Learning Shape Trends: Parameter Estimation in Diffusions on Shape Manifolds

Valentina Staneva

eScience Institute,
University of Washington

Joint work with Laurent Younes, Johns Hopkins University

July 21, 2017

Motivation

- 100 years of mathematical study of shapes and their evolution
- Few statistical methods for learning stochastic processes of shapes

Learning Stochastic Models

Goal:

- to learn stochastic models of deforming shapes from training data

Challenges:

- shapes are high-dimensional objects in a nonlinear space
- traditional time series methods fail to capture the geometry

Approach:

- define diffusion processes on the space of discretized curves
- estimate missing parameters based on likelihood-ratio techniques

Stochastic Shape Processes Through History

- "The diffusion of Shape"(Kendall'77) introduces Brownian motion on the space of landmarks [6]
- extensions to an Ornstein - Uhlenbeck process (Ball'08) [2]
- modeling biological growth by diffeomorphisms (Grenander'06) [5]
- stochastic diffeomorphic flows for shapes (Vialard'12) [10]
- noise estimation in diffusions (Sommer'15, Arnaudon'16)[1],[9]

THE STATISTICS OF SHAPE

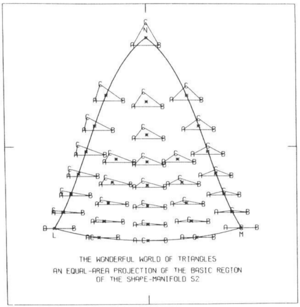


Figure 5.1 The spherical blackboard, showing 32 triangles located according to their shapes

Stochastic Shape Processes Through History

- "The diffusion of Shape"(Kendall'77) introduces Brownian motion on the space of landmarks [6]
- extensions to an Ornstein - Uhlenbeck process (Ball'08) [2]
- modeling biological growth by diffeomorphisms (Grenander'06) [5]
- stochastic diffeomorphic flows for shapes (Vialard'12) [10]
- noise estimation in diffusions (Sommer'15, Arnaudon'16)[1],[9]

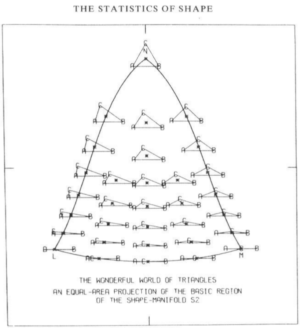
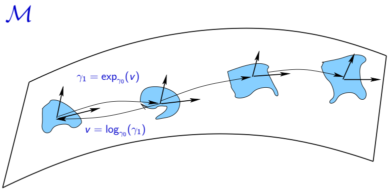


Figure 5.1 The spherical blackboard, showing 32 triangles located according to their shapes

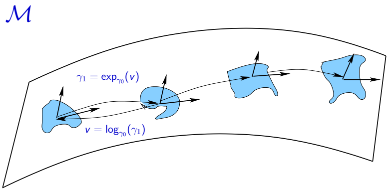
We focus on learning trends in diffusion processes on the Riemannian manifold of deformable landmarks.

The Shape Manifold



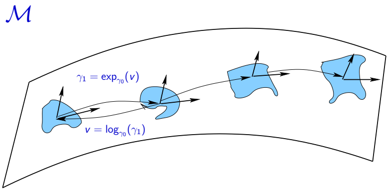
- The *shape manifold* $\mathcal{M}$ consists of simple closed plane curves γ
  - here $\gamma = \{x_1, ..., x_m \in \mathbb{R}^2, x_i \neq x_j \text{ when } i \neq j\}$

The Shape Manifold



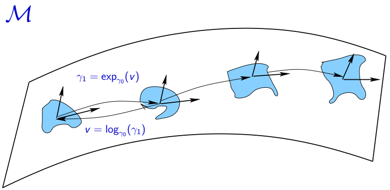
- The *shape manifold* $\mathcal{M}$ consists of simple closed plane curves γ - here $\gamma = \{x_1, ..., x_m \in \mathbb{R}^2, x_i \neq x_j \text{ when } i \neq j\}$
- The *tangent space* $\mathcal{TM}$ contains collections of 2D vectors - $\mathbf{v} = \{v_1, ..., v_m \in \mathbb{R}^2\}$

The Shape Manifold



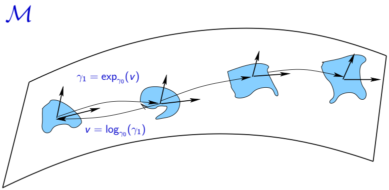
- The *shape manifold* $\mathcal{M}$ consists of simple closed plane curves γ
  - here $\gamma = \{x_1, ..., x_m \in \mathbb{R}^2, x_i \neq x_j \text{ when } i \neq j\}$
- The *tangent space* $\mathcal{TM}$ contains collections of 2D vectors - $\mathbf{v} = \{v_1, ..., v_m \in \mathbb{R}^2\}$
- *Riemannian metric:* $\|\mathbf{v}\|_\gamma = \mathbf{vK}(\gamma)^{-1}\mathbf{v}$
  - $K(x_i, x_j) = \exp(\|x_i - x_j\|^2/2\sigma^2)$

The Shape Manifold



- The *shape manifold* $\mathcal{M}$ consists of simple closed plane curves γ
  - here $\gamma = \{x_1, ..., x_m \in \mathbb{R}^2, x_i \neq x_j \text{ when } i \neq j\}$
- The *tangent space* $\mathcal{TM}$ contains collections of 2D vectors - $\mathbf{v} = \{v_1, ..., v_m \in \mathbb{R}^2\}$
- *Riemannian metric:* $\|\mathbf{v}\|_\gamma = \mathbf{v}\mathbf{K}(\gamma)^{-1}\mathbf{v}$
  - $K(x_i, x_j) = \exp(\|x_i - x_j\|^2/2\sigma^2)$
- *Exponential map:* $\gamma_1 = \exp_{\gamma_0}(\mathbf{v})$
  - can be computed by solving a system of ODE's

The Shape Manifold



- The *shape manifold* $\mathcal{M}$ consists of simple closed plane curves γ
  - here $\gamma = \{x_1, \ldots, x_m \in \mathbb{R}^2, x_i \neq x_j \text{ when } i \neq j\}$
- The *tangent space* $\mathcal{TM}$ contains collections of 2D vectors - $\mathbf{v} = \{v_1, \ldots, v_m \in \mathbb{R}^2\}$
- *Riemannian metric:* $\|\mathbf{v}\|_\gamma = \mathbf{v}\mathbf{K}(\gamma)^{-1}\mathbf{v}$
  - $K(x_i, x_j) = \exp(\|x_i - x_j\|^2/2\sigma^2)$
- *Exponential map:* $\gamma_1 = \exp_{\gamma_0}(\mathbf{v})$
  - can be computed by solving a system of ODE's
- *Logarithm map:* $\mathbf{v} = \log_{\gamma_0}(\gamma_1)$
  - we assume it exists locally

Diffusion Models

Noise Models:

Random perturbations $\rightarrow$ random walk $\rightarrow$ Brownian motion

Brownian motion is not informative!
We want to discover the 'trend' of the deformation.

Diffusion Models

Noise Models:

Random perturbations $\rightarrow$ random walk $\rightarrow$ Brownian motion

Brownian motion is not informative!
We want to discover the 'trend' of the deformation.

Diffusion Models

Noise Models:

Random perturbations $\rightarrow$ random walk $\rightarrow$ Brownian motion

Brownian motion is not informative!
We want to discover the 'trend' of the deformation.

Diffusion Models:

unknown parameters

$$d\gamma_t = \underbrace{A(\gamma_t, \theta)dt}_{drift} + \underbrace{B(\gamma_t)dW_t}_{noise}$$

Stratonovich Equations on Manifolds

Let $\{A, B_1, ..., B_n\}$ be vector fields on $\mathcal{M}$.

Stratonovich Equations on $\mathcal{M}$

The solution is a process X_t on $\mathcal{M}$, which satisfies for any smooth compactly supported f

$$f(X_T) - f(X_0) = \int_0^T Af(X_t)dt + \int_0^T \sum_{k=1}^{n} B_k f(X_t) \circ dw_k(t)$$

Stratonovich Equations on Manifolds

Let $\{A, B_1, ..., B_n\}$ be vector fields on $\mathcal{M}$.

Stratonovich Equations on $\mathcal{M}$

The solution is a process X_t on $\mathcal{M}$, which satisfies for any smooth compactly supported f

$$f(X_T) - f(X_0) = \int_0^T Af(X_t)dt + \int_0^T \sum_{k=1}^{n} B_k f(X_t) \circ dw_k(t)$$

Note: Stratonovich equations transform as tangent vectors!

Itô Equations on a Manifold

Itô equations transform according to Itô's Rule, not as tangent vectors.

Itô Equations on a Manifold

Itô equations transform according to Itô's Rule, not as tangent vectors.

$$\xcancel{f(X_T) - f(X_0) = \int_0^T Af(X_t)dt + \int_0^T \sum_{k=1}^{n} B_k f(X_t) \cdot dw_k(t)}$$

Itô Equations on a Manifold

Itô equations transform according to Itô's Rule, not as tangent vectors.

$$\xcancel{f(X_T) - f(X_0) = \int_0^T Af(X_t)dt + \int_0^T \sum_{k=1}^{n} B_k f(X_t) \cdot dw_k(t)}$$

Instead we can define them through the exponential map.

Itô Equations in Belopolskaya-Daletsky form [3]:

$$dX_t = \exp_{X_t}(A(X_t, \theta)dt + \sum_{k=1}^{n} B_k \cdot dw_k(t)),$$

Generating Diffusions

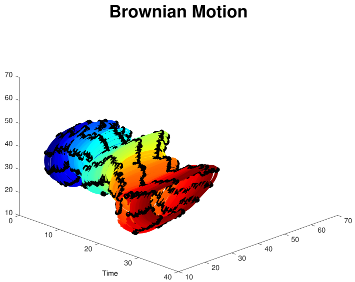
Consider a set of orthonormal vector fields: $E_1, ..., E_n$.

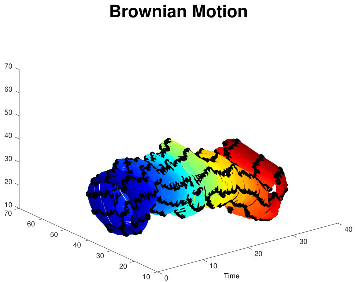
Generating Brownian motion on $\mathcal{M}$

$$\gamma_{t+dt} = \exp_{\gamma_t}\left(\sqrt{dt}\sum_{k=1}^{n} E_k(\gamma_t)\varepsilon_k\right),$$

where $\{\varepsilon_k\}_{k=1}^n \sim \mathcal{N}(\mathbf{0}, \mathbf{I_n})$.

Driftless Diffusion





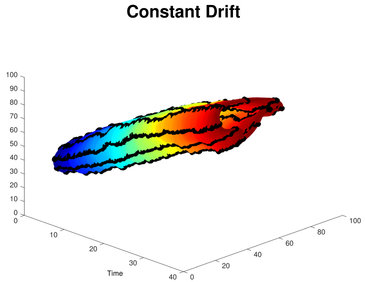
Constant drift

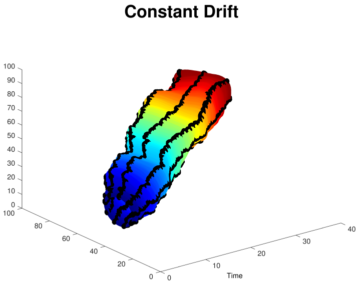
Constant with respect to the **K** basis

$$d\gamma_t = \sum_{k=1}^{n} \theta_k K(x_k, \gamma_t) dt + dW_t$$

Note: the drift will not be constant wrt to other bases.

Constant Drift Diffusion





Mean-reverting Drift

Ornstein-Uhlenbeck process on $\mathbb{R}^n$

$$dX_t = -\theta \nabla_{X_t} \left[(X_t - \mu)^2 \right] + dW_t$$

Let μ be a template shape.

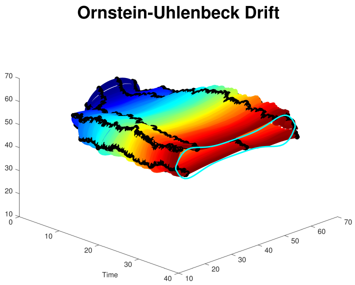
Mean-reverting process on $\mathcal{M}$

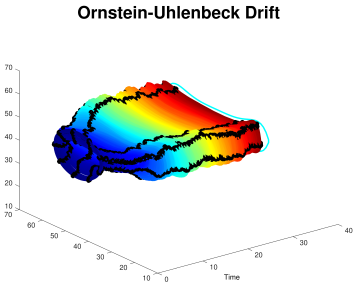
$$d\gamma_t = -\theta \nabla_{\gamma_t} dist(\gamma_t, \mu) + dW_t$$

Instead of Riemannian distance we use:

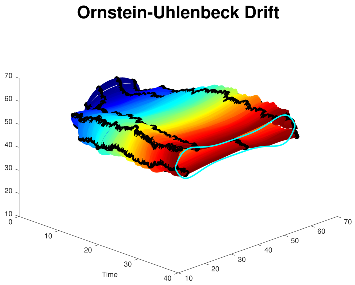
$$dist(X_t, \mu) = \text{area of mismatch of the two shapes}$$

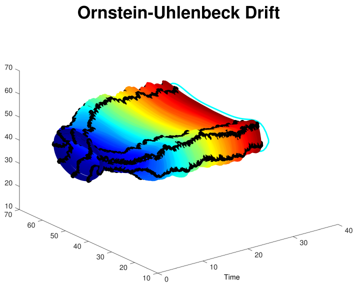
Mean-reverting drift diffusion





Mean-reverting drift diffusion





This model can be generalized for multiple templates $\mu_1, ..., \mu_p$:

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

Length-Area Drift

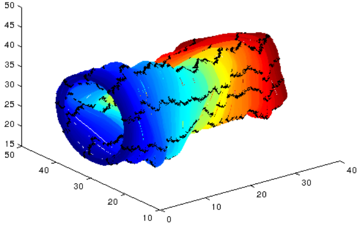
Suppose we don't have μ but only estimates for its length L_μ and area A_μ.

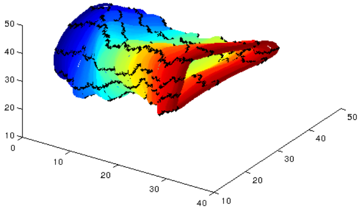
Shape process

$$dX_t = -\frac{1}{2}\theta_1 \nabla_{\gamma_t}|L_{\gamma_t} - L_\mu|^2 - \frac{1}{2}\theta_2 \nabla_{\gamma_t}|A_{\gamma_t} - A_\mu|^2 + dW_t$$

The above gradients can be explicitly computed.

Length-Area drift diffusion





Parameter Estimation in Diffusion Models

We are given a sequence of observations: $\gamma_{t_0}, ..., \gamma_{t_n}$.

We would like to maximize the joint likelihood:

$$p_\theta(\gamma_{t_1}, ..., \gamma_{t_n}) = p_\theta(\gamma_{t_n}|\gamma_{t_{n-1}})p_\theta(\gamma_{t_{n-1}}|\gamma_{t_{n-2}})...p_\theta(\gamma_{t_1}|\gamma_{t_0})p(\gamma_{t_0})$$

We do not have a closed form for the transition probability $p_\theta(\gamma_{t_i}|\gamma_{t_{i-1}})$.

Likelihood-ratio Parameter Estimation [7, 8]

Idea:

1. pick a base process which does not depend on the parameters
   - driftless process

Likelihood-ratio Parameter Estimation [7, 8]

Idea:

1. pick a base process which does not depend on the parameters
   - driftless process
2. calculate the likelihood ratio of the diffusion with respect to the base process
   - Girsanov formula

Likelihood-ratio Parameter Estimation [7, 8]

Idea:

1. pick a base process which does not depend on the parameters
   - driftless process
2. calculate the likelihood ratio of the diffusion with respect to the base process
   - Girsanov formula
3. estimate θ by maximizing the likelihood ratio
   - when the drift is linear wrt to θ the estimate has a closed form

Parameter Estimation in Diffusions on $\mathbb{R}^n$

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

$$\begin{aligned} P_X: \quad dX_t &= A(X_t, \theta)dt + dW_t \\ P_Y: \quad dY_t &= dW_t \end{aligned}$$

Girsanov Formula:

$$\frac{dP_X}{dP_Y}(X_t) = \exp\left(\int_0^T A(X_t, \theta)^T \cdot dX_t - \frac{1}{2} \int_0^T A(X_t, \theta)^T A(X_t, \theta) dt \right).$$

Parameter estimate of θ can be obtained by maximizing $\frac{dP_X}{dP_Y}(X_t)$ wrt θ.

Girsanov Formula on Manifolds (Elworthy'82 [4])

$$\frac{dP_X}{dP_Y}(X_t) = \exp\left(\int_0^T \langle A(X_t,\theta), \cdot dX_t\rangle_{X_t} - \frac{1}{2}\int_0^T \langle A(X_t,\theta), A(X_t,\theta)\rangle_{X_t} dt\right),$$

where $\cdot dX_t$ should be interpreted as $\log_{X_t}(X_{t+dt})$

Likelihood-ratio Estimators

- Constant Drift

$$\hat{\theta} = \frac{1}{Ndt} \sum_{i=1}^{N-1} K_{\gamma_i}^{-1/2} \log(\gamma_i, \gamma_{i+1})$$

- Mean-Reverting Drift

$$\hat{\theta} = \frac{\sum_{i=1}^{N-1} \langle \nabla dist(\gamma_i, \mu), \log(\gamma_i, \gamma_{i+1}) \rangle}{\sum_{i=1}^{N} \|\nabla dist(\gamma_i, \mu)\|^2 dt}$$

- Shape-Gradient Drift

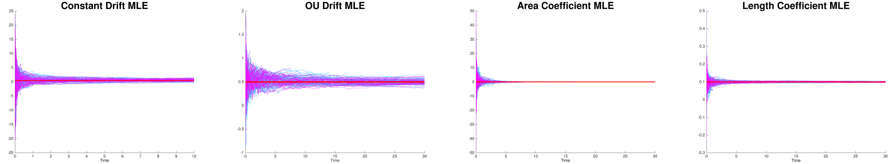
$$\begin{bmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{bmatrix} = \left(\sum_{i=1}^{N-1} M_i dt \right)^{-1} b,$$

where M_i - the Grammian matrix of $\nabla|L(\gamma_i) - L|^2$ and $\nabla|A(\gamma_i) - A|^2$, and

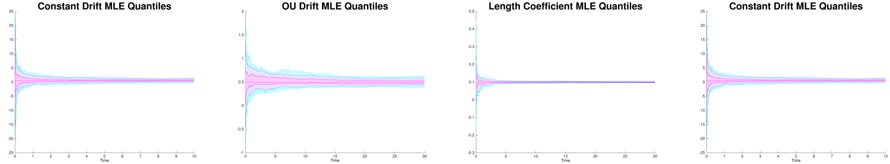
$$b = \begin{bmatrix} \sum_{i=1}^{N-1} \langle \nabla |L(\gamma_i) - L|^2, \log(\gamma_i, \gamma_{i+1}) \rangle \\ \sum_{i=1}^{N-1} \langle \nabla |A(\gamma_i) - A|^2, \log(\gamma_i, \gamma_{i+1}) \rangle \end{bmatrix}.$$

Likelihood-ratio Estimates

Independent ML Estimates



MLE Quantiles
(0.01,0.05, 0.25, 0.75,0.95,0.99)



Future Directions and Open Problems

Applications:

- parameter estimation from sparse observations
  - EM formulation to treat the missing observations
  - importance sampling approximation of the expectation step
- including higher order shape terms: e.g. curvature, torsion
- considering the distance to several template shapes - regression
- testing for different model parameters
- classification of time series

Theory:

- properties of diffusion models
  - long-term behavior
  - ergodicity
- extension to infinite dimensions: curves and surfaces
- properties of the estimators

Alexis Arnaudon, Darryl D. Holm, Akshay Pai, and Stefan Sommer.
A stochastic large deformation model for computational anatomy.
arXiv:1612.05323, 2016.

FrankG. Ball, IanL. Dryden, and Mousa Golalizadeh.
Brownian motion and Ornstein-Uhlenbeck processes in planar shape space.
Methodology and Computing in Applied Probability, 10(1):1–22, 2008.

Ya. I. Belopolskaya and Yu. L. Dalecky.
Stochastic Equations and Differential Geometry.
Springer, 1990.

K.D. Elworthy.
Stochastic Differential Equations on Manifolds.
Cambridge University Press, 1982.

Ulf Grenander, Anuj Srivastava, and Sanjay Saini.
Characterization of biological growth using iterated diffeomorphisms.
IEEE International Symposium on Biological Imaging, pages 1136–1139, 2006.

D.G. Kendall.
The diffusion of shape.
Advances in Applied Probability, 9(3):428–430, 1977.

R. S. Lipster and A. N. Shiryaev.
Statistics of Random Processes I. General Theory.
Springer-Verlag, 1977.

R. S. Lipster and A. N. Shiryaev.
Statistics of Random Processes II. Applications.
Springer-Verlag, 1977.

Stefan Sommer.
Anisotropic distributions on manifolds: Template estimation and most probable paths.
International Conference on Information Processing in Medical Imaging, 9123:193–204, 2015.