Principal Curves on Riemannian Manifolds

Søren Hauberg

Abstract—Euclidean statistics are often generalized to Riemannian manifolds by replacing straight-line interpolations with geodesic ones. While these Riemannian models are familiar-looking, they are restricted by the inflexibility of geodesics, and they rely on constructions which are optimal only in Euclidean domains. We consider extensions of Principal Component Analysis (PCA) to Riemannian manifolds. Classic Riemannian approaches seek a geodesic curve passing through the mean that optimize a criteria of interest. The requirements that the solution both is geodesic and must pass through the mean tend to imply that the methods only work well when the manifold is mostly flat within the support of the generating distribution. We argue that instead of generalizing linear Euclidean models, it is more fruitful to generalize non-linear Euclidean models. Specifically, we extend the classic *Principal Curves* from Hastie & Stuetzle to data residing on a complete Riemannian manifold. We show that for elliptical distributions in the tangent of spaces of constant curvature, the standard principal geodesic is a principal curve. The proposed model is simple to compute and avoids many of the pitfalls of traditional geodesic approaches. We empirically demonstrate the effectiveness of the Riemannian principal curves on several manifolds and datasets.

Index Terms—Principal component analysis, Principal curves, Differential geometry, Riemannian manifolds

1 Introduction

Principal component analysis (PCA) is a core tool for statistical analysis of vectorial data in $\mathbb{R}^D$. Given a data set $\mathbf{x}_{1:N} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$, PCA estimates the subspace through the mean that represent the data as well as possible,

$$\mathbf{v}_{\text{PCA}} = \operatorname*{arg\,min}_{\mathbf{v},\ \|\mathbf{v}\|=1} \sum_{n=1}^{N} \| \operatorname{proj}_{\overrightarrow{\mu\mathbf{v}}}(\mathbf{x}_n) - \mathbf{x}_n \|^2, \tag{1}$$

where $\operatorname{proj}_{\overrightarrow{\mu\mathbf{v}}}$ denotes projection onto the line spanned by the data mean μ and principal component $\mathbf{v}$. Expanding the squared norm shows that minimizing the reconstruction error (1) is equivalent to maximizing the projected variance

$$\mathbf{v}_{\text{PCA}} = \operatorname*{arg\,max}_{\mathbf{v},\ \|\mathbf{v}\|=1} \sum_{n=1}^{N} \| \operatorname{proj}_{\overrightarrow{\mu\mathbf{v}}}(\mathbf{x}_n) - \mu \|^2. \tag{2}$$

While PCA is an essential tool for data analysis in $\mathbb{R}^D$, not all data is Euclidean. Attention is growing towards data residing on Riemannian manifolds, *e.g. shape data* [1]–[6], *DTI images* [7]–[9], *image features* [10]–[12], *motion models* [13], [14], *human poses* [15]–[18], and more.

Statistical models on Riemannian manifolds often generalize Euclidean counterparts by replacing straight lines with geodesic curves in the problem formulations. The resulting models are generally referred to as *Principal Geodesic Analysis (PGA)*. In this line of thinking, Eq. 1 becomes

$$\mathbf{v}_{\text{PGA}_1} = \operatorname*{arg\,min}_{\mathbf{v}\in T_\mu\mathcal{M}} \sum_{n=1}^{N} \operatorname{dist}^2\left(\operatorname{proj}_{\gamma_\mu(\mathbf{v})}(\mathbf{x}_n),\ \mathbf{x}_n\right), \tag{3}$$

where $\gamma_\mu(\mathbf{v})\colon t \mapsto \operatorname{Exp}_\mu(t\cdot\mathbf{v})$ denotes the geodesic with starting point μ and initial velocity $\mathbf{v}$; dist is the geodesic distance; and $\operatorname{proj}_{\mathbf{c}}$ denotes projection onto the curve $\mathbf{c}$:

$$\operatorname{proj}_{\mathbf{c}}(\mathbf{x}) = \mathbf{c}(\hat{t}(\mathbf{x})) \qquad \text{where}$$
$$\hat{t}(\mathbf{x}) = \operatorname*{arg\,min}_{t} \left(\operatorname{dist}(\mathbf{c}(t),\ \mathbf{x})\right). \tag{4}$$

Similarly, Eq. 2 becomes

$$\mathbf{v}_{\text{PGA}_2} = \operatorname*{arg\,max}_{\mathbf{v}\in T_\mu\mathcal{M}} \sum_{n=1}^{N} \operatorname{dist}^2\left(\operatorname{proj}_{\gamma_\mu(\mathbf{v})}(\mathbf{x}_n),\ \mu\right). \tag{5}$$

It is worth noting that while Eq. 1 and 2 are equivalent in the Euclidean domain, the corresponding Riemannian energies (3, 5) may differ in solutions [19]. We shall return to these models in Sec. 2; for now it suffice to think of these Riemannian generalizations to PCA as replacing straight lines with geodesics to give methods in non-linear spaces.

Figure 1 provides an initial example with data in the Poincaré half-plane. PGA is unable to fit the main trend of

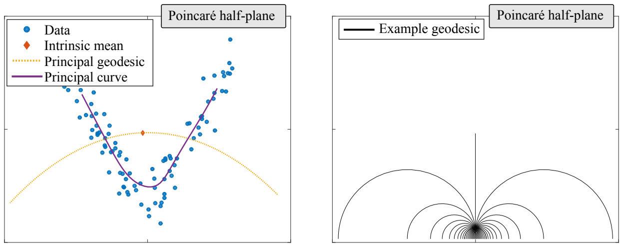

Fig. 1. *Left:* An initial example: observations (blue dots) are given in the Poincaré half-plane. The intrinsic mean (red diamond) and the principal geodesic (yellow dashed line) both do a poor job of summarizing the main trend of the data due to the restrictive structure of geodesic curves. The principal curve (purple line), in contrast, captures the main mode of variation well. *Right:* Example geodesics in the Poincaré half-plane illustrate the geometry of this space; geodesics are either vertical lines or half-circles.

• *S. Hauberg is with the Section for Cognitive Systems at the Technical University of Denmark.*
E-mail: `sohau@dtu.dk`

the data for two reasons: 1) The solution curve is required to pass through the data mean, which is outside the support of the data; and 2) the solution is required to be a geodesic, but no suitable geodesic exists in the Poincaré half-plane.

We argue for an alternative: Rather than extending *linear* Euclidean models, we find it simpler and more flexible to extend *non-linear* Euclidean techniques to Riemannian domains. In particular, this allows us to drop the two above assumptions. Figure 1 shows that this improves upon PGA.

This paper is organized as follows: The next section briefly introduce Riemannian geometry and related statistical operations. Here notation is also established. Section 3 describe principal curves and the proposed Riemannian extension. Experimental results are provided in Sec. 4, and the paper is concluded with a discussion in Sec. 5. The online supplements contain both proofs and animations.

2 Background and Related Work

In this section, we establish elementary prerequisites. More details can be found in standard references, *e.g.* [20]–[22].

2.1 Elements of Riemannian Geometry

Riemannian geometry studies smooth manifolds endowed with a smoothly changing metric. A metric on a manifold $\mathcal{M}$ is an inner product in the tangent space $T_{\mathbf{x}}\mathcal{M}$ at each point $\mathbf{x} \in \mathcal{M}$. If the inner product varies smoothly it is called a *Riemannian metric*. The tangent space $T_{\mathbf{x}}\mathcal{M}$ is a Euclidean space that locally approximates the manifold, and the tangential norm can be integrated to provide *curve lengths*. Any curve that locally minimizes its length is known as a *geodesic*. We assume that all points can be joined by a geodesic, *i.e.* we only consider *complete Riemannian manifolds*. Distances between points are then defined as the length of the shortest connecting geodesic.

Many operations are defined in the Euclidean tangent space $T_{\mathbf{x}}\mathcal{M}$, and there are mappings between this approximation of the manifold and the manifold itself. The *exponential map* at a point $\mathbf{x}$ takes a tangent vector $\mathbf{v} \in T_{\mathbf{x}}\mathcal{M}$ to

$$\mathbf{y} = \mathrm{Exp}_{\mathbf{x}}(\mathbf{v}) \in \mathcal{M} \tag{6}$$

such that the curve $t \mapsto \mathrm{Exp}_{\mathbf{x}}(t \cdot \mathbf{v})$ is a geodesic originating at $\mathbf{x}$ with initial direction $\mathbf{v}$. In general, the exponential map is only defined in a neighborhood of $\mathbf{x}$. The inverse mapping, which takes $\mathbf{y}$ to $\mathbf{v} \in T_{\mathbf{x}}\mathcal{M}$, is known as the *logarithm map* and is denoted $\mathrm{Log}_{\mathbf{x}}(\mathbf{y})$.

2.2 Intrinsic Means

Expectations are defined in the Euclidean domain through integration. This, however, generalize poorly to Riemannian domains as the expectation of a manifold-valued random variable may easily violate the constraints of the manifold [8]. It is therefore common to instead consider *intrinsic means* which are defined as the point with minimal sum-of-squared distances to the data,

$$\mu = \arg\min_{\mathbf{x}} \sum_{n=1}^{N} w_n \,\mathrm{dist}^2(\mathbf{x}_n,\ \mathbf{x}), \tag{7}$$

where w_n denotes optional weights for each observation. In Euclidean domains, this problem is uniquely solved by the empirical average. On Riemannian manifolds the solution may, however, no longer be unique; *e.g.*, consider uniformly distributed data on a unit sphere (the entire sphere solves Eq. 7). Whether μ is unique or not depends on both the spread of data and the curvature of the manifold [23], [24]. Here, we recall the uniqueness result from Kendall [24].

Definition 1 (Regular geodesic ball)**.** *The ball $\mathcal{B}(\tilde{\mathbf{x}}, r) = \{\mathbf{x} \in \mathcal{M} \mid \mathrm{dist}(\mathbf{x},\ \tilde{\mathbf{x}}) < r\}$ is "*geodesic*" if there exist a minimizing geodesic from the center $\tilde{\mathbf{x}}$ to any point in $\mathcal{B}(\tilde{\mathbf{x}}, r)$. This is* regular *if $2r\sqrt{\kappa} < \pi$, where κ is the maximum of the Riemannian curvature in the ball.*

Theorem 1 (Existence and uniqueness of means [24])**.** *Let the data be drawn from a distribution with support inside the regular geodesic ball $\mathcal{B}(\tilde{\mathbf{x}}, r)$, then there exist one and only one intrinsic mean inside $\mathcal{B}(\tilde{\mathbf{x}}, r)$.*

In practice, any concern about uniqueness is ignored and Eq. 7 is solved using gradient descent; see Algorithm 1 [22]. Empirical experience indicates that multiple modes of Eq. 7 often exist for real data, so it is potentially problematic to rely on the uniqueness of a mean.

Algorithm 1 Intrinsic mean.

Initialize μ, *e.g.* as a random data point.
repeat
  $\nabla = \sum_{n=1}^{N} w_n \,\mathrm{Log}_{\mu}(\mathbf{x}_n)$.
  $\mu = \mathrm{Exp}_{\mu}(\alpha \cdot \nabla)$.
until convergence

2.3 Principal Geodesics

As discussed in Sec. 1, standard generalizations of PCA to Riemannian manifolds seeks a geodesic curve through the mean that maximizes the projected variance (5) or minimizes the reconstruction error (3). This assumes a unique mean.

Fletcher *et al.* [25] suggest an approximate solution to (3) that computes the principal geodesic as the leading eigenvector of the covariance matrix

$$\Sigma = \frac{1}{N-1} \sum_{n=1}^{N} \mathrm{Log}_{\mu}(\mathbf{x}_n)\,\mathrm{Log}_{\mu}(\mathbf{x}_n)^{\intercal}. \tag{8}$$

In essence, Fletcher *et al.* perform Euclidean PCA in the tangent space at the mean. Sommer *et al.* [26] provide a numerical optimization scheme for solving Eq. 3 directly with no approximations. In practice, this is rarely used as it requires numerical integration of Jacobi fields and second order derivatives of geodesic families. Fletcher *et al.* [27] also consider Eq. 5, and again propose an approximation scheme that performs Euclidean PCA in the tangent space at the mean. The present author is unaware of any exact numerical solution to Eq. 5.

It is worth noting that both Eq. 3 and 5 rely upon projections of points onto a geodesic curve (4). This projection need not be unique; *e.g.* consider projecting the north pole on a sphere onto the equator (all points on the equator are solutions). Practical implementations ignore this potential issue and perform the projection through a one-dimensional optimization of Eq. 4 along the curve.

Huckemann *et al.* [19] consider Riemannian manifolds modulo isometric Lie group actions, and study an extension of Eq. 3 where the principal geodesic need not pass through the intrinsic mean. The current numerical implementation of the resulting *geodesic PCA*, however, relies on brute-force techniques that are not expected to scale gracefully.

Panaretos *et al.* [28] drop the geodesic requirement and consider smooth curves where any point on the curve has a tangent velocity vector that fits the leading eigenvector of a tangent space PCA. This increases flexibility, though the curve is required to pass through the intrinsic mean.

Jung *et al.* [29] consider data on the unit sphere and recursively discard the least interesting dimension of the data, thereby getting *principal nested spheres*. This elegantly removes the dependency on the intrinsic mean as the components are estimated before the mean. The approach is, however, dependent on the embedding of the sphere and it is unclear if the approach can generalize to other spaces.

The above methods are all derived from a least-squares perspective. Recent interest is growing toward more general probabilistic models, such as the *Probabilistic PGA* from Zhang & Fletcher [30]. Here a maximum likelihood estimate of a tangential normal distribution is estimated using a Monte Carlo EM algorithm. Attractively, this provides a noise model of the data, but the reliance on Monte Carlo makes the approach computationally challenging.

3 Principal Curves

In this section we first review the classical Euclidean principal curves for non-linear PCA. This is then extended to Riemannian manifolds in Sec. 3.2, and some links between the proposed model and PGA are established in Sec. 3.3.

3.1 Principal Curves for Euclidean Data

Informally, a *principal curve* in $\mathbb{R}^D$ is a smooth curve passing through the "middle" of a distribution [31]. Let $\mathbf{c} : \mathbb{R} \to \mathbb{R}^D$ be a one-dimensional curve, and let $\hat{t}(\mathbf{x}) \in \mathbb{R}$ be the index of the projection of $\mathbf{x} \in \mathbb{R}^D$ onto the curve $\mathbf{c}$ (see Eq. 4). A principal curve is then defined as a curve that satisfies the *self-consistency* property (see Fig. 2)

$$\mathbf{c}(t) = \mathbb{E}(\mathbf{x} \mid \hat{t}(\mathbf{x}) = t). \tag{9}$$

In words: If we consider all points that project onto $\mathbf{c}(t)$ then their expectation must equal $\mathbf{c}(t)$.

Hastie & Stuetzle [31] suggest a simple iterative algorithm for computing a principal curve from finite data. The algorithm represents the principal curve as a discrete set of points, and proceeds in two steps: first the data is projected onto the current curve estimate, and then each discrete point on the curve is updated to the average of the data which projects onto the point. To ensure that the resulting curve is smooth, a smoothing kernel is used to estimate the average.

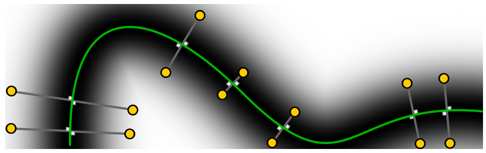

Fig. 2. *Self-consistency* (9) implies that the expectation of all points that project to the same point on a curve equals the point of projection.

Algorithm 2 Riemannian principal curve.

Initialize the principal curve as $\mathbf{c}_{1:T} = \{\mathbf{c}(1), \ldots, \mathbf{c}(T)\}$.
repeat
  $\hat{t}_n = \arg\min_t(\mathrm{dist}(\mathbf{c}(t), \mathbf{x}_n)$, $\forall n.$
  $w_{t,n} = k\left(\frac{\mathrm{dist}(\mathbf{c}(t),\ \mathbf{c}(\hat{t}_n)}{\sigma}\right)$, $\forall n, t.$
  $\mathbf{c}(t) = \arg\min_{\mathbf{x}} \sum_{n=1}^{N} w_{t,n} \,\mathrm{dist}^2(\mathbf{x}_n, \mathbf{x})$, $\forall t.$
until convergence

3.2 Principal Curves for Riemannian Data

Self-consistency (9) is extended to the Riemannian domain by replacing the expectation with an intrinsic mean (7)

$$\mathbf{c}(t) = \arg\min_{\mathbf{x}} \sum_{n=1}^{N} \delta_{t=\hat{t}_n} \,\mathrm{dist}^2(\mathbf{x}_n, \mathbf{x}), \tag{10}$$

where $\delta_{t=\hat{t}_n}$ is the Kronecker delta function, *i.e.* $\delta_{t=\hat{t}_n} = 1$ for $t = \hat{t}_n$ and 0 otherwise. Note that this definition is not enough to ensure a continuous principal curve as intrinsic means need not change continuously when the data changes continuously. In practice, the issue is resolved by representing the curve as a discrete set of points.

From the intrinsic self-consistency (10), Hastie & Stuetzle's simple algorithm generalizes directly to data residing on Riemannian manifolds. Next we describe the needed modifications, which are summarized in Algorithm 2.

3.2.1 Curve Representation

Like Hastie & Stuetzle [31], we represent a curve as a set of points, $\mathbf{c}_{1:T} = \{\mathbf{c}(1), \ldots, \mathbf{c}(T)\}$, joined by geodesic segments. This does, in principle, not give a smooth curve[1], but a fine discretization alleviates this in practice.

A point $\mathbf{x}_n$ can be projected onto such a discrete curve by computing the geodesic distance to all discrete points in $\mathbf{c}_{1:T}$, and picking the closest point as the projection, *i.e.*

$$\mathrm{proj}_{\mathbf{c}}(\mathbf{x}_n) = \mathbf{c}(\hat{t}_n) \quad \text{where} \quad \hat{t}_n = \arg\min_{t=1,\ldots,T}(\mathrm{dist}(\mathbf{c}(t), \mathbf{x}_n). \tag{11}$$

3.2.2 Kernel Smoothing

The basic algorithm for computing principal curves first projects the data $\mathbf{x}_{1:N}$ onto the discrete curve $\mathbf{c}_{1:T}$, *c.f.* Eq. 11. Each point $\mathbf{c}(t)$ on the curve is then updated as the average of the points that project onto the point, *c.f.* Eq. 10.

1. This is also the case for geodesics on manifolds where geodesics must be estimated numerically; *e.g.* Euler's method provides a polygonal curve which must be smoothed *post hoc* to ensure a smooth geodesic.

This intrinsic mean can be computed numerically using Algorithm 1. Unsurprisingly, this provides a poor estimation of the principal curve as not enough data projects onto individual points $\mathbf{c}(t)$ to provide a stable average. We, thus, replace δ in Eq. 10 with a smooth kernel k,

$$\begin{aligned} \mathbf{c}(t) &\leftarrow \operatorname*{arg\,min}_{\mathbf{x}} \sum_{n=1}^{N} w_{t,n} \operatorname{dist}^2(\mathbf{x}_n, \mathbf{x}), \\ w_{t,n} &= k\left(\frac{\operatorname{dist}(\mathbf{c}(t),\ \mathbf{c}(\hat{t}_n))}{\sigma}\right). \end{aligned} \tag{12}$$

We use a *quartic kernel* [32] $k(\Delta) = (1-\Delta^2)^2 \cdot \delta_{|\Delta|\leq 1}$, which is smooth and has finite support, but other choices are equally possible.

3.2.3 Closed Principal Curves

Algorithm 2 is an iterative procedure that preserve the connectivity of the points used to represent the principal curve. A closed principal curve can, thus, be estimated by initializing the algorithm with a closed curve. This is in contrast to PGA, which only provides a closed curve, when geodesics happen to be closed. As we shall see, this is a helpful feature for modeling periodic data, which is unavailable in alternative Riemannian models.

3.2.4 Initialization

Algorithm 2 relies on an initial estimate of the principal curve. We employ two different strategies, though alternatives are possible. When we seek an open-ended principal curve, we initialize with the principal geodesic, computed using the approximations from Fletcher *et al.* [27]. When we seek a closed principal curve, we express the data in its two leading (Euclidean) principal components in the tangent space at the intrinsic mean. We fit a circle to this data, and initialize our algorithm with its image under $\operatorname{Exp}_\mu(\cdot)$.

3.3 Relationship Between Principal Geodesics and Principal Curves

For Euclidean data, principal components are only known to be principal curves when the data follows an elliptical distribution [31]. We provide a similar result in the classic model spaces of constant curvature.

Theorem 2. *Let $\mathbf{x} \in T_\mu\mathcal{M}$, where $\mathcal{M}$ is either the Euclidean plane, the D-sphere or the hyperbolic plane, follow a zero-mean elliptical distribution with finite covariance Σ, and let $\mathbf{v}_1, \ldots, \mathbf{v}_D$ be the eigenvectors of Σ in order of decreasing eigenvalues. Then, the principal geodesic $\mathbf{c}(t) = \operatorname{Exp}_\mu(t \cdot \mathbf{v}_1)$ is a principal curve.*

Proof: See Appendix A in the supplements. □

3.4 Concerns and Benefits

While self-consistency is intuitively appealing, the principal curve does not necessarily go through the "middle" of a dataset. Hastie & Stuetzle [31] note that $\mathbf{c}$ need not be a principal curve for data $\mathbf{x}_n = \mathbf{c}(t_n) + \epsilon_n$, when the noise ϵ_n is zero-mean. Tibshirani [33] proposes a probabilistic definition of principal curves that solves this issue by modeling the data distribution as a mixture-of-Gaussians. However, standard generalizations of Gaussian distributions to Riemannian manifolds [22] does not have known normalization constants so mixture models are challenging. We, thus, stay with the original definition (9).

The original principal curves [31] are critical paths of the energy measuring sum-of-square-deviations from the data. Duchamp & Stuetzle [34], however, showed that they are only saddle points and not minimizers. Gerber & Whitaker [35] provide an excellent discussion of this phenomena and show that principal curves as defined by Hastie & Stuetzle are only well-defined due to implicit regularization, such as limiting the number of points used to represent the curve as well as the kernel smoother used for computing averages.

These concerns carry over to the Riemannian domain, where such issues are common; *e.g.* the principal geodesic is generally not an optimizer of either energies (3, 5) due to tangent space approximations. In practice, the main concern with PGA is the dependency on a global intrinsic mean, which generally does not exist, and even if it does exist may easily fall outside the support of the data distribution. Principal curves avoid these issues as the finite support of the kernel smoother implies that the average in Eq. 12 only operates on a subset of the data close to $\mathbf{c}(t)$, *i.e.* a *local average*. If the entire dataset $\mathbf{x}_{1:N}$ is in a regular geodesic ball $\mathcal{B}(\tilde{\mathbf{x}}, r)$, then the data included in the local average (12) is in another regular geodesic ball $\mathcal{B}(\hat{\mathbf{x}}, \hat{r}) \subseteq \mathcal{B}(\tilde{\mathbf{x}}, r)$ of smaller radius ($\hat{r} \leq r$). The local averages needed for computing a principal curve are then more likely to be unique than the global intrinsic average, *c.f.* Theorem 1.

4 Experiments

In this section, we evaluate the empirical performance of the principal curve on several Riemannian manifolds and datasets. We compare with the standard principal geodesic computed via the leading eigenvector of the covariance matrix in the tangent space of the intrinsic mean [27]. In all experiments, the smoothing parameter is cross-validated.

4.1 Poincaré Half-Plane

As a first illustration, we consider synthetic data distributed on the two-dimensional Poincaré half-plane $\mathbb{H}^2$ [21]. This Riemannian manifold has constant negative curvature, and a Riemannian metric given by $y^{-1}\mathbf{I}$ at $(x, y) \in \mathbb{H}^2, y > 0$. Geodesics in $\mathbb{H}^2$ are either vertical lines or circular arcs perpendicular to the x-axis. The Poincaré metric appears in information geometry as the Fisher information metric in the space of univariate normal distributions [36].

Figure 1 shows 100 samples from

$$x \sim \mathcal{U}_{[-3,3]}; \quad y \sim {}^3\!/\!_4|x| + \mathcal{N}\left(10, ({}^1\!/\!_4)^2\right), \tag{13}$$

where $\mathcal{U}_I$ denotes the uniform distribution on the interval I. As geodesics are generally circular arcs which are perpendicular to the x-axis, it is not possible to make any geodesic follow the trend of the generating distribution (13).

As a consequence, the principal geodesic provides a poor fit to the data as is evident from Fig. 1. It is worth noting that the intrinsic mean is not within the support of the distribution (13), and no smooth curve through this point will be able to capture the trend. In contrast, the principal curve follows the underlying trend of the generating distribution.

We repeat the experiment for different distributions:

$$x \sim \mathcal{U}_{[-3,3]}; \quad y \sim \mathcal{N}\left(10, (1/4)^2\right) \tag{14}$$

$$x \sim \mathcal{U}_{[-3,3]}; \quad y \sim \mathcal{N}\left(5, (1/4)^2\right) - \sin(x) \tag{15}$$

$$x \sim \mathcal{U}_{[-3,3]}; \quad y \sim 3/4 \cdot x^2 + \mathcal{N}(10, 1)\,. \tag{16}$$

Figure 3 shows the data along with the principal geodesic and the principal curve: it is evident that the first fail at capturing the main trend, while the latter succeeds.

These synthetic experiments illustrate a practical issue with the principal geodesic construction. Just as straight lines often provide a poor fit to data in Euclidean domains, geodesics cannot, in general, capture the main trend of a generating distribution. Even when a geodesic exists that captures the trend, it may not pass through the intrinsic mean, which is a practical requirement in PGA. For any but the simplest generating distributions, the principal geodesic can, thus, not be expected to capture the main trend.

4.2 The Unit Sphere

Next we consider motion capture data of a person walking in a circular pattern [37], [38]. First, we consider the global orientation of the left *femur* (*i.e. thigh bone*) represented as a unit vector, *i.e.* $\mathbf{x}_n \in \mathbb{S}^2$. The data is periodic and represented in a space with periodic geodesics, so the principal geodesic is expected to capture the main trend.

Figure 4 shows that the data approximately follow a small circle, giving an intrinsic mean at the corresponding pole. Any smooth curve passing through this pole will fail to model the trend of the data, and the principal geodesic drastically misfits the data. This illustrates a general issue with PGA. Consider data distributed uniformly over a "band" around the equator of a sphere. The intrinsic mean is given by either pole, and the principal geodesic will provide a fit which is orthogonal to the underlying trend of the data. In terms of least-squares reconstruction error, the (geodesic) equator provide a better fit to the data than the principal geodesic. The strong reliance on the intrinsic mean in PGA, thus, prevents the principal geodesic from fitting the data.

As the data captures a periodic phenomena, we fit a closed principal geodesic. The Fig. 4(left) shows that the principal curve benefits from not being forced through the global intrinsic mean. We choose the smoothing parameter σ of the principal curve with cross-validation measuring the sum-of-squared distances to the estimated curve on held-out data (with a small additive bias towards large values of σ). Figure 4(center) shows that the error measure has a well-defined optimum. While cross-validation is known to be unreliable for principal curves [34], we have found it to work well enough in our experiments. Alternatively model selection could rely on *e.g. minimum entropy* [39].

4.3 Orthogonal Matrices

In the previous section, we only considered a single bone from human motion capture data. We consider the same sequence as before, but now represent the entire skeleton as a collection of 32 rotation matrices describing the rotation of a bone relative to its parent in the kinematic tree [15]. Each data point is, thus, a member of $SO(3)^{32}$. We remove the global position and orientation from the data prior to analysis. As the underlying *walking* motion is periodic we estimate a closed principal curve.

Figure 5 shows example data points along with projections onto both the principal geodesic and the principal curve. The supplements contain the corresponding animation. It is evident that the principal geodesic fails to represent the walking motion. The principal curve captures the posture and the leg-movement, though the arm-movement is a bit too simplified. Figure 4 shows the data in the two leading principal components in the tangent space at the intrinsic mean. The figure further shows the principal geodesic and the principal curve projected into this space as well. It is evident that the ability to fit a closed (non-geodesic) curve to the data is ideal for periodic data.

4.4 Smooth Metric Learning

In the previous experiments, there was little *a priori* reason to expect a geodesic curve to be able to fit the data, which gives PGA a disadvantage. To give the principal geodesic ideal conditions, we study data on a Riemannian manifold where geodesics by definition follow local trends of the data [40]. Let $\mathbf{x}_n \in \mathbb{R}^D$ denote Euclidean observations. We now equip $\mathbb{R}^D$ with a Riemannian metric by learning metric tensors at select locations and smoothly interpolating these to form a Riemannian metric; see [40] for details. Here we learn local metrics as local inverse covariance matrices. This implies that geodesics locally follow the data.

We consider observations of hand-written one-digits from the MNIST [41] dataset (Fig. 6). For visualization, we project these into the two leading Euclidean principal components, and learn a smoothly changing metric. The left panel of Fig. 7 shows the data along with geodesics to the intrinsic mean. It is clear that geodesics follow the trend of the data. These are the ideal conditions for PGA. The center panel of Fig. 7 shows outgoing geodesics from the intrinsic mean, and the right panel shows the estimated principal geodesic. This initially follows the trend of the data, but eventually falls outside the support of the data. The principal curve, again, provides a better summary of the main trend of the data.

As we have previously seen, one problem for the principal geodesic is that it is sensitive to the location of the intrinsic mean. Here, the mean is slightly outside the support of the data, which appears to prevent the principal geodesic from fitting the data. The sensitivity of PGA to the location of the mean is concerning, when considering that intrinsic means are often not unique. The principal curve avoids this issue to some extent as it only rely on local means.

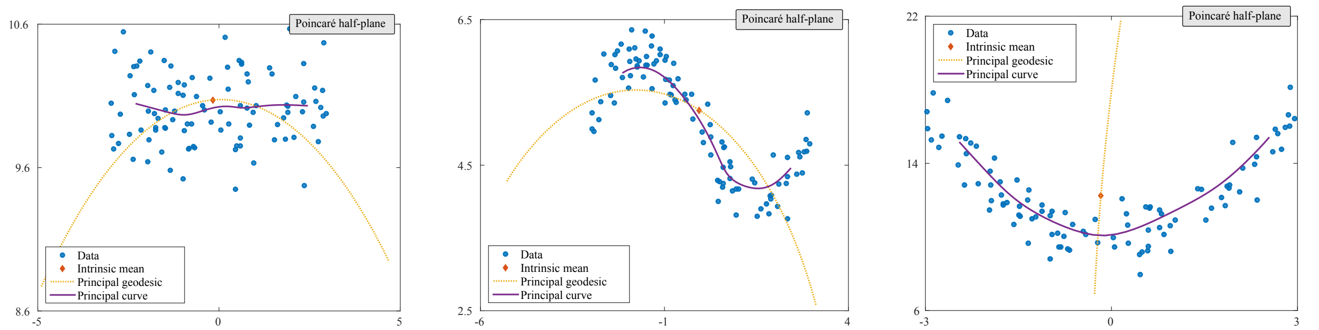

Fig. 3. Synthetic data on the Poincaré half-plane along with the principal geodesic and the principal curve. The principal curve consistently better at capturing the trend of the data.

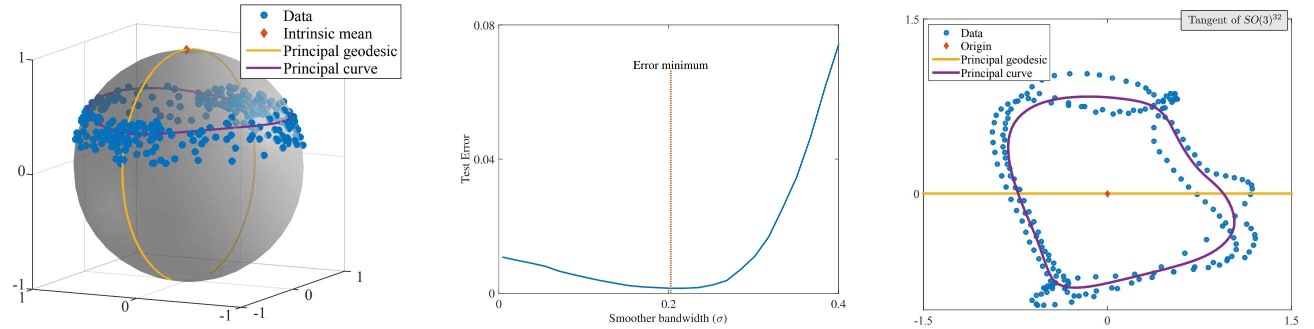

Fig. 4. *Left:* Motion capture data is represented on the power manifold $SO(3)^{32}$; here the data is shown in the first two (Euclidean) principal components in the tangent space at the intrinsic mean. The horizontal (yellow) line is then the principal geodesic. The (purple) closed curve is the principal curve mapped to the tangent space and projected into the first two principal components. The principal curve provide the best description of the data. *Center:* The cross-validation error of the smoothing parameter in the experiment corresponding to the left panel. *Right:* The relative orientation of the left femur of a person walking is represented as a point on the unit sphere. Here, the periodicity of *walking* implies that the data roughly lies on a small circle; this imply an intrinsic mean at the pole of the small circle. The principal geodesic, thus, fail to capture the trend of the data. The (closed) principal curve does better.

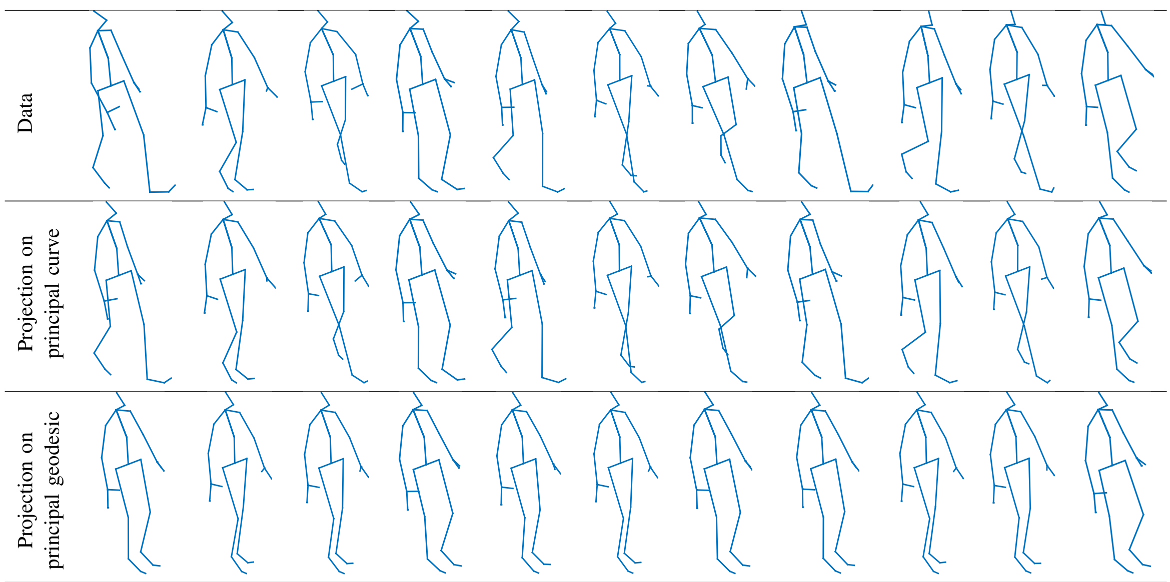

Fig. 5. Motion capture data along with the corresponding projection onto both the principal geodesic and the principal curve. The principal curve captures the structure of data much better than the principal geodesic.

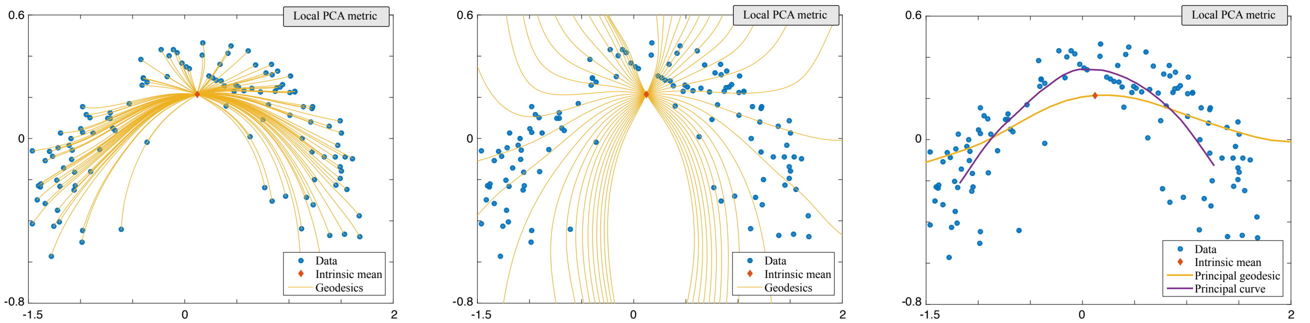

Fig. 7. MNIST images of the digit 1 is projected onto its first two (Euclidean) principal components, and a smoothly changing local diagonal metric is estimated as a local inverse covariance matrix. Left panel show geodesics between data and the intrinsic mean; geodesics generally stay with the support of the data. Center panel show geodesics originating at the intrinsic mean, shooting in different directions. Right panel show the principal geodesic and the principal curve. Even when geodesics follow the trend of the data, the principal curve captures the overall trend better than the principal geodesic.

Fig. 6. Example images from the MNIST dataset. The orientation constitute the main source of variation.

5 Summary and Discussion

Principal geodesic analysis (PGA) is one of the most widely used tools for statistical analysis of data on Riemannian manifolds. This approach seeks a geodesic through the intrinsic mean, which captures the main trend of the data. This construction is a straight-forward generalization of Euclidean PCA, which makes PGA both easy to understand and implement. The requirement that the estimated curve is geodesic and passes through the mean is at the same time the main strength and the main weakness of PGA. These strong constraints on the solution provide much needed regularization when only small quantities of data are available. However, the same constraints quickly prevent PGA from capturing the trend; in particular the requirement that the estimated curve passes through the mean is detrimental. Even when a suitable geodesic exists, it may not be found as the mean may fall outside the support of the data.

We have proposed to extend the classic *principal curves* from Hastie & Stuetzle [31] to Riemannian manifolds. The resulting model is easy to implement as it only relies on the computation of locally weighted means. As the local means are more likely to be unique than the global counterpart (Theorem 1), the model is less sensitive to curvature than PGA. The principal curve also forgoes the need for a geodesic solution as the resulting curve is flexibly represented as a piece-wise geodesic curve. This is both a strength and a limitation of the proposed approach. From small quantities of data it may be unrealistic to avoid overfitting, but when more data is available, the increased flexibility generally yields substantial improvements over the inflexible geodesic solution. This is a standard dilemma.

An added benefit of the principal curve is that it can trivially fit a closed curve to the data, even in spaces that do not have closed geodesics. We have seen how this gracefully allows for modeling periodic phenomena.

The main limitation of the principal curve framework is the curve representation. By representing the curve as a discrete set of points it is difficult to scale the approach to estimate higher dimensional surfaces. In principle, it is trivial to extend the method to estimate a D-dimensional surface [31], but since the number of needed discretization points grows exponentially with D, this quickly get impractical.

The running time of the method exceeds that of PGA as we use this for initialization. The repeated calculation of local intrinsic means may seem daunting from a computational perspective, but in practice this does not appear to be an issue. As the means are only local, they tend to not include many data points, but more importantly, they tend to be less sensitive to curvature. This implies that only few iterations of the gradient descent scheme (Algorithm 1) are needed to reach convergence. When numerical solvers are used to compute geodesics it is often more time-consuming to compute geodesics between far-away points than nearby ones. This also influence the computation of a local mean as it only relies on nearby points.

Finally, we argue that it is beneficial to generalize *non-linear* Euclidean statistics to Riemannian manifolds, rather than *linear* statistics. These non-linear models often only rely on local computations, which make them ideal for non-linear representation spaces. We further expect the principal curves to be extendable to more general metric spaces [42], [43] as they do not depend on tangent space constructions.

Acknowledgments

Søren Hauberg is funded in part by the Danish Council for Independent Research (DFF), Natural Sciences. The author is grateful to Aasa Feragen and anonymous reviewers for helpful feedback on early drafts of the manuscript.

REFERENCES

[1] D. Kendall, "Shape manifolds, Procrustean metrics, and complex projective spaces," *Bulletin of the London Mathematical Society*, vol. 16, no. 2, pp. 81–121, 1984.

[2] S. Taheri, P. Turaga, and R. Chellappa, "Towards view-invariant expression analysis using analytic shape manifolds," in *IEEE International Conference on Automatic Face & Gesture Recognition*, 2011, pp. 306–313.

[3] O. Freifeld and M. Black, "Lie bodies: A manifold representation of 3D human shape," in *European Conference on Computer Vision (ECCV)*, ser. Part I, LNCS 7572, A. Fitzgibbon et al. (Eds.), Ed. Springer-Verlag, 2012, pp. 1–14.

[4] A. Srivastava, S. H. Joshi, W. Mio, and X. Liu, "Statistical shape analysis: Clustering, learning, and testing," *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 27, no. 4, pp. 590–602, 2005.

[5] L. Younes, "Spaces and manifolds of shapes in computer vision: An overview," *Image and Vision Computing*, vol. 30, no. 6, pp. 389–397, 2012.

[6] S. Kurtek, E. Klassen, J. C. Gore, Z. Ding, and A. Srivastava, "Elastic geodesic paths in shape space of parameterized surfaces," *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 34, no. 9, pp. 1717–1730, 2012.

[7] C. Lenglet, R. Deriche, and O. Faugeras, "Inferring white matter geometry from diffusion tensor MRI: Application to connectivity mapping," in *European Conference on Computer Vision (ECCV)*, ser. Lecture Notes in Computer Science, vol. 3024, 2004, pp. 127–140.

[8] X. Pennec, P. Fillard, and N. Ayache, "A Riemannian framework for tensor computing," *International Journal of Computer Vision (IJCV)*, vol. 66, no. 1, pp. 41–66, 2006.

[9] Y. Wang, H. Salehian, G. Cheng, and B. C. Vemuri, "Tracking on the product manifold of shape and orientation for tractography from diffusion MRI," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014, pp. 3051–3056.

[10] O. Tuzel, F. Porikli, and P. Meer, "Region covariance: A fast descriptor for detection and classification," in *European Conference on Computer Vision (ECCV)*. Springer, 2006, pp. 589–600.

[11] F. Porikli, O. Tuzel, and P. Meer, "Covariance tracking using model update based on Lie algebra," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 1, 2006, pp. 728–735.

[12] O. Freifeld, S. Hauberg, and M. J. Black, "Model transport: Towards scalable transfer learning on manifolds," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014.

[13] P. K. Turaga, A. Veeraraghavan, A. Srivastava, and R. Chellappa, "Statistical computations on Grassmann and Stiefel manifolds for image and video-based recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 33, no. 11, pp. 2273–2286, 2011.

[14] H. Çetingul and R. Vidal, "Intrinsic mean shift for clustering on stiefel and grassmann manifolds," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2009, pp. 1896–1902.

[15] S. Said, N. Courty, N. Le Bihan, and S. J. Sangwine, "Exact principal geodesic analysis for data on SO(3)," in *Proceedings of the 15th European Signal Processing Conference*, 2007, pp. 1700–1705.

[16] S. Hauberg, S. Sommer, and K. S. Pedersen, "Natural metrics and least-committed priors for articulated tracking," *Image and Vision Computing*, vol. 30, no. 6-7, pp. 453–461, 2012.

[17] ——, "Gaussian-like Spatial Priors for Articulated Tracking," in *ECCV*, ser. LNCS, K. Daniilidis, P. Maragos, and N. Paragios, Eds., vol. 6311. Springer, 2010, pp. 425–437.

[18] S. Hauberg, F. Lauze, and K. S. Pedersen, "Unscented Kalman Filtering on Riemannian Manifolds," *Journal of Mathematical Imaging and Vision*, 2011.

[19] S. Huckemann, T. Hotz, and A. Munk, "Intrinsic shape analysis: geodesic PCA for Riemannian manifolds modulo isometric Lie group actions," *Statistica Sinica*, vol. 20, no. 1, pp. 1–58, 2010.

[20] J. Lee, *Introduction to Smooth Manifolds*. Springer, 2002.

[21] M. Berger, *A Panoramic View of Riemannian Geometry*. Springer, 2007.

[22] X. Pennec, "Intrinsic statistics on Riemannian manifolds: Basic tools for geometric measurements," *Journal of Mathematical Imaging and Vision*, vol. 25, no. 1, pp. 127–154, 2006.

[23] H. Karcher, "Riemannian center of mass and mollifier smoothing," *Communications on pure and applied mathematics*, vol. 30, no. 5, pp. 509–541, 1977.

[24] W. S. Kendall, "Probability, convexity, and harmonic maps with small image i: uniqueness and fine existence," *Proceedings of the London Mathematical Society*, vol. 3, no. 2, pp. 371–406, 1990.

[25] P. T. Fletcher, C. Lu, and S. Joshi, "Statistics of shape via principal geodesic analysis on lie groups," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 1, 2003, pp. 1–7.

[26] S. Sommer, F. Lauze, and M. Nielsen, "Optimization over geodesics for exact principal geodesic analysis," *Advances in Computational Mathematics*, vol. 40, no. 2, pp. 283–313, 2014.

[27] P. T. Fletcher, C. Lu, S. M. Pizer, and S. Joshi, "Principal Geodesic Analysis for the study of Nonlinear Statistics of Shape," *IEEE Transactions on Medical Imaging (TMI)*, vol. 23, no. 8, pp. 995–1005, 2004.

[28] V. M. Panaretos, T. Pham, and Z. Yao, "Principal flows," *Journal of the American Statistical Association (JASA)*, vol. 109, no. 505, pp. 424–436, 2014.

[29] S. Jung, I. L. Dryden, and J. Marron, "Analysis of principal nested spheres," *Biometrika*, vol. 99, no. 3, pp. 551–568, 2012.

[30] M. Zhang and P. Fletcher, "Probabilistic principal geodesic analysis," in *Advances in Neural Information Processing Systems (NIPS)*, 2013, pp. 1178–1186.

[31] T. Hastie and W. Stuetzle, "Principal curves," *Journal of the American Statistical Association*, vol. 84, no. 406, pp. 502–516, 1989.

[32] P. Diggle, "A kernel method for smoothing point process data," *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, vol. 34, no. 2, pp. 138–147, 1985.

[33] R. Tibshirani, "Principal curves revisited," *Statistics and Computing*, vol. 2, no. 4, pp. 183–190, 1992.

[34] T. Duchamp and W. Stuetzle, "Extremal properties of principal curves in the plane," *The Annals of Statistics*, vol. 24, no. 4, pp. 1511–1520, 08 1996.

[35] S. Gerber and R. Whitaker, "Regularization-free principal curve estimation," *Journal of Machine Learning Research (JMLR)*, vol. 14, pp. 1285–1302, 2013.

[36] S. Amari and H. Nagaoka, *Methods of information geometry*, ser. Translations of mathematical monographs; v. 191. AMS, 2000.

[37] C. Ionescu, D. Papava, V. Olaru, and C. Sminchisescu, "Human3.6m: Large scale datasets and predictive methods for 3d human sensing in natural environments," *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 36, no. 7, pp. 1325–1339, 2014.

[38] C. S. Catalin Ionescu, Fuxin Li, "Latent structured models for human pose estimation," in *International Conference on Computer Vision (ICCV)*, 2011.

[39] H. Wang and T. C. M. Lee, "Automatic parameter selection for a k-segments algorithm for computing principal curves," *Pattern Recognition Letters*, vol. 27, no. 10, pp. 1142–1150, Jul. 2006.

[40] S. Hauberg, O. Freifeld, and M. J. Black, "A geometric take on metric learning," in *Advances in Neural Information Processing Systems (NIPS) 25*, P. Bartlett, F. Pereira, C. Burges, L. Bottou, and K. Weinberger, Eds. MIT Press, 2012, pp. 2033–2041.

[41] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[42] A. Feragen, M. Owen, J. Petersen, M. Wille, L. Thomsen, A. Dirksen, and M. de Bruijne, "Tree-space statistics and approximations for large-scale analysis of anatomical trees," in *Information Processing in Medical Imaging (IPMI)*, ser. Lecture Notes in Computer Science. Springer, 2013, vol. 7917, pp. 74–85.

[43] S. Gerber, T. Tasdizen, P. T. Fletcher, S. Joshi, R. Whitaker, A. D. N. Initiative *et al.*, "Manifold modeling for brain population analysis," *Medical image analysis*, vol. 14, no. 5, pp. 643–653, 2010.

Søren Hauberg received his Ph.D. in computer science from the Univ. of Copenhagen in 2011. He has been a visiting scholar at UC Berkeley (2010), and a post doc at the Perceiving Systems department at the Max Planck Institute for Intelligent Systems (2012–2014). He is currently a post doc at the Section for Cognitive Systems, at the technical Univ. of Denmark. He works in the span of geometry and statistics.