

Bridge Simulation and Metric Estimation on Landmark Manifolds

Stefan Sommer¹, Alexis Arnaudon², Line Kuhnel¹, and Sarang Joshi³

¹ Department of Computer Science (DIKU), University of Copenhagen, Denmark,
sommer@di.ku.dk

² Department of Mathematics, Imperial College London, UK

³ Department of Bioengineering, Scientific Computing and Imaging Institute,
University of Utah, USA

Abstract. We present an inference algorithm and connected Monte Carlo based estimation procedures for metric estimation from landmark configurations distributed according to the transition distribution of a Riemannian Brownian motion arising from the Large Deformation Diffeomorphic Metric Mapping (LDDMM) metric. The distribution possesses properties similar to the regular Euclidean normal distribution but its transition density is governed by a high-dimensional PDE with no closed-form solution in the nonlinear case. We show how the density can be numerically approximated by Monte Carlo sampling of conditioned Brownian bridges, and we use this to estimate parameters of the LDDMM kernel and thus the metric structure by maximum likelihood.

Keywords: landmarks, Brownian motion, Brownian bridges, MLE

1 Introduction

Finite dimensional landmark configurations are essential in shape analysis and computational anatomy, both for marking and following anatomically important areas in e.g. changing brain anatomies and discretely represented curve outlines, and in being among the simplest non-linear shape spaces. This simplicity, in particular the finite dimensionality, makes landmarks useful for theoretical investigations and for deriving algorithms that can subsequently be generalized to infinite dimensional shape spaces.

While probability distributions in Euclidean space can often be specified conveniently from their density function, e.g. the normal distribution with the density $p_{\mu, \Sigma}(x) \propto e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)}$, the non-linear nature of shape spaces often rules out closed form functions. Indeed, a density defined in coordinates will be dependent on the chosen coordinate chart and thus not geometrically intrinsic, and normalization factors can be inherently hard to compute. A different approach defines probability distributions as transition distributions of stochastic processes. Because stochastic differential equations (SDEs) can be specified locally from their infinitesimal variations, it is natural to define them in geometric spaces. Belonging to this category, the present paper aligns with a range

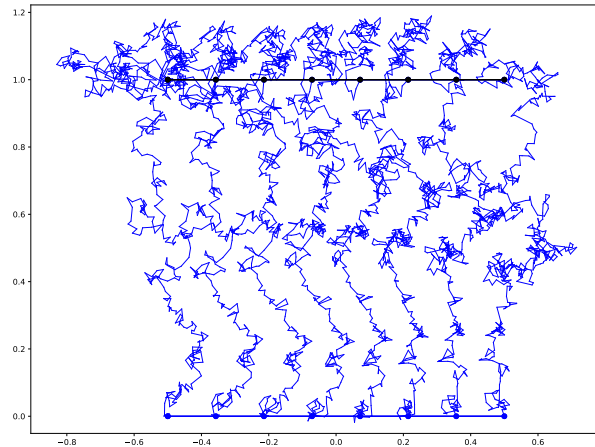


Fig. 1. A Brownian bridge connecting a configuration of 8 landmarks (blue points) to corresponding target landmarks (black points). Blue curves shows the stochastic trajectory of each landmark. The bridge arises from a Riemannian Brownian motion conditioned on hitting the target at time $T = 1$. The transition density p_T can be evaluated by taking expectation over such bridges.

of recent research activities on nonlinear SDEs in shape analysis and geometric mechanics [26,25,14,3,4].

We consider here observations distributed according to the transition distribution of a Brownian motion, which is arguably one of the most direct generalizations of the Gaussian distribution to nonlinear geometries. For the Brownian motion, each infinitesimal step defining the SDE can be considered normally distributed with isotropic covariance with respect to the Riemannian metric of the space. Then, from observations, we aim to infer parameters of this metric. In the Large Deformation Diffeomorphic Metric Mapping (LDDMM) setting, this can be framed as inferring parameters of the kernel mapping K between the dual Lie algebra \mathfrak{g}^* and the Lie algebra $\mathfrak{g} = \mathfrak{X}(\Omega)$ of the diffeomorphism group $\text{Diff}(\Omega)$ that acts on the domain Ω containing the landmarks. We achieve this by deriving a scheme for Monte Carlo simulation of Brownian landmark bridges conditioned on hitting the observed landmark configurations. Based on the Euclidean diffusion bridge simulation scheme of Delyon and Hu [6], we can compute expectation over bridges using the correction factor of a guided diffusion process to obtain the transition density of the Brownian motion. From this, we can take derivatives to obtain an iterative optimization algorithm for the most likely parameters. The scheme applies to the situation when the landmark configurations are considered observed at a fixed positive time $t = T$. The time interval $[0, T]$ will generally be sufficiently large that many time discretization points are needed to accurately represent the stochastic process.

We begin in Section 2 with a short survey of LDDMM landmark geometry, metric estimation, large deformation stochastics, and uses of Brownian motion

in shape analysis. In Section 3, we derive a scheme for simulating Brownian landmark bridges. We apply this scheme in Section 4 to derive an inference algorithm for estimating parameters of the metric. Numerical examples are presented in Section 5 before the paper ends with concluding remarks.

2 Landmarks Manifolds and Stochastic Landmark Dynamics

We start with a short survey of landmark geometry with the LDDMM framework as derived in papers including [24,7,11,5]. The framework applies to general shape spaces though we focus on configurations $\mathbf{q} = (q_1, \dots, q_N)$ of N landmarks $q_i \in \Omega \subseteq \mathbb{R}^d$. We denote the resulting manifold Q . Two sets of shapes $\mathbf{q}^0, \mathbf{q}^1$ are in LDDMM matched by minimizing the energy functional

$$E(u_t) = \int_0^1 l(u_t)dt + \frac{1}{2\lambda^2} \|g_1 \cdot \mathbf{q}^0 - \mathbf{q}^1\|^2. \quad (1)$$

The parameter of E is a time-dependent vector field $u_t \in \mathfrak{X}(\Omega)$ that via a reconstruction equation

$$\partial_t g_t = u_t \circ g_t \quad (2)$$

generates a corresponding time-dependent flow of diffeomorphisms $g_t \in \text{Diff}(\Omega)$. The endpoint diffeomorphism g_1 move the landmarks through the action $g \cdot \mathbf{q} = (g(q_1), \dots, g(q_N))$ of $\text{Diff}(\Omega)$ on Q . The right-most term of (1) measures the dissimilarity between $g_1 \cdot \mathbf{q}^0$ and \mathbf{q}^1 weighted by a factor $\lambda > 0$. In the landmark case, the squared Euclidean distance when considering the landmarks elements of \mathbb{R}^{Nd} is often used here.

The Lagrangian l on u is often in the form $l(u) = \langle u, Lu \rangle$ with the L^2 -pairing and L being a differential operator. Because $\mathfrak{X}(\Omega)$ can formally be considered the Lie algebra \mathfrak{g} of $\text{Diff}(\Omega)$, l puts the dual Lie algebra \mathfrak{g}^* into correspondence with \mathfrak{g} by the mapping $\frac{1}{2} \frac{\delta l}{\delta u} : \mathfrak{g} \rightarrow \mathfrak{g}^*$, $u \mapsto Lu$. The inverse of the mapping arise from the Green's function of L , written as the kernel K . Such l defines a right-invariant inner product on the tangent bundle $T\text{Diff}(\Omega)$ that descends to a Riemannian metric on Q . Because Q can be considered a subset of \mathbb{R}^{Nd} using the representation above, the metric structure can be written directly as a cometric

$$\langle \xi, \eta \rangle_{\mathbf{q}} = \xi^T K(\mathbf{q}, \mathbf{q}) \eta \quad (3)$$

using the kernel K evaluated on \mathbf{q} for two covectors $\xi, \eta \in T_{\mathbf{q}}^*Q$. The kernel is often specified directly in the form $K(\mathbf{q}_1, \mathbf{q}_2) = \text{Id}_d k(\|\mathbf{q}_1 - \mathbf{q}_2\|^2)$ for appropriate kernels k . One choice of k is the Gaussian kernel $k(x) = \alpha e^{-\frac{1}{2}x^T \Sigma^{-1}x}$ with matrix $\Sigma = \sigma\sigma^T$ specifying the spatial correlation structure, and $\alpha > 0$ a scaling of the general kernel amplitude.

Estimating parameters of K , with K as above α and the entries of Σ or σ , has to our knowledge previously only been treated for landmarks in the small-deformation setting [1]. While a linear vector space probability distribution is

mapped to the manifold with small deformations, this paper concerns the situation when the probability distribution is constructed directly from the Riemannian metric on the nonlinear space Q . The approach has similarities with the estimation procedures derived in [20] where a metric on a finite dimensional Lie group is estimated to optimize likelihood of data on a homogeneous space arising as the quotient of the group by a closed subgroup. Though the landmark space can be represented as $\text{Diff}(\Omega)/H$ with H the landmark isotropy subgroup [19], the approach of [20] can not directly be applied because of the infinite dimensionality of $\text{Diff}(\Omega)$.

2.1 Brownian Motion

A diffusion processes \mathbf{q}_t on a Riemannian manifold Q is said to be a Brownian motion if its generator is $\frac{1}{2}\Delta_g$ with Δ_g being the Laplace-Beltrami operator of the metric g . Such processes can be constructed in several ways, see e.g. [10]. By isometrically embedding Q in a Euclidean space \mathbb{R}^p , the process can be constructed as a process in \mathbb{R}^p that will stay on Q a.s. The process can equivalently be characterized in coordinates as being solution to the Itô integral

$$dq_t^i = g^{kl}\Gamma(\mathbf{q}_t)_{kl}{}^i dt + \sqrt{g^*(\mathbf{q}_t)}^i dW_t \quad (4)$$

where $\sqrt{g^*}$ is a square root of the cometric tensor $[g^*]^{ij} = g^{ij}$, and the drift term arise from contraction of the Christoffel symbols $\Gamma_{kl}{}^i$ with the cometric. The noise term is infinitesimal increments dW of an $\mathbb{R}^{\dim(Q)}$ -valued Brownian motion W_t . Equivalently, the Brownian motion can be constructed as a hypoelliptic diffusion processes in the orthonormal frame bundle OQ where a set of globally defined horizontal vector fields $H_1, \dots, H_{\dim(Q)} \in TOQ$ gives the Stratonovich integral equation

$$du_t = H_i(u_t) \circ W_t^i. \quad (5)$$

Note the sum over the $\dim(Q)$ horizontal fields H_i . The process $\mathbf{q}_t = \pi(u_t)$ where $\pi : OQ \rightarrow Q$ is the bundle map is then a Brownian motion. This is known as the Eells-Elworthy-Malliavin construction of Brownian motion. The fields H_i evaluated at $u \in OQ$ model infinitesimal parallel transport of the vectors comprising the frame u in the direction of the i th frame vector, see e.g. [10].

While Brownian motion is per definition isotropic with equal variation in all directions, data with nontrivial covariance can be modeled by defining the SDE (5) in the larger frame bundle FQ [17,21] using nonorthonormal frames to model the square root of the local covariance structure. In this setup, the inference problem consists of finding the starting point of the diffusion and the square root covariance matrix. Estimators are defined via a Fréchet mean like minimization in FQ with square FQ distances used as proxy for the negative log-transition density. In this paper, we remove this proxy by approximating the actual transition density, but only in the isotropic Brownian motion case.

2.2 Large Deformation Stochastics

Several papers have recently derived models for Brownian motion [13] and stochastic dynamics in shape analysis and for landmark manifolds. [25,26] considered stochastic shape evolution by adding finite and infinite dimensional noise in the momentum equation of the dynamics. In [14], noise is added to the momentum equation to make the landmark dynamics correspond to a type of heat bath appearing in statistical physics. In [3,4] a stochastic model for shape evolution is derived that descends to the landmark space in the same fashion as the right-invariant LDDMM metric descends to Q . The fundamental structure is here the momentum map that is preserved by the introduction of right-invariant noise. The approach is linked to parametric SDEs in fluid dynamics [9] and stochastic coadjoint motion [2].

3 Brownian Bridge Simulation

Brownian motion can be numerically simulated on Q using the coordinate Itô form (4). With a standard Euler discretization, the scheme becomes

$$\mathbf{q}_{t_{k+1}} = \mathbf{q}_{t_k} + K(\mathbf{q}_{t_k}, \mathbf{q}_{t_k})^{kl} \Gamma(\mathbf{q}_t)_{kl} \Delta t + \sqrt{K(\mathbf{q}_{t_k}, \mathbf{q}_{t_k})_j} \Delta W_{t_k}^j \quad (6)$$

with time discretization $t_1, \dots, t_k, t_k - t_{k-1} = \Delta t$ and discrete noise $W_{t_1}, \dots, W_{t_k} \in \mathbb{R}^{Nd}$, $\Delta W_{t_k} = W_{t_k} - W_{t_{k-1}}$. Alternatively, a Heun scheme for discrete integration of the Stratonovich equation (5) results in

$$\begin{aligned} v_{t_{k+1}} &= H_i(u_{t_k}) \Delta W_{t_k}^i \\ u_{t_{k+1}} &= u_{t_k} + \frac{v_{t_{k+1}} + H_i(u_{t_k} + v_{t_{k+1}}) \Delta W_{t_k}^i}{2}. \end{aligned} \quad (7)$$

Because the horizontal fields represent infinitesimal parallel transport, they can be expressed using the Christoffel symbols of g . The Christoffel symbols for the landmark metric are derived in [15] from which they can be directly implemented or implicitly retrieved from an automatic symbolic differentiation as done in the experiments in Section 5.

3.1 Bridge Sampling

The transition density $p_T(\mathbf{v})$ of a Brownian motion \mathbf{q}_t evaluated at $\mathbf{v} \in Q$ at time $T > 0$ can be informally obtained by taking an expectation to get the “mass” of those of the sample paths hitting \mathbf{v} at time T . We write $\mathbf{q}_t | \mathbf{v}$ for the process \mathbf{q}_t conditioned on hitting \mathbf{v} a.s. at $t = T$. Computing the expectation analytically is in nonlinear situations generally intractable. Instead, we wish to employ a Monte Carlo approach and thus derive a method for simulating from $\mathbf{q}_t | \mathbf{v}$. For this, we employ the bridge sampling scheme of Delyon and Hu [6]. We first describe the framework for a general diffusion process in Euclidean space before using it directly on the landmark manifold Q .

Let

$$dx_t = b(t, x_t)dt + \sigma(t, x_t)dW_t \quad (8)$$

be an \mathbb{R}^k valued Itô diffusion with invertible diffusion field σ . In order to sample from the conditioned process $x_t|v$, $v \in \mathbb{R}^k$, a modified processes is in [6] constructed by adding an extra drift term to the process giving the new process

$$dy_t = b(t, y_t)dt - \frac{y_t - v}{T - t}dt + \sigma(t, y_t)dW_t . \quad (9)$$

As $t \rightarrow T$, the attraction term $-(y_t - v)/(T - t)dt$ becomes increasingly strong forcing the processes to hit y at $t = T$ a.s. It can be shown that the process y_t exists when b , σ and σ^{-1} are $C^{1,2}$ with bounded derivatives. The process is then absolutely continuous with respect to the conditioned process $x_t|v$. The Radon-Nikodym derivative between the laws $P_{x_t|v}$ and P_y is

$$\frac{dP_{x|v}}{dP_y}(y) = \frac{\varphi_T(y)}{E_y[\varphi_T]}$$

with $E_y[\cdot]$ denoting expectation with respect to P_y , and the correction factor $\varphi_T(y)$ defined as the $t \rightarrow T$ limit of

$$\begin{aligned} \varphi_t(y) = \exp \left(- \int_0^t \frac{\tilde{y}_s^T A(s, y_s) b(s, y_s)}{T - s} ds \right. \\ \left. - \frac{1}{2} \int_0^t \frac{\tilde{y}_s^T (dA(s, x_s)) \tilde{y}_s + \sum_{i,j} d \langle A^{ij}(s, y_s), \tilde{y}_s^i \tilde{y}_s^j \rangle}{T - s} \right) \end{aligned} \quad (10)$$

Here $\tilde{y}_t = y_t - v$, $A = (\sigma\sigma^T)^{-1}$, and quadratic variation is denoted by $\langle \cdot, \cdot \rangle$. Then $E_x[f(x)|x_T = v] = E_{x|v}[f(x)] = E_y[f(y)\varphi_T(y)]/E_y[\varphi_T(y)]$ and

$$E[f(y)\varphi_t] = \frac{T^{k/2} e^{\frac{\|\sigma^{-1}(0, x_0)(x_0 - v)\|^2}{2T}}}{(T - t)^{k/2}} E[f(x) e^{-\frac{\|\sigma^{-1}(t, x_t)(x_t - v)\|^2}{2(T-t)}}]$$

for $t < T$. The fact that the diffusion field σ must be invertible for the scheme to work as outlined here can be seen explicitly from the use of the inverse of σ and A in these equations.

We can use the guided process (9) to take conditional expectation for general measurable functions on the Wiener space of paths $W(\mathbb{R}^k, [0, T])$ by sampling from y_t . Taking the particular choice of the constant function, the expression

$$p_T(v) = \sqrt{\frac{|A(T, v)|}{(2\pi T)^k}} e^{-\frac{\|\sigma(0, x_0)^{-1}(x_0 - v)\|^2}{2T}} E_y[\varphi_T(y)] \quad (11)$$

for the transition density of x_t arise as shown in [16]. Note that both the leading factors and the correction factor φ_T are dependent on the diffusion field σ , the starting point x_0 of the diffusion, and the drift b . Again, we can approximate the expectation in (11) by sampling from y_t .

3.2 Landmark Bridge Simulation

Because the landmark manifold has a global chart on \mathbb{R}^{Nd} from the standard representation of each landmark position in \mathbb{R}^d , we can conveniently apply the bridge construction of [6]. Writing the Itô coordinate form of the Brownian motion \mathbf{q}_t (4) in the form (8), we have $b(t, \mathbf{q}) = K(\mathbf{q}, \mathbf{q})^{kl} \Gamma(\mathbf{q})_{kl}$ and $\sigma(t, \mathbf{q}) = \sqrt{K(\mathbf{q}, \mathbf{q})}$ giving the guided SDE

$$d\mathbf{y}_t = K(\mathbf{y}_{t_k}, \mathbf{y}_{t_k})^{kl} \Gamma(\mathbf{y}_t)_{kl} dt - \frac{\mathbf{y}_t - \mathbf{v}}{T - t} dt + \sqrt{K(\mathbf{y}_{t_k}, \mathbf{y}_{t_k})} dW_t \quad (12)$$

The attraction term $-(\mathbf{y}_t - \mathbf{v})/(T - t)dt$ is the difference between the current landmark configuration \mathbf{y}_t and the target configuration \mathbf{v} . The transition density becomes

$$p_{T,\theta}(\mathbf{v}) = \frac{1}{\sqrt{|K(\mathbf{v}, \mathbf{v})|(2\pi T)^{Nd}}} e^{-\frac{\|(\mathbf{q}_0 - \mathbf{v})^T K(\mathbf{q}_0, \mathbf{q}_0)^{-1} (\mathbf{q}_0 - \mathbf{v})\|^2}{2T}} E_{\mathbf{y}_\theta}[\varphi_{\theta,T}(\mathbf{y})] \quad (13)$$

where we use the subscript θ to emphasize the dependence on the parameters \mathbf{q}_0 and the kernel K . As above, the expectation $E_{\mathbf{y}_\theta}[\varphi_{\theta,T}(\mathbf{y})]$ can be approximated by drawing samples from \mathbf{y}_θ and evaluating $\varphi_{\theta,T}(\mathbf{y})$.

Remark 1. A similar scheme is used for the bridge simulation of the stochastic coadjoint processes of [3,4]. In these cases, the flow is hypoelliptic in the phase space (\mathbf{q}, \mathbf{p}) and observations are partial in that only the landmark positions \mathbf{q} are observed. The momenta \mathbf{p} are unobserved. In addition, the fact that the landmarks can carry a large initial momentum necessitates a more general form of the guidance term $-(\mathbf{y}_t - \mathbf{v})/(T - t)dt$ that takes into account the expected value of $E_{\mathbf{y}}[\mathbf{y}_T | (\mathbf{q}_t, \mathbf{p}_t)]$ of the process at time T given the current time t position and momentum of the process.

4 Inference Algorithm

Given a set of i.i.d. observations $\mathbf{q}^1, \dots, \mathbf{q}^N$ of landmark configurations, we assume the configurations \mathbf{q}^i are distributed according to the time $t = T$ transition distribution \mathbf{q}_T of a Brownian motion on Q started at \mathbf{q}_0 . We now intend to infer parameters θ of the model. With the metric structure on Q given by (3) and kernel of the form $K(\mathbf{q}_1, \mathbf{q}_2) = \text{Id}_d k(\|\mathbf{q}_1 - \mathbf{q}_2\|^2)$, $k(x) = \alpha e^{-\frac{1}{2}x^T \Sigma x}$, parameters are the starting position \mathbf{q}_0 , α , and $\Sigma = \sigma\sigma^T$, i.e. $\theta = (\mathbf{q}_0, \alpha, \sigma)$.

The likelihood of the model given the data with respect to the Lebesgue measure on \mathbb{R}^{Nd} is

$$\mathcal{L}_\theta(\mathbf{q}^1, \dots, \mathbf{q}^N) = \prod_{i=1}^N p_{T,\theta}(\mathbf{q}^i). \quad (14)$$

Using our ability to approximate (13) by bridge sampling, we aim to find a maximum-likelihood estimate (MLE) $\hat{\theta} \in \text{argmin}_\theta \mathcal{L}_\theta(\mathbf{q}^1, \dots, \mathbf{q}^N)$. We do this by a gradient based optimization on θ , see Algorithm 1. Note that the likelihood and thus the MLE of θ are dependent on the chosen background measure, in this case coming from the canonical chart on \mathbb{R}^{Nd} .

Algorithm 1: Metric estimation: Inference of parameters θ from samples.

```

for  $l = 1$  until convergence do
  for  $i = 1$  to  $N$  do
    sample  $J$  paths from guided process  $\mathbf{y}_{\theta_l}$  hitting  $\mathbf{q}^i$ 
    compute correction factors  $\varphi_{\theta_l, T}^{i, j}$ 
  end
   $\mathcal{L}_{\theta_l}(\mathbf{q}^1, \dots, \mathbf{q}^N) \leftarrow$ 
   $\prod_{i=1}^N \frac{1}{\sqrt{|K(\mathbf{q}^i, \mathbf{q}^i)|(2\pi T)^{Nd}}} e^{-\frac{\|(\mathbf{q}_0 - \mathbf{q}^i)^T K(\mathbf{q}_0, \mathbf{q}_0)^{-1}(\mathbf{q}_0 - \mathbf{q}^i)\|^2}{2T}} \frac{1}{J} \sum_{j=1}^J \varphi_{\theta_l, T}^{i, j}$ 
   $\theta_{l+1} = \theta_l + \epsilon \nabla_{\theta_l} \mathcal{L}_{\theta_l}(\mathbf{q}^1, \dots, \mathbf{q}^N)$ 
end

```

Remark 2. The inference Algorithm 1 optimizes the likelihood \mathcal{L}_θ directly by stochastic gradient descent. A different but related approach is an Expectation-Maximization approach where the landmark trajectories between $t = 0$ and the observation time $t = T$ are considered missing data. The E -step of the EM algorithm would then involve the expectation $E_{\mathbf{x}|\mathbf{q}^i}[\log p(\mathbf{x})]$ of the landmark bridges conditioned on the data with $p(\mathbf{x})$ formally denoting a likelihood of an unconditioned sample path \mathbf{x} . This approach is used in e.g. [4]. While natural to formulate, the approach involves the likelihood $p(\mathbf{x})$ of a stochastic path which is only defined for finite time discretizations. In addition, the expected correction factor $E_{\mathbf{y}}[\varphi_T(\mathbf{y})]$ that arise when using the guided process \mathbf{y} in the estimation appears as a normalization factor in the EM Q -function. This can potentially make the scheme sensitive to the stochasticity in the Monte Carlo sampling of the expected correction $E_{\mathbf{y}}[\varphi_T(\mathbf{y})]$. While the differences between these approaches needs further investigation, we hypothesize that direct optimization of the likelihood is superior in the present context.

Remark 3. Instead of taking expectations over \mathbf{q}_T , we can identify the most probable path of the conditioned process $\mathbf{q}_t|\mathbf{v}$. This results in the Onsager-Machlup functional [8]. In [18], a different definition is given that, in the isotropic Brownian motion situation, makes the set of Riemannian geodesics from \mathbf{q}_0 to \mathbf{v} equal to the set of most probable paths of the conditioned process $\mathbf{q}_t|\mathbf{v}$. The sample Frechét mean

$$\operatorname{argmin}_{\mathbf{q}_0} \frac{1}{N} \sum_{i=1}^N d_g(\mathbf{q}_0, \mathbf{q}_i)^2 \quad (15)$$

is in that case formally also a minimizer of the negative log-probability of the most probable path to the data. Given that we are now able to approximate the density function of the Brownian motion, the MLE of the likelihood (14) with respect to \mathbf{q}_0 is equivalent to

$$\operatorname{argmin}_{\mathbf{q}_0} - \frac{2}{N} \sum_{i=1}^N \log p_{T, \mathbf{q}_0}(\mathbf{q}_i) . \quad (16)$$

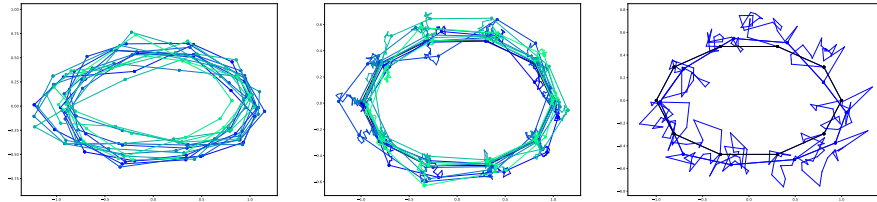


Fig. 2. (left) Samples from the transition distribution of a Brownian motion perturbing an ellipse configuration with 10 landmarks. (center) Subset of the samples shown together with trajectories of the stochastic landmark Brownian motion started at configuration \mathbf{q}_0 (black points). (right) A guided bridge from \mathbf{q}_0 (black points) to a sample (blue points).

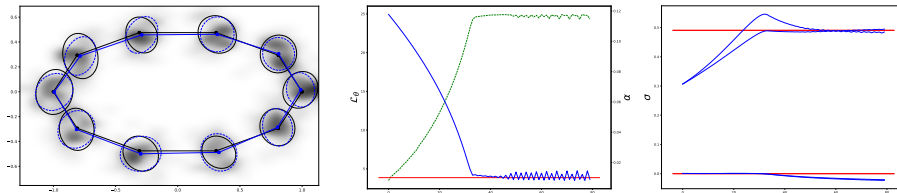


Fig. 3. (left) Result of the inference algorithm applied to the synthetic ellipse samples. The initial configuration \mathbf{q}_0 (black points) is showed along with the per-landmark sample covariance from the samples (black ellipses). The estimated initial configuration $\hat{\mathbf{q}}_0$ (blue points) is shown along with the per-landmark sample covariance from a new set of samples obtained with the inferred parameters (blue ellipses). (center) Evolution of likelihood (green) and α (blue) during the optimization. Horizontal axis shows number of iterations and red line is α ground truth. (right) Evolution of the entries of the kernel covariance σ (blue lines) during the optimization, red lines ground truth.

Compared to (15), the negative log-probability of the data is here minimized instead of the squared geodesic distance. The estimator (16) can therefore be considered a transition density equivalent of the sample Fréchet mean.

5 Numerical Experiments

We here present examples of the method on simulated landmark configurations, and an application of the method and algorithm to landmarks annotated on cardiac images of left ventricles. We aim for testing the ability of the algorithm to infer the parameters of the model given samples. We here take the first steps in this direction and leave a more extensive simulation study to future work.

For the simulated data, we compare the results against the true values used in the simulation. In addition, we do simple model checking for both experiments by simulating with the estimated parameters and comparing the per-landmark sample mean and covariance.

We use code based on the Theano library [23] for the implementation, in particular the symbolic expression and automatic derivative facilities of Theano. The

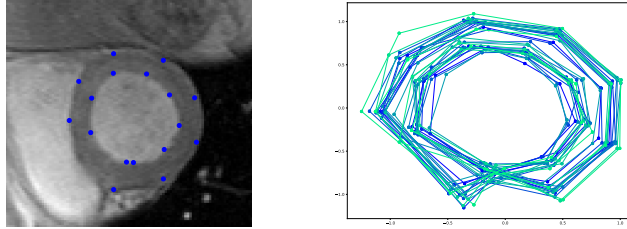


Fig. 4. (left) An image of a left cardiac ventricle with annotations. (right) The annotations from 14 cardiac images.

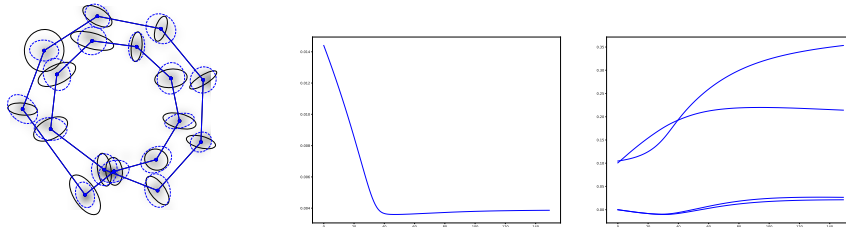


Fig. 5. Results of applying the inference algorithm to the cardiac data. Setup and subfigures as in Figure 3.

code used for the experiments is available in the software package *Theano Geometry* <http://bitbucket.org/stefansommer/theanogeometry>. The implementation and the use of Theano for differential geometry applications including landmark manifolds is described in [12].

With 10 landmarks arranged in an ellipse configuration \mathbf{q}_0 , we sample 64 samples from the transition distribution at time $T = 1$ of a Brownian motion started at \mathbf{q}_0 , see Figure 2. Parameters for the kernel are $\sigma = \text{diag}(\sigma_1, \sigma_2)$ with σ_1, σ_2 set to the average inter-point distance in \mathbf{q}_0 , and the amplitude parameter $\alpha = 0.01$.

We run Algorithm 1 with initial conditions for \mathbf{q}_0 the pointwise mean of the samples. The parameter evolution through the iterative optimization and the result of the inference can be seen in Figure 3. The algorithm is able to estimate the initial configuration and the parameters of α and Σ with a reasonable precision. The sample per-landmark covariance as measured on a new set of simulated data with the estimated parameters is comparable to the per-landmark covariance of the original dataset.

5.1 Left Cardiac Ventricles

To exemplify the approach on real data, we here use a set of landmarks obtained from annotations of the left ventricle in 14 cardiac images [22]. Each ventricle is annotated with sets of landmarks from which we select 17 from each configura-

tion for use in this experiment. Figure 4 shows an annotated image along with the sets of annotations for all images.

Figure 5 shows the results of the inference algorithm with setup equivalent to Figure 3. While the parameters converges during the iterative optimization, we here have no ground-truth comparison. A subsequent sampling using the estimated parameters allows comparison of the per-landmark sample covariance. While the new sample covariance in magnitude and to some degree shape corresponds to the sample covariance from the original data, the fact that the Brownian motion is isotropic forces the covariance to be equivalent for all landmarks as measured by the Riemannian landmark metric. Including anisotropic covariance in the distribution or the right-invariant stochastics of [3,4] would allow the per-landmark covariance to vary and result in a closer fit.

6 Conclusion

In the paper, we have derived a method for maximum likelihood estimation of parameters for the starting point of landmark Brownian motions and for the Riemannian metric structure specified from the kernel K . Using the guided process scheme of [6] for sampling conditioned Brownian bridges, the transition density is approximated by Monte Carlo sampling. With this approximation of the data likelihood, we use a gradient based iterative scheme to optimize parameters. We show on synthetic and real data sets the ability of the method to infer the underlying parameters of the data distribution and hence the metric structure of the landmark manifold.

A direct extension of the method presented here is to generalize to the anisotropic normal distributions [17] defined via Brownian motions in the frame bundle FQ . This would allow differences in the per-landmark covariance and thus improve results on datasets such as the presented left ventricle annotations. Due to the hypoellipticity of the anisotropic flows that must be conditioned on hitting fibers in FQ above points $q \in Q$, further work is necessary to adapt the scheme presented here to the anisotropic case.

Acknowledgements We are grateful for the use of the cardiac ventricle dataset provided by Jens Chr. Nilsson and Bjørn A. Grønning, Danish Research Centre for Magnetic Resonance (DRCMR).

References

1. Allasonnire, S., Amit, Y., Trouv, A.: Towards a coherent statistical framework for dense deformable template estimation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 69(1), 3–29 (Feb 2007)
2. Arnaudon, A., Castro, A.L., Holm, D.D.: Noise and dissipation on coadjoint orbits. arXiv:1601.02249 [math-ph, physics:nlin] (Jan 2016)
3. Arnaudon, A., Holm, D.D., Pai, A., Sommer, S.: A Stochastic Large Deformation Model for Computational Anatomy. *Information Processing in Medical Imaging (IPMI)* (2017)

4. Arnaudon, A., Holm, D.D., Sommer, S.: A Geometric Framework for Stochastic Shape Analysis. submitted, arXiv:1703.09971 [cs, math] (Mar 2017)
5. Beg, M.F., Miller, M.I., Trouv, A., Younes, L.: Computing Large Deformation Metric Mappings via Geodesic Flows of Diffeomorphisms. *IJCV* 61(2) (2005)
6. Delyon, B., Hu, Y.: Simulation of conditioned diffusion and application to parameter estimation. *Stochastic Processes and their Appl.* 116(11), 1660–1675 (2006)
7. Dupuis, P., Grenander, U., Miller, M.I.: Variational Problems on Flows of Diffeomorphisms for Image Matching (1998)
8. Fujita, T., Kotani, S.i.: The Onsager-Machlup function for diffusion processes. *Journal of Mathematics of Kyoto University* 22(1), 115–130 (1982)
9. Holm, D.D.: Variational principles for stochastic fluid dynamics. *Proceedings. Mathematical, Physical, and Eng. Sciences / The Royal Society* 471(2176) (2015)
10. Hsu, E.P.: Stochastic Analysis on Manifolds. American Mathematical Soc. (2002)
11. Joshi, S., Miller, M.: Landmark matching via large deformation diffeomorphisms. *Image Processing, IEEE Transactions on* 9(8), 1357–1370 (2000)
12. Kühnel, L., Arnaudon, A., Sommer, S.: Deep Learning Numerics for Differential Geometry. in preparation (2017)
13. Markussen, B.: Large deformation diffeomorphisms with application to optic flow. *Comput. Vis. Image Underst.* 106(1), 97–105 (Apr 2007)
14. Marsland, S., Shardlow, T.: Langevin equations for landmark image registration with uncertainty. arXiv:1605.09276 [math] (May 2016), arXiv: 1605.09276
15. Micheli, M.: The differential geometry of landmark shape manifolds: metrics, geodesics, and curvature. Ph.D. thesis, Brown University, Providence, USA (2008)
16. Papaspiliopoulos, O., Roberts, G.O.: Importance sampling techniques for estimation of diffusion models. In: *Statistical Methods for Stochastic Differential Equations*. Chapman & Hall/CRC Press (2012)
17. Sommer, S.: Anisotropic Distributions on Manifolds: Template Estimation and Most Probable Paths. In: *Information Processing in Medical Imaging*. Lecture Notes in Computer Science, vol. 9123, pp. 193–204. Springer (2015)
18. Sommer, S.: Anisotropically Weighted and Nonholonomically Constrained Evolutions on Manifolds. *Entropy* 18(12), 425 (Nov 2016)
19. Sommer, S., Jacobs, H.O.: Reduction by Lie Group Symmetries in Diffeomorphic Image Registration and Deformation Modelling. *Symmetry* 7(2), 599–624 (2015)
20. Sommer, S., Joshi, S.: Brownian Bridge Simulation and Metric Estimation on Lie Groups and Homogeneous Spaces. in preparation (2017)
21. Sommer, S., Svane, A.M.: Modelling Anisotropic Covariance using Stochastic Development and Sub-Riemannian Frame Bundle Geometry. Accepted for publication in *Journal of Geometric Mechanics*, arXiv:1512.08544 [math, stat] (2016)
22. Stegmann, M.B., Fisker, R., Ersbll, B.K.: Extending and applying active appearance models for automated, high precision segmentation in different image modalities. *Proc. 12th - SCIA 2001, Bergen, Norway* pp. 90–97 (2001)
23. Team, T.T.D.: Theano: A Python framework for fast computation of mathematical expressions. arXiv:1605.02688 [cs] (May 2016), arXiv: 1605.02688
24. Trouv, A.: An Infinite Dimensional Group Approach for Physics Based Models in *Patterns Recognition* (1995)
25. Trouv, A., Vialard, F.X.: Shape splines and stochastic shape evolutions: A second order point of view. *Quarterly of Applied Mathematics* 70(2), 219–251 (2012)
26. Vialard, F.X.: Extension to infinite dimensions of a stochastic second-order model associated with shape splines. *Stochastic Processes and their Applications* 123(6), 2110–2157 (Jun 2013)