work seeks to answer two pertinent questions related to landmark-based shape analysis methods, including landmark-constrained elastic shape analysis. First, in general settings, it is not clear how many landmark points should be selected to represent the shape of interest. Too few landmarks may result in the absence of important features of objects, effectively leading to biased estimation; too many may result in overfitting, a classical statistical problem. Once the number of landmarks is decided, one may also wonder where these landmarks should be located. In the case of anatomical landmarks, the points are usually selected by an expert in the application field, for example, in medical imaging, doctors manually annotate important anatomical features in an image. However, such an approach is time consuming, expensive, and prone to human error. Thus, we propose a novel automatic, model-based approach for answering these two questions under a joint framework. The Bayesian paradigm is a natural approach to infer fixed but unknown landmark locations while accounting for their associated uncertainty.

1.1. Previous Work

The proposed automatic landmark detection framework is applicable to any open or closed curve, regardless of type, dimensionality, or shape. Automatic landmark detection has been discussed mainly in the presence of specific classes of shapes (e.g., specific anatomy), and primarily from an image analysis perspective. These include the work by Chen et al. (2014) who focus on the inference of landmarks on X-ray images based on a voting scheme through displaced image patches. Facial landmarks have been the subject of many manuscripts, including those by Tie and Guan (2013), Segundo et al. (2010), and Gilani, Shafait, and Mian (2015). The latter two focus on curvature-based methods. A more general approach (applicable to an arbitrary class of shapes) was proposed by Rueda, Udupa, and Bai (2008a, 2008b). Existing frameworks are based primarily on feature detection and optimization of various application-specific criteria, and not developed from underlying statistical models; this can make uncertainty quantification challenging. To the best of our knowledge, one of the only mentions of a model-based automated landmark detection method in the Bayesian setting was presented by Domijan and Wilson (2005). However, this model was presented from an image analysis perspective, where the likelihood was based on distances between segmented image pixels and a polygon formed by landmarks; it did not account for shape-preserving transformations. In contrast, we are interested in finding landmarks directly on the given shapes (rather than their corresponding image). Note that this problem is unrelated to landmark registration of functions found in functional data analysis (see Ramsay and Silverman 2005; Gasser and Kneip 1995), which has the goal of aligning functions based on automatically detected or user-specified landmarks (i.e., local extrema). Our goal is not to register shapes, but to automatically identify these landmarks, which is often quite difficult because identifying local extrema is not trivial for shapes.

The methods of Prematilake and Ellingson (2018) and Strait and Kurtek (2016) are most relevant to this work. Prematilake and Ellingson (2018) address the problem of choosing the number and location of discretization points used to store and analyze an infinite-dimensional shape. These are distinct from landmarks, but comprise a candidate set for landmark selection. In their framework, the number of discretization points is selected by minimizing an approximation error; locations of these sampling points are then automatically determined to be equally spaced with respect to arc-length or absolute curvature. We seek to go one step further and characterize the *most* important points (landmarks) on a shape. The model proposed by Strait and Kurtek (2016) allows for landmark inference for generic shapes within the Bayesian framework. However, their model has some severe limitations which we address in the current work. First, their approach is able to identify landmarks on one shape only (rather than a collection of shapes); this allows Strait and Kurtek (2016) to use importance sampling to generate posterior samples, which is not computationally feasible as the sample size increases (see Section 1 of the supplementary material for further discussion). Second, their model requires that the number of landmarks be known. The main contribution of our work is a Bayesian model-based approach for automatic inference of landmark locations, and their number, which addresses all of the above-described limitations.

The rest of this article is organized as follows. Section 2 outlines useful background information and motivates our framework. Section 3 presents our model for a sample of shapes under the assumption that the number of landmarks is known and fixed. In Section 4, the original model is extended for an unknown number of landmarks. Simulation studies are included in Section 5. Finally, we present several applications in Section 6, and close the article with a discussion and some directions for future work in Section 7. The supplementary material includes further discussion on posterior sampling, assessments of Markov chain Monte Carlo (MCMC) convergence, discretization issues, and additional examples.

Figure 1. Nonelastic (left) and elastic (right) deformation between two shapes.

2. Background Material: Elastic Shape Analysis

We present a brief overview of topics in elastic statistical shape analysis relevant to the proposed approach. For further details, please consult Srivastava et al. (2011), Kurtek et al. (2012), and Srivastava and Klassen (2016). Let $\beta : \mathcal{D} \to \mathbb{R}^d$ be an absolutely continuous curve (corresponding to the outline of the object of interest) defined on a domain $\mathcal{D}$. For open curves, the domain is the interval $\mathcal{D} = [0, 1]$. Closed curves, defined on the unit circle $\mathcal{D} = \mathbb{S}^1$, will be equivalently represented for ease of exposition by rescaling the domain to $\mathcal{D} = [0, 1]$ and further enforcing the end-point constraint $\beta(0) = \beta(1)$ (i.e., $\mathcal{D}$ is equivalent to a circle with radius $(2\pi)^{-1}$). In subsequent sections, we focus on the case of planar curves ($d = 2$), but the presented models readily extend to higher-dimensional curves. Thus, β can be written as $\beta(t) = (\beta_x(t), \beta_y(t))^\top$, where β_x, β_y are coordinate functions mapping $\mathcal{D} \to \mathbb{R}$.

In this article, we consider shape as a property of an object invariant under its rotation, translation, scale, and reparameterization. This is called elasticity. In other words, applying any combination of these transformations to an object does not change its shape. Mathematically, a reparameterization is defined as an element of the group $\Gamma = \{\gamma : [0,1] \to [0,1] \mid \gamma(0) = 0, \gamma(1) = 1, 0 < \dot{\gamma} < \infty\}$, where $\dot{\gamma}$ is the time-derivative $\frac{d\gamma}{dt}$. A reparameterization of a curve β using a function $\gamma \in \Gamma$, given by the composition $\beta \circ \gamma$, effectively changes the rate of traversal of the curve in the same direction.

Statistical analysis on the space of elastic shapes requires a metric which respects the desired shape invariances. One way to compare shapes of two curves β_1 and β_2 is to use the $\mathbb{L}^2$ metric, defined as $\|\beta_1 - \beta_2\| = \sqrt{\int_{\mathcal{D}} |\beta_1(t) - \beta_2(t)|^2 dt}$, where $|\cdot|$ is the standard Euclidean norm in $\mathbb{R}^d$. However, this metric is not invariant to reparameterizations because the action of Γ is not distance-preserving, that is, $\|\beta_1 - \beta_2\| \neq \|\beta_1 \circ \gamma - \beta_2 \circ \gamma\|$. To fix this, Srivastava et al. (2011) instead consider a notion of distance between SRVFs.

Definition 2.1. Let $\dot{\beta}$ be the time-derivative of β. The *SRVF* of an absolutely continuous curve $\beta : \mathcal{D} \to \mathbb{R}^d$ is defined as

$$q(t) = \begin{cases} \frac{\dot{\beta}(t)}{\sqrt{|\dot{\beta}(t)|}} & \text{if } \beta \text{ is differentiable at } t \text{ and } |\dot{\beta}(t)| \neq 0 \\ 0 & \text{otherwise} \end{cases}.$$

There are several benefits to representing a curve by its SRVF. First, q encodes the instantaneous direction $\left(\frac{\dot{\beta}(t)}{|\dot{\beta}(t)|} = \frac{q(t)}{|q(t)|}\right)$ and speed ($|\dot{\beta}(t)| = |q(t)|^2$) of β. In fact, given the assumption of absolute continuity and a starting point $\beta(0)$, there is a smooth bijective mapping between q and β (Robinson 2012) given by $\beta(t) = \beta(0) + \int_0^t q(s)|q(s)|ds$. Notice that the SRVF is automatically invariant to translation due to the sole dependence on $\dot{\beta}$. This representation is also useful because it is valid for both open and closed curves of any dimension d. Most importantly, the $\mathbb{L}^2$ distance between SRVFs is preserved under common reparameterization and is equivalent to the *elastic metric* between the original curves, which measures the amount of bending and stretching required to deform one shape into another (see Younes 1998; Srivastava et al. 2011; Kurtek et al. 2012 for more details).

To formally compare two generic shapes, one must account for the variability due to shape-preserving transformations. Scale invariance is imposed by rescaling the original curves to unit length. Then, using the $\mathbb{L}^2$ distance between SRVFs, one seeks a rotation and reparameterization, which optimally register one curve to the other. This process is discussed in detail in Kurtek et al. (2012); finding the optimal reparameterization allows for flexibility in matching shape features, and thus underlies the *elastic* property of the method. Figure 1 shows a comparison of nonelastic and elastic deformations between two half circle shapes; the indentation of the bottom line segment in these two shapes is located on opposite sides. The nonelastic approach does not match the indentation feature across the two shapes; as a result, this prominent feature nearly disappears along the deformation path. The elastic path is more natural, as the indentation is simply shifted to the right due to the improved matching of this particular feature. In this article, we use the $\mathbb{L}^2$ distance between SRVFs to define the likelihood function.

3. Detection of a Fixed Number of Landmarks

We first assume that the number of landmarks in a population of shapes is known. In this section, we present a Bayesian model and associated methodology for sampling from the posterior distribution over the landmark locations. In Section 4, we extend the model to the case where the appropriate number of landmarks is unknown.

3.1. Model Specification for Shape Data

First consider the problem of identifying a fixed number, k, of landmarks, with domain locations $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k) \in \mathcal{D}^k$ for a population of shapes, subject to the constraint $\theta_1 < \cdots < \theta_k$. We assume the sample of M shapes is drawn from a homogeneous population, meaning that an arc-length parameterization is sufficient and shape registration is not necessary; in cases when this assumption does not hold, one can first perform multiple registration using the elastic metric (Srivastava et al. 2011). Let $\beta_1, \ldots, \beta_M : \mathcal{D} \to \mathbb{R}^2$ be a sample of curves formed from the outlines of these M objects. Our goal is to infer one set of landmark locations simultaneously using the sample of M curves. As a preprocessing step, we rescale the original curves to unit length to give equal weight to each shape in the sample.

We first consider reconstruction of shapes, using landmark locations, as illustrated in Figure 2. An appropriate set of landmarks can be defined as a set for which a linear reconstruction through the landmarks closely matches the original curve

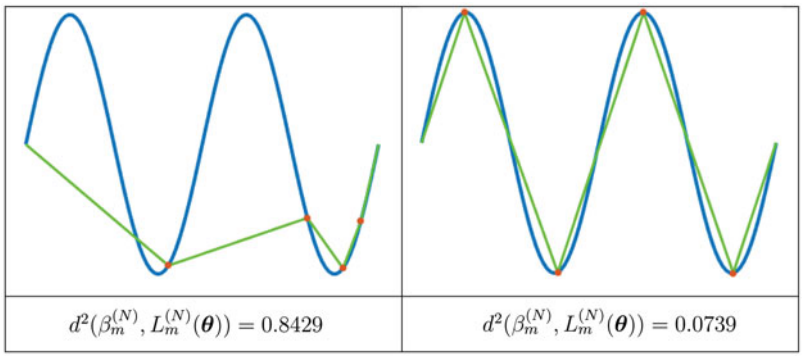


Figure 2. Poor (left) and good (right) linear reconstructions $L_m(\boldsymbol{\theta})$ (green) of β_m (blue), with squared reconstruction errors reported below. The landmarks $\boldsymbol{\theta}$ are shown as red dots.

(as shown in the left panel); the likelihood should place high probability on these configurations. More formally, we choose reconstructions of the mth curve in the sample, β_m, via a linear interpolation constructed by piecewise segments $L_m(t; \theta_i, \theta_{i+1})$ passing through the landmarks. The expression for the linear interpolator segment between θ_i and θ_{i+1} for $i = 1, \ldots, k-1$ is

$$L_m(t; \theta_i, \theta_{i+1}) = \left(1 - \frac{t-\theta_i}{\theta_{i+1}-\theta_i}\right)\beta_m(\theta_i) + \left(\frac{t-\theta_i}{\theta_{i+1}-\theta_i}\right)\beta_m(\theta_{i+1}), \quad \theta_i \le t < \theta_{i+1}. \tag{3.1}$$

Note that Equation (3.1) only gives an expression for the segment between consecutive landmarks; this does not define the entire interpolator. For open curves, the linear reconstructions must additionally connect to $\beta_m(0)$ and $\beta_m(1)$. Thus, two additional segments are required: (1) one connecting $\beta_m(0)$ and $\beta_m(\theta_1)$, and (2) one connecting $\beta_m(\theta_k)$ and $\beta_m(1)$. The expression for these segments is identical to Equation (3.1), treating the starting point of these segments as θ_i and ending point of these segments as θ_{i+1}. For closed curves, $\beta_m(0) = \beta_m(1)$, so the linear reconstruction must also be closed. This is guaranteed by forming a segment connecting $\beta_m(\theta_k)$ and $\beta_m(\theta_1)$, again using Equation (3.1). For the remainder of the article, we suppress the input t and let $L_m(\boldsymbol{\theta})$ be the full linear interpolator for the mth curve, constructed by joining these piecewise linear segments.

We observe a random sample of M shapes $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_M)^\top$, with corresponding SRVFs $q_{\beta_1}, \ldots, q_{\beta_M}$. Given landmark locations $\boldsymbol{\theta}$, we obtain the SRVFs of their linear reconstructions $q_{L_1(\boldsymbol{\theta})}, \ldots, q_{L_M(\boldsymbol{\theta})}$, and model the SRVFs $q_{\beta_1}, \ldots, q_{\beta_M}$ of our shape data as follows

$$q_{\beta_m} = q_{L_m(\boldsymbol{\theta})} + \varepsilon_m, \quad m = 1, \ldots, M, \tag{3.2}$$

where ε is a Gaussian process on $\mathcal{D}$ with zero mean function and covariance function C. Since the SRVF is related to the instantaneous velocity of the curve, it is reasonable to use a Gaussian process, as both negative and positive values of $q_{\beta_m} - q_{L_m(\boldsymbol{\theta})}$ are equally likely at each time (and one would not expect a negative value to be more likely than a positive value). Thus, the likelihood in terms of SRVFs is

$$q_{\beta_m} \mid q_{L_m(\boldsymbol{\theta})}, \boldsymbol{\theta} \sim \mathcal{GP}\left(q_{L_m(\boldsymbol{\theta})}, C\right). \tag{3.3}$$

C is a covariance function, chosen to reflect variability between q_{β_m} and $q_{L_m(\boldsymbol{\theta})}$. Since the likelihood is specified in terms of the SRVFs rather than the original curves, spatial dependence is difficult to model in the general case. Generalization of C to incorporate this spatial dependence or desired smoothness is left as future work.

3.1.1. Likelihood

For a curve β_m with reconstruction $L_m(\boldsymbol{\theta})$, we first compute the corresponding SRVFs q_{β_m} and $q_{L_m(\boldsymbol{\theta})}$. Assume that q_{β_m} and $q_{L_m(\boldsymbol{\theta})}$ are discretized using N points (call these *discretization points*). The number of discretization points is typically large, chosen externally, and assumed fixed in our model. We assume that it is defined by the resolution of the imaging device (e.g., a magnetic resonance imaging scanner) used to capture the shape of the object. Decreasing N can result in loss of shape features, while increasing N may introduce additional noise. More details on the choice of the number of discretization points for simulated data is provided in the supplementary material. We define the reconstruction error as

$$d(\beta_m^{(N)}, L_m^{(N)}(\boldsymbol{\theta})) = \left|\mathrm{vec}(q_{\beta_m}^{(N)} - q_{L_m(\boldsymbol{\theta})}^{(N)})\right|, \tag{3.4}$$

where vec is the vectorize operator which forms a vector of size $2N$ by vertically stacking the x and y coordinates of the SRVFs, $|\cdot|$ is the Euclidean norm in $\mathbb{R}^{2N}$, and the superscript (N) denotes the given function discretized using N points. (In general, if we are dealing with a d-dimensional curve, the vectorize operator will form a vector of size dN.) A small value of $d(\beta_m^{(N)}, L_m^{(N)}(\boldsymbol{\theta}))$ indicates that the landmarks $\boldsymbol{\theta}$ yield an accurate reconstruction of the mth curve, and the landmarks approximate the full object well. Figure 2 shows two landmark-based linear reconstructions of a simulated curve defined in Section 5; the left one results in a large error, as chosen landmarks do not provide a faithful reconstruction of the original curve. The right reconstruction is visually much better, with a much smaller error.

Discretizing Equation (3.3), and choosing a covariance matrix, we obtain,

$$\mathrm{vec}\left(q_{\beta_1}^{(N)} - q_{L_1(\boldsymbol{\theta})}^{(N)}\right), \ldots, \mathrm{vec}\left(q_{\beta_M}^{(N)} - q_{L_M(\boldsymbol{\theta})}^{(N)}\right) \mid \boldsymbol{\theta}, \kappa \overset{\text{iid}}{\sim} N\left(0_{2N}, \frac{1}{2\kappa} I_{2N}\right), \tag{3.5}$$

where $\kappa = \frac{1}{2\sigma^2}$ is a precision parameter. The likelihood function for the data $\boldsymbol{\beta}^{(N)} = (\beta_1^{(N)}, \ldots, \beta_M^{(N)})^\top$ is then given by

$$f(\boldsymbol{\beta}^{(N)}|\boldsymbol{\theta}, \kappa) = \pi^{-NM}\kappa^{NM} \exp\left(-\kappa \sum_{m=1}^{M} d^2(\beta_m^{(N)}, L_m^{(N)}(\boldsymbol{\theta}))\right), \tag{3.6}$$

where $d(\beta_m^{(N)}, L_m^{(N)}(\boldsymbol{\theta}))$ is defined in Equation (3.4). For a d-dimensional curve, the normal model is still appropriate, where dN replaces $2N$ in the mean and covariance. Notice that $\mathbb{L}^2$ distance between *SRVFs of curves* is used, rather than the $\mathbb{L}^2$ distance between coordinates of the curves. This likelihood model is similar to that of Cheng, Dryden, and Huang (2016) and Kurtek (2017), where it was used for registration of functional data.

3.1.2. Priors

Next, we specify prior distributions on κ and $\boldsymbol{\theta}$. Prior independence is often a reasonable assumption when the relationship between parameters is unknown a priori, for example, Bernardo and Smith (2008), so that $\pi(\kappa, \boldsymbol{\theta}) = \pi(\kappa)\pi(\boldsymbol{\theta})$. Because κ is a precision parameter (and a nuisance parameter), we choose a prior that is conditionally conjugate under our normal model (see, e.g., Gelman et al. 2004):

$$\kappa \sim \text{Gamma}(a, b). \tag{3.7}$$

Prior specification for $\boldsymbol{\theta}$ must enforce the ordering constraint on its components. To simplify this task, we transform $\boldsymbol{\theta}$ to a vector of consecutive differences between landmarks, denoted by $\boldsymbol{s}$. The dimension of $\boldsymbol{s}$ depends on whether we are detecting landmark locations on open or closed curves. For open curves, we define the components $s_i = \theta_{i+1} - \theta_i$ for $i = 1, \ldots, k-1$; we set $s_0 = \theta_1$ and $s_k = 1 - \theta_k$ (as the linear reconstruction is required to pass through the start and endpoints of the curve), and let $\boldsymbol{s} = (s_0, s_1, \ldots, s_k)$, which is $(k+1)$-dimensional. For closed curves, the components are still defined as $s_i = \theta_{i+1} - \theta_i$ for $i = 1, \ldots, k-1$; however, we let $s_k = (\theta_1 - \theta_k) \bmod 1$. There is a one-to-one correspondence between $\boldsymbol{\theta}$ and $\boldsymbol{s}$ for open curves; for closed curves, if a starting point along the curve is designated, then a one-to-one correspondence is also achieved. Thus, we proceed by using $\boldsymbol{s}$, and then recover $\boldsymbol{\theta}$ for inferential purposes. For notational simplicity, any notation which depends on $\boldsymbol{\theta}$ may also be written to depend on $\boldsymbol{s}$ instead (i.e., $d(\beta_m^{(N)}, L_m^{(N)}(\boldsymbol{\theta}))$ is equivalent to $d(\beta_m^{(N)}, L_m^{(N)}(\boldsymbol{s}))$). Notice that $\sum_i s_i = 1$ and $s_i > 0$ for all i for both open and closed curves, and by construction, $\boldsymbol{s}$ does not require the ordering constraint on its components. Thus, a natural prior for $\boldsymbol{s}$ is the Dirichlet distribution with concentration parameter α:

$$\boldsymbol{s} \sim \text{Dir}(\alpha \mathbf{1}), \tag{3.8}$$

where $\mathbf{1}$ is a vector of ones with dimension equal to $k+1$ for open and k for closed curves.

One must also choose appropriate values of prior hyperparameters. In the case of κ, we desire a diffuse prior to reflect our prior uncertainty in landmark error precision; thus, we select $a = 1$, $b = 0.01$. Another common choice for precision parameters is to set $a = b = \epsilon$ with $\epsilon \to 0$ (Gelman et al. 2004). In Section 5.3, we show that this choice, with $\epsilon = 0.01$, is appropriate. The hyperparameter for the prior on $\boldsymbol{s}$ is chosen to favor landmarks which are equally spaced to reflect our prior uncertainty about landmark locations. Setting $\alpha = 1$ is equivalent to a uniform prior on the simplex whose dimension is equal to the number of landmarks. This class of priors is commonly used for modeling multinomial distributions (Gelman et al. 2004).

3.1.3. Posterior

The density of the posterior distribution over landmark spacings $\boldsymbol{s}$ given the data $\boldsymbol{\beta}^{(N)}$ is denoted by $\pi(\boldsymbol{s}|\boldsymbol{\beta}^{(N)})$. The precision parameter κ is not of direct interest to us (since our goal is to solely infer landmark locations), so we compute the marginal likelihood,

$$\begin{aligned} f(\boldsymbol{\beta}^{(N)}|\boldsymbol{s}) &= \int_{\mathbb{R}_+} f(\boldsymbol{\beta}^{(N)}|\boldsymbol{s}, \kappa)\pi(\kappa)d\kappa \\ &= \frac{\pi^{-NM}\Gamma(a+NM)b^a}{\Gamma(a)\left(b + \sum_{m=1}^{M} d^2(\beta_m^{(N)}, L_m^{(N)}(\boldsymbol{s}))\right)^{a+NM}}, \end{aligned} \tag{3.9}$$

and use it to obtain the posterior density,

$$\pi(\boldsymbol{s}|\boldsymbol{\beta}^{(N)}) \propto f(\boldsymbol{\beta}^{(N)}|\boldsymbol{s})\pi(\boldsymbol{s}). \tag{3.10}$$

Note that if κ is of interest to the researcher, then the algorithm described in Section 3.2 can be implemented, with an additional Gibbs step to sample from the full conditional of κ. Posterior samples of $\boldsymbol{s}$ can be transformed to $\boldsymbol{\theta}$ for visualization, as described in Section 3.1.2. As posterior functionals of interest are not analytically tractable, inference will be based on approximations computed from MCMC samples. Section 1 of the supplementary material provides an alternative approach based on importance sampling. While MCMC is more appropriate for this model due to poor scaling of importance sampling with respect to the number of curves and discretization size, importance sampling can still be useful for quick posterior estimation of landmark location medians and maxima a posteriori (MAP).

3.2. Posterior Sampling via MCMC

In this section, we describe sampling strategies for obtaining estimates when posterior functionals of interest are not available analytically. For posterior inference on the locations of a fixed number of landmarks, we use the random walk Metropolis algorithm. We initialize the algorithm by sampling $\boldsymbol{s}^{[0]}$ from $\pi(\boldsymbol{s})$. The superscript $^{[t]}$ will denote the state of the Markov chain at iteration t. For a given MCMC iteration t, a proposal vector of landmarks is generated by selecting the jth component of the landmark vector $\boldsymbol{\theta}^{[t]}$ and applying a symmetric proposal distribution h. In our implementation, h is a normal probability density function with mean $\theta_j^{[t]}$ and variance v. This symmetric proposal mechanism yields a proposed vector $\boldsymbol{s}^*$. The Metropolis acceptance ratio is then defined as

$$\alpha(\boldsymbol{s}^{[t]}, \boldsymbol{s}^*) = \frac{\pi(\boldsymbol{s}^*|\boldsymbol{\beta}^{(N)})}{\pi(\boldsymbol{s}^{[t]}|\boldsymbol{\beta}^{(N)})} = \frac{f(\boldsymbol{\beta}^{(N)}|\boldsymbol{s}^*)\pi(\boldsymbol{s}^*)}{f(\boldsymbol{\beta}^{(N)}|\boldsymbol{s}^{[t]})\pi(\boldsymbol{s}^{[t]})}. \tag{3.11}$$

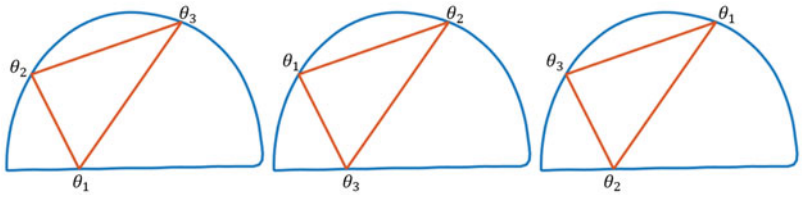


Figure 3. Example of the identifiability issue encountered when detecting landmarks on closed curves. Our model deems these three landmark configurations as equivalent.

The proposal $\boldsymbol{s}^*$ is accepted with probability $\min\{1, \alpha(\boldsymbol{s}^{[t]}, \boldsymbol{s}^*)\}$. As this procedure only updates one parameter at a time, convergence can require a large number of iterations; this is feasible because the likelihood is easy to evaluate. The algorithm is monitored for convergence, and approximate posterior samples are obtained after a suitable burn-in period with a thinning step to reduce autocorrelations. Convergence is assessed both visually (monitoring trace plots, autocorrelations, and acceptance rates) and via the Gelman–Rubin diagnostic (Gelman et al. 2004); see Section 3 in the supplementary material for more details.

Care must be taken in selecting proposal mechanisms to ensure that the Markov chain traverses the parameter space efficiently; this is important in the specified model, as the posterior can often be multimodal. In all cases, samples which violate the ordering assumption of $\boldsymbol{\theta}$ are automatically rejected; this is evident in the specification of the prior on $\boldsymbol{s}$. Care must also be taken when dealing with closed curves: since there is no designated start or endpoint, proposals must be allowed to wrap around the circular curve domain. We discuss the implementation for closed curves in more detail in the next section.

3.3. Implementation for Closed Curves

The issue of identifiability arises when dealing with closed curves since the circular domain has no natural start/endpoint. Thus, we must find a point in the domain that can be identified with $t = 0$. Figure 3 illustrates this with $k = 3$ landmarks for a half-circle; the reconstruction is invariant to how the landmarks θ_1, θ_2, θ_3 are labeled. This means that the model is exchangeable with respect to the ordering of $\boldsymbol{\theta}$.

To address this issue, we designate a reference point with parameter value θ_0 to be the point of maximal curvature on the first shape in the sample. Then, we preprocess the entire collection of curves as follows. We shift the order of points for the first curve such that $t = 0$ is identified with θ_0. Then, all subsequent curves are aligned to the first curve by finding the ordering of points which minimizes the SRVF distance to the first curve. Note that this is not the same as a full registration step in the context of elastic shape analysis; instead, we are systematically defining a starting point on each curve, since it is not well-defined for a given sample of closed curves. To visualize the posterior samples for closed curves, we post-process them to lie on the rescaled unit circle by eliminating the boundary between $\theta = 0$ and $\theta = 1$ as follows. We align all n posterior samples according to the set of locations for the first posterior sample $\boldsymbol{\theta}_1$. That is, for samples $\boldsymbol{\theta}_i$ for $i = 2, \ldots, n$, we compute the distance between corresponding components $j = 1, \ldots, k$ using the circular metric $d^{[j]}(\boldsymbol{\theta}_i, \boldsymbol{\theta}_1) = \min\{|\theta_i^{[j]} - \theta_1^{[j]}|, |\theta_i^{[j]} - 1 - \theta_1^{[j]}|, |\theta_i^{[j]} + 1 - \theta_1^{[j]}|\}$. Then, we find the landmark ordering (using circular permutations) which minimizes $d(\boldsymbol{\theta}_i, \boldsymbol{\theta}_1) = \sum_{j=1}^{k} d^{[j]}(\boldsymbol{\theta}_i, \boldsymbol{\theta}_1)$. This alignment process is done using the full configuration of $\boldsymbol{\theta}$. We have found this procedure to be robust to the choice of the first posterior sample based on many simulations and real data examples. An alternative approach is to jointly compute the posterior mean of all samples while modding out by $\mathbb{S}^1$, which we will explore as future work.

4. Estimation of Number and Location of Landmarks

Addressing how many landmarks to select on a given set of curves is a complex task and is akin to a model selection problem found in many facets of statistics, where "the number of things you don't know is one of the things you don't know" (Richardson and Green 1997). There are generally two ways in which statisticians approach this problem. One way is to develop a criterion which must be optimized while making sure to not "overfit" the model. The other is to treat the number of parameters as unknown and infer it from the data. In this section, we discuss two such approaches for selecting k, the number of landmarks.

4.1. Distance-Based Criterion

The criterion-based method for selecting k is borrowed from dimension reduction problems for high-dimensional data. One example of this is principal component analysis (PCA); it forms a much lower-dimensional space of uncorrelated modes of variation, which are first ordered by proportion of variability explained. The number of components is selected by choosing a cut-off where, at a certain point, the percentage of variation begins to "level off," as including additional components becomes unnecessary. Typically, the number of components selected is chosen based on the "elbow" of the plot of percent variation versus the number of components (known as a scree plot).

We use a similar approach to choose the number of landmarks k. For each value of k considered, we draw posterior samples $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_n$ (n is the posterior sample size of $\boldsymbol{\theta}$, each of which is a k-dimensional vector). For a posterior sample indexed by i, we form the linear reconstructions $L_{1i}^{(N)}, \ldots, L_{Mi}^{(N)}$ (for the M curves in the data) and compute the average cumulative squared distance $d_k^2 = \frac{1}{n}\sum_{i=1}^{n}\sum_{m=1}^{M} d^2(\beta_m^{(N)}, L_{mi}^{(N)}(\boldsymbol{s}))$. We repeat this procedure for multiple values of k and plot the resulting d_k^2; we expect d_k^2 to decrease as k increases, since linear reconstructions

improve as the number of landmarks increases. Then, k can be chosen at the "elbow" of this curve, which is the point at which including additional landmarks does not lead to a substantial reduction in reconstruction errors.

The distance criterion is intuitive, but is not without issues. The goal of building this model is to automatically select the number of landmarks without any manual selection. Using a plot of d_k^2 requires the user to identify the "elbow," and select that value as the desired number of landmarks. This choice may not always be immediately obvious; it certainly could be automated, but this could lead to faulty conclusions (e.g., the right panel of Figure 6). It also leaves the user unable to quantify uncertainty in choosing k, and requires performing the analysis many times with different values of k, which is computationally demanding. Thus, it may instead be better to let k be unknown and build it into the Bayesian model.

4.2. Extension of the Landmark Detection Model to Unknown k

Unknown k is considered by conditionally specifying an additional level in the Bayesian hierarchical model. The likelihood, now $\boldsymbol{\beta}^{(N)}|\boldsymbol{s}, k$, is identical to the likelihood $\boldsymbol{\beta}^{(N)}|\boldsymbol{s}$ in Section 3.1.1 (after marginalizing over the prior on κ, which is assumed independent of k). The prior on the locations $\boldsymbol{s}$ and their number k is specified as

$$\pi(\boldsymbol{s}, k) = \pi(\boldsymbol{s}|k)\pi(k). \tag{4.1}$$

The prior $\boldsymbol{s}|k$ still follows a Dirichlet distribution (as described in Section 3.1.2), where the dimension of the concentration parameter vector depends on k. We must specify a prior on k as well. Note that for open curves, $k \geq 1$ (to get a valid reconstruction); however, for closed curves, there is no start or endpoint, so $k \geq 3$. To account for this, we choose as the prior for k a shifted Poisson distribution: we assume $k = \nu + 1$ and $k = \nu + 3$ for open and closed curves, respectively. Then, the prior on ν is given by

$$\nu \sim \text{Poisson}(\lambda). \tag{4.2}$$

We justify the use of an informative Poisson prior by reminding the reader that our goal is to select a fairly low-dimensional landmark set; by setting λ to be small, this prior places high mass on small values of k. The shift guarantees that prior probabilities are greater than zero for the appropriate values of k only. Selection of λ is an interesting and difficult problem. We view λ as a regularization parameter. Because the likelihood only depends on reconstruction error, adding more landmarks will generally strengthen the likelihood relative to the prior. Thus, to avoid overfitting, the prior on k can be chosen to place most of its mass very close to zero to penalize choosing high values of k. Varying λ will therefore yield a path of posterior probability distributions; the dependence of posterior inference on the choice of λ is shown in Section 5.4.

4.3. Posterior Sampling Using Reversible Jump MCMC

Treating k as unknown complicates posterior inference on $\boldsymbol{s}, k|\boldsymbol{\beta}^{(N)}$ due to the dependence of the dimensionality of $\boldsymbol{s}$ on k: different values of k result in a different number of landmark location parameters to infer. Standard MCMC methods are defined on parameter spaces of fixed dimension. Dependent proposals between parameter spaces of different dimension k can be made via reversible jump MCMC (Green 1995; Richardson and Green 1997; Grenander and Miller 1994, referred to as RJMCMC). This type of procedure is commonly used in model selection problems, where one wants to infer model parameters as well as their number. In particular, the birth–death form of RJMCMC proposes a new parameter vector by first randomly choosing to increase the dimension of the parameter space by one (a birth), decrease the dimension by one (a death), or keep the dimension the same (a stay). In the case of a birth, a new component is added to the model according to a chosen distribution. Similarly, a component is "killed" through random selection. This extra step of selecting a move type and developing the proposal based on the selected move is accounted for in the acceptance probability of a proposal. Section 2 of the supplementary material provides a detailed RJMCMC procedure for the proposed automatic landmark detection model.

5. Model Assessments via Simulations

5.1. Detection of a Fixed Number of Landmarks

To test various properties of the proposed model, we construct a simple shape based on a sine curve with well-defined peaks and valleys. Consider the curve $\beta(t) = [t, \sin(4\pi t)]^\top$, $t \in [0, 1]$, with two peaks and two valleys, each of which appear to be optimal locations for landmark placement due to low reconstruction error, yielding a total of $k = 4$ landmarks. We first begin by drawing posterior samples under the fixed $k = 4$ model with $N = 200$ discretization points. We use the random walk Metropolis algorithm as described in Section 3.2, specifying $a = 1$, $b = 0.01$ in the prior for κ, $\alpha = 1$ in the prior for $\boldsymbol{s}$, and proposal variance $\nu = 0.02$. The chain is run for 10^6 iterations; the first 10% is discarded as burn-in, and the remaining sample is thinned by 100 to reduce autocorrelations. Trace plots used to diagnose convergence for this example are shown in Section 3 of the supplementary material; a table of the Gelman–Rubin diagnostic is also included, suggesting convergence with values close to one. A discussion of choice of N for simulated data, which in principle can be constructed at any resolution, is given in Section 4 of the supplementary material. Computing time was 1811.8 sec (MATLAB R2017b on ASUS F555UA-EH71 laptop with 8 GB RAM and Intel Core i7 processor).

The top left panel of Figure 4 shows the original curve with samples from the marginal posteriors of landmark locations plotted on top. The posterior samples from $\boldsymbol{\theta}|\beta^{(N)}$ obtained using our model coincide with the peaks and valleys of β as expected. Posterior uncertainty for each landmark is illustrated in the density plots in the bottom of Figure 4. Each density is fairly concentrated, indicating high confidence in identifying the four landmark locations. Standard posterior summaries can also be computed for $\boldsymbol{\theta}$. The top right panel of Figure 4 shows the mean, median, MAP, and 95% credible intervals for each component of $\boldsymbol{\theta}$. The mean and median are very similar; the MAP estimate is a little bit different, due to the complex dependencies

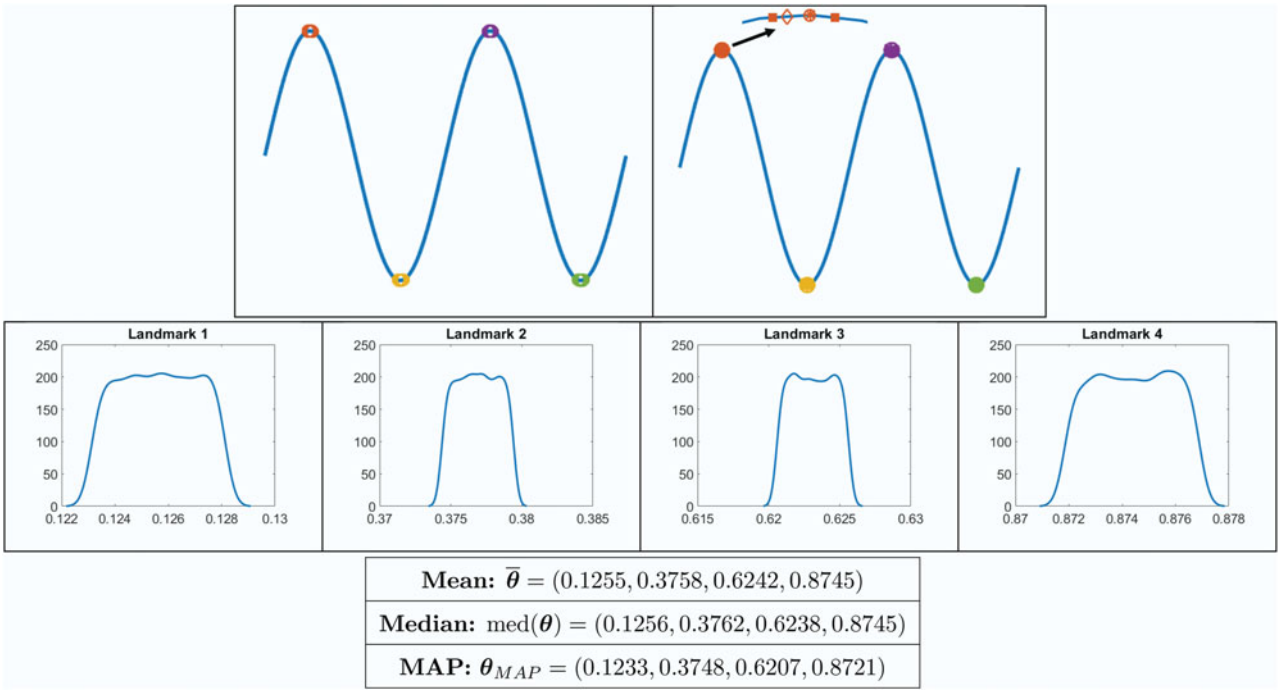


Figure 4. Top: Curve β with posterior sample of landmark locations (left, red = 1, yellow = 2, purple = 3, green = 4) and posterior landmark summaries (right, circle = mean, asterisk = median, diamond = MAP, squares = 95% credible interval). Bottom: Density plots of marginal posterior samples $\theta_i|\beta^{(N)}$, $i = 1, \ldots, 4$, and table with posterior summaries of $\boldsymbol{\theta}$.

in the landmark locations. The 95% credible intervals are narrow and disjoint, indicating precisely estimated landmark locations.

5.2. Model Invariance to Shape-Preserving Transformations

It is important to check that our inference in Section 5.1 is invariant to shape-preserving transformations, which include translation, scaling, and rotation; reparameterization is not considered here, as the given curves are always sampled using arc-length due to the population homogeneity assumption. Our models are automatically invariant to translations due to the model's dependence on the SRVF only, which is translation invariant as it is defined using the derivative of β. A rescaling of the curve should also result in no change to inference, as curves are preprocessed to have unit length. Figure 5 confirms this for β scaled by a factor of two; the resulting marginal posteriors look extremely similar to those of the original curve. The invariance to rotations is not immediately obvious. In Figure 5, we also demonstrate inference on a version of the original curve β which was rotated by 45° counter-clockwise. The marginal posteriors again appear to coincide with the original densities, and landmarks are located at the peaks and valleys as before. These experiments were conducted under the same settings as the example in Figure 4 and demonstrate that the proposed Bayesian model is invariant to shape-preserving transformations.

5.3. Sensitivity Analysis to Choice of Hyperparameters

As with any Bayesian analysis, studying the sensitivity of inference to the choice of hyperparameters is an important consideration. We assess the impact of hyperparameters a and b for the prior on the nuisance parameter κ. These two hyperparameters play prominent roles in the marginalized likelihood of $\boldsymbol{\beta}^{(N)}$. Our goal is to select priors that are weakly informative. In particular, since a is absorbed into an exponent in $f(\boldsymbol{\beta}^{(N)}|\boldsymbol{s})$ (which involves the number of points N and the number of curves M), as long as a is chosen to be small relative to NM, the posterior is not sensitive to changes in this parameter. However, the choice of b is more impactful, as large values of b will tend to dominate the linear reconstruction error term; this will result in a flattening of the posterior and greater variability in inference of $\boldsymbol{\theta}|\boldsymbol{\beta}^{(N)}$.

Table 1 shows marginal 95% posterior credible intervals of the components of $\boldsymbol{\theta}$ for the example described in Section 5.1 under different prior hyperparameter settings (as compared to the original setting of $a = 1$, $b = 0.01$). As expected, the intervals remain very similar when a is changed. In a similar fashion, as b is decreased toward 0, the credible intervals change very little from the ones obtained under the original setting. However, as b is increased, the prior becomes more informative. This increases the variance in the posterior and results in wider credible intervals for landmark locations. Nonetheless, as is evident in this table, our overall inference is robust to the choice of hyperparameters a and b.

5.4. Inference of Number and Location of Landmarks

In Section 5.1, we assumed $k = 4$ based on the number of significant features of β. However, it may be ideal to use fewer or more landmarks based on the reconstruction error. First, we select k using the distance criterion defined in Section 4.1. The left panel of Figure 6 shows the plot of the average cumulative

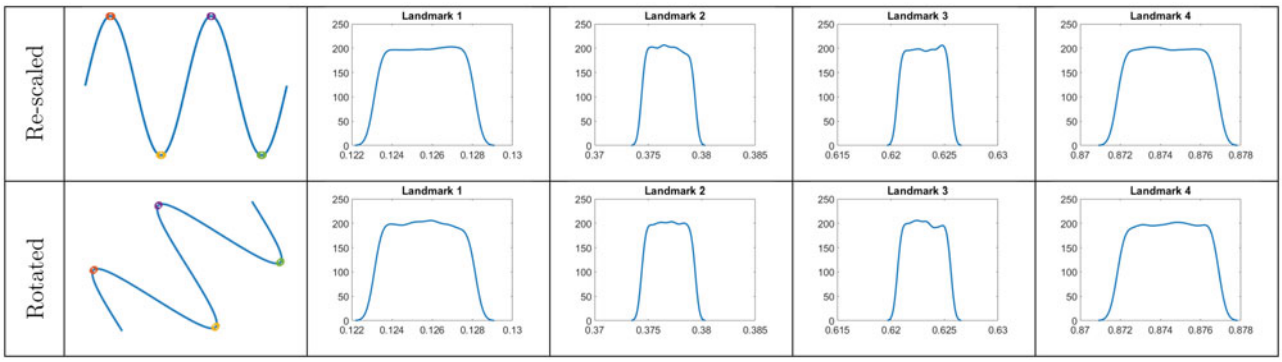


Figure 5. Left: Rescaled and rotated versions of β with posterior sample of landmark locations (red = 1, yellow = 2, purple = 3, green = 4). Right: Density plots of marginal posterior samples $\theta_i|\beta^{(N)}$, $i = 1, \ldots, 4$.

Table 1. 95% credible intervals for $\boldsymbol{\theta}|\beta^{(N)}$ under different choices of prior hyperparameters for κ, for the simulated curve example in Section 5.1.

a	b	θ_1	θ_2	θ_3	θ_4
1	0.01	(0.1215, 0.1280)	(0.3699, 0.3792)	(0.6208, 0.6297)	(0.8720, 0.8781)
0.01	0.01	(0.1217, 0.1280)	(0.3700, 0.3793)	(0.6208, 0.6300)	(0.8720, 0.8782)
0.1	0.01	(0.1219, 0.1280)	(0.3699, 0.3793)	(0.6208, 0.6301)	(0.8720, 0.8780)
3	0.01	(0.1227, 0.1280)	(0.3699, 0.3792)	(0.6208, 0.6300)	(0.8720, 0.8782)
5	0.01	(0.1226, 0.1280)	(0.3700, 0.3792)	(0.6208, 0.6301)	(0.8720, 0.8776)
1	0.0001	(0.1231, 0.1280)	(0.3700, 0.3793)	(0.6208, 0.6300)	(0.8720, 0.8769)
1	0.001	(0.1230, 0.1280)	(0.3700, 0.3792)	(0.6208, 0.6300)	(0.8720, 0.8769)
1	0.1	(0.1190, 0.1302)	(0.3697, 0.3793)	(0.6208, 0.6303)	(0.8697, 0.8810)
1	1	(0.1124, 0.1377)	(0.3629, 0.3876)	(0.6123, 0.6381)	(0.8621, 0.8882)

squared distance (d_k^2) as a function of k for $k = 1, \ldots, 10$; we used 100,000 iterations of MCMC with $a = 1$, $b = 0.01$, and $N = 100$ to generate posterior samples for each value of k. As expected, d_k^2 decreases as k increases due to the reduced reconstruction error. However, at $k = 4$, we observe a clear "elbow" after which the marginal utility of adding additional landmarks is diminished. Thus, it would appear reasonable to select $k = 4$ based on this criterion. However, as stated earlier, this process requires the user to identify this point on the curve, which may not always be obvious, and removes the automation in landmark detection. Computing time for this particular example was 1238.8 sec, using parallel computing for the MCMC chains run for all of the values of k considered; one would expect greater computational cost for more complex shapes, where it may be necessary to consider more than $k = 10$ landmarks.

Next, we treat k as a parameter and infer it using RJMCMC with the conditional model given in Section 4.3. We use Algorithm 1 (see Section 2 in the supplementary material) with the concentration parameter $\alpha = 1$ for the prior on $\boldsymbol{s}$. We select $v = 0.02$ for the variance of the normal proposal in the "Stay" step, and set $a = 1, b = 0.01$ for the prior on κ. Again, $N = 100$ discretization points are used. As stated in Section 4.2, varying λ (the prior parameter for k) changes the magnitude of the penalty on k. Thus, we present a path of posterior solutions for different values of λ, as listed in Figure 7. After running the algorithm for 100,000 iterations, we discard the first 10,000 iterations as burn-in, and take every 100th iteration to reduce autocorrelation and form the approximate posterior distribution. Convergence is diagnosed by monitoring acceptance rates, the log posterior, and examining trace plots of the parameters given values of k. The top panel of Figure 7 shows posterior histograms for $k|\beta^{(N)}$ for the different settings of λ. Computation time here was 462.0 sec with the same software and hardware, which is much shorter than the corresponding inference using the distance criterion.

As expected, as λ increases, the penalty for large values of k diminishes, and thus the posterior of $k|\beta^{(N)}$ is shifted toward higher values. Note that $\lambda = 10^{-6}$ yields a posterior mode which is consistent with the k obtained using the criterion-based approach. Posteriors with larger values of k tend to exhibit greater variability as well, since even miniscule differences in linear reconstruction error are rewarded when λ is not extremely small. Thus, controlling λ allows the user to select how detailed these linear reconstructions need to be to represent the given data: large λ will favor reconstructions which capture the majority of the high curvature points (i.e., small-scale details), while

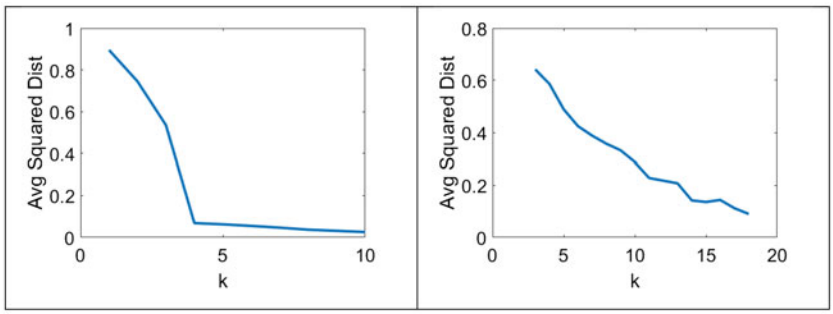


Figure 6. Plot of average cumulative squared distance versus k using MCMC sampling for the simulated curve from Section 5.1 (left) and the deer from Section 6.1 (right).

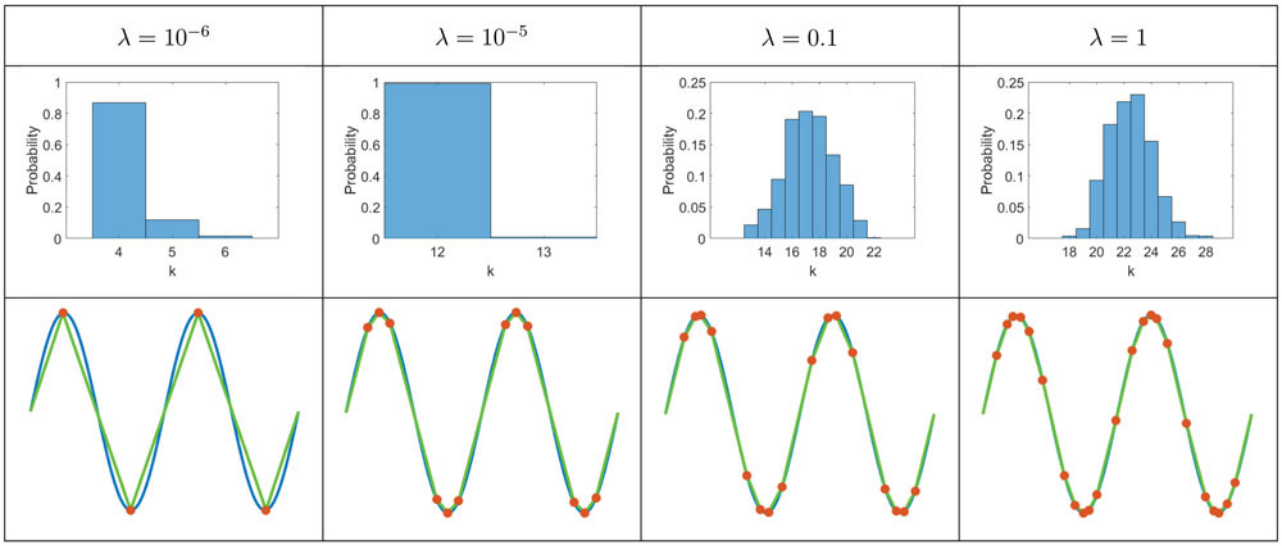


Figure 7. Inference on the number of landmarks and their locations for the example in Section 5.1. Top: Histograms of samples of $k|\beta^{(N)}$ for different settings of λ. Bottom: Conditional on the mode of the posterior of k, linear reconstructions (green) of β (blue) based on the mean configuration (red) under different values of λ.

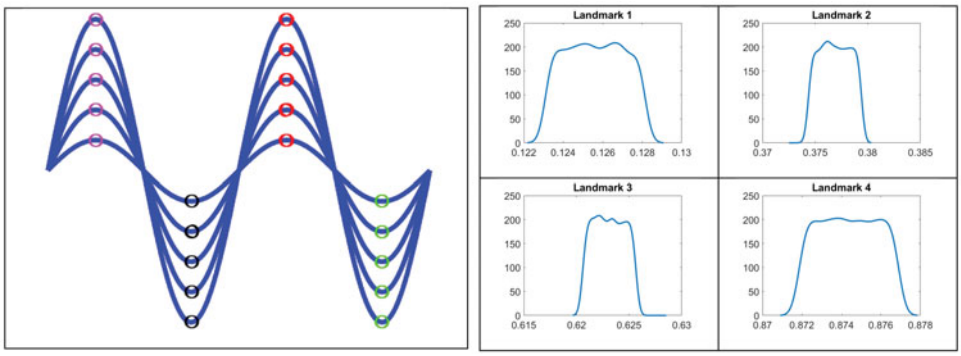


Figure 8. Left: Curves $\beta_1, \ldots, \beta_5$ with posterior sample of landmark locations (magenta = 1, black = 2, red = 3, green = 4). Right: Density plots of posterior samples $\theta_i|\boldsymbol{\beta}^{(N)}$, $i = 1, \ldots, 4$.

small λ aims for reconstructions which are more parsimonious and ignore smaller details of the shape. The magnitude of λ is dependent on the number of discretization points N and the sample size M; higher values of N or M require a "stricter" regularization (i.e., λ must be made much smaller to penalize the likelihood). This is shown in the bottom panel of Figure 7. Here, linear reconstructions of β are shown for values of k which exhibit high posterior probability under various settings of λ. The displayed reconstructions are formed from the posterior mean of $\boldsymbol{\theta}|\beta^{(N)}$ for the particular value of k. Notice that additional landmarks are placed around the detailed peaks and valleys, which are more crucial to the linear reconstruction than the other parts of β.

5.5. Inference Based on Multiple Curves

Most problems of landmark detection involve multiple curves; since the proposed model was specified for a general sample size M, this extension is simple. Consider a collection of $M = 5$ curves with $N = 200$ discretization points, each of which has two peaks and two valleys, but now with different heights: $\beta_m(t) = [t,\ m\sin(4\pi t)]^\top$, $t \in [0,1]$, $m = 1, \ldots, 5$. Our goal is to infer the locations of $k = 4$ landmarks as in Section 5.1. This set of landmarks is simultaneously inferred using the full set of five curves assuming independence across curves. The peaks and valleys in this random sample of curves occur at the same locations along each curve. Thus, we expect improved inference of landmark locations (i.e., a more precise estimation of landmarks) as compared to that in Section 5.1. However, this is not immediately obvious, because while the number of curves increased from the previous example, the cumulative linear reconstruction error will also increase. The proposal h (see Section 3.2) is chosen to be a normal density centered at the previous value of the chosen component with variance $v = 0.02$. We obtain 10^6 dependent samples via MCMC, and the approximate posterior is again formed by discarding the first 100,000 iterations for burn-in and thinning by every 100 iterations. Figure 8 shows the posterior sample of landmark locations plotted on all five curves, and density estimates for $\boldsymbol{\theta}$. These distributions are slightly more concentrated than those in Figure 4 as a result of the increased sample size.

6. Computer Vision and Medical Imaging Applications

6.1. Complex Shapes in Computer Vision

We present several examples of posterior landmark inference applied to complex shapes from the well-known MPEG-7

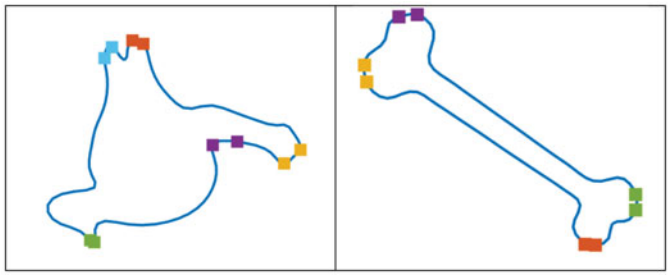


Figure 9. Posterior 95% credible intervals for landmarks on the bird (left) and bone (right). Colors for each component of θ are red = 1, yellow = 2, purple = 3, green = 4, blue = 5.

benchmark (*http://www.dabi.temple.edu/~shape/MPEG7/dataset.html*) computer vision. All of the examples involve closed curves; thus, we perform the pre and postprocessing steps described in Section 3.3.

The left panel of Figure 9 shows the outline of a bird, which contains features at different scales. The area around the feet complicates linear reconstructions using small landmark sets. We select $k = 5$ with the same MCMC settings as in Section 5.1, except with proposal variance $v = 0.04$. The 95% credible intervals are shown on the bird's outline. These intervals are very narrow in general and appear to capture the extreme points of the outline which help minimize the reconstruction error. Notice that the beak, which is an important feature but quite isolated from the other prominent features, has a very narrow credible interval. Clearly, this is an important structure that must be captured by the landmarks. The right panel of Figure 9 shows the credible intervals for a bone shape with $k = 4$. The credible intervals are again quite narrow, indicating that the four extreme points of the bone will yield a good linear reconstruction of the object. In this example, placing additional landmarks at the high absolute curvature points on the bone may be beneficial.

The MPEG-7 dataset features $M = 20$ observations of each type of shape. To examine joint inference for multiple shapes, we take a further look at posterior samples drawn for $M = 20$ bones with $k = 4$ landmarks. Figure 10 shows posterior landmark locations on the extrinsic mean (cross-sectional average at each of the N discretization points) of the 20 bones, as well as on each individual bone. Notice that the model still captures landmarks at the high absolute curvature points of the bone, even when there are abnormalities within an individual bone structure.

An interesting example to consider is the introduction of a second shape to joint posterior inference, where the second shape has a fairly different structure. Consider Figure 11; on the left are posterior locations of $k = 4$ landmarks on one half circle (i.e., $M = 1$). Notice the low variability in the landmarks on the base, and higher variability in landmarks on the curved portion of the shape, reflecting the difficulty of constructing a linear reconstruction which captures the shape's curvature. We introduce a second half circle with a large portion of the right side "missing," and perform posterior sampling with $M = 2$. This is shown in the right panel of the figure; clearly, the inference of the landmarks on the top portion of the shapes changes drastically. In particular, landmark 2 (yellow) shifts locations as compared to the left panel, and exhibits much lower variability. This is due to the large amount of curvature that occurs in the newly introduced shape, which forces a linear reconstruction to capture that particular feature.

For shapes which are even more complex, selecting the number of landmarks is not trivial. For the deer in Figure 12, the legs and antlers make it difficult for the researcher to select an appropriate number of landmarks heuristically; thus, it makes

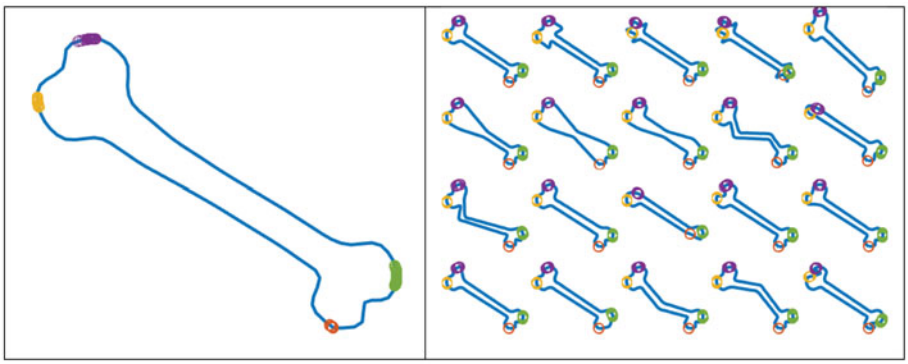


Figure 10. Posterior landmark locations plotted on the extrinsic sample mean of the 20 bones (left) and each individual bone (right); red = 1, yellow = 2, purple = 3, green = 4.

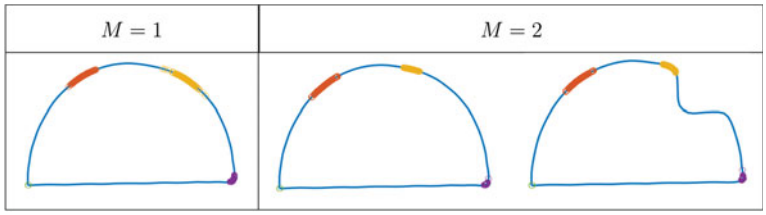


Figure 11. Posterior landmark inference for one (left) and two (right) half circles ($k = 4$). Note the change in inference for landmark 2; red = 1, yellow = 2, purple = 3, green = 4.

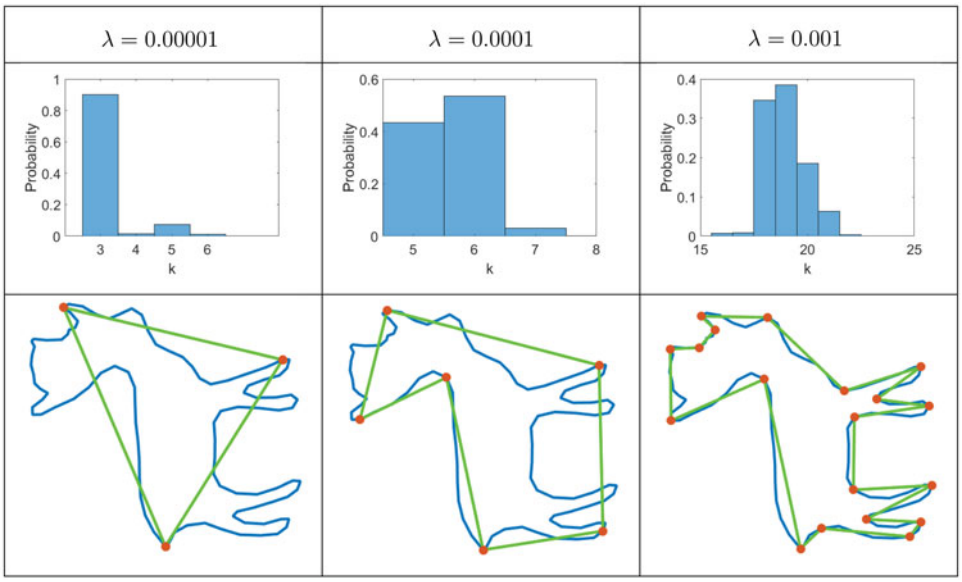


Figure 12. Top: Histograms of samples of $k|\beta^{(N)}$ for different settings of λ. Bottom: Conditional on the mode of the posterior of k, linear reconstructions (green) of β (blue) based on the median configuration (red) under different values of λ.

sense to assume k as unknown. We first attempt to estimate k for the deer outline using the distance criterion from Section 4.1; the result is shown in the right panel of Figure 6. While in the simulated open curve example there was a discernible point for which there was no benefit to adding more landmarks, there does not seem to be such a clear distinction in this case, even as we increase the number of landmarks to more than 15. As mentioned previously, this can happen, particularly with complex objects, because it is not as clear where, or how many, landmarks should be selected. This plot will begin to level off, as increasing the number of landmarks will certainly improve the linear reconstruction; it is possible that the "elbow" point has simply not occurred yet when $k = 18$. Computationally, this is extremely inefficient, and thus, it may make more sense to proceed with this problem by estimating k within the Bayesian model.

Next, we use RJMCMC to estimate the number of landmarks k on the deer example with $\alpha = 1$, $\nu = 0.05$, $a = 1$, and $b = 0.01$. As in the example in Section 5.4, Figure 12 shows posterior summaries of $k|\beta^{(N)}$ for different values of λ, which again acts as a regularizer. As expected, increasing λ shifts the marginal posterior of k to higher values and rewards better reconstructions over sufficiently small values of k. Linear reconstructions for the median configuration of landmarks are also shown in Figure 12. Notice that all three of the landmark configurations capture important features of the deer outline. In fact, our approach allows the user to control the number of landmarks selected on the shape of interest through an appropriate choice of λ. For complex shapes, such as the deer example given here, it may be beneficial to select more landmarks; on the other hand, for simpler shapes like the bone, a few landmarks are sufficient.

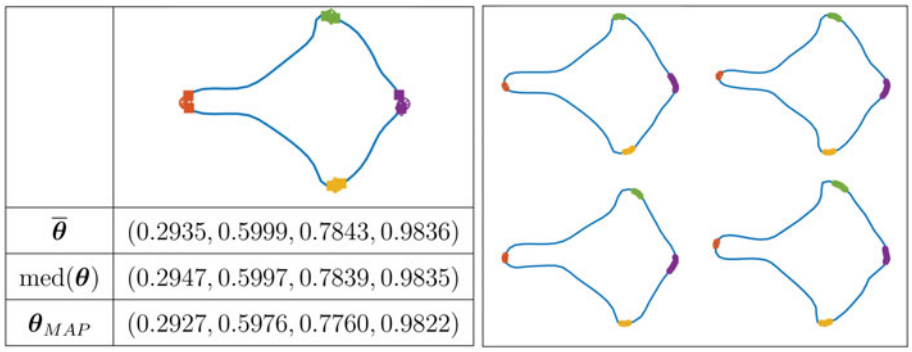
6.2. Mice Vertebrae

Anatomy is a useful application of automatic landmark detection, as existing approaches usually rely on expert knowledge. The second thoracic mice vertebrae exhibit differences in shape and size when mice are controlled for diet. In this section, we use data obtained from the R "shapes" package, as described in Dryden and Mardia (2016). Refer to Figure 3 of Strait et al. (2017) for a description of the anatomy of a mouse vertebra. Also, each data curve is sampled with the given $N = 61$ discretization points.

We begin by analyzing a single mouse vertebra; the outline appears to have four distinct landmark locations (corresponding to the neural spine, centrum, and transverse processes), so we assume $k = 4$ and proceed with the fixed k model. Once again, we perform random walk Metropolis, using the same settings as in Section 5.1. The resulting posterior summaries are shown in Figure 13. Notice that the 95% credible interval is quite narrow for estimating all of the landmarks, which appear to correspond to the anatomical meaningful neural spine, centrum, and transverse processes.

To see the impact of increased sample size on posterior inference, we also show results for $M = 4$ closed curves of mice vertebrae outlines. We use 10^6 iterations of MCMC with appropriate burn-in and thinning, and the same model parameters as in the $M = 1$ case. The posterior landmark sample locations are plotted on the right side of Figure 13. Notice that, similarly to the simulated example in Section 5.5, the variability in landmark locations is smaller when inference is based on $M = 4$ mice vertebrae rather than one; this is due to the dependence of the likelihood on the sum of reconstruction errors over all curves in the data. The estimated landmark locations appear to identify the four natural landmarks of the vertebrae as before.

This particular dataset features mice controlled for diet. We now examine posterior landmark locations for the full sample size ($M = 30$) of vertebrae from a subpopulation of mice which were not genetically selected for a large or small body weight (i.e., a control group of mice). We use the same MCMC settings as above (except with 10^5 iterations) and show the results in Figure 14. Due to the large number of samples, we



$\bar{\boldsymbol{\theta}}$	$(0.2935, 0.5999, 0.7843, 0.9836)$
$\text{med}(\boldsymbol{\theta})$	$(0.2947, 0.5997, 0.7839, 0.9835)$
$\boldsymbol{\theta}_{MAP}$	$(0.2927, 0.5976, 0.7760, 0.9822)$

Figure 13. Left: Mouse vertebra with the posterior mean, median and 95% credible intervals of $\boldsymbol{\theta}|\beta^{(N)}$ (red = 1, yellow = 2, purple = 3, green = 4; circle = mean, asterisk = median, diamond = MAP, squares = 95% credible interval). Right: Posterior sample landmark locations for $M = 4$ mice vertebrae.

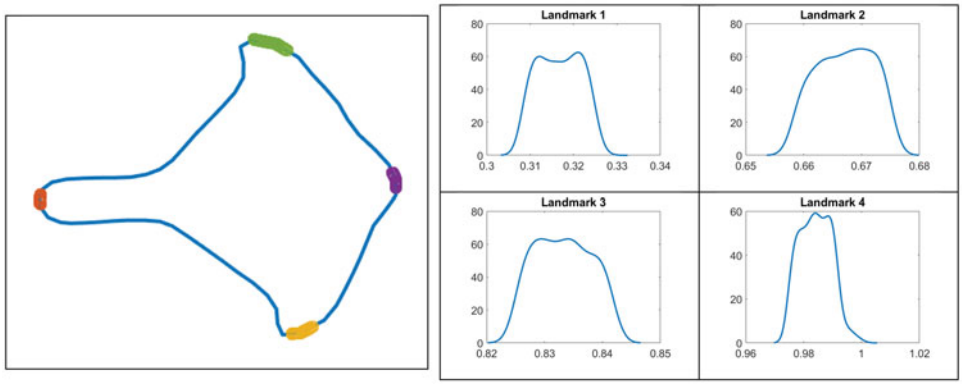


Figure 14. Left: Extrinsic mean curve $\bar{\boldsymbol{\beta}}$ for 30 control mice with posterior samples of landmark locations (red = 1, yellow = 2, purple = 3, green = 4). Right: Density plots of marginal posterior samples $\theta_i|\boldsymbol{\beta}^{(N)}$, $i = 1, \ldots, 4$.

show a plot of landmark locations on the extrinsic mean $\bar{\boldsymbol{\beta}}$ of the sample of 30 curves $\{\beta_1, \ldots, \beta_{30}\}$. As expected, landmarks are again identified at the same locations. Density plots of posterior samples are also shown in the figure.

6.3. *Brain Substructures From Magnetic Resonance Image Slice*

As mentioned in Section 1, a particular motivation for automated landmark detection arises in the field of medical imaging. Doctors are often required to manually annotate images of anatomical structures with important landmarks, which is tedious, subjective, and prone to human error. In this section, we apply our model to four different substructures (caudate, hippocampus, putamen, and thalamus) extracted from brain magnetic resonance images (MRI) of 10 different subjects. An example of a subject's original MRI slice, as well as outlines of substructures associated with all 10 subjects can be found in Kerr, Kurtek, and Srivastava (2011).

Figure 15 shows posterior landmark locations for the four substructures applied across the sample of $M = 10$ subjects; each observation was sampled with $N = 50$ points, which provided a fine resolution for landmark inference. We choose to demonstrate inference for fixed k, where $k = 3$ for the caudate, $k = 4$ for the hippocampus and thalmus, and $k = 5$ for the more structurally complex putamen. We ran MCMC for 10^6 iterations for each substructure, using a proposal variance of $\nu = 0.02$; the first 100,000 iterations were discarded for burn-in, and the remaining sample was thinned by every 100 iterations. Note the similar amount of variability for all landmarks in both the caudate and hippocampus. The putamen is interesting; with high posterior probability, three landmarks (yellow, purple, and green) are placed at the top of the structure and one landmark (red) is placed at the bottom. However, the blue landmark exhibits more variability, as it does not appear to be as necessary to the linear reconstruction as the other four landmarks. This procedure allows for automatic annotation of landmarks that doctors may otherwise be forced to do manually.

7. Discussion and Future Work

We defined a Bayesian model for inference of landmark locations given a random sample of shapes drawn from a homogeneous population. The benefits of this model include the ability to obtain automatic estimates of landmark locations along with measures of uncertainty, thereby eliminating the need for a researcher to manually annotate important features on shapes. We propose a hierarchical model for both the fixed-landmark and variable-landmark settings and describe methods for approximate sampling from the associated posterior distributions. In the variable dimension landmark setting, we discuss the impact of the regularization parameter λ on posterior inference on the number of landmarks k.

One challenge in our framework is the choice of the parameter λ in the prior distribution for the number of landmarks k in Section 4.2. As discussed and illustrated using various examples,

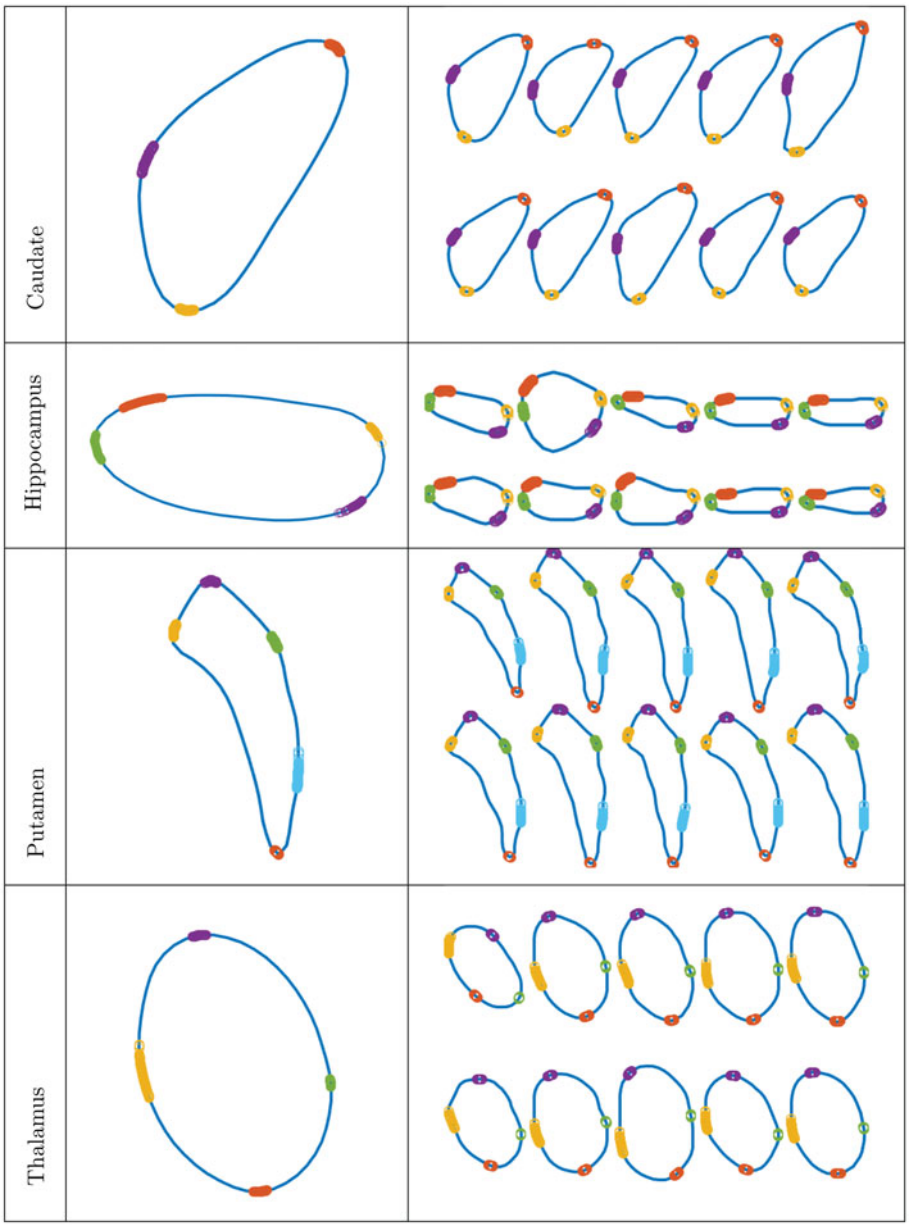


Figure 15. Posterior sample of landmarks (red = 1, yellow = 2, purple = 3, green = 4, blue = 5) plotted on the extrinsic sample mean of each substructure (left) and individually (right).

λ can be chosen to regularize the number of landmarks k; low λ typically results in the posterior for k concentrating around small values, and increasing λ yields a path of posterior probability distributions. The choice of λ, however, is not intuitive. One could potentially add a hyperprior on λ and infer this quantity; however, this is a subjective choice, and may not solve the issue. For the purpose of this problem, we desire a small number of landmarks, and so large prior mass on low values of λ would be reasonable. One could also try to infer λ through some other quantity (e.g., curvature), or even perhaps through cross-validation; reconciliation of the choice of λ is left as future work.

In this article, we present an approach to infer landmarks for two-dimensional curves; however, as briefly discussed in Section 3.1.1, this model can be easily extended to curves in any dimension. Additionally, similar ideas can be used to develop a Bayesian model for landmark inference on *surfaces*. Surfaces present new challenges for a variety of reasons. First, landmarks can be points, curves, or potentially even subsurfaces in this setting. Thus, the definition of a "linear reconstruction" no longer applies and a new measure must be developed; defining an appropriate type of reconstruction which fits the data well is a difficult problem. However, if one were to take a similar approach to the work presented here, one simplification arises in prior specification: the landmark points are no longer order-constrained on two- or higher-dimensional domains. Thus, the prior distribution can be placed without adjusting for this ordering constraint (e.g., a uniform distribution over the domain would be sufficient). While this simplification happens for landmark points, it is unclear at this stage how one could construct appropriate priors for landmark curves or subsurfaces. We leave this exciting problem for future work.

Another extension of the proposed approach is to consider heterogeneous shape populations. In the multiple curve case, introducing heterogeneity complicates inference, as values of θ may not necessarily correspond to the same feature across the sample of shapes, especially in the presence of large elastic variability or missing parts. This can be resolved by finding the optimal groupwise registration prior to landmark inference. This can be treated as a preprocessing step with direct application of methods discussed in Srivastava et al. (2011) and Kurtek et al. (2012). However, incorporating registration into the Bayesian model by conditioning on a registration function, which respects landmark locations, seems more appropriate. Cheng, Dryden, and Huang (2016) discuss a Bayesian method of function and curve registration without landmarks, which could be extended to include different types of landmark constraints for this purpose.

The last few directions for future work consider our current implementation of MCMC. First, we will explore more efficient posterior sampling strategies. Due to the high-dimensionality of the landmark detection problem, combined with the intricate geometric details of the objects under study, our current MCMC implementation based on component-wise proposals can sometimes make it challenging to explore multimodal posteriors. Other MCMC schemes, designed to more efficiently traverse multimodal posteriors, may be required for more complex shapes. Second, we will further explore different choices for model specification, including the choice of priors and hyperparameters. Finally, we will address the problem of how to evaluate posterior landmark estimates. In several examples, we have calculated the mean, median, and MAP estimates. The mean and median are found marginally for each landmark location, and thus do not account for posterior landmark dependencies; on the other hand, the MAP estimator does. Studying suitability of these estimators for shape analysis problems could be important for automatic landmark detection; this is a problem of Bayesian decision theory.

Supplementary Materials

The supplementary materials include additional discussion of important Bayesian modeling issues (e.g., with posterior sampling and diagnosis of Markov chain Monte Carlo convergence), as well as a more in-depth overview of issues associated with discretization. Extra examples are also illustrated. Additionally, code for implementation, as well as the simulated data of Section 5, are available online.

Acknowledgments

We thank the reviewers and associate editor for their comments, which greatly improved the content and delivery of this manuscript.

Funding

This work was partially supported by NSF DMS 1613054, NSF CCF 1740761, and NIH funding: NIH R37 CA214955 (SK).

References

Bauer, M., Eslitzbichler, M., and Grasmair, M. (2017), "Landmark-Guided Elastic Shape Analysis of Human Character Motions," *Inverse Problems and Imaging*, 11, 601–621. [2]

Bernardo, J., and Smith, A. (2008), *Bayesian Theory*, Chichester: Wiley. [5]

Bookstein, F. L. (1986), "Size and Shape Spaces for Landmark Data in Two Dimensions," *Statistical Science*, 1, 181–222. [1]

Chen, C., Xie, W., Franke, J., Grutzner, P., Nolte, L., and Zheng, G. (2014), "Automatic X-ray Landmark Detection and Shape Segmentation via Data-Driven Joint Estimation of Image Displacements," *Medical Image Analysis*, 18, 487–499. [2]

Cheng, W., Dryden, I. L., and Huang, X. (2016), "Bayesian Registration of Functions and Curves," *Bayesian Analysis*, 11, 447–475. [5,15]

Domijan, K., and Wilson, S. P. (2005), "A Bayesian Method for Automatic Landmark Detection in Segmented Images," in *International Conference on Machine Learning*. [2]

Dryden, I. L., and Mardia, K. V. (1992), "Size and Shape Analysis of Landmark Data," *Biometrika*, 79, 57–68. [1]

——— (2016), *Statistical Shape Analysis: With Applications in R* (2nd ed.), New York: Wiley. [1,12]

Gasser, T., and Kneip, A. (1995), "Searching for Structure in Curve Sample," *Journal of the American Statistical Association*, 90, 1179–1188. [2]

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2004), *Bayesian Data Analysis*, Boca Raton, FL: Chapman and Hall/CRC. [5,6]

Gilani, S. Z., Shafait, F., and Mian, A. (2015), "Shape-Based Automatic Detection of a Large Number of 3D Facial Landmarks," in *IEEE Conference on Computer Vision and Pattern Recognition*. [2]

Green, P. J. (1995), "Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination," *Biometrika*, 82, 711–732. [7]

Grenander, U., and Miller, M. I. (1994), "Representations of Knowledge in Complex Systems," *Journal of the Royal Statistical Society*, Series B, 56, 549–603. [7]

Kendall, D. G. (1984), "Shape Manifolds, Procrustean Metrics, and Complex Projective Shapes," *Bulletin of London Mathematical Society*, 16, 81–121. [1]

Kerr, G., Kurtek, S., and Srivastava, A. (2011), "A Joint Model for Boundaries of Multiple Anatomical Parts," in *SPIE Medical Imaging*. [13]

Klassen, E., Srivastava, A., Mio, W., and Joshi, S. H. (2004), "Analysis of Planar Shapes Using Geodesic Paths on Shape Spaces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26, 372–383. [2]

Kurtek, S. (2017), "A Geometric Approach to Pairwise Bayesian Alignment of Functional Data Using Importance Sampling," *Electronic Journal of Statistics*, 11, 502–531. [5]

Kurtek, S., Srivastava, A., Klassen, E., and Ding, Z. (2012), "Statistical Modeling of Curves Using Shapes and Related Features," *Journal of the American Statistical Association*, 107, 1152–1165. [3,15]

Liu, W., Srivastava, A., and Zhang, J. (2010), "Protein Structure Alignment Using Elastic Shape Analysis," in *ACM International Conference on Bioinformatics and Computational Biology*. [2]

Michor, P. W., Mumford, D., Shah, J., and Younes, L. (2007), "A Metric on Shape Space With Explicit Geodesics," *Matematica E Applicazioni*, 19, 25–57. [1]

Prematilake, C., and Ellingson, L. (2018), "Evaluation and Prediction of Polygon Approximations of Planar Contours for Shape Analysis," *Journal of Applied Statistics*, 45, 1227–1246. [2]

Ramsay, J., and Silverman, B. (2005), *Functional Data Analysis*, New York: Springer. [2]

Richardson, S., and Green, P. J. (1997), "On Bayesian Analysis of Mixtures With an Unknown Number of Components," *Journal of the Royal Statistical Society*, Series B, 59, 731–792. [6,7]

Robinson, D. T. (2012), "Functional Data Analysis and Partial Shape Matching in the Square Root Velocity Framework," Ph.D. thesis, Florida State University. [3]

Rueda, S., Udupa, J., and Bai, L. (2008a), "Landmark Selection for Shape Model Construction via Equalization of Variance," in *IEEE International Symposium on Biomedical Imaging*. [2]

——— (2008b), "A New Method of Automatic Landmark Tagging for Shape Model Construction via Local Curvature Scale," in *SPIE Medical Imaging*. [2]

Segundo, M. P., Silva, L., Bellon, O. R. P., and Queirolo, C. C. (2010), "Automatic Face Segmentation and Facial Landmark Detection in Range Images," *IEEE Transactions on Systems, Man, and Cybernetics B: Cybernetics*, 40, 1319–1330. [2]

Small, C. G. (1996), *The Statistical Theory of Shape*, New York: Springer. [1]

Srivastava, A., and Klassen, E. P. (2016), *Functional and Shape Data Analysis*. New York: Springer-Verlag. [3]

Srivastava, A., Klassen, E., Joshi, S. H., and Jermyn, I. H. (2011), "Shape Analysis of Elastic Curves in Euclidean Spaces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33, 1415–1428. [1,2,3,15]

Strait, J., and Kurtek, S. (2016), "Bayesian Model-Based Automatic Landmark Detection for Planar Curves," in *IEEE CVPR Workshop on Differential Geometry in Computer Vision and Machine Learning*. [2]

Strait, J., Kurtek, S., Bartha, E., and S. MacEachern (2017), "Landmark-Constrained Elastic Shape Analysis of Planar Curves," *Journal of the American Statistical Association*, 112, 521–533. [2,12]

Tie, Y., and Guan, L. (2013), "Automatic Landmark Point Detection and Tracking for Human Facial Expressions," *EURASIP Journal on Image and Video Processing*, 2013, 1–15. [2]

Younes, L. (1998), "Computable Elastic Distance Between Shapes," *SIAM Journal of Applied Mathematics*, 58, 565–586. [1,3]

Zahn, C. T., and Roskies, R. Z. (1972), "Fourier Descriptors for Plane Closed Curves," *IEEE Transactions on Computers*, 21, 269–281. [2]