Stochastic Filtering on Shape Manifolds

by

Valentina Staneva

A dissertation submitted to The Johns Hopkins University

in conformity with the requirements for the degree of

Doctor of Philosophy

Baltimore, Maryland

December, 2016

Abstract

This thesis addresses the problem of learning the dynamics of deforming objects in image time series. In many biomedical imaging and computer vision applications it is important to satisfy certain geometric constraints which traditional time series methods are not capable of handling. We focus on building topology-preserving spatio-temporal stochastic models for shape deformation, which we combine with the observed images to obtain robust object tracking. The shape of the object is modeled as obtained through the action of a group of diffeomorphisms on the initial object boundary. We formulate a state space model for the diffeomorphic deformation of the object, and implement a particle filter on this shape space to estimate the state of the shape in each video frame. We use a practical method for sampling diffeomorphic shapes in which we generate deformations via flows of finitely generated vector fields. Based on the observations and the proposed samples we obtain an approximate estimate for the posterior distribution of the shape. We present the performance of this framework on various image sequences under different scenarios.

We extend the random perturbation models to diffusion models on the manifold

of planar (discretized) shapes whose drift component represents a trend in the shape deformation. To obtain trends intrinsic to the shape, we define the drift as a gradient of appropriate functions defined over the boundary of the shape. Given a sequence of observations from the path of the suggested stochastic differential equations, we propose a likelihood-ratio-based technique to estimate the missing parameters in the drift terms. We show how to reduce the computational burden and improve the robustness of the estimators by constraining the motion of the shapes to a lower-dimensional submanifold equipped with a sub-Riemannian metric. We further discuss how to apply this methodology to obtain estimates when we have only a limited number of observations.

Primary Reader: Laurent Younes

Secondary Reader: Gregory Chirikjian

Acknowledgements

This dissertation would not have been possible without the support of many people.

I would first like to thank my advisor Dr. Laurent Younes for his continuous support in my graduate research and many enlightening discussions which steered the direction of this work.

I would like to thank my committee members, Dr. Greg Chirikjian, Dr. Daniel Naiman, Dr. Avanti Athreya, and Dr. Michael Miller for their insightful questions and discussion during the defense. In particular, I would like to thank Dr. Chirikjian for being a second reader for the dissertation, and first introducing me to stochastic processes on manifolds and Lie groups. Dr. Chirikjian and Dr. Miller also served on my graduate board exam and, although at that time I believed their questions about infinite dimensional distributions and parameter estimation were impossible, I am now happy to have the some answers. As a chair of the department, Dr. Naiman fostered a friendly atmosphere for study and research, which greatly enhanced my graduate experience.

I would especially like to thank Aarti Jajoo for reading parts of this dissertation

and providing me with invaluable feedback that contributed to the improvement of this manuscript.

I would like to thank all professors I have interacted with at Johns Hopkins University, who taught me how to do mathematics, how to teach mathematics, and how to apply it to solve important problems. I was lucky to meet great friends among fellow students at AMS and CIS, and discussions on random topics with them contributed to both my professional and personal growth. I would also like to thank my friends from around the world who remained friends without me visiting them very often, and for supporting me throughout the years.

I acknowledge the generous financial support I have received for my studies, research, and conference travel through the Rufus P. Isaacs fellowship and funding from the Office of Naval Research. Also, I am grateful to staff at CIS and AMS who made all administrative and operational hurdles invisible to me.

Finally, but most importantly, I would like to thank my family for encouraging me to follow my interests, and supporting me in all my endeavors. Their constant reminding that I have been named after the first woman in space taught me that everything is possible with hard work. I dedicate this dissertation to them.

Contents

List of Figures

List of Algorithms

Chapter 1

Introduction

In today's abundance of imaging systems recording spatio-temporal signals in a variety of settings: medical imaging, surveillance analysis, remote sensing, etc., there is a demand to develop methods which use the collected information to infer properties of the dynamical processes behind the motion of the observed objects. Inference from time-series data is not a new problem: the rapid development of the theory of stochastic processes in the 50's paired with the push for industrial applications in the post-war world yielded methods which are relevant today: the Wiener filter[95] developed to improve radar communications during WWII laid out the foundations of stochastic filtering and Gaussian process estimation; the Kalman filter developed consequently and implemented in the navigation system of the Apollo Project is now omni-present across domains; the work of Daniel Krige [55] which moved the Wiener filtering theory to the spatial domain and applied it to mining valuation pioneered the field of geostatistics, and is now applicable to any spatial data problem; Box and Jenkins' book [18] has been providing a comprehensive guide for modeling, estimation and evaluation of discrete-time stochastic processes since its first edition (now in fifth

edition). So, what is different, what more do we need? With more data available we want to know more. The objects we are observing are high dimensional, and we want to learn more details about them, and understand their complex structure and behavior. We are particularly interested in the geometry of the objects. We don't simply want to know where they are, we want to know what their shape is at a given time, we don't simply want to know whether there is a change in shape, we want to know what this change is and what has driven this change. This can allow us to answer various scientific questions:

How do leaves grow?

How do cells react to external sources?

How do worms move?

How do skulls of chimpanzees evolve?

How does a healthy human heart beat?

How do tumors spread?

How do clouds form?

Our goal is to use statistics to infer answers from the observed data without relying on the knowledge of the explicit physical or biological models describing the processes. Ideally, when present, such knowledge should be used to further refine the estimates, however, this will require customized treatment of the individual applications.

There are several common problems which need to be addressed when working with time series data: defining a mathematical representation of the objects, choosing appropriate stochastic models to describe the dynamics of the objects, estimating the states of the hidden variables from a sequence of observations, and learning the missing model parameters. As we want to develop methodology which can study objects of different geometry and variation, we need a general representation of the shape of

the object: this could be either a parameterization of its boundary, or a subset of a domain of the image. The spaces of shapes are in general infinite-dimensional and nonlinear, which makes the task of statistical inference on them hard. Classical time series methods often assume the dynamical systems of interest are driven by random variables in finite-dimensional Euclidean space and thus can often provide closed-form solutions or optimal algorithms (e.g. Kalman filter, Baum-Welch algorithm, etc.), but are not capable of preserving the geometric properties of the shapes. Problems across domains: robotics, communications, and DNA statistical mechanics (see [21] and [22] for overview of theory and applications) have steered the development of extensions of these methods to a variety of Lie groups and special manifolds (matrix groups, unimodular groups, etc.). In the same spirit, this thesis studies how time series analysis can be extended to the manifold of shapes with the goal of solving practical image analysis problems.

Thesis Overview and Contributions.

This thesis provides a statistical framework for online and offline learning of stochastic processes on the manifold of shapes. Below we briefly summarize the contributions of the thesis by chapter. Each chapter further contains a more detailed introduction section which states its contributions in the context of the relevant background and work.

- **Chapter 2** provides background on shape representations used in imaging applications which are relevant to our work. The deformable template approach of Section 2.5 is most suitable for describing smooth topology-preserving deformations, and provides foundations for our framework.

- **Chapter 3** addresses the problem of tracking the boundary of a deforming

object in a sequence of images. Mathematically, we solve a stochastic filtering problem on the manifold of shapes. By construction, our approach ensures the estimates we obtain preserve their topogy and thus avoids ambiguities caused by boundary self-intersections. We provide two algorithms to address the problem: a particle filter (Section 3.4.1) (suitable for parallel implementation) and a Resample-Move algorithm (Section 3.4.2) (capable of addressing sample impoverishment but computationally intensive). We demonstrate the performance on simulated videos under the correct model, on simulated videos with deviations from the correct model, and on natural and biomedical videos.

- **Chapter 4** studies the convergence properties of the finite-dimensional distributions introduced in Chapter 3 and their extension to infinite dimensions. We establish the L^2 convergence of Gaussian random fields indexed by contours. We identify cases (for example, polygonal curves) in which convergence in reproducing kernel Hilbert space norm does not hold.

- **Chapter 5** addresses the problem of learning Itô diffusion processes from a sequence of observed shapes. We formulate a continuous-time version of the discrete stochastic flows in Chapter 3 and show that they coincide with the formulation of a Brownian motion on the Riemannian/sub-Riemannian manifold of landmarks. We incorporate shape-dependent drift terms, derive explicit formulas for the likelihood-ratio estimates of their parameters, and simulate their performance numerically. The learned models can in turn be used for tracking, and furthermore for a variety of other tasks such as model selection, classification, feature extraction, etc.

- **Chapter 6** summarizes the results and discusses some open problems and future directions of research.

Chapter 2

Shape Representations

In the recent years there has been substantial development in the statistical analysis of shapes. It has been driven by many applications in biomedical imaging and computer vision. In practice, we never observe shapes directly. Usually, we observe images and the boundaries of the objects in them represent the shapes. There are a lot of open problems in this domain: extracting the boundaries from the images (segmentation, tracking), aligning the shapes and finding the correct correspondences of the points on different shapes (shape registration), quantifying shape differences, classification, interpolation, regression, filtering, etc.

The abundance of approaches to solving these problems partially stems from the fact that there is no unique way to represent the shapes mathematically (i.e. there is no unique parameterization). So we begin by summarizing several representations which are relevant to this thesis and discuss the spaces of shapes associated with them. We do not intend to give an extensive overview of all shape representations. For more examples and details one can refer to some review works [78, 56, 101].

2.1 Landmark spaces

In the late 70's, David Kendall [54, 53] introduces the idea of a shape space as the space of sets of labeled points (*landmarks*) after "removing" differences between them based on "size". In the past many shapes were recorded through landmarks: biologists measure key points on leaves, doctors identify anatomical landmarks on brain surfaces, etc. Often, the shapes are collected under different conditions, and the interest is in quantifying differences in the geometry of the objects (and not the variable location and size due to inconsistencies in the collection process). Formally, one considers the space of nondegenerate configurations of n points in $\mathbb{R}^d$ modulo some transformations: for example, Kendall's shape space Σ_d^n consists of all such configurations modulo translation, rotation, and scale. When $d = 2$, this space can be identified with the complex projective space $\mathbb{C}P^{n-2}$.

The main practical challenges with the landmark representation of shapes occur when there is no exact correspondence between the landmarks, for example, when the points on the shape simply represent a discretization of the boundary or they have been selected by different annotators who did not have a common selection strategy.

2.2 Signed distance functions

Implicit representations avoid the problem of representing the boundary of the shape explicitly. A shape S is considered to be a subset of $\mathbb{R}^2$. The *signed distance function* associated to S

$$\psi(x) = \begin{cases} dist(x, \partial S) & \text{if } x \in S, \\ -dist(x, \partial S) & \text{if } x \notin S, \end{cases} \tag{2.1}$$

where ∂S is the boundary of S. Given $\psi(x)$, one can obtain a representation of the shape through its zero level set:

$$S = \{x \in \mathbb{R}^2 | \psi(x) = 0\}. \tag{2.2}$$

This representation has proven extremely useful in addressing segmentation and tracking problems, and has two important features:

1. it is implicit, i.e. it eliminates the need for parameterization of the boundary shape,

2. it allows the shape to consist of multiple components and the number of components does not need to be fixed.

Many image segmentation techniques rely on minimizing an energy functional defined on the space of signed distance functions. As these functions are defined over the whole domain of the image (not restricted to the shape boundary), a lot of efficient numerical algorithms for the solution of the problems have been proposed [77].

Unfortunately, it is difficult to provide mathematical guarantees for their performance due to the complexity of the space of signed distance functions: it is non-convex so minimizers may not always exist, also, it is nonlinear so statistical formulations on this space are not always well defined.

2.3 Probability fields

Cremers [26] proposes representing the shape through a probability field $q : \mathbb{R}^2 \rightarrow [0, 1]$ which indicates the probability that a pixel in the image belongs to the shape. As the space of probability fields and the relevant segmentation functionals over this

domain are convex working with such a shape representation can guarantee global optimization. It also allows one to incorporate statistical shape priors. The representation of the shape in this case is "distributional," i.e. we cannot deterministically define the shape. When the level sets of the probability field happen to be a set of simple closed contours, then their interior could be defined as the shape. The study of the statistical properties of these shapes requires assigning statistical models on the space of probability measures, and determining the properties of their level sets which is not trivial.

2.4 Spaces of closed curves

An explicit representation usually aims to directly parameterize the boundaries of the shape. If the shape consists of one (compact) simply connected component, its boundary is a Jordan curve. Such curves can be obtained by applying transformations (with a certain level of smoothness) to the unit circle S^1. The problem of establishing point correspondence when working with landmarks transfers here to the problem of not having a unique way to parameterize a closed contour with respect to time. So one should consider spaces of curves up to reparameterization. The following two spaces have been of interest to the imaging community:

Immersed Geometric Curves:

$$B_i = (S^1, \mathbb{R}^2) = Imm(S^1, \mathbb{R}^2)/Diff(S^1). \tag{2.3}$$

Embedded Geometric Curves:

$$B_e = (S^1, \mathbb{R}^2) = Emb(S^1, \mathbb{R}^2)/Diff(S^1), \tag{2.4}$$

which describe all C^∞ immersions and embeddings of S^1 in the plane modulo reparameterizations. The main geometric difference is that immersed curves may cross themselves, while embedded curves cannot. Mathematically, the space of immersed geometric curves is not a manifold, as the action of the diffeomorphism group is not free (i.e. there exists an element $\varphi \in Diff(S^1)$ and $c \in B_i$, such that $c \circ \varphi = c$ and $\varphi \neq Id$). A restriction to only free immersions

Freely-Immersed Geometric Curves:

$$B_f = (S^1, \mathbb{R}^2) = Imm_f(S^1, \mathbb{R}^2)/Diff(S^1) \tag{2.5}$$

results in a space which is a manifold [20]. Augmenting these spaces with the L^2-metric has provided a framework for evolution of curves by gradient flows minimizing specific energy functionals and has led to various applications in computer vision. However, as shown in [65] (Section 3.10), this metric is degenerate: the distance between any two curves in this space vanishes. This urges authors to study other metrics with more appropriate geometric interpretations. Metrics introduced in [66] and [88] have led to practical applications to shape matching, segmentation, and tracking.

2.5 Shapes through group actions

An alternative way of obtaining shapes is by starting with a "template" shape and applying a sequence of transformations to deform it into a family of other shapes. These transformations usually can be encoded by a group acting on the original shape. This is known as a "deformable template" approach and has been pioneered

by Grenander [45]. For example, in tracking problems we are interested in the motion of an object so to describe the space of all its rigid transformations we need to consider the orbit of the Euclidean group. In contrast to shape matching applications in which we assume every two shapes related by a rigid transformation are the same, i.e. each orbit is a point in the space, in the deformable template setting the orbit is the entire space. It is advantageous that often theory developed for the action on one template type can be extended to more complex templates. We will next introduce the space of deformed landmarks through the action of the group of diffeomorphisms, and the methodology can be applied to contours, surfaces, etc.

2.5.1 Landmark manifold under the action of a group of diffeomorphisms

In many applications, low dimensional matrix Lie group transformations are not sufficient to describe all possible deformations of the objects of interest. Considering more general smooth actions provides greater diversity within the shape space, while restricting to a specific level of smoothness, generates classes of shapes which could be suitable for particular applications. In the past ten years there has been an extensive development in methodology and algorithms for shape analysis through diffeomorphic mappings [67, 99, 89]. We first describe the space of "deformable landmarks", which is obtained through the action of diffeomorphisms on a set of points in $\mathbb{R}^n$ (in our setting on $\mathbb{R}^2$). The main advantage of this framework is that deformations of the whole shape can be obtained by interpolating the landmarks, thus it can easily be extended to curves and surfaces (even images, diffusion tensors, etc.). We will describe this space with more detail, as it is relevant to our work.

Let χ_0 be an ordered set of n distinct points in $\mathbb{R}^2$ $(x_1, ..., x_n)$. Let's consider the

space of configurations of these points obtained by transforming the original points through a diffeomorphic mapping $\varphi \in \mathcal{G}$:

$$Lmk_n = \mathcal{G} \cdot (x_1, ..., x_n), \tag{2.6}$$

where $\mathcal{G} \subset Diff(\mathbb{R}^2)$ is a group of diffeomorphisms from $\mathbb{R}^2$ to $\mathbb{R}^2$ with some predefined level of smoothness. This space contains all ordered collections χ of n distinct points in $\mathbb{R}^2$ and is a submanifold of $\mathbb{R}^{2n}$. A path in this space can be denoted as $\chi(t) : (x_1(t), ..., x_n(t))$ and the tangent vector at χ can be represented by a set of n two-dimensional vectors: $\mathbf{c} = (c_1, c_2, ..., c_n)$. The structure of this space can be induced from the properties of deformations belonging to the group $\mathcal{G}$. The Large Deformation Diffeomorphic Mapping framework ([52],[14]) suggests to approach the problem of building deformation mappings by composing small diffeomorphisms, and in practice this can be achieved by solving the flow equation:

$$\partial\varphi_t = v(\varphi_t, t), \quad \varphi_0(x) = \chi_0, t \in [0, 1], \tag{2.7}$$

where $v(\cdot, t) : \mathbb{R}^2 \to \mathbb{R}^2$. We can see that by evaluating the vector field v at particular set of landmarks we obtain tangent vectors to the path of deformation on the landmark manifold. The properties of the space of vector fields V determine the level of smoothness of the mappings in $\mathcal{G}$. [99]. A popular approach (Chapter 9 [99]) is to assume that V is a Reproducing Kernel Hilbert Space (RKHS) with a kernel $K(\cdot, \cdot)\mathbb{I}_2$, where $K : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}^2$ is a positive-definite function and $\mathbb{I}_2$ is the 2×2 identity matrix. As shown by Aronszajn [6], the kernel function uniquely determines the reproducing kernel space. We will briefly summarize a few properties of Reproducing kernel Hilbert spaces as they will be relevant to this thesis. We will first introduce

an RKHS for scalar-valued functions, and then explain how to construct from it an RKHS for vector fields.

Definition 2.1. (RKHS) *Let V be a Hilbert space of scalar functions $v \in V$ defined on a domain D with a norm $\|v\|$. Then V a reproducing kernel Hilbert space if for each $x \in D$ the evaluation functional is continuous, i.e. there is a constant c, such that*

$$|v(x)| \leq c\|v\|, \quad \forall v \in V. \tag{2.8}$$

Proposition 2.2. (Reproducing property) *Let V be a reproducing kernel Hilbert space. Then there exists a function $K : D \times D \to \mathbb{R}$ which satisfies for each $x \in D$*

$$v(x) = \langle v, K(\cdot, x)\rangle, \qquad \forall v \in V. \tag{2.9}$$

The definition of an RKHS can be extended to vector fields.

Definition 2.3. (RKHS of vector fields) *A space V of functions $v : \mathbb{R}^2 \to \mathbb{R}^2$ is an RKHS, if for each $x \in D$ and any vector $a \in \mathbb{R}^2$, the functional $v \to a^T v(x)$ is continuous, i.e. there exists a constant c, s.t.*

$$|a^T v(x)| \leq c\|v\|, \quad \forall v \in V. \tag{2.10}$$

The *reproducing property* for vector RKHS states that there exists a function $\Gamma : \mathbb{R}^2 \to GL_2(\mathbb{R})$, such that for each $x \in D$

$$a^T v(x) = \langle v, \Gamma(\cdot, x)a\rangle, \quad \forall v \in V. \tag{2.11}$$

Now let's consider a space V of vector fields $v : \mathbb{R}^2 \to \mathbb{R}^2$ such that each coordinate function belongs to an RKHS with a kernel $K : \mathbb{R}^2 \to \mathbb{R}$. Set $\Gamma(x, y) = K(x, y)\mathbb{I}_2$, where $\mathbb{I}_2$ is the 2×2 identity matrix. We can show that V is a vector RKHS with a kernel function Γ. As the reproducing property (2.9) holds for each coordinate, we have for the standard basis $\{e_1, e_2\}$ for $\mathbb{R}^2$

$$e_j^T v(x) = \langle v, \Gamma(\cdot, x)e_j \rangle, \quad j = 1, 2, \tag{2.12}$$

hence, (2.10) holds for any a.

Definition 2.4. (Restriction to a subset) *Let $D' \subset D$. K restricted to D' is the reproducing kernel to a RKHS $V_{D'}$ consisting of the restrictions of the functions v to D' and with a norm $\|v'\|_{D'}$ equal to the minimum norm $\|v\|$ among all functions $v \in V$ which agree with v' on D':*

$$\|v'\|_{D'} = \min_{v \in V} \{\|v\| : v_{|_{D'}} = v'\}. \tag{2.13}$$

Interpolation. Let the subdomain be equal to a set of n landmarks: $D' = \chi$. Then the above optimization problem can be interpreted as an interpolation problem: we are trying to find vector field of minimum norm defined on $\mathbb{R}^2$ which evaluated at χ has certain values. The solution of this problem has an explicit form:

$$\hat{v}(x) = \sum_{i=1}^{n} K(x, x_i)p_i, \tag{2.14}$$

where the coefficients $\{p_i\}_{i=1}^n$ are such that the above system of equations holds when

evaluated at the set of landmarks in χ:

$$\hat{v}(x_j) = \sum_{i=1}^{n} K(x_j, x_i)p_i, \quad j = 1, \ldots, n. \tag{2.15}$$

Let the vector $\boldsymbol{p}$ contain the stacked coefficients p_i, $\boldsymbol{K}(\chi)$ be a $2n \times 2n$ matrix containing blocks $K(x_i, x_j)\mathbb{I}_2$ for each i and j, and $v(\chi)$ be the vector of values of v at each landmark. A closed form solution for the coefficients is

$$\boldsymbol{p} = \boldsymbol{K}(\chi)^{-1}v(\chi), \tag{2.16}$$

Riemannian Metric. The inner product on V induces an inner product on the tangent space at each point χ of Lmk_n

$$\|\mathbf{c}\|_\chi^2 = \|v\|_\chi^2 = \mathbf{c}^T\mathbf{K}^{-1}(\chi)\mathbf{c}. \tag{2.17}$$

This metric on Lmk_n has an alternative interpretation as a projection of a Riemannian metric on the group of diffeomorphisms $\mathcal{G}$ via the infinitesimal group action (as the map from $\mathcal{G}$ to Lmk_n is a Riemannian submersion). A right-invariant metric on $\mathcal{G}$ can be obtained by defining an inner product on the space of vector fields (the tangent space at the identity) and translating it to the whole group through

$$\|v\|_\varphi = \|v \circ \varphi^{-1}\|_{id}. \tag{2.18}$$

We have the following relationship between the norms of tangent vectors on Lmk_n

and tangent vectors on $\mathcal{G}$

$$\|\mathbf{c}\|_{\chi} = \|v\|_{\varphi}, \tag{2.19}$$

where $\chi = \varphi(\chi_0)$.

Geodesics. Many computer vision algorithms rely on comparing two shapes, which requires calculating the distance between them. On a Riemannian manifold this entails to finding the path of shortest length between them

$$\min_{\substack{\chi(0)=\chi_0,\\ \chi(1)=\chi_1}} \int_0^1 \|\dot{\chi}_\tau\| d\tau, \tag{2.20}$$

and it is equivalent to minimizing the energy of the path

$$\min_{\substack{\chi(0)=\chi_0,\\ \chi(1)=\chi_1}} \int_0^1 \|\dot{\chi}_\tau\|^2 d\tau, \tag{2.21}$$

which is a more tractable problem.

The geodesic equations in Lagrangian form are

$$\ddot{x}_{k,i} + \sum_{l,l'=1}^{n} \sum_{j,j'=1}^{2} \Gamma^{(k,j)}_{(l,j),(l,j')} \dot{x}_{l,j} \dot{x}_{l',j'} = 0, \tag{2.22}$$

where $x_{k,i}$ is the i'th coordinate of the k'th landmark x_k. Let's denote the Riemannian metric by g, then we can calculate the Christoffel symbols $\Gamma^{(k,j)}_{(l,j),(l,j')}$ by

$$\Gamma^{(k,j)}_{(l,j),(l',j')} = \frac{1}{2} \left(\partial_{x_{l',j'}} g_{(k,i),(l,j)} + \partial_{x_{l,j}} g_{(k,i),(l',j')} - \partial_{x_{k,i}} g_{(l,j),(l',j')} \right). \tag{2.23}$$

Computation of these symbols requires taking derivatives of the Riemannian metric g, i.e. the inverse of the matrix $\mathbf{K}(\chi)$. This task is analytically cumbersome and numerically very sensitive to the condition number of the matrix.

Equation (2.22) can be formulated as a second-order system:

$$\dot{x}_{k,i} = v_{k,j}, \tag{2.24}$$

$$\dot{v}_{k,j} = -\sum_{l,l'=1}^{n}\sum_{j,j'=1}^{2}\Gamma^{(k,j)}_{(l,j),(l,j')}\dot{x}_{l,j}\dot{x}_{l',j'}. \tag{2.25}$$

The exponential map is the solution of the above system at time one with initial conditions $\chi(0) = \chi_0$ and $v(0) = v_0$

$$\exp_{\chi_0}(v_0) = \chi(1), \tag{2.26}$$

i.e. it is a mapping from an element of the tangent bundle of the landmark manifold to a point on the manifold.

To avoid the difficulties with computing the Christoffel symbols, a Hamiltonian version of the geodesic flow can be obtained by minimizing the kinetic energy of the system with respect to the momenta of the moving landmarks. The momenta are elements of the cotangent space of the landmark manifold and there is one-to-one correspondence between tangent velocities (tangent vectors) and momenta (cotangent vectors): the momentum p corresponding to a vector v needs to satisfy

$$(p_v|v) = p_v^T v = \|v\|_K^2 = v^T\mathbf{K}(\chi)^{-1}v, \tag{2.27}$$

i.e., $p_v = \mathbf{K}(\chi)^{-1}v$.

The kinetic energy can be written as a function of the momenta: $\frac{1}{2}\|v\|^2 = \frac{1}{2}(p|\mathbf{K}(\chi)p) = \frac{1}{2}p^T\mathbf{K}(\chi)p$. The geodesic equations minimizing this energy define the following Hamiltonian flow:

$$\partial_t x_k = \sum_{i=1}^{n} K(x_k, x_l)p_l, \tag{2.28}$$

$$\partial_t p_k = -\sum_{k=1}^{n}\sum_{i,j}^{2} \nabla_1 K^{ij}(x_k, x_l)p_{k,i}p_{l,j}. \tag{2.29}$$

In matrix form the above system takes the form:

$$\partial_t \chi = \mathbf{K}(\chi)p, \tag{2.30}$$

$$\partial_t p = -\frac{1}{2}\partial_\chi(p^T\mathbf{K}(\chi)p). \tag{2.31}$$

The exponential map in Hamiltonian form (we call it the co-exponential map) is a mapping from the cotangent space to the manifold and represents the solution of the above Hamiltonian system with initial conditions $\chi(0) = \chi_0$ and $p(0) = p_0$ at time $t = 1$:

$$\exp^\flat_{\chi_0}(p_0) = \chi(1). \tag{2.32}$$

We use the flat symbol to indicate that the map is defined on the cotangent bundle.

Other geometric quantities on the deformable landmark manifold have been studied: for example, parallel transport is discussed in [102]; a formula for the sectional curvature is derived in [63].

2.5.2 Extension to contours

The above framework generalizes to curves. The reason is that the geodesic equations on the space of landmarks are induced from the geodesic equations of the group of diffeomorphisms acting on them. Since the diffeomorphisms are defined over $\mathbb{R}^2$, one could look at their action on a closed 2D contour. Let γ_0 represent the contour template and $\mathcal{G} \subset Diff(\mathbb{R}^2)$ be a group of diffeomorphisms. We can consider the space of curves belonging to the orbit of the action of $\mathcal{G}$ on the template:

$$\mathcal{M} = \mathcal{G} \cdot \gamma_0 = \{\varphi(\gamma_0) | \varphi \in \mathcal{G}\}. \tag{2.33}$$

This space can be also equipped with a Riemannian metric, induced from the metric on the space of diffeormphisms. We will consider geodesics associated with the covectors which can be expressed as an integral with respect to a measure μ on the unit circle and a vector-valued function $p(s')$ defined on the unit circle:

$$(p|v) = \int_0^{2\pi} v(\gamma_0(s'))^T p(s') d\mu(s'). \tag{2.34}$$

The geodesic flow Φ is

$$\partial_\tau \Phi(\gamma_0(s), \tau) = \int_0^{2\pi} K(\Phi(\gamma_0(s), \tau)) p(s', \tau) d\mu(s') \tag{2.35}$$

$$\partial_\tau p(s, \tau) = -\int_0^{2\pi} \nabla_1 K(\Phi(\gamma_0(s), \tau), \Phi(\gamma_0(s), \tau)) p(s, \tau)^T p(s', \tau) d\mu(s)', \tag{2.36}$$

and its existence and uniqueness properties have been justified in [39]. When the

measure is a weighted sum of Dirac delta functions centered at set of landmarks these equations coincide with (2.28). For more details and extensions one can refer to the review articles [101, 89].

Chapter 3

Diffeomorphic Shape Tracking

3.1 Introduction

Obtaining topology-preserving methods for tracking geometric objects in a video sequence is important for visual tracking since often the objects being tracked are of known unchanging topology, but their shape cannot be fully observed in the video due to noise, clutter, or occlusions. In such situations methods for extracting the boundaries of the object (edge detection or segmentation algorithms) which rely only on the data fail to preserve the original structure of the shapes. To avoid this problem one needs to also incorporate a prior dynamical model for the evolution of the shape which enforces the desired topological constraints. We propose a general-purpose stochastic model, which does not require any precise knowledge about the specific dynamics involved in the process, but can be refined if more information is available. We fuse this model with the data by constructing a state-space system which describes the dynamics of the object, and use particle filtering [49] to retrieve the deformation of the object.

3.1.1 Related work

There is an emerging literature on the combination of high-dimensional shape deformation models with particle filtering. A method combining geometric active contours with particle filtering has been proposed in [75]. Following ideas introduced in [97], the authors develop a method in which the overall deformation of the contour is split into global affine motion and local non-affine deformation. The affine transformations are estimated through particle filtering while the nonlinear deformation is approximated through a gradient descent procedure. Several extensions of this idea have been proposed in a sequence of works [92, 90, 91], which apply particle filtering to the estimation of more general deformations. In this approach the curves are explicitly parameterized and the speed of their motion in the normal direction is modeled using low-dimensional B-splines, which allows for filtering of the unknown coefficients. As opposed to parameterizing the curves, the authors in [25] represent the objects through signed distance functions, and construct dynamical systems directly on them. Although the signed distance function representation has been widely used and successful in image segmentation problems, it poses challenging problems when estimating dynamical systems since the space of signed distance functions is not linear. In an attempt to avoid some of these difficulties, the authors suggest representing the curve by a function which indicates the probability of each pixel to belong to the object [26] and thus ensure the space of shapes is convex. This probability-field representation is explored further in various dynamical models proposed for tracking curves in [71].

A recent direction in the field has been toward formulating the filtering problem

directly on the space of shapes and taking advantage of its geometric structure. This involves facing the nontrivial task of solving statistical problems on non-Euclidean spaces. Particle filtering on matrix Lie groups has been initially proposed in [23], and the idea has been later applied to video tracking of Euclidean and affine motion [58, 59]. The advantage of working with matrix groups is the existence of explicit parameterizations of their Lie algebras and hence of the groups themselves through exponentiation. A more general framework can be designed for any finite-dimensional manifold [79], but practical algorithms can only be obtained in special cases (the authors consider the space of positive-definite matrices). In order to deal with a wide range of shapes the authors in [87] define a nonlinear filter directly on the infinite-dimensional space of immersed curves by introducing a Sobolev-type metric on this space, which allows for an explicit computation of geodesics. The estimation is done by designing a prediction-correction scheme which trades off between the predicted curve and measured curves (obtained through segmentation).

3.1.2 Contribution

So far, none of the curve representations used in solving the tracking problem have an intrinsic way to prevent the curves from intersecting themselves, which can create ambiguities when trying to define the object enclosed by the curve. In our work we formulate the filtering problem on the space of curves obtained by the action of diffeomorphisms on a fixed curve, which is the natural space to work with if we want to track general smooth nonlinear deformations of shapes and simultaneously preserve their topology. The model for the evolution of curves stems from the diffeomorphic deformation approach originally proposed in [39]. Random diffeomorphic models in

this setting have been considered more recently for shape growth in [93]. In our work we restrict the deformation to particular subsets of the whole group of diffeomorphisms (using control points, in an approach similar to [4, 7, 32, 100, 81]) and provide a simple-to-implement algorithm for particle filtering on this smaller space. This approach still provides us with enough degrees of freedom to describe a wide range of shape deformations, while reducing the dimensionality of the problem and making the estimation on this space possible. In contrast to previous work we provide a shape-dependent model for the noise on the space of curves, which is independent of the choice of the parameterization of the curve. The variations of the curve are induced by the ambient space, which allows us to construct stable methods for generating diffeomorphic deformations that can be easily extended to tracking one or more objects in higher dimensions, which becomes a much harder task when working with immersed manifolds as in [87, 11].

3.1.3 Organization

There are several important subtasks in formulating and solving the dynamical inference problem for shapes, some of which are nontrivial due to the complexity of the space these shapes belong to. We describe how we have addressed them in our setting. We begin in Section 3.2 with a description of a method for constructing distributions of random diffeomorphic shapes based on Gaussian random vector fields. The approach involves moving the shape boundary along the flow of these vector fields and comes down to solving simple differential equations. An extension to this approach is discussed in Section 3.2.2. In Section 3.3 we describe the observation model we use: a basic two-class Gaussian model on the pixel intensities of the image. Given the

dynamical and observation models we estimate the state of the shape using particle filtering, as described in Section 3.4.1. The numerical performance of the proposed methods is demonstrated in Section 3.5.

3.2 Stochastic shape evolution

The shape spaces we consider in this work are equivalence classes of subsets of $\mathbb{R}^2$ under diffeomorphic transformations. Any set $\gamma^* \subset \mathbb{R}^2$ therefore generates the "shape space" $\mathit{Diff} \cdot \gamma^*$ containing all sets $\phi(\gamma^*)$ where $\phi \in \mathit{Diff}$, the set of all diffeomorphisms of $\mathbb{R}^2$. In the following, the sets γ^* will be plane curves, or finite unions of plane curves, but most of the discussion can apply to more general sets in arbitrary dimension.

We define stochastic processes $(\gamma_t, t \geq 0)$, where t is an integer representing a discrete observation time, and the γ_t's belong to the same shape space, i.e., they can be deduced from each other by a diffeomorphic transformation: $\gamma_t \in \mathit{Diff} \cdot \gamma_0$ for all t. In most cases, we will couple the shape variables with an auxiliary process, $(\chi_t, t \geq 0)$, and we will refer to $S_t = (\gamma_t, \chi_t)$ as the state variable. Our construction will also include a control variable, denoted by α_t. We will consider evolution schemes in the form of

$$\begin{cases} S_{t+1} = F(S_t, \alpha_t) \\ \\ \alpha_t \sim P(\cdot \mid S_t) \end{cases} \tag{3.1}$$

where F is a deterministic function and P a probability distribution. The second equation should be interpreted as: the conditional distribution of α_t given $\alpha_0, \ldots, \alpha_{t-1}$ and $S_0, \ldots, S_t$ only depends on S_t and is given by $P(\cdot \mid S_t)$ (we will later on consider extensions in which the conditional distribution also depends on α_{t-1}).

3.2.1 Discrete stochastic flows

To obtain a diffeomorphic evolution, we will assume that the shape evolves, between states t and $t+1$, according to a motion driven by an ordinary differential equation, which will be associated with a smooth vector field on $\mathbb{R}^2$, $v_t : \mathbb{R}^2 \to \mathbb{R}^2$ interpreted as a shape velocity. We will let $(x, \tau) \mapsto \Phi^{v_t}(x, \tau)$ represent the flow associated to v_t, defined by

$$\partial_\tau \Phi^{v_t}(x, \tau) = v_t(\Phi^{v_t}(x, \tau)), \qquad \text{with } \Phi(x, 0) = x. \tag{3.2}$$

(Note that t is fixed in this equation.) Given v_t (which will be defined as a deterministic function, H, of the state and control variables), we set the shape transformation to be

$$\gamma_{t+1} = \Phi^{v_t}(\gamma_t, 1)$$

(i.e., the image of γ_t by the transformation $x \mapsto \Phi^{v_t}(x, 1)$). The transformation of the second component, χ_t, of the state variable can vary, but, in the simplest setting in which χ_t also is a subset of $\mathbb{R}^2$, one can use the same definition

$$\chi_{t+1} = \Phi^{v_t}(\chi_t, 1).$$

In all cases, we have the sequence of deterministic transformations $(S_t, \alpha_t) \to v_t \to S_{t+1}$ that specifies the function F in (3.1). We can also rewrite (3.1) by explicitly

including v_t, yielding

$$\begin{cases} S_{t+1} = G(S_t, v_t) \\ v_t = H(S_t, \alpha_t) \\ \alpha_t \sim P(\cdot \mid S_t) \end{cases} \tag{3.3}$$

Under certain smoothness conditions on v_t, the solution of (3.2) is such that $x \mapsto \Phi^{v_t}(x,t)$ is a diffeomorphism at all times over which it is defined. We will design the distribution $P(\cdot \mid S_t)$ so that a solution exists at all times with probability one.

3.2.1.1 Stationary random fields

In this first model, we let $S_t = \gamma_t$ (no auxiliary state component), $\alpha_t = v_t$ and assume that v_t is a centered Gaussian random field (GRF) on $\mathbb{R}^2$ with a fixed (state independent) distribution. We assume that its covariance function takes the form $C(\cdot,\cdot)\mathbb{I}_2$ where $C : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ is some positive-definite function and $\mathbb{I}_2$ is the identity matrix in $\mathbb{R}^2$ so that, for all $x, y \in \mathbb{R}^2$,

$$\mathbb{E}[v_t(x)v_t(y)^T] = C(x,y)\mathbb{I}_2. \tag{3.4}$$

(In particular, the coordinates of v_t are independent). Since v_t is a Gaussian field, its properties are completely determined by its covariance function. A natural choice is to let C be radial so that the correlation between vector field values depends only on the distance between their positions (this makes v_t stationary and rotation invariant). For most of our applications we assume that C is proportional to a Gaussian function

over $\mathbb{R}^2$:

$$C(x, y) \propto e^{-\|x-y\|_2^2/2\sigma^2}, \tag{3.5}$$

where $\sigma > 0$ is a parameter regulating how quickly the correlation decreases with the increase of the distance between the points at which the vector field is evaluated. Note that, in this case, $P(\cdot \mid S_t)$ does not depend on S_t.

The well-posedness of the model is ensured by the following result.

Proposition 3.1. *Let v be a GRF defined over $\mathbb{R}^2$, with zero mean and covariance function $C(x, y) = e^{-\|x-y\|_2^2/2\sigma^2}\mathbb{I}_2$. Then the solutions of (3.2) exist and are diffeomorphisms for any time with probability one.*

Proof. Since the covariance $C(x, y)$ is analytic, the realizations of v are analytic and hence continuously differentiable almost surely [16]. This implies that each realization of v defines a local flow of diffeomorphisms. To ensure that the solution can be extended to arbitrary time intervals, one needs to control the growth of the vector field. Let ξ denote one of the coordinates of v (therefore with covariance $e^{-\|x-y\|_2^2/2\sigma^2}$). To prove that the flow associated with ξ is complete, it is, for example, sufficient to show that [40]

$$|\xi(x)| \leq c(1 + \|x\|) \tag{3.6}$$

for some constant c and for almost all realizations of the random field. In [72], it is shown that isotropic sample-path-continuous GRF ζ with covariance $(x, y) \mapsto r(x-y)$

satisfying $|r(z)| = o(2/\log(|z|))$ as $z \to \infty$, are such that

$$\sup_{x \in D_k} \zeta(x) \to 2\sqrt{\log(k)} \qquad as \;\; k \to \infty \;\; a.s., \tag{3.7}$$

where D_k is the square of size k centered at 0. Applying this result to ξ, we find that, for almost all realizations of ξ, there exists a constant c such that $|\xi(x)| \leq c\sqrt{\log(1 + \|x\|)}$. This proves that solutions of (3.2) exist at all times. Thus the associated flow is a global flow of diffeomorphisms. □

3.2.1.2 Projected random field

Even if conceptually simple, the previous approach is computationally challenging since it requires, after discretization, sampling a random vector whose dimension equals the size of the discretization grid. Of course, since the displacements are only computed along the shape by integrating the flow over a finite time, it suffices to sample its values in a neighborhood of the shape, but the size of this neighborhood cannot be decided a priori. To avoid this, we modify the previous construction, and restrict v_t to a finite-dimensional space, which will depend on the auxiliary state variable, χ (which is made explicit below). More precisely, we assume that one attaches to each instance of χ a finite collection of vector fields on $\mathbb{R}^2$, denoted $u_1(\cdot, \chi), \ldots, u_n(\cdot, \chi)$. At a given discrete time t we let v_t take the form

$$v_t(x) = \sum_{k=1}^{n} u_k(x, \chi_t)\alpha_{t,k}.$$

We will denote $V(\chi_t) = \text{span}\{u_k(x, \chi_t), k = 1, \ldots, n\}$ so that, by construction, $v_t \in V(\chi_t)$. To simplify, we will also assume that the u_k's are always linearly independent,

so that the α_k's are uniquely defined (and will provide the control variables described in (3.1)).

To specify the rest of system (3.1), we need to define the conditional distribution $P(\cdot \mid S_t)$ of the control variables α_t. This distribution is chosen so that the resulting distribution of v_t is similar to the one of a random field with covariance C, where C is fixed (i.e., shape-independent), which can be done in the following way.

Assume that GRF's with covariance C belong with probability one to a topological vector space V, and that $V(\chi) \subset V$ for all χ. Assign to each χ a family of linear forms $\eta_1(\chi), \ldots, \eta_n(\chi) \in V^*$ (the topological dual of V) such that the matrix

$$\mathbf{K}(\chi) = ((\eta_k(\chi) \mid u_l(\cdot, \chi)), k, l = 1, \ldots, n)$$

is always invertible, which ensures a one-to-one mapping between the basis functions u_l's and the linear forms η_k's. Recall, that by definition the covariance of a centered Gaussian vector in V is a bilinear form such that

$$Cov(\eta, \tilde{\eta}) = \mathbb{E}[(\eta|v)(\tilde{\eta}|v)^T], \quad \eta, \tilde{\eta} \in V. \tag{3.8}$$

Since $V(\chi)$ is finite-dimensional, it is sufficient to consider the action of the covariance only on $\{\eta_k\}_{k=1}^n$:

$$Cov_v(\eta_k(\chi), \eta_l(\chi)) = \mathbb{E}[(\eta_k(\chi)|v)(\eta_l(\chi)|v)^T]. \tag{3.9}$$

We would like it to agree with the action of the covariance of a zero Gaussian vector

field $\xi \in V$ with covariance function C on $\{\eta_k\}$:

$$Cov_v(\eta_k(\chi), \eta_l(\chi)) = Cov_\xi(\eta_k(\chi), \eta_l(\chi)). \tag{3.10}$$

The covariance function $C(\cdot, \cdot)$ relates to the covariance form in the following way:

$$Cov(\eta, \tilde{\eta}) = (\eta | \tilde{\eta} C), \tag{3.11}$$

where $\eta C(y) = (\eta | C(y, \cdot))$. We conclude that

$$Cov_v(\eta_k, \eta_l) = \mathbb{E}[(\eta_k(\chi)|v)(\eta_l(\chi)|v)^T] = (\eta_| k(\eta_l | C(\cdot, \cdot)) = \mathbf{C}(\chi)_{kl}, \tag{3.12}$$

where $\mathbf{C}(\chi)_{kl} = Cov_\xi(\eta_k, \eta_l)$.

We can now identify what distribution of α_t would define a covariance of v_t such that as if v_t were a GRF with covariance C. Let's expand for each $k = 1, \ldots, n$:

$$(\eta_k(\chi_t) \mid v_t) = \left(\eta_k(\chi_t) \Big| \sum_{l=1}^{n} u_l(x, \chi_t) \alpha_{t,l} \right) = \sum_{l=1}^{n} (\eta_k(\chi_t) \mid u_l(x, \chi_t)) \, \alpha_{t,l} = \sum_{l=1}^{n} \mathbf{K}(\chi)_{kl} \alpha_{t,l}. \tag{3.13}$$

We would like the following equality to hold

$$\mathbb{E}[(\eta_j(\chi_t) \mid v_t)(\eta_k(\chi_t) \mid v_t)^T] = \mathbf{C}(\chi)_{j,k}. \tag{3.14}$$

Hence, we have

$$\mathbb{E}\left[\sum_{l=1}^{n}\mathbf{K}(\chi)_{jl}\alpha_{t,l}\left(\sum_{l'=1}^{n}\mathbf{K}(\chi)_{kl'}\alpha_{t,l'}\right)^{T}\right] = \mathbf{C}(\chi)_{j,k} \tag{3.15}$$

$$\mathbb{E}\left[\sum_{l=1}^{n}\sum_{l'=1}^{n}\mathbf{K}(\chi)_{jl}\alpha_{t,l}\alpha_{t,l'}^{T}\mathbf{K}(\chi)_{kl'}\right] = \mathbf{C}(\chi)_{j,k} \tag{3.16}$$

$$\mathbf{K}(\chi)\mathbb{E}[\alpha_t\alpha_t^T]\mathbf{K}^T(\chi) = \mathbf{C}(\chi), \tag{3.17}$$

and we conclude that α_t should be a Gaussian vector with covariance matrix

$$\mathbf{\Sigma}(\chi_t) = \mathbf{K}(\chi_t)^{-1}\mathbf{C}(\chi_t)\mathbf{K}(\chi_t)^{-T}. \tag{3.18}$$

In the simplest setting, we can choose V to be the space of all functions over $\mathbb{R}^2$ with the pointwise convergence topology and the linear forms to be evaluation functionals at control points attached to χ, as described in the following example.

3.2.1.3 Control points and the associated RKHS

Control points.

Let $\chi = \{x_1, ..., x_n\}$ be a finite subset of $\mathbb{R}^2$. The basis elements of $V(\chi)$ will be numbered with double indices, (k, j) for $k = 1, \ldots, n$ and $j = 1, 2$. Let $K : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ be a symmetric positive definite function (a reproducing kernel), and $u_{kj}(x, \chi) = K(x, x_k)e_j$, $k = 1, \ldots, n$, $j = 1, 2$ where (e_1, e_2) is the canonical basis of $\mathbb{R}^2$. Define the corresponding linear form

$$(\eta_{kj}(\chi) \mid v) := e_j^T v(x_k)$$

as an evaluation functional. Elements of $V(\chi)$ take the form

$$v(\cdot) = \sum_{k=1}^{n} \sum_{j=1}^{2} K(\cdot, x_k) e_j \alpha_{k,j} = \sum_{k=1}^{n} K(\cdot, x_k) \alpha_k \tag{3.19}$$

where $\alpha_k = (\alpha_{k,1}, \alpha_{k,2})^T \in \mathbb{R}^2$. It will be convenient to rewrite (3.19) in vector form

$$v(\cdot) = \mathbf{K}(\cdot, \chi)\boldsymbol{\alpha}$$

where, for $x \in \mathbb{R}^2$, $\mathbf{K}(x, \chi)$ is a 2 by $2n$ matrix formed by aligning 2 by 2 blocks $K(x, x_k)\mathbb{I}_2$, $k = 1, \ldots, n$ and $\boldsymbol{\alpha}$ is a $2n$ column vector stacking up the vectors α_k, $k = 1, \ldots, n$. More generally, if $\chi' = (x'_1, \ldots, x'_m)$, we will denote by $\mathbf{K}(\chi', \chi)$ the $2m$ by $2n$ matrix formed with 2 by 2 blocks $K(x'_k, x_l)\mathbb{I}_2$, and let $\mathbf{K}(\chi) = \mathbf{K}(\chi, \chi)$. Similarly, let $\mathbf{C}(\chi)$ consists of 2 by 2 blocks of $C(x_k, x_l)\mathbb{I}_2$. With this notation, the covariance matrix $\boldsymbol{\Sigma}$ is still given by (3.18), and the covariance function of v is

$$C_\chi(x, y) = \boldsymbol{K}(x, \chi)\boldsymbol{K}(\chi)^{-T}\boldsymbol{C}(\chi)\boldsymbol{K}(\chi)^{-1}\boldsymbol{K}(\chi, y).$$

Using this approach, we bias the model toward a limited, small-dimensional, class of diffeomorphisms. However, it turns out (from experimental evidence) that even with a small number of control points this class can provide a wide variety of deformations.

The structure of $V(\chi)$. The assumption that K is positive definite has several useful implications. As shown by Aronszajn in [6] (page 344) we can construct a unique reproducing kernel Hilbert space V_K consisting of vector fields over $\mathbb{R}^2$, whose reproducing kernel is $K(\cdot, \cdot)\mathbb{I}_2$. Clearly $V(\chi)$ is a subspace of this space, and the

realizations of the random fields v defined by (3.19) are in V_K.

The inner product between an element $v \in V_K$ and an element $w \in V(\chi)$, with $w(\cdot) = \sum_{k=1}^{n} K(\cdot, x_k)\beta_k$ is given by

$$\langle v, w \rangle_{V_K} = \langle v, \sum_{k=1}^{n} K(\cdot, x_k)\beta_k \rangle = \sum_{k=1}^{n} \beta_k^T v(x_k) = \sum_{k=1}^{n} \sum_{j=1}^{2} \beta_{kj}(\eta_{kj}(\chi) \mid v). \tag{3.20}$$

The orthogonal projection of an element v in V_K onto $V(\chi)$ needs to satisfy for every basis element $K(\cdot, x_k)e_j$

$$\langle v, K(\cdot, x_k)e_j \rangle = \langle \bar{v}, K(\cdot, x_k)e_j \rangle. \tag{3.21}$$

As the above equality is equivalent to

$$(\eta_{kj}(\chi)|v) = (\eta_{kj}(\chi)|\bar{v}), \tag{3.22}$$

the previous construction can be interpreted as letting the distribution of v_t coincide with the distribution of the orthogonal projection of a GRF with covariance C onto the finite-dimensional space $V(\chi_t)$ (only loosely, since such a GRF does not belong to V_K in general).

This remark justifies the statement of the following proposition that assesses the consistency of the distribution of random fields v with respect to their projections on low dimensional subspaces.

Proposition 3.2. ***(Consistency with projections)*** *Suppose χ_m and χ_n are two point sets such that $\chi_n \subset \chi_m$, and let $V(\chi_n)$ and $V(\chi_m)$ be the corresponding reproducing kernel Hilbert spaces restricted to those subsets. A random field $v_m \in V(\chi_m)$*

is defined to be consistent with the random field $v_n \in V(\chi_n)$ *when its orthogonal projection* $\bar{v}_m$ *onto* $V(\chi_n)$ *satisfies*

$$Cov(\bar{v}_m) = Cov(v_n). \tag{3.23}$$

Let $K(x, y)$ *be the reproducing kernel associated with the inner product on* V *and let* $C(x, y)$ *be another positive definite function. Then selecting the covariance of* v_n *to be*

$$C_n(x, y) = \boldsymbol{K}(x, \chi_n)\boldsymbol{K}(\chi_n)^{-T}\boldsymbol{C}(\chi_n)\boldsymbol{K}(\chi_n)^{-1}\boldsymbol{K}(\chi_n, y) \tag{3.24}$$

ensures that v_n *is consistent with* v_m*.*

Proof. Let $v_m(\cdot) = \sum_{i=1}^{m} K(\cdot, x_i)\bar{\alpha}_i^{(m)}$ and $v_n(\cdot) = \sum_{i=1}^{n} K(\cdot, x_i)\bar{\alpha}_i^{(n)}$. Denote the projection of v_m onto $V(\chi_n)$ by $\bar{v}_m(\cdot) = \sum_{i=1}^{n} K(\cdot, x_i)\bar{\alpha}_i$. The orthogonality condition requires that

$$\langle \bar{v}_m(\cdot), K(\cdot, x_k)e_p\rangle_V = \langle v_m(\cdot), K(\cdot, x_k)e_p\rangle_V \qquad \text{for } \; k = 1, ..., n,$$

where $\{e_p\}_{p=1}^{2}$ is the canonical basis of $\mathbb{R}^2$. After substituting the representations of v_m and $\bar{v}_m$ this condition becomes

$$\langle \sum_{i=1}^{n} K(\cdot, x_i)\alpha_i^{(m)}, K(\cdot, x_k)\rangle_V = \langle \sum_{j=1}^{n} K(\cdot, x_j)\bar{\alpha}_j, K(\cdot, x_k)\rangle_V \qquad \text{for } \; k = 1, ..., n.$$

Using the reproducing property of the inner product we obtain:

$$\sum_{i=1}^{m} K(x_k, x_i)\alpha_i^{(m)} = \sum_{j=1}^{n} K(x_k, x_j)\bar{\alpha}_j.$$

This can written in a block form

$$\mathbf{K}(\chi_n, \chi_m)\boldsymbol{\alpha}^{(m)} = \mathbf{K}(\chi_n)\bar{\boldsymbol{\alpha}} \tag{3.25}$$

Since

$$E(\bar{v}_m(x)\bar{v}_m(y)^T) = \mathbf{K}(x, \chi_n)E(\bar{\boldsymbol{\alpha}}\bar{\boldsymbol{\alpha}}^T)\mathbf{K}(y, \chi_n)^T$$

with a similar expression for v_n, it suffices to prove that $E(\bar{\boldsymbol{\alpha}}\bar{\boldsymbol{\alpha}}^T) = E(\boldsymbol{\alpha}^{(n)}(\boldsymbol{\alpha}^{(n)})^T)$. Combining (3.18) and (3.25), this is equivalent to

$$\mathbf{K}_{n,m}\mathbf{K}_{m,m}^{-1}\mathbf{C}_{m,m}\mathbf{K}_{m,m}^{-1}\mathbf{K}_{m,n} = \mathbf{C}_{n,n} \tag{3.26}$$

where we have written, for short, $\mathbf{K}_{n,m} = \mathbf{K}(\chi_n, \chi_m)$, $\mathbf{K}_{m,m} = \mathbf{K}(\chi_m)$ etc.

We can write out $\mathbf{K}_{m,m}$ as

$$\mathbf{K}_{m,m} = \begin{bmatrix} \mathbf{K}_{n,n} & \mathbf{K}_{n,m-n} \\ \mathbf{K}_{m-n,n} & \mathbf{K}_{m-n,m-n} \end{bmatrix}. \tag{3.27}$$

Letting $\mathbf{M} = (\mathbf{K}_{n-m,n-m} - \mathbf{K}_{m-n,n}\mathbf{K}_{n,n}^{-1}\mathbf{K}_{n,m-n})^{-1}$, we have

$$\mathbf{K}_{m,m}^{-1} = \begin{bmatrix} \mathbf{K}_{n,n}^{-1} + \mathbf{K}_{n,n}^{-1}\mathbf{K}_{n,m-n}M\mathbf{K}_{m-n,n}\mathbf{K}(\chi_n)^{-1} & -\mathbf{K}_{n,n}^{-1}\mathbf{K}_{n,m-n}\mathbf{M} \\ -\mathbf{M}\mathbf{K}_{m-n,n}\mathbf{K}_{n,n}^{-1} & \mathbf{M} \end{bmatrix}, \tag{3.28}$$

so that

$$\begin{aligned} & \mathbf{K}_{n,m}\mathbf{K}_{m,m}^{-1} = \\ = & \begin{bmatrix} \mathbf{K}_{n,n} & \mathbf{K}_{n,m-n} \end{bmatrix} \begin{bmatrix} \mathbf{K}_{n,n}^{-1} + \mathbf{K}(\chi_n)^{-1}\mathbf{K}_{n,n}\mathbf{M}\mathbf{K}_{m-n,n}\mathbf{K}_{n,n}^{-1} & -\mathbf{K}_{n,n}^{-1}\mathbf{K}_{n,m-n}\mathbf{M} \\ -\mathbf{M}\mathbf{K}_{m-n,n}\mathbf{K}_{n,n}^{-1} & \mathbf{M} \end{bmatrix} \\ = & \begin{bmatrix} \mathbb{I} + \mathbf{K}_{n,m-n}\mathbf{M}\mathbf{K}_{m-n,n}\mathbf{K}_{n,n}^{-1} - \mathbf{K}_{n,m-n}\mathbf{M}\mathbf{K}_{n,m-n}\mathbf{K}_{n,n}^{-1} & -\mathbf{K}_{n,m-n}\mathbf{M} + \mathbf{K}_{n,m-n}\mathbf{M} \end{bmatrix} \\ = & [\mathbb{I} \;\; \mathbf{O}] \end{aligned} \tag{3.29}$$

This yields

$$\mathbf{K}_{n,m}\mathbf{K}_{m,m}^{-1}\mathbf{C}_{m,m}\mathbf{K}_{m,m}^{-1}\mathbf{K}_{m,n} = [\mathbb{I} \;\; \mathbf{0}] \begin{bmatrix} \mathbf{C}_{n,n} & \mathbf{C}_{n,m-n} \\ \mathbf{C}_{m-n,n} & \mathbf{C}_{m-n,m-n} \end{bmatrix} [\mathbb{I} \;\; \mathbf{0}]^T = \mathbf{C}_{n,n} \tag{3.30}$$

as needed. □

3.2.1.4 Examples

We now describe several different ways of selecting the control points and show the deformations they yield in Figure 3.1. For our experiments we select $K(x, y) = e^{-\|x-y\|_2^2/2\sigma_K^2}$, where σ_K is a parameter which determines the level of fineness of the deformations: large σ_K favors almost rigid motion, while small σ_K allows for more elaborate evolution of the boundary.

Grid-based deformations. A possible approach is to approximate the stationary GRF model using a dense grid of points over the domain of the image. As already noted, this can be a computationally heavy model. This model is unbiased, in the sense that the set χ and the resulting distribution are independent of the shape, which is not necessarily a desirable feature.

Boundary-based deformations. As we remarked, only values of the vector field at points close to the boundary of the shape will affect the deformation. This suggests placing the control points along the boundary of the shape. This defines a vector field over the whole domain, such that its values are small far from the shape. In this case, the set χ is attached to the shape and moves following the same flow, resulting in a shape-dependent transition probability in the system. We have used this model in most of our applications.

Knowledge-based deformations. We can modify this model, if we know that only certain parts of the shape are driving the deformation, and select the control points in these areas. Thus we can afford having a very low-dimensional representation of the vector space, by selecting its basis in an "intelligent" way. For example, possible locations could be along the medial axis of the shape, as illustrated in the third panel of Figure 3.1.

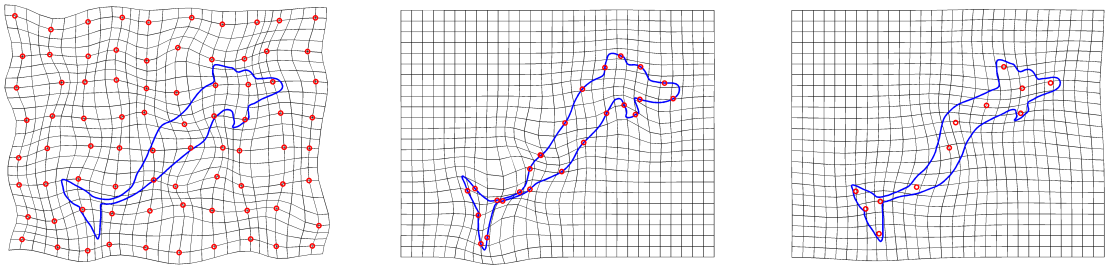


Figure 3.1: Left: a grid-based model - unbiased, but numerically impractical; middle: a boundary-based model - useful for small perturbations of the contour; right: knowledge-based model - useful for bigger deformations but with some prior knowledge on the locations of the control points.

3.2.2 Second-order dynamics via geometric formulation

In the previous models, the velocity, v_t, was either independent from the past, or only depended on it via the current state S_t (as a deterministic function of the control variables). This provides what is usually referred to as a first-order model. The goal of this section is to generalize these models by allowing v_t to also depend on its previous value, v_{t-1}, which would constitute a second-order model. This is written compactly as

$$\begin{cases} S_{t+1} = G(S_t, v_t) \\ v_t = H(S_t, \alpha_t, v_{t-1}) \\ \alpha_t \sim P(\cdot \mid S_t) \end{cases} \tag{3.31}$$

A natural model to consider is

$$v_t = v_{t-1} + \xi_t, \quad \text{with} \quad \xi_t = \sum_{k=1}^{n} \alpha_k u_k(\cdot, \chi_t) \in V(\chi_t), \tag{3.32}$$

where $\boldsymbol{\alpha}$ follows a Gaussian distribution as defined before. However, recall that v_{t-1}

is a sum of rapidly decreasing radial-basis functions (RBF's) centered on χ_{t-1}, and since it is kept fixed during the integration of (3.2), the associated shape evolution slows down as soon as the curve moves away from the original position of the control points. This is not a major issue for the first-order models, since the underlying covariance, C, can be scaled up or down to allow for large changes, but this is more problematic for the second-order models, since the vector field v_{t-1} evaluated at χ_t will always be reduced in magnitude compared to its original value at χ_{t-1}.

To address this problem we replace (3.2) by another evolution equation that generates a diffeomorphism as a flow of a time-dependent vector field initialized at v, but follows the motion of the control points.

In Section 2 we introduced the equations for the geodesic flow on the manifold of n landmarks. Given the initial velocity v, represented as $v = \mathbf{K}(\cdot, \chi)\boldsymbol{\alpha}$, we let the control points evolve according to this flow (in Hamiltonian form)

$$\begin{cases} \dfrac{d\chi^v}{d\tau} = \mathbf{K}(\chi^v, \chi^v)\boldsymbol{\beta}^v \\ \\ \dfrac{d\boldsymbol{\beta}^v}{d\tau} = -D_1(\mathbf{K}(\chi^v, \chi^v)\boldsymbol{\beta}^v)^T\boldsymbol{\beta}^v \end{cases} \tag{3.33}$$

where D_1 represents differentiation with respect to the first χ^v variable in $\mathbf{K}(\chi^v, \chi^v)$, β^v is an auxiliary variable $\chi^v(0) = \chi$ and $\boldsymbol{\beta}^v(0) = \boldsymbol{\alpha}$. We can define a time-dependent vector field $w : \mathbb{R}^2 \times [0, 1] \to \mathbb{R}^2$

$$w^v(\cdot, \tau) = \mathbf{K}(\cdot, \chi^v(\tau))\boldsymbol{\beta}^v(\tau), \quad \tau \in [0, 1]$$

which therefore satisfies $w^v(\cdot, 0) = v$, and a time-dependent diffeomorphism $\Psi^v(\cdot, \tau)$

solution of

$$\partial_\tau \Psi^v(x,\tau) = w^v(\Psi^v(x,\tau),\tau), \quad x \in \mathbb{R}^2. \tag{3.34}$$

The deformation driven by Ψ interpolates the geodesic flow restricted on the space of the control points to the whole plane and hence to any contour in $\mathbb{R}^2$. It turns out that the paths generated by this flow correspond to geodesics on the sub-Riemannian manifold of (discrete) curves.

We can formulate a new first-order model given by $\gamma_{t+1} = \Psi^{v_t}(\gamma_t, 1)$, $\chi_{t+1} = \chi^v(1) = \Psi^{v_t}(\chi_t, 1)$ with $v_t = \mathbf{K}(\cdot, \chi_t)\boldsymbol{\alpha}_t$, $\alpha_t \sim \mathcal{N}(0, \boldsymbol{\Sigma}(\chi_t))$. This model only differs from the previous one in that Ψ^v replaces Φ^v. In practice, they are very similar for small deformations. In this construction the evolution $\tau \mapsto \Psi^v(\cdot,\tau)$ can be interpreted as a geodesic in a group of diffeomorphisms equipped with a right-invariant metric for which the tangent space at the identity is the RKHS associated to K. This implies, in particular, that $\|w^v(\cdot,\tau)\|_K$ remains constant over time, which suggests introducing the second-order model in which one sets $v_t = w^{v_{t-1}}(\cdot, 1) + \xi_t$, or, equivalently,

$$\alpha_t \sim \mathcal{N}(\boldsymbol{\beta}^{v_{t-1}}(1), \boldsymbol{\Sigma}(\chi_t)).$$

3.2.3 Sub-Riemannian point of view

In this section we put the model described in Section 3.2.2 in the context of sub-Riemannian geometry. Equation (3.33) defines a geodesic evolution equation for a Riemannian metric in the small-dimensional space of control points (in which control points form a homogeneous space under the action of diffeomorphisms). One may argue, however, that the Riemannian manifold of interest should be a similar

homogeneous space under the action of diffeomorphisms, but for planar curves. In this setting, control points constrain the set of allowed directions of motion on this manifold, which is exactly the situation studied in sub-Riemannian geometry, when one is interested in finding shortest paths on manifolds subject to constraints.

We first introduce a few notions from sub-Riemannian geometry. A *sub-Riemannian manifold* is a manifold M equipped with:

- a *distribution* Δ, which is a family of spaces $(\Delta_\gamma, \gamma \in M)$ indexed by the manifold, such that each Δ_γ is a k-dimensional subspace of the tangent space $T_\gamma M$ that can locally be represented as the span of k smooth vector fields evaluated at γ (this representation is actually global in the situation we consider);
- an inner product $\langle \cdot, \cdot \rangle$ on Δ_γ.

A smooth curve $c : [0, 1] \to M$ is called *a horizontal curve* on M if its tangent vector belongs to Δ_γ at each γ. The inner product on Δ_γ determines the length of a horizontal path:

$$l(c) = \int_0^1 \langle c'(\tau), c'(\tau) \rangle^{\frac{1}{2}} d\tau, \tag{3.35}$$

wich is used to define a distance on M called the *Carnot-Caratheodory distance*, which is equal to the length of the shortest horizontal path connecting two points on M. The path obtaining this minimum is called a *geodesic* on M:

$$d(\gamma_0, \gamma_1) = \inf\{l(c) | c(0) = \gamma_0, c(1) = \gamma_1, c'(\tau \in \Delta_{c(\tau)})\}; \tag{3.36}$$

i.e. it equals the infimum of the length over all horizontal paths connecting two points on M. A path which attains the minimum is called a *geodesic* on M. As in

the case of classical Riemannnian geometry, one can equivalently obtain geodesics on a sub-Riemannian manifold, by minimizing the energy of the path:

$$E(c) = \int_0^1 \langle c'(\tau), c'(\tau) \rangle d\tau. \tag{3.37}$$

To simplify the presentation in our setting we will treat the shape space of curves as a finite- (but high-) dimensional space, in which curves are discretized over finite sets of m points (so that $M \subset \mathbb{R}^{2m}$). Given an RKHS V_K, this space can be equipped with a Riemannian manifold structure, when the metric associated to a curve $\gamma = (x_1, \ldots, x_m)$ is defined by

$$\|\xi\|_\gamma^2 = \min\{\|v\|_{V_K}^2 : v(x_j) = \xi_j, j = 1, \ldots, m\} = \xi^T \mathbf{K}(\gamma)^{-1} \xi$$

for $\xi \in T_\gamma M \sim \mathbb{R}^{2m}$, where the second identity comes from standard reproducing kernel arguments [6].

The introduction of control points to constrain evolution directly provides M with a sub-Riemannian structure. More precisely, associate to each curve γ a set of n control points, $\chi = \chi_\gamma$, typically with n much smaller than m. We will assume that the control points form a subset of the discrete representation of the curve ($\chi_\gamma \subset \gamma$) (if not we can augment γ with these points). We want to restrict the evolution

$$\partial_\tau x_k = v(x_k, \tau), x_k \in \gamma$$

to vector fields taking the form

$$v(\cdot, \tau) = \mathbf{K}(\cdot, \chi_\gamma) \boldsymbol{\beta}(\tau)$$

with $\boldsymbol{\beta}(\tau) \in \mathbb{R}^{2n}$. In terms of the manifold structure, this is equivalent to requiring that

$$\partial_\tau \gamma = \xi(\tau)$$

with $\xi(\tau) = \mathbf{K}(\gamma, \chi_\gamma)\boldsymbol{\beta}(\tau)$, i.e., to constraining the evolution to horizontal curves associated to the distribution

$$\Delta_\gamma = \{\mathbf{K}(\gamma, \chi_\gamma)\boldsymbol{\beta}, \boldsymbol{\beta} \in \mathbb{R}^{2n}\}.$$

Solutions of this minimization problem are described by Pontryagin's maximum principle [35, 2, 17] and are called *sub-Riemannian geodesics*. The following result states that the geodesics defined in the previous section are consistent with the sub-Riemannian interpretation.

Proposition 3.3. *The paths generated in* (3.33) *are normal sub-Riemannian geodesics along the distribution* Δ_γ.

Proof. There are two types of paths which can arise as minimizers: *normal* geodesics, which are solutions of a system of ODE's similar to the classical geodesic equations, and *abnormal* geodesics, which are a result of certain singularities of the system. For our purpose it is sufficient to describe only normal geodesics on a sub-Riemannian manifold. Let $\mathbf{p}$ be a costate variable which is of the same dimension as the state, γ. We define the following Hamiltonian function

$$H(\gamma, \mathbf{p}, \boldsymbol{\beta}) \quad = \quad \mathbf{p}^T \mathbf{K}(\gamma, \chi_\gamma)\boldsymbol{\beta} - \frac{1}{2}\boldsymbol{\beta}^T \mathbf{K}(\chi_\gamma, \chi_\gamma)\boldsymbol{\beta}. \tag{3.38}$$

The Pontryagin maximum principle states that the optimal control for a normal

geodesic satisfies

$$\boldsymbol{\beta}^* = \arg\max_{\boldsymbol{\beta}} H(\gamma, \mathbf{p}, \boldsymbol{\beta}) \tag{3.39}$$

$$= \boldsymbol{K}(\chi_\gamma, \chi_\gamma)^{-1} \boldsymbol{K}(\chi_\gamma, \gamma)\mathbf{p}, \tag{3.40}$$

and that the solution of the constrained optimization problem is

$$\begin{cases} \partial_\tau \gamma^* = \nabla_p H(\gamma^*, \mathbf{p}^*, \boldsymbol{\beta}^*) = \boldsymbol{K}(\gamma^*, \chi_{\gamma^*})\boldsymbol{\beta}^* \\ \\ \partial_\tau \mathbf{p}^* = -\nabla_\gamma H(\gamma^*, \mathbf{p}^*, \boldsymbol{\beta}^*). \end{cases} \tag{3.41}$$

Since we assume that χ_γ is a subset of γ, the evolution equation for χ in (3.33) is derived directly from restricting the first equation of (3.41). Writing $\gamma = (x_1, \ldots, x_m)$ and assuming, without loss of generality, that the points are ordered so that $\chi_\gamma = (x_1, \ldots, x_n)$, the explicit form of the second equation in (3.41) is

$$\begin{aligned} \partial_\tau p_k &= -\sum_{j=1}^{m} \nabla_1 K(x_k(\tau), x_j(\tau))\beta_k(\tau)^T p_j(\tau) - \sum_{i=1}^{n} \nabla_2 K(x_i(\tau), x_k(\tau))\beta_i(\tau)^T p_k(\tau) + \\ &\quad + \sum_{j=1}^{n} \nabla_1 K(x_k(\tau), x_j(\tau))\beta_k(\tau)^T \beta_j(\tau), \quad \text{for} \quad k = 1, ..., n \end{aligned} \tag{3.42}$$

$$\partial_\tau p_k = -\sum_{i=1}^{n} \nabla_2 K(x_i(\tau), x_k(\tau))\beta_i(\tau)^T p_k(\tau), \quad \text{for} \quad k = n+1, ..., m. \tag{3.43}$$

From this, we can observe that solutions initialized with $p_k = 0$ for $k > n$ satisfy

$p_k = 0$ at all times. For these solutions, we have

$$\boldsymbol{K}(\chi_\gamma, \gamma)\mathbf{p} = \boldsymbol{K}(\chi_\gamma, \chi_\gamma)\mathbf{p}_n$$

where $\mathbf{p}_n$ refers to the $\mathbf{p}$ restricted to its n first coordinates. From equation (3.39), we obtain $\boldsymbol{\beta}^* = \mathbf{p}_n$. Given this, we have, for $k \leq n$,

$$\begin{aligned}
\partial_\tau \beta_k^* &= -\sum_{j=1}^{n} \nabla_1 K(x_k(\tau), x_j(\tau))\beta_k^*(\tau)^T \beta_j^*(\tau) - \sum_{i=1}^{n} \nabla_2 K(x_i(\tau), x_k(\tau))\beta_i^*(\tau)^T \beta_k^*(\tau) + \\
&\quad + \sum_{j=1}^{n} \nabla_1 K(x_k(\tau), x_j(\tau))\beta_k(\tau)^T \beta_j(\tau) \\
&= -\sum_{j=1}^{n} \nabla_1 K(x_k(\tau), x_j(\tau))\beta_k^*(\tau)^T \beta_j^*(\tau)
\end{aligned}$$

which is the second equation in (3.33). □

3.2.4 Stochastic models for affine motion

Our framework allows us to model any smooth invertible deformation through flows of diffeomorphisms; however, it is clear that in practice it is better to model any "big motion" separately (as suggested in [97]), since there are simpler ways to describe such deformations, as they usually involve fewer parameters and are easier to generate. We propose two different methods to achieve this.

3.2.4.1 Decomposition of the diffeomorphism

The simplest approach is to assume that affine and diffeomorphic actions operate in turn, yielding a transition taking the form

$$\gamma_t = \varphi_t \cdot A_t \gamma_{t-1}.$$

with $A_t \in GAff^+(2)$, the subgroup of the affine group of $\mathbb{R}^2$ containing transformations with positive determinant. The matrix A can be easily parametrized through the matrix exponential map: if $\{E_i\}_{i=1}^6$ is a basis for the Lie algebra of $GAff(2)$, then A is given by $A = \exp\left(\sum_{i=1}^6 c_i E_i\right)$, where $\mathbf{c} = \begin{bmatrix} c_1 & \dots & c_6 \end{bmatrix}^T$, and we will write for short $A = \exp(\mathbf{c})$. A random affine transformation can be obtained by assuming that these parameters are independent normally distributed random variables with distribution $\mathbf{c} \sim \mathcal{N}(0, \mathbb{I}_6)$. (Assuming spherical covariance is no loss of generality since this can always be achieved by a proper selection of the basis.) The diffeomorphism φ_t can be generated by any of the previous models, although it seems reasonable to restrict it to be first order while allowing the affine part to be first or second order.

For example, a possible first-order model involving control points would take the form

$$\begin{cases} \gamma_{t+1} = \varphi_t \cdot A_t \gamma_t \\ \chi_{t+1} = \varphi_t \cdot A_t \chi_t \\ \varphi_t = \Phi^{v_t}(\cdot, 1), \;\; A_t = \exp(\mathbf{c}_t) \\ \mathbf{c}_t \sim \mathcal{N}(0, \mathbb{I}_6), \;\; v_t = \mathbf{K}(\cdot, A_t \chi_t)\boldsymbol{\alpha}_t \\ \boldsymbol{\alpha}_t \sim \mathcal{N}(0, \boldsymbol{\Sigma}(\chi_t)) \end{cases}$$

For a second-order model in the affine component, one can simply replace the fourth equation by

$$A_t = \exp(\mathbf{c}_t) A_{t-1}$$

3.2.4.2 Decomposition of the vector field

A second option is to apply an affine component at the vector field level, extending the definition of v in order to allow for the juxtaposition of an affine component and a sum of RBF's. In this setting, we consider vector fields taking the form

$$v_{\mathfrak{aff}}(x) = v(x) + Mx + b,$$

where $M \in \mathbb{R}^{2\times 2}$ is a two-by-two matrix and b is a two-dimensional vector. This decomposition is unique if v is assumed to vanish at infinity, which applies in particular to the case in which v belongs to an RKHS, V_K, generated by a kernel K such as the ones we consider here. An interesting feature is that vector fields such as $v_{\mathfrak{aff}}$ also form an RKHS. In particular, if $E_1, \ldots, E_6$ form a basis for $\mathfrak{aff}(2) = \mathbb{R}^{2\times 2} \times \mathbb{R}^2$, each one being interpreted as an affine transformation $x \mapsto E_i x$, one can define the affine kernel

$$K_{\mathfrak{aff}}(y, x) = K(y, x)\mathbb{I}_2 + \sum_{i=1}^{6} (E_i y)(E_i x)^T \tag{3.44}$$

combining the Hilbert structure on V_K and the Euclidean structure on $\mathfrak{aff}(2)$ for which $E_1, \ldots, E_6$ form an orthonormal basis (this comes as a direct consequence of the reproducing property). For example, if one chooses the norm on $\mathfrak{aff}(2)$ to take the form

$$\|(M, b)\|^2 = \lambda \mathrm{tr}(M^T M) + \mu \|b\|^2,$$

one finds

$$K_{aff}(y,x) = (K(y,x) + \lambda x^T y + \mu)\mathbb{I}_2$$

as derived in [99]. More interestingly, one can separate, in the Lie algebra of the affine group, the rotation, scaling, shearing and translation parts, introducing the basis

$$\sqrt{2}E_1 = \mathbb{I}_2, \sqrt{2}E_2 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \sqrt{2}E_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \sqrt{2}E_4 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

of $\mathbb{R}^{2\times 2}$ and letting the norm on $\mathfrak{aff}(2)$ be given by

$$\|(M,b)\|^2 = \lambda_1 a_1^2 + \lambda_2 a_2^2 + \lambda_3 a_3^2 + \lambda_4 a_4^2 + \mu\|b\|^2$$

if $M = a_1 E_1 + \cdots + a_4 E_4$. In this case, (3.44) gives

$$\begin{aligned} K_{aff}(y,x) &= K(y,x)\mathbb{I}_2 + \frac{\lambda_1}{2} yx^T + \frac{\lambda_2}{2}(x^T y\,\mathbb{I}_2 - xy^T) \\ &+ \frac{\lambda_3}{2}\begin{bmatrix} x_1y_1 & -x_2y_1 \\ -x_1y_2 & x_2y_2 \end{bmatrix} + \frac{\lambda_4}{2}\begin{bmatrix} x_2y_2 & x_1y_2 \\ x_2y_1 & x_1y_1 \end{bmatrix} + \mu\mathbb{I}_2 \end{aligned}$$

Note that K_{aff} is not a scalar kernel in general unless all λ's are equal.

This kernel can be directly plugged into the models that were introduced in Sections 3.2.1 or 3.2.2. Although the interpretation of this model is more subtle, in practice we can generate deformations having interesting combined affine and non-linear properties. Note also that the construction in this section closely relates to multi-scale diffeomorphisms, as introduced in [76].

3.2.5 Generalizing the Riemannian metric formulation

In the suggested models we have used the inner product of the RKHS associated to K (the kernel that was used to discretize the vector fields using control points) to define orthogonal projections or the Riemannian structure of geodesics. This was convenient because it led to simpler mathematical expressions and computation. Conceptually, one may argue that the two are different objects, the first one defining a numerical discretization scheme of the vector fields and their associated stochastic models, while the second one, via the RKHS structure, representing the energy, or cost, that can be associated to variations of diffeomorphisms. In view of this, one can build more general schemes, within the discussion of Section 3.2.1.2.

Fixing the reproducing kernel K, we can define the linear forms, η_k, and the basis functions, u_k, consistently by ensuring that

$$u_k(x, \chi) = \Big(\eta_k(\chi) \mid K(\cdot, x)\Big).$$

Assume, to simplify matters, that $(\eta_k(\chi), k = 1, \ldots, n)$ is a collection of vector measures. In this case, the entries of the matrix $\mathbf{K}(\chi)$ are provided by integrals

$$(\mathbf{K}(\chi))_{k,l} = \int \int (K(x, y) d\eta_k(\chi)(x)) \cdot d\eta_l(\chi)(y) \tag{3.45}$$

(recall that K is matrix valued). The control-point example is a special case in which

$$\int v(x) \cdot d\eta_{k,j}(\chi)(dx) = e_j^T v(x_k).$$

The interpretation of the construction as an orthogonal projection remains appropriate if the coefficients of the matrix $\mathbf{C}(\chi))$ are defined as in Section 3.2.1.2, namely

$$(\mathbf{C}(\chi))_{k,l} = \int\int (C(x,y)d\eta_k(\chi)(x)) \cdot d\eta_l(\chi)(y). \tag{3.46}$$

For such a construction to be feasible, however, the integrals in (3.45) and (3.46) should be easily computable. This can be achieved, for example, if K and C are Gaussian RBF's (as in our experiments) and if $\eta_1, \ldots, \eta_n$ are Gaussian measures (possibly singular, like Dirac measures). Examples of some explicit computations in similar settings can be found in [7, 100].

3.3 Observation models for shapes in images

In the previous section we have described several stochastic models for the evolution of shapes. Our goal is to infer about a particular realization of these models based on an observed image sequence. For that we need to specify an observation model which gives the relationship between the state of the shape and the image. We describe two models: one based directly on the image intensities, and another one based on specific features extracted from the images. Alternative observation models can be easily constructed and incorporated depending on the type of data available.

3.3.1 Region-based observation likelihood

Here we consider the following standard region-based approach. Let γ be the contour which represents the shape in the image I. We associate with γ an ideal (continuous) image $\mathcal{I}$ which is constant over the interior region, R_{in} of γ, and over its exterior region R_{out}, with constant values given by μ_{in} and μ_{out} respectively. We assume that the

observed image, I, is obtained by, first, discretizing $\mathcal{I}$, then, adding independent noise to each discrete value, the noise variances being denoted σ^2_{in} and σ^2_{out} in the inside and the outside regions. The conditional observation density as the joint density is then given by

$$p(I|\gamma) = \text{const} \prod_{c_j \in R_{in}} e^{-(I_j - \mu_{in})^2/2\sigma^2_{in}} \times \prod_{c_j \in R_{out}} e^{-(I_j - \mu_{out})^2/2\sigma^2_{out}}, \tag{3.47}$$

where c_j is the central point of the jth pixel and I_j the corresponding image intensity. We note that it is explicitly computable given a curve and an image. This model can be directly extended to multiple contours delimiting multiple regions (like in the case of double contours in Figure 3.3, or of the cardiac sequences in Figure 3.9a).

3.3.2 Feature-based observation likelihood

The approach described above is particularly useful when pixels in the regions inside and outside of the contour have relatively uniform grayscale values, for example, when we have a light object on a darker background. However, in many practical applications, the background, and sometimes the object itself, are very non-homogeneous. For instance, in the presence of clutter, we need a different observation model. For such situations it is better to assume that the final observations of the system are features in the image obtained through an edge detection algorithm. The number of features is random and they could be close to the boundary of the object - "real" features, or further in the background - a result of noise or clutter.

The form of the features suggests that it is natural to model them as a spatial point process X defined on a bounded domain $\Omega \subset \mathbb{R}^2$. X takes values in the space of finite configurations of points in Ω, i.e. $X : \Omega \to Y$, where $Y = \cup_{n=1}^{\infty} \mathbb{R}^n$. We

will denote any $y \in Y$ as $\{n; x_1, ..., x_n\}$ to emphasize that the number of points may vary. We can further assume that X is a Poisson point process with an intensity function $\lambda : \Omega \to [0, \infty)$, such that $\int_\Omega \lambda(x)dx < \infty$. Define a measure Λ on Ω by $\Lambda(B) = \int_B \lambda(x)dx$ for any $B \subset \Omega$, and a random counting measure X on Ω by $X(B) = (\#\text{ points of } X \text{ in } B)$. The definition of a Poisson point process states that $X(B)$ has Poisson distribution with a parameter $\Lambda(B)$ (therefore $\mathbb{E}[X(B)] = \Lambda(B)$), and that $X(B_1)$ and $X(B_2)$ are independent random variables for disjoint sets B_1 and B_2. An interesting property of X (which can sometimes be used as a definition of a Poisson point process [69]) is that conditional on $X(B)$, the points of X in B are i.i.d. random variables with density $\lambda(x)/\Lambda(B)$. This allows us to describe the density of X:

$$f_n(x_1, ..., x_n) = \frac{e^{-\Lambda(\Omega)}}{n!}\lambda(x_1)...\lambda(x_n), \tag{3.48}$$

which is crucial to define an observation likelihood.

To model the features in the image, we consider the following generalization of a Poisson point process. Let X_1 and X_2 be two independent Poisson point processes on Ω with corresponding intensity functions λ_1 and λ_2. Define a spatial point process X as $X(B) = X_1(B) \sqcup X_2(B)$ for every B in Ω: X is the superposition of X_1 and X_2, assuming that points do not coincide. The number of points of X in B is then equal to the total number of points of X_1 and X_2 in B, i.e. it follows Poisson distribution with a parameter $\Lambda_1(\Omega)+\Lambda_2(\Omega)$, where Λ_1 and Λ_2 are the corresponding measures. In our context, X_1 corresponds to the process generated by the boundary of the object, so we model the dependence of λ_1 on the curve γ in the following way:

$$\lambda_1(x) = Ce^{-\text{dist}(x,\gamma)^2/2\sigma^2}, \tag{3.49}$$

where σ is a positive parameter and C is a constant which ensures that $\int_\Omega \lambda_1(x)dx$ is equal to the total number of points generated by X_1. For a distance function we use the minimal distance from the point x to a point on γ. The second process X_2 describes all the remaining features, which are either features generated by the texture of the object (belonging to the region inside the contour R_{in}), or features generated by the clutter in the background (belonging to the region outside the contour R_{out}). Let λ_{in} and λ_{out} be the corresponding densities of the features in the two regions. We define λ_2 to be piecewise constant on Ω: $\lambda_2(x) = \lambda_{in}\mathbf{1}_{R_{in}}(x) + \lambda_{out}\mathbf{1}_{R_{out}}(x)$. The intensity function of all features then is

$$\lambda(x) = Ce^{-\mathrm{dist}(x,\gamma)^2/2\sigma^2} + \lambda_{in}\mathbf{1}_{R_{in}}(x) + \lambda_{out}\mathbf{1}_{R_{out}}(x). \tag{3.50}$$

Now the observation likelihood given γ can be written as:

$$\begin{aligned} p(\{n; x_1, ...x_n\}|\gamma) &= \frac{e^{-\Lambda(\Omega)}}{n!} \prod_{x_i \in R_{in}} \left(Ce^{-\mathrm{dist}(x_i,\gamma)^2/2\sigma^2} + \lambda_{in} \right) \times \\ &\times \prod_{x_i \in R_{out}} \left(Ce^{-\mathrm{dist}(x_i,\gamma)^2/2\sigma^2} + \lambda_{out} \right). \end{aligned} \tag{3.51}$$

The above model describes well the most likely contour in the image when the boundary features are dense relative to the ones generated by clutter. However, when there are a lot of "false" edges close to the boundary, more information is required in order to distinguish the real edges from them. What is important to know in such cases is the orientation of the edges. A simple estimate for the orientation of an edge can be obtained from the gradient of the image. Most edge detection algorithms involve the computation of the gradient so we can have the edge orientations for free. We can easily incorporate these measurements in our statistical model using

a marked point process. A marked point process $X_m : \Omega \to Y$ is a standard point process X with marks attached to each point: $y = \{n; x_1, ..., x_n; m_1, ..., m_n\}$. We take the marks to represent the angle each edge makes with the x-axis, so the space of marks is $\mathbb{S}^1$. The distribution of the marks needs to be defined on the unit circle and it has to depend on the discrepancy between the orientation of the edge and the orientation of the boundary. We assume a mark m follows the Von Mises distribution with parameters μ and κ:

$$p(m) = \frac{e^{\kappa \cos(m-\mu)}}{2\pi I_0(\kappa)}, \tag{3.52}$$

where I_0 is the modified Bessel function of the first kind. This distribution has the property to be uniform on the circle for $\kappa = 0$ and peaked around μ for κ big. Let μ be the orientation of the point on the boundary closest to x. Since the orientation of the boundary should match that one of the edges close to the boundary, we select κ to be proportional to the inverse of the distance between x and γ. Thus edges far from the boundary will have relatively uniform orientation, which is consistent with the fact that the orientation of clutter features does not depend on γ. Assuming that the marks are independent given their position, the density of the marked point process takes the form

$$f_n(x_1, ..., x_n, m_1, ..., m_n) = \frac{e^{-\Lambda(\Omega)}}{n!} \prod_{i=1}^{n} \lambda(x_i) \prod_{i=1}^{n} p(m_i|x_i). \tag{3.53}$$

The final form of the likelihood is

$$p(\{n; x_1, ...x_n\}|\gamma) \propto e^{-\Lambda(\Omega)} \prod_{x_i \in R_{in}} \left(Ce^{-\text{dist}(x_i,\gamma)^2/2\sigma^2} + \lambda_{in} \right) \times \\ \times \prod_{x_i \in R_{out}} \left(Ce^{-\text{dist}(x_i,\gamma)^2/2\sigma^2} + \lambda_{out} \right) \times \prod_{i=1}^{n} \frac{e^{\frac{\cos(m_i - \mu)}{\text{dist}(x_i,\gamma)}}}{I_0\left(\frac{1}{\text{dist}(x_i,\gamma)}\right)}. \tag{3.54}$$

Other features can be used to build observation models. For example, an appearance model based on color histograms has been extremely successful for tracking in color videos [24]; the probabilistic formulation for optical flow provides an observation model which incorporates spatio-temporal information.

3.4 Particle filtering in shape space

3.4.1 Particle filtering

Particle filtering was first introduced in the computer vision field through the Condensation algorithm [49] which opened a path for developing practical tracking algorithms. This method, which relies on Monte Carlo simulations, is more versatile than model-based methods, like the Kalman filter, and is well adapted to handle nonlinear shape deformation models like the ones we consider here. Based on the stochastic dynamical model from Section 3.2 and the observation model from Section 3.3, we can construct a state-space system for the evolution of a shape boundary (Figure 3.2),

that we summarize as:

$$\begin{cases} S_t \sim P_S(\cdot \mid S_{t-1}), \\ I_t \sim P_I(\cdot \mid \gamma_t), \end{cases}$$

where P_S is the transition probability for our state variable $S = (\gamma, \chi)$ and P_I is the observation law. In tracking we are interested in sequentially estimating the posterior probability $\nu(B) = Pr(S_1, ..., S_t \in B \mid I_1, ..., I_t)$ where B is a subset of configurations for $\{S_1, ..., S_t\}$. There is generally no closed-form expression for ν in a nonlinear and non-Gaussian situation such as ours, and one needs to rely on approximation methods. Monte Carlo methods, and in particular particle filtering, allow for the construction of an estimate of η by an empirical measure, based on importance sampling from tractable distributions.

The idea is to represent the posterior by a weighted set of particles. The algorithm consists of two main steps which are iterated over time: generating a sample of states $\{S_t^{(i)}\}_{i=1}^N$ (the particles) according to $S_t \sim P_S(\cdot \mid S_{t-1}^{(i)})$, and attaching a weight $w_t^{(i)}$ to each particle which relies on the existence of the observation likelihood $p_I(I_t \mid S_t^{(i)})$. The weighted sample $\{S_t^{(i)}, w_t^{(i)}\}_{i=1}^N$ can then be used to approximate the posterior at time t. The algorithm often requires an additional step of resampling the particles according to their weights, which results in obtaining an unweighted set of particles $\{\hat{S}_t^i\}$. The reader can refer to a broad overview of the theory and applications of the particle filter and its extensions in [29].

We give the steps of this Importance Sampling-Resampling algorithm for estimation with a first-order dynamical model on the nonlinear deformations in Table 3.1. The structure of the algorithm allows for an easy parallelization which significantly speeds up the performance. The last resampling step is necessary to prevent the

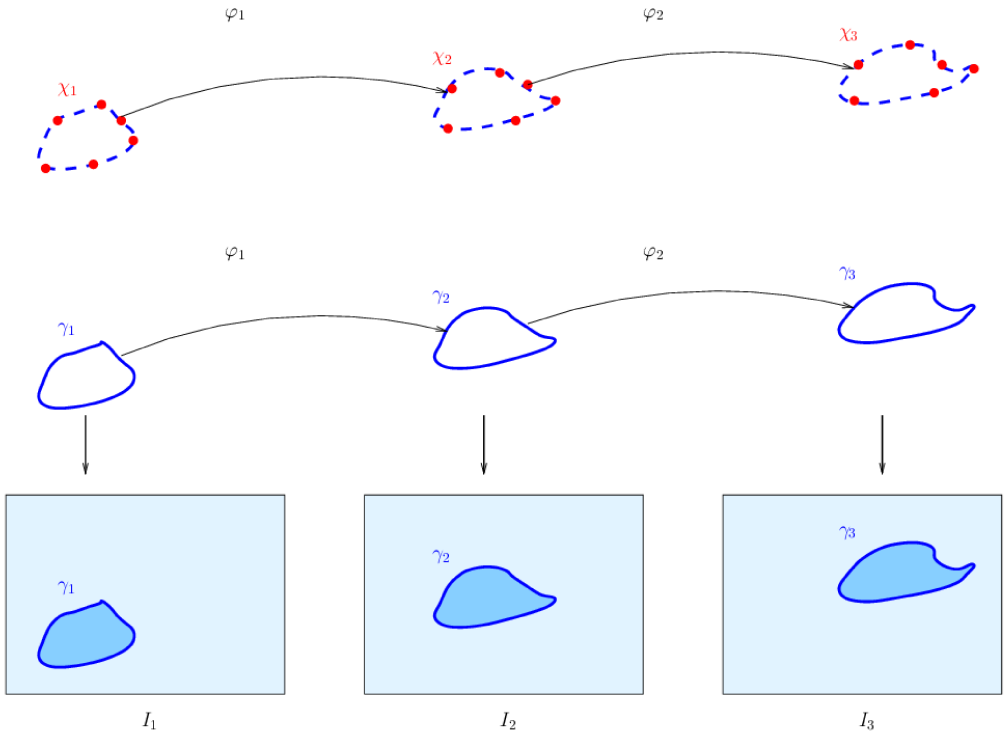


Figure 3.2: This figure describes the components of our dynamical system. In the top row we see how the control points are evolving based on the diffeomorphisms dependent on their position at a fixed state. This process induces the deformation of the contours in the middle row. In the last row we see the sequence of observations which are the images capturing the motion of the object defined by the contours.

weights from becoming degenerate, i.e. almost all of them becoming zero. After re-sampling particles with small weights are eliminated and ones with large weights can appear more than once in the new sample. We used a systematic resampling scheme

Table 3.1: Importance Sampling-Resampling Algorithm

Importance Sampling-Resampling Algorithm

0. Initialize $\gamma_0^{(i)} = \gamma_0$, $\chi_0^{(i)} = \chi_0$ for $i = 1, ..., N$.
1. For each t, construct $\mathbf{\Sigma}^{(i)} = \mathbf{K}(\chi_{t-1}^{(i)})^{-1}\mathbf{C}(\chi_{t-1}^{(i)})\mathbf{K}(\chi_{t-1}^{(i)})^{-1}$.
2. Sample $\boldsymbol{\alpha}_{t-1}^{(i)} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}^{(i)})$.
3. Evolve the boundary points through
$$\partial_\tau \Phi(\gamma_{t-1}^{(i)}, \tau) = K(\Phi(\gamma_{t-1}^{(i)}, \tau), \chi_{t-1}^{(i)})\boldsymbol{\alpha}_t,$$
and set $\gamma_t^{(i)} = \Phi(\gamma_{t-1}^{(i)}, T)$.
4. Evolve the control points through
$$\partial_\tau \Phi(\chi_{t-1}^{(i)}, \tau) = K(\Phi(\chi_{t-1}^{(i)}, \tau), \chi_{t-1}^{(i)})\boldsymbol{\alpha}_t,$$
and set $\chi_t^{(i)} = \Phi(\chi_{t-1}^{(i)}, T)$.
5. Compute the weights by $w_t^{(i)} = p(I_t|\gamma_t^{(i)}) / \sum_{i=1}^N p(I_t|\gamma_t^{(i)})$
6. Resample $\{\gamma_t^{(i)}\}_{i=1}^N$ according to $\{w_t^{(i)}\}_{i=1}^N$.

[8] due to its simplicity and relatively good performance. First we generate N random numbers:

$$u_j = u + \frac{j-1}{N} \qquad \text{for } j = 1, ..., N, \tag{3.55}$$

where $u \sim \mathcal{U}(0, 1/N)$. Then, we construct the cumulative distribution F associated with the weights $\{w_i\}_{i=1}^N$, and set $x_j = x_{F^{-1}(u_j)}$. The steps of the algorithm are as

follows:

Table 3.2: Systematic Resampling

Systematic Resampling

1. set $c_1 = w_1$
2. for $i = 2 : N$, $c_i = c_{i-1} + w_i$ ($\{c_i\}_{i=1}^N$ stores the cdf of the particles)
3. sample $u \sim \mathcal{U}(0, 1/N)$
4. for $j = 1 : N$
   $u_j = u + (j-1)/N$
   while $u_j > c_j$, $i = i + 1$
   $x_j^* = x_i$, $w_j^* = 1/N$
   endfor
5. replace $\{x_i, w_i\}$ with $\{x_j^*, w_j^*\}$

The performance of this method is comparable to other popular resampling schemes like multinomial, stratified, and residual resampling [47], however, its asymptotic analysis is harder since it produces dependent particle positions [28]. At the cost of increasing the computational complexity, this algorithm can be easily modified to stratified or residual resampling which enjoy more theoretical results.

Practice shows that the resampling step is inevitable. The restriction of having finite and, often due to computational constraints, limited-sized samples, requires high care in deciding what the criterion for resampling should be and what resampling

scheme should be used.

3.4.2 Resample-Move algorithm

The main advantage of the particle filter for solving the tracking problem is that it can sequentially update the estimate of the posterior distribution as new observations become available. Also the method is very easy to implement and is naturally parallelizable. Unfortunately, when used for large systems, it usually requires a substantial number of particles in order for the estimate to converge to the true one [80], so care should be applied when working with high-dimensional states or long-time sequences. In this section we describe how the Resample-Move [38] algorithm can be applied in our context to improve the performance of the particle filter.

MCMC methods provide alternatives to importance sampling which have better convergence properties, since they employ dependent samples which move in the direction of the true posterior. This, however, makes them iterative in nature and hard to parallelize. Also, to apply such methods to the problem of tracking an object over time, one needs to reestimate the whole posterior at each step, which is extremely time- and memory-consuming and hence impractical. To unite the power of both approaches, Gilks et. al. [38] propose the Resample-Move algorithm which addresses the problem of sample impoverishment occurring after the resampling step. The method has a broader applicability: it can also be used to battle the inaccuracy of the proposed samples caused by the limitations of finite sampling or errors in the

prior model, which makes it very suitable for our setting.

Let the set of particles $\{S_{1:t+1}^{(i)}\}_{i=1}^{N}$ define the empirical estimate for the posterior based on the sequential importance sampling-resampling procedure. We can improve this estimate using an MCMC approach. It is important to note that although $S_{1:t+1}$ is a sequence of infinite-dimensional objects, their distribution is completely determined by the initial curve γ_0, the initial location of the control points χ_0 (both being assumed to be deterministic), and the sequence of controls $\boldsymbol{\alpha}_{0:t}$, and the latter admit probability density functions. Assuming that the estimate $\{\boldsymbol{\alpha}_{0:t-1}^{(i)}\}_{i=1}^{N}$ is reliable, one only modifies $\boldsymbol{\alpha}_t^{(i)}$ into ${\boldsymbol{\alpha}_t}^{*(i)}$ for each $i = 1, ..., N$, so that the new estimate $\{\boldsymbol{\alpha}_{0:t-1}^{(i)}, {\boldsymbol{\alpha}}_t^{*(i)}\}_{i=1}^{N}$ becomes closer to the target posterior distribution. This is done using MCMC sampling, based on the Metropolis-Hastings algorithm, for the posterior distribution of interest, initialized with $\boldsymbol{\alpha}_t^{(i)}$. The procedure is repeated for each $i = 1, ..., N$ and provides the new sample $\{\boldsymbol{\alpha}_t^{*(i)}\}_{i=1}^{N}$.

The Metropolis-Hastings algorithm iterates the following sequence of operations. Let $\boldsymbol{\alpha}_t'$ be the current state. The algorithm first generates a new state $\boldsymbol{\alpha}_t''$ with a proposal density $q(\boldsymbol{\alpha}_t''|\boldsymbol{\alpha}_t')$, which is then accepted with probability $\min\{r(\boldsymbol{\alpha}_t''|\boldsymbol{\alpha}_t'), 1\}$, where

$$r(\boldsymbol{\alpha}_t''|\boldsymbol{\alpha}_t') = \frac{p(\boldsymbol{\alpha}_{1:t-1}, \boldsymbol{\alpha}_t''|I_{1:t+1})q(\boldsymbol{\alpha}_t'|\boldsymbol{\alpha}_t'')}{p(\boldsymbol{\alpha}_{1:t-1}, \boldsymbol{\alpha}_t'|I_{1:t+1})q(\boldsymbol{\alpha}_t''|\boldsymbol{\alpha}_t')},$$

$p(\boldsymbol{\alpha}_{1:t-1}, \cdot | I_{1:t+1})$ being the posterior density treated as a function of $\boldsymbol{\alpha}_t$ (which determines the state γ_{t+1}).By Bayes rule we have that

$$p(\boldsymbol{\alpha}_{0:t-1}, \boldsymbol{\alpha} | I_{1:t+1}) = \frac{p(I_{t+1}|\boldsymbol{\alpha}_{0:t-1}, \boldsymbol{\alpha}) p(\boldsymbol{\alpha}|\boldsymbol{\alpha}_{0:t-1}) p(\boldsymbol{\alpha}_{0:t-1}|I_{1:t})}{p(I_{t+1}|I_{1:t})}, \tag{3.56}$$

which allows us to rewrite the expression for r using more familiar quantities:

$$r(\boldsymbol{\alpha}_t''|\boldsymbol{\alpha}_t') = \frac{p(I_{t+1}|\boldsymbol{\alpha}_{0:t-1}, \boldsymbol{\alpha}_t'') p(\boldsymbol{\alpha}_t''|\boldsymbol{\alpha}_{0:t-1}) q(\boldsymbol{\alpha}_t'|\boldsymbol{\alpha}_t'')}{p(I_{t+1}|\boldsymbol{\alpha}_{0:t-1}, \boldsymbol{\alpha}_t') p(\boldsymbol{\alpha}_t'|\boldsymbol{\alpha}_{0:t-1}) q(\boldsymbol{\alpha}_t''|\boldsymbol{\alpha}_t')}. \tag{3.57}$$

We observe that $p(I_{t+1}|\boldsymbol{\alpha}_{0:t-1}, \boldsymbol{\alpha}_t)$ is equivalent to the observation likelihood $p(I_{t+1}|\gamma_{t+1})$, while $p(\boldsymbol{\alpha}_t|\boldsymbol{\alpha}_{0:t-1})$ is the prior density, and both distributions are computable. For the proposal density, q, we consider two options: a random walk proposal and a Langevin proposal.

Random Walk Proposal. Update the state by

$$\boldsymbol{\alpha}_t'' = \boldsymbol{\alpha}_t' + \delta\boldsymbol{\varepsilon}, \tag{3.58}$$

para where $\boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_\varepsilon)$. To ensure that we preserve the smoothness properties of the deformations which we generate through $\boldsymbol{\alpha}_t''$, we select $\boldsymbol{\Sigma}_\varepsilon = \boldsymbol{\Sigma}(\chi_t)$. This choice also sets the scaling of the proposal covariance (up to a constant) to that of the prior model, and allows one to tune the acceptance rate through δ. Since this proposal

density is symmetric, r reduces to

$$r(\boldsymbol{\alpha}_t''|\boldsymbol{\alpha}_t') = \frac{p(I_{t+1}|\boldsymbol{\alpha}_{0:t-1}, \boldsymbol{\alpha}_t'')p(\boldsymbol{\alpha}_t''|\boldsymbol{\alpha}_{0:t-1})}{p(I_{t+1}|\boldsymbol{\alpha}_{0:t-1}, \boldsymbol{\alpha}_t')p(\boldsymbol{\alpha}_t'|\boldsymbol{\alpha}_{0:t-1})}. \tag{3.59}$$

Langevin Proposal. Unless the importance sampling estimate is already good, the random walk proposal can require a lot of iterations to converge. A Langevin proposal can be more efficient, since it moves the particles toward regions of higher posterior probability. For that, the update step involves the gradient of the log-posterior density with respect to $\boldsymbol{\alpha}_t'$:

$$\boldsymbol{\alpha}_t'' = \boldsymbol{\alpha}_t' + \frac{\delta^2}{2}\nabla_{\boldsymbol{\alpha}_t'} \log p(\boldsymbol{\alpha}_{0:t-1}, \boldsymbol{\alpha}_t'|I_{1:t+1}) + \delta\boldsymbol{\varepsilon}. \tag{3.60}$$

A representation for the logarithm of the posterior can be obtained from (3.56)

$$\log p(\boldsymbol{\alpha}_{0:t-1}, \boldsymbol{\alpha}_t'|I_{1:t+1}) = \log p(I_{t+1}|\boldsymbol{\alpha}_{0:t-1}, \boldsymbol{\alpha}_t') + \log p(\boldsymbol{\alpha}_t'|\boldsymbol{\alpha}_{0:t-1}) + \text{const}, \tag{3.61}$$

where the final term is a constant with respect to $\boldsymbol{\alpha}_t'$. Define $L(\boldsymbol{\alpha}_t') = \log p(I_{t+1}|\boldsymbol{\alpha}_{0:t-1}, \boldsymbol{\alpha}_t')$. In our models, this likelihood is a function of the curve γ_{t+1}' which is uniquely determined as a function of S_t and the new control $\boldsymbol{\alpha}_t'$. Since S_t is fixed in this part of the algorithm, we have that $L(\boldsymbol{\alpha}_t') = \tilde{L}(\gamma_{t+1}'(\boldsymbol{\alpha}_t'))$, and the gradient of interest is

$$\nabla_{\boldsymbol{\alpha}_t'} L(\boldsymbol{\alpha}_t') = \left(\frac{\partial \gamma_{t+1}'}{\partial \boldsymbol{\alpha}_t'}\right)^T \nabla_{\gamma_{t+1}'} \tilde{L}(\gamma_{t+1}'). \tag{3.62}$$

We recognize that $\tilde{L}(\gamma'_{t+1})$ is the observation likelihood of γ'_{t+1}, and we have an explicit formula for it based on the observation model, from which it is easy to compute a gradient. The curve γ'_{t+1} is obtained from $\boldsymbol{\alpha}'_t$ via the solution of a differential equation, and below we show how the gradient with respect to $\boldsymbol{\alpha}'_t$ can also be obtained by solving a differential equation. Note that the proposal density in this case is not symmetric, and all terms in (3.57) need to be included. Note also that, if we set the level of the noise ε to zero, we obtain a gradient descent method which minimizes the negative log-posterior likelihood.

Log-likelihood gradient. Here we derive the gradient of the log-likelihood for the Langevin proposal. We recall the form of the observation likelihood

$$\begin{aligned} L(\gamma'_{t+1}) &= \log(p(I_{t+1}|\gamma'_{t+1})) = \\ &= -\sum_{x_j \in R_{in}} \left(\frac{|I_{t+1}(x_j) - \mu_{in}|^2}{2\sigma_{in}^2} - \frac{\log(2\pi\sigma_{in}^2)}{2} \right) + \\ &\quad -\sum_{x_j \in R_{out}} \left(\frac{|I_{t+1}(j) - \mu_{out}|^2 - \log(2\pi\sigma_{out}^2)}{2\sigma_{out}^2} \right) = \\ &= -\sum_{x_j \in R_{in}} \left(\frac{|I_{t+1}(x_j) - \mu_{in}|^2}{2\sigma_{in}^2} - \frac{\log(2\pi\sigma_{in}^2)}{2} \right) + \\ &\quad +\sum_{x_j \in R_{in}} \left(\frac{|I_{t+1}(x_j) - \mu_{out}|^2}{2\sigma_{out}^2} - \frac{\log(2\pi\sigma_{out}^2)}{2} \right) + \text{const.} \end{aligned} \tag{3.63}$$

We approximate the above sum by an integral over R_{in}:

$$L(\gamma'_{t+1}) = \int_{R_{in}} \left[-\frac{|I_{t+1}(x) - \mu_{in}|^2}{2\sigma_{in}^2} + \frac{|I_{t+1}(x) - \mu_{out}|^2}{2\sigma_{out}^2} + \log\left(\frac{\sigma_{in}}{\sigma_{out}}\right) \right] dx + \text{const.} \tag{3.64}$$

Defining the function f as

$$f(x) = \frac{|I_{t+1}(x) - \mu_{out}|^2}{2\sigma_{out}^2} - \frac{|I_{t+1}(x) - \mu_{in}|^2}{2\sigma_{in}^2} - \log\left(\frac{\sigma_{in}}{\sigma_{out}}\right), \tag{3.65}$$

we can write the log-likelihood as

$$L(\gamma'_{t+1}) = \int_{R_{in}} f(x)dx, \tag{3.66}$$

and its gradient as

$$\nabla_{\gamma'_{t+1}} L(\gamma'_{t+1}) = \int_{\gamma'_{t+1}} f\nu ds, \tag{3.67}$$

with ν being the normal to the curve γ'_{t+1}.

To compute $\frac{\partial \gamma'_{t+1}}{\partial \boldsymbol{\alpha}'_t}$, we observe that γ'_{t+1} is the end-point solution of the ODE

$$\frac{\partial \gamma_\tau(\boldsymbol{\alpha}'_t)}{\partial \tau} = \sum_{i=1}^{n} K(\gamma_\tau(\boldsymbol{\alpha}'_t), x_i)\alpha'_i, \tag{3.68}$$

initialized at γ_t. Thus we have for $j = 1, ..., n$

$$\frac{\partial}{\partial \alpha_j}\left(\frac{\partial \gamma_\tau(\boldsymbol{\alpha}'_t)}{\partial \tau}\right) = \frac{\partial}{\partial \alpha_j}\left(\sum_{i=1}^{n} K(\gamma_\tau(\boldsymbol{\alpha}'_t), x_i)\alpha_i\right), \tag{3.69}$$

$$\frac{\partial}{\partial \tau}\left(\frac{\partial \gamma_\tau(\boldsymbol{\alpha}'_t)}{\partial \alpha_j}\right) = \sum_{i=1}^{n} \frac{\partial}{\partial \alpha_j} K(\gamma_\tau(\boldsymbol{\alpha}'_t), x_i)\alpha_i + \sum_{i=1}^{n} K(\gamma_\tau(\boldsymbol{\alpha}'_t), x_i)\frac{\partial \alpha_i}{\partial \alpha_j}, \tag{3.70}$$

which defines a differential equation for $\frac{\partial \gamma_\tau}{\partial \alpha_j}$. Integrating this equation together with (3.68) up to time T provides us with $\frac{\partial \gamma'_{t+1}}{\partial \boldsymbol{\alpha}'_t}$.

The last term in $\log p(\boldsymbol{\alpha}_{1:t}, \boldsymbol{\alpha}'_t | I_{1:t+1})$ is

$$\log p(\boldsymbol{\alpha}'_t | \boldsymbol{\alpha}_{0:t-1}) = -\frac{1}{2}\boldsymbol{\alpha}_t'^T \boldsymbol{\Sigma}(\chi_t)^{-1}\boldsymbol{\alpha}'_t, \tag{3.71}$$

so the final form of the gradient is:

$$\nabla_{\boldsymbol{\alpha}'_t} \log p(\boldsymbol{\alpha}_{0:t-1}, \boldsymbol{\alpha}'_t | I_{1:t+1}) =$$

$$= \int_{\gamma'_{t+1}} \left[\frac{|I(x) - \mu_{out}|^2}{2\sigma_{out}^2} - \frac{|I(x) - \mu_{in}|^2}{2\sigma_{in}^2} + \log\left(\frac{\sigma_{in}}{\sigma_{out}}\right)\right]\left(\frac{\partial \gamma'_{t+1}}{\partial \boldsymbol{\alpha}'_t}\right)^* \nu - \boldsymbol{\Sigma}(\chi_t)^{-1}\boldsymbol{\alpha}'_t. \tag{3.72}$$

This derivation is also used it when we implement a direct gradient ascent procedure to maximize this likelihood in Section 3.5.

In Table 3.3 we provide the steps of the Resample-Move algorithm when we use a random walk proposal. The number of MCMC moves is set to M (in practice M is

taken to be small, for example, equal to the number of available processors).

3.5 Numerical experiments

3.5.1 Initialization

We have initialized all our tracking experiments with an estimate for the boundary of the object in the initial frame, which was obtained by hand segmentation (in some cases it can be sufficient to use a reliable automatic segmentation method for the first frame). Based on this initial boundary we estimated the parameters of the observation likelihood, i.e. we calculated the means and variances of the image intensity values in the regions determined by the boundary.

3.5.2 Importance sampling-resampling

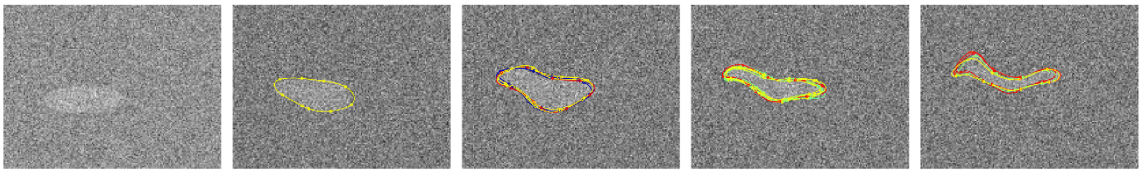
We demonstrate the performance of the importance sampling-resampling algorithm on simulated sequences (Figure 3.3), as well as on real videos (Figure 3.5 and Figure 3.6). The simulated sequences are created using the deformation model described in Section 3.2. We have generated 50 frames, however we display only a subset of them. We have used 1000 particles in the sampling scheme and we see that the estimated positions of the curves are consistent with the region boundaries. Usually only a few particles survive the resampling scheme and those are the ones plotted in the figure. The algorithm is also shown to be robust to occlusions and non-Gaussian

noise.

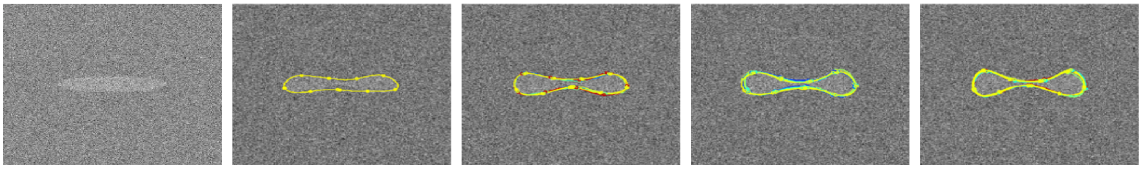
In Figure 3.4 we display how the standard Chan-Vese segmentation algorithm performs on the dumbbell example. The images have been processed using the algorithm provided online through [37]. It can be seen that without extra constraints on the shape, the segmentation splits the dumbbell into two parts.

In Figure 3.5 we compare the proposed method with a particle filter allowing only for affine transformations. The sequence (180 frames) on which we perform the comparison displays the motion of a paramecium which cannot be naturally described by affine transformations. Paramecia often change their direction of motion based on the objects they touch. Therefore, in our deformation model we have placed the control points equidistantly along the boundary. Using this model as a prior in the particle filtering scheme allows for good reconstruction of the boundary of the object.

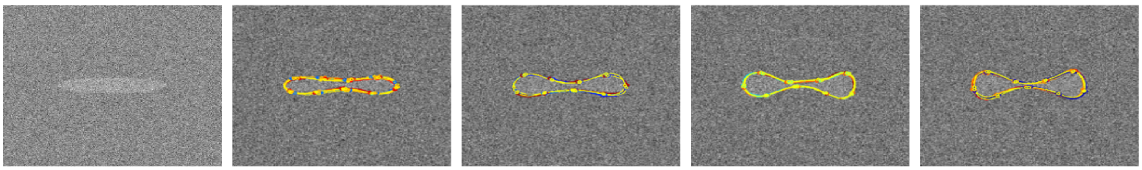
In Figure 3.6 we combine affine and non-affine models to track the motion of a fish in an aquarium. Main components of the fish motion are translation, rotation and scaling, and we preestimate them using an affine particle filter (any other method for affine registration could have been used). Then we incorporate these affine transformations in the proposed model (by composing them with the nonlinear diffeomorphism) and use importance sampling to estimate the final curves. We show that this procedure also works well under noise and occlusion.



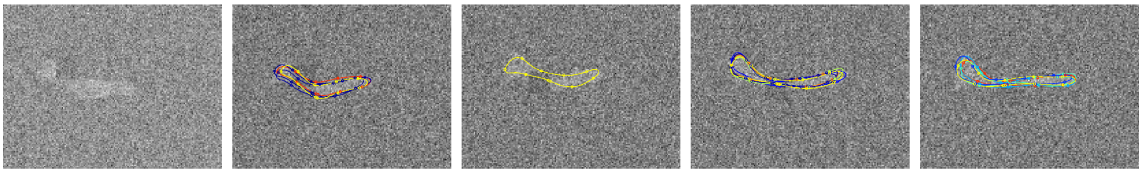
(a) **Tracking of a single contour.** We have simulated a sequence of random deformations of an ellipse and have added Gaussian noise to the shape and the background. The boundaries are successfully tracked despite the noise and the large deformation.



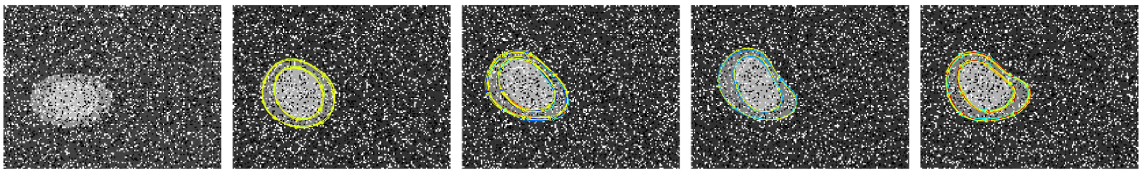
(b) **Tracking an almost self-intersecting contour.**We have generated a sequence of deterministic deformation from an ellipse to a dumbbell. The diffeomorphic model ensures the contour does not intersect itself.



(c) **Tracking with few control points.** We reduce the number of control points to 10 and still succeed in capturing the shape of the dumbbell.



(d) **Tracking under occlusions.** In this sequence a single contour is deformed and in some frames it is occluded by a dark object having a color similar to the background's color. The constraints on the deformation in the model allow to preserve the object's shape, even though the image frames are missing that information.



(e) **Tracking under noise.** We have simulated a sequence of random deformations of two closed contours. 25% of salt & pepper noise has been added to the images. The algorithm is robust to non-Gaussian noise and the two contours do not intersect each other.

Figure 3.3: Tracking simulated objects

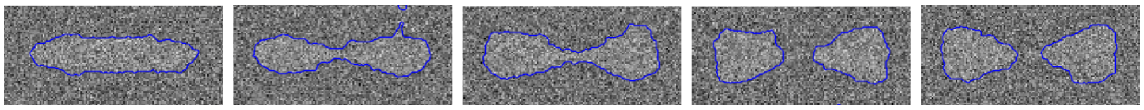
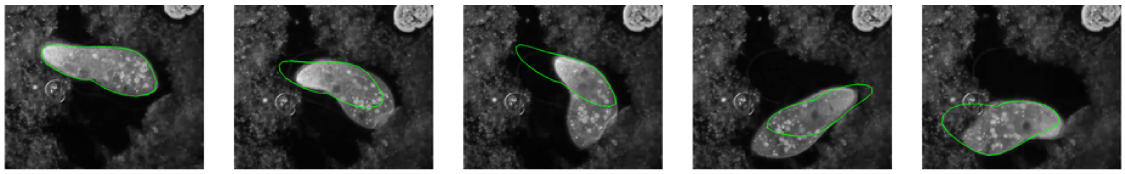
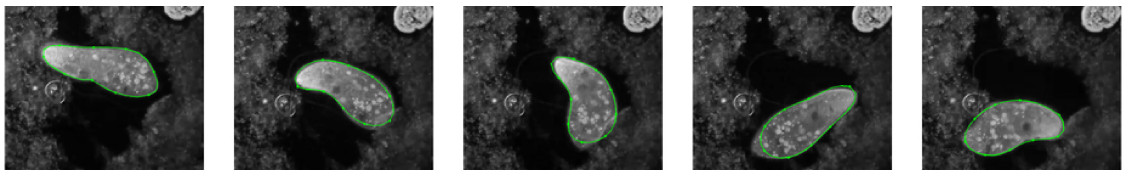


Figure 3.4: Chan-Vese static segmentation of the dumbbell sequence. The topology of the dumbbell shape is not preserved.



(a) **Affine model tracking.** The above sequence displays the motion of a paramecium. We have applied a particle filter with a simple affine transformation prior model to estimate the boundary of the paramecium. We can see that an affine transformation is insufficient to describe its motion.

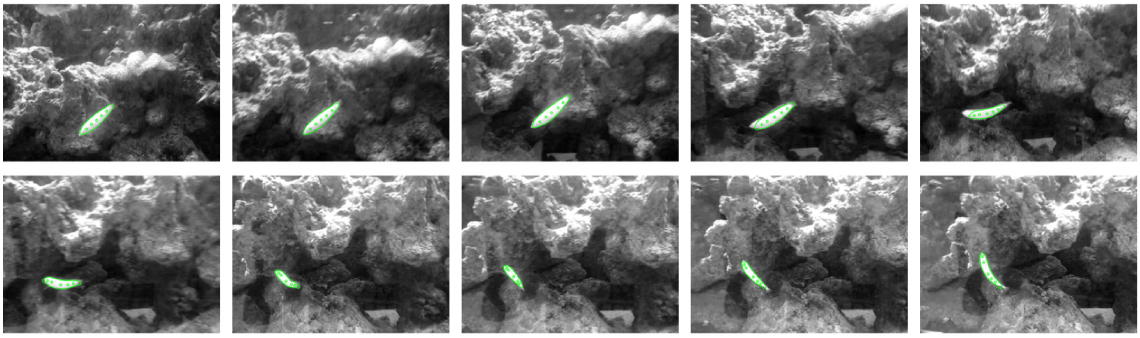


(b) **General nonlinear model tracking.** In the above sequence we show the performance of the particle filter with the deformation model in Section 3.2. Here we have plotted the average positions of the points on the curves in the final sample.

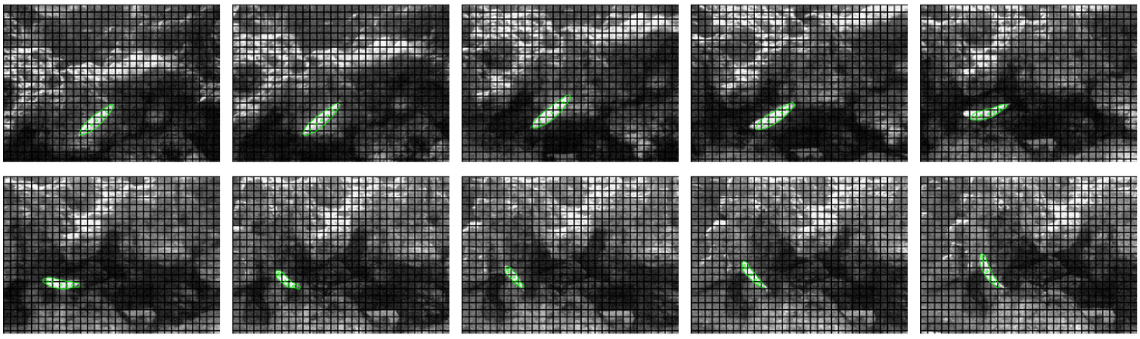
Figure 3.5: Tracking with an affine model vs. tracking with a general nonlinear model

3.5.3 Resample-Move

In Figure 3.8, we add a significant amount of Gaussian noise to the paramecium video, and we only use every 10^{th} frame of the sequence. The task is to track larger deformations with less clear object boundaries. The poor performance of the particle filter is improved by incorporating MCMC moves, resulting in a diversified sample, containing particles that are more representative of the solution. We also compare with a direct segmentation technique which maximizes the logarithm of the posterior given in 3.61 with respect to $\boldsymbol{\alpha_t}$ by using a gradient ascent method. The algorithm



(a) To track the movements of the fish, we first estimate its affine motion (using particle filtering), and then we incorporate it in our nonlinear model. When tracking the deformations, we select the control points to be along the medial axis of the shape, since this naturally describes the possible motions of the fish.



(b) Here we have added additional noise and an artificial grid to the image frames. The algorithm performs sufficiently well in the presence of these obstacles.

Figure 3.6: Combining affine and non-affine transformations

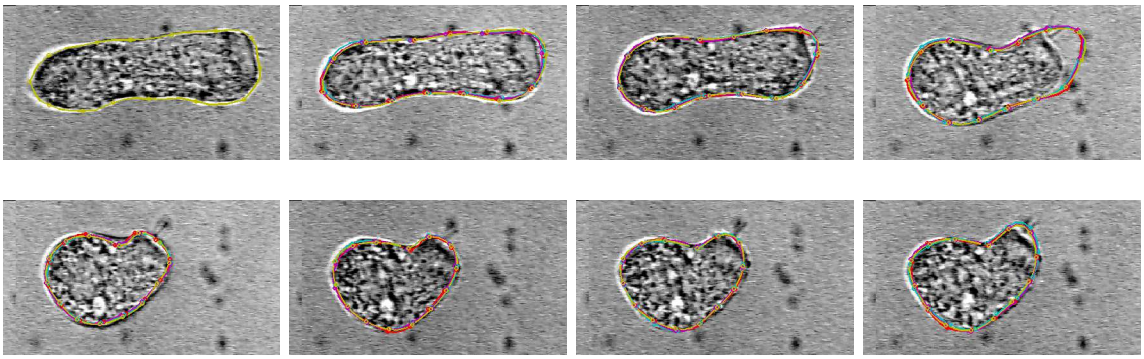
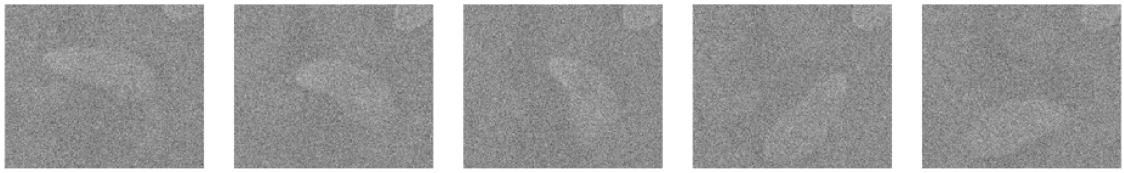


Figure 3.7: The above sequence displays contraction of a haircell as a response to a stimulant (image courtesy of J. Tilak Rathanater). The interior of the object consists of irregular texture of varying color which is unsuitable for intensity-based models. Particle filter (with MCMC) with an edge-based observation likelihood manages to track the deformation of the cell. There is a slight lag in frame four (where the contraction is fastest), but the algorithm manages to escape getting trapped by background false edges and extracts the correct boundary in the consequent frames.

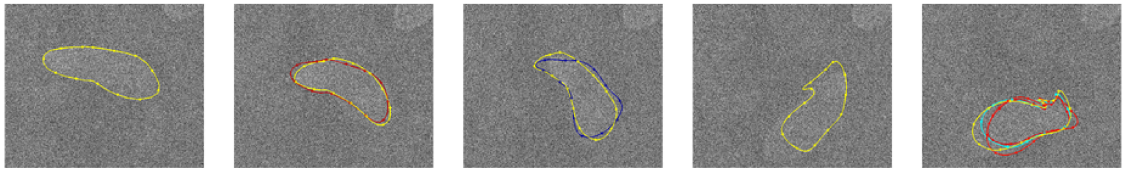
is prone to getting trapped in local maxima and fails to track the shape.

3.5.4 Tracking cardiac motion

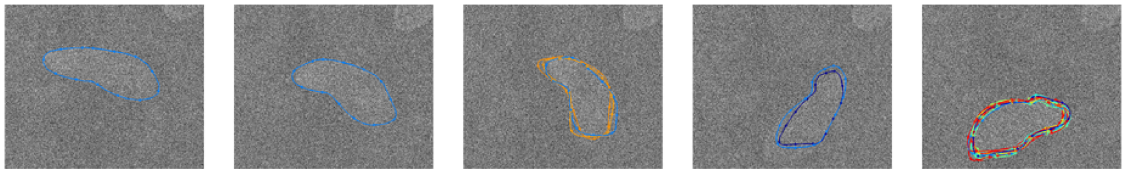
In this section we study the performance of these methods on the task of tracking the deformation of the left ventricle of a human heart. Cardiac MRIs usually consist of slices of the heart measured at different times of the cardiac cycle. Therefore, the left ventricle can be represented by its outermost (epicardium) and innermost (endocardium) layers. Single contour evolution methods have been previously used to either track only the endocardium, or track the endocardium first and separately track the epicardium afterwards. With our representation of the shape we can track both the endocardium and the epicardium simultaneously. We use data from the Sunnybrook Cardiac MR Database [74]. We apply the particle filter to a healthy



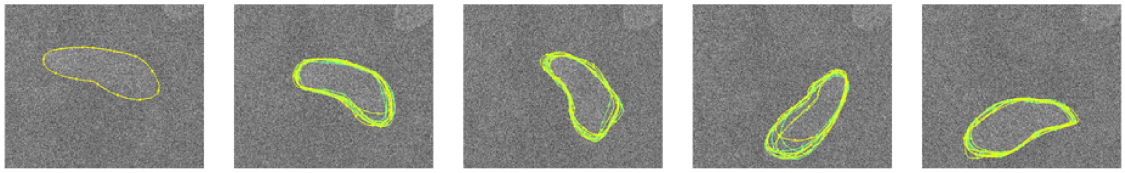
(a) Noisy sequence: we corrupt the paramecium sequence with a significant amount of noise.



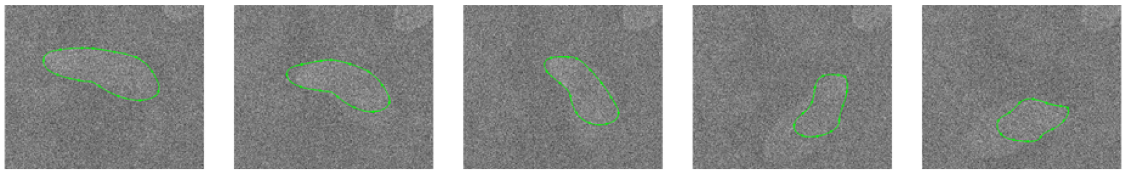
(b) Importance Sampling - Resampling I: although we take into account the larger frame rate and adjust the length of the evolution of the random deformations, the particle filter with 1000 particles does not perform well on this video sequence.



(c) Importance Sampling - Resampling II: even when we increase the number of particles to 10 000, the method creates artifacts.



(d) Importance Sampling - Resample Move: this approach succeeds in tracking the paramecium, since it generates a diverse sample of contours which allows it to capture the details of the shape properly.



(e) Likelihood Maximization: direct optimization fails due to local extrema.

Figure 3.8: Improving tracking through MCMC moves

heart image sequence in short-axis view by slightly modifying the model in Section 3.2.

Of particular interest when tracking heart sequences is measuring the width of the left-ventricle wall. Because of this, constraining the motion of the walls by a rigid prior can result in failing to extract important information from the images. On the other hand, a small correlation between the boundary points yields deformations which are not smooth enough to represent the walls of the heart. To achieve a compromise between these two situations we use a non-homogeneous covariance for the vector field values at the control points. In general, we expect points belonging to the epicardium to be highly correlated between each other, and the same should hold for points on the endocardium. However, we would like the two contours to move relatively freely with respect to each other, subject to the constraint that they do not intersect. We accomplish this by considering two different kernels K_{small} and K_{big}, where they both are assumed to be Gaussian kernels and σ_{small} (the width of K_{small}) is smaller than σ_{big} (the width of K_{big}). We construct the following matrices: $K_{\sigma_{big}}(\chi_{endo})$, $K_{\sigma_{big}}(\chi_{epi})$ (where χ_{endo} and χ_{epi} are the corresponding control points on the endocardium and epicardium) and $K_{small}(\chi)$ (χ contains all the control points as usual). Then we set the covariance of the Gaussian distribution defined over those control points to be:

$$C(\chi) = \frac{1}{2}K_{small}(\chi) + \frac{1}{2}\begin{bmatrix} K_{big}(\chi_{endo}) & 0 \\ 0 & K_{big}(\chi_{epi}) \end{bmatrix}, \tag{3.73}$$

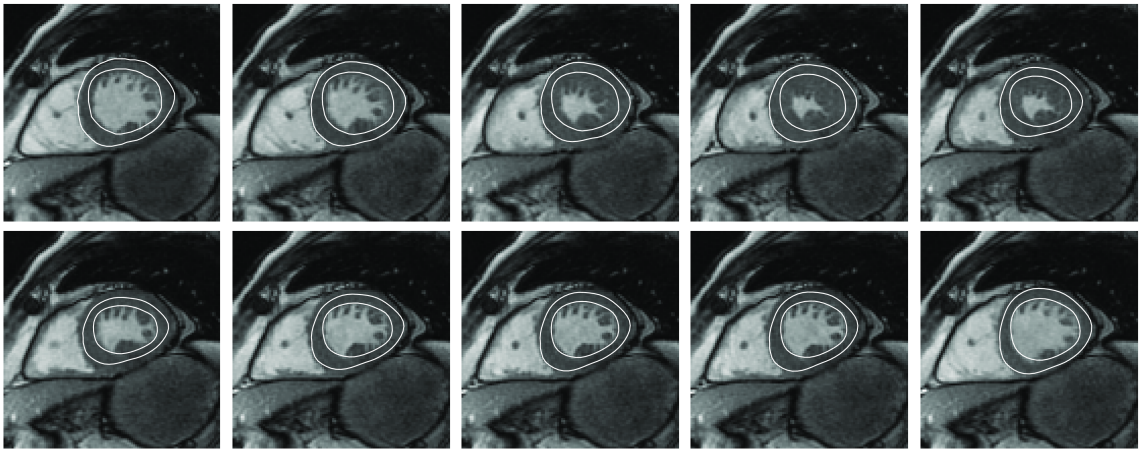
i.e. if x_i and x_j belong to the same contour, then

$$\mathbb{E}[v(x_i)v(x_j)^T] = \frac{1}{2}(K_{small}(x_i, x_j) + K_{big}(x_i, x_j))\mathbb{I}_2, \tag{3.74}$$

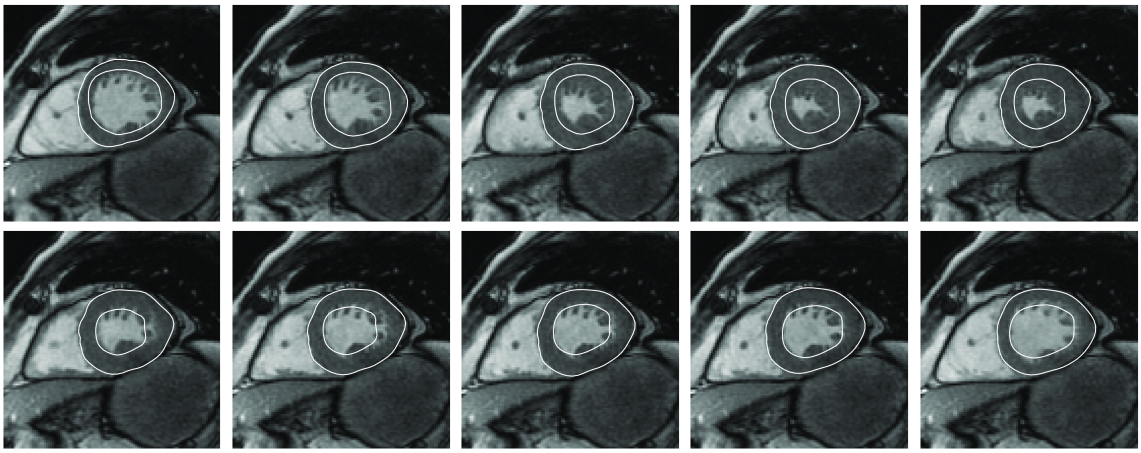
and if they belong to different contours, then $\mathbb{E}[v(x_i)v(x_j)^T] = \frac{1}{2}K_{small}(x_i, x_j)\mathbb{I}_2$. It turns out that such a covariance, although harder to analyze analytically, has the desired properties for tracking the heart. We display the result of tracking with such a model in Figure 3.9b. As compared to the performance of the regular particle filter (Figure 3.9a) we see that the modified model does not exhibit the problem of the epicardium sticking close to the endocardium and missing the correct boundary. We also provide the segmentation obtained through the direct likelihood maximization technique as described in the previous section (Figure 3.10a) and the built-in segmentation technique from the freely available software Segment [82] based on a 3D volume (2D+T) active contour segmentation method (Figure 3.10b). We observe that both of these methods tend to rely more on the image intensities, which results in a failure to separate the heart walls from the papillary muscles, a phenomenon common in heart segmentation. The particle filter relies on a sample of contours and thus is more robust to this type of issues.

We also note that diffeomorphic tracking algorithms provide us with additional information describing the deformation of the object: i.e. we can obtain trajectories for the points on the object which are not only part of the boundary. This allows us to study additional features: like expansion and contraction, by calculating the

Jacobian of the deformation. For the cardiac sequence, we propagate the inner part of the left ventricle using the deformation field of a single particle and calculate the Jacobian of the transformation. The result is displayed in Figure 3.11.

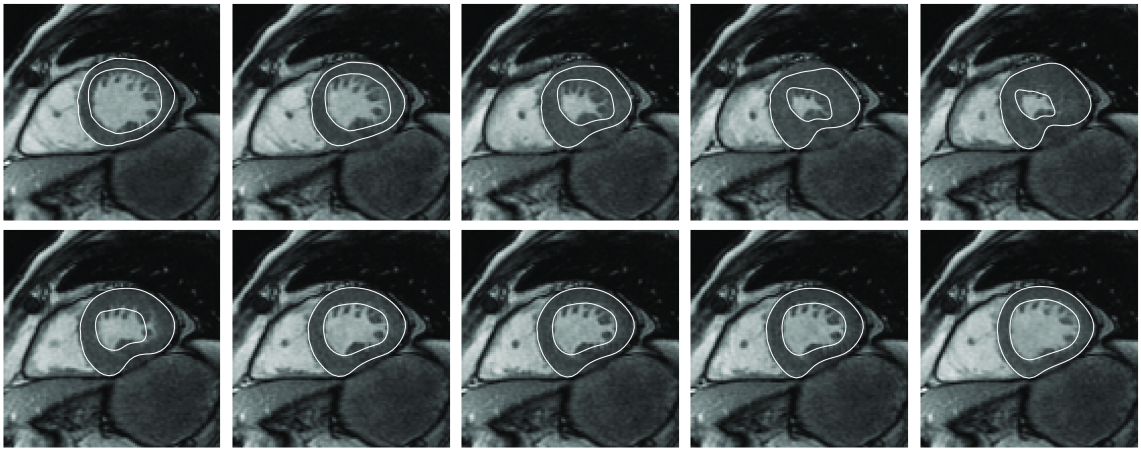


(a) Importance sampling-resampling - homogeneous covariance ($\sigma = 40, \sigma_K = 30$): the homogeneous covariance prevents the contours from separating and the expansion of the left ventricle wall is not captured (this is more visible in frames 5 and 6).

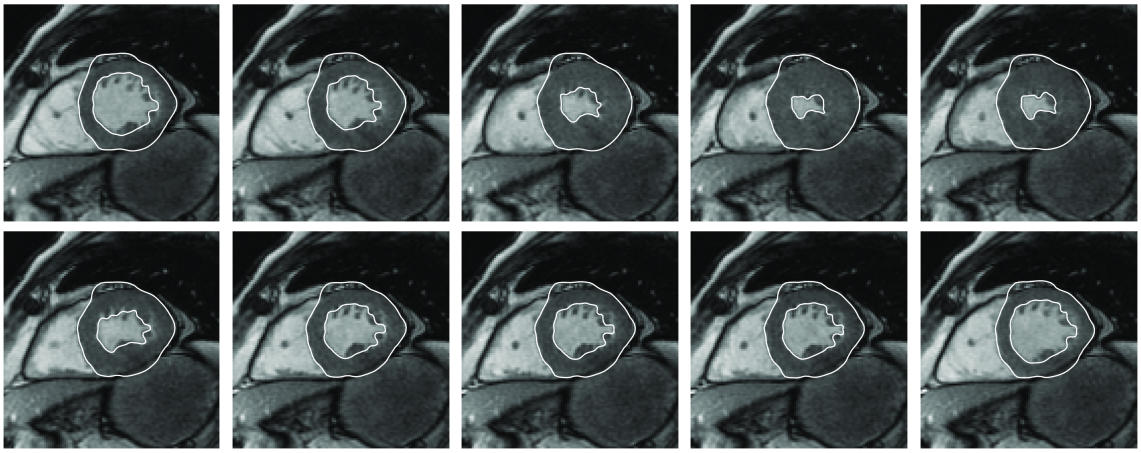


(b) Importance sampling-resampling - non-homogeneous covariance ($\sigma_{small} = 40, \sigma_{big} = 50, \sigma_K = 30$): the contours have more freedom to move independently and manage to follow the walls of the ventricle (compare the positions of the epicardium with those in the homogeneous case).

Figure 3.9: Tracking of a human heart left ventricle using particle filtering



(a) Segmentation with a prior: here we simply maximize the observation likelihood subject to a Gaussian prior constraint on the vector field. The outer contour gets trapped by few darker pixels (frame 5).



(b) Segmentation using the Segment software (2D+T): the package provides satisfactory automatic (no initial hand-segmentation needed) left ventricle segmentations. Note that without a prior the final contours are not necessarily smooth.

Figure 3.10: Segmentation of a human heart left ventricle using deterministic methods

Table 3.3: Resample-Move Algorithm

Resample-Move Algorithm

FOR $i = 1, ..., N$,

0. Set $\boldsymbol{\alpha}_0 = \boldsymbol{\alpha}_t^{(i)}$, $\gamma_0 = \gamma_{t+1}^{(i)}$.

FOR $k = 1, ..., M$,

1. Sample $\boldsymbol{\alpha}_{temp} = \boldsymbol{\alpha}_{k-1} + \delta\boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$.

2. Evolve the boundary points through

$$\partial_\tau \Phi(\gamma_t^{(i)}, \tau) = K(\Phi(\gamma_t^{(i)}, \tau), \chi_t^{(i)})\boldsymbol{\alpha}_{temp},$$

and set $\gamma_{temp} = \Phi(\gamma_t^{(i)}, T)$.

3. Evolve the control points through

$$\partial_\tau \Phi(\chi_t^{(i)}, \tau) = K(\Phi(\chi_t^{(i)}, \tau), \chi_t^{(i)})\boldsymbol{\alpha}_{temp},$$

and set $\chi_{temp} = \Phi(\chi_t^{(i)}, T)$.

4. Evaluate

$$r = \frac{p(I_{t+1}|\gamma_{temp})}{p(I_{t+1}|\gamma_k)} \frac{e^{-\boldsymbol{\alpha}_{temp}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\alpha}_{temp}/2}}{e^{-\boldsymbol{\alpha}_{k-1}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\alpha}_{k-1}/2}}.$$

5. Sample $u \sim \mathcal{U}(0, 1)$.

6. If $u < \min(r, 1)$, set $\boldsymbol{\alpha}_k = \boldsymbol{\alpha}_{temp}$, $\gamma_k = \gamma_{temp}$, $\chi_k = \chi_{temp}$, else, set $\boldsymbol{\alpha}_k = \boldsymbol{\alpha}_{k-1}$, $\gamma_k = \gamma_{k-1}$, $\chi_k = \chi_{k-1}$.

END

7. Set $\boldsymbol{\alpha}_t^{(i)} = \boldsymbol{\alpha}_M$, $\gamma_{t+1}^{(i)} = \gamma_M$, $\chi_{t+1}^{(i)} = \chi_M$.

END

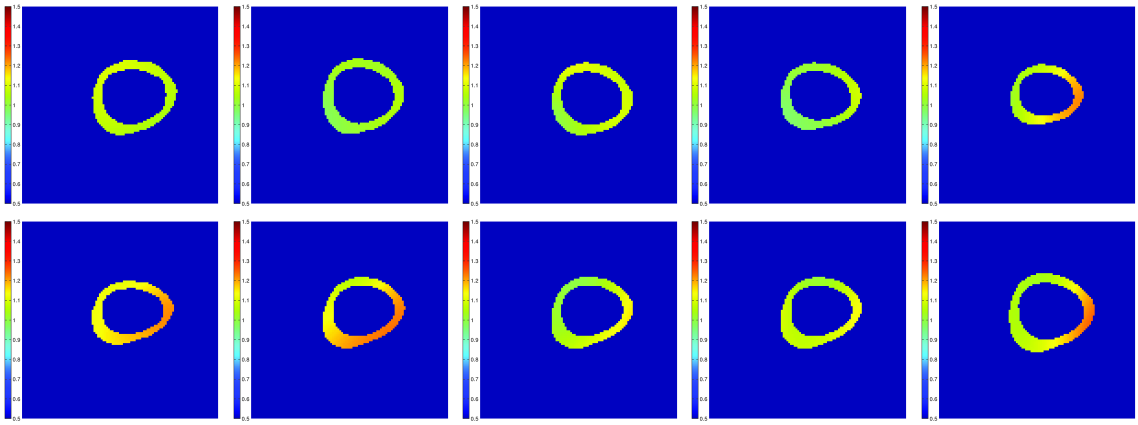


Figure 3.11: Jacobian field - in this figure we display how the tracking based on the curve affects the points in the domain of the image. The color field indicates the value of the determinant of the Jacobian of the deformation with respect to the first frame. This can provide us with additional knowledge about the geometry of the heart motion that is usually not available with static segmentation algorithms.

Chapter 4

Convergence of Gaussian Random Fields Indexed by Curves

4.1 Introduction

In the previous chapter we constructed distributions of random shapes by defining finitely-generated Gaussian random vector fields and deforming template shapes along their flow. Although for numerical reasons we used a finite number of control points to generate the vector fields, it is natural to ask the following question:

How do the random shapes behave as we increase the number of control points?

Mathematically, to answer this question we need to study the properties of the

distributions of the random vector fields generating the random shape deformations as we place the control points in a way that they better approximate the boundary of the shape of interest. Ideally, if we place the control points densely along a smooth closed curve, we would like the limiting distribution of the vector fields to coincide with some meaningful infinite-dimensional measure defined over the curve. Or, if we reverse the problem, we would like to first define a distribution for the vector fields defined on the contour, with properties which ensure that the deformations of the contour along its flow are diffeomorphisms, and then consider approximations of it through finitely-generated vector fields.

Contribution. In this chapter we study the convergence of a sequence of Gaussian random fields indexed by a closed curve with covariance of the form as in (3.24). We prove weak convergence with respect to L^2-norm holds. We show some interesting examples in which the convergence in RKHS norm does not hold: for example, when the curve contains a flat region, or when the kernel of the RKHS and the covariance are both Gaussian functions.

4.2 Convergence on $L^2(\gamma)$

Let's denote the curve by $\gamma \subset \mathbb{R}^2$. Define a sequence of finite sets χ_n on γ such that $\cup_{n=1}^{\infty}\chi_n$ is dense in γ (each set contains n distinct points) and $\chi_n \subset \chi_m$ for $n \leq m$. Let $K(\cdot,\cdot)$ be a reproducing kernel on γ. By restricting this kernel to the sets χ_n we obtain a sequence of reproducing kernel Hilbert spaces $V(\chi_n)$. We assume that we are given

a sequence of Gaussian random fields over $\mathbb{R}^2$ $(\xi_n, n > 1)$ which are centered and their covariances $C_n(x, y)$ are such that the random fields are consistent with respect to their finite-dimensional projections on given reproducing kernel Hilbert spaces (note we are going to establish convergence properties for random fields with realizations in $\mathbb{R}$, i.e. for each coordinate of a random vector field). For consistency, we want projection of the random field ξ_m with measure P_m onto $V(\chi_n)$ (where $n < m$) to have the same distribution as ξ_n with measure P_n. We have already established two cases when this is true. The first case is when we derive the covariance directly from the reproducing kernel:

$$\text{Case I}: \qquad C_n(x, y) = \mathbf{K}(x, \chi_n)\mathbf{K}^{-1}(\chi_n, \chi_n)\mathbf{K}(\chi_n, y). \tag{4.1}$$

For points $x, y \in \chi_n$, $C_n(x, y) = K(x, y)$. In general the choice of the kernel for the norm of the reproducing kernel Hilbert space is not necessary related to the choice of the covariance. For the second case, we assume $C(x, y)$ is a covariance reflecting the properties of the Gaussian random field defined over γ. We can build approximate Gaussian random fields ξ_n by requiring that $C_n(x, y) = C(x, y)$ for points $x, y \in \chi_n$. If we ensure that the consistency property holds, we obtain the following form of the covariance:

$$\text{Case II:} \quad C_n(x, y) = \mathbf{K}(x, \chi_n)\mathbf{K}(\chi_n, \chi_n)^{-1}\mathbf{C}(\chi_n, \chi_n)\mathbf{K}(\chi_n, \chi_n)^{-1}\mathbf{K}(\chi_n, y). \tag{4.2}$$

Although the first case is covered by the second case when $K = C$, we present them separately as the treatment in the first case is more straightforward and easier to follow.

Theorem 4.1. *Under the following assumptions:*

Case I: K is continuous;

Case II: K, C are continuous and $V(C) \subset V(K)$,

P_n converge weakly to a Gaussian measure on $L^2(\gamma)$.

Before we begin the proof, we provide the following corollary when the covariance and the reproducing kernel are both Gaussian.

Corollary 4.2. *Let $C(x,y) = \frac{1}{2\pi\sigma_1^2} e^{-\|x-y\|^2/2\sigma_1^2}$ and $K(x,y) = \frac{1}{2\pi\sigma_0} e^{-\|x-y\|^2/2\sigma_0^2}$. For $\sigma_0 \leq \sigma_1$ the sequence of measures P_n converge weakly to a Gaussian measure P on $L^2(\gamma)$.*

Proof. (Corollary 4.2) We need to show that the conditions of Theorem 4.1 are satisfied. The first case is trivial, as the Gaussian kernel is continuous. For the second case, an equivalent condition for the RKHS with kernel C to belong to a RKHS with kernel K is that there a positive constant B such that $BK - C$ is nonnegative definite [6](Corollary IV_2, p. 383). Set $B = \sigma_1/\sigma_0$, and consider the inverse Fourier transform

of this difference:

$$\mathcal{F}^{-1}\{BK - C\} = 2\pi\sigma_1^2\mathcal{F}^{-1}\left\{\frac{1}{2\pi\sigma_0^2}e^{-\|x-y\|_2^2/2\sigma_0^2} - \frac{1}{2\pi\sigma_1^2}e^{-\|x-y\|_2^2/2\sigma_1^2}\right\} = \quad (4.3)$$

$$= 2\pi\sigma_1^2(e^{-2\sigma_0^2\|x-y\|_2^2} - e^{-2\sigma_1^2\|x-y\|_2^2}). \quad (4.4)$$

Clearly when $\sigma_0 < \sigma_1$, the right side is positive so by Bochner's theorem we conclude that $K - C$ is a positive definite function, and $V(C) \subset V(K)$. □

Proof. (Theorem 4.1)

Pointwise Covariance Convergence. We will first show using properties of reproducing kernel Hilbert spaces that $C_n(x, y) \to C(x, y)$ as $n \to \infty$ for x, y on γ.

Case I: We recall that the form of the orthogonal projection of $K(\cdot, x)$ on $V(\chi_n)$ is

$$\pi_{V(\chi_n)}[K(\cdot, x)] = \mathbf{K}(\chi_n, x). \quad (4.5)$$

Therefore, we can rewrite $C_n(x, y)$ as

$$C_n(x, y) = \mathbf{K}(x, \chi_n)\mathbf{K}^{-1}(\chi_n, \chi_n)\mathbf{K}(\chi_n, y) = \quad (4.6)$$

$$= \langle \pi_{V(\chi_n)(K)}[K(\cdot, x)], \pi_{V(\chi_n)(K)}[K(\cdot, y)]\rangle_{V_\gamma(K)}. \quad (4.7)$$

By Theorem 6E of (Parzen, 1959 [73]) we have that

$$\langle \pi_{V(\chi_n)(K)}[K(\cdot,x)], \pi_{V(\chi_n)(K)}[K(\cdot,y)] \rangle_{V_\gamma(K)} \to \langle K(\cdot,x), K(\cdot,y) \rangle_{V(\gamma)}, \tag{4.8}$$

and we conclude

$$\lim_{n\to\infty} C_n(x,y) = \langle K(\cdot,x), K(\cdot,y) \rangle_{V(\gamma)}. \tag{4.9}$$

When x or y belongs to γ, $C_n(x,y) \to K(x,y)$.

Case II: Theorem 1 in [30] states that $V_\gamma(C) \subset V_\gamma(K)$ implies also the existence of a nonnegative self-adjoint bounded linear operator $G : V_\gamma(K) \to V_\gamma(C)$ which satisfies

$$G[K(\cdot,x)] = C(\cdot,x) \quad \forall x \in \gamma, \tag{4.10}$$

and $\|G\| \leq B$, such that $BK - C \geq 0$. If we apply this operator to an element of $V_{\chi_n}(K)$

which takes the form $v(\cdot) = \sum_{i=1}^n K(\cdot,x_i)\alpha_i$, we obtain

$$Gv = G\left[\sum_{i=1}^n K(\cdot,x_i)\alpha_i\right] = \sum_{i=1}^n C(\cdot,x_i)\alpha_i. \tag{4.11}$$

Also, recall that the vector of coefficients of the projection of $K(\cdot,y)$ onto $V_{\chi_n}(K)$ is

$K(\chi_n, \chi_n)^{-1} K(\chi_n, y)$. Thus we obtain

$$\begin{aligned} G[\pi_{V_{\chi_n}(K)}[K(\cdot, y)]] &= G\left[\sum_{i=1}^{n} K(\cdot, x_i)[K(\chi_n, \chi_n)^{-1} K(\chi_n, y)]_i\right] = \\ &= C(\cdot, \chi_n) K(\chi_n, \chi_n)^{-1} K(\chi_n, y). \end{aligned} \tag{4.12}$$

This yields an alternative representation of the covariance:

$$C_n(x, y) = \langle \pi_{V_{\chi_n}(K)}\{K(\cdot, x)\}, G[\pi_{V_{\chi_n}(K)}\{K(\cdot, y)\}]\rangle_{V(\gamma)}. \tag{4.13}$$

Since G is self-adjoint and nonegative, we can define the following inner product:

$$\langle\langle f, g\rangle\rangle = \langle f, Gg\rangle_{V(\gamma)}(K). \tag{4.14}$$

The corresponding inner product space is a reproducing kernel Hilbert space, and we denote its reproducing kernel by $\hat{K}(x, y)$. With the above notation,

$$C_n(x, y) = \langle\langle \pi_{V_{\chi_n}(K)}\{K(\cdot, x)\}, \pi_{V_{\chi_n}(K)}\{K(\cdot, y)\}\rangle\rangle. \tag{4.15}$$

We will show that $C_n(x, y) \to \langle\langle K(\cdot, x), K(\cdot, y)\rangle\rangle$ as $n \to \infty$. For that we will use the following theorem by Parzen (Theorem 6D,[73]), which we summarize using our notation

Theorem 4.3. *Let $\{V_n, n = 1, 2, ...\}$ be a sequence of Hilbert subspaces of V which are either (i) monotone non-decreasing; that is, $V_n \subset V_{n+1}$, or (ii) monotone non-increasing; that is $V_n \supset V_{n+1}$. Define V_∞ to be, in case (i), the Hilbert subspace of V spanned by the union $\cup_{n=1}^{\infty} V_n$, and, in case (ii), the intersection $\cap_{n=1}^{\infty} V_n$. Let $v_1, v_2, ...$ be a sequence of vectors in V such that for every integer m and n*

$$\pi_{V_m}[v_n] = v_m \quad \textit{if} \quad m \leq n. \tag{4.16}$$

Then there is a unique vector v in V_∞ such that $v_n = \pi_{V_n}[v] \quad \forall n$ and

$$\lim_{n\to\infty} \|v_n - v\| = 0 \quad \textit{iff} \quad \lim_{n\to\infty} \|v_n\|^2 \leq \infty. \tag{4.17}$$

If w is a vector in V such that for all n

$$v_n = \pi_{V_n}[w], \tag{4.18}$$

then

$$v = \pi_{V_\infty}[w]. \tag{4.19}$$

We will set $V = V_\gamma(\hat{K})$ (the vector space with the newly introduced inner product), and for $V_1, V_2, ...$ we will take the corresponding sequence of restricted subspaces $V_{\chi_1}(\hat{K}), V_{\chi_2}(\hat{K}), ...$ They are clearly nested and non-decreasing. We consider the

sequence of vectors $\left\{v_n | v_n = \pi_{V_{\chi_n}(K)}[K(\cdot, x)]\right\}$. We will show that $\pi_{V_{\chi_m}(\hat{K})} v_n = v_m$ for $m \leq n$.

$$\begin{aligned}
\pi_{V_{\chi_m}(\hat{K})}[\pi_{V_{\chi_n}(K)}[K(\cdot,x)]] &= \pi_{V_{\chi_m}(\hat{K})}\left[\sum_{i=1}^{n} K(\cdot, x_i)[K(\chi_n,\chi_n)^{-1}K(\chi_n,x)]_i\right] = \\
&= \sum_{j=1}^{m} \hat{K}(\cdot,x_j)\left[\hat{K}(\chi_m,\chi_m)^{-1}\sum_{i=1}^{n} K(\chi_m,x_i)[K(\chi_n,\chi_n)^{-1}K(\chi_n,x)]_i\right]_j = \\
&= \sum_{j=1}^{m} \hat{K}(\cdot,x_j)\left[K(\chi_m,\chi_m)^{-1}K(\chi_m,\chi_n)K(\chi_n,\chi_n)^{-1}K(\chi_n,x)\right]_j = \\
&= \sum_{j=1}^{m} \hat{K}(\cdot,x_j)\left[K(\chi_m,\chi_m)^{-1}K(\chi_m,x)\right]_j = \\
&= \pi_{V_{\chi_m}(\hat{K})}[K(\cdot,x)]. \qquad (4.20)
\end{aligned}$$

Consider $v = \pi_\gamma(\hat{K})[K(\cdot, x)]$. It, of course, belongs to $V_\gamma(\hat{K})$, and we have

$$v_n = \pi_{V_{\chi_m}(\hat{K})}[K(\cdot,x)] = \pi_{V_{\chi_m}(\hat{K})}[\pi_{V_\gamma(\hat{K})}K(\cdot,x)] = \pi_{V_{\chi_m}(\hat{K})}[v], \qquad (4.21)$$

thus showing that (4.18) is satisfied, and $v_n \to v$ in the norm of $V_\gamma(\hat{K})$. From

$$\|\pi_{V_{\chi_n}(\hat{K})}[K(\cdot,x)] - \pi_{V_\gamma(\hat{K})}[K(\cdot,x)]\|_{V_\gamma(\hat{K})} \to 0, \qquad (4.22)$$

$$\|\pi_{V_{\chi_n}(\hat{K})}[K(\cdot,y)] - \pi_{V_\gamma(\hat{K})}[K(\cdot,y)]\|_{V_\gamma(K)} \to 0, \qquad (4.23)$$

and using the polarization identity, we conclude that

$$\langle\langle \pi_{V_{\chi n}(\hat{K})}[K(\cdot,x)], \pi_{V_{\chi n}(\hat{K})}[K(\cdot,y)]\rangle\rangle \to \langle\langle \pi_{V_\gamma(\hat{K})}[K(\cdot,x)], \pi_{V_\gamma(\hat{K})}[K(\cdot,y)]\rangle\rangle, \quad (4.24)$$

and hence

$$\begin{aligned} \lim_{n\to\infty} C_n(x,y) &= \langle\langle \pi_{V_\gamma(\hat{K})}[K(\cdot,x)], \pi_{V_\gamma(\hat{K})}[K(\cdot,y)]\rangle\rangle = \\ &= \langle \pi_{V_\gamma(K)}[K(\cdot,x)], \pi_{V_\gamma(K)}[K(\cdot,y)]\rangle_{V_\gamma(K)}. \end{aligned} \quad (4.25)$$

Clearly, when $x, y \in \gamma$, $\lim_{n\to\infty} C_n(x,y) = C(x,y)$.

Boundedness. Here we will also show that for each n $C_n(x,y)$ is bounded by a function which integrable on γ.

Case I: By Cauchy-Schwartz

$$\begin{aligned} C_n(x,y) &= \langle \pi_{V_{\chi n}(K)}[K(\cdot,x)], \pi_{V_{\chi n}(K)}[\pi_{V_\chi}(K)(\cdot,y)]\rangle_{V_\gamma(K)} \leq \\ &\leq \|\pi_{V_{\chi n}(K)}[K(\cdot,x)]\|_{V_\gamma(K)} \|\pi_{V_{\chi n}(K)}[K(\cdot,y)]\|_{V_\gamma(K)} \leq \\ &\leq \|K(\cdot,x)\|_{V_\gamma(K)} \|K(\cdot,y)\|_{V_\gamma(K)} = K(x,x)K(y,y), \end{aligned} \quad (4.26)$$

which is a constant.

Case II: Similarly,

$$\begin{aligned} C_n(x,y) &= \langle \pi_{V_{\chi_n}(K)}[K(\cdot,x)], G[\pi_{V_{\chi_n}(K)}[K(\cdot,y)]]\rangle_{V(\gamma)} \leq \\ &\leq \|\pi_{V_{\chi_n}(K)}[K(\cdot,x)]\|_{V_\gamma(K)} \|G[\pi_{V_{\chi_n}(K)}[K(\cdot,y)]]\|_{V_\gamma(K)} \leq \\ &\leq \|K(\cdot,x)\|_{V_\gamma(K)} \|G\| \|K(\cdot,y)\|_{V_\gamma(K)} \leq BK(x,x)K(y,y). \end{aligned} \tag{4.27}$$

Weak Convergence. Now we would like to show that for a general choice of $C_n(x,y) \to C(x,y)$, the corresponding probability measures converge weakly. There are different ways to formulate the condition for weak convergence of Gaussian measures on a Hilbert space. It usually comes down to showing the convergence of the means and the covariance operators, and also showing relative weak compactness of the measures. The following set of conditions in (Baushev, 1986 [12]) are necessary and sufficient for the convergence of a sequence of Gaussian measures P_n to a Gaussian measure P on $L^2(\gamma)$

(i) $\|\mu_n - \mu\| \to 0$, where μ_n and μ are the means corresponding to the measures P_n and P

(ii) $\langle C_n e_i^*, e_j^* \rangle \to \langle C e_i^*, e_j^* \rangle$, where C_n and C are the corresponding covariance operators and $\{e_i^*\}$ is the dual basis in $L^2(\gamma)^*$

(iii) $\|x_n - x\| \to 0$, where $x_n = \sum_{k=1}^{\infty} \langle C_n e_k^*, e_k^* \rangle^{1/2} e_k$ and $x = \sum_{k=1}^{\infty} \langle C e_k^*, e_k^* \rangle^{1/2} e_k$.

In our case the measures have zero mean so the first condition is trivially satisfied.

To proceed we need to determine the form of the covariance operators C_n. First note that the dual of $L^2(\gamma)$ can be identified with $L^2(\gamma)$, so C_n can be treated as an operator on $L^2(\gamma)$. Let f, g be arbitrary functions in $L^2(\gamma)$, and μ and ν be Lebesgue measures on γ. Then the covariance operator is such that

$$\begin{aligned}
\langle C_n f, g\rangle_{L^2(\gamma)} &= \mathbb{E}\left[\langle f, \xi_n\rangle_{L^2(\gamma)}\langle g, \xi_n\rangle_{L^2(\gamma)}\right] = \mathbb{E}\left[\int_\gamma f(x)\xi_n(x)d\mu(x)\int_\gamma g(y)\xi_n(y)d\nu(y)\right] = \\
&= \int_\gamma\int_\gamma f(x)\mathbb{E}[\xi_n(x)\xi_n(y)]g(x)d\mu(x)d\nu(y) = \\
&= = \int_\gamma\int_\gamma f(x)C_n(x,y)g(x)d\mu(x)d\nu(y) = \\
&= \int_\gamma g(x)\left[\int_\gamma C_n(x,y)f(y)d\nu(y)\right]d\mu(x). \qquad (4.28)
\end{aligned}$$

Thus we see that the action of the operator on a function f is

$$C_n[f](\cdot) = \int_\gamma C_n(\cdot, y)f(y)d\nu(y). \qquad (4.29)$$

Our first step is to show that $\langle C_n f, f\rangle_{L^2(\gamma)} \to \langle Cf, f\rangle_{L^2(\gamma)}$ for an arbitrary function f.

$$\begin{aligned}
\lim_{n\to\infty}\langle C_n f, f\rangle_{L^2(\gamma)} &= \lim_{n\to\infty}\int_\gamma C_n[f](x)f(x)d\mu(x) = \\
&= \lim_{n\to\infty}\int_\gamma\int_\gamma C_n(x,y)f(y)f(x)d\mu(x)d\nu(y). \qquad (4.30)
\end{aligned}$$

We have shown that $\lim_{n\to\infty} C_n(x,y) = C(x,y)$ for every x and y, and each $C_n(x,y)$ is bounded by a constant (we will denote this bound by B_1) so we can apply the dominated convergence theorem in (4.30). First,

$$\begin{aligned}\int_\gamma\int_\gamma |C_n(x,y)f(x)f(y)|d\mu(x)d\nu(y) &= \int_\gamma\int_\gamma |C_n(x,y)||f(x)||f(y)| \leq \\ &\leq \int_\gamma\int_\gamma B_1|f(x)||f(y)|d\mu(x)d\nu(y) = \\ &= B_1\|f\|^2_{L^1(\gamma)} \leq B_1\mu(\gamma)\|f\|^2_{L^2(\gamma)} < \infty,\end{aligned} \tag{4.31}$$

where the last quantity is finite since γ is bounded and $f \in L^2(\gamma)$. Next,

$$\int_\gamma\int_\gamma |C(x,y)||f(x)||f(y)|d\mu(x)d\nu(y) = \int_\gamma\int_\gamma B_2|f(x)||f(y)|d\mu(x)d\nu(y) < \infty \tag{4.32}$$

nonumber (4.33)

(4.34)

where B_2 is the upper bound of $C(x,y)$ which exists since it is a continuous function on a compact domain.

Therefore, $C_n(x,y)f(x)f(y)$ is bounded by the integrable function $B_1|f(x)||f(y)|$

and by DCT we can interchange the limit with the integral and obtain

$$\begin{aligned}\lim_{n\to\infty}\langle C_n f, f\rangle_{L^2(\gamma)} &= \lim_{n\to\infty}\int_\gamma\int_\gamma C_n(x,y)f(y)f(x)d\mu(x)d\nu(y) = \\ &= \int_\gamma\int_\gamma C(x,y)f(x)f(y)d\mu(x)d\nu(y) = \langle Cf, f\rangle_{L^2(\gamma)}.\end{aligned} \tag{4.35}$$

Now let's define $f = e_i - e_j$. By the above conclusion we have

$$\lim_{n\to\infty}\langle C_n(e_i - e_j), e_i - e_j\rangle = \langle C(e_i - e_j), e_i - e_j\rangle \tag{4.36}$$

Therefore,

$$\lim_{n\to\infty}\langle C_n e_i, e_i\rangle + \langle C_n e_j, e_j\rangle - 2\langle C_n e_i, e_j\rangle = \langle Ce_i, e_i\rangle + \langle Ce_j, e_j\rangle - 2\langle Ce_i, e_j\rangle, \tag{4.37}$$

and since the first two terms on the right hand side converge to the first two terms on the left hand side, we have (ii)

$$\lim_{n\to\infty}\langle C_n e_i, e_j\rangle = \langle Ce_i, e_j\rangle. \tag{4.38}$$

The third condition turns out to be equivalent to the convergence of the covariance

operators in nuclear norm:

$$\|x_n - x\|^2 = \left\| \sum_{k=1}^{\infty} \langle (C_n - C)e_k, e_k \rangle^{1/2} e_k \right\|^2 = \sum_{k=1}^{\infty} \langle (C_n - C)e_k, e_k \rangle = tr(C_n - C) \quad (4.39)$$

First we need to check whether C and C_n are trace class operators. The answer is positive due to the continuity of $C(x, y)$ and the compactness of γ. By Mercer's theorem we can obtain the following representation for the kernel

$$C(x, y) = \sum_{k=1}^{\infty} \lambda_k \varphi_k(x) \varphi_k(y), \quad \text{for } x, y \in \gamma \quad (4.40)$$

where $\{\lambda_k\}$ and $\{\varphi_k\}$ are the eigenvalues and corresponding eigenfunctions of the integral operator

$$[Cf](\cdot) = \int_{\gamma} C(\cdot, x) f(x) dx, \quad (4.41)$$

and this convergence is uniform. Let's consider the form of the trace in this case

$$\begin{aligned} tr(C) &= \sum_{k=1}^{\infty} \langle C\varphi_k, \varphi_k \rangle = \sum_{k=1}^{\infty} \langle \lambda_k \varphi_k, \varphi_k \rangle = \sum_{k=1}^{\infty} \lambda_k \int_{\gamma} \varphi(x) \varphi_k(x) d\mu(x) = \\ &= \int_{\gamma} \left[\sum_{k=1}^{\infty} \lambda_k \varphi(x) \varphi_k(x) \right] d\mu(x) = \int_{\gamma} C(x, x) d\mu(x) \leq \infty. \end{aligned} \quad (4.42)$$

In the above calculation we were allowed to exchange the infinite sum with the integral thanks to the uniform convergence. C_n is continuous on γ as well, as it depends on

x and y only through the functions $K(\chi_n, x)$ and $K(\chi_n, y)$ which are continuous as well. Therefore, by the above trace argument, we have

$$tr(C_n) = \int_\gamma C_n(x,x)d\mu(x) \leq \int_\gamma C(x,x)d\mu(x) \leq \infty. \tag{4.43}$$

Since $C_n(\cdot,\cdot)$ is bounded we can apply DCT again:

$$\lim_{n\to\infty} tr(C_n) = \lim_{n\to\infty} \int_\gamma C_n(x,x)d\mu(x) = \int_\gamma \lim_{n\to\infty} C_n(x,x)d\mu(x) = tr(C). \tag{4.44}$$

Thus (iii) is established and hence $P_n \xrightarrow{w} P$, where P is a Gaussian measure on L^2_γ corresponding to the Gaussian random field defined over γ with zero mean and covariance $C(x,y)$. □

4.3 Convergence in RKHS norm

In the previous section we showed the convergence of the random vector fields ξ_n with covariance $C_n(x,y)$ to the random vector field ξ with a covariance $C(x,y)$ on γ. We are interested in the properties of these random vector fields and how they affect the types of shapes which can be obtained by evolving γ along their realizations. We have already shown in Proposition 3.1 that the flow associated with stationary Gaussian random vector fields with Gaussian-kernel covariance is diffeomorphic. We are also interested in deformations generated by time-dependent vector fields such as in (3.33).

It has been shown in [99] (Theorem 8.7 p.165), that if the vector field ξ belongs to a RKHS with a sufficiently nice kernel, the solutions of the flow are diffeomorphisms, thus the random shapes have the same topology as the original shape. Clearly, every χ_n as an element of $V(\chi_n)$ belongs to V_γ since it can be written as a finite linear combination of the kernel functions with normally distributed coefficients. The question is whether their limit ξ belongs to $V(\gamma)$ too, i.e. whether these vector fields converge in the norm of the reproducing kernel Hilbert space.

In short, we would like to know whether the realizations of a Gaussian random field ξ with covariance given by the kernel C belongs to a reproducing kernel Hilbert space with a kernel K. The circumstances under which this is true have been studied in [30]. The following zero-one law holds for any continuous kernels C and K and continuous realizations of the random vector fields (Theorem 3 in [30])

$$P(\xi \in V_\gamma(K)) = 1 \quad or \quad P(\xi \in V_\gamma(K)) = 0, \tag{4.45}$$

and the probability is 1 when

$$\sup_n \operatorname{tr}(C(\chi_n)K^{-1}(\chi_n)) < \infty, \tag{4.46}$$

where χ_n is a countably dense subset of points on γ. We can see that when $C = K$ the matrix product is an "infinite-dimensional" identity matrix and the supremum is infinite, so the realizations of a Gaussian random field are never in the RKHS

corresponding to its covariance.

The condition in (4.46) can be formulated in terms of the operator G defined in the previous section: (4.46) is equivalent to G being a trace class operator. To obtain a form for the trace of G first we need to select an orthonormal basis for $V_\gamma(K)$. Recall that in the previous section we constructed an orthonormal basis for $L^2(\gamma)$ using the eigenfuctions $\psi_1, \psi_2, ...$ associated with the kernel operator K. One can verify that the inner product on $V_\gamma(K)$ can be written in terms of inner products on $L^2(\gamma)$:

$$\langle v_1, v_2 \rangle_{V_\gamma} = \langle K^{-1} v_1, v_2 \rangle_{L^2_\gamma} = \sum_{k=1}^{\infty} \frac{\langle v_1, \psi_k \rangle_{L^2_\gamma} \langle v_2, \psi_k \rangle_{L^2_\gamma}}{\lambda_k}, \tag{4.47}$$

and thus we can show that $\{\sqrt{\lambda_k}\psi_k\}_{k=0}^{\infty}$ form an orthonormal basis for $V_\gamma(K)$. We have

$$\langle \sqrt{\lambda_i}\psi_i, \sqrt{\lambda_j}\psi_j \rangle_{V_\gamma(K)} = \sum_{k=1}^{\infty} \frac{\langle \sqrt{\lambda_i}\psi_i, \psi_k \rangle_{L^2(\gamma)} \langle \sqrt{\lambda_j}\psi_j, \psi_k \rangle_{L^2(\gamma)}}{\lambda_k}. \tag{4.48}$$

Since ψ_k's are orthonormal on $L^2(\gamma)$, the only nonzero terms in the above sum are the ones for which $i = j = k$, and then $\langle \sqrt{\lambda_i}\psi_i, \sqrt{\lambda_i}\psi_i \rangle_V = 1$. We can define the trace of G as

$$\begin{aligned} tr(G) &= \sum_{i=1}^{\infty} \langle G\sqrt{\lambda_i}\psi_i, \sqrt{\lambda_i}\psi_i \rangle_{V_\gamma(K)} = \sum_{i=1}^{\infty} \lambda_i \langle G\psi_i, \psi_i \rangle_{V_\gamma(K)} = \\ &= \sum_{i=1}^{\infty} \lambda_i \sum_{k=1}^{\infty} \frac{\langle G\psi_i, \psi_k \rangle_{L^2_\gamma} \langle \psi_i, \psi_k \rangle_{L^2_\gamma}}{\lambda_k}. \end{aligned} \tag{4.49}$$

The only nonzero term in the second sum is when $i = k$, so we conclude

$$tr(G) = \sum_{i=1}^{\infty} \langle G\psi_i, \psi_i \rangle_{L^2(\gamma)}. \tag{4.50}$$

We see that the trace of G as an operator on $V_\gamma(K)$ is the same as the trace of G as an operator on $L^2(\gamma)$.

Another formulation can be obtained from the following relationship between the norms

$$\langle Gf, g \rangle_{V_\gamma(C)} = \langle f, g \rangle_{V_\gamma(K)} \tag{4.51}$$

Similarly to K, we can define an operator C, and an inner product associated with it:

$$\langle f, g \rangle_{L^2(\gamma)} = \langle Cf, g \rangle_{V_\gamma(C)} \tag{4.52}$$

We can obtain several alternative representations of the trace:

$$tr(G) = \sum_{i=1}^{\infty} \langle G\psi_i, \psi_i \rangle_{L^2(\gamma)} = \sum_{i=1}^{\infty} \langle G\psi_i, C\psi_i \rangle_{V_\gamma(C)} = \sum_{i=1}^{\infty} \langle \psi_i, C\psi_i \rangle_{V_\gamma(K)}. \tag{4.53}$$

Let $\{\rho_i\}_{i=1}^{\infty}$ be the eigenvalues associated with C. In the special case when the

eigenfunctions of C and K coincide, the trace reduces to

$$tr(G) = \sum_{i=1}^{\infty} \langle \psi_i, C\psi_i \rangle_{V_\gamma(K)} = \sum_{i=1}^{\infty} \langle \psi_i, \rho_i \psi_i \rangle_{V_\gamma(K)} = \sum_{i=1}^{\infty} \frac{\rho_i}{\lambda_i}. \tag{4.54}$$

Example 1 (Circle domain): Let γ be S^1 embedded in $\mathbb{R}^2$. Then the eigenfunctions of the Gaussian kernel are the spherical harmonics and do not depend on the width of the kernel: i.e., they are the same for K and C. The corresponding eigenvalues can be derived from the Funk-Hecke formula [68]:

$$\lambda_k = e^{2/\sigma^2} I_k(2/\sigma^2), \tag{4.55}$$

where I_k are the modified Bessel functions of the first kind. For $\sigma_0 < \sigma_1$, the sum

$$tr(L) = \sum_{k=1}^{\infty} \frac{e^{2/\sigma_1^2} I_k(2/\sigma_1^2)}{e^{2/\sigma_0^2} I_k(2/\sigma_0^2)} \tag{4.56}$$

converges. If we substitute the definitions of the Bessel functions we obtain

$$tr(L) = (e^{2/\sigma_1^2 - 2/\sigma_0^2}) \sum_{k=0}^{\infty} \frac{\frac{1}{\sigma_1^{2k}} \sum_{l=0}^{\infty} \frac{(1/\sigma_1)^{2l}}{l!(k+l)!}}{\frac{1}{\sigma_0^{2k}} \sum_{l=0}^{\infty} \frac{(1/\sigma_0)^{2l}}{l!(k+l)!}} = (e^{2/\sigma_1^2 - 2/\sigma_0^2}) \sum_{k=0}^{\infty} \left(\frac{\sigma_0}{\sigma_1}\right)^{2k} \frac{\sum_{l=0}^{\infty} \frac{(1/\sigma_1)^{2l}}{l!(k+l)!}}{\sum_{l=0}^{\infty} \frac{(1/\sigma_0)^{2l}}{l!(k+l)!}}. \tag{4.57}$$

Since $\sigma_0 < \sigma_1$,

$$tr(L) \leq (e^{2/\sigma_1^2 - 2/\sigma_0^2}) \sum_{k=0}^{\infty} \left(\frac{\sigma_0}{\sigma_1}\right)^{2k} \frac{\sum_{l=0}^{\infty} \frac{(1/\sigma_0)^{2l}}{l!(k+l)!}}{\sum_{l=0}^{\infty} \frac{(1/\sigma_0)^{2l}}{l!(k+l)!}} = (e^{2/\sigma_1^2 - 2/\sigma_0^2}) \sum_{k=0}^{\infty} \left(\frac{\sigma_0}{\sigma_1}\right)^{2k} < \infty. \tag{4.58}$$

Thus the realizations of a Gaussian random field along the circle with covariance C belong to $V_\gamma(K)$.

Example 2 (Euclidean plane): Let's first consider the case when the domain of restriction is $\mathbb{R}^2$. The form of the operator G can be easily identified on $\mathbb{R}^2$, since we can resort to properties of Fourier transforms. It can be written as an integral transform on $\mathbb{R}^2$

$$[Gf](x) = \int_{\mathbb{R}^2} G(x,y) f(y) dy, \tag{4.59}$$

where $G(x,y) = e^{-\frac{\|x-y\|^2}{2(\sigma_0^2 - \sigma_1^2)}}$.

We note that

$$\int_{\mathbb{R}^2} \int_{\mathbb{R}^2} G(x,y)^2 dx dy = 1, \tag{4.60}$$

so $G(x,y)$ is a Hilbert-Schmidt kernel and the corresponding operator is compact.

Applying Mercer's theorem again, we obtain

$$tr(G) = \int_{\mathbb{R}^2} G(x,x)dx = \infty. \tag{4.61}$$

We conclude that G is not trace class and realizations of random vector fields with Gaussian covariance are never in an RKHS with a Gaussian kernel (no matter how the kernel widths are chosen).

Example 3 (Line segment domain): Let the domain be $I = (-1, 1)$. For simplicity we will use $\sigma = 1$. First consider the Taylor expansion of $K(\cdot, x)$ at 0 ($x \in I$):

$$K(\cdot, x) = \sum_{k=0}^{\infty} \frac{\partial^k K(\cdot, x)}{\partial x^k}\Big|_{x=0} \frac{x^k}{k!}. \tag{4.62}$$

We would like to know if this series converges in the V-norm.

First we note that the derivatives of the kernel belong to V, and their norm has an explicit form. By the reproducing property we have for any function f in V:

$$f(x) = \langle f(\cdot), K(\cdot, x)\rangle_V. \tag{4.63}$$

Therefore,

$$\begin{aligned} \frac{\partial f(x)}{\partial x} &= \lim_{h\to 0} \frac{f(x+h) - f(x)}{h} = \lim_{h\to 0} \frac{\langle f(\cdot), K(\cdot, x+h) - K(\cdot, x)\rangle}{h} \\ &= \left\langle f(\cdot), \frac{\partial K(\cdot, x)}{\partial x} \right\rangle_V . \end{aligned} \tag{4.64}$$

Repeating the same argument we have that

$$\frac{\partial^k f(x)}{\partial x^k} = \left\langle f(\cdot), \frac{\partial^k K(\cdot, x)}{\partial x^k} \right\rangle_V . \tag{4.65}$$

Let's take $f(x) = \frac{\partial^k K(x,y)}{\partial y^k}$ and obtain

$$\frac{\partial^k}{\partial x^k}\left(\frac{\partial^k K(x,y)}{\partial y^k}\right) = \left\langle \frac{\partial^k K(\cdot, y)}{\partial y^k}, \frac{\partial^k K(\cdot, x)}{\partial x^k} \right\rangle_V \tag{4.66}$$

Since the kernel is translation invariant,

$$\frac{\partial^k}{\partial x^k}\left(\frac{\partial^k K(x,y)}{\partial y^k}\right) = (-1)^k \frac{\partial^{2k} K(x,y)}{\partial x^{2k}}, \tag{4.67}$$

and we conclude that

$$\begin{aligned} \left\| \frac{\partial^k K(\cdot, x)}{\partial x^k} \right\|_V^2 &= (-1)^k \frac{\partial^{2k} K(x,y)}{\partial x^{2k}}\Big|_{y=x} = H_{2k}(x-y)K(x,y)|_{y=x} = \\ &= (-1)^k (-1)^k 1 \times 3 \times ... \times (2k-1) = \\ &= 1 \times 3 \times ... \times (2k-1), \end{aligned} \tag{4.68}$$

where $H_k(x)$ is the Hermite polynomial of order k.

We can use the above expression to show that the series in converges absolutely:

$$\begin{aligned}\sum_{k=0}^{\infty}\left\|\frac{\partial^k K(\cdot,x)}{\partial x^k}\Big|_{x=0}\frac{x^k}{k!}\right\|_V &= \sum_{k=0}^{\infty}\left\|\frac{\partial^k K(\cdot,x)}{\partial x^k}\Big|_{x=0}\right\|_V\frac{|x|^k}{k!} = \\ &= \sum_{k=0}^{\infty}[1\times 3\times ...\times(2k+1)]^{\frac{1}{2}}\frac{|x|^k}{k!}. \qquad (4.69)\end{aligned}$$

The ratio of two consecutive terms of this series is $\sqrt{2k+1}\frac{|x|}{k}$ which goes to zero as k goes to infinity, and thus the series converges. The absolute convergence implies the convergence in the V-norm. Since it also converges pointwise to $K(x,y)$, we have established the Taylor series expansion of $K(\cdot,x)$ in V.

Since we have picked $0\in I$ we see that all functions $\frac{\partial^k K(\cdot,x)}{\partial x^k}\big|_{x=0}$ belong to V_I, so $K(\cdot,x)$ can be written as a limit of partial sums in V_I and hence belongs to V_I itself. Since V is generated by $K(\cdot,x)$ for $x\in\mathbb{R}$, we conclude that $V\equiv V_I$. Thus the operator $G:V_K(I)\to V_C(I)$ is equivalent to $G:V_K(\mathbb{R})\to V_C(\mathbb{R})$ and it is not trace class.

This result can be extended to justify that the RKHS restricted to an arbitrary flat segment in $\mathbb{R}^2$ is equivalent to the RKHS over the whole line tangent to this segment. Thus when γ contains any flat region, the realizations of the random vector fields over γ do not belong to an RKHS with a Gaussian kernel.

So far we have seen some examples in which the realization of the random field belong to a RKHS (the circle), and some examples in which it does not (line segment, $\mathbb{R}^2$). What can be said about general contours? Clearly, these contours cannot contain

a flat region. Thus a minimum requirement is that the tangent to the curve (if it exists) should be nonzero at every point on the contour.

We conjecture that the realizations of the random field belong to the RKHS when the domain is an analytic curve, i.e. the parameterization $\gamma(t)$ is an analytic function with respect to t.

Chapter 5

Parameter Estimation in Diffusions on the Space of Shapes

5.1 Introduction

We address the problem of learning the dynamics of shapes from a sequence of observations. Our goal is to build algorithms which capture intrinsic features of the shape changes and which can be used in a variety of applications: tracking, classification, regression. By building more informative prior distributions for these complex processes we can address a wide variety of statistical problems with more precision and less computation. Once we have learned the underlying models we can incorporate them in a filtering algorithm and estimate the positions of the shapes when direct observations are not available: for example, when we are given only a sequence of

images without segmented shapes.

We split the task in two stages: first we need to construct appropriate parametric models for the evolution of the shapes; next, we would like to estimate the missing model parameters from a sequence observations. We select to model the evolution of the shapes with diffusion processes which on one hand are flexible enough to describe a wide variety of shape deformations, and on the other hand posses some general theory and well-established properties. As the structure on the space of shapes is non-Euclidean, we need to resort to working with diffusions on manifolds.

5.1.1 Related work

The subject of studying diffusions of shapes dates back to the work of Kendall [54], 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. [9] 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 are interested in shapes which do not change their topology so their deformations can be appropriately described by stochastic flows of diffeomorphisms [57]. Such processes have been studied in the context of images in [19], in the context of landmarks in [93], and extended to the infinite-dimensional spaces of curves and surfaces in [94].

5.1.2 Contribution

While these previous models concentrate on modeling general random shape perturbations and their properties, we focus on the tasks of constructing more informative parametric deformations for shapes (here 2D contours) and estimating their underlying parameters from a discrete sequence of observations.

We extend the model proposed in Section 3.2.1 and [83], which already includes a shape-based noise term, by introducing additional drift terms describing various shape motions. In the process we formally define a diffusion process on the landmark manifold. We justify that the selected noise model yields a well-posed process on the manifold and that it coincides with the sub-Riemannian formulation of Brownian motion on this space. We provide a procedure for simulating diffusion sample paths, derive explicit formulas for the likelihood-ratio estimates of their drift parameters from the observed shapes, and demonstrate their numerical performance when true parameters are known.

5.1.3 Organization

We begin with introducing in Section 5.2.1 several different ways for constructing diffusions on manifolds. Some of those are better suited for numerical simulation, others are easier to study analytically or provide more intuitive interpretation. Next in Section 5.3, 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 5.4 we introduce several drift models and show sequences of shapes they can yield in Section 5.5. In Section 5.6 we address the task of estimating the missing diffusion parameters. Finally, in Section 5.7 we discuss the properties of the solutions of the proposed diffusion stochastic differential equations.

5.2 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 contours $\mathcal{M}$ forms a $2m$-dimensional manifold as described in Section 2.5.1.

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

$$d\chi_t = A(\chi_t, \theta)dt + B(\chi_t)dW_t, \tag{5.1}$$

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 a to 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.

5.2.1 Diffusions on manifolds

There are several approaches for formulating diffusion processes (and their corresponding SDE's) on manifolds. Here we concentrate on the ones relevant to our work; for a more extensive treatment of the topic one can refer to the rich literature in [33, 48, 15]. 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 $T_{X_t}\mathcal{M}$ $(n \leq d)$ at each X_t, and define the stochastic differential equation on M as

$$dX_t = A(X_t)dt + B(X) \circ dW_t, \tag{5.2}$$

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 (5.2) 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

$$dX^i(t) = a^i(X_t)dt + \sum_{k=1}^{n} b_k^i(X_t) \circ dw_k(t), \quad i = 1, ..., d, \tag{5.3}$$

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

$$d\varphi^i(X_t) = \sum_{j=1}^{n} \partial_j \varphi^i a^j(\varphi(X_t)) + \sum_{j=1}^{n}\sum_{k=1}^{n} \partial_j \varphi^i b_k^j(\varphi(X_t)) \circ dw_k(t). \tag{5.4}$$

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

$$dX_t = A(X_t)dt + \sum_{k=1}^{n} B_k(X_s) \circ dw_k(s), \tag{5.5}$$

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

$$f(X_t) - f(X_0) = \int_0^t Af(X_s)ds + \int_0^t \sum_{k=1}^{n} B_k f(X_s) \circ dw_k(s), \tag{5.6}$$

or equivalently

$$df(X_t) = Af(X_t)dt + \sum_{k=1}^{n} B_k f(X_t) \circ dw_k(t). \tag{5.7}$$

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 $E_1(X), ..., E_d(X)$ on M. Then the local representation becomes global.

Fortunately this is true in for $M = \mathcal{M}$. 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 m smooth independent 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

$$d\chi_t = \sum_{k=1}^{2m} \alpha_k(\chi_t)E_k(\chi_t)dt + \sum_{k=1}^{2m} E_k(\chi_t) \circ dw_k(t). \tag{5.8}$$

Itô SDE's on Manifolds. In a given coordinate chart we can define the following Itô equation [50]:

$$d\chi^i(t) = \hat{a}^i(X_t)dt + \sum_{k=1}^{n} b_k^i(X_t) \cdot dw_k(t), \quad i = 1, ..., 2m, \tag{5.9}$$

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:

$$d\varphi^i(X_t) = \sum_{j=1}^{d} \partial_j\varphi^i \hat{a}^j(X_t)dt + \frac{1}{2}\sum_{j=1}^{d}\sum_{k=1}^{n}\sum_{l=1}^{n} \partial_{kl}\varphi^i b_k^j(X_t)b_l^j(X_t)dt + \sum_{j=1}^{d}\sum_{k=1}^{n} \partial_j\varphi^i b_k^j(X_t) \cdot dw_k(t), \tag{5.10}$$

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

$$d\chi^i(t) = \left[\hat{a}^i(X_t) - \frac{1}{2}\sum_{j=1}^{d}\sum_{k=1}^{n} b_k^j(X_t)\partial_j b_k^i(X_t)\right]dt + \sum_{k=1}^{n} b_k^i(X_t) \circ dw_k(t), \quad i = 1, ..., d. \tag{5.11}$$

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. A global definition of Itô equations can be given by introducing a special bundle (Itô bundle [15, 41]), and then Itô equations can be defined as its sections.

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[13], Belopolskaya-Daletsky[15] forms). Let $\exp_X : T_X\mathcal{M} \to \mathcal{M}$ be the Riemannian exponential map on $\mathcal{M}$ and consider the equation

$$dX_t = \exp_{X_t}(A(X_t)dt + B(X_t)dW_t), \tag{5.12}$$

where the *forward stochastic differential*

$$A(X_t)dt + B(X_t)dW_t \tag{5.13}$$

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

$$u(t+s) = \int_t^{t+s} \tilde{A}(u_\tau)d\tau + \int_t^{t+s} \tilde{B}(u_\tau)dw_\tau, \tag{5.14}$$

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

Definition 5.1. (p.153 7.28 [41]) *A solution of (5.12) is a process X_t, for which at every X_t there exists a neighborhood in which X_{t+s} (for $s \geq 0$ such that X_{t+s} is in the neighborhood) coincides with a process from the class* $\exp_{X_t}(A(X_t)dt + B(X_t)dW_t)$ *a.s..*

Let's consider the Taylor series expansion of the exponential map. For any curve $X(t)$ on $\mathcal{M}$ in a local chart we have:

$$X(t) = X(0) + t\dot{X}(0) + \frac{1}{2}t^2\ddot{X}(0) + o(t^2). \tag{5.15}$$

Thus

$$\exp_X(tv) \quad = \quad X + tv - \frac{1}{2}t^2\mathbf{\Gamma}_X(v,v) + o(t^2), \tag{5.16}$$

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 $\mathcal{M}$. Using this expansion one can obtain a local chart formulation of (5.12):

$$dX_t = A(X_t)dt - \frac{1}{2}\sum_{k=1}^{d}\Gamma_{ij}^k \sum_{l=1}^{n} b_l^i(X_t)b_l^j(X_t)\frac{\partial}{\partial x^k}dt + \sum_{k=1}^{n} b_k(X_t)\cdot dw_k(t). \tag{5.17}$$

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

$$\hat{a}_k = a_k - \frac{1}{2}\Gamma_{ij}^k \sum_{l=1}^{n} b_l^i(X_t)b_l^j(X_t). \tag{5.18}$$

Diffusions as a Limit of a Random Walk on a Manifold. Intuitively, we would want a diffusion to be a limit of small steps on the manifold in a given direction with noise added to them. Baxendale [13] introduces an approach closest to this idea. First we recall the definition of a Wiener process on an infinite-dimensional space.

Definition 5.2. (Wiener process on an infinite-dimensional space) *A process W_t on a separable Fréchet space E is called a Wiener process on E generated by a Gaussian measure μ, if it satisfies the following properties:*

(a) it has continuous sample paths

(b) it has independent increments

(c) the distribution of $W_{t+s} - W_t$ *is independent of* s

(d) $W_0 = 0$

(e) the distribution of W_1 *is* μ*.*

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

$$X^\pi_{t_{j+1}} = \exp_{X^\pi_{t_j}}(\Delta_j U_t(X_{t_j})), \quad j = 0, ..., N-1. \tag{5.19}$$

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

$$dX_t = dU_t. \tag{5.20}$$

Finally, this approach provides us with a meaningful discretization scheme for small values of dt

$$X_{t+dt} = \exp_{X_t}(A(X_t)dt + B(X_t)W_t), \tag{5.21}$$

where W_t is a Wiener process on $\mathbb{R}^n$, or equivalently

$$X_{t+dt} = \exp_{X_t}\left(A(X_t)dt + \sum_{k=1}^{n} B_k(X_t)\varepsilon_k(t)\sqrt{dt}\right), \tag{5.22}$$

where $\varepsilon_k(t)$'s are independent standard normally distributed r.v.'s.

Since in our setting the exponential map can be numerically computed, defining diffusion processes through it is preferable for simulating diffusion paths and eliminates the need to resort to Stratonovich equations or computing correction terms.

Diffusions through Infinitesimal Generators. We introduce here one more approach to defining diffusions which facilitates the interpretation of the properties of the processes we construct. In the second part of his work on stochastic differential equations on manifolds [51] Itô discusses the topic of building a diffusion process on a manifold whose infinitesimal generator coincides with a given second order elliptic operator. The infinitesimal generator of a Markov process is an operator L acting on the space of compactly supported twice-differentiable functions on $\mathcal{M}$ in the following way

$$Lf(x) = \lim_{h\to 0}\frac{1}{h}[\mathbb{E}[f(X_h)|X(0)=x)] - f(x)] = \frac{\partial}{\partial h}\mathbb{E}[f(X_h)|X(0)=x]_{|_{h=0}}. \tag{5.23}$$

If we denote the transition semigroup of the process as

$$[P_h f](x) = \mathbb{E}[f(X_h)|X(0)=x)], \tag{5.24}$$

we have $Lf = \partial_h P_h f_{|_{h=0}}$ and $LP_t f = \partial_h P_{t+h} f_{|h=0}$. Applying Dynkin's formula we obtain:

$$P_t f(X_t) - P_0 f(X_0) = \int_0^t LP_\tau f d\tau \tag{5.25}$$

$$\mathbb{E}_x f(X_t) - f(x) = \int_0^t L\mathbb{E}_x f(X_\tau) d\tau \tag{5.26}$$

$$\mathbb{E}_x[f(X_t) - f(x) - \int_0^t Lf(X_\tau) d\tau] = 0, \tag{5.27}$$

and we conclude that $f(X_t) - f(x) - \int_0^t Lf(X_\tau) d\tau$ is a martingale. We can alternatively take this result as a definition of the generator L of the Markov process.

The question of existence of a process for which (5.27) holds is called the Martingale Problem [85, 86], and is a coordinate-free approach of studying properties of SDE's on manifolds. The problem is well studied in the case of diffusion processes. Let's consider the Stratonovich differential equation (5.2)

$$df(X_t) = (B_k f)(X_t) \circ dW_t^k + (Af)(X_t) dt \tag{5.28}$$

$$df(X_t) = (B_k f)(X_t) \cdot dW_t^k + \frac{1}{2} d(B_k f)(X_t) \cdot dW_t^k + (Af)(X_t) dt \tag{5.29}$$

$$df(X_t) = (B_k f)(X_t) \cdot dW_t^k + \frac{1}{2} B_j (B_k f)(X_t) dt + (Af)(X_t) dt. \tag{5.30}$$

Let $L = \frac{1}{2}B^2 + A$. Then we have

$$f(X_t) - f(X_0) = \int_0^t B_k f(X_\tau) \cdot dW_\tau^k + \int_0^t L(f(X_\tau))d\tau. \tag{5.31}$$

We observe that the first term on the right-hand side is an Itô integral with respect to a Brownian motion so it is a local martingale. Hence, $f(X_t) - f(X_0) - \frac{1}{2}\int_0^t L(f(X_\tau)d\tau$ is a local martingale as well, and we conclude that L is the generator of the diffusion process. Its form in local coordinates is:

$$\begin{aligned} L &= \frac{1}{2}\sum_{k=1}^n B_k^2 + A = \frac{1}{2}\sum_{k=1}^n b_k^i(X_t)\frac{\partial}{\partial x_i}\left(b_k^j(X_t)\frac{\partial}{\partial x_j}\right) + a^i\frac{\partial}{\partial x_i} = \\ &= \frac{1}{2}\sum_{k=1}^n b_k^i(X_t)b_k^j(X_t)\frac{\partial}{\partial x_i}\frac{\partial}{\partial x_j} + \frac{1}{2}\sum_{k=1}^n b_k^i(X_t)\frac{\partial}{\partial x_i}b_k^j(X_t)\frac{\partial}{\partial x_j} + a^i\frac{\partial}{\partial x_i} = \\ &= \frac{1}{2}B_{ij}\frac{\partial}{\partial x_i}\frac{\partial}{\partial x_j} + \hat{a}^i\frac{\partial}{\partial x_i}, \end{aligned} \tag{5.32}$$

where $B_{ij} = \sum_{k=1}^n b_k^i(X_t)b_k^j(X_t)$ and $\hat{a}$ is the drift of the diffusion equation in Itô form (5.9).

Alternatively, taking any positive semidefinite matrix B_{ij} and a vector a we can define a second-order (semi-elliptic) operator L and construct a diffusion which has such a generator (under certain regularity conditions discussed in Section 5.7).

Let's define an operator in local coordinates:

$$(Lf)(X_t) = \sum_{j=1}^{n} a^j(\chi_t)\partial_j f(X_t) + \frac{1}{2}\sum_{jk} B_{jk}(X_t)\partial_{kj} f(X_t). \tag{5.33}$$

The diffusion process generated by it satisfies

$$\lim_{h\to 0} \mathbb{E}(X_i(t+h)) = a_i(X(t)), \tag{5.34}$$

$$\lim_{h\to 0} \mathbb{E}[(X_i(t+h) - X_i(t))(X_j(t+h) - X_j(t))|X(s) : 0 \le s \le t] = B_{ij}(X(t)), \tag{5.35}$$

and the vector a and the matrix B are defined as the *infinitesimal mean* and the *infinitesimal covariance* of the process. If we can find a smooth $b_k^i(X_t)$ such that

$$B_{ij}(X_t) = \sum_{k=1}^{n} b_k^i(X_t) b_k^j(X_t), \tag{5.36}$$

then the stochastic process takes the form:

$$dX^i(t) = a^i(X_t)dt + \sum_{k=1}^{n} b_k^i(X_t) \cdot dw_k(t), \quad i = 1, ..., d. \tag{5.37}$$

To fully determine a diffusion process on the shape manifold we need to define its infinitesimal mean and covariance and we discuss appropriate choices for them in the next two sections.

5.3 Noise models

In this section we discuss the choice of an infinitesimal covariance, which in turn determines the form of vector fields $B_1, ..., B_n$, and the covariance of the noise in the diffusion equations. To focus on the properties of the noise we will consider diffusions for which 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. Motivated by the consistency arguments in Section 3.2.1.2 we would like the correlation between two points to be

$$C(x, y) = K(x, \chi_n)K^{-1}(\chi_n, \chi_n)K(\chi_n, y), \tag{5.38}$$

where $K(x, y) = e^{-\frac{\|x-y\|^2}{2\sigma^2}}$. In this case the correlation between two points in χ_n is $K(x, y)$. Evaluating at each $x \in \chi_m$ the full diffusion matrix becomes:

$$C(\chi_m, \chi_m) = K(\chi_m, \chi_n)K(\chi_n, \chi_n)^{-1}K(\chi_n, \chi_m). \tag{5.39}$$

This choice, in addition to possessing the consistency property, is also driven by geometric motivation as it relates to the Brownian motion on the manifold. Before we address this connection, however, we first want to ensure this covariance gives rise to a well-defined process.

Well-posedness. Let's consider a process with zero infinitesimal mean and infinitesimal covariance as in (5.39). In order to define an associated SDE, we need to be able to write $C(\chi_m, \chi_m)$ as a product $C(\chi_m, \chi_m) = \Sigma(\chi_m)\Sigma(\chi_m)^T$, where $\Sigma(\chi_m)$ is

smooth in some appropriate sense to allow to define vector fields on $\mathcal{M}$. First we note that $K(\chi_n, \chi_n)$ is a positive definite matrix whenever all points in χ_m are distinct. Therefore, there exists a unique positive definite square root of $K(\chi_m, \chi_m)$ which we will denote by $K(\chi_m, \chi_m)^{1/2}$. We can then decompose the covariance as:

$$C(\chi_m, \chi_m) = (K(\chi_m, \chi_n)K(\chi_n, \chi_n)^{-1/2})(K(\chi_m, \chi_n)K(\chi_n, \chi_n)^{-1/2})^T \tag{5.40}$$

The entries of the matrices $K(\chi_m, \chi_n)$ and $K(\chi_n, \chi_n)$ are an analytic function of χ_m, so we can state that $K(\chi_m, \chi_m)$ is continuously differentiable as a function of χ_m (in the space of matrices with some matrix norm). Below we justify that the square root is continuously differentiable too.

Proposition 5.3. ***(Smoothness of matrix square roots)*** *Let $A \in \mathcal{S}_+^n$, where $\mathcal{S}_+^n$ denotes the space of n-dimensional real positive definite matrices. Consider the unique symmetric positive definite square root of A: $S = \sqrt{A}$ ($S^2 = A, S \in \mathcal{S}_+^n$). The function $f(A) = \sqrt{A}$ is continuously differentiable.*

Proof. We will employ the inverse function theorem. Let's define the map $g : \mathcal{S}_+^n \to \mathcal{S}_+^n$ satisfying $g(S) = S^2$. First we note that the existence of the symmetric positive definite square root justifies that g is onto while its uniqueness ensures g is 1-1, hence g is a bijection on $\mathcal{S}_+^n$.

Next we observe that g is a smooth map and we look at its derivative. We recall that S_+^n is a differentiable manifold. The tangent space at a point S on this manifold

is equivalent to the space of symmetric matrices. We consider a small perturbation of S in the direction of the tangent vector H: for t small enough $S + tH$ stays on S_+^n. Then

$$dg(S)H = \frac{d}{dt} g(S + tH)|_{t=0} = S^2 + tSH + tHS + t^2H^2|_{t=0} = SH + HS \quad (5.41)$$

We can show that dg has a full rank on $\mathcal{S}_+^n$, i.e. $SH + HS = 0$ implies $H = 0$ for any $S \in \mathcal{S}_+^n$. Let's take an eigenvalue-eigenvector pair for S: λ, v (clearly λ is positive). Then

$$SH + HS = 0 \Rightarrow SHv + HSv = 0 \Rightarrow SHv = -\lambda Sv \quad (5.42)$$

If $Hv \neq 0$ we have that λ and $-\lambda$ are both eigenvalues for S, and since S is positive definite we reach a contradiction, hence $Hv = 0$. Since this is true for any eigenvector of S, and we can select the eigenvectors of S to form a basis for R^n we show that H is equal to zero in this basis.

Since dg is full rank, by the inverse function theorem g is a local C^1-diffeomorphism at each point on S_+^n. As we also know that g is a bijection, we conclude that g is a global C^1-diffeomorphism, and that the square root function is continuously differentiable. □

We conclude that, since $K(\chi_n, \chi_n)^{1/2}$ is continuously differentiable and both matrix inversion and multiplication are smooth operations,

$\Sigma(\chi_m) = K(\chi_m, \chi_n)K(\chi_n, \chi_n)^{-1/2}$ is continuously differentiable as well, and the process (written in Belopolskaya-Daletsky form)

$$d\chi_t = \exp_{\chi_t}(\Sigma(\chi_t)dW_t) \tag{5.43}$$

is well-posed on the manifold.

In practice, we do not implement the exponential map directly. Instead, we use the Hamiltonian formulation of the geodesic equations and use the co-exponential map (2.32), evaluated at the initial momentum corresponding to the differential $\Sigma(\chi_t)dW_t$. When we expand Σ, the differential becomes $K(\chi_m, \chi_n)K(\chi_n, \chi_n)^{-1/2}dW_t$, and we realize that its covector formulation is simply $K(\chi_n, \chi_n)^{-1/2}dW_t$. So we have

$$d\chi_t = \exp^{\flat}_{\chi_t}\left(K(\chi_n(t), \chi_n(t))^{-1/2}dW_t\right). \tag{5.44}$$

Brownian Motion on a Riemannian manifold

We discuss here the most famous example of a diffusion – the Brownian motion and its generator – the Laplace-Beltrami operator. For that we require the manifold to possess a Riemannian structure. Let's denote the metric coefficients by g_{ij} and those of its inverse by g^{ij}. Assume that the vector fields $B_1, ..., B_m$ form an orthonormal frame with respect to this metric; this implies that in local coordinates: $b^i_k g_{ij} b^j_l = \delta_{k,l}$, and $b^i_k b^j_k = g^{ij}$. The matrix B_{ij} appearing in the infinitesimal generator becomes the

inverse of the metric tensor. The full form of the infinitesimal generator is

$$L = \frac{1}{2} g^{ij} \frac{\partial}{\partial x_i} \frac{\partial}{\partial x_j} + \frac{1}{2} \Gamma_{ij}^k g^{ij} \frac{\partial}{\partial x_k}. \tag{5.45}$$

Using the direct relationship between the Christoffel symbols and the metric

$$\Gamma_{ij}^k = \frac{1}{2} g^{kl} \left(\frac{\partial g_{li}}{\partial x^j} + \frac{\partial g_{lj}}{\partial x^i} - \frac{\partial g_{ij}}{\partial x^m} \right) \tag{5.46}$$

one can verify that the generator can also be written in the more familiar way

$$L = \frac{1}{2} g^{ij} \frac{\partial}{\partial x_i} \frac{\partial}{\partial x_j} + \frac{1}{2} |g|^{-1/2} \frac{\partial}{\partial x_i} (|g|^{1/2} g^{ij}) \frac{\partial}{\partial x_j} = \frac{1}{2} |g|^{-1/2} \frac{\partial}{\partial x_i} \left(|g|^{1/2} g^{ij} \frac{\partial}{\partial x_i} \right) = \frac{1}{2} \nabla \cdot \nabla, \tag{5.47}$$

which is the traditional definition of the Laplace-Beltrami operator as the divergence of the gradient.

We observe that the Laplace-Beltrami operator has an additional first order term (containing the Christoffel symbols) in contrast to the classical Laplacian which has only the second order term. Clearly they coincide when the manifold is flat and the Christoffel symbols are zero. On a general manifold one uses the exponential map to generate Brownian motion, which yields an extra correction term to the drift. So if we evaluate the exponential map at $A(X_t, t) = 0$, we obtain the following local

coordinate formulation of the SDE ($i = 1, ..., d$)

$$dX_i = \underbrace{-\frac{1}{2}\sum_{k=1}^{d}\Gamma_{ij}^{k}\sum_{l=1}^{d} b_l^i(X_t)b_l^j(X_t)dt}_{\textit{Riemannian correction drift}} + \underbrace{\sum_{k=1}^{d} b_k(X_t)dw_k(t)}_{\textit{Brownian motion}}. \tag{5.48}$$

Thus, in local coordinates the Brownian motion has a nonzero drift, and the infinitesimal generator has an associated first order term. Although we do not use this representation for sampling, the form of the equations in local coordinates will be useful later in Section 5.7.

Brownian Motion on a Sub-Riemannian Manifold

Note that when $n \leq d$ the $B_i(X)$'s form a basis only for an n-dimensional subspace of $T_X\mathcal{M}$ (they form a basis for the distribution $\mathcal{H}_X$). The notion of a metric does not exist on a sub-Riemannian manifold (the inner product is not defined for vectors not belonging to the distribution), therefore, the above definition of Brownian motion cannot be automatically extended to the sub-Riemannian case. Instead, we resort to the notion of a cometric [70] which corresponds to the inverse of the metric in Riemannian geometry. We first introduce the bundle map: $\tilde{\beta} : T^*\mathcal{M} \to T\mathcal{M}$ in the following way:

- $im(\tilde{\beta}_X) = \mathcal{H}_X$
- $p(v) = \langle \tilde{\beta}_X(p), v \rangle$ for $v \in \mathcal{H}_X$ and $p \in T_X^*M$.

Now we can define the cometric as the contravariant tensor $\beta : T^*\mathcal{M} \times T^*\mathcal{M} \to \mathbb{R}$

satisfying

$$\beta(p_1, p_2) = p_1(\tilde{\beta}(p_2)) = \langle \tilde{\beta}(p_1), \tilde{\beta}(p_2) \rangle, \quad \text{for } p_1, p_2 \in T^*\mathcal{M}. \tag{5.49}$$

It is clear that β is degenerate since $\tilde{\beta}$ is not onto.

In local coordinates, we can refer to the cometric as $g^{ij}(x)$ (and $g_{ij}(x)$ is not well defined). Formulas in Riemannian geometry which can be written in terms of the inverse of the Riemannian metric can be generalized to the sub-Riemannian case. We observe that the Laplace-Beltrami operator contains the Christoffel symbols, whose direct computation requires the derivatives of the Riemannian metric, and thus cannot be automatically generalized to sub-Riemannian manifolds. We define the raised Christoffel symbols (as introduced by [46], [84]):

$$\Gamma^{kpq} = \frac{1}{2}\left(g^{jp}\frac{\partial g^{kq}}{\partial x^j} + g^{jq}\frac{\partial g^{kp}}{\partial x^j} - g^{jk}\frac{\partial g^{pq}}{\partial x^j} \right). \tag{5.50}$$

The classical Christoffel symbols specify the covariant derivative and appear in the geodesic equations on a Riemannian manifold (2.22). Similarly, the raised Christoffel symbols appear in the Hamiltonian formulation of the geodesic equations on a sub-Riemannian manifold:

$$\ddot{x}^i(s) = \Gamma^{ijk}(x(s))p_j(s)p_k(s), \quad i = 1, ..., d \tag{5.51}$$

for some initial conditions on the covectors $p_j(0)$ for $j = 1, ..., d$. We recall that there is no one-to-one correspondence between covectors and horizontal vectors.

Differentiating both sides of the equality $g^{il}g_{lj} = \delta_{ij}$, we obtain:

$$\frac{\partial g^{il}}{\partial x} g_{lj} = -g^{il} \frac{g_{lj}}{\partial x}. \tag{5.52}$$

The following relationship between the classical and raised Christoffel symbols is deduced

$$\begin{aligned} g^{ij}\Gamma^k_{ij} &= g^{ij}g^{km}\left(\frac{\partial g_{mi}}{\partial x^j} + \frac{\partial g_{mj}}{\partial x^i} - \frac{\partial g_{ij}}{\partial x^m}\right) = & (5.53)\\ &= -g^{ij}\frac{\partial g^{km}}{\partial x^j} g_{mi} - g^{ij}\frac{\partial g^{km}}{\partial x^i} g_{mj} + g^{km}\frac{\partial g^{ij}}{\partial x^m} g_{ij} = & (5.54)\\ &= -g^{im}\frac{\partial g^{kj}}{\partial x^m} g_{ji} - g^{mj}\frac{\partial g^{ki}}{\partial x^m} g_{ij} + g^{km}\frac{\partial g^{ij}}{\partial x^m} g_{ij} = & (5.55)\\ &= -g_{ij}\left(g^{mi}\frac{\partial g^{kj}}{\partial x^m} + g^{mj}\frac{\partial g^{ki}}{\partial x^m} - g^{mk}\frac{\partial g^{ij}}{\partial x^m}\right) = -g_{ij}\Gamma^{ijk}. & (5.56) \end{aligned}$$

The third equality was obtained by renaming the indices of the first two terms, the next one – by using the symmetry of the Riemannian metric.

Therefore, the Laplace-Beltrami operator can be written in terms of the raised Christoffel symbols as:

$$L = \frac{1}{2}\left(g^{ij}\frac{\partial}{\partial x_i}\frac{\partial}{\partial x_j} - g_{ij}\Gamma^{ijk}\frac{\partial}{\partial x_k}\right). \tag{5.57}$$

This representation improves upon (5.45) as it does not require calcultion of derivatives of the metric, but still depends on g_{ij} which is not well defined on a sub-Riemannian manifold, so more intrinsic formulation should be sought.

On a Riemannian manifold the definition of the Laplace-Beltrami operator as the divergence of the gradient is possible due to the existence of the Riemannian volume form and this definition is intrinsic. On a sub-Riemannian manifold there is no natural way to define a volume form, hence, there does not exist a unique definition of the corresponding sub-Laplacian operator (the generator of the sub-Riemannian Brownian motion). There has been a lot of recent work toward providing an appropriate formulation for the sub-Laplacian: based on the Hausdorff volume [1], based on Popp's volume (introduced by Popp and discussed for the first time in [70]), etc [42, 43, 3].

Rather than settling on a choice of volume, in a recent work Gordina, et. al. [43] suggest choosing the sub-Laplacian operator so that the resulting process is a limit of a suitably defined random walk on the manifold. The proposed definition is based on a Riemannian metric which is compatible with the sub-Riemannian metric on the manifold. A metric g_{ij} on $\mathcal{M}$ is called compatible if when restricted to the horizontal distribution $\mathcal{H}$ coincides with the sub-Riemannian metric: $\langle v, w\rangle_{\mathcal{M}} = \langle v, w\rangle_{\mathcal{H}}$ for $v, w \in \mathcal{H}$. In our particular setting there is a natural choice for a Riemannian metric on the manifold: the one induced by the reproducing kernel evaluated at χ_m. The Riemannian metric provides a one-to-one correspondence between the tangent and

cotangent spaces of the manifold, thus given a vector in horizontal distribution we can identify a unique covector associated to it. Let $\Phi^g_{t+s}(x,v)$ be the Hamiltonian flow for g on M with an initial condition x and $p = g(v)$ (the map g here indicates the correspondence between v and p based on the metric). For our choice of a metric this corresponds to solving the exponential map as defined in (3.33). Gordina, et al. define the sub-Laplacian operator in the following way:

$$Lf(x) = \int_{\mathcal{S}_{\mathcal{H}_x}} \left\{ \frac{\partial}{\partial t}\frac{\partial}{\partial s} f(\Phi^g_{t+s}(x,v)) \Big|_{s=0,t=0} \mathbb{U}_x(dv) \right\}, \tag{5.58}$$

where the integral is calculated over the unit sphere in the horizontal distribution $\mathcal{S}_{\mathcal{H}_\S}$, Φ^g_t is the flow along a horizontal geodesic, and $\mathbb{U}_x$ is the rotationally invariant measure on the distribution. In local coordinates the generator takes the form

$$L = \frac{1}{m}\sum_{i,j=1}^{m} \left\{ \beta^{ij}\frac{\partial^2}{\partial x^i \partial x^j} - \sum_{k,q,r=1}^{m} \Gamma^{ijk} g_{iq}\beta^{qr} g_{rj}\frac{\partial}{\partial x^k} \right\}, \tag{5.59}$$

where $\beta^{ij} = \langle \beta(dx^i), \beta(dx^j)\rangle$. Clearly, on a Riemannian manifold, where $\beta^{ij} = g^{ij}$, this reduces to a scaled version of the Laplace-Beltrami operator (5.57). The corresponding SDE in local coordinates is

$$dX_i = \underbrace{-\frac{1}{2}\sum_{k=1}^{d}\Gamma^{ijk}\sum_{q,r=1}^{d} g_{iq}\beta^{qr}g_{rj}dt}_{\textit{sub-Riemannian correction drift}} + \underbrace{\frac{1}{2}\sum_{k=1}^{d} b_k dw_k(t)}_{\textit{Brownian motion}}, \tag{5.60}$$

where $\beta^{ij} = \sum_{l=1}^{n} b_l^i b_l^j$.

In our setting the matrix with entries β^{qr} takes the form:

$$\beta = \begin{bmatrix} \mathbf{K}(\chi_n)^{-1} & 0 \\ 0 & 0 \end{bmatrix}. \tag{5.61}$$

Thus $g_{iq}\beta^{qr}g_{rj}$ is the ij's entry of the matrix

$$\mathbf{K}(\chi_m) \begin{bmatrix} \mathbf{K}(\chi_n)^{-1} & 0 \\ 0 & 0 \end{bmatrix} \mathbf{K}(\chi_m) = \mathbf{K}(\chi_m, \chi_n)\mathbf{K}(\chi_n, \chi_n)^{-1}\mathbf{K}(\chi_n, \chi_m) = \mathbf{C}(\chi_m). \tag{5.62}$$

Further, $b_l^i(\chi_t)$ is the il'th entry of $\mathbf{\Sigma}(\chi_m) = \mathbf{K}(\chi_m, \chi_n)\mathbf{K}(\chi_n, \chi_n)^{-1/2}$.

Therefore, we conclude that the model we have proposed in (5.43) coincides with this definition of Brownian motion on sub-Riemannian manifold.

5.4 Drift models

In this section we discuss various models for the drift in the diffusion.

5.4.1 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 $\mathcal{M}$:

$$A(\chi_t, \theta) = \sum_{i=1}^{n} E_i(\chi_t)\theta_i. \tag{5.63}$$

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):

$$A(\chi_t, \boldsymbol{\theta}) = \sum_{i=1}^{n} K(\chi_t, x_i)\theta_i = \mathbf{K}(\chi_t)\boldsymbol{\theta}, \tag{5.64}$$

or an orthonormal basis as defined in Section 5.3

$$A(\chi_t, \theta) = \boldsymbol{\Sigma}(\chi_t)\theta. \tag{5.65}$$

As usual, 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 Belopolskaya-Daletskiy form

$$d\chi_t = \exp_{\chi_t}(\mathbf{K}(\chi_t)\boldsymbol{\theta}dt + \boldsymbol{\Sigma}(\chi_t)dW_t), \tag{5.66}$$

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

5.4.2 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

$$d\chi_t = \theta \nabla U(\chi_t) + B(\chi_t) dW_t, \tag{5.67}$$

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:

$$(dU|v)_\chi = \langle \nabla U, v \rangle_\chi, \tag{5.68}$$

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

$$\langle \nabla U, v \rangle_\chi = (\nabla U_\chi^{\mathcal{M}})^T \mathbf{K}(\chi, \chi)^{-1} \mathbf{v}_\chi, \tag{5.69}$$

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

$$(dU|v)_\chi = \nabla U_\chi^T \mathbf{v}_\chi. \tag{5.70}$$

The condition in (5.68) becomes

$$\nabla U_\chi^T \mathbf{v}_\chi = (\nabla U_\chi^{\mathcal{M}})^T \mathbf{K}(\chi, \chi)^{-1} \mathbf{v}_\chi, \tag{5.71}$$

hence the form of the Riemannian gradient is

$$\nabla U_\chi^{\mathcal{M}} = \mathbf{K}(\chi, \chi) \nabla U_\chi. \tag{5.72}$$

The evaluation at an individual point is

$$\nabla_{x_i}^{\mathcal{M}} U = \sum_{j=1}^{m} K(x_i, x_j) \nabla_{x_i} U. \tag{5.73}$$

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

$$(dU|v) = \langle \nabla^{\mathcal{H}} U, v \rangle_{\mathcal{H}}. \tag{5.74}$$

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 (5.74) the coefficients of the horizontal gradient need to satisfy

$$\nabla U^T_{\chi_m} \mathbf{K}(\chi_m, \chi_n)\boldsymbol{\beta} = \boldsymbol{\alpha}^T \mathbf{K}(\chi_n, \chi_n)\boldsymbol{\beta}, \tag{5.75}$$

i.e.

$$\boldsymbol{\alpha} = \mathbf{K}(\chi_n, \chi_n)^{-1} \mathbf{K}(\chi_n, \chi_m) \nabla_{\chi_m} U, \tag{5.76}$$

and therefore

$$\nabla^{\mathcal{H}}_{\chi_m} U = \mathbf{K}(\chi_m, \chi_n)\mathbf{K}(\chi_n, \chi_n)^{-1}\mathbf{K}(\chi_n, \chi_m)\nabla_{\chi_m} U. \tag{5.77}$$

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

5.4.2.1 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 $\mathbb{R}$:

$$dX_t = \theta(\mu - X_t) + dW_t, \quad \theta > 0. \tag{5.78}$$

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:

$$dX_t = \theta \nabla_{X_t} \left[-\frac{(X_t - \mu)^2}{2} \right] + dW_t. \tag{5.79}$$

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:

$$dist(\chi_t, \mu) = \min_{v:\, \chi_t = \exp_\mu(v)} \|v\| \tag{5.80}$$

The gradient of the distance reflects the rate of change of the minimizer of (5.80) with respect to a change in the shape χ_t, which is not a trivial problem.

Instead, we define U to be a function which simply 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 pre-determined size B_{χ_t} and B_μ which are nonzero in the interior of the corresponding contours. Then we let $U = |B_{\chi_t} - B_\mu|$, 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}):

$$U(\chi_t) = \int_{\mathbb{R}^2} |B_{\chi_t} - B_\mu| \propto \int_{\Omega_{\chi_t}} (\underbrace{|1 - B_\mu(x)| - |0 - B_\mu(x)|}_{F(x)}) dx. \tag{5.81}$$

Now we simply 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 5.4.2.3.

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

$$d\chi_t = -\theta \nabla^{\mathcal{M}} U(\chi_t)dt + B(\chi_t)dW_t. \tag{5.82}$$

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

$$d\chi_t = -\sum_{i=1}^{p} \theta_i \nabla^{\mathcal{M}} dist(\chi_t, \mu_i) + B(\chi_t)dW_t. \tag{5.83}$$

We call this a "regression drift".

5.4.2.2 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

$A(\chi_t, \theta) = \theta_1 \nabla^{\mathcal{M}} U_1 + \theta_2 \nabla^{\mathcal{M}} U_2$:

$$d\chi_t = -\frac{1}{2}\theta_1 \nabla^{\mathcal{M}} |L_{\chi_t} - L_\mu|^2 - \frac{1}{2}\theta_2 \nabla^{\mathcal{M}} |A_{\chi_t} - A_\mu|^2 + B(\chi_t) dW_t, \tag{5.84}$$

$$d\chi_t = -\theta_1 (L_{\chi_t} - L_\mu) \nabla^{\mathcal{M}} L(\chi_t) - \theta_2 (A_{\chi_t} - A_\mu) \nabla^{\mathcal{M}} A(\chi_t) + B(\chi_t) dW_t, \tag{5.85}$$

$$d\chi_t = -[(L_{\chi_t} - L_\mu) \nabla^{\mathcal{M}} L(\chi_t) \;\; (A_{\chi_t} - A_\mu) \nabla^{\mathcal{M}} A(\chi_t)] \begin{bmatrix} \theta_1 \\ \\ \theta_2 \end{bmatrix} dt + B(\chi_t) dW_t. \tag{5.86}$$

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

$$U(\chi_t) = \sum_{i=1}^{p} \theta_i dist(m^i(\chi_t), m^i_\mu)^2, \tag{5.87}$$

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

5.4.2.3 Discretized gradients

In this section we provide explicit calculations of the gradients appearing in the shape models described in the previous two sections and discuss some of their properties. We first obtain the Euclidean gradients and then Riemannian/sub-Riemannian gradients

can be obtained by equations (5.72) and (5.77).

Template mismatch:

The mismatch from the template is calculated according to

$$U = \int_{\Omega_{\chi_t}} F(x)dx, \tag{5.88}$$

where $F(x) = |1 - B_\mu(x)| - |B_\mu(x)| = 1 - 2B_\mu(x)$ (B_μ takes values 0 or 1). As this function is discontinuous at the boundary of the template, a smoothed version of B_μ should be used instead. Now that F is differentiable (and also bounded) the gradient of U can be obtained using the divergence theorem

$$\nabla U = \int_{\Omega_{\chi_t}} \nabla F(x)dx = \int_{\chi_t} F(x) \cdot \nu dx, \tag{5.89}$$

where ν is the outer pointing unit normal along the boundary of the shape.

We recall that χ_t represents the polygon determined by the points $x_1(t), ..., x_m(t) \in \chi_t$. As χ_t is piecewise smooth, we can write the integral as

$$\int_{X_t} F(x) \cdot \nu dx = \sum_{i=1}^{m} \int_{x_i x_{i+1}} F(x) \cdot \nu dx, \tag{5.90}$$

where $x_i x_{i+1}$ indicates the line segment from x_i to x_{i+1} and $x_{m+1} = x_1$. We assume that $F(x)$ is close to constant along the line segments (here we clearly introduce an error since we know that $F(x)$ depends on the binary image B_μ and can change values

rapidly along its boundary) and thus approximate

$$\nabla U \approx \sum_{i=1}^{m} F(x_i) N_i |x_i x_{i+1}|, \tag{5.91}$$

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. In our implementation we first reparameterize the contour by arc length, and then evaluate the function and the normal at the newly created points to simplify the integral computation.

Area:

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

$$A = \frac{1}{2} \sum_{i=1}^{m-1} \det(x_i, x_{i+1}). \tag{5.92}$$

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

$$\nabla_{x_i} A = \frac{1}{2} \begin{bmatrix} x_{i+1}^{(2)} - x_{i-1}^{(2)} \\ x_{i-1}^{(1)} - x_{i+1}^{(1)} \end{bmatrix}, \tag{5.93}$$

where the superscripts indicate the coordinates of the points.

Using the relationship between Euclidean and Riemannian gradients in (5.72) we

can derive the explicit form of the Riemannian area gradient:

$$\nabla^{\mathcal{M}}_{x_i} A = \frac{1}{2} \begin{bmatrix} \sum_{j=1}^{m-1} K(x_i, x_j)(x^{(2)}_{j+1} - x^{(2)}_{j-1}) \\ \sum_{j=1}^{m-1} K(x_i, x_j)(x^{(1)}_{i-1} - x^{(1)}_{i+1}) \end{bmatrix}. \tag{5.94}$$

Length:

We represent the length in the following way

$$L(\chi) = \sum_{i=1}^{m-1} \|x_i - x_{i+1}\|, \tag{5.95}$$

where $\|\cdot\|$ is the Euclidean norm. To obtain the derivative we consider a displacement $h_1, ..., h_m$ and calculate

$$\begin{aligned} dL(h) &= \frac{d}{dt} \sum_{i=1}^{m-1} \|x_i + th_i - x_{i+1} - th_{i+1}\|_{|t=0} \\ &= \sum_{i=1}^{m-1} \frac{(x_i - x_{i+1})^T (h_i - h_{i+1})}{\|x_i - x_{i+1}\|}. \end{aligned} \tag{5.96}$$

The length gradient is

$$\nabla L = \frac{x_i - x_{i+1}}{\|x_i - x_{i+1}\|} - \frac{x_{i-1} - x_i}{\|x_{i-1} - x_i\|}. \tag{5.97}$$

We make the important observation that the gradient of the length becomes ill-defined when points coincide. We introduce a better behaved shape descriptor with

similar properties in the next section.

Energy:

The energy of a contour γ is defined as

$$E(\gamma) = \int_{\gamma} \|\gamma'\|^2 ds. \tag{5.98}$$

Its polygonal approximation is

$$E(X) = m \sum_{i=1}^{m} \|x_i - x_{i+1}\|^2, \tag{5.99}$$

assuming that the coordinates of the set of landmarks are stored in X (which is an $n \times 2$ matrix). Let's look at its vector form. The squared length function is

$$E(X) = \|X - M_1 X\|^2, \tag{5.100}$$

where

$$M_1 = \begin{bmatrix} 0 & 1 & 0 & \dots \\ 0 & 0 & 1 & 0 \\ \vdots & \vdots & 0 & \ddots \\ 1 & 0 & \dots & 0 \end{bmatrix} \tag{5.101}$$

We define the following matrix

$$L = (I - M_1)^T (I - M_1), \tag{5.102}$$

so we have $E(X) = X^T L X$ and $\nabla E = LX$.

We first note that the energy of a contour is not invariant under reparameterization. Further, while the length of the discrete contour does not depend on the position of the contour in the plane, the discrete energy grows when the contour moves further from the origin. To have a function which is invariant of the position of the shape, we introduce X' as the coordinates of X after subtracting the center of mass:

$$X' = X - \mathbf{1}X/n, \tag{5.103}$$

where $\mathbf{1}$ is the matrix of ones. We see that $X' = M_2 X$, with

$$M_2 = I - \frac{1}{n}\mathbf{1}. \tag{5.104}$$

The modified energy is

$$E(X') = X'^T L X' = X^T M_2^T L M_2 X, \tag{5.105}$$

and the gradient is

$$\nabla E' = M_2^T L M_2 X. \tag{5.106}$$

5.5 Simulation of shape paths

We generate artificial observations from the suggested dynamical shape models by numerically integrating the corresponding SDEs.

The initial shape is a circle of radius 10. The diffusion is simulated up to time $T = 30$ and the time step is $dt = 0.05$ (except (d) $dt = 0.1$). The deformation kernel width is set to $\sigma = 10$. Below we display the sequences on a 3D plot in which the 3rd dimension corresponds to time. The black stripes correspond to the locations of the control points.

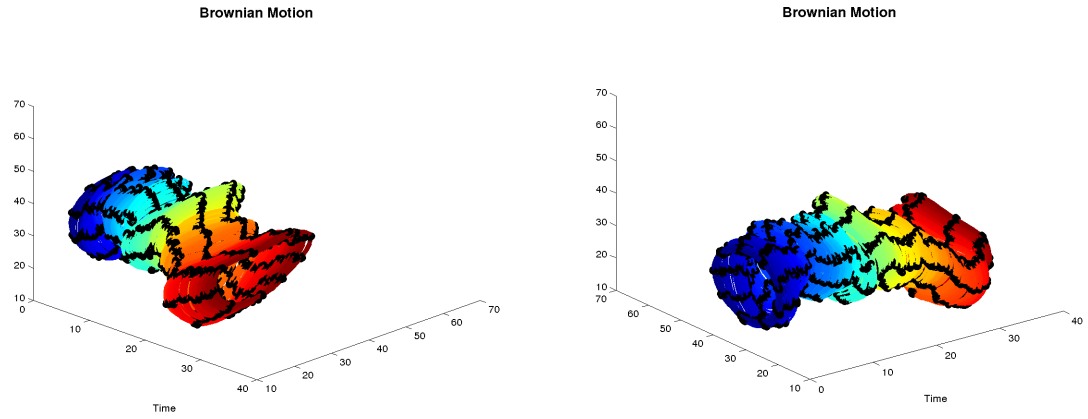


Figure 5.1: Driftless Diffusion

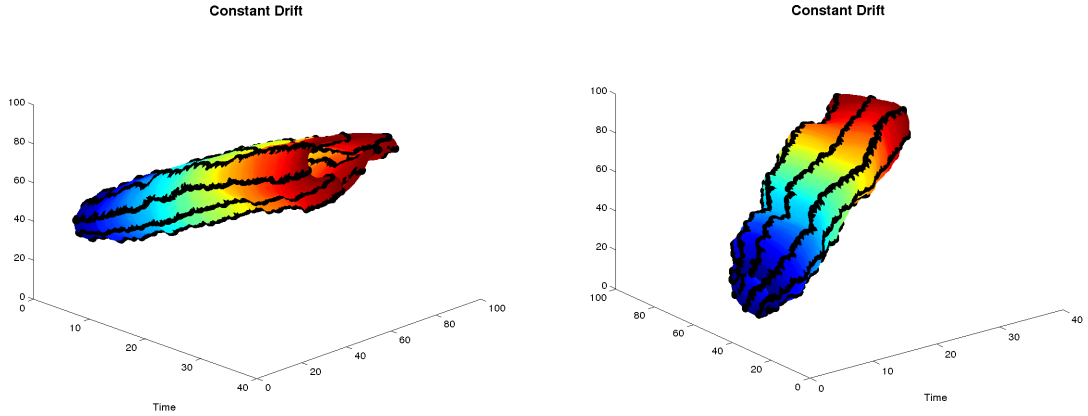


Figure 5.2: Constant Drift Diffusion

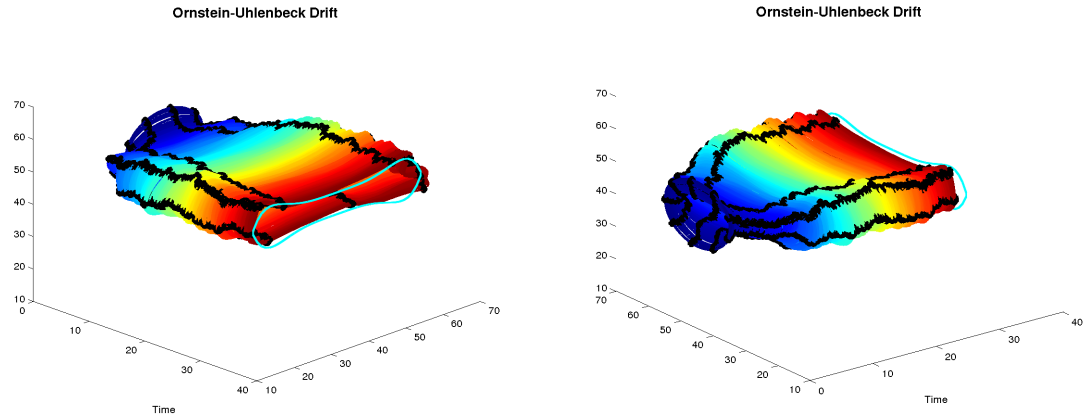


Figure 5.3: Mean-reverting Diffusion (initial shape is a circle, template shape is a dumbbell)

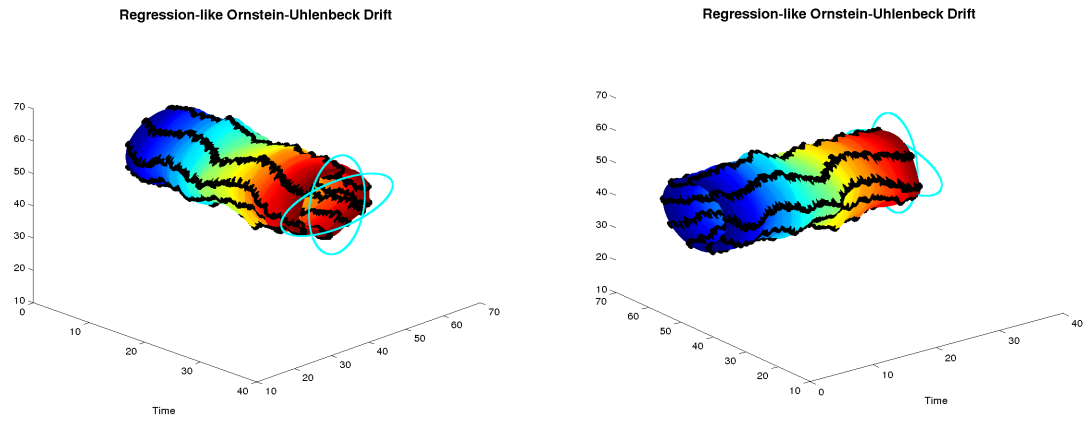


Figure 5.4: "Regression-like" Diffusion (with two template shapes: one vertical ellipse and one horizontal ellipse)

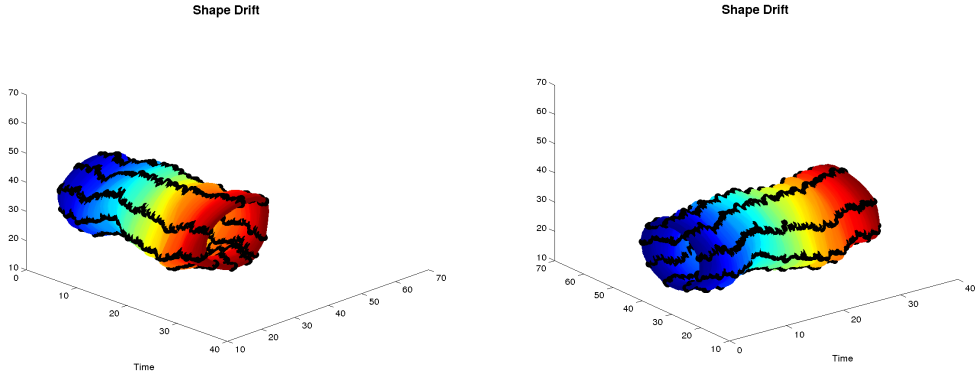


Figure 5.5: Shape Descriptor Drift Diffusion (with length and area terms)

5.6 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

$$d\chi_t = A(\chi_t, \theta)dt + B(\chi_t)dW_t, \quad \chi_0 = x, \tag{5.107}$$

(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 provide some background of the methodology for processes in Euclidean space.

5.6.1 Maximum likelihood estimation for processes on $\mathbb{R}^n$

In this section we discuss the topic of obtaining maximum likelihood estimates for drift parameters in diffusion processes on $\mathbb{R}^n$:

$$dX_t = A(X_t, \theta)dt + B(X_t)dW_t, \quad X_0 = x, \quad t \in [0, T]. \tag{5.108}$$

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 (5.108) 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

$$\mathbb{E}_\theta \exp\left(\int_0^T A(X_t, \theta)^T C(X_t)^{-1} A(X_t, \theta)dt\right) < \infty, \tag{5.109}$$

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

$$\frac{dP_\theta}{dP}(X) = \int_0^T A(X_t,\theta)^T C(X_t)^{-1} dX_t - \frac{1}{2}\int_0^T A(X_t,\theta)^T C(X_t)^{-1} A(X_t,\theta) dt. \tag{5.110}$$

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 [98]. 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. Specific properties of the estimator can be obtained in many situations. For example, when X_t is one-dimensional and the drift is $A(X_t,\theta) = \theta a(X_t)$, the form of the MLE estimate is:

$$\hat{\theta}(\xi) = \frac{\int_0^T a(X_t)X_t dt}{\int_0^T a^2(X_t)dt}, \tag{5.111}$$

and following results hold under additional regularity conditions ([61], Theorem 17.2, p. 202):

$$\text{Bias}(\theta, T) = \frac{d}{d\theta}\mathbb{E}_\theta\left(\int_0^T a^2(X_t)dt\right)^{-1}, \tag{5.112}$$

$$\text{MSE}(\theta, T) = \frac{d}{d\theta}\mathbb{E}_\theta\left(\int_0^T a^2(X_t)dt\right)^{-1} + \frac{d^2}{d\theta^2}\mathbb{E}_\theta\left(\int_0^T a^2(X_t)dt\right)^{-2}. \tag{5.113}$$

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

$$l_{h,N}(\theta, X) = \sum_{i=1}^{N} A(X_{t_{i-1}}, \theta)^T C(X_{t_{i-1}})^{-1}(X_{t_i} - X_{t_{i-1}}) - \\ -\frac{h}{2}\sum_{i=1}^{N} A(X_{t_{i-1}}, \theta)^T C(X_{t_{i-1}})^{-1} A(X_{t_{i-1}}, \theta). \tag{5.114}$$

which can be maximized to obtain an MLE estimate for θ. The MLE estimator has many desirable properties: under the condition that $Nh^3 \to 0$ (*moderately increasing design*), it can be shown that it is consistent, asymptotically normal and efficient [98].

5.6.2 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 Belopolskaya form of the Itô equation, we can write the process as

$$\chi_{t_{j+1}} = \exp_{\chi_{t_j}}\left(\Delta_j A(\chi_{t_j}, \theta) + \sqrt{\Delta_j}\sum_{i=1}^{n} B_i(\chi_{t_j})\varepsilon_i(t_j)\right), \quad j = 0, ..., N-1, \tag{5.115}$$

where B_i's represents a (not necessarily orthonormal) basis for $T_{\chi_{t_j}}M$, and ε_i's are independent standard normally distributed r.v.'s.

We are interested in $p_\theta(\chi_{0:N})$ and each $\chi_{t_{j+1}} = F(\chi_j, \varepsilon)$ and F represents the transformation in (5.115). Let's write for short $\chi_{0:N} = \Phi(\varepsilon)$, then

$$p_{\chi_{0:N},\theta}(x) = p_{\varepsilon,\theta}(\Phi^{-1}(x))|D\Phi^{-1}(x)|. \tag{5.116}$$

First let's compute $p_{\varepsilon,\theta}(\Phi(\chi_{0:N}))$. We can rewrite (5.115) in terms of ε_i's, by introducing the Riemannian logarithm map $\log_{\chi_{t_j}}(\chi_{t_{j+1}}) = \exp^{-1}_{\chi_{t_j}}(\chi_{t_{j+1}})$ (we assume that if the observations are not far apart the inverse of the exponential map is well defined):

$$\sum_{i=1}^{n} B_i(\chi_{t_j})\varepsilon_i(t_j) = \log_{\chi_{t_j}}(\chi_{t_{j+1}})/\sqrt{\Delta_j} - \sqrt{\Delta_j}A(\chi_{t_j},\theta). \tag{5.117}$$

The matrix form of the above equation is

$$\boldsymbol{\varepsilon}_{t_j} = \mathbf{B}(\chi_{t_j})^{-1}\left(\log_{\chi_{t_j}}(\chi_{t_{j+1}})/\sqrt{\Delta_j} - \sqrt{\Delta_j}A(\chi_{t_j},\theta)\right), \tag{5.118}$$

$$\Phi^{-1}(\chi_{0:N}) = \mathbf{B}(\chi_{t_j})^{-1}\left(\log_{\chi_{t_j}}(\chi_{t_{j+1}})/\sqrt{\Delta_j} - \sqrt{\Delta_j}A(\chi_{t_j},\theta)\right). \tag{5.119}$$

Since the $\boldsymbol{\varepsilon}_{t_j}$'s are normally distributed, their joint density given the observed

path is

$$p_{\varepsilon,\theta}(\Phi^{-1}(\chi_{0:N}))) \propto \prod_{j=0}^{N} e^{-\boldsymbol{\varepsilon}_{t_j}^T \boldsymbol{\varepsilon}_{t_j}/2} =$$

$$= \exp\Bigg(\sum_{k=1}^{N} - \log_{\chi_{t_j}}(\chi_{t_{j+1}})^T \mathbf{B}(\chi_{t_j})^{-T} \mathbf{B}(\chi_{t_j})^{-1} \log_{\chi_{t_j}}(\chi_{t_{j+1}})/2\Delta_j +$$

$$+ \log_{\chi_{t_j}}(\chi_{t_{j+1}})^T \mathbf{B}(\chi_{t_j})^{-T} \mathbf{B}(\chi_{t_j})^{-1} A(\chi_{t_j}, \theta) -$$

$$- \frac{\Delta_j}{2} A(\chi_{t_j}, \theta)^T \mathbf{B}(\chi_{t_j})^{-T} \mathbf{B}(\chi_{t_j})^{-1} A(\chi_{t_j}, \theta) \Bigg). \tag{5.120}$$

Next, we look at the form of the Jacobian:

$$|D\Phi^{-1}(x)| = |D\Phi(\Phi^{-1}(x))|^{-1}, \tag{5.121}$$

where

$$[D\Phi(\varepsilon)]_{j+1} = \partial_1 F(\chi_i, \varepsilon_i)\delta\chi_i + \partial_2 F(\chi_j, \varepsilon_j)\delta\varepsilon_j. \tag{5.122}$$

We observe that the Jacobian of Φ is triangular, and its determinant is

$$|D\Phi^{-1}(\chi_{0:N})| = \prod_{j=1}^{N} \det(\partial_2 F(\chi_j, \varepsilon_j)), \tag{5.123}$$

with

$$\partial_F(\chi_j, \varepsilon_j)_u = D \exp_{\chi_j}(\log_{\chi_j} \chi_{j+1})(\sqrt{\Delta_j} B(\chi_j)u). \tag{5.124}$$

This transformation does not depend on θ, and will cancel in the likelihood ratio.

Define the driftless process

$$\chi_{t_{j+1}} = \exp_{\chi_{t_j}} \left(\sqrt{\Delta_j} \sum_{i=1}^{n} B_i(\chi_{t_j})\varepsilon_i(t_j) \right), \quad j = 0, ..., N-1. \tag{5.125}$$

The density of $\boldsymbol{\varepsilon}$ for the driftless process is

$$p_{\boldsymbol{\varepsilon},\theta}(\Phi^{-1}(\chi_{0:N})) \propto \exp\left(\sum_{j=1}^{N} - \log_{\chi_{t_j}}(\chi_{t_{j+1}})^T \mathbf{B}(\chi_{t_j})^{-T} \mathbf{B}(\chi_{t_j})^{-1} \log_{\chi_{t_j}}(\chi_{t_{j+1}})/2\Delta_j \right), \tag{5.126}$$

and the density of $\chi_{0:N}$ is

$$p_{\chi_{0:N}}(\Phi^{-1}(\chi_{0:N})) \propto p_{\boldsymbol{\varepsilon},\theta}(\Phi^{-1}(\chi_{0:N}))|D\Phi^{-1}(\chi_{0:N})|, \tag{5.127}$$

where has $|D\Phi^{-1}(\chi_{0:N})|$ the same form as the determinant of the Jacobian for process with drift, and these terms will cancel when we calculate the likelihood ratio.

The final form of the likelihood ratio is

$$\frac{p_\theta(\chi_{1:N})}{p(\chi_{1:N})} \propto \exp\Bigg(\sum_{j=0}^{N} \log_{\chi_{t_j}}(\chi_{t_{j+1}})^T \mathbf{B}(\chi_{t_j})^{-T} \mathbf{B}(\chi_{t_j})^{-1} A(\chi_{t_j},\theta) - \\ - \frac{\Delta_j}{2} A(\chi_{t_j},\theta)^T \mathbf{B}(\chi_{t_j})^{-T} \mathbf{B}(\chi_{t_j})^{-1} A(\chi_{t_j},\theta)\Bigg). \tag{5.128}$$

If $\boldsymbol{\alpha}(\chi_{t_j},\theta)$ contains the coefficients of $A(\chi_{t_j},\theta)$ in the basis $\mathbf{B}(\chi_{t_j})$, then

$$\frac{p_\theta(\chi_{1:N})}{p(\chi_{1:N})} \propto \exp\left(\sum_{j=0}^{N} \log_{\chi_{t_j}}(\chi_{t_{j+1}})^T \mathbf{B}(\chi_{t_j})^{-T} \boldsymbol{\alpha}(\chi_{t_j},\theta) - \tfrac{\Delta_j}{2} \boldsymbol{\alpha}(\chi_{t_j},\theta)^T \boldsymbol{\alpha}(\chi_{t_j},\theta)\right). \tag{5.129}$$

5.6.3 Girsanov theorem on manifolds

Girsanov theorem has been generalized to differentiable manifolds by Elworthy in [33] (p. 263). Let X_t and Y_t be two processes on an m-dimensional Riemannian manifold $\mathcal{M}$ (with a metric $\langle , \rangle$):

$$dX_t = A(X_t,\theta)dt + \sum_{k=1}^{m} B_k(X_t) \circ dw_k(t), \tag{5.130}$$

$$dY_t = \sum_{k=1}^{m} B_k(Y_t) \circ dw_k(t). \tag{5.131}$$

Let's denote by P_X and P_Y the measures corresponding to X_t and Y_t. Let's also assume that $B_1, ..., B_n$ are orthonormal. Under the Novikov condition

$$\mathbb{E}_{P_X} \exp\left(\frac{1}{2}\int_0^T \langle A(X_t,\theta), A(X_t,\theta)\rangle dt\right) < \infty, \tag{5.132}$$

Girsanov theorem states that $P_X \sim P_Y$ and

$$\frac{dP_X}{dP_Y}(X) = \exp\left(\int_0^T \langle A(X_t,\theta), dX_t\rangle - \frac{1}{2}\int_0^T \langle A(X_t,\theta), A(X_t,\theta)\rangle dt\right). \tag{5.133}$$

In coefficient form the likelihood ratio is:

$$\frac{dP_X}{dP_Y}(X) = \exp\left(\int_0^T \boldsymbol{\alpha}(X_t,\theta)^T \mathbf{B}(X_t)^{-1} dX_t - \frac{1}{2}\int_0^T \boldsymbol{\alpha}(X_t,\theta)^T \boldsymbol{\alpha}(X_t,\theta) dt\right). \tag{5.134}$$

Let $\rho \in C([0,T], M)$, i.e. it is a continuous path on M. Then we would like to maximize the function $l_\rho(\theta)$:

$$l_\rho(\theta) = \mathbb{E}\left[\frac{dP_X}{dP_Y}(X)\middle| X = \rho\right] \tag{5.135}$$

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

5.6.4 Likelihood ratio estimates

We derive the likelihood ratio function for the drift models proposed in Section 5.4. 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.

5.6.4.1 Constant drift

We write the drift in a matrix form

$$A(\chi_t, \theta) = K(\chi_t)\theta = K_{\chi t}\theta \tag{5.136}$$

$$\begin{aligned}
\frac{d\mu_\chi}{d\mu_{\chi_0}}(\chi) &= \exp\left\{\int_0^T \langle A(\chi_t,\theta), d\chi_t\rangle - \frac{1}{2}\int_0^T \langle A(\chi_t,\theta), A(\chi_t,\theta)\rangle dt\right\} = \\
&= \exp\left\{\int_0^T \langle (K_{\chi_t}\theta, d\chi_t\rangle - \frac{1}{2}\int_0^T \langle K_{\chi_t}\theta, K_{\chi_t}\theta\rangle dt\right\} = \\
&= \exp\left\{\int_0^T \theta^T d\chi_t - \frac{1}{2}\int_0^T \theta^T K_{\chi_t}\theta dt\right\}.
\end{aligned} \tag{5.137}$$

Given the observations $\chi_{t_0}, ..., \chi_{t_N}$, we can approximate the differential $d\chi_{t_j} \approx \log(\chi_{t_j}, \chi_{t_{j+1}})$ (that such a discretization is valid is justified in Theorem 7.37 [34]). Since we usually work with the Hamiltonian formulation of the exponential map, we can first find the initial momentum α_j which maps χ_{t_j} to $\chi_{t_{j+1}}$, and then set $\log(\chi_{t_j}, \chi_{t_{j+1}}) = K(\chi_{t_j})\alpha_j$. This allows us to obtain an approximation to the likelihood ratio:

$$\frac{d\mu_\chi}{d\mu_{\chi_0}}(\chi) \approx \exp\left\{\sum_{j=0}^{N-1} \theta^T \log(\chi_{t_j}, \chi_{t_{j+1}}) - \frac{1}{2}\sum_{j=1}^{N-1} \Delta_j \theta^T K_{\chi_{t_j}}\theta dt\right\}, \tag{5.138}$$

and then optimize with respect to θ

$$\hat{\theta} = \frac{1}{T}\sum_{i=1}^{n} K_{\chi_{t_i}}^{-1} \log(\chi_{t_i}, \chi_{t_{i+1}}). \tag{5.139}$$

If we actually obtain the true piecewise momenta which connect the data points we would have

$$\hat{\theta} = \frac{1}{T}\sum_{j=0}^{N-1} K_{\chi_{t_j}}^{-1} \log(\chi_{t_j}, \chi_{t_{j+1}}) = \tag{5.140}$$

$$\hat{\theta} = \frac{1}{T}\sum_{j=0}^{N-1} K_{\chi_{t_j}}^{-1} K_{\chi_{t_j}} \alpha_j = \tag{5.141}$$

$$\hat{\theta} = \frac{1}{T}\sum_{j=0}^{N-1} \alpha_j \tag{5.142}$$

Thus the estimator can be interpreted as a time average of the initial momenta connecting consecutive observations. As the time step is assumed to be small, the differential can be further approximated by $d\chi_{t_j} \approx \chi_{t_{j+1}} - \chi_{t_j}$:

$$\hat{\theta} = \frac{1}{T}\sum_{j=1}^{N-1} K_{\chi_{t_j}}^{-1} (\chi_{t_{i+1}} - \chi_{t_i}). \tag{5.143}$$

Note that when the metric is the identity matrix, then the exponential map is simply

addition, and we have

$$\hat{\theta} = \frac{1}{T}\sum_{j=0}^{N-1}(\chi_{t_{j+1}} - \chi_{t_j}) = \frac{1}{T}(\chi(T) - \chi(0)). \tag{5.144}$$

5.6.4.2 Mean-reverting drift

The likelihood ratio is

$$\frac{d\mu_\chi}{d\mu_{\chi_0}}(\chi) = \exp\left(\theta\int_0^T \langle\nabla dist(\chi_t,\mu), d\chi_t\rangle - \frac{1}{2}\theta^2\int_0^T \|\nabla dist(\chi_t,\mu)\|^2 dt\right), \tag{5.145}$$

which gives an estimate for θ

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

5.6.4.3 Shape descriptor drift

The likelihood ratio is

$$\begin{aligned}\frac{d\mu_\chi}{d\mu_{\chi_0}}(\chi) &= \exp\left(\int_0^T [\theta_1 \;\; \theta_2] \begin{bmatrix} (L_{\chi_t} - L_\mu)\nabla L(\chi_t)^T \\ (A_{\chi_t} - A_\mu)\nabla A(\chi_t)^T \end{bmatrix} d\chi_t - \right. \\ &- \left. \frac{1}{2}\int_0^T [\theta_1 \;\; \theta_2] \begin{bmatrix} (L_{\chi_t} - L_\mu)\nabla L(\chi_t)^T \\ (A_{\chi_t} - A_\mu)\nabla A(\chi_t)^T \end{bmatrix} \begin{bmatrix} (L_{\chi_t} - L_\mu)\nabla L(\chi_t)^T \\ (A_{\chi_t} - A_\mu)\nabla A(\chi_t)^T \end{bmatrix}^T \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} dt \right).\end{aligned} \tag{5.146}$$

Setting the derivative to zero we obtain

$$\int_0^T \begin{bmatrix} (L_{\chi_t} - L_\mu)\nabla L(\chi_t)^T \\ (A_{\chi_t} - A_\mu)\nabla A(\chi_t)^T \end{bmatrix} \begin{bmatrix} (L_{\chi_t} - L_\mu)\nabla L(\chi_t)^T \\ (A_{\chi_t} - A_\mu)\nabla A(\chi_t)^T \end{bmatrix}^T dt \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} = \begin{bmatrix} \int_0^T (L_{\chi_t} - L_\mu)\nabla L(\chi_t)^T d\chi_t \\ \int_0^T (A_{\chi_t} - A_\mu)\nabla A(\chi_t)^T d\chi_t \end{bmatrix}, \tag{5.147}$$

and the estimate for θ becomes

$$\begin{bmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{bmatrix} = \left(\int_0^T \begin{bmatrix} (L_{\chi_t} - L_\mu)\nabla L(\chi_t)^T \\ (A_{\chi_t} - A_\mu)\nabla A(\chi_t)^T \end{bmatrix} \begin{bmatrix} (L_{\chi_t} - L_\mu)\nabla L(\chi_t)^T \\ (A_{\chi_t} - A_\mu)\nabla A(\chi_t)^T \end{bmatrix}^T dt\right)^{-1} \begin{bmatrix} \int_0^T (L_{\chi_t} - L_\mu)\nabla L(\chi_t)^T d\chi_t \\ \int_0^T (A_{\chi_t} - A_\mu)\nabla A(\chi_t)^T d\chi_t \end{bmatrix} \tag{5.148}$$

To simplify the notation we define M_j as the Grammian matrix of $\nabla|L(\chi_{t_j}) - L|^2$ and $\nabla|A(\chi_{t_j}) - A|^2$, and set

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

Then estimate based on the discrete observations is

$$\begin{bmatrix} \hat{\theta}_1 \\ \\ \hat{\theta}_2 \end{bmatrix} = \left(\sum_{j=0}^{N-1} M_j \Delta_j \right)^{-1} b. \tag{5.149}$$

5.6.5 Estimation results

We present the performance of the likelihood ratio estimator for the different types of diffusion drifts. It is known for Euclidean diffusions that if $dt \to 0$, and $T \to \infty$ the likelihood-ratio estimator is consistent. We look at the numerical convergence of the estimates in our case as time increases while keeping the time step fixed. We perform the estimation experiment 100 times and plot how each estimate of θ changes with time. The results are displayed in the figures below: 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.

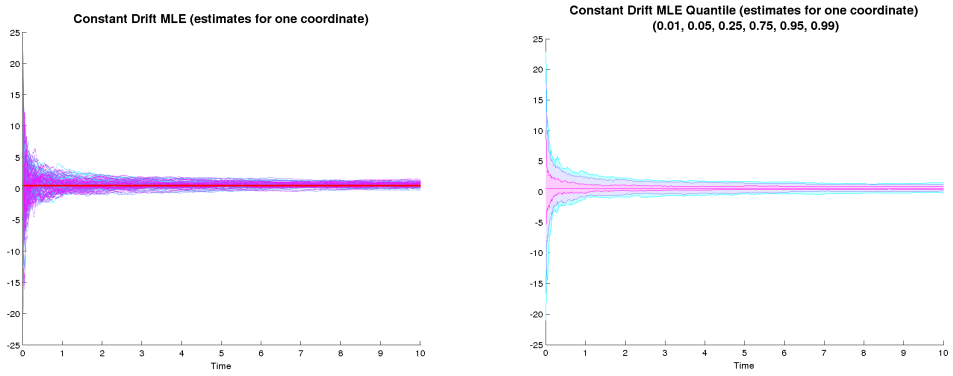


Figure 5.6: Estimation of a constant drift (first parameter)

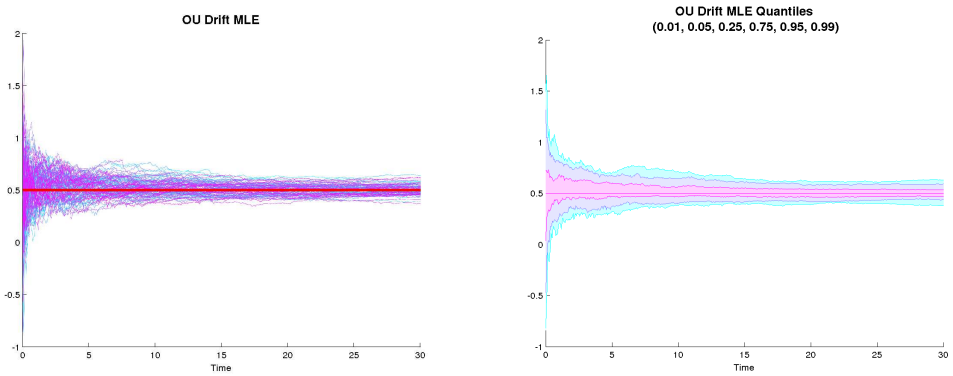
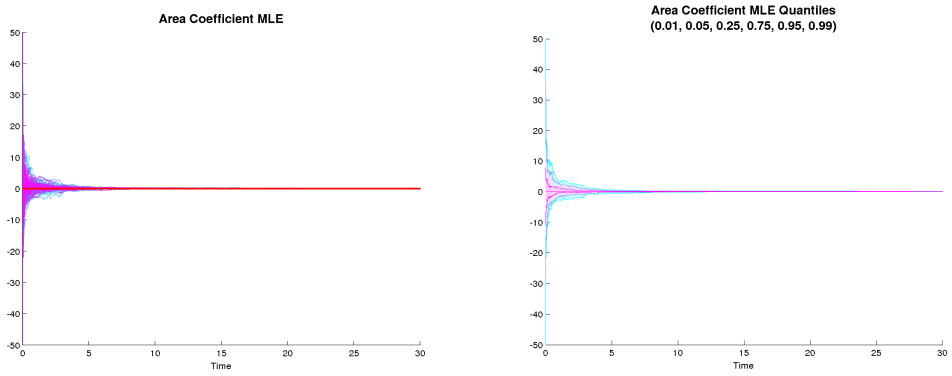
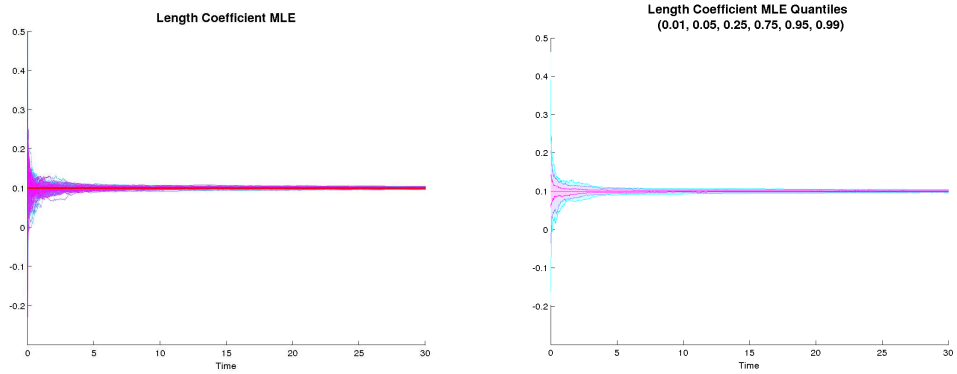


Figure 5.7: Estimation of the coefficient in a mean-reverting drift (the initial shape is a circle and the template shape is a dumbbell)

Remark: in the estimation procedure we use the true momenta $\boldsymbol{\alpha}_t's$ and the true control points. In real applications these will not be available and they need to be computed using the logarithm map or its approximation when it does not exist (the question of existence is further discussed in Section 6.2).



(a) Estimation of the coefficient corresponding to the gradient of the distance to fixed area



(b) Estimation of the coefficient corresponding to the gradient of the distance to fixed length

Figure 5.8: 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 above plots, but with time they quickly approach the true parameter

5.7 On the properties of the solutions of shape diffusion equations

In this section we discuss the properties of the solutions of the proposed diffusions.

5.7.1 Definitions

Weak vs. Strong Solutions. There are two distinct notions of a solution of a stochastic differential equation. A process X_t adapted to a filtration $\mathcal{F}_t$ associated with some probability space paired with a $\mathcal{F}_t$-adapted m-dimensional Brownian motion which together satisfy (5.6):

$$f(X_t) - f(X_0) = \int_0^t Af(X_s)ds + \int_0^t \sum_{k=1}^{n} B_k f(X_s) \circ dw_k(s), \tag{5.150}$$

for any compactly supported function $f \in \mathcal{M}$ is called the *weak solution* of (5.8). If X_t is adapted to the filtration $\mathcal{F}_t^W$ generated by W_t, then it is a *strong solution* on the probability space associated to $\mathcal{F}_t^W$. A defining property of strong solutions is that the process at any time can be written as a function of the initial condition and the Brownian motion: $X(\cdot) = F(X_0, W(\cdot))$ a.s.. The existence of such a function and its smoothness are important in the study of the flow of the shape over time, so we are interested in strong solutions.

Pathwise Uniqueness vs. Uniqueness in Law. *Pathwise uniqueness* requires that any two solutions are equal a.s., and since this concept is not affected by the choice of filtration, it applies to both weak and strong solutions. Another notion of uniqueness is *uniqueness in law*, which requires that any two solutions have the same probability distribution (under the assumption that they have the same initial

distribution). Since this definition directly deals with the laws of the processes, it does not depend on the probability spaces they are defined on, and hence it also applies to strong and weak solutions.

Important results by Yamada & Watanabe [96](Proposition 1 and Corollary 3) state

$$\textit{pathwise uniqueness} \Rightarrow \textit{uniqueness in law},$$

$$\textit{weak solution + pathwise uniqueness} \Rightarrow \textit{(unique) strong solution.}$$

Other relationships exist: for example, when the coefficients are bounded and the mixing matrix has a bounded inverse (Theorem 4.2 [10])

$$\text{uniqueness in law + strong solution} \Rightarrow \text{pathwise uniqueness.}$$

We will discuss different conditions more thoroughly in the next section.

5.7.2 Conditions for existence and uniqueness

In this section we include some results on existence and uniqueness of SDEs on manifolds. Some of the earliest work on this subject was done by Itô, who formulates sufficient conditions for existence and uniqueness of stochastic differential equations in Itô form on a general differentiable manifold [50]. Consider the stochastic differential equation in local coordinates:

$$dX^i(t) = \hat{a}^i(X_t)dt + \sum_{k=1}^{n} b_k^i(X_t) \cdot dw_k(t), \quad i = 1, ..., n. \tag{5.151}$$

Assume that $\hat{a}^i$ and b^i_k are bounded in the following sense: there exists a coordinate system in which $|\hat{a}^i(X_t)| < K$ and $|b^i_k(X_t)| \leq K$ for some constant K (chart-independent), such that in that coordinate system the neighborhood of each point on the manifold is mapped to the interior of the unit sphere in $\mathbb{R}^n$ and the point is mapped to the center of this sphere. If additionally the coefficients are continuously differentiable, then there exists only one solution to the stochastic differential equation given some initial condition (Theorem 3.1 [50]). The solution is strong and pathwise unique in the sense of the definitions above. Alternatively, we can directly analyze the coefficients of the generator of the process (5.33). When the coefficients $\hat{a}^j$, B_{jk} and the coefficients of the inverse of the matrix B are all bounded and continuously differentiable, there exists a continuous Markov process with the desired generator (Theorem [51]). Global existence and uniqueness conditions are provided by Gliklikh [41] (Theorem 7.36, Remark 7.39) : a strongly unique solution of (5.12) exists if the norms of the tangent vectors $A(X_t, \theta)$ and the linear operators $B(X_t)$ are bounded and

$$\|tr\Gamma_{m'}(B(m'), B(m'))\| \leq C \tag{5.152}$$

for $m' \in V_m(r)$ holds on the balls $V_m(r)$ in the charts of a uniform Riemannian atlas where the bound C is independent of the ball and the chart.

In general, boundedness or Lipschitz continuity are not always required to obtain

existence and uniqueness of solutions of SDEs. One can obtain similar results under weaker conditions, as long as the coefficients are sufficiently smooth and their growth is controlled. For example, local Lipschitz continuity and linear growth can be sufficient (Durrett (page 190, Theorem 3.1)) [31]. Under the following conditions:

$$|b_k^i(x) - b_k^i(y)| \leq K_n|x-y|, \quad i = 1, \ldots, d, \tag{5.153}$$

$$|\hat{a}_i(x) - \hat{a}_i(x)| \leq K_n|x-y|, \quad i = 1, \ldots, d, \tag{5.154}$$

when $|x|, |y| \leq n$, and

$$\sum_{i=1}^{d} 2x_i\hat{a}_i(x) + B_{ii}(x) \leq C(1+|x|^2), \tag{5.155}$$

the SDE in (5.9) (as an equation on $\mathbb{R}^n$) has a unique strong solution.

In general, to show existence of solutions on a manifold, one can either directly justify the global conditions, or show existence in local coordinates and justify that solutions in different charts can be pieced together (or that there is a global chart).

5.7.3 Solutions on the landmark manifold

We recall that $\mathcal{M}$ is an open subset of $\mathbb{R}^{2m}$ (the subset on which the landmarks are all different), so any diffusion can be written in a global chart on this subset. We can show the existence of local solutions on the landmark manifold by showing

the existence of local solutions in this chart, and further, if we can justify that the solutions can be extended to the whole chart without leaving it (i.e. landmarks do not coincide), we obtain global solutions. We discuss here the case when the drift is zero and $n = m$. Note that still an extra drift term appears in the local representation resulting from the Riemannian correction 5.48. So one needs to study the properties both of the noise terms and the Riemannian correction drift term.

Properties of the noise term. As we have shown in Section 5.3 the smoothness and the positive-definiteness of the covariance matrix ensure its square roots are smooth which allows us to define the diffusion process with smooth coefficients (they correspond to smooth vector fields on the landmark manifold). Further, the covariance matrix consists of the evaluations of the Gaussian kernel at the landmark points and the matrix is positive definite. Since the Gaussian kernel is Lipschitz, the coefficients of the matrix are Lipschitz. A positive definite matrix with Lipschitz coefficients has a unique positive square root which also has Lipschitz coefficients [36].

Properties of the Riemannian correction term. We recall that the correction term on a Riemannian manifold ($m = n$) takes the form

$$\begin{aligned} corr_k(\chi) = \frac{1}{2}\Gamma^k_{ij}\sum_{l=1}^{m} b^i_l(\chi)b^j_l(\chi) &= \frac{1}{2}\Gamma^k_{ij}g^{ij}(\chi) = \frac{1}{2}|g|^{-1/2}\frac{\partial}{\partial x_i}(|g|^{1/2}g^{ik}) = \\ &= \frac{1}{2}\frac{\partial}{\partial x_i}g^{ik} + \frac{1}{2}|g|^{-1/2}\frac{\partial}{\partial x_i}(|g|^{1/2})g^{ik} = \\ &= \frac{1}{2}\underbrace{\frac{\partial}{\partial x_i}B_{ik}}_{f_k} + \frac{1}{2}\underbrace{|B|^{1/2}\frac{\partial |B|^{-1/2}}{\partial x_i}B_{ik}}_{h_k}, \end{aligned} \tag{5.156}$$

where B is the cometric. First we note that it is a sum of product of differentiable functions, so the coefficients are differentiable and, hence, locally Lipschitz. This, together with the smoothness of the noise coefficients, justifies the existence of a local solution. Further, we observe that the first term in the above sum consists of derivatives of the Gaussian kernel, so it is bounded with bounded derivatives, i.e. it is globally Lipschitz. However, it is harder to say at first look what the properties of the second term are.

The term simplifies in the case of two landmarks: x_1 and x_2. The form of the cometric is

$$B = \begin{bmatrix} 1 & e^{-\frac{\|x_1-x_2\|^2}{2\sigma^2}} \\ e^{-\frac{\|x_1-x_2\|^2}{2\sigma^2}} & 1 \end{bmatrix} \tag{5.157}$$

and hence the metric is

$$B^{-1} = \frac{1}{1-e^{-\frac{\|x_1-x_2\|^2}{\sigma^2}}} \begin{bmatrix} 1 & -e^{-\frac{\|x_1-x_2\|^2}{2\sigma^2}} \\ -e^{-\frac{\|x_1-x_2\|^2}{2\sigma^2}} & 1 \end{bmatrix}. \tag{5.158}$$

The first term is easy to calculate:

$$f_j(x_1, x_2) = \frac{\partial}{\partial x_i} B_{ij} = \frac{x_j - x_{j'}}{\sigma^2} e^{-\frac{\|x_j-x_{j'}\|^2}{2\sigma^2}}, \tag{5.159}$$

where j' corresponds to the coordinate not equal to j.

The j'th coefficient of the second term in the correction is

$$\begin{aligned} h_j(x_1,x_2) &= |B|^{1/2}\frac{\partial |B|^{-1/2}}{\partial x_i}g^{ij} = -\frac{1}{2}|B|^{1/2}|B|^{-3/2}\frac{\partial |B|}{\partial x_i}g^{ij} = -\frac{1}{2}|B|^{-1}\frac{\partial |B|}{\partial x_i}g^{ij} = \\ &= -\frac{1}{2}\frac{1}{1-e^{-\frac{\|x_1-x_2\|^2}{\sigma^2}}}\frac{\partial}{\partial x_i}\left(\frac{\|x_1-x_2\|^2}{\sigma^2}\right)e^{-\frac{\|x_1-x_2\|^2}{\sigma^2}}g^{ij} = \\ &= -\frac{1}{2}\frac{1}{e^{\frac{\|x_1-x_2\|^2}{\sigma^2}}-1}\frac{\partial}{\partial x_j}\left(\frac{\|x_1-x_2\|^2}{\sigma^2}\right) - \\ &\quad -\frac{1}{2}\frac{1}{e^{\frac{\|x_1-x_2\|^2}{\sigma^2}}-1}\frac{\partial}{\partial x_{j'}}\left(\frac{\|x_1-x_2\|^2}{\sigma^2}\right)e^{-\frac{\|x_1-x_2\|^2}{2\sigma^2}} \end{aligned}$$

. (5.160)

Expanding the derivatives, we obtain

$$\begin{aligned} h_1(x_1,x_2) &= -\frac{1}{2}\frac{1}{e^{\frac{\|x_1-x_2\|^2}{\sigma^2}}-1}\frac{x_1-x_2}{\sigma^2} - \frac{1}{2}\frac{1}{e^{\frac{\|x_1-x_2\|^2}{\sigma^2}}-1}\frac{x_2-x_1}{\sigma^2}e^{-\frac{\|x_1-x_2\|^2}{2\sigma^2}} = \\ &= \frac{1}{2}\frac{x_1-x_2}{\sigma^2}\frac{e^{-\frac{\|x_1-x_2\|^2}{2\sigma^2}}-1}{e^{\frac{\|x_1-x_2\|^2}{\sigma^2}}-1} = \frac{x_1-x_2}{2\sigma^2}e^{-\frac{\|x_1-x_2\|^2}{2\sigma^2}}\frac{1-e^{\frac{\|x_1-x_2\|^2}{2\sigma^2}}}{e^{\frac{\|x_1-x_2\|^2}{\sigma^2}}-1} = \\ &= \frac{x_1-x_2}{2\sigma^2}e^{-\frac{\|x_1-x_2\|^2}{2\sigma^2}}\frac{-1}{e^{\frac{\|x_1-x_2\|^2}{2\sigma^2}}+1} \end{aligned} \quad (5.161)$$

and

$$h_2(x_1,x_2) = \frac{x_1-x_2}{\sigma^2}e^{-\frac{\|x_1-x_2\|^2}{2\sigma^2}}\frac{1}{1+e^{\frac{\|x_1-x_2\|^2}{2\sigma^2}}}. \quad (5.162)$$

So for the case of two landmarks the correction term is well defined even if the

two landmarks coincide. It is also continuous and bounded. We can see that as $\|x_1 - x_2\| \to \infty$, the product of the first two terms goes to zero, and the last term goes to zero, as well. Further, the derivatives are bounded too. Let $r = (x_2 - x_1)\sqrt{2}\sigma$. Then,

$$h_1(x_1, x_2) = \frac{r}{\sqrt{2}\sigma} \frac{e^{-r^2}}{1 + e^{r^2}}, \tag{5.163}$$

and

$$\frac{\partial h_1}{\partial r} = \frac{1}{\sqrt{2}\sigma} \frac{e^{-r^2} - 1 + 2r - 3re^{-r^2}}{(1 + e^{r^2})^2}, \tag{5.164}$$

and the full derivative is

$$\frac{\partial h_1}{\partial x_1} = -\frac{1}{\sqrt{2}\sigma} \frac{e^{-\|x_1 - x_2\|^2} - 1 + 2(x_2 - x_1) - 3(x_2 - x_1)e^{-\|x_1 - x_2\|^2}}{(1 + e^{\|x_1 - x_2\|^2})^2}. \tag{5.165}$$

The denominator goes to infinity as $r \to \infty$ much faster than the nominator, so the derivative is bounded, from which we conclude that h_1 is Lipschitz continuous. Same can be shown for h_2.

We conclude that all the coefficients in the drift and noise terms for two-landmark diffusion equations are Lipschitz, and hence the solution exists at all times and is a homeomorphic flow. This property guarantees that the solution also never leaves the chart.

Proposition 5.4. *Suppose* (5.43) *in local coordinates (as an SDE on* $\mathbb{R}^4$*) has a global solution defined for which an associated global flow of homeomorphisms exists. Then no two landmarks meet along the path of the solution and the solution stays on the landmark manifold for all times.*

Proof. (of Proposition 5.4) Let's denote this flow starting from time 0 and ending at time t by $F_{0,t}$. Assume the two landmarks x_1 and x_2 meet at some time t. Denote the hyperplane defined by $x_1 = x_2$ by H_{12}. Therefore, we have assumed that the stochastic flow maps a set of points outside of H_{12} to a set of points on H_{12}: $F_{0,t}(\chi) = \bar{\chi}$ where $\chi \notin H_{12}$ and $\bar{\chi} \in H_{12}$. We can check that once two landmarks coincide, their equations in the system become identical, and any solution restricted to H_{12} is equivalent to the solution of the stochastic differential equation with one of the two points x_1 or x_2 dropped, i.e. the flow on the corresponding lower-dimensional space. As this flow is also a homeomorphism, $F_{0,t}^{-1}(\bar{\chi})$ should map to an element on H_{12}. Therefore, we reach to a contradiction with the assumption that there exists $\chi \notin H_{12}$, such that $F(\chi) = \bar{\chi}$ and we conclude that the landmarks cannot meet. □

Extensions to higher dimensions. On manifolds of more than two landmarks, the formula for the correction term does not automatically simplify, and one needs to consider the properties of the term in the limit of landmarks coinciding, and establish the behavior of the solutions when approaching the boundary of the manifold.

Chapter 6

Conclusion and Future Directions

In this thesis we addressed diverse aspects of statistical inference of stochastic processes on the manifold of shapes: modeling, online filtering, offline learning, and applications. Our work reveals both the opportunities for advancement in this underexplored area and the challenges associated with statistics on manifolds, and high-dimensional inference. Below we outline several future directions of research.

6.1 Customized tracking models

We demonstrated that our general particle filter formulation is useful for tracking of wide variety of objects in videos. What is more important is that the framework allows for easy incorporation of constraints on the evolution models and the properties of the image observations, which can lead to highly customizable algorithms for special applications. Next we describe how this can be done in two possible scenarios: in

case of heart tracking and cell tracking.

6.1.1 Multi-region cardiac tracking

In Section 3.5.4 we presented results of tracking the motion of the left ventricle of the heart in MRI sequences. One of the biggest challenges of this task is the detection of the epicardium (the outer wall of the chamber) as it often gets blended with the background when expanded. We have modeled the image as a three-region representation of the heart: left ventricle, left ventricle wall, and outer background. However, we know that this is a very simplified representation of all other objects observed in the image: right ventricle, apex, pericardium, interventricular septum, other organs, muscles, and tissue. The cumulative motion of these components follow the topology-preserving assumptions we have already made. We also have the additional knowledge that certain parts of the background do not move. A customized model which incorporates this information would include the following

1. extract the boundaries of all the regions of interest in the initial frame and extract summary statistics (needed for the observation model in Section 3.3.1)

2. specify a random diffeomorphic model to deform the set of boundaries: this will be essentially the same as the models used before - the difference that some of the boundaries are not closed contours does not pose a problem as the deformations are defined everywhere on $\mathbb{R}^2$.

3. add the additional constraints that a set of points does not move, i.e. $v(x) = 0$

for $x \in R_{static}$. Points with zero initial momenta remain static according to our models.

Such a system will be best to implement with a interactive interface within which a user can annotate the initial boundaries and set some immobility constraints, and then run a cardiac tracking algorithm.

6.1.2 Organelle tracking

The availability of high-resolution microscopic images opens interesting problems: *Can we track the location and shape of individual cellular organelles within a cell? Is it possible to do this simultaneously for many cells interacting with each other?*

Usually individual organelles appear to float freely within a cell, however, they are constrained by by a couple of rules: they cannot leave the cell and cannot cross boundaries, i.e. their motion preserves the topology. We have already proposed a model which allows the epicardium and endocardium of the heart to move more freely with respect to each other in Section 3.5.4. Less restrictive deformation vector fields for multiple shapes have been proposed in [5] in the context of registration. Stochastic models in this framework are yet to be developed and tested.

6.2 Controllability

In Section 3.2.3 we have presented the sub-Riemannian interpretation of the flows driven by control points. Experimentally we have observed that we can generate a

wide variety of shapes using this procedure. However, a side question is

Can we generate all possible shapes?

In the context of the manifold of deformable landmarks (Section 2.5.1), a shape is a set of m landmarks. Therefore, we would like to know whether starting from an initial configuration of landmarks we can deform them into another arbitrary configuration of landmarks through a sub-Riemannian flow driven by n landmarks where $n \leq m$, i.e. whether there any two points on the landmark manifold can be connected through a horizontal curve. In control theory, this concept is known as controllability. Let $R_\Delta(x_0)$ be the reachability set associated with the dynamical system in (3.2.3), i.e. the set of points which can be reached by moving along the distribution Δ. A dynamical system is controllable if $R_\Delta(x_0) = \mathcal{M}$. Chow's theorem [70](p. 9–10) provides a condition for controllability based on the properties of the vector fields in the horizontal distribution. If the vector fields and their brackets span the whole tangent bundle of the manifold (the distribution is bracket-generating), then any two points can be joined through a path in the distribution. Although we did not require to establish controllability for forward simulation of shapes, the property becomes important when we need to calculate logarithm maps between two shapes (controllability establishes whether the logarithm exists). Obtaining such results for various types of kernels can have applications outside of the image processing domain.

6.3 Convergence in RKHS norm

In Section 4 we raised the question of convergence of random vector fields along a contour in RKHS norm with Gaussian kernels for which the covariance is also a Gaussian kernel. Because the Gaussian kernel is not square integrable on $\mathbb{R}^2$ (which is also true for any other radially symmetric kernel), the trace condition in (4.46) does not hold for domains containing an open interval or an open ball and convergence fails. Furthermore, Mercer's theorem does not hold for non-square-integrable kernels, which does not allow us to construct eigenbasis for the RKHS to facilitate our analysis. However, the Gaussian kernel is not the only kernel which provides sufficient smoothness conditions for generating diffemorphic shapes. Other Sobolev norms and corresponding kernels have been used for diffeomorphic shape matching [64, 62], so studying the random field convergence properties for a broad class of kernels can guide to designing better numerical approximation schemes and statistical models.

6.4 Estimation of diffusion parameters from sparse observations

We have only scratched the surface of learning dynamical processes of shapes from observations. In Section 5.6.2, we have addressed the parameter estimation problem when the observations $X_0, ..., X_T$ are closely observed and when the direct approximation of the stochastic integrals in the likelihood ratio (5.128) is reliable (for example,

when working with videos or dense image sequences). However, for many biomedical applications it is impossible or expensive to obtain high temporal resolution (for example, in longitudinal studies). Fortunately, the proposed approach can be further extended to the case when only sparse observations are observed. As before, the maximum likelihood estimate for θ maximizes $P_\theta(X_1 = x_1, ..., X_T = x_T)$. As solving the optimization problem directly is intractable, we introduce the full path of X_t as a hidden variable Z_t, and take an E-M approach to obtain an estimate for θ. Of course, the likelihood of the full path is not well defined, so we need to consider the likelihood ratio with respect to some process not depending on θ. As usual we pick this process to be the one corresponding to the driftless diffusion. Let's define

$$L(\theta; Z, X) = \frac{dP_{Z_t}}{dP_{W_t}}(Z, X) \tag{6.1}$$

(this is the likelihood ratio evaluated at the random path Z with discrete observations X). The steps of the E-M algorithm are:

$$\text{E-step} \qquad Q(\theta|\theta_{old}) = \mathbb{E}_{Z|X,\theta_{old}} \log L(\theta; Z, X) \tag{6.2}$$

$$\text{M-step} \qquad \theta_{new} = \arg\max_\theta Q(\theta|\theta_{old}). \tag{6.3}$$

First, we note that we have a closed form formula for the full likelihood ratio

$L(\theta; Z, X)$. Nevertheless, the expectation of this function cannot be computed explicitly, and we need to resort to a Monte Carlo approximation of Q based on samples from $P_{Z|X}$. We observe that $Z|X$ is a diffusion process constrained to hit the observations $x_0, ..., x_T$. To sample from this process, we need to sample a sequence of subpaths between every two observations, which reduces the problem to sampling from the corresponding diffusion bridge. Luckily, a diffusion bridge on manifold is a diffusion process itself (as is true for diffusions in Euclidean space), so as long as we establish the form of its drift and noise terms, we can simulate it using the exponential map. We outline the formulation of the diffusion bridge on $\mathbb{R}^n$ to reveal what kind of steps will be necessary to achieve a solution on a manifold.

Sampling Diffusion Bridges on $\mathbb{R}^n$. For simplicity, we assume that we have only two observations x_0 and x_T, so we want to sample from

$$dZ_t = A(Z_t)dt + B(Z_t)dW_t, \quad Z_0 = x_0, Z_T = x_T. \tag{6.4}$$

This is equivalent to sampling from the unconstrained diffusion

$$d\bar{Z}_t = A(\bar{Z}_t)dt + \frac{1}{2}B(\bar{Z}_t)B(\bar{Z}_t)^*\nabla_z \log p(t, \bar{Z}_t, T, x_T)dt + B(\bar{Z}_t)dW_t, \tag{6.5}$$

where $p(t, Z_t, T, x_T)$ is the transition probability from Z_t at time t to x_T at the final time. Unfortunately, the dependency of the drift and the diffusion coefficients on the state make the diffusion non-Gaussian and we do not have a closed form

for $p(t, Z_t, T, x_T)$ which prevents us from sampling the paths using a simple Euler-Maruyama scheme. Instead, we can achieve the sampling by importance sampling: we sample from a process whose paths are easy to generate, and associate a weight to each sample path which is equal to the Radon-Nikodym derivative of the target distribution with respect to the proposal distribution evaluated at this path. The proposal processes suggested in [27] provide a good starting point.

6.5 Diffusion properties

In Section 5.7 we showed that while existence of local solutions of the proposed diffusions follows directly from the smoothness of the metric, establishing global properties is a non-trivial task due to the boundary of the landmark manifold. We have established the global existence for the very special case of a two-landmark manifold, and it is an interesting question how to set up an inductive argument to extend this result to any dimension. In high dimensions, we need to consider all possible combinations of landmarks approaching each other, so direct Taylor-expansion calculations might not be feasible. As many conditions for proving properties of SDE solutions are sufficient but not necessary, different approaches should be compared to obtain strongest results. For example, going a step further and studying the properties of the curvature of the manifold may allow showing that the diffusions have global smooth flow [60]. Further study of the gradient drifts could reveal whether the processes possess ergodic properties, which in turn can inform about appropriate parameter estimation

procedures and their statistics.

6.6 Toward a unified framework

A unified framework for stochastic filtering of shapes can be built by considering a hidden diffusion process on the infinite-dimensional manifold of closed curves with discrete-time image observations. The advantage of working directly with curves is that we can eliminate errors due to discretization or spurious landmark correspondence. Although diffusions on infinite-dimensional manifolds have been studied thoroughly by Belopolskaya and Daletsky in [15], estimation on these spaces (from discrete observations) is an open and exciting topic to be explored in the future.

Bibliography

[1] A. Agrachev, D. Barilari, and U. Boscain, "On the Hausdorff volume in sub-Riemannian geometry," *Calculus of Variations and Partial Differential Equations*, vol. 43, no. 3-4, pp. 355–388, 2011.

[2] ——, *Introduction to Riemannian and Sub-Riemannian geometry.* Preprint SISSA, 2013.

[3] A. Agrachev, U. Boscain, R. Neel, and L. Rizzi, "Intrinsic random walks in riemannian and sub-riemannian geometry via volume sampling," *arXiv:1601.03304*, 2016.

[4] S. Allassonniére, A. Trouvé, and L. Younes, "Geodesic shooting and diffeomorphic matching via textured meshes," in *Energy Minimization Methods in Computer Vision and Pattern Recognition*, ser. Lecture Notes in Computer Science, A. Rangarajan, B. Vemuri, and A. Yuille, Eds. Springer Berlin Heidelberg, 2005, vol. 3757, pp. 365–381. [Online]. Available: http://dx.doi.org/10.1007/11585978_24

[5] S. Arguillère, E. Trélat, A. Trouvé, and L. Younes, "Multiple shape registration using constrained optimal control," *SIAM J. Imaging Sci.*, vol. 9.

[6] N. Aronszajn, "Theory of reproducing kernels," *Transactions of the American Mathematical Society*, vol. 68, no. 3, p. 337, 1950.

[7] F. Arrate, T. Ratnanather, and L. Younes, "Diffeomorphic active contours," *SIAM J. Imaging Sci.*, vol. 3, no. 2, pp. 176–198, 2010.

[8] S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-gaussian Bayesian tracking," *IEEE Transactions on Singal Processing*, vol. 50, no. 2, pp. 174–188, 2002.

[9] F. Ball, I. Dryden, and M. Golalizadeh, "Brownian motion and Ornstein-Uhlenbeck processes in planar shape space," *Methodology and Computing in Applied Probability*, vol. 10, no. 1, pp. 1–22, 2008. [Online]. Available: http://dx.doi.org/10.1007/s11009-007-9042-6

[10] R. Bass, *Diffusions and Elliptic Operators*, ser. Probability and Its Applications. Springer, 1997.

[11] M. Bauer, P. Harms, and P. Michor, "Sobolev metrics on shape space of surfaces in n-space," *Journal of Geometric Mechanics*, 2010.

[12] A. N. Baushev, "On weak convergence of Gaussian measures," *Theory Prob. Applicaitons*, vol. 32, no. 4, 1987.

[13] P. Baxendale, "Measures and Markov processes on function spaces," *Mémoires de la Société Mathématique de France*, vol. 46, pp. 131–141, 1976.

[14] M. F. Beg, M. Miller, A. Trouve, and L. Younes, "Landmark matching via large deformation diffeomorphisms," *International Journal of Computer Vision*, vol. 61, no. 2, pp. 139–157, 2005.

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

[16] Y. K. Belyaev, "Analytic random processes," *Probability Theory and its Applications*, vol. 2, no. 4, pp. 402–407, 1959.

[17] B. Bonnard and M. Chyba, *Singular trajectories and their role in control theory*. Springer, 2003.

[18] G. Box and G. Jenkins, *Time Series Analysis: Forecasting and Control*. Holden-Day, San Francisco, 1970.

[19] A. Budhiraja, P. Dupuis, and V. Maroulas, "Large deviations for stochastic flows of diffemophisms," *Bernoulli*, vol. 16, no. 1, pp. 234–257, 2010.

[20] V. Cervera, F. Mascaró, and P. Michor, "The action of the diffeomorphism group on the space of immersions," *Differential Geom. Appl.*, no. 1, pp. 391–401, 1991.

[21] G. S. Chirikjian, *Stochastic Models, Information Theory, and Lie Groups, Volume 1*. Birkhäuser, 2009, vol. 1.

[22] ——, *Stochastic Models, Information Theory, and Lie Groups, Volume 2*. Birkhäuser, 2012, vol. 2.

[23] A. Chiuso and S. Soatto, "Monte Carlo filtering on Lie groups," *Proceedings of IEEE Conference on Decision and Control*, pp. 304–309, 2000.

[24] D. Comaniciu, V. Ramesh, and P. Meer, "Real-time tracking of non-rigid objects using mean shift," *Proc. CVPR*, vol. 2, pp. 142–149, 2000.

[25] D. Cremers, "Dynamical statistical shape priors for level-set based tracking," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 28, no. 8, pp. 1262–1273, 2006.

[26] D. Cremers, F. R. Schmidt, and F. Barthel, "Shape priors in variational image segmentation: Convexity, Lipschitz continuity and globally optimal solutions," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2008.

[27] B. Delyon and Y. Hu, "Simulation of conditioned diffusion and application to parameter estimation," *Stochastic Processes and their Applications*, vol. 116, pp. 1660–1675, 2006.

[28] R. Douc, O. Cappé, and E. Moulines, "Comparison of resampling schemes for particle filtering," *Proc. of Image and Signal Processing and Analysis*, pp. 64–69, 2005.

[29] A. Doucet, N. de Freitas, and N. Gordon, *Sequential Monte Carlo Methods in practice*. Springer, 2006.

[30] M. F. Driscoll, "The reproducing kernel Hilbert space structure of the sample paths of a Gaussian process," *Z. Wahrscheinlichkeistheorie verw. Geb.*, vol. 26, pp. 309–316, 1973.

[31] R. Durrett, *Stochastic Calculus: A Practical Introduction*, ser. Probability and Stochastics Series. CRC Press, 1996.

[32] S. Durrleman, M. Prastawa, G. Gerig, and S. Joshi, "Optimal data-driven sparse parameterization of diffeomorphisms for population analysis," in *Information Processing in Medical Imaging*, ser. Lecture Notes in Computer Science, G. Székely and H. K. Hahn, Eds. Springer Berlin Heidelberg, 2011, vol. 6801, pp. 123–134. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-22092-0_11

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

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

[35] W. Fleming and R. Rishel, *Deterministic and Stochastic Optimal Control*. Springer-Verlag, 1975.

[36] M. I. Freidlin, "On the factorization of non-negative definite matrices," *Theory of Probability and its Applications*, vol. 13, no. 2, 1968.

[37] P. Getreuer, "Chan-Vese segmentation," *Image Processing On Line*, 2012.

[38] W. Gilks and C. Berzuini, "Following a moving target - Monte Carlo inference for dynamic Bayesian models," *Journal of the Royal Statistical Society B*, vol. 63, pp. 127–146, 2001.

[39] J. Glaunès, A. Trouvé, and L. Younes, "Modeling planar shape variation via Hamiltonian flows of curves," in *Analysis and Statistics of Shapes, Modeling and Simulation in Science, Engineering and Technology, chapter 14. Birkhäuser*. Springer - Verlag, 2005.

[40] Y. Gliklikh, "Necessary and sufficient conditions for global-in-time existence of solutions of ordinary, stochastic, and parabolic differential equations," *Abstract and Applied Analysis*, 2006.

[41] Y. E. Gliklikh, *Global and Stochastic Analysis with Applications to Mathematical Physics*. Springer, 2011.

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

[43] ——, "Weak convergence to Brownian motion on sub-Riemannian manifolds," *arXiv:1403.0142*, 2014.

[44] N. J. Gordon, D. J. Salmond, and A. F. M. Smith, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," *IEE Proceedings-F (Radar and Signal Processing)*, vol. 140(2), pp. 107–113, 1993.

[45] U. Grenander, *General Pattern Theory*. Oxford University Press, 1993.

[46] N. C. Günther, "Hamiltonian mechanics and optimal control," Ph.D. dissertation, Harvard University, 1982.

[47] J. D. Hol, T. B. Schön, and F. Gustafsson, "On resampling algorithms for particle filters," *IEEE Nonlinear Statistical Signal Processing Workshop*, pp. 79–82, 2006.

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

[49] M. Isard and A. Blake, "Condensation - conditional density propagation for visual tracking," *Int'l. J. Computer Vision*, vol. 29, no. 1, pp. 5–28, 1998.

[50] K. Itô, "Stochastic differential equations in a differentiable manifold," *Nagoya Math. J.*, vol. 1, pp. 35–47, 1950.

[51] ——, "Stochastic differential equations in a differentiable manifold (2)," *Mem. College Sci. Univ. Kyoto Ser. A Math.*, vol. 28, pp. 1–85, 1953.

[52] S. Joshi and M. Miller, "Landmark matching via large deformation diffeomorphisms," *IEEE Transactions on Image Processing*, vol. 9, no. 8, pp. 1357–1370, 2000.

[53] D. G. Kendall, *Statistical Science*, vol. 4, no. 2, pp. 87–99, 1989.

[54] D. Kendall, "The diffusion of shape," *Advances in Applied Probability*, vol. 9, no. 3, pp. 428–430, 1977.

[55] D. G. Krige, "A statistical approach to some mine valuations and allied problems at the witwatersrand," Master's thesis, University of Witwatersrand, 1951.

[56] H. Krim and A. Yezzi, *Statistics and Analysis of Shapes*, ser. Modeling and Simulation in Science, Engineering and Technology. Birkhäuser, 2006.

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

[58] J. Kwon, M. Choi, C. Chun, and F. Park, "Particle filtering on the Euclidean group," *IEEE International Conference on Robotics and Automation*, 2007.

[59] J. Kwon and F. Park, "Visual tracking via particle filtering on the affine group," *The International Journal of Robotics Research*, vol. 29, pp. 198–217, 2010.

[60] X.-M. Li, "Strong p-completeness of stochastic differential equations and the existence of smooth flows on noncompact manifolds," *Probab. Theory Relat. Fields*, vol. 100, pp. 485–511, 1994.

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

[62] M. Micheli and J. A. Glaunès, "Matrix-valued kernels for shape deformation analysis," *arXiv:1308.5739*, 2013.

[63] 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*, vol. 5, pp. 394–433, 2012.

[64] ——, "Sobolev metrics on diffeomorphism groups and the derived geometry on spaces of submanifolds," *Izvestiya: Mathematics*, vol. 77, no. 3, pp. 541–570, 2013.

[65] P. Michor and D. Mumford, "Riemannian geometries on spaces of plane curves," *J. Eur. Math.*, no. 8, pp. 1–48, 2006.

[66] P. W. Michor, D. Mumford, J. Shah, and L. Younes, "A metric on shape spaces with explicit geodesics," *Rendiconti Lincei - Matematicae Applicazioni*, vol. 8, pp. 25–57, 2008.

[67] M. Miller, "Computational anatomy: shape, growth, and atrophy comparison via diffeomorphism," *NeuroImage*, vol. 23, pp. S19–S33, 2004.

[68] H. Q. Minh, "Reproducing kernel Hilbert space in learning theory: the sphere and the hypercube," 2000.

[69] J. Møller, *Statistical Inference and Simulation for Spatial Point Processes*. Chapman & Hall/CRC, 2004.

[70] R. Montgomery, *A Tour of Subriemannian Geometries, Their Geodesics and Applications*. American Mathematical Society, 2002.

[71] I. J. Ndiour, O. Arif, J. Teizer, and P. A. Vela, "A probabilistic observer for visual tracking," *IEEE American Control Conference*, 2010.

[72] J. Ortega, "Asymptotic behavior of Gaussian random fields," *Probability theory and related fields*, vol. 59, no. 2, pp. 169–177, 1982.

[73] E. Parzen, "Statistical inference on time series by RKHS methods," *Technical Report, Stanford University*, 1970.

[74] P. Radau, Y. Lu, K. Connelly, G. Paul, A. Dick, and G. Wright", "Evaluation framework for algorithms segmenting short axis cardiac MRI."," *The Midas Journal*, 07 2009, cardiac MR Left Ventricle Segmentation Challenge.

[75] Y. Rathi, N. Vaswani, A. Tannenbaum, and A. Yezzi, "Particle filtering for geometric active contours with application to tracking moving and deforming objects," *Computer Vision and Pattern Recognition*, vol. 2, pp. 2–9, 2005.

[76] L. Risser, F.-X. Vialard, R. Wolz, M. Murgasova, D. D. Holm, D. Rueckert, and THE ALZHEIMER'S DISEASE NEUROIMAGING INITIATIVE, "Simultaneous multi-scale registration using large deformation diffeomorphic metric mapping," *Medical Imaging, IEEE Transactions on*, vol. 30, no. 10, pp. 1746 –1759, 2011.

[77] J. A. Sethian, "Fast marching methods," *SIAM Review*, vol. 41, pp. 199–235, 1998.

[78] C. G. Small, *The Statistical Theory of Shape*. Springer, 1996.

[79] H. Snoussi and C. Richard, "Monte Carlo tracking on the Riemannian manifold of multivariate normal distributions," *IEEE Digital Signal Processing Workshop*, 2009.

[80] C. Snyder, T. Bengtsson, P. Bickel, and J. Anderson, "Obstacles to high-dimensional particle filtering," *Monthly weather review*, vol. 136, pp. 4629–4640, 2008.

[81] S. Sommer, M. Nielsen, S. Darkner, and X. Pennec, "Higher-order momentum distributions and locally affine LDDMM registration," *SIAM Journal on Imaging Sciences*, vol. 6, no. 1, pp. 341–367, 2013.

[82] H. Soneson, J. F. Ubachs, M. Ugander, H. kan Arheden, and E. Heiberg, "An improved method for automatic segmentation of the left ventricle in myocardial perfusion SPECT," *J. Nucl. Med.*, vol. 50, no. 2, pp. 205–213, 2009.

[83] V. Staneva and L. Younes, "Modeling and estimation of shape deformation for topology-preserving object tracking," *SIAM Journal on Imaging Sciences*, vol. 7, no. 1, pp. 427–455, 2014.

[84] R. S. Strichartz, "Sub-Riemannian geometry," *J. Differerential Geometry*, vol. 24, pp. 221–263, 1986.

[85] D. Stroock and S. Varadhan, "Diffusion processes with continuous coefficients, I," *Comm. Pure Appl. Math.*, vol. 22, no. 3, pp. 345–400, 1969.

[86] ——, "Diffusion processes with continuous coefficients, II," *Comm. Pure Appl. Math.*, vol. 22, no. 4, pp. 479–530, 1969.

[87] G. Sundaramoorthi, A. Mennucci, S. Soatto, and A. Yezzi, "Tracking deforming objects by filtering and prediction in the space of curves," in *Proceedings of the 48th IEEE Conference on Decision and Control, 2009 held jointly with the 2009 28th Chinese Control*, 2009, pp. 2395–2401.

[88] G. Sundaramoorthi, A. Yezzi, A. Mennucci, and G. Sapiro, "New possibilities with Sobolev active contours," *Int. J. Comput. Vis.*, vol. 84, pp. 113–129, 2009.

[89] A. Trouvé and L. Younes, "Shape spaces," in *Handbook of Mathematical Methods in Imaging*, O. Scherzer, Ed. Springer New York, 2011, pp. 1309–1362. [Online]. Available: http://dx.doi.org/10.1007/978-0-387-92920-0_30

[90] N. Vaswani, "Particle filters for infinite (or large) dimensional state spaces-part 2," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 3, 2006, pp. 29–32.

[91] N. Vaswani, Y. Rathi, A. Yezzi, and A. Tannenbaum, "Deform PF-MT: Particle filter with mode tracker for tracking non-affine countour deformations," *IEEE Trans. Image Process.*, vol. 19, no. 4, pp. 841–857, April 2010.

[92] N. Vaswani, A. Yezzi, Y. Rathi, and A. Tannenbaum, "Particle filters for infinite (large) dimensional state spaces - part 1," *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, 2006.

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

[94] F.-X. Vialard, "Extension to infinite dimensions of a stochastic second-order model associated with shape splines," *Stochastic Processes and their Applications*, vol. 123, no. 6, pp. 2110 – 2157, 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0304414913000215

[95] N. Wiener, *The interpolation, extrapolation and smoothing of stationary time series.* MIT Press, 1964.

[96] T. Yamada and S. Watanabe, "On the uniqueness of solutions of stochastic differential equations," *J. Math. Kyoto Univ.*, no. 11-1, pp. 155–167, 1971.

[97] A. Yezzi and S. Soatto, "Deformotion: deforming motion, shape average and the joint segmentation and registration of images," *Intl. J. of Comp. Vis.*, vol. 53(2), pp. 153–167, 2003.

[98] N. Yoshida, "Estimation of diffusion processes from discrete observation," *Journal of Multivariate Analysis*, vol. 41, pp. 220–242, 1992.

[99] L. Younes, *Shapes and Diffeomorphisms*. Springer, 2010.

[100] ——, "Constrained diffeomorphic shape evolution," *Foundations of Computational Mathematics*, vol. 12, no. 3, pp. 295–325, 2012.

[101] ——, "Spaces and manifolds of shapes in computer vision: an overview," *Image and Vision Computing*, vol. 30, pp. 389–397, 2012.

[102] L. Younes, F. Arrate, and M. Miller, "Evolution equations in computational anatomy," *Neuroimage*, vol. 45, pp. S40–S50, 2009.

Appendix A

Sequential Importance Sampling

We describe the basic structure of the Sequential Importance Sampling algorithm, which is also known in the literature as a particle filter [44], or the CONDENSATION algorithm [49].

Consider a hidden Markov model for a dynamical system, in which the sequence of states $x_0, x_1, ..., x_t, ...$ cannot be observed directly, but we have available a sequence of observations $y_0, y_1, ..., y_t, ...$, which provide us with some information about the unknown states. In the following discussion we will abuse notation by denoting the conditional density of a sequence of random variables $x_0, ..., x_t$ on a sequence of random variables $y_0, ..., y_t$ by $p(x_{0:t}|y_{0:t})$. We assume that

(i) the states follow a first order Markov process:

$$p(x_{t+1}|x_{0:t}) = p(x_{t+1}|x_t) \tag{A.1}$$

(ii) given the state sequence, the observations are independent mutually and with respect to the dynamical process:

$$p(y_{1:t}, x_{t+1}|x_{0:t}) = p(x_{t+1}|x_{0:t}) \prod_{i=1}^{t} (y_i|x_i) \tag{A.2}$$

(iii) we can sample from the initial state density $p(x_0)$.

(iv) we can sample from the transition density $p(x_{t+1}|x_t)$.

(v) we can evaluate pointwise the observation density $p(y_t|x_t)$.

The first two assumptions simplify certain densities which will be useful later on. First note that, if we integrate (A.2) with respect to x_{t+1}, we obtain:

$$p(y_{1:t}|x_{0:t}) = \int p(y_{1:t}, x_{t+1}|x_{0:t}) dx_{t+1} = \underbrace{\int p(x_{t+1}|x_{0:t}) dx_{t+1}}_{1} \prod_{i=1}^{t} p(y_i|x_i) = \prod_{i=1}^{t} p(y_i|x_i).$$

Therefore,

$$p(x_{t+1}|x_{0:t}, y_{1:t}) = \frac{p(y_{1:t}, x_{t+1}|x_{0:t})}{p(y_{1:t}|x_{0:t})} = \frac{p(x_{t+1}|x_{0:t}) \prod_{i=1}^{t} (y_i|x_i)}{\prod_{i=1}^{t} (y_i|x_i)} = p(x_{t+1}|x_t). \tag{A.3}$$

Also we have

$$\begin{aligned} p(y_{1:t+1}|x_{0:t+1}) &= p(y_{t+1}|y_{1:t}, x_{0:t+1})p(y_{1:t}|x_{0:t+1}) = \\ &= \frac{p(y_{t+1}|y_{1:t}, x_{0:t+1})p(y_{1:t}, x_{t+1}|x_{0:t})}{p(x_{t+1}|x_{0:t})} = \\ &= \frac{p(y_{t+1}|y_{1:t}, x_{0:t+1})p(x_{t+1}|x_{0:t})\prod_{i=1}^{t} p(y_i|x_i)}{p(x_{t+1}|x_{0:t})} \\ &= p(y_{t+1}|y_{1:t}, x_{0:t+1})\prod_{i=1}^{t} p(y_i|x_i). \end{aligned} \tag{A.4}$$

One the other hand, by (A.3)

$$p(y_{1:t+1}|x_{0:t+1}) = \prod_{i=1}^{t+1} p(y_i|x_i), \tag{A.5}$$

so

$$\prod_{i=1}^{t+1}(y_i|x_i) = p(y_{t+1}|y_{1:t}, x_{0:t+1})\prod_{i=1}^{t} p(y_i|x_i), \tag{A.6}$$

from where we conclude that

$$p(y_{t+1}|y_{1:t}, x_{0:t+1}) = p(y_{t+1}|x_{t+1}). \tag{A.7}$$

We are interested in the posterior density $p(x_{0:t}|y_{1:t})$ or in some cases only in its marginal, the filtering density $p(x_t|y_{0:t})$. Our goal is to estimate these densities online, i.e. as each new observation becomes available we would like to be able to update the estimates for the target densities of the new state. In order to do that we would need the following recursive formula for the posterior density:

$$p(x_{0:t+1}|y_{1:t+1}) = \frac{p(y_{t+1}|x_{0:t+1}, y_{1:t})p(x_{0:t+1}|y_{1:t})}{p(y_{t+1}|y_{1:t})} \quad (A.8)$$

$$= k_t p(y_{t+1}|x_{0:t+1}, y_{1:t})p(x_{t+1}|x_{0:t}, y_{1:t})p(x_{0:t}|y_{1:t}) \quad (A.9)$$

$$= k_t p(y_{t+1}|x_{t+1})p(x_{t+1}|x_t)p(x_{0:t}|y_{1:t}), \quad (A.10)$$

where $k_t = p(y_{t+1}|y_{1:t})$ is independent of x. To derive a similar formula for $p(x_{t+1}|y_{1:t+1})$, we need to integrate both sides of this equality with respect to $x_{0:t}$:

$$\int p(x_{0:t+1}|y_{1:t})dx_{0:t} = \int k_t p(y_{t+1}|x_{t+1})p(x_{t+1}|x_t)p(x_{0:t}|y_{1:t})dx_{0:t} \quad (A.11)$$

$$p(x_{t+1}|y_{1:t}) = k_t p(y_{t+1}|x_{t+1}) \int p(x_{t+1}|x_t)p(x_t|y_{1:t})dx_t \quad (A.12)$$

$$= k_t p(y_{t+1}|x_{t+1})p(x_{t+1}|y_{1:t}). \quad (A.13)$$

Since these formulas involve integrals which cannot be evaluated analytically, it is reasonable to try to approximate them through Monte Carlo methods. However, a

direct Monte Carlo approach would not be tractable since sampling from the posterior density is usually impossible. Instead, we select a proposal density $\pi(x_{0:t}|y_{1:t})$, from which it is easy to sample and whose support is included in the support of the posterior density. Then, we could use the samples generated from $\pi(x_{0:t}|y_{1:t})$ to estimate the target densities. This approach is known as importance sampling, or, when it is applied recursively to a sequence of random variables, as sequential importance sampling.

Suppose we are trying to estimate the expected value of some function of the joint state $f_t(x_{0:t})$ at time t with respect to the posterior density:

$$I(f_t) = \int f_t(x_{0:t}) p(x_{0:t}|y_{1:t}) dx_{0:t}. \tag{A.14}$$

Let $w(x_{0:t}) = \frac{p(x_{0:t}|y_{1:t})}{\pi(x_{0:t}|y_{1:t})}$. Now we can write $I(f_t)$ as

$$I(f_t) = \frac{\int f_t(x_{0:t}) w(x_{0:t}) \pi(x_{0:t}|y_{1:t}) dx_{0:t}}{\int w(x_{0:t}) \pi(x_{0:t}|y_{1:t}) dx_{0:t}}. \tag{A.15}$$

If we draw a sample $\{x^i_{0:t}\}_{i=1}^N$ of size N from the proposal density, we can approximate $I(f_t)$ by

$$\hat{I}_N(f_t) = \frac{\sum_{i=1}^N \frac{1}{N} f_t(x^i_{0:t}) w(x^i_{0:t}) dx^i_{0:t}}{\sum_{i=1}^N \frac{1}{N}. w(x^i_{0:t}) dx^i_{0:t}} \tag{A.16}$$

In order to simplify the calculations we choose the prior as our proposal density, i.e.

$$\pi(x_{0:t}|y_{1:t}) = p(x_{0:t}). \tag{A.17}$$

From the Markov property, we have

$$p(x_{0:t+1}) = p(x_{0:t})p(x_{t+1}|x_{0:t}) = p(x_{0:t})p(x_{t+1}|x_t) = p(x_0)\prod_{k=0}^{t} p(x_{k+1}|x_k), \tag{A.18}$$

which gives the following recursive relation for the proposal density:

$$\pi(x_{0:t+1}|y_{1:t+1}) = p(x_{t+1}|x_t)\pi(x_{0:t}|y_{1:t}). \tag{A.19}$$

From equation (A.16) we can see that knowing the values of the weights $w(x^i_{0:t})$ for each particle can give us the value of $\hat{I}_N(f_t)$. At the first step $\pi(x_0) = p(x_0)$, so each weight $w(x^i_{0:t})$ equals 1, and the normalized weights $\tilde{w}^i_0$ equal $1/N$. For the subsequent steps we obtain

$$\tilde{w}^i_{t+1} \propto w(x^i_{0:t+1}) = \frac{p(x^i_{0:t+1}|y_{1:t+1})}{\pi(x^i_{0:t+1}|y_{1:t+1})} \propto \frac{p(x^i_{t+1}|x^i_t)p(y_{t+1}|x^i_{t+1})p(x^i_{0:t}|y_{1:t})}{p(x^i_{t+1}|x^i_t)\pi(x^i_{0:t}|y_{1:t})} \tag{A.20}$$

$$\tilde{w}^i_{t+1} \propto p(y_{t+1}|x^i_{t+1})\tilde{w}^i_t. \tag{A.21}$$

Since we can evaluate the transition density pointwise and we can generate new sample states from the transition density, we can evolve the weights over time, and

thus obtain an estimate for the posterior. From the strong Law of Large Numbers we have

$$\lim_{N\to\infty} \hat{I}_N(f_t) = I(f_t), \tag{A.22}$$

and by the Central Limit Theorem

$$\lim_{N\to\infty} \sqrt{N}(\hat{I}_N(f_t) - I(f_t)) \sim \mathcal{N}(0, \sigma^2_{f_t}), \tag{A.23}$$

where $\sigma^2_{f_t} = \frac{1}{N} \int \frac{p^2(x_{0:t}|y_{1:t})}{\pi(x_{0:t}|y_{1:t})}(f_t(x_{0:t}) - I(f_t))^2 dx_{0:t}$.

Vita

Valentina Staneva holds a B.S. in Mathematics from Concord University and M.S.E. in Applied Mathematics & Statistics from Johns Hopkins University. She got first involved in image processing research during the RIPS summer program at the Institute for Pure and Applied Mathematics at University of California, Los Angeles, and since then has worked on various problems in the field. She spent 1.5 years at Los Alamos National Laboratory working on optimization algorithms for image denoising and segmentation, and on the fundamental theory of nonconvex compressed sensing. Her graduate research at the Center for Imaging Science at Johns Hopkins University focused on developing methods for statistical inference of stochastic processes on shape manifolds and their applications to object tracking in computer vision and biomedical imaging. Valentina is currently working as a Senior Data Scientist at the University of Washington's eScience Institute as part of the Moore–Sloan Data Science Initiative.