

Anisotropic Distributions on Manifolds: Template Estimation and Most Probable Paths

Stefan Sommer

DIKU, University of Copenhagen, Denmark, sommer@diku.dk

Abstract. We use anisotropic diffusion processes to generalize normal distributions to manifolds and to construct a framework for likelihood estimation of template and covariance structure from manifold valued data. The procedure avoids the linearization that arise when first estimating a mean or template before performing PCA in the tangent space of the mean. We derive flow equations for the most probable paths reaching sampled data points, and we use the paths that are generally not geodesics for estimating the likelihood of the model. In contrast to existing template estimation approaches, accounting for anisotropy thus results in an algorithm that is not based on geodesic distances. To illustrate the effect of anisotropy and to point to further applications, we present experiments with anisotropic distributions on both the sphere and finite dimensional LDDMM manifolds arising in the landmark matching problem.

Keywords: template estimation, manifold, diffusion, geodesics, frame bundle, most probable paths, anisotropy

1 Introduction

Among the most important tasks in medical imaging is building statistical models that best describe a set of observed data. For data in Euclidean space, a basic approach is finding the mean and performing PCA on the centered data, a procedure that can be considered fitting a normal distribution to observations [20]. This approach can be extended to non-Euclidean spaces, in particular to Riemannian manifolds, by finding a Fréchet mean on the manifold and performing PCA in the tangent space of the mean (PGA, [2]), or by formulating a latent variable model in a tangent space (Probabilistic PGA, [24]). The difference between the mean and the observed data is generally represented by tangent vectors that specify the initial velocity of geodesic paths with endpoints close to the observations. This approach effectively linearizes the curved manifold [18].

In this paper, we will advocate for a different approach: we will generalize the fit of normal distributions to data without linearizing to a tangent space. Euclidean normal distributions with anisotropic covariances can be generalized to manifolds by taking limits of stochastic paths starting at a source point and propagating with stochastic velocities sampled from normal distributions. Avoiding the linearization of the manifold, the proposed procedure uses this class of

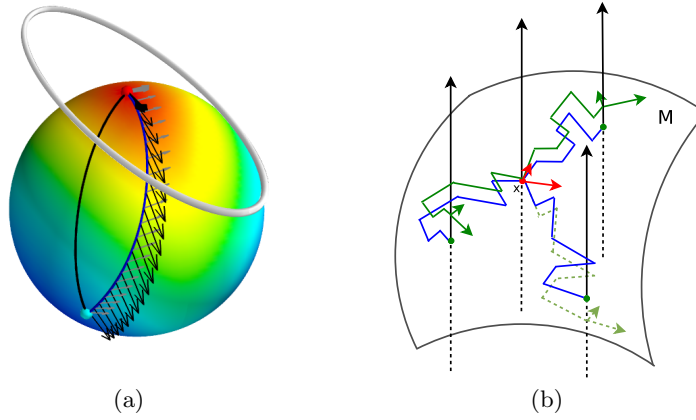


Fig. 1. (a) The sphere \mathbb{S}^2 and optimal paths for different costs: (black curve) a geodesic between the north pole and a point on the southern hemisphere; (blue curve) a most probable path (MPP) between the points for an anisotropic diffusion starting at the north pole with covariance $\text{diag}(2, .5)$ as indicated by the horizontal ellipsis at the north pole. The covariance is proportional to the squared length of the vectors that constitute a frame for $T_{(0,0,1)}\mathbb{S}^2$. The frame is parallel transported along the MPP (arrows). The MPP increases the movement in the direction of minor variance as it reaches the southern hemisphere. Compare with Figure 2 (a,b). (b) Frame bundle diffusion: for each $x \in M$, the fibers (black arrows) contain frames X_α for $T_x M$. The stochastic paths from x (blue) transport the initial frame X_α (red vectors) horizontally in the frame bundle by parallel transport (green curves and vectors). The map $\pi : FM \rightarrow M$ sends a point and frame $(z, Z_\alpha) \in FM$ to the base point $z \in M$.

distributions to find the most likely source point and covariance structure for explaining the data. The intrinsic definition of the statistical model has a significant side effect: while tangent vectors of geodesics are conventionally used as placeholders of data because geodesics are locally length minimizing, the *most probable paths* between source and data points under the anisotropic model are not geodesics. They are curves that depend on the covariance structure and the curvature of the manifold, and they provide a natural representation of data as the most probable ways of reaching observations given the statistical model.

1.1 Content and Outline

We construct an intrinsic statistical approach to modeling anisotropic data in manifolds using diffusions processes. We describe how the Eells-Elworthy-Malliavin construction of Brownian motion defines a space of probability distributions and how the non-linear template estimation problem can be recast using this construction. We argue for why the *most probable paths* are more natural than geodesics for representing data under the model and for approximating

likelihood, and we derive analytical expressions for the flow equations by identifying the paths as projections of extremal curves on a sub-Riemannian manifold. The paper ends with numerical experiments on the sphere for visualization and further on the LDDMM manifold of landmarks that is widely used in medical imaging and computational anatomy.

The paper thus presents an entirely new approach to one of the most important tasks in medical imaging: template and covariance estimation. The focus is on the theoretical insight behind the construction. The experiments presented indicate the effect of the new model without attempting to constitute a thorough evaluation on medical data. We will explore that further in future work.

While the paper necessarily uses differential geometry, we seek to only include technical details strictly needed for the exposition. Likewise, in order to clearly express the differential geometric ideas, analytical details of stochastic differential equations will be omitted.

2 Background

The data observed in a range of medical imaging applications can only be properly modeled in nonlinear spaces. Differentiable manifolds are therefore used for shape models from the Kendall shape spaces [8], to embedded curves [11] and m-reps and related models [15]. In computational anatomy, the LDDMM framework [23] provides a range of nonlinear geometries dependent on the data types to be matched: landmarks, curves, surfaces, images and tensors lead to different manifolds.

The goal of template estimation is to find a representative for a population of observations [6], for example as a Fréchet mean. Template estimation can be considered a non-linear generalization of the Euclidean space mean estimation. The Principal Geodesic Analysis (PGA, [2]) goes beyond the mean or template by modeling the covariance structure and linearizing data to the tangent space of the mean. Geodesic PCA (GPCA, [5]) and Horizontal Component Analysis (HCA, [16]) likewise provide low dimensional models of manifold valued data by seeking to generalize the Euclidean principal component analysis procedure. Probabilistic PGA [24] builds a latent variable model similar to [20] by the geodesic spray of linearly transformed latent variables. Generalizing Euclidean statistical tools to properly handling curved geometries [14] remain a challenging problem.

The frame bundle was used for dimensionality reduction with HCA [16] and later for Diffusion PCA (DPCA, [17]). While the current paper builds a statistical model similar to the DPCA construction, the focus is on the influence of anisotropy, in particular on the optimal curves, and on likelihood approximation.

For template estimation [6] and first order models such as [21], geodesic curves play a fundamental role. The Fréchet mean minimizes squared geodesic distances, and the momentum representation in the Lie algebra, the tangent space at identity, uses momenta or velocity vectors as representatives of the observed samples. Geodesics and geodesic distances take the role of difference

vectors and norms in Euclidean space because differences between the template and the observed data cannot be represented by vectors in the nonlinear geometry. Geodesic paths are used because they are locally length minimizing. The fundamental argument in this paper is that in the case of anisotropy, a different family of curves and distance metrics can be used: the most probable paths for anisotropic distributions and log-likelihoods, respectively.

3 A Class of Anisotropic Distributions

In Euclidean space, the normal distribution $\mathcal{N}(\mu, \Sigma)$ can be defined as the transition probability of a driftless diffusion process with stationary generator Σ stopped at time $t = 1$, i.e. as the solution of the stochastic differential equation $dX_t = \Sigma^{1/2} \circ X_t$. Here, the probability density $p_{\mu, \Sigma}(y)$ is given by the marginal density of all sample paths ending at y . Formally, $p_{\mu, \Sigma}(y) = \int_{x_0=\mu, x_1=y} \tilde{p}_{\Sigma}(x_t) dx_t$ where the path density $\tilde{p}_{\Sigma}(x_t) = \int_0^1 f_{\Sigma}(\dot{x}_t) dt$ integrates the density of a normal distribution $\mathcal{N}(0, \Sigma)$ on the infinitesimal steps \dot{x}_t .¹

In the Euclidean case, the distribution of $y \sim \mathcal{N}(\mu, \Sigma)$ can equivalently be specified both using the difference vector $y - \mu \sim \mathcal{N}(0, \Sigma)$ between observations and mean and using the density $p_{\mu, \Sigma}(y)$ that comes from the path density \tilde{p}_{Σ} as above. The latter definition is however much more amenable to the nonlinear situation where a displacement vector $y - \mu$ does not have intrinsic meaning. In a differentiable manifold M , vector space structure is lost but infinitesimal displacements remain in the form of tangent vectors in the tangent spaces $T_x M$, $x \in M$. The path density \tilde{p}_{Σ} therefore allows a more natural generalization to the nonlinear situation.

We here outline a construction of anisotropic normal distributions using diffusion processes. We present the discussion in the continuous limit omitting the stochastic analysis needed when the time steps of the stochastic paths goes to zero.

3.1 Stochastic Diffusion Processes on Manifolds

Let M be a manifold with affine connection ∇ , and possibly Riemannian metric g . Isotropic diffusions can then be constructed as eigenfunctions of the Laplace-Beltrami operator. The Eells-Elworthy-Malliavin construction of Brownian motion (see e.g. [4]) provides an equivalent construction² that also extends to anisotropic diffusions. The starting point is the frame bundle FM that consists of pairs (x, X_{α}) of points $x \in M$ and frames X_{α} at x . A frame at x is a collection of $n = \dim M$ tangent vectors in $T_x M$ that together constitute a basis for the tangent space. The frame can thereby be considered an invertible linear map $\mathbb{R}^n \rightarrow T_x M$. If (x, X_{α}) is a point in FM , the frame X_{α} can be considered

¹ We use the notation $\tilde{p}(x_t)$ for path densities and $p(y)$ for point densities.

² If M has more structure, e.g. being a Lie group or a homogeneous space, additional equivalent ways of defining Brownian motion and heat semi-groups exist.

a representation of the square root covariance matrix $\Sigma^{1/2}$ that we used above and thus a representation for a diffusion process starting at x . The frame being invertible implies that the square root covariance has full rank.

The difficulty in defining a diffusion process from (x, X_α) lies in the fact that the frame X_α cannot in general be defined globally on the manifold. A frame $X_{\alpha,t}$ can however be defined at a point x_t close to x by parallel transport along a curve from x to x_t . It is a fundamental property of curvature denoted holonomy [10] that this transport depends on the chosen curve, and the transport is therefore not unique. The Eells-Elworthy-Malliavin construction handles this by constructing the diffusion in FM directly so that each stochastic path carries with it the frame by parallel transport. If dU_t is a solution to such a diffusion in FM , U_t projects to a process X_t in M by the projection $\pi : FM \rightarrow M$ that maps (x, X_α) to the base point x . Different paths reaching x hit different elements of the fiber $\pi^{-1}(x)$ due to the holonomy, and the projection thus essentially integrates out the holonomy. The solutions at $t = 1$ which we will use are stochastic variables, and we write $\int_{\text{Diff}}(x, X_\alpha)$ for the map from initial conditions to X_1 . If M has a volume form, the stochastic variable X_1 has a density which generalizes the Euclidean density $p_{\mu,\Sigma}$ above.

3.2 Parallel Transport and Horizontallity

The parallel transport of a frame X_α from x to a frame $X_{\alpha,t}$ on x_t defines a map from curves x_t on M to curves $(x_t, X_{\alpha,t})$ in FM called a horizontal lift. The tangent vectors of lifted curves $(x_t, X_{\alpha,t})$ define a subspace of the frame bundle tangent space $T_{(x_t, X_{\alpha,t})}FM$ that is denoted the horizontal subspace $H_{(x_t, X_{\alpha,t})}FM$. This space of dimension n is naturally isomorphic to $T_{x_t}M$ by the map $\pi_* : (\dot{x}_t, \dot{X}_{\alpha,t}) \mapsto \dot{x}_t$ and it is spanned by n vector fields $H_i(x_t, X_{\alpha,t})$ called horizontal vector fields. Using the horizontal vector fields, the FM diffusion U_t is defined by the SDE

$$dU_t = \sum_{i=1}^n H_i(U_t) \circ W_t \quad (1)$$

where W_t is an \mathbb{R}^n valued Wiener process. The diffusion thereby only flows in the horizontal part of the frame bundle, see Figure 1 (b). Note that if M is Euclidean, the solution X_t is a stationary diffusion process.

This way of linking curves and transport of frames is also denoted “rolling without slipping” [4]. The procedure is defined only using the connection ∇ and its parallel transport. A Riemannian metric is therefore not needed at this point (though a Riemannian metric can be used to define a connection). A volume form on M , for example induced by a Riemannian metric, is however needed when we use densities for likelihoods later.

3.3 Why Parallel Transport?

In Euclidean space, the matrix $\Sigma^{1/2}$ being the square root of the covariance matrix specifies the diffusion process. In the manifold situation, the frame $X_{\alpha,0}$

takes the place of $\Sigma^{1/2}$ by specifying the covariance of infinitesimal steps. The parallel transport provides a way of linking tangent spaces, and it therefore links the covariance at x_0 with the covariance at x_t . Since the acceleration $\nabla_{\dot{x}_t} X_{\alpha,t}$ vanishes when $X_{\alpha,t}$ is parallel transported, this can be interpreted as a transport of covariance between two tangent spaces with no change seen from the manifold, i.e. no change measured by the connection ∇ . While $\Sigma^{1/2}$ can be defined globally in Euclidean space, each sample path will here carry its own transport of $X_{\alpha,0}$, and two paths ending at $y \in M$ will generally transport the covariance differently. The global effect of this stochastic holonomy can seem quite counterintuitive when related to the Euclidean situation but the infinitesimal effect is natural as a transport with no change seen from ∇ .

4 Model and Estimation

The Euclidean procedure of mean estimation followed by PCA can be given a probabilistic interpretation [20] as a maximum likelihood estimate (MLE) of the matrix W in the latent variable model $y = Wx + \mu + \epsilon$. Here x are isotropically normally distributed $\mathcal{N}(0, I)$ latent variables, and ϵ is noise $\epsilon \sim \mathcal{N}(0, \sigma^2 I)$. The marginal distribution of y is with this model normal $\mathcal{N}(\mu, C_\sigma)$, $C_\sigma = WW^T + \sigma^2 I$. The MLE for μ is the sample mean, and finding the MLE for W is equivalent to performing PCA when $\sigma \rightarrow 0$. See also [24].

We here setup the equivalent model under the assumption that $y \in M$ has distribution $y \sim \tilde{X}_1$ resulting from a diffusion (1) in FM with $X_1 = \int_{\text{Diff}}(x, X_\alpha)$ for some $(x, X_\alpha) \in FM$ and with $\tilde{X}_1 = X_1 / \int_M X_1$ the probability distribution resulting from normalizing X_1 . The source (x, X_α) of the diffusion is equivalent to the pair (μ, W) in the Euclidean situation. If M has a volume form, for example resulting from a Riemannian metric, we let $p_{(x, X_\alpha)}$ denote the density of \tilde{X}_1 . Then following [17], we define the log-likelihood

$$\ln \mathcal{L}(x, X_\alpha) = \sum_{i=1}^N \ln p_{(x, X_\alpha)}(y_i)$$

of a set of samples $\{y_1, \dots, y_N\}$, $y \in M$. Maximum likelihood estimation is then a search for $\arg \max_{(x, X_\alpha) \in FM} \ln \mathcal{L}(x, X_\alpha)$. The situation is thus completely analogous to the Euclidean situation the only difference being that the diffusion is performed in FM and projected to M . If M is Euclidean, the PCA result is recovered.³

4.1 Likelihood Estimation

Computing the transition probabilities $p_{(x, X_\alpha)}(y_i)$ directly is computationally demanding as it amounts to evaluating the solution of a PDE in the frame

³ In both cases, the representation is not unique because different matrices W /frames X_α can lead to the same diffusion. Instead, the MLE can be specified by the symmetric positive matrix $\Sigma = WW^T$ or, in the manifold situation, be considered an element of the bundle of symmetric positive covariant 2-tensors.

bundle FM . Instead, a simulation approach can be employed [13] or estimates of the mean path among the stochastic paths reaching y_i can be used [17]. Here, we will use the probability of the most probable stochastic paths reaching y_i as an approximation.

The template estimation algorithm is often considered a Fréchet mean estimation, i.e. a minimizer for the geodesic distances $d(x, y_i)^2$. With isotropic covariance, $d(x, y_i)^2$ corresponds the conditional path probability $\tilde{p}_{(x, X_\alpha)}(x_t | x_1 = y_i)$ defined below. We extend this to the situation when X_α defines an anisotropic distribution by making the approximation

$$p_{(x, X_\alpha)}(y_i) \approx \arg \max_{c_t, c_0=x, c_1=y_i} \tilde{p}_{(x, X_\alpha)}(c_t) . \quad (2)$$

We will compute the extremizers, the most probable paths, in the next section.

4.2 Metric and Rank

The covariance X_α changes the transition probability and the energy of paths. It does not in general define a new Riemannian metric. As we will see below, it defines a sub-Riemannian metric on FM but this metric does not in general descend to M . The reason being that different paths transport X_α differently to the same point, and a whole distribution of covariances thus lies in the fiber $\pi^{-1}(x)$ of each $x \in M$. A notable exception is if M already has a Riemannian metric g and X_α is unitary in which case the metric does descend to g itself.

This fact also highlights the difference with metric estimation [22] in that the proposed frame estimation does not define a metric. Here, the emphasis is placed on probabilities of paths from a source point in contrast to the symmetric distances arising from a metric.

The model above is described for a full rank $n = \dim M$ frame X_α . The MLE optimization will take place over FM which is $n + n^2$ dimensional. The frame can be replaced by $k < n$ linearly independent vectors, and a full frame can be obtained by adding these to an orthogonal basis arising from a metric g . This corresponds to finding the first k eigenvectors in PCA. We use this computationally less demanding approach for the landmark matching examples.

5 Most Probable Paths

We will here derive flow equations for a formal notion of most probable paths between source and observations under the model. We start by defining the log-probability of observing a path.

In the Euclidean situation, for any finite number of time steps, the sample path increments $\Delta x_{t_i} = x_{t_{i+1}} - x_{t_i}$ of a Euclidean stationary driftless diffusion are normally distributed $\mathcal{N}(0, (\delta t)^{-1} \Sigma)$ with log-probability

$$\ln \tilde{p}_\Sigma(x_t) \propto -\frac{1}{\delta t} \sum_{i=1}^{N-1} \Delta x_{t_i}^T \Sigma^{-1} \Delta x_{t_i} - c_\Sigma .$$

Formally, the log-probability of a differentiable path can in the limit be written

$$\ln \tilde{p}_\Sigma(x_t) \propto - \int_0^1 \|\dot{x}_t\|_\Sigma^2 dt - c_\Sigma \quad (3)$$

with the norm $\|\cdot\|_\Sigma$ given by the inner product $\langle v, w \rangle_\Sigma = \langle \Sigma^{-1/2}v, \Sigma^{-1/2}w \rangle$. This definition transfers to paths for manifold valued diffusions. Let $(x_t, X_{\alpha,t})$ be a path in FM and recall that $X_{\alpha,t}$ represents $\Sigma^{1/2}$ at x_t . Since $X_{\alpha,t}$ defines an invertible map $\mathbb{R}^n \rightarrow T_{x_t}M$ by virtue of being a basis for $T_{x_t}M$, the norm $\|\cdot\|_\Sigma$ in (3) has an analogue in the norm $\|\cdot\|_{X_{\alpha,t}}$ defined by the inner product

$$\langle v, w \rangle_{X_{\alpha,t}} = \langle X_{\alpha,t}^{-1}v, X_{\alpha,t}^{-1}w \rangle_{\mathbb{R}^n} \quad (4)$$

for vectors $v, w \in T_{x_t}M$. The norm defines a corresponding path density $\tilde{p}_{(x, X_\alpha)}(x_t)$. The transport of the frame in effect defines a transport of inner product along the sample paths: the paths carry with them the inner product defined by the covariance $X_{\alpha,0}$ at x_0 . Because $X_{\alpha,t}$ is parallel transported, the derivative of the inner product by means of the connection is zero. If M has a Riemannian metric g and $X_{\alpha,0}$ is an orthonormal basis, the new inner product will be exactly g . The difference arise when $X_{\alpha,0}$ is anisotropic.

The Onsager-Machlup functional [3] for most probable paths on manifolds differ from (4) in including a scalar curvature correction term. Defining the path density using (4) corresponds to defining the probability of a curve through its anti-development in \mathbb{R}^n . In [1], the relation between the two notions is made precise for heat diffusions. Taking the maximum of (3) gives a formal interpretation of geodesics as most probable paths for the pull-back path density of the stochastic development when Σ is unitary. We here proceed using (4) without the scalar correction term.

5.1 Flow Equations

Recall that geodesic paths for a Riemannian metric g are locally length and energy minimizing, and they satisfy the flow equation $\ddot{x}_t^k + \Gamma_{ij}^k \dot{x}_t^i \dot{x}_t^j = 0$. The super and subscripts indicate coordinates in a coordinate chart, and the Einstein summation convention implies a sum over repeated indices. We are interested in computing the most probable paths between points $x, y \in M$, i.e. the mode of the density $\tilde{p}_{(x, X_\alpha)}(x_t)$ where (x, X_α) is the starting point of a diffusion (1). By (3) and (4), this corresponds to finding minimal paths of the norm $\|\cdot\|_{X_{\alpha,t}}$, i.e. minimizers

$$x_t = \arg \min_{c_t, c_0=x, c_1=y} \int_0^1 \|\dot{c}_t\|_{X_{\alpha,t}}^2 dt. \quad (5)$$

This turns out to be an optimal control problem with nonholonomic constraints [19]. The problem can be approached by using the horizontal lift between tangent vectors in $T_{x_t}M$ and curves in $H_{(x_t, X_{\alpha,t})}FM$ coming from the map π_* : For vectors $\tilde{v}, \tilde{w} \in H_{(x_t, X_{\alpha,t})}FM$, the inner product $\langle v, w \rangle_{X_{\alpha,t}}$ lifts to the inner product

$$\langle \tilde{v}, \tilde{w} \rangle_{HFM} = \langle X_{\alpha,t}^{-1}\pi_*(\tilde{v}), X_{\alpha,t}^{-1}\pi_*(\tilde{w}) \rangle_{\mathbb{R}^n}$$

on $H_{(x_t, X_{\alpha,t})} FM$. This defines a (non bracket-generating) sub-Riemannian metric on TFM that we denote G , and (5) is equivalent to solving

$$(x_t, X_{\alpha,t}) = \arg \min_{(c_t, C_{\alpha,t}), c_0=x, c_1=y} \int_0^1 \|(\dot{c}_t, \dot{C}_{\alpha,t})\|_{HFM}^2 dt, \quad (\dot{c}_t, \dot{C}_{\alpha,t}) \in H_{(c_t, C_{\alpha,t})} FM. \quad (6)$$

The requirement on the frame part $\dot{C}_{\alpha,t}$ to be horizontal is the nonholonomic constraint. Solutions of (6) are sub-Riemannian geodesics. Sub-Riemannian geodesics satisfy the Hamilton-Jacobi equations [19]

$$\dot{y}_t^k = G^{kj}(y_t) \xi_{t,j} \quad , \quad \dot{\xi}_{t,k} = -\frac{1}{2} \frac{\partial G^{pq}}{\partial y^k} \xi_{t,p} \xi_{t,q} \quad (7)$$

similar to the geodesic equations for a Riemannian metric. In our case, y_t will be points in the frame bundle $y_t = (x_t, X_{\alpha,t})$ and $\xi_t \in T^*FM$ are covectors or momenta. Just as the geodesic equations can be written using a coordinate chart (x^1, \dots, x^n) , the Hamilton-Jacobi equations have concrete expressions in coordinate charts. The first part of the system is in our case

$$\dot{x}^i = W^{ij} \xi_j - W^{ih} \Gamma_h^{j\beta} \xi_{j\beta} \quad , \quad \dot{X}_\alpha^i = -\Gamma_h^{i\alpha} W^{hj} \xi_j + \Gamma_k^{i\alpha} W^{kh} \Gamma_h^{j\beta} \xi_{j\beta} \quad (8)$$

where (x^i) are coordinates on M , X_α^i are coordinates for the frame X_α , and $\Gamma_h^{i\alpha}$ are the Christoffel symbols following the notation of [12]. The matrix W encodes the inner product in coordinates by $W^{kl} = \delta^{\alpha\beta} X_\alpha^k X_\beta^l$. The corresponding equations for $(\xi_{i,t}, \xi_{i\alpha,t})$ are

$$\begin{aligned} \dot{\xi}^i &= W^{hl} \Gamma_{l,i}^{k\delta} \xi_h \xi_{k\delta} - \frac{1}{2} \left(\Gamma_{k,i}^{h\gamma} W^{kh} \Gamma_h^{k\delta} + \Gamma_k^{h\gamma} W^{kh} \Gamma_{h,i}^{k\delta} \right) \xi_{h\gamma} \xi_{k\delta} \\ \dot{\xi}^{i\alpha} &= \Gamma_{k,i\alpha}^{h\gamma} W^{kh} \Gamma_h^{k\delta} \xi_{h\gamma} \xi_{k\delta} - \left(W^{hl}_{,i\alpha} \Gamma_l^{k\delta} + W^{hl} \Gamma_{l,i\alpha}^{k\delta} \right) \xi_h \xi_{k\delta} \\ &\quad - \frac{1}{2} \left(W^{hk}_{,i\alpha} \xi_h \xi_k + \Gamma_k^{h\gamma} W^{kh}_{,i\alpha} \Gamma_h^{k\delta} \xi_{h\gamma} \xi_{k\delta} \right) \end{aligned} \quad (9)$$

where $\Gamma_{l,i}^{k\delta}$ denotes derivatives of the Christoffel symbols with respect to the i th component. We denote the concrete form (8),(9) of the Hamilton-Jacobi equations (7) for MPP equations.

Note that while the geodesic equations depend on $2n$ initial conditions, the flow equations for the most probable paths dependent on $2(n+n^2)$ initial conditions. The first n encode the starting position on the manifold and the next n^2 the starting frame. The remaining $n+n^2$ coordinates for the momentum allows the paths to “twist” more freely than geodesics as we will see in the experiments section. Alternatively, a system of size $2(n+kn)$ can be obtained using the low-rank approach described above resulting in reduced computational cost.

6 Experiments

The aim of the following experiments is to both visualize the nature of the most probable paths and how they depend on the initial covariance of the diffusion

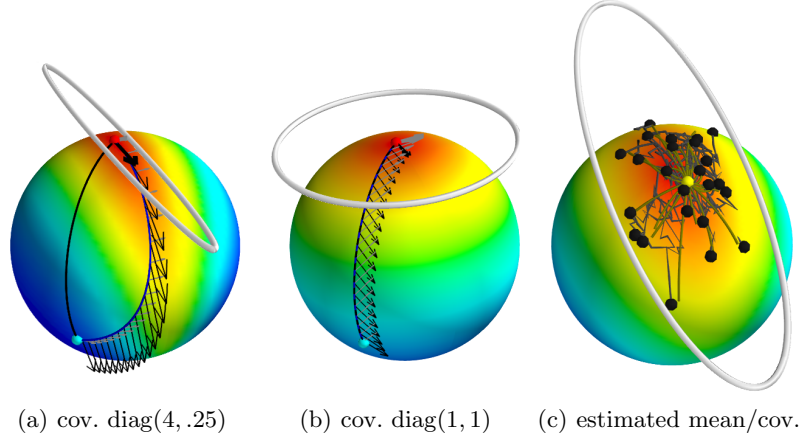


Fig. 2. Diffusions on the sphere \mathbb{S}^2 with different covariances. Compare with Figure 1 (a). (a,b) The MPPs are geodesics only with isotropic covariance, and they deviate from geodesics with increasing anisotropy. (c) Estimated mean (yellow dot) from samples (black) with stochastic paths (thin, gray) and MPPs (thick, lighter gray). Ellipsis shows the estimated covariance. Surfaces are colored by estimated densities.

and to show that the framework can be applied on a manifold that is widely used in medical imaging. We perform experiments on the sphere to satisfy the first goal, and we implement the MPP equations for the finite dimensional manifold used for LDDMM landmark matching for the second goal. The code and scripts for producing the figures is available at <http://github.com/stefansommer/dpca>.

6.1 Sphere

In the Figures 1 and 2, we visualize the difference between MPPs and geodesics on the 2-sphere \mathbb{S}^2 . The computations are performed in coordinates using the stereographic projection, and the MPP equations (8),(9) are numerically integrated. Optimization for finding MPPs between points and for finding the MLE template and covariance below is performed using a standard `scipy` optimizer. The figures clearly show how the MPPs deviate from being geodesic as the covariance increases. In the isotropic case, the curve is geodesic as expected.

Figure 2 (c) shows 24 points sampled from a diffusion with anisotropic covariance. The samples arise as endpoints of stochastic paths. We perform MLE of the most likely mean x and covariance represented by a frame X_α at x . The figure shows the sample paths together with the MPPs under the model determined by the MLE. The difference between the estimated and true covariance is $< 5\%$.

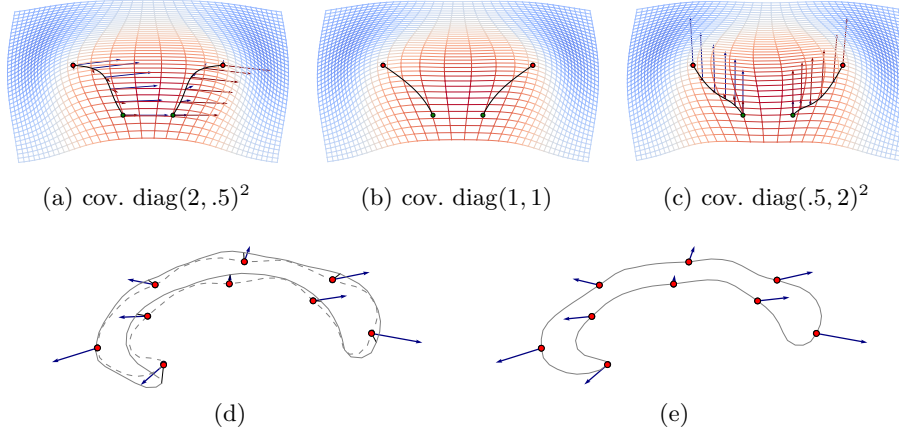


Fig. 3. Anisotropic matching and mean/covar. estimation on the LDDMM landmark manifold. The differences between the optimal paths (black curves) for the upper row matching problem is a result of anisotropy. Arrows show the covariance frame X_α . Subfigure (d) shows the MPP that move one of the ten corpora callosa into correspondence with the estimated mean (e). Arrows show the estimated major variance direction that corresponds to the Euclidean first PCA eigenvector.

6.2 Landmarks

The landmark matching problem [7] consists in finding paths ϕ_t of diffeomorphisms of a domain Ω that move finite sets of landmarks (p_1, \dots, p_m) to different sets of landmarks (q_1, \dots, q_m) through the action $\phi_t \cdot p = (\phi(p_1), \dots, \phi(p_m))$. Given a metric on $\text{Diff}(\Omega)$, the problem reduces to finding geodesics on a finite dimensional manifold Q of point positions [23]. We will show examples of landmark geodesics and MPPs, and we will show how mean shape and covariance can be estimated on a dataset of outlines of corpora callosa.

Here, $Q = (x_1^1, x_1^2, \dots, x_m^1, x_m^2)$ has metric $g(v, w) = \sum_{i,j=1}^m v K^{-1}(x_i, x_j) w$ where K^{-1} is the inverse of the matrix $K_j^i = K(x_i, x_j)$ of the Gaussian kernel. The Christoffel symbols for this metric are derived in [9].

Figure 3, upper row, shows three examples of matching two landmarks with fixed anisotropic covariance (a,c) and the geodesic isotropic case (b). The frame X_α is displayed with arrows on (a,c). The anisotropy allows less costly movement in the horizontal direction (a) and vertical direction (c) resulting in different optimal paths. In Figure 3, lower row, the result of template estimation on a set of ten outlines of corpora callosa represented by landmarks is shown. The landmarks are sampled along the outlines, and the MLE for the mean and major variance direction is found (e). Subfigure (d) shows the MPP that move one of the shapes towards the mean, and the major variance vector is displayed by arrows on the shapes.

References

1. Andersson, L., Driver, B.K.: Finite Dimensional Approximations to Wiener Measure and Path Integral Formulas on Manifolds. *Journal of Functional Analysis* 165(2), 430–498 (Jul 1999)
2. Fletcher, P., Lu, C., Pizer, S., Joshi, S.: Principal geodesic analysis for the study of nonlinear statistics of shape. *Medical Imaging, IEEE Transactions on* (2004)
3. Fujita, T., Kotani, S.i.: The Onsager-Machlup function for diffusion processes. *Journal of Mathematics of Kyoto University* 22(1), 115–130 (1982)
4. Hsu, E.P.: *Stochastic Analysis on Manifolds*. American Mathematical Soc. (2002)
5. Huckemann, S., Hotz, T., Munk, A.: Intrinsic shape analysis: Geodesic PCA for Riemannian manifolds modulo isometric Lie group actions. *Statistica Sinica* (2010)
6. Joshi, S., Davis, B., Jomier, B.M., B, G.G.: Unbiased diffeomorphic atlas construction for computational anatomy. *NeuroImage* 23, 151–160 (2004)
7. Joshi, S., Miller, M.: Landmark matching via large deformation diffeomorphisms. *Image Processing, IEEE Transactions on* 9(8), 1357–1370 (2000)
8. Kendall, D.G.: Shape Manifolds, Procrustean Metrics, and Complex Projective Spaces. *Bull. London Math. Soc.* 16(2), 81–121 (Mar 1984)
9. Micheli, M.: The differential geometry of landmark shape manifolds: metrics, geodesics, and curvature. Ph.D. thesis, Brown University, Providence, USA (2008)
10. Michor, P.W.: *Topics in Differential Geometry*. American Mathematical Soc. (2008)
11. Michor, P.W., Mumford, D.: Riemannian Geometries on Spaces of Plane Curves. *J. Eur. Math. Soc.* 8, 1–48 (2004)
12. Mok, K.P.: On the differential geometry of frame bundles of Riemannian manifolds. *Journal Fur Die Reine Und Angewandte Mathematik* 1978(302), 16–31 (1978)
13. Nye, T.: Construction of Distributions on Tree-Space via Diffusion Processes. *Mathematisches Forschungsinstitut Oberwolfach* (2014), http://www.mfo.de/document/1440a/preliminary_OWR_2014.44.pdf
14. Pennec, X.: Intrinsic Statistics on Riemannian Manifolds: Basic Tools for Geometric Measurements. *J. Math. Imaging Vis.* 25(1), 127–154 (2006)
15. Siddiqi, K., Pizer, S.: *Medial Representations: Mathematics, Algorithms and Applications*. Springer, 1 edn. (Dec 2008)
16. Sommer, S.: Horizontal Dimensionality Reduction and Iterated Frame Bundle Development. In: *Geometric Science of Information*. pp. 76–83. LNCS, Springer (2013)
17. Sommer, S.: Diffusion Processes and PCA on Manifolds. *Mathematisches Forschungsinstitut Oberwolfach* (2014), http://www.mfo.de/document/1440a/preliminary_OWR_2014.44.pdf
18. Sommer, S., Lauze, F., Hauberg, S., Nielsen, M.: Manifold Valued Statistics, Exact Principal Geodesic Analysis and the Effect of Linear Approximations. In: *ECCV 2010*. vol. 6316. Springer (2010)
19. Strichartz, R.S.: Sub-Riemannian geometry. *Journal of Differential Geometry* 24(2), 221–263 (1986), <http://projecteuclid.org/euclid.jdg/1214440436>
20. Tipping, M.E., Bishop, C.M.: Probabilistic Principal Component Analysis. *Journal of the Royal Statistical Society. Series B* 61(3), 611–622 (Jan 1999)
21. Vaillant, M., Miller, M., Younes, L., Trounev, A.: Statistics on diffeomorphisms via tangent space representations. *NeuroImage* 23(Supplement 1), S161–S169 (2004)
22. Vialard, F.X., Risser, L.: Spatially-Varying Metric Learning for Diffeomorphic Image Registration: A Variational Framework. In: *MICCAI*. Springer (2014)
23. Younes, L.: *Shapes and Diffeomorphisms*. Springer (2010)
24. Zhang, M., Fletcher, P.: Probabilistic Principal Geodesic Analysis. In: *NIPS*. pp. 1178–1186 (2013)