Learning Shape Trends: Parameter Estimation in Diffusions on Shape Manifolds

Valentina Staneva
University of Washington
Seattle, WA, USA
vms16@uw.edu

Laurent Younes
Johns Hopkins University
Baltimore, MD, USA
laurent.younes@jhu.edu

Abstract

Learning the dynamics of shape is at the heart of many computer vision problems: object tracking, change detection, longitudinal shape analysis, trajectory classification, etc. In this work we address the problem of statistical inference of diffusion processes of shapes. We formulate a general Itô diffusion on the manifold of deformable landmarks and propose several drift models for the evolution of shapes. We derive explicit formulas for the maximum likelihood estimators of the unknown parameters in these models, and demonstrate their convergence properties on simulated sequences when true parameters are known. We further discuss how these models can be extended to a more general non-parametric approach to shape estimation.

1. Introduction

The subject of studying diffusions of shapes dates back to the work of Kendall [13], where Brownian motion is considered on the space of points in $\mathbb{R}^n$ after excluding similarity transformations. A more recent work by Ball et. al. [3] adds an additional drift term to the random perturbations of the shapes to construct Ornstein-Uhlenbeck processes in the appropriate Kendall and Goodal-Mardia coordinates. The authors obtain the stationary distributions of the proposed processes which facilitates the parameter estimation. In our work we focus on shapes which do not change their topology so their deformations can be appropriately described by stochastic flows of diffeomorphisms [14]. Such processes have been studied in the context of images in [6], in the context of landmarks in [22], and extended to the infinite-dimensional spaces of curves and surfaces in [21]. Stochastic fluid flows derived from variational principles are introduced in [10], and [1] addresses the problem of noise estimation in such models. In the context of template and variance estimation, the authors in [16] and [17] propose to look for most probable paths as realizations of a diffusion process on the manifold of landmarks to eliminate the need to resort to linearization of the nonlinear space.

While these previous works are motivated by the computational anatomy problems of registration under uncertainty and variance estimation, and concentrate on obtaining random shape perturbations as deviations from geodesic paths, we focus on the task of constructing more informative parametric deformations for shapes (here 2D contours) and estimating their underlying parameters from a sequence of observations. Our approach is geometric, we define a diffusion process on the landmark manifold whose drift terms describe the shape's intrinsic 'trend' and are obtained as a gradient of a function over the shape. The unknown parameters represent the strength of the effect of the drift term and Girsanov formula provides the form of the likelihood ratio with respect to the law of a process with no drift, which eventually is maximized to obtain their estimates[1].

Organization. We begin with a brief introduction of the geometry of the landmark manifold. In Section 2.1 we describe several different ways for constructing diffusions on manifolds and highlight the ones which are suited for numerical simulation in our context. Next in Section 3, we define the noise model of our choice and relate the corresponding diffusion to Brownian motion on a Riemannian or sub-Riemannian manifold. In Section 4 we introduce several drift models and in Section 5 we show sample shape paths they can yield. In Section 6 we address the task of estimating the missing diffusion parameters.

2. Diffusions of Shapes

We represent the boundary of the shape by a sequence of m distinct points in $\mathbb{R}^2$ denoted by χ. The space of all such discretized contours $\mathcal{M}$ forms a $2m$-dimensional manifold as described in Section 30.3.2 [20]. Fortunately, many geometric objects on this manifold can be derived and computed: geodesics, exponential map, parallel transport [25], curvature [15], etc. We define the Riemannian metric on the manifold of landmarks by

$$\|\mathbf{c}\|_\chi^2 = \mathbf{c}^T \mathbf{K}(\chi)^{-1} \mathbf{c}, \tag{1}$$

[1] this work appears in [18]

where $\mathbf{c}$ is a tangent vector on the shape manifold. We assume that the matrix $\mathbf{K}(\chi)$ is block-diagonal and each block consists of the evaluations $K(x_i, x_j) = e^{-\|x_i - x_j\|/2\sigma^2}\mathbb{I}_2$, where x_i and x_j are two points in $\mathbb{R}_2$, and $\mathbb{I}_2$ is the 2×2 identity matrix. Note that this representation does not include invariance under transformations such as translation/scale. Often when we observe a sequence of shapes, we are interested in their motion or growth, which requires studying both affine and more general nonlinear deformations.

Our goal is to define diffusion equations on $\mathcal{M}$ of the form

$$d\chi_t = A(\chi_t, \theta)dt + B(\chi_t)dW_t, \tag{2}$$

where χ_t specifies a process on $\mathcal{M}$, $A(\chi_t, \theta)$ is an element of $T_{\chi_t}\mathcal{M}$ with a parameter θ, and $B(\chi_t)dW_t$ corresponds to a Brownian motion on $\mathcal{M}$ with a mixing matrix $B(\chi_t)$ whose details we will specify later. In the following sections, we will denote by M a general d-dimensional Riemannian manifold, with a generic element X or x. Specializing the discussion to landmark manifold results, we take $M = \mathcal{M}$, $d = 2m$, and a generic element will be denoted by χ.

2.1. Diffusions on Manifolds

There are several approaches for formulating diffusion processes (and their corresponding SDE's) on manifolds. Here we introduce the ones relevant to our work; for a more extensive treatment of the topic one can refer to the rich literature in [7, 11, 5]. We discuss both Stratonovich and Itô formulations.

Stratonovich SDE's on Manifolds. Since Stratonovich calculus follows classical differentiation rules, it is easy to define Stratonovich SDE's in local coordinates, which would appropriately transform under change of coordinates. For that we simply define A to be a smooth vector field on a d-dimensional manifold M and $B(X_t)$ to be a mapping from $\mathbb{R}^n$ to $T_{X_t}M$ ($n \leq d$) at each X_t, and define the stochastic differential equation on M as

$$dX_t = A(X_t)dt + B(X) \circ dW_t, \tag{3}$$

where W_t is an n-dimensional Brownian motion, and B converts a Brownian motion on $\mathbb{R}^n$ to a process on the tangent bundle of M (a more general formulation allows for a time-dependent drift $A(X_t, t)$, but will focus on the time-independent case). A solution of (3) is any process X_t which satisfies the above equation in any local chart. Let $\{a^i\}_{i=1}^d$ and $\{b_k^i\}_{i=1}^d$ be the coefficients of the vector fields $A, B_1, ...B_n$ in the local chart. The Stratonovich equation then takes the form

$$dX^i(t) = a^i(X_t)dt + \sum_{k=1}^{n} b_k^i(X_t) \circ dw_k(t), \qquad i = 1, ..., d, \tag{4}$$

and under change of parameterization $\varphi : \mathbb{R}^d \to \mathbb{R}^d$ the equations transform according to

$$\begin{aligned} d\varphi^i(X_t) &= \sum_{j=1}^{n} \partial_j \varphi^i a^j(\varphi(X_t)) + \\ &+ \sum_{j=1}^{n}\sum_{k=1}^{n} \partial_j \varphi^i b_k^j(\varphi(X_t)) \circ dw_k(t). \end{aligned} \tag{5}$$

Alternatively, we can consider n smooth vector fields on M denoted as $B_1, ..., B_n$ ($B_i : M \to TM$ for $i = 1, ..., n$) and define the SDE

$$dX_t = A(X_t)dt + \sum_{k=1}^{n} B_k(X_s) \circ dw_k(s), \tag{6}$$

whose solution satisfies for any smooth function with compact support on $\mathcal{M}$

$$f(X_t) - f(X_0) = \int_0^t Af(X_s)ds + \int_0^t \sum_{k=1}^{n} B_k f(X_s) \circ dw_k(s). \tag{7}$$

Selecting $n = d$ is not necessary, however, this choice becomes important in the special case when the manifold of interest is parallelizable, i.e. when there exists a global frame of vector fields on M. Then the local representation becomes global. Fortunately, this is true when $M = \mathcal{M}$. By evaluating the kernel at each individual point on the curve, we obtain a basis of vector fields on the tangent space: $\{K(\chi, x_1)e_p, ..., K(\chi, x_m)e_p\}$ ($p = 1, 2$), which varies smoothly when changing the points, i.e. we have $2m$ smooth vector fields which when evaluated at a fixed point form a basis for the tangent space at that point. We denote them by $E_1(\chi), ..., E_{2m}(\chi)$. Then we can write the above SDE in this basis

$$d\chi_t = \sum_{k=1}^{2m} \alpha_k(\chi_t)E_k(\chi_t)dt + \sum_{k=1}^{2m} E_k(\chi_t) \circ dw_k(t). \tag{8}$$

Itô SDE's on Manifolds. In a given coordinate chart we can define the following Itô equation [12]:

$$dX^i(t) = \hat{a}^i(X_t)dt + \sum_{k=1}^{n} b_k^i(X_t) \cdot dw_k(t), \quad i = 1, ..., 2m, \tag{9}$$

where the pairing $b \cdot dw$ corresponds to the classical Itô differential. Under change of coordinates, Itô equations are required to satisfy Itô's formula:

$$\begin{aligned} d\varphi^i(X_t) &= \sum_{j=1}^{d} \partial_j \varphi^i \hat{a}^j(X_t)dt + \\ &+ \frac{1}{2}\sum_{j=1}^{d}\sum_{k=1}^{n}\sum_{l=1}^{n} \partial_{kl}\varphi^i b_k^j(X_t)b_l^j(X_t)dt + \\ &+ \sum_{j=1}^{d}\sum_{k=1}^{n} \partial_j \varphi^i b_k^j(X_t) \cdot dw_k(t), \end{aligned} \tag{10}$$

and the equation can be converted from Itô to Stratonovich form and vice versa using the standard rules

$$\begin{aligned} dX^i(t) &= \left[\hat{a}^i(X_t) - \frac{1}{2}\sum_{j=1}^{d}\sum_{k=1}^{n} b_k^j(X_t)\partial_j b_k^i(X_t)\right]dt + \\ &+ \sum_{k=1}^{n} b_k^i(X_t)\circ dw_k(t), \quad i = 1, ..., d. \end{aligned} \tag{11}$$

One can observe that in order for the Itô formula to be satisfied, under change of coordinates the transformation of $\hat{a}$ has to depend on the b_k's, i.e. it cannot be defined as a vector field on the manifold.

Diffusions through the Riemannian Exponential Map. An alternative approach to defining Itô equations resorts to the Riemannian structure on the manifold of interest and the associated exponential map (Baxendale[4], Belopolskaya-Daletsky[5] forms). Let $\exp_X : T_X M \to M$ be the Riemannian exponential map on M and consider the solution of equation

$$dX_t = \exp_{X_t}(A(X_t)dt + B(X_t)dW_t), \tag{12}$$

where the *forward stochastic differential*

$$A(X_t)dt + B(X_t)dW_t \tag{13}$$

corresponds to the class of diffusion processes in $T_{X_t}M$ whose drift and noise terms coincide locally with A and B, i.e. they satisfy the equation

$$u(t+s) = \int_t^{t+s} \tilde{A}(u_\tau)d\tau + \int_t^{t+s} \tilde{B}(u_\tau)dw_\tau, \tag{14}$$

where $\tilde{A}(u_\tau)$ is a vector field on $T_{X_t}M$, and $\tilde{B}(u_\tau) : \mathbb{R}^n \to T_{X_t}M$, $\tilde{A}(0) = A(X_t)$ and $\tilde{B}(0) = B(X_t)$ in a neighborhood around the origin of $T_{X_t}M$ and zero outside.

The Taylor series expansion for any curve $X(t)$ on M in local chart is

$$X(t) = X(0) + t\dot{X}(0) + \frac{1}{2}t^2\ddot{X}(0) + o(t^2), \tag{15}$$

which provides an approximation for the exponential map

$$\exp_X(tv) = X + tv - \frac{1}{2}t^2\mathbf{\Gamma}_X(v, v) + o(t^2), \tag{16}$$

where $\mathbf{\Gamma}_X\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right) = \Gamma_{ij}^k \frac{\partial}{\partial x^k}$ and Γ_{ij}^k the Christoffel symbols associated with the connection on M. Using this expansion one can obtain a local chart formulation of (12):

$$\begin{aligned} dX_t = A(X_t)dt - \frac{1}{2}\sum_{k=1}^{d}\Gamma_{ij}^k \sum_{l=1}^{n} b_l^i(X_t)b_l^j(X_t)\frac{\partial}{\partial x^k}dt + \\ + \sum_{k=1}^{n} b_k(X_t)\cdot dw_k(t). \end{aligned} \tag{17}$$

We observe that the drift coefficients of the Itô equation contain a correction term due to the non-flatness of the manifold:

$$\hat{a}^k = a^k - \frac{1}{2}\Gamma_{ij}^k \sum_{l=1}^{n} b_l^i(X_t)b_l^j(X_t). \tag{18}$$

Diffusions as a Limit of a Random Walk on a Manifold. Intuitively, a diffusion should be a limit of small steps on the manifold in a given direction with noise added to them. Baxendale [4] introduces an approach closest to this idea.

Let μ be a zero mean Gaussian measure on $C(TM)$ (the continuous vector fields on M), and let W_t be the associated Wiener process. Let A be a smooth vector field on M. Set $V_t = tA + W_t$. Define a partition $\pi = \{t_0 = 0, t_1, ..., t_N = T\}$. Suppose $X_{t_j}^\pi$ satisfies

$$X_{t_{j+1}}^\pi = \exp_{X_{t_j}^\pi}(\Delta_j V_t(X_{t_j})), \quad j = 0, ..., N-1. \tag{19}$$

It can be shown that under suitable conditions $X_{t_j}^\pi$ converges to a Markov process X_t on $[0, T]$ as the mesh π becomes denser, which corresponds to the Itô SDE:

$$dX_t = dV_t = Adt + dW_t. \tag{20}$$

and justifies the discretization scheme below for small values of dt

$$X_{t+dt} = \exp_{X_t}(A(X_t)dt + B(X_t)W_t). \tag{21}$$

On the shape manifold we can numerically compute the exponential map by solving a system of ordinary differential equations (in Hamiltonian form [20], eq. 30.5). Thus the Baxendale/Belopolskaya/Daletsky approach of defining a diffusion process through the exponential map is preferable for simulating diffusion paths and eliminates the need to resort to Stratonovich equations or computing correction terms.

3. Noise Models

In this section we define the form of the vector fields $B_1, ..., B_n$, which in turn determines the covariance of the noise in the diffusion equations. We temporarily assume the drift $A(\chi_t)$ is zero, i.e. the motion of individual points is driven only by the mixing of the individual Brownian motions. There is zero mixing when $B_1, ..., B_n$ form an orthonormal frame, and in that case the process corresponds to a standard Brownian motion on the manifold. This requires that the local coordinates satisfy $b_k^i g_{ij} b_l^j = \delta_{k,l}$. When $n = m$ this is equivalent to $b_k^i b_k^j = g^{ij}$, where g^{ij} are the coefficients of the inverse of the Riemannian metric. Let $b_l^i(\chi_t)$ be the il'th entry of $\mathbf{\Sigma}(\chi_m)$, then in matrix form this entails to:

$$\mathbf{\Sigma}(\chi_m)\mathbf{\Sigma}(\chi_m)^T = \mathbf{K}(\chi_m, \chi_m). \tag{22}$$

Since $\mathbf{K}$ is symmetric positive-definite, a unique symmetric positive-definite square root exists:

$$\mathbf{\Sigma}(\chi_m) = \mathbf{K}(\chi_m, \chi_m)^{1/2}. \tag{23}$$

Using the exponential map formulation the Brownian motion on the shape manifold can be written as

$$d\chi_t = \exp_{\chi_t}(\mathbf{\Sigma}(\chi_t)dW_t). \tag{24}$$

Suppose χ_n is a subset of the points in χ_m. We can reduce the computational burden of computing the exponential map by restricting the random deformations to be driven by the motion of these points (hence called them control points [2, 24, 19]). For that we require that the vector fields driving the deformations belong to the tangent bundle of the submanifold associated with this subset of points. Geometrically, we restrict to horizontal paths on a sub-Riemannian manifold, and if we set $\mathbf{\Sigma}(\chi_m)$ to be the positive-semidefinite square root of $\mathbf{K}(\chi_m, \chi_n)\mathbf{K}(\chi_n, \chi_n)^{-1}\mathbf{K}(\chi_n, \chi_m)$:

$$\mathbf{\Sigma}(\chi_m) = \mathbf{K}(\chi_m, \chi_n)\mathbf{K}(\chi_n, \chi_n)^{-1/2}, \tag{25}$$

(24) is the horizontal Brownian motion on the shape manifold as defined in [9].

4. Drift Models

Next we propose several models for the diffusion drifts.

4.1. Constant Drift

A simple drift formulation assumes the coefficients of the vector field are constant with respect to some basis of vector fields $E_1, ...E_n$ on $\mathcal{M}$:

$$A(\chi_t, \theta) = \sum_{i=1}^{n} E_i(\chi_t)\theta_i. \tag{26}$$

The choice of a basis is important: a drift constant with respect to one basis may not be constant with respect to another one. We will consider two special cases: the basis obtained by evaluating the kernel at a set of control points (i.e. the covector associated with the drift is fixed):

$$A(\chi_t, \boldsymbol{\theta}) = \sum_{i=1}^{n} K(\chi_t, x_i)\theta_i = \mathbf{K}(\chi_t)\boldsymbol{\theta}, \tag{27}$$

or an orthonormal basis as defined in Section 3

$$A(\chi_t, \theta) = \mathbf{\Sigma}(\chi_t)\theta. \tag{28}$$

We allow for the vector fields $E_1(\chi_t), ..., E_n(\chi_t)$ to span only a subspace of the tangent space when $n < m$.

The diffusion process with this drift written in exponential map form

$$d\chi_t = \exp_{\chi_t}(\mathbf{K}(\chi_t)\boldsymbol{\theta}dt + \mathbf{\Sigma}(\chi_t)dW_t), \tag{29}$$

can be interpreted as a random walk with a fixed trend.

4.2. Shape Gradient Drifts

Although intuitively simple, the constant drift model does not preserve its properties under change of coordinates. Thus, we are urged to construct models which possess drift terms with intrinsic properties. A natural approach is to consider a "potential" function $U :\to \mathbb{R}$, and the corresponding stochastic gradient flow with a drift $-\nabla U$. Thus the drift will "push" the process toward the minimizer of the potential function. The "strength" of this push can be determined by a parameter θ, yielding a process

$$d\chi_t = \exp_{\chi_t}(\theta\nabla U(\chi_t) + \mathbf{\Sigma}(\chi_t)dW_t), \tag{30}$$

where $B(\chi_t)$ can represent the Brownian motion coefficients in a Riemannian or sub-Riemannian sense, and the diffusion is in Belopolskaya-Daletsky form.

Riemannian gradient. As U is a function on the Riemannian manifold $\mathcal{M}$, the gradient is a vector field on $\mathcal{M}$ and satisfies for any other vector field v:

$$(dU|v)_\chi = \langle \nabla U, v\rangle_\chi, \tag{31}$$

where $(dU|v)_\chi$ represents the action of the differential of U on v evaluated at χ.

In local coordinates the inner product can be written as

$$\langle \nabla U, v\rangle_\chi = (\nabla U_\chi^{\mathcal{M}})^T\mathbf{K}(\chi, \chi)^{-1}\mathbf{v}_\chi, \tag{32}$$

where by $\nabla U_\chi^{\mathcal{M}}$ we denote the evaluation of the Riemannian gradient to distinguish from the Euclidean gradient ∇U_χ. Further, the action of the differential can be written in terms of the Euclidean inner product

$$(dU|v)_\chi = \nabla U_\chi^T\mathbf{v}_\chi. \tag{33}$$

The condition in (31) becomes

$$\nabla U_\chi^T\mathbf{v}_\chi = (\nabla U_\chi^{\mathcal{M}})^T\mathbf{K}(\chi, \chi)^{-1}\mathbf{v}_\chi, \tag{34}$$

hence the form of the Riemannian gradient is

$$\nabla U_\chi^{\mathcal{M}} = \mathbf{K}(\chi, \chi)\nabla U_\chi. \tag{35}$$

Horizontal gradient. When working on a sub-Riemannian manifold and equipped only with a sub-Riemannian metric, we resort to the definition of a horizontal gradient: a vector field in the distribution $\mathcal{H}$ which satisfies for every horizontal vector field v

$$(dU|v) = \langle \nabla^{\mathcal{H}} U, v\rangle_{\mathcal{H}}. \tag{36}$$

Let's assume the form of the horizontal gradient in local coordinates is $\nabla^{\mathcal{H}}_{\chi_m} U = \mathbf{K}(\chi_m, \chi_n)\boldsymbol{\alpha}$ and the form of an arbitrary vector field in local coordinates is $\mathbf{v} = \mathbf{K}(\chi_m, \chi_n)\boldsymbol{\beta}$. According to (36) the coefficients of the horizontal gradient need to satisfy

$$\nabla U_{\chi_m}^T\mathbf{K}(\chi_m, \chi_n)\boldsymbol{\beta} = \boldsymbol{\alpha}^T\mathbf{K}(\chi_n, \chi_n)\boldsymbol{\beta}, \tag{37}$$

i.e.

$$\boldsymbol{\alpha} = \mathbf{K}(\chi_n, \chi_n)^{-1}\mathbf{K}(\chi_n, \chi_m)\nabla_{\chi_m} U, \tag{38}$$

and therefore

$$\nabla^{\mathcal{H}}_{\chi_m} U = \mathbf{K}(\chi_m, \chi_n)\mathbf{K}(\chi_n, \chi_n)^{-1}\mathbf{K}(\chi_n, \chi_m)\nabla_{\chi_m} U. \quad (39)$$

This formulation will be used in the models in the next two sections.

4.2.1 Mean-reverting drift

In this section we mathematically describe a process for which the shape is free to vary from step to step but in the long run does not deviate much from a fixed template shape. We are motivated by the following definition of the Ornstein-Uhlenbeck process on $\mathbb{R}$:

$$dX_t = \theta(\mu - X_t)dt + dW_t, \quad \theta > 0. \quad (40)$$

Like Brownian motion, this process is Gaussian and Markovian, however, it also admits a stationary distribution (and is the unique process which possesses these three properties simultaneously). The stationary behavior can be understood by rewriting the drift of the process as a gradient of a function:

$$dX_t = \exp_{\chi_t}(\theta\nabla_{X_t}\left[-\frac{(X_t-\mu)^2}{2}\right] + dW_t). \quad (41)$$

The drift of the process can now be interpreted as a force pushing toward the minimizer of the squared distance between X_t and the fixed element μ, and, due to the stochastic effect of the Brownian motion term, the process ends up oscillating around the mean (hence the commonly used name "mean-reverting" process). To transfer this idea to the space of shapes, we define the potential function as $U = dist(\chi_t, \mu)$ where μ is a mean shape (in practice, it can be represented by a template shape calculated from a set of training data). In a Riemannian framework it is natural to take this distance to represent the length of the geodesic path connecting X_t and μ, i.e. solution of the following minimization problem:

$$dist(\chi_t, \mu) = \min_{v:\ \chi_t = \exp_\mu(v)} \|v\| \quad (42)$$

The calculation of the gradient of the distance involves computing the logarithm map between χ_t and μ, which requires solving an optimization problem, and relies on having the correct correspondence of the landmarks of the two shapes. To simplify this step, we instead define U to be a function which measures the area of mismatch between the shapes determined by χ_t and μ. Since area is invariant under parameterization, it is an intrinsic geometric property. For that first we can construct two binary images of the same predetermined size I_{χ_t} and I_μ which are nonzero in the interior of the corresponding contours. Then we let $U = |I_{\chi_t} - I_\mu|$, i.e. the number of mismatched pixels. The continuous version of U would give us the area of mismatch of the two regions, and can be written as an integral of a function over the region enclosed by χ_t (denoted by Ω_{χ_t}):

$$U(\chi_t) = \int_{\mathbb{R}^2} |I_{\chi_t} - I_\mu| dx \quad (43)$$

$$\propto \int_{\Omega_{\chi_t}} \underbrace{(|1 - I_\mu(x)| - |0 - I_\mu(x)|)}_{F(x)} dx. \quad (44)$$

We observe that the function $U(\chi_t)$ can be written as an integral of a function over the domain of the shape χ_t: $\int_{\Omega_{\chi_t}} F(x)dx$. We can obtain the gradient by applying the divergence theorem to convert the integral to one over the boundary of the region which can be further discretized to obtain an explicit form. We provide details in Section 4.2.3.

The parameter θ determines how strongly the shape is attracted to the mean shape:

$$d\chi_t = \exp_{\chi_t}\left(-\theta\nabla^{\mathcal{M}} U(\chi_t)dt + \mathbf{\Sigma}(\chi_t)dW_t\right). \quad (45)$$

This model can easily be generalized to the case when we have multiple template shapes $\mu_1, ..., \mu_p$ and we would like to learn how the object is attracted to each of them. We can consider

$$d\chi_t = \exp_{\chi_t}\left(-\sum_{i=1}^{p}\theta_i\nabla^{\mathcal{M}} dist(\chi_t, \mu_i) + \mathbf{\Sigma}(\chi_t)dW_t\right). \quad (46)$$

We call this a "regression drift".

4.2.2 Shape descriptor drifts

In the absence of a template shape, we consider more general characteristics of the shape. For example, suppose that we have knowledge about the average length L_μ and area A_μ of the object. Since these are scalar measures, the potential function can simply be set to the quadratic deviation of the shape's length and area from the mean values: we set $U_1(\chi_t) = -\frac{1}{2}|L_{\chi_t} - L_\mu|^2$, $U_2(\chi_t) = -\frac{1}{2}|A_{\chi_t} - A_\mu|^2$, and define $A(\chi_t, \theta) = \theta_1\nabla^{\mathcal{M}} U_1(\chi_t) + \theta_2\nabla^{\mathcal{M}} U_2(\chi_t)$:

$$d\chi_t = \exp_{\chi_t}\left(\theta_1\nabla^{\mathcal{M}} U_1(\chi_t) + \theta_2\nabla^{\mathcal{M}} U_2(\chi_t) + \mathbf{\Sigma}(\chi_t)dW_t\right). \quad (47)$$

The gradients of these functions are also computable and we provide the derivations in the next section.

We can generalize to p shape descriptors m^i with average values m^i_μ by defining a potential function

$$U(\chi_t) = \sum_{i=1}^{p}\theta_i dist(m^i(\chi_t), m^i_\mu)^2, \quad (48)$$

where the distance is appropriate for the space each shape descriptor is defined in.

4.2.3 Discretized gradients

In this section we provide explicit calculations of the gradients appearing in the shape models described in the previous two sections. We first obtain the Euclidean gradients and then Riemannian/sub-Riemannian gradients can be obtained by equations (35) and (39).

Template mismatch:
The mismatch from the template is calculated according to

$$U = \int_{\Omega_{\chi_t}} F(x)dx, \tag{49}$$

where $F(x) = |1 - B_\mu(x)| - |B_\mu(x)| = 1 - 2B_\mu(x)$. The gradient can be approximated by

$$\nabla U \approx \sum_{i=1}^{m} F(x_i)N_i|x_i x_{i+1}|, \tag{50}$$

where N_i indicates the outward normal (when points are ordered clockwise) at the midpoint of the line segment $x_i x_{i+1}$, and $|x_i x_{i+1}|$ is the length of the segment.

Area:
The area of a polygonal curve with points $x_1 ..., x_m$ is

$$A = \frac{1}{2}\sum_{i=1}^{m-1} \det(x_i, x_{i+1}). \tag{51}$$

The Euclidean gradient of the discretized area with respect to the x_i'th point takes the form:

$$\nabla_{x_i} A = \frac{1}{2}\begin{bmatrix} x_{i+1}^{(2)} - x_{i-1}^{(2)} \\ x_{i-1}^{(1)} - x_{i+1}^{(1)} \end{bmatrix}, \tag{52}$$

where the superscripts indicate the coordinates of the points.

Length:
We represent the length in the following way

$$L(\chi) = \sum_{i=1}^{m-1} \|x_i - x_{i+1}\|, \tag{53}$$

where $\|\cdot\|$ is the Euclidean norm. The length gradient is

$$\nabla_{x_i} L = \frac{x_i - x_{i+1}}{\|x_i - x_{i+1}\|} - \frac{x_{i-1} - x_i}{\|x_{i-1} - x_i\|}. \tag{54}$$

5. Simulation of Shape Paths

We generate realizations of the proposed dynamical shape models by numerically integrating the corresponding SDEs. The initial shape is a circle of radius 10 represented by 63 points and the number of control points is 9. The deformation kernel width is set to $\sigma = 10$. In Figure 1 we display the simulated sequences on a 3D plot in which the third dimension corresponds to time. The diffusion is simulated up to time $T = 30$. The black stripes correspond to the locations of the control points. The template in the Ornstein-Uhlenbeck drift is a dumbbell, and the process eventually starts to oscillate around this shape. Note, since the distance to the template is based on the area of overlap, the number of points on the ellipse does not need to be equal to the number of points on the circle. The templates in the regression drift are two ellipses: as a result the contour stretches along its path both vertically and horizontally. The shape drift contains terms forcing the deformed shapes to have length and area similar to those of the original circle, thus the simulation path stays rigid.

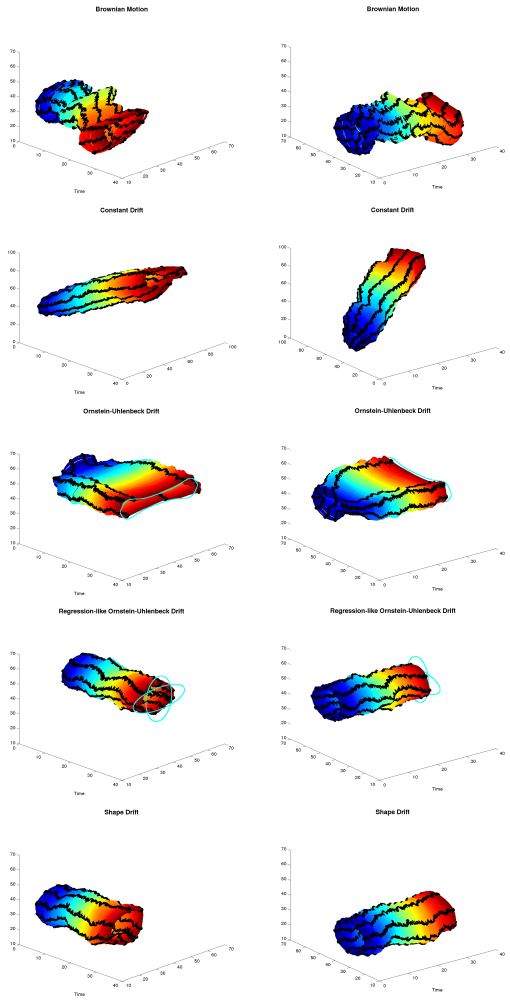


Figure 1: Simulated 2D shape diffusion paths: the third dimension indicates time (the black traces correspond to the positions of the control points)

6. Estimation of Drift Parameters in Shape Diffusions

Given observations $\{\chi_t, t \in [0,T]\}$ from the Itô process χ_t on $\mathcal{M}$ satisfying

$$d\chi_t = \exp_t(A(\chi_t,\theta)dt + B(\chi_t)dW_t), \tag{55}$$

(note we assume we have continuous measurements) we would like to find an estimate for the drift parameters stored in θ. Although in practice we can never obtain observations in continuous time, methods for estimating the parameters in this case lead to natural approaches to the problem in the case when observations occur at discrete times. We consider likelihood-based estimation, but before addressing how to solve the problem on the space of shapes, we briefly review the methodology for processes in Euclidean space.

6.1. Maximum Likelihood for Processes on $\mathbb{R}^n$

Let X_t be a diffusion process on $\mathbb{R}^n$:

$$dX_t = A(X_t,\theta)dt + B(X_t)dW_t, \quad t \in [0,T]. \tag{56}$$

Let P_θ be the measure generated by the process X_t. A likelihood function for θ is obtained by introducing the measure P corresponding to process (56) when $A(X_t,\theta) = 0$ (driftless diffusion), and considering the Radon-Nikodym derivative of P_θ with respect to P. Assume that the matrix $C(x) = B(x)B(x)^T$ is non-singular. Girsanov theorem states that under the following Novikov condition

$$\mathbb{E}_\theta \exp\left(\int_0^T A(X_t,\theta)^T C(X_t)^{-1} A(X_t,\theta)dt\right) < \infty, \tag{57}$$

P_θ is absolutely continuous with respect to P and the corresponding Radon-Nikodym derivative takes the form

$$\frac{dP_\theta}{dP}(X) = \int_0^T A(X_t,\theta)^T C(X_t)^{-1} dX_t - \\ - \frac{1}{2}\int_0^T A(X_t,\theta)^T C(X_t)^{-1} A(X_t,\theta)dt. \tag{58}$$

When C is not invertible (which is always the case when $n < m$), C^{-1} can be substituted with $B((B^TB)^\dagger)^2B^T$, where $(B^TB)^\dagger$ is the generalized pseudo-inverse of B^TB [23]. A likelihood function can be defined $l(\theta,X) = \frac{dP_\theta}{dP}(X)$, and an MLE estimate for θ can be obtained by maximizing l with respect to θ.

When measurements are observed at discrete equally sampled times: $t_0, t_1, ..., t_N$ with $h = t_{i+1} - t_i$, the approximation of the likelihood ratio is

$$l_{h,N}(\theta,X) = \sum_{i=1}^N A(X_{t_{i-1}},\theta)^T C(X_{t_{i-1}})^{-1}(X_{t_i} - X_{t_{i-1}}) - \\ -\frac{h}{2}\sum_{i=1}^N A(X_{t_{i-1}},\theta)^T C(X_{t_{i-1}})^{-1} A(X_{t_{i-1}},\theta). \tag{59}$$

which can be maximized to obtain an MLE estimate for θ.

6.2. Discrete Likelihood Ratio

To obtain intuition of what the likelihood ratio represents for diffusions on a manifold, we look at its approximation by considering a discretized version of the diffusion evaluated at finitely many time points $t_0, ..., t_N$ with distance between them $\Delta_j = t_{j+1} - t_j$. Using the exponential map form of the Itô equation, we can write the process as

$$\chi_{t_{j+1}} = \exp_{\chi_{t_j}}\left(\Delta_j A(\chi_{t_j},\theta) + \sqrt{\Delta_j}\sum_{i=1}^n B_i(\chi_{t_j})\varepsilon_i(t_j)\right), \tag{60}$$

where $j = 0, ..., N-1$, and ε_i's are independent standard normally distributed r.v.'s.

We are interested in the ratio of the joint probability density function of $\chi_{t_0}, ..., \chi_{t_N}$ denoted by $p_\theta(\chi_{0:N})$, and $p(\chi_{0:N})$, the probability density function of a diffusion process without a drift

$$\chi_{t_{j+1}} = \exp_{\chi_{t_j}}\left(\sqrt{\Delta_j}\sum_{i=1}^n B_i(\chi_{t_j})\varepsilon_i(t_j)\right). \tag{61}$$

After some algebraic manipulations we obtain

$$\frac{p_\theta(\chi_{1:N})}{p(\chi_{1:N})} \propto \exp\Bigg(\sum_{j=0}^N \log_{\chi_{t_j}}(\chi_{t_{j+1}})^T \mathbf{B}(\chi_{t_j})^{-T}\mathbf{B}(\chi_{t_j})^{-1}A(\chi_{t_j},\theta) - \\ - \frac{\Delta_j}{2}A(\chi_{t_j},\theta)^T\mathbf{B}(\chi_{t_j})^{-T}\mathbf{B}(\chi_{t_j})^{-1}A(\chi_{t_j},\theta)\Bigg).$$

6.3. Girsanov Theorem on Manifolds

Girsanov theorem has been generalized to differentiable manifolds by Elworthy in [7] (p. 263). Let X_t and Y_t be two processes on an m-dimensional Riemannian manifold $\mathcal{M}$ (with a metric $\langle,\rangle$):

$$dX_t = A(X_t,\theta)dt + \sum_{k=1}^m B_k(X_t)\circ dw_k(t), \tag{62}$$

$$dY_t = \sum_{k=1}^m B_k(Y_t)\circ dw_k(t). \tag{63}$$

Let's denote by P_X and P_Y the measures corresponding to X_t and Y_t. Girsanov theorem states that under sufficient regularity conditions P_X is absolutely continuous with respect to P_Y, and for orthonormal $B_1, ..., B_n$

$$\frac{dP_X}{dP_Y}(X) = \exp\Bigg(\int_0^T \langle A(X_t,\theta), dX_t\rangle - \\ - \frac{1}{2}\int_0^T \langle A(X_t,\theta), A(X_t,\theta)\rangle dt\Bigg). \tag{64}$$

Let $\rho \in C([0,T],M)$, i.e. it is a continuous path on M. To find an estimate for the unknown parameters we need to maximize the function $l_\rho(\theta)$:

$$l_\rho(\theta) = \mathbb{E}\left[\frac{dP_X}{dP_Y}(X)\middle| X = \rho\right] \tag{65}$$

with respect to θ, where ρ is the observed process.

6.4. Likelihood-ratio Estimators

We derive the likelihood ratio function for the drift models proposed in Section 4. Given observations $\chi_{t_0}, ..., \chi_{t_N}$, we approximate the integrals in the likelihood ratio by sums and the differential $d\chi_{t_j}$ by $\log(\chi_{t_j}, \chi_{t_{j+1}})$, where $\log(\chi_{t_j}, \cdot)$ is the Riemannian logarithm map at χ_{t_j}. The validity of such discretization is justified in Theorem 7.37 [8]). We note that in all the considered cases the drift is linear with respect to the parameter θ. This simplifies the likelihood maximization and we provide explicit MLE estimates (denoted by $\hat{\theta}$ below).

Constant drift:

$$\hat{\theta} \quad = \quad \frac{1}{T}\sum_{j=1}^{N-1} K(\chi_{t_j})^{-1} \log(\chi_{t_j}, \chi_{t_{i+1}}) \tag{66}$$

Mean-reverting drift:

$$\hat{\theta} = \frac{\sum_{j=0}^{N-1} \langle \nabla dist(\chi_{t_j}, \mu), \log(\chi_{t_i}, \chi_{t_{i+1}}) \rangle}{\sum_{j=0}^{N-1} \|\nabla dist(\chi_{t_j}, \mu)\|^2 dt} \tag{67}$$

Shape descriptor drift:

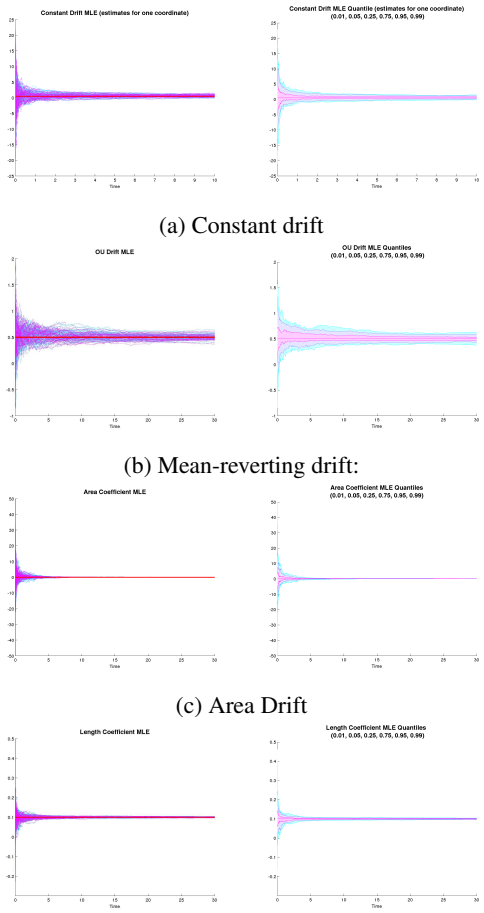
$$\begin{bmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{bmatrix} = \left(\sum_{j=0}^{N-1} M_j \Delta_j \right)^{-1} b, \tag{68}$$

where M_j as the Grammian matrix of $\nabla |L(\chi_{t_j}) - L|^2$ and $\nabla |A(\chi_{t_j}) - A|^2$, and

$$b = \begin{bmatrix} \sum_{j=0}^{N-1} \langle \nabla |L(\chi_{t_j}) - L|^2, \log(\chi_{t_j}, \chi_{t_{j+1}}) \rangle \\ \sum_{j=0}^{N-1} \langle \nabla |A(\chi_{t_j}) - A|^2, \log(\chi_{t_j}, \chi_{t_{j+1}}) \rangle \end{bmatrix} \tag{69}$$

6.5. Estimation results

We present the performance of the likelihood ratio estimator for the different types of diffusion drifts in Figure 2. We look at the empirical convergence of the estimates as time increases while keeping the time step fixed. We perform the experiment 100 times and plot how each estimate of θ changes with time. On the left the parameter estimates for each experiment are plotted against the true parameter represented by a red line; on the right we summarize the distribution by displaying the quantiles for the sample at different levels. In all cases, we observe that as time increases, the average of the MLE estimates approaches the true parameter. The code for the experiments can be found at https://github.com/valentina-s/ShapeDriftEstimation.



(a) Constant drift

(b) Mean-reverting drift:

(c) Area Drift

(d) Length Drift

Figure 2: Empirical convergence of parameter estimates: (a) estimation of one of the constant drift's parameters; (b) estimation of a mean reverting-drift parameter - initial shape is a circle, template shape is a dumbbell; (c,d) estimation of the coefficients in a shape descriptor drift; there is significant variation in the initial estimates of the coefficients, some of which are outside of the vertical range of the plots, but with time they quickly approach the true parameter.

7. Conclusion

We have proposed methodology for modeling and estimation of drifts in diffusion processes of shapes. The approach provides many opportunities for describing shape variations: from ones constrained to oscillate around a fixed shape, to ones which only restrict the geometric properties of the shape (such as length and area) without imposing specific structure. It is trivial to include higher-order shape descriptors, such as curvature and torsion, and as each closed contour can be represented by its moments, this can provide a nonparametric framework for learning shape evolution. The promising numerical convergence of the parameter estimates on simulated shapes motivates future work to establish their theoretical properties and test the performance on contours extracted from image segmentations. Going beyond estimation, the likelihood ratio can be used in testing and classification, and can turn into a powerful tool for the statistical analysis of shape time series.

References

[1] A. Arnaudon, D. D. Holm, A. Pai, and S. Sommer. A stochastic large deformation model for computational anatomy. *arXiv:1612.05323*, 2016. 1

[2] F. Arrate, T. Ratnanather, and L. Younes. Diffeomorphic active contours. *SIAM J. Imaging Sci.*, 3(2):176–198, 2010. 4

[3] F. Ball, I. Dryden, and M. Golalizadeh. Brownian motion and Ornstein-Uhlenbeck processes in planar shape space. *Methodology and Computing in Applied Probability*, 10(1):1–22, 2008. 1

[4] P. Baxendale. Measures and Markov processes on function spaces. *Mémoires de la Société Mathématique de France*, 46:131–141, 1976. 3

[5] Y. I. Belopolskaya and Y. L. Dalecky. *Stochastic Equations and Differential Geometry*. Springer, 1990. 2, 3

[6] A. Budhiraja, P. Dupuis, and V. Maroulas. Large deviations for stochastic flows of diffemophisms. *Bernoulli*, 16(1):234–257, 2010. 1

[7] K. Elworthy. *Stochastic Differential Equations on Manifolds*. Cambridge University Press, 1982. 2, 7

[8] M. Emery. *Stochastic Calculus on Manifolds*. Springer-Verlag Berlin Heidleberg, 1989. 8

[9] M. Gordina and T. Laetsch. Sub-Laplacians on sub-Riemannian manifolds. *arXiv:1412.0155*, 2014. 4

[10] D. Holm. Variational principles for stochastic fluid dynamics. *arXiv:1410.8311v3*, 2015. 1

[11] N. Ikeda and S. Watanabe. *Stochastic Differential Equations and Diffusion Processes*. North-Holland/Kodansha, 1989. 2

[12] K. Itô. Stochastic differential equations in a differentiable manifold. *Nagoya Math. J.*, 1:35–47, 1950. 2

[13] D. Kendall. The diffusion of shape. *Advances in Applied Probability*, 9(3):428–430, 1977. 1

[14] H. Kunita. *Stochastic flows and stochastic differential equations*. Cambridge University Press, 1990. 1

[15] M. Micheli, P. W. Michor, and D. Mumford. Sectional curvature in terms of the cometric, with applications to the Riemannian manifolds of landmarks. *SIAM Journal on Imaging Sciences*, 5:394–433, 2012. 1

[16] S. Sommer. Anisotropic distributions on manifolds: Template estimation and most probable paths. *International Conference on Information Processing in Medical Imaging*, 9123:193–204, 2015. 1

[17] S. Sommer and A. M. Svane. Modelling anisotropic covariance using stochastic development and sub-riemannian frame bundle geometry. *arXiv:1512.08544*, 2016. 1

[18] V. Staneva. *Stochastic Filtering on Shape Manifolds*. PhD thesis, Johns Hopkins University, 2016. 1

[19] V. Staneva and L. Younes. Modeling and estimation of shape deformation for topology-preserving object tracking. *SIAM Journal on Imaging Sciences*, 7(1):427–455, 2014. 4

[20] A. Trouvé and L. Younes. Shape spaces. In O. Scherzer, editor, *Handbook of Mathematical Methods in Imaging*, pages 1309–1362. Springer New York, 2011. 1, 3

[21] F.-X. Vialard. Extension to infinite dimensions of a stochastic second-order model associated with shape splines. *Stochastic Processes and their Applications*, 123(6):2110 – 2157, 2013. 1

[22] F.-X. Vialard and A. Trouvé. Shape splines and stochastic shape evolutions: A second-order point of view. *Quarterly of Applied Mathematics*, 70:219–251, 2010. 1

[23] N. Yoshida. Estimation of diffusion processes from discrete observation. *Journal of Multivariate Analysis*, 41:220–242, 1992. 7

[24] L. Younes. Constrained diffeomorphic shape evolution. *Foundations of Computational Mathematics*, 12(3):295–325, 2012. 4

[25] L. Younes, F. Arrate, and M. Miller. Evolution equations in computational anatomy. *Neuroimage*, 45:S40–S50, 2009. 1