

# Nonparametric Statistical Methods on Manifolds

I.L. Dryden, H. Le, S.P. Preston, A.T.A. Wood  
School of Mathematical Sciences, University of Nottingham

One of the many fundamental contributions that Rabi Bhattacharya, together with his co-authors, has made is the development of a general nonparametric theory of statistical inference on manifolds, in particular related to both intrinsic and extrinsic Fréchet means of probability distributions thereon (cf. [3], [4] and [5]). With the increasing importance of statistical analysis for non-Euclidean data in many applications, there is much scope for further advances related to this particular broad area of research. In the following, we concentrate on two particular important themes in data analysis on manifolds: nonparametric bootstrap methods and nonparametric curve fitting.

## 1 Bootstrap methods

A major application of the central limit results in [4] and [5] for intrinsic and extrinsic means on general manifolds is to statistical inference, e.g. the construction of confidence regions and multi-sample tests via bootstrap methods.

The bootstrap, see [8], provides a way of estimating, on the basis of an observed sample, the sampling distribution of a statistic  $T(S; F)$ , where  $S = \{X_1, \dots, X_n\}$  is an independent random sample from a probability distribution  $F$ . Here we will assume that  $F$  is a distribution on a Riemannian manifold  $\mathcal{M}$  and that the target for inference is a population location parameter  $\mu = \mu(F) \in \mathcal{M}$ . Let  $\hat{F}$  denote the empirical distribution function (EDF) based on the sample  $S$ ; so for  $A \subset \mathcal{M}$ ,

$$\hat{F}(A) = n^{-1} \sum_{i=1}^n \delta_{X_i}(A),$$

where  $\delta_{X_i}(A) = 1$  if  $X_i \in A$  and zero otherwise. Let  $S^* = \{X_1^*, \dots, X_n^*\}$  denote a sample of size  $n$  drawn randomly with replacement from the original sample  $S$ , with corresponding EDF  $\hat{F}^*$ . Then Efron's bootstrap principle is that the distribution of  $\hat{\mu} = \mu(\hat{F})$  based on an i.i.d. sample  $\{X_1, \dots, X_n\}$  from  $F$  is approximately the same as the distribution of  $\hat{\mu}^* = \mu(\hat{F}^*)$  based on an i.i.d. sample  $\{X_1^*, \dots, X_n^*\}$  from  $\hat{F}$ . The usefulness of this result is that the distribution of  $\hat{\mu}^*$  can be approximated with arbitrary accuracy by simulation and hence used as a basis for inference about  $\mu$ .

Two very basic but commonly arising inference problems concerning  $\mu$  are as follows: (i) given a sample  $S = \{X_1, \dots, X_n\}$ , construct a confidence region for  $\mu \in \mathcal{M}$ ; and (ii) given  $k \geq 2$  independent samples  $S_1, \dots, S_k$ , test the hypothesis that  $\mu_1 = \dots = \mu_k$ , where  $\mu_i = \mu(F_i)$  and  $F_i$  is the population distribution from which the sample  $S_i = \{X_{i,1}, \dots, X_{i,n_i}\}$  is drawn.

There are many possible ways to construct bootstrap procedures to address (i) and (ii), but doing so using *pivotal* statistics has well-known advantages; see, for example, [10] and [12]. A pivotal statistic is one whose asymptotic distribution does not depend on unknown parameters. Pivotal statistics for bootstrapping in the setting where  $\mu$  is an extrinsic mean and  $\mathcal{M}$  is the unit sphere were devised in [10], and the approach was generalised in [4] and [5] to general  $\mathcal{M}$ .

Consider the following strategy for constructing a pivotal statistic for an extrinsic mean. Suppose that  $\{X_1, \dots, X_n\}$  is an i.i.d. sample from a subset of a finite-dimensional linear space, represented as  $R^q$  or  $C^q$ , in which  $\mathcal{M}$  is embedded. Assume that a location parameter of interest can be written as

$$\mu = \phi(\Xi) \quad \text{where} \quad \Xi = E[X_1],$$

where  $\phi(\Xi)$  is a smooth map with codomain of dimension  $t$ , say. Typically, extrinsic means can be written in this form.

Suppose that the sample mean  $\hat{\Xi} = n^{-1} \sum_{i=1}^n X_i$  satisfies a central limit theorem, so that as  $n \rightarrow \infty$ ,

$$n^{1/2}(\hat{\Xi} - \Xi) \rightarrow N_q(0, V)$$

in distribution. Since  $\phi(\cdot)$  is smooth, then, by the delta method,

$$n^{1/2}\{\phi(\hat{\Xi}) - \phi(\Xi)\} \approx L(\hat{\Xi} - \Xi),$$

where  $L$  is a matrix of derivatives, and

$$n^{1/2}\{\phi(\hat{\Xi}) - \phi(\Xi)\} \rightarrow N_t(0, LVL^\top)$$

in distribution. Then, provided  $LVL^\top$  has full rank,

$$n\{\phi(\hat{\Xi}) - \phi(\Xi)\}^\top (LVL^\top)^{-1} \{\phi(\hat{\Xi}) - \phi(\Xi)\}$$

converges in distribution to  $\chi_t^2$  when  $\Xi = E[X_1]$ . In practice it is convenient to replace  $LVL^\top$  with its asymptotically equivalent sample analogue  $\hat{L}\hat{V}\hat{L}^\top$ , leading to the pivotal statistic

$$T(\mu) = n\{\phi(\hat{\Xi}) - \mu\}^\top (\hat{L}\hat{V}\hat{L}^\top)^{-1} \{\phi(\hat{\Xi}) - \mu\}. \quad (1)$$

A bootstrap version of this statistic is

$$T^*(\mu) = n\{\phi(\hat{\Xi}^*) - \mu\}^\top (\hat{L}^*\hat{V}^*\hat{L}^{*\top})^{-1} \{\phi(\hat{\Xi}^*) - \mu\}, \quad (2)$$

which can be used to construct bootstrap algorithms to address problems (i) and (ii) as follows.

**Algorithm 1: Bootstrap confidence region for  $\mu \in \mathcal{M}$ .**

**Step 1.** Starting with the original sample  $S = \{X_1, \dots, X_n\}$ , calculate  $\hat{\mu} = \mu(\hat{F})$ .

**Step 2.** Generate  $B$  bootstrap resamples  $S_1^*, \dots, S_B^*$  randomly with replacement from the original sample  $S$ . For  $b = 1, \dots, B$ , calculate  $c_b = T_b^*(\hat{\mu})$ , the

value of the bootstrap version of the pivotal statistic based on resample  $S_b^*$  evaluated at  $\hat{\mu}$ .

**Step 3.** Order the  $c$ -values to obtain  $c_{(1)} \leq \dots \leq c_{(B)}$ .

**Step 4.** Return the approximate  $100(1 - \alpha)\%$  confidence region

$$\{\mu \in \mathcal{M} : T(\mu) \leq c_{([B(1-\alpha)]+1)}\},$$

where  $T(\mu)$  is the pivotal statistic based on the original sample  $S$ .

**Algorithm 2: Bootstrap test for equality of means**  $\mu_1 = \dots = \mu_k$ .

**Step 1.** Given samples  $S_1, \dots, S_k$ , calculate the quantities needed to evaluate the pivotal statistics  $T_1(\mu), \dots, T_k(\mu)$ .

**Step 2.** Find  $\hat{\mu}_{\text{pooled}}$  to minimise  $\sum_{j=1}^k T_j(\mu)$  over  $\mu$ , and write

$$\hat{\tau} = \sum_{j=1}^k T_j(\hat{\mu}_{\text{pooled}}).$$

**Step 3.** Set up the bootstrap null hypothesis  $H_{\text{Boot}}$  by adjusting the empirical distribution functions  $\hat{F}_1, \dots, \hat{F}_k$  to  $\hat{F}_1^{\text{adj}}, \dots, \hat{F}_k^{\text{adj}}$ , in such a way that

$$\mu(\hat{F}_1^{\text{adj}}) = \dots = \mu(\hat{F}_k^{\text{adj}}) = \hat{\mu}_{\text{pooled}}.$$

One can adjust the  $\hat{F}_j$  either by transforming the samples in some way, or by resampling with non-uniform resampling probabilities, or maybe a combination of the two.

**Step 4.** Generate  $B$  independent resamples under the bootstrap null hypothesis  $H_{\text{Boot}}$  to obtain resamples  $\{S_{11}^*, \dots, S_{k1}^*\}, \dots, \{S_{1B}^*, \dots, S_{kB}^*\}$ . For  $b = 1, \dots, B$ , perform Step 2 to obtain  $\hat{\mu}_{\text{pooled},b}$ , and calculate

$$\tau_1^* = \sum_{j=1}^k T_{j1}^*(\hat{\mu}_{\text{pooled},1}^*), \dots, \tau_B^* = \sum_{j=1}^k T_{jB}^*(\hat{\mu}_{\text{pooled},B}^*),$$

where  $T_{jb}^*$  is the pivotal statistic based on  $S_{1b}^*, \dots, S_{kb}^*$ .

**Step 5.** Calculate the bootstrap  $p$ -value

$$\frac{1}{B} \#\{b : \hat{\tau} > \hat{\tau}_b\}.$$

Types of manifold-valued data for which bootstrap algorithms of this kind have been developed include directions and axes and 2D and 3D shape (cf. [1], [2], [10], [23] and [24]). In 2D shape analysis, for example, preshapes (configurations of landmarks with location and scale information removed) can be written as complex unit vectors  $Z_1, \dots, Z_n$ . Taking  $X_i = Z_i Z_i^*$ , where  $*$  denotes conjugate transpose, further removes rotation information, then the space of all

possible  $X_i$  can be identified with 2D similarity shape space (cf. [5]). The population and sample extrinsic mean shapes are the unit eigenvectors corresponding to the largest eigenvalues of  $\Xi = E(X_1)$  and  $\hat{\Xi} = n^{-1} \sum X_i$ , respectively. Defining  $\phi(\hat{\Xi})$  as the maximum eigenvalue of  $\hat{\Xi}$  leads to straightforward calculations for  $\hat{L}$  and  $\hat{V}$  in (1) (cf. [2]). In 3D shape analysis with  $p$  landmarks, preshapes can be written as 3-by- $(p-1)$  matrices,  $Z_i$ , satisfying  $\text{trace}(Z_i^\top Z_i) = 1$ . Taking  $X_i = Z_i^\top Z_i$  removes rotation (as well as reflection) information. A map  $\phi(\hat{\Xi})$  defined to project  $\hat{\Xi}$  onto the space of positive-definite matrices of rank  $\leq 3$  can be interpreted as an extrinsic mean reflection shape (cf. [7]). Calculations for  $\hat{L}$  and  $\hat{V}$  are lengthier but tractable. An important issue, however, is that the limiting  $\chi^2$  distribution of (1) has  $3p - 7$  degrees of freedom, and unless sample sizes are very large then  $\hat{L}^* \hat{V}^* \hat{L}^{*\top}$  can be singular or ill-conditioned under bootstrap resampling. This leads to poor coverage accuracy even with reasonably large sample sizes. Using regularised test statistics, such as (1) and (2) with adjustments made to the smaller eigenvalues of  $LVL^\top$ , appears necessary, and numerical evidence suggests such an approach is very effective (cf. [23] and [24]).

In summary, the pivotal statistic for Algorithms 1 and 2 requires: a central limit theorem for the extrinsic mean (as provided by [5] in a general manifold setting), and a smooth map  $\phi(\cdot)$  that defines a meaningful location parameter, for which the calculations of  $L$  and  $V$  are tractable. If any ingredient is missing, one may still develop bootstrap approaches which are nonpivotal; see examples for analysis of projective shape in [21], and planar curves in [9].

A broadly applicable nonpivotal bootstrap approach addressing problem (i) above is the following, in which  $d$  is a metric on  $\mathcal{M}$ .

---



---

**Algorithm 3: Nonpivotal bootstrap confidence region for  $\mu \in \mathcal{M}$**

**Step 1:** Starting with the original sample  $S = \{X_1, \dots, X_n\}$ , calculate  $\hat{\mu} = \mu(\hat{F})$ .

**Step 2:** Generate  $B$  independent resamples

$$S_1^* = \{X_{11}^*, \dots, X_{1n}^*\}, \dots, S_B^* = \{X_{B1}^*, \dots, X_{Bn}^*\},$$

sampled randomly with replacement from the original sample  $S$ , and calculate  $\hat{\mu}_b^* = \mu(\hat{F}_b^*)$ , where  $\hat{F}_b^*$  is the empirical distribution function based on resample  $S_b^*$ .

**Step 3:** order the values  $d(\hat{\mu}, \hat{\mu}_1^*), \dots, d(\hat{\mu}, \hat{\mu}_B^*)$  to obtain  $c_{(1)} \leq c_{(2)} \leq \dots \leq c_{(B)}$ .

**Step 4:** return the (approximate)  $100(1 - \alpha)\%$  confidence region

$$\mathcal{R}_\alpha = \{\mu \in \mathcal{M} : d(\mu, \hat{\mu}) \leq c_{([B(1-\alpha)]+1)}\},$$

where  $[.]$  denotes integer part.

---

It remains an open question how best to develop effective bootstrap algorithms on more general spaces, such as stratified manifolds and, more generally,

various types of metric space of potential interest in applications. For the example of “open books” (disjoint copies of half-spaces glued along their boundary hyperplanes), the central limit theorem has a limit distribution which is a mixture of components (cf. [16]), which is a non-standard setting from a bootstrap perspective.

## 2 Curve fitting

Other classical statistical methodology, such as the widely-used method of principal component analysis, can be adapted and developed to analyse the variability of manifold-valued data further. The most straightforward way to apply principal component analysis to manifold-valued data is to perform the standard principal component analysis on the tangent space at the Fréchet mean of the data, i.e. obtain the eigendecomposition of the sample covariance of tangent coordinates, with eigenvectors giving the principal components ordered by the corresponding eigenvalues. In other words, the data are first transferred, using the inverse exponential map or other similar maps, to the tangent space at their Fréchet mean, and then principal component analysis is applied to the transferred data to find the lower dimensional subspace in that tangent space that maximizes the variance of the projection of the transferred data. This method, combined with generalised Procrustean analysis, is widely used in statistical shape analysis (cf. [6]). If the first principal component obtained in this way has a sufficiently high eigenvalue, a unit vector in the resulting 1-dimensional sub-tangent space determines a geodesic which often gives a good approximation to indicate the variability of the data.

This idea of using geodesics to model the variability of the data can be refined by replacing the use of the tangent space by working directly on the manifold, leading to the concept of *principal geodesic component analysis*, as introduced in [14]. Applications of principal geodesic component analysis so defined to Kendall’s shape spaces can be found in [13] & [15], and to medially-defined anatomical shapes can be found in [11]. For a set of data  $\{X_1, \dots, X_n\}$  in a given Riemannian manifold  $\mathcal{M}$  with induced metric  $d$ , the first principal geodesic  $\gamma_0$  to this set of data is defined to be

$$\gamma_0 = \arg \min_{\gamma \in G(\mathcal{M})} \sum_{i=1}^n d(X_i, \gamma)^2,$$

where  $G(\mathcal{M})$  denotes the set of all possible maximal geodesics in  $\mathcal{M}$  and the distance between a given point and a geodesic is defined as the minimum of the distances between that given point and points on the geodesic. To find such an optimal  $\gamma_0$ , we note that, up to re-parametrization and translation along its curve, any geodesic  $\gamma$  can be expressed, using the exponential map, in terms of a point  $x$  on  $\gamma$  and a unit tangent vector  $v$  at  $x$  as  $\gamma(t) = \exp_x(tv)$ . Then, the

above minimization problem can be expressed in a more tractable way as

$$\begin{aligned}
& \min_{x \in \mathcal{M}, v \in T_x(\mathcal{M}), \|v\|=1} \sum_{i=1}^n d(X_i, \gamma)^2 \\
= & \min_{x \in \mathcal{M}, v \in T_x(\mathcal{M}), \|v\|=1} \sum_{i=1}^n \min_{t_i \in \mathbb{R}} d(\exp_x(t_i v), X_i)^2 \\
= & \min_{x \in \mathcal{M}, v \in T_x(\mathcal{M}), \|v\|=1, t_i \in \mathbb{R}} \sum_{i=1}^n d(\exp_x(t_i v), X_i)^2.
\end{aligned}$$

This will allow us to use iterative computer algorithms to approximate the first principal geodesic. However, on account of the ambiguity in the choice of reference point  $x$ , mentioned above, the solution to such a minimization problem is no longer unique.

We can also consider searching for an optimal curve among other prescribed sets of curves to capture the main features of the data, for example the use of the set of small circles for data lying on a sphere. The method of *principal nested spheres* (PNS), proposed in [17], introduces a general framework for a novel non-geodesic decomposition of variability of data lying on high-dimensional spheres. Instead of searching for an optimal small circle directly, it decomposes a high-dimensional sphere into a sequence of sub-manifolds with decreasing intrinsic dimensions, which can be interpreted as an analogue of principal component analysis. The procedure for finding the PNS involves iterative reduction of the data dimension. To describe it more explicitly, we assume that  $\{X_1, \dots, X_n\}$  is a sample in the unit  $m$ -sphere  $\mathcal{S}^m$ , where  $m > 1$ . Then the best fitting sub-sphere for this set of data is defined as the sub-sphere of dimension  $m - 1$  in  $\mathcal{S}^m$  minimizing, among all possible such sub-spheres, the sum of the squares of the distances of the data points to it. Since any sub-sphere  $A_{m-1}$  of dimension  $m - 1$  can be characterized by an  $r \in (0, \pi/2]$  and a direction  $x \in \mathcal{S}^m$  as

$$A_{m-1}(x, r) = \{x' \in \mathcal{S}^m \mid d(x, x') = r\},$$

where  $d$  denotes the intrinsic distance on  $\mathcal{S}^m$ , the problem of finding the best fitting sub-sphere becomes searching for  $(\hat{x}, \hat{r})$  which solves

$$\arg \min_{x \in \mathcal{S}^m, r \in (0, \pi/2]} \sum_{i=1}^n (d(X_i, x) - r)^2.$$

Then, each step of the iterative procedure for fitting principal nested spheres repeats this, after rescaling to standardize the radius of the spheres, for the (orthogonally) projected data onto the best fitting sub-sphere obtained in the previous step. If  $\hat{r} = \pi/2$ , then the best fitting sub-sphere is a great sub-sphere, so that this method generalises the method of finding principal geodesics, similar to those from previous approaches to manifold principal component analysis.

The procedure for implementing PNS makes it clear that the method can also be used for dimension reduction of spherical data. Moreover, although the PNS method is primarily proposed for data lying on spheres as suggested by its name, the idea could possibly be generalised to fit more general manifold-valued data using principal nested sub-manifolds defined by a sequence of constraints.

For example, [17] did include principal nested 2D shape spaces, which is a simple generalisation. This would expand the range of techniques for the analysis of variability of such data. Nevertheless, this would require the understanding of the geometry of the underlying manifold. One of the challenging issues here is the careful choice of the class of sub-manifolds so that it is possible to implement the method and the interpretation is meaningful.

The methods mentioned above all use some form of orthogonal projection from the observed data to an estimated curve. However, this is not always adequate for the interpolation of time-indexed observed data on manifolds, so that it is necessary to adapt other classical techniques, such as generalising Euclidean cubic spline fitting. Recall first that, for a given data set  $\{X_1, \dots, X_n\}$  in  $\mathbb{R}^m$ , where  $X_j$  is observed at time  $t_j \in T$ ,  $j = 0, \dots, n$ , the cubic spline in  $\mathbb{R}^m$  fitted to this dataset with smoothing parameter  $\lambda$  is the function  $f(\cdot, \lambda) : T \rightarrow \mathbb{R}^m$  that minimizes

$$\sum_{i=0}^n \|f(t_j, \lambda) - X_j\|^2 + \lambda \int_T \|f''(t, \lambda)\|^2 dt, \quad (3)$$

among all  $\mathcal{C}^2$ -functions, where  $T$  is a time interval containing all the time points.

One way to generalise the Euclidean cubic spline to manifolds is to use *parallel transport* to transfer data to tangent spaces, preserving the inter-relationships among the data as much as possible, and then to use the known procedure in Euclidean space to find the cubic spline for the transported data. More precisely, for a given dataset  $S = \{X_1, \dots, X_n\}$  in a manifold  $\mathcal{M}$ , where  $X_j$  is observed at time  $t_j$ , and smoothing parameter  $\lambda$ , the  $\mathcal{M}$ -valued smoothing spline fitted to  $S$  with parameter  $\lambda$  is defined to be the  $\mathcal{C}^2$ -function

$$\gamma(\cdot, \lambda) : [t_0, t_n] \rightarrow \mathcal{M}$$

such that its unrolling  $\gamma^\dagger$  onto the tangent space of  $\mathcal{M}$  at  $\gamma(t_0, \lambda)$  is the cubic smoothing spline fitted to the data  $S^\dagger$  obtained by unwrapping  $S$  at times  $t_j$ , with respect to  $\gamma$ , into the tangent space of  $\mathcal{M}$  at  $\gamma(t_0, \lambda)$ . For a more formal definition, see [18]. The terms *unrolling* and *unwrapping* are both intuitive descriptions of moving the curve and general points on the manifold using the concept of parallel transport. The  $\mathcal{M}$ -valued smoothing spline defined in this way is, except at the data times, the solution to the 4th order differential equation

$$\nabla^4 f = 0,$$

where  $\nabla$  denotes the covariant derivative (cf. [18]). For a given data set, the search for the  $\mathcal{M}$ -valued smoothing spline involves a straightforward iterative algorithm: given an estimate  $\gamma^{(i)}$  at the  $i$ th stage, fit a Euclidean smoothing spline  $(\gamma^{(i+1)})^\dagger$  with parameter  $\lambda$ , in the tangent space at  $\gamma^{(i)}(t_0)$ , to the unwrapped data  $S^\dagger$  with respect to  $\gamma^{(i)}$ ; and then wrap  $(\gamma^{(i+1)})^\dagger$  at the corresponding times back onto the manifold with respect to  $\gamma^{(i)}$  to define  $\gamma^{(i+1)}$ . This method was introduced in [18] for solving a non-parametric smoothing problem on the sphere. Subsequent developments have included applications to regression problems and extensions to more complicated manifolds(cf. [19] and [20]). However, for the construction of  $\mathcal{M}$ -valued splines of this type to be feasible, it is crucial to have knowledge of parallel transport in the manifold  $\mathcal{M}$ .

A more direct generalisation of the Euclidean spline to manifold-valued data  $\{X_1, \dots, X_n\}$  in  $\mathcal{M}$ , where  $X_i$  is observed at time  $t_i$ , is to find the solution to the analogue for manifolds of the minimization problem (3), i.e. to find the solution to the problem of minimizing

$$E(\gamma) = \sum_{i=0}^n d(\gamma(t_i, \lambda), X_i)^2 + \lambda \int_T \left\| \frac{\nabla^2 \gamma(t, \lambda)}{dt^2} \right\|^2 dt$$

within an certain set of  $\mathcal{C}^2$ -curves on  $\mathcal{M}$ . Then, the minimizing function is four-times differentiable and satisfies the differential equation

$$\nabla^4 f + \|\nabla f\|^2 \nabla^2 f - \langle \nabla^2 f, \nabla f \rangle \nabla f = 0,$$

except at the data times. Clearly, this differential equation is heavily entangled with the geometry of the manifold, and so to find the minimizing curve requires full knowledge of that geometry. Rather than directly solving such a, usually complicated, differential equation, it is possible to search for the minimizer using the steepest-descent direction iteratively (cf. [25]): at each step  $\gamma$  is replaced by  $\gamma^\tau$ , where

$$\gamma^\tau(t) = \exp_{\gamma(t)}(-\tau \text{grad}(E(\gamma))(t))$$

and  $\tau$  is a pre-determined positive constant. To be able successfully to implement this procedure, the crucial feature is the use of the second order Palais metric defined by

$$\begin{aligned} \langle\langle v, w \rangle\rangle_\gamma &= \langle v(0), w(0) \rangle_{\gamma(0)} + \left\langle \frac{\nabla v}{dt}(0), \frac{\nabla w}{dt}(0) \right\rangle_{\gamma(0)} \\ &\quad + \int_0^T \left\langle \frac{\nabla^2 v}{dt^2}, \frac{\nabla^2 w}{dt^2} \right\rangle_{\gamma(t)} dt, \end{aligned}$$

for any tangent vector fields  $v$  and  $w$  along  $\gamma$ . With this metric, the gradient of  $E$  has a relatively simple closed expression in terms of the curvature tensor, together with the parallel transport. Compared with the previous method, this requires more detailed knowledge of the geometry of manifolds. However, it has proved possible to carry it out (cf. [26]). In general, the model choice appropriate to the problem needs to take into account the balance between the need to respect to geometry of the manifold in order to preserve the inter-relation among the data and the requirement to know that geometry in detail.

If the data points are not indexed in some natural order, it is also possible to fit a curve to capture patterns of non-local variation by using the differential equations, as implied by the method of *principal flows* in [22]. The idea behind the method is that, at each point on the curve, the derivative of the fitted curve is the first principal component generated by a *local* tangent principal component, so that the curve always follows the direction of maximal variability. For a given set of data  $\{X_1, \dots, X_n\}$  in  $\mathcal{M}$ , the method relies on the introduction of the localised version of the tangent covariance matrix which is a tensor on a suitably restricted open set of  $\mathcal{M}$  defined by

$$\Sigma_\lambda(x) = \frac{1}{\sum_{i=1}^n \kappa_\lambda(X_i, x)} \sum_{i=1}^n \{ \exp_x^{-1}(X_i) \otimes \exp_x^{-1}(X_i) \} \kappa_\lambda(X_i, x).$$

The positive constant  $\lambda$  is used to control the size of the neighbourhood and  $\kappa_\lambda(X_i, x) = K(\lambda^{-1}d(X_i, x))$  for a smooth non-increasing univariate kernel  $K$  on  $[0, \infty)$ . Then, the local tangent principal component is the vector field  $W(x)$  defined to be the first principal unit eigenvector of  $\Sigma_\lambda(x)$ . Assuming that  $\Sigma_\lambda$  is defined and has distinct first and second eigenvalues on an open set containing the Fréchet mean  $\hat{x}$  of the data set, then the principal flow is the curve  $\gamma$  that solves

$$\dot{\gamma}(t) = W(\gamma(t)), \quad \gamma(0) = \hat{x}.$$

The principal flow  $\gamma$  can be extended in both directions at  $\hat{x}$  by choosing the opposite sign for  $W(\hat{x})$ : otherwise the sign of  $W$  is determined by requiring  $\dot{\gamma}$  to be smooth. Note that  $W(x)$  so defined is always tangent to the manifold and so the solution  $\gamma$  lies in  $\mathcal{M}$ . The rigidity or flexibility of the principal flows can be controlled by varying  $\lambda$ .

Over the last decade there have been many models proposed for curve fitting techniques on manifolds: the preceding selection is by no means complete. On the other hand, although some statistical analysis, such as inference, has been carried out for the estimated curves (cf. [19]), their asymptotic properties have so far been relatively less explored. Whereas similar behaviour to their Euclidean counterparts is expected, the extent of the role played by the geometry of the underlying manifolds and its effect on that behaviour is certainly unclear. Methodology based on extensions of the work of Rabi Bhattacharya, e.g. the bootstrap in [4] and [5], is one plausible avenue to explore.

### Acknowledgements

The authors acknowledge funding from the Engineering and Physical Sciences Research Council.

## References

- [1] Amaral, G.A., Dryden, I.L. & Wood, A.T.A. (2007). Pivotal bootstrap methods for  $k$ -sample problems in directional statistics and shape analysis. *Journal of the American Statistical Association*, **102** 695-707.
- [2] Amaral, G.J., Dryden, I.L., Patrangenaru, V. & Wood, A.T.A. (2010). Bootstrap confidence regions for the planar mean shape. *Journal of Statistical Planning and Inference*, **140** 3026-3034.
- [3] Bhattacharya, A. & Bhattacharya, R. (2012). *Nonparametric Inference on Manifolds: With Applications to Shape Spaces*. CUP.
- [4] Bhattacharya, R. & Patrangenaru, V. (2003). Large sample theory of intrinsic and extrinsic sample means on manifolds-I. *Ann. Statist.*, **31** 1-29.
- [5] Bhattacharya, R. & Patrangenaru, V. (2005). Large sample theory of intrinsic and extrinsic sample means on manifolds-II. *Ann. Statist.*, **33** 1225-1259.
- [6] Dryden, I.L. & Mardia, K.V. (1998). *Statistical Shape Analysis*. Wiley: Chichester.

- [7] Dryden, I.L., Kume, A., Le, H. & Wood, A. T. (2008). A multi-dimensional scaling approach to shape analysis. *Biometrika*, **95** 779-798.
- [8] Efron, B. (1979). Bootstrap methods: another look at the jackknife. *Ann. Statist.*, **7** 1-26.
- [9] Ellingson, L., Patrangenaru, V. & Ruymgaart, F. (2013). Nonparametric estimation of means on Hilbert manifolds and extrinsic analysis of mean shapes of contours. *Journal of Multivariate Analysis*, **122** 317-333.
- [10] Fisher, N.I., Hall, P., Jing, B.Y. & Wood, A.T.A. (1996). Improved pivotal methods for constructing confidence regions with directional data. *Journal of the American Statistical Association*, **91** 1062-1070.
- [11] Fletcher, P.T., Lu, C., Pizer, S.M. & Joshi, S.C. (2004). Principal geodesic analysis for the study of nonlinear statistics of shape. *IEEE Trans. Med. Imaging*, **23** 995-1005.
- [12] Hall, P.G. (1992). *The Bootstrap and Edgeworth Expansion*. Springer: New York.
- [13] Huckemann, S. & Hotz, T. (2009). Principal components geodesics for planar shape. *Journal of Multivariate Analysis*, **100** 699-714.
- [14] Huckemann, S. & Ziezold, H. (2006). Principal component analysis for Riemannian manifolds with an application to triangular shape spaces. *Adv. Appl. Prob.*, **38** 299-319.
- [15] Huckemann, S., Hotz, T. & Munk, A. (2010). Intrinsic shape analysis: geodesic principal component analysis for Riemannian manifolds modulo Lie group actions. *Statistica Sinica*, **20** 1-100.
- [16] Hotz, T., Huckemann, S., Le, H., Marron, J. S., Mattingly, J. C., Miller, E., Nolen, J., Owen, M., Patrangenaru, V. & Skwerer, S. (2013). Sticky central limit theorems on open books. *Ann. Appl. Probab.*, **23** 2238-2258.
- [17] Jung, S., Dryden, I.L. & Marron, J.S. (2012). Analysis of principal nested spheres. *Biometrika*, **99** 551-568.
- [18] Jupp, P.E. & Kent, J.T. (1987). Fitting smooth paths to spherical data. *Appl. Statist.*, **36** 34-46.
- [19] Kume, A., Dryden, I.L. & Le, H. (2007). Shape-space smoothing splines for planar landmarks. *Biometrika*, **94** 513-528.
- [20] Le, H. (2003). Unrolling shape curves. *J. London Math. Soc.*, **68** 511-526.
- [21] Mardia, K. V. & Patrangenaru, V. (2005). Directions and projective shapes. *Ann. Statist.*, **33** 1666-1699.
- [22] Panaretos, V.M., Pham, T. & Yao, Z. (2014). Principal flows. *Journal of the American Statistical Association*, **109** 424-436.
- [23] Preston, S.P. & Wood, A.T.A. (2010). Two-sample bootstrap hypothesis tests for three-dimensional labelled landmark data. *Scandinavian Journal of Statistics*, **37** 568-587.

- [24] Preston, S.P. & Wood, A.T.A. (2011). Bootstrap inference for mean reflection shape and size-and-shape with three-dimensional landmark data. *Biometrika*, **98** 49-63.
- [25] Samir, C., Absil, P.-A., Srivastava, A. & Klassen, E. (2012). A gradient-descent method for curve fitting on Riemannian manifolds. *Found. Comput. Math.*, **12** 49-73.
- [26] Su, J., Dryden, I.L., Klassen, E., Le, H. & Srivastava, A. (2012). Fitting optimal curves to time indexed, noisy observations of stochastic processes on nonlinear manifolds. *Journal of Image and Vision Computing*, **30** 428-442.